Communicating research data

From principles to figures

HoloGen Retreat 4, Leuven · October 2026

Leo Lahti (Department of Computing, University of Turku, Finland)

datascience.utu.fi | @antagomir



Turun yliopisto
University of Turku



Plan for today

Part I: Visualising research data

Team work: figure critique

Ten simple rules for better figures; Tufte's principles

Break

15 min

Part II: Communicating research

Seven principles; group discussion

Why and to whom: principles and audiences

Related HoloGen training: Retreat 1, statistical concepts · Retreat 3, storyboarding · Retreat 4 (now), data visualisation techniques (Aerts) and scientific writing (Hall)

Source: HoloGen Doctoral Network (n.d.)

Part I: Visualising research data

Communicating research data

HoloGen Retreat 4, Leuven · October 2026

Leo Lahti (Department of Computing, University of Turku, Finland)

datascience.utu.fi | @antagomir

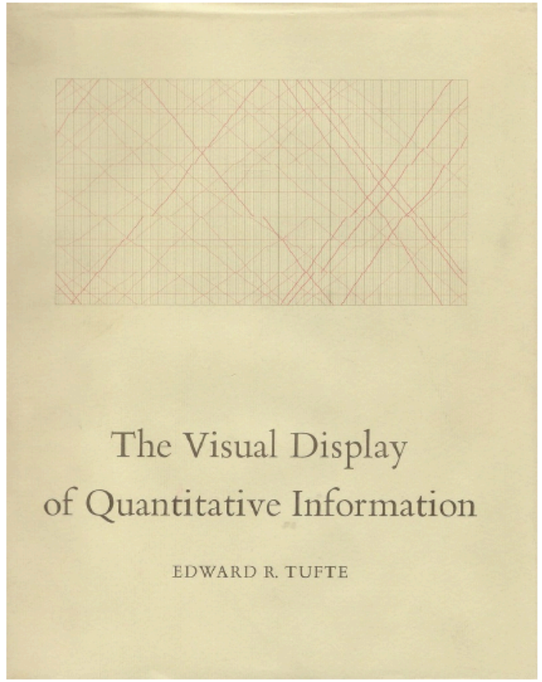


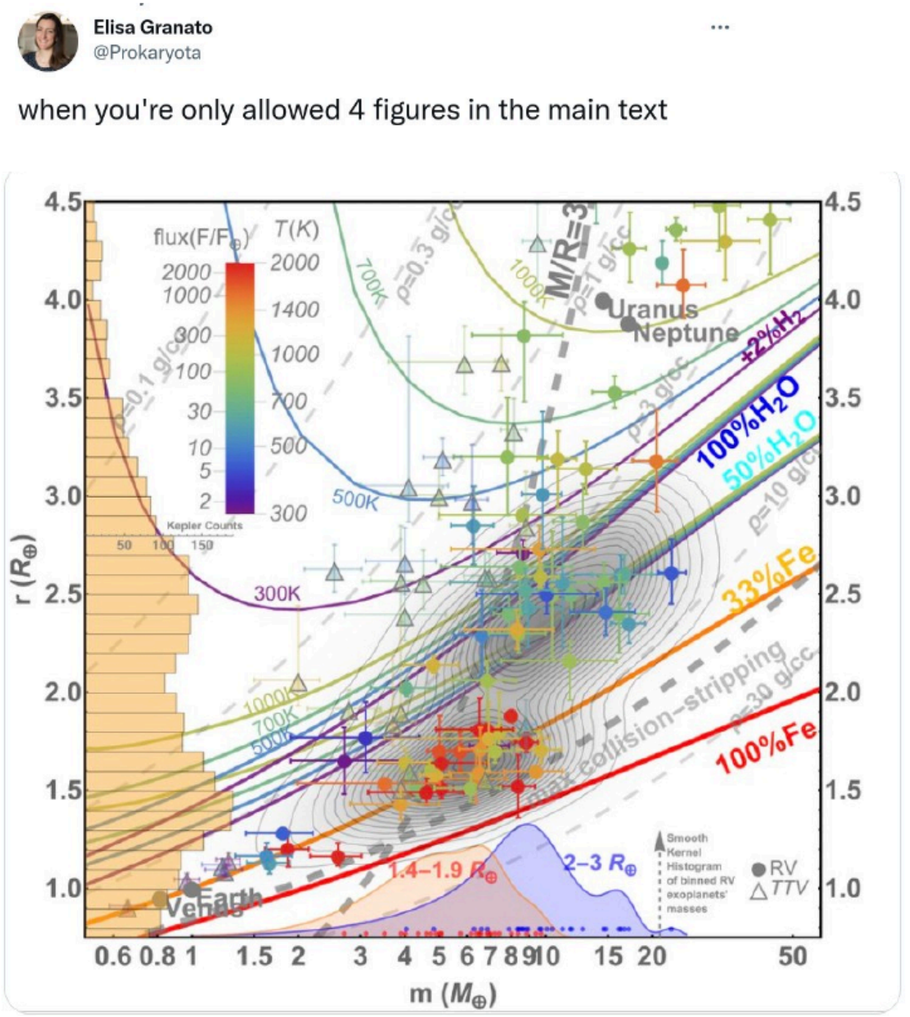
Turun yliopisto
University of Turku

Visualization principles (Tufte, 1983)

Excellence in statistical graphics:

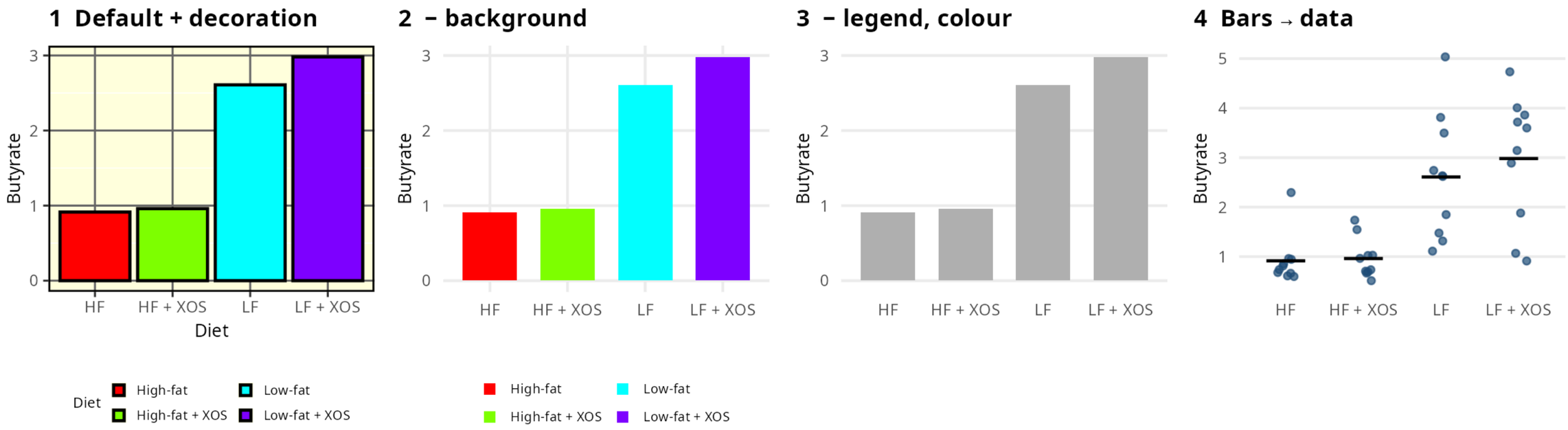
complex ideas communicated with clarity, precision, and efficiency





Source: Tufte (1983); example figure shared on social media by Elisa Granato (@Prokaryota); slide from Lahti (2022)

> *“Perfection is achieved, not when there is nothing more to add, but when there is nothing more to take away.”*
>
> — Antoine de Saint Exupéry



Data: HintikkaXOData (Hintikka et al. 2021) via mia

Terre des hommes (1939); quoted in Publications Office of the European Union (2019); erasing redesigns after Tufte (1983) (“maximise the data-ink ratio”); data: Hintikka et al. (2021)

Ten simple rules and Tufte’s nine principles

Ten simple rules for better figures

1. Know your audience
2. Identify your message
3. Adapt the figure to the medium
4. Captions are not optional
5. Do not trust the defaults
6. Use colour effectively
7. Do not mislead the reader
8. Avoid chartjunk
9. Message trumps beauty
10. Get the right tool

Rougier, Droettboom & Bourne (2014)

Tufte: graphical displays should…

1. Show the data
2. Make the viewer think about the substance
3. Avoid distorting what the data say
4. Present many numbers in a small space
5. Make large data sets coherent
6. Encourage the eye to compare
7. Reveal several levels of detail
8. Serve a clear purpose
9. Integrate with statistics and text

Tufte (1983), chapter 1

Sources: Rougier, Droettboom, and Bourne (2014); Tufte (1983)

Team work: figure critique (30 min)

1 · In teams of 3–4 (10 min)

Choose one figure to criticise: your own, or from a recent paper in your field. Then:

1. What is the **message**? Is it stated in the title or caption?
2. Which **encodings** carry the data? Are they the most accurate ones?
3. Reflect it against the **ten simple rules** and/or **Tufte's nine principles** (references below)
4. Is it **accessible**: colour, text size, alt text?
5. **Sketch** an improved version on paper

2 · Presentations (20 min)

Five teams × 4 min: each team shows its figure

- the figure and its intended message
- the two most important problems
- the improved sketch, and what it changes

Class: one question or suggestion per team

Ten simple rules: Rougier NP, Droettboom M, Bourne PE (2014). Ten simple rules for better figures. *PLOS Computational Biology* 10(9): e1003833. doi:10.1371/journal.pcbi.1003833

Tufte's principles: Tufte ER (1983). *The Visual Display of Quantitative Information*, chapter 1. Cheshire, CT: Graphics Press.

Optional checklist: Evergreen & Emery, *Data Visualization Checklist* (datavizchecklist.stephanieevergreen.com) (via Publications Office of the European Union 2019)

Rougier, Droettboom, and Bourne (2014); Tufte (1983)

Tufte’s nine principles: graphical displays should…

1 Show the data

Rule 8: bars → data points

2 Make the viewer think about the substance, not methodology or design

3 Avoid distorting what the data say

Rule 7: truncated axis

4 Present many numbers in a small space

Rule 1: heatmap

5 Make large data sets coherent

6 Encourage the eye to compare

Common baselines, small multiples

7 Reveal the data at several levels of detail, from overview to fine structure

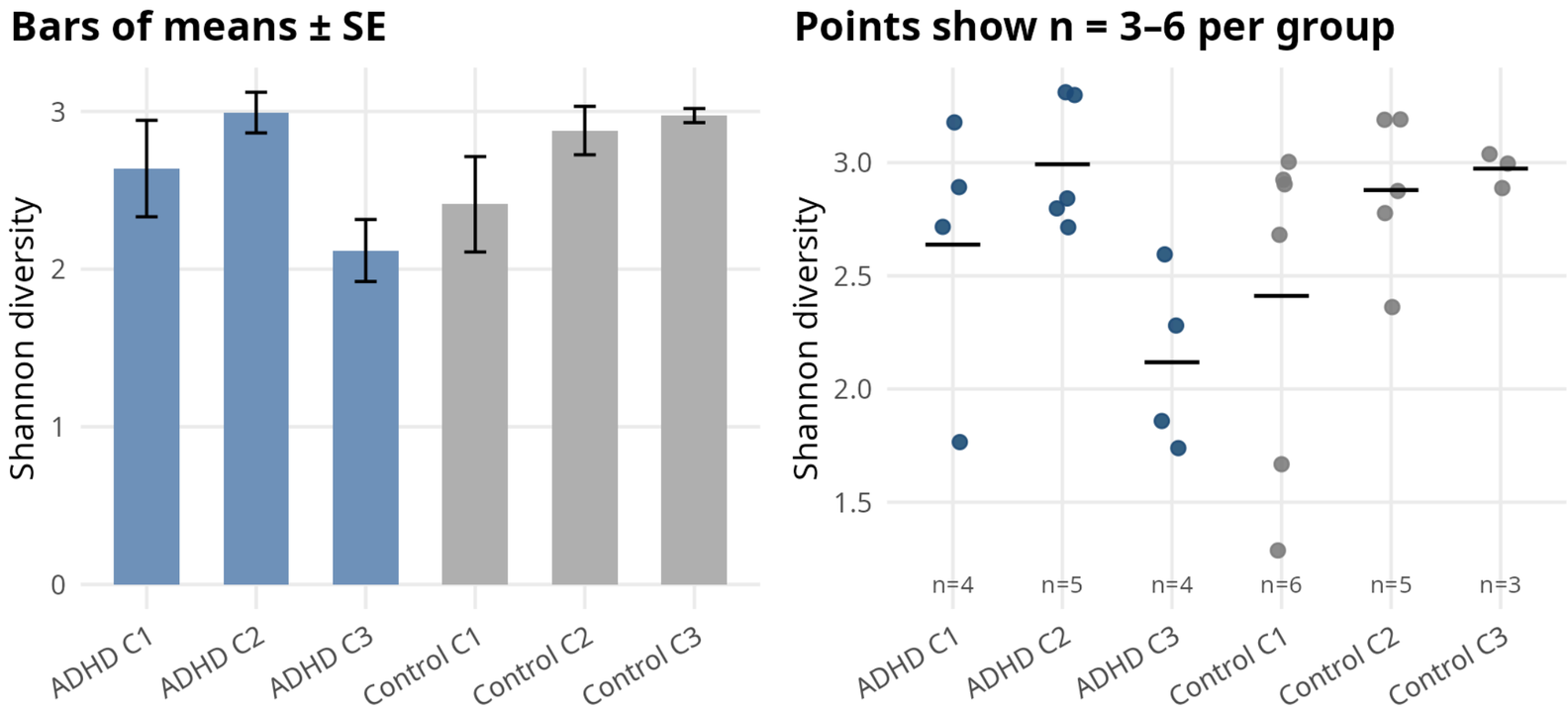
8 Serve a clear purpose: description, exploration, tabulation or decoration

9 Be integrated with the statistical and verbal descriptions

Rule 4: captions

Source: Tufte (1983) (chapter 1); one slide per principle follows; rule examples after Rougier, Droettboom, and Bourne (2014)

Tufte 1: Show the data

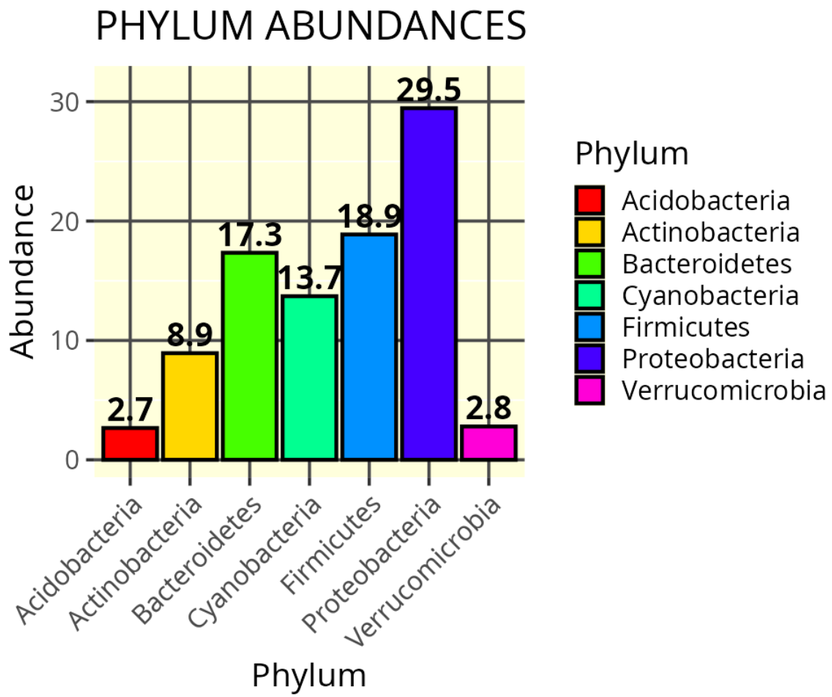


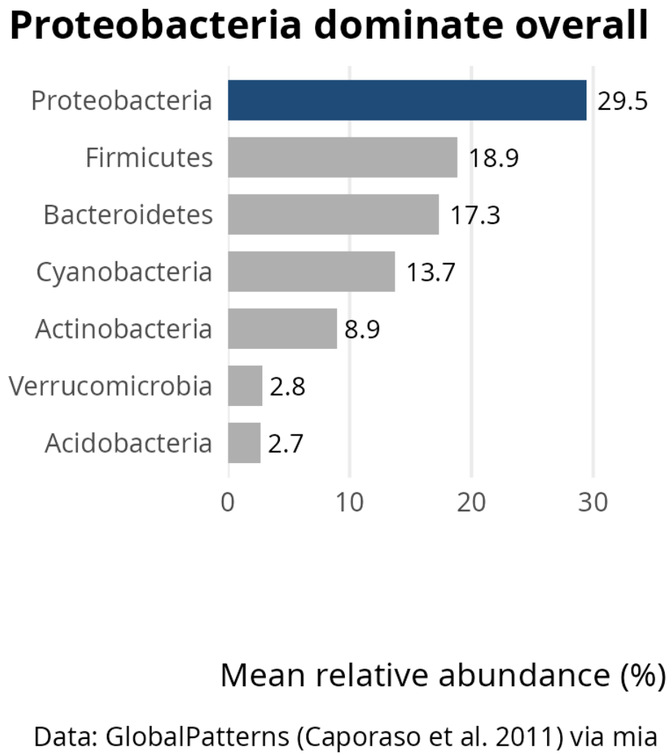
Data: Tengeler2020 (Tengeler et al. 2020) via mia; ADHD patients vs controls, 3 cohorts

With *n* = 3–6 per group, a bar of means hides the data; the points show them.

Tufte (1983) (chapter 1); Weissgerber et al. (2015); data: Tengeler et al. (2020)

Tufte 2: Make the viewer think about the substance

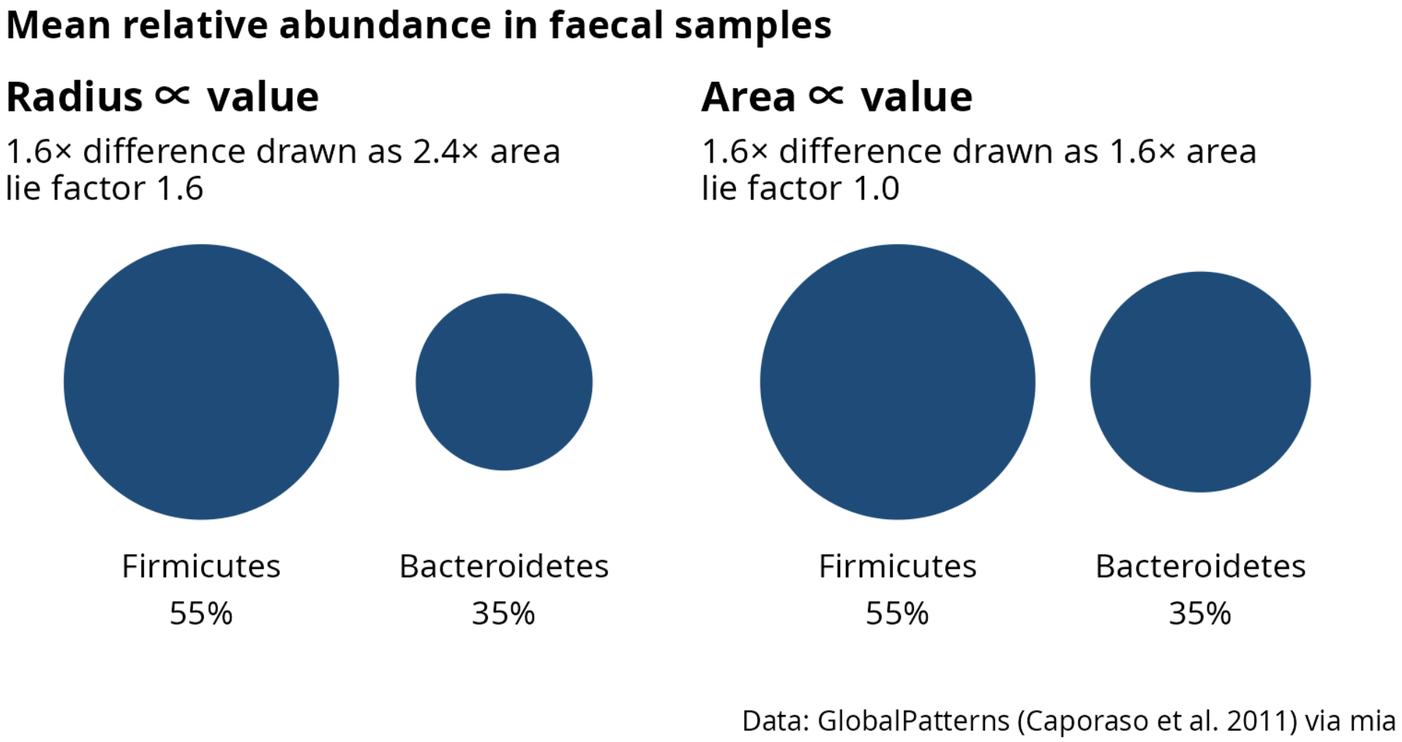




Rather than about the methodology, graphic design or technology: plain design and a message title point to the result.

Tufte (1983) (chapter 1); data: Caporaso et al. (2011)

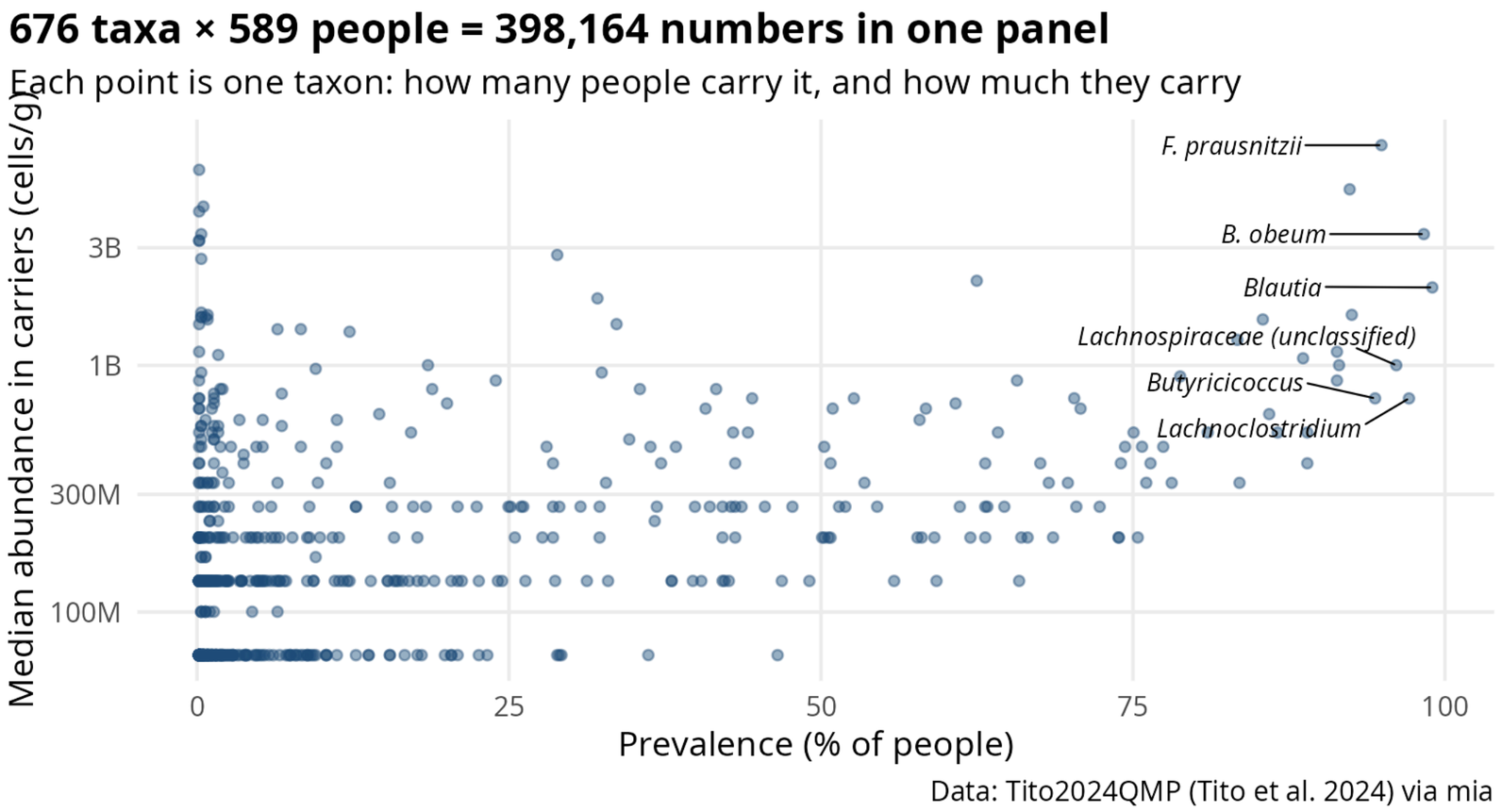
Tufte 3: Avoid distorting what the data say



Lie factor = size of the effect shown in the graphic ÷ size of the effect in the data. It should be close to 1.

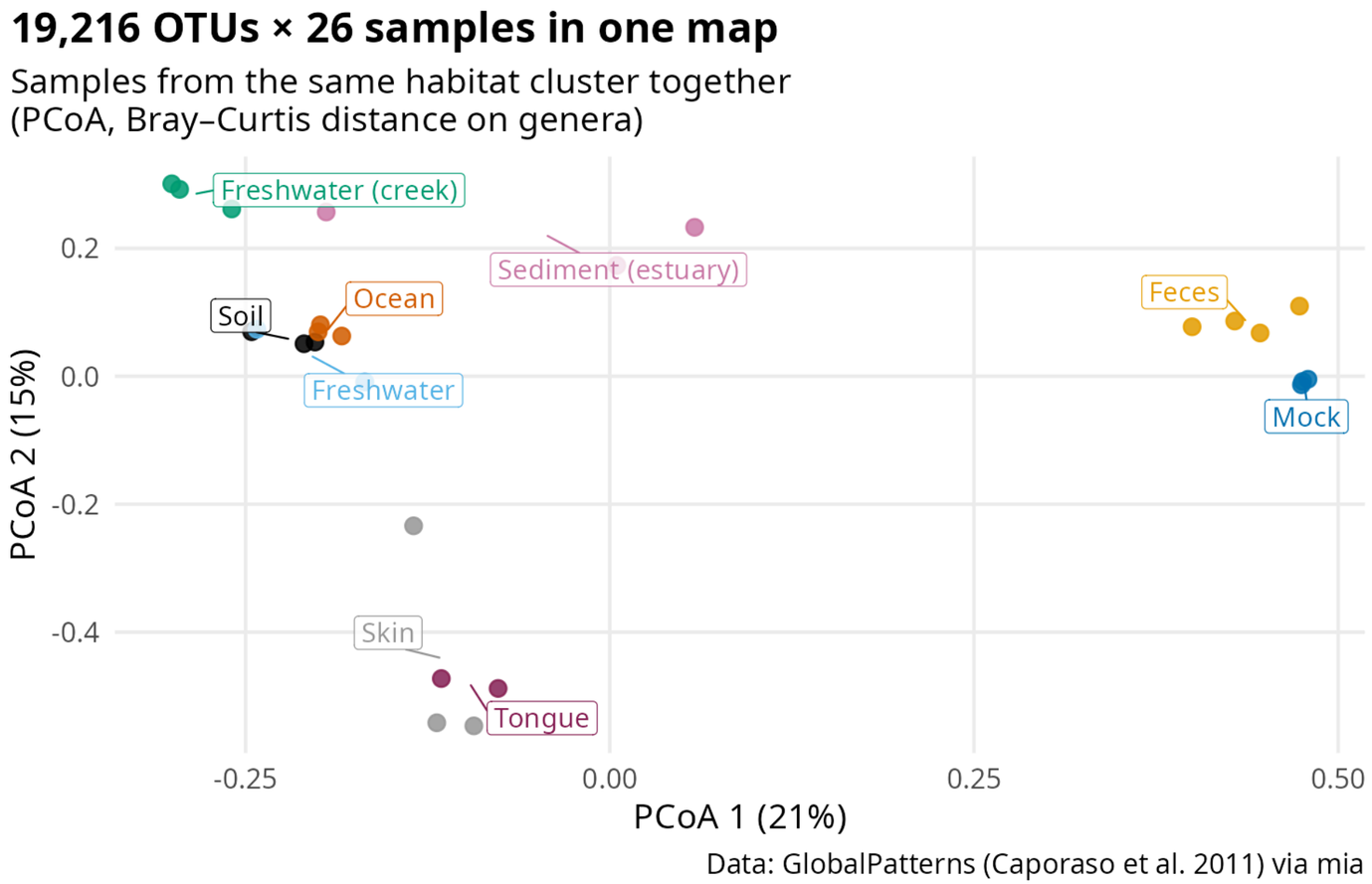
Tufte (1983) (chapter 1); data: Caporaso et al. (2011)

Tufte 4: Present many numbers in a small space



Tufte (1983) (chapter 1); data: Tito et al. (2024)

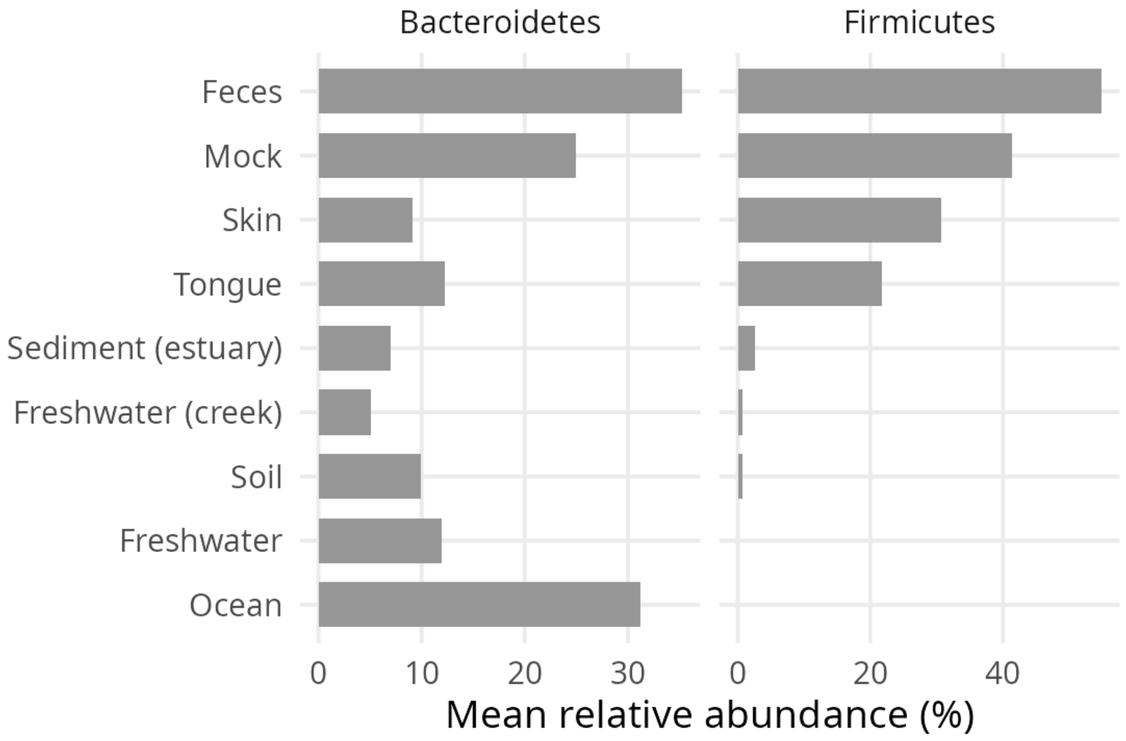
Tufte 5: Make large data sets coherent



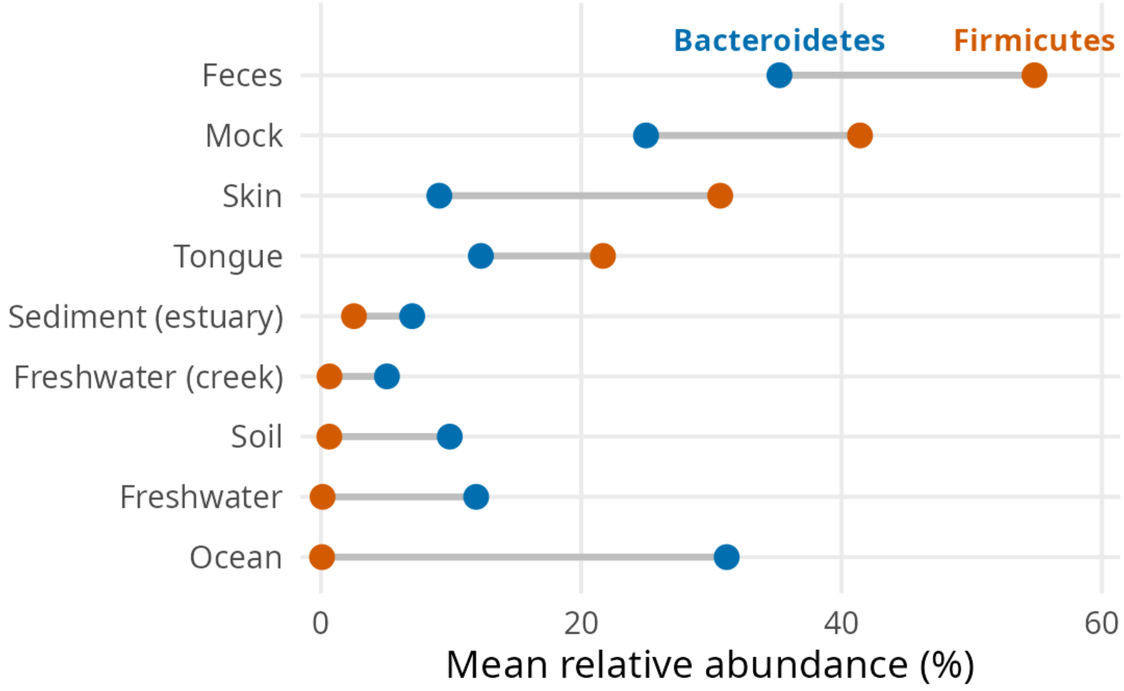
Tufte (1983) (chapter 1); data: Caporaso et al. (2011)

Tufte 6: Encourage the eye to compare

Hard: separate panels, different scales



Easy: side by side on one common axis



Data: GlobalPatterns (Caporaso et al. 2011) via mia

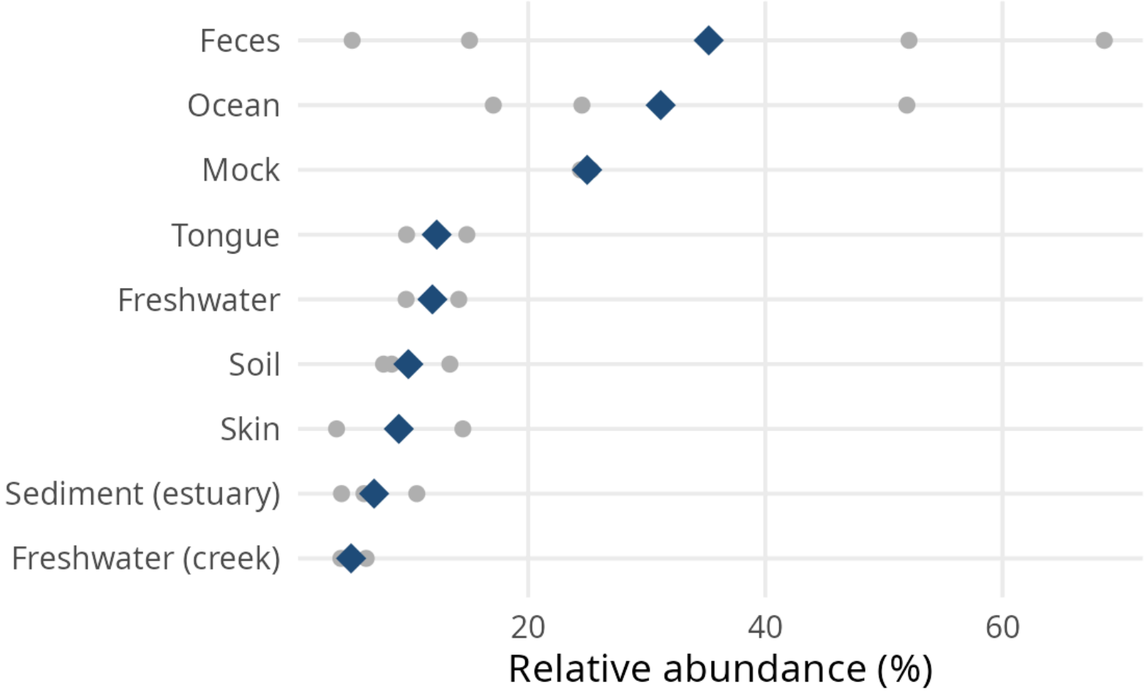
Put what should be compared **side by side, on a common scale.**

Tufte (1983) (chapter 1); Cleveland and McGill (1984); data: Caporaso et al. (2011)

Tufte 7: Reveal the data at several levels of detail

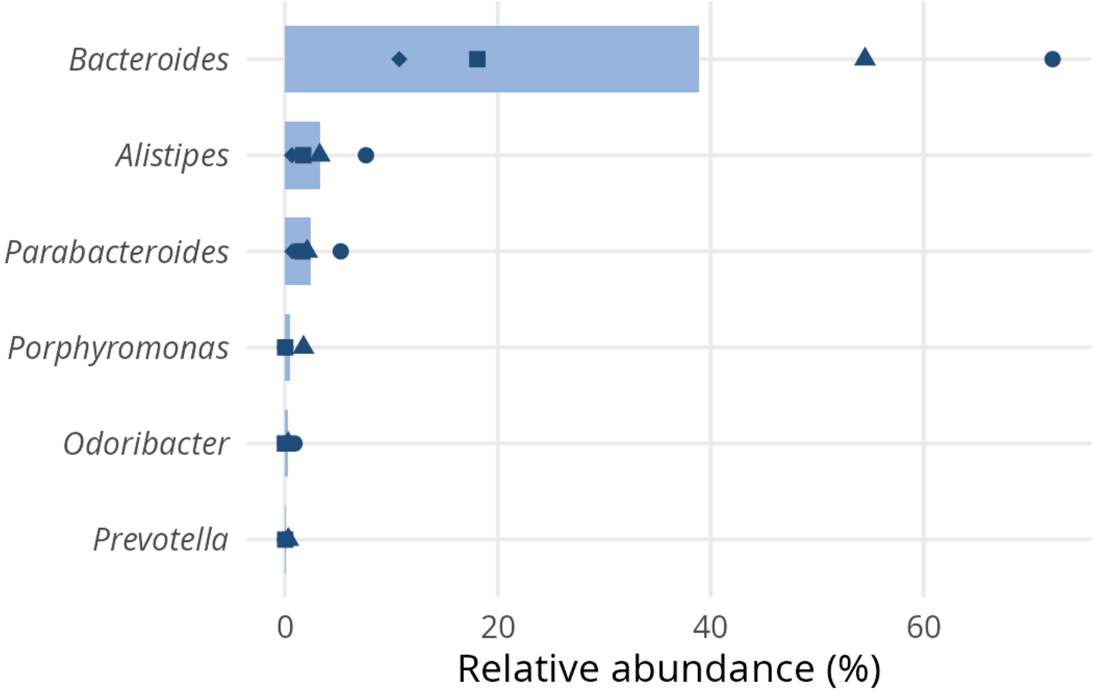
Overview + samples: Bacteroidetes by habitat

Diamond: habitat mean; grey: each sample



Fine structure: its genera in faeces

Bars: mean; points: the four faecal samples



Data: GlobalPatterns (Caporaso et al. 2011) via mia

From a broad overview to the fine structure: group means, individual samples, then the genera within one phylum.

Tufte (1983) (chapter 1); data: Caporaso et al. (2011)

Tufte 8: Serve a clear purpose

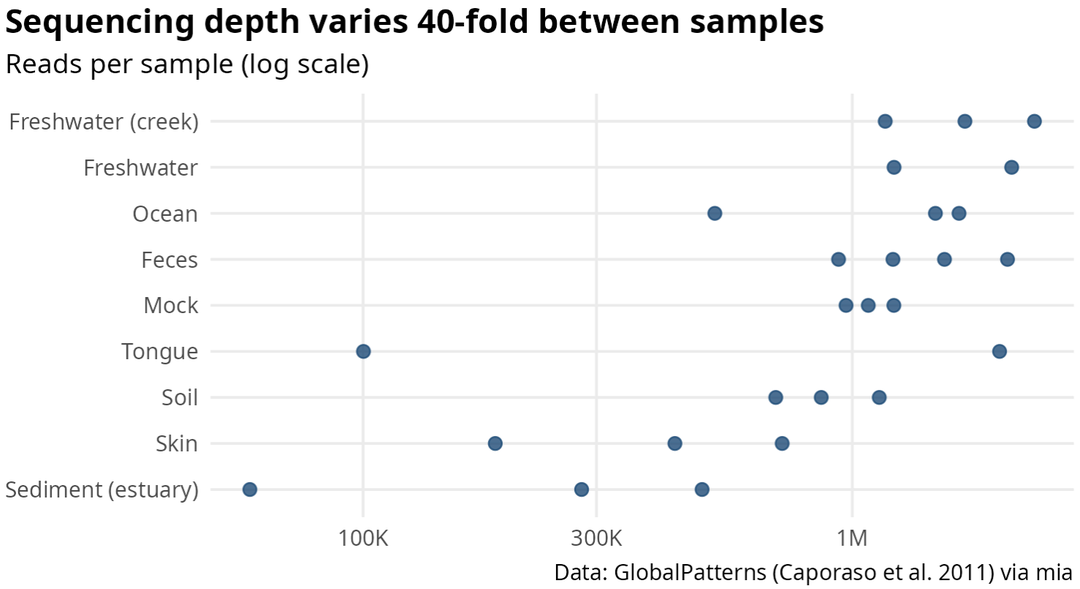
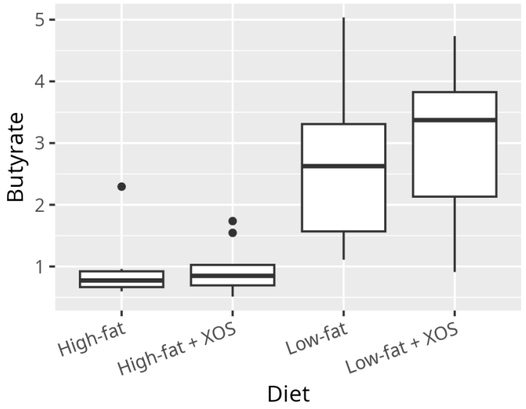
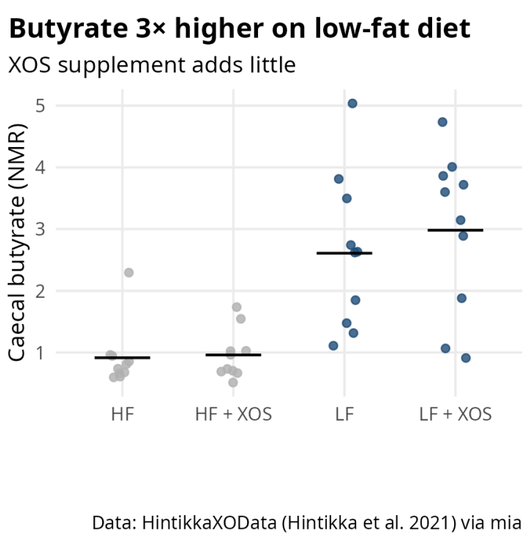


Figure 2. Butyrate





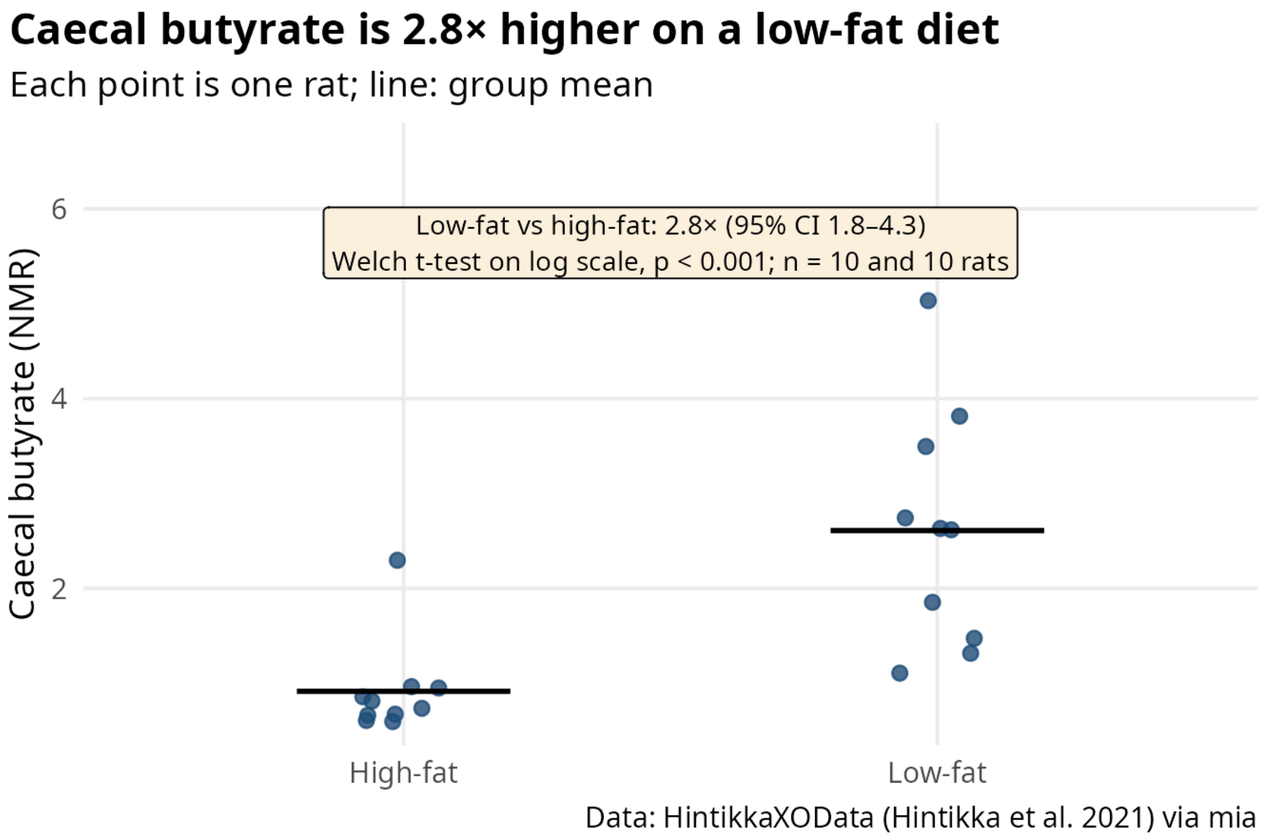
Exploration and quality control: for the analyst

Description: one message for the reader

Description, exploration, tabulation or decoration: decide which one a figure is for, and design for that purpose.

Tufte (1983) (chapter 1); data: Caporaso et al. (2011); Hintikka et al. (2021)

Tufte 9: Integrate with the statistical and verbal descriptions



Figure, statistics and text tell one story

On the figure: estimate, interval, test, *n*

In the caption: what each mark is, where the data come from

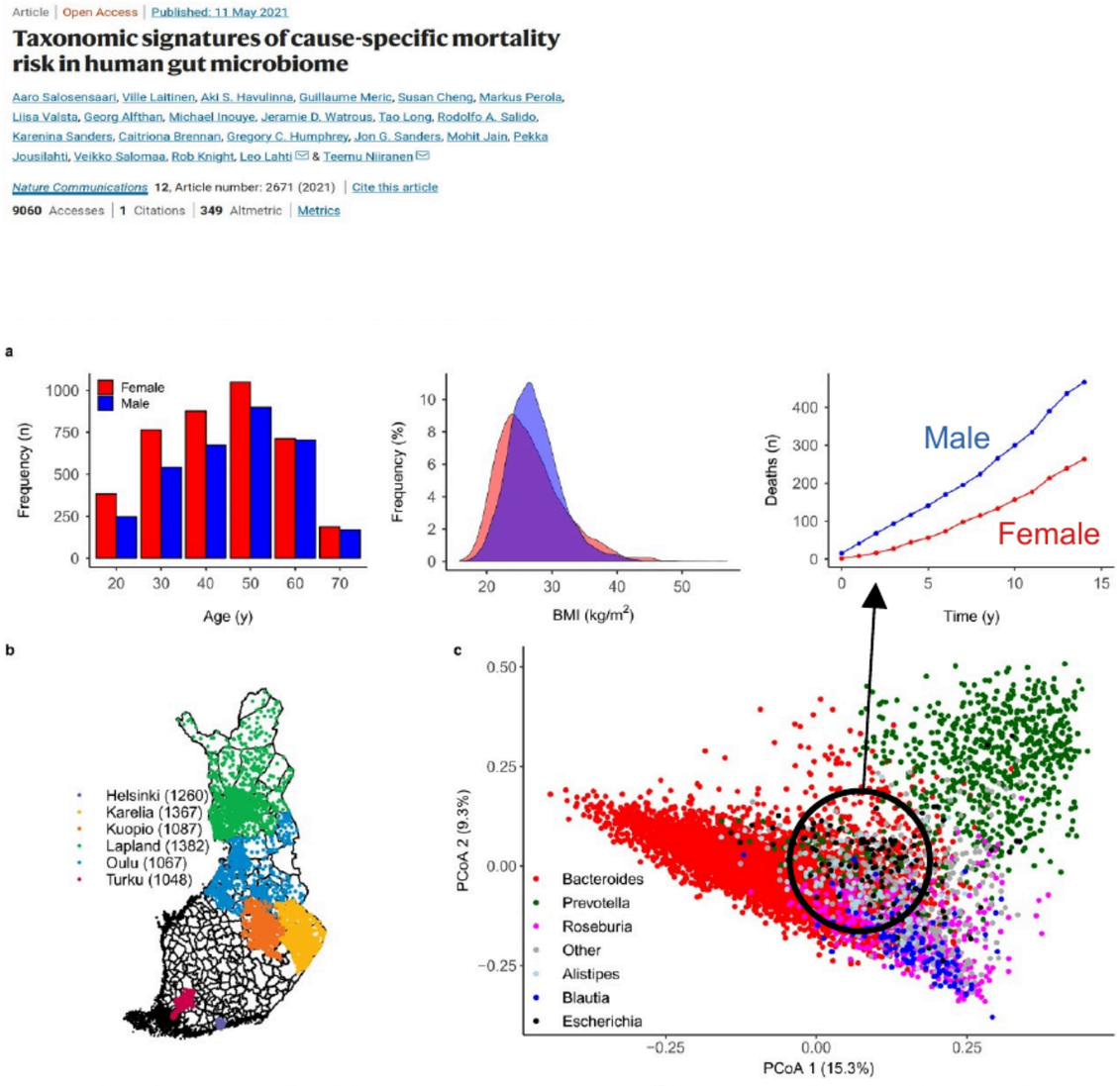
In the text: *"Caecal butyrate was 2.8-fold higher on the low-fat diet (95% CI 1.8–4.3; p < 0.001; n = 10 and 10)."*

Tufte (1983) (chapter 1); data: Hintikka et al. (2021)

Tufte visualization principles:

- make large data sets coherent
- show the data
- induce the viewer to think about the *substance* rather than methodology, graphic design, or something else
- avoid distorting what the data has to say
- present many numbers in a small space
- encourage the eye to compare different pieces of data
- reveal the data at several levels of detail, from a broad overview to the fine structure
- serve a reasonably clear purpose: description, exploration, tabulation, or decoration
- be closely integrated with the statistical and verbal descriptions of a data set.

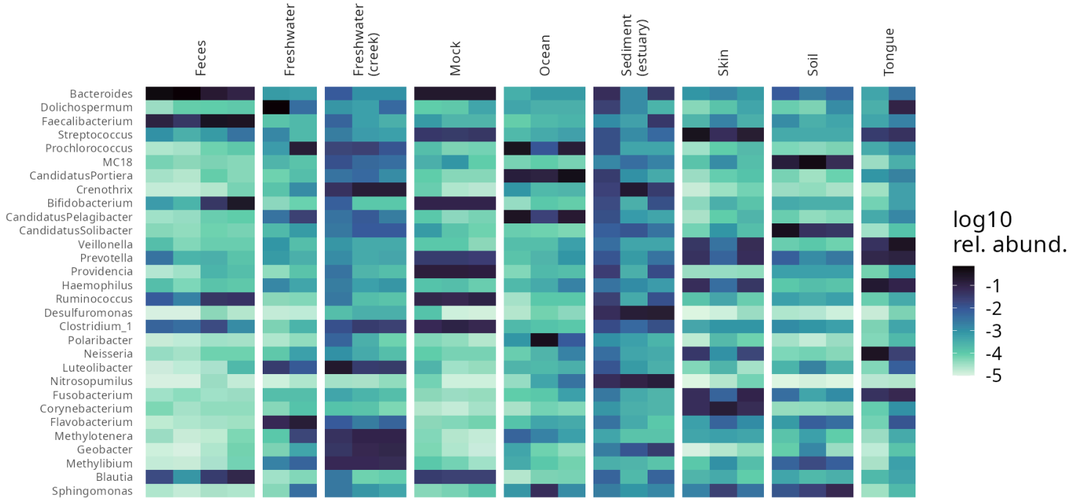
Graphics reveal data. Graphics can be more precise and revealing than conventional statistical computations.



Source: Tufte (1983); example: Salosensaari et al. (2021)

Rule 1: Know your audience

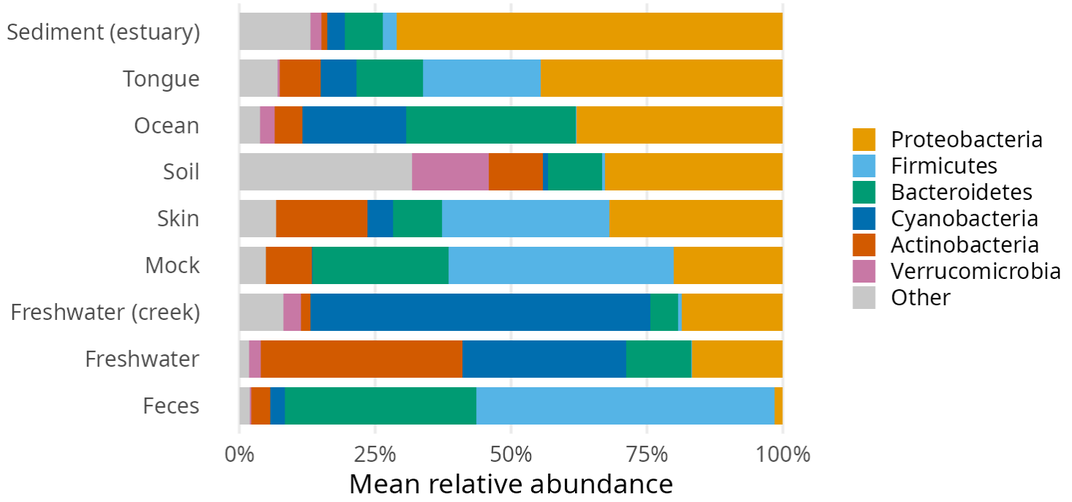
780 numbers at a glance: top 30 genera × 26 samples



Data: GlobalPatterns (Caporaso et al. 2011) via mia

For specialists: dense, many numbers

Microbiome composition by habitat

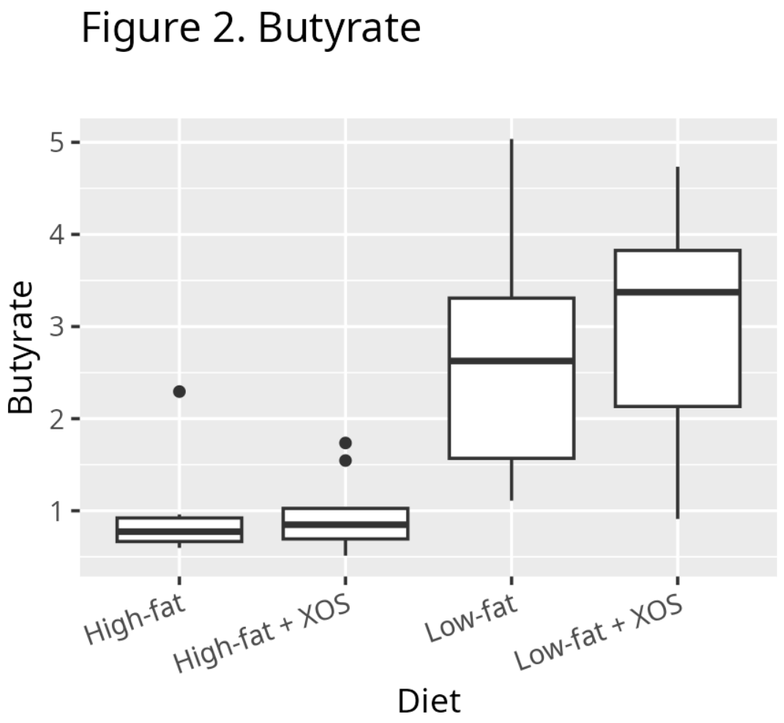


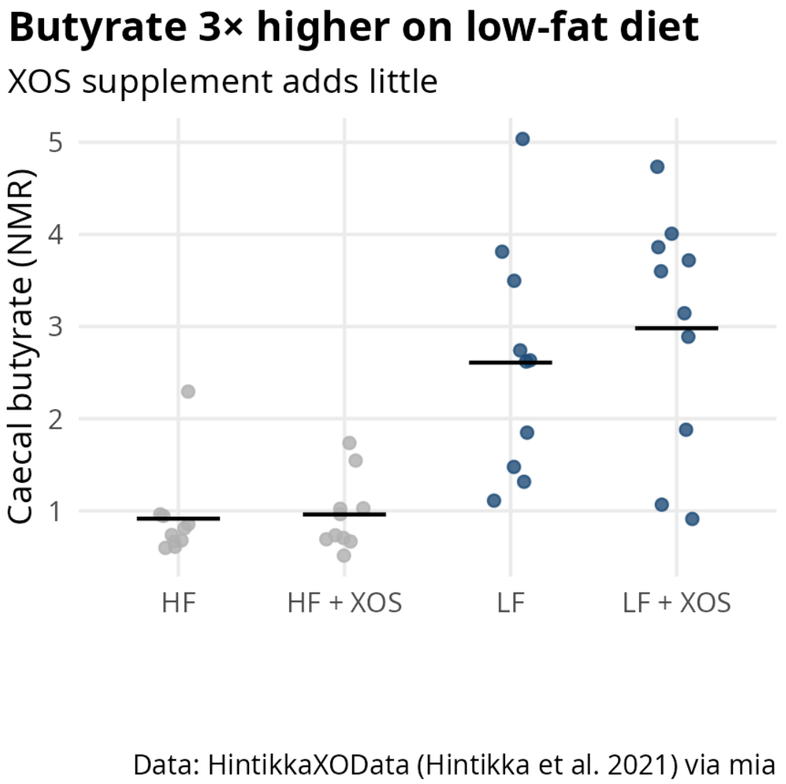
Data: GlobalPatterns (Caporaso et al. 2011) via mia; code: figures.R

For a general audience: one level, few categories

Rougier, Droettboom, and Bourne (2014) (rule 1); data: Caporaso et al. (2011)

Rule 2: Identify your message



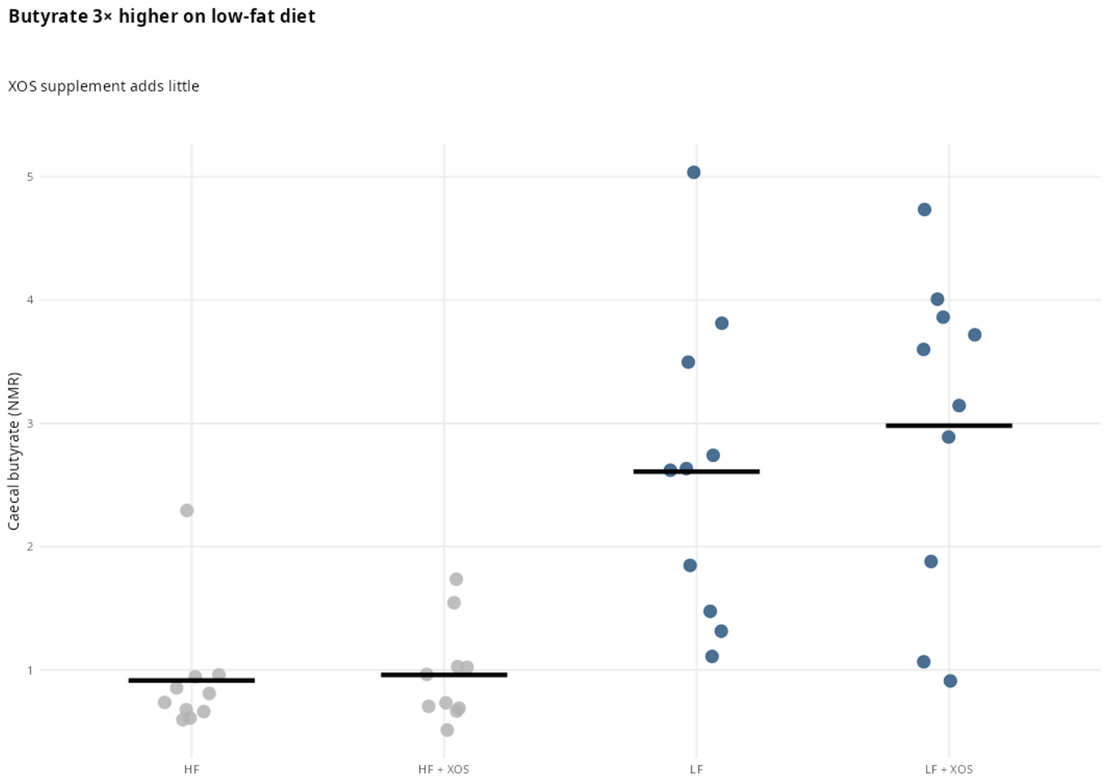


Data: HintikkaXOData (Hintikka et al. 2021) via mia

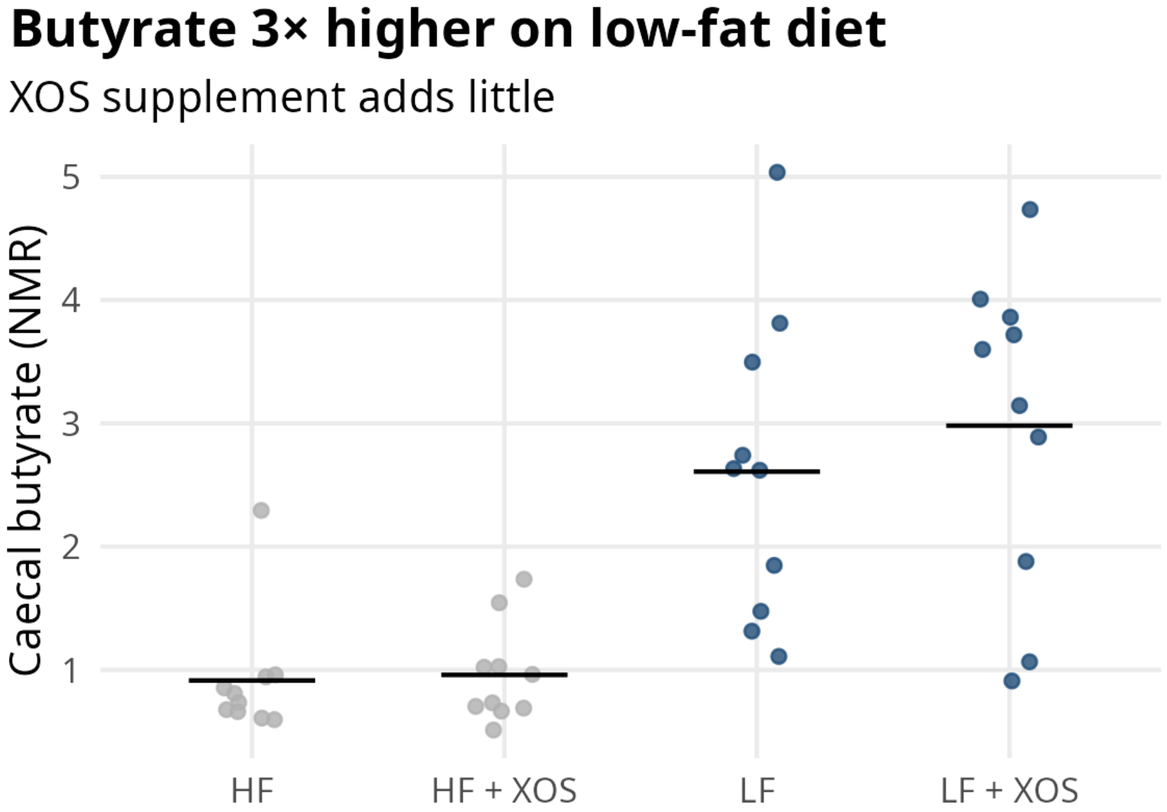
Write the take-home message in **one sentence** first, then put it in the title. *"Message trumps beauty."*

Sources: Rougier, Droettboom, and Bourne (2014) (rules 2, 9); Publications Office of the European Union (2019); data: Hintikka et al. (2021)

Rule 3: Adapt the figure to the medium



Sized for a printed page (7 pt)



Sized for a slide (20 pt)

A figure made for a printed page is unreadable when projected; posters need larger text still.

Rougier, Droettboom, and Bourne (2014) (rule 3); data: Hintikka et al. (2021)

Rule 4: Captions are not optional

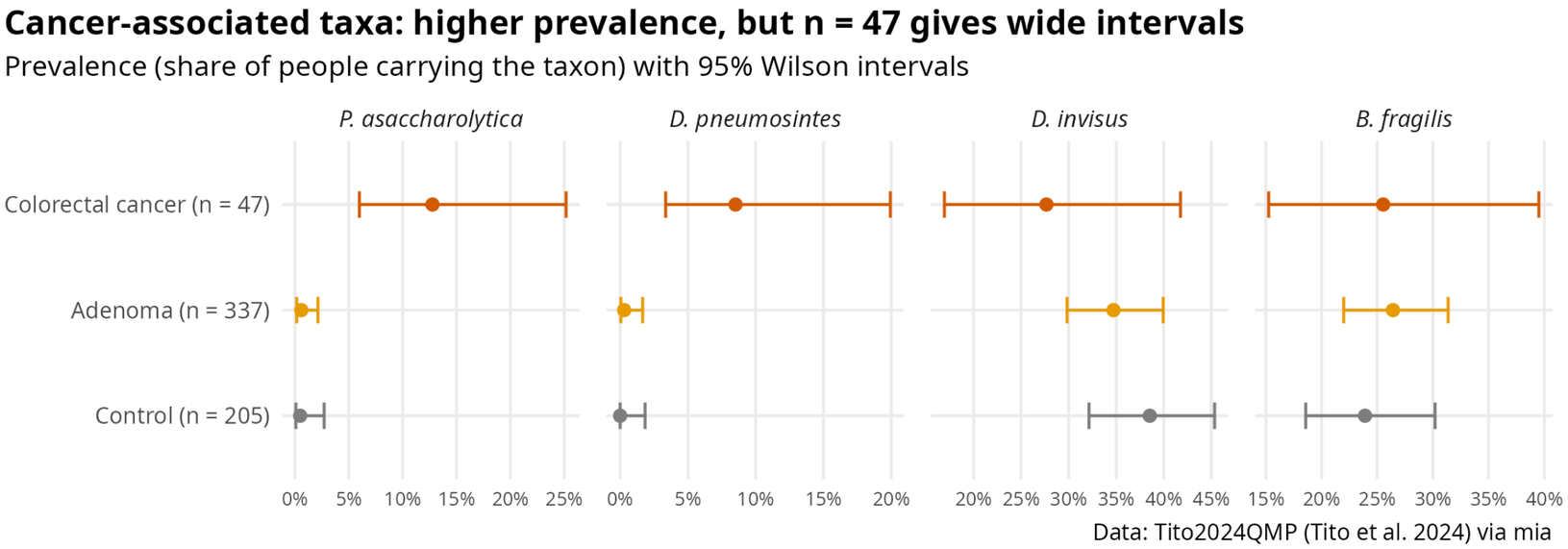
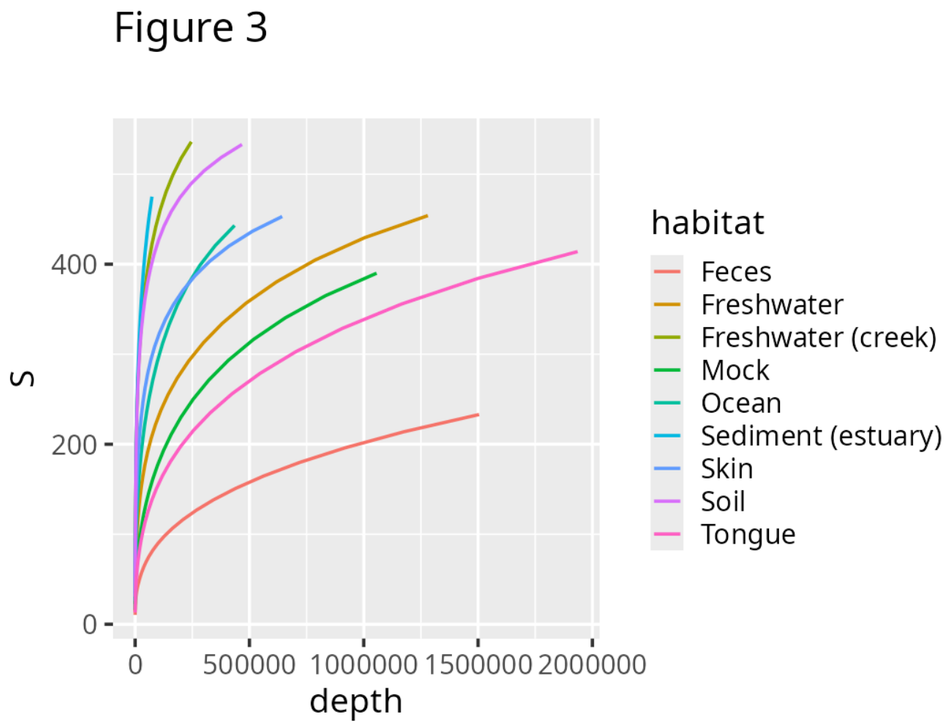
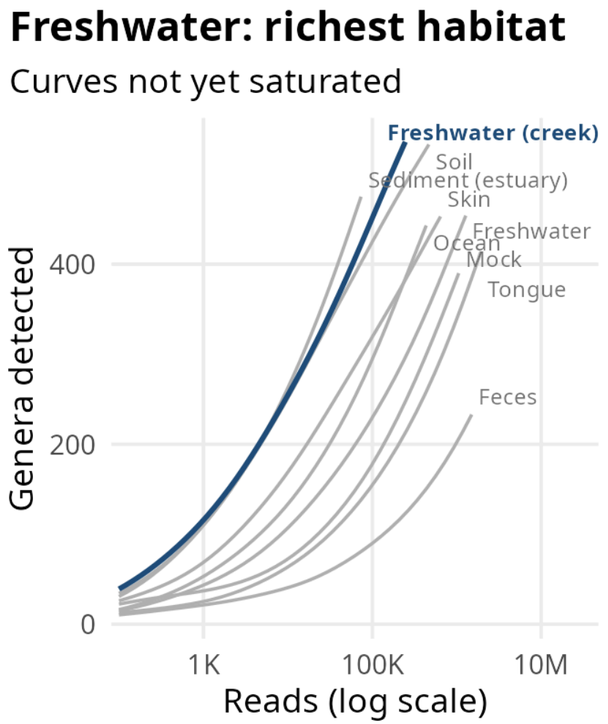


Figure 1. Taxa linked to colorectal cancer are more prevalent in patients, with wide uncertainty. Points: share of people carrying each taxon; bars: 95% Wilson intervals; group sizes in the labels. Data: 589 Belgian adults (Tito et al. 2024).

Rougier, Droettboom, and Bourne (2014) (rule 4); data: Tito et al. (2024)

Rule 5: Do not trust the defaults



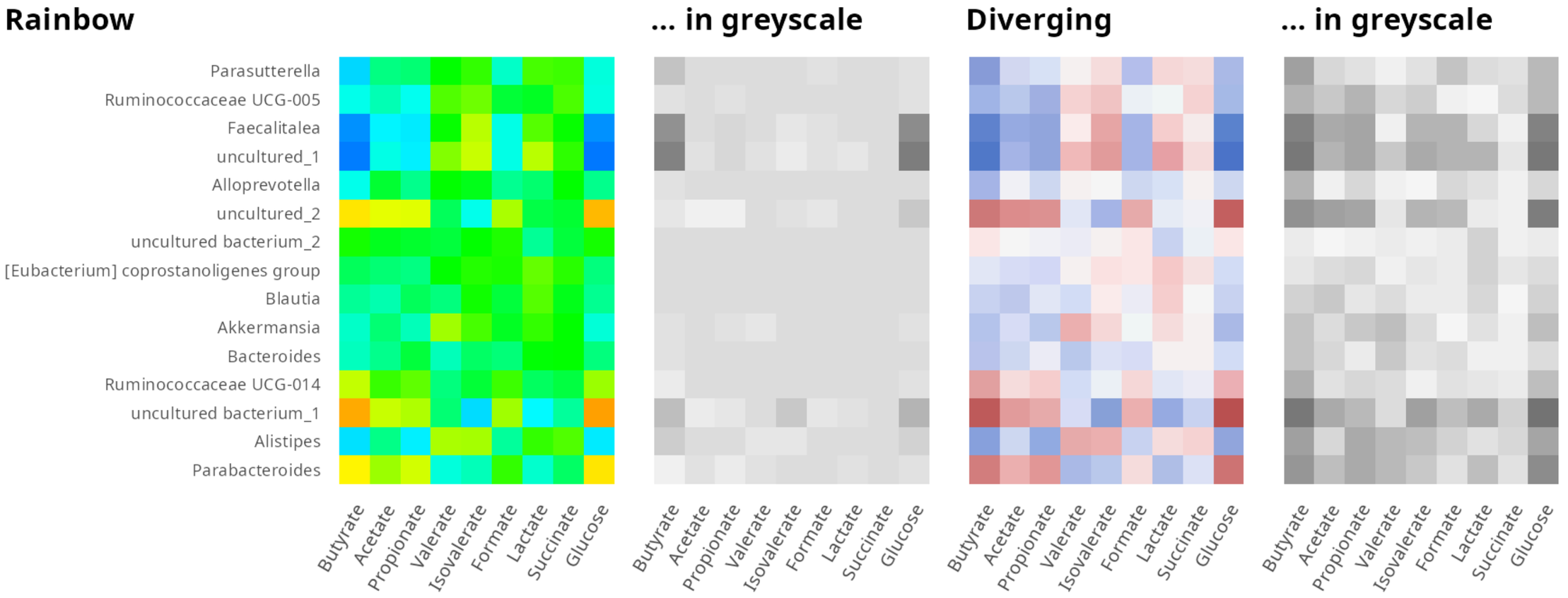


Data: GlobalPatterns (Caporaso et al. 2011) via mia

Rougier, Droettboom, and Bourne (2014) (rule 5); data: Caporaso et al. (2011)

Rule 6: Use colour effectively

Genus–metabolite correlations: where is zero? Are +0.5 and −0.5 equally salient?

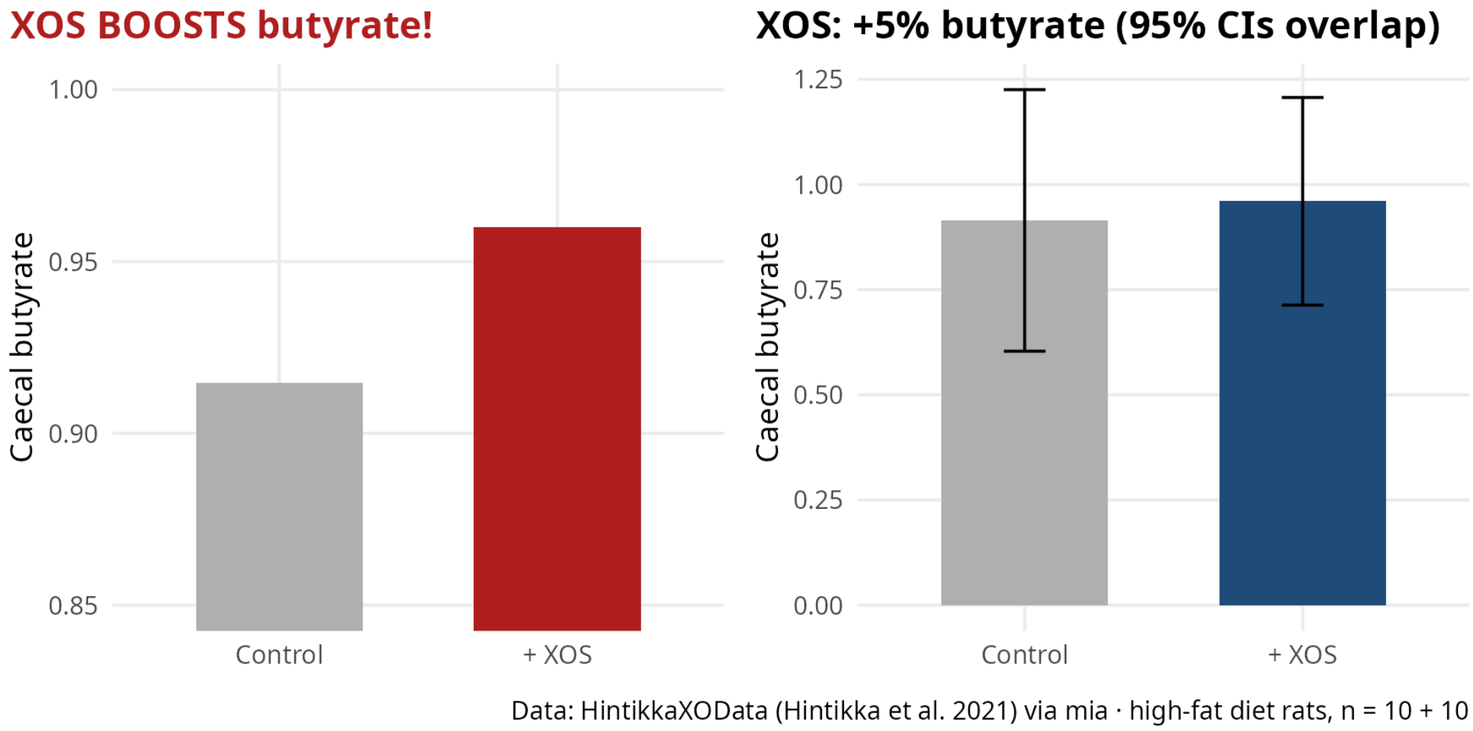


Data: HintikkaXOData (Hintikka et al. 2021) via mia

Use a palette that matches the data type: here a **diverging** scale for correlations, centred on zero.

Rougier, Droettboom, and Bourne (2014) (rule 6); Crameri, Shephard, and Heron (2020); data: Hintikka et al. (2021)

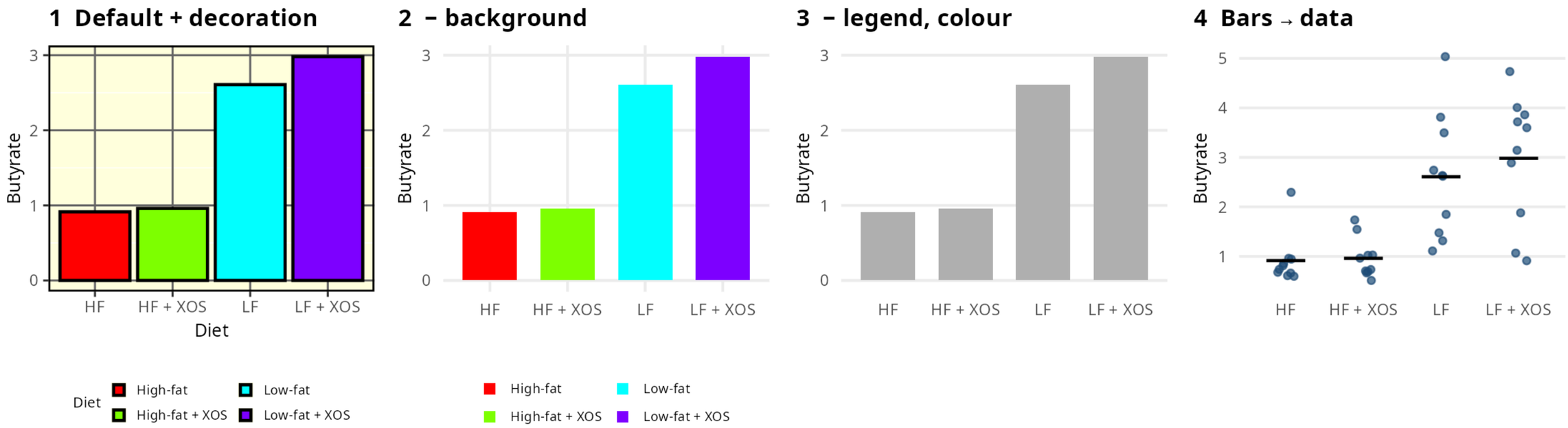
Rule 7: Do not mislead the reader



A truncated axis turns +5% into “BOOSTS”.

Rougier, Droettboom, and Bourne (2014) (rule 7); data: Hintikka et al. (2021)

Rule 8: Avoid chartjunk



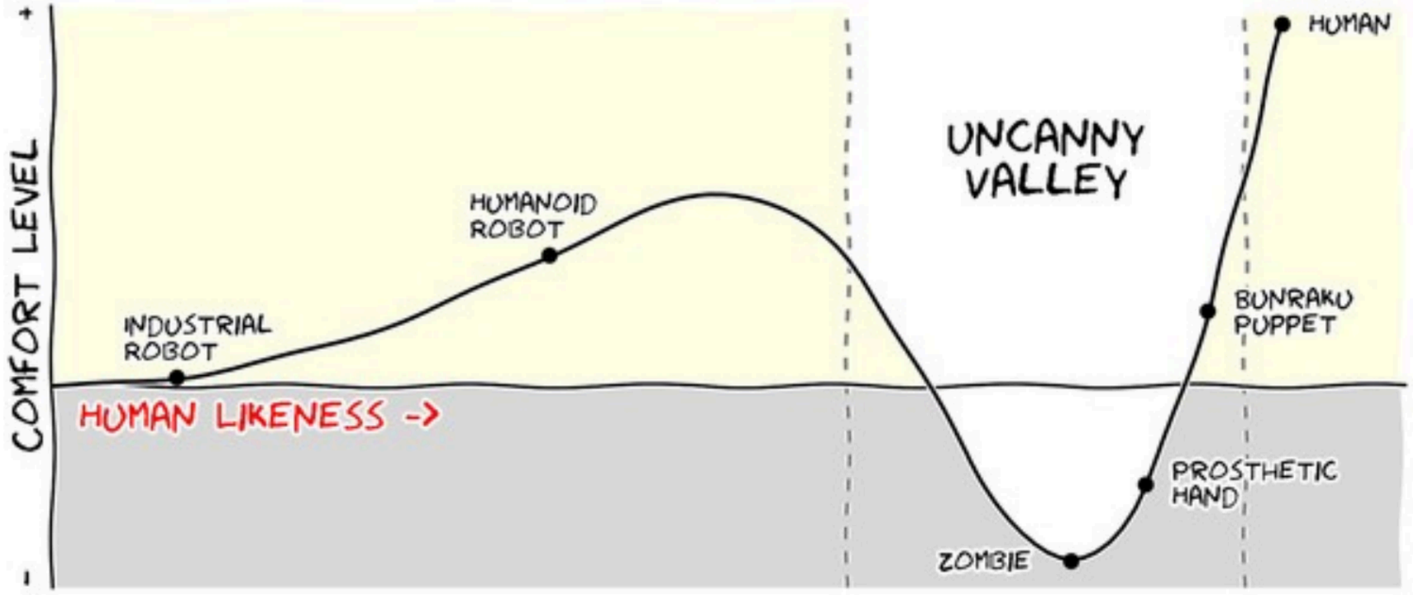
Data: HintikkaXOData (Hintikka et al. 2021) via mia

Rougier, Droettboom, and Bourne (2014) (rule 8); data: Hintikka et al. (2021)

Rule 9: Message trumps beauty

Message trumps beauty.

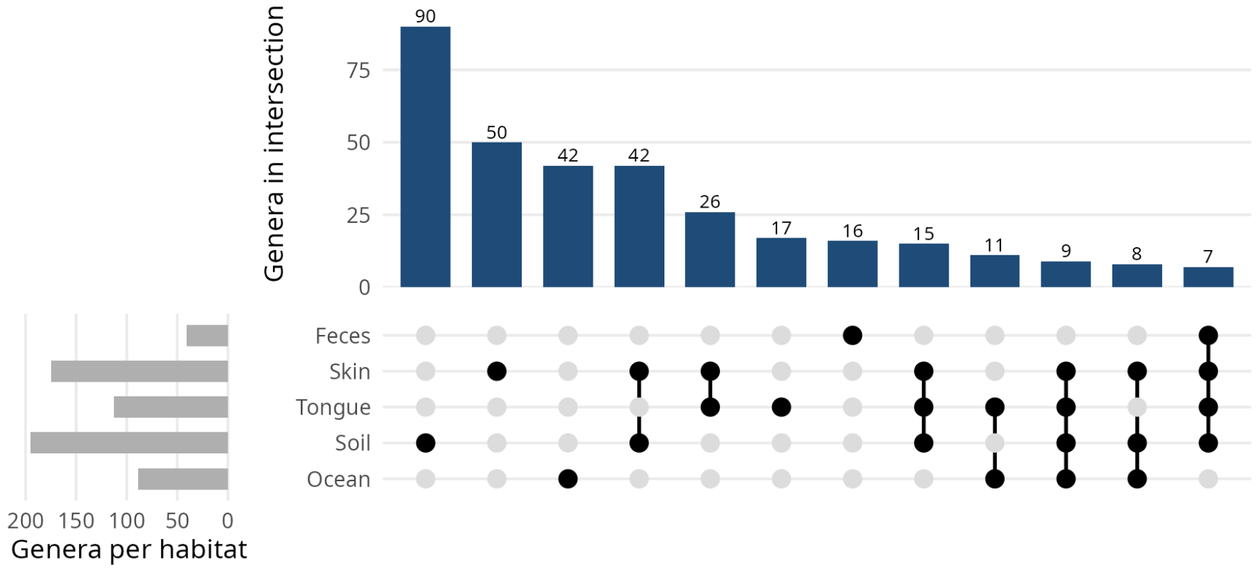
Ten Simple Rules for Better Figures
Nicolas P. Rougier, Michael Droettboom, Philip E. Bourne
Published: September 11, 2014 • https://doi.org/10.1371/journal.pcbi.1003833



Source: Rougier, Droettboom, and Bourne (2014) (rule 9); slide from Lahti (2022)

Rule 10: Get the right tool

Which bacterial genera are shared across habitats?



Data: GlobalPatterns (Caporaso et al. 2011) via mia; genus detected if >0.01% in ≥50% of samples

Tools fit the task

- A dedicated chart type (UpSet) beats a 5-set Venn diagram
- Domain tools (`mia`, `miaViz`) for omics data structures
- A vector editor (e.g. Inkscape) only for final touches, never for data

Rougier, Droettboom, and Bourne (2014) (rule 10); Lex et al. (2014); data: Caporaso et al. (2011)

Break · 15 min

Part II: Communicating research

Communicating research data

HoloGen Retreat 4, Leuven · October 2026

Leo Lahti (Department of Computing, University of Turku, Finland)

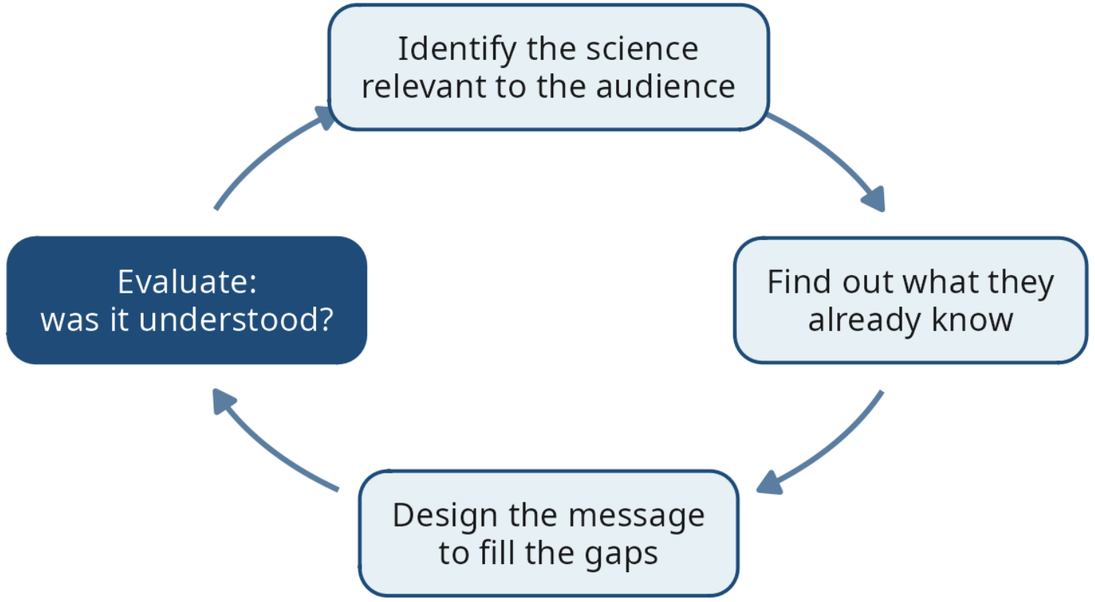
datascience.utu.fi | @antagomir



Turun yliopisto
University of Turku

Know your audience

- **Who** is the reader? Peer, reviewer, policymaker, patient, the public?
- **What** do they already know and need to decide?
- **Where** will they see it? Paper, poster, slide, web?
- Communication is designed **and evaluated**, like any intervention



Sources: Fischhoff (2013); Rougier, Droettboom, and Bourne (2014) (rules 1–3)

Inform, not persuade

COMMENT | 18 November 2020

Five rules for evidence communication

Avoid unwarranted certainty, neat narratives and partisan presentation; strive to inform, not persuade.

By Michael Blastland, Alexandra L. J. Freeman ✉, Sander van der Linden, Theresa M. Marteau & David Spiegelhalter ✉

Consider what information — in what format — would best support your audiences' decisions.



Screenshots, photo and caption: Blastland et al. (2020) (Credit: Matthew Horwood/Getty)

Five rules for evidence communication

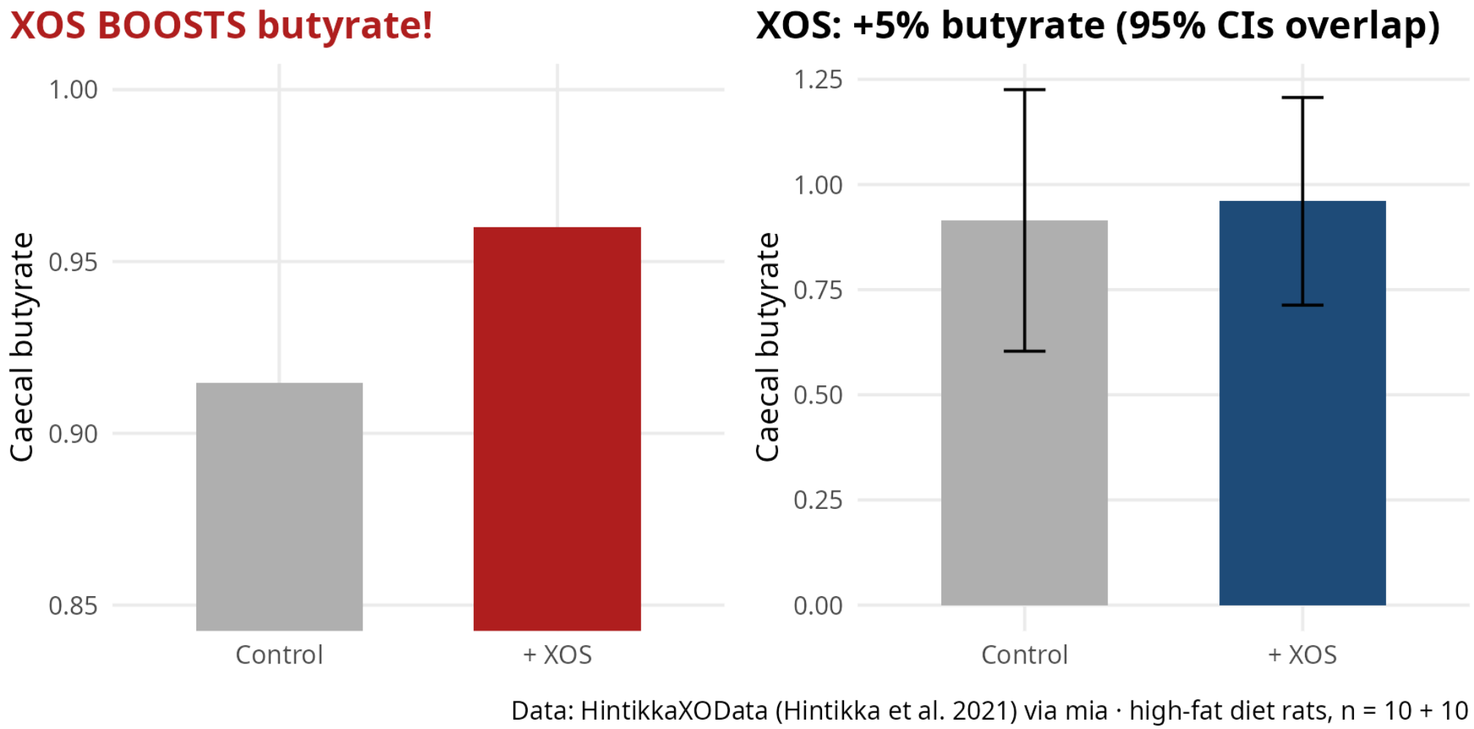
Blastland, Freeman, van der Linden, Marteau & Spiegelhalter (2020), *Nature*

1. **Inform**, not persuade
2. Offer **balance**, not *false* balance
3. Disclose **uncertainties**
4. State **evidence quality**
5. **Inoculate** against misinformation

One slide per rule follows, with our own data.

Source: Blastland et al. (2020); see also Fischhoff (2013); Bles et al. (2019)

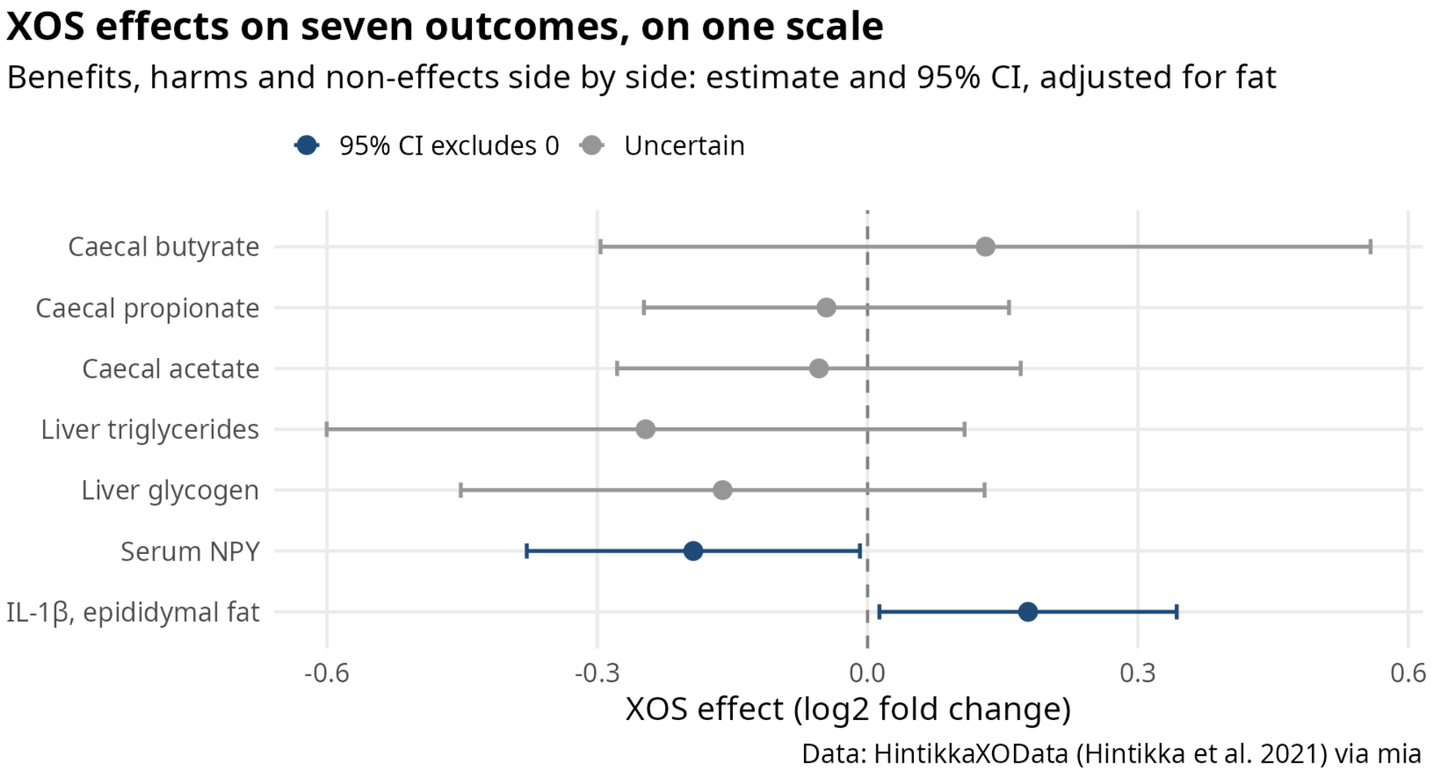
Evidence rule 1: Inform, not persuade



Aim to help people **decide**, not to change what they believe or do. Persuasive framing is easy to spot, and it costs trust.

Blastland et al. (2020) (rule 1); data: Hintikka et al. (2021)

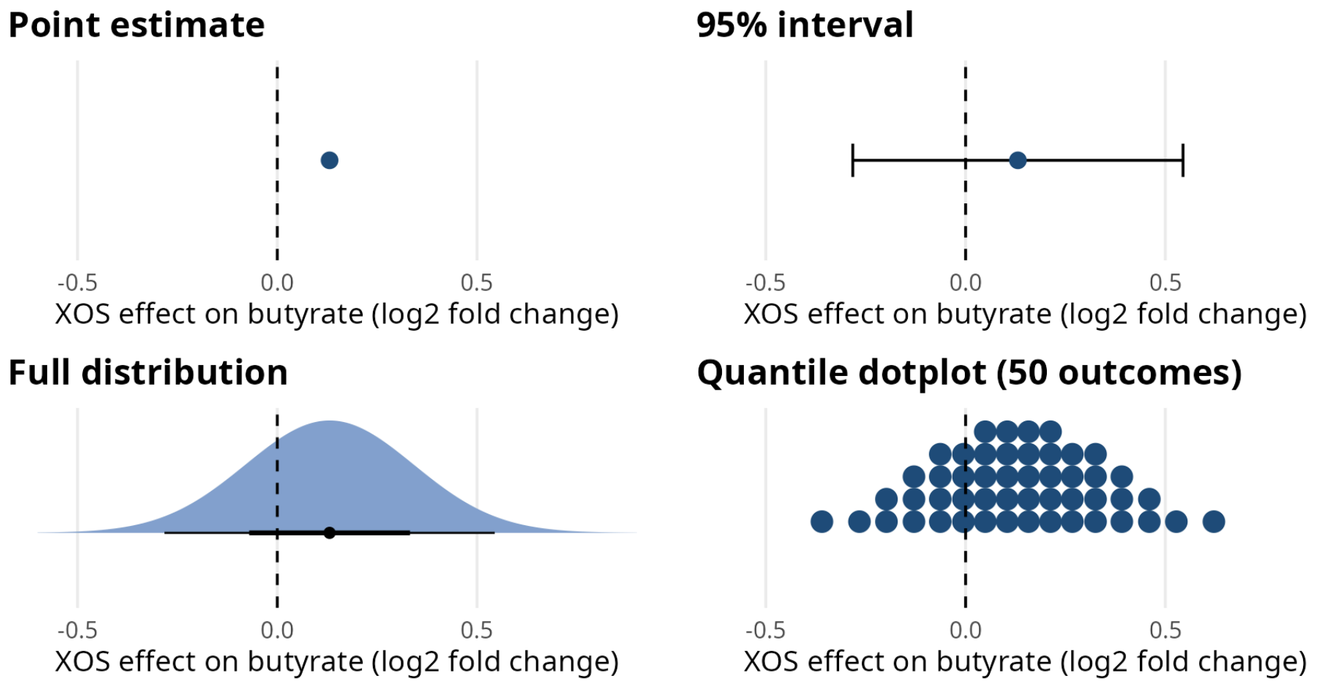
Evidence rule 2: Offer balance, not false balance



Show possible benefits **and** harms together, in the **same format**; do not cherry-pick the outcome that moved. *False balance*: giving fringe views equal weight to the evidence. (Seven outcomes, no multiplicity correction: exploratory.)

Blastland et al. (2020) (rule 2); data: Hintikka et al. (2021)

Evidence rule 3: Disclose uncertainties



Data: HintikkaXOData (Hintikka et al. 2021) via mia · linear model adjusted for fat

Say what is **not** known and how large the range is: in words, numbers or graphics. A point estimate alone hides the uncertainty.

Blastland et al. (2020) (rule 3); Bles et al. (2019); data: Hintikka et al. (2021)

Evidence rule 4: State evidence quality

Claim: *“A low-fat diet raises caecal butyrate 2.8-fold.”* How strong is the evidence?

Design

Controlled diet experiment, one study

Size

n = 10 rats per group; 95% CI 1.8–4.3

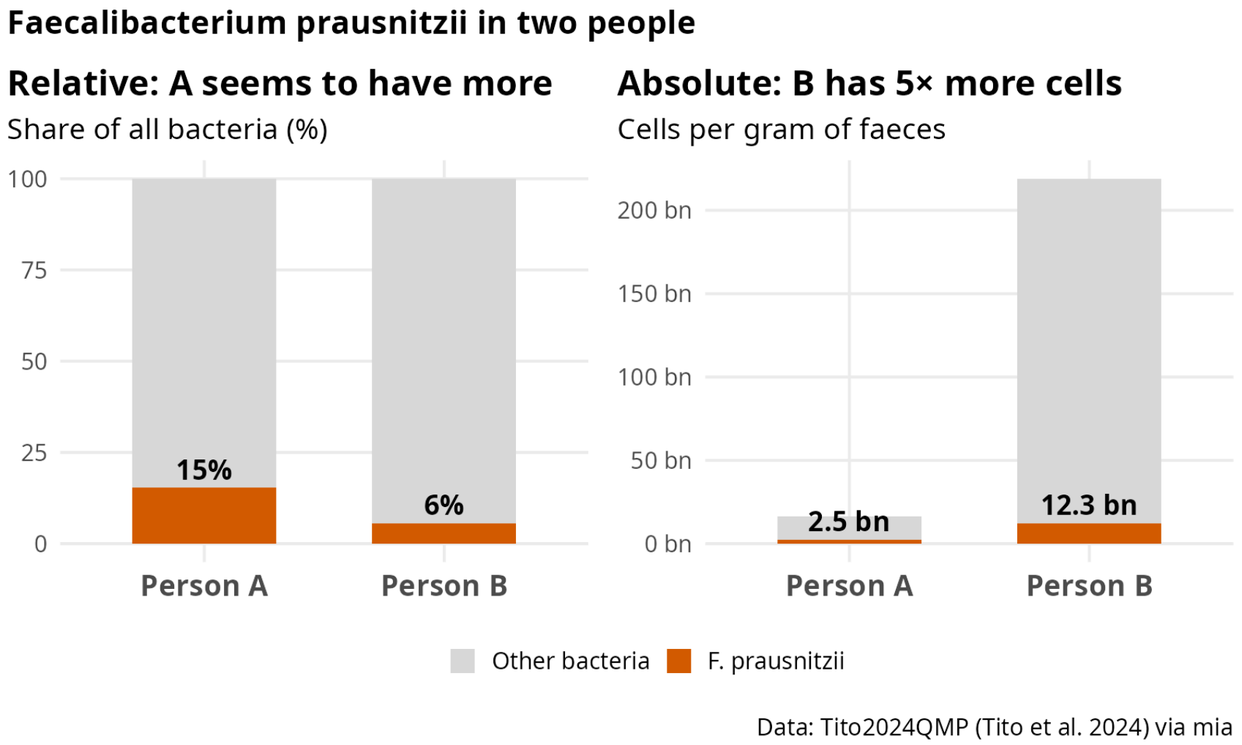
Applicability

Rats, not humans; one laboratory; not replicated

Say it: a clear effect *in these rats*, from one small study; **unknown** for humans. Frameworks such as GRADE rate evidence quality explicitly (Guyatt et al. 2008).

Blastland et al. (2020) (rule 4); Guyatt et al. (2008); data: Hintikka et al. (2021)

Evidence rule 5: Inoculate against misinformation



Data: Tito2024QMP (Tito et al. 2024) via mia

Prebunk the likely misreading

Misreading: *"Person B has less of this beneficial bacterium."*

Warn in advance: *"Relative abundance is a share. B has a lower share (6% vs 15%) but 5× more cells, because B carries more bacteria in total."*

Anticipate how a result can be misread, and say it **before** others do.

Blastland et al. (2020) (rule 5); Linden et al. (2017); data: Tito et al. (2024)

Principles for research communication

A research result is communicated through **what is said** and **who says it**

What is said

Correctness: facts conveyed accurately
Relevance: what the audience needs; no needless alarm
Uncertainty: limitations, open questions
Scientific status: peer-reviewed? consensus?

Who says it, and why

Authorship: own research or others'?
Perspective: expertise or opinion?
Transparency: funding, interests, partners

Cf. research integrity codes: ALLEA (2023); Finnish National Board on Research Integrity TENK (2023)

Source: Lund University (n.d.)

Principle 1: Correctness

Content “must be correct”: the audience should understand the results, and the research behind them, correctly.

Incorrect

“Gut bacteria cause colorectal cancer”

Correct

“Some gut bacteria are more common in people with colorectal cancer”

An association in one cohort is not a cause. Check the claim against the study design before it leaves the lab.

Lund University (n.d.); example: Tito et al. (2024)

Principle 2: Relevance

Include the facts needed to avoid misreading or overstatement. “Avoid communicating alarming messages unnecessarily.”

Alarming

“Cancer bacterium found in patients”

With the relevant facts

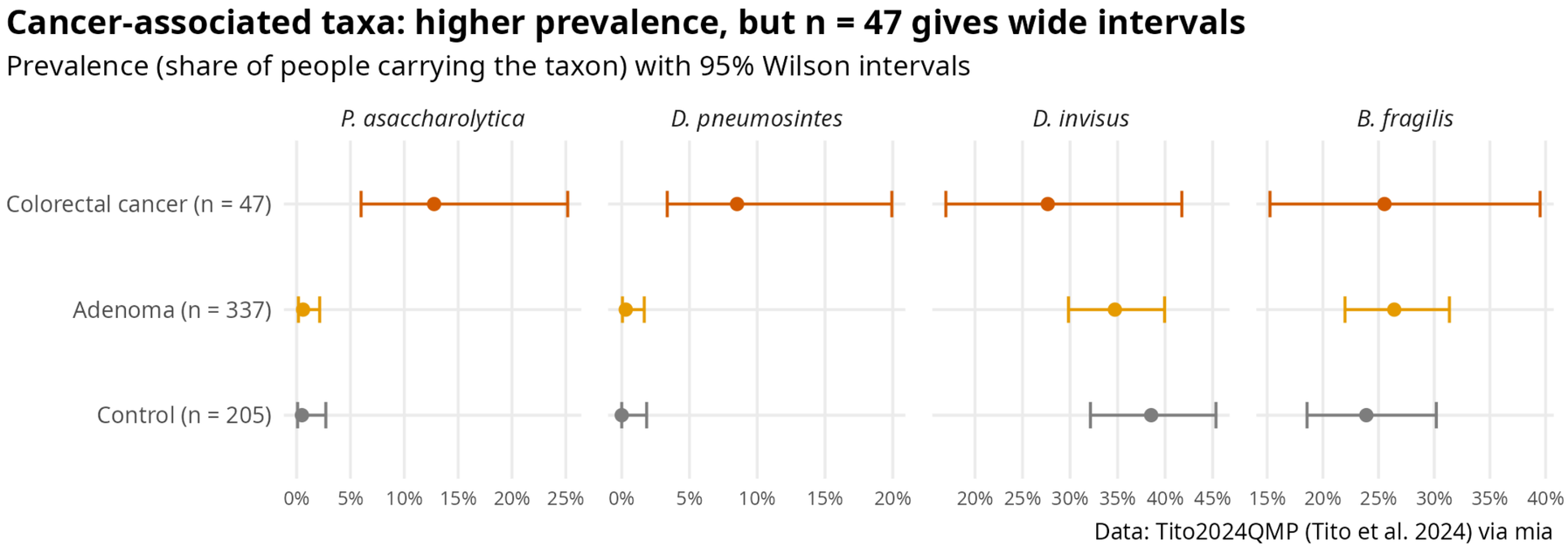
Porphyromonas asaccharolytica in **13%** of colorectal cancer patients (6 of 47) and **0.5%** of controls (1 of 205): most patients do not carry it

Give the numbers people need to judge the size and meaning of a result.

Lund University (n.d.); data: Tito et al. (2024)

Principle 3: Uncertainty

“Is there any uncertainty?” State the limitations, weaknesses and uncertainties of the research and its interpretation.



With 47 patients, the intervals are wide: say so in the text, and show it in the figure.

Lund University (n.d.); data: Tito et al. (2024)

Principle 4: Scientific status

Is the result broadly supported, or does it deviate from consensus? Is it peer-reviewed, or preliminary?

Preliminary

Hypotheses, preprints, a reanalysis on a slide, a single small study

Established

Peer-reviewed and replicated; consistent with other studies and reviews

In this lecture: the datasets are peer-reviewed publications; the figures are **exploratory reanalyses** made for teaching.

Lund University (n.d.); data: Hintikka et al. (2021); Tito et al. (2024); Caporaso et al. (2011)

Principle 5: Authorship

Is this the researcher's own research, or someone else's?

Others' work

Published data and results by other groups: name them and cite them

Own work

Own analyses, figures and interpretation: say that they are yours

Every example in this lecture names whose data it shows and who made the figure.

Lund University (n.d.)

Principle 6: Perspective

Is the topic within the researcher's expertise, or is it other knowledge or a personal opinion?

Outside expertise

"As a data scientist, I think this diet will prevent cancer"

Within expertise

"The statistical association is clear in this cohort; whether it is causal needs clinical studies"

Mark the switch from expert knowledge to personal view explicitly.

Lund University (n.d.)

Principle 7: Transparency

“State any vested interests!” Funding, company interests and collaboration partners.

Disclose

Funding · company interests · collaboration partners · data and code availability

Example statement

“Funded by [funder]. The authors declare no competing interests. Data and code: [repository].”

Disclosure lets the audience judge the result on its merits. Integrity codes require it (ALLEA 2023; Finnish National Board on Research Integrity TENK 2023).

Lund University (n.d.); ALLEA (2023); Finnish National Board on Research Integrity TENK (2023)

Group discussion (20 min)

In groups of 3–4 (10 min):

1. Pick a recent result **from your own work**.
2. Write the **one-sentence message** for (a) a peer and (b) a non-expert.
3. Which of the **seven principles** is **hardest** to satisfy for this result? Why?
4. What is the main **uncertainty**, and how would you communicate it?

Report back (10 min): one example per group.

Principles: Lund University (n.d.)

Seven principles

What is said: correctness · relevance · uncertainty · scientific status

Who says it: authorship · perspective · transparency

Lund University, *Principles for research communication*: https://www.staff.lu.se/research-and-education/research-support/communicating-your-research/principles-research-communication