UVA DS2003 Communicating with Data
Fall 2026

Science Visualizations
Telling a story you can trust

Sarah Groves
based on materials by Hannah Bako, Rafael Alvarado, & Kevin Janes

What you will learn today

- What makes a scientific audience different?
- How do you build a multi-panel figure that tells a story?
- What makes a figure trustworthy, and what makes it feel like spin?

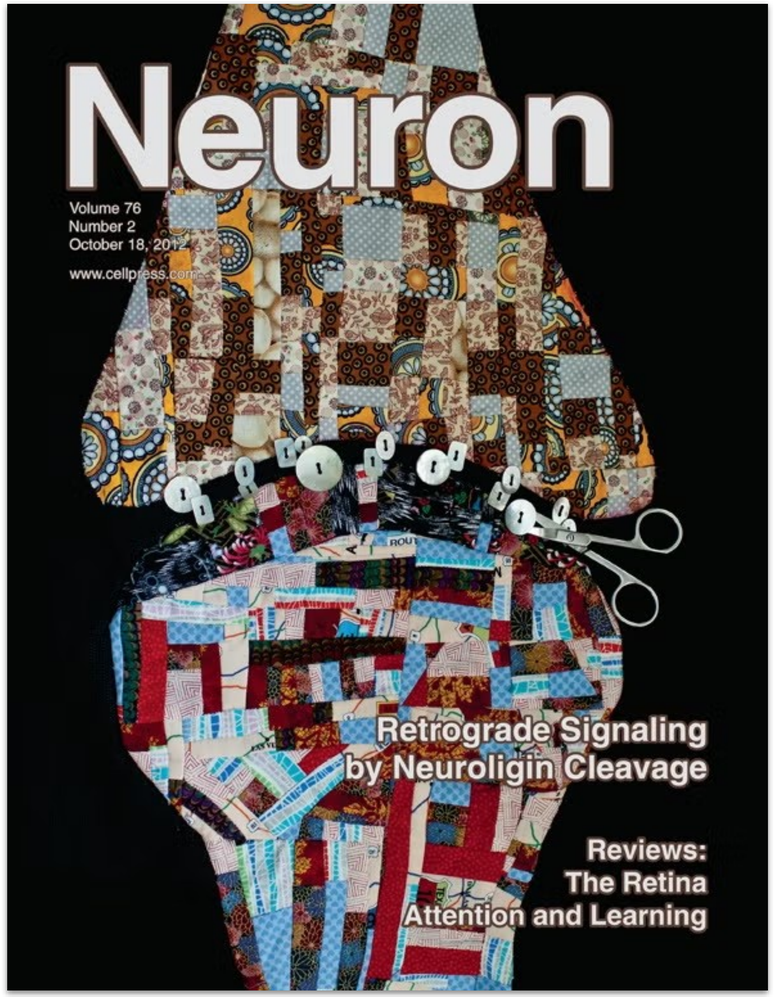
Your reader is a skeptic

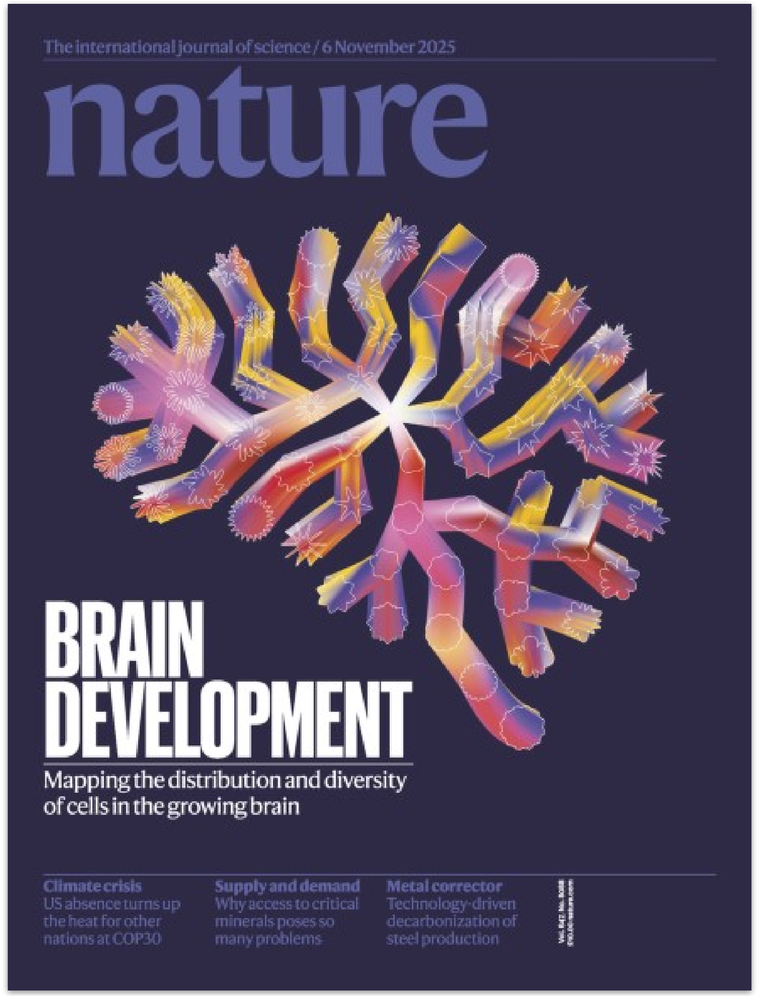
- Your job as author: show enough that they can check the conclusion themselves
  - Is it real? Could it be noise? Is this the whole picture?
- “The first principle is that you must not fool yourself—and you are the easiest person to fool.” – Richard Feynman, 1974

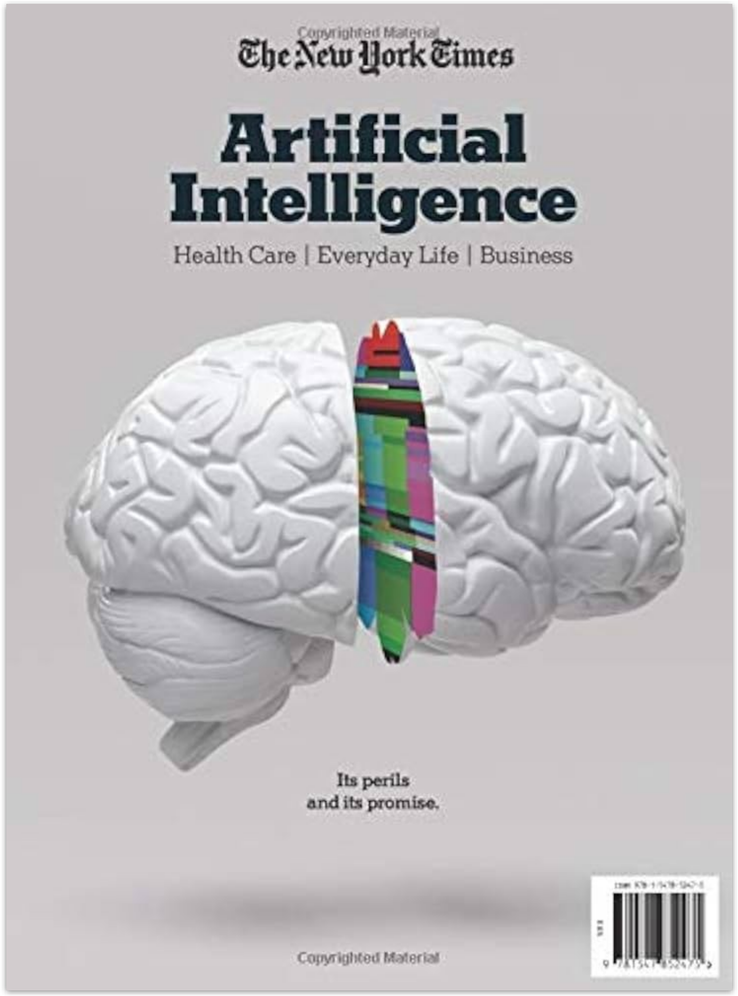


A lawyer vs. a scientist

- A lawyer shows the evidence that helps the client
- A scientist shows the evidence, including what doesn't fit
- What would each one leave out of a figure?
- Where does journalism fall? Where does your Project 2 fall?
- Ask of every panel: am I arguing a case or showing the evidence?







SPECIALIST

GENERAL

Scientific vs. information visualization

- Scientific visualization (SciVis): data with real spatial dimensions, e.g. maps, CT scans, molecules
- Information visualization (InfoVis): abstract data, e.g. charts, networks, statistical plots
- A “science visualization” for Project 2 can be either/both
  - For this class and Project 2, we are considering a scientific visualization to be a combination of these two

Design Principles for SciVis

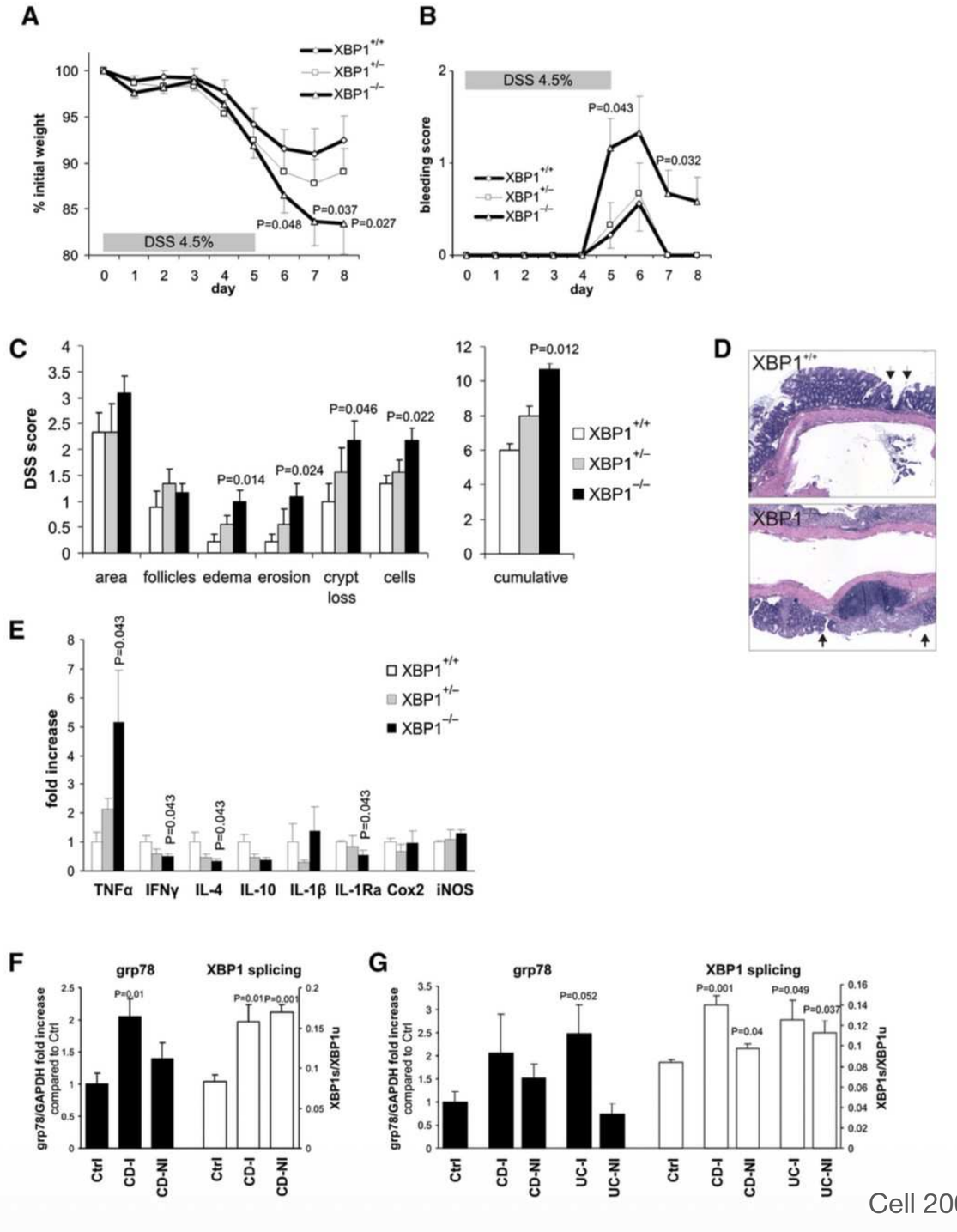
from Wong, 2011, *Nature Methods*
https://www.nature.com/articles/nmeth0211-101

"Points of View" Nature Methods series

Overall advice for storytelling in science visualizations

One figure, many mistakes

Getting rid of a protein called XBP1 exacerbates a kind of colitis in mice and correlates with endoplasmic reticulum (ER) stress markers in human inflammatory bowel disease (IBD)



Cell 2008

Unfold the story, don’t just dump data

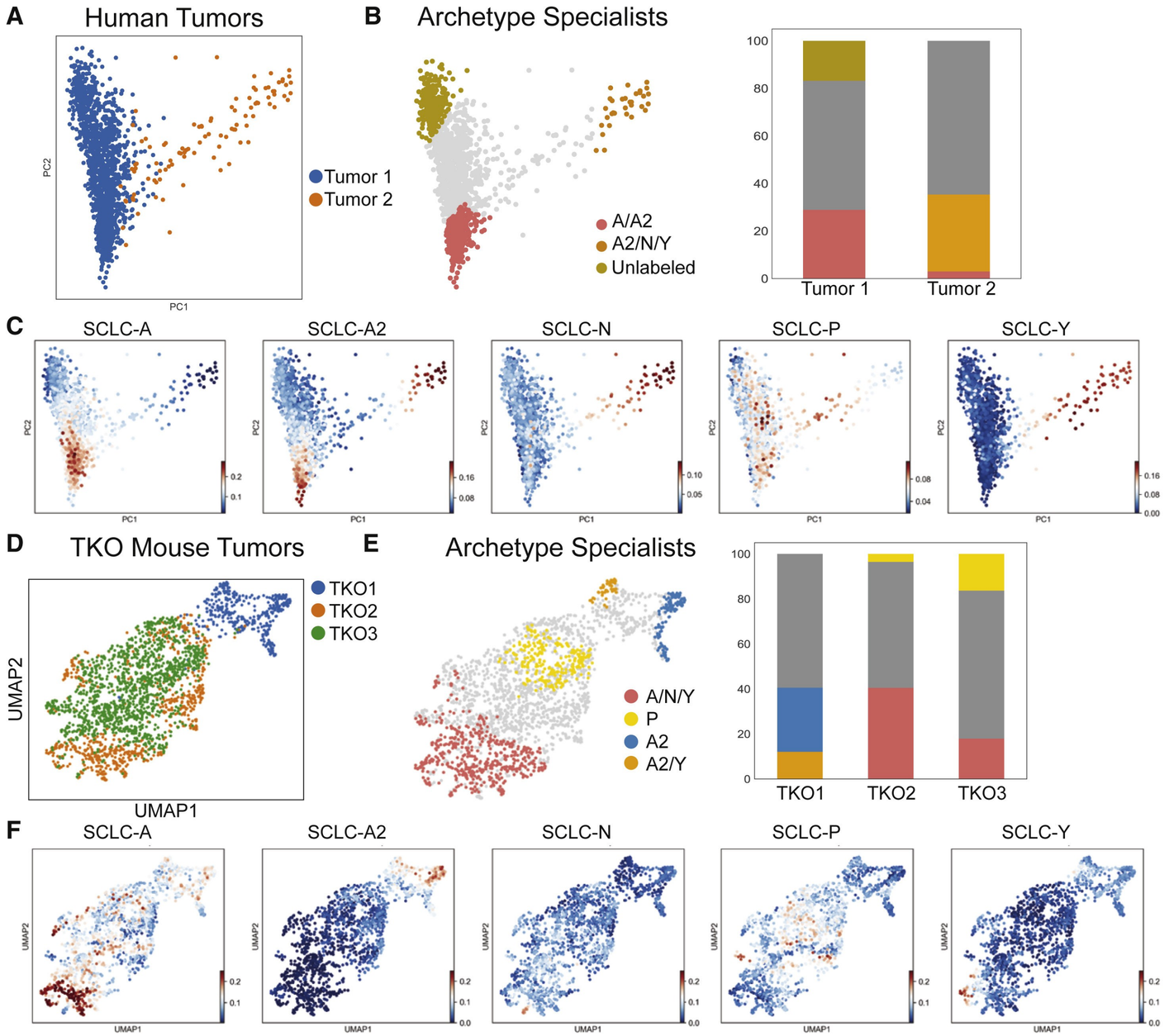
- Krzywinski & Cairo argue that multi-panel figures can follow a story arc:
  - introduction → question → conflict → buildup → resolution
- Each panel has one clear job
- Why is “I did X, then Y, then Z” not compelling?

Krzywinski, M. & Cairo, A. Storytelling. Nature Methods 10, 687 (2013)

“I did X, and then Y, and then Z…”

Background
Each dot is one cell; similar cells sit close together

- Same analysis on different datasets
- No new conclusions; the only point was to show I did it
- No logical flow from coloring by sample to coloring by “archetype” to archetype scoring

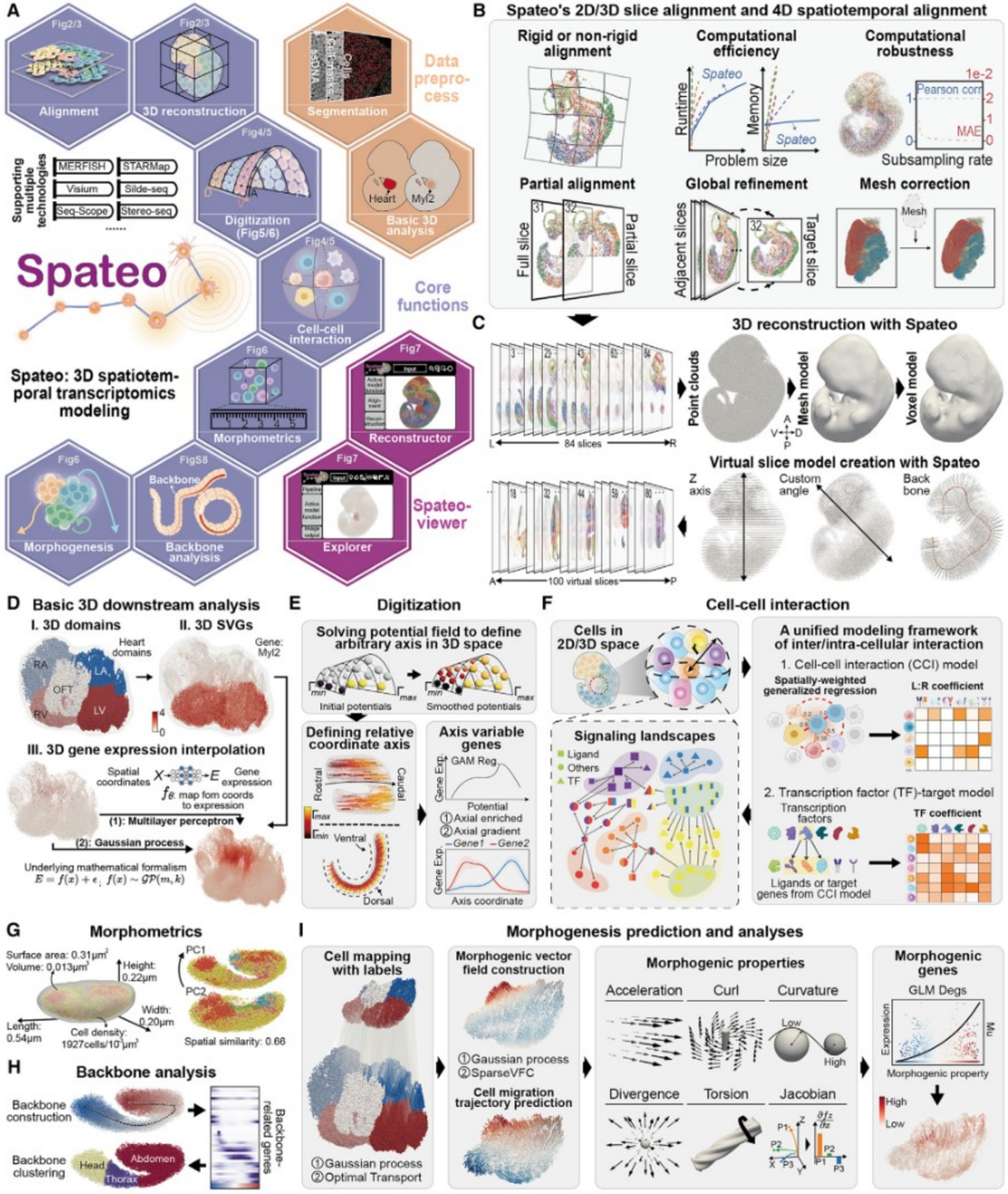


Groves et al. Cell Systems 2022

Poor information display in the modern era

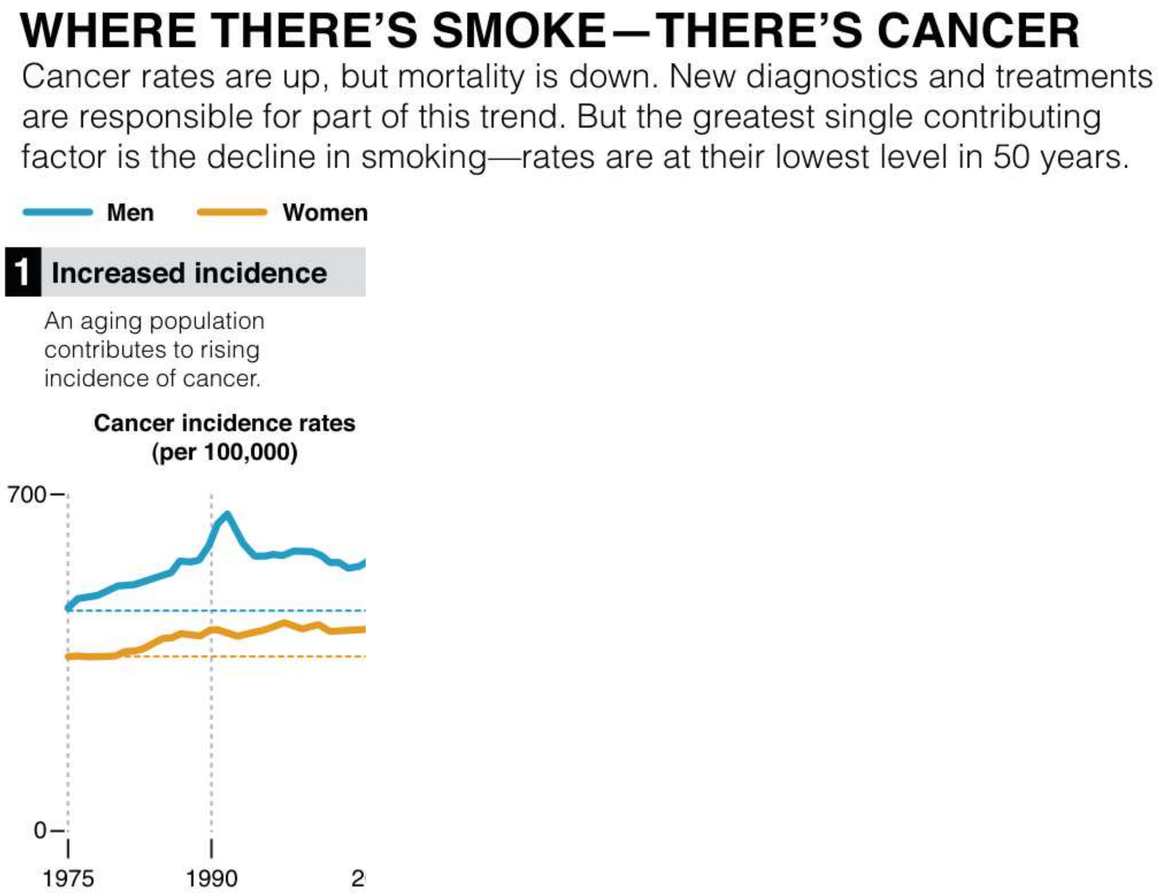
Where does your eye land first?
Can you state the finding?

“Data advertisement, not data display”



Cell (2025)

A figure that unfolds



Krzywinski, M. & Cairo, A. Storytelling. Nature Methods 10, 687 (2013), Fig. 2. Data: American Cancer Society; Monitoring the Future

Figures as montages

A: Two contrasting representative examples of simulated droplets

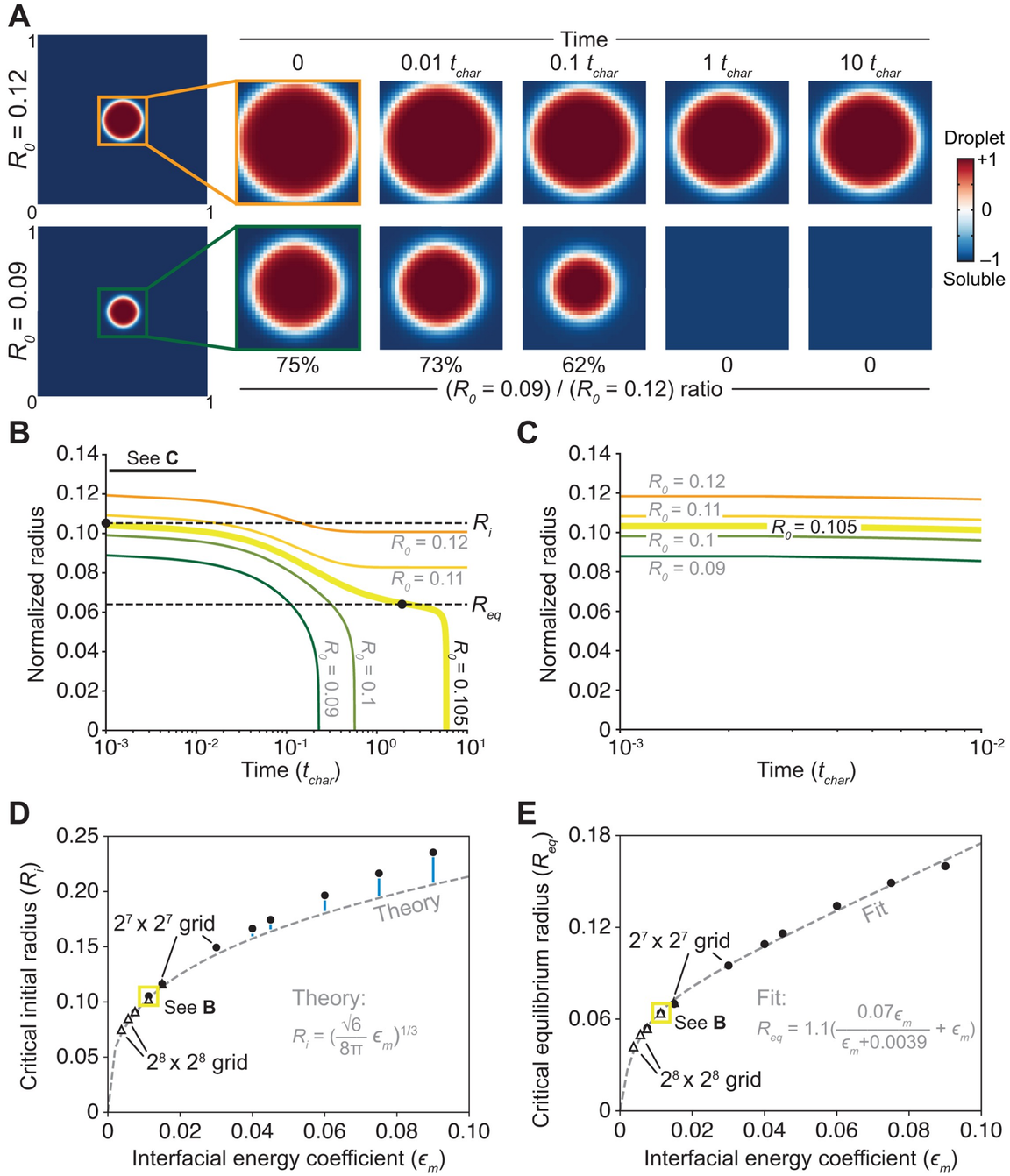
→ B: Plot summarizing evolution of radius over time, for many different starting radii

→ C: Zoomed in version of B (was important for our story, which was that you need a longer simulation, so you need an efficient method like the one we built)

→ D: Taking a point in B (“critical initial radius”) and showing how it changes based on a parameter in the model; highlight connects D back to B

→ E: same concept for “critical equilibrium radius”

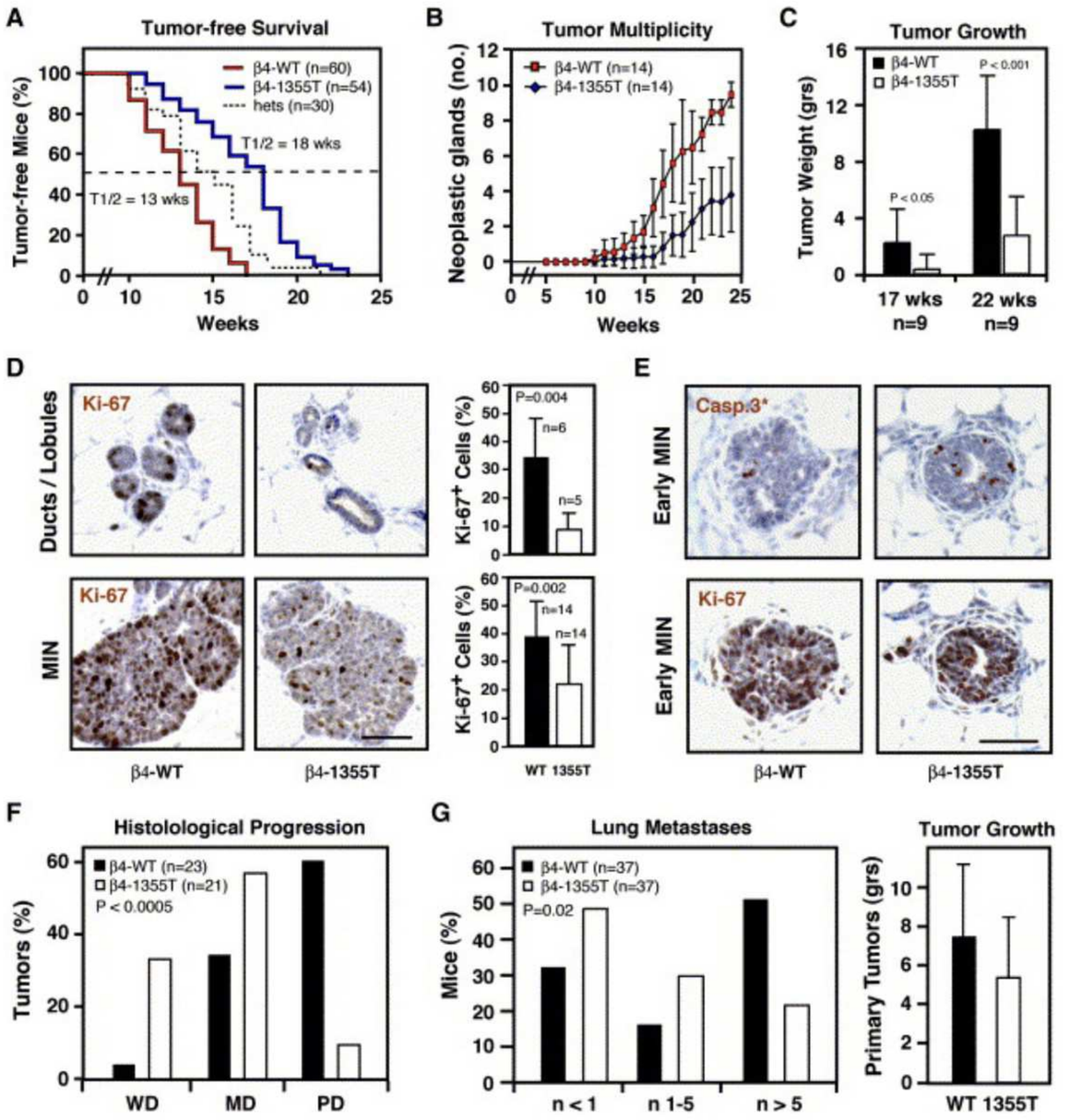
how big does a droplet have to be to survive?



Groves et al 2026

Figures as montages

Each piece is understandable on its own
What is each row's job?
What single sentence ties them together?



Guo et al., *Cell* **126**:489 (2008)

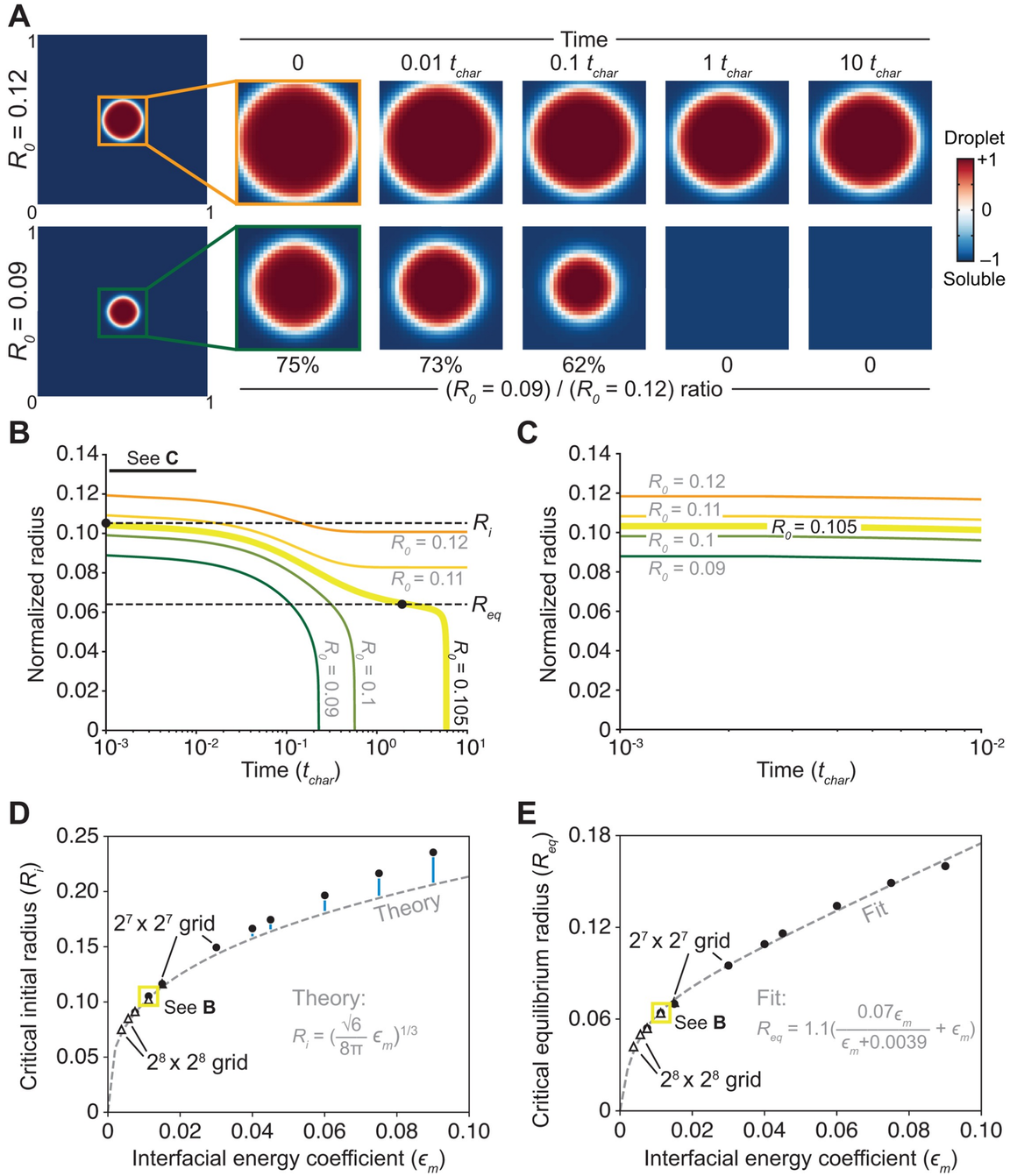
The case against storytelling

- Katz (2013) objected: stories want a hero, a goal, and a satisfying ending. Science often has none of these.
- What gets lost when you force results into a neat story?
- In a real paper, where should the “bad data” go?
- Takeaway: **structure** the evidence; don't spin it

Katz, Y. Against storytelling of scientific results. Nature Methods 10, 1045 (2013); Krzywinski, M. & Cairo, A. reply, Nature Methods 10, 1046 (2013)

This was not how the study proceeded; a lot of the work is “hidden” from this figure

- Which panel do you think I made first? Which last?
- What got left out, and where did it go?
- Would anything hidden change the conclusion?



Groves et al 2026

Think | Pair | Share

- Where is the line between telling a story and spinning one?
- Is it spin to leave out a panel?
- To show one example instead of all of them?
- To put the most convincing panel first?
- To pick the axis range?

Building the layout of figure panels in a science visualization

Building a multi-panel figure

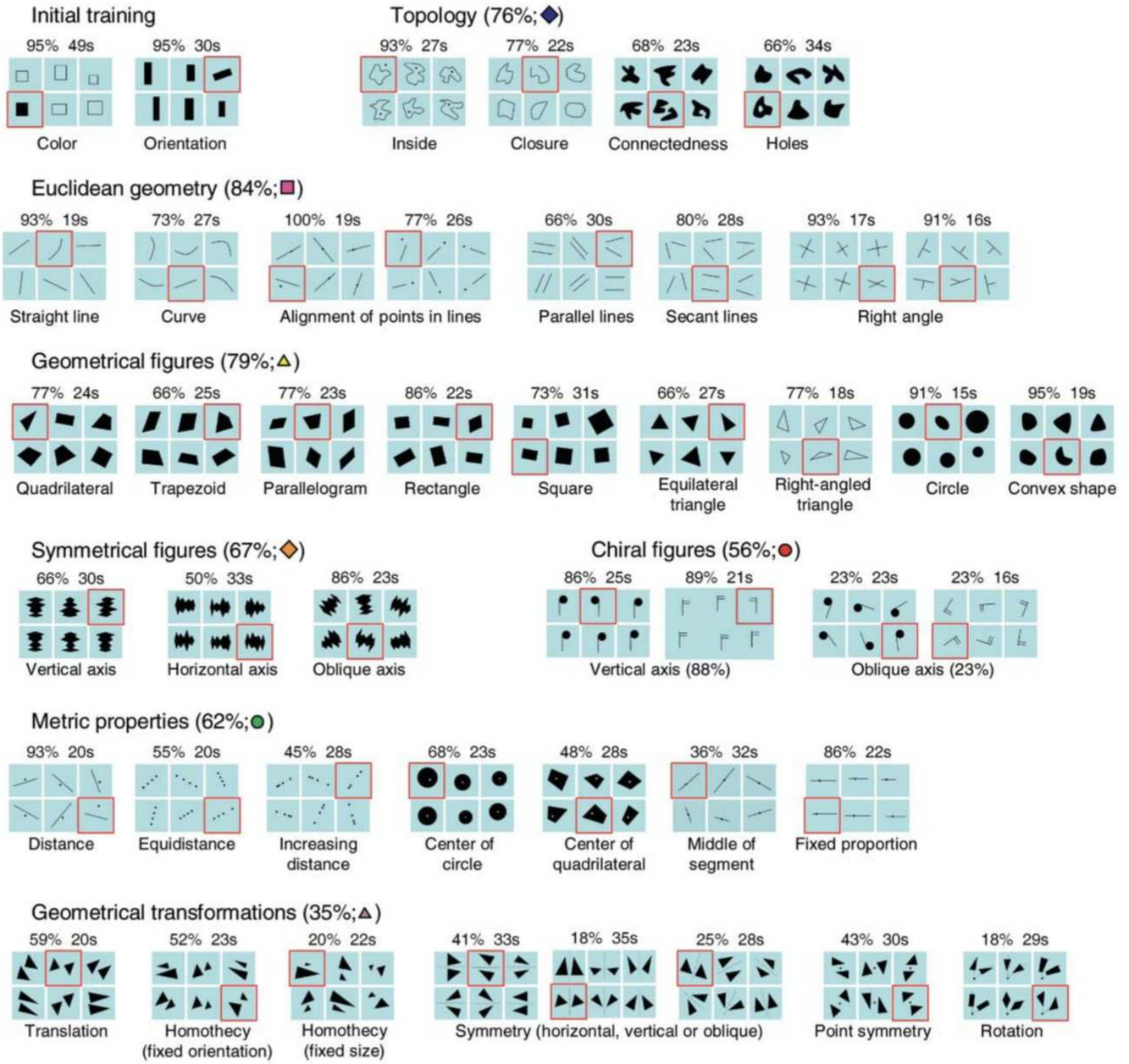
- A common arc for scientific evidence

context → one raw example → all the data → model → check → the claim

- Why start with one raw example instead of the model?
- ALWAYS SKETCH/PLAY WITH LAYOUT FIRST! Just like an infographic

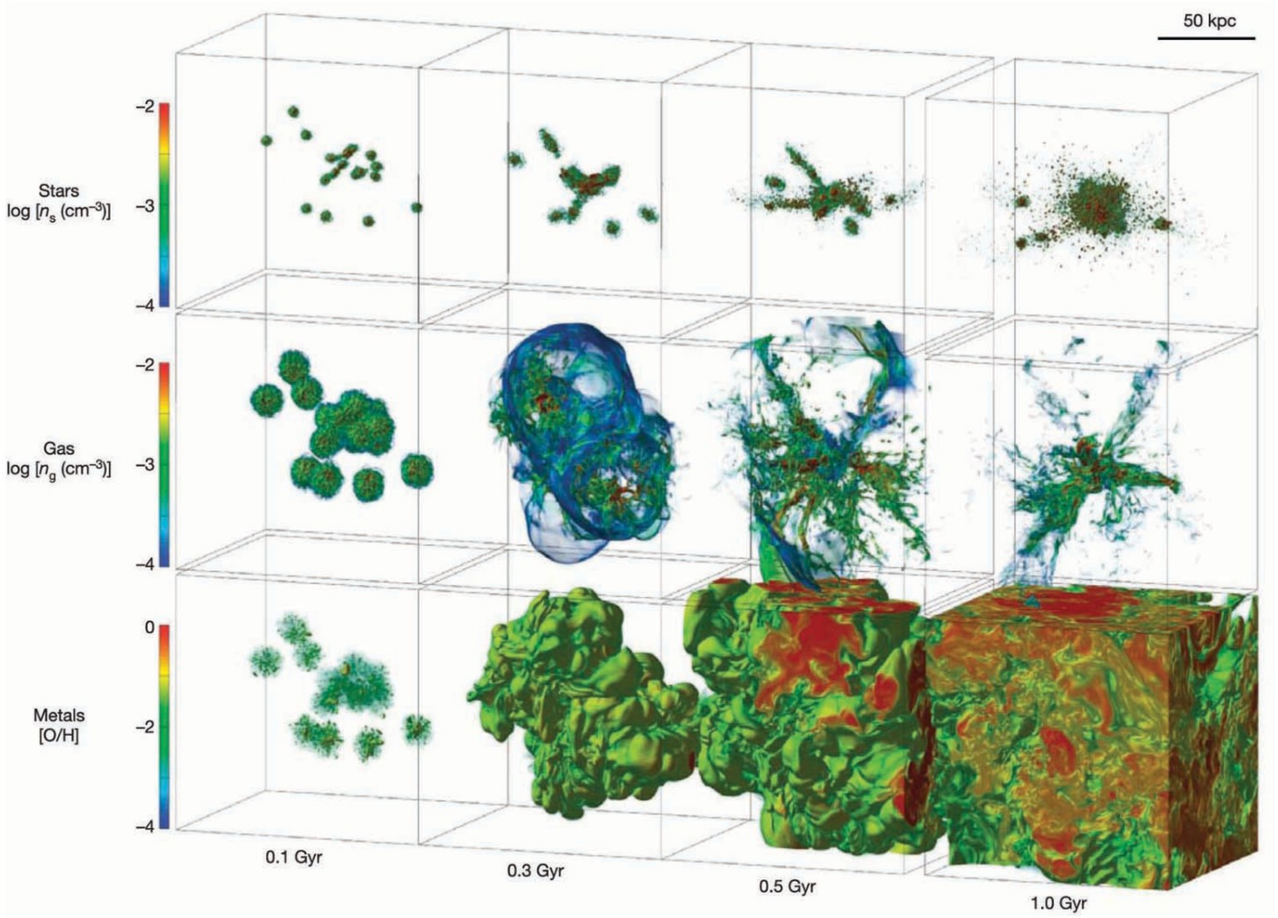
See also Wong, B. Points of View: Layout. Nature Methods 8, 783 (2011)

Small multiples as an effective information design tool



Dehaene et al., *Science* **311**:381 (2006)

Small multiples share one structure



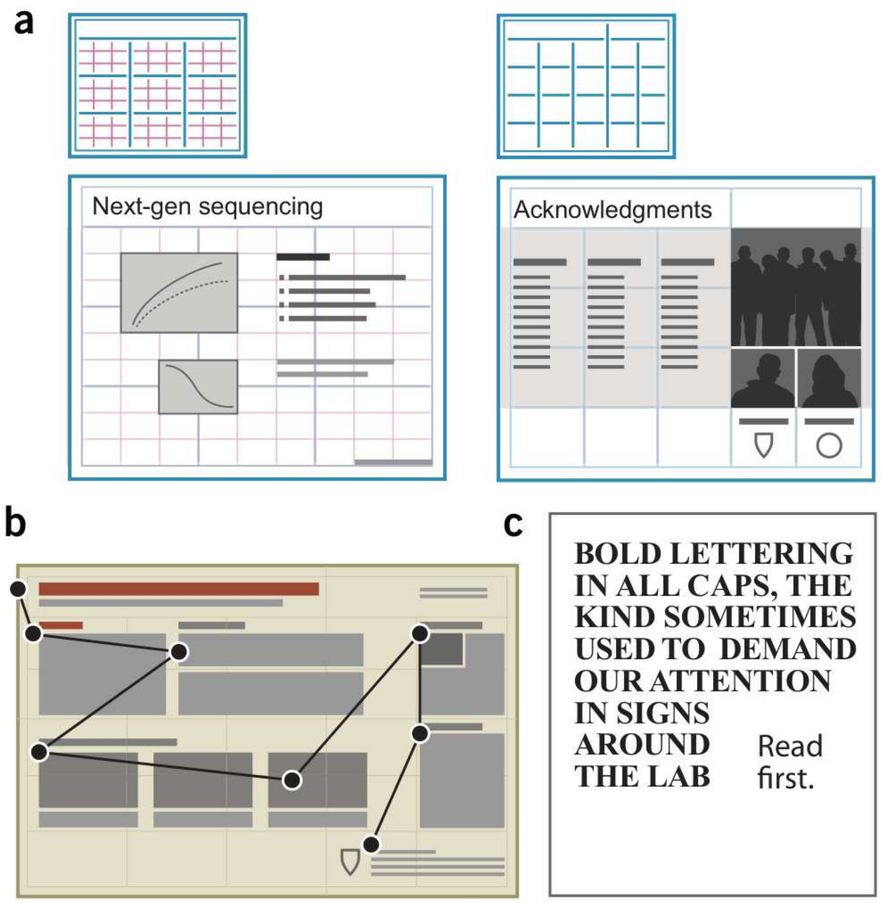
Mori et al., *Nature* **440**:644 (2006)

What changes from row to row?
Column to column?

Why keep the same frame in every panel?

From K. Janes, Technical information design (lecture), University of Virginia. Figure: Mori, M. & Umemura, M. Nature 440, 644 (2006)

Layout sets the reading order



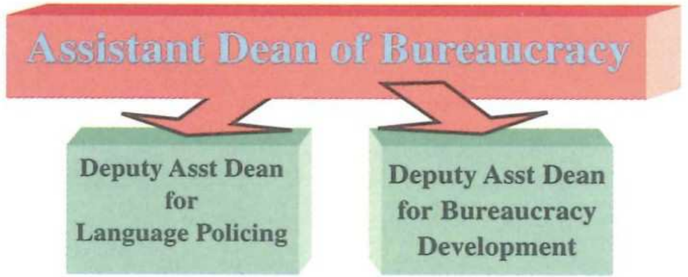
(a) A grid keeps panels aligned and comparable

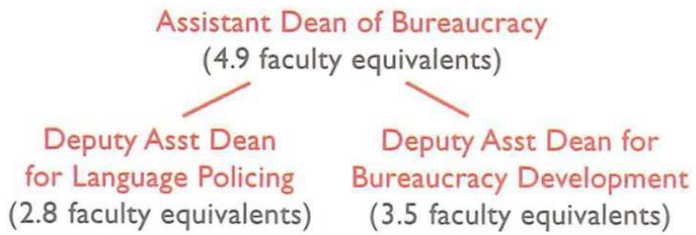
(b) Readers follow a path: put panels in the order they should be read

(c) Contrast and white space decide what is read first

Wong, B. Points of View: Layout. Nature Methods 8, 783 (2011), Fig. 2

Negative space





Tufte, *Beautiful Evidence* (2006)

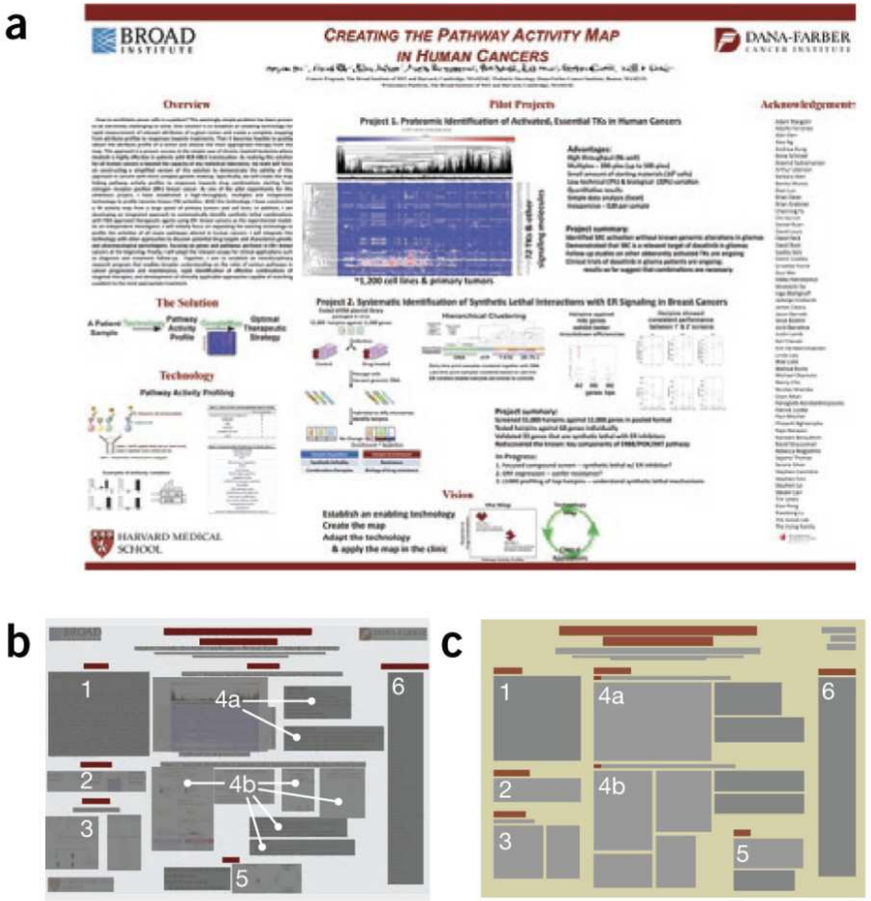
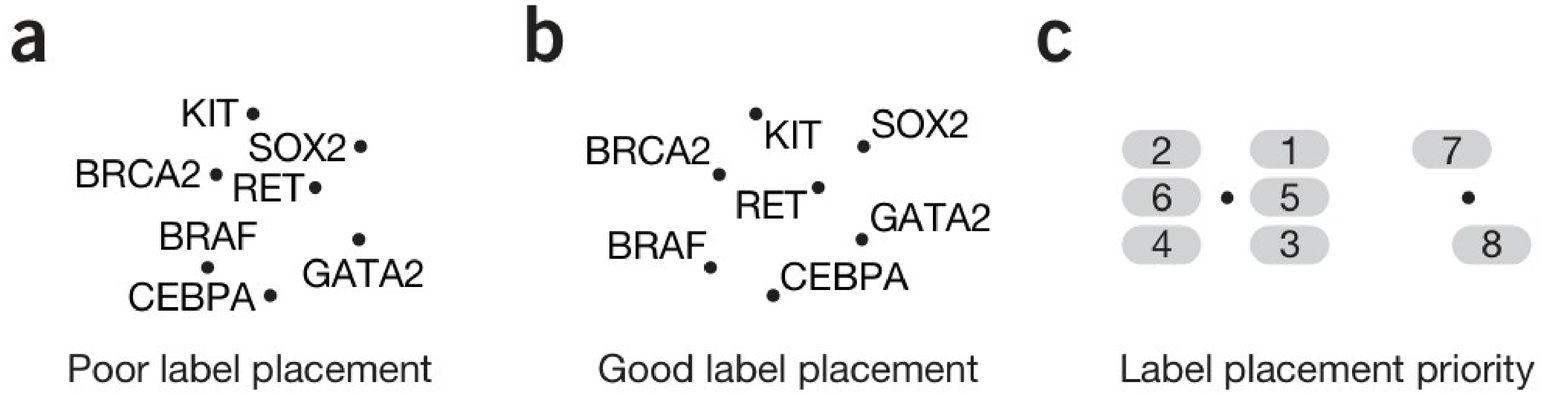


Figure 2 | Whitespace can be used to structure content. (**a**) An example of a scientific poster. (**b**) A space study reveals that contents in sections 1–6 are scattered and whitespace is fragmented. (**c**) An example of consolidated whitespace organizing contents.

Wong, *Nat Methods* **8**:1 (2011)

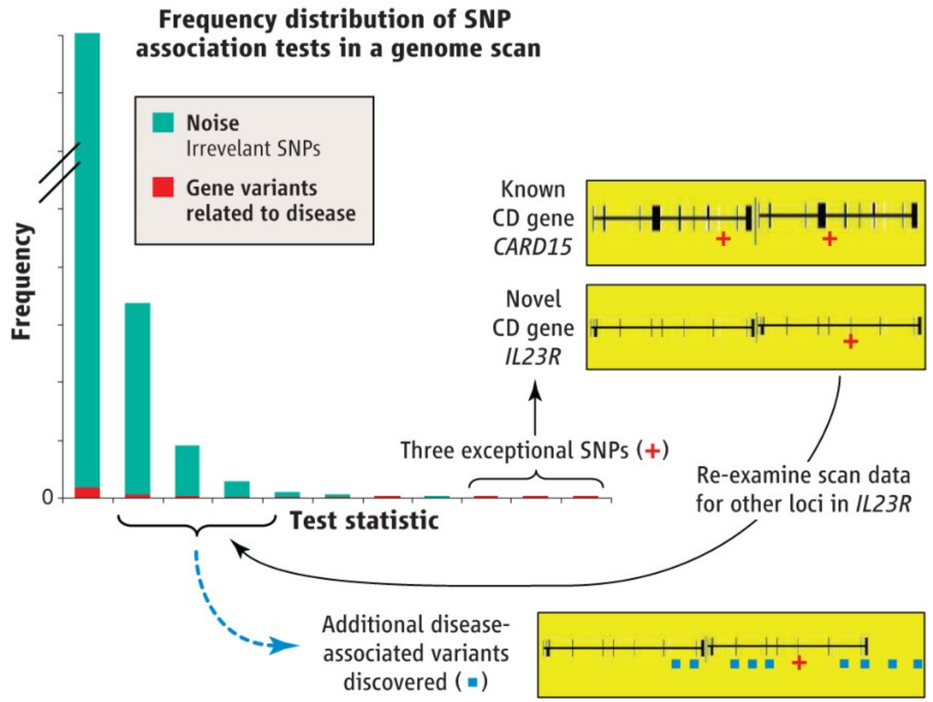
Label directly and consistently



a and b label the same points. Which is easiest to read? Why?

Krzywinski, M. Points of View: Labels and callouts. Nature Methods 10, 275 (2013), Fig. 1

Arrows and callouts tell the story



Cascade of discovery. Common human diseases are thought to have few disease loci (red) relative to the number of single-nucleotide polymorphisms (SNPs) tested (green), and most of the true loci will have a small effect. The greatest statistical power is thus attributed to loci that occur least frequently in the genome. Duerr *et al*. found three SNPs associated with Crohn's disease (CD), two in a known disease gene and one in the *IL23R* gene. This allowed them to extract other real SNP-disease associations embedded in the noise of the initial genome-wide scan.

Cardon, *Science* **314**:1403 (2006)

Follow the arrows: what is the argument, step by step?

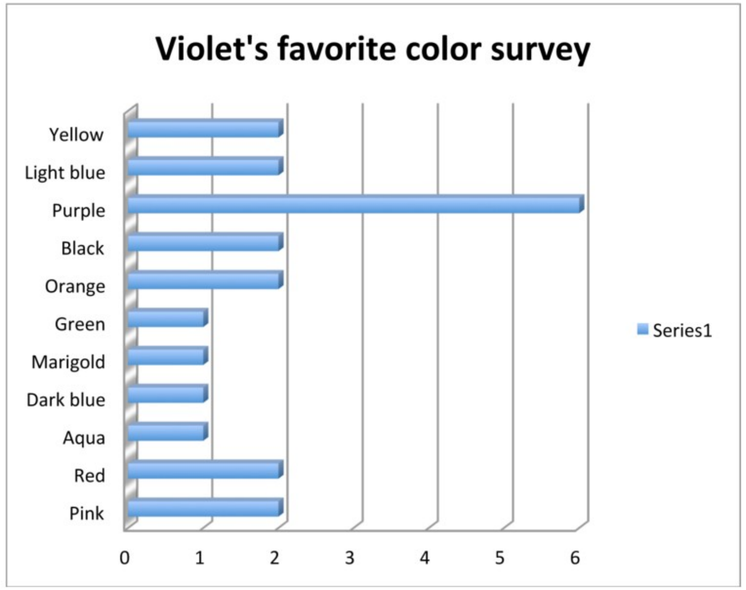
What is each callout pointing at?

Cardon, L.R. Science 314, 1403 (2006)

Maximize data, minimize clutter



22 data points, 47 visual elements



22 data points, 171 visual elements

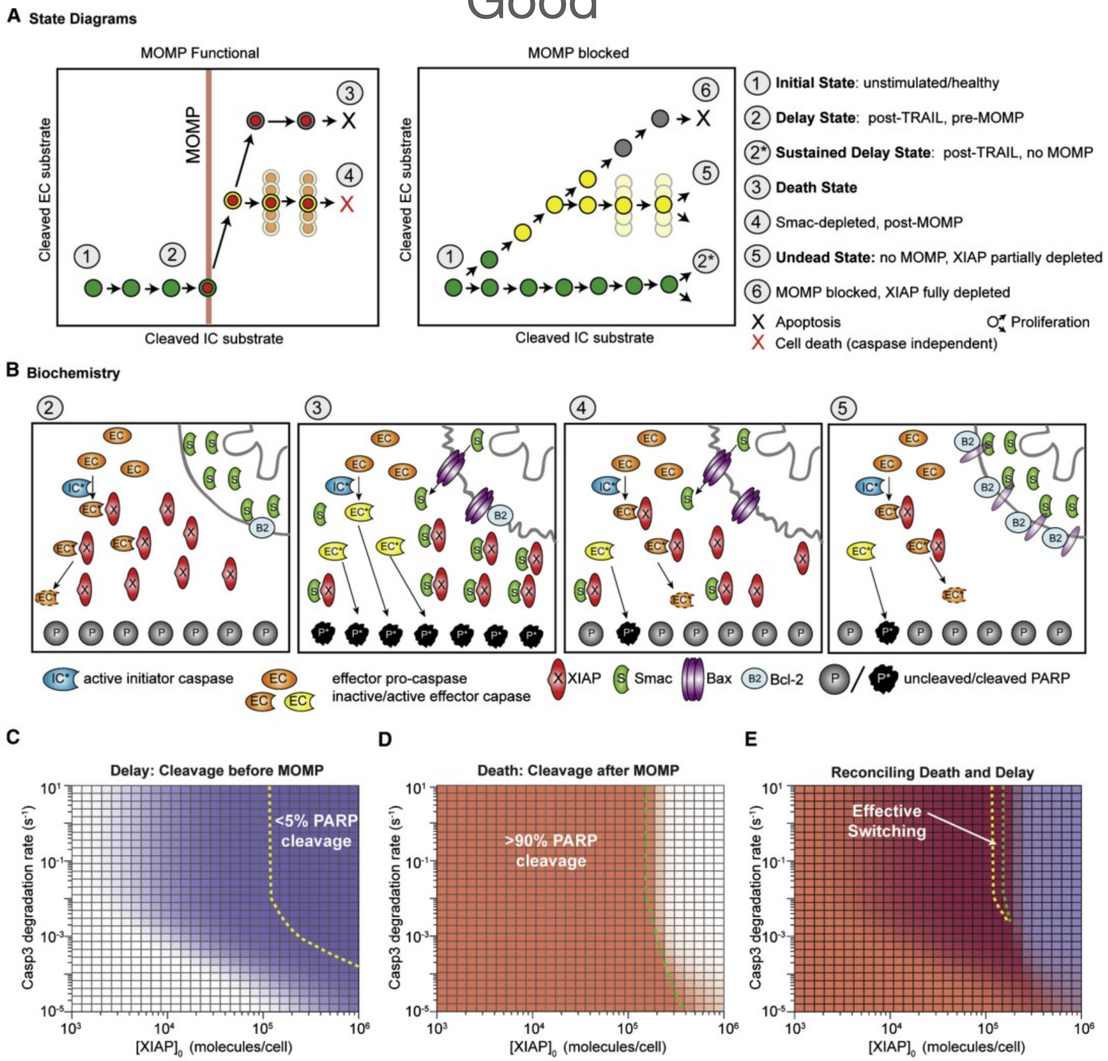
Both: 22 data points

How much of each chart is data?

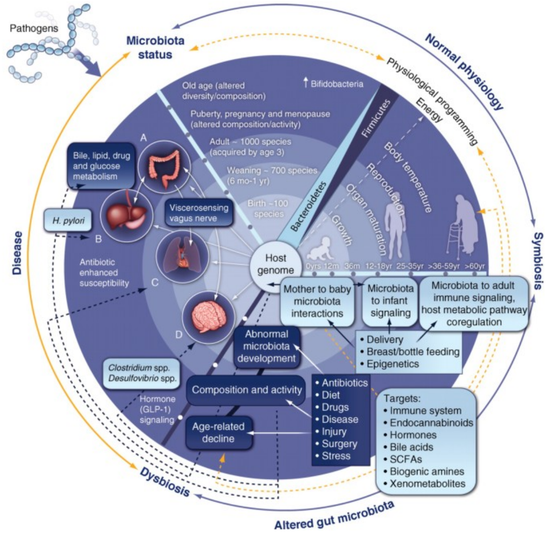
From K. Janes, Technical information design (lecture), University of Virginia. After Tufte, edwardtufte.com

Model figures

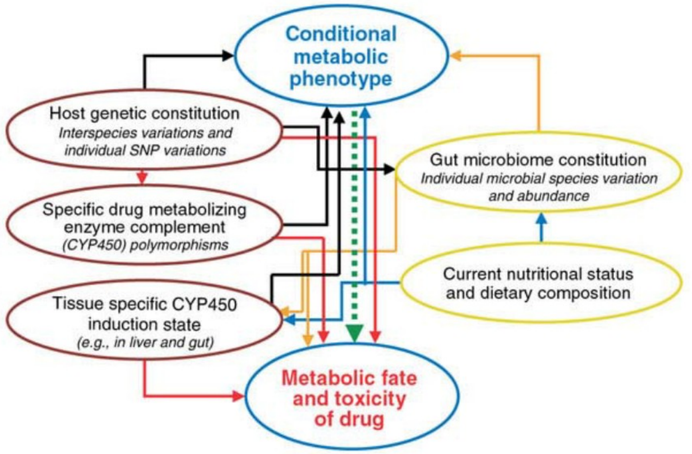
Good



Less Good



Science (2012)

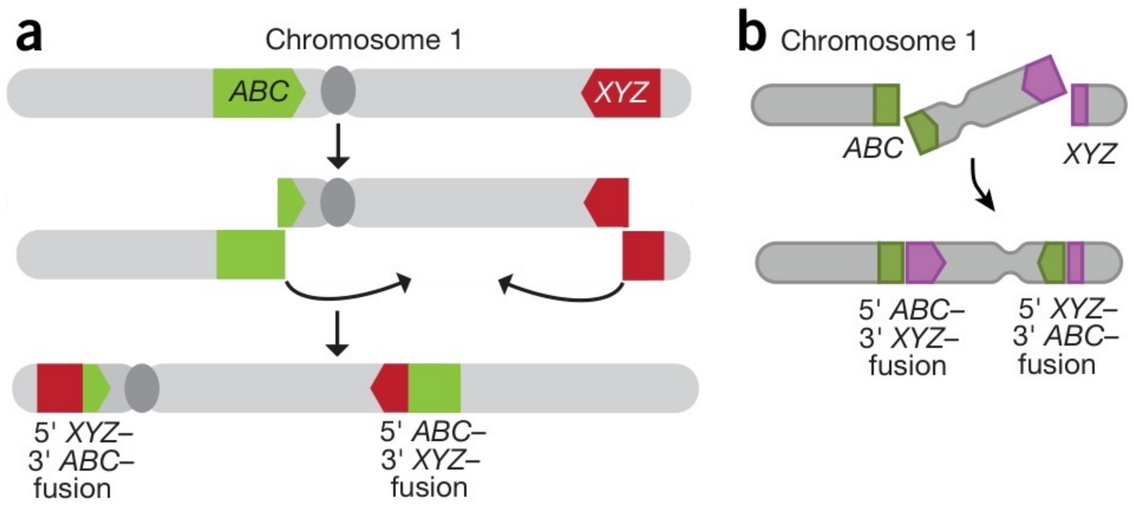


Nat Biotechnol (2004)

Albeck et al., *Mol Cell* **30**:11 (2008)

From K. Janes, Technical information design (lecture), University of Virginia. Left: Albeck, J.G. et al. Mol. Cell 30, 11 (2008). Right: Science (2012); Nat. Biotechnol. (2004)

Simplify to clarify



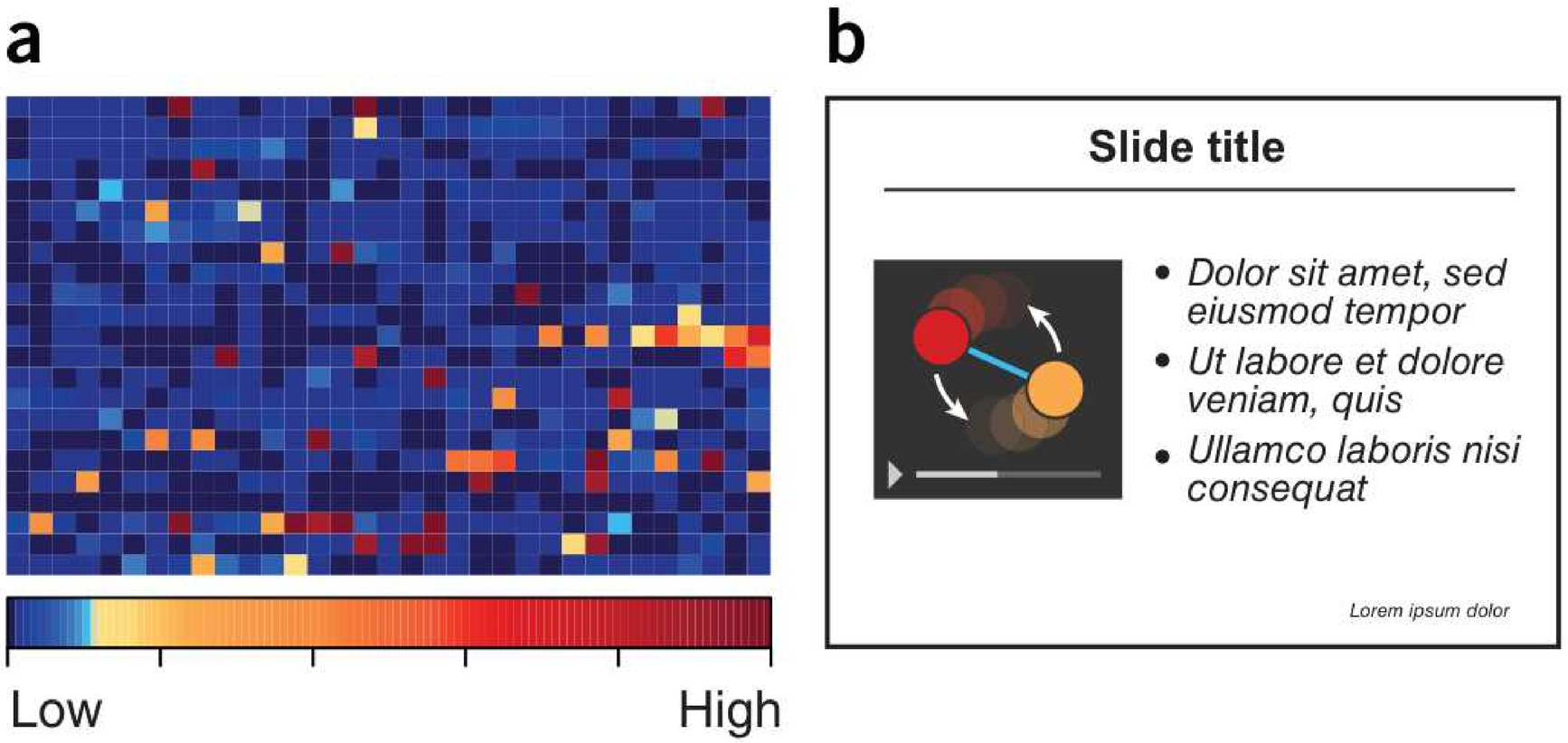
a and b show the same process

What did b leave out? Does the reader need it?

Wong, B. Points of View: Simplify to clarify. Nature Methods 8, 611 (2011), Fig. 1

How to build trust and communicate honestly in a science visualization

Emphasis should match importance



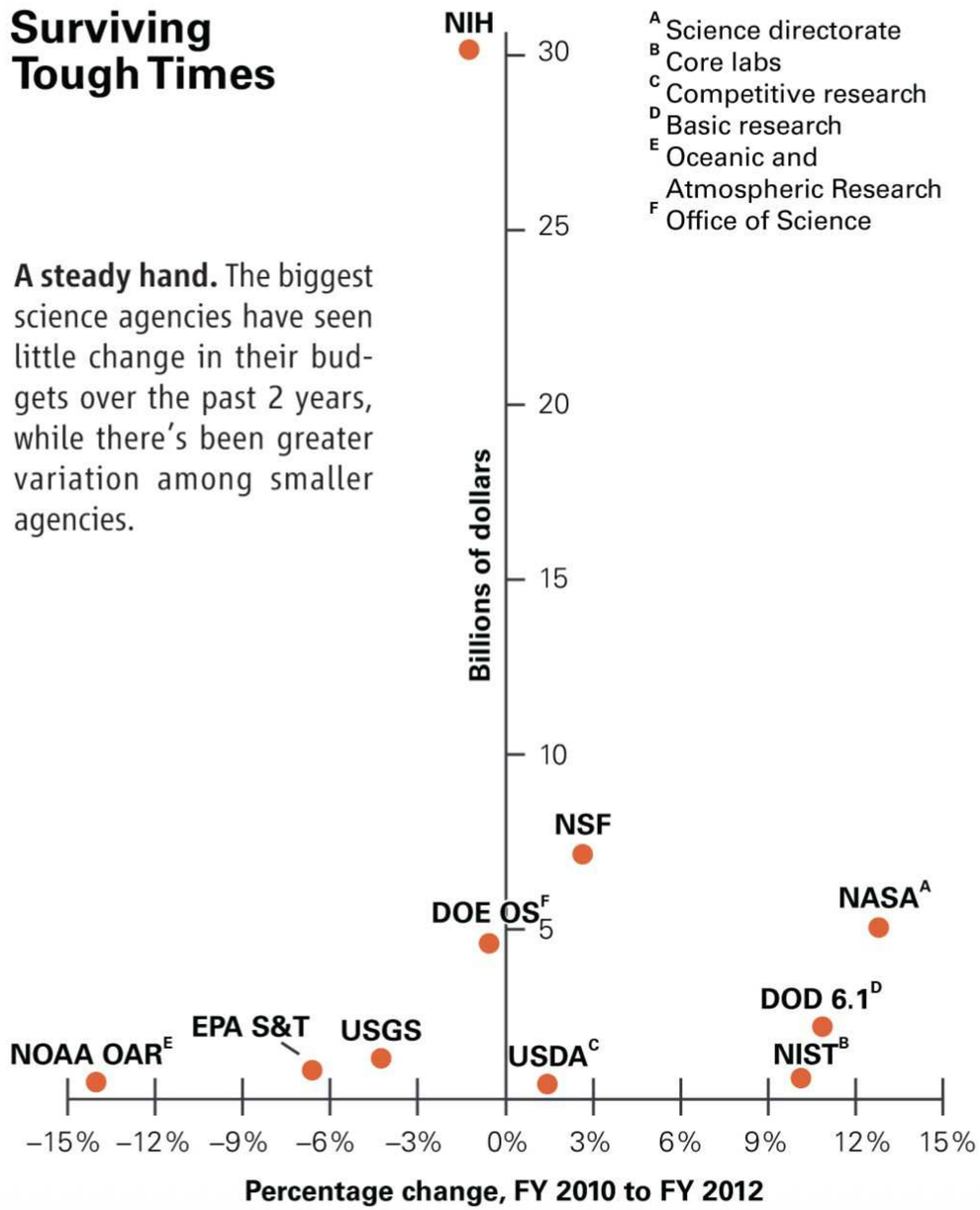
Where does your eye go first in a?

Is that the most important value?

Make the most important thing the most noticeable

Wong, B. Points of View: Salience to relevance. Nature Methods 8, 889 (2011), Fig. 2

Keep independent and dependent variables separate



Science (2012)

Go beyond bar graphs

OPEN ACCESS

PERSPECTIVE

Beyond Bar and Line Graphs: Time for a New Data Presentation Paradigm

Tracey L. Weissgerber, Natasa M. Milic, Stacey J. Winham, Vesna D. Garovic

Published: April 22, 2015 • https://doi.org/10.1371/journal.pbio.1002128

Article | Authors | Metrics | Comments | Media Coverage

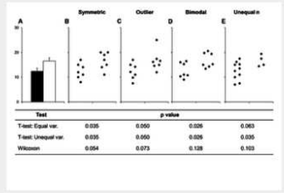
Abstract
Introduction
Are Your Figures Worth a Thousand Words?
Summary Statistics Are Only Meaningful When There Are Enough Data to Summarize
Recommendations for a New Data Presentation Paradigm
Conclusions
Supporting Information
References

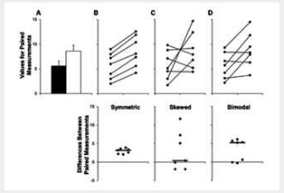
Reader Comments
Figures

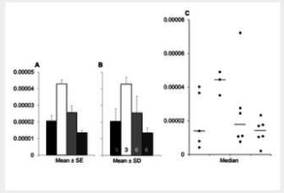
Abstract

Figures in scientific publications are critically important because they often show the data supporting key findings. Our systematic review of research articles published in top physiology journals (n = 703) suggests that, as scientists, we urgently need to change our practices for presenting continuous data in small sample size studies. Papers rarely included scatterplots, box plots, and histograms that allow readers to critically evaluate continuous data. Most papers presented continuous data in bar and line graphs. This is problematic, as many different data distributions can lead to the same bar or line graph. The full data may suggest different conclusions from the summary statistics. We recommend training investigators in data presentation, encouraging a more complete presentation of data, and changing journal editorial policies. Investigators can quickly make univariate scatterplots for small sample size studies using our Excel templates.

Figures







A review of 703 physiology papers: most showed continuous data only as bars or lines

What would you use instead in your project?

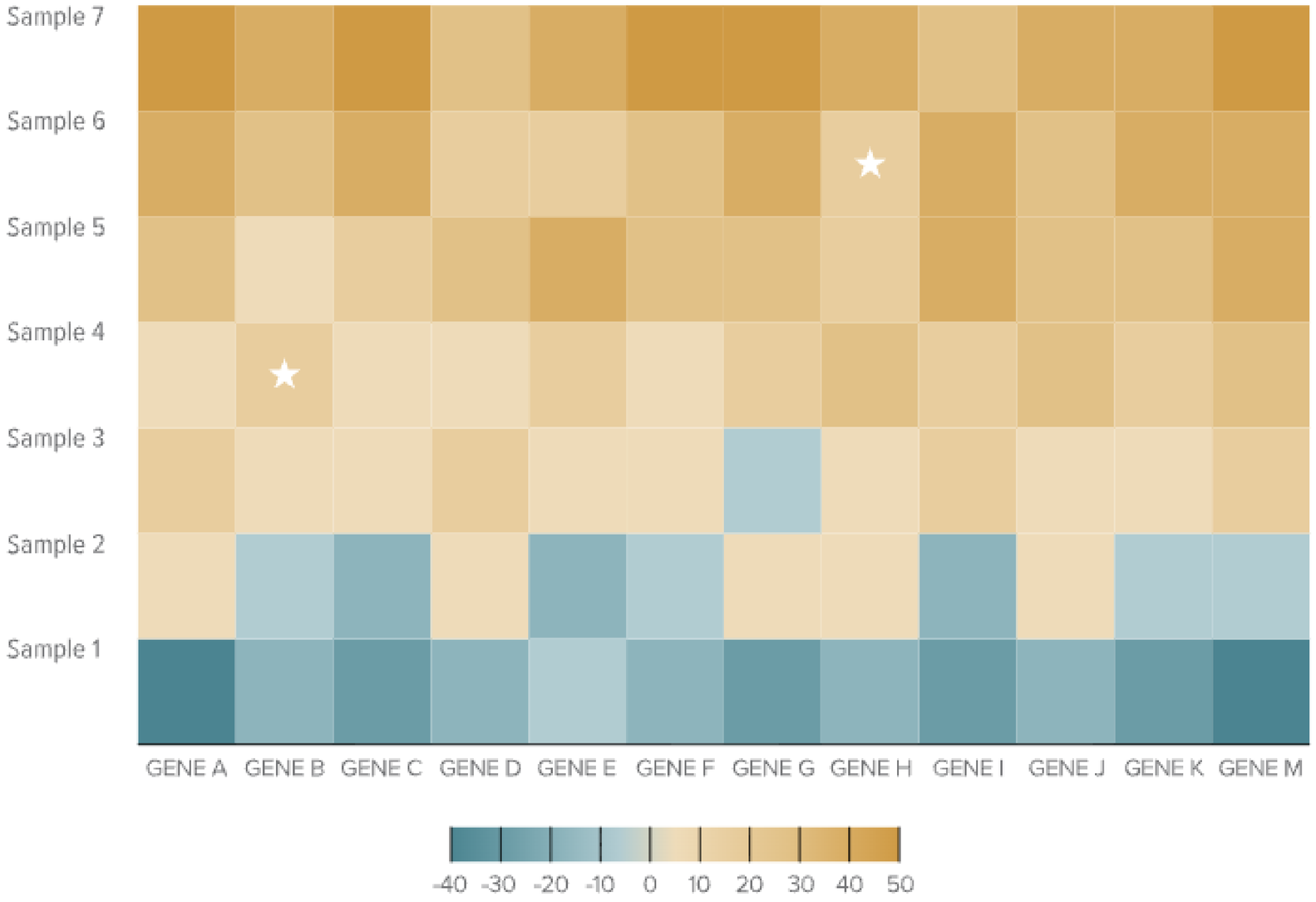
If the distribution is key to your conclusion, how would you show it?

Weissgerber, T.L., Milic, N.M., Winham, S.J. & Garovic, V.D. (2015)

Uncertainty, examples, and axes

- Uncertainty: SD, SE, or 95% CI? Does it matter which?
- Examples: How do you decide what is representative?
- Axes: when is zooming in honest, and when is it spin?

Always consider color effectiveness



Which of the two starred squares are darker?

Use caution when a heat map relies only on color intensity

A skeptic's checklist for your Project 2 deliverable

- 1. Could a reader reach my conclusion from the figure alone?
- 2. Is the most eye-catching thing also the most important thing?
- 3. Can they see the data, its spread, and the uncertainty?
- 4. Is my example typical, and do I say how I chose it?
- 5. What did I leave out, and would it change their mind?
- Use this on your Project 2 final piece if you are doing a science visualization

Infographic vs scivis

Pros and cons of each. You should choose your project 2 deliverable based on your **results** and **audience, not just what you're drawn to**

Infographic

- WHY: storytelling to a general audience
- More flexible, more creative
- Incorporate visual elements that are NOT charts: icons, text, visually appealing embellishments
- Layout is flexible; consider where the reader's eye will go to tell the story, but the reader may wander/explore
- No caption: text is in graphic

In between author driven and reader driven

SciVis

- WHY: Communicate a scientific result from data analysis
- More structured, concise
- Should NOT embellish; arrows/annotations are good for clarity, but limit icons, etc. except in model figures/overview panels
- Layout is constrained; label panels, consider whitespace, clear progression through figure
- Caption with figure title and info for each panel

Author driven

Summary

- A scientific reader is a skeptic: give them what they need to check you
- Tell a story by ordering the evidence: unfold it, don't dump it
- Trust comes from showing the data, the uncertainty, and the loose ends
- Structure, not spin: no distortion, embellishment, concealment, or unrepresentative displays