OBESITY EXPERTISE

A first-principles curriculum in 196 books

BOOK 9 10th of 196 in the series

Scientific writing, visualisation and public communication · Rung 1

Writing a paper as an argument, and drawing a figure that tells the truth

ONE IDEA PER PARAGRAPH, HONEST AXES, AND A READER WHO CAN STATE YOUR ARGUMENT AFTER ONE PASS

Dr. Harsh Maheshwari

MD Community Medicine Resident, AIIMS Raipur

drharshmaheshwari.com · version 1.0

OBESITY EXPERTISE · BOOK 9

Scientific writing, visualisation and public communication · Rung 1

Writing a paper as an argument, and drawing a figure that tells the truth

Dr. Harsh Maheshwari

MD Community Medicine Resident, AIIMS Raipur

contact@drharshmaheshwari.com · drharshmaheshwari.com · ORCID 0000-0001-8848-1918

SERIES	Obesity Expertise: A first-principles curriculum in 196 books
BOOK	Book 9, Scientific writing, visualisation and public communication · Rung 1 (10th of 196 in the series)
VERSION	1.0
LAST UPDATED	3 October 2026
AUTHOR	Dr. Harsh Maheshwari, MD Community Medicine Resident, AIIMS Raipur
CONTACT	contact@drharshmaheshwari.com
ORCID	https://orcid.org/0000-0001-8848-1918
NEWEST VERSION	drharshmaheshwari.com

© 2026 Harsh Maheshwari.

This book is shared free, for anyone to read, under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International licence (CC BY-NC-SA 4.0). You may copy, share and adapt it for non-commercial purposes, if you credit the author and share what you make under the same licence.

It quotes and adapts material from OpenStax textbooks (Rice University), which are licensed CC BY-NC-SA 4.0; that is why this book carries the same licence. It also quotes short passages from laws, reports and other works, each credited where it is used; those passages remain under their owners' terms. Fonts: Source Serif 4, Inter and JetBrains Mono, SIL Open Font License 1.1.

HOW THIS BOOK WAS MADE

Written for the author's own study, then shared. Drafted and checked with AI assistance (Claude, by Anthropic) under the author's direction. Every factual claim was checked against a source the checker had open, every quotation against its source text, and every worked answer recomputed. Mistakes can remain; the feedback page says how to report one.

NOT ADVICE

This book is for learning. It is not medical, legal or financial advice, and it does not replace a qualified professional who knows your situation.

About the author



Dr. Harsh Maheshwari

MD Community Medicine Resident, AIIMS Raipur

Harsh Maheshwari completed his MBBS at King George's Medical University (KGMU), Lucknow, and is a resident in MD Community Medicine at the All India Institute of Medical Sciences (AIIMS), Raipur. His MD thesis, on chrononutrition and obesity, was selected by the Indian Council of Medical Research (ICMR) for funding in 2026.

He learns best from first principles: rebuilding an idea from what came before it until the formula feels like the only reasonable answer. This series grew out of that habit.

His long-term aims are to reduce obesity in India and to help make India healthy.

Email	contact@drharshmaheshwari.com
Website	drharshmaheshwari.com
ORCID	0000-0001-8848-1918

Introduction

Why this book exists

Most researchers learn to write by copying the papers they read, and those papers are an uneven model. This book treats a paper as an argument made to a reader with little time, and a figure as a claim that has to be true to its numbers. It teaches the sentence and the paragraph as the units of clarity, the number written with its unit and its count, and the plot with honest axes and nothing on it that does no work. The test it sets is simple to state: a colleague reads your draft once and can tell you your argument.

How to read it

Read Book 0 first, or at least the sections this book names as it goes. Then read this one in order: the paper first, then the sentence and the paragraph, then the figure. The book asks you to build one thing: one section of a manuscript or thesis, rewritten so that each paragraph carries one idea, and shown to a reader. Choose the section before you start, work on it as you read, and draw one figure for it when you reach the figures.

Every section has the same shape. **Definition** states the idea exactly. **In plain terms** says it again in everyday words. **Illustration** works through one real case, usually with a source you can open yourself. **Where this picture breaks** says where the illustration stops being true. **Must-know points** are what should change what you do. Then come **exercises** and, in most sections, a graded **practice set**.

Do the exercises before you look anything up. The worked answers are in the appendix at the back, under the same numbers. The references for each Part follow that Part. The glossary lists every term of art in the plain words it was first given. The last pages list all 196 books in the series and show where this one sits.

If you are new to this

Book 0, in particular decimals and rounding (A3), percentages (A4), ratios (A5), units (B1), graphs (C5), average and spread (D5), reading a table and a figure (F1, F2) and the shape of an argument (F4). The mathematics is arithmetic with a calculator: percentages, ratios, and putting counts into a formula stated in words. No algebra beyond that, and no statistics beyond a mean and a standard deviation. You need a piece of your own writing to work on, such as a thesis chapter or a manuscript section, and any program that draws a chart; a spreadsheet will do if you are prepared to change nearly all of its defaults.

On a first reading, read each section's definition, plain-terms explanation, illustration and must-know points, and do the exercises. Worked derivations can be skimmed and revisited once the idea feels familiar. Keep a calculator beside you: every worked answer can be followed with pen, paper and a calculator. Anything beyond the stated prerequisites is explained where it first appears, or marked as beyond this book. The **Symbols used in this book** page lists every sign and Greek letter with how to read it and where it is first used.

How to send feedback

If something is wrong, unclear, or could be better, please write to **contact@drharshmaheshwari.com**. Put the book number, the version and the section in the subject line (for example: *Book 9, v1.0, section C5*). Say what the text says, what you think it should say, and, if you can, the source.

Corrections are made in new versions. The version number and the date of the last update are on the copyright page of every book, and the newest version of every book is at **drharshmaheshwari.com**.

> **WHERE THIS BOOK SITS**
> Book 9 of the series' 196 books (Book 0 and 195 rung books). Before it: Book 8, Teaching, mentorship and coalition building · Rung 1. After it: Book 10, Epistemics: philosophy of science and calibration · Rung 1. The full list is at the back.

Symbols used in this book

Every sign and Greek letter this book prints, how to read it aloud, what it means in this series, and the section where it first appears. Each is also explained where it is first used.

Symbol	Read as	Meaning	First in
SIGNS			
≥	greater than or equal to	An inequality that includes its boundary	8 · Words: the everyday word, one term for one thing, abbreviations that cost the reader
±	plus or minus	A range either side of a value: 11.8 ± 0.6 means 11.2 to 12.4	9 · Numbers in a sentence: the unit, the count behind the percentage, and no more digits than the data hold
÷	divided by	Division	9 · Numbers in a sentence: the unit, the count behind the percentage, and no more digits than the data hold
≈	approximately equal to	The two sides are close but not exactly equal	15 · Choosing the mark: what the eye reads accurately, and the shape of the data
–	minus	Subtraction, or a negative number	17 · Honest axes: where an axis starts, what range it shows, and measuring the distortion
×	times	Multiplication	17 · Honest axes: where an axis starts, what range it shows, and measuring the distortion
√	square root of	The number that, multiplied by itself, gives the value: √9 = 3	20 · Words on a figure: axis titles with units, direct labels, and a caption that stands alone
MARKS ON LETTERS			
x^2, 10^{-3}	superscript	Small raised characters are powers: x^2 = x × x, and 10^{-3} = 0.001	Before you start

Contents

Before you start 9

Rung 1 — Introductory 10

1 · What a research paper is: its parts and what each part is for 10
2 · A paper is an argument, not a report of activities 14
3 · The one message: your argument in one or two sentences, written first 18
4 · The argument across sections: from the gap to the answer and its limits 22
5 · Citing and referencing: the original source, in the order cited, in one format 27
6 · The parts of a sentence a writer works with 32
7 · The sentence as the unit of clarity 36
8 · Words: the everyday word, one term for one thing, abbreviations that cost the reader 42
9 · Numbers in a sentence: the unit, the count behind the percentage, and no more digits than the data hold 48
10 · One idea per paragraph: the topic sentence, its support, and the reverse outline 56
11 · Measuring a draft: words per sentence, a readability score, and what they cannot show 61
12 · Cutting: deleting the words, sentences and paragraphs that do no work 70
13 · Showing a draft to a reader: the one-pass test 75
14 · A figure makes one claim 78
15 · Choosing the mark: what the eye reads accurately, and the shape of the data 84
16 · Show the data: when a bar of means hides what the points would show 90
17 · Honest axes: where an axis starts, what range it shows, and measuring the distortion 95
18 · Decoration that does no work 105
19 · Colour: for meaning only, readable by everyone, and in grey 110
20 · Words on a figure: axis titles with units, direct labels, and a caption that stands alone 116
21 · Making a table that stands alone 122
22 · Making the plot: from a table of numbers to a finished figure 127
23 · Worked journey: One finding, from the fact sheet to a paragraph a colleague can restate 132
References 142

Appendix · Worked answers 168

Glossary 201

The series: all 196 books 207

Before you start

This book uses the following Book 0 sections. Each is summarised where it is first needed; work through Book 0 itself if a summary is not enough.

- Book 0, A3 · Decimals, rounding, and significant figures
- Book 0, A4 · Percentages, and the percentage of a percentage
- Book 0, A5 · Ratios, rates and proportions
- Book 0, A7 · Logarithms — what they are for before how they work
- Book 0, B1 · What a unit is, and the SI base and derived units
- Book 0, B2 · Converting units as multiplication by one
- Book 0, B6 · Body-size units: kg, m, cm and kg/m^2
- Book 0, C2 · Rearranging an equation
- Book 0, C5 · Graphs: axes, scale, slope and intercept
- Book 0, D1 · What a probability is
- Book 0, D3 · Conditional probability through the two-way table
- Book 0, D4 · Variation: what differs and by how much
- Book 0, D5 · Average and spread, and why the average is not the person
- Book 0, D6 · Sampling: how a part can tell you about a whole
- Book 0, F1 · Reading a table
- Book 0, F2 · Reading a figure, and the ways a figure misleads
- Book 0, F3 · Checking a reference
- Book 0, F4 · Argument: premise, inference, conclusion

Rung 1 — Introductory

1 · What a research paper is: its parts and what each part is for

Checked against the sources as they stood on 2026-10-02.

DEFINITION. An original research article is a journal paper that reports a study for the first time. The International Committee of Medical Journal Editors (ICMJE) is a group of medical journal editors. Its Recommendations, updated January 2026, say the text of such an article is usually divided into Introduction, Methods, Results and Discussion (IMRAD).

Around the text sit a title page, which carries the title; an abstract; and a numbered list of references.

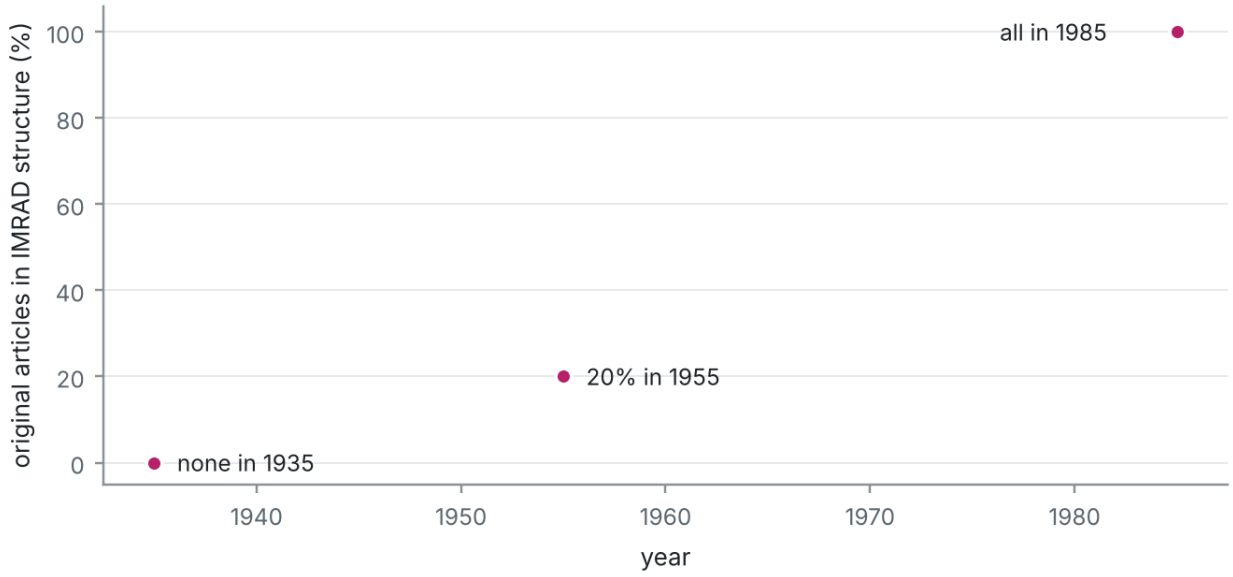
IMRAD is a convention that editors adopted, not a fixed law of science. In four leading journals of internal medicine it was absent from original articles in 1935, reached 80 per cent of them in the 1970s, and was the only pattern by the 1980s (Sollaci and Pereira, 2004). [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23]

IN PLAIN TERMS. A research paper reports one study. Most medical journals want it in the same four parts, in the same order.

The introduction says why the study was needed and what question it asked. The methods say what was done, so that someone else could check it or do it again. The results say what was found. The discussion says what the findings mean, and where they stop.

Before these four come a title and an abstract, a short summary of the whole paper. After them comes the list of references, the sources the paper leans on.

The four-part shape has a name, IMRAD, from the first letters of the parts. It is a habit that editors spread in the twentieth century, not a rule of nature. So the journal you send your paper to has the last word on its shape.



The four-part IMRAD structure is a recent convention: none of the original articles used it in 1935, 20 per cent did in 1955, and all of them did in 1985. Sollaci and Pereira (2004) counted 1,297 original articles in the British Medical Journal, JAMA, The Lancet and the New England Journal of Medicine, every five years from 1935 to 1985. Only the years their text states are drawn; their over 80 per cent in the 1970s names no single year, so it is not plotted.

Illustration 1. Here is the opening of a draft paper. It is made up for this section, but both of its faults are common. The study measured women aged 15 to 49 who came to one urban health centre. Read each sentence and ask which question it answers for a reader: why the study was done, how it was done, what was found, or what it means.

> ***Introduction.*** *(1) Overweight and obesity are common among Indian women: in the fifth National Family Health Survey (NFHS-5, 2019-21), 24.0 per cent of women aged 15-49 had overweight or obesity, against 20.6 per cent in the fourth (2015-16). (2) A national figure does not tell a health centre how many of the women it serves are affected. (3) In our survey, 31 per cent of the women had overweight or obesity. (4) We set out to estimate the share of women aged 15-49 attending the centre who have overweight or obesity.*
>
> ***Methods.*** *(5) We weighed and measured consecutive women attending the centre. (6) Of the 412 women approached, 380 agreed to be measured.*

The numbers in sentences 3 and 6 are made up. The two survey figures in sentence 1 are real.

Start with sentence 3. It is a finding of this study, sitting in the introduction. ICMJE's Recommendations say the introduction should not "include data or conclusions from the work being reported". A reader who meets 31 per cent here has not yet been told how it was measured. So they cannot judge it, and they will meet it again in the results. Move it to the results.

Now sentence 6. It reads like method, because it is about how the women were found. But you learnt 412 and 380 only by doing the study. The Recommendations put everything obtained during the study in the results; the methods hold only what was known when the protocol, the study's written plan, was drawn up. So the methods say how women were approached. The count of who agreed goes to the results.

Then label the four that stay. Sentence 1 is context. Sentence 2 is the gap, what is not yet known. Sentence 4 is the purpose. Sentence 5 is method. Each now answers the question its section exists to answer, and a reader who goes to one section looks in the right place.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. Each misplaced sentence here has one plain home. A real draft often runs a known fact and the study's own number into one sentence. Or it joins a result and what you think it means: the number goes to the results, the reading to the discussion. Split it first, then send each half to its section.

The pattern is for original research. A case report (the account of one or a few patients), a narrative review (a summary of published work on a topic) or an editorial (a journal's opinion piece) has its own shape, and in every case the instructions of the journal you have chosen win over this pattern.

Illustration 2. You will hear that papers have always been written in these four parts. Check it.

Open Sollaci and Pereira's survey of the structure, free to read at https://pmc.ncbi.nlm.nih.gov/articles/PMC442179/. Read the Methods first. They took one issue in every ten of four leading journals of internal medicine: the British Medical Journal, the Journal of the American Medical Association (JAMA), The Lancet and the New England Journal of Medicine. They looked every five years from 1935 to 1985, and counted 1,297 original articles.

Now read the first paragraph of the Results and write down each share the text states. In 1935 they found no article in the four-part structure: 0 per cent. By 1955 the share had reached 20 per cent. By the 1970s it was over 80 per cent. Then read the next paragraph: the last two journals fully adopted it in 1985. So by 1985 every original article in the four journals used it: 100 per cent.

So the claim fails. The structure that looks permanent became the usual one in these journals only in the 1970s. The Discussion says why it may have spread. Editors pushed for it, to help readers and reviewers. And readers do not read a paper from start to finish; they jump to the section where they expect to find what they want.

That last reason is the one to keep. A reader who wants your sample size goes to the methods. A reader who wants your main number goes to the results. Put each thing where they will look.

Figures quoted in this illustration are from [25].

WHERE THIS PICTURE BREAKS. These are four journals of internal medicine, counted every five years. The authors themselves warn against extending their findings to other journals. The year-by-year shares are in the paper's own figure, which is a picture and is not used here; only the shares its text states are.

MUST KNOW POINTS FOR YOU.

- Before you draft, open the instructions for authors of the journal you are aiming at. Note its article types, its abstract format and its word limits.

- The two commonest misplacements are your own result in the introduction and a number you obtained during the study in the methods. A reader who meets a result before its method cannot judge it.
- "Papers have always been written this way" is wrong. In four leading medical journals the IMRAD structure was absent in 1935. ICMJE calls it a reflection of how discovery works. Neither makes it a rule for every kind of article.
- Do not force a case report, a narrative review or an editorial into the four parts. ICMJE says such articles may be less structured or unstructured, and that a meta-analysis (a pooling of the results of several studies) may need a different format.
- Readers jump to the section where they expect to find a thing. So put it there.
- The four parts tell you where each kind of sentence goes. They do not tell you what the paper claims, or whether the data carry that claim.
- When you turn a thesis chapter into a paper, read the journal's instructions afresh.

COMMON MISREADING. It is tempting to read the IMRAD order as the order in which to write: introduction first, then methods, then results, then discussion. It is the order in which a reader looks things up. ICMJE's Recommendations say which section holds what, and nothing about which to write first. Mensh and Kording, who wrote ten simple rules for structuring papers, suggest starting elsewhere: build the outline around a description of each result. Those descriptions can become the headings of your results section.

EXERCISE 1 (retrieval). From memory, name the four sections of an original research article in the IMRAD structure, and the question each one answers for a reader. Then name the three parts that sit around them.

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (critique). A colleague's draft has these four sentences in its methods section. The numbers are made up.

> *(a) Adults aged 30 and over who came to the diabetes and blood pressure clinic on Mondays were eligible. (b) Waist was measured with a tape that does not stretch, at the level set in the protocol. (c) Of 260 adults who were eligible, 214 were measured. The rest said no or left first. (d) Waist was above the clinic's cut-off in 58 per cent of those measured, more than we expected.*

Say which sentences belong in the methods, where the others go, and why.

EXERCISE 3 (teaching). For a first-year resident, ten minutes with a whiteboard. She asks why journals insist on four fixed sections when a paper would read better as a story. Answer her, and tell her the kinds of article that need not use the four sections.

2 · A paper is an argument, not a report of activities

DEFINITION. A paper is written as an argument when its text states a claim and gives the evidence and the reasoning that support it, in the order the reasoning needs. In the terms of Book 0, F4 (argument), the claim is the conclusion, the data are the premises, and the reasoning that carries one to the other is the inference.

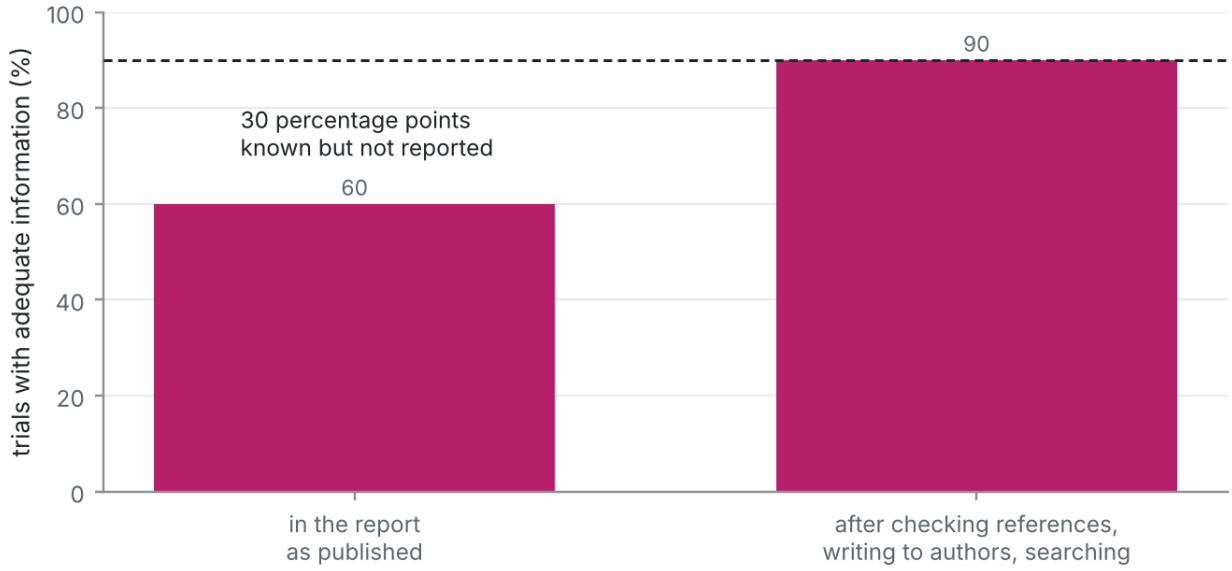
A report of activities records what the authors did, in the order they did it.

The two forms can hold the same facts. They differ in what each sentence is for. [26, 27, 28, 29, 30, 30, 30, 31, 31, 32, 33, 34, 35]

IN PLAIN TERMS. When you write up a study, the easy thing is to tell it the way it happened. That is a diary.

Your reader does not want the diary. They want two things: what you found, and why they should believe it.

The steps you took still matter, because a reader has to be able to check them. They go in the methods, once, and not scattered through the results.



For about 30 trials in every 100, the authors knew what treatment they had tested and how it was given, but their report did not say it: adequate information was in around 60 per cent of reports as published, and could be found for 90 per cent of the trials by checking references, writing to authors and searching further. The dashed line marks 90 per cent. Figures from one study of trial reports, as reported by Chalmers and Glasziou (2009); this book has not read the original study.

Illustration 1. Start with a results paragraph written as a diary. It is made up for this section, but its numbers are real: they are the "overweight or obese" rows of the National Family Health Survey (NFHS) fact sheet for India.

> *We opened the NFHS-5 India fact sheet and went to the section on the nutritional status of adults. We first read the row for women aged 15-49 and found that 24.0 per cent had overweight or obesity. We then read the row for men and found 22.9 per cent. Next we read the NFHS-4 column, which gave 20.6 per cent for women and 18.9 per cent for men. Finally, we subtracted each NFHS-4 figure from its NFHS-5 figure.*

Every sentence is true. Now ask of each one: what claim does this support? You cannot answer, because no sentence states a claim. The paragraph ends at the subtraction and never gives its result. The reader has to do the arithmetic and draw the conclusion, which is the writer's job.

So ask the question the diary never answers: what did you find? Write the answer as one sentence. Here it is that the share of adults with overweight or obesity rose between the two surveys, in women and in men.

Put that sentence first. Then keep only the evidence it needs, in the order it needs it: women with both surveys, then men with both surveys. Do the subtraction yourself and print the answer.

```
24.0 minus 20.6 = 3.4
22.9 minus 18.9 = 4.0
```

The differences are in percentage points, the plain difference between two percentages, as Book 0 taught it. Here is the paragraph rewritten.

> *The share of adults aged 15-49 with overweight or obesity rose between the two surveys, in women and in men. In women it rose from 20.6 per cent in NFHS-4 (2015-16) to 24.0 per cent in NFHS-5 (2019-21), by 3.4 percentage points. In men it rose from 18.9 to 22.9 per cent, by 4.0 percentage points.*

The steps did not vanish. "We took the share with overweight or obesity for each sex from the NFHS-4 and NFHS-5 India fact sheets and subtracted the first from the second" is one sentence, and it goes in the methods.

Last, say what the claim does not cover. It is about adults aged 15-49 only, and the women's rows leave out women who were pregnant or had given birth in the last two months. Overweight and obesity are counted together, because the fact sheet gives no separate row for obesity. And two surveys of different people show that the share rose, not why it rose.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. This paragraph carries one claim with four numbers, so the claim-first order is easy to see. A paragraph whose claim needs a chain of steps, such as a check that a measurement worked before it is used, keeps those steps in the order the logic needs. That order is still the argument's, not the diary's.

The numbers hold for India as a whole, for the two survey rounds named. They are not a figure for any one state or district.

Illustration 2. Here is what a report written for its writer costs a reader. Chalmers and Glasziou's 2009 viewpoint on avoidable waste in research is at https://doi.org/10.1016/S0140-6736(09)60329-9. Find the section headed "Biased or unusable reports of research".

There they cite a study of reports of clinical trials — studies that give a treatment to people and compare what happens. This book has not read that study; the figures are as Chalmers and Glasziou report them. Adequate information on the treatment tested was in around 60 per cent of the reports. By checking references, writing to authors and searching further, it could be found for 90 per cent of the trials. Work out the gap.

```
90 minus 60 = 30
```

For about 30 trials in every 100, the authors knew what they had done. Their reports did not say it, so a clinician could not use the result without detective work. That is a loss made in the writing, not in the research.

Notice what kind of loss it is. The missing pieces were methods: what the treatment was and how it was given. Writing a paper as an argument does not mean leaving out what you did. It means putting it where the reader needs it, complete enough to judge and repeat.

Figures quoted in this illustration are from [36].

WHERE THIS PICTURE BREAKS. The 60 and 90 per cent come from one study of trial reports, which Chalmers and Glasziou cite rather than carried out. They describe treatments in trials, not every kind of paper, and they are not a measurement of Indian journals.

MUST KNOW POINTS FOR YOU.

- "The results should follow the order in which the work was done" is the error. They follow the order the argument needs.
- Before you write a results paragraph, write its claim as one sentence. Keep only the sentences that support that claim, and put the claim first. This rule is for a results paragraph. Mensh and Kording instead open it on its question and end it on the answer. Either way the claim is printed, not left to the reader.
- If a paragraph ends on a step, such as "we then compared the two groups", the reader has to do your conclusion for you. Finish the step: print the result of the comparison and say what it shows.
- Write the methods so a reader could use your result. In one study that Chalmers and Glasziou cite, only around 60 per cent of trial reports described the treatment adequately. The information existed for 90 per cent. The difference was lost in the writing.
- Cut any table or figure that supports no claim.
- When a trainee's draft reads "first we..., then we..., next we...", do not correct the grammar. Ask what they found, have them write that as the first sentence, and move the steps to the methods.
- Writing your paper as an argument makes your claim and its support easy to see. It cannot make a weak claim strong.

COMMON MISREADING. "Argument" can sound like persuasion: state your claim, and leave out whatever does not fit it. That reading is wrong. It leads straight to biased reporting.

The International Committee of Medical Journal Editors (ICMJE) asks in its Recommendations for data on every main and secondary outcome named in the methods. Chalmers and Glasziou give two examples of biased reporting. Harms of a treatment are left out. The main outcome is changed between the plan and the report. An argument in a paper states its claim. Then it shows all the evidence on it, including the evidence against it.

REPORTING IT. Between NFHS-4 (2015-16) and NFHS-5 (2019-21), the share of adults aged 15-49 in India with overweight or obesity rose in both sexes. It rose from 20.6 to 24.0 per cent in women and from 18.9 to 22.9 per cent in men.

EXERCISE 1 (retrieval). From memory: in the terms of Book 0, F4 (argument), what are a paper's claim, its data, and the reasoning between them? And what are the two separate ways the argument can fail?

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (critique). A resident's clinic audit has this results paragraph. The numbers are made up.

> *We reviewed the outpatient registers of the non-communicable disease clinic for January to March. First we counted the adults registered, who numbered 640. We then checked how many had a weight recorded, which was 410. After that we looked for a height, which was recorded for 152. Finally, we checked how many had a body mass index written in the register, and found 96.*

Rewrite it as an argument. State the claim first, keep the evidence it needs, and say where the first sentence goes.

EXERCISE 3 (design). Take one results paragraph from the section you chose for this book's build: one section of your own manuscript or thesis, chosen before you start, as "How to read" asks. Mark each sentence with a letter. Write C if it states a claim, E if it is evidence for a claim, S if it describes a step of the work. A sentence that reports a number from a step is E; the step itself goes to the methods. Then rewrite the paragraph as an argument.

EXERCISE 4 (teaching). For a first-year resident, ten minutes with a whiteboard. Her co-resident says the results section should walk through the analysis step by step, "so the examiner can see the work". Tell her what is right and what is wrong in that.

3 · The one message: your argument in one or two sentences, written first

DEFINITION. A paper's one message is its central claim, stated in one or two sentences: what the study found, and why that matters.

The message is written before the draft, and it is a test. It passes only if it is a claim, which would be false had the data come out differently, and if it answers the reader's "so what?".

The title is the message in its shortest form. The abstract is the argument in small: the context, the gap, what was done, the main result, and what it means. [37, 38, 38, 39, 40, 41, 42, 43, 44, 45, 46, 46, 46, 47, 48, 49, 50, 31, 51, 52]

IN PLAIN TERMS. Before you write the paper, write one or two sentences that say what you found and why it matters. That is your one message.

Test it two ways. First, would the sentence be false if your data had come out differently? If not, it says nothing yet. Second, does it answer "so what?" for someone who reads only that sentence? That is, does it say what the finding changes for someone: a programme, a clinician, the next study?

The abstract is the message with its reasons: what was known, what was missing, what you did, what you found, and what it means. Check that the title, the abstract and the paper all say the same thing.

Illustration 1. Take the running case of this book: the National Family Health Survey's figures for adults with overweight or obesity, in the fourth round (NFHS-4, 2015-16) and the fifth (NFHS-5, 2019-21). Try to write its one message. Here are three attempts, each made up for this section. Test each one by asking: would this sentence be false if the data had come out the other way?

> ***Attempt 1.*** *This paper is about overweight and obesity in Indian adults, using NFHS data.*

True whatever the data showed. It names a topic and claims nothing. A reader who stops here learns what you looked at, not what you found.

> ***Attempt 2.*** *We report the share of adults with overweight or obesity by sex, by urban and rural residence, and in two survey rounds.*

Still true whatever the data showed. It is a list of contents. Five findings may be hiding in it, and the reader is not told which one the paper is for.

> ***Attempt 3.*** *Between 2015-16 and 2019-21, the share of Indian adults aged 15-49 with overweight or obesity rose in women and in men, to about one woman in four and more than one man in five.*

This one could be false. Had the share fallen, the sentence would be wrong, so it says something. Check its two size phrases against the figures, 24.0 per cent of women and 22.9 per cent of men.

```
1 divided by 4 = 0.25, which is 25 per cent
1 divided by 5 = 0.2, which is 20 per cent
```

24.0 per cent is close to 25, so "about one woman in four" holds. 22.9 per cent is above 20, so "more than one man in five" holds.

Now the "so what?" test: does the message say what the finding changes for someone? The size phrases say how big the rise is, not what it changes. So add a second sentence.

> *So a programme that planned its reach on the 2015-16 shares would now plan for too few people.*

With it, attempt 3 passes both tests.

Last, say what the message does not claim. It covers ages 15 to 49 only. Its women's figures leave out pregnant women and women who gave birth in the previous two months. It counts overweight and obesity together. And it shows that the share rose, not why it rose. Do not let the title or the abstract claim more.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. This case has a thin gap: anyone can read the two columns of the fact sheet. A real paper's message reports something nobody had measured, and its "so what?" usually needs a sentence of its own. The test is the same.

The figures hold for India as a whole, for adults aged 15-49, in the two rounds named.

Illustration 2. Now build the title and the abstract from that message, and check them against each other.

Start with a title that fails: "NFHS-5 and overweight". It names a topic, not a finding. It also opens on an abbreviation, which the Recommendations of the International Committee of Medical Journal Editors (ICMJE) ask authors to keep out of a title. A searcher who types "National Family Health Survey" may not find it.

Cut the message down instead, keeping the finding and the words a searcher would use.

> *Overweight and obesity rose among women and men aged 15-49 in India between 2015-16 and 2019-21: a comparison of two National Family Health Surveys*

The part after the colon names the design, which some journals require. Suppose your journal also wants a short title of no more than 40 characters, letters and spaces both counted. Try "Overweight and obesity rose, ages 15-49" and count it; the comma, digits and hyphen count too.

```
10 plus 3 plus 7 plus 5 plus 4 plus 5 = 34 characters without spaces
34 plus 5 spaces = 39 characters
```

It fits, and it still carries the finding and the age range.

Then the abstract, one sentence for each part of the argument.

> *The National Family Health Survey reports the share of women and men aged 15-49 with overweight or obesity [context]. The fifth round's India fact sheet prints the fourth round's figures beside its own but does not state the change [gap]. We compared the two rounds for each sex [what was done]. The share rose from 20.6 to 24.0 per cent in women and from 18.9 to 22.9 per cent in men, by 3.4 and 4.0 percentage points [result]. So a programme that planned its reach on the 2015-16 shares would now plan for too few people; the surveys show the rise, not its cause [meaning].*

The labels in square brackets are there to show the parts; a real abstract leaves them out, or uses the headings its journal asks for.

Last, the check that catches the commonest fault. Suppose your first draft of the abstract said "by 3.6 percentage points in women", written before you redid the subtraction. That 3.6 is planted here; the results give 3.4. ICMJE notes that information in abstracts often differs from that in the text. So list every number in the title and abstract, find each one in the results, and make them agree before anyone else reads the draft.

Figures quoted in this illustration are from [53].

WHERE THIS PICTURE BREAKS. A five-sentence abstract fits this small case. A journal may ask for set headings and a word limit, and then its format wins; the parts of the argument stay the same.

Counting characters by hand works for a short title. For a long one, let the word processor count, and check whether your journal counts spaces.

MUST KNOW POINTS FOR YOU.

- Write your one message before you draft any section. If you cannot write it, you are not ready to draft.
- "This study is about..." is a topic, not a message.
- Make the title the message in its shortest form, with no abbreviations, which ICMJE's Recommendations ask authors to avoid there.
- Before anyone else reads your draft, list every number in the title and abstract and find each one in the results.
- Mensh and Kording apply this rule, one of their first four, to grants and posters too. In this book, carry the same message into the talk and the answer to a journalist as well.
- A clear message is not a true one. Writing the message does not test whether the data carry it. Check that separately, and if they carry less, cut the message to fit them.

COMMON MISREADING. "One message" can be read as "report one result and drop the rest". That reading is wrong. ICMJE's Recommendations ask for data on every main and secondary outcome named in the methods. Mensh and Kording allow one contribution to have several parts. The one message decides what the paper is for. The other results stay in, each placed where it supports the message or marks a limit on it.

REPORTING IT. Between 2015-16 and 2019-21, the share of Indian adults aged 15-49 with overweight or obesity rose in women and in men. By 2019-21 it was about one woman in four and more than one man in five. The two surveys show the rise, not its cause.

EXERCISE 1 (retrieval). From memory: what are the two tests a one-message sentence must pass? And what are the parts of an abstract's argument, in order?

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (critique). Three residents offer the one message of their theses. All three are made up. Say which is a message, what is wrong with the other two, and what the message does not show.

> ***(a)*** *This thesis studies waist and body mass index in adults who come to a rural health centre.*

> ***(b)*** *We measured waist, body mass index, blood pressure and fasting blood sugar in 300 adults. We report each by age and sex.*

> ***(c)*** *In 300 adults at a rural health centre, waist put more people above its cut-off than body mass index did. So screening by body mass index alone would miss some of the people waist would flag.*

EXERCISE 3 (design). Write the one message of the section you chose for this book's build, in one or two sentences. Test it both ways. Then write a title from it, and an abstract of five sentences, one for each part of the argument.

EXERCISE 4 (teaching). A journalist rings and asks: "Is obesity rising in India?" You have three minutes and the NFHS-4 and NFHS-5 figures. Give the one quotable sentence, and say how sure the evidence lets you be.

4 · The argument across sections: from the gap to the answer and its limits

DEFINITION. In a paper that reports original research, each of the four main sections does one job in the paper's argument.

The introduction moves from what is known to what is not yet known, the gap. It ends on the study's specific purpose: the question it asks, its research objective, or the hypothesis it tests (a statement the study is designed to support or refute).

The discussion opens by stating the main findings briefly, as the answer to the introduction's question. [54, 55, 56, 57, 58, 8, 59, 10, 60, 23, 31, 61, 62, 63, 64, 12, 65, 66, 67, 22, 68, 28, 69, 70, 52]

IN PLAIN TERMS. The introduction says what is known, then what is not known, and ends on your question. The methods say what you did and why, so that a reader can judge it and could repeat it. The results answer the question in steps, one claim at a time, each with its numbers. The discussion gives the answer in its first lines, then what might explain it, what could make it wrong, and what it means.

Illustration 1. Start with a plan that fails. A resident is turning her thesis work into a short paper on the national survey's figures for adults with overweight or obesity. Her plan is a list of headings. The plan is invented for this lesson.

section	the resident's plan
Introduction	obesity: definition, history, causes, complications; the burden worldwide and in India
Methods	about the survey; sampling; body mass index; the urban figures we noticed later
Results	Table 1; Table 2; Figure 1
Discussion	obesity is a growing problem; our results agree with other studies

Read each heading and ask what it claims. None of them claims anything. "Table 1" tells the reader where to look, not what they will find there. The introduction covers everything known about obesity and never reaches a question, so the reader cannot tell what the paper will answer. "The urban figures we noticed later" is something learnt during the work, so it does not belong in the methods.

Now write the same paper as an outline: one sentence for each planned paragraph, and each sentence a claim. First get the data. Open the India fact sheet of the fifth National Family Health Survey (NFHS-5) at https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf and find indicators 88 and 89, under "Nutritional Status of Adults (age 15-49 years)". Each row gives four numbers, all in per cent: NFHS-5 (2019-21) urban, rural and total, then the NFHS-4 (2015-16) total. Overweight or obesity means a body mass index of 25.0 kg/m^2 or more, the measure Book 0 built (Book 0, B6).

paragraph	the claim it will make
Introduction 1	India's NFHS reports, in its fourth and fifth rounds, the share of women and of men aged 15-49 with overweight or obesity.
Introduction 2	[What the papers you have read say is known, and what they leave open. Each sentence carries its reference. This outline cannot fill it for you.]
Introduction 3	We asked how that share changed between NFHS-4 (2015-16) and NFHS-5 (2019-21), and how it differs between urban and rural India in NFHS-5.
Methods 1	We used indicators 88 and 89 of the NFHS-5 India fact sheet, and fixed two comparisons before reading the numbers: between the rounds, and between urban and rural areas.
Results 1	A larger share of women aged 15-49 had overweight or obesity in NFHS-5 than in NFHS-4: 24.0 against 20.6 per cent.
Results 2	The same held for men: 22.9 against 18.9 per cent.
Results 3	In NFHS-5, overweight or obesity was commoner in urban than in rural adults: 33.2 against 19.7 per cent of women, and 29.8 against 19.3 per cent of men.
Discussion 1	Between the two rounds, the share of adults aged 15-49 with overweight or obesity rose for women and for men, and in NFHS-5 it was higher in urban than in rural India for both.
Discussion 2	[Possible explanations, each from a paper you cite, and each called possible. The fact sheet offers none.]
Discussion 3	The rows cover only ages 15-49, put overweight and obesity together, leave out pregnant women and women with a birth in the last two months, and come from two separate surveys, so they cannot show what caused the change.
Discussion 4	[What the answer means for research or for programmes, claiming no more than the data support.]

Now read the right-hand column alone, top to bottom. Introduction 3 asks a question with two parts. Results 1 and 2 answer the first part, and Results 3 answers the second. Discussion 1 answers the whole question in one sentence. Discussion 3 says how far that answer reaches. The argument is there before a single paragraph exists.

Each results sentence would also work as a subheading: it states a finding, where "Table 1" only names a place. And notice what the outline stops you claiming. The fact sheet's NFHS-4 column gives the total only, so you cannot say whether the urban-rural gap widened. You find that out here, in one line, rather than after writing three paragraphs about it.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. This outline fits a short paper with one question and three results. A paper on an experiment, or on people followed over years, has more steps, such as showing that a measure works before using it. Its results section is longer, but each step is still one claim.

The numbers are shares of adults aged 15-49 in India in two survey rounds, overweight and obesity together. They do not show why the share rose. They say nothing about adults over 49, and nothing about any one state or district.

The bracketed rows are not decoration. An outline with them left empty is a plan for a paper whose gap and explanations are still unsupported. Fill them from papers you have read and can cite.

Illustration 2. The second move is putting each sentence in the section whose job it does. Here are seven sentences from a draft paper on a district survey of women's weight. The draft and its sentences are invented. Cover the last two columns, decide where each sentence goes, then check.

sentence in the draft	the draft put it in	it belongs in	why
Obesity is a major public health problem all over the world.	Introduction	nowhere, as written	It names no gap and leads to no question. A context sentence has to narrow towards this study's question.
We found that one woman in four had overweight or obesity.	Introduction	Results	The introduction carries no data or conclusions from the work being reported.
Some of the women we approached declined to be weighed.	Methods	Results	It was learnt during the study. The methods hold only what was known when the protocol (the written plan of the study, made before it began) was written.
We planned to compare women in urban and rural blocks.	Results	Methods	It was decided before the data came in.
Overweight or obesity was commoner in women in urban blocks than in rural blocks.	Discussion, paragraph 3	Results, as a subheading; and again in Discussion, paragraph 1, as part of the answer	It is a finding. It is also part of the answer, so the discussion opens with it.
Our survey covered one district, so it cannot speak for the state.	Results	Discussion	It is a limit on how far the answer reaches.
A district programme for this would pay for itself.	Discussion	nowhere, unless the paper has cost data	A claim about money needs data on money.

Two of these moves surprise people. The first is the women who declined. It feels like a fact about method, but you learnt it while doing the study. The Recommendations of the International Committee of Medical Journal Editors (ICMJE; section IV.A.3.d, at https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html) put everything obtained during the study in the results. A reader can then see what you planned and, separately, what happened.

The second is the finding that appears twice. The results state it with its numbers. The discussion's first paragraph states it again, briefly, as the answer to the question the introduction asked. That is the one repetition the argument needs.

WHERE THIS PICTURE BREAKS. The sorting rule for methods and results follows ICMJE's general requirement. A journal's own instructions, or a reporting guideline for one kind of study, may

ask for a particular item in a particular section; follow them where they speak. Reporting guidelines are taught in the course's book on protocols, reporting standards and research ethics.

Some sentences do two jobs, as the urban-rural finding does. The test is the job the sentence does where it stands, not its topic.

MUST KNOW POINTS FOR YOU.

- Write the outline before the prose: one sentence for each planned paragraph, each sentence the claim that paragraph makes. Then read the sentences alone. If they do not take a reader from your question to your answer, fix them there.
- An introduction that reviews everything known about obesity buries the gap. A reader should be able to underline the sentence that names the gap, and the one that states your question.
- Mensh and Kording end the introduction on a summary of the results; ICMJE says the introduction holds no data or conclusions from the work. For a medical journal that follows ICMJE, end on the question or research objective. Then read the journal's instructions, which decide.
- The methods hold what you knew when you wrote the protocol. How many people declined, who was excluded once the data came in, and anything you changed along the way are results.
- Give the results subheadings that state findings, not topics: "Overweight or obesity was commoner in urban than in rural women", not "Urban-rural comparison". Without subheadings, the same claim opens the paragraph, as section 2 teaches for every results paragraph. This section's question-then-answer order governs only the introduction's end and the discussion's start.
- "The discussion begins with background and other people's studies" is the error. Begin with your main finding, briefly, as the answer to your question.
- Claim no more than your data carry. A new idea the data suggest is written as a hypothesis, and labelled as one: "We suggest, as a hypothesis for a later study, that …".
- "Stating limitations weakens the paper" is the error. A limitation you leave out is still there. Name each limit, say which way it could push the answer, and say what would settle it. For example, if women with higher weight more often declined to be weighed, the share you report is too low.
- The structure shows the logic; it cannot add evidence. If an outline sentence has no data behind it, either cut the claim or get the data.

COMMON MISREADING. "My paper has an Introduction, Methods, Results and Discussion, so it is structured." The four headings are containers. A paper can carry all four and still be a diary of what was done. Its introduction never names a gap. Its results list tables. Its discussion opens

with other people's studies. The structure that matters is the chain inside the headings: the gap, the question, the evidence in steps, the answer and its limits.

REPORTING IT. This analysis compared the share of adults aged 15-49 in India with overweight or obesity in NFHS-4 (2015-16) and in NFHS-5 (2019-21). It also compared urban and rural areas in NFHS-5.

EXERCISE 1 (retrieval). Without looking back: give the one job of each of the four main sections in a sentence each. Then say in which section the number of people who declined to take part belongs, and why.

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (critique). Here is the first paragraph of the discussion in an invented paper on a survey of overweight or obesity among adults in one rural block. The paper asked how common overweight or obesity was among adults in the block, and whether it differed by age group.

> *Obesity is a major public health problem in India and the world. Several studies from different states have reported different levels of overweight. Our study was conducted in a rural block of Chhattisgarh. Table 2 shows the distribution of body mass index by age group. Further research is needed.*

Say what is wrong with it. Then rewrite it so that it does the first paragraph's job. Where you need the study's own results, write a labelled blank rather than a number.

EXERCISE 3 (design). Take the manuscript or thesis section you chose to rewrite for this book. Write its outline: one sentence for each planned paragraph, each sentence the claim the paragraph will make. Then read the sentences alone and answer three questions. Where is the gap? Which sentence answers the question? Which claims have no data or reference behind them yet?

EXERCISE 4 (teaching). A first-year resident shows you the introduction to her draft paper. It runs to four pages and reviews obesity from its definition to its complications, worldwide and in India. Ten minutes, with a whiteboard: explain what an introduction is for, and how she should cut hers.

5 · Citing and referencing: the original source, in the order cited, in one format

Checked against the sources as they stood on 2026-10-02.

DEFINITION. This section follows two documents. The first is the Recommendations of the International Committee of Medical Journal Editors (ICMJE), updated January 2026, section IV.A.3.g, References. The committee was once known as the Vancouver Group. The second document is the style guide those Recommendations point to for format: Citing Medicine, 2nd edition, from the US National Library of Medicine (NLM).

What to cite. Cite original research directly where you can, and the published article rather than its abstract. Where you cannot obtain the original, this book's advice is to cite the source you read and say in your sentence whose figure it is. Do not cite articles in predatory or pseudo-journals.

Mark a preprint as a preprint.

The authors must check each reference against a bibliographic database such as PubMed or against the original.

How to number. References are numbered consecutively in the order in which they are first mentioned in the text. They are identified in the text, tables and figure legends by Arabic numerals (1, 2, 3) in parentheses. A reference cited only in a table or a figure legend takes its number from the point where the text first mentions that table or figure.

How to format a journal article. Citing Medicine lists the elements in the order they appear. For an ordinary journal article these are the authors, the article title, the abbreviated journal title, the date of publication, the volume, the issue in parentheses, and the pages.

Authors are entered surname first, with at most two initials. Only the first word of an article title, its proper nouns and adjectives, acronyms and initialisms take capitals. Journal titles are abbreviated as the NLM Catalog lists them. A page range drops the repeated digits: 123-125 becomes 123-5. [1, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102, 103, 104, 19]

IN PLAIN TERMS. Every sentence in your paper that rests on someone else's work carries a number, such as (3). The number points to one entry in the list at the end of the paper, the reference list. Number your sources in the order the reader first meets them. A source you use again keeps the number it got first.

Cite what you actually read. If you read a preprint, a paper posted publicly before a journal has checked and accepted it, cite the preprint and say it is one.

A reference manager, a program that stores your references and types out the list, can do the typing.

ICMJE asks you not to cite predatory or pseudo-journals, and ICMJE's references section does not define them. In this book, read the phrase as journals that present themselves as checking what they publish and do not. How to recognise one is taught at the next rung.

Illustration 1. Start with a shortcut that fails. You want to cite a paper on figures that this book uses later. Its journal page offers a ready-made line, headed "Citation". Copy it into your reference list, and the entry reads:

> *Rougier NP, Droettboom M, Bourne PE (2014) Ten Simple Rules for Better Figures. PLoS Comput Biol 10(9): e1003833. https://doi.org/10.1371/journal.pcbi.1003833*

That line is the journal's own style, not NLM's. Put it among entries in NLM format and it stands out. The year sits after the authors, every main word of the title has a capital, and the punctuation differs. Rebuild it one element at a time, in the order Citing Medicine lists them. The journal page also says "Published: September 11, 2014".

element	the journal's line	the NLM rule	your entry
authors	Rougier NP, Droettboom M, Bourne PE	surname first, at most two initials; a full stop after the last author	Rougier NP, Droettboom M, Bourne PE.
article title	Ten Simple Rules for Better Figures	a capital only on the first word and on proper nouns; a full stop	Ten simple rules for better figures.
journal title	PLoS Comput Biol	the abbreviation in the NLM Catalog; a full stop	PLoS Comput Biol.
date	(2014), after the authors	year, month, day; the month in its first three letters; a semicolon	2014 Sep 11;
volume and issue	10(9):	the volume, then the issue in parentheses, then a colon	10(9):
pages	e1003833	where the pages go; a full stop	e1003833.

Put the right-hand column together:

> *Rougier NP, Droettboom M, Bourne PE. Ten simple rules for better figures. PLoS Comput Biol. 2014 Sep 11;10(9):e1003833.*

Two checks remain. First, look the journal up in the NLM Catalog, https://www.ncbi.nlm.nih.gov/nlmcatalog/journals, and confirm that its abbreviation is the one the journal printed. Do not take the journal's line as proof of that. Second, this journal gives "e1003833" where a page range would go, in its line and so in yours. If your target journal asks for the DOI, the paper's permanent identifier, add it at the end, as Citing Medicine's example 71 does: doi:10.1371/journal.pcbi.1003833.

Citing Medicine is at https://www.ncbi.nlm.nih.gov/books/NBK7256/, and its chapter 1 covers journals. ICMJE's references section is at https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html.

Figures quoted in this illustration are from [105].

WHERE THIS PICTURE BREAKS. This rebuilds one kind of source: a journal article with three authors and no page range. A book, a report such as the National Family Health Survey (NFHS) fact sheet, a web page or a dataset follows other chapters of Citing Medicine, which this section does not teach. Open the right chapter before you format one.

A journal may also ask for its own variant: fewer authors before "et al." (Latin for "and others"), no day of the month, no issue number, or the DOI. Its instructions for authors win over this table.

This book's own reference lists are printed in the series' house style, and are not a model. Your references follow NLM, as this section teaches.

Illustration 2. Now the numbers in the text. Your draft introduction cites four sources, called A, B, C and D here, and has one table. Go through the text from the top. Each source gets the next free number the first time the text mentions it. A source cited only inside the table gets its number where the text first mentions the table.

where in the text	what it cites	number in the text
sentence 1	A	A is (1)
sentence 2	B	B is (2)
sentence 3	A again	A is (1)
sentence 4, which first mentions Table 1	Table 1 cites D	D is (3)
sentence 5	C	C is (4)

Two things here catch people. A cited again in sentence 3 is still (1): a source keeps its first number however often you return to it. And D gets 3, before C, although D appears only in the table. The table entered the text at sentence 4, so D's number comes from there. C, in sentence 5, gets 4. The reference list then runs 1 to 4 in that order: A, B, D, C.

Now move sentence 5 above sentence 4 in a later draft. C becomes (3) and D becomes (4), in the text, in the table and in the list. Let your reference manager do the renumbering, then read the numbers once more against the list.

WHERE THIS PICTURE BREAKS. This numbering is the system ICMJE describes, with the numbers in parentheses. A journal's instructions may show the numbers in another form, or may not number references at all; then its instructions tell you how to order the list.

This book numbers each passage it checked, not each source, and uses square brackets, so that every number names the exact words behind a claim. A journal paper follows ICMJE.

Illustration 3. Here is the trap with a number found second-hand. Suppose you read in a review that about a quarter of Indian women have overweight or obesity, citing the national survey. The review is invented for this example. Do not cite the review for the number. Do not cite the survey on the review's word either.

Open the NFHS-5 India fact sheet at https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf and find indicator 88, women with overweight or obesity. The row

reads 33.2, 19.7, 24.0 and 20.6: NFHS-5 urban, rural and total, then the NFHS-4 total, in per cent of women aged 15-49. “About a quarter” was the total, 24.0 per cent. Write your sentence from the row you read, with the age range the row carries, and cite the fact sheet.

The same rule covers versions of one paper. NLM’s page of sample references lists a preprint, “Ten simple rules for structuring papers”, posted on bioRxiv, a website for preprints. Its authors are given as “Kording KP, Mensh B”, and it is dated 2016 and marked “[Preprint]”. The paper this book cites under that title is the one published in 2017 in PLoS Computational Biology. Its copyright line reads “Mensh, Kording”: the other name comes first. If you read the preprint, cite the preprint and mark it as one. If you read the journal’s version, cite that. A reader who searches for one with the details of the other may not find it.

Figures quoted in this illustration are from [24, 106, 107].

WHERE THIS PICTURE BREAKS. Opening the original tells you what it says. It does not tell you that it is right, or that a later paper has not overturned it; Book 0’s section on checking a reference (Book 0, F3) stops at the same place.

The fact sheet’s figure is a share of women aged 15-49 in 2019-21. It is not a figure for all adult women, and the review’s “about a quarter” hides that.

MUST KNOW POINTS FOR YOU.

- Open the original before you cite it. If you met a finding in a review, find the paper the review cites and read the passage. Cite that paper only if it says what you claim.
- “My reference manager makes my references correct” is the error. It formats what it was given, and what it was given can be wrong. Check every entry against PubMed or the paper itself, and check that the paper supports the sentence it is attached to.
- Before you submit, check that no reference is a retracted paper. Search PubMed for “Retracted publication [pt]”; “pt” stands for publication type. Or look each paper up in PubMed.
- Cite something not formally published by what it is. A preprint carries “[Preprint]”. A paper accepted but not yet out is “forthcoming”, NLM’s preferred word, or “in press”.
- Never cite an artificial intelligence tool as the source of a fact. If a tool gives you a reference, find the paper, open it, and read the passage before you cite it. If you cannot find it, do not cite it.
- Do not cite a paper from a predatory or pseudo-journal; ICMJE asks you to avoid them.
- How many authors to list is the journal’s choice. Read the journal’s instructions, set your reference manager to match, and check the first entry with more than six authors by eye.
- A reference in perfect NLM format can still point to a paper that does not say what your sentence says. When you check a student’s thesis, open three of its references at random and read the passages they point to.

COMMON MISREADING. “Number the references in order” read as sorting the list, by author or by year. The order is the order in which the text first mentions each source. A source cited again

keeps its first number. A source cited only in a table takes its number from where the text first mentions the table. The reference list is then printed in number order.

REPORTING IT. Among women aged 15-49 in India, 24.0 per cent had overweight or obesity in NFHS-5 (2019-21), against 20.6 per cent in NFHS-4 (2015-16) (1).

EXERCISE 1 (retrieval). From memory: in what order does ICMJE number references, and how does it show them in the text? Then list, in order, the elements of an NLM reference to an ordinary journal article.

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (critique). Your target journal follows ICMJE and NLM, and lists up to six authors before "et al.". A co-author hands you these two entries. Rewrite each in NLM format. The second paper is in volume 13, issue 4. Its four authors are Tracey L Weissgerber, Natasa M Milic, Stacey J Winham and Vesna D Garovic.

1. Cole, T. J. (2015). Too Many Digits: The Presentation of Numerical Data. Archives of Disease in Childhood, 100(7), 608-609.
2. Weissgerber TL et al. Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm. PLoS Biol 2015; 13: e1002128

EXERCISE 3 (interpretation). A draft paragraph cites sources P, Q, R and S. Sentence 1 cites P as (1). Sentence 2 cites Q as (2). Sentence 3 first mentions Table 1, which cites R; the draft numbers R as (5), "because it is only in a table". Sentence 4 cites S as (3). Sentence 5 cites P again, as (4). Which numbers are wrong under ICMJE's rule, and what should every number be?

EXERCISE 4 (teaching). A first-year resident asks: "I found the NFHS-5 figure for women in a review article. Why can I not cite the review, when it gives the same number?" Ten minutes, with a whiteboard.

6 · The parts of a sentence a writer works with

DEFINITION. A sentence is built from clauses. A clause is a group of words with a subject and a verb. A subordinate clause also has a subject and a verb, but it cannot stand alone. Words such as although, because, even if, when, since or while open one.

The verb expresses an action, an occurrence or a state of being. A verb may carry a small word with it, as in "left out"; the two together are the verb.

The subject names what the clause is about: the thing the verb is said of. Its main word is the noun the verb is said of, and the rest of the phrase describes that noun.

The object is the noun that receives the action of the verb.

The doer is whoever or whatever performs the action, whether or not it is the subject.

Voice says which of them is the subject. In the active voice, the subject performs the action of the verb. In the passive voice, the subject receives the action: the active sentence's object becomes the subject. The doer moves into a phrase beginning "by", or is left out. A passive verb has a form of "to be" and a past participle, the form that follows "has" or "was": measured, given, left. [108, 109, 110, 110, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119]

IN PLAIN TERMS. The verb is the word for what happens: measured, rose, is.

Ask "who or what" in front of the verb: who measured? The answer is the subject.

In "We measured the weight", the doer is "we", and "we" is also the subject. That is the active voice.

Turn it round: "The weight was measured". Now the thing measured is the subject, and the doer has gone. That is the passive voice. Its sign is a form of "to be" with a verb ending, usually, in -ed.

A main clause can stand as a sentence on its own. A clause that starts with "because", "when" or "although" cannot.

Illustration 1. Take a sentence of the kind found in the methods chapter of almost any thesis. This one is made up.

> *Written informed consent was obtained from all participants by the investigator before the interview.*

Find the verb first. Ask what happens in the sentence: something was obtained. The verb is "was obtained". It carries both signs of the passive: a form of "to be", "was", and a past participle, "obtained".

Now find the subject. Ask what was obtained: written informed consent. That is the subject, and it is not the doer. Consent did nothing. It received the action.

Now find the doer. Look for "by". The investigator obtained the consent.

Turn the sentence round so that the doer is the subject.

> *The investigator obtained written informed consent from all participants before the interview.*

The subject is now "the investigator", the verb is "obtained", and the object is "written informed consent". The facts have not changed. Neither has the tense: both sentences are in the past. Only the voice has changed.

Now try a harder one, from the same made-up chapter.

> *Data were collected using a pretested questionnaire and entered in a spreadsheet.*

The verbs are "were collected" and "(were) entered". Both are passive. Look for "by". There is none. The sentence never says who collected the data, or who typed them in. To turn it into the active voice, you have to know the answer. "Two trained interviewers collected the data" is a different sentence from "I collected the data", and a reader judging the study needs to know which one is true. The passive let the writer skip that question. The reader cannot skip it.

WHERE THIS PICTURE BREAKS. The two signs of the passive are a test, not a proof. The guideline that lists them says both may not appear in every passive sentence. "The forms got lost" is passive with no form of "to be". And "the clinic was closed on Sunday" may describe how the clinic was, not something anybody did to it. When the signs mislead, ask the plain question: did the subject do the action, or receive it?

Turning a sentence into the active voice is a test you can run, not an order to obey. Some passive sentences are the better sentence. The next section shows when.

Illustration 2. The subject is not always one word. Read this sentence. It is made up, but its figure is the survey's own.

> *In the fifth National Family Health Survey (NFHS-5), the proportion of women aged 15 to 49 in India who had overweight or obesity was 24.0 per cent.*

First take off any opening phrase that ends with a comma: "In the fifth National Family Health Survey (NFHS-5)". Then find the verb. It is "was".

Ask what was 24.0 per cent. The answer is the whole of "the proportion of women aged 15 to 49 in India who had overweight or obesity". That long phrase is the subject. Its main word is "proportion". The rest tells you which proportion.

Count the words between "proportion" and "was". There are 13. The reader holds all 13 in mind while waiting for the verb. The next section calls this the gap between subject and verb, and asks you to close it.

Inside that subject sits a clause of its own: “who had overweight or obesity”. It opens with “who”, a word the list above does not name. It has a subject, “who”, and a verb, “had”. Read it alone and it is not a sentence. So it is a subordinate clause.

Now a sentence with two clauses side by side.

> *Because the fact sheet’s adult rows cover ages 15 to 49 only, they say nothing about women of 50 or older.*

Read each clause alone. “They say nothing about women of 50 or older” stands as a sentence: it is the main clause. “Because the fact sheet’s adult rows cover ages 15 to 49 only” does not stand: it is subordinate. The main point of the sentence belongs in the main clause, and here it is.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. “Take off the opening phrase, then find the verb” works for most sentences in a paper. It fails on a question (“Did the proportion rise?”), where the verb comes first, and on a sentence that starts “There were”, where “there” only fills the place of the subject. In “There were three villages without a road”, the subject is “villages”.

The count of 13 words is for this one sentence. It shows what a long subject costs. It is not a limit.

MUST KNOW POINTS FOR YOU.

- “The passive voice is the past tense” is the error. Voice says whether the subject did the action or received it. Tense says when. When an editor tells you to remove the passive, check voice, not tense.
- Verb first. Then ask “who or what” in front of it to find the subject. Then look for “by”, or ask who did it, to find the doer.
- A passive sentence with no “by” hides who did the work. Before you submit, find each passive verb and ask whether the reader needs the doer. If yes, name it.
- The two signs of the passive, a form of “to be” and an -ed word, are a quick test that can mislead. When the test is unsure, ask whether the subject did the action or received it.
- Put the point of a sentence in its main clause. If your finding sits in the subordinate clause, the reader will weigh it less than you meant.

COMMON MISREADING. “The passive hides the doer, so every passive sentence is wrong and must go.” It does not follow. A passive sentence is right when the doer does not matter, or when the paragraph is telling the story of the thing acted on. Gopen and Swan’s example is “Pollen is

dispersed by bees", which is the better sentence in a paragraph about pollen. The passive is a choice of subject. The next section shows how to make that choice.

REPORTING IT. The investigator obtained written informed consent from every participant before the interview.

EXERCISE 1 (retrieval). Without looking back, name the two signs of a passive verb. Then say what is wrong with the statement "the passive voice is the past tense", with one example.

EXERCISE 2 (interpretation). The three sentences below are made up. In each one, mark the verb, the subject and its main word, the object if there is one, and the voice. Name the doer, or say that the sentence does not name one. Where there are two clauses, say which is the main clause.

1. The auxiliary nurse midwife weighed every child at the anganwadi, the government child care centre of the village.
2. Three villages were left out because the road was closed.
3. There were many households without a weighing scale.

EXERCISE 3 (critique). Open the manuscript section you chose for this book's build. Find every passive verb in it. For each one, write down the doer, if you know it. Then mark the ones where a reader judging your work would need to know who did it.

7 · The sentence as the unit of clarity

DEFINITION. A sentence is clear when a reader takes from it, on one reading, the point the writer meant, and gives each part of it the weight the writer meant. Readers look for each kind of information in a fixed place in the sentence. Five principles put it there. They come from Gopen and Swan's reader-expectation approach (1990) and the US Federal Plain Language Guidelines (2011). Gopen and Swan call theirs principles, not rules.

1. One idea per sentence. Each sentence makes a single point.
2. The subject close to its verb. Readers expect the verb straight after the subject.
3. The action in the verb. A hidden verb is an action turned into a noun, such as "increase", "assessment" or "measurement", carried by an empty verb such as "was observed", "was done" or "make".
4. Old information first, new information last. The start of a sentence, the topic position, holds the information that links back to what the reader already has, and names whose story the sentence tells. The end, the stress position, holds the new information the reader should weigh most.
5. The doer as the subject where the doer matters, in the active voice, with "we" for the authors' own actions. The passive is kept where the doer does not matter, or where the sentence carries on the story of the thing acted on. [120, 121, 122, 123, 124, 125, 126, 127, 127, 128, 120, 123, 129, 118, 130, 130, 131, 131, 132, 132, 116, 117, 133, 134, 135, 136]

IN PLAIN TERMS. A reader reads your sentence once and decides what it said.

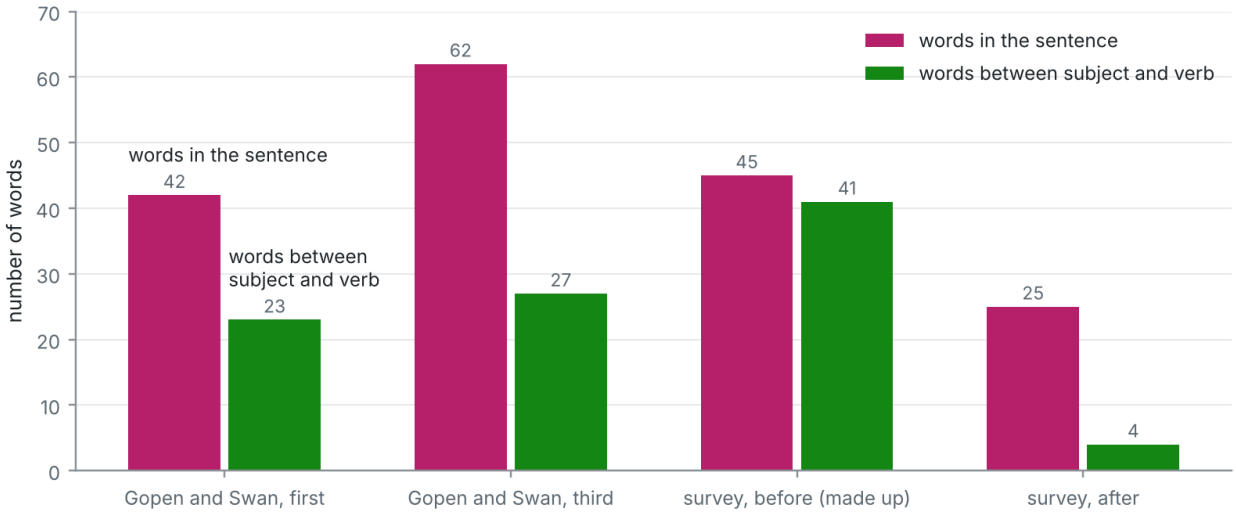
Say one thing in each sentence.

Put the verb soon after the subject.

Put the action in the verb. Write "the proportion rose", not "an increase was observed".

Start with what the reader already knows, and end with what is new.

Name who did it when that matters. If it was you, write "we".



The four sentences of illustration 1 below: how many words each holds, and how many sit between its subject and its verb. Gopen and Swan's first sentence has 42 words, 23 of them between subject and verb; their third has 62, with 27 between (their counts; the sentences are in their article). The made-up survey sentence has 45 words, 41 between, before mending, and 25 words, 4 between, after.

Illustration 1. Start with a sentence that fails. It is made up, but its numbers are from the National Family Health Survey (NFHS), and sentences like it fill Indian thesis chapters. NFHS-4 is the round of the survey before NFHS-5.

> *An increase in the proportion of women in the age group of 15 to 49 years with overweight or obesity, from 20.6 per cent in NFHS-4 (2015-16) to 24.0 per cent in NFHS-5 (2019-21), as per the India fact sheet of the survey, was observed.*

Mend it in five steps.

1. Find the subject and the verb. The verb is "was observed", the last two words. Ask what was observed: an increase. The subject's main word is "increase", the second word. Count the words after it and before the verb, taking "was" as part of the verb. A figure or a bracketed year counts as one word. There are 41. The reader carries 41 words, waiting to learn what the increase did. When the verb arrives, it says only that somebody saw it.

2. Find the action. What happened? A proportion rose. The action is hidden in the noun "increase", and an empty verb, "was observed", carries it. Turn the noun back into a verb: rose.

3. Find whose story it is. Nobody in the sentence did anything worth naming. The story is the proportion's. Make it the subject, and put "rose" straight after it.

4. Put old information first and new information last. The reader of your results chapter already knows the study is about women aged 15 to 49. That goes first. The new thing is where the proportion got to, 24.0 per cent. That goes last, where a reader puts the most weight.

5. Take out what belongs somewhere else. "As per the India fact sheet of the survey" is a source. It becomes a reference number, as the section on citing set out (section 5).

> *Among women aged 15 to 49, the proportion with overweight or obesity rose between NFHS-4 (2015-16) and NFHS-5 (2019-21), from 20.6 to 24.0 per cent.*

Now count both sentences, and set them beside a published one. Gopen and Swan found the same fault in a paper on genes. Its first sentence was 42 words long. Search their article for "23 words".

> *Note that the grammatical subject ("the smallest") is separated from its verb ("has been identified") by 23 words, more than half the sentence.*

The third sentence of the same passage was 62 words long, with 27 words between its subject, "experiments", and its verb, "have indicated".

sentence (Gopen and Swan's counts are their own)	words in the sentence	words between subject and verb
Gopen and Swan, first	42	23
Gopen and Swan, third	62	27
survey, before (made up)	45	41
survey, after	25	4

The mended sentence is shorter. That is a side effect, not the aim. The gap between subject and verb fell by 37 words, from 41 to 4, and the number you want remembered now sits at the end.

Figures quoted in this illustration are from [24, 137].

WHERE THIS PICTURE BREAKS. The steps worked because this sentence had one point. Gopen and Swan's own rewrite of their gene passage was not shorter than the original, and they say sheer length was neither the problem nor the solution. If a sentence holds two points that each deserve weight, it needs two sentences, and the paragraph gets longer.

"Old information first" depends on what the reader already has. The same sentence, opening the first paragraph of a paper, comes before the reader knows the study is about women aged 15 to 49. There it may need a sentence of context before it.

Illustration 2. Now the other direction. The active voice is not always the better sentence. Gopen and Swan give a pair: "Bees disperse pollen" and "Pollen is dispersed by bees". The first is a sentence about bees, the second a sentence about pollen. In a paragraph telling the story of pollen, the passive one is better.

Try the same test on a methods paragraph. This one is made up.

> *Each woman was weighed twice. The two readings were averaged.*

Ask of each verb: does the reader need to know who did it? For the weighing, yes. A reviewer judging the measurement needs to know whether trained staff weighed the women, or the women read their own scales. So name the doer.

> *Two trained field staff weighed each woman twice. The two readings were averaged.*

For the averaging, no. The average comes out the same whoever works it out, and the sentence carries on the story of the readings. It can stay passive.

When the doer is you, write "we". Gopen and Swan rewrote a chemistry passage so that it opened "We have directly measured", because the authors' own measurement was the news. In their words, "since the authors performed the first direct measurement, recognizing their agency in the topic position may well be appropriate."

WHERE THIS PICTURE BREAKS. Whether "we" is welcome is a matter for the journal or the university. Look at its instructions and at papers it has recently published. The test in this illustration still applies inside the passive: if a reader needs the doer, a passive sentence must name it after "by".

"Does the reader need to know who did it?" has no answer fixed for all time. A step that seems to need no doer, such as averaging, needs one the day someone asks whether the averages were checked.

Illustration 3. Last, a sentence with too many points. It is made up, of a kind common in methods chapters.

> *Height and weight were measured by trained field staff using a stadiometer and a digital scale respectively, and body mass index (BMI) was calculated and women were classified as having overweight or obesity using a cut-off of 25.0 kg/m^2.*

Count the points that deserve weight. Who measured, with what, how BMI was worked out, and where the cut-off sits. That is four, and a sentence has one end. Split it, giving each point a sentence of its own.

> *Trained field staff measured each woman's height with a stadiometer and her weight with a digital scale. From these we worked out her body mass index (BMI). We counted a BMI of 25.0 kg/m^2 or more as overweight or obesity.*

Here is the tempting wrong fix. It compresses instead of splitting.

> *Trained-staff stadiometer and digital-scale height-weight measurement-based BMI cut-off classification was done.*

It is shorter, and it is worse. It turns every action into a noun, stacks the nouns, and ends on an empty verb. The US Federal Plain Language Guidelines warn that "Readability suffers when three words that are ordinarily separate nouns follow in succession." This one stacks more than three.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. Splitting can cut a link the reader needs. An "if" and its "then" belong together. When you split them, carry the link into the second sentence with a word such as "in this case", as the US Federal Plain Language Guidelines do with "suppose" and "in this case". Splitting is right when the points can stand apart; when they cannot, keep the link word.

The cut-off of 25.0 kg/m^2 is the one the NFHS-5 fact sheet uses for adults aged 15 to 49. A thesis that uses another cut-off states its own.

MUST KNOW POINTS FOR YOU.

- Find the subject and the verb, and count the words between them.
- "Long sentences are the problem, so cut words" is the error. The test is how many points in the sentence deserve weight.
- Search your draft for the empty verbs that carry hidden actions: "was done", "was carried out", "was observed", "was performed".
- End a results sentence on the number or finding you want remembered.
- In a methods section, name the doer of every step a reviewer would judge: who measured, who interviewed, who entered the data. Keep the passive only where the result is the same whoever did it.
- Split a sentence with several points into several sentences. A chain of three or more nouns makes the reader guess how they relate.
- These are principles, and they stop working when followed blindly. If a rewrite makes a sentence worse for its paragraph, as "Bees disperse pollen" would in a paragraph about pollen, keep the original.

COMMON MISREADING. "One idea per sentence" read as "keep every sentence short", so that a draft is chopped until it scores well. That is not the principle. The test is the number of points that deserve weight, not the number of words. A 30-word sentence with one point and its subject next to its verb is clear. A 10-word sentence with two points is not. Chopping a draft into fragments also cuts the links that tell the reader how one sentence leads to the next.

REPORTING IT. Among women aged 15 to 49 in India, the proportion with overweight or obesity rose between 2015-16 (NFHS-4) and 2019-21 (NFHS-5), from 20.6 to 24.0 per cent.

EXERCISE 1 (retrieval). Without looking back: where in a sentence does a reader put the most weight? Where does a reader look for the link to what came before? What should you put in each place?

EXERCISE 2 (critique). This sentence is made up, from a clinic audit at a primary health centre.

> *Assessment of the adherence of the patients attending the hypertension clinic to their prescribed medicines by pill count was done at every visit by the staff nurse.*

Find the subject's main word and the verb, and count the words between them. Name the hidden verb and the doer. Then rewrite the sentence so that the doer is the subject and the action is the verb.

EXERCISE 3 (teaching). A first-year resident brings you this pair from her thesis methods. She says you told her always to use the active voice, so she changed the first into the second.

> *Each blood sample was centrifuged within one hour of collection. Each blood sample was labelled with the participant's code.*
>
> *The laboratory technician centrifuged each blood sample within one hour of collection. The laboratory technician labelled each blood sample with the participant's code.*

You have ten minutes and a whiteboard. Explain when the active voice helps here and when it does not.

EXERCISE 4 (critique). Take five sentences from the results part of the manuscript section you chose for this book's build. For each, write down its subject, its verb, the number of words between them, and the last three words of the sentence. Rewrite any sentence whose gap is long, whose action hides in a noun, or whose last words are not the new information.

8 · Words: the everyday word, one term for one thing, abbreviations that cost the reader

Checked against the sources as they stood on 2026-10-02.

DEFINITION. Four practices govern the choice of words in a scientific paper.

The everyday word. Of two words that carry the same meaning, the familiar, short and concrete one is used.

One term for one thing. Once a concept has a name, the same name is used for it throughout.

Abbreviations. An abbreviation is a shortened form of a word or phrase. The International Committee of Medical Journal Editors (ICMJE) asks for standard abbreviations only, and none in the title. At first mention it asks for the full term followed by the abbreviation in brackets, except for a standard unit of measurement. The US Federal Plain Language Guidelines, written for government writing, advise no more than three abbreviations in a document, and preferably two. For papers, Barnett and Doubleday (2020) cite a recommendation of at most three acronyms (Talk Medicine BMJ, 2019). In this book, the limit counts a paper's nonstandard abbreviations. A long document, such as a thesis, may keep more if it spells each out again in every chapter, as this book does in every section.

Statistical words. "Random", "normal", "significant", "correlation" and "sample" have exact meanings in statistics. ICMJE asks authors not to use them in their everyday senses.

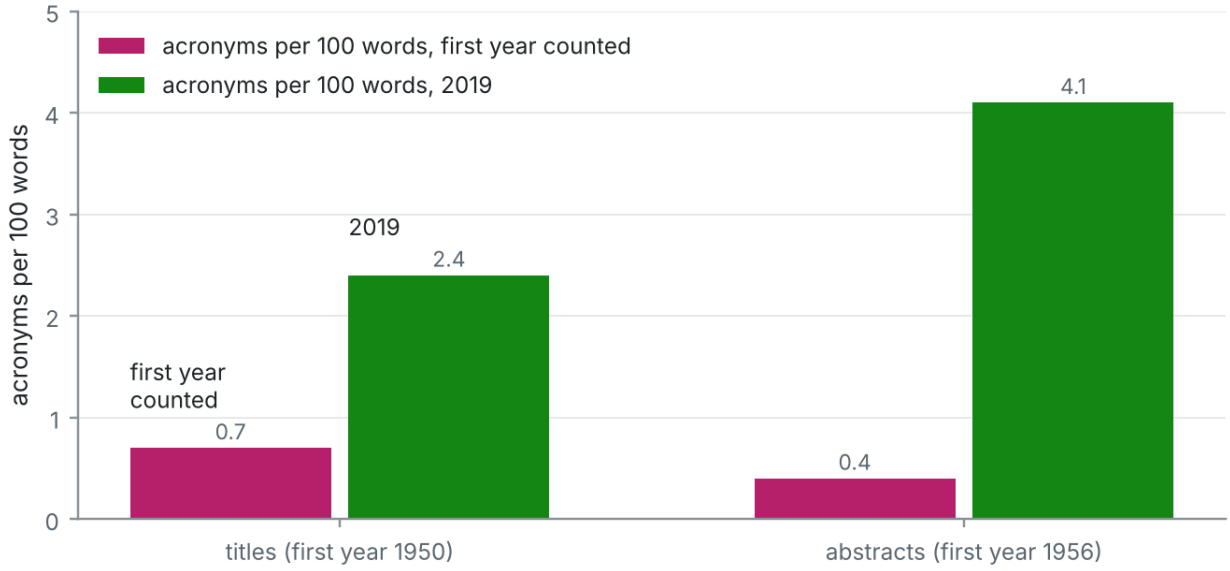
In this section, an acronym is the kind of abbreviation Barnett and Doubleday counted: a word in which half or more of the characters are capital letters. They counted acronyms in 24,873,372 titles and 18,249,091 abstracts of English-language papers in PubMed, published from 1950 to 2019. Their yearly figures for abstracts start in 1956, when enough abstracts became available.

The study counted acronyms in titles and abstracts. It did not test whether readers understood less. [138, 139, 140, 141, 140, 142, 142, 143, 144, 145, 146, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159]

IN PLAIN TERMS. Use the word a reader would say out loud. "Many", not "a number of".

Once you have named a thing, keep the name.

Spell out an abbreviation the first time you use it. Then ask whether it earns its place.



Barnett and Doubleday (2020): acronyms per 100 words in titles and abstracts in PubMed. In titles, 0.7 in 1950 and 2.4 in 2019. In abstracts, 0.4 in 1956 and 4.1 in 2019. A count of use in titles and abstracts, not a measure of how well readers understood them.

Illustration 1. Start with a results paragraph that changes its words. It is made up, but its numbers are those of India's National Family Health Survey (NFHS).

> *Among women aged 15 to 49, overweight or obesity rose from 20.6 per cent in NFHS-4 to 24.0 per cent in NFHS-5. Obesity was more common in urban areas (33.2 per cent) than in rural areas (19.7 per cent). Excess weight was also found in 22.9 per cent of men. High BMI therefore affects about a quarter of Indian adults.*

Count the names for the condition. There are four: overweight or obesity, obesity, excess weight, high BMI (body mass index). A reader meeting "obesity" in the second sentence will take it for a smaller group than "overweight or obesity" in the first. Is it?

Open the NFHS-5 India fact sheet at https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf and find indicator 88.

> *1. Women who are overweight or obese (BMI ≥25.0 kg/m2)21 (%) 33.2 19.7 24.0 20.6*

The urban and rural figures come from that same row. The adult rows of the table give overweight and obesity together, and none gives obesity alone. So the second sentence claims something the fact sheet does not report. The change of word did that, not the numbers.

Now look at the group. The paragraph ends with "Indian adults". The table's adult rows sit under the heading "Nutritional Status of Adults (age 15-49 years)". The group's name drifted too, and the last sentence now covers people the survey's rows do not.

Pick one term and define it once. Then use it in every sentence, even when it feels repetitive.

> *Among women aged 15 to 49, the proportion with overweight or obesity (a body mass index of 25.0 kg/m^2 or more) rose from 20.6 per cent in NFHS-4 to 24.0 per cent in NFHS-5. Overweight or obesity was more common among urban women (33.2 per cent) than among rural women (19.7 per cent). Among men aged 15 to 49, the proportion with overweight or obesity was 22.9 per cent.*

Mensh and Kording put the rule plainly: "There is nothing wrong with using the same word multiple times in a sentence or paragraph."

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. One term for one thing holds only while the thing stays the same. When you do mean a different group, such as obesity alone from a source that reports it, a different term is right, and it needs its own definition.

The fact sheet's row says "women who are overweight or obese". This book writes "women with overweight or obesity", putting the person first. Choose your wording once, quote the source's label once where you define the term, and keep your own wording from then on.

Illustration 2. Now the abbreviations. Here is a made-up sentence of a kind found in Indian theses. It uses five: accredited social health activist (ASHA), primary health centre (PHC), waist circumference (WC), cardiovascular disease (CVD) and non-communicable disease (NCD).

> *ASHAs of the PHC were trained on WC measurement for CVD risk screening in the NCD clinic.*

That is five abbreviations in 17 words. Put three questions to each one. Is it standard for the reader who will read this paper? Does it save the reader anything, or only ink? Will it appear often enough to be worth learning?

"PHC" can become "the primary health centre" once and "the centre" after that. The US Federal Plain Language Guidelines call this a nickname, and prefer it to an abbreviation. "WC", "CVD" and "NCD" each appear once, so spell them out. "ASHA" is the one worth keeping if the chapter is about these workers. An Indian reader knows it. A reviewer abroad may not, so spell it out at first mention.

> *We trained the accredited social health activists (ASHAs) of one primary health centre to measure waist circumference, as part of the centre's screening for cardiovascular disease.*

Is this a real problem, or a matter of taste? Barnett and Doubleday counted acronyms in titles and abstracts in PubMed, the National Library of Medicine's index of medical journals, from 1950 to 2019. Search their article (https://doi.org/10.7554/eLife.60080) for "per 100 words".

> *We found that the proportion of acronyms in titles increased from 0.7 per 100 words in 1950 to 2.4 per 100 words in 2019*

> *the proportion of acronyms in abstracts also increased, from 0.4 per 100 words in 1956 to 4.1 per 100 words in 2019.*

where	acronyms per 100 words, first year counted	acronyms per 100 words, 2019
titles (first year 1950)	0.7	2.4
abstracts (first year 1956)	0.4	4.1

Work out how many times as dense the acronyms became, as a ratio (Book 0, A5).

```
2.4 divided by 0.7 = 3.43
4.1 divided by 0.4 = 10.25
```

In abstracts, acronyms became about ten times as dense. Now search for "only once".

> *with 30% occurring only once, and 49% occurring between two and ten time times. Only 0.2% of acronyms (just over 2,000) occurred more than 10,000 times.*

Most of the 1.1 million different acronyms were rare: 30 per cent appeared once, and another 49 per cent two to ten times. A year after first use, only 11 per cent had been used again in another paper in the same journal.

Acronyms also collide. The same paper reports, from Lang (2019), that "UA" has 18 different meanings in medicine. The authors' own example is heart rate (HR), which can also mean hazard ratio, or hour.

Figures quoted in this illustration are from [160].

WHERE THIS PICTURE BREAKS. Barnett and Doubleday counted acronyms. They did not test whether readers understood less. Their own abstract says acronyms "are not the biggest current problem in science communication". Read the rise as a fact about how papers are written, not as a measure of how much each acronym costs a reader.

The count covers titles and abstracts in English-language PubMed, not full texts, and not Indian journals on their own. The authors expect it to be an undercount, because their method missed acronyms with symbols and lower-case letters.

The ratio of about ten compares 1956 with 2019. It says nothing about any one paper, or about your field.

Illustration 3. Last, four statistical words used in their everyday sense. Each sentence is made up.

> *A significant number of women had overweight or obesity.*

In a paper, "significant" reads as the verdict of a statistical test: roughly, a difference too large to put down to chance. This book does not teach the test. No test was run here, so write the number: "Nearly a quarter of women aged 15 to 49 (24.0 per cent) had overweight or obesity."

> *Households were selected randomly.*

Suppose the field team walked down each lane and called at whichever houses were open. That is not random. As Book 0, D6 showed, a random sample needs a real chance process, so that every household has the same chance of being picked. That section also gives "sample" its exact sense: the part of a population that is measured. Write what was done: "The team visited every house that was open on the survey day."

> *Most men had a normal BMI.*

In statistics, "normal" also names the shape of a distribution, which is beyond this book. As a category of BMI, it means nothing until you give its cut-offs. The fact sheet's own row 86 shows the habit to copy: it says "below normal" and gives the cut-off beside it, "BMI <18.5 kg/m2". Name the range your source uses.

> *Screen time was correlated with BMI.*

A correlation is a number worked out from paired measurements, and this book does not teach it. If you did not compute one, describe what you saw: "Children who reported more screen time tended to have a higher BMI."

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. The rule is against the everyday sense, not against the words. If you did run a test, or did use a chance process, or did compute a correlation, the technical word is the right one. Name the test, the process or the number beside it, so the reader can check it.

"Nearly a quarter" is fair for 24.0 per cent of women aged 15 to 49. It is not a figure for all adults, nor for men, whose own figure is 22.9 per cent.

MUST KNOW POINTS FOR YOU.

- "Using a different word each time makes writing less repetitive, and so better" is the error. A reader who meets a new word assumes a new thing.
- Do not write "obesity" for the NFHS-5 row that reports "overweight or obese". The fact sheet's adult rows give the two together, at a body mass index of 25.0 kg/m^2 or more, and none gives obesity alone.
- Put three questions to every abbreviation in your draft. Is it standard for the reader of this paper? Does it save the reader anything but ink? Will it appear often? In a paper, keep no more than three nonstandard ones, spell each out at first mention, and use none in the title.

- "Significant", "random", "normal", "correlation" and "sample" make a statistical claim when a reviewer reads them. If you ran no test, drew no sample by a chance process, or computed no correlation, use other words, or state the number itself.
- In PubMed abstracts, acronyms rose from 0.4 to 4.1 per 100 words between 1956 and 2019, about ten times as dense (Barnett and Doubleday, 2020). The count measured use, not understanding, and only in English-language titles and abstracts. When it matters, as in a report that ASHAs or patients will read, give the text to one such reader and watch where they stop.
- An abbreviation every Indian community medicine reader knows, such as ASHA, may be unknown to a journal's reviewer abroad.

COMMON MISREADING. "Once an abbreviation is defined at first use, it costs the reader nothing." It still costs. A reader who meets it three pages later has to stop and look back. The US Federal Plain Language Guidelines and Mensh and Kording both name that cost. Many acronyms are rare, so a reader may well not know the one in front of them: Barnett and Doubleday found 30 per cent used only once, and 49 per cent two to ten times. Defining one is the minimum, not a licence.

REPORTING IT. In this chapter, overweight or obesity means a body mass index of 25.0 kg/m^2 or more, the NFHS-5 cut-off for adults aged 15 to 49.

EXERCISE 1 (retrieval). Without looking back: what does the International Committee of Medical Journal Editors ask for at an abbreviation's first mention? Where does it ask you to use none? Name four statistical words that should not be used in their everyday sense.

EXERCISE 2 (critique). This results paragraph is made up, but its two numbers are the survey's own. Find every word problem in it and rewrite it.

> *Urban men showed significantly higher obesity (29.8%) than rural males (19.3%), and high BMI was correlated with urbanisation.*

EXERCISE 3 (critique). A colleague's slide reads: "A study of 24 million papers proved that acronyms make science harder to understand." Say what the study measured, and what the slide claims that it did not.

EXERCISE 4 (design). Make the word list for the manuscript section you chose for this book's build. List every abbreviation in it and every term for its main outcome. Decide which abbreviations stay, and which single term you will use for the outcome. Write the sentence that defines that term.

9 · Numbers in a sentence: the unit, the count behind the percentage, and no more digits than the data hold

DEFINITION. A number in a sentence is reported in full when the reader can check it. A percentage carries the two counts it was worked out from: the numerator, how many had the thing, and the denominator, how many were counted.

A difference between two percentages is stated in percentage points. A relative change divides that difference by the starting percentage, and is stated in per cent.

A number shows false precision when it carries more digits than the data can support.

Cole (2015), a statistical reviewer for medical journals, gives these rules for how many digits to report, as recommendations, not requirements:

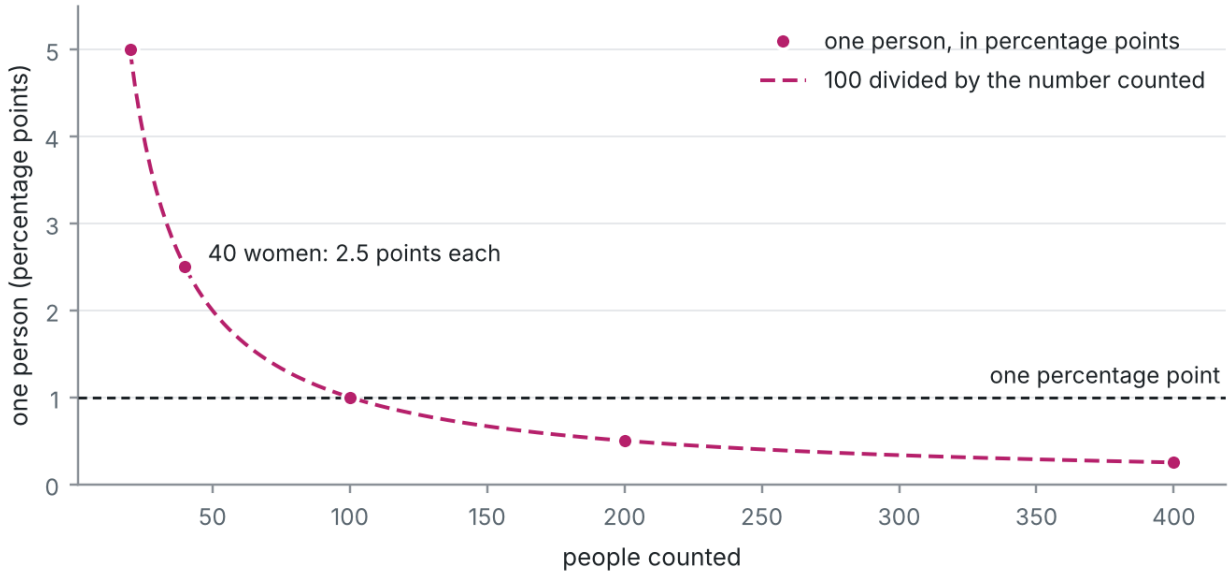
statistic	how many digits to report
mean	enough decimal places to give the standard deviation (SD) two significant figures, or the standard error (SE) one
percentage	whole numbers, or one decimal place for values under 10%
risk ratio, odds ratio, or any ratio of two percentages	two significant figures if the first digit that is not zero is 4 or more, otherwise three: the rule of four
P value	one significant figure, never 0.000

By the SD rule, the mean stops at the place of the SD's second significant figure. A risk ratio divides the risk in one group by the risk in another. Illustration 3 explains odds ratios and P values.

The SAMPL guidelines (Statistical Analyses and Methods in the Published Literature) are by Lang and Altman. They prefer a confidence interval to a P value. If a P value is given, they ask for an equality ($P = 0.03$, not $P < 0.05$, read "P below 0.05"), to one or two decimal places. They do not say when to use one and when two. The smallest that need be reported is $P < 0.001$. They ask for a mean and its SD in the form mean (SD), not mean ± SD (read "plus or minus").

When n people are counted, one person moves a percentage by 100 divided by n (written $100 \div n$) percentage points. [161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172]

IN PLAIN TERMS. Ask three questions of every number you write. In what unit? Out of how many? How finely do the data let you say it?



How far one person moves a percentage: 100 divided by the number counted, in percentage points. With 40 women, each is worth 2.5 points; with 400, 0.25. Below 100 people, one person is worth more than one percentage point, and a decimal place describes less than one person.

Illustration 1. Start with a sentence from a district survey report. The sentence and its numbers are made up for this example.

> *The prevalence of overweight was found to be 37.50 per cent.*

Prevalence, as used in Book 0, D3, is how many people in a group have the condition, here given as a share.

Now ask the three questions. In what unit? A per cent, but of whom, and by what cut-off? Out of how many? The sentence does not say. How finely? To a hundredth of a percentage point.

Go back to the survey's own counts. It weighed and measured 40 women. Fifteen had a body mass index (BMI) of 25.0 kg/m^2 or more, BMI as built in Book 0, B6.

```
15 divided by 40 = 0.375
0.375 times 100 = 37.5
```

Now see how far one woman moves that figure. One woman out of 40 is 100 divided by 40 percentage points.

```
100 divided by 40 = 2.5
16 divided by 40 = 0.4
0.4 times 100 = 40
```

Each woman is worth 2.5 percentage points. With one more woman the figure jumps from 37.5 per cent to 40 per cent. Nothing in between can happen. So "37.50 per cent" claims a fineness that 40 women cannot give. Cole's rule for a percentage is a whole number, or one decimal place below 10 per cent. Here 37.5 rounds to 38.

Rewrite the sentence with its count first and its cut-off named.

> *Of the 40 women weighed and measured, 15 (38 per cent) had a body mass index of 25.0 kg/m^2 or more.*

Journals print the sign, %. This book writes the words. Follow your journal.

The reader can now check the percentage, and can see how small the group was. The Recommendations of the International Committee of Medical Journal Editors (ICMJE) ask for "the absolute numbers" behind every percentage. The SAMPL guidelines ask for the numerator and denominator.

Watch the step shrink as the count grows.

```
100 divided by 20 = 5
100 divided by 100 = 1
100 divided by 200 = 0.5
100 divided by 400 = 0.25
```

Below 100 people, each person is worth more than a whole percentage point. A decimal place on such a percentage describes less than one person. The count is what carries the information.

Now a national figure, from the fifth National Family Health Survey (NFHS-5, 2019-21). It reports that 24.0 per cent of women aged 15-49 had overweight or obesity, meaning a BMI of 25.0 kg/m^2 or more. The fact sheet prints no count beside it.

Its introduction says the survey gathered information from 724,115 women. That is not the denominator. A footnote on the women's rows leaves out pregnant women and women who gave birth in the two months before. So quote what the fact sheet gives: the percentage, the survey and its years, the age range and the cut-off. Do not make a count by multiplying 724,115 by 0.24.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. The step of 100 ÷ n is the smallest move the count allows. It is not how uncertain the percentage is. A second random sample of 40 women would give a different count, and that sampling variation (Book 0, D6) is a separate matter. So your last digit should be no finer than the step; meeting that is necessary, not sufficient.

The rule of the count binds your own data. For a figure you quote, you can only give what the source gives.

Illustration 2. NFHS-4 (2015-16) put the same row for women at 20.6 per cent. NFHS-5 (2019-21) put it at 24.0 per cent. Work out the change both ways, as in Book 0, A4.

```
24.0 minus 20.6 = 3.4
3.4 divided by 20.6 = 0.165
```

A rise of 3.4 percentage points. That is a rise of about 17 per cent on the NFHS-4 figure. Both are true, and they are different quantities. A reader who meets "a 17 per cent rise" alone may take it for 17 points.

So write the absolute change first. If you add the relative one, label it.

> *Among women aged 15-49, the share with a BMI of 25.0 kg/m^2 or more rose from 20.6 per cent in NFHS-4 (2015-16) to 24.0 per cent in NFHS-5 (2019-21). That is a rise of 3.4 percentage points.*

Do the same for men.

```
22.9 minus 18.9 = 4.0
4.0 divided by 18.9 = 0.212
```

Men rose by 4.0 points, about 21 per cent on their NFHS-4 figure.

Now the digits. On its own, Cole's rule would print 24.0 per cent as 24 per cent. But you are comparing groups, and rounding must not blur the gaps between them. For that case Cole suggests a rule built on the range, the gap from the lowest value to the highest. In his words: "if the range is 10% or more use whole numbers, if less than 1% use two decimal places, and otherwise one". Your four figures run from 18.9 to 24.0.

```
24.0 minus 18.9 = 5.1
```

The range is 5.1, between 1 and 10, so keep one decimal place, as the fact sheet does. Round the women's two figures to whole numbers and they become 21 and 24. Their gap would read as 3, not the 3.4 the survey gives.

Absolute and relative changes in risk get their full treatment in the second rung of this subject. Here the habit is enough: points first, per cent labelled.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. The two surveys measured different women, each a sample of its own time. The change is a difference between two estimates, and each has its own sampling variation. A rise of 3.4 points describes the two surveys. It does not say why the share rose.

The row covers ages 15-49 only, and counts overweight and obese together. It is not a figure for all Indian women, nor for obesity alone.

Illustration 3. Statistical software prints every digit it has. Here is made-up output from a survey of 400 women: mean BMI 24.6318 kg/m^2, SD 4.1237 kg/m^2. The SD is the standard deviation, the spread built in Book 0, D5.

Cole's rule for a mean says: round the SD to two significant figures, then give the mean the same number of decimal places.

```
4.1237 rounds to 4.1
24.6318 rounds to 24.6
```

Write it in the SAMPL form: mean BMI 24.6 kg/m^2 (SD 4.1). Cole's other rule uses the standard error (SE): the SD divided by the square root of the count, as built in Book 0, D6.

```
square root of 400 = 20
4.1237 divided by 20 = 0.206
```

The SE to one significant figure is 0.2. That is one decimal place again, so 24.6. Now suppose the survey had 40,000 women and the same SD.

```
square root of 40,000 = 200
4.1237 divided by 200 = 0.0206
```

The SE is now 0.02, so the SE rule allows two decimal places: 24.63. The SD rule still gives 24.6. The two rules answer different questions. The SD describes how the women differ. The SE describes how well the mean is known, and it shrinks as the count grows. Say which rule you used.

Next, a ratio. Cole's own example is an odds ratio of 22.68. An odds ratio divides the odds of an outcome in one group by the odds in another, with odds as built in Book 0, D1. The paper gave it with a 95 per cent confidence interval (CI) of 7.51 to 73.67. A CI is a range, worked out from the data, that shows how precisely the ratio is known. How it is worked out is beyond this book.

Apply the rule of four. The first digit of 22.68 is 2, below 4, so keep three significant figures: 22.7. The first digits of 7.51 and 73.67 are both 7, so keep two: 7.5 and 74. The result is 22.7 (95 per cent CI 7.5 to 74). Cole also accepts one figure fewer throughout, which gives 23 (8 to 70).

Last, the P value, the number a statistical test prints beside a comparison. What it means is beyond this book; how many digits to give it is not. Here the two guides differ, so look at both. The four values are made up.

software prints	Cole: one significant figure	SAMPL: one or two decimal places
0.38	0.4	0.38 or 0.4
0.047	0.05	0.05
0.0026	0.003	no equality fits: 0.00 would be false
0.000	< 0.001	< 0.001

The two rules differ only for 0.0026, a value below 0.01. Two decimal places turn 0.0026 into 0.00, which says the P value is zero. Cole's point is that any rule counting decimal places fails there. Both forbid 0.000. Your journal's instructions for authors may name a rule. Follow them, and use one rule through the whole paper.

Figures quoted in this illustration are from [173].

WHERE THIS PICTURE BREAKS. Cole writes "round up" for P values. For each value in the table, rounding up and rounding to the nearest give the same figure. For a value such as 0.032 they do not: up gives 0.04, and nearest gives 0.03. Say which you did.

These rules decide how many digits to print. They do not make the mean, the ratio or the P value right. A mean printed to the right digits can still come from 400 women who do not stand for the district.

MUST KNOW POINTS FOR YOU.

- More decimal places look more careful, and that is the error. Round once, when you write the sentence, to the digits Cole's rules give.
- Put the count before the percentage: "15 of 40 (38 per cent)".
- A change between two percentages has two sizes. For women in NFHS it is 3.4 percentage points, or about 17 per cent of the starting figure. Say the points first.
- When fewer than 100 people are counted, one person is worth more than a percentage point. In a small study the count is the result.
- Cole and SAMPL give different rules for P values. Neither allows 0.000.
- Rounding rules are for the page, not for the working. A figure rounded early and then used in a later sum carries its rounding into the answer.
- These rules decide how many digits you print. They cannot tell you whether the number is right.
- The NFHS fact sheet prints percentages with no counts. Never make a count by multiplying the survey's 724,115 women by the percentage.

COMMON MISREADING. "24.0 per cent of Indian women are obese." The fact sheet's row is narrower on three counts. It covers women aged 15-49 only. It counts overweight and obese together, at a BMI of 25.0 kg/m^2 or more, and it has no row for obesity alone. And it leaves out pregnant women and women who gave birth in the two months before. Read the row's own words and its footnote before you write the sentence.

REPORTING IT. In NFHS-5 (2019-21), 24.0 per cent of women aged 15-49 had a body mass index of 25.0 kg/m^2 or more. In NFHS-4 (2015-16) the figure was 20.6 per cent, so the share rose by 3.4 percentage points.

EXERCISE 1 (retrieval). From memory: name the three things a reader needs with every number in a results sentence. Then state Cole's rule for the digits of a mean and of a percentage. Last, name the one value a P value may never be printed as.

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (critique). Here is a made-up results paragraph from a thesis on a school survey. Find every number that is incomplete or carries false precision. Then rewrite the paragraph so that it makes one point.

> *A total of 87 students participated. Overweight was seen in 19.54 per cent of students. Mean age was 13.4782 years ± 1.2066. Girls had 9.2 per cent more overweight than boys (p = 0.000).*

The ± figure is the SD. The counts behind the percentages are 17 of 87 in all, 11 of 42 girls and 6 of 45 boys. The cut-off was the one in the survey's protocol, which the methods section names.

EXERCISE 3 (teaching). A journalist rings you: "So obesity among Indian women is up 17 per cent in five years?" You have three minutes and no jargon. Give the answer you would say, ending in one sentence she can quote.

PRACTICE. Thirteen problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. Write each number with two significant figures: 0.04372, 3465.31, 22.68, 0.5049.

2. 37 of 412. Work out the percentage, then write it as Cole's rule for percentages asks, with its count.

3. A percentage goes from 12.5 per cent to 15.0 per cent. Give the change in percentage points and in per cent.

4. Round each mean by Cole's rule, using the SD beside it.

mean	SD
52.381	11.46
3.0538	0.5712
0.8147	0.0236

5. Round each ratio by the rule of four: 0.3172, 1.0462, 4.183, 11.27, 0.0482.

6. NFHS-5 (2019-21) gives the share of women aged 15-49 with a BMI of 25.0 kg/m^2 or more. It is 33.2 per cent in urban India and 19.7 per cent in rural India. Give the gap in percentage points and the urban figure as a ratio of the rural one, by the rule of four. Then use Cole's suggestion for comparing groups to decide how many decimal places the two percentages need. Write the sentence.

7. Cole quotes a paper: "The mean birth weight of babies was 3.05±0.57 (95% CI 2.95 to 3.15) kg". The 0.57 is the SD. Rewrite the mean and SD in grams, in the SAMPL form, to the digits Cole's rule for a mean gives.

8. Made-up output: mean age 34.2871 years, SD 9.8364 years. Give the mean by Cole's SD rule. Then give it by his SE rule for a sample of 100 and for a sample of 10,000. Say which you would print for 100 people.

9. Made-up software output gives four P values: 0.67, 0.0086, 0.096 and 0.00004. Write each by Cole's rule and by SAMPL's.

10. Here is a worked answer. The counts are made up. Find the step that broke.

```
village A: 15 of 40 women with a BMI of 25.0 kg/m2 or more = 37.50%
village B: 18 of 40 women = 45.00%
45.00 minus 37.50 = 7.50
so village B has 7.50% more women in this group than village A
```

11. Cole cites a review that gives a hazard ratio of 0.03, rounded to two decimal places. A hazard ratio compares how fast an event happens in two groups; here the event was updating a reference work's citations. Here is a worked answer. Find the step that broke.

```
hazard ratio = 0.03
turned the other way round: 1 divided by 0.03 = 33.33
so the other reference work updates its citations 33.33 times as fast
```

12. A news report says: "One in four Indian women is now obese. That is about 1.7 lakh of the 7.2 lakh women surveyed, up 16.5 per cent since the last survey." A lakh is 100,000. Decide what to compute, compute it, and say what the claim does not establish.

13. A made-up project report says: "We visited 3 schools (n = 87). After the new menu, fewer students bought a sugar-sweetened drink at the canteen: 49 per cent, down from 62 per cent (p = 0.000). That is a 21 per cent reduction." Decide what to compute, compute it, and say what the claim does not establish.

10 · One idea per paragraph: the topic sentence, its support, and the reverse outline

DEFINITION. A paragraph is a group of sentences that makes one point. Its topic sentence states that point, often first. Mensh and Kording use a second shape: context, content, conclusion. The first sentence sets the topic, the body carries the content, and the last sentence gives the conclusion to remember. Illustration 2 shows both shapes.

A paragraph zig-zags when it leaves a topic and comes back to it later.

In this book, a reverse outline is the outline of section 4 made after the draft instead of before it. A paragraph whose point cannot be written as one sentence without “and” carries more than one point. [174, 175, 176, 177, 22]

IN PLAIN TERMS. Write the point of each paragraph in the margin, in one sentence. Then read the margin alone. If a margin note needs an “and”, that paragraph is two.



Before the rewrite: the topic of each of the 7 sentences of the made-up paragraph in illustration 1, below, in reading order. Topic 1 is how common and how it changed, 2 is towns against villages, 3 is what the survey counted, 4 is what to do. The topic changes 6 times, and the reader is taken to topics 1, 2 and 3 twice each.



After: the rewrite in illustration 1, below, drawn the same way. Topics 3, 1 and 2 are each in one place, and topic 4 is cut. The topic changes 2 times, and each change falls at a paragraph break.

Illustration 1. Start with a paragraph that fails. Its wording is made up for this example; its numbers are from the India fact sheet of the fifth National Family Health Survey (NFHS-5). Its sentences are numbered so you can follow them.

> 1. *Overweight and obesity are becoming commoner among Indian women. (2) NFHS-5 measured women aged 15-49 and counted those with a body mass index (BMI) of 25.0 kg/m^2 or more.*
> 2. *Urban women were more affected than rural women. (4) The share rose from 20.6 per cent in NFHS-4 to 24.0 per cent in NFHS-5.*
> 3. *Pregnant women and women who had given birth in the previous two months were left out.*
> 4. *In urban areas the figure was 33.2 per cent, against 19.7 per cent in rural areas. (7) This shows that action is needed.*

Every sentence is clear on its own. The paragraph is still hard to follow. Find out why.

First, try the reverse outline. Write the paragraph's one point in a single sentence. You cannot do it without "and": the share rose, and it is higher in towns, and here is what the survey counted, and action is needed. Four points.

Next, give each topic a number. Topic 1 is how common it is and how it changed. Topic 2 is towns against villages. Topic 3 is what the survey counted. Topic 4 is what to do. Label each sentence. Then rewrite the paragraph with each topic in one place, and label the new sentences the same way.

sentence	topic, before	topic, after
1	1	3
2	3	3
3	2	1
4	1	1
5	3	1

sentence	topic, before	topic, after
6	2	2
7	4	2

Count the changes of topic down each column. Before, the topic changes 6 times in 7 sentences. The reader is taken to topics 1, 2 and 3 twice each, and to topic 4 once. After, it changes 2 times, and each change falls at a paragraph break. That is Mensh and Kording's rule: each subject in one place, so the reader changes subject as few times as possible.

Here is the rewrite, in three paragraphs.

> *NFHS-5 counted a woman aged 15-49 as having overweight or obesity when her BMI was 25.0 kg/m^2 or more. It left out pregnant women and women who had given birth in the two months before.*
>
> *Among women aged 15-49, overweight and obesity became commoner between NFHS-4 and NFHS-5. The share rose from 20.6 per cent in 2015-16 to 24.0 per cent in 2019-21. That is a rise of 3.4 percentage points.*
>
> *The share was higher in towns and cities than in villages. It was 33.2 per cent among urban women and 19.7 per cent among rural women.*

Sentence 7 of the old paragraph is gone. Its "this" pointed at nothing in particular, and "action is needed" is a claim no sentence in the paragraph supports. The first sentence of each new paragraph states its point, and the rest support it.

Now write the reverse outline of the rewrite: one sentence per paragraph, in the margin.

paragraph	its one point, written in the margin
1	what NFHS-5 counted as overweight or obesity
2	the share of women in that group rose between the two surveys
3	the share was higher in towns and cities than in villages

Read the margin alone. Notes 2 and 3 are findings. Note 1 is not: it says how the measurement was made. In a paper, that paragraph goes to the methods, where section 4 put everything a reader needs to judge how the study was done. The reverse outline found that in one reading of three short lines.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. The topic numbers are your own labels. Another reader might draw the line between topics in a different place. So the count of changes is a check on a draft, not a score to compare across writers.

Two topics side by side are sometimes the point. A discussion paragraph that weighs two reasons for one interpretation holds both. Mensh and Kording ask for similar ideas to come "one immediately after the other". What fails is leaving a topic and coming back to it.

Illustration 2. The book has now shown a one-point paragraph in two shapes. The claim came first in section 2, the shape for a results paragraph and the abstract. Mensh and Kording open a results paragraph with its question and close it with the answer; section 4 keeps that order for the end of the introduction and the start of the discussion. Here is one paragraph of the rewrite in each shape.

```
33.2 minus 19.7 = 13.5
```

Point first:

> *The share with a BMI of 25.0 kg/m^2 or more was higher among urban women than rural women. In NFHS-5 it was 33.2 per cent among urban women aged 15-49 and 19.7 per cent among rural women, a gap of 13.5 percentage points.*

Question first, answer last:

> *Did the share differ between towns and villages? In NFHS-5 it was 33.2 per cent among urban women aged 15-49 and 19.7 per cent among rural women, a gap of 13.5 percentage points. So the share was higher in towns and cities.*

Both make one point, and both give the same margin note. The first tells a hurried reader the finding at once. The second leads a careful reader to it. Mensh and Kording say each such paragraph "convinces the reader of the answer given in its last sentence". That makes a doubtful conclusion easy to find and check.

Choose one shape for a section and keep it; in this book's results sections, that is point first. Mensh and Kording ask that "parallel messages should be communicated with parallel form". A reader who has learnt where the point sits finds it faster in every paragraph after.

The margin notes also tell you when to split or merge. A note that needs "and" is two paragraphs. Two notes next to each other that say the same thing are one paragraph. A note that serves no claim on the way to your one message, from section 3, belongs elsewhere or nowhere.

Length follows the support, not a rule. The OpenStax writing guide puts it both ways. "When a paragraph is too short, it often lacks enough evidence and examples to back up your claims. When a paragraph is too long, readers can lose the point you are making."

WHERE THIS PICTURE BREAKS. The question-first shape needs a question the reader already wants answered. Opened with a question nobody asked, it delays the point for nothing; then put the point first.

A margin note can be one tidy sentence and still be wrong. The reverse outline tests where the points are, not whether the evidence below each one supports it.

MUST KNOW POINTS FOR YOU.

- "One idea per paragraph" read as "short paragraphs" is the error.

- Split a paragraph whose note needs "and". Merge two neighbours whose notes say the same thing. Move or cut one whose note serves no claim.
- The commonest fault is not a long paragraph but a topic visited twice. Before you edit sentences, label each one with its topic and gather each topic into one place.
- The two shapes of a one-point paragraph, point first and answer last, are both sound. Choose one for a section and use it in every paragraph, so the reader knows where to look.
- When each paragraph has one conclusion, the reviewer can find the one paragraph behind a doubtful claim.
- The reverse outline checks where each point is. It cannot check that the point is true, or that the sentences under it support it.
- When you correct a trainee's thesis chapter, do not rewrite their paragraphs. Ask them to write the margin note for each one.

COMMON MISREADING. "One idea per paragraph" is often read as "keep paragraphs short". A paragraph cut short has its point and too little support for it. One idea means one point with all the support it needs, and no second point. A long paragraph that makes one point passes; a three-sentence paragraph that makes two fails.

EXERCISE 1 (retrieval). From memory: what are the three parts of a paragraph in Mensh and Kording's scheme? What is a reverse outline, and what are the four things it can tell you to do with a paragraph?

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (critique). Here is a made-up discussion paragraph from a thesis on a district survey of women aged 15-49. The district's figures are made up; the national figure is from NFHS-5.

> 1. *The district's share of women with a BMI of 25.0 kg/m^2 or more was higher than the national figure. (2) Our sample came from two blocks only. (3) Urban wards had the highest share, as in NFHS-5. (4) Weight and height were measured by trained staff, which is a strength. (5) The national figure in NFHS-5 was 24.0 per cent. (6) Screening in urban wards should come first. (7) Our response rate was 82 per cent.*

Label each sentence with its topic, count the changes of topic, write the reverse outline you would want, and rewrite the paragraph.

EXERCISE 3 (critique). Take the section of your own manuscript or thesis that you chose for this book's build. Write its reverse outline: one sentence per paragraph, in the margin or on a separate sheet. Read the sentences alone. Mark every paragraph to split, merge, move or cut, and rewrite the two worst.

EXERCISE 4 (teaching). A first-year resident says: "I read my chapter three times and it reads fine. Why do I need a reverse outline?" You have ten minutes and a whiteboard. Explain.

11 · Measuring a draft: words per sentence, a readability score, and what they cannot show

Checked against the sources as they stood on 2026-10-02.

DEFINITION. A readability formula estimates how hard a passage is to read from three counts: words, sentences and syllables. A syllable is one beat of a spoken word: “row” has one, “maintain” two, “dictionary” four.

The counts give two averages. The average sentence length (ASL) is the number of words divided by the number of sentences. The average syllables per word (ASW) is the number of syllables divided by the number of words.

The Flesch Reading Ease score (FRE), published by Rudolf Flesch in 1948, is 206.835 minus 1.015 times ASL minus 84.6 times ASW. Flesch wrote the second term as 0.846 times the syllables per 100 words, which gives the same score. A higher score means easier text. For almost all ordinary prose the score falls between 0 and 100, but the formula can go below 0 or above 100.

Every worked score below uses ASL, ASW and FRE, so this section keeps all three. The limit on abbreviations in section 8 counts a paper’s nonstandard ones.

Kincaid and colleagues (1975) gave a grade-level formula: 0.39 times ASL, plus 11.8 times ASW, minus 15.59. Its result is a United States school grade, not an Indian class. Later writers call the formula the Flesch–Kincaid grade level.

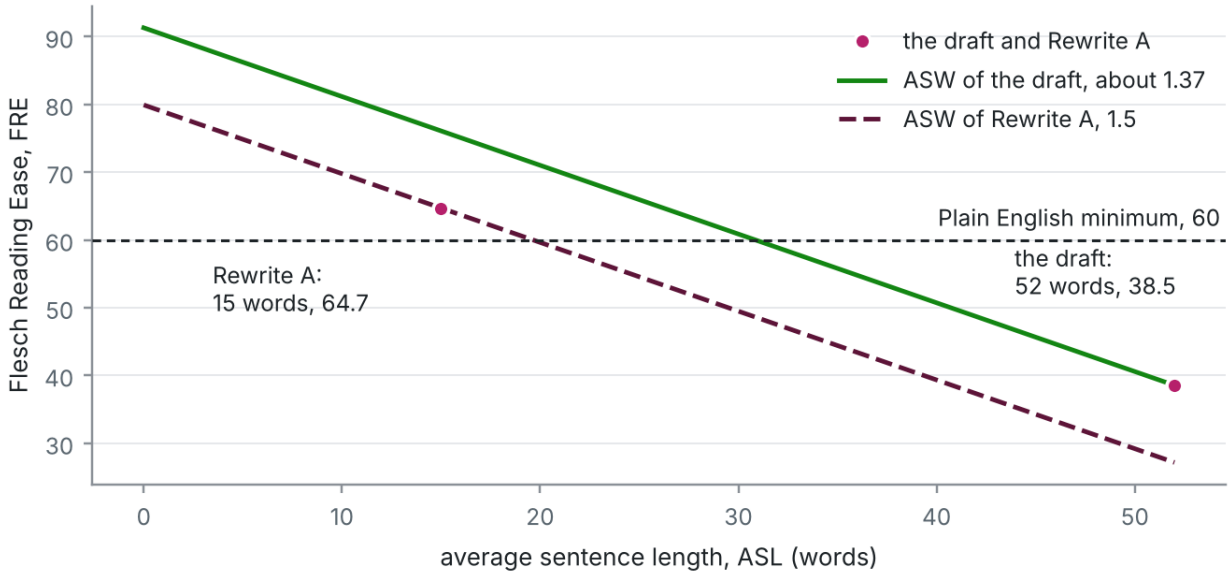
Both formulas count lengths. Neither reads meaning, order or the logic between sentences.

Measured this way, scientific abstracts have grown harder to read. Plavén-Sigray and colleagues (2017) scored 709,577 abstracts from 123 journals, published from 1881 to 2015. [178, 179, 180, 180, 181, 180, 180, 180, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 180, 180, 180, 195, 196, 180, 180, 197, 198, 199, 200, 201, 202, 203]

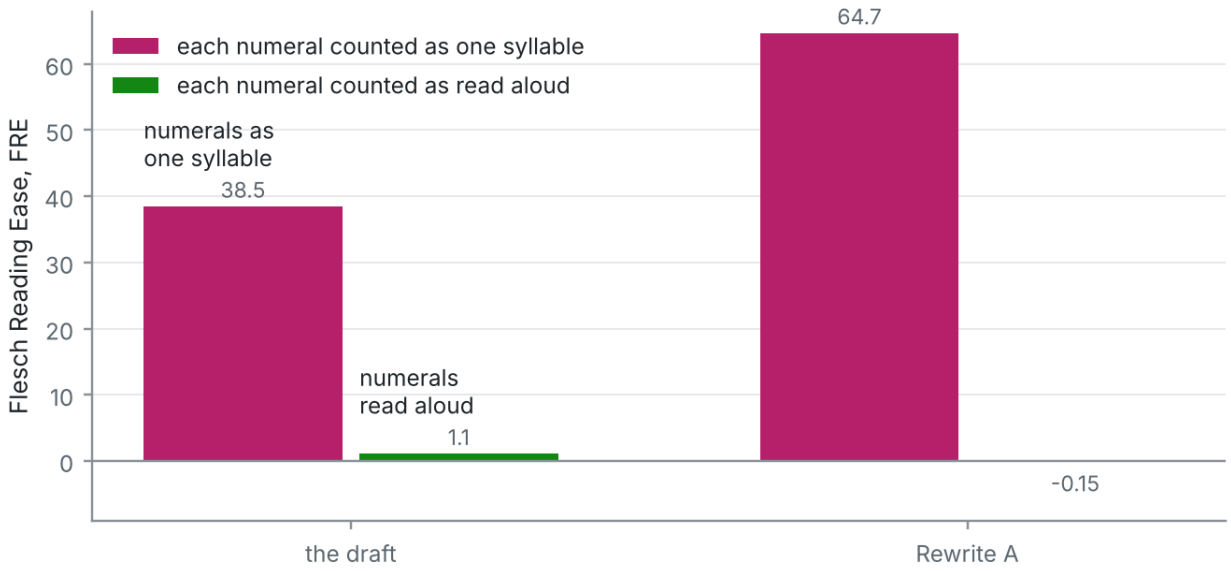
IN PLAIN TERMS. A readability formula reads your draft with a ruler, not with a mind. Long sentences and long words push the score towards hard.

Use it as a trip-wire.

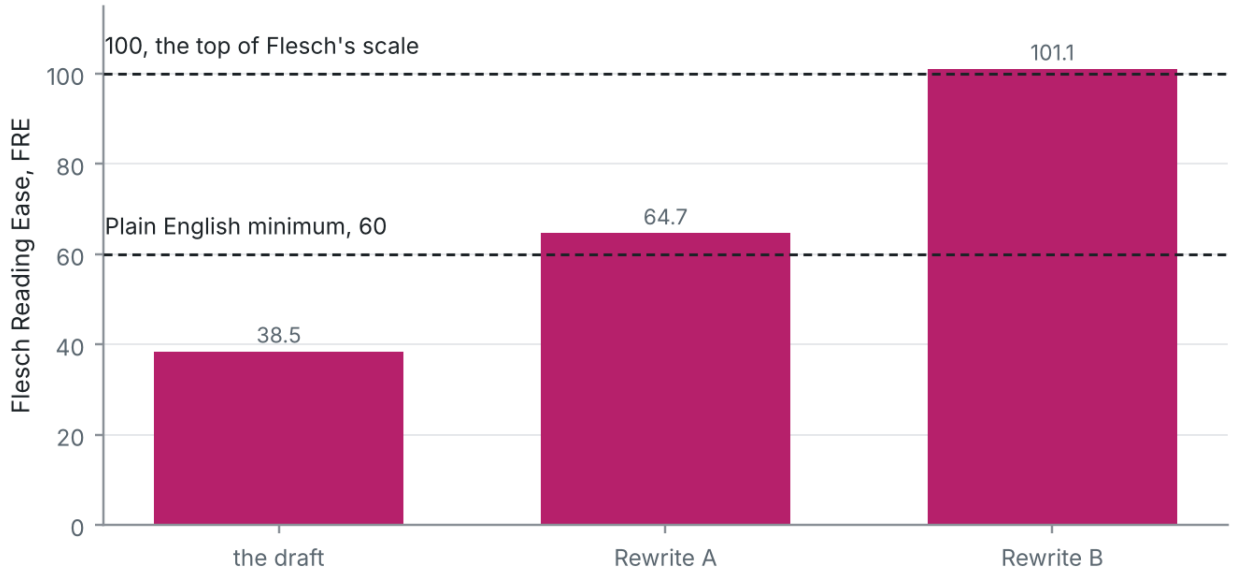
Never use it as a target. A row of four-word fragments scores as very easy, and it can still leave out the one thing the reader needed.



The Flesch Reading Ease score falls by 1.015 points for each extra word in the average sentence. Each line holds one version's average syllables per word: about 1.37 for the 52-word draft of illustration 2, below, and 1.5 for its Rewrite A. The draft sits at 38.5; Rewrite A at 64.7, above Flesch's Plain English minimum of 60. The whole gain comes from shorter sentences: Rewrite A's words are a little longer, so its line sits lower.



The draft and Rewrite A of illustration 2, below, scored under Flesch's two rules for numerals (illustration 3). Counting each numeral as one syllable, the draft scores 38.5 and Rewrite A 64.7. Reading each aloud, they score 1.1 and -0.15. The rule for numerals moved each score more than the rewrite did.



The three versions of illustration 2, below, scored by the same rule. Rewrite B scores 101.1, above the 100 at the top of Flesch's scale, and it is the worst of the three: it splits one group in two and drops which survey each number came from. Rewrite A, at 64.7, keeps everything the draft said. The score counts lengths; only reading tells you which to keep.

Illustration 1. Start with a sentence short enough to count on your fingers. In *How to Write Plain English* (1979), Rudolf Flesch scores three ways of saying one thing. A copy of his chapter is at https://pages.stern.nyu.edu/%7Ewstarbuc/Writing/Flesch.htm. Search for profound.

> *Now let's say the same thing with a little more sophistication - "John has a profound affection for Mary." This adds some complexity to the idea. It rates "Plain English"-a score of 67*

Score it yourself. First count. The sentence has 7 words. Now count the beats in each word, saying it aloud.

word	John	has	a	profound	affection	for	Mary
syllables	1	1	1	2	3	1	2

Flesch gives a quicker way to count. Every word has at least one syllable, so start from the number of words, 7. Then add the extra beats in the longer words: 1 for profound, 2 for affection, 1 for Mary.

```
7 plus 1 plus 2 plus 1 = 11
```

One sentence, 7 words, 11 syllables. Now the two averages. Each is one count divided by another (Book 0, A5), and the average sentence length is a mean (Book 0, D5).

```
7 divided by 1 = 7
11 divided by 7 = 1.571428571
```

So ASL is 7 and ASW, rounded to four decimal places, is 1.5714. Keep four places from here. Now the formula, in Flesch's own order. Multiply ASL by 1.015. Multiply ASW by 84.6. Add the two. Subtract the sum from 206.835.

```
1.015 times 7 = 7.105
84.6 times 1.5714 = 132.94
7.105 plus 132.94 = 140.045
206.835 minus 140.045 = 66.79
```

66.79 rounds to 67, Flesch's score. Now the grade level, for the same counts. Multiply ASL by 0.39 and ASW by 11.8, add them, and take away 15.59.

```
0.39 times 7 = 2.73
11.8 times 1.5714 = 18.54
2.73 plus 18.54 minus 15.59 = 5.68
```

A grade level of about 5.7. Now Flesch's first version, "John loves Mary.": 3 words, 4 syllables, one sentence.

```
4 divided by 3 = 1.333333333
1.015 times 3 = 3.045
84.6 times 1.3333 = 112.80
206.835 minus 3.045 minus 112.80 = 90.99
```

Flesch prints 92 for it. The formula gives 91.0. Printed scores and your own can differ by a point, so say how you counted, and trust the arithmetic you can show.

The short version scores 24 points easier. It also says something different: loving someone is not the same as having a profound affection for them. The formula cannot see that.

Figures quoted in this illustration are from [204].

WHERE THIS PICTURE BREAKS. One sentence is far too small a sample. Flesch's own instructions take samples of 100 words, three to five of them from an article. Score whole paragraphs before you act on a number.

The syllable count is a judgement, not a measurement. If you say "family" in two beats where someone else says three, your total moves. In a 7-word sentence one syllable moves the score by about 12 points. In a paragraph of 100 words it moves it by less than 1.

Illustration 2. Now a sentence of the kind a thesis is full of. It is invented, but you have read many like it. It reports the National Family Health Survey (NFHS), whose fourth round, NFHS-4, ran in 2015-16 and whose fifth, NFHS-5, in 2019-21.

> *It was observed in the present study, which was based on an analysis of the data of the National Family Health Survey, that the proportion of women in the age group of 15 to 49 years who were overweight or obese showed an increase from 20.6% in NFHS-4 to 24.0% in NFHS-5.*

Count it. Take each numeral, abbreviation or symbol (such as 20.6%, NFHS or NFHS-4) as one word of one syllable. That is Flesch's own rule from 1979. Where a word can be said two ways, this book counts the fuller one (Flesch in 1979 took the shorter); state your choice. There are 52 words. The words with more than one syllable add 19 extra beats: observed 1, present 1, study 1, analysis 3, data 1, National 2, Family 2, Survey 1, proportion 2, women 1, overweight 2, obese 1, increase 1.

```
52 plus 19 = 71
71 divided by 52 = 1.365384615
1.015 times 52 = 52.78
84.6 times 1.3654 = 115.51
206.835 minus 52.78 minus 115.51 = 38.545
0.39 times 52 = 20.28
11.8 times 1.3654 = 16.11
20.28 plus 16.11 minus 15.59 = 20.80
```

An FRE of 38.5 and a grade of 20.8. Its words are short. The one long sentence does the damage.

Here are two rewrites. Score both before you read on.

> *Rewrite A. We analysed data from the National Family Health Survey. Among women aged 15 to 49, the share with overweight or obesity rose from 20.6% in NFHS-4 to 24.0% in NFHS-5.*

> *Rewrite B. We used NFHS. Women were measured. They were 15 to 49. Some were overweight. Some were obese. It was 20.6%. It is now 24.0%.*

Rewrite A has 30 words in 2 sentences, with 15 extra beats: analysed 2, data 1, National 2, Family 2, Survey 1, Among 1, women 1, overweight 2, obesity 3. Rewrite B has 24 words in 7 sentences, with 5 extra beats: women 1, measured 1, overweight 2, obese 1.

```
30 plus 15 = 45
30 divided by 2 = 15
45 divided by 30 = 1.5
206.835 minus 1.015 times 15 minus 84.6 times 1.5 = 64.71
24 plus 5 = 29
24 divided by 7 = 3.428571429
29 divided by 24 = 1.208333333
206.835 minus 1.015 times 3.4286 minus 84.6 times 1.2083 = 101.13
```

version	words	sentences	syllables	ASL	ASW	FRE
the draft	52	1	71	52	1.3654	38.5
Rewrite A	30	2	45	15	1.5	64.7
Rewrite B	24	7	29	3.4286	1.2083	101.1

Rewrite B scores best, above even the 100 at the top of Flesch's scale. It is also the worst of the three. It splits "overweight or obese", one group, into two, as if each had its own number. It drops

which survey each number came from. It says "now" of a survey that ended in 2021. The score rose because the sentences got shorter, and it would rise just as far whatever they said.

Rewrite A keeps everything the draft said and scores 64.7, above Flesch's minimum of 60 for Plain English. It is the one to keep, and the score is not what tells you so. Reading it is.

Figures quoted in this illustration are from [24, 204].

WHERE THIS PICTURE BREAKS. These are single sentences and two short paragraphs, chosen to show the arithmetic. A score for a whole results section, a few hundred words, is steadier. The ranking of A and B would not change: a formula that counts lengths cannot tell a clear sentence from a short broken one.

The counts here take each numeral as one syllable. The next illustration counts the same sentences the other way.

Illustration 3. Here is the trap with counting numerals. Flesch himself gave two rules. In 1948 he counted a numeral or symbol the way it is read aloud: "two for $ ("dollars") and four for 1918". So did Kincaid and colleagues. In 1979 he counted every numeral, symbol and abbreviation as one syllable. Plavén-Sigray and colleagues, scoring 709,577 abstracts by computer, removed the numbers altogether.

Count the draft and Rewrite A again, reading each numeral and abbreviation aloud. "15" is fifteen, 2 beats. "49" is forty-nine, 3. "20.6%" is twenty point six per cent, 6. "24.0%" is twenty-four point zero per cent, 8. Each "NFHS" with its number is four letters and a numeral, said one by one: 5. Each of these words had 1 beat before.

```
1 plus 2 plus 5 plus 7 plus 4 plus 4 = 23
71 plus 23 = 94
94 divided by 52 = 1.807692308
206.835 minus 1.015 times 52 minus 84.6 times 1.8077 = 1.12
45 plus 23 = 68
68 divided by 30 = 2.266666667
206.835 minus 1.015 times 15 minus 84.6 times 2.2667 = -0.15
```

The same draft scores 38.5 or 1.1. The same rewrite scores 64.7 or -0.15. Nothing in either sentence changed. Only the counting rule did.

So a score means nothing until you know how it was counted. A results paragraph is thick with numerals, and the rule for numerals can move its score more than any rewrite. When you compare two drafts, score both with the same program or the same rule. When you read someone else's score, look for how they counted numerals before you compare it with yours.

Figures quoted in this illustration are from [205].

WHERE THIS PICTURE BREAKS. The gap is this large because these sentences are thick with numerals: 6 of the draft's 52 words. In a paragraph of discussion with no numerals, symbols or abbreviations in it, the two rules give the same score. Most word processors can score readability; one 2022 study used Microsoft Word for both scores. Yours may follow yet another rule, and may

not say which. So test it: score one sentence with numerals by hand, and see which rule its answer matches.

MUST KNOW POINTS FOR YOU.

- "The score went up, so the text is clearer" is the error. The formula counts lengths and nothing else.
- Before you compare two scores, find out how each counted numerals. Flesch in 1948 read them aloud, Flesch in 1979 counted each as one syllable, and one large study dropped them.
- Use the score as a trip-wire on your own draft, never as a target. Where one scores far harder than the rest, read its longest sentences and rewrite them for the reader.
- A grade level is carried over from the people the formulas were fitted on: Flesch's schoolchildren and Kincaid's 531 Navy personnel. Flesch wrote that his score tracks a school grade only up to about seventh grade.
- These formulas were fitted on schoolchildren's test passages and Navy training manuals, so they say nothing about a Hindi or Tamil translation. For a health leaflet in any language, give it to a few people from the audience and ask them what it tells them to do.
- Mind the unit of the syllable term. Flesch's 1948 article multiplies syllables per 100 words by 0.846; his 1979 book multiplies syllables per word by 84.6. Mix the two and the score comes out in the thousands below zero.
- Quote the rise in abstracts scoring below 0 on the FRE (14 per cent in 1960, 22 per cent in 2015, in one study of 709,577) as a rise in long words and long sentences. It does not show that specialists understand the newer abstracts less well.

COMMON MISREADING. "A Flesch–Kincaid grade of 12 means a reader needs twelve years of school to understand this." That is not what the number is. It is what a formula predicts from two lengths.

Kincaid's numbers were fitted to United States Navy personnel reading Navy training manuals, in a 1975 report. Flesch's own numbers were fitted to schoolchildren answering questions on short reading passages. He wrote that children's grades were not the best yardstick for how readable a text is in general. Neither formula was tested on your readers. A grade is a guess about difficulty from the outside of the words, not a measured need for schooling.

REPORTING IT. "We scored each leaflet with the Flesch Reading Ease score [ref] and the Flesch–Kincaid grade level [ref], using [program and version]. The program counts each numer-

al as [one syllable / the syllables it has when read aloud]. Both scores estimate difficulty from the length of words and sentences alone. We did not test our readers' understanding."

EXERCISE 1 (retrieval). Without looking back, write down the two averages the Flesch Reading Ease formula uses. Say how you get each from the counts, and which way the score moves when each one grows. Then say what the formula cannot see.

EXERCISE 2 (critique). A colleague's paper on patient leaflets at a district hospital reports: "Mean Flesch–Kincaid grade 11.2. The leaflets are therefore suitable only for patients who have completed class 11." Find what is wrong with the conclusion, and say what the paper would need to show it.

EXERCISE 3 (teaching). A journalist rings you: "Is it true scientific papers are getting harder to read?" You have three minutes. Give one quotable sentence, and say how sure the evidence lets you be.

PRACTICE. Eleven problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. A paragraph has 120 words, 6 sentences and 180 syllables. Find its average sentence length and its average syllables per word.

2. A passage has an average sentence length of 20 words and an average of 1.5 syllables per word. Find its Flesch Reading Ease score and its Flesch–Kincaid grade level, each to one decimal place.

3. A passage averages 1.6 syllables per word. What average sentence length, to the nearest whole word, gives it a Flesch Reading Ease score of 50?

4. In 1948 Flesch scored an issue of the New Yorker magazine from October 1946. His counts were an average sentence length of 20 words and 148 syllables per 100 words. He printed a reading ease of 61. Recompute the score from his counts, to one decimal place. Then find the Flesch–Kincaid grade level for the same counts.

5. Flesch (1979) gives this sentence a score of 32.

> *Even though John is not normally given to a display of his deeper emotions, he allegedly has developed a profound affection for Mary, as compared to the more equable feelings he seems to have for Lucy, Fran and, to a lesser extent, Sue.*

Count its words and syllables, and compute its Flesch Reading Ease score to one decimal place. Then work out how far one syllable more or less in your count moves the score.

6. Suppose the 43-word sentence of the last problem could be split into three sentences. No word is added or removed, so it keeps every syllable you counted in the last problem. Compute its new Flesch Reading Ease score, and how much it changed.

7. Score this results sentence twice: once taking each numeral as one syllable, and once counting each numeral as it is read aloud. Give both Flesch Reading Ease scores to one decimal place.

> *Among men aged 15 to 49, 22.9% had overweight or obesity in NFHS-5, against 18.9% in NFHS-4.*

8. Here is a worked answer. Find the step that broke.

```
a paragraph: 200 words, 10 sentences, 300 syllables
ASL = 200 divided by 10 = 20
syllables per 100 words = 300 divided by 200 times 100 = 150
FRE = 206.835 minus 1.015 times 20 minus 84.6 times 150 = -12503.465
so the paragraph is unreadable
```

9. Here is a student's worked score of a two-sentence paragraph. Find the step that broke.

```
the paragraph has 60 words and 90 syllables
it has 2 full stops and 3 more clauses joined by "and", so 5 sentences
ASL = 60 divided by 5 = 12; ASW = 90 divided by 60 = 1.5
FRE = 206.835 minus 12.18 minus 126.9 = 67.755
a score above 60, so it is Plain English
```

10. In *How to Write Plain English* (1979), Flesch writes this.

> *Conversational English for consumers should score at least 80, or about 15 words per sentence and 1 1/2 syllables per word.*

A colleague writes a patient leaflet to exactly those two averages and says it is now conversational English. Decide what to compute, compute it, and say what your answer does not establish.

11. A newspaper column says: "A study of 709,577 scientific abstracts found that more than a fifth of them now score below zero for readability. So a fifth of science cannot be understood by anyone." Decide what to compute, compute it, and say what your answer does not establish.

12 · Cutting: deleting the words, sentences and paragraphs that do no work

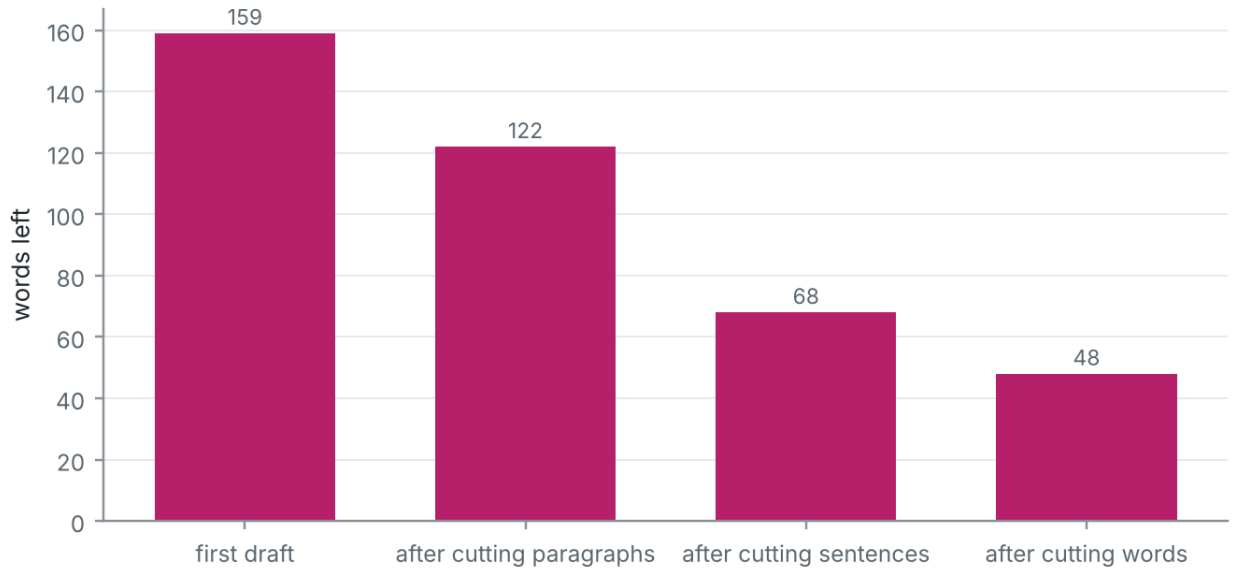
DEFINITION. Cutting is deleting words, sentences or paragraphs that do no work for the reader, so that what remains says the same thing in fewer words. A unit does no work when deleting it loses nothing the reader needs.

The test for a word is whether the sentence means the same without it.

Cutting is not compressing. Compressing keeps everything and packs it into fewer, longer sentences. [206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216]

IN PLAIN TERMS. First ask of each paragraph whether it serves your one message. Then ask it of each sentence.

Big cuts first, because polishing a sentence you later delete is work thrown away. In a padded draft, most of what you lose goes in whole sentences and paragraphs.



The draft of illustration 1, below, cut in three passes. It falls from 159 words to 122 when one paragraph goes, to 68 when three sentences go, and to 48 after the word-level cuts. Whole paragraphs and sentences removed 91 of the 111 words; single words removed 20.

Illustration 1. Here is a draft of two paragraphs from a thesis discussion. It is invented, but you will recognise it. The section's one message, written first as section 3 asks, is this: among women

aged 15 to 49 in India, overweight or obesity was commoner in the fifth National Family Health Survey (NFHS-5) than in the fourth (NFHS-4), and commoner in urban than in rural areas.

> 1. *Many definitions of obesity have been used in the literature over the years, and each of them has its own advantages and disadvantages. A detailed discussion of these definitions is beyond the scope of the present chapter.*
> 2. *Obesity is a major public health problem all over the world, and India is no exception to this. In the present study, it was observed that the proportion of women aged 15 to 49 years with overweight or obesity increased from 20.6% in NFHS-4 to 24.0% in NFHS-5. Thus, it can be seen that there has been an increase in the proportion of women with overweight or obesity. It is important to note that at this point in time the proportion was higher in urban areas (33.2%) than in rural areas (19.7%). The survey also collected a large amount of data on a number of other important topics. Further studies are needed in order to find out the reasons for this increase.*

Count the words: 159. Now cut in three passes, biggest first.

Pass 1, paragraphs. Write the one point of each paragraph in the margin, the reverse outline of section 10. Paragraph 1: "definitions of obesity vary". That serves no part of the message. The paragraph even says so: its subject is "beyond the scope". Delete all 37 words.

```
159 minus 37 = 122
```

Pass 2, sentences. Hold each sentence of paragraph 2 against the message.

- "Obesity is a major public health problem all over the world" is true of every thesis on obesity. It says nothing about these women. Delete: 18 words.
- "Thus, it can be seen that there has been an increase" says the sentence before it again. Delete: 20 words.
- "The survey also collected a large amount of data" serves no claim. Delete: 16 words. It contains "a number of", a phrase the Federal Plain Language Guidelines tell you to shorten to "several". Had you shortened it first, the work would now be thrown away.

```
18 plus 20 plus 16 = 54
122 minus 54 = 68
```

Pass 3, words. Three sentences are left. Go through them word by word.

- "In the present study, it was observed that" delays the subject by 8 words. Delete them.
- "It is important to note that at this point in time" is 11 words before the subject. The guidelines would turn "at this point in time" into "now". But "now" is wrong: NFHS-5 ended in
  1. The phrase was hiding a missing fact, which survey. Replace the 11 words with "In NFHS-5," which is 2.

• "in order to find out" becomes "to find": 3 words go.

> *The proportion of women aged 15 to 49 years with overweight or obesity increased from 20.6% in NFHS-4 to 24.0% in NFHS-5. In NFHS-5, the proportion was higher in urban areas (33.2%) than in rural areas (19.7%). Further studies are needed to find the reasons for this increase.*

```
8 plus 11 minus 2 plus 3 = 20
68 minus 20 = 48
159 minus 48 = 111
```

stage	words left	words removed in this pass
first draft	159	0
after cutting paragraphs	122	37
after cutting sentences	68	54
after cutting words	48	20

From 159 words to 48: 111 gone. Whole paragraphs and sentences took 91 of them, and single words only 20. Every fact the message needs is still there, and one fact is there that was missing before.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. The split of 91 words to 20 belongs to this draft, which was padded on purpose to show each pass. A tight draft may lose little at the first two passes, and that is a result, not a failure.

The last sentence, "Further studies are needed", survived because it states what the data leave open. Some guides would cut it too. Whether a sentence serves the message is a judgement you make against your own message, not a rule this example settles.

Illustration 2. There are two ways to cut that look like the work above and are not.

The first is compressing. A writer who wants fewer words fuses the three sentences into one.

> *Overweight or obesity in women aged 15 to 49 rose from 20.6% (NFHS-4) to 24.0% (NFHS-5), higher in urban (33.2%) than rural (19.7%) areas, reasons needing further study.*

That is 28 words against 48. It is also one sentence carrying three ideas, and the reader must hold all of them to the full stop. The Federal Plain Language Guidelines name this temptation: "Resist the temptation to put everything in one sentence". Measure it as in section 11. The average sentence length goes from 48 words over 3 sentences to 28 words in one.

```
48 divided by 3 = 16
28 minus 16 = 12
1.015 times 12 = 12.18
```

With the same syllables per word, the Flesch Reading Ease score falls by about 12 points. The word count went down and the sentence got harder.

The second is cutting a word that does work. A writer in a hurry trims the first sentence to "The proportion of women with overweight increased from 20.6% to 24.0%." Two cuts changed the claim. Without "aged 15 to 49", the sentence speaks for all women, and the survey measured only women of 15 to 49. Without "or obesity", it speaks of overweight alone, and the survey's figure covers overweight and obesity together. Without the survey names, the reader cannot tell which figure came first. Each cut word looked like padding. Each was a limit on the claim.

WHERE THIS PICTURE BREAKS. Not every long sentence is compressed. A sentence with one idea and a long name in it, such as a survey's full title, is long and still fine. The fault is several ideas in one sentence, not length alone.

The words that limit a claim differ from paper to paper: an age range, a place, a year, a denominator, "or obesity". Read each candidate word and ask whether the claim gets wider without it. If it does, keep it.

MUST KNOW POINTS FOR YOU.

- Cut in order: paragraphs first, then sentences, then words.
- Test each paragraph against your one message with a reverse outline. Where a margin note serves no part of the message, delete that paragraph.
- "Fewer words is better, so join short sentences into one" is the error. Cut whole units instead, and keep one idea a sentence.
- Cutting has no test for truth. Before you delete any word, ask whether the claim gets wider without it. If it does, the word stays.
- This book's advice: before you cut, open a second file and paste every deleted passage into it.
- Delete the opener that delays the subject: "It was observed that", "It is important to note that", "In the present study".
- When you correct a trainee's thesis chapter, mark the paragraphs and sentences to delete before you touch a single word.

COMMON MISREADING. "Cutting means making every sentence shorter." In the draft above, deleting one paragraph and three sentences removed 91 words, and word-level cuts removed 20. A

writer who only trims words, sentence by sentence, polishes text that should not be there. And fusing sentences to save words makes the page harder, not easier.

EXERCISE 1 (retrieval). From memory: in what order do you cut, and what question does each unit have to fail before you delete it? Then say how cutting differs from compressing.

EXERCISE 2 (critique). Cut this methods paragraph from a district survey by deleting words only. It is invented. You may delete a whole phrase or clause, and change a capital or a comma; nothing else. Say what you cut and why, and count the words before and after.

> *The present study was conducted in the Department of Community Medicine of a tertiary care centre in central India. A cross-sectional study design was used for the purpose of this study. The study was carried out over a period of six months, from January to June 2025. The study participants were adults aged 18 years and above who were residing in the field practice area. Each and every participant was weighed with a digital weighing scale, and the weight was recorded to the nearest 0.1 kg.*

EXERCISE 3 (design). Take the manuscript or thesis section you chose for this book's build. Write its one message at the top. Make a reverse outline. Then cut it in three passes, paragraphs, sentences, words, pasting every deletion into a second file. Record the word count after each pass.

13 · Showing a draft to a reader: the one-pass test

DEFINITION. The one-pass test checks whether a draft carries its argument to a reader. It has five steps.

1. Before the test, write your argument in one or two sentences, your one message from section 3, and put it away where the reader cannot see it.
2. Give the draft to a reader who does not know the work. They read it once, at their normal pace, without going back, and mark each place where they stopped or read a passage again.
3. Without looking at the draft, the reader writes in one or two sentences what it argues.
4. Compare the reader's sentences with yours, part by part: whom or what the claim is about, what was found, compared with what, and how far it reaches. How far it reaches means the claim's limits: the ages, place and years it covers, and what it does not claim, such as a cause.
5. Revise the paragraphs that the comparison and the marks point to, and test again with a new reader.

The draft passes when the reader's sentences carry every part of your message and add no part of their own. A part is kept when the reader's words say the same thing at the same strength, no narrower, wider or stronger. The wording need not match. [217, 218, 219, 220, 221, 38, 222, 223, 207, 42, 224]

IN PLAIN TERMS. You cannot tell whether your draft is clear by reading it yourself. You know what it means, so you read that meaning into it.

ILLUSTRATION. First, write your one message and seal it in an envelope, or in a file the reader will not see.

> *Among women aged 15 to 49 in India, overweight or obesity was commoner in the fifth National Family Health Survey (NFHS-5) than in the fourth (NFHS-4), and commoner in urban than in rural areas. These survey data cannot say why.*

Next, find a reader outside the work: a resident in Physiology, say, who knows what a survey is and nothing about yours. The draft, the readers and their sentences here are made up.

She hands back one mark, beside paragraph 3, and this sentence.

> *Obesity is increasing among Indian women because of urbanisation.*

Now open the envelope and compare, part by part.

part	your message	her sentence	verdict
about whom	women aged 15 to 49 in India	Indian women	widened: the age range is lost

part	your message	her sentence	verdict
what was found	overweight or obesity, commoner in NFHS-5 than in NFHS-4	obesity is increasing	changed: overweight dropped, and two surveys turned into a trend
compared with what	urban higher than rural	because of urbanisation	added: a difference turned into a cause
how far it reaches	the data cannot say why	nothing	lost, and contradicted

The added cause is the worst failure. Find where it came from. Her mark sits beside paragraph 3. You open it: the urban and rural figures come straight after two sentences on diet and work changing in Indian cities. The order invites the reader to join them.

Revise only what the test pointed to. Move the urban and rural figures into the paragraph that reports the findings.

Then test again, with a new reader. A resident in Microbiology reads it once, makes no mark, and writes this.

> *Among women aged 15 to 49 in India, overweight or obesity went up between the two surveys and is higher in towns. The survey can't tell us why.*

Every part is there and nothing is added. The draft passes.

WHERE THIS PICTURE BREAKS. One reader is one reader.

The test shows that the argument gets across. It does not show that the argument is right.

MUST KNOW POINTS FOR YOU.

- Never count your guide's or co-author's approval as the test. They know the work, so they read your meaning into the draft, as you do.
- Write your own one-message sentence and put it away before the reader starts.
- "Did you like it?" and "Any comments?" are the wrong questions. Ask for the one sentence and the marks.
- Treat a part the reader added as worse than a part the reader lost. A reader repeats an added cause as the paper's claim; a lost part is only missing. Find the paragraph whose order or wording invited it, and change that paragraph.
- Use a new reader for each round.
- A pass shows that one reader, minutes after reading, could state your argument.

COMMON MISREADING. "My reader got it wrong, so my reader was careless." The test assumes a reader who reads once and does not know the work, because that is how most of your readers will read. Mensh and Kording note that for most readers the abstract is the only part of a paper

they read at all. If one careful reader outside the work misses or changes a part, the draft let them. Revise the draft, not the reader.

EXERCISE 1 (retrieval). From memory, list the five steps of the one-pass test. Then say why the reader must not know your work, and why each round needs a new reader.

EXERCISE 2 (build). This book's build: one section of a manuscript or thesis, rewritten so that each paragraph carries one idea, and shown to a reader. Use the section you have been working on since the start of the book.

1. Write the section's one message at the top of a separate page, and put the page away.
2. Make a reverse outline. Rewrite until each paragraph has one point, stated in its first sentence.
3. Cut it in three passes, paragraphs, sentences, words, keeping the cuts in a second file.
4. Score each paragraph for readability with one program, or by hand with the formula of section 11, and say which. Reread the longest sentences of any paragraph that scores far harder than the rest.
5. Run the one-pass test with a reader outside the work. Fill in the four-part comparison.
6. Revise what the test pointed to, and run it again with a new reader until the draft passes.

Keep the dated versions, the reverse outlines, the readers' sentences and their marks.

EXERCISE 3 (teaching). A first-year resident hands you her thesis introduction and asks, "Can you read it and tell me if it's OK?" You have ten minutes and a whiteboard. Explain what you will do instead of reading it for her, and why.

14 · A figure makes one claim

DEFINITION. A figure in a paper is drawn to make one claim: one statement about the data that a reader should be able to repeat after looking at it. The claim is written as a sentence before the kind of chart is chosen. One figure carries one claim.

A result belongs in a sentence when it rests on a few numbers. It belongs in a table when the reader needs exact values, or will look up one row among many. It belongs in a figure when the claim is a pattern among many values, such as a trend, a gap or a ranking, that words would take too long to carry. The same numbers are not shown in both a table and a figure.

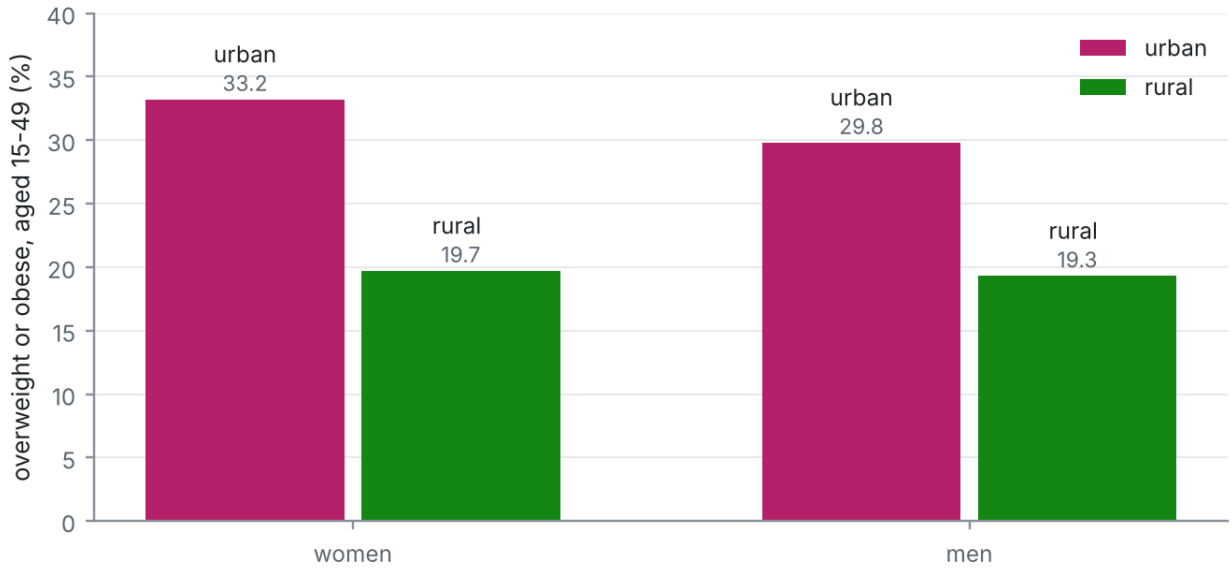
The caption is the text printed with a figure. The International Committee of Medical Journal Editors (ICMJE) calls it the legend. Its first sentence states the finding, not the topic.

A figure on a printed page can be studied for as long as the reader likes. A figure on a slide is seen for a few seconds, from across a room, while someone talks. The two are made separately. A poster is a third medium, and gets its own figure too. [225, 226, 227, 228, 229, 32, 230, 231, 232, 233, 234, 235, 224, 236]

IN PLAIN TERMS. Before you draw anything, write down the one sentence you want a reader to take away from the figure. If you cannot write it, you do not have a figure yet.

Then ask whether the claim needs a picture at all. Four numbers read well in a sentence. Forty do not.

Last, put the sentence where the reader meets it first: the opening line of the caption.



In NFHS-5 (2019-21), a larger share of urban than of rural adults aged 15-49 had overweight or obesity, in women (33.2% against 19.7%) and in men (29.8% against 19.3%). Bars show the per cent with a body mass index of 25.0 kg/m^2 or more. Source: NFHS-5 India Fact Sheet, indicators 88 and 89.

Illustration 1. Start with a figure that fails. A draft results chapter of a thesis has this caption. The draft is made up; the numbers under it are real.

> *Figure 3.2. Overweight and obesity in India, NFHS-4 and NFHS-5.*

Above it are eight bars in one cluster, with a key in the corner saying which colour is which. Find where the eight numbers came from. Open the *National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet*, at https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf. Search for overweight or obese. Two rows come up.

> *1. Women who are overweight or obese (BMI ≥25.0 kg/m2)21 (%) 33.2 19.7 24.0 20.6*

> *1. Men who are overweight or obese (BMI ≥25.0 kg/m2) (%) 29.8 19.3 22.9 18.9*

The number that opens each row is its indicator number. Read the rows as Book 0 taught you to read a table (Book 0, F1). Each row is adults aged 15-49. BMI is the body mass index (Book 0, B6), and the sign ≥ means "greater than or equal to": the rows count a BMI of 25.0 kg/m^2 or more. The four columns are NFHS-5 (2019-21) urban, rural and total, then NFHS-4 (2015-16) total, all in per cent. The 21 after the bracket is a footnote mark.

Now cover the figure and write the one sentence a reader should take away from it. Most people who try write two.

> 1. *Between NFHS-4 and NFHS-5, the share of adults aged 15-49 with overweight or obesity rose, in women and in men.*

> 1. *In NFHS-5, the share was higher in urban than in rural India, in women and in men.*

Each is a claim. The figure tries to carry both, so it carries neither. The reader faces eight bars and no hint of which comparison matters. Choose the claim your chapter argues: the one message you wrote for it (section 3). Say the chapter is about the gap between town and village. Then (b) is this figure's claim.

Claim (a) still has a home. Ask whether it needs a figure at all. It rests on four numbers, and a sentence holds four numbers easily.

> *Between NFHS-4 (2015-16) and NFHS-5 (2019-21), the share of adults aged 15-49 with overweight or obesity rose from 20.6 to 24.0 per cent in women, and from 18.9 to 22.9 per cent in men.*

That is a rise of 3.4 percentage points in women and 4.0 in men. No figure is needed, and the ICMJE Recommendations agree: "Restrict tables and figures to those needed to explain the argument of the paper".

So where does each kind of result go? Ask how many numbers the result rests on, and what the reader has to do with them.

the result rests on	the reader has to	put it in
a few numbers	read them in passing	a sentence
many exact values	look one up, or check it	a table
many values	see a pattern: a trend, a gap, a ranking	a figure

Claim (b) also rests on four numbers, and in a paper it could be a sentence too. It becomes a figure's claim when the values multiply. Suppose your chapter compares urban and rural shares in every state. That is two numbers per state, too many for a sentence. If the reader needs to find their own state's values, use a table. If they need to see that urban sits above rural almost everywhere, use a figure. Never both: the same Recommendations say "do not duplicate data in graphs and tables".

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. Four national numbers are a sentence's job in a paper. The figure in this section draws claim (b) to show what a one-claim figure looks like, not because four numbers need one.

The rows cover adults aged 15-49 only, and count overweight and obesity together. Whatever claim you choose has to stay inside those limits. Neither row says why the shares differ.

Illustration 2. Now give claim (b) its figure and its caption. Start with the caption, because it is where the claim is written down. Here is the draft's caption again, with the one this section builds beside it. The first is made up; the second uses the fact sheet's numbers.

before	after
Figure 3.2. Overweight and obesity in India, NFHS-4 and NFHS-5.	Figure 3.2. In NFHS-5 (2019-21), a larger share of urban than of rural adults aged 15-49 had overweight or obesity, in women and in men. Bars show the per cent with a body mass index of 25.0 kg/m^2 or more. Source: NFHS-5 India Fact Sheet, indicators 88 and 89.

The first names a topic. A reader who looks at it learns what the figure is about, and must work out the point alone. The second states the finding first, then says what the bars are, then where the numbers came from.

Put the finding in the caption, not across the top of the drawing. The ICMJE Recommendations say: "Titles and detailed explanations belong in the legends—not on the illustrations themselves." Legend is ICMJE's word for the caption.

Next, take off the drawing everything the claim does not use. The claim compares urban with rural, so the two totals go, and so do the two NFHS-4 bars. Four bars remain: urban and rural, for women and for men. Claus Wilke, in his book *Fundamentals of Data Visualization* (2019), puts the rule plainly: "We don't need to visualize data dimensions that are tangential to our story". Tangential means off to one side of it.

Then ask the harder question. Would anything you removed make a reader doubt the claim? The totals would not: they sit between urban and rural, as they must. Rural NFHS-4 numbers might. If the rural share had been rising faster than the urban one, a reader would want to know. This fact sheet gives NFHS-4 only as a total, so you cannot check it here. Say so in the text, and find the NFHS-4 report if the chapter's argument depends on it.

Last, the medium. For your thesis, the figure sits on a page, with its caption, and the examiner can study it. For a talk, make it again. Rougier and colleagues say a slide figure needs thicker lines, bigger text and strong contrast, because the audience sees it from a distance for a short time. They also say to stop "extracting a figure from your article to be put, as is, in your oral presentation".

WHERE THIS PICTURE BREAKS. Removing the totals is safe here only because each total must lie between its urban and rural values. A dropped group that could fall outside the others, such as another survey round, is a different matter: check it before you cut it.

The caption's first sentence is a claim about these two rows. It says nothing about any one state, about people over 49, or about why urban shares are higher.

MUST KNOW POINTS FOR YOU.

- Write the figure's claim as one sentence before you open any program that draws charts.
- "A result looks more serious as a figure" is the error. A few numbers belong in a sentence, where the reader takes them in without stopping.
- Never show the same numbers in a table and in a figure.
- Make the first sentence of every caption the finding, not the topic. "Overweight and obesity, NFHS-4 and NFHS-5" names a topic. "A larger share of urban than of rural adults had overweight or obesity" is a finding.
- Do not lift a figure from your paper onto a slide. Make it again, with fewer elements, thicker lines and bigger text.
- The one-claim rule lets you cut data that are off to the side of the claim. It never lets you cut data that would weaken it.
- Wilke recommends three to six figures for a scientific paper, with the rest moved to an appendix or supplement. The journal's instructions to authors set the actual limit; read them first.
- A clear figure persuades whether or not its claim is true.

COMMON MISREADING. "One figure, one claim" is easily read as "one figure, few numbers". It limits what the figure says, not how much data it shows. A figure claiming that urban shares are above rural ones in every state shows two numbers for each state, and it still makes one claim. A figure with only four bars can make three claims at once and fail.

REPORTING IT. Figure 1. In NFHS-5 (2019-21), a larger share of urban than of rural adults aged 15-49 had overweight or obesity, in both women and men. The shares were 33.2 against 19.7 per cent in women, and 29.8 against 19.3 per cent in men.

EXERCISE 1 (retrieval). From memory: when does a result belong in a sentence, when in a table, and when in a figure? Where does a figure's claim go, and what must never be shown twice?

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (critique). A colleague's clinic audit has this caption over a figure of three panels. The audit is made up.

> *Figure 2. Clinic attendance, BMI categories and waist circumference, by age group and sex, April 2024 to March 2025.*

Panel A shows monthly attendance as a line. Panel B shows the share of patients in each BMI category, by age group, as stacked bars. Panel C shows mean waist circumference by sex, as two bars. The text says the audit's main finding is that the share of patients with a BMI of 25.0 kg/m^2 or more rises with age. Say what is wrong and what you would do.

EXERCISE 3 (design). Take the manuscript or thesis section you are rewriting for this book. List every result in it. For each, write the one-sentence claim, then decide: sentence, table or figure. For each figure, write the first sentence of its caption. Check that no numbers appear in both a table and a figure.

EXERCISE 4 (teaching). A first-year resident wants to keep all eight NFHS bars in one figure, "so the reader has everything". Ten minutes, with a whiteboard. Explain why one claim per figure, and what happens to the numbers she takes off.

15 · Choosing the mark: what the eye reads accurately, and the shape of the data

Checked against the sources as they stood on 2026-10-02.

DEFINITION. A mark is anything drawn that stands for one number (Book 0, F2). The reader gets the number back by a judgment the eye makes about the mark. Cleveland and McGill (1984) called these judgments elementary perceptual tasks, and named ten. They are position along a common scale, position along scales that are not lined up, length, direction (such as the slope of a line), angle, area, volume, curvature, shading and colour saturation. A common scale is one axis that every mark being compared is read against.

They proposed an order, from the judgment people make most accurately to the least. First comes position along a common scale, then positions along scales not lined up, such as the same measure in two panels side by side, each with its own axis. Length, direction and angle share third place. Area is fourth. Their rule for design follows: a graph should use judgments as high in this order as it can.

The order is a hypothesis. Their two experiments tested two parts of it, position against length and position against angle. Position won both times.

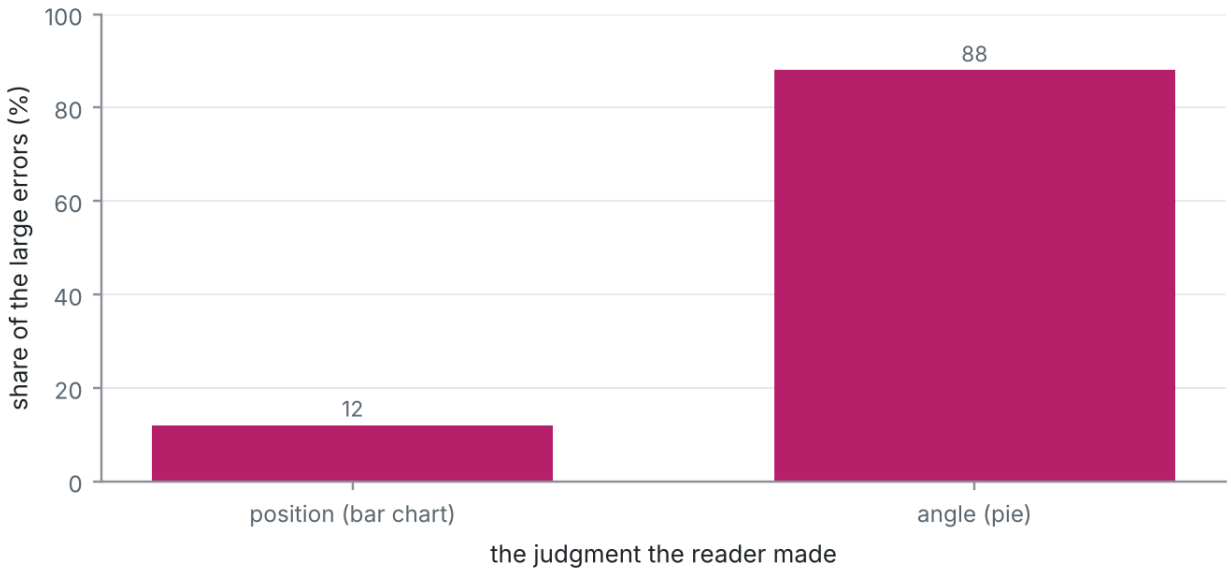
The mark is chosen by the shape of the data as well. Amounts for a set of groups take bars, or dots where the bars would end. Many values from each group take a chart of their distribution. Parts of one whole take side-by-side bars or, for a few parts, a pie. One measured quantity against another takes a scatterplot. Change over time takes a line. [237, 237, 237, 237, 237, 237, 237, 237, 237, 237, 237, 237, 237, 237, 237, 237, 237, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 224, 236, 248, 249, 250, 251, 252, 253]

IN PLAIN TERMS. Your eye reads some marks better than others. So put the numbers a reader must compare on one shared axis, side by side.

Many values from one group take a chart of how they spread, which the next section is about.



Cleveland and McGill (1984), first experiment. Position judgments were 60% of all judgments but drew 22% of the large errors; length judgments were 40% of judgments and drew 78%. Per judgment, a large error was 5.3 times as frequent on length as on position.



Cleveland and McGill (1984), second experiment. Each set of numbers was drawn once as bars and once as a pie, so position and angle judgments were equal in number. The pie drew 88% of the large errors and the bars 12%: per judgment, a large error was about seven times as likely on the pie (88 ÷ 12 is about 7.33; the paper prints 7.3).

Illustration 1. Start with a chart that fails. A draft district report puts four numbers from the last section into a pie, from the fifth National Family Health Survey (NFHS-5): urban women, rural women, urban men, rural men. The draft is made up; the numbers are the fact sheet's.

Before you look at the angles, add up the slices.

```
33.2 plus 19.7 plus 29.8 plus 19.3 = 102.0
```

A pie says its slices are parts of one whole. These four are not. Each is a share of a different group of people: urban women, rural women, and so on. Their sum, 102.0, means nothing. The program drew the pie anyway: it made 102.0 into a full circle, and each slice became a share of that meaningless total.

Now suppose the numbers were parts of one whole. The pie would still ask the reader to compare angles and areas. Wilke gives the reason this goes badly: "human perception primarily judges distances and not areas".

The fix is the figure of the last section (section 14). There the four values stand as bars from one shared start, read against one axis. The reader compares where each bar ends. That is position along a common scale, the judgment people make best.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. A pie is not always wrong. When a few parts make one whole, such as how a clinic's patients split across four body mass index (BMI) categories, a pie shows that the parts add up to the whole. Wilke says that is what pies do well, and that side-by-side bars still compare the parts better.

The sum of 102.0 is a quick test, not a proof. Four shares of different groups could add to 100 by chance and still not be parts of one whole. Ask what group each number is a share of.

Illustration 2. How much worse are the other judgments? Open Cleveland and McGill, "Graphical perception: theory, experimentation, and application to the development of graphical methods", *Journal of the American Statistical Association* 1984, at https://doi.org/10.1080/01621459.1984.10478080.

In their first experiment, people looked at bar charts with two marks on them. They wrote down what per cent the smaller marked value was of the larger. In some charts the two values could be read against one axis: position. In others they could only be compared as lengths, such as two pieces stacked inside a bar. Search for the summary of the experiments.

> *In the first experiment, po- sition judgments were more accurate than length judg- ments by factors varying from 1.4 to 2.5. In the second experiment, position judgments were 1.96 times as ac- curate as angle judgments.*

The broken words are the scanned page's line breaks. The second experiment drew the same numbers as a bar chart and as a pie.

Now look at the worst mistakes, the ones the authors call large errors. Search for Seventy-eight.

> *Seventy-eight percent of the large errors occurred for the length judgments; since there were three position judgments for each two length judgments, the rate of occurrence of large errors for length judgments is 5.3 times that for position judgments.*

Rebuild the 5.3 yourself. It is a ratio of two rates (Book 0, A5). Length judgments made 2 of every 5 judgments, which is 40 per cent, and position judgments 3 of every 5, which is 60 per cent.

Share out the large errors, then divide each share by how many judgments of that kind there were. The sign ≈ means “is approximately equal to”.

```
100 minus 78 = 22
78 divided by 2 = 39
22 divided by 3 ≈ 7.33
39 divided by 7.33 ≈ 5.32
```

The second experiment drew each set of numbers once as bars and once as a pie. So position and angle judgments were equal in number, and the shares can be compared directly.

```
100 minus 88 = 12
88 divided by 12 ≈ 7.33
```

The paper prints 7.3. Per judgment, a large error was about seven times as likely on the pie as on the bars beside it. In only 3 of 40 comparisons was the pie more accurate on average than the bar chart.

Heer and Bostock (2010) ran the first experiment again, online, with 50 people per chart. Position again beat length. They also tested angle and area with the same question. Area did worse than angle, and both did worse than position.

Figures quoted in this illustration are from [237, 254].

WHERE THIS PICTURE BREAKS. Both experiments asked one question: what per cent is this value of that one? A paper’s reader may instead need to see which group is largest, or that the parts make a whole. The experiments do not measure those.

The 5.3 and the 7.3 come from two different experiments with different charts. Do not set them side by side to rank angle below length. Heer and Bostock, testing both the same way, found angle no worse than length.

Cleveland and McGill say so themselves: the order “does not provide a complete prescription for how to make a graph”.

Illustration 3. The order of judgments tells you how to draw. The shape of the data tells you what to draw. Before you choose a chart, say which of five kinds of numbers you have. The table puts chapter 5 of Wilke’s *Fundamentals of Data Visualization* (https://clauswilke.com/dataviz/) in this book’s words, with examples from your own work.

what the numbers are	an example from your work	marks that suit it
amounts, one for each group	per cent with overweight or obesity, urban and rural	bars from zero, or dots on one axis
many values in each group	the BMI of every adult in a district survey	a chart of the distribution, such as a strip chart or box plot
parts of one whole	how a clinic’s patients split across BMI categories	side-by-side bars; a pie only for a few parts
one measured quantity against another	waist circumference against BMI, one point per adult	a scatterplot

change over time	one indicator across survey rounds	a line through the values

Try it on the running case. In the two fact-sheet rows of the last section, the last number is NFHS-4 (2015-16) and the third is NFHS-5 (2019-21). So you have the share of women and of men aged 15-49 with overweight or obesity, in two survey rounds. That is change over time, for two groups.

So draw a line for women and a line for men, each joining its two survey rounds, on one shared axis. Each point is read by its position. The women's line runs from 20.6 to 24.0 per cent and the men's from 18.9 to 22.9. The lines climb by 3.4 and 4.0 percentage points.

The value axis of this chart need not start at zero. It could start at 15 per cent, a little below the lowest point. For points and lines that is allowed, because each value is carried by where the point sits, not by a length from zero (Book 0, F2). Wilke makes the same move for amounts that need a logarithmic axis (Book 0, A7), on which zero lies infinitely far away. He drops the bars and places "a dot at the appropriate location along the scale".

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. Two survey rounds give two points, and a straight line between them says nothing about the years in between. The line shows the change from one round to the next, not a steady rise.

Each NFHS round measures different people. The line joins two population shares, not the same people measured twice.

MUST KNOW POINTS FOR YOU.

- Put the values a reader must compare on one shared axis, next to each other: bars from one shared start, or dots. Avoid making the reader compare pieces stacked inside bars, or slices of a pie. Those are lengths without a shared start, and angles.
- In Cleveland and McGill's experiments, large errors were 5.3 times as frequent per judgment on length as on position.
- "The ranking of marks was proved by experiment" is the error. The full order of ten judgments is Cleveland and McGill's hypothesis.
- Before you draw a pie, check that its numbers are parts of one whole: shares of the same group of people, adding to 100 per cent. Prevalences in different groups are not. Neither are conditions a person can have together, such as stunting and underweight in the same child.
- When a trainee shows you a pie chart or a stacked bar chart, ask for the same numbers as dots or grouped bars on one axis.
- Name the shape of your numbers before you choose the chart.
- When an amount needs a logarithmic axis, draw dots, not bars: zero lies infinitely far down it. A ratio there takes bars from 1 (section 17).
- Use the order as a guide, as its authors did, not as a rule that settles every choice.

COMMON MISREADING. "Position beats length, so bar charts are second-rate." A bar chart whose bars stand on one shared start is read by where each bar ends, against the shared axis: that is position. Cleveland and McGill replace the pie with a bar chart for exactly this reason, "thus replacing angle judgments by position judgments". What does badly is a length with no shared start, such as the upper pieces of a stacked bar.

EXERCISE 1 (retrieval). From memory: list the first four places in Cleveland and McGill's order of judgments, from most accurate down. Which comparisons did their two experiments actually test, and which judgment won?

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (critique). A district nutrition report shows one pie chart of children under 5, titled "Child undernutrition and overweight, India, NFHS-5". Its four slices are stunted (short for age) 35.5%, wasted (thin for height) 19.3%, underweight (light for age) 32.1% and overweight (heavy for height) 3.4%, each judged against the World Health Organization growth standard. The numbers are the NFHS-5 India Fact Sheet's totals for indicators 81, 82, 84 and 85. Say what is wrong and what you would draw instead.

EXERCISE 3 (design). Take the results you listed for figures in your own section. For each, name the shape of the numbers from the table in this section, then choose the mark. Say which judgment the reader will make to compare the values you most want compared.

EXERCISE 4 (teaching). A district health officer wants a pie chart for the monthly review meeting. It is to show the share of adults with overweight or obesity in each of the district's six blocks. One page for her, answer first.

16 · Show the data: when a bar of means hides what the points would show

Checked against the sources as they stood on 2026-10-02.

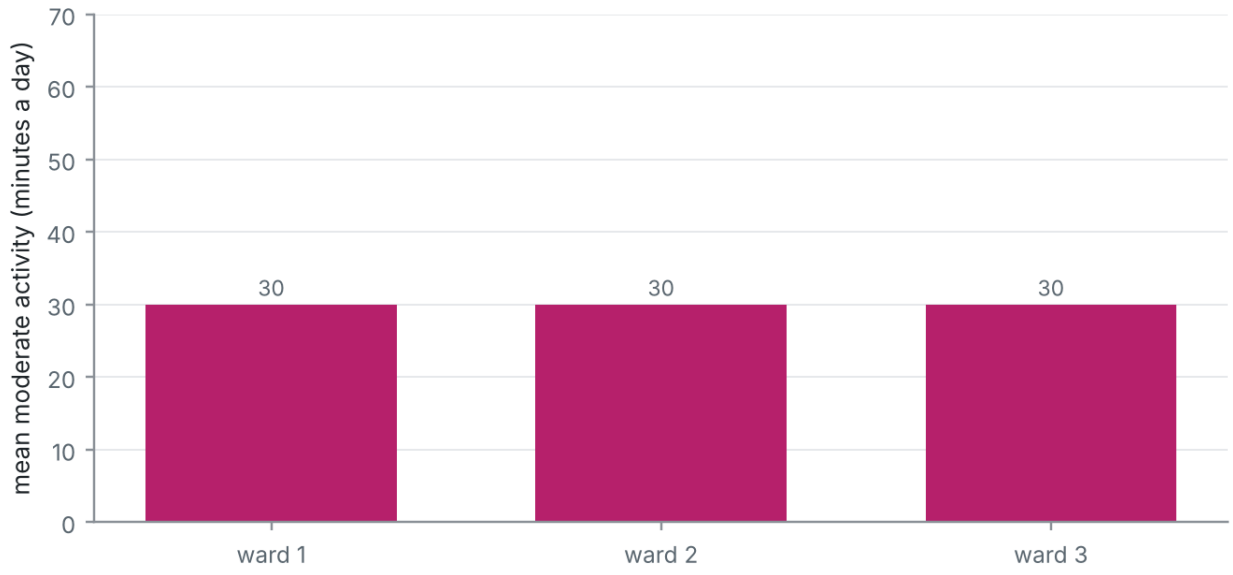
DEFINITION. A bar chart of a measured quantity draws each group as one bar, the height of its mean.

Many different sets of values give the same bar and the same error bar. An error bar is a line above and below the top of the bar, usually one standard deviation or one standard error each way (Book 0, D5 and Book 0, D6). The values may be spread evenly, or bunched with one far from the rest, or split into two clusters. There may be only three of them.

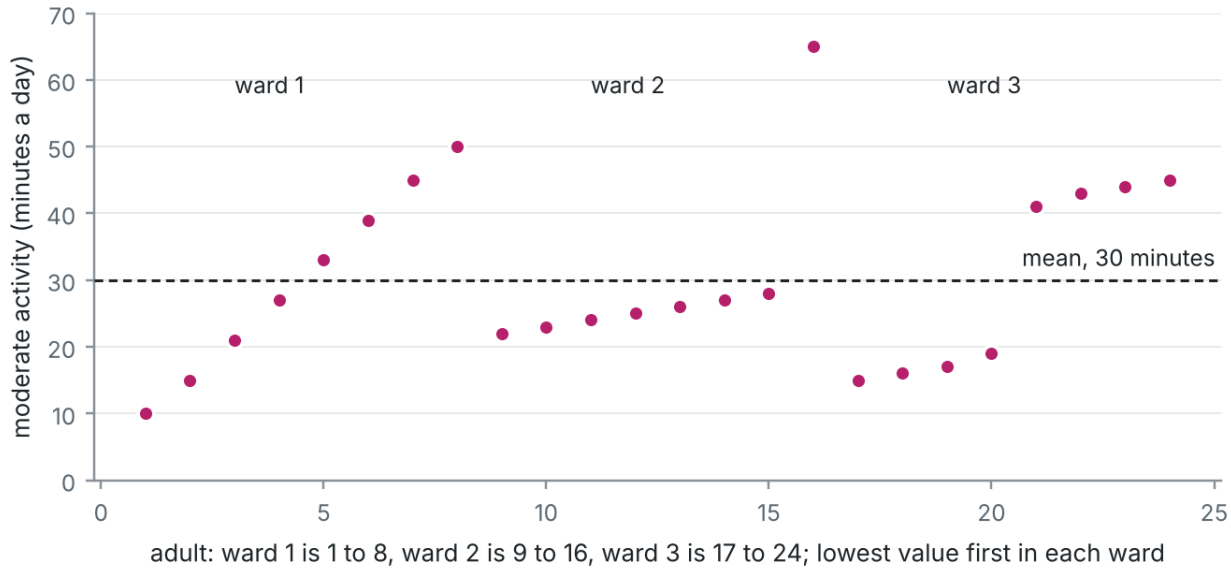
Showing the data means drawing every value. A strip chart puts each value as a dot along the value axis, one column of dots per group; Weissgerber and colleagues call it a univariate scatterplot. [255, 255, 256, 257, 258, 259, 259, 259, 260, 261, 262, 263, 264, 264, 265, 265, 266, 267, 266, 268]

IN PLAIN TERMS. A bar chart of a measured quantity shows one number per group: the average. It hides everything else.

So when your groups are small, draw every value as a dot. The reader sees how many people there were, how they spread, and whether one person is pulling the average.



Three wards look alike as bars of means. Made-up survey of 8 adults in each of three wards; each bar is a ward's mean minutes of moderate activity a day. All three are 30 minutes, and each ward's standard deviation is 14.3 minutes, so the bars and any error bars would match.



The same three means of 30 minutes hide three different wards. One dot per adult, 24 made-up values, set out ward by ward; within each ward the eight adults go from left to right, lowest value first, so no dot hides another. The dots are not joined, because the adults are not pairs. Ward 1 is spread fairly evenly from 10 to 50 minutes. In ward 2, seven adults report 22 to 28 and one reports 65. Ward 3 falls into two clusters, 15 to 19 and 41 to 45. The dashed line is the mean of every ward, 30 minutes.

Illustration 1. Start with a figure that fails. A draft survey report shows three bars of equal height: the mean minutes of moderate activity a day reported by adults in three wards of a town. Under it the draft says, “Activity was similar across the three wards.” The survey is made up, and so are its 24 values, eight adults per ward.

ward	each adult’s minutes of moderate activity a day
ward 1	10, 15, 21, 27, 33, 39, 45, 50
ward 2	22, 23, 24, 25, 26, 27, 28, 65
ward 3	15, 16, 17, 19, 41, 43, 44, 45

Check the means first. Do ward 2.

```
22 plus 23 plus 24 plus 25 plus 26 plus 27 plus 28 plus 65 = 240
240 divided by 8 = 30
```

Wards 1 and 3 also add to 240, so every mean is 30 minutes. Now the standard deviation, as Book 0 built it (Book 0, D5), again for ward 2. Subtract the mean from each value, square, add, divide by one less than the count.

```
8 squared plus 7 squared plus 6 squared plus 5 squared plus 4 squared plus 3 squared
plus 2 squared plus 35 squared = 1428
1428 divided by 7 = 204
```

The square root of 204 is 14.28. Wards 1 and 3 come to 14.29 and 14.25. To one decimal, all three are 14.3 minutes. So the three bars match, and so would their error bars.

ward	adults	mean (minutes)	standard deviation (minutes)	median (minutes)
ward 1	8	30	14.3	30
ward 2	8	30	14.3	25.5
ward 3	8	30	14.3	30

Now draw every value as a dot and look again. In the second figure, the adults are numbered 1 to 24, ward by ward, lowest value first. They are set out from left to right, so that no dot hides another. Ward 1 is spread fairly evenly from 10 to 50. In ward 2, seven adults report 22 to 28 minutes, and one reports 65. That one value lifts the mean to 30, though seven of the eight report less; the median is 25.5. In ward 3 the adults fall into two clusters, 15 to 19 and 41 to 45. Nobody in ward 3 is closer than 11 minutes to the mean.

“Activity was similar across the three wards” is true of the means and false of the people. Weissgerber and colleagues make the same point with four example datasets behind one bar chart: “many different data distributions can lead to the same bar or line graph”.

WHERE THIS PICTURE BREAKS. These values were made up to share a mean and a standard deviation exactly. Real groups will not match so neatly. The lesson is that matching bars do not show matching groups, not that real wards look like these.

The single 65 in ward 2 is not wrong only because it is far from the rest. Check it against the form before you decide anything. If it was recorded correctly, it stays in the figure.

Illustration 2. Now choose how to show the values. The choice turns on how many values each group has.

With a handful per group, draw a strip chart: every value as a dot, one column per group. The figure for the three wards does the same job with each ward’s dots spread out in order. Add a short line at each group’s median or mean, and say which in the caption. Weissgerber and colleagues found that in 75% of the papers they reviewed, the smallest group in a figure had between two and six values. For groups that small, they call this chart “the best choice”, because box plots and histograms “would be difficult to interpret”.

Watch what a box plot does to ward 3. Work its quartiles the way Book 0 did: the median of the lower four values and the median of the upper four.

```
(16 plus 17) divided by 2 = 16.5
(43 plus 44) divided by 2 = 43.5
```

A box plot draws a box from the lower to the upper quartile, a line at the median, and whiskers out towards the far values. Ward 3’s box runs from 16.5 to 43.5 minutes, with its median line at 30. A reader sees a wide, even spread with a middle at 30. No adult in ward 3 reports anything between 19 and 41. Wilke says it plainly: violin plots, a smoothed outline of where the values fall, “will accurately represent bimodal data whereas a boxplot will not”. Bimodal means two clusters.

With dozens or hundreds of values per group, as in a district survey, dots in one column sit on top of each other. Spread them a little sideways at random, which is jittering. Or draw a box plot,

a violin or a histogram. Before you trust a histogram, redraw it with two or three different bin widths, the width of each bar. A pattern that comes and goes with the bin width is not a pattern in the data.

Last, look at how the data were collected. If the same adults were measured twice, before and after something, the two sets of values are pairs. Two bars, one for before and one for after, hide whether each person went up or down. Draw each person's two values joined by a line.

WHERE THIS PICTURE BREAKS. The cut between "a handful" and "many" is not a fixed number. Draw the dots. If they still show the shape and the count, keep them. If they pile into a smear, change the chart.

All of this is for measured quantities. A share of people with or without something, such as the National Family Health Survey rows in section 14 and section 15, is one number per group. A bar or a dot is the right mark for it.

MUST KNOW POINTS FOR YOU.

- "The bar chart summarises the group" is the error. Very different sets of values give the same mean and the same error bar: an even spread, one far value, two clusters.
- When each group has only a handful of values, draw every value as a dot, with a short line at the median or mean. Say in the caption which the line is, and how many values each group has.
- A box plot is a summary too. It can draw two clusters as one wide box with a median where nobody is.
- When the same people are measured twice, never draw the two occasions as two bars. Join each person's two values with a line, so the reader sees who went up, who went down and who stayed the same.
- Weissgerber and colleagues' systematic review (2015) covered 703 papers in leading physiology journals from early 2014. Of these, 85.6% had at least one bar chart, and only 13.4% had a chart showing every value.
- When you review a colleague's draft with bars of means for small groups, ask for the values as dots.
- Weissgerber and colleagues counted physiology journals only; this book holds no such count for Indian journals. Leaf through a recent issue of the journal you will submit to and count for yourself before you assume its habits.
- Showing every value shows the shape and the count. It does not tell you whether a gap between groups is real or large enough to matter. And it is for measured quantities. A share of people with a condition is one number per group, and a bar or dot is right for it.

COMMON MISREADING. "Show the data" is easily read as "never show a mean". A strip chart with a short line at each group's mean shows the summary and what it summarises, side by side. What the rule forbids is the mean on its own, standing in for values the reader never sees.

REPORTING IT. Each dot is one participant's value; the short horizontal line marks the median of each group. We show every value because no group had more than ten participants.

EXERCISE 1 (retrieval). From memory: name three different ways a group of values can produce the same bar of the mean and the same error bar. Which chart do Weissgerber and colleagues recommend for small groups, and why not a box plot?

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (interpretation). A made-up student survey compares hours of sleep a night in two hostels, six students each. The draft shows two bars of equal height and says, "Sleep did not differ between hostels."

hostel	each student's hours of sleep a night
hostel A	6.5, 7.0, 7.0, 7.5, 7.5, 8.0
hostel B	5.0, 5.5, 5.5, 6.0, 6.0, 15.5

Work out each hostel's mean and median. Sketch the values as a strip chart. Say what the bars hid and what you would do before writing the sentence again.

EXERCISE 3 (design). Find one figure in your own section, or in a recent paper in your field, that shows a measured quantity as bars of means. Redraw it by hand as a strip chart, or say exactly what you would need from the authors to do so. Write the caption's sentence that says what the dots and the line mean.

EXERCISE 4 (teaching). A first-year resident has drawn her thesis results as bars of means, eight patients per group, and asks why her guide wants "the dots". Ten minutes, with a whiteboard.

17 · Honest axes: where an axis starts, what range it shows, and measuring the distortion

DEFINITION. The axis start, written s, is the value at which the value axis begins.

The principle of proportional ink, so named by Bergstrom and West, says that a shaded area must be in proportion to the value it stands for. On an ordinary, linear axis a bar, or an area shaded down to the axis, therefore starts at zero. Ink here means everything drawn on the figure: lines, bars, shading. On a logarithmic axis a bar stands for a ratio, and it starts at 1, because a ratio of 1 means no difference, and the logarithm of 1 is 0 (Book 0, A7). A point or a line carries its value by position alone.

Take two bars with values a and b, a the smaller, drawn up from an axis start s above zero. Their drawn heights are a − s and b − s, read “a minus s” and “b minus s”. The drawn ratio is (b − s) ÷ (a − s), where ÷ is read “divided by”. The data ratio is b ÷ a.

Tufte’s Lie Factor (LF) compares a change as the figure draws it with the same change in the data. Each change is a percentage change: the change divided by the value it started from, times 100.

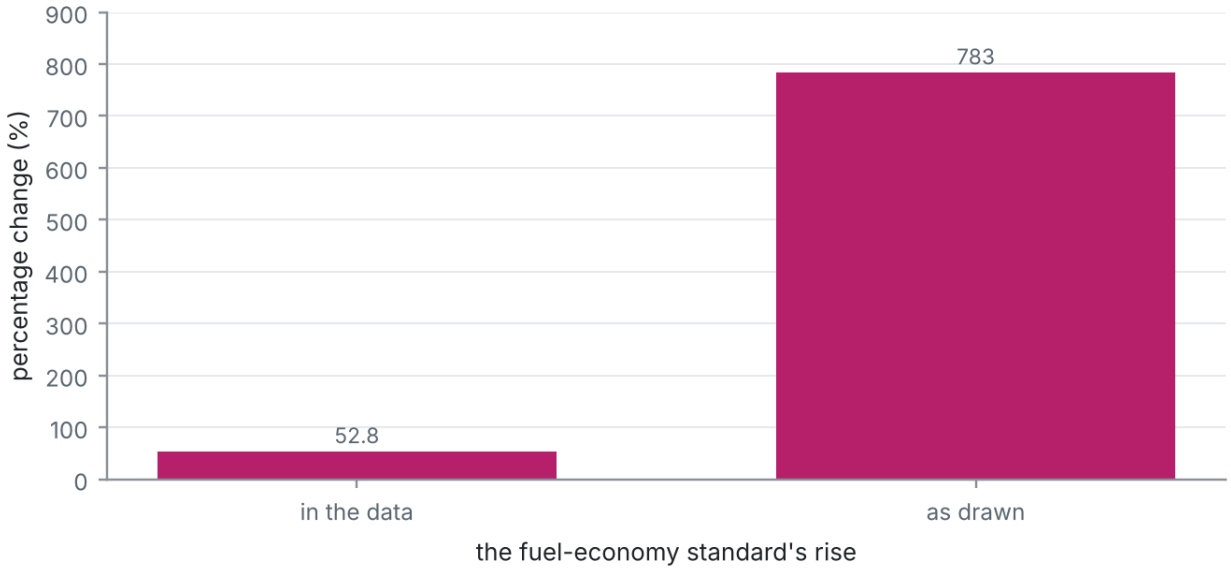
```
LF = (percentage change drawn in the figure) ÷ (percentage change in the data)
```

An LF of 1 draws the change at its true size. Tufte counts an LF above 1.05 or below 0.95 as substantial distortion.

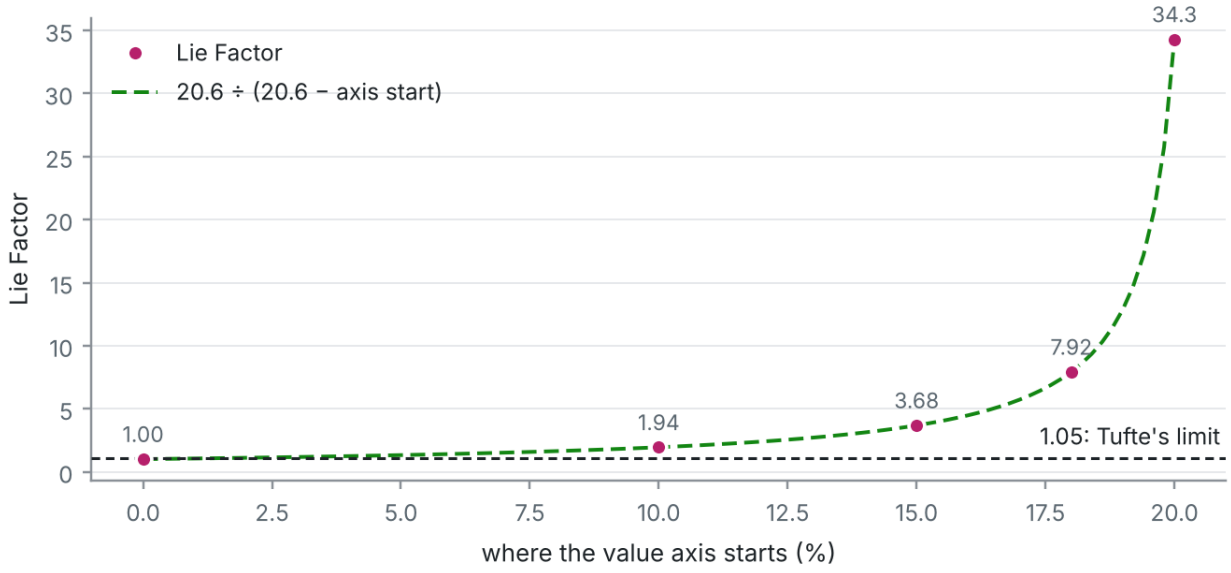
For two bars, measure both changes from the bar the change starts from. The LF then comes to a ÷ (a − s) for a rise from a, and b ÷ (b − s) for a fall from b. Two groups side by side have no start: measure from the group you compare against, and say which.

Starting a bar chart’s axis above zero is called truncating it. Truncation makes readers judge a difference as more severe. [269, 270, 271, 272, 273, 274, 275, 275, 275, 275, 276, 277, 278]

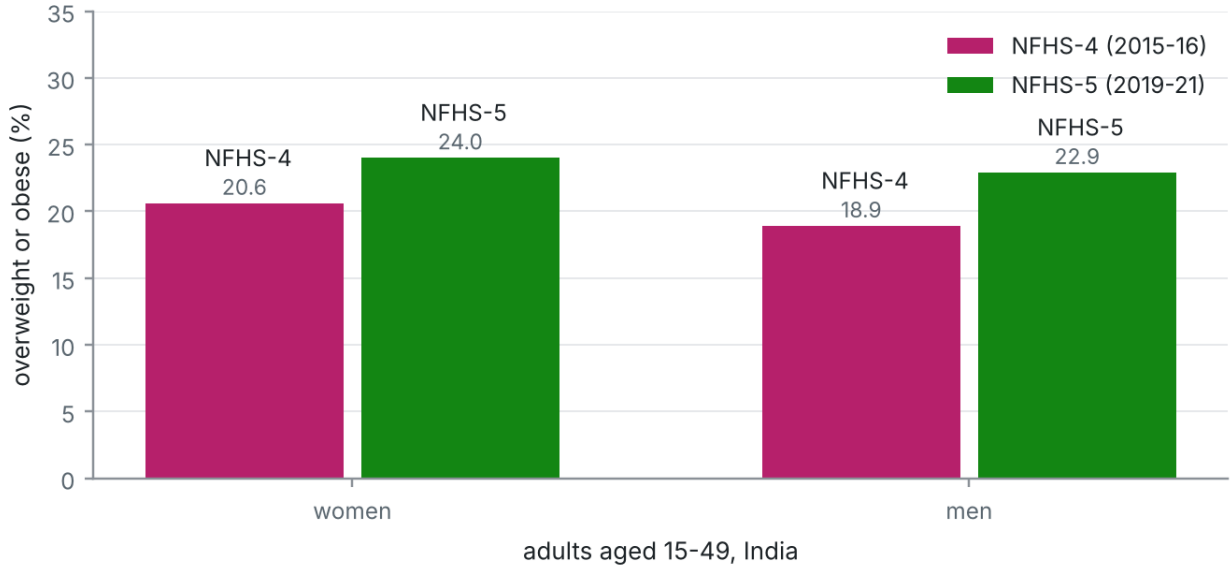
IN PLAIN TERMS. A bar is a length, and a reader reads a length as a value. A small rise then looks like a leap.



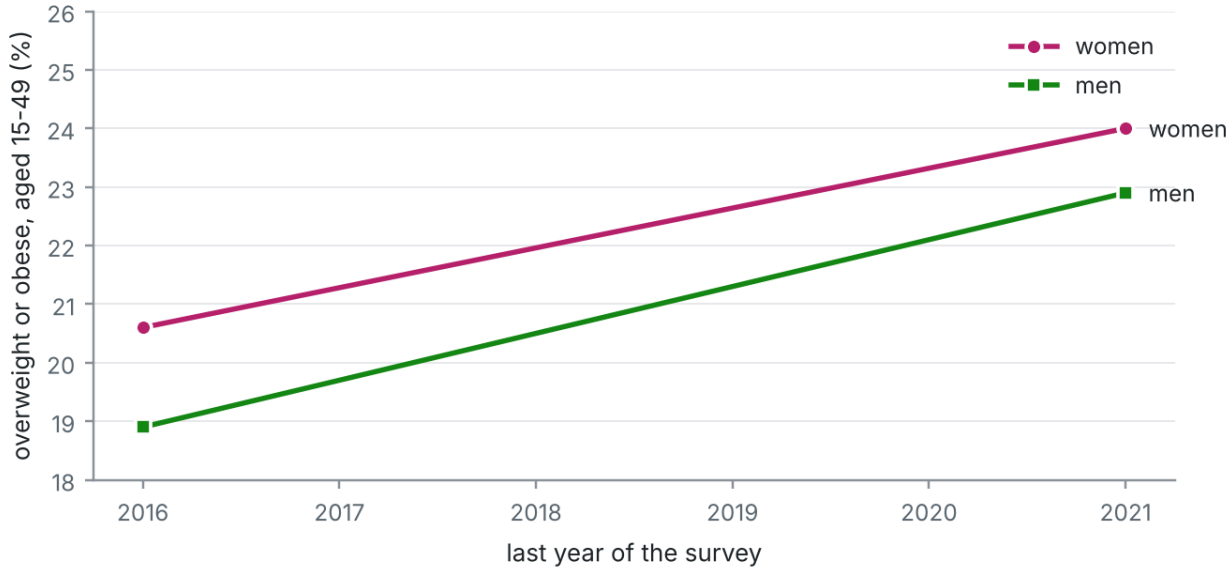
The newspaper's fuel-economy chart drew a rise of 52.8 per cent as a rise of 783 per cent. The standard rose from 18.0 to 27.5 miles per gallon; the lines that drew it grew from 0.6 to 5.3 inches on the page, as Tufte measured them. Lie Factor about 14.8.



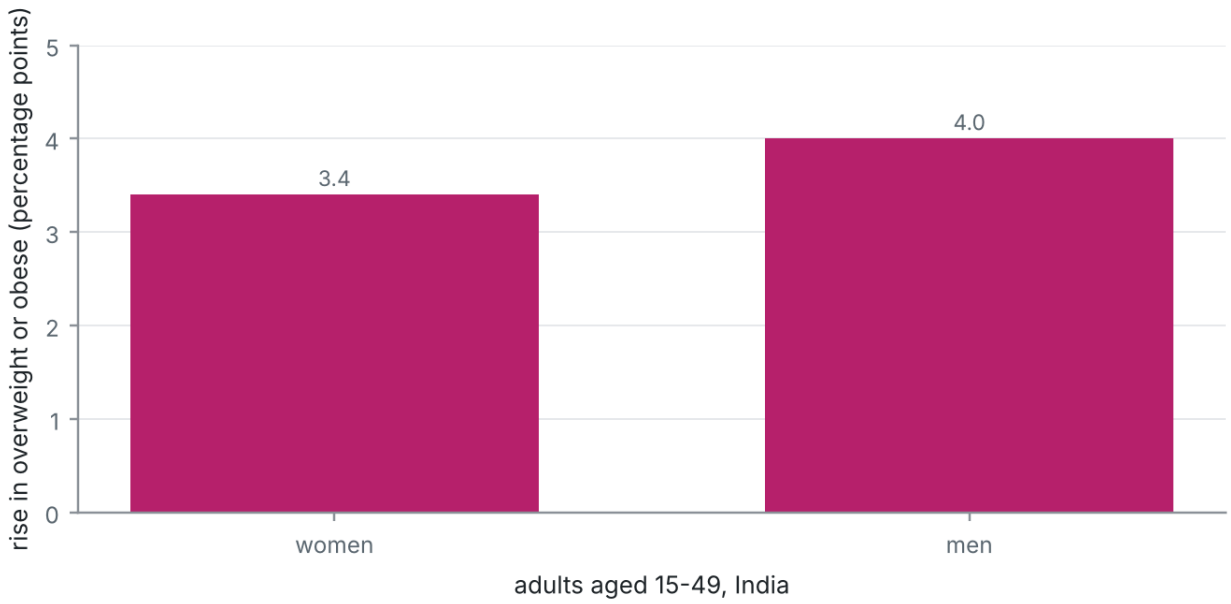
The Lie Factor climbs fastest as the axis start nears the smaller bar. Here it is for a bar slide of women aged 15-49 with overweight or obesity, 20.6% in NFHS-4 and 24.0% in NFHS-5, against where the value axis starts. It is 20.6 ÷ (20.6 − s): 1.00 from zero, 1.94 from 10, 3.68 from 15, 7.92 from 18 and 34.3 from 20. The dashed line is Tufte's 1.05.



Women and men aged 15-49 with overweight or obesity, India, NFHS-4 (2015-16) and NFHS-5 (2019-21). Bars from zero: women 20.6% and 24.0%, men 18.9% and 22.9%. The rise is drawn at its true size beside the level.



The same four values as points joined by lines, each survey placed at its last year: 2016 for NFHS-4 (2015-16), and 2021 for NFHS-5, whose fieldwork ended on 30 April 2021. The side axis starts at 18%, not zero: a point carries its value by position, so this is honest when the caption says so. Women rose by 3.4 percentage points, men by 4.0.



The rise itself, drawn as bars from zero: between NFHS-4 (2015-16) and NFHS-5 (2019-21), overweight or obesity among adults aged 15-49 rose by 3.4 percentage points in women and 4.0 in men. Each bar's length is the change, so the change is drawn at its true size. The two rises come from separate samples. The figure does not show that the rise was larger in men.

Illustration 1. Start with a real chart that went wrong. A newspaper drew the fuel-economy standards the United States set for new cars. The standard rose from 18 miles per gallon in 1978 to 27.5 in 1985. Each year's standard was drawn as a line, longer for a higher standard. Tufte measured the lines in *The Visual Display of Quantitative Information* (1983), on page 57.

> *This line, representing 18 miles per gallon in 1978, is 0.6 inches long.*

> *This line, representing 27.5 miles per gallon in 1985, is 5.3 inches long.*

Work out the change in the data first, as a percentage change.

```
27.5 minus 18.0 = 9.5
9.5 divided by 18.0 = 0.528
0.528 times 100 = 52.8
```

About 53 per cent, as Tufte gives it. Now the change in the drawing, measured on the page.

```
5.3 minus 0.6 = 4.7
4.7 divided by 0.6 = 7.83
7.83 times 100 = 783
```

The lines grew by 783 per cent. Divide the drawn change by the change in the data.

```
783 divided by 53 is about 14.8
```

The Lie Factor is about 14.8. The chart drew a rise of 53 per cent as a rise of 783 per cent. Nothing on it was a wrong number. The lengths were wrong.

Figures quoted in this illustration are from [275].

WHERE THIS PICTURE BREAKS. Tufte's Lie Factor needs a length or an area you can measure on the page. It works for bars, lines whose length is the value, shaded areas and pictures. It does not apply to a point on a line chart, which carries its value by position: there is no drawn length to measure.

The 14.8 is this one chart's. It says nothing about newspaper charts in general.

Illustration 2. Now run the same check on the running case of this book. Open the NFHS-5 India fact sheet at https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf and find indicator 88. In NFHS-5 (2019-21), 24.0% of women aged 15 to 49 had overweight or obesity. In NFHS-4 (2015-16) the figure was 20.6%.

Suppose a district slide, made up for this example, draws the two values as bars on an axis that starts at 20. Each bar runs from 20 up to its value.

```
20.6 minus 20 = 0.6
24.0 minus 20 = 4.0
4.0 divided by 0.6 = 6.67
```

The NFHS-5 bar is 6.67 times the NFHS-4 bar. The data ratio is much smaller.

```
24.0 divided by 20.6 = 1.165
```

Now the Lie Factor. The change in the data is 3.4 percentage points on a start of 20.6. The change in the drawing is 3.4 points of height on a start of 0.6.

```
3.4 divided by 20.6 = 0.16505
3.4 divided by 0.6 = 5.6667
5.6667 divided by 0.16505 = 34.33
```

A Lie Factor of about 34.3, far past Tufte's 1.05.

Look at what happened in those lines. The 3.4 sat on top of both fractions and dropped out. What was left was 20.6 divided by 0.6. That holds for any rise between two bars. Written as fractions, the change drawn is (b − a) ÷ (a − s), and the change in the data is (b − a) ÷ a. Divide the first by the second. Dividing by a fraction is multiplying by it turned upside down. The b − a then cancels, top and bottom.

```
(b - a) ÷ (a - s) × a ÷ (b - a) = a ÷ (a - s)
LF = a ÷ (a - s)
```

So, for a rise from a, the Lie Factor depends only on the starting bar and where the axis starts. Try it on the slide.

```
20.6 divided by 0.6 = 34.33
```

The same answer. Now move the axis start and watch the Lie Factor grow.

axis start (%)	NFHS-4 bar drawn (points)	NFHS-5 bar drawn (points)	drawn ratio	Lie Factor
0	20.6	24.0	1.17	1.00
10	10.6	14.0	1.32	1.94
15	5.6	9.0	1.61	3.68
18	2.6	6.0	2.31	7.92
20	0.6	4.0	6.67	34.3

Starting the axis halfway to the smaller value, at 10, already draws the rise almost twice its size. The last few points before 20.6 do most of the damage.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. The shortcut a ÷ (a − s) holds for a rise from a. For a fall from b it is b ÷ (b − s). For two groups, work the two percentage changes out in full from the group you compare against, and say which. With several bars, work it out for each pair you expect the reader to compare.

The NFHS values are real; the slide is made up. The fact sheet's figures are for ages 15 to 49, and "overweight or obese" is one combined row.

Illustration 3. You have a slide that exaggerates. Here are four honest ways to redraw the same numbers. Which one you choose depends on the one claim the figure makes (section 14).

First, bars from zero. The NFHS-4 and NFHS-5 bars stand at 20.6 and 24.0 for women, and at 18.9 and 22.9 for men (indicator 89). The rise now looks modest. It is modest beside the level: about one adult in five had overweight or obesity in NFHS-4, and nearly one in four in NFHS-5. Bars from zero carry the level.

Second, points joined by a line, on an axis near the data. A point is read by its position, so its axis may start above zero. Bergstrom and West say how to choose the range: "usually by allowing the axis to span the region not much larger than the range of the data values". The four values run from 18.9 to 24.0. An axis from 18 to 26 holds them with a little room, and it carries the change. Say in the caption that the axis starts at 18. Along the bottom, put each round at its last year: 2016 for NFHS-4, the end of its 2015-16 label, and 2021 for NFHS-5, whose fieldwork ran to 30 April 2021.

Third, bars of the change itself, from zero. Wilke allows this: bars may show a difference "as long as we make it explicit which differences we are showing". Women rose by 3.4 percentage points and men by 4.0. Two bars of 3.4 and 4.0, from zero, show the rise at its true size. The two rises come from separate samples. The bars do not show that the rise was larger in men.

Fourth, for a ratio, a logarithmic axis. Overweight or obesity was found in 33.2% of urban women and 19.7% of rural women in NFHS-5.

```
33.2 divided by 19.7 = 1.685
19.7 divided by 33.2 = 0.593
```

"Urban about 1.69 times rural" and "rural about 0.593 times urban" are the same fact said two ways. On a log axis that starts at 1, the two bars have the same length and point opposite ways: $\log_{10}$ of 1.685 is 0.227 and $\log_{10}$ of 0.593 is −0.227. Drawn from 0 on a linear axis, the first bar would be almost three times the second. Wilke: "Bars on a log scale represent ratios and must always start at 1".

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. The four redraws are honest about lengths. None of them settles which message is right. Bars from zero say "about one adult in four or five"; a line from 18 says "it rose". Both are true. Choosing between them is the claim of the figure, and the caption should state the numbers either way.

The log axis is for ratios. Never draw an amount, such as a percentage, as a bar on a log axis: zero lies infinitely far down it, so any start you pick is arbitrary.

Illustration 4. Does any of this change what readers think? Correll, Bertini and Franconeri tested it. Open "Truncating the Y-Axis: Threat or Menace?", the authors' preprint of their CHI 2020 paper, at https://arxiv.org/abs/1907.02035. In their first experiment, 40 people rated how big the change in each chart looked, on a scale from 1 to 5. The same data were drawn with the axis starting at 0, 25 or 50%. Search for 0.36.

> *compared to an increase of 0.36 for starting the y-axis at 25% rather than 0%*

Starting the axis a quarter of the way up raised the rating by 0.36 on the 5-point scale. Bar charts and line charts behaved alike.

Next they tried the usual fix: a break mark on the axis and on the bars. Search for Merely.

> *Merely indicating that truncation has occurred, even in a prominent and unambiguous way, may not be sufficient to "de-bias" viewers of truncated charts.*

The break mark did not stop the exaggeration. In a third experiment, people wrote down the first and last values before rating. Their estimate of the change was no worse on a cut axis. The exaggeration stayed.

> *This bias is not merely a misreading of values, but seems to be connected to the visual magnification of differences.*

So labelling a cut axis does not repair a bar chart. Changing the lengths does.

Figures quoted in this illustration are from [279].

WHERE THIS PICTURE BREAKS. These were paid online participants rating plain charts with no subject matter, 40, 32 and 25 people in the three experiments. The size of the shift will differ for a district officer reading a slide about her own district.

The authors do not conclude that every axis must start at zero. They write that "There is no a priori, domain-agnostic ground truth" for how severe an effect ought to look. The axis start is a choice, and the person who makes it owns the impression it gives.

MUST KNOW POINTS FOR YOU.

- "Every number on the axis is correct, so the chart cannot mislead" is the error. Readers judge the lengths. In Correll's experiments the exaggeration stayed even when readers reported the numbers accurately. Fix the drawing, not the labels.
- Before you show or accept a bar chart whose axis starts above zero, work out the LF from the bar the change starts from: a ÷ (a − s) for a rise from a, b ÷ (b − s) for a fall from b. For two groups, take the group you compare against, and say which. If the LF is above 1.05, redraw the chart before anyone discusses the size of the gap.
- A break mark on the axis does not make a truncated bar chart honest.
- Shading counts as ink. If you fill the area under a line, take the axis down to zero, or leave the line unfilled.
- To show a small change honestly, draw the change itself as bars from zero, or draw points on a stated range near the data.
- Draw a ratio on a logarithmic axis, with bars from 1. Then "twice" and "half" have the same length, pointing opposite ways, as the same fact said two ways should.
- The range you pick for a line chart still sets how big the change feels. So state the axis start in the caption, and give the change in numbers.
- Report the change in percentage points and in per cent, with both values.

COMMON MISREADING. "The bars start at 20, but the axis says so, so anyone who reads the numbers is fine." The numbers are right and the picture is still wrong. Readers judge the drawn lengths, and in Correll and colleagues' experiments the exaggeration stayed even for readers

who reported the numbers accurately. The axis label does not change the Lie Factor. Only the lengths do.

REPORTING IT. The vertical axis of the figure starts at 18%, not at zero. The share of women aged 15-49 with overweight or obesity rose from 20.6% in NFHS-4 (2015-16) to 24.0% in NFHS-5 (2019-21). That is a rise of 3.4 percentage points.

EXERCISE 1 (retrieval). From memory: what does the Lie Factor divide by what? What value means the drawing is honest, and from what value does Tufte call it substantial distortion? What does it come to for two bars drawn from an axis start s?

EXERCISE 2 (design). Take one result from the manuscript or thesis section you are rewriting for this book. Plan its figure's value axis.

Write down four things. The mark you will use, and why. Where the value axis starts, and the range it shows. If you use bars, the Lie Factor of that start. The sentence of the caption that tells the reader where the axis starts and how big the change is in numbers.

PRACTICE. Fourteen problems on this technique, easiest first. Work them on paper before turning to the answers, which are in the appendix at the back under this section's name.

1. Two bars stand for 40 and 50. The value axis starts at 30. Work out the height each bar is drawn above the axis start, the drawn ratio and the data ratio.

2. Two bars stand for 40 and 50, and the value axis starts at 30. Work out the percentage change in the data, the percentage change in the drawing, and the Lie Factor.

3. A value rises from 10 to 20. An infographic shows each value as a picture, and doubles the picture's height and its width for the second value. Work out how many times larger the second picture's area is, and the Lie Factor if readers judge by area.

4. Two bars stand for 60 and 66. On the page, the second bar is three times as tall as the first. Where does the value axis start?

5. Two ratios, 2 and 0.5, are drawn as bars on a logarithmic axis. Work out each bar's length on the log scale when the bars start at 1: the $\log_{10}$ of its ratio. Then work them out when the bars start at 0.1.

6. In NFHS-4 (2015-16), 18.9% of men aged 15-49 in India had overweight or obesity. In NFHS-5 (2019-21) it was 22.9%. A slide draws the two as bars on an axis starting at 18%. Work out the drawn heights, the drawn ratio, the data ratio and the Lie Factor.

7. Bergstrom and West describe a bar chart of jobs in Tennessee. The value for 2014 is about 1.08 times the value for 2010, but the 2014 bar uses about 2.7 times as much ink. The bars are equally wide. Work out the Lie Factor. Then work out where the axis started, as a fraction of the 2010 value.

8. In NFHS-5, 33.2% of urban women aged 15-49 and 19.7% of rural women had overweight or obesity. A slide draws the two as bars on an axis starting at 15%. Work out the Lie Factor of the comparison.

9. You want one figure for the change in overweight or obesity in adults aged 15-49 between NFHS-4 and NFHS-5. Women went from 20.6% to 24.0%, and men from 18.9% to 22.9%. You

choose points joined by lines. Choose the value axis's range, say why, and say what would happen if a colleague drew bars on the same axis.

10. In NFHS-5, 29.8% of urban men aged 15-49 and 19.3% of rural men had overweight or obesity. For women the two figures were 33.2% and 19.7%. Draw the urban-to-rural ratio for women and for men as bars on a logarithmic axis. Work out each ratio, each bar's length on the log scale (the $\log_{10}$ of its ratio), and the ratio of the two lengths.

11. Here is a worked answer. Find the step that broke.

```
women overweight or obese: 20.6% in NFHS-4, 24.0% in NFHS-5
bars drawn from an axis start of 20
drawn heights: 0.6 and 4.0
drawn ratio: 4.0 divided by 0.6 = 6.67
data ratio: 24.0 divided by 20.6 = 1.165
Lie Factor: 6.67 divided by 1.165 = 5.73
```

12. Here is a worked answer. Find the step that broke.

```
the slide: bars for 20.6% and 24.0% on an axis from 20, Lie Factor 34.3
the fix: a zig-zag break mark added to the axis and to both bars
the axis numbers are correct, and the break tells the reader the axis was cut
so readers now read the values, not the lengths: the Lie Factor is 1
```

13. Correll and colleagues reproduce a bar chart shown on Fox News. Their caption says the second bar is 6 times taller than the first. It goes on: "even though there is only a 4.6% increase in tax rate (ratio of 1.13 to 1)". A colleague says: "So the chart made a 4.6% rise look like a 500% rise."

Decide what to compute, compute it, and say what the chart's critics cannot claim from it.

14. A district officer presents NFHS results with a bar slide and says: "Look at this. Overweight and obesity in men almost doubled between the two surveys."

Decide what to compute, compute it, and say what the claim and the slide do not establish.

18 · Decoration that does no work

Checked against the sources as they stood on 2026-10-02.

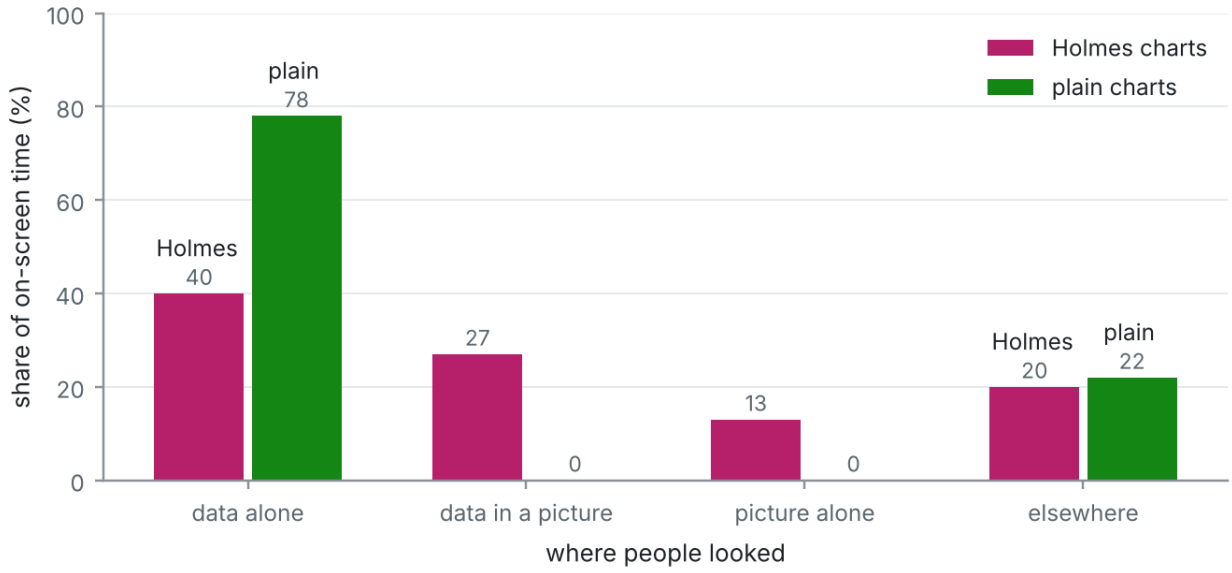
DEFINITION. Tufte divides the ink on a figure into two kinds. Data-ink is the ink that changes when the data change: the bars, the points, the line. Everything else is non-data ink: frames, grid lines, background fills, shadows. The data-ink ratio is the data-ink divided by all the ink used to print the figure.

Tufte's rule is to maximise this ratio "within reason": erase non-data ink, and erase data-ink that repeats what other ink already shows. He gives no one-sentence definition of chartjunk. He introduces it as decoration that "does not tell the viewer anything new".

In this book, decoration is any element that tells the reader nothing about the data or about how to read them. A shadow, a coloured background behind a single plot, a frame round the whole figure and a dark grid are decoration. So is a 3-D effect on flat data: numbers that vary in one direction only, such as one value per group. Axis titles, tick labels and a light grid or reference line are context. So is a direct label: a group's name written beside its data, in place of a key saying which colour is which (drawing programs call it the legend). Context is non-data ink, and it stays (Wilke). [275, 275, 275, 275, 275, 280, 281, 282, 283, 284, 285, 286, 275, 275, 287, 288, 289, 290, 291, 292, 293, 294]

IN PLAIN TERMS. Every mark on a figure either tells the reader something or gets in the way. The bars and points tell. So do the axis titles and the labels: without them the bars mean nothing.

So ask one question of each mark: what does this tell the reader? If the answer is "nothing", delete it.



The pictures took looking time. On the Holmes charts, built into cartoon pictures, 40% of on-screen time went on data alone, 27% on data drawn as part of a picture, 13% on picture alone and 20% elsewhere. On the plain charts, 78% went on data and 22% elsewhere; they have no picture, so those two bars are 0. Bateman and colleagues, "Useful junk?": an eye tracker, 20 students, untimed viewing.

Illustration 1. Here is a figure of the kind spreadsheet defaults and slide templates produce. It is made up for this example, from real numbers. It shows the share of women and men aged 15-49 with overweight or obesity in the National Family Health Survey (NFHS-5), 24.0% and 22.9%. The bars are 3-D columns on a shaded background. A dark grid sits behind them, and a frame runs round the whole. A key box says which colour is which. The value is printed on top of each bar.

Go through it one element at a time. For each, ask: what does this tell the reader about the data, or about how to read them?

element	what it tells the reader	what to do
3-D columns	nothing; the third dimension carries no data	draw flat bars
shaded background	nothing	delete
dark grid	where values fall, but it competes with the bars	keep a few light grey lines, or none
frame round the figure	nothing	delete
key box	which bar is which, but the eye must travel to it	direct labels: "women" and "men" under the bars
value on each bar	the exact number, a second time	keep only if readers need the exact values
axis title with its unit	what the bars measure	keep

The 3-D columns are not only useless. They bend the data. Open Wilke's *Fundamentals of Data Visualization*, chapter 26, at https://clauswilke.com/dataviz/no-3d.html, and search for 322.

> *there were 322 passengers total traveling in 1st class, yet Figure 26.2 suggests that the number was less than 300.*

A 3-D bar's top floats away from the back wall where the grid lines are drawn. Readers read it against the wrong line.

The value printed on each bar is a smaller case. Tufte counts six ways a plain labelled bar shows one number: its left edge, its right edge, its shading, its top line, the position of the label and the label itself. Any five can go and the sixth still shows it. Keep the printed value when readers will want the exact number. Otherwise the bar is enough.

Now stop. Wilke shows the opposite mistake: a figure whose axis labels are so faint, and whose grid is so bare, that the points "seem to float in space". Tufte's own rule ends "within reason". The axis title, the tick labels and the direct labels stay.

Figures quoted in this illustration are from [24, 295].

WHERE THIS PICTURE BREAKS. The table works for a figure read on a page. A light grid that is clutter in a paper can help on a figure people use to look up values, and Tufte himself says a grid may help "When a graphic serves as a look-up table". Rougier and colleagues put it generally: "an element that could be considered chartjunk in one figure can be justified in another". Ask what the reader will do with the figure before you delete anything.

The made-up slide uses real NFHS numbers. The Titanic figure is Wilke's, and the 322 is his.

Illustration 2. Now the case for decoration, and it is a real one. Open Bateman and colleagues (2010), "Useful junk? The effects of visual embellishment on comprehension and memorability of charts", at https://doi.org/10.1145/1753326.1753716. Read the abstract.

> *We found that people's accuracy in describing the embellished charts was no worse than for plain charts, and that their recall after a two-to-three-week gap was significantly better.*

Before you use this, read how it was done. Twenty students saw 14 charts by the graphic artist Nigel Holmes, whose charts are built into cartoon pictures, and plain versions of the same charts. Each person saw one version of each chart. They described each chart aloud, with as much time as they wanted. Then 10 of them were tested on the charts after five minutes, and 10 after two to three weeks.

An eye tracker recorded where they looked while the charts were on screen. Search for 67%.

> *participants spent 67% of their on-screen time looking at data or data and embellishments for Holmes charts as compared to 78% for Plain charts.*

> *For Holmes charts, users spent 27% of their on-screen time looking at areas that contained both data and embellishments, and 13% of their time looking at embellishments that did not communicate any data.*

Split the Holmes charts' time four ways.

```
67 minus 27 = 40
100 minus 67 minus 13 = 20
40 plus 27 plus 13 plus 20 = 100
100 minus 78 = 22
```

On the Holmes charts, 40% of the time went on data alone, 27% on data drawn as part of a picture, 13% on picture alone and 20% elsewhere on the screen. On the plain charts, 78% went on data and 22% elsewhere. The pictures took looking time, and still the descriptions were no worse.

Now read what else they found. Search for value message.

> *with better descriptions of a value message for the Holmes charts than the plain charts*

A value message is the opinion a chart carries: "costs are monstrous". The pictures carried one, and readers picked it up. The authors write that "we do not advocate this strategy as a general principle".

So here is what the study allows. A picture tied to the message may help a poster or a public leaflet be remembered. It does not allow a 3-D bar, a shadow or a background in a paper's figure. Those carry no message and some bend the data. And it does not show that decoration is harmless when a reader studies exact values against a clock.

Figures quoted in this illustration are from [296].

WHERE THIS PICTURE BREAKS. One study, 20 students at one university, 14 charts by one artist, and 10 people in each memory group. Viewing was untimed and the task was to describe each chart aloud. The paper reports t-tests and p values in its text, not the size of the difference in points. A reader of a methods section, checking a confidence interval against an axis, is doing a different task.

Plain charts are not neutral either. The authors point out that a minimal chart can still mislead through "the scale of axes" and the choice of data. Removing decoration does not fix an axis; section 17 does.

MUST KNOW POINTS FOR YOU.

- "The cleanest figure is the one with the least ink" is the error. Delete decoration; keep context.
- Ask one question of every element of a figure: what does it tell the reader about the data, or about how to read them? Do this before you change anything else.
- Never draw flat data in 3-D. The third dimension carries no data and it bends the bars: in Wilke's example, 322 first-class passengers read as fewer than 300.
- Use direct labels: write each line's or group's name beside its data, and delete the key.
- "Useful junk?" found that cartoon charts were described no worse, and were remembered better two to three weeks later. That was 20 students, with no time limit. It

does not show that decoration is harmless when a reader needs exact values, and the authors say so. Keep it out of a paper's figure.

- A picture on a chart carries an opinion. On a public chart, choose a picture only if you are willing to say its message out loud.
- Removing decoration cannot fix a dishonest axis. A plain bar chart that starts above zero is still wrong. So fix the axis first, then the decoration.
- When a trainee brings you a chart straight from a spreadsheet, do not redraw it for them. Have them delete one element at a time, and ask after each whether the reader lost anything.

COMMON MISREADING. "A study showed decorated charts are remembered better, so decoration is fine." The study had 20 students. They had no time limit, and they described each chart aloud. Each picture was tied to its chart's subject. The study found no loss for that task, and better memory weeks later.

It also found that the pictures carried the designer's opinion. It says nothing for 3-D, shadows or coloured backgrounds, which carry no message. And it says nothing for a figure read for exact values.

EXERCISE 1 (retrieval). From memory: what is data-ink, what is the data-ink ratio, and what two words does Tufte attach to his rule to maximise it? Name two elements that are not data-ink but should stay.

EXERCISE 2 (critique). A made-up district report has a figure for six blocks. It shows the share of children under 5 weighed last month at the anganwadi, the village child-care centre. The figure has:

- 3-D cylinders in six colours, with a shadow under each;
- a key matching the colours to block names;
- a dark grid every 5%;
- a picture of a weighing scale behind the bars;
- no axis title;
- a value axis running from 60% to 100%.

List what you would change, in the order you would change it, and say why for each.

EXERCISE 3 (teaching). A first-year resident has read the abstract of "Useful junk?" and says: "So chartjunk is fine now. Decorated charts are remembered better." You have ten minutes and a whiteboard. Say what you would teach.

19 · Colour: for meaning only, readable by everyone, and in grey

DEFINITION. In a data figure, colour does one of three jobs. It separates groups that have no order, using a qualitative scale: a small set of colours that look clearly different from each other and equally strong, so that none stands out. Groups that have an order, such as age groups, take steps of one colour from light to dark. Colour also stands for a value, using a sequential scale, which runs from light to dark. Or it picks out one element: one strong accent colour, with the rest in grey. Colour that does none of these jobs is decoration (section 18).

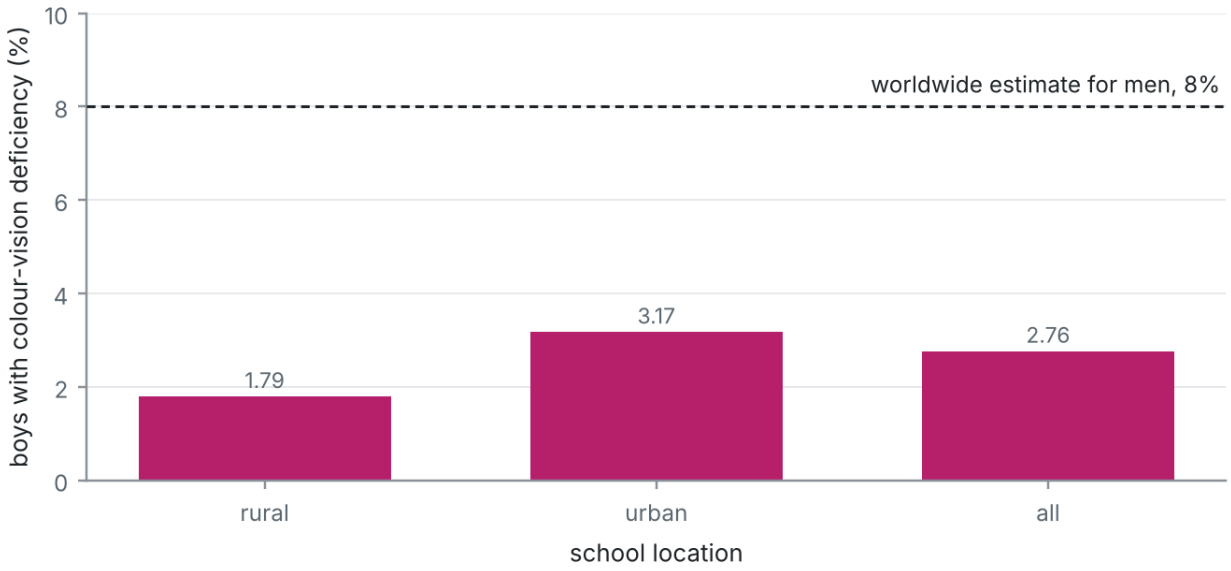
A colour scale shows values fairly when it is perceptually uniform: the same step in the data looks like the same step in colour anywhere along the scale. Its lightness should also change in one direction only, so that the order of the values survives when the colour is removed. The rainbow scale fails both. Crameri, Shephard and Heron (2020) report that its distortion can exceed seven per cent of the range of the data shown.

Colour-vision deficiency is a reduced ability to tell certain colours apart, most often red from green. [297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321]

IN PLAIN TERMS. Colour in a chart should have a job. It can tell groups apart, stand for a number, or point at one thing. If it does none of these, take it out and draw in black or grey.

Some of your readers cannot tell red from green. Some copies of what you write will be printed or photocopied in black and white. A grey copy keeps only how light or dark each colour is. So two colours of the same lightness turn into the same grey, and the reader loses what they meant.

So test every colour figure twice before it leaves you. Look at it in grey. Then run it through a program that shows it as a reader with colour-vision deficiency sees it.



In one district, about one boy in 36 had colour-vision deficiency: 2.76% of 74986 boys aged 11 to 17 screened in schools of Kanchipuram district, Tamil Nadu (Krishnamurthy and colleagues, 2021); 1.79% of 21911 in rural schools and 3.17% of 53075 in urban schools. The dashed line is the general worldwide estimate for men, 8%, which the study did not measure. Girls were not screened.

Illustration 1. Start with a chart that fails. You draw the National Family Health Survey (NFHS-5) results for adults aged 15 to 49. In urban India, 33.2 per cent of women had overweight or obesity, and 19.7 per cent in rural India. For men the two numbers were 29.8 and 19.3 per cent. You draw four bars, grouped by sex. You colour urban red and rural green, because those are the first two colours your program offers. A key in the corner says which is which.

Now test it two ways.

First, print it on a black-and-white printer, or photocopy it. A grey copy keeps only how light or dark each colour is. A red and a green of about the same lightness print as the same grey. The key now tells the reader nothing.

Second, think of the reader who cannot tell red from green. Crameri, Shephard and Heron, “The misuse of colour in science communication” (Nature Communications, 2020, https://pmc.ncbi.nlm.nih.gov/articles/PMC7595127/), list four checks any reader can run on a colour figure. One of them:

> *The colour map should not contain red and green at a similar luminosity*

Luminosity there means how bright or dark a colour looks. Run your chart through a colour-vision-deficiency simulator. Claus Wilke’s book *Fundamentals of Data Visualization* (2019, free online at https://clauswilke.com/dataviz/) says several exist, online and as programs for your computer. In the simulated copy, the red and green bars look alike.

Now fix it, in this order.

1. Ask what the colour is for. Here it separates two groups, urban and rural. That is a real job, so the colour may stay.

2. Put the words on the chart. Write “urban” and “rural” beside the bars and delete the key. This is the direct labelling of the last section. The chart now reads with no colour at all.
3. Make the two colours differ in lightness, not only in hue. Hue is the name of the colour: red, green, blue. A dark blue for urban and a light blue for rural survives a grey copy.
4. Or take two colours from a palette built for colour-vision deficiency. Wilke uses the Okabe–Ito palette throughout his book. Its orange is #E69F00 and its blue is #0072B2, codes most programs accept. Keep large bars soft: Wilke warns against “large filled areas of overly saturated colors”.
5. Print it in grey again, and run it through the simulator again. In both copies a reader should still be able to say which bar is urban.

Step 2 did most of the work. Once the words are on the chart, colour only helps; nothing depends on it. Wilke calls this redundant coding: the same information carried two ways.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. Labels beside the bars work when there are few groups and room for the words. Wilke’s rule of thumb is that colours for groups work best with three to five groups. At eight to ten or more, matching colours to groups becomes too hard. Then label directly, or label only the groups the text talks about.

A palette built for colour-vision deficiency is not a pass. Wilke shows that colours are harder to tell apart on thin lines and small points than on large bars, and harder still under colour-vision deficiency. Test the figure itself, at its printed size, not the palette.

Illustration 2. Now colour does a different job: it stands for a number. Say you want a map of your state’s districts, each shaded by the per cent of women with overweight or obesity. NFHS-5 publishes a fact sheet for every district, so the numbers exist.

Your program offers a rainbow: blue for low, through green and yellow, to red for high. It looks lively. Turn the map grey and see what happens. Wilke describes the rainbow scale in grey:

> *The scale goes from medium dark to light to very dark and back to medium dark*

So in grey the order is lost. A district at the bottom of the scale and one at the top come out about equally dark. In colour the rainbow fails another way. Its yellow is the brightest colour and pulls the eye, wherever it falls on the scale. Its greens cover a wide band in which real differences are hard to see. Crameri and colleagues put a size on the damage: the colours can misrepresent the data “by more than seven percent of the displayed data variation”.

Use instead:

- For a value that runs from low to high, a sequential scale. One colour going from light to dark will do. Scales built to be even, such as viridis or cividis, are listed with their sources in Crameri and colleagues’ Box 2.
- For a change that can go either way, such as each district’s rise or fall from NFHS-4 to NFHS-5, a diverging scale. Put its light middle colour at zero, where the change is nothing.

Wilke's warning: "Diverging scales need to be balanced". The colours should darken at the same rate on both sides, or a rise will look bigger than a fall of the same size.

- Always print the colour bar: the strip, like an axis, that shows which colour stands for which value. Crameri and colleagues say that leaving it out is like leaving the labels and tick marks off an axis.

Then turn the map grey again. A sequential scale that runs from light to dark keeps its order in grey.

Figures quoted in this illustration are from [322].

WHERE THIS PICTURE BREAKS. An even colour scale makes equal differences look equal. It does nothing for the numbers themselves. A district's percentage from a survey carries its own error, and the colour hides it: two neighbouring shades may differ by less than that error. If the reader needs the numbers, give them in a table as well, as Wilke suggests for any figure whose data the reader may want in full.

Illustration 3. How many of your readers is this about? One figure appears almost everywhere. Crameri and colleagues give it as "the general estimate":

> *worldwide 0.5% of women and 8% of men are subject to a colour-vision deficiency*

It is a general estimate they cite, not something they measured. They add that it is lower in sub-Saharan Africa and probably higher in some European populations.

Now look for an Indian measurement. Krishnamurthy and colleagues, "Prevalence of color vision deficiency among school-going boys in South India" (Indian Journal of Ophthalmology, 2021, https://pmc.ncbi.nlm.nih.gov/articles/PMC8482944/), screened every boy in classes 6 to 12, aged 11 to 17, in the schools they visited in Kanchipuram district, Tamil Nadu. Boys who failed a first set of colour plates were tested on Ishihara's plates to confirm. Their Table 1 gives the counts by school location:

school location	boys screened	colour-vision deficiency	per cent
rural	21911	392	1.79
urban	53075	1681	3.17
all	74986	2073	2.76

Check the total row, and turn it into "one boy in how many".

2073 divided by 74986 is 0.0276, to four decimal places.

```
0.0276 times 100 = 2.76
100 divided by 2.76 = 36.2
100 divided by 8 = 12.5
```

So about one boy in 36 in this district, against one man in 12 or 13 in the general estimate.

Now say what the 2.76 is. It is boys, at school, in one district, found by colour plates. Girls were not screened. Within the same district, the share differed between blocks, from 1.12 per cent in one to 3.40 per cent in another. So it is not a rate for India. The authors go further than their data in one sentence:

> *Extrapolating this, there could be 17 million boys affected with CVD in the country.*

That sentence assumes one district's boys stand for every boy in India. Book 0's section on sampling (Book 0, D6) says why a sample speaks only for the population it was drawn from. This study did not sample India.

For your figure, none of this changes the rule. Whether it is one man in 12 or one boy in 36, some of the people on your thesis committee, or among a journal's reviewers, may not tell red from green. Designing for them costs you two tests and a few labels.

Figures quoted in this illustration are from [322, 323].

WHERE THIS PICTURE BREAKS. Both numbers are counts of people who failed colour plates. The 8 per cent is a figure the authors cite from others, for the world as a whole; it is not an Indian measurement. The 2.76 per cent is a measurement, but of boys aged 11 to 17 in one district, screened in school corridors in daylight. Neither tells you how many of your own readers, adults of both sexes, have colour-vision deficiency.

MUST KNOW POINTS FOR YOU.

- "More colours make a figure clearer" is the error. Colour with no job gives the reader a puzzle to solve.
- Each job takes its own kind of colour. A rainbow fits none of them.
- Red against green is the tempting pair: good and bad, urban and rural, before and after. At similar lightness it fails twice: in a grey copy and for readers with red-green colour-vision deficiency, the commonest kind. Change one of the pair, or make them differ in lightness.
- The general estimate cited worldwide is 8 per cent of men and 0.5 per cent of women with colour-vision deficiency.
- The one Indian measurement here found 2.76 per cent of 74,986 school boys in Kanchipuram district, Tamil Nadu. It is boys in one district, with block figures from 1.12 to 3.40 per cent.
- Write "urban" on the urban bar so that nothing depends on colour at all.
- A palette sold as colour-blind safe stops being safe on thin lines and small points. Wilke shows colours are harder to tell apart there, and harder still under colour-vision deficiency. So test the finished figure at its printed size, not the palette.
- When you review a colleague's map or heat map (a table whose cells are shaded by value), run Crameri and colleagues' checks aloud. Is there a colour bar? Does the scale get steadily lighter or darker? Are red and green at similar lightness? Is it a rainbow?

COMMON MISREADING. "Eight per cent of men are colour-blind, so 8 per cent of Indian men are." The 8 per cent is a general estimate for the world that Crameri and colleagues cite; they did not measure it. The one Indian measurement in this section found 2.76 per cent, in school boys of one district of Tamil Nadu. Even there, the share varied from block to block. Neither number is a rate for Indian adults. Your figure needs the same two tests whichever number is true.

REPORTING IT. Colour in Figure 2 only separates urban from rural residence; each bar is also labelled, so the figure reads the same in black and white.

EXERCISE 1 (retrieval). Without looking back: name the three jobs colour can do in a data figure, and the kind of colour each job takes. Then name the two tests a colour figure should pass before it leaves you.

EXERCISE 2 (critique). A colleague's thesis draft has this figure, described in words. It is made up for this exercise.

Figure 4 is a map of the districts of a state. Each district is shaded on a rainbow scale, blue through green and yellow to red, by the per cent of women with overweight or obesity. There is no colour bar. The caption ends "Red districts have the highest burden."

Figure 5 is a line chart of the same percentage in four age groups over two survey rounds. The four lines are red, green, brown and orange, all thin and of the same thickness. A key in the corner names the age groups in the order youngest to oldest.

List what is wrong with the colour, and how you would fix each problem.

EXERCISE 3 (teaching). A first-year resident says: "In that Tamil Nadu study only about 3 boys in 100 were colour-blind. Why should I redraw my red and green slides for so few people?" You have ten minutes and a whiteboard. Answer her.

EXERCISE 4 (design). Take the figure you are drawing for your build. For each colour in it, write down the job it does: separating groups, standing for a value, or picking out one thing. Remove every colour that has no job. Then print the figure in grey and run it through a colour-vision-deficiency simulator, and write down what you changed after each test.

20 · Words on a figure: axis titles with units, direct labels, and a caption that stands alone

DEFINITION. The words on a figure let a reader understand it without the text of the paper.

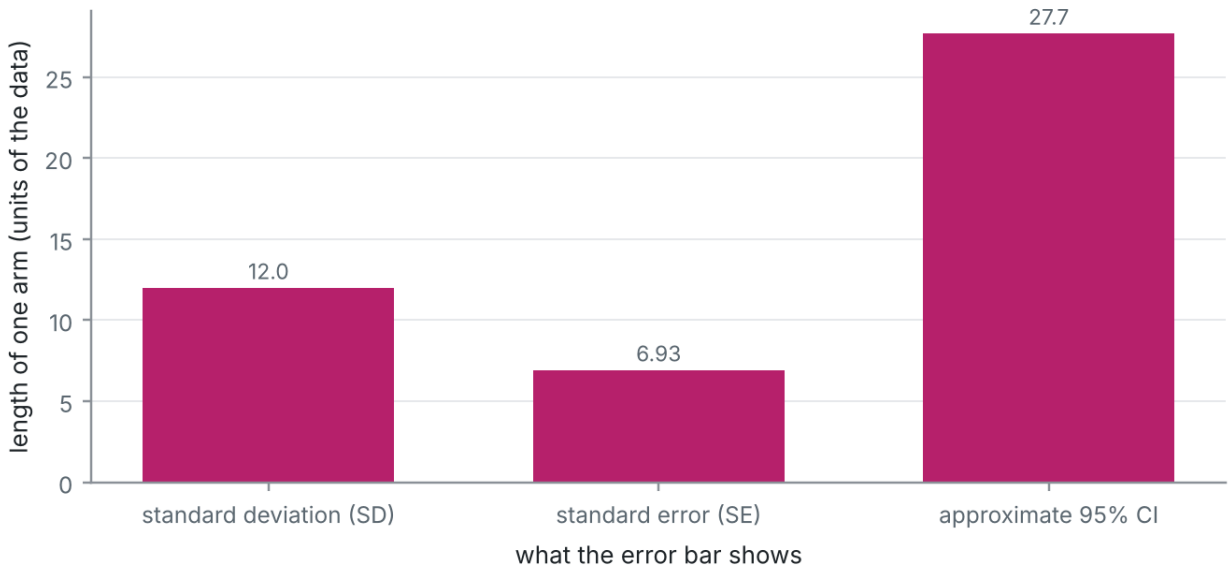
The caption is the text printed with a figure, and journals call it the figure legend. Its first sentence is the figure's title, which states the finding. The title is not text drawn across the figure. A table's title, by contrast, sits above the table. The caption then says what is shown, in whom, where and when; how many independent people or units were measured, written n; what any error bar shows; and where the data came from.

An error bar is a line drawn through a value, usually a mean, to show a range around it. Descriptive error bars show how the data are spread: the range, from the lowest value to the highest, or the standard deviation (SD). Inferential error bars show how precisely the mean is estimated: the standard error of the mean (SE), which is the SD divided by the square root of n, or a 95% confidence interval (CI). A 95% CI is a range worked out from the sample. If the study were repeated many times, 95 per cent of such ranges would contain the true mean. When n is 10 or more, the 95% CI is approximately the mean plus or minus 2 SE.

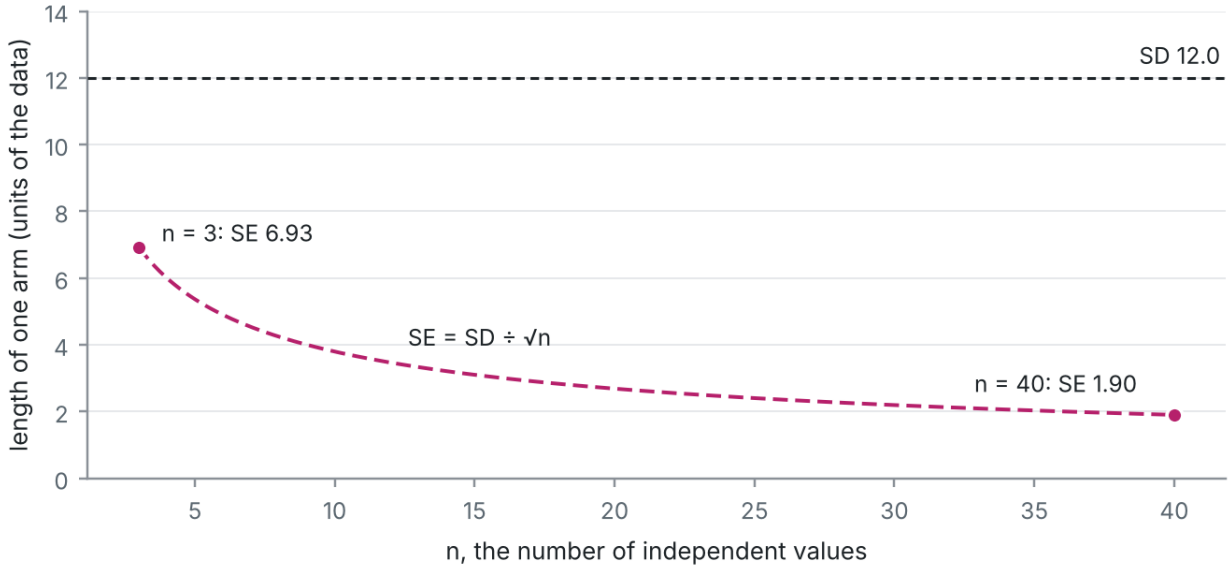
In n, independent means separate people or units. [324, 325, 326, 327, 328, 329, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345]

IN PLAIN TERMS. A reader may look at your figure before reading a word of your paper. So the figure has to explain itself. Each group has its name written beside it.

The same line through a mean can show how spread out the people are, or how sure you are of the mean itself. These are different things, and they look the same. So the caption has to say which one it is, every time.



One set of data, three error bars. Cumming, Fidler and Vaux's three measurements, 28.7, 38.7 and 52.6, have a mean of 40.0. Each bar's height is the length of one arm; the error bar itself would run that far above and below 40.0: 12.0 for the standard deviation, 6.93 for the standard error, and 27.7 for an approximate 95% confidence interval (4 times the standard error, their rule for three values). A reader cannot tell these apart unless the caption names the bar.



The standard error shrinks as n grows; the standard deviation does not. If more values had the same spread as Cumming's three, an SD of 12.0, the SE would be 12.0 divided by the square root of n: 6.93 at n = 3, and 1.90 at n = 40. A short SE bar says the mean is precise, not that the people are alike.

Illustration 1. The claim for the figure of urban and rural adults in the fifth National Family Health Survey (NFHS-5) came in section 14, as the first sentence of its caption. The rest of the words on that figure are still to do. Start from a draft as a spreadsheet leaves it. The draft below is made up for this example; the numbers on it are the real ones.

It has four bars. The side axis, which carries the values (the value axis), runs from 0 to 35 with no title. The bottom axis says "Women" and "Men". A key in the corner says "Series1" and "Series2".

Hold it up and ask what a reader who has only this page can tell. Series1 of what? Per cent of whom, measured how, in which survey?

Fix each element in turn.

element	as the spreadsheet left it (made up)	finished
value axis title	none	adults with body mass index (BMI) 25.0 kg/m^2 or more (%)
bottom axis	Women, Men	women, men; no axis title needed
key	Series1, Series2	none; “urban” and “rural” written beside the bars
values	none	each bar’s per cent printed on it
text size	readable only with the figure filling the screen	readable at its printed width

The value axis title names the quantity and its unit. BMI is in kilograms per square metre (Book 0, B6). “Women” and “men” are categories, so that axis needs no unit and no title. The key goes, because the bars now carry their own names: the direct labelling of section 18. The values go on the bars because the reader will want them. Rougier and colleagues: “If the numeric values are important, they must be provided elsewhere in your article or be written very clearly on the figure.”

For text size, Wilke’s test is to shrink the figure on screen to three to five inches wide. One inch is 2.54 cm.

```
3 times 2.54 = 7.62
5 times 2.54 = 12.7
```

So look at it about 8 to 13 cm wide, or from across the room. If you cannot read the tick labels there, neither can a reader of the printed page.

Now finish the caption. Put the claim from section 14 first, then everything a reader needs to read the bars without your paper:

> *Figure 2. A larger share of urban than of rural adults aged 15–49 had overweight or obesity, in women and in men (NFHS-5, 2019–21). Bars show the per cent of women and of men with a body mass index (BMI) of 25.0 kg/m^2 or more, by place of residence. Women who were pregnant or had given birth in the preceding 2 months are excluded. n: [number of women and of men weighed and measured]. Source: NFHS-5 India Fact Sheet, indicators 88 and 89.*

The bracket is deliberate. The fact sheet says the survey gathered information from 724,115 women and 101,839 men. Those are the people interviewed. It does not say how many were weighed and measured, and that is the n for these bars. So the fact sheet alone cannot finish this caption. Find n in a source that gives it, before the figure goes out.

There are no error bars on this figure, because the fact sheet gives no confidence intervals. Do not draw bars you do not have.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. This figure has four bars and two groups, so every label fits. With dozens of points, label only the ones the text talks about, as Wilke does, and keep the full values for a table.

A caption that names everything makes the figure readable. It does not make the numbers right. If n is wrong, or the exclusion is misstated, the caption carries the error with full confidence.

Illustration 2. Now an error bar. Cumming, Fidler and Vaux, "Error bars in experimental biology" (Journal of Cell Biology, 2007, https://pmc.ncbi.nlm.nih.gov/articles/PMC2064100/), work one small case: three measurements, 28.7, 38.7 and 52.6. Work its bars yourself.

The mean and the standard deviation are Book 0's (Book 0, D5): add the squared deviations, divide by one less than the count, take the square root.

```
28.7 plus 38.7 plus 52.6 = 120.0
120.0 divided by 3 = 40.0
28.7 minus 40.0 = -11.3
38.7 minus 40.0 = -1.3
52.6 minus 40.0 = 12.6
-11.3 times -11.3 = 127.69
-1.3 times -1.3 = 1.69
12.6 times 12.6 = 158.76
127.69 plus 1.69 plus 158.76 = 288.14
288.14 divided by 2 = 144.07
```

The square root of 144.07 is 12.0. So the mean is 40.0 and the SD is 12.0, as the paper reports. The standard error is the SD divided by the square root of n, written √n. Here n is 3, and √3 is 1.732.

```
12.0 divided by 1.732 = 6.93
```

For a 95% confidence interval, Cumming and colleagues give a rule of thumb: when n is 10 or more, double the SE; when n is 3, multiply it by 4. The exact multiplier comes from a table of the t statistic, which is beyond this book.

```
4 times 6.93 = 27.7
```

Now set the three side by side. Each number is the length of one arm of the bar, drawn above and below the mean of 40.0.

what the bar shows	one arm of the bar
standard deviation (SD)	12.0
standard error (SE)	6.93
approximate 95% CI	27.7

The same three numbers can carry a bar four times as long as another. Shown a mean of 40.0 with an arm of 12.0, a reader cannot tell which bar it is. If it is the SD, it describes the people: the

authors note that about two thirds of the data lie within one SD of the mean. If it is the SE or the CI, it describes the mean: how far from the true mean this estimate might be.

So the caption names the bar and gives n, every time. The authors' first two rules say exactly that:

> *Rule 1: when showing error bars, always describe in the figure legends what they are.*

> *Rule 2: the value of n (i.e., the sample size, or the number of independently performed experiments) must be stated in the figure legend.*

Their "figure legends" are what this book calls captions.

Figures quoted in this illustration are from [346].

WHERE THIS PICTURE BREAKS. Three values are a teaching case, not a study. With n as small as 3, Cumming and colleagues advise plotting the individual values instead of any error bar. The doubling rule for a 95% CI holds only when n is 10 or more; at n = 3 the multiplier is about 4, and in between it lies between the two.

The SD stays about the same as n grows, while the SE and the CI shrink. So a short SE bar from a large sample says the mean is precise. It says nothing about how alike the people are.

MUST KNOW POINTS FOR YOU.

- "An error bar shows the error" is the error. Write in every caption whether a bar is the SD, the SE or a 95% CI, and give n beside it.
- Title every measured axis with the quantity and its unit: "body mass index (kg/m^2)", not "BMI".
- Write the caption in a fixed order. The finding comes first. Then what each mark shows; who, where and when; n; what any error bar is; and the source.
- n counts separate people or units, not measurements. Weigh 40 children three times each and you have 120 readings and an n of 40.
- "Legend" means two things. To a journal, the figure legend is the caption: ICMJE asks for "legends for illustrations" on a separate page. On the figure itself, drawing programs call the key to the colours a legend; this book calls it the key. When an author's instructions ask for legends, send captions.
- To show how spread out people are, draw the SD or, better, the points themselves. To show how precisely a mean or a prevalence is known, draw a 95% CI.
- Before a figure leaves you, shrink it to 8 to 13 cm wide on screen and read every label.
- Bars on the same group measured at different times cannot show whether the group changed (Cumming, Fidler and Vaux). For that, work out each person's change, then the mean change and its interval.

COMMON MISREADING. "Short error bars mean the people were much alike." Only an SD bar describes the people. An SE or CI bar describes how precisely the mean is known. It shrinks as n grows, even when the people differ as much as before. A large survey can show a tiny CI around a mean of people who vary widely. Read the caption for which bar it is before you read its length.

REPORTING IT. Points show group means and error bars show 95% confidence intervals. Each child counts once in n, however often measured.

EXERCISE 1 (retrieval). From memory: list what a figure caption should say, in order. Then name the two error bars that describe the data and the two that describe how precisely the mean is known.

EXERCISE 2 (interpretation). The numbers here are made up. A figure shows mean waist circumference in two groups, 92 cm and 88 cm. Each mean has an error bar whose arm is 3 cm long. The caption uses "standard error of the mean (SEM)", another name for the SE, and the sign ±, read "plus or minus". It ends: "Values are mean ± SEM; n = 40 per group."

What does a 3 cm arm tell you about how much waist circumference varies among the people? What does it tell you about how precisely each mean is known?

EXERCISE 3 (critique). A draft thesis figure, made up for this exercise, has a value axis titled "BMI", no title on the bottom axis, and this caption:

> *Figure 3. Graph showing BMI of participants in the two groups. Error bars represent variation. *p<0.05.*

The asterisk marks a comparison whose P value was below, "<", 0.05.

List everything a reader cannot tell from it, and rewrite what you can.

EXERCISE 4 (teaching). A first-year resident shows you two slides made from the same data. On one, the error bars are long; on the other, short. She asks which slide is right. You have ten minutes and a whiteboard.

EXERCISE 5 (design). Take the figure you are drawing for your build. Write its axis titles with units. Replace any key with direct labels where you can. Write its full caption in the order this section gives. Then cover the rest of your text. Ask a colleague to tell you, from the figure and its caption alone, what was measured, in whom, how many, and what the finding is.

21 · Making a table that stands alone

DEFINITION. A table stands alone when a reader can understand it without going back to the text. Its title is short but self-explanatory, and sits above the table: what was measured, in whom, where and when.

Each column has a short heading with the unit. Explanations, exclusions and nonstandard abbreviations (check the target journal's instructions for authors) go in footnotes below the table, marked with letters or symbols, not in the headings.

Each heading is aligned like the column under it. [347, 348, 349, 350, 230, 32, 32, 31, 351, 352, 353, 354, 355, 356, 357, 358, 358]

IN PLAIN TERMS. A table is for the reader who wants exact numbers, or wants to find one row. It has to make sense on its own, because a reader may look at the tables before reading the text.

Keep it plain. No grid of lines, no lines between rows. Numbers line up on the right and keep the same number of decimal places.

Illustration 1. Start with a published table and test it. Krishnamurthy and colleagues' paper on colour-vision deficiency in school boys, met in section 19 (Indian Journal of Ophthalmology, 2021, https://pmc.ncbi.nlm.nih.gov/articles/PMC8482944/), prints this as its first table. Its title:

> *Association between color vision deficiency, age groups, location and type of schools*

Here are its first four columns. A fifth, of odds ratios and P values, is left out.

Demographics	Total (n)	Color vision deficiency (n)	Prevalence (%)
Total	74986	2073	2.76
Class groups			
Middle	58907	1608	2.73
Higher secondary	16079	465	2.89
Type of school			
Private	21422	535	2.50
Public	53564	1538	2.87
Location of school			
Rural	21911	392	1.79
Urban	53075	1681	3.17

Cover the paper and ask Book 0's four questions of the table alone (Book 0, F1). What does a row stand for? A group of children, but which children? The title does not say they are boys, aged 11 to 17, in schools of one district of Tamil Nadu. It says "age groups", while the rows say "Class groups", and nothing gives the ages. A reader holding only the table cannot say whom it describes. It fails the first test of a table that stands alone.

Now see what it does well. Every percentage has its two counts beside it. The unit sits in the heading, "(%)", not in every cell. The per cents all carry two decimal places. By the rule in section 9, one decimal place would be enough for values under 10 per cent. But the column is consistent, and consistency is what lets the eye compare down it.

Last, hold the table against the paper's text. The Results say there were:

> *21422 (28.57%) and 53564 (71.43%) in public and private schools, respectively*

The table puts 21422 against Private and 53564 against Public. One of the two is wrong. The text's own percentages, 2.87 for public schools and 2.50 for private, match the table's labels. So the slip is probably in the sentence. Either way, a reader holding both cannot be sure. This is one reason not to give the same numbers twice.

Rewrite the title so the table stands alone:

> *Table 1. Colour-vision deficiency confirmed in boys aged 11–17 years screened in schools of Kanchipuram district, Tamil Nadu, by school level, type and location*

Then add a footnote for what the headings cannot carry: "Middle school: classes 6–10, ages 11–15 years; higher secondary: classes 11–12, ages 16–17 years."

Figures quoted in this illustration are from [323].

WHERE THIS PICTURE BREAKS. This critique is of the table's words and layout. It says nothing about whether the counts are right; for that you would need the study's data. A table can stand alone perfectly and still carry a wrong number.

The fifth column, of odds ratios and P values, is left out because this book does not teach odds ratios. Its heading, "Odds ratio (95%Confidence Interval) (P)", packs three things into one heading. That is the explanatory matter the International Committee of Medical Journal Editors (ICMJE) asks you to put in footnotes.

Illustration 2. Now build one. Say you are writing a report for district programme officers. Each will set their own district's figures beside the national ones, so they need the exact values. That is a job for a table. For the same rows, section 14 chose a figure when the claim was a pattern, urban above rural. A paper uses one or the other for these numbers, never both.

Take the National Family Health Survey (NFHS-5) India Fact Sheet, indicators 88 and 89. Build the table in five steps.

1. Write the title above the table: what, in whom, where, when. Spell out the survey's name once.

2. Give each column a short heading with its unit, "(%)". Put the survey round in the heading, not in a sentence above.
3. Keep the decimals the same all the way down. The fact sheet prints women's NFHS-5 total as 24.0. A spreadsheet will try to show it as 24; type the zero back.
4. Move everything else into footnotes. The fact sheet's own footnote to the women's row says pregnant women and women who gave birth in the preceding 2 months are excluded. That goes in a footnote, marked a. Abbreviations are spelled out there too.
5. Add a source line, and say what the source does not give.

The result:

> *Table 1. Women and men aged 15–49 years with overweight or obesity (body mass index 25.0 kg/m^2 or more), India, National Family Health Survey 2019–21 (NFHS-5), by residence, with the 2015–16 (NFHS-4) total*

group	NFHS-5 urban (%)	NFHS-5 rural (%)	NFHS-5 total (%)	NFHS-4 total (%)
women (a)	33.2	19.7	24.0	20.6
men	29.8	19.3	22.9	18.9

> *(a) Excludes pregnant women and women who gave birth in the preceding 2 months. Counts of women and men weighed and measured are not given in the source; NFHS-5 interviewed 724,115 women and 101,839 men. Source: NFHS-5 India Fact Sheet, indicators 88 and 89 [reference number].*

Now the lines and the alignment, which the page here cannot show. In your word processor, draw one line under the headings and one above and below the whole table. Delete every other line. Wilke's verdict on his word processor's default table: "if you pick the default, you will end up with a poorly formatted table every time." Align "women" and "men" left. Align every number column right, so the decimal points line up, and align each heading like its column.

Last, cite it in the text, and say only the most important thing the table shows: "Overweight or obesity was more common in urban than in rural India, in women and in men (Table 1)." Do not walk the reader through all eight numbers.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. Here the source gave no counts, so the table says so instead of inventing them. With your own data you have the counts, and they go beside each percentage, as section 9 showed: "1608 of 58907 (2.7)" or a separate column for n.

This table has two rows. With every state and union territory as a row, it is still a table if the reader will look up a state. If the point is the pattern across states, ICMJE offers a graph as the alternative to a table with many entries.

The fact sheet is a public government report. ICMJE asks you to acknowledge such a source fully and to obtain permission for data from another source. Whether a public report needs written permission depends on its terms, which the next rung takes up. Unpublished data, such as a colleague's clinic register, needs the owner's permission before you use it at all.

MUST KNOW POINTS FOR YOU.

- "Put the numbers in a table and in a figure, to be safe" is the error. Choose one home for each result. The text then states the most important observation and cites it.
- Write the title above the table, short and self-explanatory: what was measured, in whom, where and when.
- Keep headings short, with the unit in brackets.
- Give every percentage its count, and name every measure of variation. A column of "mean ± value" (± is read "plus or minus") is unfinished until its heading or a footnote names the value.
- When your source gives no counts, as the NFHS fact sheet does not, say so in a footnote.
- Delete the word processor's default grid. No vertical lines, no lines between rows; a line under the headings and one above and below the table. Align numbers right with the same decimal places, and text left.
- "Many rows, so a table" is a half-truth. A table is for exact values and for a reader who looks up rows. For a pattern among many values, ICMJE offers a graph instead. In this book's advice, two or three numbers belong in a sentence.
- Keep the same decimal places down a column of like values, such as the per cents in one table: that is Wilke's rule. Cole adds a warning. A column of ratios or P values may need more digits in some rows than in others, by the rules of section 9. Give each value the digits it needs, and line the column up on the decimal point. When a column's decimal places are the same, right alignment lines up the decimal points.
- Acknowledge every table's outside data in a source line, every time. For published data, the source's licence decides whether you also need permission (next rung). For unpublished data, such as a colleague's register, always get the owner's permission.

COMMON MISREADING. "A good results section walks the reader through the table." It does the opposite. ICMJE asks you not to repeat all of a table's data in the text. State only the most important observations, and cite the table. A paragraph that reads out every cell makes the reader check two copies of each number. And, as Krishnamurthy and colleagues' paper shows, two copies can disagree.

REPORTING IT. Overweight or obesity was more common in urban than in rural India, in women and in men (Table 1).

EXERCISE 1 (retrieval). From memory: what must a table's title do, where do explanations and abbreviations go, and what are the rules for lines and for aligning numbers?

EXERCISE 2 (critique). A resident's thesis draft has the table below, made from the real NFHS-5 India Fact Sheet numbers for anaemia; the layout is made up for this exercise. Haemoglobin,

written Hb, is the oxygen-carrying protein of the blood, measured in grams per decilitre (g/dl; a decilitre is a tenth of a litre). The table has a border round every cell, and every number is centred.

The table has no title. Its first column heading reads “Prevalence of anaemia among non pregnant women (Hb<12) and men (Hb<13) as per NFHS 5 survey conducted in 2019-21 in India (capillary blood)”; < is read “below”. Its rows are:

	Urban	Rural	Total
Women	54.1%	58.7	57.2
Men	20.4%	27.4	25

Under the table, the draft reads out all six numbers in a sentence: “Anaemia was 54.1% in urban women, 58.7% in rural women …” and so on. A bar chart of the same six numbers follows. List what you would change.

EXERCISE 3 (interpretation). Krishnamurthy and colleagues write that the 74,986 boys they screened had a “mean age: 13.75 ± 1.91”, without saying what 1.91 is. Can you tell whether it is a standard deviation or a standard error? Show your reasoning. Then say how the authors should have written it in a table.

EXERCISE 4 (teaching). A first-year resident brings you a thesis table in her word processor’s default style. It has a grid of lines round every cell, centred numbers, and a heading three lines long. She says the grid “makes it easier to read”. You have ten minutes. Answer her.

EXERCISE 5 (design). Take one result from your build section that belongs in a table: exact values a reader will look up. Lay it out as a table that stands alone. It needs a title above, short headings with units, counts beside percentages, the measure of variation named, footnotes and a source line. Then write the one sentence in your text that cites it.

22 · Making the plot: from a table of numbers to a finished figure

DEFINITION. Making a figure is a sequence of decisions, taken in a fixed order, which any program can be made to carry out. The first four decide the figure's content. They are the claim, the mark that suits the shape of the data, the range of the value axis (the side axis that carries the numbers), and the words on it, with each axis titled and given its unit.

A default is a setting a program uses when you have not chosen one. Defaults are made to suit any plot, so they suit no particular plot well.

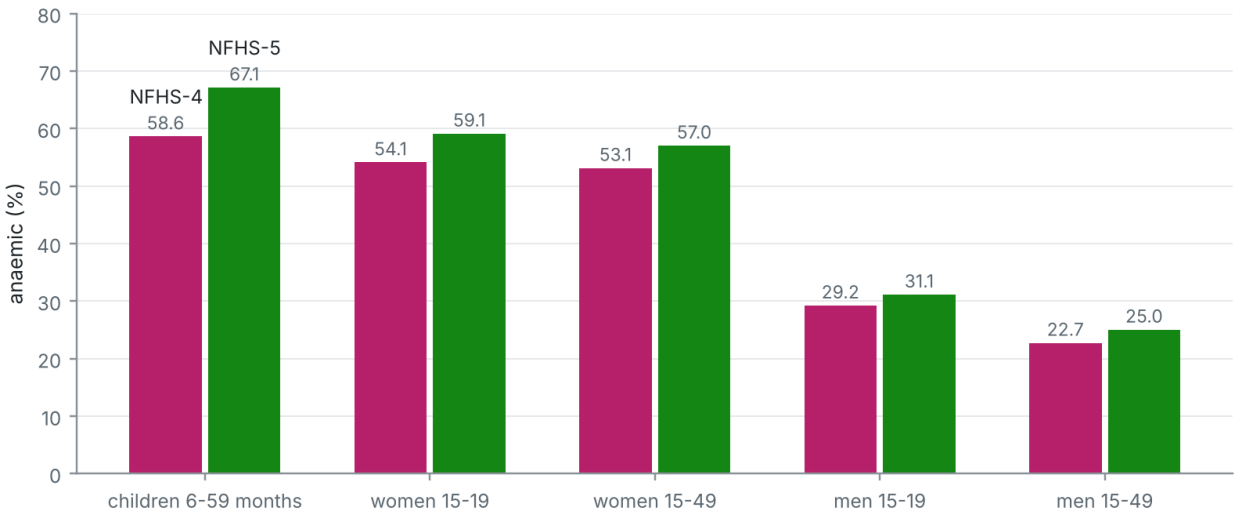
In this book, a figure sheet is the written list of the decisions for one figure, kept beside the data.

An image is saved either as a bitmap, a grid of coloured dots called pixels, or as a vector graphic, which stores the shapes themselves and is redrawn at any size without losing sharpness. A png or jpeg file is a bitmap; a pdf can hold a vector graphic. [359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380]

IN PLAIN TERMS. A program will draw a chart from your numbers in one click. That chart is the program's guess, made for every chart at once.

So decide first, then draw. Write your decisions down in order.

Keep that list. When a number changes, or a reviewer asks for a change, you can draw the figure again in minutes and get the same result.



Between NFHS-4 (2015-16) and NFHS-5 (2019-21), the share with anaemia rose in each of the five groups shown, by 1.9 to 8.5 percentage points. Bars show the per cent anaemic, by each group's own haemoglobin cut-off; ages in years. Counts are not given in the source. Source: NFHS-5 India Fact Sheet, indicators 92 and 95 to 98.

ILLUSTRATION. Here is the trap with the chart a program offers first. It looks finished.

Open the *National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet* at https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf. Search for anaemic. The fact sheet counts a person as anaemic when their haemoglobin, measured in grams per decilitre (g/dl), is below a cut-off for their group. For women it is 12.0 g/dl, or 11.0 if pregnant (indicators 93 and 94). Five of its rows are these ([22] marks a footnote).

> *1. Children age 6-59 months who are anaemic (<11.0 g/dl)[22] (%) 64.2 68.3 67.1 58.6*

> *1. All women age 15-49 years who are anaemic[22] (%) 53.8 58.5 57.0 53.1*

> *1. All women age 15-19 years who are anaemic[22] (%) 56.5 60.2 59.1 54.1*

> *1. Men age 15-49 years who are anaemic (<13.0 g/dl)[22] (%) 20.4 27.4 25.0 22.7*

> *1. Men age 15-19 years who are anaemic (<13.0 g/dl)[22] (%) 25.0 33.9 31.1 29.2*

The four columns are NFHS-5 (2019-21) urban, rural and total, then NFHS-4 (2015-16) total, as in the overweight rows you have used all through this book. Take the two totals for each group into a table of your own.

group	NFHS-4, 2015-16 (%)	NFHS-5, 2019-21 (%)
children 6-59 months	58.6	67.1
women 15-19	54.1	59.1
women 15-49	53.1	57.0
men 15-19	29.2	31.1
men 15-49	22.7	25.0

Now suppose you select the table and ask your program for a column chart.

It has a title across the top, “Chart 1”. A key in a box says “Series1” and “Series2”. The value axis starts at 20. The text is small and grey. The bars are red and green. A dark grid line runs every 5 points. It saves as a jpeg.

Do not start fixing what you see.

1. The claim. Cover the chart and write the one sentence it should make. Here: between NFHS-4 and NFHS-5, the share with anaemia rose in every group.

Keep every row, even one that would weaken the claim. Here none does: indicators 93 and 94, for non-pregnant and pregnant women, rose too. The fact sheet prints no confidence intervals. So it cannot show whether the smallest rises are more than sampling variation.

2. The mark. The shape of the data is amounts for a set of groups, at two times. That takes bars, or dots where the bars would end, on one shared axis (section 15). Put each group's two bars side by side, because the claim compares the two rounds within a group.

3. The axis. Bars start at zero, so the starting point of 20 goes.

4. The words. Title the value axis with the quantity and its unit: "anaemic (%)". Delete "Chart 1": the finding goes in the caption, not across the drawing.

5. The defaults. Go through the rest one at a time and ask of each what it tells the reader (section 18). The dark grid goes, or becomes faint. Make the small grey text larger and darker.

6. Colour. Its job here is to separate two rounds. Red against green fails in grey and for a reader with colour-vision deficiency (section 19). Choose two colours that differ clearly in lightness. Where you cannot, keep the order fixed and name each round on the figure.

7. The caption. Finding first, then what the bars are, who, when, and the source (section 20).

8. Saving. Read the journal's instructions on file type and size first. Never use a jpeg for a chart: Wilke says to avoid it for "images containing line drawings or text".

9. The checks. Shrink the figure on screen to about 8 to 13 cm wide, the size of one column of a printed page, and read every label (section 20).

Here is the figure sheet you have written. It records the data as well as the nine decisions.

step	decision for this figure
claim	anaemia rose between NFHS-4 and NFHS-5 in every group the fact sheet reports
data	five rows of the NFHS-5 India fact sheet, NFHS-4 and NFHS-5 totals; pregnant and non-pregnant women left out, contained in all women
mark	bars, the two rounds side by side within each group
axis	value axis from 0 to 80, titled "anaemic (%)"
words	groups under the bars with ages in years; no title on the drawing
defaults changed	no key; faint grid; larger, darker text
colour	this book's two colours merge in grey, so NFHS-4 always on the left and each round named over the first pair
caption	finding first; what the bars show; ages; no counts in the source; source and indicator numbers
saving	the journal's format and size; otherwise pdf, and png where a bitmap is required
checks	read at 8 cm wide; in grey; through a simulator

The figure drawn from this sheet is the one at the start of this section. That is the point: any program that will take these decisions can draw it again.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. The sheet makes the figure repeatable only as far as you wrote it down. A change you make by dragging a label with the mouse and forget to record is lost the next time.

Each group has its own cut-off for haemoglobin, so the bars compare shares, not haemoglobin levels.

MUST KNOW POINTS FOR YOU.

- Write the figure's decisions down before you open the program, in order: claim, mark, axis, words, defaults, colour, caption, saving, checks. If you start by fixing what the first chart shows, you fix what you happen to notice.
- "The program drew it, so it is a neutral picture of the numbers" is the error. Every setting you did not choose was chosen for any chart at all. Check each one, the axis start first.
- A spreadsheet will do only if you change nearly every default. If a spreadsheet is what you have, keep a figure sheet beside the file, and record every change as you make it.
- Never finish a figure by hand in a drawing program and leave no record. When a reviewer asks for one changed number, every hand edit has to be done again, and some will be forgotten.
- Read the journal's instructions for figures before you save. Where they leave it to you, save a vector graphic such as a pdf for print. Use a png where a bitmap is required, and never a jpeg for a chart.
- Check the figure at the size it will be printed, not at the size of your screen. Shrink it to one column's width, then view it in grey and through a simulator.
- When a trainee shows you a chart, ask for the figure sheet before you look at the chart.
- The procedure makes sure each decision is taken on purpose. A perfectly made figure of a wrong table is a wrong figure, so check the data against their source before step 1.

COMMON MISREADING. "Wilke says Excel is not recommended, so a figure made in a spreadsheet is wrong." A figure is right or wrong by what it shows: its claim, its mark, its axis and its words. Wilke's objection is that a figure made by hand cannot be made again exactly. In his own words, "The best visualization software is the one that allows you to make the figures you need." A spreadsheet figure made from a written figure sheet, with every default checked, can be a

good figure. It is harder to repeat than one drawn by code. Drawing figures by code is taught in the course's book on R and reproducibility (S52).

REPORTING IT. Figures were drawn in [program and version] from the data in [table or file]. The decisions for each figure, including axis ranges and colours, are listed in [supplementary file].

EXERCISE 1 (retrieval). From memory: list the nine decisions for making a figure, in order. Which four decide what the figure shows, and which come after?

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (critique). A colleague hands you the figure sheet below for a figure in her thesis. It is invented. Say what is wrong with each line you would change, and what you would write instead.

step	decision
claim	anaemia in NFHS-5
mark	3-D columns, the program's first offer
axis	the program's own range, which starts at 50
words	title on top: "Anaemia"; axis untitled
colour	red for women, green for men
saving	jpeg, so the file is small

EXERCISE 3 (design). Take the figure you are drawing for your build section. Write its figure sheet, the data and all nine decisions, before you open any program. Then draw it, and note every place where you had to change something the sheet did not mention. Add those to the sheet.

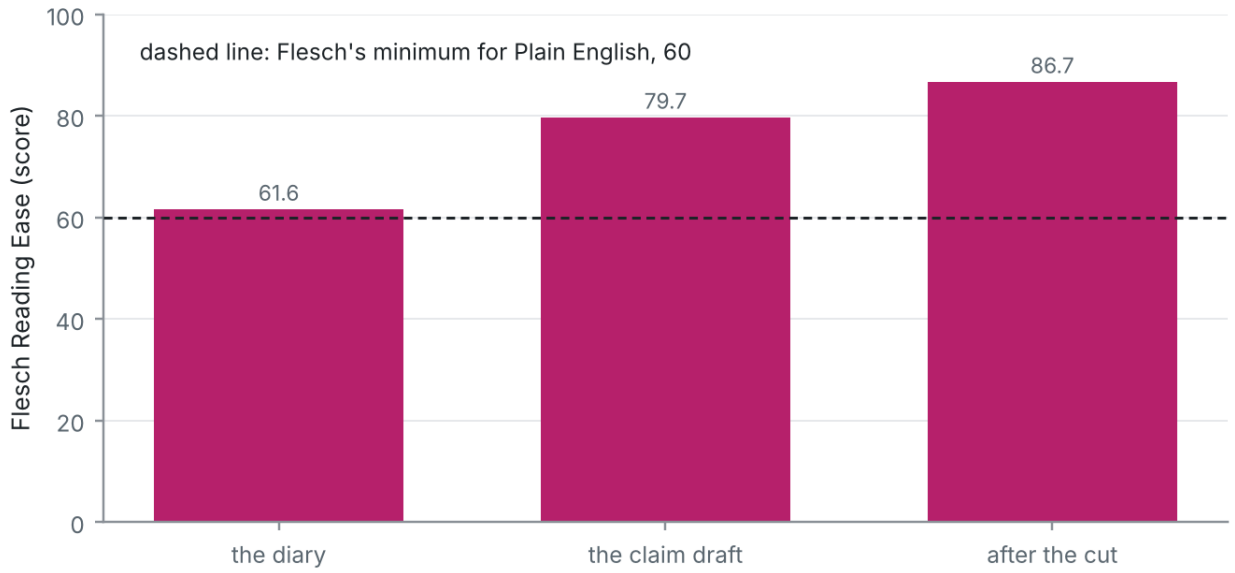
EXERCISE 4 (teaching). A first-year resident says: "I only have a spreadsheet, and my guide says it is fine. Why should I bother with all these steps?" Ten minutes, with a whiteboard.

23 · Worked journey: One finding, from the fact sheet to a paragraph a colleague can restate

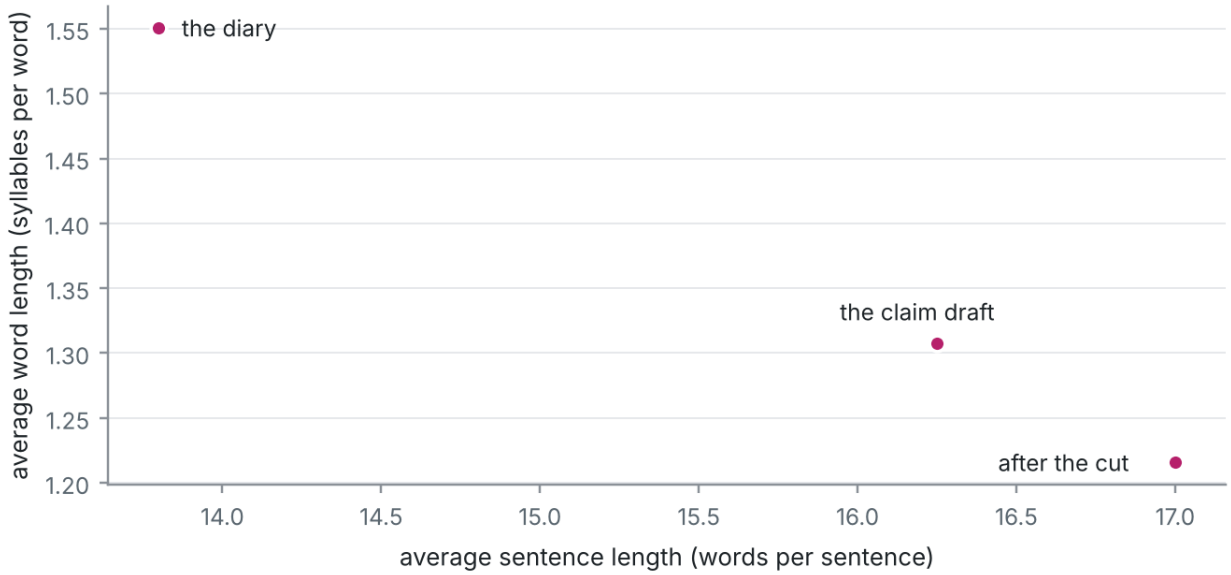
DEFINITION. In this book, a finding's journey is the order of work that takes one result from its source to a results paragraph a reader can restate after one reading.

1. The question, and the data read from the source with every limit the source states.
2. The one message: the claim in one or two sentences, false if the data had come out otherwise.
3. The results paragraph, rewritten from an account of what was done into the claim and its evidence.
4. The numbers: units, counts where the source gives them, differences in percentage points, and no more digits than the data hold.
5. A readability score taken as a trip-wire, and a cut.
6. One home for the numbers, chosen by what the reader has to do with them. The paper keeps one of the two, never both.
7. A numbered reference to the source, formatted from the rules for that kind of source.
8. A one-pass test with a reader who does not know the work, and a revision.
9. A statement of what the finding does not mean. [381, 382, 383, 384, 385, 385, 386, 387, 359, 388, 389, 390, 390, 391, 392, 393, 394, 395, 396]

IN PLAIN TERMS. Each earlier section taught one tool. Real work uses them one after another, on one result. The result is the share of Indian women and men aged 15 to 49 with overweight or obesity in two national surveys.



The worst draft passes: the diary scores 61.6 in Flesch Reading Ease, above Flesch's minimum of 60 for Plain English (dashed line). The claim draft scores 79.7, and the cut version 86.7. Each figure in a draft is counted as one word of one syllable.



The cut scored higher than the diary although its sentences were longer: 17 words a sentence against 13.8. Its words were shorter, 1.2157 syllables a word against 1.5507. The long words went, and the score rose. These are points, so neither axis starts at zero.

Illustration 1. Suppose you are writing a short paper, a research letter, on one question: did the share of Indian women and men aged 15 to 49 with overweight or obesity change between the fourth and fifth rounds of the National Family Health Survey (NFHS)?

The data. Open the *National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet* at https://dhs-program.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf. Find the heading "Nutritional Status of Adults (age 15-49 years)". Two rows under it carry your question.

> *88. Women who are overweight or obese (BMI ≥25.0 kg/m2)21 (%) 33.2 19.7 24.0 20.6*

> *89. Men who are overweight or obese (BMI ≥25.0 kg/m2) (%) 29.8 19.3 22.9 18.9*

Read them as section 14 did. The four columns are NFHS-5 (2019-21) urban, rural and total, then NFHS-4 (2015-16) total, in per cent. The rows count a body mass index (BMI) of 25.0 kg/m^2 or more ("kg/m2" on the sheet), overweight and obese together. The "21" after the women's unit is the fact sheet's footnote 21. It excludes pregnant women and women with a birth in the preceding 2 months. The fact sheet prints no counts of the people weighed and measured.

Write those limits down now, before you write anything else. Every later step has to stay inside them.

The one message. Keep the message section 3 wrote for these rows.

> *Between 2015-16 and 2019-21, the share of Indian adults aged 15-49 with overweight or obesity rose in women and in men, to about one woman in four and more than one man in five.*

Test it once more. Had either share fallen, the sentence would be false, so it claims something. "About one woman in four" stands for 24.0 per cent, close to 25. "More than one man in five" stands for 22.9 per cent, above 20. Put the sentence at the top of a page and keep it in view. Every step below is checked against it.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. This question is answered by reading two columns, so the gap a real paper fills is thin here. The order of work is the same when the data are your own; then the limits come from your methods, not from a footnote.

Illustration 2. The paragraph. Here is a first draft of the results, of the kind a thesis is full of. It is invented; its numbers are the fact sheet's.

> *Data on the nutritional status of adults were taken from the NFHS-5 India fact sheet. The proportion of overweight and obese individuals was examined in both sexes. The NFHS-4 figures were also noted for the purpose of comparison. It was observed that the proportion was 24.0% in females and 22.9% in males in NFHS-5, as compared to 20.6% and 18.9% respectively in NFHS-4. Thus an increasing trend was seen.*

Mark what is wrong before you fix it. The first three sentences are a diary: taken, examined, noted (section 2). The claim arrives in the last five words, and "trend" claims more than two surveys can show. "Females" and "males" drift from the fact sheet's women and men (section 8). "Respectively" makes the reader pair four numbers by position. The age range and the cut-off are gone.

Put the claim first, then the evidence it needs. Work out the two differences, as section 9 taught, in percentage points.

```
24.0 minus 20.6 = 3.4
22.9 minus 18.9 = 4.0
```

Here is the claim version, as first drafted. It is invented too.

> *It can be noted that the share of adults aged 15-49 who were overweight or obese rose between NFHS-4 (2015-16) and NFHS-5 (2019-21), in women and in men. In women it rose from 20.6 to 24.0 per cent, by 3.4 percentage points. In men it rose from 18.9 to 22.9 per cent, by 4.0 percentage points. This clearly shows that obesity is increasing in India.*

The numbers. Three checks. Units: every figure says per cent or percentage points. Counts: the source gives none, so you cannot put a count before each percentage. Say so where the numbers first appear, in the table's footnote. Never make a count by multiplying the 724,115 women NFHS-5 interviewed by a percentage: they are not the women weighed for this row. Digits: the four totals run from 18.9 to 24.0.

```
24.0 minus 18.9 = 5.1
```

Cole's rule for comparing groups gives one decimal place for a range between 1 and 10. Keep the fact sheet's one decimal.

The score. Now score the two drafts for Flesch Reading Ease (FRE) and Flesch-Kincaid grade level (FKGL), from the average sentence length (ASL) and average syllables per word (ASW), as section 11 did, taking each figure as one word of one syllable. The diary has 69 words in 5 sentences. Its words of more than one syllable add 38 beats: Data 1, nutritional 3, status 1, adults 1, taken 1, India 2, proportion 2 twice, overweight 2, obese 1, individuals 4, examined 2, sexes 1, figures 1, also 1, noted 1, purpose 1, comparison 3, observed 1, females 1, compared 1, respectively 3, increasing 2. The claim draft has 65 words in 4 sentences, with 20 extra beats: noted 1, adults 1, overweight 2, obese 1, between 1, women 1 twice, percentage 2 twice, clearly 1, obesity 3, increasing 2, India 2.

```
69 plus 38 = 107
69 divided by 5 = 13.8
107 divided by 69 = 1.550724638
206.835 minus 1.015 times 13.8 minus 84.6 times 1.5507 = 61.64
0.39 times 13.8 plus 11.8 times 1.5507 minus 15.59 = 8.09
65 plus 20 = 85
65 divided by 4 = 16.25
85 divided by 65 = 1.307692308
206.835 minus 1.015 times 16.25 minus 84.6 times 1.3077 = 79.71
0.39 times 16.25 plus 11.8 times 1.3077 minus 15.59 = 6.18
```

Look at the first score. The diary scores 61.6, above Flesch's minimum of 60 for Plain English. By the number, it passes. By reading, it fails: no claim until the end, and the wrong claim when it comes. The score counts the length of words and sentences. It cannot see where the claim is, so it could not catch this draft.

The cut. Cut the claim draft in section 12's order, largest unit first. The last sentence restates the paragraph, and widens it: "obesity" alone, "India" with no age range, and "increasing" as if the rise went on. Delete it. Then the empty opener, "It can be noted that". What is left has 51 words in 3 sentences, with 11 extra beats: adults 1, overweight 2, obese 1, between 1, women 1 twice, percentage 2 twice.

```
51 plus 11 = 62
51 divided by 3 = 17
62 divided by 51 = 1.215686275
206.835 minus 1.015 times 17 minus 84.6 times 1.2157 = 86.73
0.39 times 17 plus 11.8 times 1.2157 minus 15.59 = 5.39
```

draft	words	sentences	syllables	ASL	ASW	FRE	FKGL
the diary	69	5	107	13.8	1.5507	61.6	8.1
the claim draft	65	4	85	16.25	1.3077	79.7	6.2
after the cut	51	3	62	17	1.2157	86.7	5.4

The cut version's sentences are longer on average than the diary's, and it scores higher. The long words went with the diary, and so did its missing claim. Keep the cut version for now.

Figures quoted in this illustration are from [24, 204].

WHERE THIS PICTURE BREAKS. Three short drafts are too small a sample for the score to mean much; section 11 asks for passages of about 100 words. The point here is the direction of the error: a bad paragraph can pass, so the score is never the reason to keep one.

The syllable counts are judgements. If you say "India" in two beats, the diary scores a little higher still; the comparison does not change.

Illustration 3. The table. The paper's readers include programme officers who will set their own state's figures beside the national ones. They need exact values, including urban and rural. Build the table as section 21 did: title above, short headings with the unit, a footnote for the exclusion and for the missing counts, a source line with its reference number.

> *Table 1. Women and men aged 15–49 years with overweight or obesity (body mass index 25.0 kg/m^2 or more), India, National Family Health Survey 2019–21 (NFHS-5), by residence, with the 2015–16 (NFHS-4) total*

group	NFHS-5 urban (%)	NFHS-5 rural (%)	NFHS-5 total (%)	NFHS-4 total (%)
women (a)	33.2	19.7	24.0	20.6
men	29.8	19.3	22.9	18.9

> *(a) Excludes pregnant women and women who gave birth in the preceding 2 months. Counts of women and men weighed and measured are not given in the source. Source: NFHS-5 India Fact Sheet, indicators 88 and 89 (1).*

Check the digits again, because the table holds eight numbers, not four. They run from 18.9 to 33.2.

```
33.2 minus 18.9 = 14.3
```

A range of 14.3 is 10 or more, so Cole's range rule alone would print whole numbers. Then the women's totals read 21 and 24, and the paragraph's rise of 3.4 points could not be read off the table. Cole says his rules are "recommendations not requirements". Keep one decimal place in the table and the text alike, so the two never disagree.

The figure. Now make a figure of the claim, by the steps of section 22. The claim is a rise in each sex. The shape of the data is amounts for two groups at two times, so the mark is bars, each sex's two rounds side by side.

The axis comes next. Suppose your program starts the value axis at 18, just below the smallest value. Work out the Lie Factor (LF) of section 17 for each pair, with the smaller bar, a ÷ (a − s).

```
20.6 minus 18 = 2.6
20.6 divided by 2.6 = 7.923076923
18.9 minus 18 = 0.9
18.9 divided by 0.9 = 21
```

The program's chart drew the women's rise about 7.9 times its true size and the men's 21 times. Tufte counts anything above 1.05 as substantial distortion, as section 17 showed. Start the bars at zero, and the Lie Factor is 1 for both.

Then the words: the value axis titled "overweight or obese (%)", the groups named under the bars, the rounds named with their years. Colour separates the two rounds, but should not carry the difference alone: keep NFHS-4 on the left in each pair, and print each bar's value on it. Drawn this way, from zero, it is the figure of these four values that section 17 drew as bars from zero, so it is not drawn again here. Its caption, finding first:

> *Between NFHS-4 (2015-16) and NFHS-5 (2019-21), the share of adults aged 15-49 in India with overweight or obesity rose in women and in men, by 3.4 and 4.0 percentage points. Bars from zero show the per cent with a body mass index of 25.0 kg/m^2 or more. Counts are not given in the source. Source: NFHS-5 India Fact Sheet, indicators 88 and 89.*

Shrink it to one column's width and read it; view it in grey.

One home. You now hold a table and a figure of the same numbers, and the International Committee of Medical Journal Editors (ICMJE) Recommendations say "do not duplicate data in graphs and tables". Choose by what the reader has to do (section 14). The claim rests on four totals, which the paragraph can carry. The officers need eight exact values, which only the table gives. The figure adds nothing a paper reader needs, so the paper keeps the table. The figure is not wasted: it goes to your talk, made again for a slide, with bigger text and fewer words.

The table now holds the levels, so the paragraph need not read them out. ICMJE asks the text to "emphasize or summarize only the most important observations". Keep the claim and the two rises, and cite the table.

WHERE THIS PICTURE BREAKS. The choice of table over figure follows from these readers and these eight numbers. With every state as a row and a pattern to show, such as urban above rural almost everywhere, the figure would win and the table would go to a supplement.

The Lie Factor checks only bars drawn from an axis start. It says nothing about whether a rise of 3.4 points matters for health.

Illustration 4. The reference. The table's source line and the paragraph cite the fact sheet as reference (1), the first source the text mentions (section 5). Now write entry 1 in National Library of Medicine (NLM) format.

The NLM pattern in section 5 is for a journal article. The fact sheet is not one, and Citing Medicine's chapter on journals says "Journal references omit information on place of publication and publisher, whereas book references carry these details." A short report read as a PDF on the Internet is cited under chapter 22, "Books and Other Individual Titles on the Internet", at https://

www.ncbi.nlm.nih.gov/books/NBK7269/. The chapter says such a title "may also be smaller works such as a brochure, single-page fact sheet, or brief treatise".

NLM prints the chapter's general format as a picture, and this book does not hold it. Its examples are text. Example 7, a title on the Internet with an organisation as author, is the nearest to the fact sheet:

> *Kaiser Commission on Medicaid and the Uninsured. The uninsured and their access to health care [Internet]. Washington: Henry J. Kaiser Family Foundation; 2006 Oct [cited 2006 Nov 3]. 2 p. Available from: http://www.kff.org/uninsured/upload/The-Uninsured-and-Their-Access-to-Health-Care-Oct-2004.pdf*

Read its parts in order: the organisation as author; the title, with a capital only on its first word and on proper nouns, then "[Internet]"; the place and the publisher; the date of publication; the date the citer saw it, "[cited ...]"; the number of pages; "Available from:" and the full address.

Now collect every element the fact sheet itself prints, as the cover gives it:

element	as the fact sheet prints it
issued by	Ministry of Health and Family Welfare
prepared by	International Institute for Population Sciences (Deemed University)
title	National Family Health Survey - 5, 2019-21, India Fact Sheet
where you read it	https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf
when you read it	the date you opened it, from your own notes

Set them in example 7's order. The organisation that prepared it is the author. The chapter defines the publisher: "A publisher is defined as the individual or organization issuing the book". Two elements are not on the cover, the place and the year of publication, so they stay as marked gaps until you find them in the full document. Here is entry 1, modelled on example 7; the date cited is the day this book read the fact sheet, so give your own.

> 1. *International Institute for Population Sciences. National Family Health Survey - 5, 2019-21: India fact sheet [Internet]. place: Ministry of Health and Family Welfare; year [cited 2026 Sep 24]. Available from: https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf*

Replace *place* and *year* from the document, and count its pages for the extent if you want it. Until then the entry is unfinished, and should look it.

WHERE THIS PICTURE BREAKS. The entry is modelled on an example, not built from NLM's general format, which this book has not seen. The "(US)" that NLM's other examples put after a government body's name, and how an entry is shortened when one body both prepares and issues a

report, follow specific rules this book does not hold. Check them in chapter 22, and follow your journal's instructions where they differ.

Illustration 5. The reader. Write your one message on a separate page and put it away, as section 13 sets out. Give a colleague from another department the cut paragraph, with its last change made, and Table 1. She reads once. She marks one stop, at "NFHS-4": your draft began at the results, so give the next reader the methods sentence that names the survey. Then she writes what the paper says. Her sentence is invented here, but it is the kind readers write.

> *Obesity went up in Indian adults between the two surveys, a little more in men.*

Compare it with your message part by part.

part	your message	her sentence	verdict
about whom	adults aged 15-49 in India	Indian adults	widened: the age range is lost
what was found	the share with overweight or obesity rose, to about one woman in four and more than one man in five	obesity went up	changed: overweight dropped, and the levels lost
compared with what	NFHS-4 against NFHS-5, in each sex	the two surveys, and men against women	added: a comparison of the sexes the paper does not make
how far it reaches	a rise between two surveys, not a trend or its cause	a rise between the two surveys	kept

The added comparison is the worst fault, because readers repeat additions. Find the words that invited each fault. The age range sat in the middle of a long noun phrase. The draft's "overweight or obese" never said where overweight begins, so she heard "obese". The levels had gone to the table, so the paragraph gave her none to repeat. And the two rises stood side by side, inviting a comparison. Two estimates from two samples, 3.4 and 4.0 points, do not show that men rose faster.

Revise the paragraph:

> *Among adults aged 15-49 in India, the share with overweight or obesity rose between NFHS-4 (2015-16) and NFHS-5 (2019-21) (Table 1). It rose by 3.4 percentage points in women and by 4.0 percentage points in men. By 2019-21, about one woman in four and more than one man in five had a body mass index of 25.0 kg/m^2 or more.*

Give it, with the methods sentence, to a new reader, who writes:

> *Among people aged 15 to 49 in India, the share with a BMI of 25 or more went up between the two surveys, to about a quarter of women and over a fifth of men.*

Every part is kept at the same strength, and nothing is added. The paragraph passes.

What the finding does not mean. Write these into the discussion, in plain words.

- It is about adults aged 15 to 49. It says nothing of older adults, who are not in these rows.
- It counts overweight and obese together, at a BMI of 25.0 kg/m^2 or more. The fact sheet has no row for obesity alone.
- It leaves out pregnant women and women with a birth in the preceding 2 months.
- It shows that the share rose between two surveys of different people. It does not show why, and two surveys are not a trend.
- It does not show that the rise was larger in men than in women.
- It describes 2019-21, with fieldwork ending on 30 April 2021. It is not a figure for now.
- It is a national figure. It is not a figure for any state or district.

Figures quoted in this illustration are from [24].

WHERE THIS PICTURE BREAKS. Two readers are the smallest test that can work. A pass shows that one new reader could restate this paragraph minutes after reading it. It does not show that the claim is true, or that the survey's methods were sound.

MUST KNOW POINTS FOR YOU.

- Work in this order: the message, then the paragraph, then the table or figure, then the reference, then the reader.
- Give every number one home. Decide by what the reader has to do with the numbers, then cut the other homes.
- A passing score never tells you a paragraph works. Read it for its claim before you look at the number.
- Check the axis of every bar chart a program draws for you. Work out a ÷ (a − s) with the smaller bar; if it is above 1.05, start the bars at zero.
- Use one number of decimal places for the same figures in the text and the table. Two copies of one number with different digits look like two different numbers.
- When the source gives no counts, say so in a footnote. Never make one by multiplying a percentage by the survey's total sample. NFHS-5's 724,115 women are those it gathered information from, not the women weighed for this row.
- Format each kind of source from its own chapter of Citing Medicine. Where you have only an example to work from, say the entry is modelled on it. Leave a marked gap for any element you have not found.
- Watch for a reader's restatement that changes or adds a part, such as "obesity" for "overweight or obesity", or a comparison you did not make. Readers will repeat it. Find the words in your paragraph that invited it and change them, then test again with a new reader.
- This finding covers adults aged 15 to 49, counts overweight and obese together, and shows a rise between two surveys, not its cause. It describes 2019-21, not now.

COMMON MISREADING. "A finished result has a paragraph, a table and a figure, because the journey made all three." The journey made a table and a figure so that you could choose between them, and the paper kept the table. The figure went to the talk. Making both and keeping both is the duplication the ICMJE Recommendations ask you to avoid.

REPORTING IT. Among adults aged 15-49 in India, the share with overweight or obesity rose between NFHS-4 (2015-16) and NFHS-5 (2019-21) (Table 1). It rose by 3.4 percentage points in women and by 4.0 percentage points in men. By 2019-21, about one woman in four and more than one man in five had a body mass index of 25.0 kg/m^2 or more.

EXERCISE 1 (retrieval). From memory, list the steps of a finding's journey in order, from the question to what the finding does not mean. For each, name one check you make before moving on.

Record your confidence as a percentage before turning to the answer.

EXERCISE 2 (critique). A newspaper headline reads: "Obesity in Indian women up 17%, says national survey". Decide what was computed to get 17, compute it, and say what the headline gets wrong.

EXERCISE 3 (teaching). A journalist rings: "What does the national survey say about obesity in India?" You have three minutes. Give her one sentence she can quote, and say how sure you are.

EXERCISE 4 (design). Take one finding from the section you have been building for this book. Run it through the nine steps of the journey, on one sheet of paper.

Write down the limits of the data and the one message. Show the paragraph before and after, with its numbers checked, its score and its cut. Say which home the numbers get, and why. Write the reference. Compare a reader's sentence with yours. End with what the finding does not mean.

References

1. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. Recommendations home page, version line. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).
2. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.1 General Principles, first sentence. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).
3. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.1 General Principles, second sentence. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).
4. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.1 General Principles, last sentence of the first paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).
5. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.a Title Page, first sentence. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).
6. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.b Abstract, end of the second paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).
7. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.c Introduction, second sentence. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).
8. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.c Introduction, last sentence. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).
9. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.d Methods, first paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).
10. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.d Methods, first paragraph, third sentence. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).
11. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.e Results, first sentence. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

12. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.f Discussion, first paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

13. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.g References, i. General Considerations, the paragraph on accuracy. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

14. Sollaci, Luciana B and Pereira, Mauricio G. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. 2004. Abstract, Results. https://pmc.ncbi.nlm.nih.gov/articles/PMC442179/ (accessed 2026-10-02).

15. Sollaci, Luciana B and Pereira, Mauricio G. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. 2004. Results, first paragraph, second sentence. https://pmc.ncbi.nlm.nih.gov/articles/PMC442179/ (accessed 2026-10-02).

16. Sollaci, Luciana B and Pereira, Mauricio G. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. 2004. Results, first paragraph, fifth sentence. https://pmc.ncbi.nlm.nih.gov/articles/PMC442179/ (accessed 2026-10-02).

17. Sollaci, Luciana B and Pereira, Mauricio G. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. 2004. Results, the paragraph after Figure 1. https://pmc.ncbi.nlm.nih.gov/articles/PMC442179/ (accessed 2026-10-02).

18. Sollaci, Luciana B and Pereira, Mauricio G. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. 2004. Methods, first paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC442179/ (accessed 2026-10-02).

19. Sollaci, Luciana B and Pereira, Mauricio G. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. 2004. Discussion, third paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC442179/ (accessed 2026-10-02).

20. Sollaci, Luciana B and Pereira, Mauricio G. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. 2004. Discussion, fourth paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC442179/ (accessed 2026-10-02).

21. Sollaci, Luciana B and Pereira, Mauricio G. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. 2004. Discussion, last paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC442179/ (accessed 2026-10-02).

22. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 9, third paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

23. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 9, second paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

24. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. figures quoted in the text. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

25. Sollaci, Luciana B and Pereira, Mauricio G. The introduction, methods, results, and discussion (IMRAD) structure: a fifty-year survey. 2004. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC442179/ (accessed 2026-10-02).

26. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 3, last paragraph (the chronological-structure passage). https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

27. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 3, last paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

28. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Introduction, third paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

29. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 7, heading. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

30. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 7, the paragraph on results paragraphs. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

31. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.e Results, first paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

32. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.e Results, second paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

33. Chalmers, Iain and Glasziou, Paul. Avoidable waste in the production and reporting of research evidence. PMID 19525005. Viewpoint. Not openly licensed. Quoted from the publisher's full-text web page. Opening paragraph, first sentence. https://doi.org/10.1016/S0140-6736(09)60329-9 (accessed 2026-09-25).

34. Chalmers, Iain and Glasziou, Paul. Avoidable waste in the production and reporting of research evidence. PMID 19525005. Viewpoint. Not openly licensed. Quoted from the publisher's full-text web page. Biased or unusable reports of research, first paragraph. https://doi.org/10.1016/S0140-6736(09)60329-9 (accessed 2026-09-25).

35. Chalmers, Iain and Glasziou, Paul. Avoidable waste in the production and reporting of research evidence. PMID 19525005. Viewpoint. Not openly licensed. Quoted from the publisher's full-text web page. Biased or unusable reports of research, second paragraph. https://doi.org/10.1016/S0140-6736(09)60329-9 (accessed 2026-09-25).

36. Chalmers, Iain and Glasziou, Paul. Avoidable waste in the production and reporting of research evidence. PMID 19525005. Viewpoint. Not openly licensed. Quoted from the publisher's full-text web page. figures quoted in the text. https://doi.org/10.1016/S0140-6736(09)60329-9 (accessed 2026-09-25).

37. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 1, heading. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

38. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 1, first paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

39. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 1, fourth paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

40. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 1, third paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

41. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 3, first paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

42. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 5, first paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

43. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 5, paragraphs on context, content and conclusion. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

44. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 5, last paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

45. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Table 1, row for Rule 1. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

46. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.a Title Page, Article title. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

47. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.k Abbreviations and Symbols. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

48. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.b Abstract, first paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

49. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.b Abstract, second paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

50. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.b Abstract, second paragraph, first sentence. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

51. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Introduction, fourth paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

52. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Footnote 21 (to indicator 88). https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

53. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. figures quoted in the text. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

54. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 6, first paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

55. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 6, second paragraph (the progression of gaps), last sentence. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

56. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 6, third paragraph, the sentence on literature review. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

57. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 6, last paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

58. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.c Introduction. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

59. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.d Methods, first paragraph, first two sentences. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

60. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical

Journal. 2026. IV.A.3.d.iii Statistics, first sentence. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

61. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.e Results, first paragraph, third sentence. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

62. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 7, third paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

63. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 7, sixth paragraph (each subsequent results paragraph). https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

64. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 7, sixth paragraph, last sentences. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

65. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.f Discussion, second paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

66. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 8, first paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

67. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Table 1, row 8 (Discussion), the sign it is violated. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

68. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 9, third paragraph, last sentence. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

69. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Key indicators table, heading of the adult rows and indicators 88-89. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

70. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Key indicators table, column heading. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

71. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.g.i General Considerations, first paragraph, first two sentences. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

72. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.g.i General Considerations, first paragraph, predatory journals and preprints. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

73. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.g.i General Considerations, first paragraph, review articles. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

74. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.g.i General Considerations, second paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

75. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.g.i General Considerations, fourth paragraph (personal communication). ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

76. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.g.i General Considerations, fifth paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

77. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.g.i General Considerations, sixth paragraph, first two sentences. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

78. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.g.i General Considerations, sixth paragraph, retracted articles. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

79. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.g.i General Considerations, sixth paragraph, the PubMed search. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

80. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.g.i General Considerations, seventh paragraph (numbering). ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

81. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.g.i General Considerations, eighth paragraph (tables, legends, journal titles). ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

82. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.g.i General Considerations, eighth paragraph, last sentences. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

83. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.g.ii Style and Format. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

84. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1 Journals, A. Sample Citation and Introduction, important points. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

85. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1, Citation Rules with Examples for Journal Articles, the element list. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

86. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1, General Rules for Author. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

87. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1, General Rules for Author, initials. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

88. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1, General Rules for Author, number of authors. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

89. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1, Examples, 3. Journal article with optional limit to the number of authors, the six-author form. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

90. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. NLM, Samples of Formatted References for Authors of Journal Articles (a separate NLM page, last reviewed 2 June 2026), item 1. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

91. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. NLM, Samples of Formatted References, item 1, the six-author example. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

92. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1, General Rules for Article Title. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

93. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1, General Rules for Journal Title. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

94. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1, General Rules for Date of Publication. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

95. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1, General Rules for Date of Publication, months. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

96. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1, General Rules for Issue Number. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

97. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1, General Rules for Location (Pagination). https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

98. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1, Examples, 1. Standard journal article, first example. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

99. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. NLM, Samples of Formatted References, item 34 Forthcoming and Preprints, the note. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

100. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. NLM, Samples of Formatted References, item 34, the bioRxiv preprint of Ten simple rules for structuring papers. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

101. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1, Examples, 71. Journal article with DOI provided. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

102. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Article page, citation line. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

103. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Article page, date line. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

104. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Front matter, copyright line. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

105. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. figures quoted in the text. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

106. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. figures quoted in the text. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

107. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. figures quoted in the text. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

108. Robinson, Michelle Bachelor and Jerskey, Maria and Fulwiler, Toby. Writing Guide with Handbook. 2021. Section 6.6 Editing Focus: Subject-Verb Agreement, opening paragraph. https://openstax.org/books/writing-guide/pages/handbook (accessed 2026-10-02).

109. Robinson, Michelle Bachelor and Jerskey, Maria and Fulwiler, Toby. Writing Guide with Handbook. 2021. Handbook H3 Clear and Effective Sentences, Variety, Simple Sentences. https://openstax.org/books/writing-guide/pages/handbook (accessed 2026-10-02).

110. Robinson, Michelle Bachelor and Jerskey, Maria and Fulwiler, Toby. Writing Guide with Handbook. 2021. Handbook H3 Clear and Effective Sentences, Variety, Complex Sentences. https://openstax.org/books/writing-guide/pages/handbook (accessed 2026-10-02).

111. Robinson, Michelle Bachelor and Jerskey, Maria and Fulwiler, Toby. Writing Guide with Handbook. 2021. Handbook H3 Clear and Effective Sentences, Active Voice, first paragraph. https://openstax.org/books/writing-guide/pages/handbook (accessed 2026-10-02).

112. Robinson, Michelle Bachelor and Jerskey, Maria and Fulwiler, Toby. Writing Guide with Handbook. 2021. Section 3.6 Editing Focus: Sentence Structure, Revising Sentences to Change There are / There is or It is. https://openstax.org/books/writing-guide/pages/handbook (accessed 2026-10-02).

113. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.1.i Use active voice, p. 20. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

114. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.1.i Use active voice, How do you identify passive sentences?, p. 21. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

115. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.1.i Use active voice, How do you identify passive sentences?, second bullet, p. 21. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

116. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.1.i Use active voice, first paragraph, p. 20. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

117. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.1.i Use active voice, Use passive voice when the law is the actor, p. 21. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

118. Gopen, George D and Swan, Judith A. The science of scientific writing. 1990. The Topic Position, second paragraph (bees and pollen). https://www.usenix.org/sites/default/files/gopen_and_swan_science_of_scientific_writing.pdf (accessed 2026-10-02).

119. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Key indicators table, heading of the adult rows (above indicators 86-91). https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

120. Gopen, George D and Swan, Judith A. The science of scientific writing. 1990. Subject-Verb Separation, second paragraph. https://www.usenix.org/sites/default/files/gopen_and_swan_science_of_scientific_writing.pdf (accessed 2026-10-02).

121. Gopen, George D and Swan, Judith A. The science of scientific writing. 1990. Subject-Verb Separation, last paragraph. https://www.usenix.org/sites/default/files/gopen_and_swan_science_of_scientific_writing.pdf (accessed 2026-10-02).

122. Gopen, George D and Swan, Judith A. The science of scientific writing. 1990. The Stress Position, first paragraph. https://www.usenix.org/sites/default/files/gopen_and_swan_science_of_scientific_writing.pdf (accessed 2026-10-02).

123. Gopen, George D and Swan, Judith A. The science of scientific writing. 1990. The Stress Position, the paragraph beginning 'When is a sentence too long?'. https://www.usenix.org/sites/default/files/gopen_and_swan_science_of_scientific_writing.pdf (accessed 2026-10-02).

124. Gopen, George D and Swan, Judith A. The science of scientific writing. 1990. The Topic Position, the note of clarification. https://www.usenix.org/sites/default/files/gopen_and_swan_science_of_scientific_writing.pdf (accessed 2026-10-02).

125. Gopen, George D and Swan, Judith A. The science of scientific writing. 1990. The Topic Position, second paragraph. https://www.usenix.org/sites/default/files/gopen_and_swan_science_of_scientific_writing.pdf (accessed 2026-10-02).

126. Gopen, George D and Swan, Judith A. The science of scientific writing. 1990. Locating the Action, third paragraph. https://www.usenix.org/sites/default/files/gopen_and_swan_science_of_scientific_writing.pdf (accessed 2026-10-02).

127. Gopen, George D and Swan, Judith A. The science of scientific writing. 1990. Writing and the Scientific Process, the paragraph after the seven principles. https://www.usenix.org/sites/default/files/gopen_and_swan_science_of_scientific_writing.pdf (accessed 2026-10-02).

128. Gopen, George D and Swan, Judith A. The science of scientific writing. 1990. Perceiving Logical Gaps, revision tactic 1. https://www.usenix.org/sites/default/files/gopen_and_swan_science_of_scientific_writing.pdf (accessed 2026-10-02).

129. Gopen, George D and Swan, Judith A. The science of scientific writing. 1990. The Stress Position, the paragraph after the authors' revision of the gene passage. https://www.usenix.org/sites/default/files/gopen_and_swan_science_of_scientific_writing.pdf (accessed 2026-10-02).

130. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.b.1 Write short sentences, p. 50. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

131. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.b.2 Keep subject, verb, and object close together, p. 52. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

132. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.1.iii Avoid hidden verbs, What are hidden verbs?, p. 23. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

133. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.2.i Don't turn verbs into nouns, p. 29. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

134. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.b.1 Write short sentences, the royalty example, p. 50. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

135. Robinson, Michelle Bachelor and Jerskey, Maria and Fulwiler, Toby. Writing Guide with Handbook. 2021. Handbook H3 Clear and Effective Sentences, Emphasis. https://openstax.org/books/writing-guide/pages/handbook (accessed 2026-10-02).

136. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Key indicators table, column headings. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

137. Gopen, George D and Swan, Judith A. The science of scientific writing. 1990. figures quoted in the text. https://www.usenix.org/sites/default/files/gopen_and_swan_science_of_scientific_writing.pdf (accessed 2026-10-02).

138. Barnett, Adrian and Doubleday, Zoe. The growth of acronyms in the scientific literature. PMID 32701448, PMC7556863. CC BY 4.0. Quoted from the PubMed Central open-access text. Results, first paragraph. https://doi.org/10.7554/eLife.60080 (accessed 2026-10-02).

139. Barnett, Adrian and Doubleday, Zoe. The growth of acronyms in the scientific literature. PMID 32701448, PMC7556863. CC BY 4.0. Quoted from the PubMed Central open-access text. Materials and methods, Data extraction, first sentence. https://doi.org/10.7554/eLife.60080 (accessed 2026-10-02).

140. Barnett, Adrian and Doubleday, Zoe. The growth of acronyms in the scientific literature. PMID 32701448, PMC7556863. CC BY 4.0. Quoted from the PubMed Central open-access text. Results, second paragraph. https://doi.org/10.7554/eLife.60080 (accessed 2026-10-02).

141. Barnett, Adrian and Doubleday, Zoe. The growth of acronyms in the scientific literature. PMID 32701448, PMC7556863. CC BY 4.0. Quoted from the PubMed Central open-access text. Figure 1, legend. https://doi.org/10.7554/eLife.60080 (accessed 2026-10-02).

142. Barnett, Adrian and Doubleday, Zoe. The growth of acronyms in the scientific literature. PMID 32701448, PMC7556863. CC BY 4.0. Quoted from the PubMed Central open-access text. Results, the paragraph on three-letter acronyms and re-use. https://doi.org/10.7554/eLife.60080 (accessed 2026-10-02).

143. Barnett, Adrian and Doubleday, Zoe. The growth of acronyms in the scientific literature. PMID 32701448, PMC7556863. CC BY 4.0. Quoted from the PubMed Central open-access text. Introduction, second paragraph. https://doi.org/10.7554/eLife.60080 (accessed 2026-10-02).

144. Barnett, Adrian and Doubleday, Zoe. The growth of acronyms in the scientific literature. PMID 32701448, PMC7556863. CC BY 4.0. Quoted from the PubMed Central open-access text. Discussion, third paragraph. https://doi.org/10.7554/eLife.60080 (accessed 2026-10-02).

145. Barnett, Adrian and Doubleday, Zoe. The growth of acronyms in the scientific literature. PMID 32701448, PMC7556863. CC BY 4.0. Quoted from the PubMed Central open-access text. Discussion, second paragraph. https://doi.org/10.7554/eLife.60080 (accessed 2026-10-02).

146. Barnett, Adrian and Doubleday, Zoe. The growth of acronyms in the scientific literature. PMID 32701448, PMC7556863. CC BY 4.0. Quoted from the PubMed Central open-access text. Limitations. https://doi.org/10.7554/eLife.60080 (accessed 2026-10-02).

147. Barnett, Adrian and Doubleday, Zoe. The growth of acronyms in the scientific literature. PMID 32701448, PMC7556863. CC BY 4.0. Quoted from the PubMed Central open-access text. Materials and methods, Exclusion reasons and numbers. https://doi.org/10.7554/eLife.60080 (accessed 2026-10-02).

148. Barnett, Adrian and Doubleday, Zoe. The growth of acronyms in the scientific literature. PMID 32701448, PMC7556863. CC BY 4.0. Quoted from the PubMed Central open-access text. Abstract, last sentence. https://doi.org/10.7554/eLife.60080 (accessed 2026-10-02).

149. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. Section IV.A.3.k Abbreviations and Symbols. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

150. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. Section IV.A.3.e Results, last paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

151. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.2.iii Minimize abbreviations, If you must abbreviate, pp. 33-34. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

152. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.2.iii Minimize abbreviations, If you must abbreviate, p. 34. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

153. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.2.iii Minimize abbreviations, first paragraph, p. 33. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

154. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.2.iii Minimize abbreviations, Use nicknames, p. 33. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

155. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.3.i Use short, simple words, Fowler's list, p. 36. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

156. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.3.iv Use the same term consistently for a specific thought or object, p. 45. https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

157. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 4, Using parallelism, last sentences. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

158. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 2, second paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

159. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Key indicators table, indicator 88. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

160. Barnett, Adrian and Doubleday, Zoe. The growth of acronyms in the scientific literature. PMID 32701448, PMC7556863. CC BY 4.0. Quoted from the PubMed Central open-access text. figures quoted in the text. https://doi.org/10.7554/eLife.60080 (accessed 2026-10-02).

161. Cole, T J. Too many digits: the presentation of numerical data. 2015. Body text, paragraph defining decimal places and significant digits. https://pmc.ncbi.nlm.nih.gov/articles/PMC4483789/ (accessed 2026-10-02).

162. Cole, T J. Too many digits: the presentation of numerical data. 2015. Table 1, Rounding rules for summary statistics, row Mean. https://pmc.ncbi.nlm.nih.gov/articles/PMC4483789/ (accessed 2026-10-02).

163. Cole, T J. Too many digits: the presentation of numerical data. 2015. Table 1, row Percentage. https://pmc.ncbi.nlm.nih.gov/articles/PMC4483789/ (accessed 2026-10-02).

164. Cole, T J. Too many digits: the presentation of numerical data. 2015. Body text, paragraph on the rule of four. https://pmc.ncbi.nlm.nih.gov/articles/PMC4483789/ (accessed 2026-10-02).

165. Cole, T J. Too many digits: the presentation of numerical data. 2015. Table 1, row p value. https://pmc.ncbi.nlm.nih.gov/articles/PMC4483789/ (accessed 2026-10-02).

166. Cole, T J. Too many digits: the presentation of numerical data. 2015. Body text, last paragraph before Table 1. https://pmc.ncbi.nlm.nih.gov/articles/PMC4483789/ (accessed 2026-10-02).

167. Cole, T J. Too many digits: the presentation of numerical data. 2015. Body text, paragraph on intermediate calculations. https://pmc.ncbi.nlm.nih.gov/articles/PMC4483789/ (accessed 2026-10-02).

168. Lang, Thomas A and Altman, Douglas G. Basic Statistical Reporting for Articles Published in Biomedical Journals: The ``Statistical Analyses and Methods in the Published Literature'' or The SAMPL Guidelines. 2013. General Principles for Reporting Statistical Results, Reporting numbers and descriptive statistics, third bullet. EQUATOR Network PDF; also in Smart P, Maisonneuve H, Polderman A (eds), Science Editors' Handbook, European Associ-

ation of Science Editors, 2013. https://www.equator-network.org/wp-content/uploads/2013/07/SAMPL-Guidelines-6-27-13.pdf (accessed 2026-10-02).

169. Lang, Thomas A and Altman, Douglas G. Basic Statistical Reporting for Articles Published in Biomedical Journals: The ``Statistical Analyses and Methods in the Published Literature'' or The SAMPL Guidelines. 2013. Reporting numbers and descriptive statistics, fourth bullet. EQUATOR Network PDF; also in Smart P, Maisonneuve H, Polderman A (eds), Science Editors' Handbook, European Association of Science Editors, 2013. https://www.equator-network.org/wp-content/uploads/2013/07/SAMPL-Guidelines-6-27-13.pdf (accessed 2026-10-02).

170. Lang, Thomas A and Altman, Douglas G. Basic Statistical Reporting for Articles Published in Biomedical Journals: The ``Statistical Analyses and Methods in the Published Literature'' or The SAMPL Guidelines. 2013. Reporting hypothesis tests, the bullet on P values. EQUATOR Network PDF; also in Smart P, Maisonneuve H, Polderman A (eds), Science Editors' Handbook, European Association of Science Editors, 2013. https://www.equator-network.org/wp-content/uploads/2013/07/SAMPL-Guidelines-6-27-13.pdf (accessed 2026-10-02).

171. Lang, Thomas A and Altman, Douglas G. Basic Statistical Reporting for Articles Published in Biomedical Journals: The ``Statistical Analyses and Methods in the Published Literature'' or The SAMPL Guidelines. 2013. Reporting hypothesis tests, the bullet on P values, last sentence. EQUATOR Network PDF; also in Smart P, Maisonneuve H, Polderman A (eds), Science Editors' Handbook, European Association of Science Editors, 2013. https://www.equator-network.org/wp-content/uploads/2013/07/SAMPL-Guidelines-6-27-13.pdf (accessed 2026-10-02).

172. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.e Results, second paragraph, first sentence. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

173. Cole, T J. Too many digits: the presentation of numerical data. 2015. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC4483789/ (accessed 2026-10-02).

174. Robinson, Michelle Bachelor and Jerskey, Maria and Fulwiler, Toby. Writing Guide with Handbook. 2021. Handbook H2, Effective Paragraphs, Developing a Main Point. https://openstax.org/books/writing-guide/pages/handbook (accessed 2026-10-02).

175. Robinson, Michelle Bachelor and Jerskey, Maria and Fulwiler, Toby. Writing Guide with Handbook. 2021. Handbook H2, Effective Paragraphs, Supporting Evidence and Analysis, first sentence. https://openstax.org/books/writing-guide/pages/handbook (accessed 2026-10-02).

176. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 3, second-last paragraph (C-C-C at the paragraph scale). https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

177. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 4, Avoiding zig-zag, first paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

178. Kincaid, J. Peter and Fishburne, Jr., Robert P. and Rogers, Richard L. and Chissom, Brad S. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. 1975. Appendix B, Instructions for Recalculated Flesch Formula, step 3, the syllable examples. https://stars.library.ucf.edu/istlibrary/56 (accessed 2026-10-02).

179. Kincaid, J. Peter and Fishburne, Jr., Robert P. and Rogers, Richard L. and Chissom, Brad S. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. 1975. Appendix B, steps 4 and 5, the two averages. https://stars.library.ucf.edu/istlibrary/56 (accessed 2026-10-02).

180. Flesch, Rudolf. A New Readability Yardstick. 1948.

181. Flesch, Rudolf. How to Write Plain English: A Book for Lawyers and Consumers. 1979. Chapter 2, Let's Start With the Formula, Step 6, the formula in words. https://pages.stern.nyu.edu/%7Ewstarbuc/Writing/Flesch.htm (accessed 2026-10-02).

182. Kincaid, J. Peter and Fishburne, Jr., Robert P. and Rogers, Richard L. and Chissom, Brad S. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy En-

listed Personnel. 1975. Appendix B, step 6, the grade-level equation (report p. 39; checked against the page image at intake). https://stars.library.ucf.edu/istlibrary/56 (accessed 2026-10-02).

183. Kincaid, J. Peter and Fishburne, Jr., Robert P. and Rogers, Richard L. and Chissom, Brad S. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. 1975. Table 3, footnote. https://stars.library.ucf.edu/istlibrary/56 (accessed 2026-10-02).

184. Edwards, Caleb Simpeh and Ammanuel, Simon Gashaw and Silva, Ogonna N Nnamani and Greeneway, Garret P and Bunch, Katherine M and Meisner, Lars W and Page, Paul S and Ahmed, Azam S. Academics versus the Internet: Evaluating the readability of patient education materials for cerebrovascular conditions from major academic centers. 2022. Introduction, paragraph 3. https://pmc.ncbi.nlm.nih.gov/articles/PMC9479524/ (accessed 2026-10-02).

185. Kincaid, J. Peter and Fishburne, Jr., Robert P. and Rogers, Richard L. and Chissom, Brad S. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. 1975. Summary, the derivation sample. https://stars.library.ucf.edu/istlibrary/56 (accessed 2026-10-02).

186. Kincaid, J. Peter and Fishburne, Jr., Robert P. and Rogers, Richard L. and Chissom, Brad S. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. 1975. Summary, the test passages. https://stars.library.ucf.edu/istlibrary/56 (accessed 2026-10-02).

187. Kincaid, J. Peter and Fishburne, Jr., Robert P. and Rogers, Richard L. and Chissom, Brad S. Derivation of New Readability Formulas (Automated Readability Index, Fog Count and Flesch Reading Ease Formula) for Navy Enlisted Personnel. 1975. Summary, what the recalculated formulas are for. https://stars.library.ucf.edu/istlibrary/56 (accessed 2026-10-02).

188. Plav'en-Sigray, Pontus and Matheson, Granville James and Schiffler, Bj"orn Christian and Thompson, William Hedley. The readability of scientific texts is decreasing over time. 2017. Abstract. https://pmc.ncbi.nlm.nih.gov/articles/PMC5584989/ (accessed 2026-10-02).

189. Plav'en-Sigray, Pontus and Matheson, Granville James and Schiffler, Bj"orn Christian and Thompson, William Hedley. The readability of scientific texts is decreasing over time. 2017. Results, the yearly trend in FRE. https://pmc.ncbi.nlm.nih.gov/articles/PMC5584989/ (accessed 2026-10-02).

190. Plav'en-Sigray, Pontus and Matheson, Granville James and Schiffler, Bj"orn Christian and Thompson, William Hedley. The readability of scientific texts is decreasing over time. 2017. Discussion, first paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC5584989/ (accessed 2026-10-02).

191. Plav'en-Sigray, Pontus and Matheson, Granville James and Schiffler, Bj"orn Christian and Thompson, William Hedley. The readability of scientific texts is decreasing over time. 2017. Abstract, the explanation. https://pmc.ncbi.nlm.nih.gov/articles/PMC5584989/ (accessed 2026-10-02).

192. Plav'en-Sigray, Pontus and Matheson, Granville James and Schiffler, Bj"orn Christian and Thompson, William Hedley. The readability of scientific texts is decreasing over time. 2017. Discussion, the limits of readability formulas. https://pmc.ncbi.nlm.nih.gov/articles/PMC5584989/ (accessed 2026-10-02).

193. Plav'en-Sigray, Pontus and Matheson, Granville James and Schiffler, Bj"orn Christian and Thompson, William Hedley. The readability of scientific texts is decreasing over time. 2017. Discussion, the limits of readability formulas, first sentences. https://pmc.ncbi.nlm.nih.gov/articles/PMC5584989/ (accessed 2026-10-02).

194. Plav'en-Sigray, Pontus and Matheson, Granville James and Schiffler, Bj"orn Christian and Thompson, William Hedley. The readability of scientific texts is decreasing over time. 2017. Discussion, on specialists. https://pmc.ncbi.nlm.nih.gov/articles/PMC5584989/ (accessed 2026-10-02).

195. Flesch, Rudolf. How to Write Plain English: A Book for Lawyers and Consumers. 1979. Chapter 2, Step 2, figures as one syllable. https://pages.stern.nyu.edu/%7Ewstarbuc/Writing/Flesch.htm (accessed 2026-10-02).

196. Plav'en-Sigray, Pontus and Matheson, Granville James and Schiffler, Bj"orn Christian and Thompson, William Hedley. The readability of scientific texts is decreasing over time. 2017. Materials and methods, Language preprocessing, and the syllable count. https://pmc.ncbi.nlm.nih.gov/articles/PMC5584989/ (accessed 2026-10-02).

197. Flesch, Rudolf. How to Write Plain English: A Book for Lawyers and Consumers. 1979. Chapter 2, the first tip after the score table. https://pages.stern.nyu.edu/%7Ewstarbuc/Writing/Flesch.htm (accessed 2026-10-02).

198. Flesch, Rudolf. How to Write Plain English: A Book for Lawyers and Consumers. 1979. Chapter 2, the score-to-school-level table and its lead sentence. https://pages.stern.nyu.edu/%7Ewstarbuc/Writing/Flesch.htm (accessed 2026-10-02).

199. Flesch, Rudolf. How to Write Plain English: A Book for Lawyers and Consumers. 1979. Chapter 2, after the formula, the Plain English minimum. https://pages.stern.nyu.edu/%7Ewstarbuc/Writing/Flesch.htm (accessed 2026-10-02).

200. Plav'en-Sigray, Pontus and Matheson, Granville James and Schiffler, Bj"orn Christian and Thompson, William Hedley. The readability of scientific texts is decreasing over time. 2017. Discussion, on the top of the scale. https://pmc.ncbi.nlm.nih.gov/articles/PMC5584989/ (accessed 2026-10-02).

201. Plav'en-Sigray, Pontus and Matheson, Granville James and Schiffler, Bj"orn Christian and Thompson, William Hedley. The readability of scientific texts is decreasing over time. 2017. Discussion, last paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC5584989/ (accessed 2026-10-02).

202. Edwards, Caleb Simpeh and Ammanuel, Simon Gashaw and Silva, Ogonna N Nnamani and Greeneway, Garret P and Bunch, Katherine M and Meisner, Lars W and Page, Paul S and Ahmed, Azam S. Academics versus the Internet: Evaluating the readability of patient education materials for cerebrovascular conditions from major academic centers. 2022. Methods, how the scores were generated. https://pmc.ncbi.nlm.nih.gov/articles/PMC9479524/ (accessed 2026-10-02).

203. Flesch, Rudolf. How to Write Plain English: A Book for Lawyers and Consumers. 1979. Chapter 2, Step 2, words with two pronunciations. https://pages.stern.nyu.edu/%7Ewstarbuc/Writing/Flesch.htm (accessed 2026-10-02).

204. Flesch, Rudolf. How to Write Plain English: A Book for Lawyers and Consumers. 1979. figures quoted in the text. https://pages.stern.nyu.edu/%7Ewstarbuc/Writing/Flesch.htm (accessed 2026-10-02).

205. Plav'en-Sigray, Pontus and Matheson, Granville James and Schiffler, Bj"orn Christian and Thompson, William Hedley. The readability of scientific texts is decreasing over time. 2017. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC5584989/ (accessed 2026-10-02).

206. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.3.ii Omit unnecessary words, first paragraph (p. 38). https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

207. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.3.ii Omit unnecessary words, second paragraph (p. 38). https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

208. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.3.ii Omit unnecessary words, the redundant-words example (p. 38). https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

209. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.3.ii Omit unnecessary words, the “Don’t say / Say” table of wordy phrases (p. 38). https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

210. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.3.ii Omit unnecessary words, excess modifiers and doublets (p. 39). https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

211. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.a.3.ii Omit unnecessary words, the worked example (p. 39). https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

212. Plain Language Action and Information Network. Federal Plain Language Guidelines. 2011. III.b.1 Write short sentences, second paragraph (p. 50). https://wid.org/wp-content/uploads/2022/03/FederalPLGuidelines.pdf (accessed 2026-10-02).

213. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 4, Avoiding zig-zag. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

214. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 9, on outlining before writing. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

215. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 9, the same paragraph, on wasted wordsmithing. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

216. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 10, first paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

217. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Table 1, row for Rule 1, the sign the rule is violated. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

218. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 2, first paragraph. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

219. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 10, the paragraph after Table 1. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

220. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Table 1, row for Rule 4. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

221. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 10, the paragraph after Table 1, on kinds of feedback. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

222. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 10, first paragraph, the writer's own check. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

223. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 10, last sentences of the paragraph after Table 1. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

224. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Key indicators, indicator 88 (women aged 15-49 overweight or obese). https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

225. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Rule 2, Identify Your Message, first sentence. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

226. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Rule 2, Identify Your Message, the sentence after the reference to Figure 2. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

227. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Rule 1, Know Your Audience, second sentence. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

228. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Rule 3, Adapt the Figure to the Support Medium, last sentence. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

229. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Rule 3, Adapt the Figure to the Support Medium, first sentence. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

230. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical

Journal. 2026. IV.A.3.h Tables, first paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

231. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.i Illustrations (Figures), third paragraph, last sentence. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

232. Wilke, Claus O. Chapter 29: Telling a story and making a point. 2019. Chapter 29, section 29.1 What is a story?, third paragraph. https://clauswilke.com/dataviz/telling-a-story.html (accessed 2026-10-02).

233. Wilke, Claus O. Chapter 29: Telling a story and making a point. 2019. Chapter 29, section 29.2 Make a figure for the generals, the closing paragraph on the airline figures. https://clauswilke.com/dataviz/telling-a-story.html (accessed 2026-10-02).

234. Wilke, Claus O. Chapter 29: Telling a story and making a point. 2019. Chapter 29, section 29.5 Be consistent but don't be repetitive, last paragraph. https://clauswilke.com/dataviz/telling-a-story.html (accessed 2026-10-02).

235. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Rule 7, Do Not Mislead the Reader, first sentence. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

236. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Key indicators, indicator 89 (men aged 15-49 overweight or obese). https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

237. Cleveland, William S. and McGill, Robert. Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods. 1984.

238. Heer, Jeffrey and Bostock, Michael. Crowdsourcing graphical perception: using Mechanical Turk to assess visualization design. Copyright 2010 ACM. Quoted from the authors' copy on the Stanford Visualization Group site (PDF text layer). Experiment 1A, Method, second paragraph. http://vis.stanford.edu/files/2010-MTurk-CHI.pdf (accessed 2026-10-02).

239. Heer, Jeffrey and Bostock, Michael. Crowdsourcing graphical perception: using Mechanical Turk to assess visualization design. Copyright 2010 ACM. Quoted from the authors' copy on the Stanford Visualization Group site (PDF text layer). Experiment 1A, Results, second paragraph. http://vis.stanford.edu/files/2010-MTurk-CHI.pdf (accessed 2026-10-02).

240. Heer, Jeffrey and Bostock, Michael. Crowdsourcing graphical perception: using Mechanical Turk to assess visualization design. Copyright 2010 ACM. Quoted from the authors' copy on the Stanford Visualization Group site (PDF text layer). Experiment 1A, Results, the paragraph on angle and circular area. http://vis.stanford.edu/files/2010-MTurk-CHI.pdf (accessed 2026-10-02).

241. Wilke, Claus O. Chapter 5: Directory of visualizations. 2019. Chapter 5, section 5.1 Amounts. https://clauswilke.com/dataviz/directory-of-visualizations.html (accessed 2026-10-02).

242. Wilke, Claus O. Chapter 5: Directory of visualizations. 2019. Chapter 5, section 5.3 Proportions. https://clauswilke.com/dataviz/directory-of-visualizations.html (accessed 2026-10-02).

243. Wilke, Claus O. Chapter 5: Directory of visualizations. 2019. Chapter 5, section 5.4 x-y relationships, first paragraph. https://clauswilke.com/dataviz/directory-of-visualizations.html (accessed 2026-10-02).

244. Wilke, Claus O. Chapter 5: Directory of visualizations. 2019. Chapter 5, section 5.4 x-y relationships, third paragraph. https://clauswilke.com/dataviz/directory-of-visualizations.html (accessed 2026-10-02).

245. Wilke, Claus O. Chapter 5: Directory of visualizations. 2019. Chapter 5, section 5.2 Distributions, second paragraph. https://clauswilke.com/dataviz/directory-of-visualizations.html (accessed 2026-10-02).

246. Wilke, Claus O. Chapter 17: The principle of proportional ink. 2019. Chapter 17, section 17.3 Direct area visualizations, second paragraph. https://clauswilke.com/dataviz/proportional-ink.html (accessed 2026-10-02).

247. Wilke, Claus O. Chapter 17: The principle of proportional ink. 2019. Chapter 17, section 17.2 Visualizations along logarithmic axes, the GDP example. https://clauswilke.com/dataviz/proportional-ink.html (accessed 2026-10-02).

248. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Key indicators, indicators 81, 82, 84 and 85 (children under 5), for the critique exercise. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

249. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Key indicators, indicators 82, 84 and 85. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

250. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Key indicators, indicator 84. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

251. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Key indicators, indicator 85. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

252. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Footnotes 18 and 20 to the key indicators (the cut-offs for indicators 81-85). https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

253. Wilke, Claus O. Chapter 17: The principle of proportional ink. 2019. Chapter 17, section 17.2 Visualizations along logarithmic axes, the rule at the end of the section. https://clauswilke.com/dataviz/proportional-ink.html (accessed 2026-10-02).

254. Heer, Jeffrey and Bostock, Michael. Crowdsourcing graphical perception: using Mechanical Turk to assess visualization design. Copyright 2010 ACM. Quoted from the authors' copy on the Stanford Visualization Group site (PDF text layer). figures quoted in the text. http://vis.stanford.edu/files/2010-MTurk-CHI.pdf (accessed 2026-10-02).

255. Weissgerber, Tracey L and Milic, Natasa M and Winham, Stacey J and Garovic, Vesna D. Beyond bar and line graphs: time for a new data presentation paradigm. PMID 25901488, PMC4406565. Open access under CC BY 4.0. Quoted from the PMC open-access text. Introduction, second paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC4406565/ (accessed 2026-10-02).

256. Weissgerber, Tracey L and Milic, Natasa M and Winham, Stacey J and Garovic, Vesna D. Beyond bar and line graphs: time for a new data presentation paradigm. PMID 25901488, PMC4406565. Open access under CC BY 4.0. Quoted from the PMC open-access text. Abstract. https://pmc.ncbi.nlm.nih.gov/articles/PMC4406565/ (accessed 2026-10-02).

257. Weissgerber, Tracey L and Milic, Natasa M and Winham, Stacey J and Garovic, Vesna D. Beyond bar and line graphs: time for a new data presentation paradigm. PMID 25901488, PMC4406565. Open access under CC BY 4.0. Quoted from the PMC open-access text. Fig 1 caption, Panels C to E. https://pmc.ncbi.nlm.nih.gov/articles/PMC4406565/ (accessed 2026-10-02).

258. Weissgerber, Tracey L and Milic, Natasa M and Winham, Stacey J and Garovic, Vesna D. Beyond bar and line graphs: time for a new data presentation paradigm. PMID 25901488, PMC4406565. Open access under CC BY 4.0. Quoted from the PMC open-access text. The systematic review, the paragraph describing its scope. https://pmc.ncbi.nlm.nih.gov/articles/PMC4406565/ (accessed 2026-10-02).

259. Weissgerber, Tracey L and Milic, Natasa M and Winham, Stacey J and Garovic, Vesna D. Beyond bar and line graphs: time for a new data presentation paradigm. PMID 25901488, PMC4406565. Open access under CC BY 4.0. Quoted from the PMC open-access text. Are Your Figures Worth a Thousand Words?, first paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC4406565/ (accessed 2026-10-02).

260. Weissgerber, Tracey L and Milic, Natasa M and Winham, Stacey J and Garovic, Vesna D. Beyond bar and line graphs: time for a new data presentation paradigm. PMID 25901488, PMC4406565. Open access under CC BY 4.0.

Quoted from the PMC open-access text. Summary Statistics Are Only Meaningful When There Are Enough Data to Summarize, first paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC4406565/ (accessed 2026-10-02).

261. Weissgerber, Tracey L and Milic, Natasa M and Winham, Stacey J and Garovic, Vesna D. Beyond bar and line graphs: time for a new data presentation paradigm. PMID 25901488, PMC4406565. Open access under CC BY 4.0. Quoted from the PMC open-access text. Recommendations for a New Data Presentation Paradigm, Encourage a more complete presentation of data. https://pmc.ncbi.nlm.nih.gov/articles/PMC4406565/ (accessed 2026-10-02).

262. Weissgerber, Tracey L and Milic, Natasa M and Winham, Stacey J and Garovic, Vesna D. Beyond bar and line graphs: time for a new data presentation paradigm. PMID 25901488, PMC4406565. Open access under CC BY 4.0. Quoted from the PMC open-access text. Introduction, second paragraph, the second problem. https://pmc.ncbi.nlm.nih.gov/articles/PMC4406565/ (accessed 2026-10-02).

263. Weissgerber, Tracey L and Milic, Natasa M and Winham, Stacey J and Garovic, Vesna D. Beyond bar and line graphs: time for a new data presentation paradigm. PMID 25901488, PMC4406565. Open access under CC BY 4.0. Quoted from the PMC open-access text. Are Your Figures Worth a Thousand Words?, first paragraph, last sentence. https://pmc.ncbi.nlm.nih.gov/articles/PMC4406565/ (accessed 2026-10-02).

264. Wilke, Claus O. Chapter 9: Visualizing many distributions at once. 2019. Chapter 9, section 9.1, the paragraph on strip charts. https://clauswilke.com/dataviz/boxplots-violins.html (accessed 2026-10-02).

265. Wilke, Claus O. Chapter 9: Visualizing many distributions at once. 2019. Chapter 9, Figure 9.2 caption, anatomy of a boxplot. https://clauswilke.com/dataviz/boxplots-violins.html (accessed 2026-10-02).

266. Wilke, Claus O. Chapter 9: Visualizing many distributions at once. 2019. Chapter 9, section 9.1, the paragraph on violin plots. https://clauswilke.com/dataviz/boxplots-violins.html (accessed 2026-10-02).

267. Wilke, Claus O. Chapter 9: Visualizing many distributions at once. 2019. Chapter 9, section 9.1, first paragraph, the problems with points and error bars. https://clauswilke.com/dataviz/boxplots-violins.html (accessed 2026-10-02).

268. Wilke, Claus O. Chapter 7: Visualizing distributions: Histograms and density plots. 2019. Chapter 7, section 7.1, on bin widths. https://clauswilke.com/dataviz/histograms-density-plots.html (accessed 2026-10-02).

269. Bergstrom, Carl T and West, Jevin. The principle of proportional ink. 2016. Opening paragraph, the statement of the rule. Calling Bullshit (course website), Tools. https://callingbullshit.org/tools/tools_proportional_ink.html (accessed 2026-10-02).

270. Wilke, Claus O. Chapter 17: The principle of proportional ink. 2019. Ch. 17, The principle of proportional ink, opening section (block quote naming the principle). https://clauswilke.com/dataviz/proportional-ink.html (accessed 2026-10-02).

271. Wilke, Claus O. Chapter 17: The principle of proportional ink. 2019. §17.1 Visualizations along linear axes, the rule set apart after the Hawaii example. https://clauswilke.com/dataviz/proportional-ink.html (accessed 2026-10-02).

272. Bergstrom, Carl T and West, Jevin. The principle of proportional ink. 2016. Section Line graphs, second paragraph. Calling Bullshit (course website), Tools. https://callingbullshit.org/tools/tools_proportional_ink.html (accessed 2026-10-02).

273. Wilke, Claus O. Chapter 17: The principle of proportional ink. 2019. §17.2 Visualizations along logarithmic axes, the rule set apart at the end of the section. https://clauswilke.com/dataviz/proportional-ink.html (accessed 2026-10-02).

274. Bergstrom, Carl T and West, Jevin. The principle of proportional ink. 2016. Section Line graphs, first paragraph. Calling Bullshit (course website), Tools. https://callingbullshit.org/tools/tools_proportional_ink.html (accessed 2026-10-02).

275. Tufte, Edward R. The Visual Display of Quantitative Information. 1983.

276. Correll, Michael and Bertini, Enrico and Franconeri, Steven. Truncating the Y-Axis: Threat or Menace?. 2020. Experiment One, Results, first hypothesis. https://arxiv.org/abs/1907.02035 (accessed 2026-10-02).

277. Correll, Michael and Bertini, Enrico and Franconeri, Steven. Truncating the Y-Axis: Threat or Menace?. 2020. DISCUSSION, first paragraph. https://arxiv.org/abs/1907.02035 (accessed 2026-10-02).

278. Correll, Michael and Bertini, Enrico and Franconeri, Steven. Truncating the Y-Axis: Threat or Menace?. 2020. Conclusion. https://arxiv.org/abs/1907.02035 (accessed 2026-10-02).

279. Correll, Michael and Bertini, Enrico and Franconeri, Steven. Truncating the Y-Axis: Threat or Menace?. 2020. figures quoted in the text. https://arxiv.org/abs/1907.02035 (accessed 2026-10-02).

280. Wilke, Claus O. Chapter 23: Balance the data and the context. 2019. Ch. 23 Balance the data and the context, opening paragraph. https://clauswilke.com/dataviz/balance-data-context.html (accessed 2026-10-02).

281. Bateman, Scott and Mandryk, Regan L and Gutwin, Carl and Genest, Aaron and McDine, David and Brooks, Christopher. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. 2010. Abstract. https://doi.org/10.1145/1753326.1753716 (accessed 2026-10-02).

282. Bateman, Scott and Mandryk, Regan L and Gutwin, Carl and Genest, Aaron and McDine, David and Brooks, Christopher. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. 2010. COMPARISON OF PLAIN AND EMBELLISHED CHARTS, Participants and Apparatus. https://doi.org/10.1145/1753326.1753716 (accessed 2026-10-02).

283. Bateman, Scott and Mandryk, Regan L and Gutwin, Carl and Genest, Aaron and McDine, David and Brooks, Christopher. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. 2010. Procedure, Part 2: Recall Task. https://doi.org/10.1145/1753326.1753716 (accessed 2026-10-02).

284. Bateman, Scott and Mandryk, Regan L and Gutwin, Carl and Genest, Aaron and McDine, David and Brooks, Christopher. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. 2010. RESULTS, Recall, paragraph on the long-term group. https://doi.org/10.1145/1753326.1753716 (accessed 2026-10-02).

285. Bateman, Scott and Mandryk, Regan L and Gutwin, Carl and Genest, Aaron and McDine, David and Brooks, Christopher. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. 2010. DISCUSSION, on why interpretation accuracy was not reduced. https://doi.org/10.1145/1753326.1753716 (accessed 2026-10-02).

286. Bateman, Scott and Mandryk, Regan L and Gutwin, Carl and Genest, Aaron and McDine, David and Brooks, Christopher. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. 2010. DISCUSSION, Design Questions and Implications. https://doi.org/10.1145/1753326.1753716 (accessed 2026-10-02).

287. Wilke, Claus O. Chapter 26: Don't go 3D. 2019. Ch. 26 Don't go 3D, §26.1 Avoid gratuitous 3D, first paragraph. https://clauswilke.com/dataviz/no-3d.html (accessed 2026-10-02).

288. Wilke, Claus O. Chapter 20: Redundant coding. 2019. Ch. 20 Redundant coding, §20.2 Designing figures without legends, the rule set apart. https://clauswilke.com/dataviz/redundant-coding.html (accessed 2026-10-02).

289. Wilke, Claus O. Chapter 23: Balance the data and the context. 2019. Ch. 23, figure 23.3 caption, the over-minimal figure. https://clauswilke.com/dataviz/balance-data-context.html (accessed 2026-10-02).

290. Bateman, Scott and Mandryk, Regan L and Gutwin, Carl and Genest, Aaron and McDine, David and Brooks, Christopher. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. 2010. RESULTS, Gaze Detection. https://doi.org/10.1145/1753326.1753716 (accessed 2026-10-02).

291. Bateman, Scott and Mandryk, Regan L and Gutwin, Carl and Genest, Aaron and McDine, David and Brooks, Christopher. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. 2010. RESULTS, Description, the value message. https://doi.org/10.1145/1753326.1753716 (accessed 2026-10-02).

292. Bateman, Scott and Mandryk, Regan L and Gutwin, Carl and Genest, Aaron and McDine, David and Brooks, Christopher. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. 2010. INTRODUCTION, last paragraph. https://doi.org/10.1145/1753326.1753716 (accessed 2026-10-02).

293. Bateman, Scott and Mandryk, Regan L and Gutwin, Carl and Genest, Aaron and McDine, David and Brooks, Christopher. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. 2010. DISCUSSION, The wider problem of bias in charts. https://doi.org/10.1145/1753326.1753716 (accessed 2026-10-02).

294. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Rule 8: Avoid "Chartjunk". https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

295. Wilke, Claus O. Chapter 26: Don't go 3D. 2019. figures quoted in the text. https://clauswilke.com/dataviz/no-3d.html (accessed 2026-10-02).

296. Bateman, Scott and Mandryk, Regan L and Gutwin, Carl and Genest, Aaron and McDine, David and Brooks, Christopher. Useful junk? The effects of visual embellishment on comprehension and memorability of charts. 2010. figures quoted in the text. https://doi.org/10.1145/1753326.1753716 (accessed 2026-10-02).

297. Wilke, Claus O. Chapter 4: Color scales. 2019. Chapter 4, Color scales, opening paragraph. https://clauswilke.com/dataviz/color-basics.html (accessed 2026-10-02).

298. Wilke, Claus O. Chapter 4: Color scales. 2019. Chapter 4, section 4.1 Color as a tool to distinguish, first paragraph. https://clauswilke.com/dataviz/color-basics.html (accessed 2026-10-02).

299. Wilke, Claus O. Chapter 4: Color scales. 2019. Chapter 4, section 4.2 Color to represent data values, paragraph on diverging scales. https://clauswilke.com/dataviz/color-basics.html (accessed 2026-10-02).

300. Wilke, Claus O. Chapter 4: Color scales. 2019. Chapter 4, section 4.3 Color as a tool to highlight, last paragraph. https://clauswilke.com/dataviz/color-basics.html (accessed 2026-10-02).

301. Crameri, Fabio and Shephard, Grace E and Heron, Philip J. The misuse of colour in science communication. PMID 33116149, PMC7595127. Open access under CC BY 4.0. Quoted from the PMC open-access text. The importance of colour to prevent data distortion, first paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC7595127/ (accessed 2026-10-02).

302. Crameri, Fabio and Shephard, Grace E and Heron, Philip J. The misuse of colour in science communication. PMID 33116149, PMC7595127. Open access under CC BY 4.0. Quoted from the PMC open-access text. The importance of colour to prevent data distortion, first paragraph (the size of the error). https://pmc.ncbi.nlm.nih.gov/articles/PMC7595127/ (accessed 2026-10-02).

303. Wilke, Claus O. Chapter 19: Common pitfalls of color use. 2019. Chapter 19, section 19.2 Using non-monotonic color scales to encode data values. https://clauswilke.com/dataviz/color-pitfalls.html (accessed 2026-10-02).

304. Crameri, Fabio and Shephard, Grace E and Heron, Philip J. The misuse of colour in science communication. PMID 33116149, PMC7595127. Open access under CC BY 4.0. Quoted from the PMC open-access text. Introduction, paragraph on red and green and colour-vision deficiency. https://pmc.ncbi.nlm.nih.gov/articles/PMC7595127/ (accessed 2026-10-02).

305. Crameri, Fabio and Shephard, Grace E and Heron, Philip J. The misuse of colour in science communication. PMID 33116149, PMC7595127. Open access under CC BY 4.0. Quoted from the PMC open-access text. Box 1, Human colour perception, third paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC7595127/ (accessed 2026-10-02).

306. Krishnamurthy, Sruthi Sree and Rangavittal, Subhiksha and Chandrasekar, Ambika and Narayanan, Anuradha. Prevalence of color vision deficiency among school-going boys in South India. PMID 34304169, PMC8482944. Open access under CC BY-NC-SA 4.0. Quoted from the PMC open-access text. Abstract, Results. https://pmc.ncbi.nlm.nih.gov/articles/PMC8482944/ (accessed 2026-10-02).

307. Krishnamurthy, Sruthi Sree and Rangavittal, Subhiksha and Chandrasekar, Ambika and Narayanan, Anuradha. Prevalence of color vision deficiency among school-going boys in South India. PMID 34304169, PMC8482944. Open access under CC BY-NC-SA 4.0. Quoted from the PMC open-access text. Abstract, Results, first sentence. https://pmc.ncbi.nlm.nih.gov/articles/PMC8482944/ (accessed 2026-10-02).

308. Krishnamurthy, Sruthi Sree and Rangavittal, Subhiksha and Chandrasekar, Ambika and Narayanan, Anuradha. Prevalence of color vision deficiency among school-going boys in South India. PMID 34304169, PMC8482944. Open access under CC BY-NC-SA 4.0. Quoted from the PMC open-access text. Methods, Definitions, first paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC8482944/ (accessed 2026-10-02).

309. Krishnamurthy, Sruthi Sree and Rangavittal, Subhiksha and Chandrasekar, Ambika and Narayanan, Anuradha. Prevalence of color vision deficiency among school-going boys in South India. PMID 34304169, PMC8482944. Open access under CC BY-NC-SA 4.0. Quoted from the PMC open-access text. Methods, first paragraph (the setting). https://pmc.ncbi.nlm.nih.gov/articles/PMC8482944/ (accessed 2026-10-02).

310. Crameri, Fabio and Shephard, Grace E and Heron, Philip J. The misuse of colour in science communication. PMID 33116149, PMC7595127. Open access under CC BY 4.0. Quoted from the PMC open-access text. Design and implementation of scientifically derived colour maps, paragraph on universal readability. https://pmc.ncbi.nlm.nih.gov/articles/PMC7595127/ (accessed 2026-10-02).

311. Wilke, Claus O. Chapter 19: Common pitfalls of color use. 2019. Chapter 19, section 19.3 Not designing for color-vision deficiency, last two paragraphs. https://clauswilke.com/dataviz/color-pitfalls.html (accessed 2026-10-02).

312. Crameri, Fabio and Shephard, Grace E and Heron, Philip J. The misuse of colour in science communication. PMID 33116149, PMC7595127. Open access under CC BY 4.0. Quoted from the PMC open-access text. Introduction, fourth paragraph (what is wrong with rainbow colour maps). https://pmc.ncbi.nlm.nih.gov/articles/PMC7595127/ (accessed 2026-10-02).

313. Crameri, Fabio and Shephard, Grace E and Heron, Philip J. The misuse of colour in science communication. PMID 33116149, PMC7595127. Open access under CC BY 4.0. Quoted from the PMC open-access text. Design and implementation of scientifically derived colour maps, paragraph on the colour bar. https://pmc.ncbi.nlm.nih.gov/articles/PMC7595127/ (accessed 2026-10-02).

314. Crameri, Fabio and Shephard, Grace E and Heron, Philip J. The misuse of colour in science communication. PMID 33116149, PMC7595127. Open access under CC BY 4.0. Quoted from the PMC open-access text. Box 2, Freely available scientifically derived colour maps, the MPL and Cividis entries. https://pmc.ncbi.nlm.nih.gov/articles/PMC7595127/ (accessed 2026-10-02).

315. Crameri, Fabio and Shephard, Grace E and Heron, Philip J. The misuse of colour in science communication. PMID 33116149, PMC7595127. Open access under CC BY 4.0. Quoted from the PMC open-access text. How to recognise an unscientific colour map, second check. https://pmc.ncbi.nlm.nih.gov/articles/PMC7595127/ (accessed 2026-10-02).

316. Wilke, Claus O. Chapter 19: Common pitfalls of color use. 2019. Chapter 19, section 19.1 Encoding too much or irrelevant information, second paragraph. https://clauswilke.com/dataviz/color-pitfalls.html (accessed 2026-10-02).

317. Wilke, Claus O. Chapter 19: Common pitfalls of color use. 2019. Chapter 19, section 19.1, the boxed rule on saturated colours. https://clauswilke.com/dataviz/color-pitfalls.html (accessed 2026-10-02).

318. Wilke, Claus O. Chapter 19: Common pitfalls of color use. 2019. Chapter 19, section 19.3, paragraph after Table 19.1. https://clauswilke.com/dataviz/color-pitfalls.html (accessed 2026-10-02).

319. Wilke, Claus O. Chapter 20: Redundant coding. 2019. Chapter 20, Redundant coding, opening paragraph. https://clauswilke.com/dataviz/redundant-coding.html (accessed 2026-10-02).

320. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Introduction, fieldwork paragraph, last sentence. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

321. Krishnamurthy, Sruthi Sree and Rangavittal, Subhiksha and Chandrasekar, Ambika and Narayanan, Anuradha. Prevalence of color vision deficiency among school-going boys in South India. PMID 34304169, PMC8482944. Open access under CC BY-NC-SA 4.0. Quoted from the PMC open-access text. Abstract, Methods. https://pmc.ncbi.nlm.nih.gov/articles/PMC8482944/ (accessed 2026-10-02).

322. Crameri, Fabio and Shephard, Grace E and Heron, Philip J. The misuse of colour in science communication. PMID 33116149, PMC7595127. Open access under CC BY 4.0. Quoted from the PMC open-access text. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC7595127/ (accessed 2026-10-02).

323. Krishnamurthy, Sruthi Sree and Rangavittal, Subhiksha and Chandrasekar, Ambika and Narayanan, Anuradha. Prevalence of color vision deficiency among school-going boys in South India. PMID 34304169, PMC8482944. Open access under CC BY-NC-SA 4.0. Quoted from the PMC open-access text. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC8482944/ (accessed 2026-10-02).

324. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Rule 4: Captions Are Not Optional. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

325. Wilke, Claus O. Chapter 22: Titles, captions, and tables. 2019. Chapter 22, section Figure titles and captions, last paragraph. https://clauswilke.com/dataviz/figure-titles-captions.html (accessed 2026-10-02).

326. Wilke, Claus O. Chapter 22: Titles, captions, and tables. 2019. Chapter 22, section Axis and legend titles, second paragraph. https://clauswilke.com/dataviz/figure-titles-captions.html (accessed 2026-10-02).

327. Wilke, Claus O. Chapter 22: Titles, captions, and tables. 2019. Chapter 22, section Axis and legend titles, second paragraph, last sentence. https://clauswilke.com/dataviz/figure-titles-captions.html (accessed 2026-10-02).

328. Wilke, Claus O. Chapter 20: Redundant coding. 2019. Chapter 20, section 20.2 Designing figures without legends, second paragraph. https://clauswilke.com/dataviz/redundant-coding.html (accessed 2026-10-02).

329. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.i Illustrations (Figures), third paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

330. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.i Illustrations (Figures), sixth paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

331. Cumming, Geoff and Fidler, Fiona and Vaux, David L. Error bars in experimental biology. PMID 17420288, PMC2064100. CC BY-NC-SA 4.0 (after the first six months). Quoted from the PMC open-access text. What are error bars for?, first paragraph. https://pmc.ncbi.nlm.nih.gov/articles/PMC2064100/ (accessed 2026-10-02).

332. Cumming, Geoff and Fidler, Fiona and Vaux, David L. Error bars in experimental biology. PMID 17420288, PMC2064100. CC BY-NC-SA 4.0 (after the first six months). Quoted from the PMC open-access text. Table I, Common error bars, standard error row. https://pmc.ncbi.nlm.nih.gov/articles/PMC2064100/ (accessed 2026-10-02).

333. Cumming, Geoff and Fidler, Fiona and Vaux, David L. Error bars in experimental biology. PMID 17420288, PMC2064100. CC BY-NC-SA 4.0 (after the first six months). Quoted from the PMC open-access text. Table I, Common error bars, confidence interval row. https://pmc.ncbi.nlm.nih.gov/articles/PMC2064100/ (accessed 2026-10-02).

334. Cumming, Geoff and Fidler, Fiona and Vaux, David L. Error bars in experimental biology. PMID 17420288, PMC2064100. CC BY-NC-SA 4.0 (after the first six months). Quoted from the PMC open-access text. Figure 2 caption, Confidence intervals. https://pmc.ncbi.nlm.nih.gov/articles/PMC2064100/ (accessed 2026-10-02).

335. Cumming, Geoff and Fidler, Fiona and Vaux, David L. Error bars in experimental biology. PMID 17420288, PMC2064100. CC BY-NC-SA 4.0 (after the first six months). Quoted from the PMC open-access text. Inferential error bars, Rule 1. https://pmc.ncbi.nlm.nih.gov/articles/PMC2064100/ (accessed 2026-10-02).

336. Cumming, Geoff and Fidler, Fiona and Vaux, David L. Error bars in experimental biology. PMID 17420288, PMC2064100. CC BY-NC-SA 4.0 (after the first six months). Quoted from the PMC open-access text. Replicates or independent samples—what is n?, Rule 2. https://pmc.ncbi.nlm.nih.gov/articles/PMC2064100/ (accessed 2026-10-02).

337. Cumming, Geoff and Fidler, Fiona and Vaux, David L. Error bars in experimental biology. PMID 17420288, PMC2064100. CC BY-NC-SA 4.0 (after the first six months). Quoted from the PMC open-access text. Replicates or inde-

pendent samples—what is n?, paragraph after Rule 2. https://pmc.ncbi.nlm.nih.gov/articles/PMC2064100/ (accessed 2026-10-02).

338. Wilke, Claus O. Chapter 24: Use larger axis labels. 2019. Chapter 24, Use larger axis labels, the boxed rule. https://clauswilke.com/dataviz/small-axis-labels.html (accessed 2026-10-02).

339. Cumming, Geoff and Fidler, Fiona and Vaux, David L. Error bars in experimental biology. PMID 17420288, PMC2064100. CC BY-NC-SA 4.0 (after the first six months). Quoted from the PMC open-access text. Descriptive error bars, paragraph giving the SD formula. https://pmc.ncbi.nlm.nih.gov/articles/PMC2064100/ (accessed 2026-10-02).

340. Cumming, Geoff and Fidler, Fiona and Vaux, David L. Error bars in experimental biology. PMID 17420288, PMC2064100. CC BY-NC-SA 4.0 (after the first six months). Quoted from the PMC open-access text. What type of error bar should be used?, Rule 4. https://pmc.ncbi.nlm.nih.gov/articles/PMC2064100/ (accessed 2026-10-02).

341. Cumming, Geoff and Fidler, Fiona and Vaux, David L. Error bars in experimental biology. PMID 17420288, PMC2064100. CC BY-NC-SA 4.0 (after the first six months). Quoted from the PMC open-access text. Confidence interval (CI), paragraph on the critical value of t. https://pmc.ncbi.nlm.nih.gov/articles/PMC2064100/ (accessed 2026-10-02).

342. Cumming, Geoff and Fidler, Fiona and Vaux, David L. Error bars in experimental biology. PMID 17420288, PMC2064100. CC BY-NC-SA 4.0 (after the first six months). Quoted from the PMC open-access text. Figure 4 caption, Inferential error bars. https://pmc.ncbi.nlm.nih.gov/articles/PMC2064100/ (accessed 2026-10-02).

343. Cumming, Geoff and Fidler, Fiona and Vaux, David L. Error bars in experimental biology. PMID 17420288, PMC2064100. CC BY-NC-SA 4.0 (after the first six months). Quoted from the PMC open-access text. Repeated measurements of the same group, Rule 8. https://pmc.ncbi.nlm.nih.gov/articles/PMC2064100/ (accessed 2026-10-02).

344. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.d Methods, iii Statistics. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

345. Wilke, Claus O. Chapter 24: Use larger axis labels. 2019. Chapter 24, Use larger axis labels, first paragraph. https://clauswilke.com/dataviz/small-axis-labels.html (accessed 2026-10-02).

346. Cumming, Geoff and Fidler, Fiona and Vaux, David L. Error bars in experimental biology. PMID 17420288, PMC2064100. CC BY-NC-SA 4.0 (after the first six months). Quoted from the PMC open-access text. figures quoted in the text. https://pmc.ncbi.nlm.nih.gov/articles/PMC2064100/ (accessed 2026-10-02).

347. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.h Tables, second paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

348. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.h Tables, third paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

349. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.h Tables, third paragraph, last sentence. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

350. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. IV.A.3.h Tables, fourth paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

351. Wilke, Claus O. Chapter 22: Titles, captions, and tables. 2019. Chapter 22, section Tables, the numbered rules for table layout. https://clauswilke.com/dataviz/figure-titles-captions.html (accessed 2026-10-02).

352. Wilke, Claus O. Chapter 22: Titles, captions, and tables. 2019. Chapter 22, section Tables, the numbered rules, rule 6. https://clauswilke.com/dataviz/figure-titles-captions.html (accessed 2026-10-02).

353. Wilke, Claus O. Chapter 22: Titles, captions, and tables. 2019. Chapter 22, section Tables, last paragraph. https://clauswilke.com/dataviz/figure-titles-captions.html (accessed 2026-10-02).

354. Wilke, Claus O. Chapter 22: Titles, captions, and tables. 2019. Chapter 22, section Tables, first paragraph. https://clauswilke.com/dataviz/figure-titles-captions.html (accessed 2026-10-02).

355. Krishnamurthy, Sruthi Sree and Rangavittal, Subhiksha and Chandrasekar, Ambika and Narayanan, Anuradha. Prevalence of color vision deficiency among school-going boys in South India. PMID 34304169, PMC8482944. Open access under CC BY-NC-SA 4.0. Quoted from the PMC open-access text. Results, Table 1 title. https://pmc.ncbi.nlm.nih.gov/articles/PMC8482944/ (accessed 2026-10-02).

356. Krishnamurthy, Sruthi Sree and Rangavittal, Subhiksha and Chandrasekar, Ambika and Narayanan, Anuradha. Prevalence of color vision deficiency among school-going boys in South India. PMID 34304169, PMC8482944. Open access under CC BY-NC-SA 4.0. Quoted from the PMC open-access text. Results, first paragraph (school type counts in the text). https://pmc.ncbi.nlm.nih.gov/articles/PMC8482944/ (accessed 2026-10-02).

357. Krishnamurthy, Sruthi Sree and Rangavittal, Subhiksha and Chandrasekar, Ambika and Narayanan, Anuradha. Prevalence of color vision deficiency among school-going boys in South India. PMID 34304169, PMC8482944. Open access under CC BY-NC-SA 4.0. Quoted from the PMC open-access text. Methods, Definitions, last paragraph (school levels). https://pmc.ncbi.nlm.nih.gov/articles/PMC8482944/ (accessed 2026-10-02).

358. Cole, T J. Too many digits: the presentation of numerical data. 2015. Body, the paragraph after the definitions of decimal places and significant digits. https://pmc.ncbi.nlm.nih.gov/articles/PMC4483789/ (accessed 2026-10-02).

359. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Rule 2: Identify Your Message. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

360. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Rule 5: Do Not Trust the Defaults. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

361. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Rule 5: Do Not Trust the Defaults, second sentence. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

362. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Rule 5: Do Not Trust the Defaults, the list of packages. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

363. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Rule 7: Do Not Mislead the Reader. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

364. Rougier, Nicolas P and Droettboom, Michael and Bourne, Philip E. Ten simple rules for better figures. 2014. Rule 10: Get the Right Tool. https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1003833 (accessed 2026-10-02).

365. Wilke, Claus O. Preface. 2019. Preface, the paragraph on interactive plot programs. https://clauswilke.com/dataviz/preface.html (accessed 2026-10-02).

366. Wilke, Claus O. Preface. 2019. Preface, the paragraph on automation. https://clauswilke.com/dataviz/preface.html (accessed 2026-10-02).

367. Wilke, Claus O. Preface. 2019. Preface, the paragraph on R and ggplot2. https://clauswilke.com/dataviz/preface.html (accessed 2026-10-02).

368. Wilke, Claus O. Chapter 28: Choosing the right visualization software. 2019. Chapter 28, opening section, the sentence set on its own. https://clauswilke.com/dataviz/choosing-visualization-software.html (accessed 2026-10-02).

369. Wilke, Claus O. Chapter 28: Choosing the right visualization software. 2019. §28.3 Separation of content and design, first paragraph. https://clauswilke.com/dataviz/choosing-visualization-software.html (accessed 2026-10-02).

370. Wilke, Claus O. Chapter 28: Choosing the right visualization software. 2019. §28.3 Separation of content and design, first paragraph, next sentence. https://clauswilke.com/dataviz/choosing-visualization-software.html (accessed 2026-10-02).

371. Wilke, Claus O. Chapter 28: Choosing the right visualization software. 2019. §28.3, last paragraph. https://clauswilke.com/dataviz/choosing-visualization-software.html (accessed 2026-10-02).

372. Wilke, Claus O. Chapter 28: Choosing the right visualization software. 2019. §28.2 Data exploration versus data presentation, the paragraph on manual sprucing up. https://clauswilke.com/dataviz/choosing-visualization-software.html (accessed 2026-10-02).

373. Wilke, Claus O. Chapter 24: Use larger axis labels. 2019. Chapter 24, first paragraph. https://clauswilke.com/dataviz/small-axis-labels.html (accessed 2026-10-02).

374. Wilke, Claus O. Chapter 27: Understanding the most commonly used image file formats. 2019. §27.1 Bitmap and vector graphics, first paragraph. https://clauswilke.com/dataviz/image-file-formats.html (accessed 2026-10-02).

375. Wilke, Claus O. Chapter 27: Understanding the most commonly used image file formats. 2019. §27.1, the paragraph after Table 27.1. https://clauswilke.com/dataviz/image-file-formats.html (accessed 2026-10-02).

376. Wilke, Claus O. Chapter 27: Understanding the most commonly used image file formats. 2019. Chapter 27, second paragraph. https://clauswilke.com/dataviz/image-file-formats.html (accessed 2026-10-02).

377. Wilke, Claus O. Chapter 27: Understanding the most commonly used image file formats. 2019. §27.2 Lossless and lossy compression, the paragraph on avoiding jpeg. https://clauswilke.com/dataviz/image-file-formats.html (accessed 2026-10-02).

378. Wilke, Claus O. Chapter 27: Understanding the most commonly used image file formats. 2019. §27.3 Converting between image formats. https://clauswilke.com/dataviz/image-file-formats.html (accessed 2026-10-02).

379. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. Section IV.A.3.i Illustrations (Figures), first paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

380. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. Section IV.A.3.i Illustrations (Figures), third paragraph. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

381. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. India Key Indicators table, indicator 88. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

382. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. India Key Indicators table, indicator 89. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

383. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. India Key Indicators table, the heading above indicator 86. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

384. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Footnote 21 to indicators 86 and 88. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

385. International Institute for Population Sciences and Ministry of Health and Family Welfare. National Family Health Survey (NFHS-5), 2019-21: India Fact Sheet. 2021. Introduction, fieldwork dates and sample. https://dhsprogram.com/pubs/pdf/OF43/India_National_Fact_Sheet.pdf (accessed 2026-09-24).

386. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 10: Get feedback to reduce, reuse, and recycle the story. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

387. Mensh, Brett and Kording, Konrad. Ten simple rules for structuring papers. 2017. Rule 9: Allocate time where it matters, the paragraph on outlines. https://doi.org/10.1371/journal.pcbi.1005619 (accessed 2026-10-02).

388. Cole, T J. Too many digits: the presentation of numerical data. 2015. Body, the paragraph on comparing groups (the range rule for percentages). https://pmc.ncbi.nlm.nih.gov/articles/PMC4483789/ (accessed 2026-10-02).

389. Cole, T J. Too many digits: the presentation of numerical data. 2015. Body, the paragraph introducing Table 1. https://pmc.ncbi.nlm.nih.gov/articles/PMC4483789/ (accessed 2026-10-02).

390. International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals. IV.A: Preparing a Manuscript for Submission to a Medical Journal. 2026. Section IV.A.3.e Results. ICMJE website. https://www.icmje.org/recommendations/browse/manuscript-preparation/preparing-for-submission.html (accessed 2026-10-02).

391. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 1 Journals, part A, Examples of Citations to Journal Articles, first list. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

392. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 22, part A, introduction, first paragraph. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

393. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 22, part A, Examples, 7. Book on the Internet with an organization(s) as author, second item. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

394. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 22, part A, Citation Rules, General Rules for Publisher. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

395. Patrias, Karen. Citing Medicine: The NLM Style Guide for Authors, Editors, and Publishers. 2007. Chapter 4, part A, Citation Rules, General Rules for Report Title. https://www.ncbi.nlm.nih.gov/books/NBK7256/ (accessed 2026-10-02).

396. Flesch, Rudolf. How to Write Plain English: A Book for Lawyers and Consumers. 1979. Chapter 2, the score bands. https://pages.stern.nyu.edu/%7Ewstarbuc/Writing/Flesch.htm (accessed 2026-10-02).

Appendix · Worked answers

1 · What a research paper is: its parts and what each part is for

EXERCISE 1. Introduction: why was the study done, and what did it set out to find? It gives the context and states the purpose or hypothesis.

Methods: how was it done? Enough detail for others with the data to reproduce the results, and only what was known when the protocol was written.

Results: what was found? The findings in logical order, the most important first.

Discussion: what does it mean, and where does it stop? The main findings, set among the other evidence, with the study's limitations and what follows from it.

Around them: the title page with the title, the abstract, and the list of references.

EXERCISE 2. Sentences (a) and (b) belong. Who was eligible and how waist was measured were both fixed when the protocol was written.

Sentence (c) goes to the results. You learn that 214 of the 260 were measured only by doing the study. The Recommendations put everything obtained during the study in the results. The methods may say how adults were approached; the count of who took part is a result.

Sentence (d) splits. "Waist was above the clinic's cut-off in 58 per cent of those measured" is a finding and goes to the results. The words "more than we expected" are interpretation and go to the discussion, where they need a reason: expected from what?

EXERCISE 3. Start with how papers are read. Most readers do not go from start to finish. They read the title and abstract, then jump to the part they want. They go to the methods to judge how it was done, and to the results for the numbers.

Mensh and Kording note that the title, abstract and figures are seen by far more people than the rest. The methods are read least of all. The four fixed sections let each reader find their part. Editors, by one account, pushed the structure to help readers and reviewers.

Then the history, so she does not mistake it for a law. In four leading medical journals it was absent in 1935 and became the only pattern by the 1980s. It is a convention, and a useful one.

Draw the four boxes and write one question in each: why, how, what was found, what it means. Tell her a sentence belongs where its question is.

Then the exceptions. A case report, a narrative review or an editorial need not follow it. A meta-analysis may need a different format. The journal's own instructions for her article type decide.

2 · A paper is an argument, not a report of activities

EXERCISE 1. The claim is the conclusion. The data are the premises. The reasoning that carries the data to the claim is the inference.

It fails if a premise is false: the data are wrong, for example mismeasured or miscounted. It also fails if the inference does not carry: the data are right but do not support the claim. The two are separate questions, and both must hold.

EXERCISE 2. First find the claim. The counts fall at each step, and height is where most are lost. A body mass index needs both weight and height, so it could be worked out for at most 152 of the 640.

Work out the shares, so the reader does not have to.

```
410 divided by 640 = 0.641, about 64 per cent
152 divided by 640 = 0.2375, about 24 per cent
96 divided by 640 = 0.15, which is 15 per cent
```

> *Most adults attending the clinic could not have their body mass index worked out from the register, because their height was not recorded. Of 640 adults registered from January to March, 410 (64 per cent) had a weight recorded. Only 152 (24 per cent) had a height. A body mass index was written for 96 (15 per cent).*

The first sentence of the old paragraph, which registers were reviewed and for which months, goes to the methods.

EXERCISE 3. Check your rewrite four ways.

- Its first sentence states what you found. A reader who stops there knows your claim.
- Every other sentence is evidence for that claim, in the order the claim needs. No sentence is left without a job.
- No S sentence remains. Each one is in the methods, once, or deleted because the methods already say it.
- Any comparison you made is finished. The paragraph prints its result and does not leave the reader to subtract.

If your paragraph had two claims, it is now two paragraphs.

EXERCISE 4. Right: the examiner must be able to see the work. Wrong: the results are not where it goes.

Draw two boxes, methods and results. The methods say what was done, in enough detail to judge and repeat it. That is where the examiner checks the work.

The results say what was found. Walking through the steps in the results makes the examiner read the work twice and still hunt for the findings. Mensh and Kording put it plainly: readers care about the claim and the logic supporting it, not the path you took. In this book, each results paragraph opens on its finding, then gives the numbers that support it.

Then show her one paragraph both ways: first the diary, then the claim with its evidence.

3 · The one message: your argument in one or two sentences, written first

EXERCISE 1. First, it must be a claim: it would be false had the data come out differently. Second, it must answer "so what?" for a reader told only that sentence.

The abstract: the context, the gap, what was done, the main result, and what it means.

EXERCISE 2. (c) is a message. Had body mass index flagged as many people as waist did, or more, its first sentence would be false. So it claims something. Its second sentence answers "so what?". It follows from the first. If waist flags more people than body mass index does, some of the people waist flags must be missed by body mass index.

Message (a) is a topic. It is true whatever the thesis found.

Message (b) is a list of what was measured. It is also true whatever was found, and it does not say which finding the thesis is for.

What (c) does not show: that the extra people flagged by waist are at higher risk of disease. That needs data on what happens to them, which this design does not collect.

EXERCISE 3. Check your work in this order.

- Message: would it be false had your data come out differently? Does it say why it matters? Does it claim no more than the data carry, with no cause where you measured only an association?
- Title: does it state the finding, with no abbreviation, in words a searcher would type? Does it name the design, if your journal asks?
- Abstract: is there one sentence each for the context, the gap, what you did, the main result and the meaning? Does the meaning say what the finding changes for someone, and answer the question the context posed?
- Agreement: is every number in the title and abstract the same as in your results?

EXERCISE 4. The quotable sentence: "In the national family health surveys, a larger share of adults aged 15 to 49 had overweight or obesity in 2019-21 than in 2015-16. By 2019-21 the share was about one woman in four, and more than one man in five."

Then the confidence, in plain words. These are large national surveys, and the rise appears in both women and men. But the figures count overweight and obesity together. They cover ages 15 to 49 only. They tell us the share rose, not why.

If asked "why", say that these surveys cannot answer it, and do not guess.

4 · The argument across sections: from the gap to the answer and its limits

EXERCISE 1. Introduction: move from what is known to the gap, and end on the study's question, research objective or hypothesis.

Methods: say how and why the study was done. Give enough detail to judge it and, with the data, to reproduce it. Include only what was known when the protocol was written.

Results: the findings as a sequence of claims, each with its data, the main findings first.

Discussion: open with the main finding, as the answer. Then give possible explanations, the other evidence, the limits, and what it means, claiming no more than the data carry.

The number who declined goes in the results. It was learnt during the study, and ICMJE puts everything obtained during the study there.

EXERCISE 2. It never answers the question. The first two sentences are background, which belongs in the introduction, if anywhere. The third repeats the methods. The fourth points to a table instead of saying what the table shows. The last sentence asks for more research without saying what research, or why. A reader finishes the paragraph not knowing what the study found.

A rewrite opens with the answer and makes one point:

1. "In this rural block, [the share of adults with overweight or obesity, with its count and denominator]."
2. "[Whether and how it differed by age group, with the numbers]."
3. "This is [above, below or close to] the rural figure for the state in NFHS-5 (reference)."
4. "It adds [what this answer adds to what was known]."

Every bracket is filled from the paper's own results or from a cited source. The explanations, limits and the "further research" go in later paragraphs. Each of those names the specific question that remains open.

EXERCISE 3. Judge your outline with four checks.

- Every sentence is a claim. A sentence such as "Table 2" or "Sociodemographic profile" is a topic. Rewrite it as what the table shows.
- The introduction's sentences narrow to one gap, and its last sentence states the question, research objective or hypothesis.
- Each results sentence answers part of that question, in an order where later sentences lean on earlier ones. Nothing learnt during the study sits in the methods.
- The discussion's first sentence answers the whole question. A later sentence names each limit and which way it could push the answer.

Mark every claim that has no data or reference behind it yet. Those marks are your work list. Do not write the paragraphs until the list is empty or the claims are cut.

EXERCISE 4. Draw a funnel on the board. At the top write what is known; below it, what is not known; at the bottom, her question. An introduction is that funnel and nothing else. It earns the reader's attention by showing a gap that matters, then says what the study asks about it.

Then work through her four pages with her. For each paragraph, ask one question: does this lead to the gap? The definition of obesity and its complications do not, unless her question is about them. A sentence on worldwide trends usually does not. What leads to the gap is what is known about her exact question, and the specific thing nobody has measured.

Tell her what ICMJE asks: directly pertinent references only, and no data or conclusions from her own study. Her last paragraph states the research objective. Mensh and Kording's own example narrows to its gap in three paragraphs: the field, the subfield, then the untested idea. The cut pages are not wasted. Whatever is needed to explain her findings goes to the discussion, and the rest goes into a file of her notes.

5 · Citing and referencing: the original source, in the order cited, in one format

EXERCISE 1. References are numbered in the order the text first mentions them. A source cited again keeps its first number. A source cited only in a table or a figure legend gets its number where the text first mentions that table or figure. In the text, each number is in parentheses, such as (3).

The elements, in order: authors; article title; journal title, abbreviated; date; volume; issue, in parentheses; pages.

EXERCISE 2. Entry 1:

> *Cole TJ. Too many digits: the presentation of numerical data. Arch Dis Child. 2015;100(7):608-9.*

The initials lose their full stops and come straight after the surname. The year moves to the date position. The title keeps a capital only on its first word. The journal title is abbreviated; confirm "Arch Dis Child" in the NLM Catalog. The page range 608-609 becomes 608-9. No month is given, so the date is the year alone, as in NLM's sample entry "Brain Res. 2002;935(1-2):40-6."

Entry 2:

> *Weissgerber TL, Milic NM, Winham SJ, Garovic VD. Beyond bar and line graphs: time for a new data presentation paradigm. PLoS Biol. 2015;13(4):e1002128.*

Four authors is fewer than six, so all four are listed and "et al." goes. The title drops its capitals after the first word, keeping the colon. The issue, 4, was missing; it goes in parentheses after the volume. A full stop follows the journal title, and the spaces round the semicolon go. Confirm "PLoS Biol" in the NLM Catalog.

Last, check both against the papers themselves or PubMed, not against your co-author's list.

EXERCISE 3. Three numbers are wrong. R is numbered where the text first mentions Table 1, which is sentence 3. R is the third source met, so it is (3), not (5). S, printed (3), is (4): it is first met in sentence 4, after R. And P in sentence 5 is a source already cited, so it keeps its first number, (1), not (4).

The correct numbers are P (1), Q (2), R (3), S (4), and P again (1). The reference list runs P, Q, R, S.

EXERCISE 4. Start with what a reference promises. It tells the reader where the evidence for this sentence is. If you cite the review, you send them to someone else's account of the survey. ICMJE warns that reviews do not always report the original accurately.

Then show the difference on the board. Write "about a quarter of Indian women have overweight or obesity", the kind of line a review might carry. Under it, write what the fact sheet's indicator 88 says. It gives 24.0 per cent of women aged 15-49 in 2019-21: 33.2 per cent in urban areas and 19.7 per cent in rural areas. The age range, the dates and the urban-rural split were lost in the second-hand line. Ask her which sentence she would rather defend in her viva.

Then the rule: open the fact sheet, read the row, write the sentence from it, and cite the fact sheet. Cite the review only for what the review itself adds, such as its own judgement across several studies.

6 · The parts of a sentence a writer works with

EXERCISE 1. The first sign is a form of "to be": is, was, were, has been. The second is a past participle, the form that follows "has" or "was", which usually ends in -ed. Both may not appear in every passive sentence.

Voice and tense are different things. Voice says whether the subject did the action or received it. Tense says when. "Weight is measured" is passive and present. "We measured weight" is active and past.

EXERCISE 2. Sentence by sentence:

1. Verb: "weighed". Subject: "the auxiliary nurse midwife", main word "midwife". Object: "every child". Active voice. The doer is the subject, the midwife. One clause.
2. Two clauses. The main clause is "Three villages were left out". Its verb is "were left out", its subject "three villages", main word "villages". It is passive, and it names no doer: the sentence does not say who left the villages out.

   The subordinate clause is "because the road was closed". Its subject is "the road" and its verb "was closed". It may be passive (someone closed the road) or describe how the road was (shut). The sentence does not tell you which.

   A reader judging the survey would want to know who decided to drop the villages.
3. Verb: "were". The word "there" stands where a subject usually stands, but it names nothing. Ask what there were: households. The subject is "many households without a weighing scale", main word "households". No object, and no action, so no doer and no voice to change. A rewrite gives the sentence a real subject and verb: "Many households had no weighing scale."

EXERCISE 3. There is no single right list, so check your own against three tests.

- You found the verb first, then the subject, then looked for "by". A passive with no "by" is the kind most worth marking.
- For each marked verb, the doer is something a reviewer could ask about. Who measured? Who interviewed? Who entered or cleaned the data? Who decided which participants to leave out?
- Verbs whose doer does not change the result are left unmarked. "The two readings were averaged" gives the same average whoever did the sum.

If you could not name the doer of a marked verb, that is a gap in your methods, not only in your sentence. Find out, and write it down.

7 · The sentence as the unit of clarity

EXERCISE 1. The most weight goes on the end of the sentence, the stress position. Put there the new information you want the reader to weigh most.

The link to what came before is looked for at the start, the topic position. Put there the old information, which links back to what the reader already has, and the person or thing whose story the sentence tells.

EXERCISE 2. The verb is “was done”. Ask what was done: assessment. The subject’s main word is “Assessment”, the first word. Count the words between “Assessment” and “was done”. There are 17: “of the adherence of the patients attending the hypertension clinic to their prescribed medicines by pill count”.

The hidden verb is “assess”, turned into the noun “assessment” and carried by the empty verb “was done”. There is a second one: “adherence” hides “adhere”, to take the medicines as prescribed. Both sit between “the” and “of”, where hidden verbs often sit.

The doer is the staff nurse, waiting at the end after “by”.

One rewrite:

> *At every visit, the staff nurse counted each patient’s remaining pills to check whether they were taking their medicines as prescribed.*

The subject is now “the staff nurse”, the doer, and the verb “counted” is the action. What the count is for comes last. Other rewrites are as good if they put the doer first and the action in the verb.

EXERCISE 3. Start with the test, not the rule. Ask of each verb: does a reader judging the study need to know who did it?

For centrifuging within an hour, a reader may want to know who was responsible for the timing, because a delay changes the result. If that matters, the doer belongs in the sentence, and the active voice puts it there.

For labelling, the doer matters less. Whoever writes the code, the label is the same. The passive keeps the story on the samples, which is what this paragraph is about.

Then show what her change cost. Both sentences now begin “The laboratory technician”, so the paragraph has become a story about the technician instead of the samples. Gopen and Swan’s pair makes the same point. “Bees disperse pollen” and “Pollen is dispersed by bees” are each right in a paragraph about their own first word.

One good version: “The laboratory technician centrifuged each blood sample within one hour of collection. Each sample was then labelled with the participant’s code.” The doer is named where it matters, and the second sentence carries on the samples’ story.

EXERCISE 4. Check your rewrites against four questions.

- Is the subject next to its verb, or close to it, with any list or bracket moved to the end or to its own sentence?
- Is the action a verb (“rose”, “fell”, “differed”) rather than a noun carried by “was observed” or “was found”?
- Does each sentence end on the new information, usually the number or the finding, and not on a source, a table number or “was seen”?
- Does each sentence carry one point? If a rewrite holds two, split it, and carry any “if” or “because” link into the second sentence.

Read your five rewrites aloud in order. If they now read as a sequence, with each start linking to the end of the sentence before, they pass.

8 · Words: the everyday word, one term for one thing, abbreviations that cost the reader

EXERCISE 1. At first mention, the term spelled out, followed by the abbreviation in brackets. A standard unit of measurement is the exception. No abbreviations in the title.

Any four of: random, normal, significant, correlation, sample.

EXERCISE 2. Five problems.

1. "Significantly" makes a statistical claim. No test is reported, so it has to go, or the test has to be named.
2. "Obesity" is not what the numbers measure. Indicator 89 of the fact sheet is "Men who are overweight or obese (BMI ≥25.0 kg/m2)". These are the urban and rural figures for overweight or obesity together.
3. "Men", then "males": two words for one group. Keep one.
4. "High BMI" is a third name for the same condition.
5. "Correlated with urbanisation" claims a correlation was computed. Two numbers, urban and rural, cannot give one. Describe the difference instead.

A rewrite:

> *Among men aged 15 to 49, overweight or obesity was more common in urban areas (29.8 per cent) than in rural areas (19.3 per cent).*

The rewrite says less than the original, and all of it is true.

EXERCISE 3. Barnett and Doubleday counted acronyms in 24,873,372 titles and 18,249,091 abstracts in PubMed from 1950 to 2019. They measured how often acronyms appear (per 100 words), how many different ones there are, and how often each is used again. In abstracts, acronyms rose from 0.4 per 100 words in 1956, the first year with enough abstracts, to 4.1 in 2019.

They did not test any reader. "Harder to understand" is a reasonable worry, and the authors give examples and a reason, but the study does not measure it. Their own abstract says acronyms are "not the biggest current problem in science communication".

Two smaller errors. The 24 million are titles; abstracts were about 18 million. And the count covers titles and abstracts in English, not whole papers.

A fair slide: "Acronyms in medical abstracts grew about ten times as dense between 1956 and 2019, and most are used only a few times."

EXERCISE 4. Check your list against four tests.

- No more than three abbreviations stay, each standard for your reader or used often in the section. Each is spelled out at first mention, and none is in the title.
- Every abbreviation that appears only once or twice is spelled out instead, or given a short plain name such as "the centre" or "the survey".
- The outcome has one term. Every other word you found for it (excess weight, high BMI, obesity when you mean overweight or obesity) is replaced.

- The defining sentence gives the term, its cut-off or rule, and its source. The sentence under "Reporting it" in this section is one model.

Then search the section for "significant", "random", "normal" and "correlat". Each hit either names its test, process or number, or is rewritten.

9 · Numbers in a sentence: the unit, the count behind the percentage, and no more digits than the data hold

EXERCISE 1. Its unit; for a percentage, the counts it came from (numerator and denominator); and no more digits than the data support.

A mean gets enough decimal places to give the SD two significant figures, or the SE one. A percentage is a whole number, or has one decimal place below 10 per cent. A P value is never printed as 0.000.

EXERCISE 2. Work out what the counts give.

```
17 divided by 87 is about 0.1954
11 divided by 42 is about 0.2619
6 divided by 45 is about 0.1333
26.19 minus 13.33 = 12.86
```

Four faults. First, "19.54 per cent" has no count, and two decimal places from 87 students. Each student is worth more than one point, so write "17 of 87 (20 per cent)". Second, the mean age has four decimal places and the "±" form. The SD 1.2066 to two significant figures is 1.2, so write "13.5 years (SD 1.2)".

Third, "9.2 per cent more" matches neither the points nor the per cent the counts give. The counts give 26 per cent of girls and 13 per cent of boys, about 13 percentage points apart. Fourth, "p = 0.000" is never printed. If the test gave less than 0.001, write P < 0.001.

A rewrite with one point:

> *Of the 87 students, 17 (20 per cent) were at or above the protocol's cut-off for overweight. They were 11 of 42 girls (26 per cent) and 6 of 45 boys (13 per cent). Their mean age was 13.5 years (SD 1.2).*

The comparison of girls and boys belongs with its test, reported to one rule, in a sentence of its own.

EXERCISE 3. Not quite, and the difference matters. The survey measured women aged 15 to 49, and it counts overweight and obesity together; it has no figure for obesity alone. In the 2015-16 survey, 20.6 in every hundred women of that age were in that group. In the 2019-21 survey it was 24.0 in every hundred. So about 3 or 4 more women in every hundred. Your 17 per cent is that rise measured against the starting figure. It is true, but it sounds far bigger than it is.

The sentence to quote: "In the 2019-21 survey, about one woman in four aged 15 to 49 had overweight or obesity. In 2015-16 it was about one in five."

Calibrated: both figures come from national surveys. They tell you how the share changed, not why.

1.

```
0.04372 rounds to 0.044
3465.31 rounds to 3500
22.68 rounds to 23
0.5049 rounds to 0.50
```

The leading zeros of 0.04372 do not count. Its first two significant figures are 4 and 3, and the 7 after them rounds the 3 up to 4. In 0.50 the final zero is significant, so it is kept.

2.

```
37 divided by 412 is about 0.0898
0.0898 times 100 = 8.98
```

8.98 per cent is under 10 per cent, so it keeps one decimal place: 9.0 per cent. Write “37 of 412 (9.0 per cent)”.

3.

```
15.0 minus 12.5 = 2.5
2.5 divided by 12.5 = 0.2
0.2 times 100 = 20
```

A rise of 2.5 percentage points, which is a rise of 20 per cent on the starting figure.

4. Round each SD to two significant figures, then give the mean the same decimal places.

```
11.46 rounds to 11, no decimal places, so 52.381 rounds to 52
0.5712 rounds to 0.57, two decimal places, so 3.0538 rounds to 3.05
0.0236 rounds to 0.024, three decimal places, so 0.8147 rounds to 0.815
```

52 (SD 11), 3.05 (SD 0.57) and 0.815 (SD 0.024).

5. Look at the first digit that is not zero. If it is 4 or more, keep two significant figures; otherwise keep three.

```
0.3172: first digit 3, keep three, 0.317
1.0462: first digit 1, keep three, 1.05
4.183: first digit 4, keep two, 4.2
11.27: first digit 1, keep three, 11.3
0.0482: first digit 4, keep two, 0.048
```

6.

```
33.2 minus 19.7 = 13.5
33.2 divided by 19.7 = 1.685
```

The ratio’s first digit is 1, so keep three significant figures: 1.69. The range across the two groups is 13.5, which is 10 or more. So Cole’s suggestion gives whole numbers: 33 per cent and 20 per cent.

> *In NFHS-5 (2019-21), 33 per cent of urban women aged 15-49 had a BMI of 25.0 kg/m^2 or more. Among rural women it was 20 per cent. The gap was about 13 percentage points.*

Whole numbers blur the gap a little: 33 minus 20 is 13, against 13.5 from the printed figures. If the gap is your point, keep the fact sheet’s one decimal place and write 13.5 points. Either way, write the gap in points before any ratio.

7. One kilogram is 1000 grams.

```
3.05 times 1000 = 3050
0.57 times 1000 = 570
```

The SD of 570 g has two significant figures, the 5 and the 7, and its last significant digit is in the tens. So the mean goes to the nearest 10 g: 3050 g. Write "3050 g (SD 570 g)". Cole reaches the same place: birth weight in grams "ought to be rounded to the nearest 10 g".

8. The SD rule: 9.8364 to two significant figures is 9.8, one decimal place, so 34.3 years.

The SE rule, with the SE as the SD divided by the square root of the count.

```
square root of 100 = 10
9.8364 divided by 10 = 0.98364
square root of 10,000 = 100
9.8364 divided by 100 = 0.098364
```

For 100 people the SE is 0.98, which is 1 to one significant figure. That allows no decimal places: 34 years. For 10,000 people the SE is 0.1 to one significant figure: 34.3 years.

For 100 people, print 34.3 years (SD 9.8) if you describe the people, or 34 years if you follow the SE rule. SAMPL notes that mean age "can often be rounded to the nearest year", which gives 34. Say which rule you used.

9. Each value by each rule:

software prints	Cole: one significant figure	SAMPL: one or two decimal places
0.67	0.7	0.67 or 0.7
0.0086	0.009	0.01
0.096	0.1	0.10 or 0.1
0.00004	< 0.001	< 0.001

The two rules differ only for 0.0086. Cole's one significant figure keeps it below 0.01; SAMPL's two decimal places round it up to 0.01. For 0.00004, Cole allows a lower limit than 0.001 but his examples stop there, and SAMPL's smallest is P < 0.001.

10. The first three lines are right, apart from their digits.

```
15 divided by 40 = 0.375
18 divided by 40 = 0.45
45 minus 37.5 = 7.5
```

The last line broke. The difference of 7.5 is in percentage points. As a per cent of village A's figure it is larger.

```
7.5 divided by 37.5 = 0.2
```

Village B's share is 7.5 percentage points higher, or 20 per cent higher than village A's. The digits are false too. With 40 women each is worth 2.5 points, so write 15 of 40 (38 per cent) and 18 of 40 (45 per cent). And the whole difference is three women.

Why it looks reasonable: both numbers are percentages, so "per cent" feels like the unit of their difference.

11. The arithmetic is right. The step that broke is treating 0.03 as exact. Rounded to two decimal places, 0.03 stands for any value from 0.025 up to, but not including, 0.035.

```
1 divided by 0.035 = 28.6
1 divided by 0.025 = 40
```

Turned round, the ratio could be anywhere from about 29 to 40. Cole makes the same point about this example. Four significant figures, 33.33, cannot come from a one-figure input. Write "about 30 times, somewhere between 29 and 40", or go back to the unrounded ratio.

Why it looks reasonable: the calculator prints 33.33 and nothing on the screen says the input was rounded.

12. Find where each figure came from.

```
724,115 times 0.24 is about 173,788
3.4 divided by 20.6 = 0.165
```

The 1.7 lakh is the 724,115 women multiplied by 24.0 per cent. The 16.5 per cent is the relative change. The arithmetic holds, and almost every word around it is wrong.

What it does not establish. The row is for women aged 15-49, not Indian women. It counts overweight and obese together, and there is no figure for obesity alone.

The 724,115 women are everyone the survey gathered information from. The row leaves out pregnant women and women who gave birth in the two months before. So 1.7 lakh is not a count the survey gives. The rise is 3.4 percentage points, and "up 16.5 per cent" said alone sounds far larger. Last, two surveys say nothing about why the share rose.

13. Compute the change both ways, and what one student is worth.

```
62 minus 49 = 13
13 divided by 62 is about 0.21
100 divided by 87 is about 1.15
```

The share fell 13 percentage points. The 21 per cent is the relative change. Each student is worth more than one point, so the report should give counts. If the same 87 students were counted both times, 54 of 87 and 43 of 87 would fit. The report does not say they were. And "p = 0.000" is never printed: at most, $P < 0.001$.

What it does not establish. A before-and-after count in three schools has no comparison group, so it cannot show the menu caused the fall. The schools were the ones visited, not a sample of a district. And a fall in buying at the canteen says nothing about what the students drank elsewhere.

10 · One idea per paragraph: the topic sentence, its support, and the reverse outline

EXERCISE 1. Context, content, conclusion: the first sentence sets the topic, the body carries the content, and the last sentence gives the conclusion to remember.

A reverse outline is one sentence per paragraph, stating its one point, written from the finished draft and read alone. It tells you four things. Split a paragraph whose note needs "and". Merge two neighbours whose notes say the same thing. Move or cut one whose note serves no claim.

EXERCISE 2. There are three topics. A is how the district compares with the nation: sentences 1, 3 and 5. B is what the method can and cannot support: sentences 2, 4 and 7. C is what to do: sentence 6. In order they run A, B, A, B, A, C, B. That is 6 changes of topic in 7 sentences.

The reverse outline you want, one paragraph per point, in the order a discussion needs:

paragraph	its one point
1	the district's share was above the nation's, highest in urban wards
2	how far the survey's methods let us trust that
3	what the district should do first

A rewrite:

> *In the district, the share of women aged 15-49 with a BMI of 25.0 kg/m^2 or more was higher than the national figure. That figure was 24.0 per cent in NFHS-5. As in NFHS-5, the share was highest in urban wards.*

> *Two features of the survey bear on that comparison. Weight and height were measured by trained staff, and 82 per cent of the women approached took part. But the sample came from two blocks only, so the figure may not hold for the whole district.*

> *If the pattern holds, screening should start in the urban wards.*

The third paragraph now says what it rests on. A real discussion would also give the district's own figure, with its count.

EXERCISE 3. Check your work in four ways.

- Every margin note is one sentence, with no "and" joining two separate points.
- Read alone, the notes take a reader from your question to your answer, in the order the argument needs. A note that describes how the work was done belongs in the methods.
- No topic appears in two paragraphs that are not next to each other.
- In each rewritten paragraph the point is stated once, first or last, in the same place as in its neighbours. Every other sentence supports it.

If you could not write a note for a paragraph at all, the paragraph has no point yet. Decide what it is for before you touch its sentences.

EXERCISE 4. Draw a column of boxes, one per paragraph, and a column beside it for the margin notes.

Reading the chapter again tests sentences. You wrote them, so they all make sense to you, and each one reads fine. The faults that lose a reader are between sentences: a paragraph that makes two points, a topic left and taken up again three paragraphs later. You cannot see those while you read sentence by sentence, because you already know the argument.

The reverse outline hides the sentences. You write one line per paragraph and read only the lines. Now a topic that turns up twice is two lines with the same words. A paragraph with two points is a line you cannot write without "and". A methods point among the results is a line that says how, not what.

Then write two lines on the board: "split, merge, move or cut". Every fault the margin shows leads to one of those four actions.

11 · Measuring a draft: words per sentence, a readability score, and what they cannot show

EXERCISE 1. The average sentence length, words divided by sentences, and the average syllables per word, syllables divided by words. As either grows, the score falls, which means harder. The formula cannot see meaning, order, or whether the sentences connect: it counts lengths only.

EXERCISE 2. Three things are wrong. First, the grade is a school grade predicted by a formula fitted on United States Navy personnel reading Navy manuals. Nothing has shown that it matches a class in an Indian school. Second, the formula counts word and sentence length only. A leaflet can score grade 11 and still be clear, or score grade 6 and leave out what the patient must do. Third, "suitable" is a claim about understanding, and nobody's understanding was tested.

The paper would need to give the leaflets to patients from the clinic and test what they understood. For example, ask each one what the leaflet tells them to do. Mark the answers against a key written in advance. It should also say which program computed the grade, and how that program counts numerals and abbreviations.

EXERCISE 3. The quotable sentence: "Scientific abstracts have been getting harder to read for decades, by a measure that counts long words and long sentences. Part of the reason is that scientists use more of their own jargon."

Then the size, in words a reader can picture. One study scored over seven lakh abstracts from 123 journals. The share scoring below zero on that scale rose from about one in seven in 1960 to more than one in five in 2015.

Then the limit, plainly. The measure counts lengths, not understanding. It was built on schoolchildren's reading tests and refitted on Navy personnel, not on scientists reading their own field. So the trend is well measured, but whether experts understand papers less well is not shown. Rewriting abstracts more plainly did help academics understand them, in one study the authors cite.

1. Divide the words by the sentences, then the syllables by the words.

```
120 divided by 6 = 20
180 divided by 120 = 1.5
```

ASL is 20 words a sentence. ASW is 1.5 syllables a word.

2. The Flesch Reading Ease score first.

```
1.015 times 20 = 20.3
84.6 times 1.5 = 126.9
20.3 plus 126.9 = 147.2
206.835 minus 147.2 = 59.635
```

FRE 59.6. Then the grade level.

```
0.39 times 20 = 7.8
11.8 times 1.5 = 17.7
7.8 plus 17.7 = 25.5
25.5 minus 15.59 = 9.91
```

Grade level 9.9.

3. Work out the part of the formula you know, then undo the rest (Book 0, C2).

```
84.6 times 1.6 = 135.36
206.835 minus 135.36 = 71.475
```

So 71.475 minus 1.015 times ASL must equal 50. The part taken away is the difference.

```
71.475 minus 50 = 21.475
21.475 divided by 1.015 = 21.16
```

About 21 words a sentence. Longer sentences than that would push the score below 50.

4. Turn syllables per 100 words into syllables per word first.

```
148 divided by 100 = 1.48
1.015 times 20 = 20.3
84.6 times 1.48 = 125.208
206.835 minus 20.3 minus 125.208 = 61.327
```

FRE 61.3, which rounds to Flesch's 61. His 1948 form gives the same: 0.846 times 148 is also 125.208.

```
0.39 times 20 = 7.8
11.8 times 1.48 = 17.464
7.8 plus 17.464 minus 15.59 = 9.674
```

Grade level 9.7. The grade formula was made 27 years after Flesch's article, for Navy personnel. It gives a number for a 1946 magazine, but nobody tested that number on the magazine's readers.

5. Count the words: 43, in one sentence. Then the extra beats in the longer words.

word	even	normally	given	display	deeper	emotions	allegedly	developed	profound
extra beats	1	2	1	1	1	2	3	2	1

word	affection	Mary	compared	equable	feelings	Lucy	lesser	extent
extra beats	2	1	1	2	1	1	1	1

```
1 plus 2 plus 1 plus 1 plus 1 plus 2 plus 3 plus 2 plus 1 plus 2 plus 1 plus 1 plus
2 plus 1 plus 1 plus 1 plus 1 = 24
43 plus 24 = 67
67 divided by 43 = 1.558139535
1.015 times 43 = 43.645
84.6 times 1.5581 = 131.82
206.835 minus 43.645 minus 131.82 = 31.37
```

FRE 31.4, against Flesch's 32. Now one syllable. It changes ASW by 1 divided by 43.

```
1 divided by 43 = 0.02325581395
84.6 times 0.0233 = 1.97
```

About 2 points. If you say allegedly in three beats, not four, your score is about 33.3. Flesch's 32 sits between the two counts. Careful counters differ by a syllable or two, so a score is good to a point or two, not to the decimal.

6. Only the average sentence length changes.

```
43 divided by 3 = 14.33333333
1.015 times 14.33 = 14.545
206.835 minus 14.545 minus 131.82 = 60.47
60.47 minus 31.37 = 29.1
```

FRE 60.5, up 29.1 points, with the same words. Two full stops moved the sentence from Flesch's "college" band (30 to 50) to just above his Plain English minimum of 60. That is why Flesch's own advice for a higher score is to "sprinkle periods". Whether the three sentences are clearer depends on how you split them, and the score cannot tell you.

7. 17 words, one sentence. With numerals as one syllable, the extra beats are: among 1, overweight 2, obesity 3, against 1.

```
17 plus 7 = 24
24 divided by 17 = 1.411764706
1.015 times 17 = 17.255
84.6 times 1.4118 = 119.44
206.835 minus 17.255 minus 119.44 = 70.14
```

FRE 70.1. Now read the numerals aloud. Fifteen adds 1 beat. Forty-nine adds 2. "22.9%", twenty-two point nine per cent, is 7 beats, so it adds 6. "18.9%", eighteen point nine per cent, is 6, so it adds 5. Each "NFHS" with its number is 5 beats, adding 4 each.

```
1 plus 2 plus 6 plus 5 plus 4 plus 4 = 22
24 plus 22 = 46
46 divided by 17 = 2.705882353
84.6 times 2.7059 = 228.92
206.835 minus 17.255 minus 228.92 = -39.34
```

FRE -39.3. The same 17 words score 70.1 or -39.3. State your counting rule beside any score you report.

8. The step that broke is the fourth. The number 84.6 goes with syllables per word. With syllables per 100 words, Flesch's 1948 number is 0.846. Use either pair, not a mix.

```
300 divided by 200 = 1.5
84.6 times 1.5 = 126.9
206.835 minus 20.3 minus 126.9 = 59.635
0.846 times 150 = 126.9
```

FRE 59.6, just under Flesch's Plain English minimum of 60.

Why the wrong answer looks reasonable. Both forms of the formula are in print, one per 100 words and one per word. They differ only by where the decimal point sits. A score in the thousands below zero should have stopped the writer: almost all ordinary prose scores between 0 and 100.

9. The step that broke is the second. Flesch says: "don't break up sentences that are joined by conjunctions like and or but." The paragraph has 2 sentences.

```
60 divided by 2 = 30
1.015 times 30 = 30.45
206.835 minus 30.45 minus 126.9 = 49.485
```

FRE 49.5, in Flesch's band for college readers (30 to 50), not Plain English.

Why the wrong answer looks reasonable. The rules do count some marks inside a sentence as ending one. Flesch in 1979 counted a colon, a semicolon or a dash as a sentence end. And a clause joined by "and" looks like a sentence of its own. But a reader still has to carry the whole thing to the full stop.

10. What to compute: the score those two averages give, from Flesch's own formula.

```
1.015 times 15 = 15.225
84.6 times 1.5 = 126.9
206.835 minus 15.225 minus 126.9 = 64.71
```

64.7, not 80. The two averages do not give the score they are quoted for. Work back to see what would.

```
206.835 minus 15.225 = 191.61
191.61 minus 80 = 111.61
111.61 divided by 84.6 = 1.3193
```

At 15 words a sentence, a score of 80 needs about 1.32 syllables a word, nearer 1 1/3 (1.33) than 1 1/2. The same page's other pair works: 20 words and 1 1/2 syllables give 59.6, about the 60 it is quoted for. So one number in the 80 sentence is probably a slip, in Flesch's book or in the retyped copy online. The arithmetic shows a mismatch; it cannot show which number is wrong.

What this does not establish. Even a leaflet that does score 80 has not been shown to read as conversation, or to be understood by its readers. The score counts lengths. To know whether the leaflet works, give it to people it was written for and ask them what it tells them to do.

11. What to compute: what a score below 0 takes, and whether the formula was built for these readers.

Take an abstract with sentences of 25 words. Find the syllables per word that would bring its score to 0.

```
1.015 times 25 = 25.375
206.835 minus 25.375 = 181.46
181.46 divided by 84.6 = 2.1449
```

About 2.14 syllables a word, at 25 words a sentence. Many medical words have three or four syllables: proportion, obesity, prevalence. In the study, 14 per cent of abstracts scored below 0 in 1960 and 22 per cent in 2015. So "more than a fifth" is right for 2015.

What this does not establish. A score below 0 means long words in long sentences. It does not mean nobody can understand the text. Flesch fitted his formula on children's reading passages and said its link to school grades holds only up to about seventh grade. Neither formula was fitted on specialists reading their own field. The authors themselves say the scores are an estimate, not a measure of how well a text will be understood.

What the study does support is narrower, and still worth saying. Abstracts have grown harder by this measure, partly from more general scientific jargon. In one study the authors cite, abstracts rewritten to read more easily were better understood by academics.

12 · Cutting: deleting the words, sentences and paragraphs that do no work

EXERCISE 1. Paragraphs, then sentences, then words. A paragraph or a sentence goes when it serves no part of the piece's one message, or says again what is already said. A word goes when the sentence means the same without it.

Compressing keeps all the content and packs it into fewer, longer sentences. Cutting deletes what does no work and leaves each sentence carrying one idea.

EXERCISE 2. The paragraph has 86 words in five sentences: 19, 12, 16, 18 and 21.

- "for the purpose of this study" adds nothing: the design is this study's. Delete 6 words.
- "over a period of six months," repeats the dates that follow. Delete 6 words.
- "who were" can go: "adults … residing in the field practice area" means the same. Delete 2.
- "and every" is a doublet: "each" says it. Delete 2.
- "weighing" before "scale" repeats "weighed". Delete 1.
- ", and the weight was recorded" says what weighing to the nearest 0.1 kg already says. Delete 5.

```
6 plus 6 plus 2 plus 2 plus 1 plus 5 = 22
86 minus 22 = 64
```

> *The present study was conducted in the Department of Community Medicine of a tertiary care centre in central India. A cross-sectional study design was used. The study was carried out from January to June 2025. The study participants were adults aged 18 years and above residing in the field practice area. Each participant was weighed with a digital scale to the nearest 0.1 kg.*

64 words, every fact kept. Two things must stay: "18 years and above" and "to the nearest 0.1 kg" each limit what the reader may conclude. A further rewrite into the active voice, as in section 7, could shorten it more, but that is rewriting, not cutting. section 7 would also have you name who weighed the participants; that too is a rewrite, for after this exercise.

EXERCISE 3. Check your work against four things.

- The message is written before any cut, and every paragraph left has a margin note that serves it.
- The counts fall at each pass, and you can say which pass removed most. If the word pass removed most, look again at whole sentences: one that restates its neighbour is easy to miss.
- No sentence left carries more ideas than before. If two sentences became one, undo it.
- Every age range, place, year, denominator and qualifier in your claims is still there.

A section that loses little has passed, not failed. Keep the second file until the paper is accepted.

13 · Showing a draft to a reader: the one-pass test

EXERCISE 1. Write your one message and put it away. A reader outside the work reads the draft once, without going back, marking where they stop or reread. The reader writes in one or two sentences what it argues. You compare their sentences with yours, part by part. You revise the paragraphs the comparison and the marks point to, and test again with a new reader.

A reader who knows your work reads your argument into the draft, so the test would pass whatever the draft says. A reader who has done the test once knows your argument, so a second summary tests their memory and not the draft.

EXERCISE 2. Check your pack against this list.

- The one message was written before the rewrite. It names whom the claim is about, what was found, compared with what, and how far it reaches.
- Each paragraph's first sentence states its point, and the reverse outline of the final draft reads as the argument in order.
- Word counts after each cutting pass are recorded. No two sentences were joined to save words, and no age range, place, year or qualifier was cut.
- The readability scores came from one program or the hand formula, and you say which and how it counts figures. No sentence was chopped only to raise a score.
- Each reader's sentence is kept beside yours, with the four parts marked kept, lost, changed or added. Each change in the next draft is traced to one of them.
- The last reader's sentence carries every part of your message and adds none.

The build is met when the last condition holds. That is this book's gate too: a colleague reading your draft can state your argument after one pass.

EXERCISE 3. Start with why "is it OK?" fails. If you read it as her senior, you will mark words and say it is fine or not fine. Neither tells her whether a reader gets her argument, and comments that are general usually mean the reader missed the main line.

Draw the five steps on the board. Ask her first to write, on a separate page, the one sentence her introduction exists to say, and to fold it away. Then you, or better a resident from another department, read it once without going back, marking where you stop. Then write what it argues, before you talk.

Then compare the two sentences together, part by part. An introduction's message has its own parts (section 4): whom it is about, what is known, what is missing, and what the study will do. Show her that a part you lost points to a paragraph that hid it. A part you added points to a paragraph that invited it.

Last, tell her the rule that makes the test work: a new reader each round. And offer to be her reader once, not every time.

14 · A figure makes one claim

EXERCISE 1. A sentence, when the result rests on a few numbers. A table, when the reader needs exact values or will look up one row among many. A figure, when the claim is a pattern among many values: a trend, a gap, a ranking.

The claim goes in the first sentence of the caption, which ICMJE calls the legend, not on the drawing.

The same numbers are never shown in both a table and a figure.

EXERCISE 2. The caption names three topics and states no finding. So the figure has no claim, or three.

The audit's finding is about BMI by age, so panel B is the figure. Its caption opens with the finding: "The share of patients with a BMI of 25.0 kg/m^2 or more rose with age group." Then it says what the bars show and where the data came from.

Panel A, monthly attendance, is not part of that claim. It belongs in a sentence in the methods or results, such as how many patients were seen, or nowhere.

Panel C rests on two numbers. A sentence holds them, with each mean and its count.

Check one thing before cutting. Would attendance or waist circumference make a reader doubt the BMI finding? If attendance by older patients fell sharply in some months, for example, the reader needs to know who was measured. Say that in the text.

EXERCISE 3. Check your list against four tests.

- Every result has a claim written as a sentence with a verb, not a topic. "BMI by age group" fails; "The share with a BMI of 25.0 kg/m^2 or more rose with age" passes.
- Every result resting on a few numbers is in a sentence.
- Every figure's claim is a pattern among many values, and each figure has one claim. Two claims in one figure means two figures, or one claim moved to the text.
- No numbers appear in both a table and a figure. Where they do, keep the table if readers will look values up, and the figure if they need to see the pattern.

If you have more than about six figures, ask which ones the argument needs. Move the others to a supplement, or cut them.

EXERCISE 4. Start with the reader. Ask her to look at her eight bars for five seconds, then say what the figure shows. She will name a topic, or two findings. That is the problem: a reader given everything has to choose the finding alone, and may choose the wrong one.

Write the two candidate claims on the board: the rise between survey rounds, and the gap between urban and rural. Ask which one her chapter argues. That one gets the figure.

Then show where the other numbers go. The rise rests on four numbers, so it becomes a sentence. The totals sit between urban and rural, so they add nothing to the gap. Nothing is hidden: every number is still in the chapter, in the place that suits it.

End with the caption. Write the finding as its first sentence, and show her how the figure now reads in two seconds.

15 · Choosing the mark: what the eye reads accurately, and the shape of the data

EXERCISE 1. Position along a common scale; positions along scales that are not lined up; length, direction and angle, tied; area.

The experiments tested position against length, and position against angle. Position won both. The rest of the order is the authors' hypothesis.

EXERCISE 2. Add the slices first.

```
35.5 plus 19.3 plus 32.1 plus 3.4 = 90.3
```

The total is 90.3, not 100. That alone shows the slices are not parts of one whole, and the pie has stretched 90.3 to fill a circle.

The deeper problem would remain even if they added to 100. Each number is the share of all children under 5 with one condition, and one child can have more than one. A child can be stunted and underweight at once. So the four groups overlap, and a pie, which shows parts that do not overlap, cannot hold them.

Draw the four values as dots or bars on one axis from 0 to 100 per cent, one row each. The reader then compares positions on a shared scale. The caption states the finding the report needs, for example that stunting and underweight each affect about a third of children.

EXERCISE 3. Check each choice in three steps.

- The shape is named: amounts for groups, many values per group, parts of one whole, one quantity against another, or change over time.
- The mark fits the shape. Amounts take bars from zero or dots. Many values per group take a chart of the distribution. Parts of a whole take side-by-side bars. Two measured quantities take a scatterplot. Time takes a line.
- The comparison that carries your claim is made by position on one shared axis. Suppose the reader must compare two pieces stacked in a bar, or two slices. Then redraw it so the two values sit side by side on one axis.

EXERCISE 4. Answer first: a pie is the wrong chart for these numbers. I will send a chart of six dots on one scale, sorted from highest to lowest block.

Why. Each block's figure is a share of that block's own adults. The six shares are not parts of one whole, so slices of a circle misstate them. Even when a pie is the right kind of chart, people compare its slices less accurately than points on one scale. In one experiment, large errors were about seven times as frequent per judgment on a pie as on bars.

What the chart will show. Each block as one dot, on one axis marked in per cent, labelled with the block's name and its value. The meeting can then see at once which blocks sit highest.

What it will not show. Why the blocks differ, or whether a small difference between two blocks is real. For that we need each block's sample size beside its share.

16 · Show the data: when a bar of means hides what the points would show

EXERCISE 1. Values spread evenly; values bunched together with one far from the rest; values in two separate clusters. A group of very few values is a fourth.

For small groups, a univariate scatterplot, which this book calls a strip chart: every value as a dot. A box plot or a histogram is hard to read with so few values. A box plot can also draw two clusters as one wide box.

EXERCISE 2. The means.

```
6.5 plus 7.0 plus 7.0 plus 7.5 plus 7.5 plus 8.0 = 43.5
43.5 divided by 6 = 7.25
5.0 plus 5.5 plus 5.5 plus 6.0 plus 6.0 plus 15.5 = 43.5
43.5 divided by 6 = 7.25
```

The medians: with six values, the mean of the third and fourth.

```
(7.0 plus 7.5) divided by 2 = 7.25
(5.5 plus 6.0) divided by 2 = 5.75
```

The strip chart shows hostel A from 6.5 to 8.0 hours. In hostel B, five students sleep 5.0 to 6.0 hours and one value sits at 15.5. That single value lifts hostel B's mean to equal A's.

First, check 15.5 against the student's form. A value of 5.5 typed as 15.5 is one possibility; a student who really reported 15.5 is another. Do not delete it because it is inconvenient.

Then write what the values show. Five of six students in hostel B slept less than everyone in hostel A, and the medians are 5.75 and 7.25 hours. Six students per hostel cannot tell you whether this holds for the hostels as a whole.

EXERCISE 3. A good redraw keeps the groups on one shared axis and puts every value as a dot. A short line marks the median or mean. The caption says which the line is and how many values each group has.

If you could not redraw it, you needed the individual values, which a bar chart never shows. That is the point: the bar chart gave you no way to check the claim it was drawn to support. Asking authors for their data, or for the values in a supplement, is a reasonable request for a reviewer to make.

If the same people were measured twice, your redraw joins each person's two values with a line.

EXERCISE 4. Draw three bars of the same height on the board, and write 30 above each. Ask her what she can say about the three groups. She will say they are the same.

Now draw three columns of dots behind the bars: one spread evenly, one bunched with a single dot far above, one in two clusters. Every column has the same mean. Ask again.

Make three points. First, with eight values a group, the reader can see every value; a bar shows one number in place of eight. Second, one far value can make a group's mean match another's, and only the dots show it. Third, her examiner will want to check her claim, and the dots are what lets them.

End with what to draw: one dot per patient, a short line at each group's median, and a caption that says so.

17 · Honest axes: where an axis starts, what range it shows, and measuring the distortion

EXERCISE 1. The percentage change drawn in the figure, divided by the percentage change in the data.

An LF of 1 draws the change at its true size. Tufte counts above 1.05, or below 0.95, as substantial distortion.

For two bars a and b, a the smaller, drawn from an axis start s: work out the LF from the bar the change starts from. It is a ÷ (a − s) for a rise from a, b ÷ (b − s) for a fall from b. For two groups, take the group you compare against, and say which. With s at zero it is 1.

EXERCISE 2. A good plan decides the axis from the mark, and the mark from the one claim.

The mark. Bars if the claim is about the level of something, such as "a quarter of women". Points, joined by a line where time runs along the bottom, if the claim is about a change. The change itself as bars, if the change is the claim.

The axis. Bars start at zero, so their Lie Factor is 1. Points may start near the data: a range not much wider than the values, with round tick marks. A ratio goes on a log axis, with bars from 1.

The caption sentence names the start whenever it is not zero, and gives the values. For example: "The vertical axis starts at 18%. The share of women with overweight or obesity rose from 20.6% to 24.0%, a rise of 3.4 percentage points."

If your draft has bars from an axis above zero, work out the LF and write it down: a ÷ (a − s) for a rise from a. A plan that leaves an LF above 1.05 in place has not finished.

1.

```
40 minus 30 = 10
50 minus 30 = 20
20 divided by 10 = 2
50 divided by 40 = 1.25
```

Drawn heights of 10 and 20. The second bar is drawn twice as tall as the first. The data ratio is 1.25.

2. The data.

```
50 minus 40 = 10
10 divided by 40 = 0.25
```

A rise of 25 per cent. The drawing: the heights are 10 and 20.

```
20 minus 10 = 10
10 divided by 10 = 1
```

A rise of 100 per cent. Divide the drawn change by the change in the data.

```
100 divided by 25 = 4
40 divided by 10 = 4
```

The Lie Factor is 4. The second line checks it with a ÷ (a − s), for a rise from a: 40 divided by 10, where 10 is 40 minus 30.

3. The area is height times width, and both doubled.

```
2 times 2 = 4
```

The second picture covers 4 times the area. As percentage changes:

```
4 minus 1 = 3
3 times 100 = 300
20 minus 10 = 10
10 divided by 10 = 1
1 times 100 = 100
300 divided by 100 = 3
```

The area grew by 300 per cent for a rise of 100 per cent. The Lie Factor by area is 3.

4. Call the axis start s. The drawn heights are 60 − s and 66 − s, and the second is three times the first.

```
66 - s = 3 × (60 - s)
66 - s = 180 - 3 × s
```

Add 3 × s to both sides, and take 66 from both sides.

```
3 × s - s = 2 × s
180 minus 66 = 114
114 divided by 2 = 57
```

The axis starts at 57. Check it.

```
60 minus 57 = 3
66 minus 57 = 9
9 divided by 3 = 3
```

5. Starting at 1, a bar's length is the $\log_{10}$ of the ratio.

```
log10 of 2 = 0.301
log10 of 0.5 = -0.301
```

Equal lengths, pointing opposite ways. That is right: 2 is “twice” and 0.5 is “half”.

Starting at 0.1, each bar runs from 0.1 to its value, so its length is the $\log_{10}$ of the value divided by 0.1.

```
2 divided by 0.1 = 20
0.5 divided by 0.1 = 5
log10 of 20 = 1.301
log10 of 5 = 0.699
1.301 divided by 0.699 = 1.86
```

Both bars now point up, so the half reads as a gain. And the bar for 2 is less than twice the bar for 0.5, though 2 is four times 0.5.

6.

```
18.9 minus 18 = 0.9
22.9 minus 18 = 4.9
4.9 divided by 0.9 = 5.44
22.9 divided by 18.9 = 1.21
```

Drawn heights of 0.9 and 4.9 percentage points. The NFHS-5 bar is 5.44 times the NFHS-4 bar, for a data ratio of 1.21. Now the Lie Factor of the rise from 18.9.

```
18.9 divided by 0.9 = 21.0
```

A Lie Factor of 21. The rise of 4.0 percentage points, about 21 per cent, is drawn as a rise of about 444 per cent.

7. As percentage changes: the data rose by 8 per cent, the ink by 170 per cent.

```
1.08 minus 1 = 0.08
2.7 minus 1 = 1.7
1.7 divided by 0.08 = 21.25
```

A Lie Factor of about 21.

For the axis start, measure everything in units of the 2010 value, so a = 1 and b = 1.08. The drawn heights are 1 − s and 1.08 − s, and the second is 2.7 times the first.

```
1.08 - s = 2.7 × (1 - s)
1.08 - s = 2.7 - 2.7 × s
```

Add 2.7 × s to both sides, and take 1.08 from both sides.

```
2.7 minus 1 = 1.7
2.7 minus 1.08 = 1.62
1.62 divided by 1.7 = 0.953
```

The axis started at about 0.953 of the 2010 value: about 95 per cent of the way up the first bar. Check with the shortcut for a rise: 1 ÷ (1 − 0.953) is about 21, as before.

8. Compare urban against rural, so measure both percentage changes from rural, 19.7.

```
33.2 minus 19.7 = 13.5
13.5 divided by 19.7 = 0.685
19.7 minus 15 = 4.7
33.2 minus 15 = 18.2
18.2 minus 4.7 = 13.5
13.5 divided by 4.7 = 2.87
2.87 divided by 0.685 = 4.19
```

The urban figure is 68.5 per cent above the rural one. The drawing shows 287 per cent. The Lie Factor of urban against rural is about 4.19, the same as 19.7 ÷ (19.7 − 15).

9. The four values run from 18.9 to 24.0.

```
24.0 minus 18.9 = 5.1
```

A spread of 5.1 percentage points. An axis from 18 to 26, marked every 2 points, holds all four with a little room at each end. It is not much wider than the data, which is what Bergstrom and West advise for a line. Points are read by position, so this is honest, if the caption says the axis starts at 18.

Bars on that axis would not be honest. Measure each rise from NFHS-4:

```
20.6 minus 18 = 2.6
20.6 divided by 2.6 = 7.92
18.9 minus 18 = 0.9
18.9 divided by 0.9 = 21.0
```

Lie Factors of about 7.9 for women and 21 for men. If the message is the level, about one adult in four or five, use bars from zero instead.

10.

```
33.2 divided by 19.7 = 1.685
29.8 divided by 19.3 = 1.544
log10 of 1.685 = 0.227
log10 of 1.544 = 0.189
0.227 divided by 0.189 = 1.20
```

The women's bar is about 1.2 times the men's bar, both drawn up from 1. Urban women are about 1.69 times rural women; urban men are about 1.54 times rural men. Had the bars started at 0 on a linear axis, they would stand for amounts, not ratios. A ratio below 1 could not point the other way.

11. The first five lines are right. The last line broke. The Lie Factor divides percentage changes, not ratios. Take 1 from each ratio to turn it into a change.

```
4.0 divided by 0.6 = 6.6667
6.6667 minus 1 = 5.6667
24.0 divided by 20.6 = 1.16505
1.16505 minus 1 = 0.16505
5.6667 divided by 0.16505 = 34.33
```

The Lie Factor is about 34.3, as a ÷ (a − s) gives for a rise from a: 20.6 divided by 0.6.

Why it looks reasonable: the ratio of the two ratios is a real comparison of picture against data. Bergstrom and West use ink ratios that way. But it is not Tufte's Lie Factor, and Tufte's 1.05 belongs to the Lie Factor. Here the wrong method understates the distortion six times over.

12. The fourth line broke. The Lie Factor belongs to the drawing. The bars are still 0.6 and 4.0 high, so it is still 20.6 divided by 0.6.

```
20.6 divided by 0.6 = 34.33
```

Readers did not switch to the numbers either. In Correll and colleagues' second experiment a break mark on the axis and the bars did not stop truncation raising how severe the difference looked. In their third, it stayed even when readers reported the numbers accurately.

An honest fix changes the lengths: bars from zero, points on a stated range, or bars of the 3.4 percentage-point change from zero.

Why it looks reasonable: the break mark is true, and it is what many style guides advise. It tells the reader a fact; the picture tells them a different one, and the picture wins.

13. Compute the Lie Factor from the two ratios, and check what "4.6%" is.

```
6 minus 1 = 5
1.13 minus 1 = 0.13
5 divided by 0.13 = 38.5
```

The bars changed by 500 per cent for a rise of 13 per cent: a Lie Factor of about 38.

The colleague's "4.6%" is a rise of 4.6 percentage points, not a 4.6 per cent rise. A ratio of 1.13 with a gap of 4.6 points puts the first rate near 4.6 divided by 0.13.

```
4.6 divided by 0.13 = 35.4
```

About 35%. So the true rise is 13 per cent, and the sentence should say "a 4.6-point rise, about 13 per cent".

What it does not establish. The chart's numbers were right; only its lengths were wrong. The Lie Factor says nothing about whether a 4.6-point rise in a tax rate matters: that is a judgement about tax, not about charts. And it does not show the designer meant to mislead.

14. Compute the real change both ways, then find the axis start that would make the bars look doubled. The men's figures are 18.9% in NFHS-4 and 22.9% in NFHS-5.

```
22.9 minus 18.9 = 4.0
4.0 divided by 18.9 = 0.212
22.9 divided by 18.9 = 1.21
```

A rise of 4.0 percentage points, about 21 per cent. Not a doubling. Now the start that draws a doubling: the NFHS-5 height, 22.9 − s, equals twice the NFHS-4 height, 2 × (18.9 − s).

```
22.9 - s = 37.8 - 2 × s
37.8 minus 22.9 = 14.9
```

Add 2 × s to both sides and take 22.9 from both sides: s = 14.9. Any bar slide whose axis starts near 15% will show the men's bar about twice as tall. The officer read the lengths.

What it does not establish. The figures are for men aged 15 to 49 only, with overweight and obesity combined in one row. NFHS-4 and NFHS-5 are two separate samples, so this is a change in the share, not in any group of men followed over time. Nothing in the two numbers says why it rose. And whether a rise of 4 points matters is a question for the programme, not for the chart.

18 · Decoration that does no work

EXERCISE 1. Data-ink is the ink that changes when the data change: the bars, points and lines. The data-ink ratio is the data-ink divided by all the ink used to print the figure.

Tufte's rule is to maximise it "within reason".

Any two of: the axis titles with their units, the tick labels, a light grid or reference line, direct labels. They are context. Without them the data cannot be read.

EXERCISE 2. Start with the axis, because no amount of tidying fixes it. Bars from 60% exaggerate every gap between blocks (section 17). Start the axis at zero, or draw the six values as dots on a stated range.

Then the elements, asking of each what it tells the reader.

The 3-D cylinders and their shadows tell nothing and bend the heights against the grid. Draw flat bars or dots.

The six colours tell nothing the block names do not. Use one colour, and write each block's name under its bar. The key then goes.

The picture of a scale behind the bars is decoration, and it competes with the data. Delete it.

The dark grid: keep a few light lines, or none.

Add what is missing. An axis title with its unit: "children under 5 weighed last month (%)". The month and the source in the caption.

Last, sort the blocks from highest to lowest if the order of blocks means nothing.

EXERCISE 3. Start with what the study found, fairly. In 20 students, charts built into cartoon pictures were described as accurately as plain charts. Ten of them, tested two to three weeks later, remembered the decorated charts better.

Then draw the design on the board. Fourteen charts by one illustrator. Untimed viewing. The task was to describe each chart aloud, not to read off exact values. Ten people per memory group. The paper reports t-tests, not how many points better.

Then the second finding: readers saw the designer's opinion, the "value message", more often in the decorated charts. A picture argues.

Then what it does not cover. The pictures were tied to each chart's subject. A 3-D effect, a shadow or a background fill carries no message at all, and 3-D bends the bars. Nothing in the study speaks for those.

End with the rule. In a paper's figure, delete what tells the reader nothing and keep the labels. On a poster, a picture tied to the one message may help people remember it, if you accept the opinion it adds. The authors themselves do not advocate it "as a general principle".

19 · Colour: for meaning only, readable by everyone, and in grey

EXERCISE 1. To separate groups with no order, use a few colours that look clearly different and equally strong. That is a qualitative scale. To stand for a value, use colours that run from light to dark: a sequential scale. If the value can go either way from a middle such as zero, join two of them at a light middle colour: a diverging scale. To pick out one thing, use one strong colour and make the rest grey.

The two tests: look at the figure in grey, and run it through a simulator of colour-vision deficiency. If its meaning survives both, the colour is doing its job.

EXERCISE 2. Figure 4.

- The rainbow scale is not even. Yellow pulls the eye wherever it falls, and the wide green band hides real differences. In grey the order of districts is lost. Use a sequential scale that runs from light to dark.
- There is no colour bar. Without it a reader cannot turn any shade into a number. Print one, with values marked.
- The caption asks the reader to find "red". A reader with red-green colour-vision deficiency may not be able to. On a light-to-dark scale the caption can say "darker districts".

Figure 5.

- Red, green, brown and orange at similar lightness are hard to tell apart for many readers with colour-vision deficiency, and thin lines make it worse.
- The key sends the reader back and forth. Write each age group's name at the end of its line instead. If a key stays, put its entries in the order the lines appear.
- The age groups have an order, so use four steps of one colour, lightest for the youngest. Then check the figure in grey and in a simulator.

EXERCISE 3. Start with what the number is. It is 2.76 per cent of 74,986 boys aged 11 to 17 in one district, about one boy in 36. It is not a rate for Indian adults, and girls were not tested. The general estimate cited worldwide is higher for men, 8 per cent, about one in 12 or 13.

Then say why the number does not decide the question. Any room of examiners or reviewers may hold one person who cannot tell red from green. And some copies of her slides and thesis will be printed or photocopied in black and white. There, a red and a green of the same lightness become the same grey. That second problem affects every reader of a grey copy.

Then show the cost of fixing it. Draw a red and green pair on the board, then the same chart with "urban" and "rural" written beside the bars. The labelled version needs no colour at all. Changing one colour and checking in grey takes minutes.

EXERCISE 4. A good record of this has three parts. A list of colours, each with its job, or "none, removed". A grey copy in which every group or value can still be told apart. Usually that is because the groups are labelled directly or the colours differ in lightness. And a note of what the simulator showed, with the change you made.

If a colour stands for a value, check three more things. The scale runs steadily from light to dark. It is not a rainbow. A colour bar is printed. If you used a palette built for colour-vision deficiency, check thin lines and small points at the printed size; the palette alone does not guarantee them.

20 · Words on a figure: axis titles with units, direct labels, and a caption that stands alone

EXERCISE 1. The caption starts with the finding, as its title. Then it says what each mark shows, and who, where and when. Then n, what any error bar is, and the source of the data.

Describing the data: the range and the standard deviation. Describing the precision of the mean: the standard error and the 95% confidence interval. All four are drawn the same way.

EXERCISE 2. The bar is the SE, so it is about the mean, not the people. To see the people's spread, undo the SE: the SE is the SD divided by √n, so the SD is the SE times √n.

```
square root of 40 = 6.32
3 times 6.32 = 18.96
```

The SD is about 19 cm. Waist circumference varies a great deal among these people, far more than the 3 cm bar suggests at first glance.

For the precision of the mean, n is 40, so the 95% CI is roughly the mean plus or minus 2 SE.

```
2 times 3 = 6
92 minus 6 = 86
92 plus 6 = 98
88 minus 6 = 82
88 plus 6 = 94
```

Each mean is known to within about 6 cm either way: roughly 86 to 98 cm and 82 to 94 cm. The caption did its job: because it said SEM and gave n, you could work out both.

EXERCISE 3. The problems:

- The first sentence names a topic, "graph showing BMI". It should state the finding, such as which group's mean BMI was higher and by how much.
- The value axis has no unit. It should read "body mass index (kg/m^2)".
- The bottom axis gives no clue what the two groups are, unless their names sit on it. Name them on the axis or beside the bars.
- "Variation" does not say whether the bars are the range, the SD, the SE or a 95% CI.
- n is missing, for each group.
- Who the participants were, where and when are missing.
- The asterisk is explained only as "p<0.05", with no comparison named. Say which two values were compared, and by what test, or leave it out. P alone does not show how big the difference is.

A rewrite states the finding, names the bar and gives n. For example: "Mean body mass index was higher in group A than in group B. Points are means, error bars 95% confidence intervals; n = [number] in each group." The real numbers come from your data.

EXERCISE 4. Both may be right; they may be showing different bars. Ask her what each bar is. She may have drawn the SD on one slide and the SE on the other.

On the board, take three numbers, 28.7, 38.7 and 52.6. Their mean is 40.0 and their SD 12.0. The SE is the SD divided by √3, about 6.93. An approximate 95% CI for three values is about 4 times the SE, 27.7. Draw the three bars on the same mean.

Then the rule. The SD tells you how spread out the people are. The SE and the CI tell you how precisely the mean is known. Neither is wrong. What is wrong is a slide that does not say which it shows. Write it in the caption, with n, on both slides.

EXERCISE 5. Check your figure against five questions. Does every measured axis carry its unit? Does every group carry its name beside its data, or is there a reason a key must stay? Is the first sentence of the caption the finding? Are who, where, when and n all there? If there are error bars, does the caption say which kind?

If your colleague hesitated on any of the four things you asked, the words on the figure are not finished. Fix the one they hesitated on, then ask someone else.

21 · Making a table that stands alone

EXERCISE 1. The title goes above the table. It is short and says what was measured, in whom, where and when. A reader should not need the text to understand it. Notes, exclusions and abbreviations go in footnotes under the table, not in the headings.

Draw no lines down the table and none between rows. One line under the headings, and one above and below the table, are enough. Words line up on the left. Numbers line up on the right, with the same number of decimal places all the way down. Each heading lines up like its column.

EXERCISE 2. Seven changes:

- Add a title above the table: "Table 2. Non-pregnant women and men aged 15–49 years with anaemia, India, NFHS-5 (2019–21), by residence."
- Cut the long heading. The column headings become "urban (%)", "rural (%)", "total (%)". The rest goes into footnotes. They give the cut-offs: haemoglobin below 12.0 g/dl for non-pregnant women and below 13.0 g/dl for men. They also expand Hb and the survey's name.
- The fact sheet's own footnote warns that NFHS-5 measured anaemia in capillary blood, so its results "need not be compared with other surveys using venous blood". That caution belongs in a footnote too, because a reader will compare.
- Put "%" in the headings, not in some cells. Write 25 as 25.0, the fact sheet's figure, so every number has one decimal place.
- Remove the cell borders. Keep one line under the headings and lines above and below the table. Align the numbers right and the words left.
- Add a source line, and say that the fact sheet gives no counts.
- Cut the duplication. Keep the table or the bar chart, not both. Replace the sentence that lists six numbers with the one observation that matters, citing the table. For example: anaemia was more than twice as common in women as in men.

The two cut-offs and the capillary-blood caution are from the fact sheet's indicators 93 and 97 and its footnote 22.

EXERCISE 3. Try each reading. If 1.91 years were the standard error, the standard deviation would be the standard error times the square root of n.

```
square root of 74986 = 273.84
1.91 times 273.84 = 523.0
```

A standard deviation of 523 years is impossible for boys aged 11 to 17. So 1.91 must be the standard deviation.

That you could work it out does not excuse the omission. In a table, the heading should read "age, mean (SD), years", or a footnote should say that values are mean ± standard deviation. ICMJE asks authors to identify statistical measures of variation.

EXERCISE 4. Agree with her aim, then show her the cost. The lines are meant to help the eye follow a row. But we do not rule lines between the lines of a paragraph, and we read those easily. Every extra line is something the eye has to look past to reach the numbers.

Then show her. Delete the vertical lines and the lines between rows. Keep one line under the headings and one above and below the table. Right-align the numbers so the decimal points line up, and the column can be compared at a glance. Put the three-line heading's explanation in a footnote, so the heading shrinks to a few words and a unit.

Then the reason that matters most: an examiner or reviewer may look at the table before reading the text. Ask her to cover the text and read the table alone. If she cannot say whom the rows describe, the title needs work before the lines do.

EXERCISE 5. Check it against this list. The title says what, in whom, where and when. Each heading is a few words and a unit. Every percentage has its count. Any "±" or bracketed value is named as standard deviation, standard error or confidence interval. Exclusions and abbreviations are in footnotes. A source line names any outside data.

There are no lines down the table or between rows. Numbers are right-aligned with the same decimals. The table is not repeated in a figure or read out in the text. Your one sentence states the most important observation and ends "(Table 1)".

Then cover the text and give the table alone to a colleague. If they cannot say whom it describes, fix the title first.

22 · Making the plot: from a table of numbers to a finished figure

EXERCISE 1. Claim, mark, axis, words; then defaults, colour, caption, saving, checks.

The first four decide what the figure shows: the claim, the mark, the axis range, and the axis titles with units. The rest come after. Check the defaults one by one. Give colour a job. Write the caption. Choose the file type and size. Last, check it at printed size, in grey and through a simulator.

EXERCISE 2. Every line needs changing.

- Claim: "anaemia in NFHS-5" is a topic. Write the finding, for example "anaemia was more than twice as common in women as in men aged 15-49 in NFHS-5". Then decide what to draw.
- Mark: a 3-D effect adds depth that is not in the data (section 18). Use flat bars, or dots.
- Axis: bars start at zero. Men aged 15-49 were at 25.0 per cent in NFHS-5, so an axis from 50 cannot show the men's bar at all. Take the axis from 0, and title it "anaemic (%)".
- Words: the title moves into the caption, as a finding. The value axis needs its quantity and unit.
- Colour: red against green fails in grey and for readers with the commonest colour-vision deficiency. Two groups with no order take two colours of clearly different lightness, or direct labels and one colour.
- Saving: jpeg blurs lines and text. Use the journal's format; otherwise pdf, or png where a bitmap is required.

Three lines are missing: no step for checking the defaults that are left, none for the caption, and none for the checks at printed size and in grey.

EXERCISE 3. Judge your sheet by whether a colleague could draw the same figure from it and your data without asking you anything. Check four lines in particular.

- The claim is a sentence that would be false if the data had come out otherwise.

- The axis line gives the start and the end. Bars start at zero; any other start for points is stated, with the reason.
- The defaults line names each default you changed, not "tidied up".
- The checks line records what you did: the width you shrank it to, and that you viewed it in grey and through a simulator.

Every change you had to add after drawing is a decision you were about to make by accident. The list of them is worth keeping for your next figure.

EXERCISE 4. Agree with the guide first: a spreadsheet can draw a good figure. Wilke's own test is whether the software lets you make the figure you need.

Then draw the program's first chart on the board, from her own data if she has them. Give it a title on top, a key saying "Series1", an axis starting above zero, and small text. Ask her what each element tells the reader. Most tell nothing, and the axis start misleads. Rougier and colleagues' point is that defaults suit every chart and so suit none.

Then the real cost. Ask what happens when her guide finds one wrong value in the data. If she changed fifteen things by hand, she must remember and repeat all fifteen. Show her the figure sheet: the data and nine decisions, written once, kept beside the file. With it, the redraw takes minutes.

End on what the sheet does not do. It does not make her data right. She checks those first.

23 · One finding, from the fact sheet to a paragraph a colleague can restate

EXERCISE 1. In order:

1. The question and the data: read the source's row, columns and footnotes, and write its limits down.
2. The one message: would it be false had the data come out otherwise?
3. The paragraph: does it open on the claim, with the steps moved to the methods?
4. The numbers: units, counts or a note that there are none, points for differences, digits by Cole's rules.
5. Score and cut: the score as a trip-wire only; cut the largest units first.
6. One home: table or figure, by what the reader does. For a figure, the steps of making the plot, and the Lie Factor of any bars.
7. The reference: numbered by first mention, in the format for that kind of source.
8. The reader: one pass, compared part by part with your hidden message.
9. What it does not mean: population, measure, time, place and cause.

EXERCISE 2. The 17 is a relative change: the rise in percentage points divided by the NFHS-4 figure.

```
24.0 minus 20.6 = 3.4
3.4 divided by 20.6 = 0.165
```

About 17 per cent of the starting figure, so the arithmetic is right. Three things are wrong.

- "Obesity": the row counts overweight and obese together, at a BMI of 25.0 kg/m^2 or more. There is no row for obesity alone.
- "Indian women": the row covers women aged 15 to 49, excluding pregnant women and women with a birth in the preceding 2 months.
- "Up 17%" with no base: a reader may take it as 17 points. Say the absolute change first, 3.4 percentage points, from 20.6 to 24.0 per cent, and label any relative change.

The headline also says nothing of when: the survey ended in 2021.

EXERCISE 3. Correct the question gently first. The survey whose fact sheet you hold, NFHS-5, ran in 2019-21, and its fact sheet reports overweight and obesity together, for adults aged 15 to 49.

Then give her the sentence to quote.

> *In the national survey of 2019-21, about one woman in four and more than one man in five aged 15 to 49 had overweight or obesity.*

Add, if she asks about change: "That is more than in the survey of 2015-16."

How sure: be exact about what you know. The figures are the survey's own, read from its fact sheet. The fact sheet gives no counts and no margins of error for these rows, so do not tell her how precise they are. What the survey cannot tell her is why the share rose, or anything about people over 49. If she wants a number for obesity alone, this fact sheet does not have one.

EXERCISE 4. Judge your sheet by three tests.

- Can you point, for each step, to a change it made? A step that changed nothing was either done earlier or skipped.
- Does every number in the paragraph appear once only, in the paragraph, the table or the figure, with the same digits wherever it is repeated?
- Does your "does not mean" list name the population, the measure, the time, the place and the cause? Most findings need all five.

Your reference may rest on a kind of source this book did not cover. Then the sheet should say which chapter of Citing Medicine you opened for it.

Glossary

Every term of art this book teaches, in the plain words it gets where it is first used, and the section that teaches it.

Term	Plain words	Section
abbreviation	a shortened form of a word or phrase	8
abstract (of a paper)	a short summary of the whole paper, printed before it; its parts are the context, the gap, what was done, the main result and what it means (3)	1, 3
acronym (as Barnett and Doubleday count it)	a word in which half or more of the characters are capital letters	8
active voice	the subject of the sentence performs the action of the verb: "We measured the weight"	6
anaemic (as the NFHS fact sheet counts it)	haemoglobin below a cut-off for the person's group	22
average sentence length (ASL)	the number of words divided by the number of sentences	11
average syllables per word (ASW)	the number of syllables divided by the number of words	11
axis start	the value at which the value axis begins, written s	17
axis title	the words on an axis that name its quantity and its unit; an axis of categories needs no unit	20
bimodal	having two clusters of values	16
bin width (of a histogram)	the width of each bar	16
bitmap	an image saved as a grid of coloured dots called pixels; a png or jpeg file is one	22
box plot	a box from the lower to the upper quartile, a line at the median, and whiskers out towards the far values	16
caption	the text printed with a figure; journals call it the figure legend	14
chartjunk	Tufte's word for decoration on a figure that "does not tell the viewer anything new"	18
claim (of a paper)	what the paper's evidence and reasoning are offered to support; in Book 0's terms, the conclusion, with the data as the premises	2
clause	a group of words with a subject and a verb	6
clinical trial	a study that gives a treatment to people and compares what happens	2
colour bar	the strip, like an axis, that shows which colour stands for which value	19
colour-vision deficiency	a reduced ability to tell certain colours apart, most often red from green	19

Term	Plain words	Section
colour-vision-deficiency simulator	a program that shows a figure as a reader with colour-vision deficiency sees it	19
common scale	one axis that every mark being compared is read against	15
compressing (a draft)	keeping everything and packing it into fewer, longer sentences; not the same as cutting	12
confidence interval (CI)	a range, worked out from the data, that shows how precisely an estimate such as a ratio is known; how it is worked out is beyond this book (9); for a 95% CI, if the study were repeated many times, 95 per cent of such ranges would contain the true value (20)	9, 20
context (on a figure)	the non-data ink that stays: axis titles, tick labels, a light grid or reference line, direct labels	18
correlation	a number worked out from paired measurements; this book does not teach it	8
cutting (a draft)	deleting words, sentences or paragraphs that do no work for the reader, biggest units first, so that what remains says the same in fewer words	12
data ratio	the ratio of the two values themselves, b ÷ a, as against the drawn ratio	17
data-ink	the ink on a figure that changes when the data change: the bars, the points, the line	18
data-ink ratio	the data-ink divided by all the ink used to print the figure	18
decoration (on a figure)	any element that tells the reader nothing about the data or about how to read them	18
default (of a program)	a setting a program uses when you have not chosen one	22
direct label	a group's name written beside its data, in place of a key saying which colour is which	18
diverging colour scale	for a change that can go either way: a light middle colour put at zero, the colours darkening at the same rate on both sides	19
doer	whoever or whatever performs the action, whether or not it is the subject	6
DOI	a paper's permanent identifier	5
drawn ratio	the ratio of two bars' heights as drawn from an axis start s: (b − s) ÷ (a − s)	17
editorial	a journal's opinion piece	1
elementary perceptual task	Cleveland and McGill's name for a judgment the eye makes about a mark to get its number back: position, length, angle, area and others	15

Term	Plain words	Section
error bar	a line above and below the top of a bar, usually one standard deviation or one standard error each way (16); drawn through a value, usually a mean, to show a range around it: descriptive bars (the range, the SD) show how the data are spread, inferential bars (the SE, a 95% CI) how precisely the mean is estimated (20)	16, 20
et al.	Latin for "and others"; it follows the authors a reference lists	5
false precision	a number carrying more digits than the data can support	9
figure sheet	(this book's term) the written list of the decisions for one figure, kept beside the data	22
flat data	numbers that vary in one direction only, such as one value per group	18
Flesch Reading Ease (FRE)	206.835 minus 1.015 times ASL minus 84.6 times ASW; a higher score means easier text	11
Flesch–Kincaid grade level (FKGL)	Kincaid and colleagues' formula, 0.39 times ASL plus 11.8 times ASW minus 15.59; its result is a United States school grade	11
footnote (of a table)	a note below a table, marked with a letter or symbol, carrying explanations, exclusions and nonstandard abbreviations	21
forthcoming (in a reference)	a paper accepted but not yet out; NLM's preferred word, also "in press"	5
gap (in a paper)	what is not yet known; the introduction moves to it from what is known (4)	1, 4
grey copy	a copy printed or photocopied in black and white; it keeps only how light or dark each colour is	19
haemoglobin (Hb)	the oxygen-carrying protein of the blood, measured in grams per decilitre (g/dl)	21
hazard ratio	compares how fast an event happens in two groups	9
heat map	a table whose cells are shaded by value	19
hidden verb	an action turned into a noun ("increase", "assessment") and carried by an empty verb ("was observed", "was done")	7
hue	the name of a colour: red, green, blue	19
hypothesis	a statement the study is designed to support or refute	4
IMRAD structure	the four parts the text of an original research article is usually divided into: Introduction, Methods, Results and Discussion	1
ink (on a figure)	everything drawn on the figure: lines, bars, shading	17
jittering	spreading dots a little sideways at random, so that they do not sit on top of each other	16

Term	Plain words	Section
journey (of a finding)	the order of work that takes one result from its source to a results paragraph a reader can restate after one reading	23
key (on a figure)	the box in a corner of a figure that says which colour is which	14
legend	(1) to a journal, a figure's caption (14); (2) drawing programs' word for the key, which says which colour is which (18)	14, 18
Lie Factor (LF)	Tufte's measure: the percentage change drawn in the figure divided by the percentage change in the data; 1 draws the change at its true size	17
luminosity (of a colour)	how bright or dark a colour looks	19
main clause	a clause that can stand as a sentence on its own; the main point of the sentence belongs in it	6
n (in a caption)	how many independent, that is separate, people or units were measured; not the number of measurements	20
narrative review	a summary of published work on a topic	1
nickname (for a long name)	a short plain name used in place of an abbreviation, such as "the centre" for the primary health centre	8
non-data ink	everything on a figure that is not data-ink: frames, grid lines, background fills, shadows	18
object (of a verb)	the noun that receives the action of the verb	6
odds ratio	the odds of an outcome in one group divided by the odds in another	9
one message (of a paper)	a paper's central claim in one or two sentences, what the study found and why that matters, written before the draft	3
one-pass test	a reader who does not know the work reads a draft once and writes in one or two sentences what it argues; you compare that, part by part, with your own message	13
original research article	a journal paper that reports a study for the first time	1
outline (of a paper)	one sentence for each planned paragraph, each sentence the claim that paragraph makes, written before the prose	4
P value	the number a statistical test prints beside a comparison; what it means is beyond this book	9
pairs (of values)	two sets of values from the same people measured twice, before and after something	16
passive voice	the subject receives the action; the verb has a form of "to be" and a past participle, and the doer moves into a phrase beginning "by" or is left out	6
past participle	the form of a verb that follows "has" or "was": measured, given, left	6
perceptually uniform (colour scale)	the same step in the data looks like the same step in colour anywhere along the scale	19

Term	Plain words	Section
predatory or pseudo-journal	(this book's reading; ICMJE's references section does not define it) a journal that presents itself as checking what it publishes and does not	5
preprint	a paper posted publicly before a journal has checked and accepted it	5
principle of proportional ink	Bergstrom and West: a shaded area must be in proportion to the value it stands for	17
qualitative colour scale	a small set of colours that look clearly different from each other and equally strong, for groups that have no order	19
readability formula	a formula that estimates how hard a passage is to read from three counts: words, sentences and syllables	11
redundant coding	the same information carried two ways, such as words on the chart and colour	19
reference list	the list at the end of a paper; each number in the text points to one entry in it	5
reference manager	a program that stores your references and types out the list	5
relative change	a difference between two percentages divided by the starting percentage, stated in per cent	9
report of activities	a write-up that records what the authors did, in the order they did it	2
reverse outline	the outline of (4) made after the draft instead of before it: the point of each paragraph in one sentence, written in the margin and read alone	10
risk ratio	the risk in one group divided by the risk in another	9
rule of four	give a risk ratio, odds ratio or ratio of two percentages two significant figures if its first digit that is not zero is 4 or more, otherwise three	9
sequential colour scale	colours running from light to dark, for a value that runs from low to high	19
short title	a shortened title a journal may want, such as one of no more than 40 characters, letters and spaces both counted	3
showing the data	drawing every value	16
significant (in statistics)	the verdict of a statistical test: roughly, a difference too large to put down to chance; this book does not teach the test	8
stress position	the end of a sentence, which holds the new information the reader should weigh most	7
strip chart	each value drawn as a dot along the value axis, one column of dots per group; also called a univariate scatterplot	16
subject (of a sentence)	what the clause is about, the thing the verb is said of; find the verb, then ask "who or what" in front of it	6

Term	Plain words	Section
subordinate clause	a clause that has a subject and a verb but cannot stand alone; words such as although, because, when or while open one	6
syllable	one beat of a spoken word	11
tense	when the action happens; not the same as voice	6
title (of a figure)	the caption's first sentence, which states the finding; not text drawn across the figure	20
topic position	the start of a sentence, which links back to what the reader already has and names whose story the sentence tells	7
topic sentence	the sentence that states a paragraph's one point, often first	10
truncating an axis	starting a bar chart's value axis above zero	17
value message	the opinion a chart carries ("costs are monstrous"), in Bateman and colleagues' sense	18
vector graphic	an image saved as the shapes themselves, redrawn at any size without losing sharpness; a pdf can hold one	22
verb	the word for what happens: an action, an occurrence or a state of being	6
violin plot	a smoothed outline of where the values fall	16
voice	which of the doer and the thing acted on is the subject: whether the subject did the action or received it; tense says when	6
zig-zag (of a paragraph)	leaving a topic and coming back to it later	10

The series: all 196 books

In build order: Book 0 first, then every Introductory rung, then every Intermediate rung, and so on; within a level, a subject comes after the subjects it depends on. "done" means that book's version 1 is finished. This is Book 9.

GROUND FLOOR		
0	Ground floor	done
INTRODUCTORY		
1	Energy balance and dynamic weight-change modelling · Rung 1	done
2	Mathematics for modelling and inference · Rung 1	done
3	Qualitative and mixed methods · Rung 1	done
4	Food systems, value chains and Indian food policy · Rung 1	done
5	Policy process, framing and advocacy craft · Rung 1	done
6	R, reproducibility engineering and data management · Rung 1	done
7	Question selection and research strategy · Rung 1	done
8	Teaching, mentorship and coalition building · Rung 1	done
9	**Scientific writing, visualisation and public communication · Rung 1 ◀ you are here**	done
10	Epistemics: philosophy of science and calibration · Rung 1	
11	Adjacent specialist domains: recognise and route · Rung 1	
12	Probability and statistical inference · Rung 1	
13	Integrative metabolism, insulin resistance and energy expenditure regulation · Rung 1	
14	Commercial determinants and corporate political activity · Rung 1	
15	System dynamics and complexity concepts · Rung 1	
16	Indian regulatory architecture and the live policy agenda · Rung 1	
17	Comparative and international policy · Rung 1	
18	Python for simulation, scraping and tooling · Rung 1	
19	Regression modelling for health outcomes · Rung 1	
20	Potential outcomes, identification and target trial emulation · Rung 1	
21	Adipose tissue biology, ectopic fat and the South Asian phenotype · Rung 1	
22	Appetite, reward and the determinants of intake · Rung 1	
23	Nutrient science for obesity and cardiometabolic risk · Rung 1	
24	Physical activity and sedentary behaviour science · Rung 1	
25	Sleep science and obstructive sleep apnoea · Rung 1	
26	Participatory systems mapping and group model building · Rung 1	
27	Agent-based and network modelling · Rung 1	
28	Clustered, longitudinal, hierarchical and small-area models · Rung 1	
29	Evidence synthesis and meta-analysis · Rung 1	
30	Prediction modelling and machine learning · Rung 1	
31	Measurement error theory and correction · Rung 1	
32	Causal diagrams, structural bias and sensitivity analysis · Rung 1	
33	Study and trial design, power and missing data · Rung 1	
34	Clinical assessment, staging and secondary causes · Rung 1	
35	Indian dietary patterns, guidelines, food composition and diet cost · Rung 1	
36	Circadian biology and chrononutrition · Rung 1	

37	Behavioural theory, choice architecture and intervention design · Rung 1
38	Measurement instruments: body composition, expenditure, diet, activity, sleep · Rung 1
39	G-methods, time-varying treatment and mediation · Rung 1
40	Instrumental variables and Mendelian randomisation · Rung 1
41	Quasi-experimental policy evaluation · Rung 1
42	Genetics, developmental origins and life-course epidemiology · Rung 1
43	Social epidemiology, equity measurement and surveillance in India · Rung 1
44	Nutritional epidemiology and its pathologies · Rung 1
45	Lifestyle, dietary and behavioural treatment · Rung 1
46	Pharmacotherapy of obesity · Rung 1
47	Metabolic and bariatric surgery: selection and co-management · Rung 1
48	Eating behaviour, mental health and weight stigma · Rung 1
49	Food environment measurement, retail and digital delivery · Rung 1
50	Food demand, elasticity and the economics of externality · Rung 1
51	Implementation science and programme craft · Rung 1
52	Protocol, preregistration, reporting standards and research ethics · Rung 1
53	Life-stage obesity: paediatric, adolescent and pregnancy · Rung 1
54	Ultra-processed food and the effects of processing · Rung 1
55	Activity measurement and compositional 24-hour analysis · Rung 1
56	Fiscal instruments: taxation and subsidy design in India · Rung 1
57	Burden of disease, economic evaluation and decision-analytic modelling · Rung 1
58	The Indian health system and its financing mechanics · Rung 1
59	India's data infrastructure, access and complex survey analysis · Rung 1
60	Ethics of obesity intervention and industry engagement · Rung 1
61	Health financing, HTA and pharmaceutical access in India · Rung 1
INTERMEDIATE	
62	Energy balance and dynamic weight-change modelling · Rung 2
63	Mathematics for modelling and inference · Rung 2
64	Qualitative and mixed methods · Rung 2
65	Food systems, value chains and Indian food policy · Rung 2
66	Policy process, framing and advocacy craft · Rung 2
67	R, reproducibility engineering and data management · Rung 2
68	Question selection and research strategy · Rung 2
69	Teaching, mentorship and coalition building · Rung 2
70	Scientific writing, visualisation and public communication · Rung 2
71	Epistemics: philosophy of science and calibration · Rung 2
72	Probability and statistical inference · Rung 2
73	Integrative metabolism, insulin resistance and energy expenditure regulation · Rung 2
74	Commercial determinants and corporate political activity · Rung 2
75	System dynamics and complexity concepts · Rung 2
76	Indian regulatory architecture and the live policy agenda · Rung 2
77	Comparative and international policy · Rung 2
78	Python for simulation, scraping and tooling · Rung 2
79	Regression modelling for health outcomes · Rung 2
80	Potential outcomes, identification and target trial emulation · Rung 2
81	Adipose tissue biology, ectopic fat and the South Asian phenotype · Rung 2

82	Appetite, reward and the determinants of intake · Rung 2
83	Nutrient science for obesity and cardiometabolic risk · Rung 2
84	Physical activity and sedentary behaviour science · Rung 2
85	Sleep science and obstructive sleep apnoea · Rung 2
86	Participatory systems mapping and group model building · Rung 2
87	Agent-based and network modelling · Rung 2
88	Clustered, longitudinal, hierarchical and small-area models · Rung 2
89	Evidence synthesis and meta-analysis · Rung 2
90	Prediction modelling and machine learning · Rung 2
91	Measurement error theory and correction · Rung 2
92	Causal diagrams, structural bias and sensitivity analysis · Rung 2
93	Study and trial design, power and missing data · Rung 2
94	Clinical assessment, staging and secondary causes · Rung 2
95	Indian dietary patterns, guidelines, food composition and diet cost · Rung 2
96	Circadian biology and chrononutrition · Rung 2
97	Behavioural theory, choice architecture and intervention design · Rung 2
98	Measurement instruments: body composition, expenditure, diet, activity, sleep · Rung 2
99	G-methods, time-varying treatment and mediation · Rung 2
100	Instrumental variables and Mendelian randomisation · Rung 2
101	Quasi-experimental policy evaluation · Rung 2
102	Genetics, developmental origins and life-course epidemiology · Rung 2
103	Social epidemiology, equity measurement and surveillance in India · Rung 2
104	Nutritional epidemiology and its pathologies · Rung 2
105	Lifestyle, dietary and behavioural treatment · Rung 2
106	Pharmacotherapy of obesity · Rung 2
107	Metabolic and bariatric surgery: selection and co-management · Rung 2
108	Eating behaviour, mental health and weight stigma · Rung 2
109	Food environment measurement, retail and digital delivery · Rung 2
110	Food demand, elasticity and the economics of externality · Rung 2
111	Implementation science and programme craft · Rung 2
112	Protocol, preregistration, reporting standards and research ethics · Rung 2
113	Life-stage obesity: paediatric, adolescent and pregnancy · Rung 2
114	Ultra-processed food and the effects of processing · Rung 2
115	Activity measurement and compositional 24-hour analysis · Rung 2
116	Fiscal instruments: taxation and subsidy design in India · Rung 2
117	Burden of disease, economic evaluation and decision-analytic modelling · Rung 2
118	The Indian health system and its financing mechanics · Rung 2
119	India's data infrastructure, access and complex survey analysis · Rung 2
120	Ethics of obesity intervention and industry engagement · Rung 2
121	Health financing, HTA and pharmaceutical access in India · Rung 2
ADVANCED	
122	Energy balance and dynamic weight-change modelling · Rung 3
123	Mathematics for modelling and inference · Rung 3
124	Qualitative and mixed methods · Rung 3
125	Food systems, value chains and Indian food policy · Rung 3
126	Policy process, framing and advocacy craft · Rung 3

127	R, reproducibility engineering and data management · Rung 3
128	Question selection and research strategy · Rung 3
129	Teaching, mentorship and coalition building · Rung 3
130	Scientific writing, visualisation and public communication · Rung 3
131	Epistemics: philosophy of science and calibration · Rung 3
132	Probability and statistical inference · Rung 3
133	Integrative metabolism, insulin resistance and energy expenditure regulation · Rung 3
134	Commercial determinants and corporate political activity · Rung 3
135	System dynamics and complexity concepts · Rung 3
136	Indian regulatory architecture and the live policy agenda · Rung 3
137	Regression modelling for health outcomes · Rung 3
138	Potential outcomes, identification and target trial emulation · Rung 3
139	Adipose tissue biology, ectopic fat and the South Asian phenotype · Rung 3
140	Appetite, reward and the determinants of intake · Rung 3
141	Nutrient science for obesity and cardiometabolic risk · Rung 3
142	Physical activity and sedentary behaviour science · Rung 3
143	Sleep science and obstructive sleep apnoea · Rung 3
144	Participatory systems mapping and group model building · Rung 3
145	Clustered, longitudinal, hierarchical and small-area models · Rung 3
146	Evidence synthesis and meta-analysis · Rung 3
147	Measurement error theory and correction · Rung 3
148	Causal diagrams, structural bias and sensitivity analysis · Rung 3
149	Study and trial design, power and missing data · Rung 3
150	Clinical assessment, staging and secondary causes · Rung 3
151	Indian dietary patterns, guidelines, food composition and diet cost · Rung 3
152	Circadian biology and chrononutrition · Rung 3
153	Behavioural theory, choice architecture and intervention design · Rung 3
154	Measurement instruments: body composition, expenditure, diet, activity, sleep · Rung 3
155	G-methods, time-varying treatment and mediation · Rung 3
156	Instrumental variables and Mendelian randomisation · Rung 3
157	Quasi-experimental policy evaluation · Rung 3
158	Genetics, developmental origins and life-course epidemiology · Rung 3
159	Social epidemiology, equity measurement and surveillance in India · Rung 3
160	Nutritional epidemiology and its pathologies · Rung 3
161	Lifestyle, dietary and behavioural treatment · Rung 3
162	Pharmacotherapy of obesity · Rung 3
163	Metabolic and bariatric surgery: selection and co-management · Rung 3
164	Eating behaviour, mental health and weight stigma · Rung 3
165	Food environment measurement, retail and digital delivery · Rung 3
166	Food demand, elasticity and the economics of externality · Rung 3
167	Implementation science and programme craft · Rung 3
168	Protocol, preregistration, reporting standards and research ethics · Rung 3
169	Life-stage obesity: paediatric, adolescent and pregnancy · Rung 3
170	Ultra-processed food and the effects of processing · Rung 3
171	Activity measurement and compositional 24-hour analysis · Rung 3
172	Fiscal instruments: taxation and subsidy design in India · Rung 3
173	Burden of disease, economic evaluation and decision-analytic modelling · Rung 3

174	The Indian health system and its financing mechanics · Rung 3
175	India's data infrastructure, access and complex survey analysis · Rung 3
176	Ethics of obesity intervention and industry engagement · Rung 3

EXPERT

177	Energy balance and dynamic weight-change modelling · Rung 4
178	R, reproducibility engineering and data management · Rung 4
179	Question selection and research strategy · Rung 4
180	Probability and statistical inference · Rung 4
181	Indian regulatory architecture and the live policy agenda · Rung 4
182	Regression modelling for health outcomes · Rung 4
183	Potential outcomes, identification and target trial emulation · Rung 4
184	Adipose tissue biology, ectopic fat and the South Asian phenotype · Rung 4
185	Appetite, reward and the determinants of intake · Rung 4
186	Measurement error theory and correction · Rung 4
187	Causal diagrams, structural bias and sensitivity analysis · Rung 4
188	Circadian biology and chrononutrition · Rung 4
189	Measurement instruments: body composition, expenditure, diet, activity, sleep · Rung 4
190	Genetics, developmental origins and life-course epidemiology · Rung 4
191	Nutritional epidemiology and its pathologies · Rung 4
192	Eating behaviour, mental health and weight stigma · Rung 4
193	Activity measurement and compositional 24-hour analysis · Rung 4
194	India's data infrastructure, access and complex survey analysis · Rung 4
195	Ethics of obesity intervention and industry engagement · Rung 4

OBESITY EXPERTISE · BOOK 9

Scientific writing, visualisation and public communication · Rung 1

A short book on the first craft of scientific communication. It covers what a research paper is and why it is an argument rather than a diary of what was done, how to write the one sentence your paper exists to say, the sentence and the paragraph as units of clarity, how to report a number, how to cite a source, how to measure and cut a draft, how to make a table that stands alone, and how to choose, draw and caption a figure whose axes tell the truth. It ends with one real Indian survey result taken from a table to a paragraph and a figure a colleague can restate.

Book 9 of the series' 196 books, free to read at drharshmaheshwari.com.



Dr. Harsh Maheshwari
MD Community Medicine Resident, AIIMS Raipur
contact@drharshmaheshwari.com
ORCID 0000-0001-8848-1918
drharshmaheshwari.com
Version 1.0 · 3 October 2026

CC BY-NC-SA 4.0 · free to share, not to sell