

Guidelines for good visualization

History time!

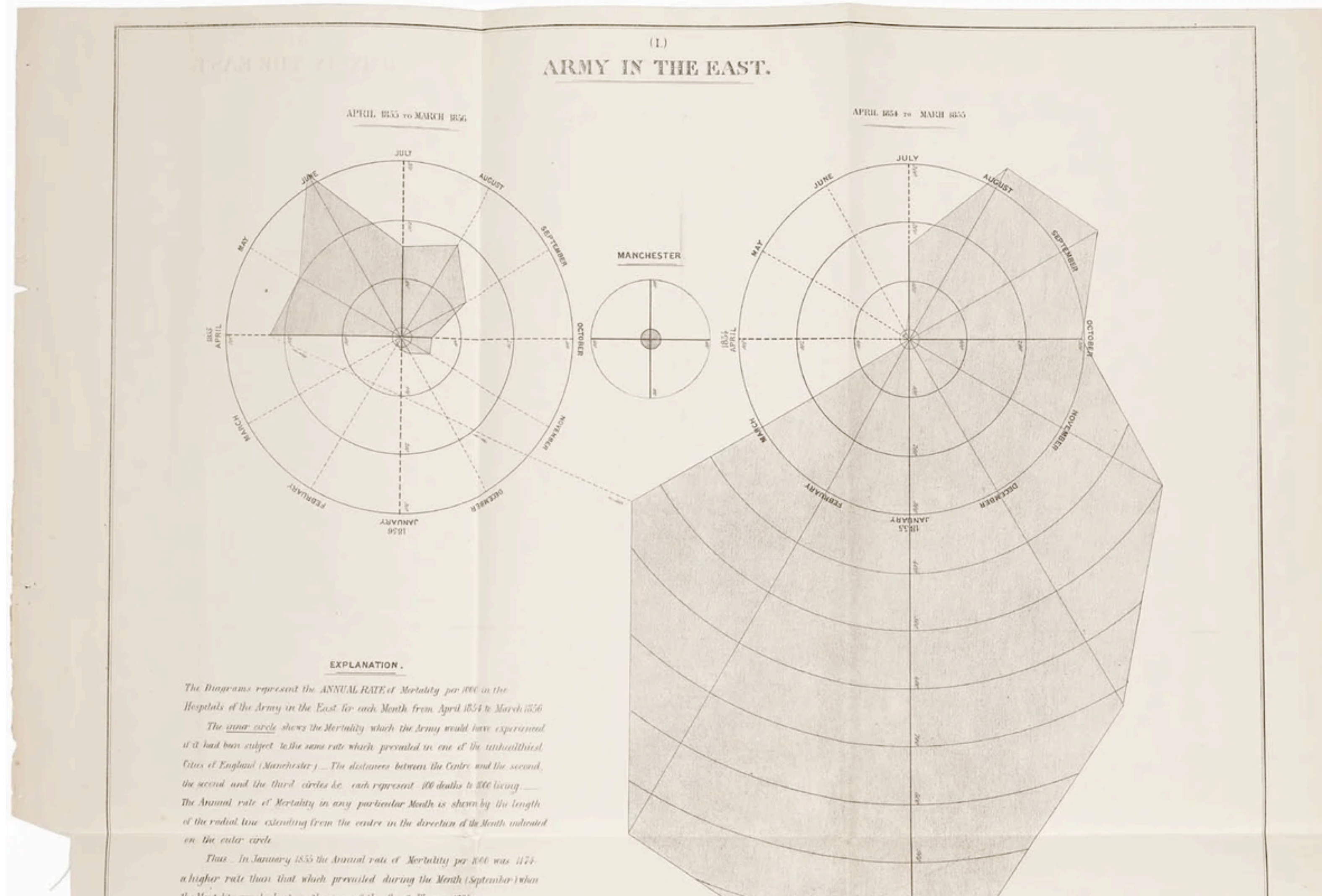
Florence Nightingale



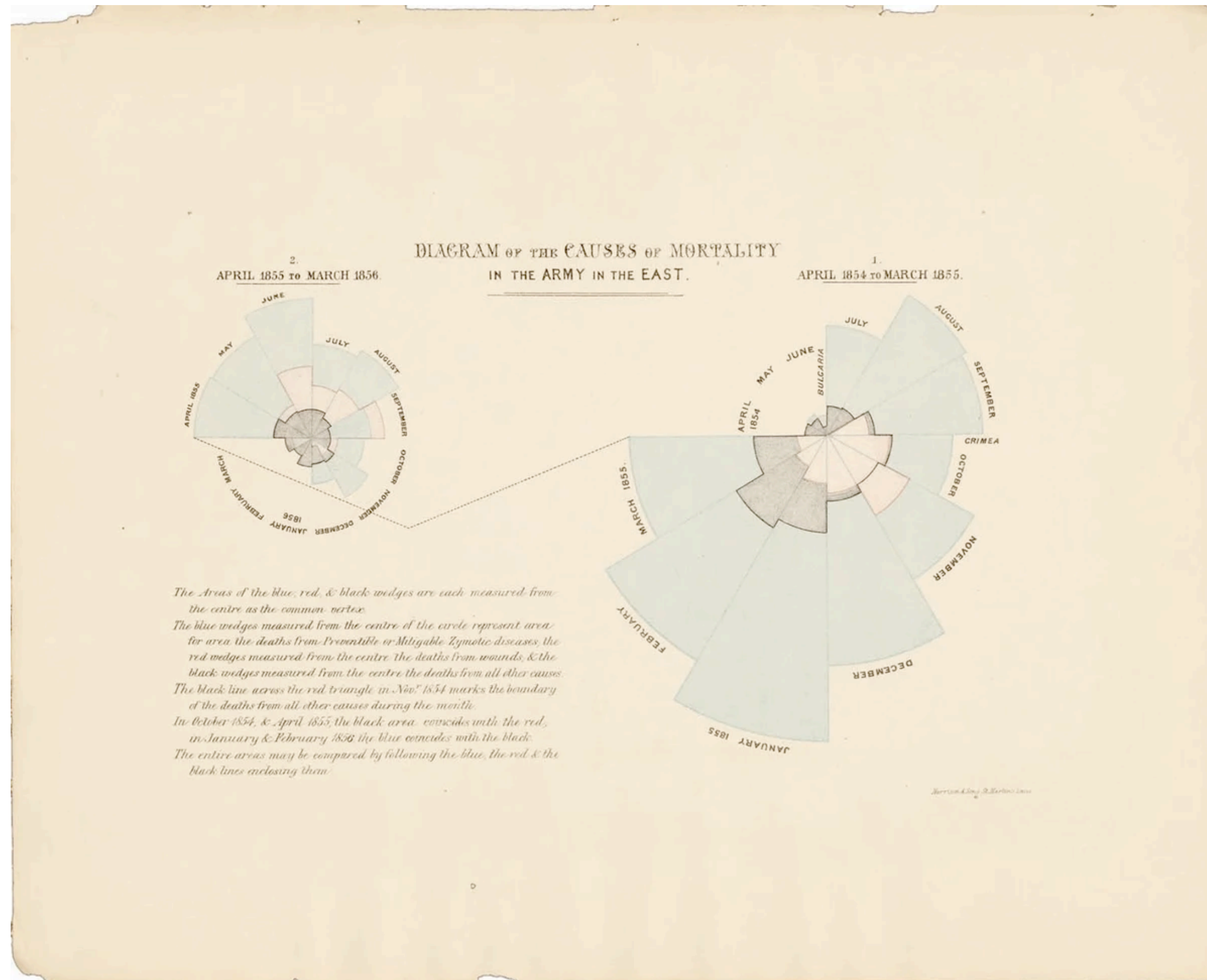
1854-1856: The Crimean War



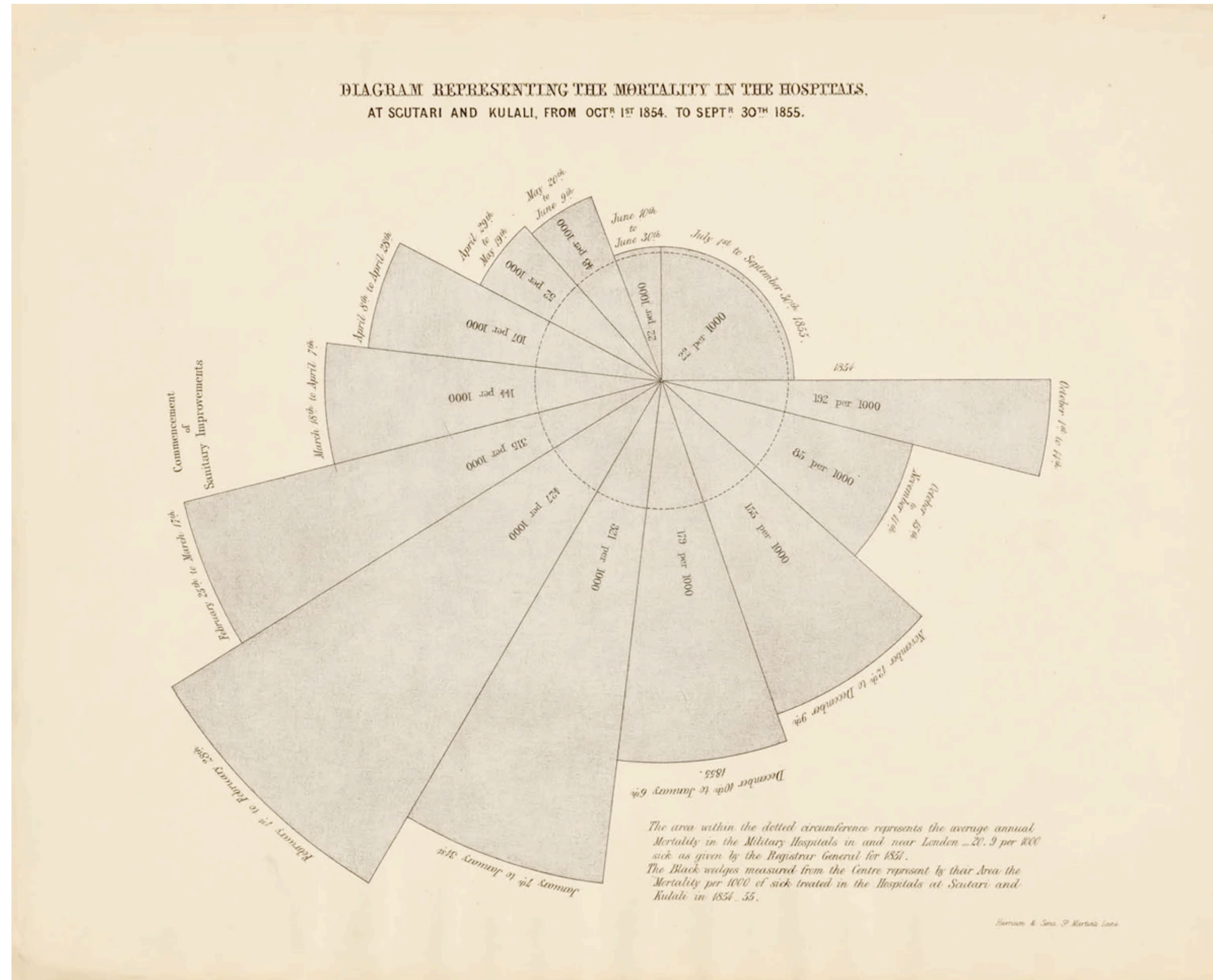
Tracking Mortality



Tracking Causes



Implementing Solutions



Legacy



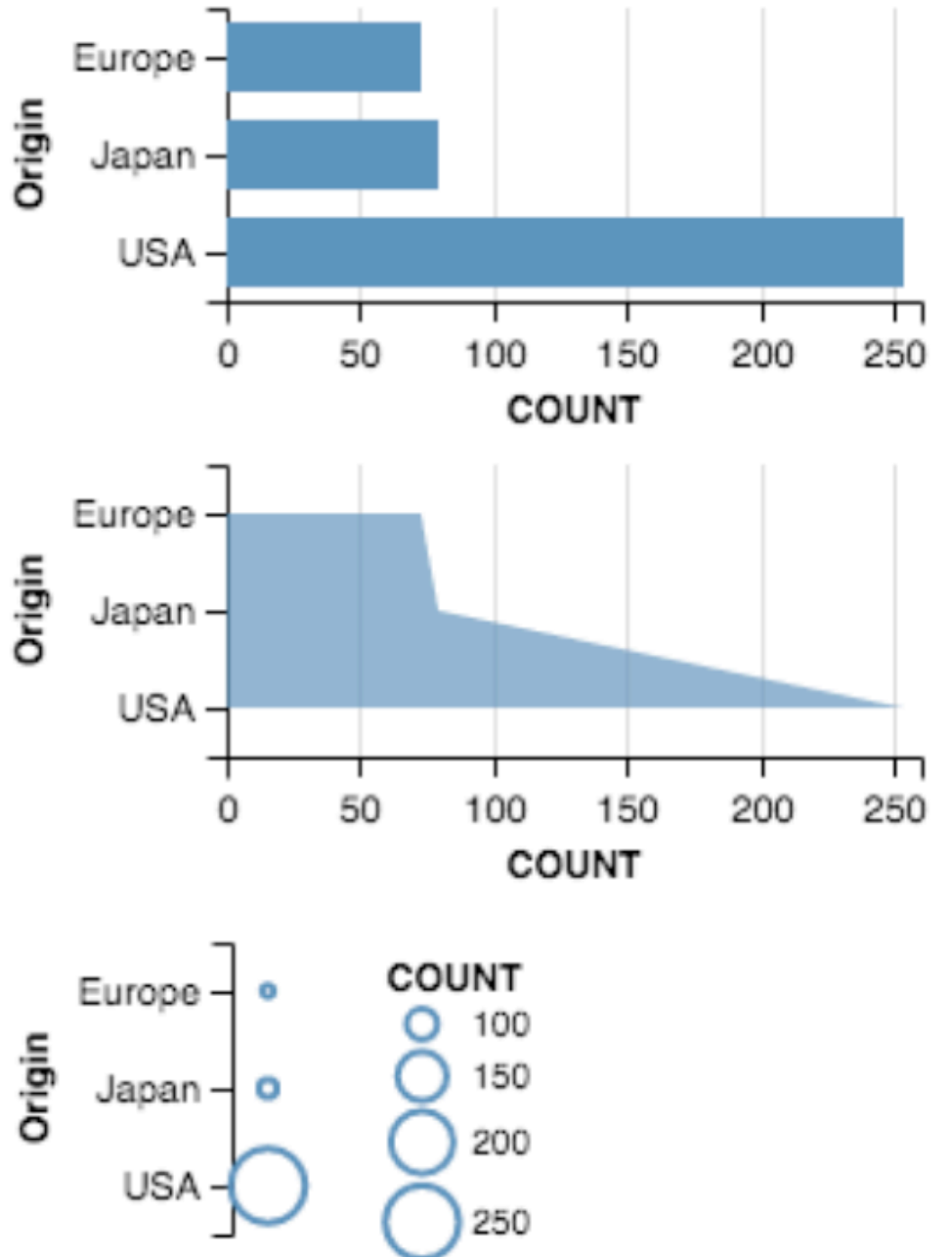
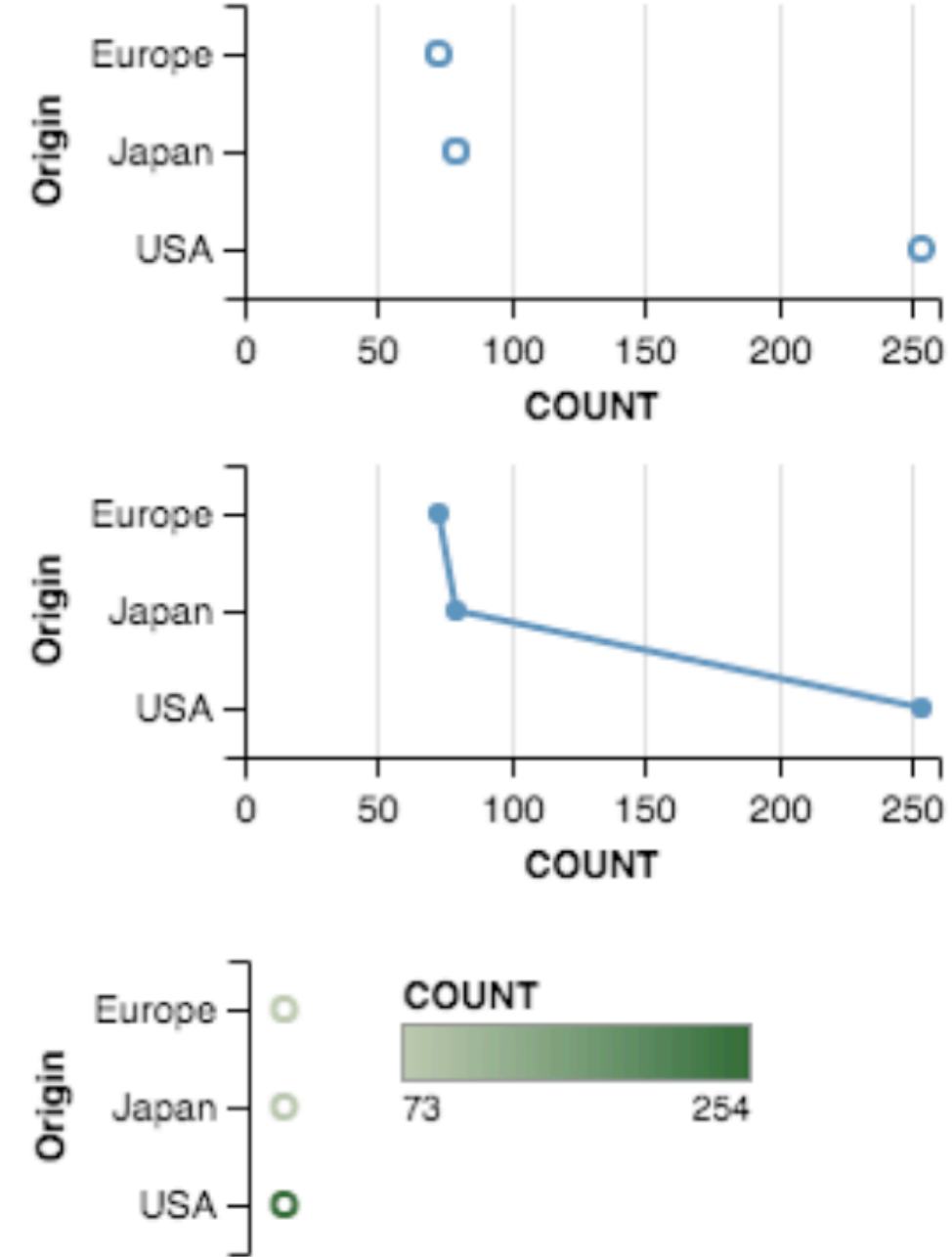
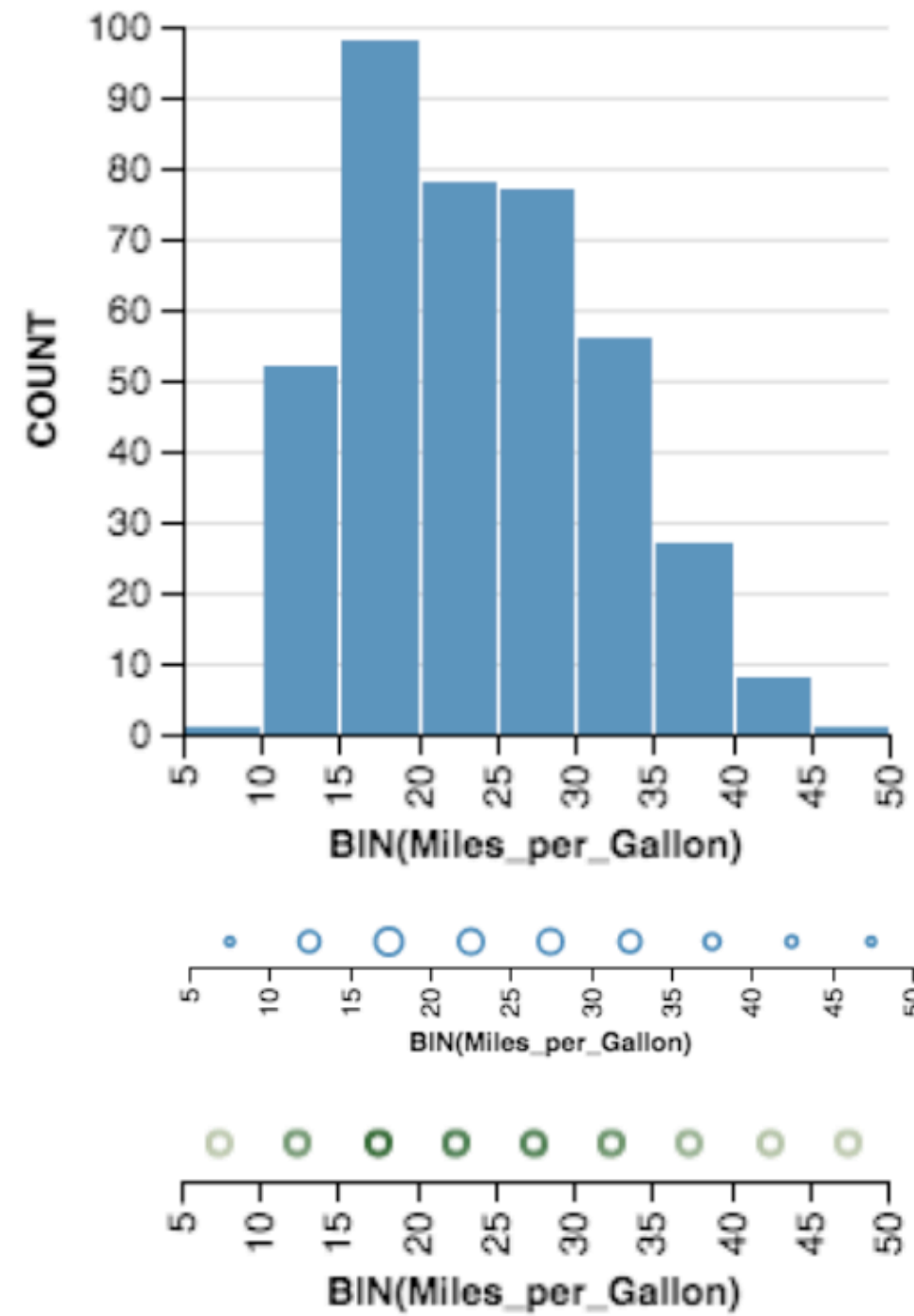
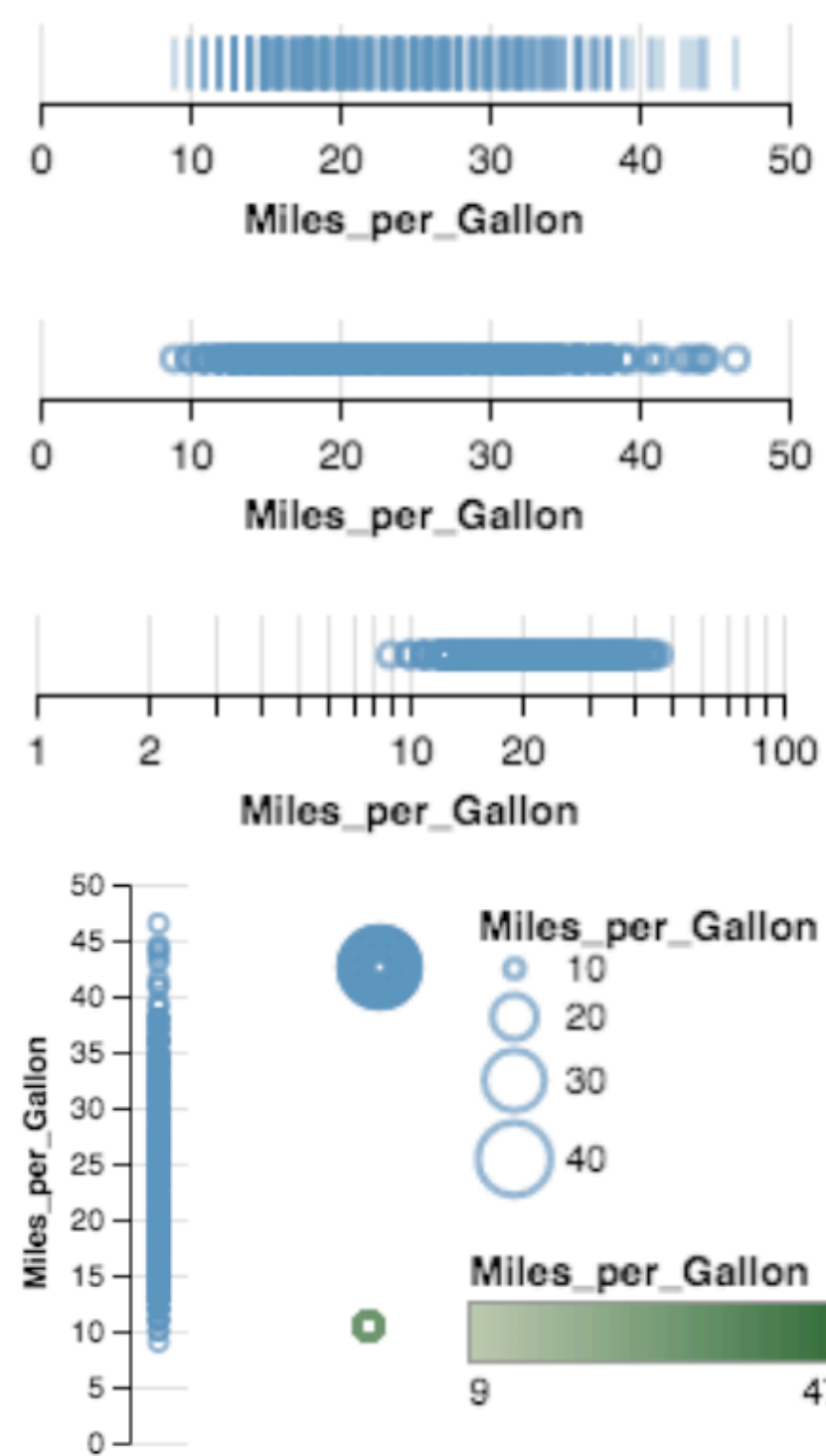
Grammar of Graphics

Components of the grammar of graphics

Stretching the metaphor: The *parts of speech* for our visual language

- Data
- Aesthetic mappings
- Geometries
- Transforms
- Scales
- Coordinate systems
- Faceting systems
- Annotations

Large space of visualizations!



variables × aesthetics × geoms × scales...

**Our framework for good
visualization**

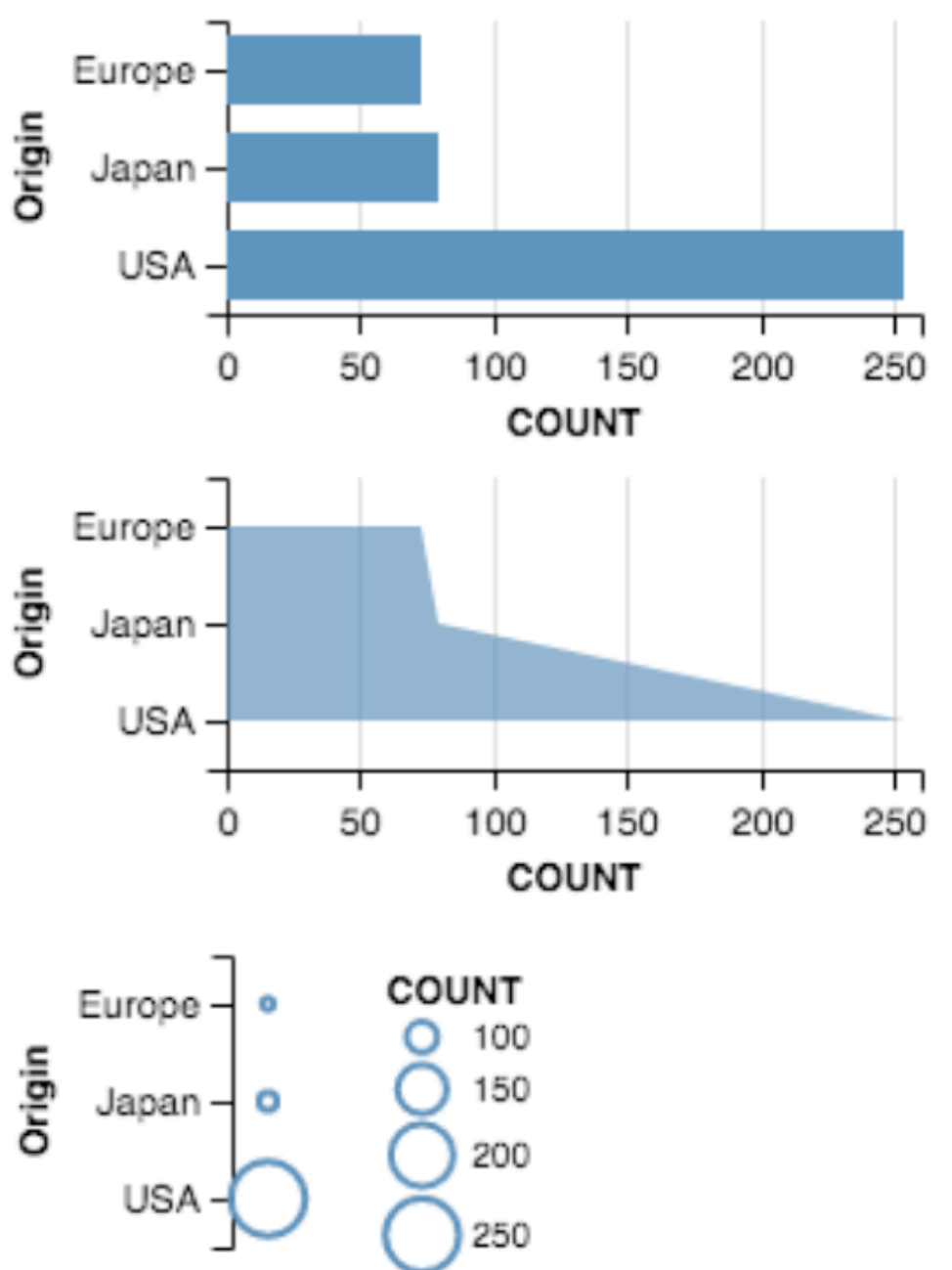
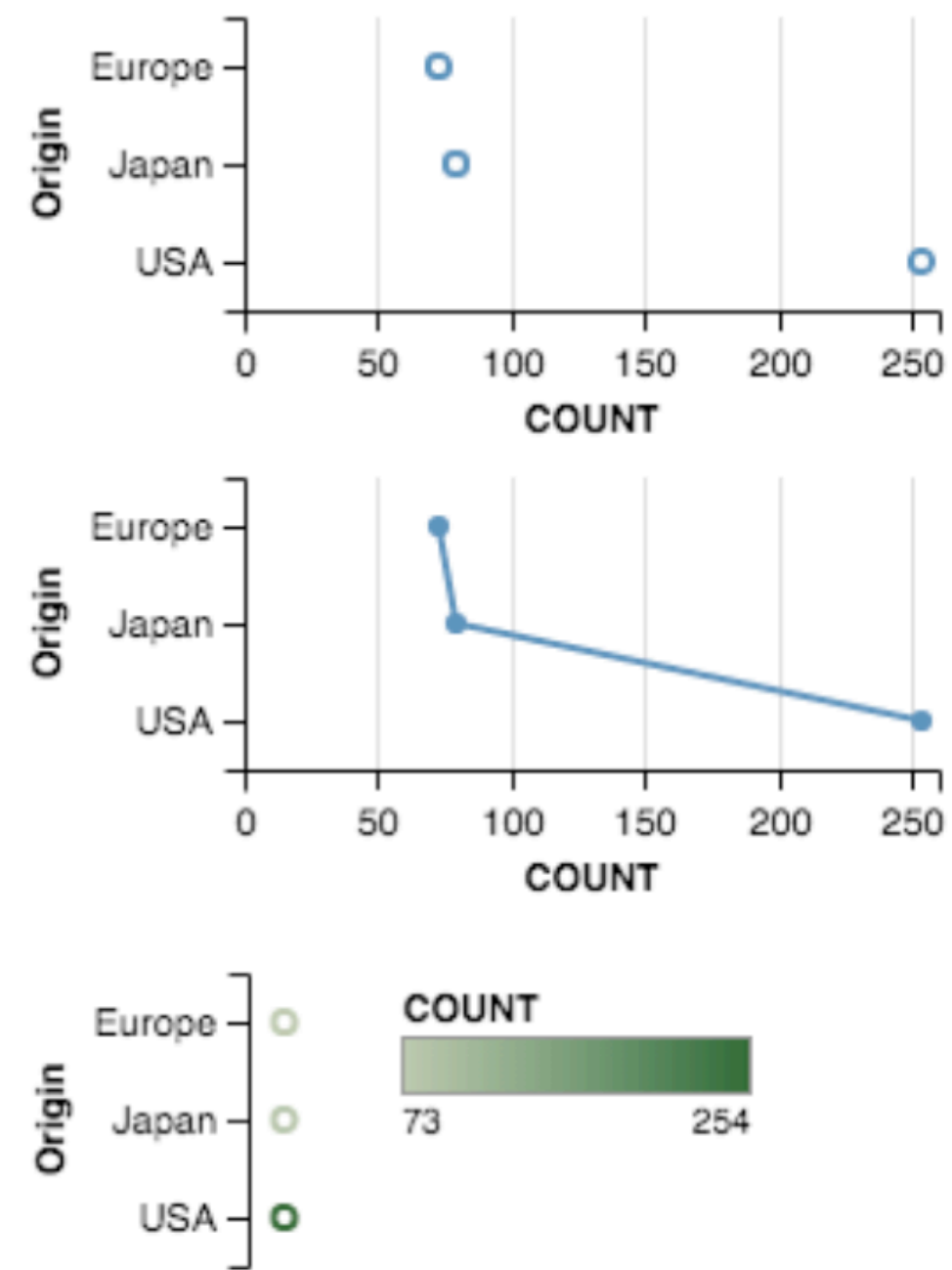
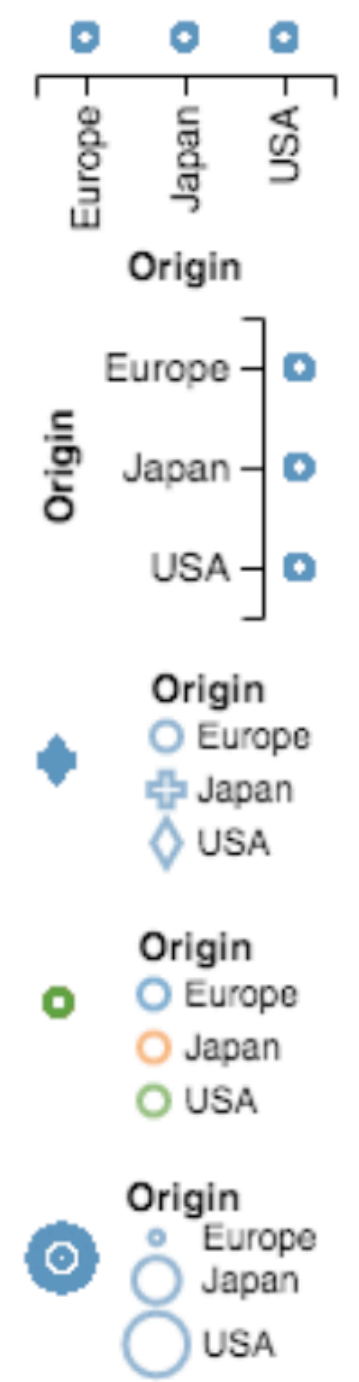
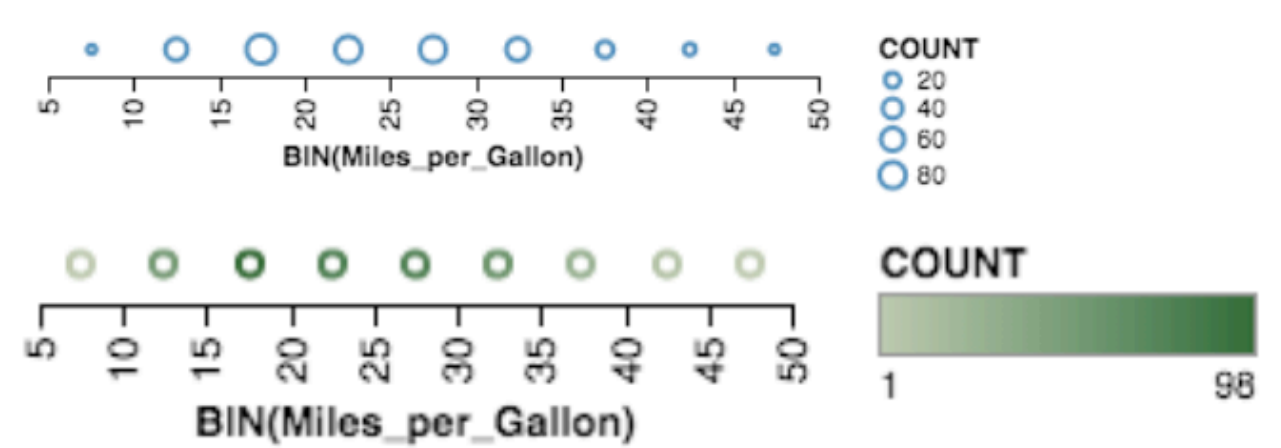
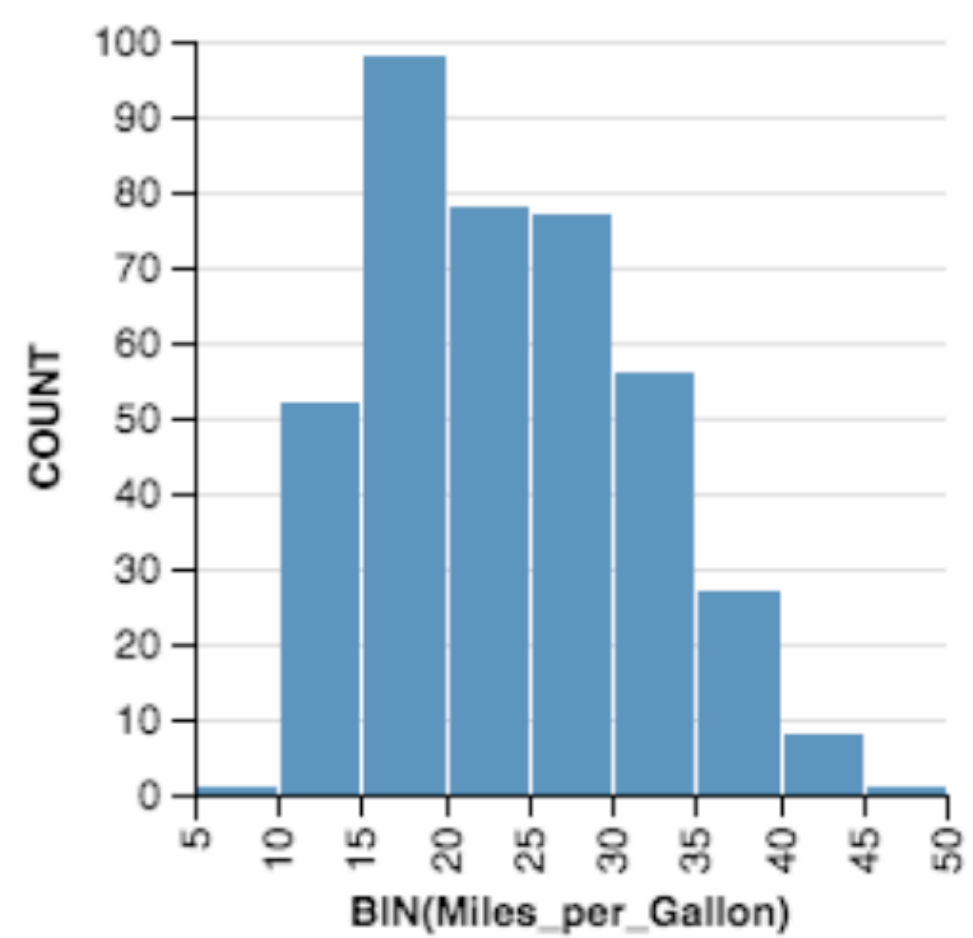
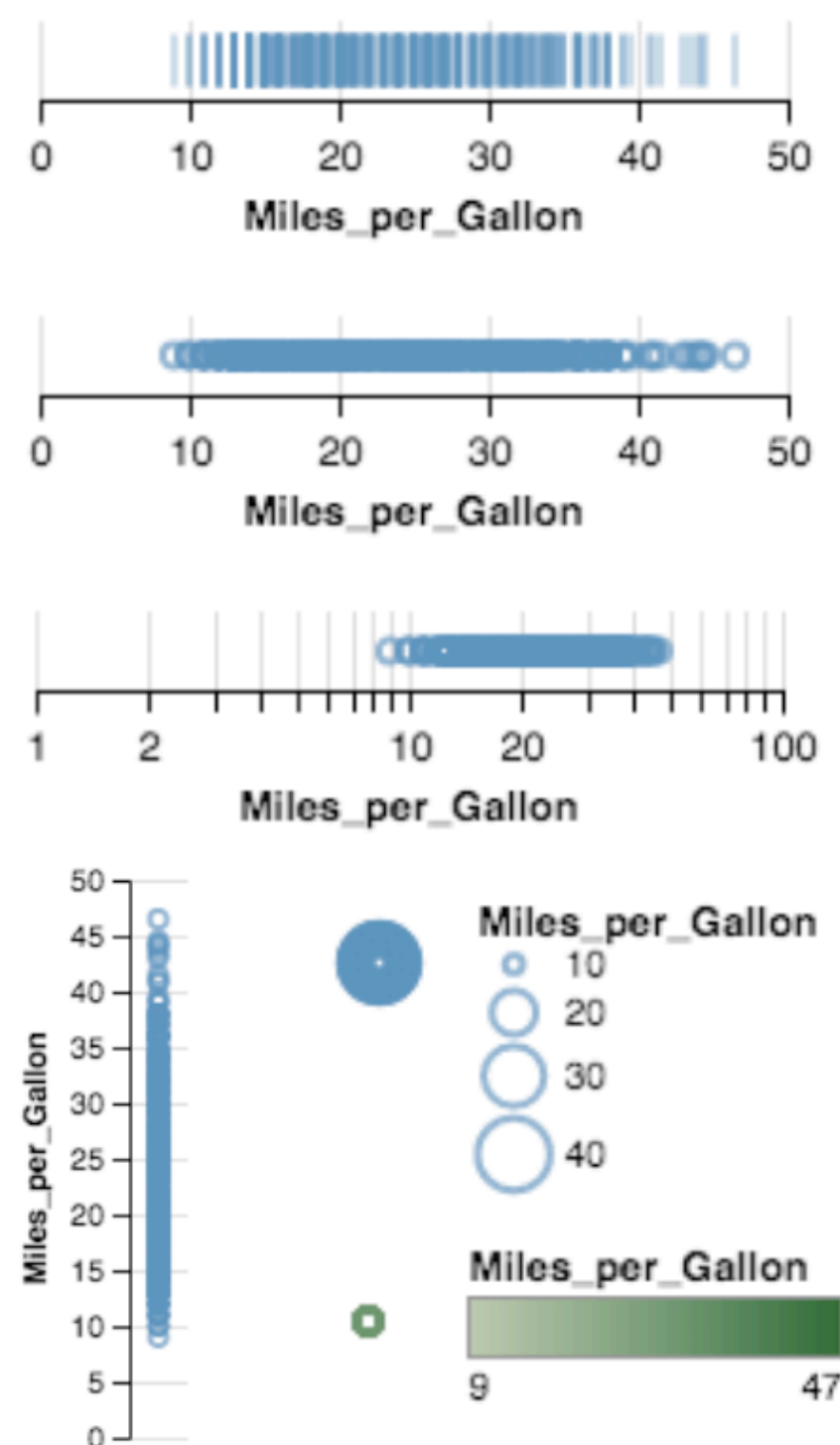
9 Criteria for good visualizations

- Expressiveness
- Effectiveness
- Intuitiveness
- Honesty
- Accessibility
- Impact
- Insight
- Writing
- Engineering

9 Criteria for good visualizations

- Expressiveness
- Effectiveness
- Intuitiveness
- Honesty
- Accessibility
- Impact
- Insight
- Writing
- Engineering

Can we automate choices?



variables × aesthetics × geoms × scales...

Can we automate choices?

To answer this question we first need to determine:

- How can we efficiently search the space of possible visualizations?
- What criteria do we use?

In the 1980s **Jock MacKinlay** was one of the first to tackle these questions.



MackKinley's Criteria

*“**Expressiveness** criteria identify graphical languages that express the desired information.”*

*“**Effectiveness** criteria identify which of these graphical languages, in a given situation, is the most effective at exploiting the capabilities of the output medium and the human visual system.”*

What do these mean?

Expressiveness

Our first criteria: expressiveness

MacKinley's definition:

*“A set of facts is **expressible** in a [visual] language if it contains a sentence [visualization] that:”*

- *“Encodes **all** the facts in the set”*
- *“Encodes **only** the facts in the set”*

Expressiveness



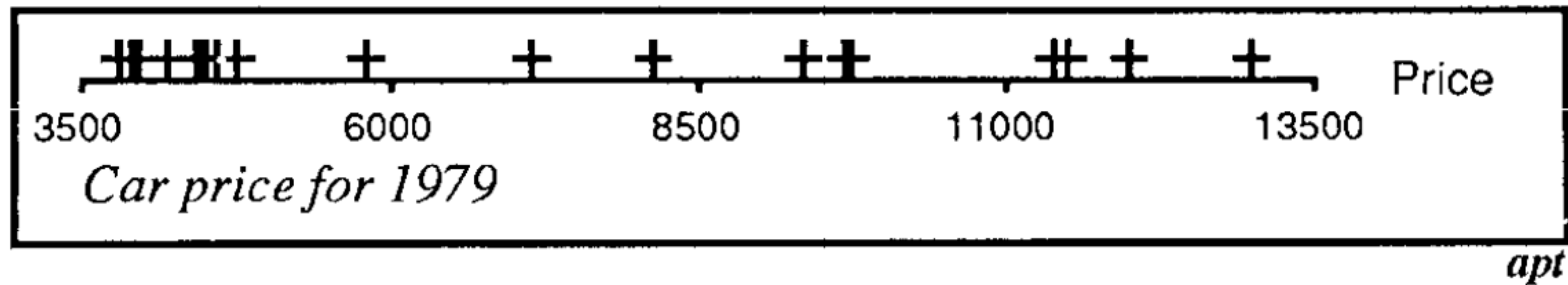
Tell the whole truth and nothing but the truth

Intuitiveness & honesty



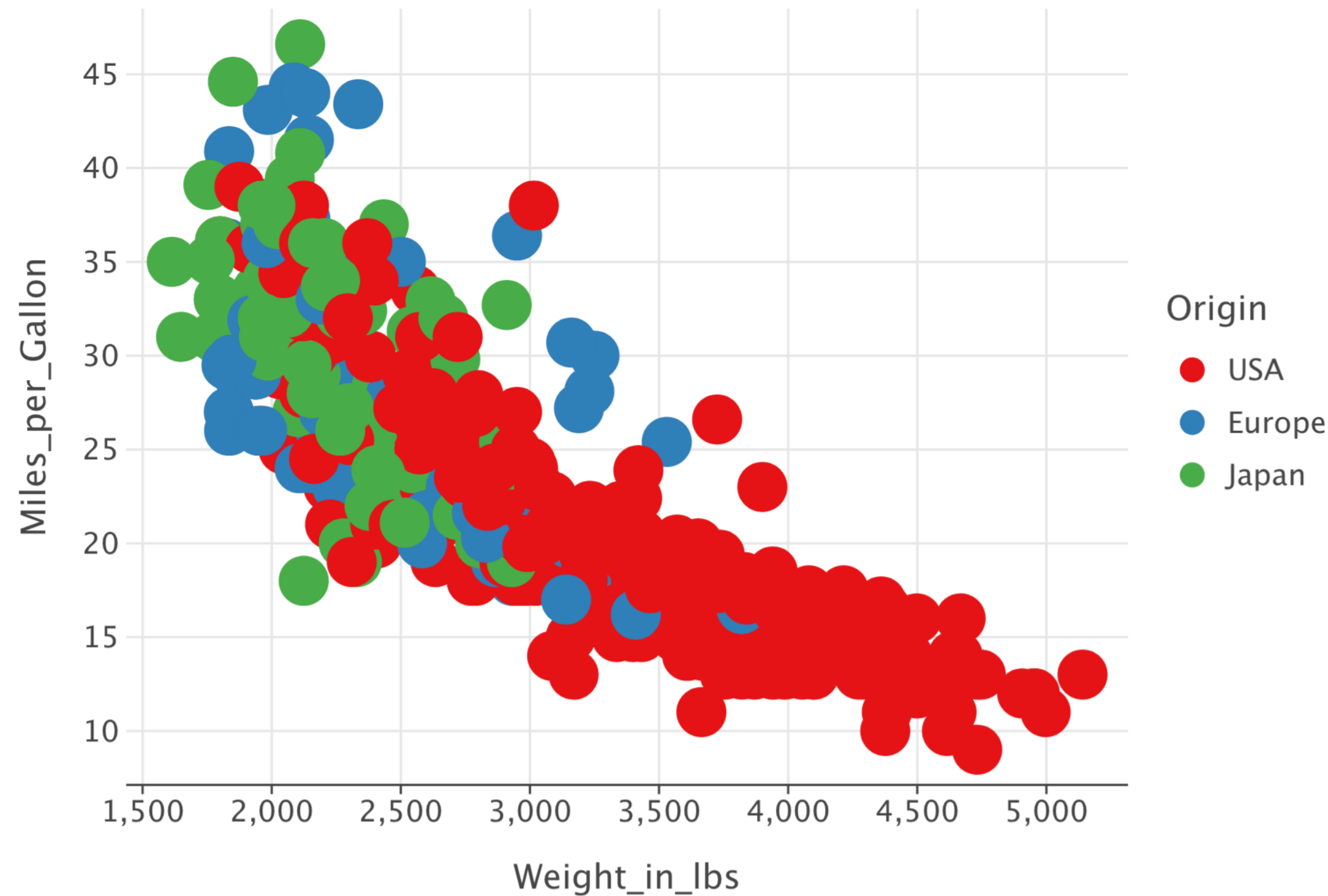
***Our definition will only
be the first half***

Express all the facts: Overplotting

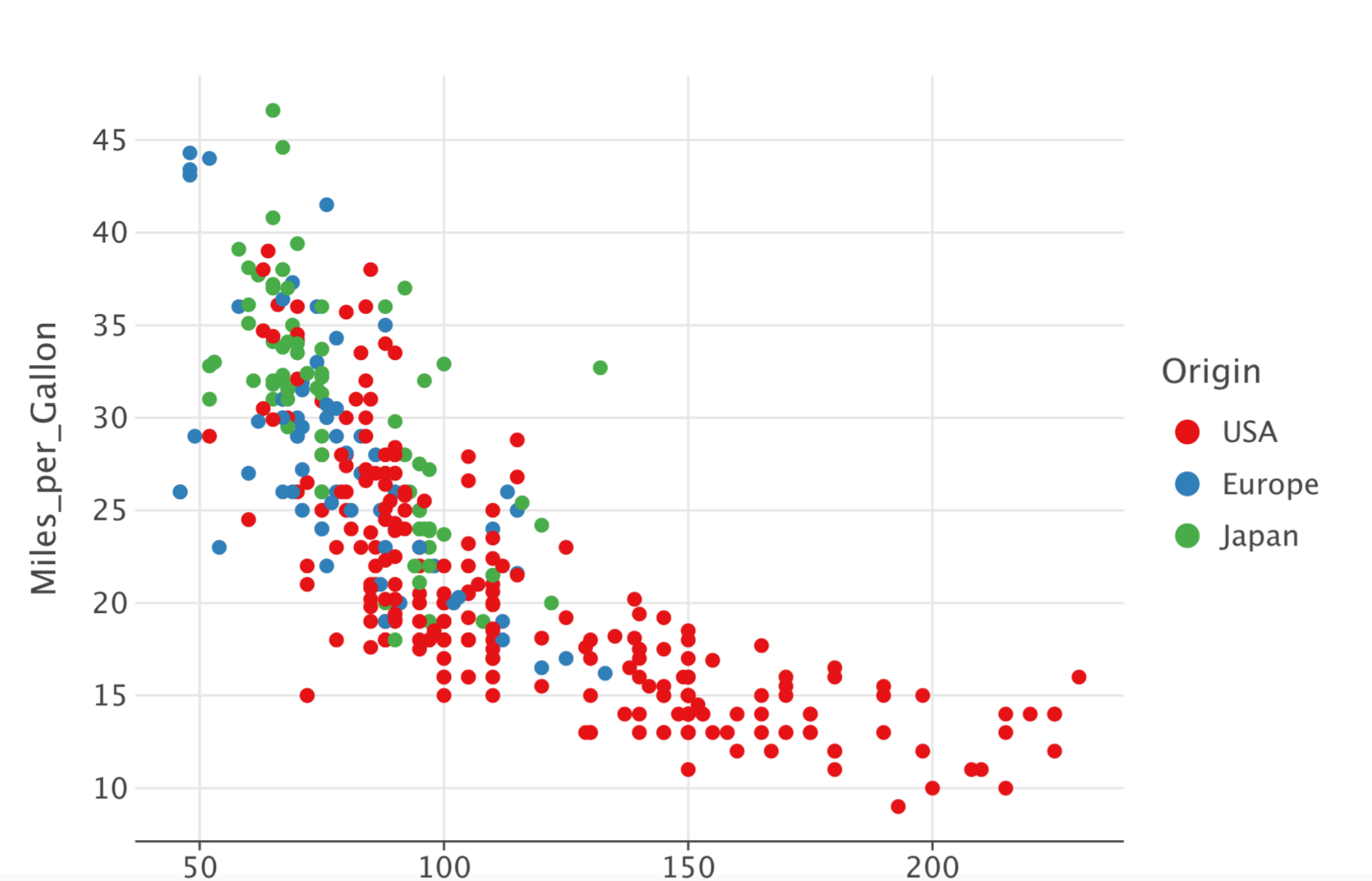


How many cars are in this dataset?

Express all the facts: Overplotting

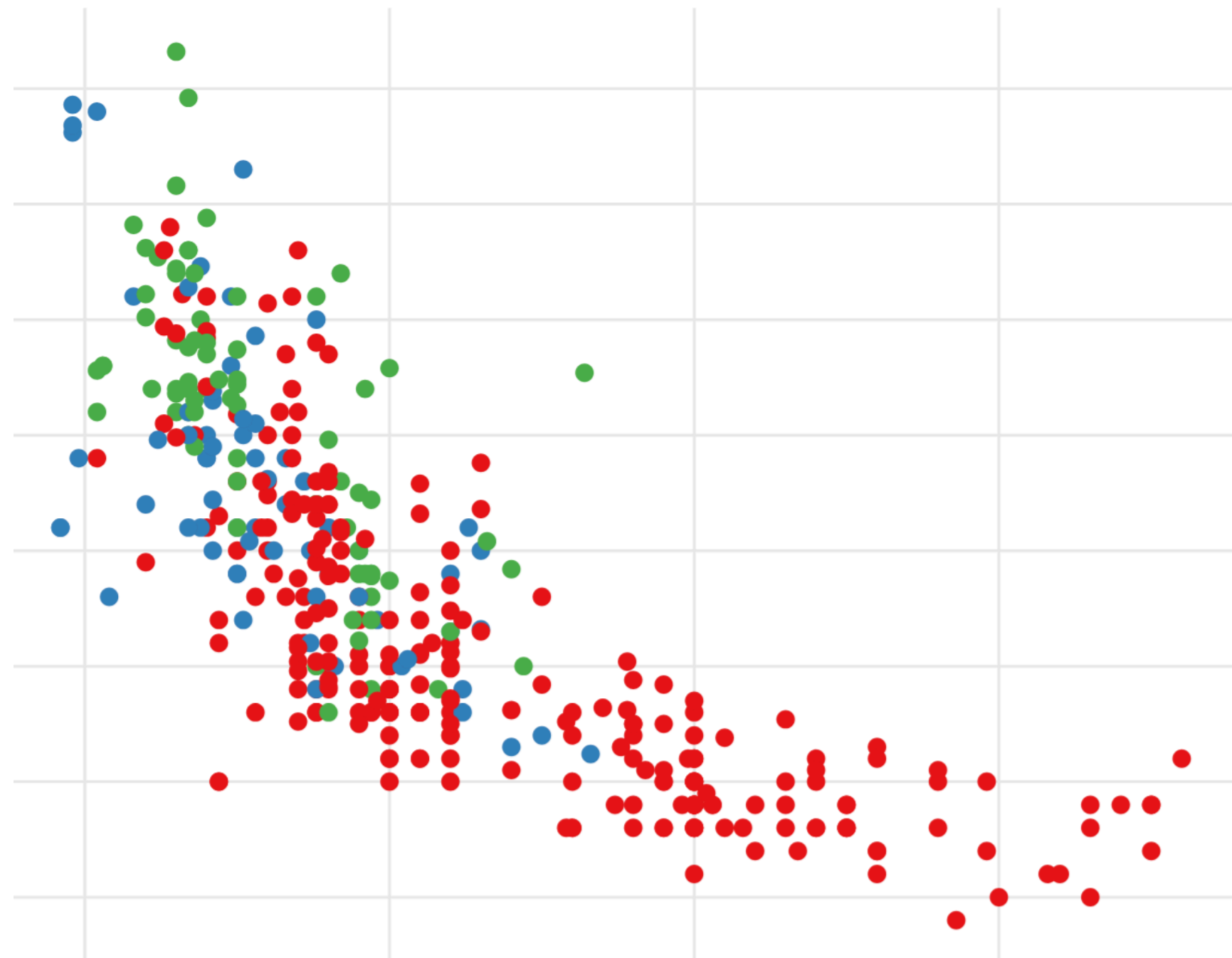


Not expressive

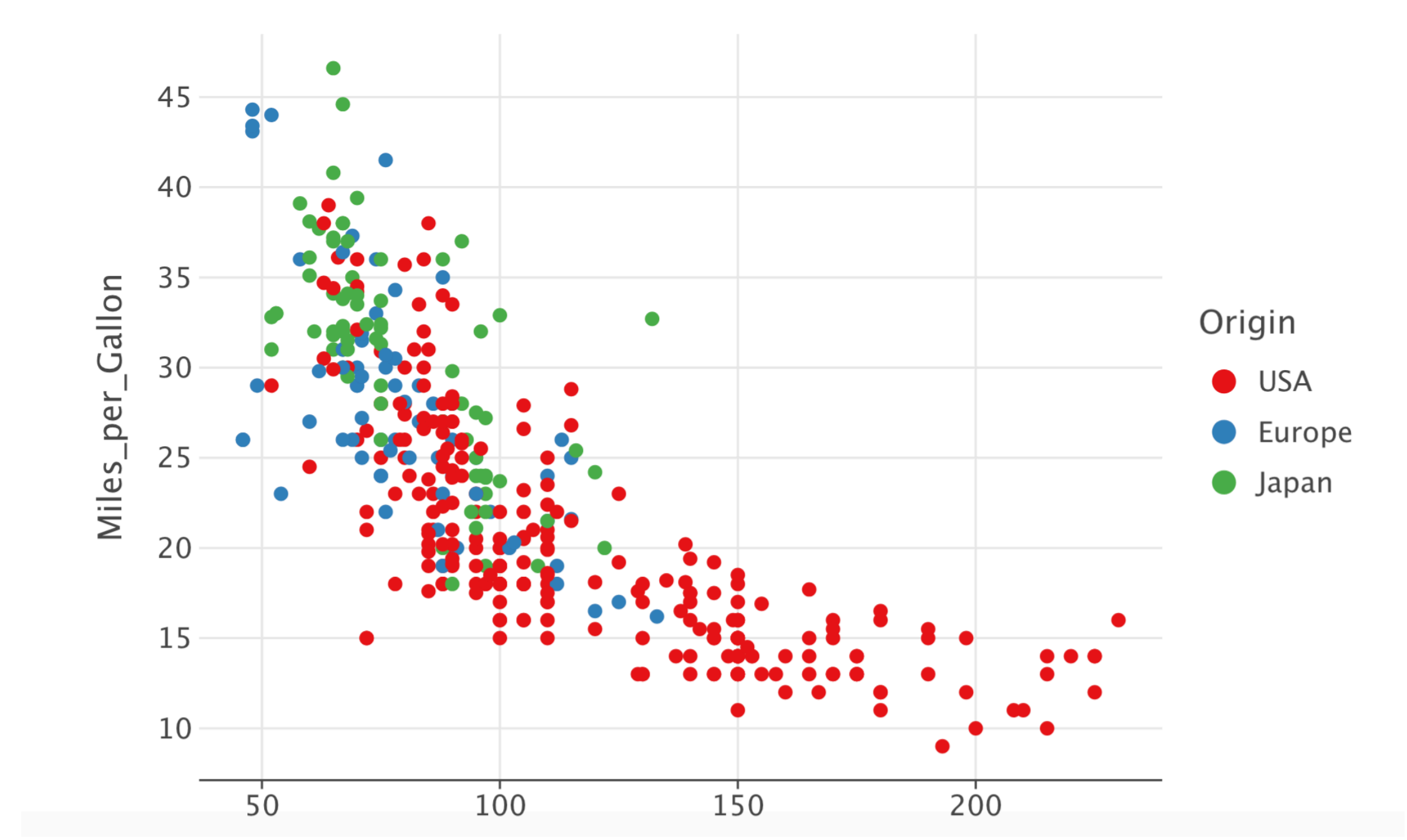


A bit better

Express all the facts: Axes and legends

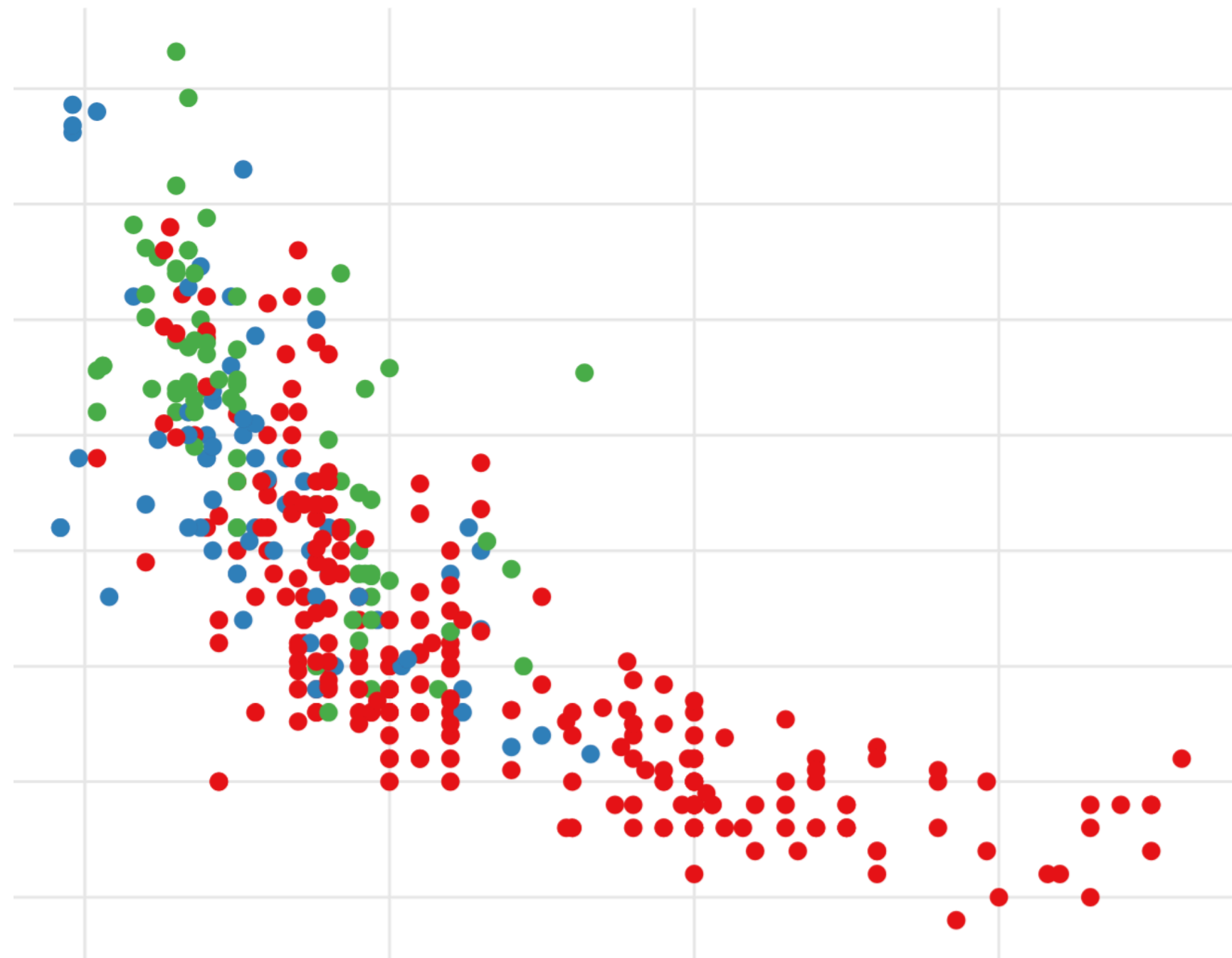


Not expressive

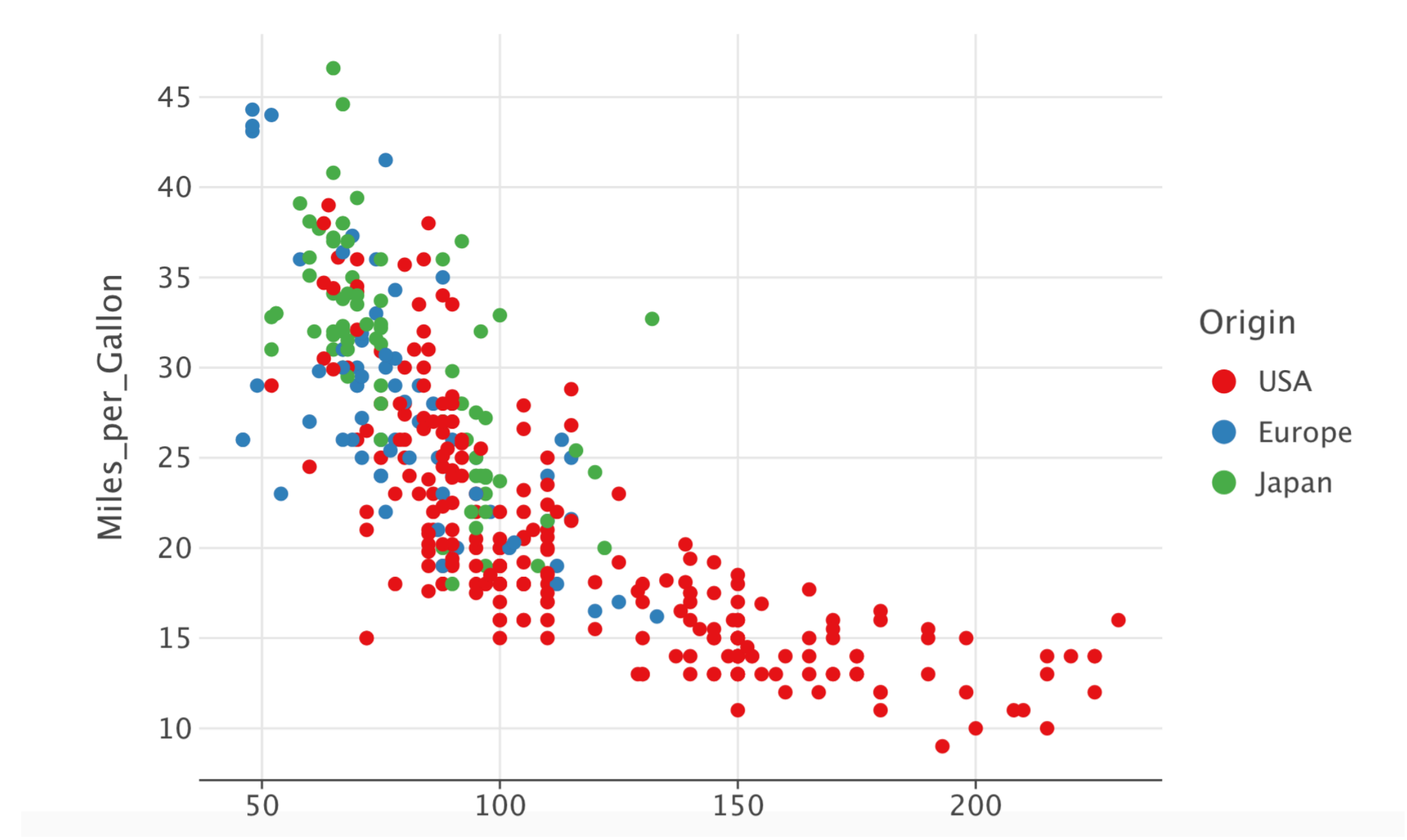


A bit better

Express all the facts: Axes and legends

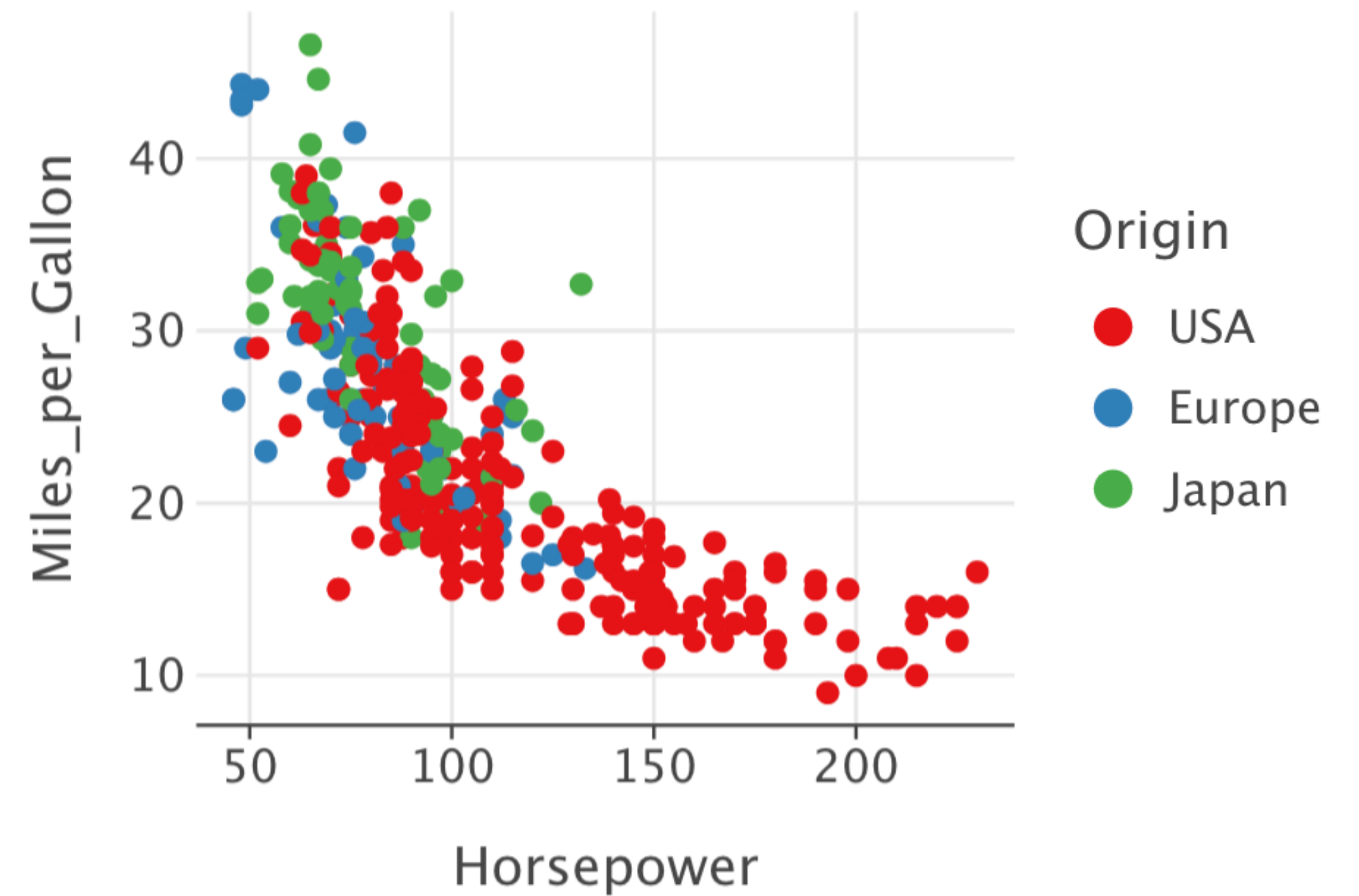
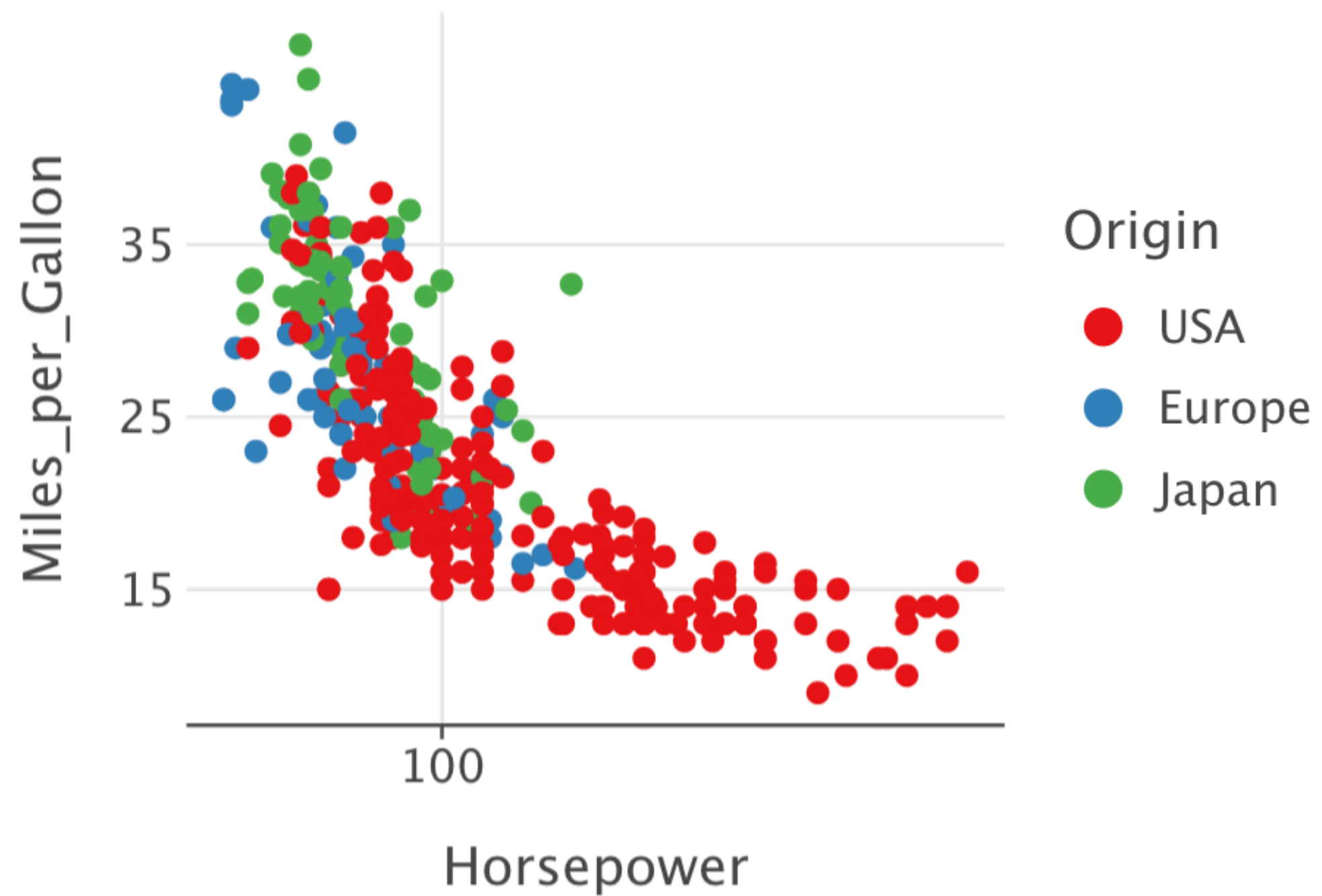


Not expressive



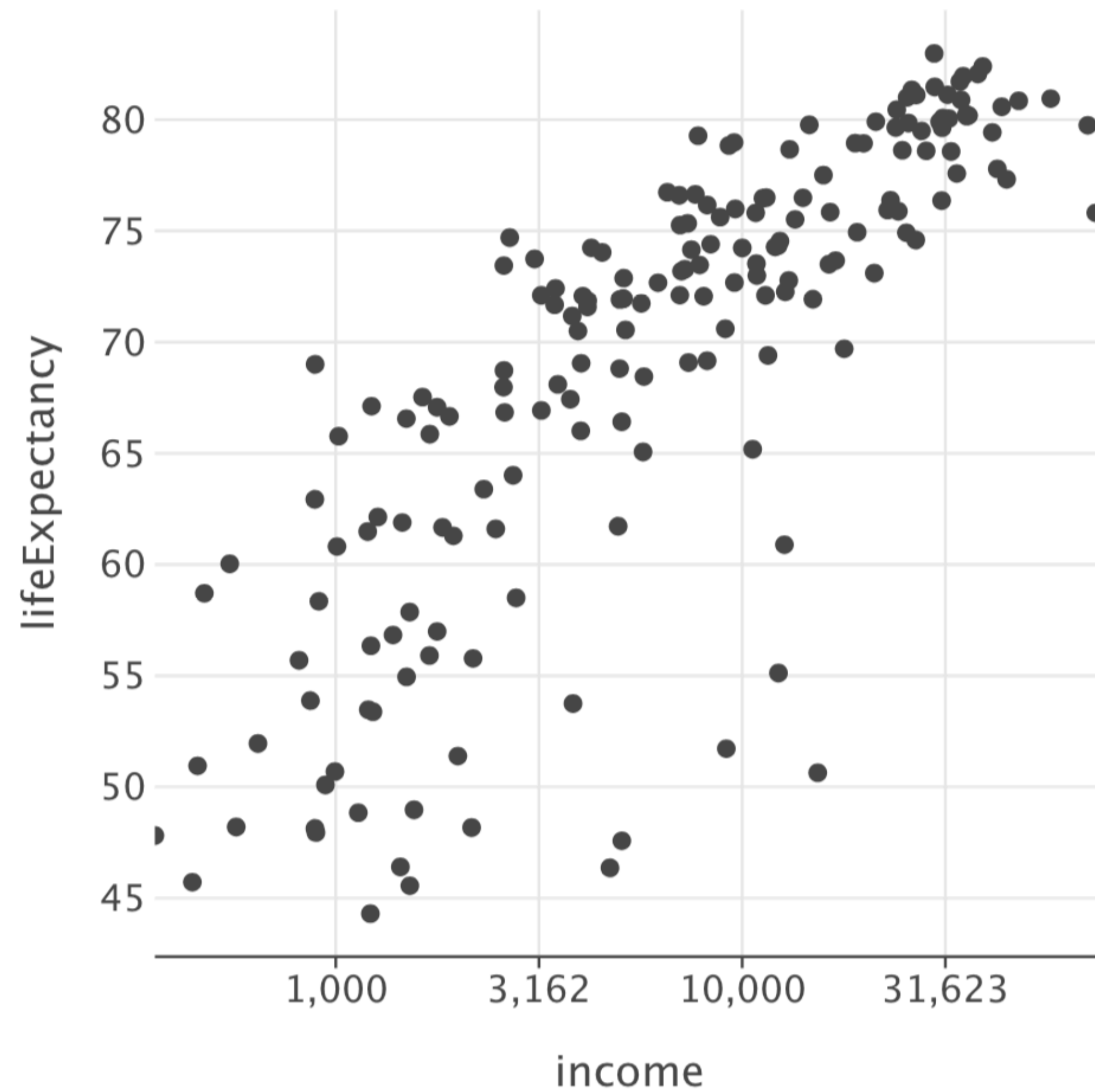
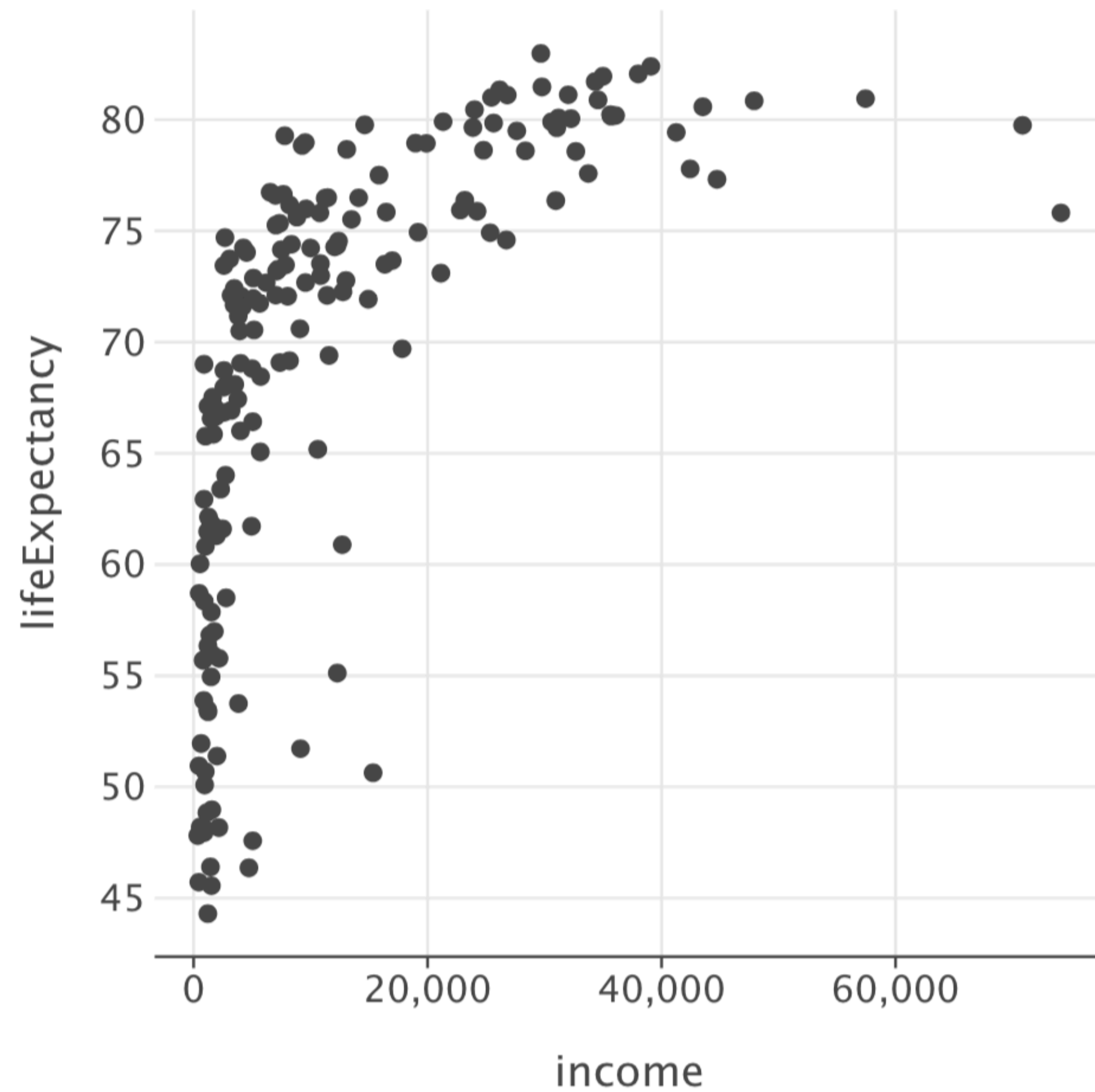
A bit better

Express all the facts: Axes and legends

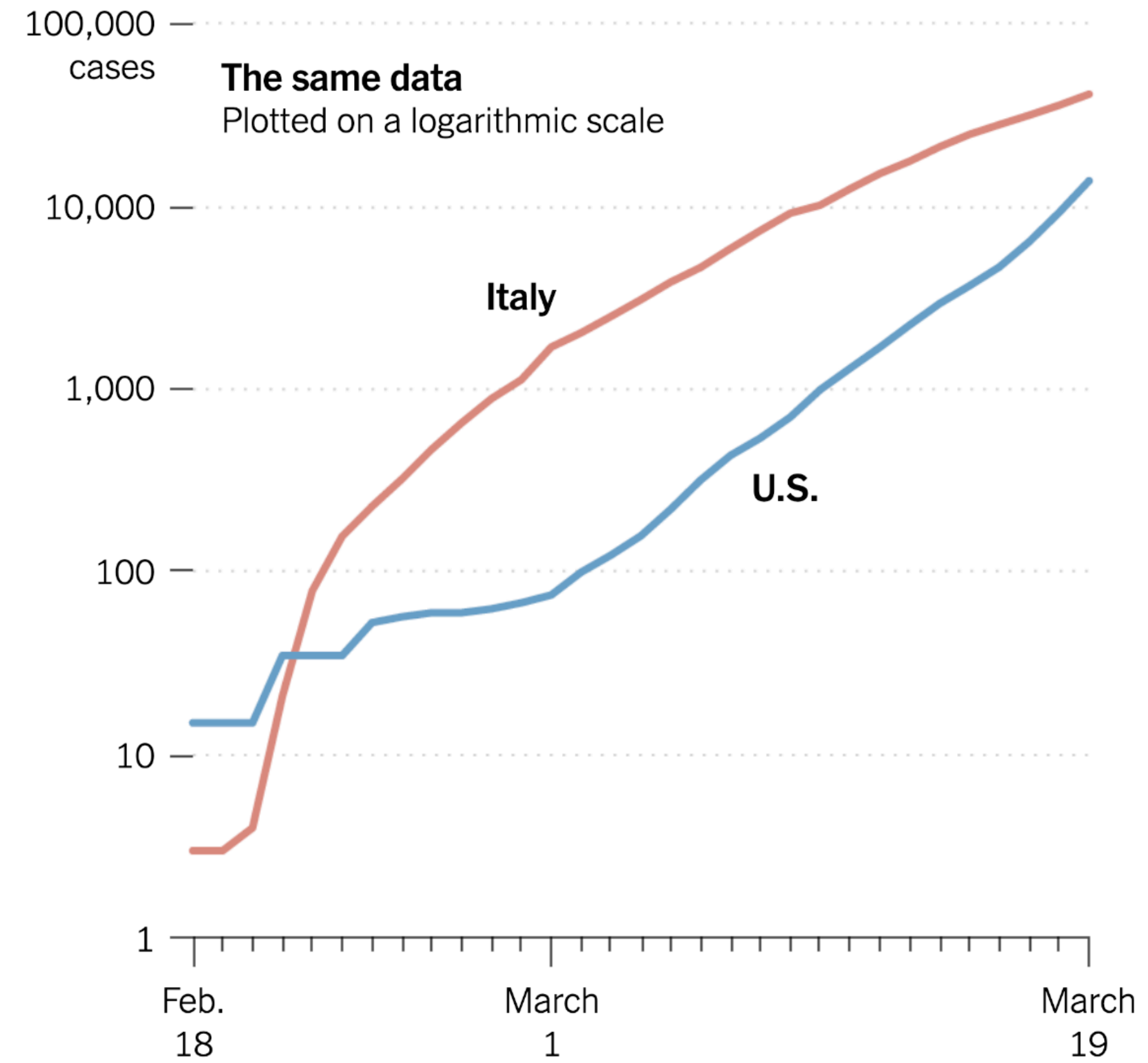
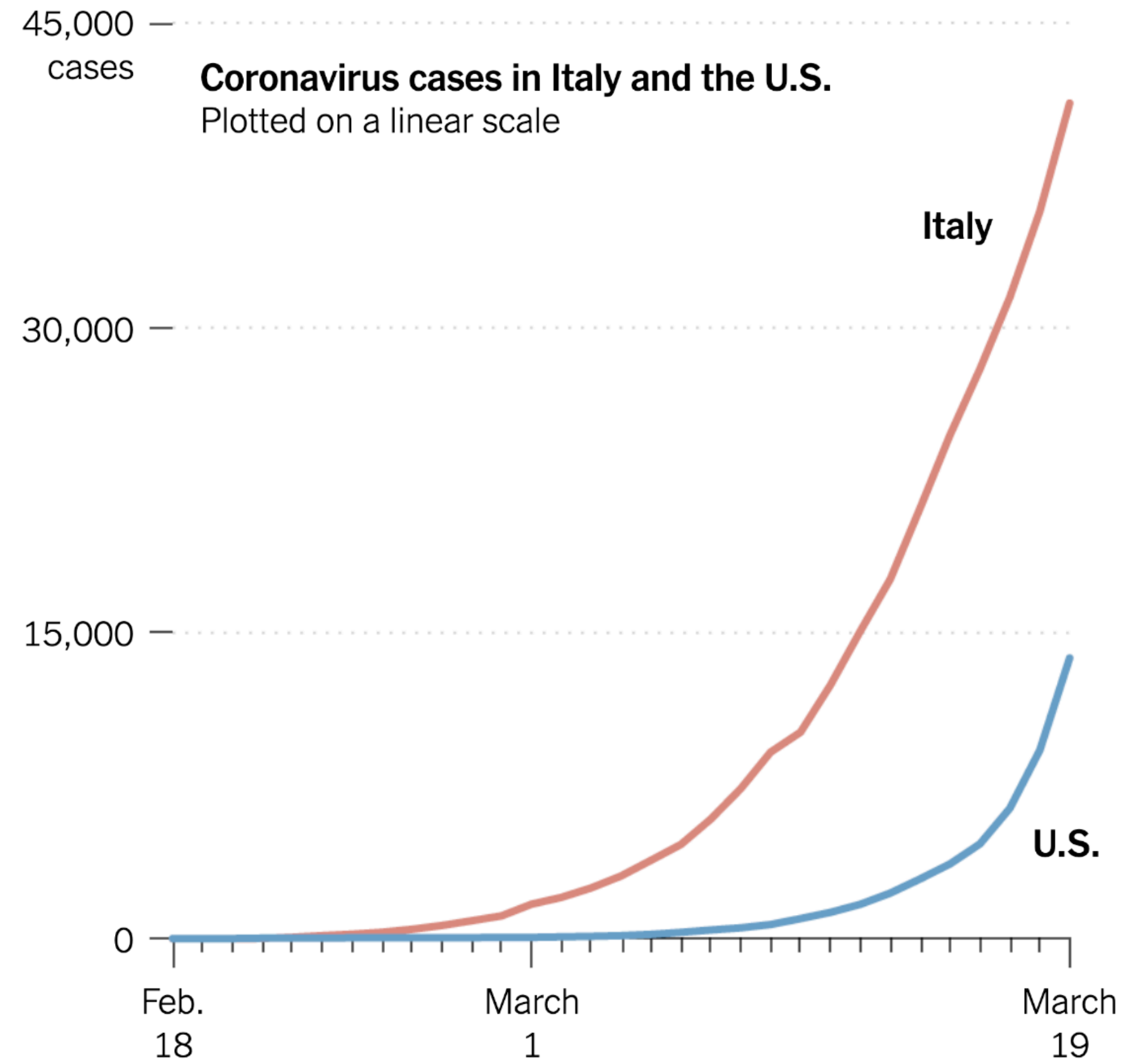


What's the issue here?

Express all the facts: scales



Express all the facts: scales



Effectiveness

Effectiveness: How do we choose good aesthetic mappings?

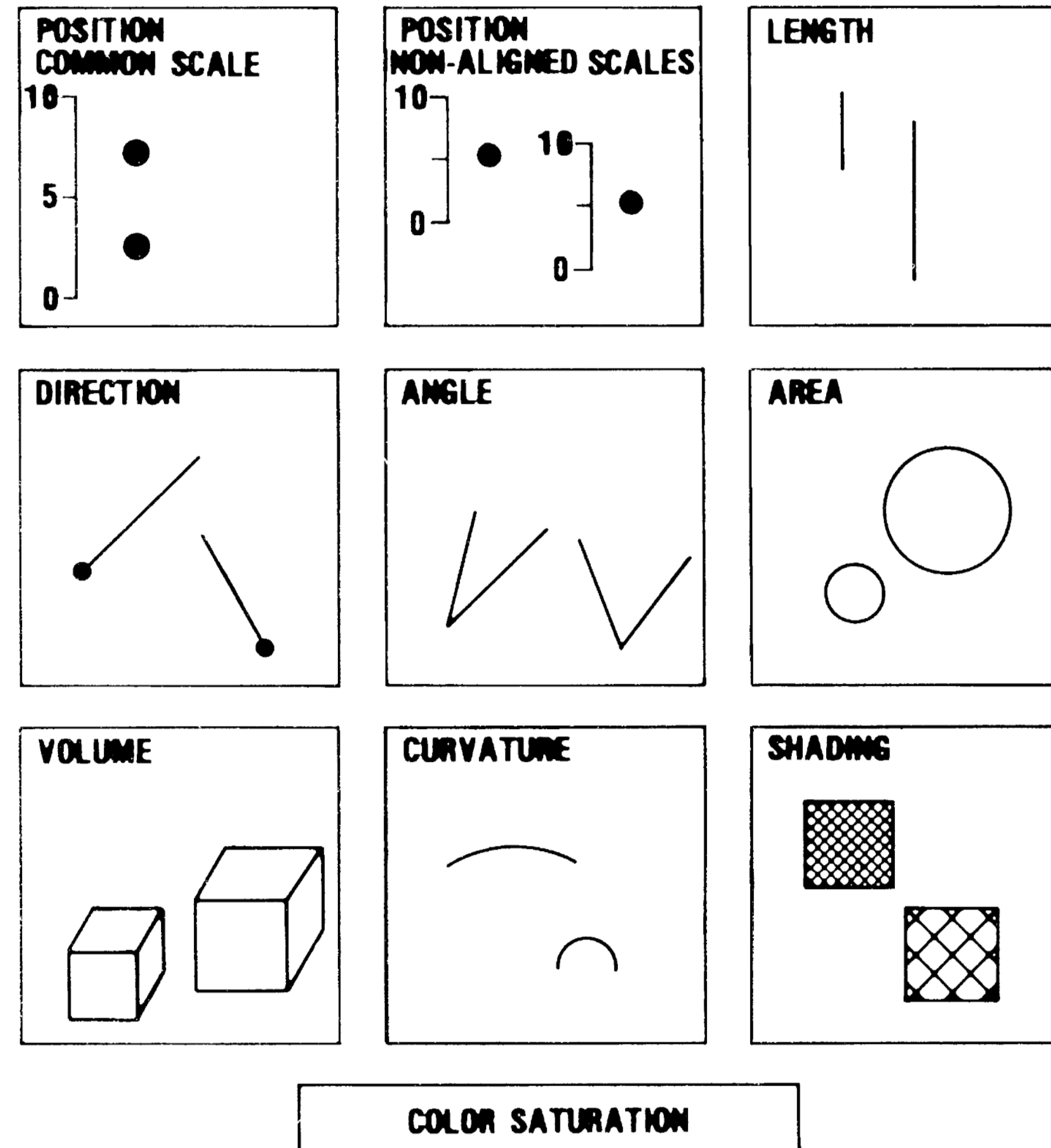
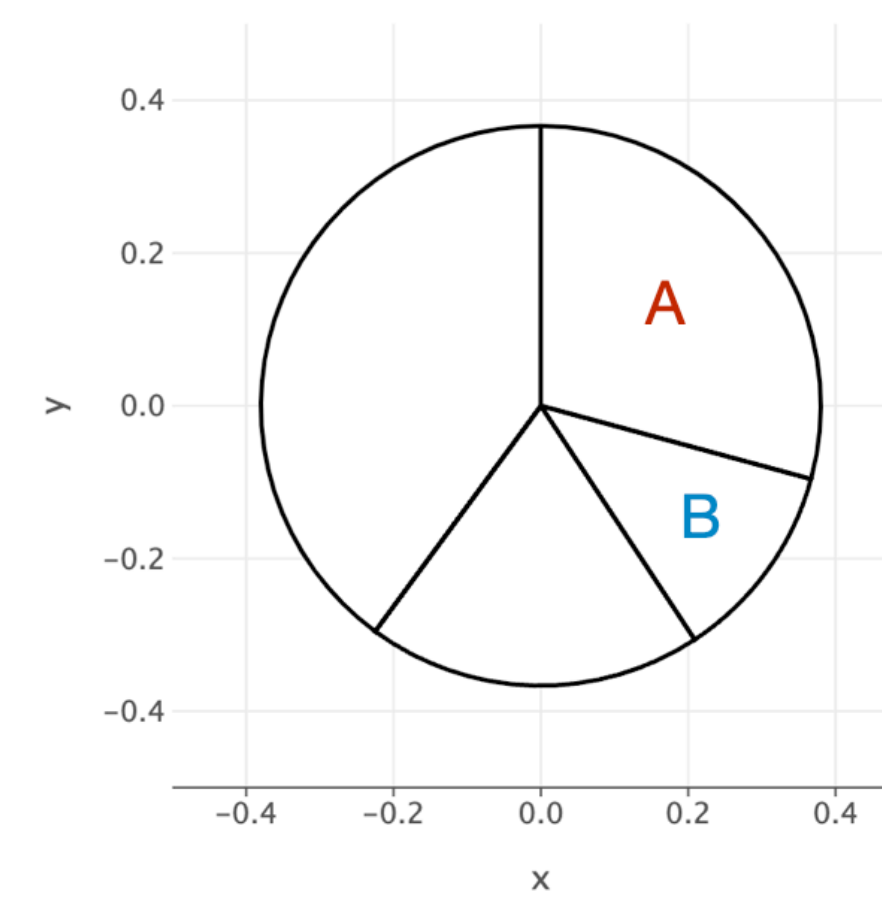
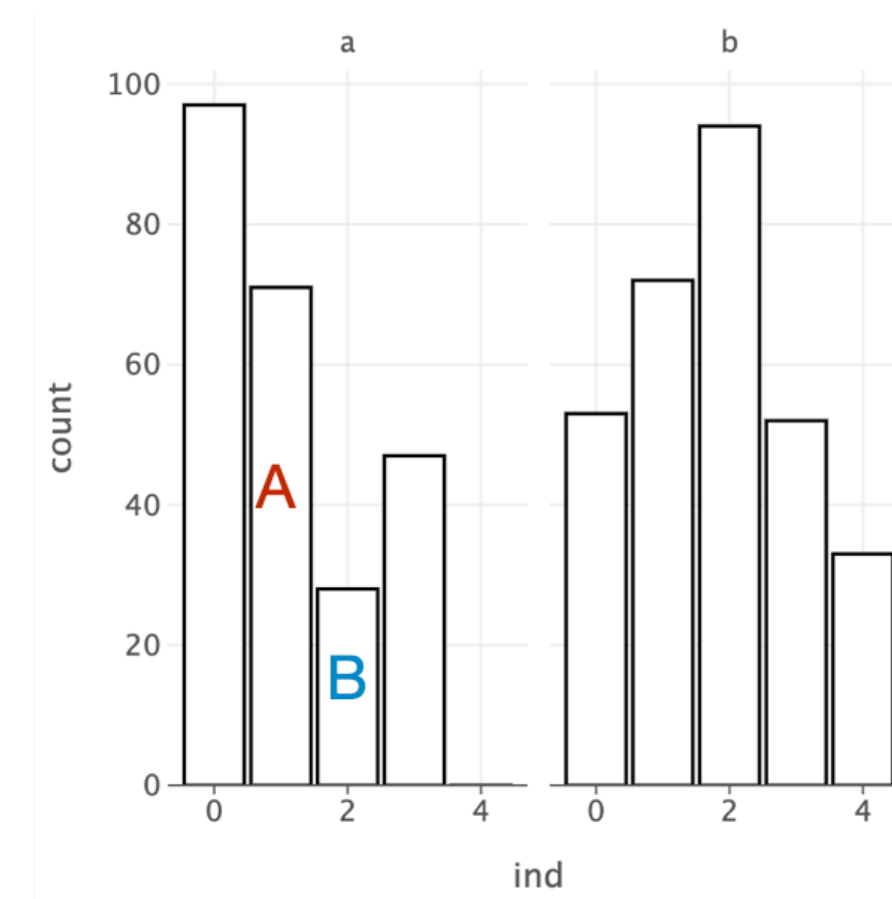
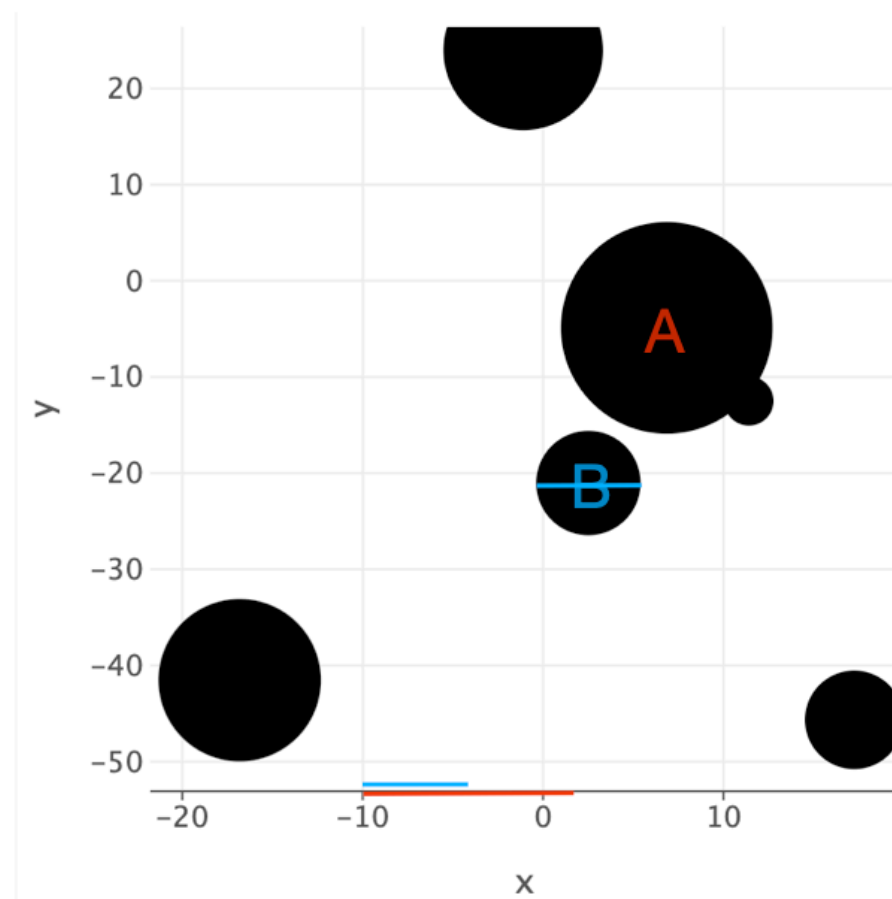
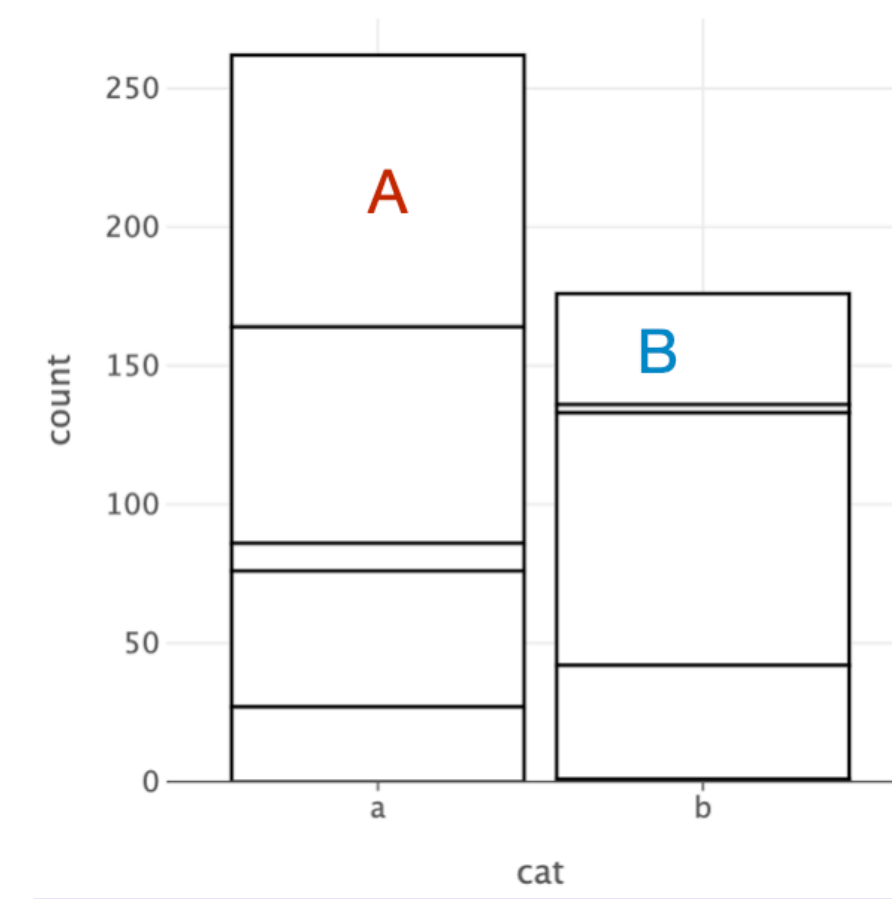
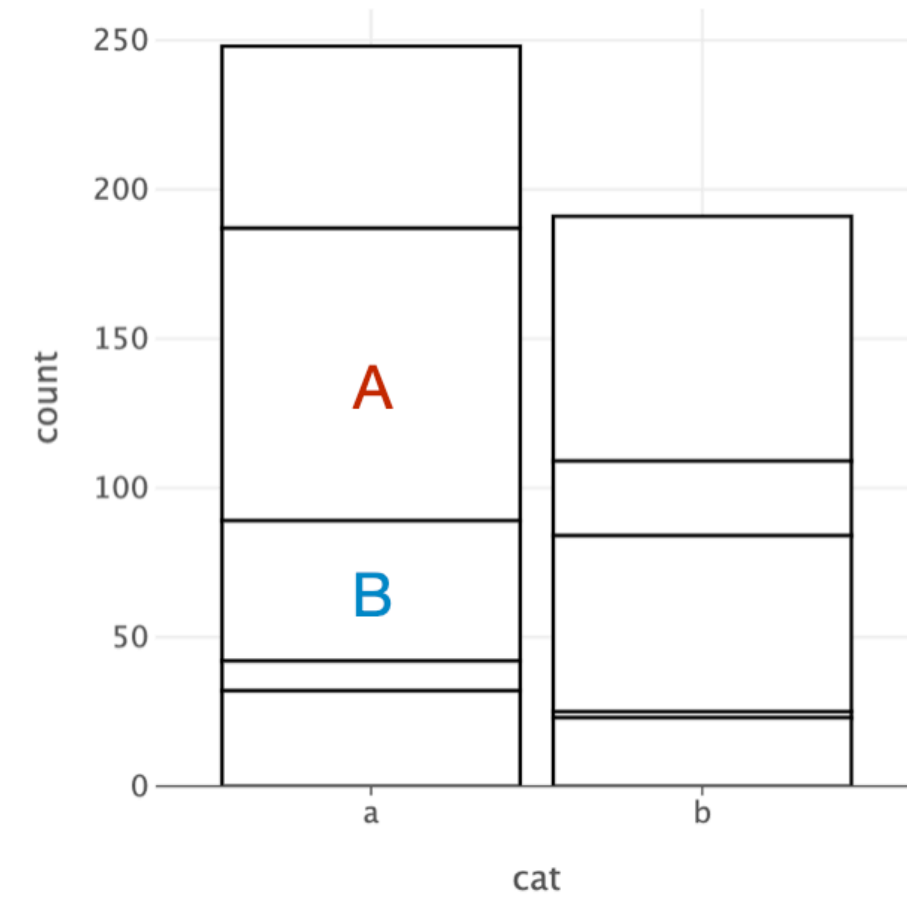
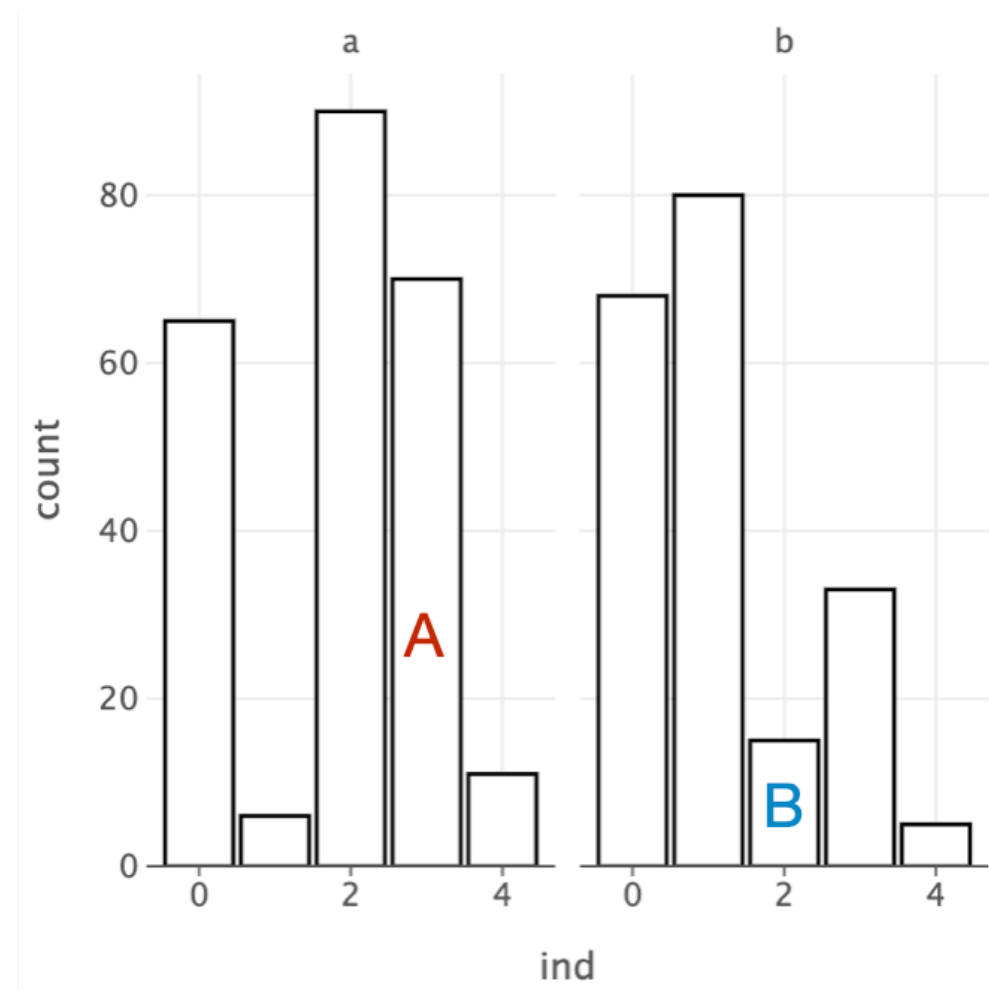


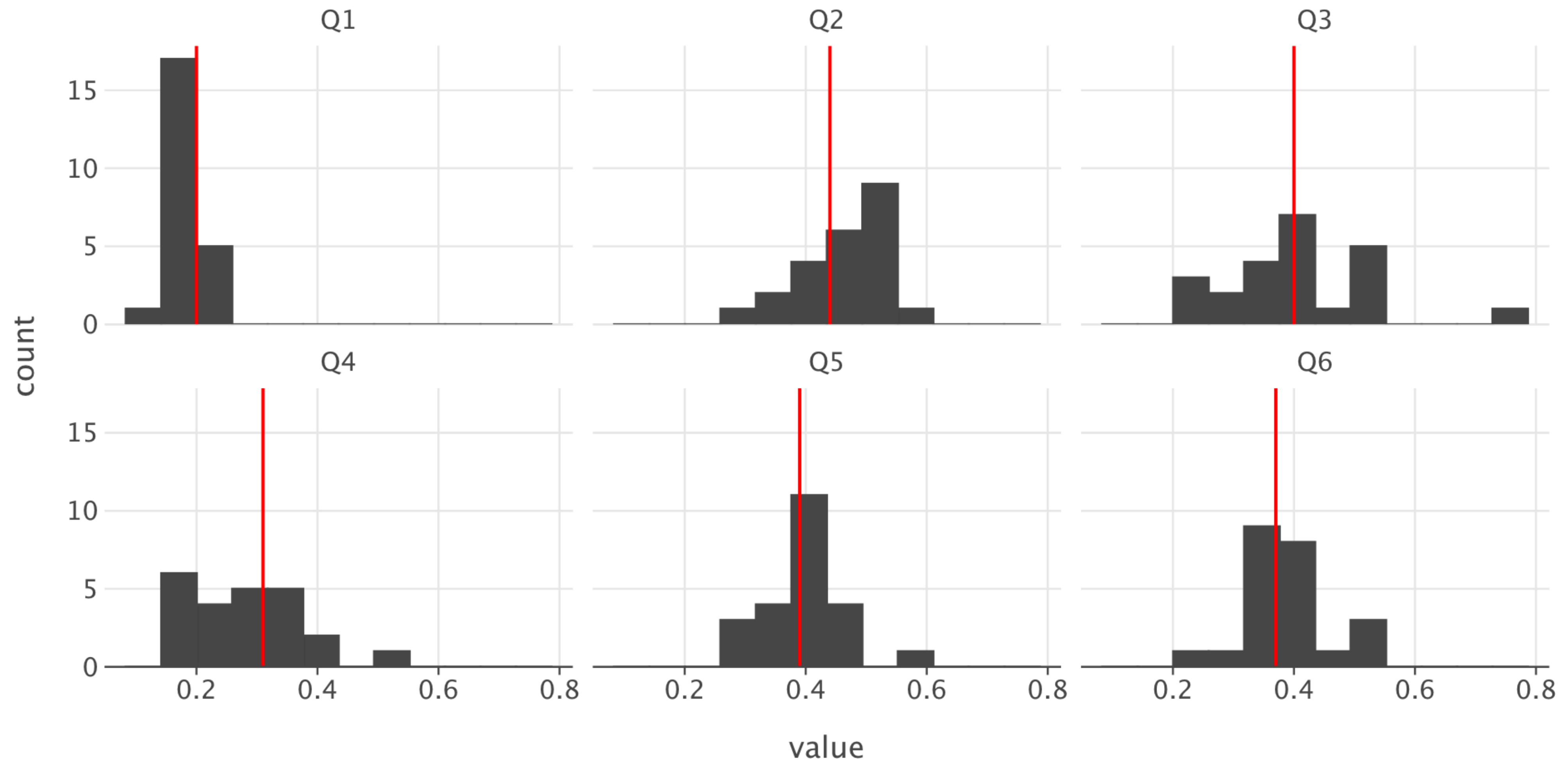
Figure 1. Elementary perceptual tasks.

Exercise: Survey

What did we find?

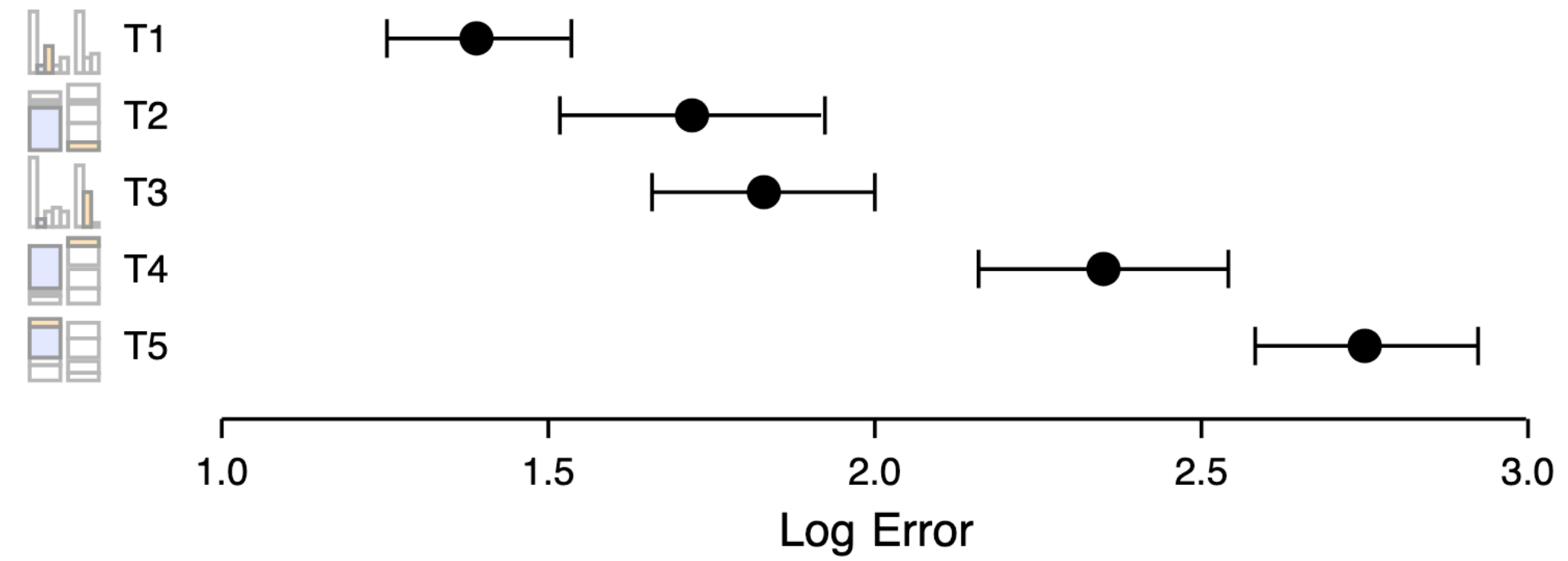


Last semester's results

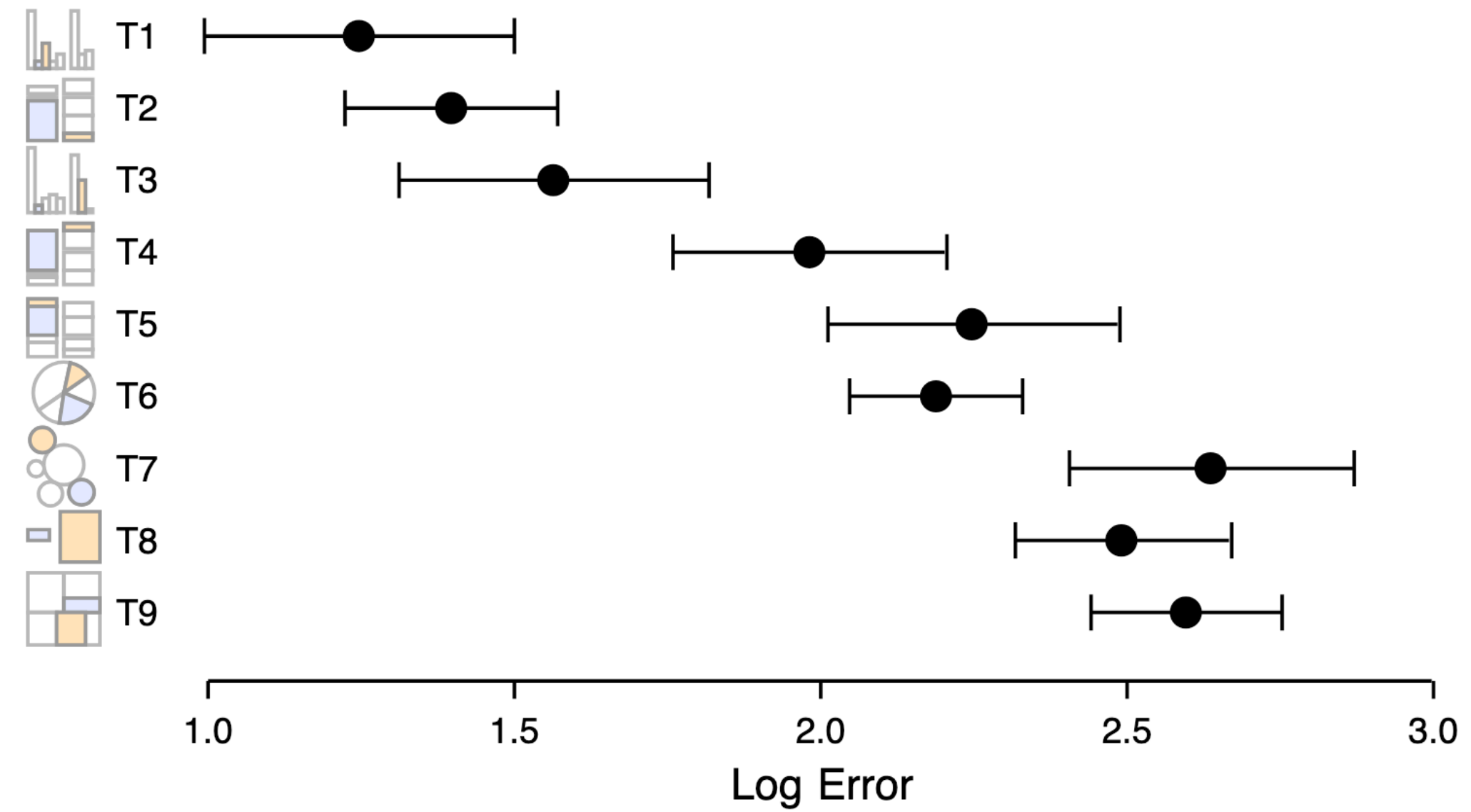


Proper studies

Cleveland & McGill's Results



Crowdsourced Results



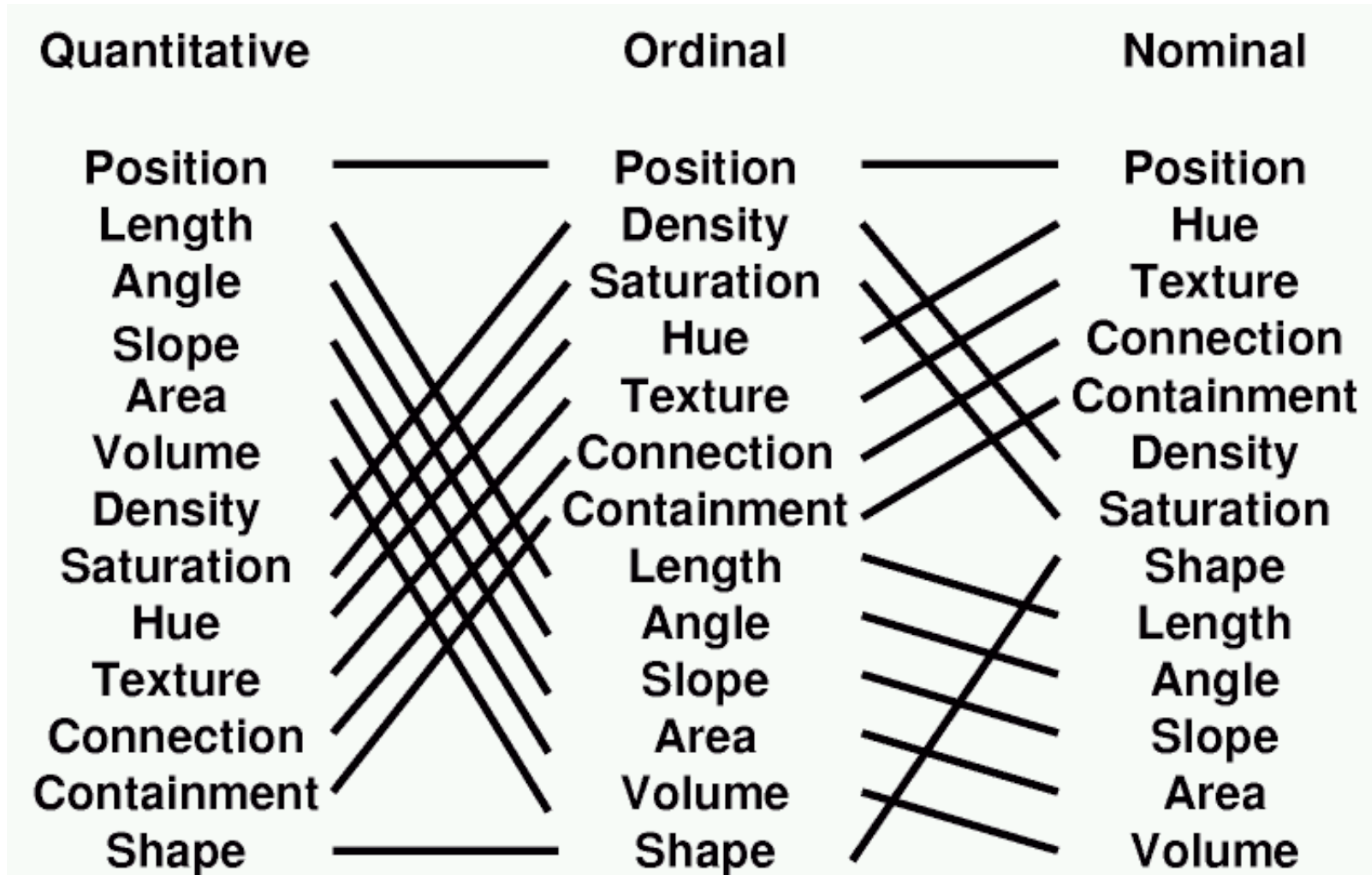
A first principle for effectiveness

Importance ordering

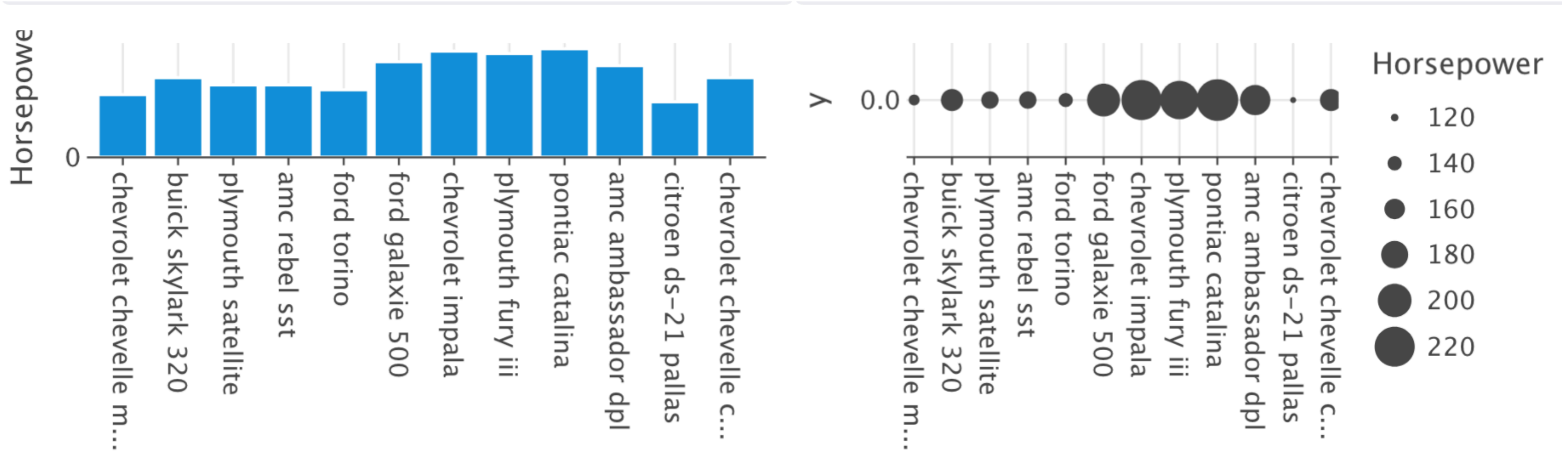
“Encode more important information more effectively”

- A visualization is more **effective** than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.

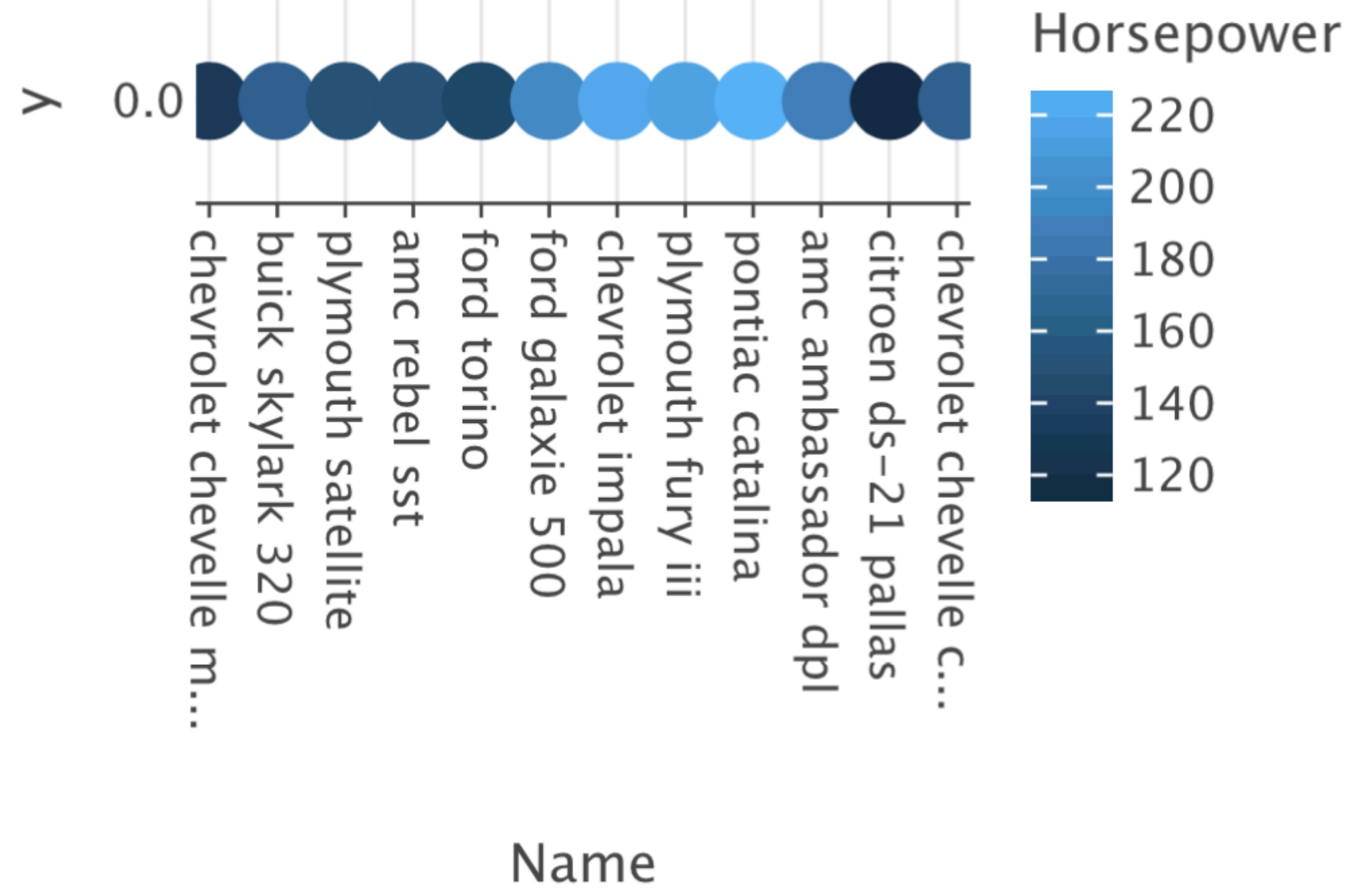
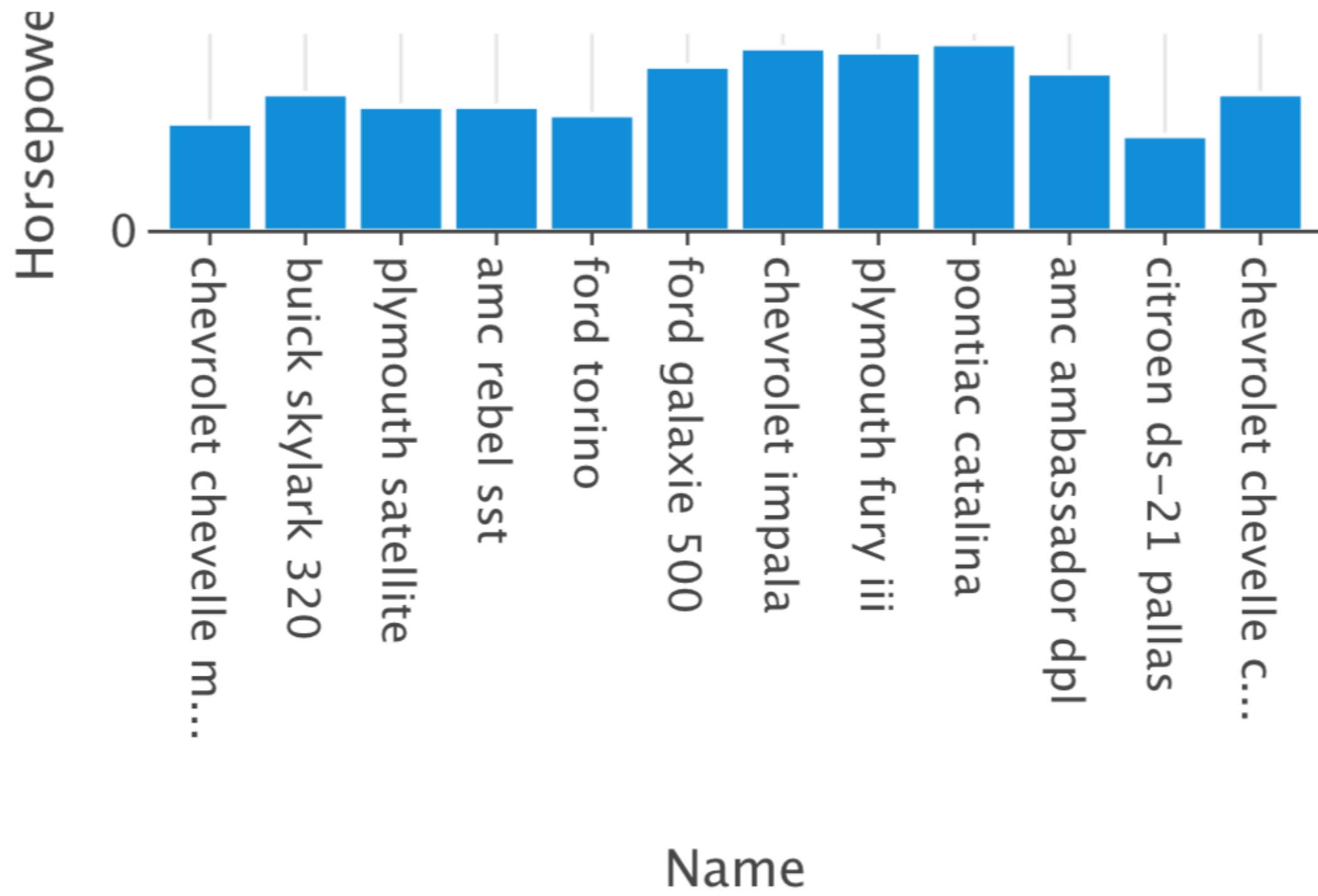
MacKinley's Ranking



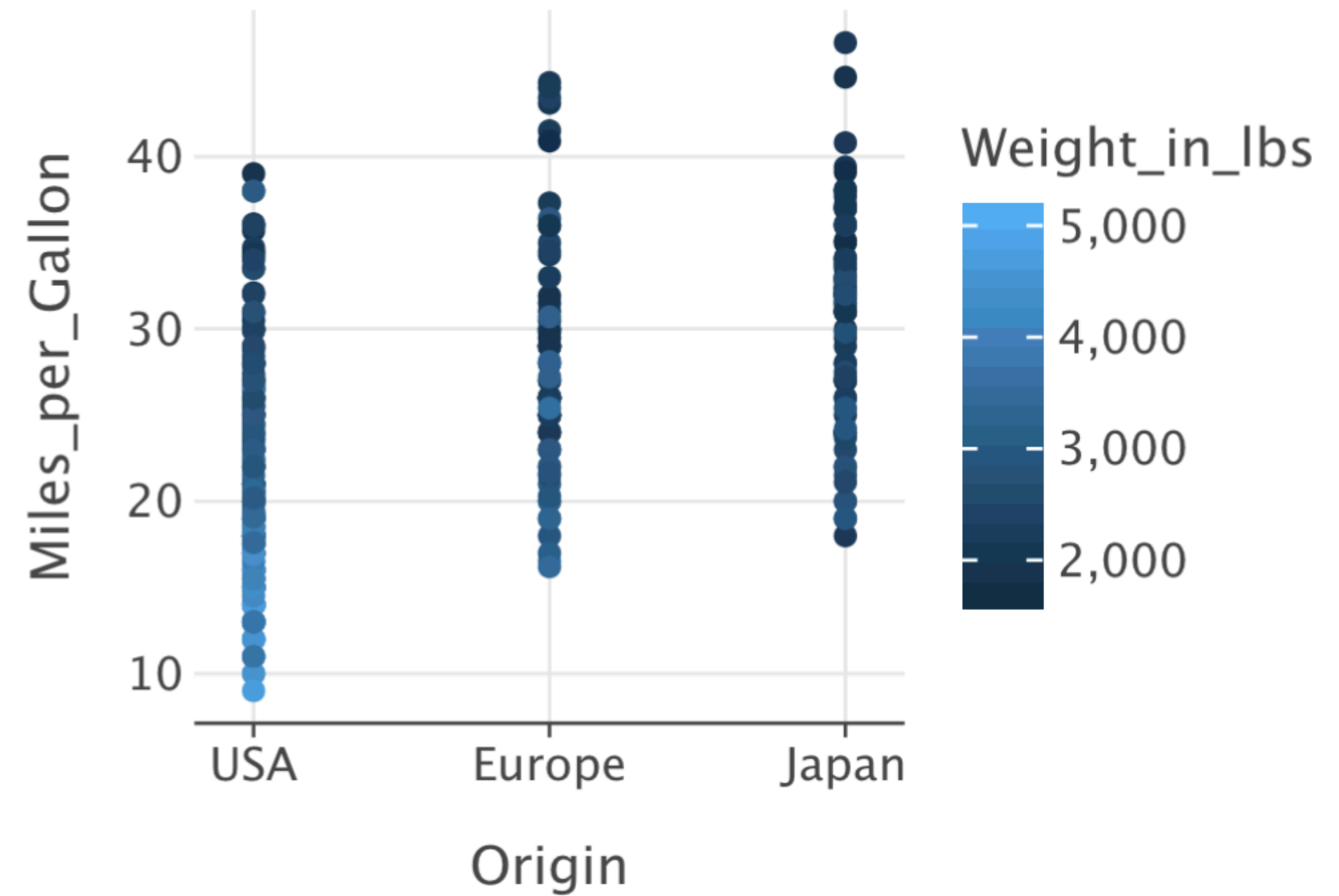
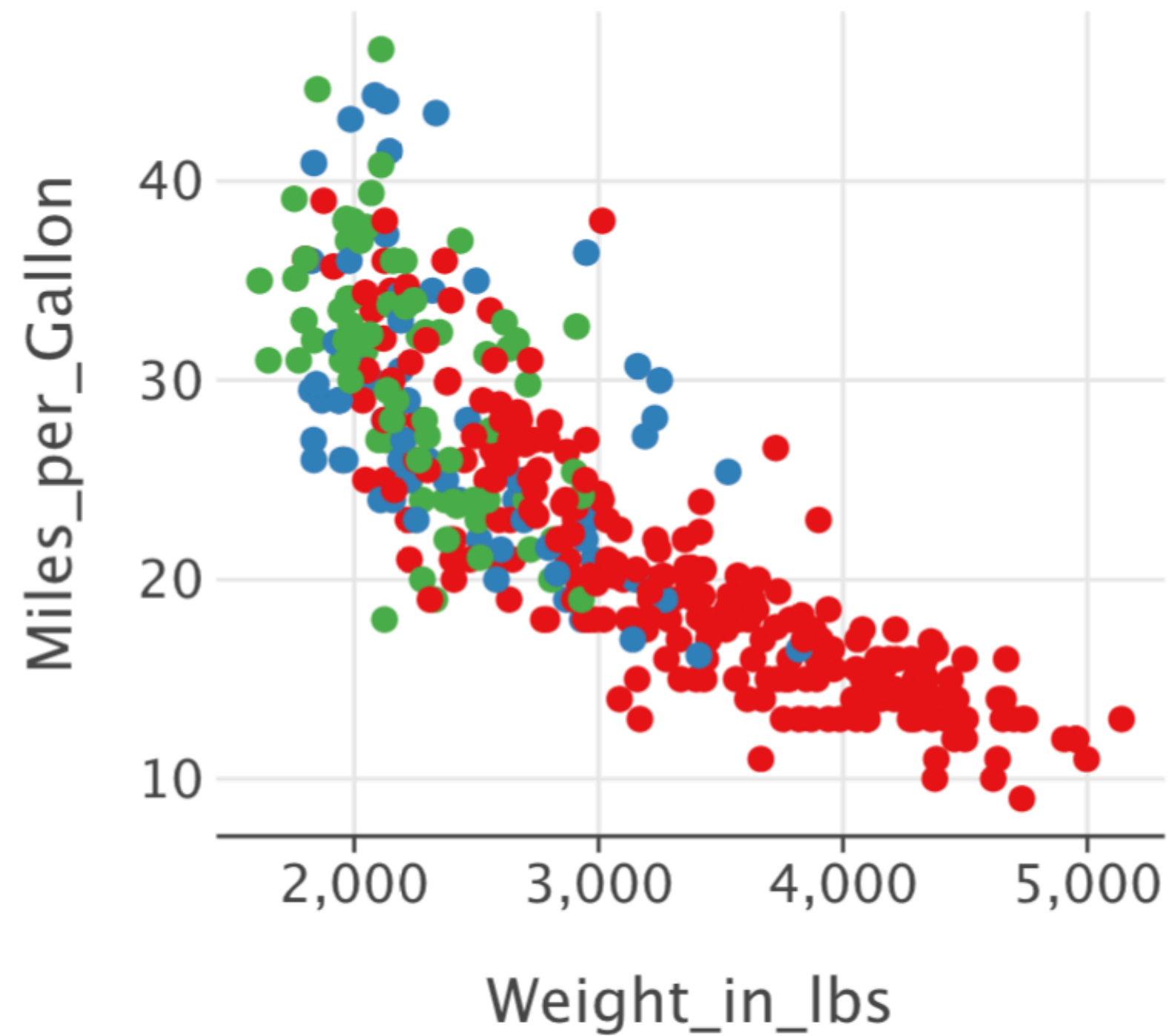
Single variable example



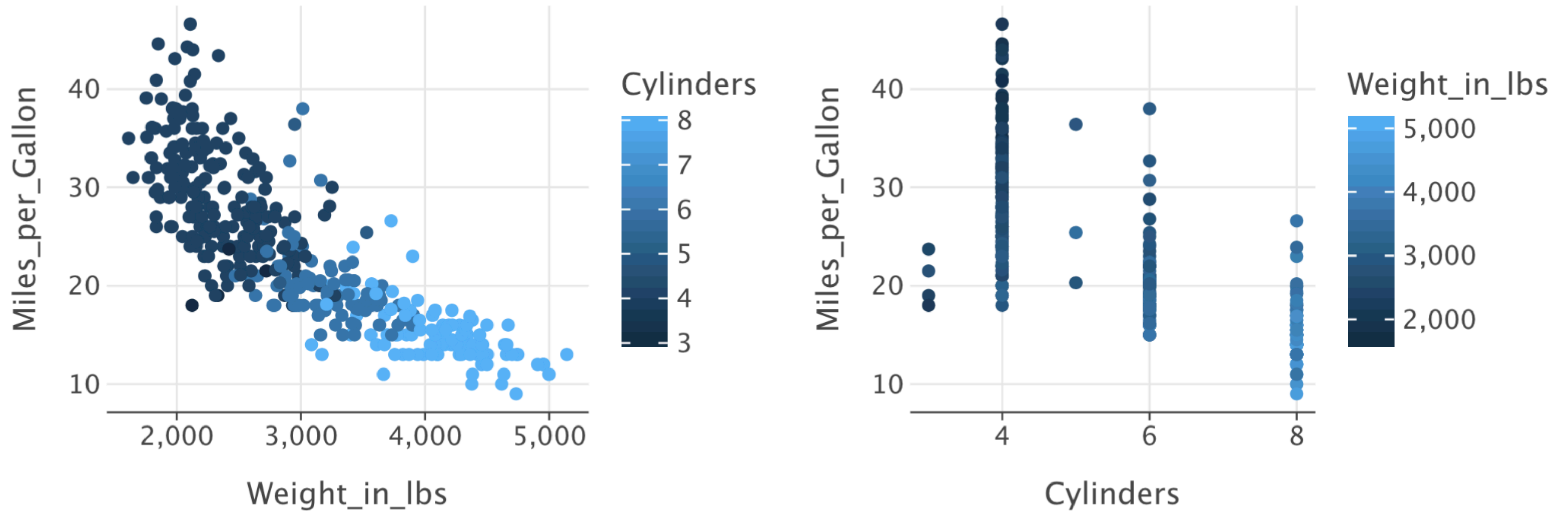
Single variable example



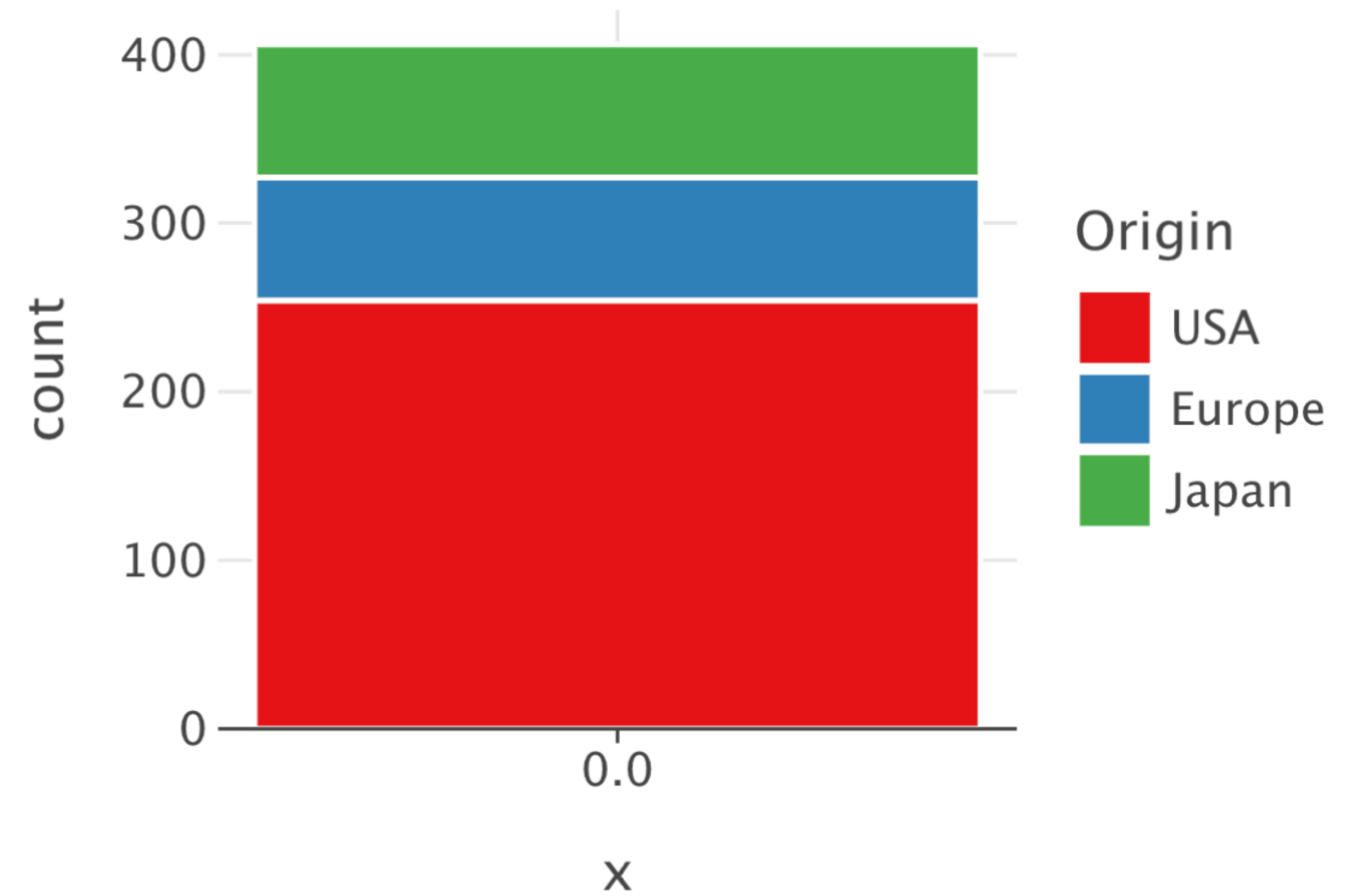
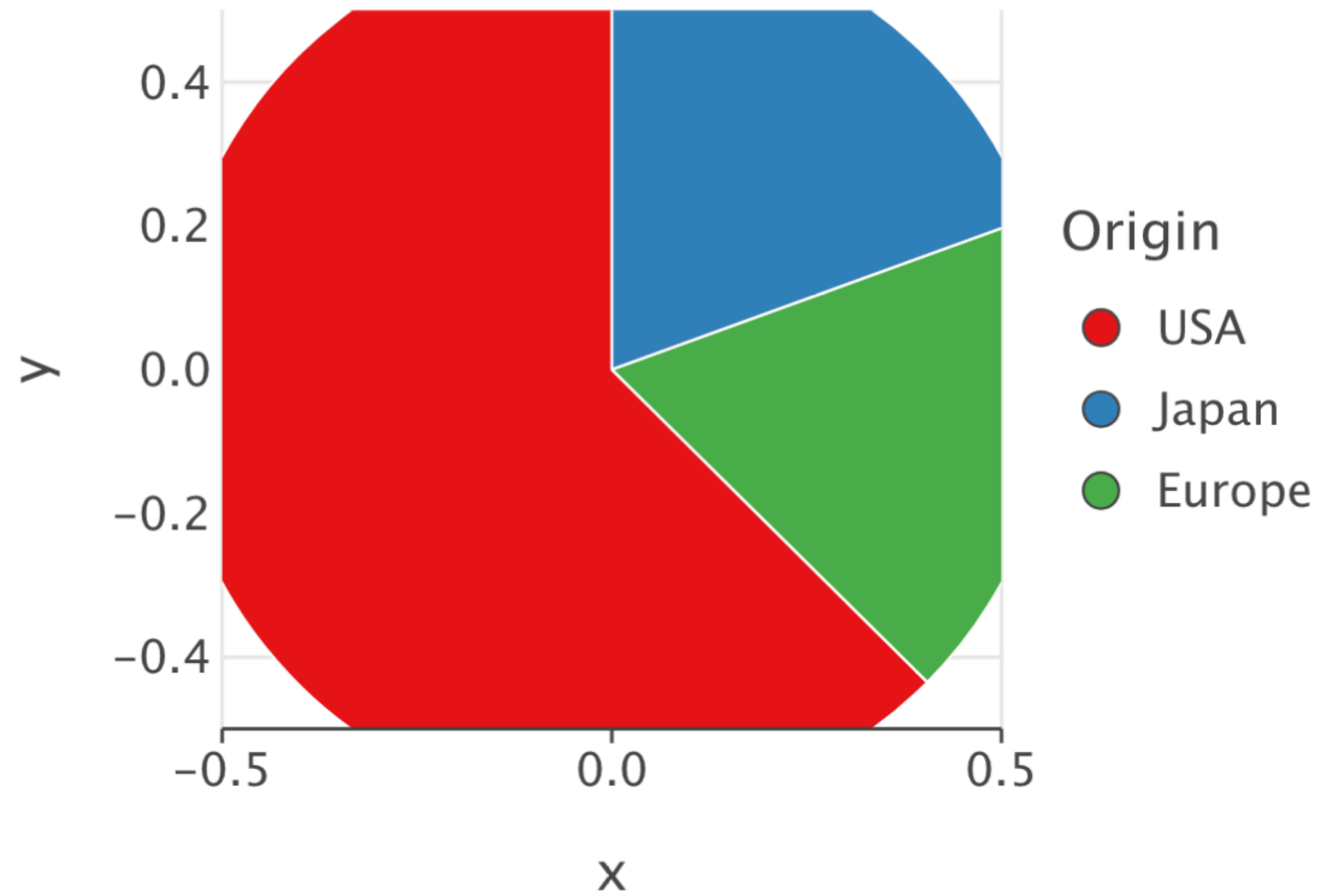
Multi-Variable example



Multi-Variable example



A debate



A solution

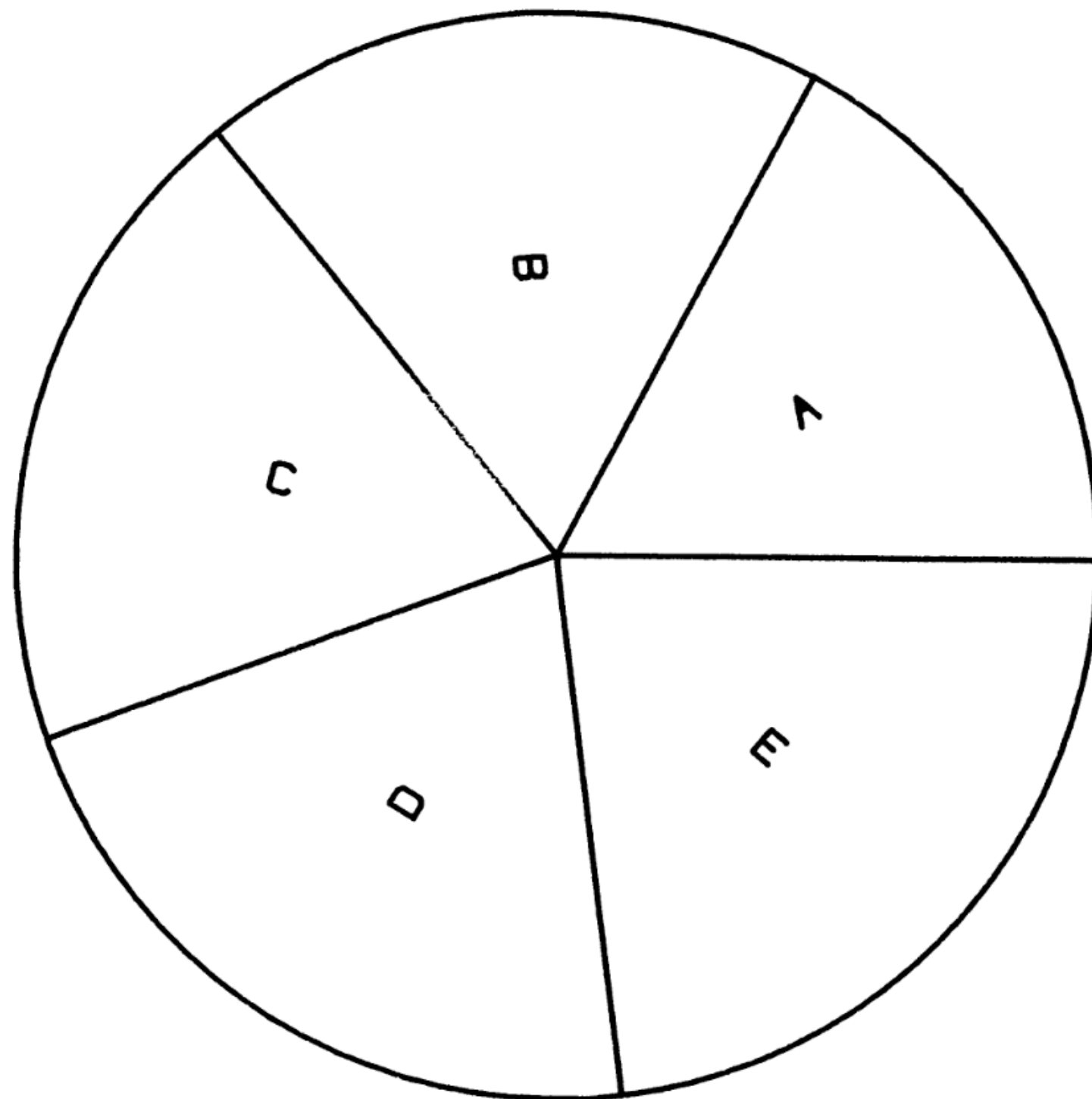


Figure 22. Pie chart.

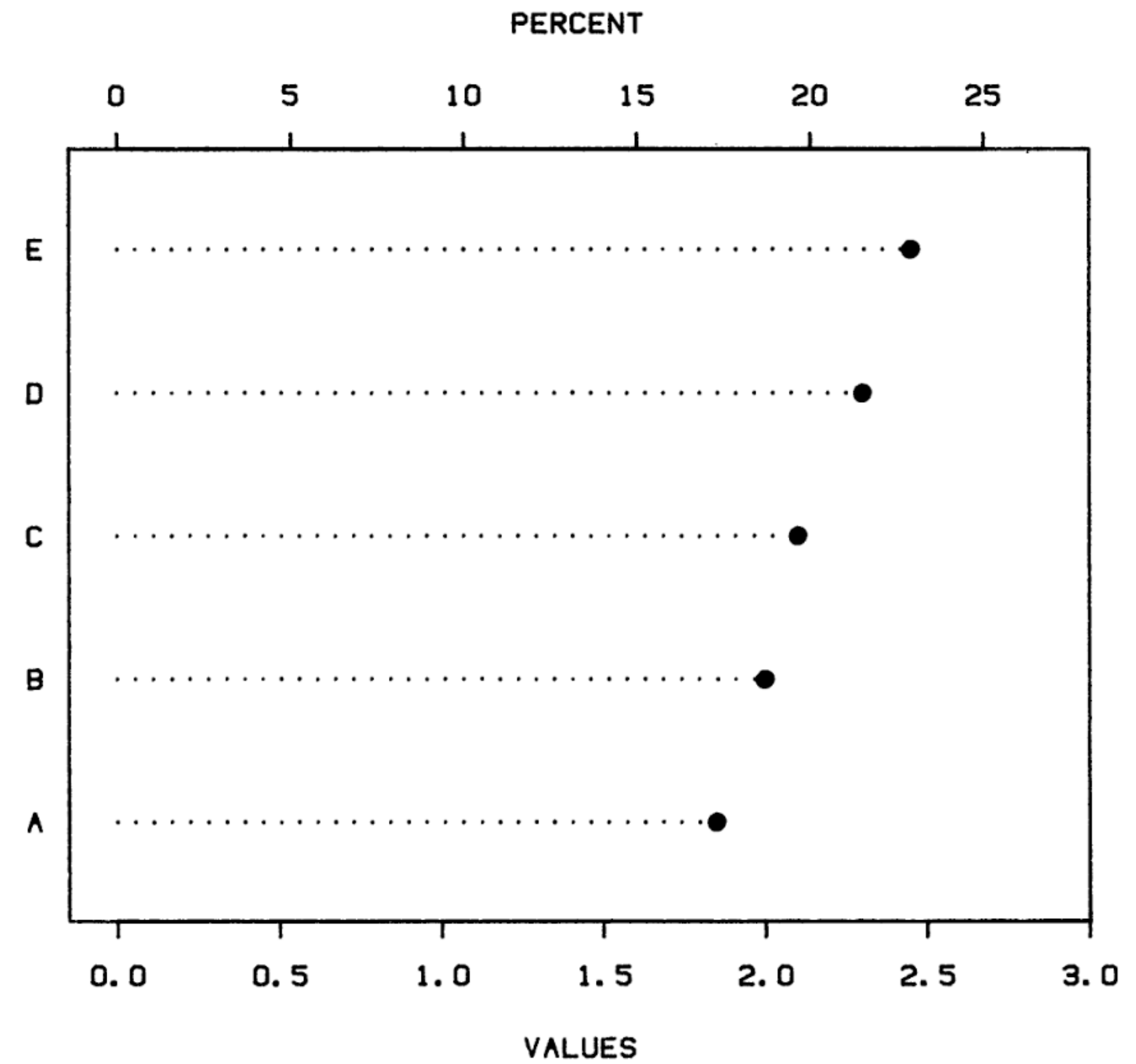


Figure 23. Dot chart.

Caveat: People don't read by angle!

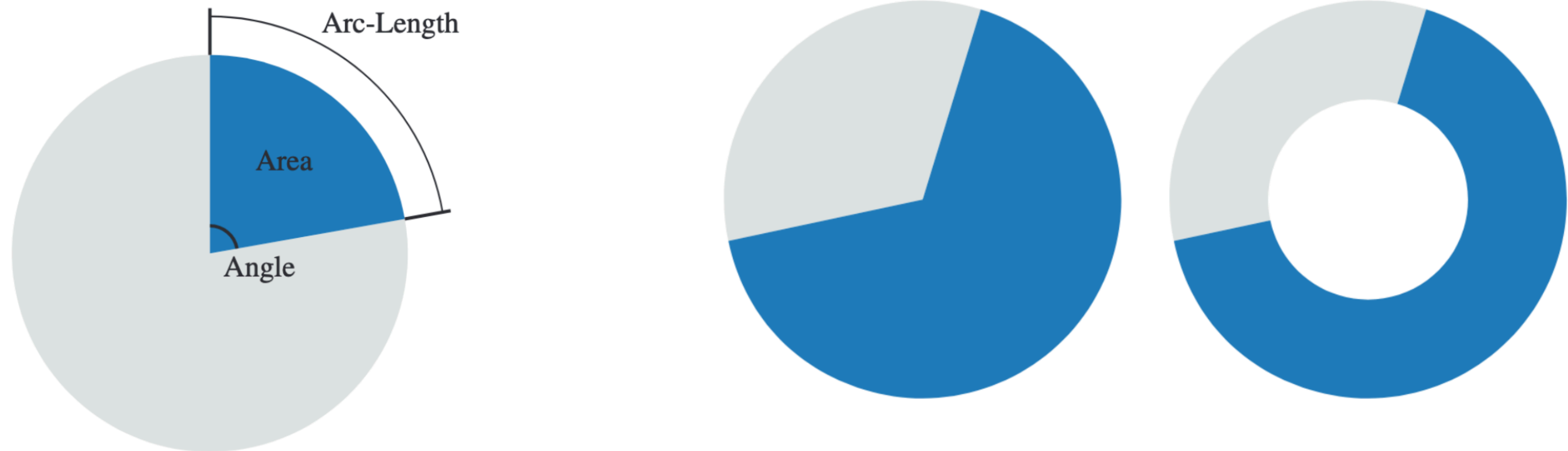
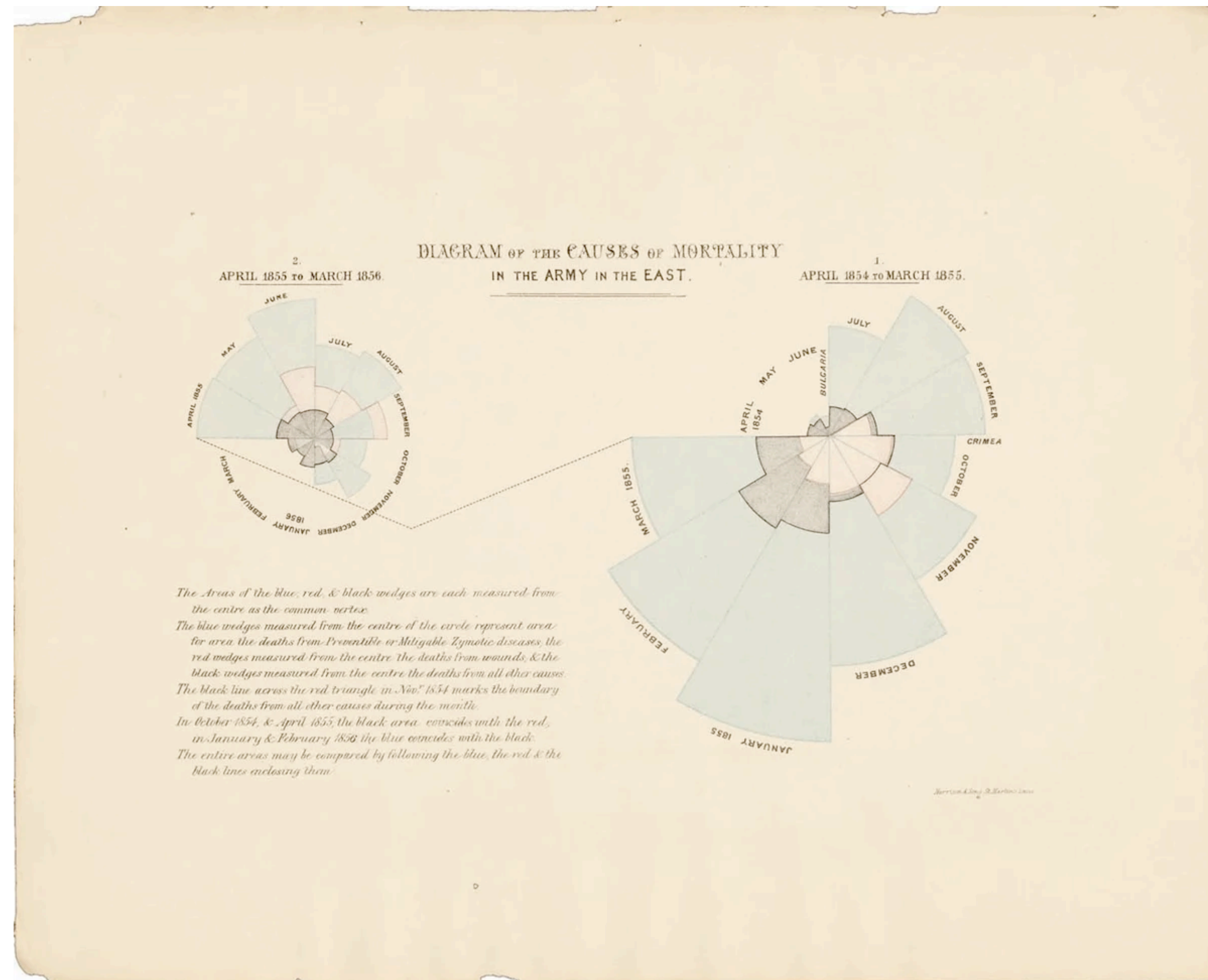


Figure 2: The three different encodings representing data in a pie or donut chart: central angle, wedge area, and arc length.

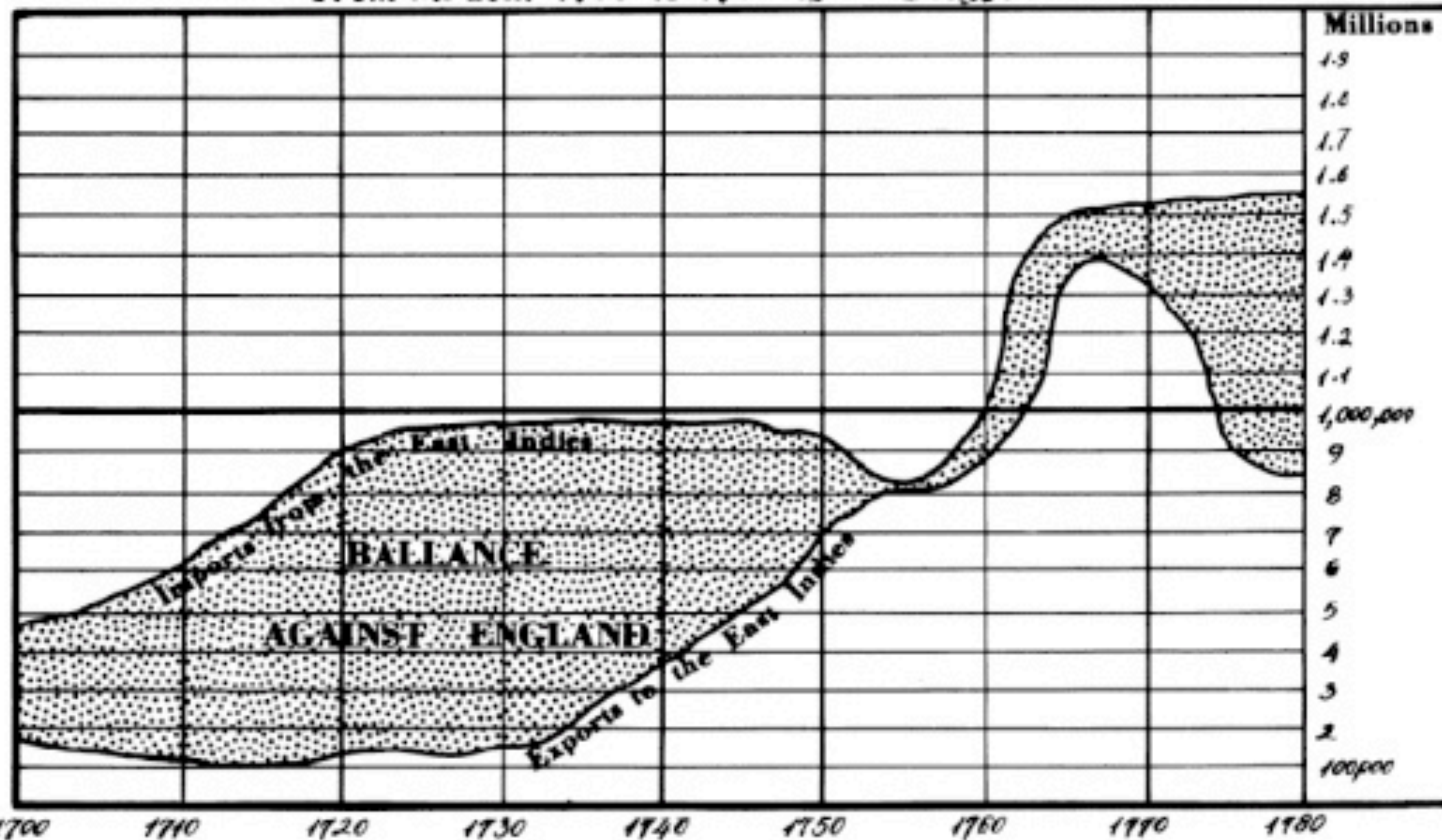
Thus donut charts can be just as effective!

Nightingale's Polar Area Charts



William Playfair's plot

CHART of EXPORTS and IMPORTS to and from the EAST INDIES
From the Year 1700 to 1780 by W. Playfair



The Bottom Line is Divided into Years the Right hand Line into HUNDRED THOUSAND POUNDS
Simple & Right *is from Page 31st* *Published in the Act & since 16th Aug. 1785*

Which is more effective?

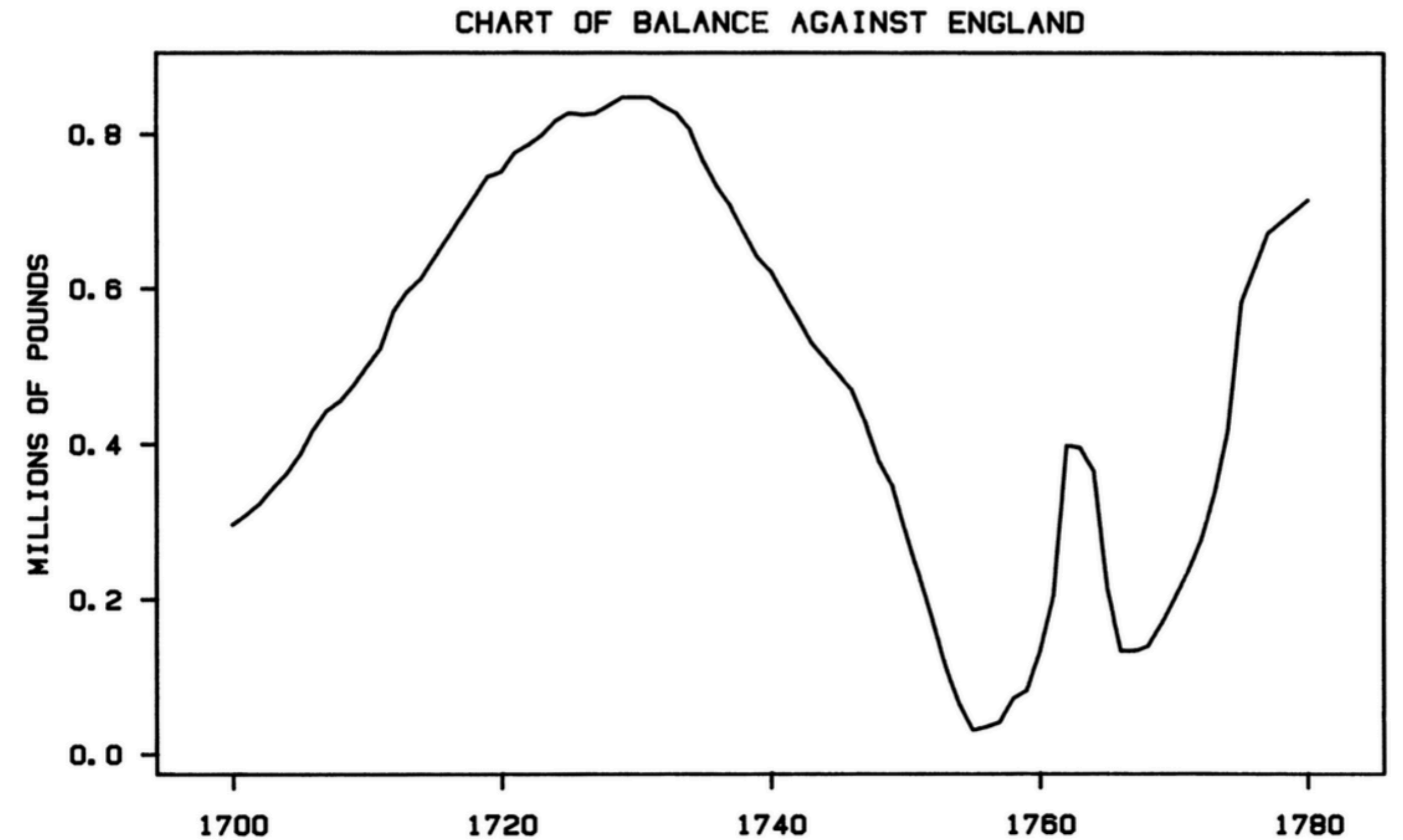
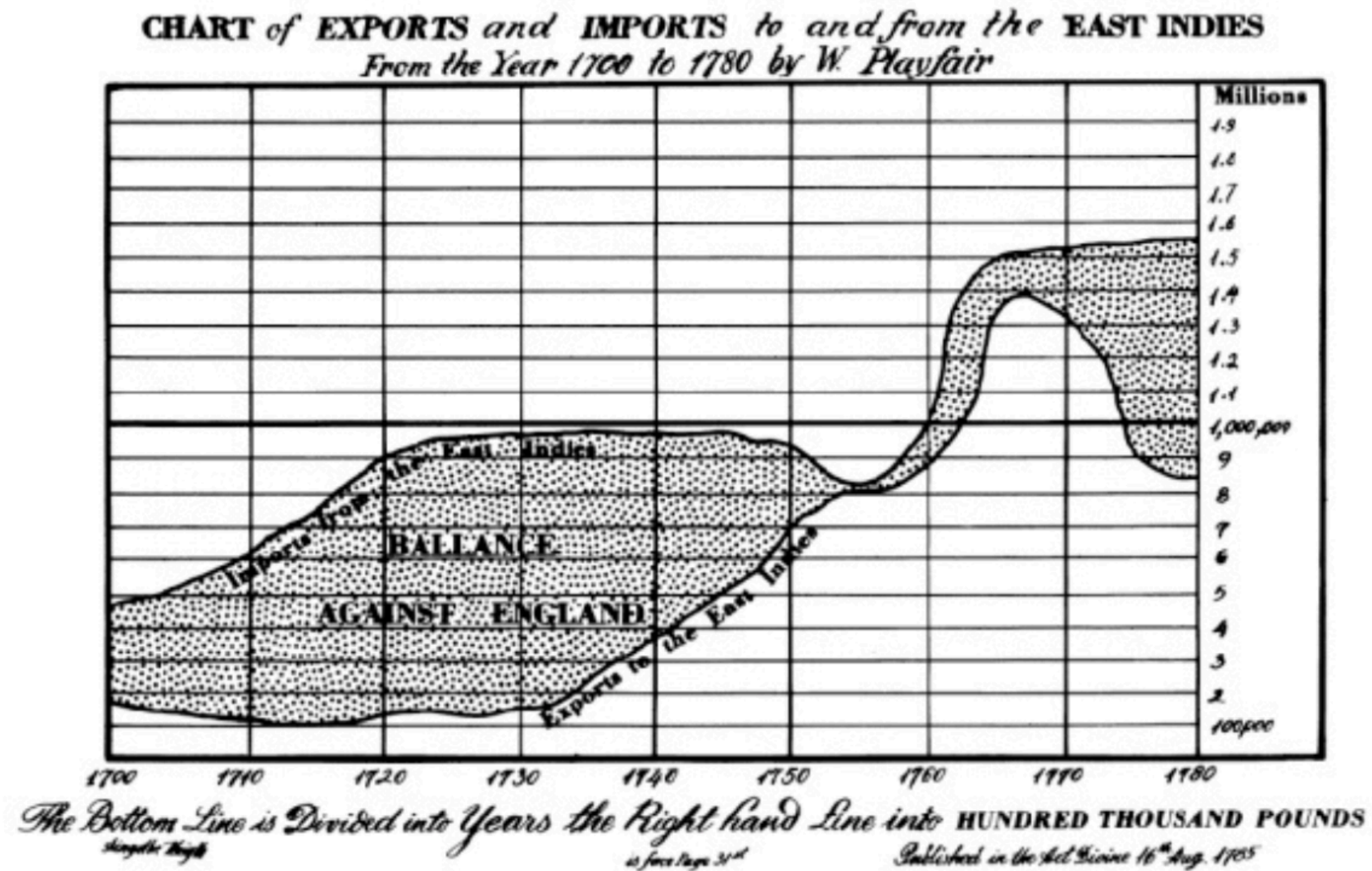


Figure 28. Playfair data.

Curve differences

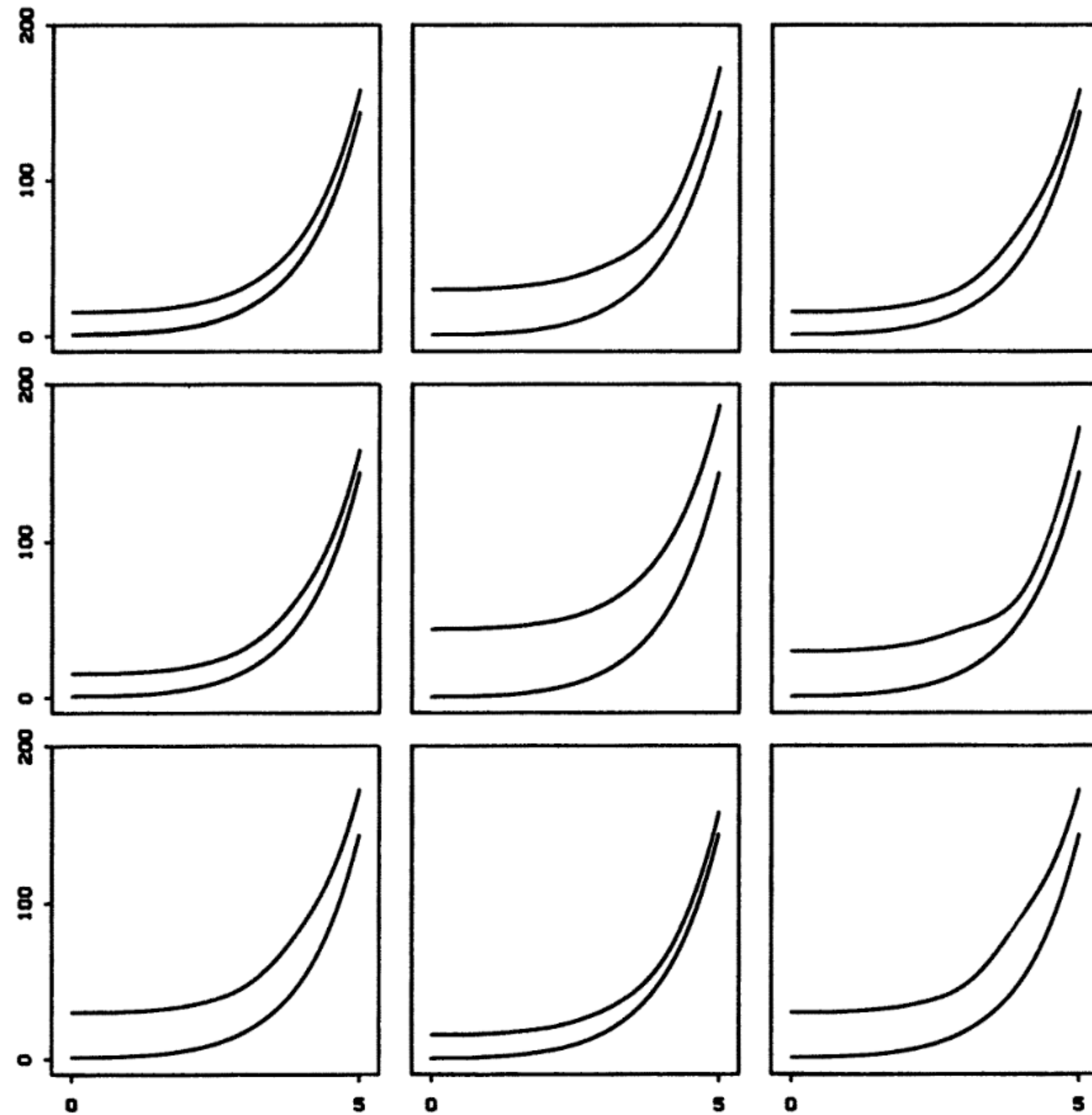


Figure 26. Curve-difference chart.

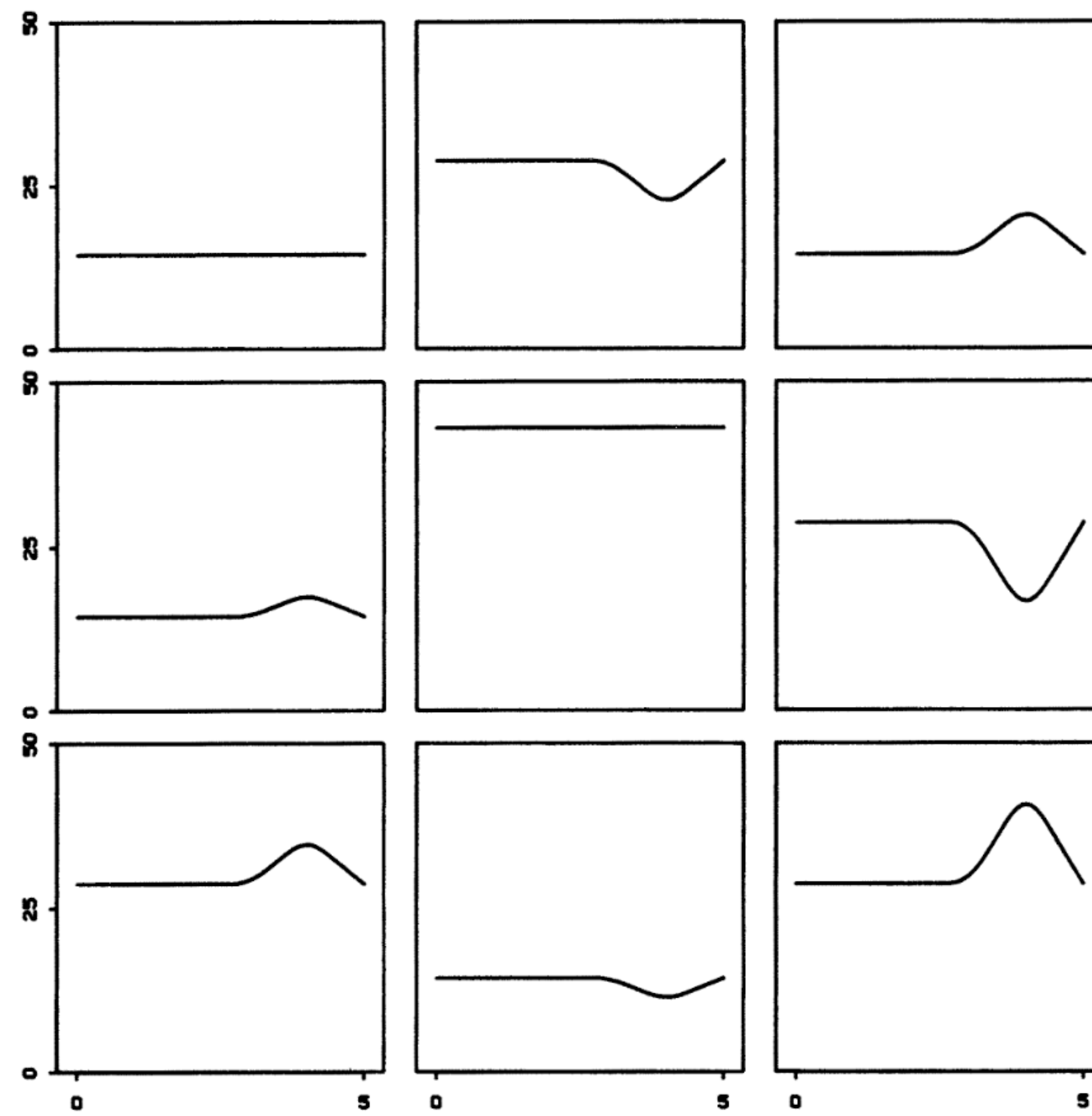
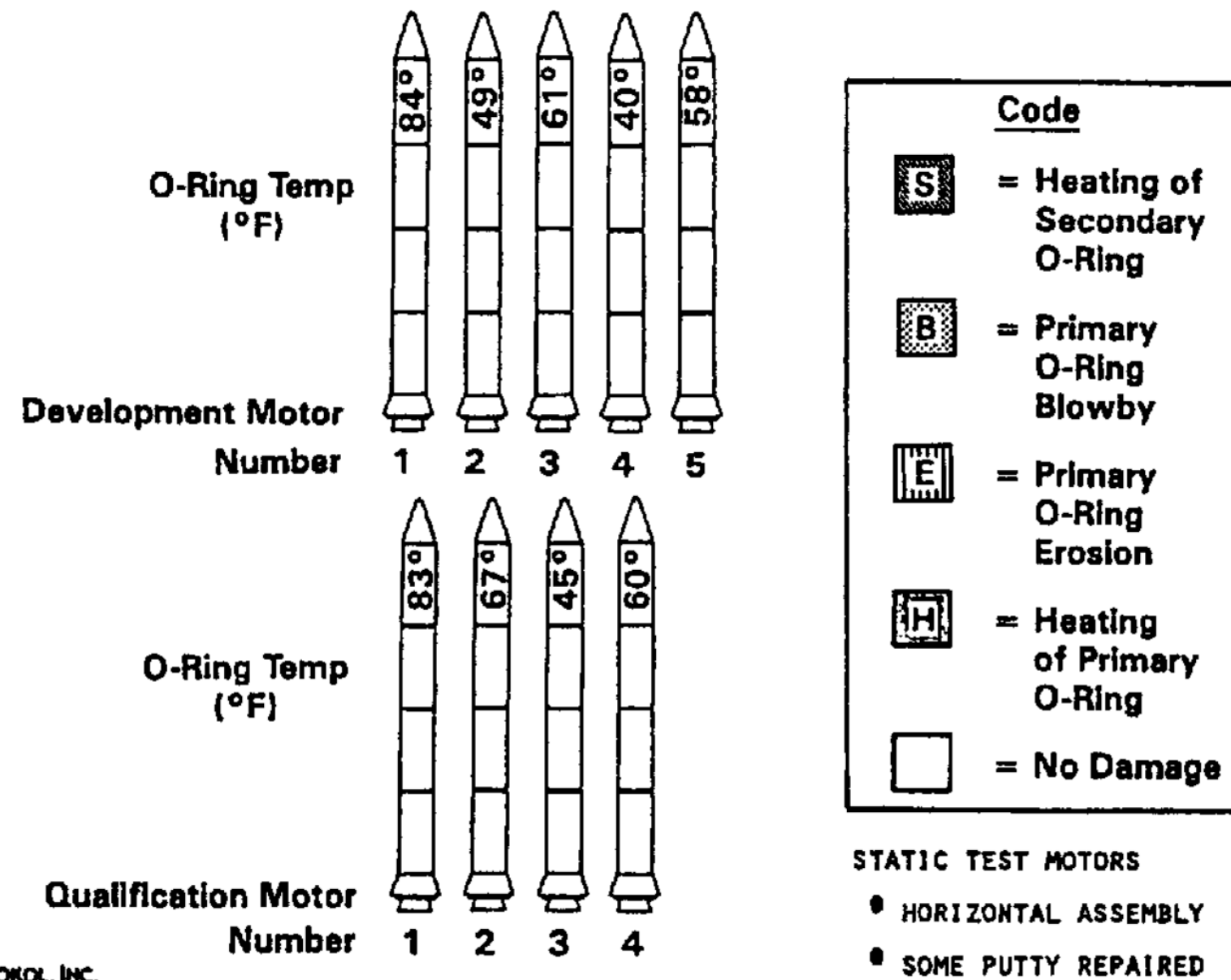


Figure 27. Curve differences.

Back to the challenger disaster!

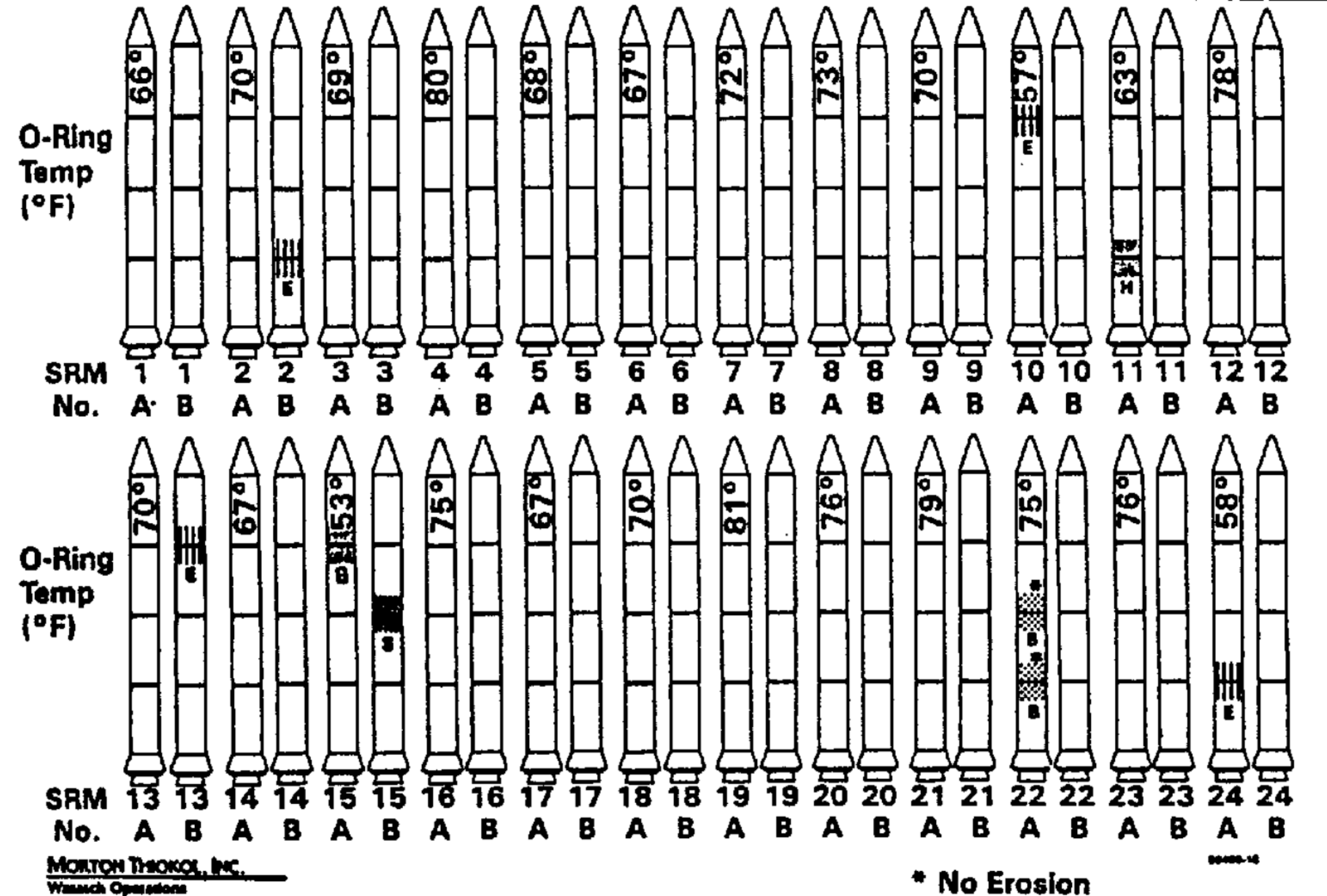
History of O-Ring Damage in Field Joints



MORTON THIKOL, INC.
Watch Operations

INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION

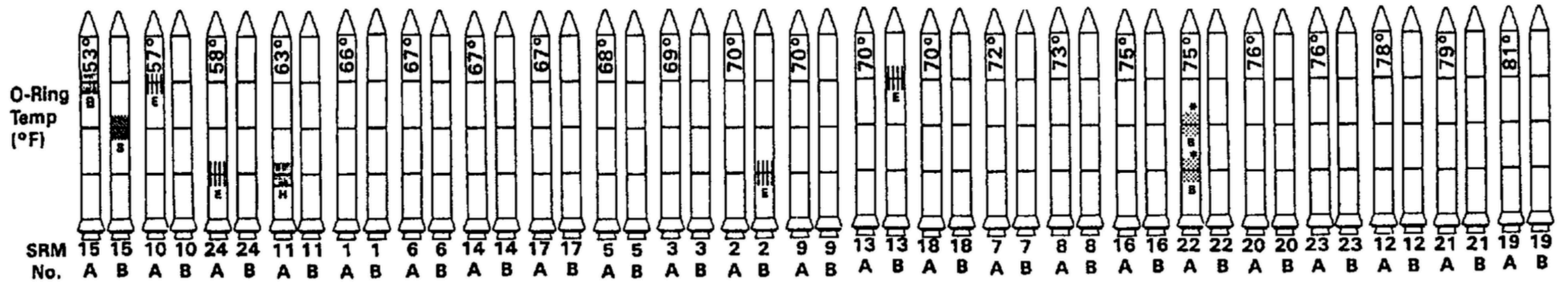
History of O-Ring Damage in Field Joints (Cont)



INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION

Tufte's Fix

Temperature is more important than date!



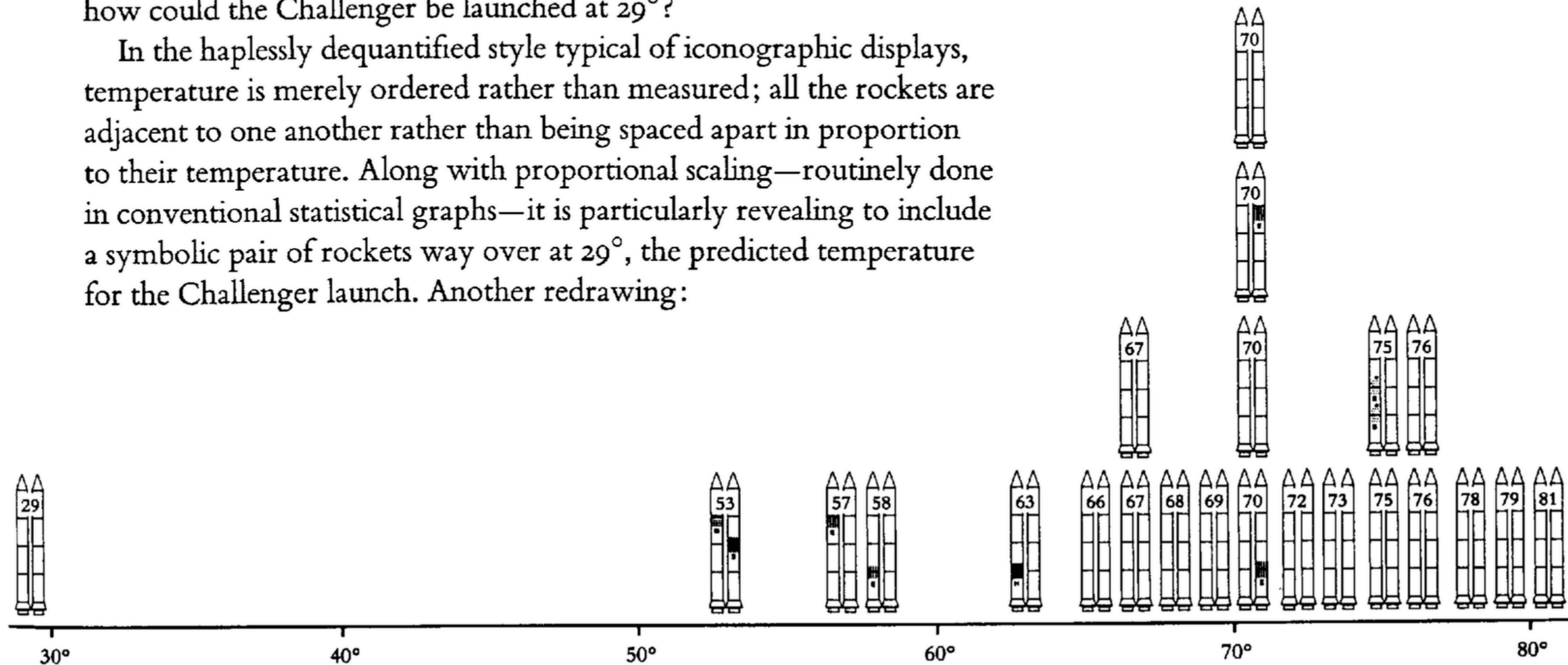
* No Erosion

Order isn't enough!

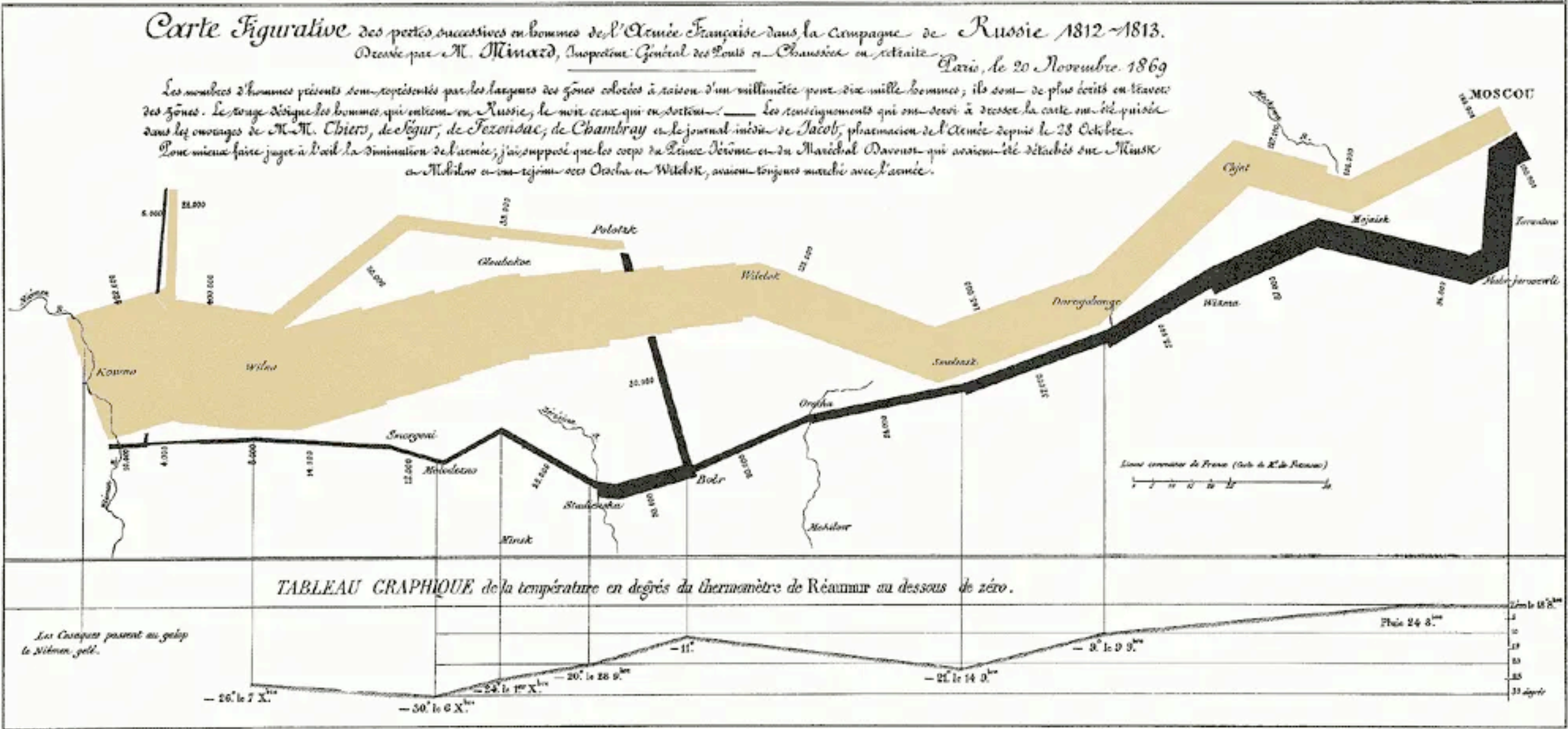
Ordering alone doesn't effectively convey differences!

temperature. The ~~most~~ most O-ring damage occurs at cooler temperatures. Given this evidence, how could the Challenger be launched at 29°?

In the haplessly dequantified style typical of iconographic displays, temperature is merely ordered rather than measured; all the rockets are adjacent to one another rather than being spaced apart in proportion to their temperature. Along with proportional scaling—routinely done in conventional statistical graphs—it is particularly revealing to include a symbolic pair of rockets way over at 29°, the predicted temperature for the Challenger launch. Another redrawing:



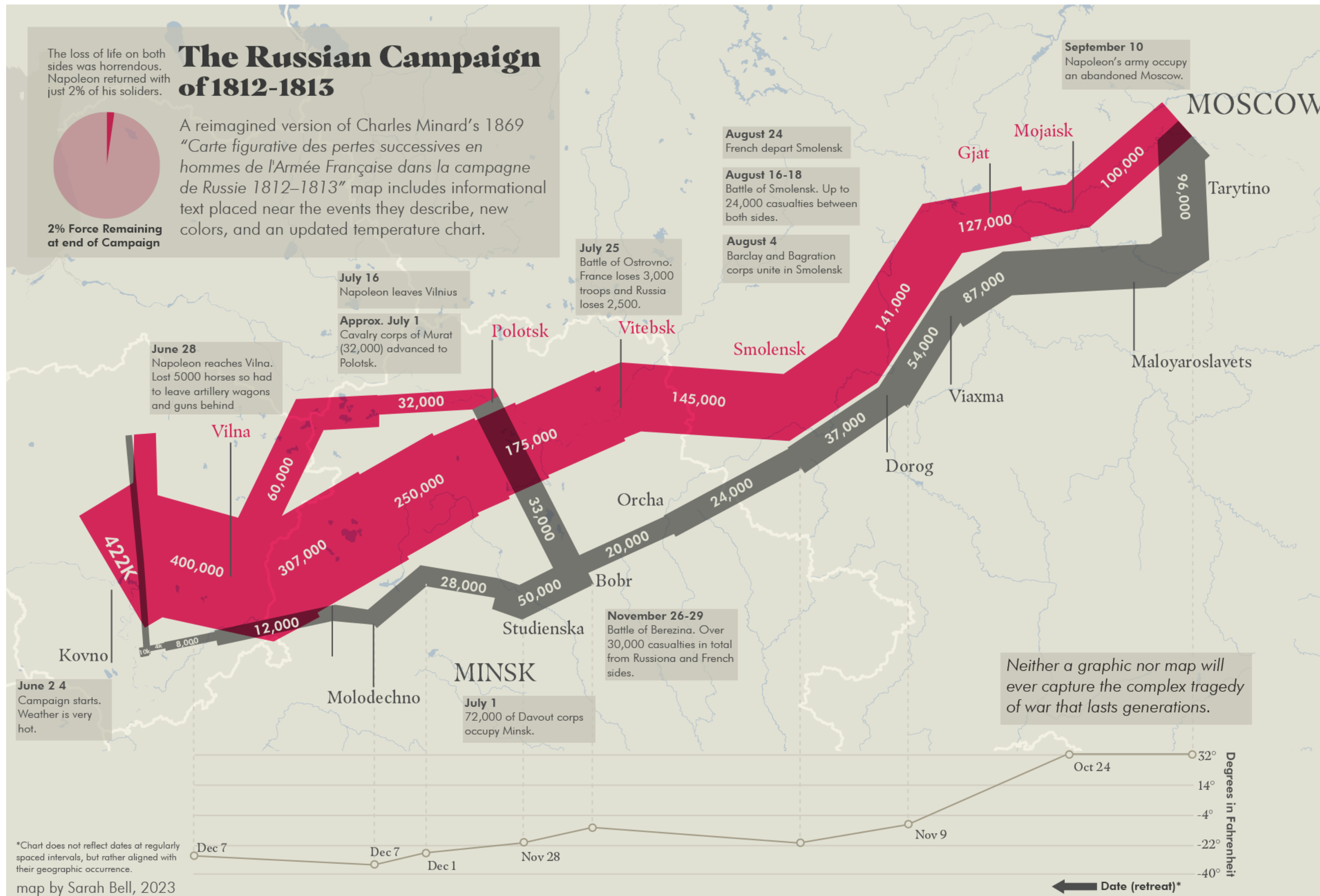
Tufte's favorite visualization



Impr. par Regnier, 1. Rue St Marc 37 27^e à Paris.

Imp. Ind. Regnier et Cie.

Warm-up exercise



Translated version by Sarah Bell

- *What are the variables shown?*
- *What are the aesthetic mappings for these variables?*
- *What do you like or dislike about this visualization?*

Back to Playfair

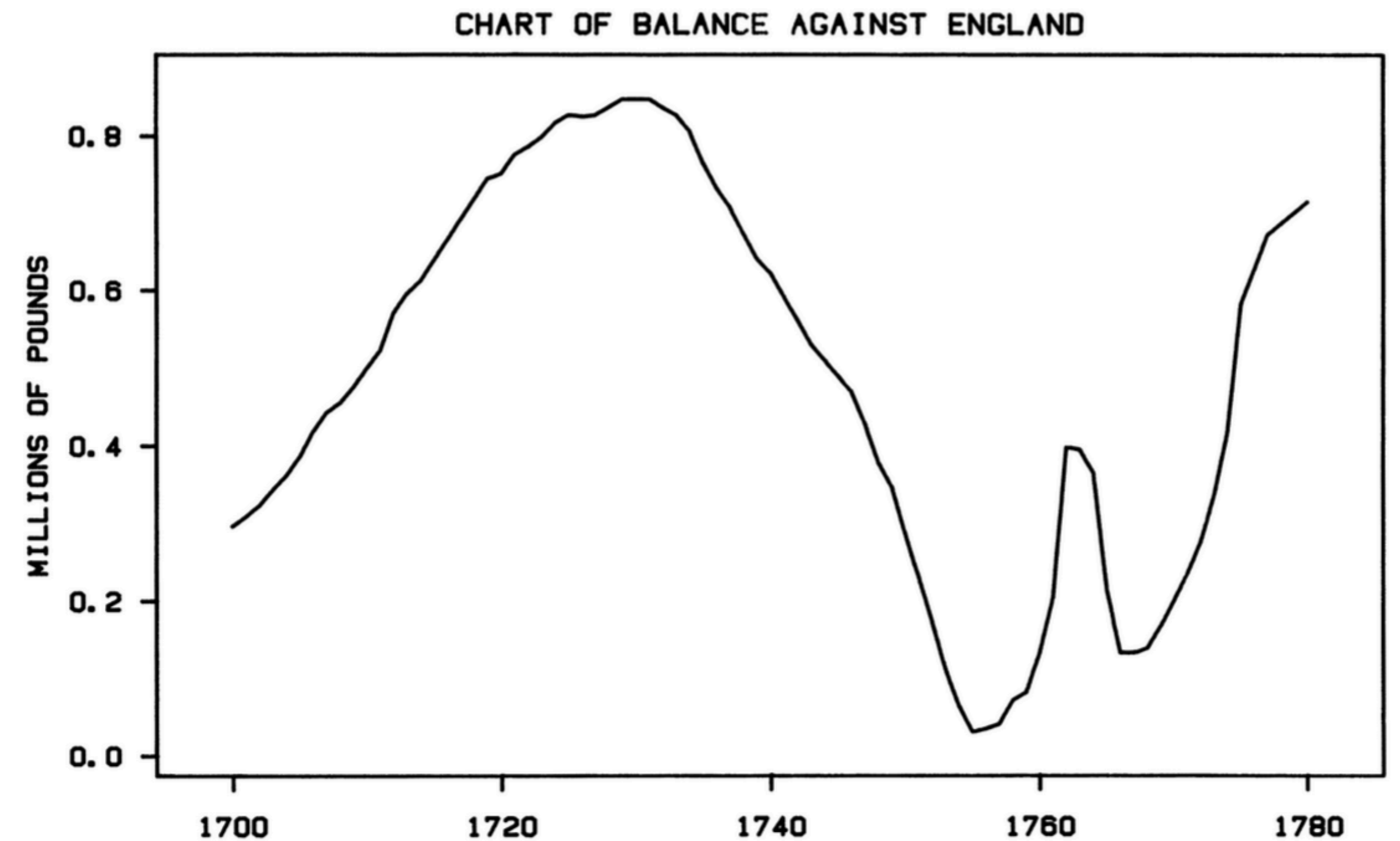
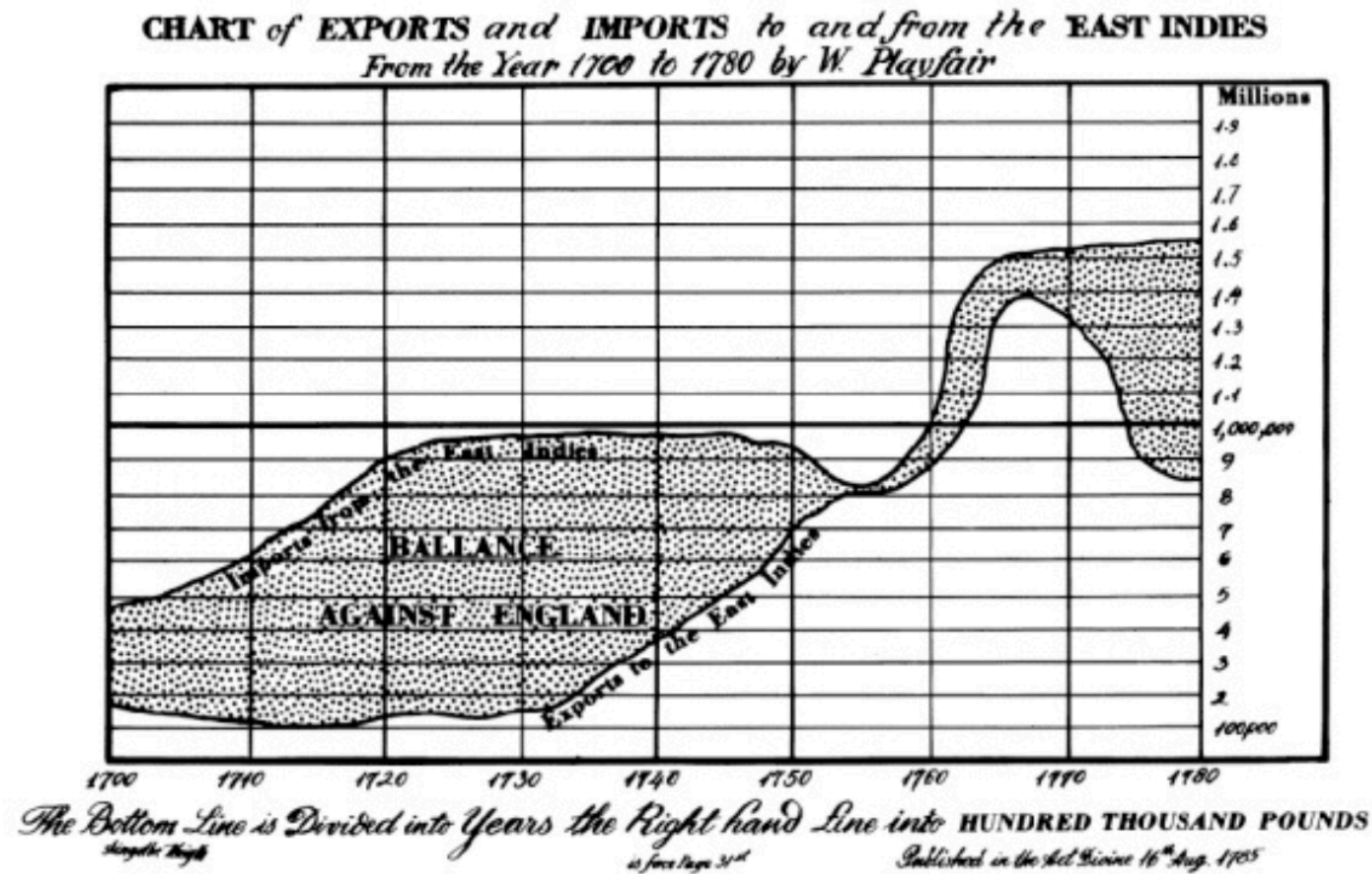


Figure 28. Playfair data.

Exercise:

How would you update the right chart to re-incorporate the missing information?

9 Criteria for good visualizations

- Expressiveness
- Effectiveness
- Intuitiveness
- Honesty
- Accessibility
- Impact
- Insight
- Writing
- Engineering

9 Criteria for good visualizations

- Expressiveness
- Effectiveness
- Intuitiveness
- Honesty
- Accessibility
- Impact
- Insight
- Writing
- Engineering

Intuitiveness (i.e. clarity)

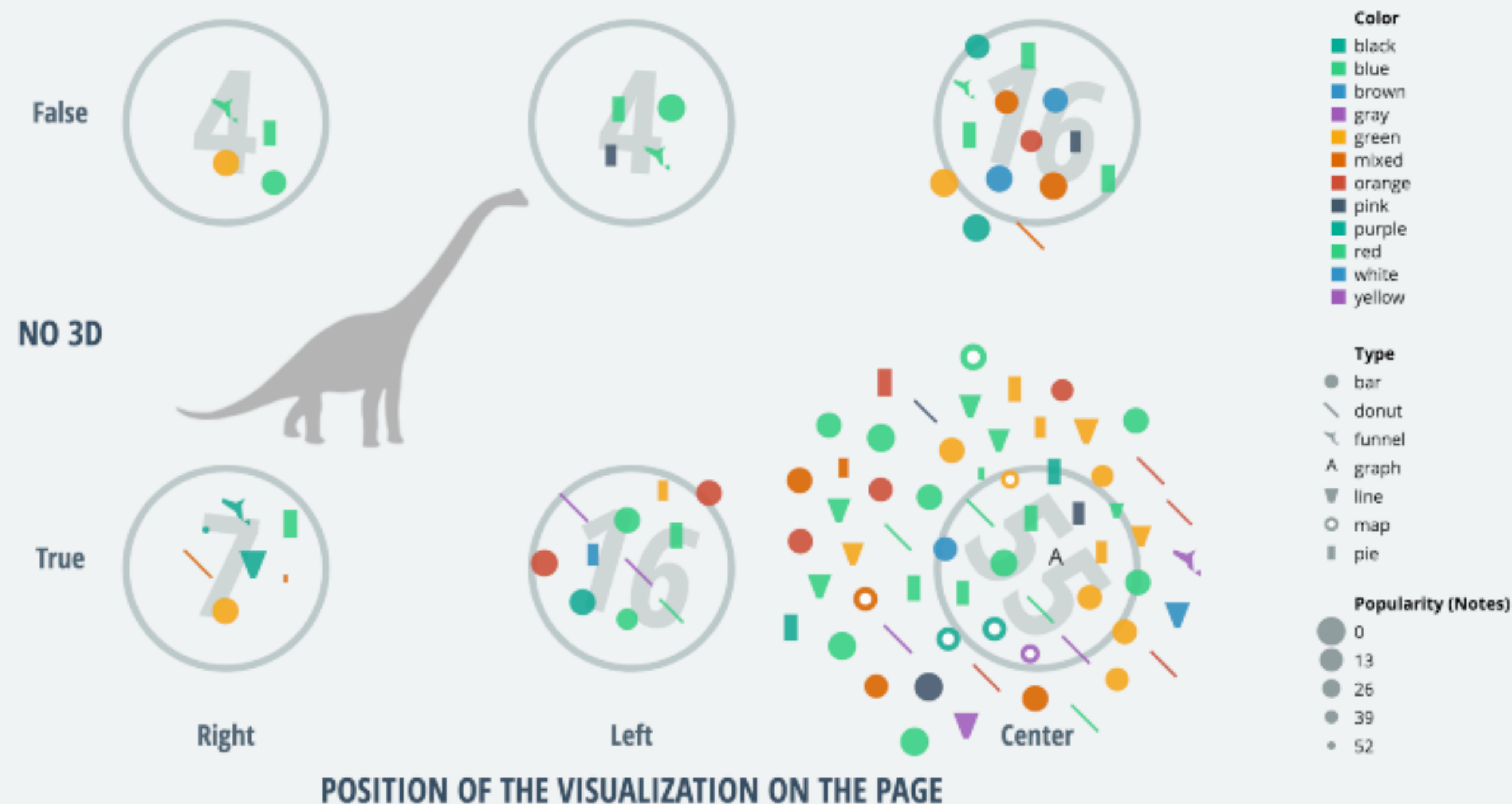
How easy is it to understand what is (*and is not!*) being shown in a visualization?

A basic example

MAKING TERRIBLE CHARTS GREAT AGAIN! #BESTWORSTVIZ

Each mark in this visualization represents a visualization from [viz.wtf](#), a well-known guide to best practice in data visualization. These 100 visualizations are carefully collected, i.e., scroll down and enter data into a spreadsheet until I reach the 100th. The variables are *popularity* (number of notes), *position* of the visualization on the page, *main color*, *visualization type* and whether it *uses 3D*.

You can clearly see the emerging pattern from this carefully encoded visualization, that is nothing can be interpreted without blowing your mind first. Let's see if you can figure out everything that is wrong with this chart. Raw data are also available ([data.pdf](#)).



Too many encodings!

- More information isn't always better
- Focus on the important variables
- Highlight the takeaways

Guidelines: Tufte's Principles

Minimize distraction

- Above all else show the data
- Maximize the **data-ink ratio**
- Erase non-data-ink
- Erase redundant data-ink
- Revise and edit



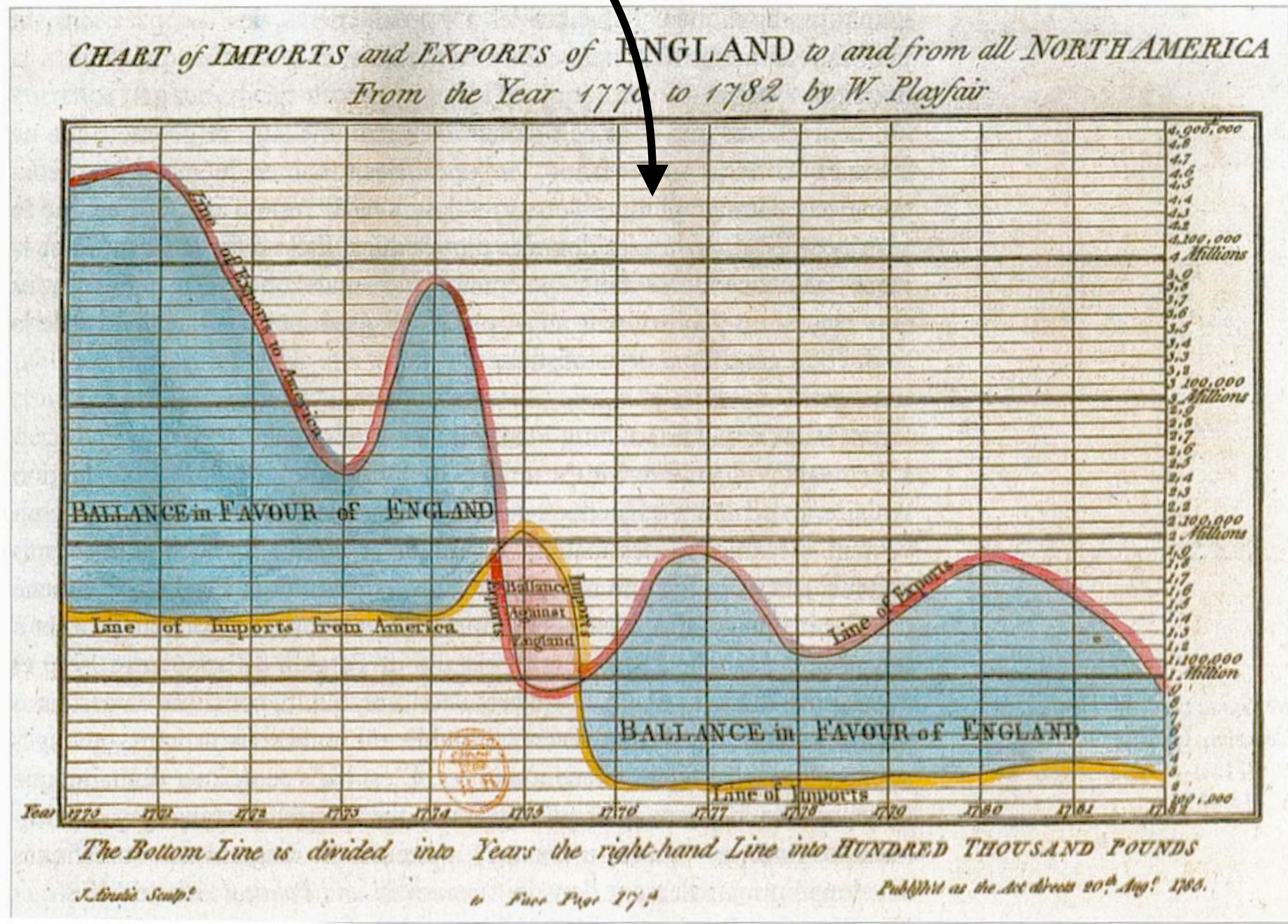
Data-ink ratio

Heuristic for visual complexity

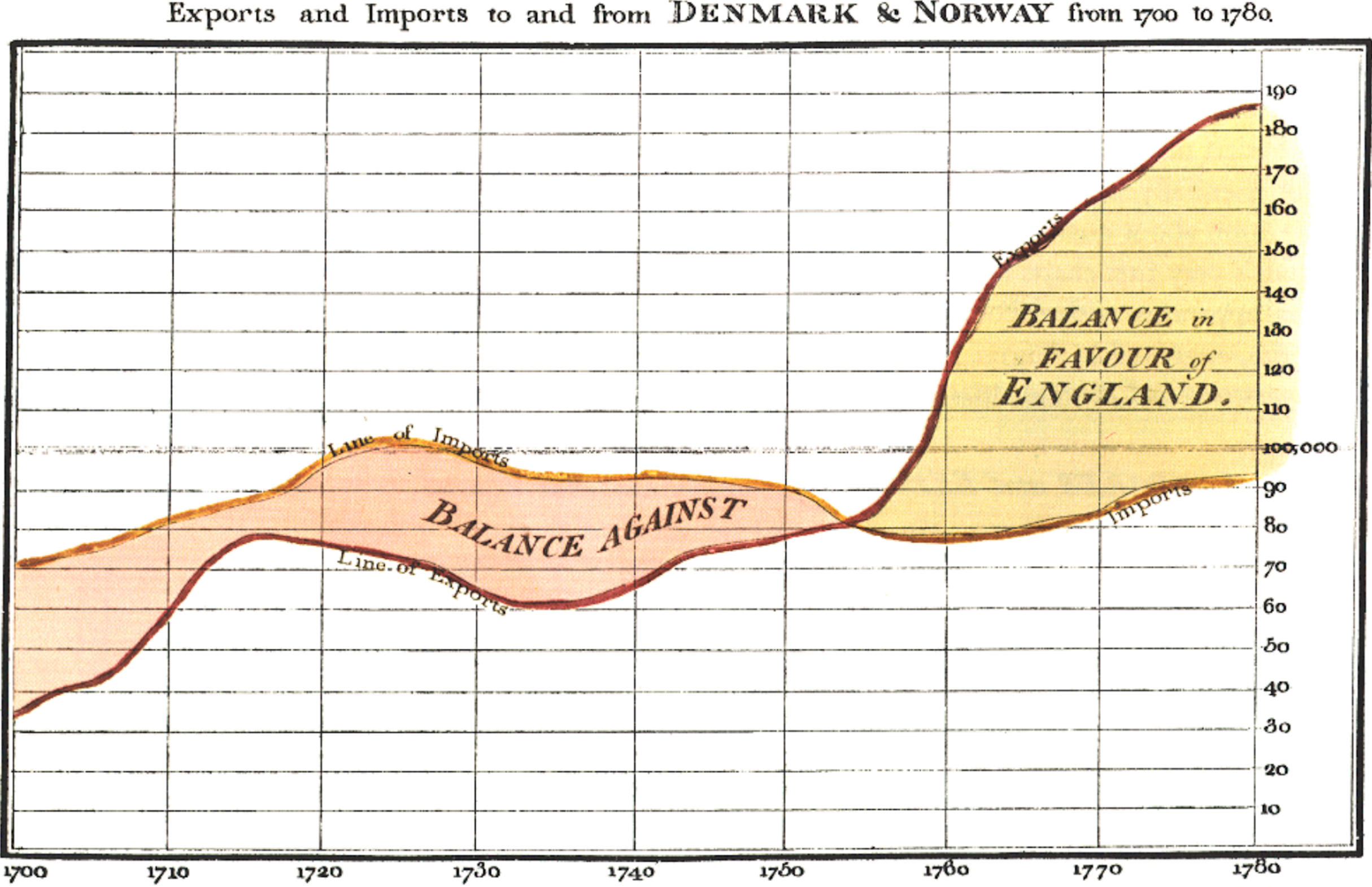
$$\begin{aligned} \text{Data-ink ratio} &= \frac{\text{data-ink}}{\text{total ink used to print the graphic}} \\ &= \text{proportion of a graphic's ink devoted to the} \\ &\quad \text{non-redundant display of data-information} \\ &= 1.0 - \text{proportion of a graphic that can be erased} \\ &\quad \text{without loss of data-information.} \end{aligned}$$

Unnecessary grid lines!

Data-ink ratio



Low ratio: more ink for the same amount of data



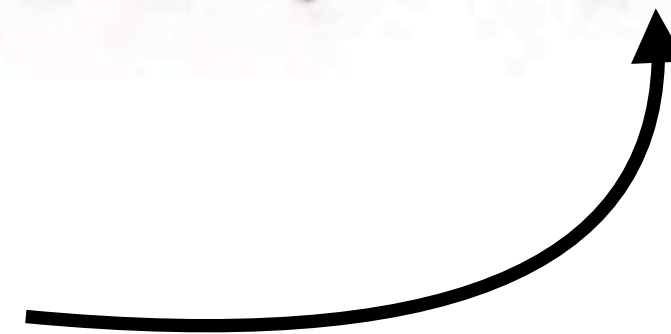
High ratio: less ink for the same amount of data

Maximal data-ink?

EEG Data

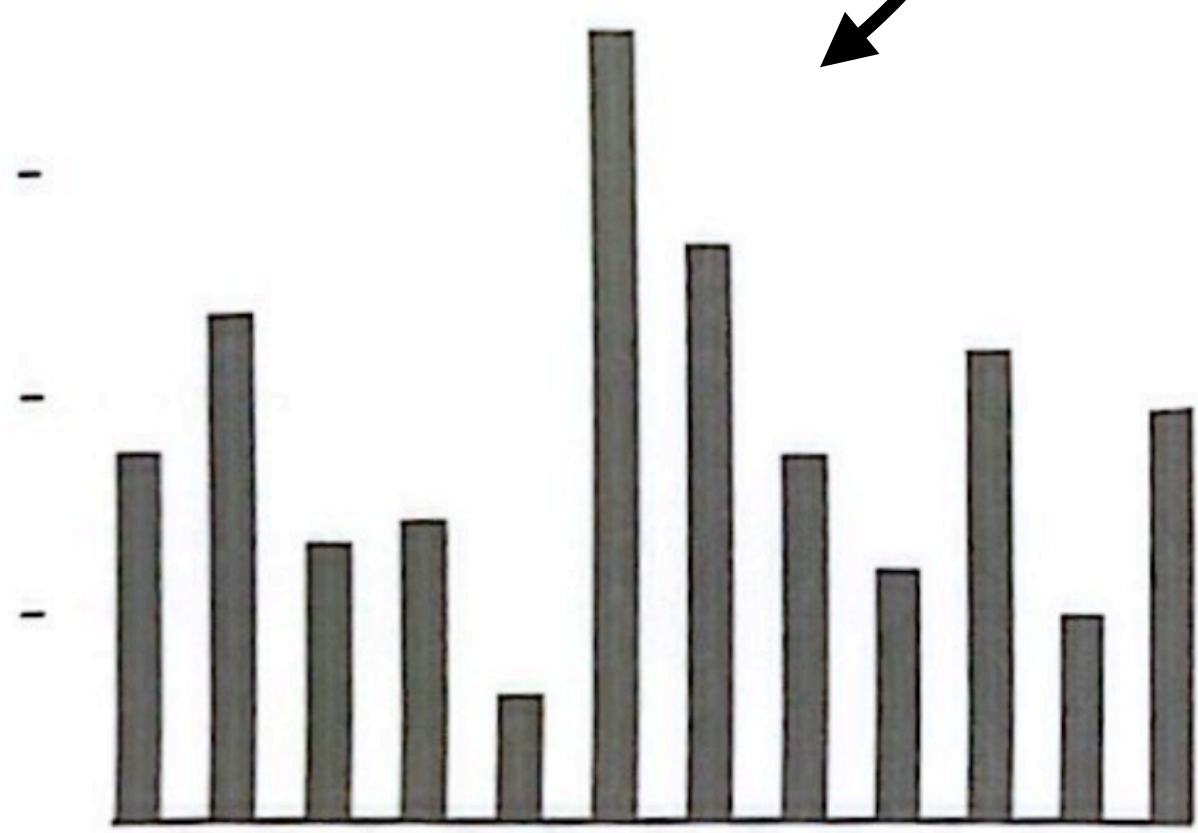
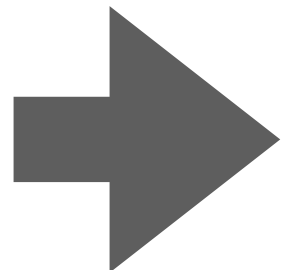
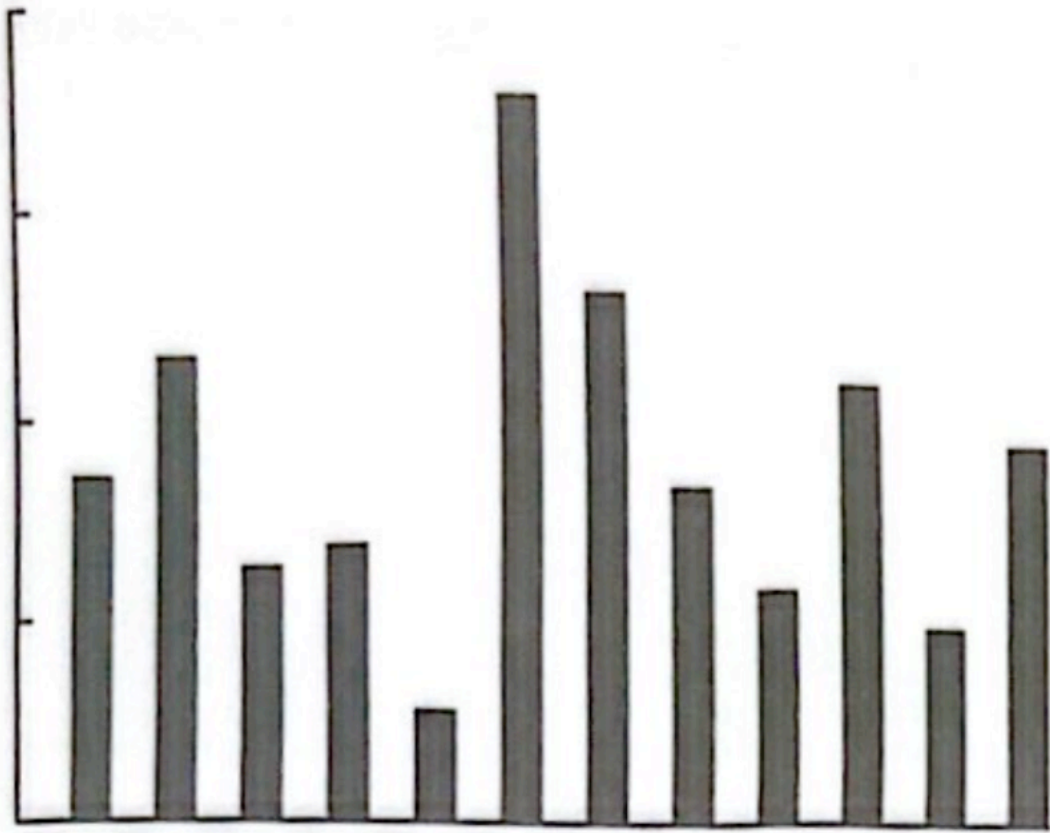
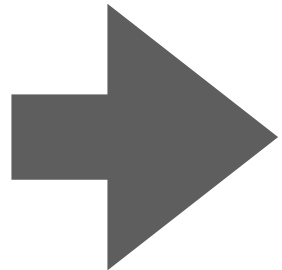
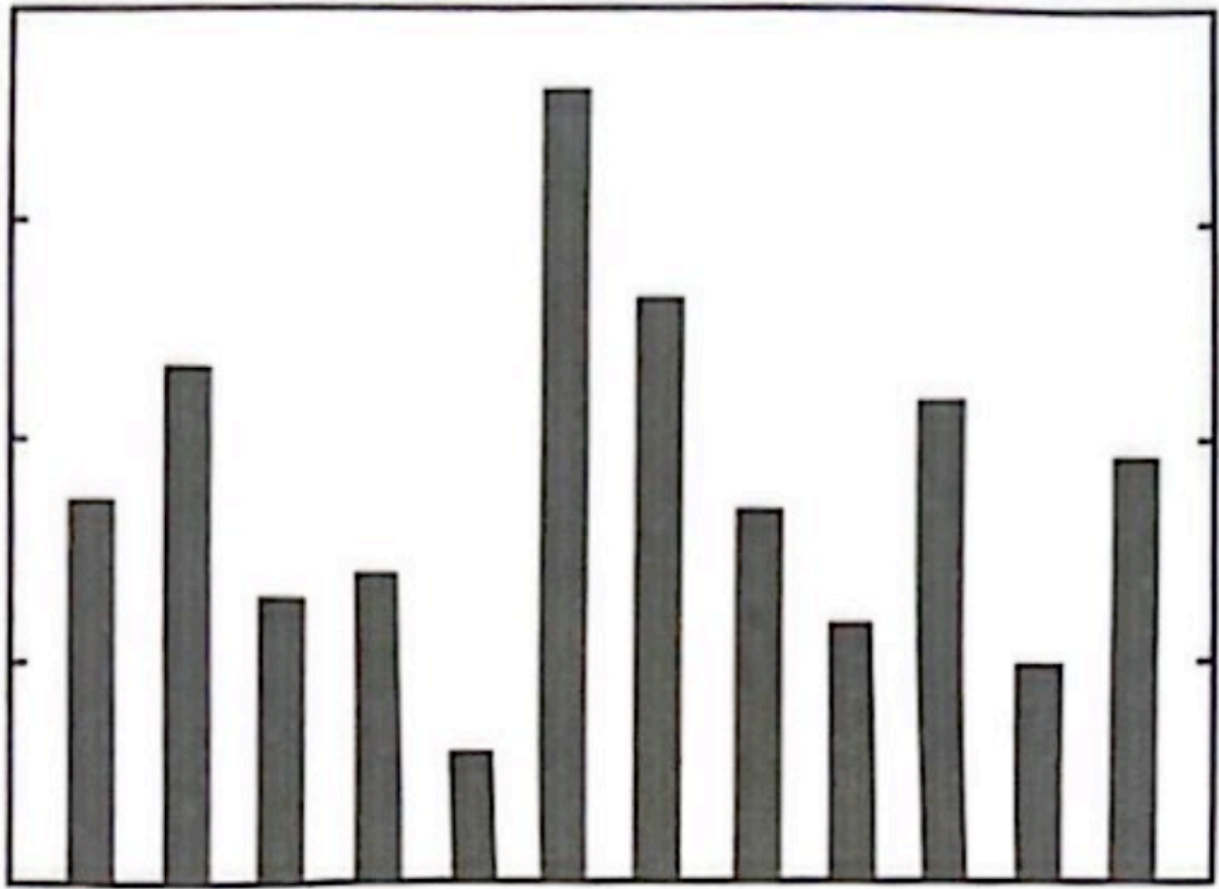


Can't reduce ink
without losing data!

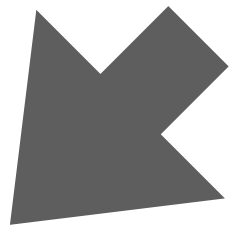


Revise and edit!

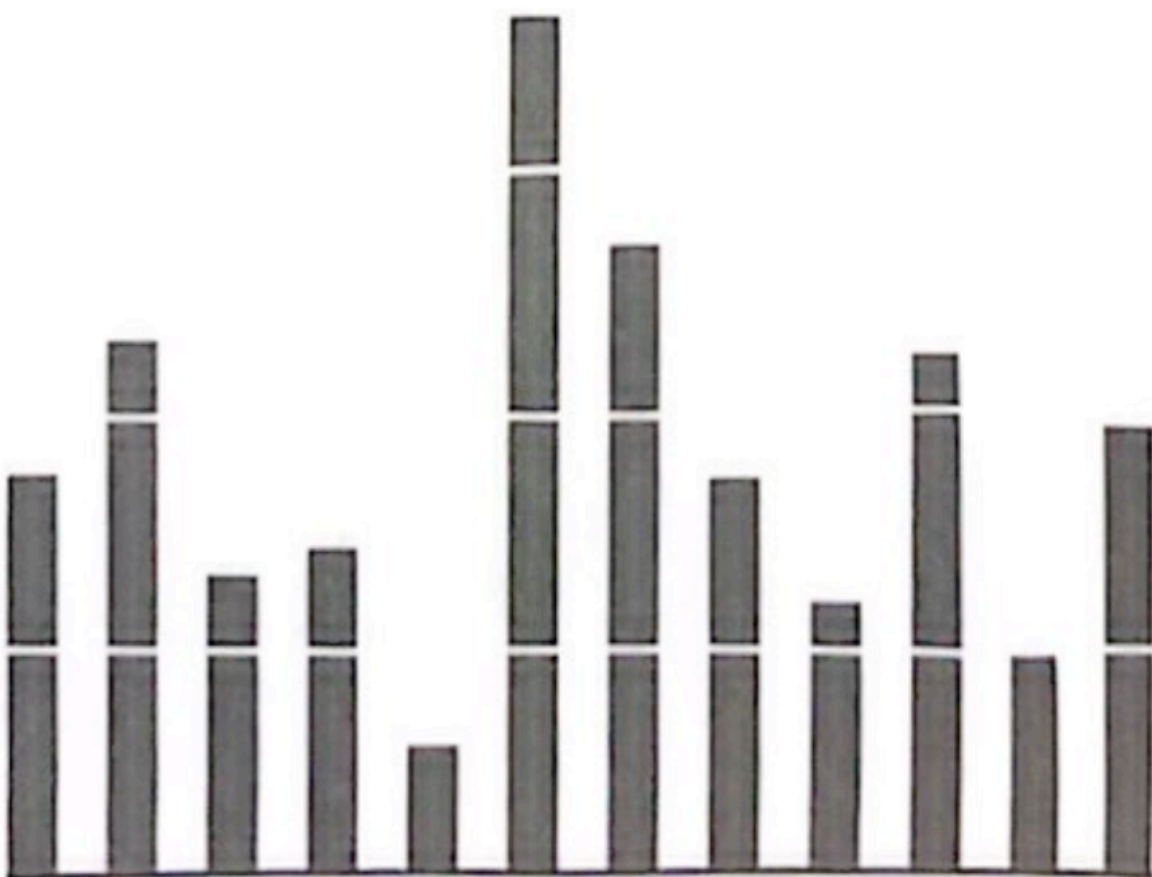
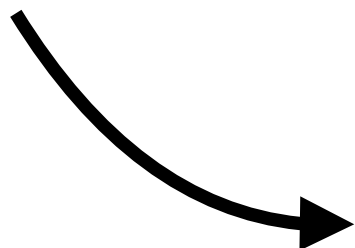
Start



Remove unneeded elements



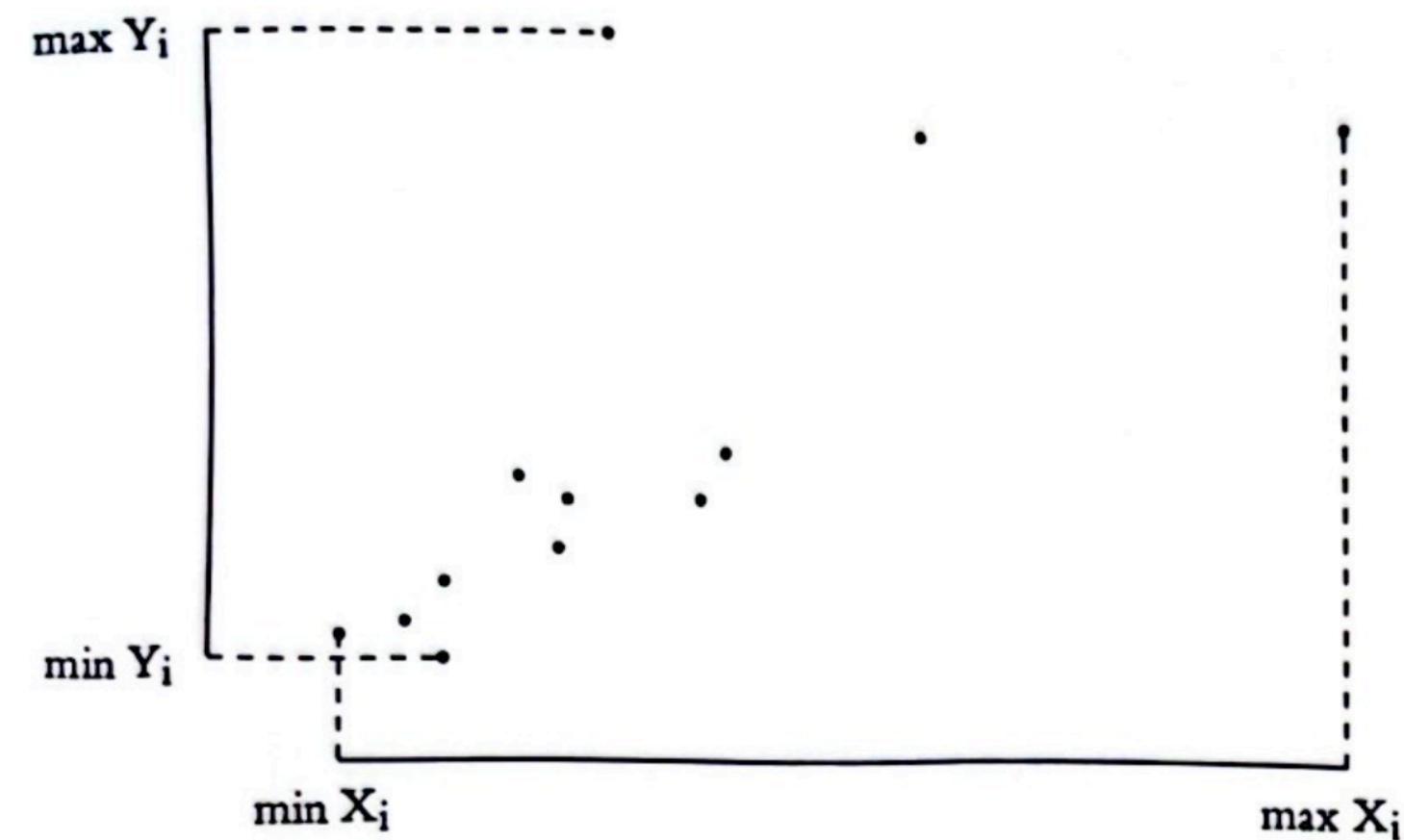
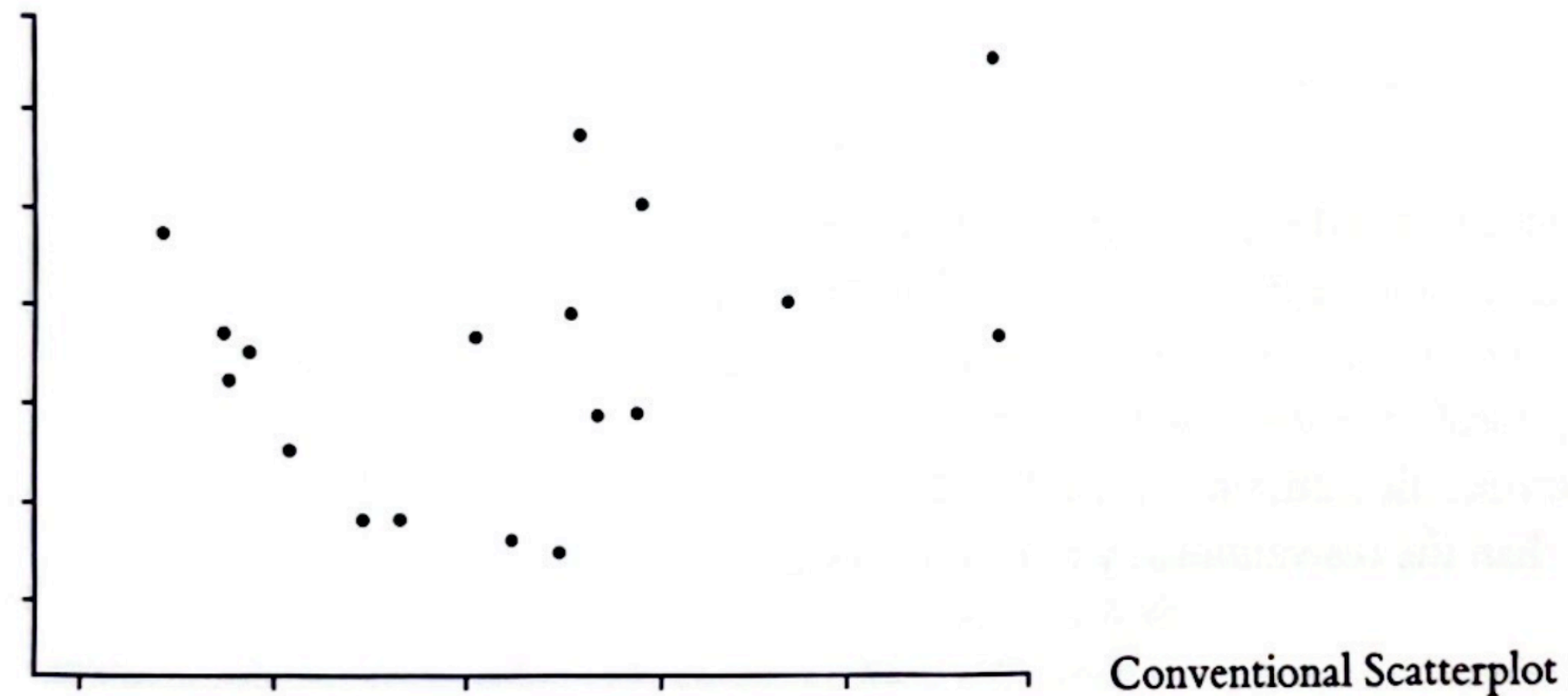
Encode scale with whitespace!



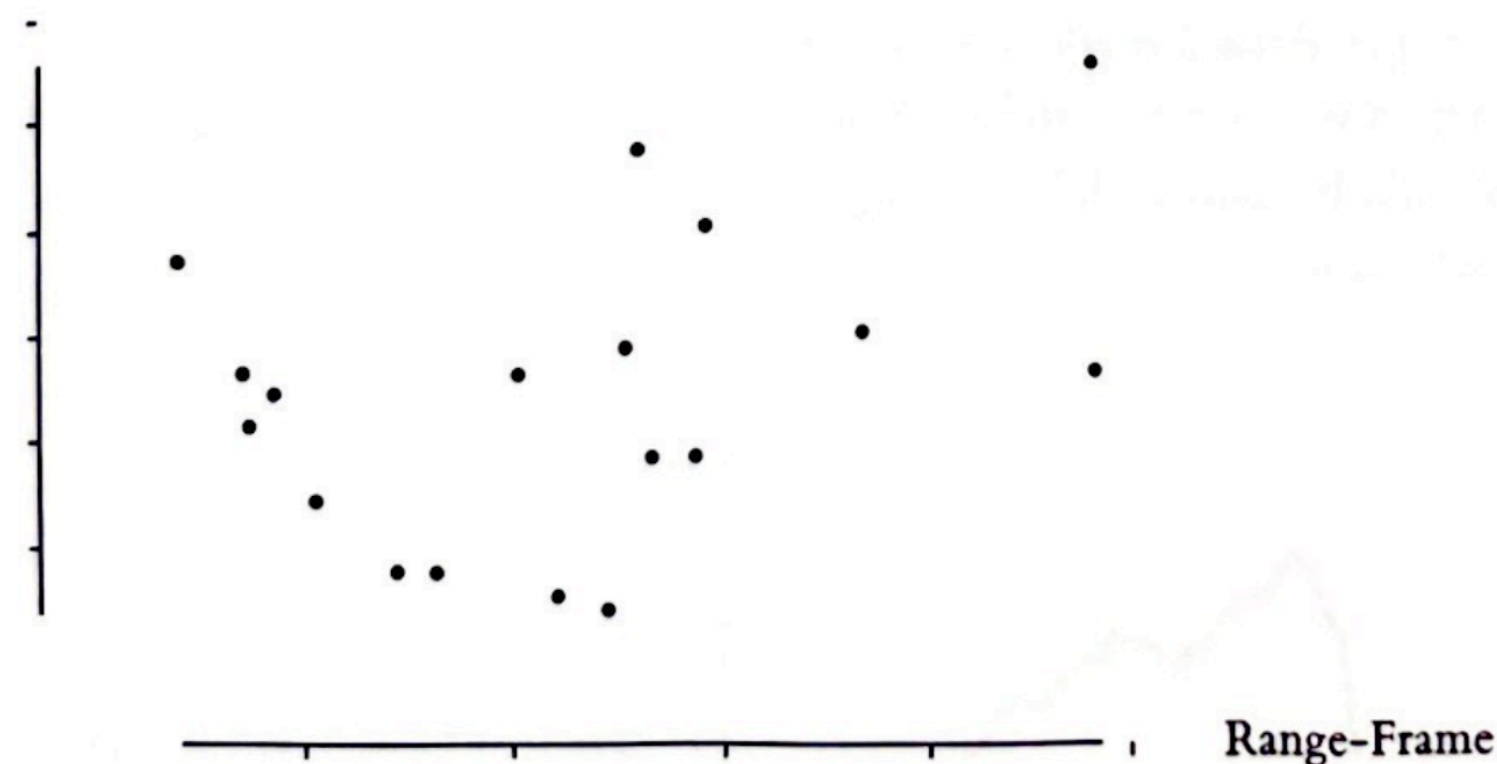
Possibly also more effective?

Data-ink benefits

Improving data-ink can also improve *effectiveness*

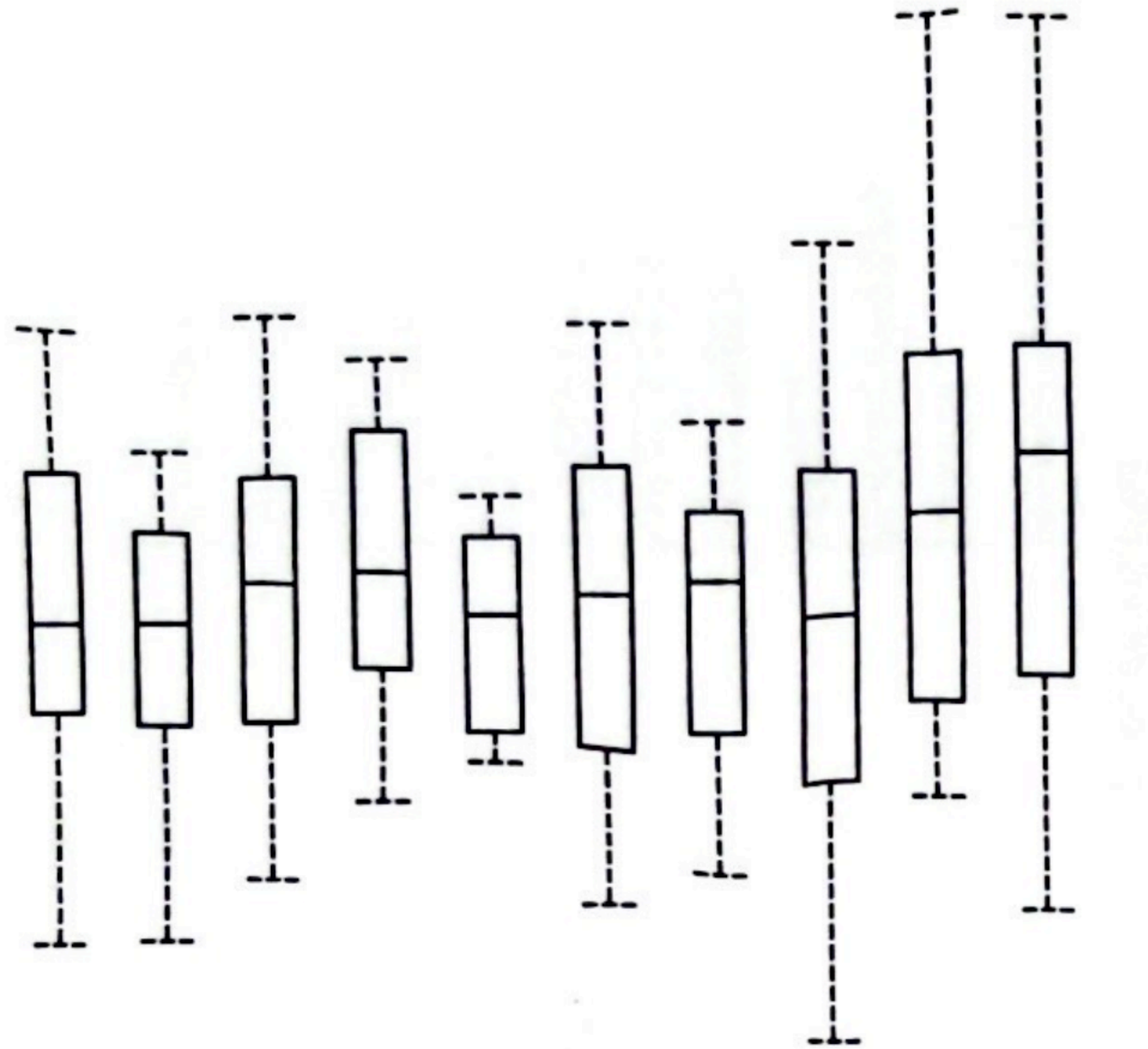


Example: range-frame scatterplot shows extent of data

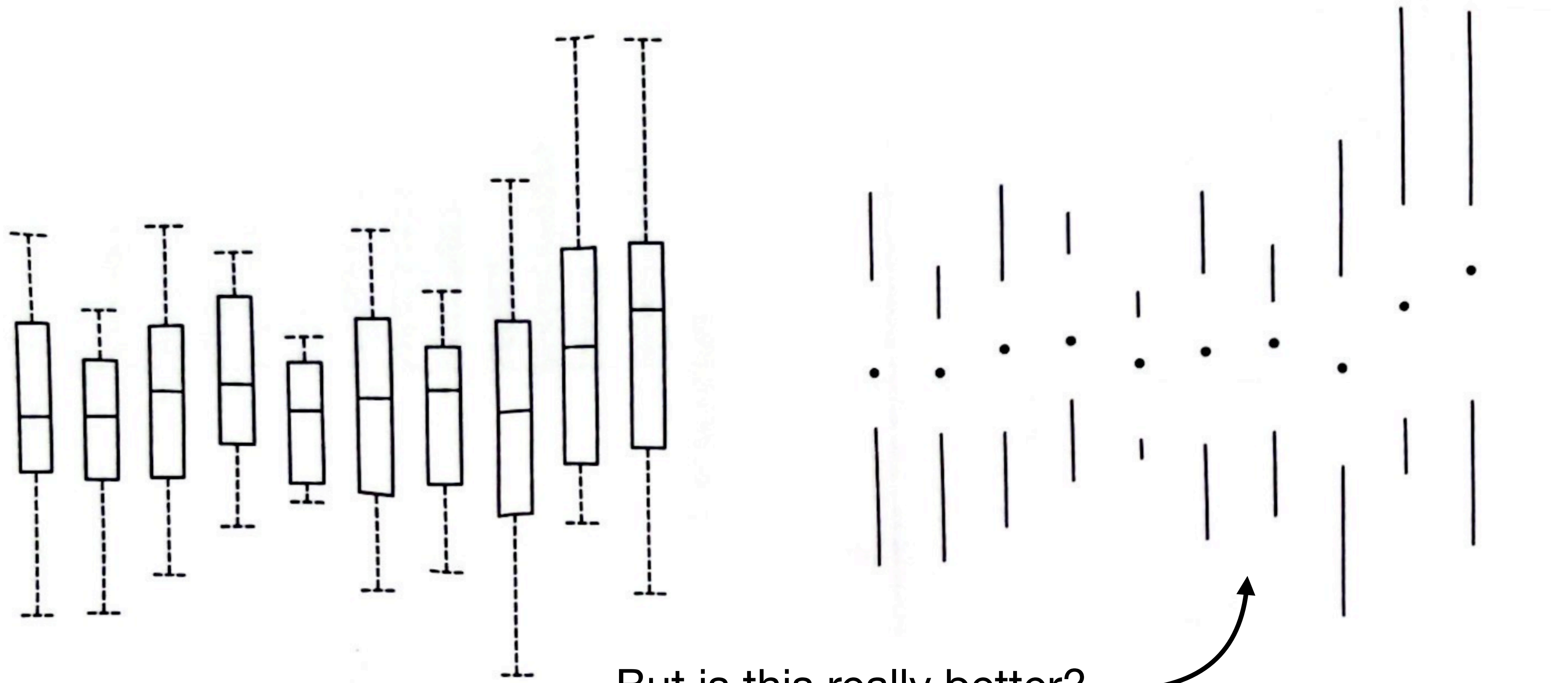


Data-ink: deriving new forms

Exercise: Can you simplify Tukey's classic box-and-whisker plot?



An example



But is this really better?

Intuitiveness: Respect established conventions!

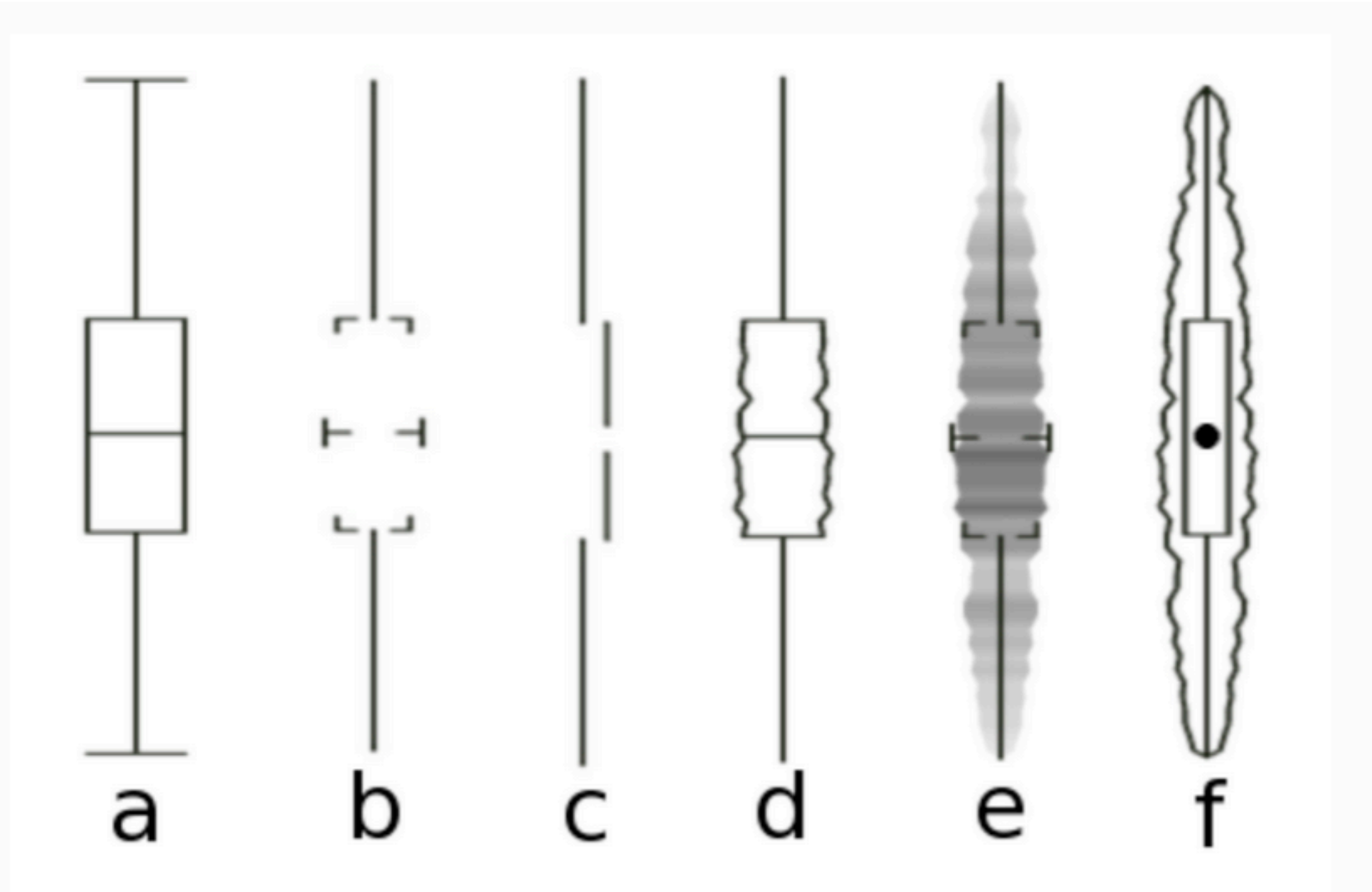


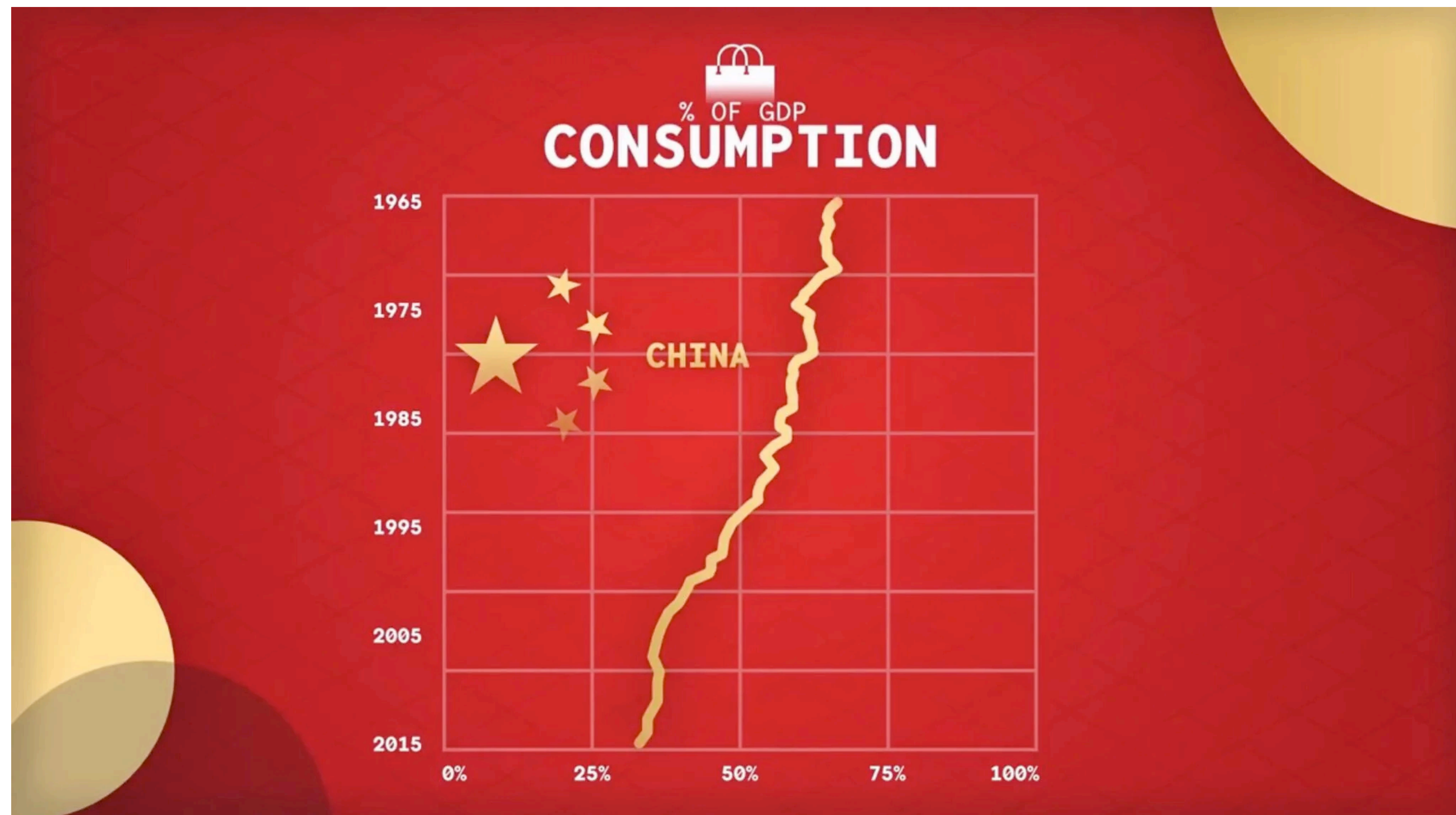
Figure 1.7: Six kinds of summary boxplots. Type (c) is from Tufte.

Saving ink is counter-productive if users need to spend more time understanding what the plot is showing!

Prof. Gabe's take: Unless you have good reason, just go with a!

Respect established conventions

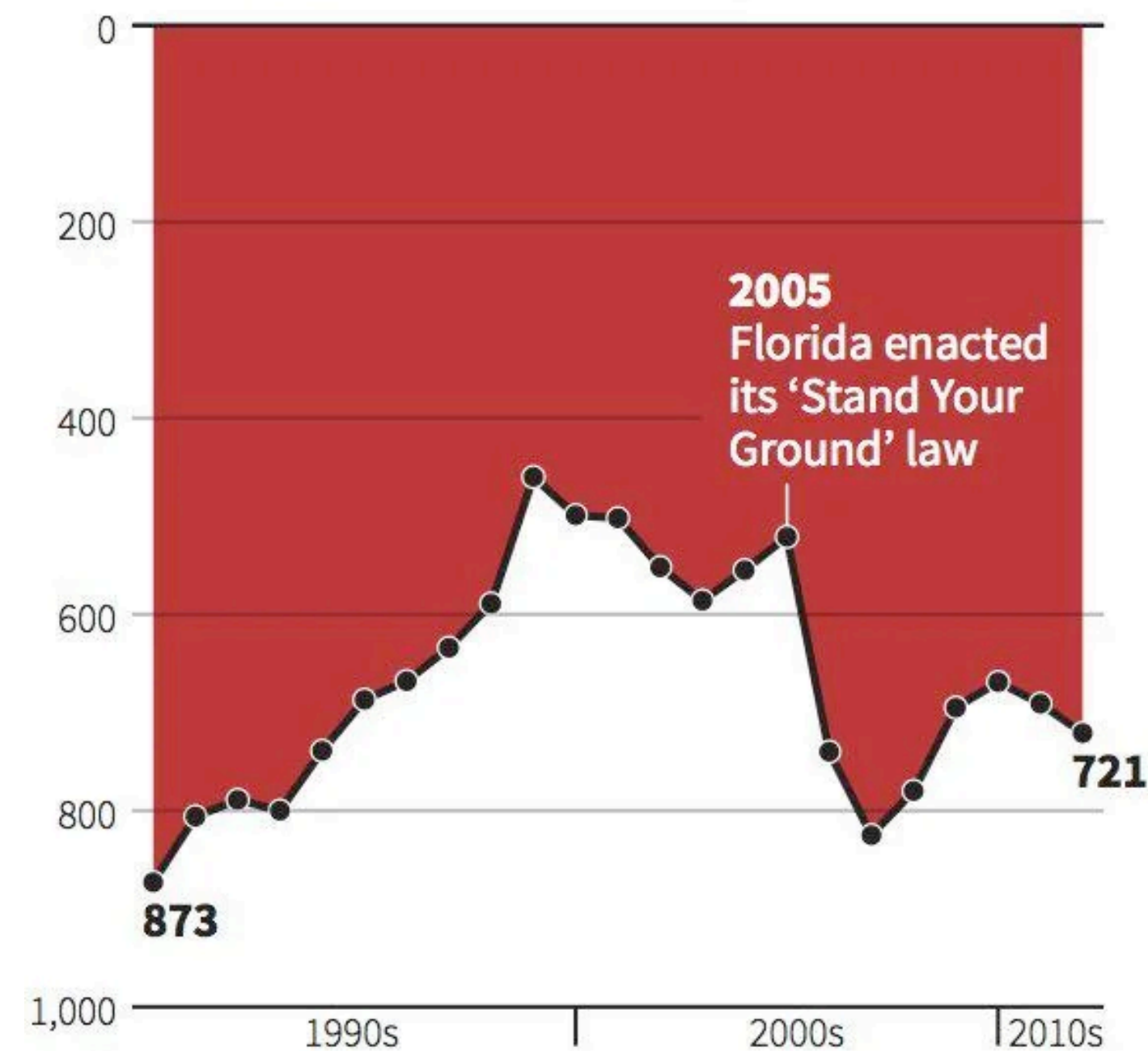
Example: Time on the x-axis!



Respect established conventions

Gun deaths in Florida

Number of murders committed using firearms



Source: Florida Department of Law Enforcement

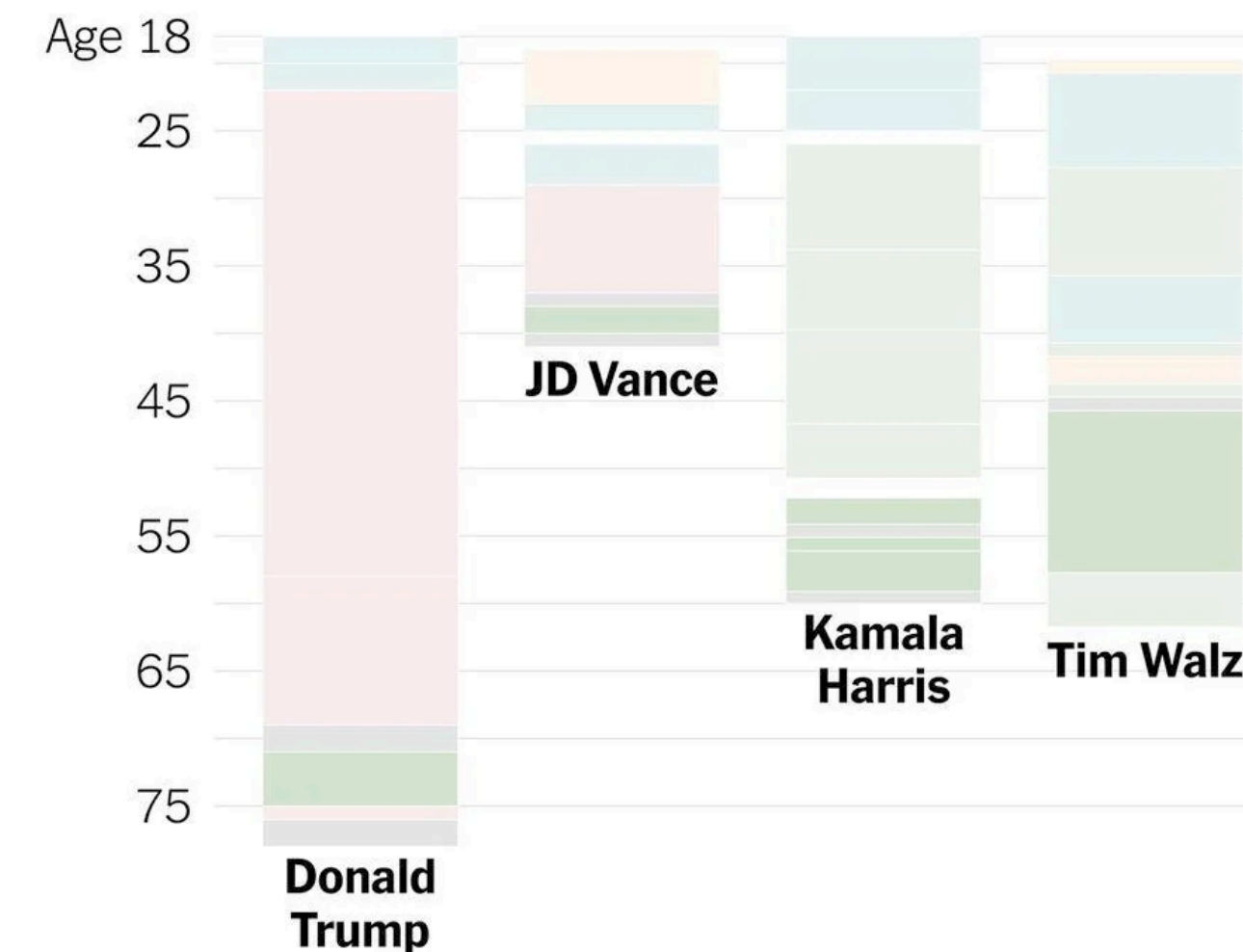
C. Chan 16/02/2014

REUTERS

2024

How Trump-Vance and Harris-Walz Made It to the Presidential Ticket

College Private sector Federal government
Military Public service or politics Candidate for federal office

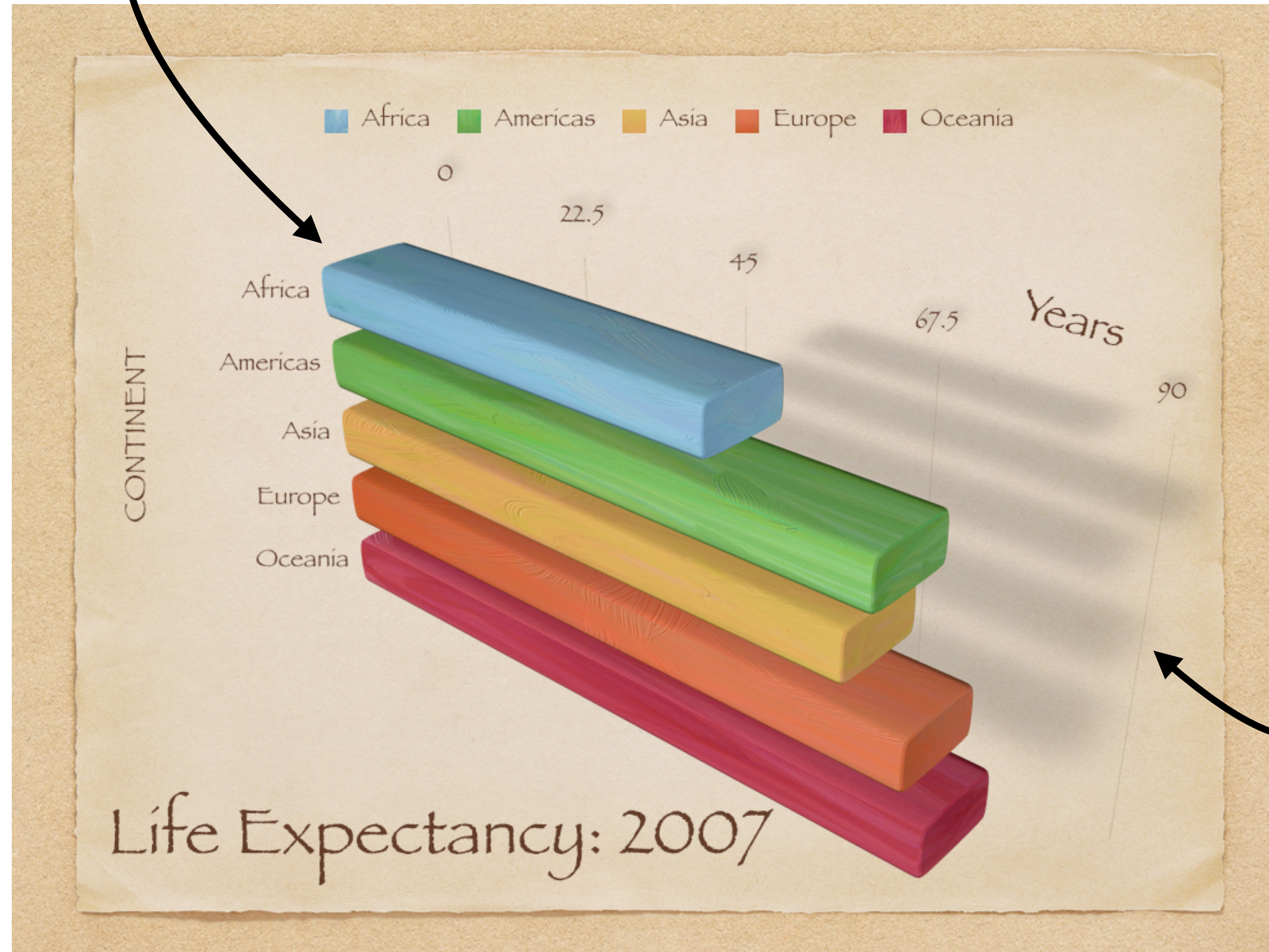


Higher numbers at the top/right!

3-d rotation inhibits effectiveness

“Chart junk”

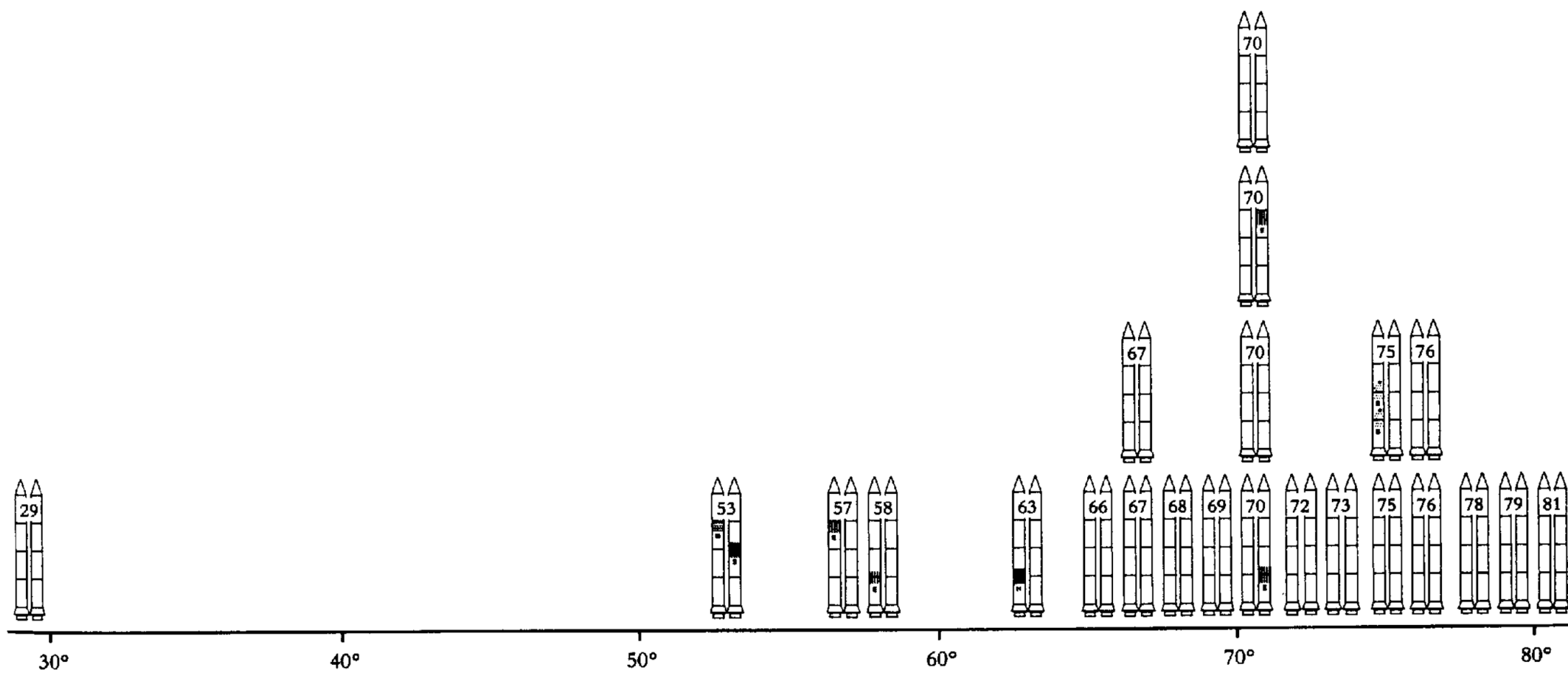
The antithesis of Tufte’s minimalism principles



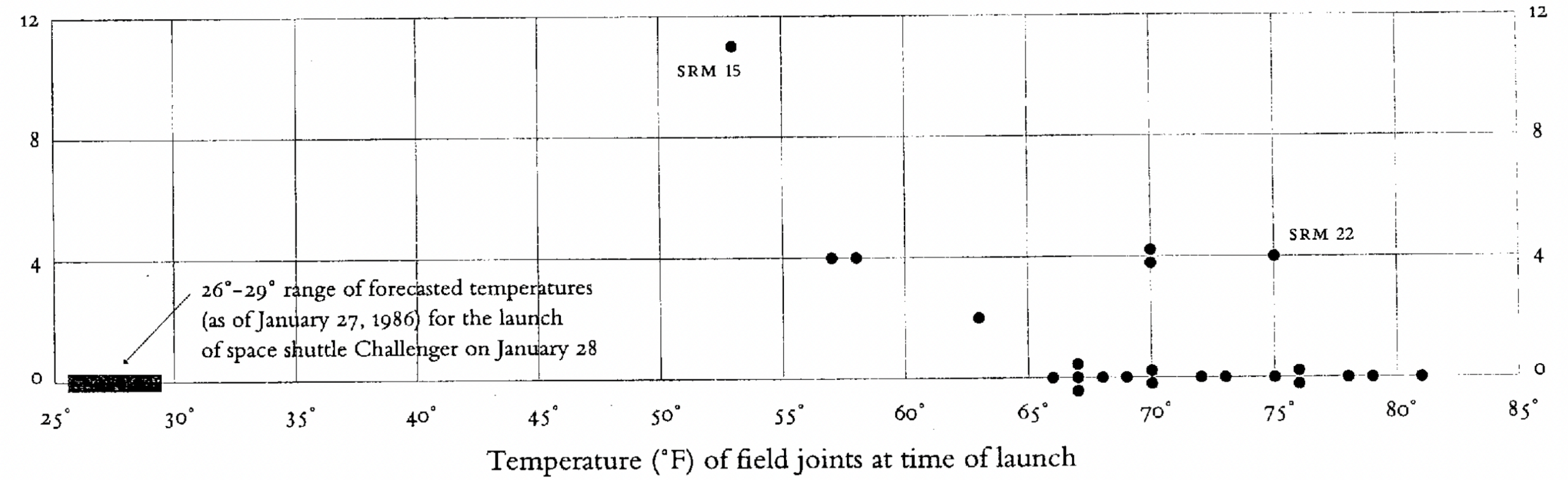
- Distracting patterns
- Unnecessary artistic embellishment

Shadows and background are useless distractions

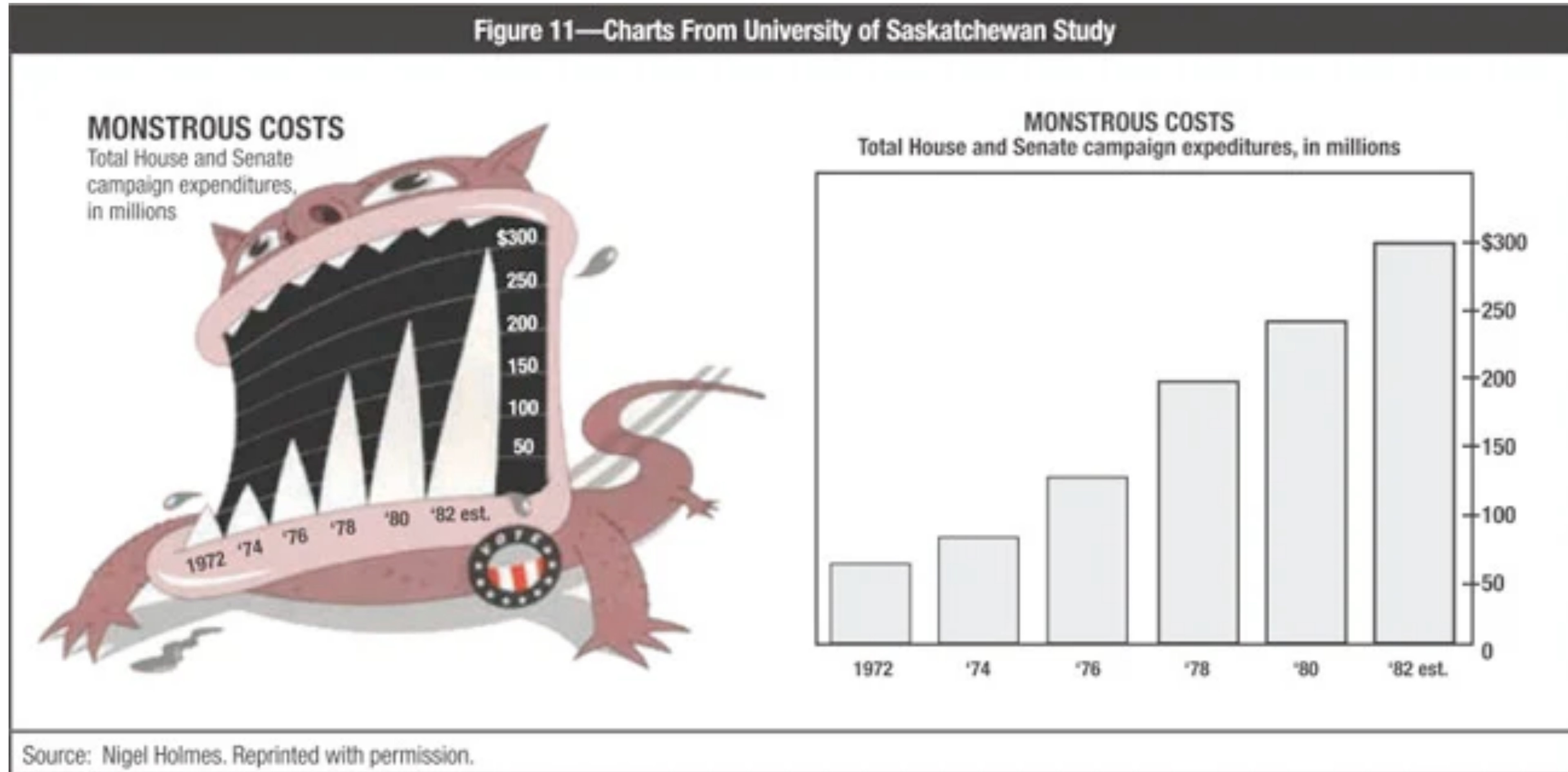
Unnecessary artistic embellishment



O-ring damage index, each launch



But is Tufte always right?



Two versions of the same plot, which is more intuitive?

Which is more memorable?

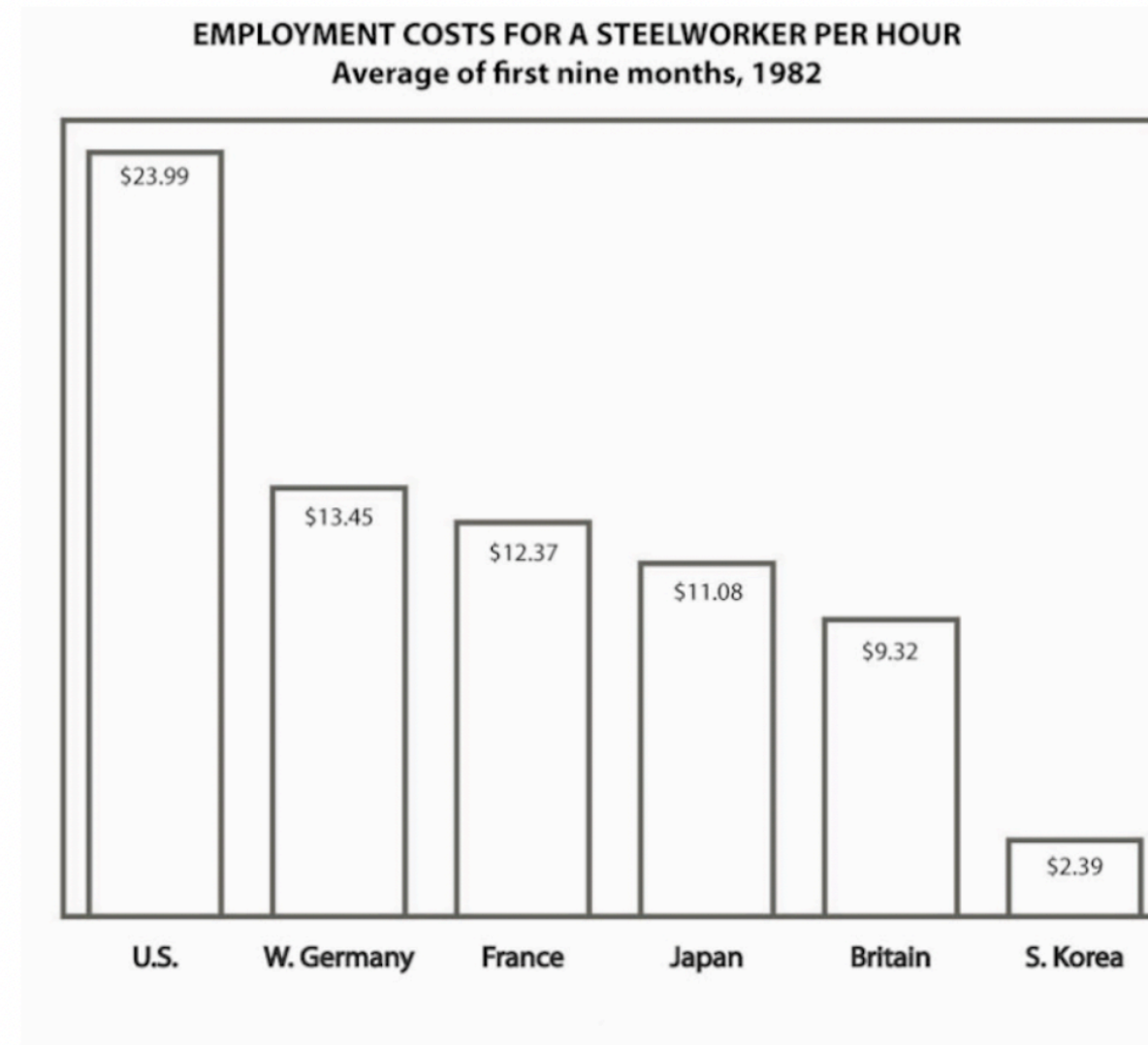
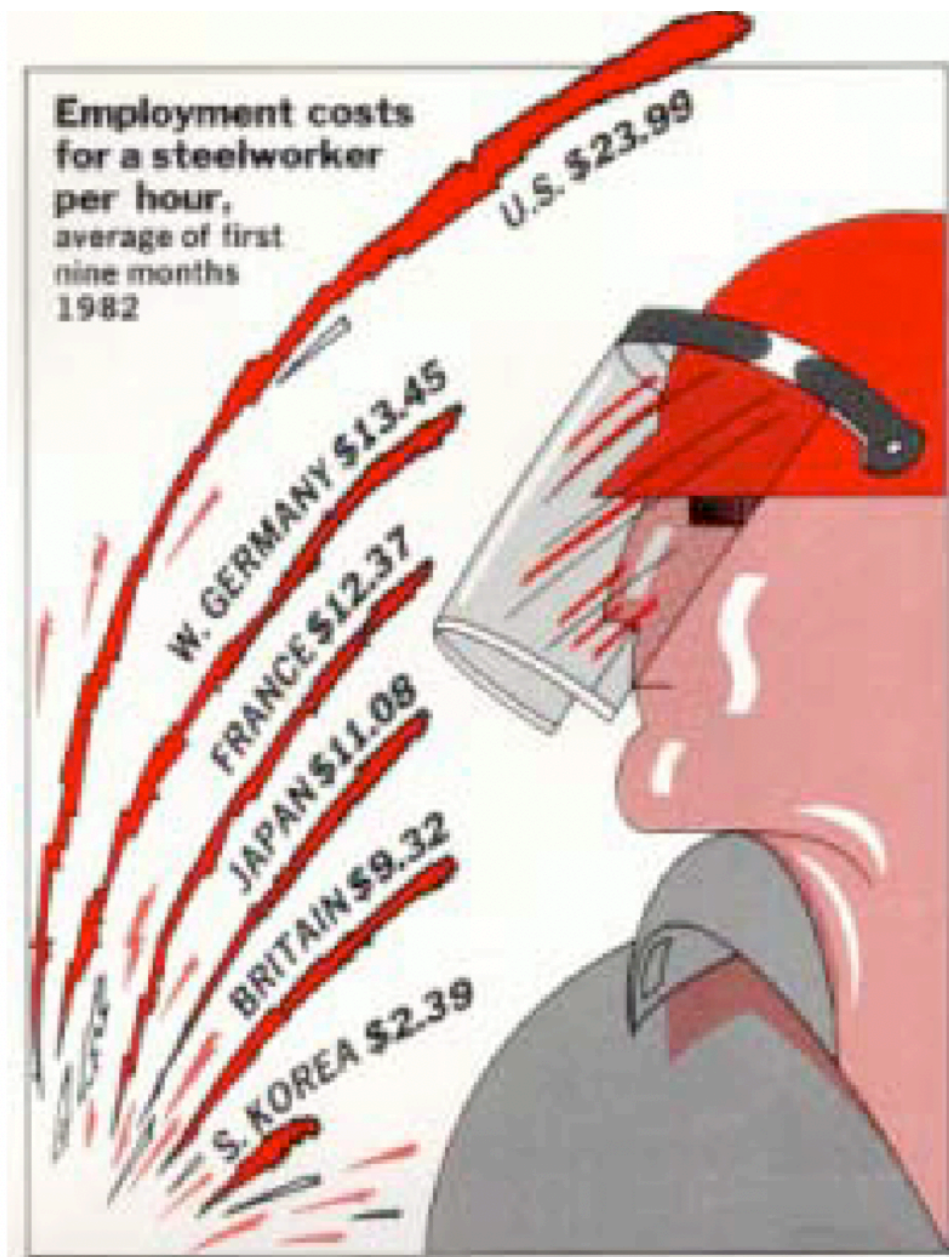
9 Criteria for good visualizations

- Expressiveness
- Effectiveness
- Intuitiveness
- Honesty
- Accessibility
- Impact
- Insight
- Writing
- Engineering

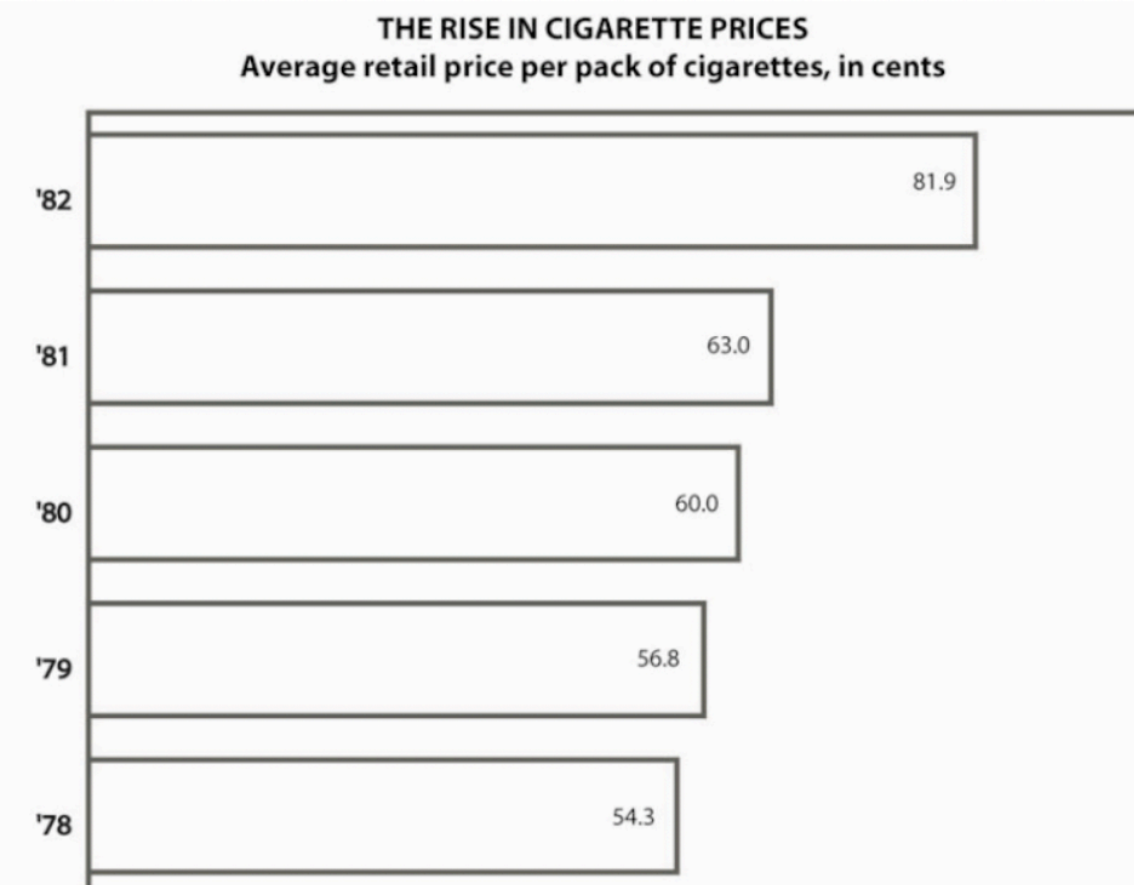
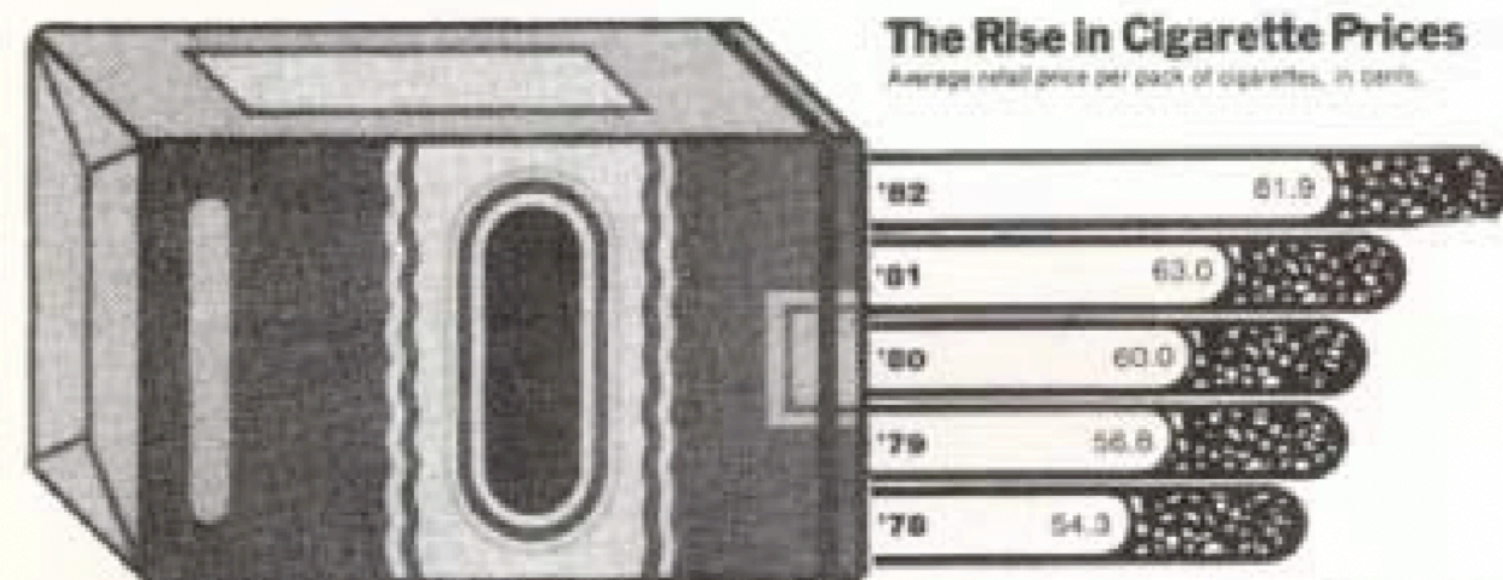
Impact

Will readers *internalize and remember* the information we present to them?

A perceptual study



Bateman et. Al. tested artistic visualizations by Nigel Holmes against simplified versions



A perceptual study

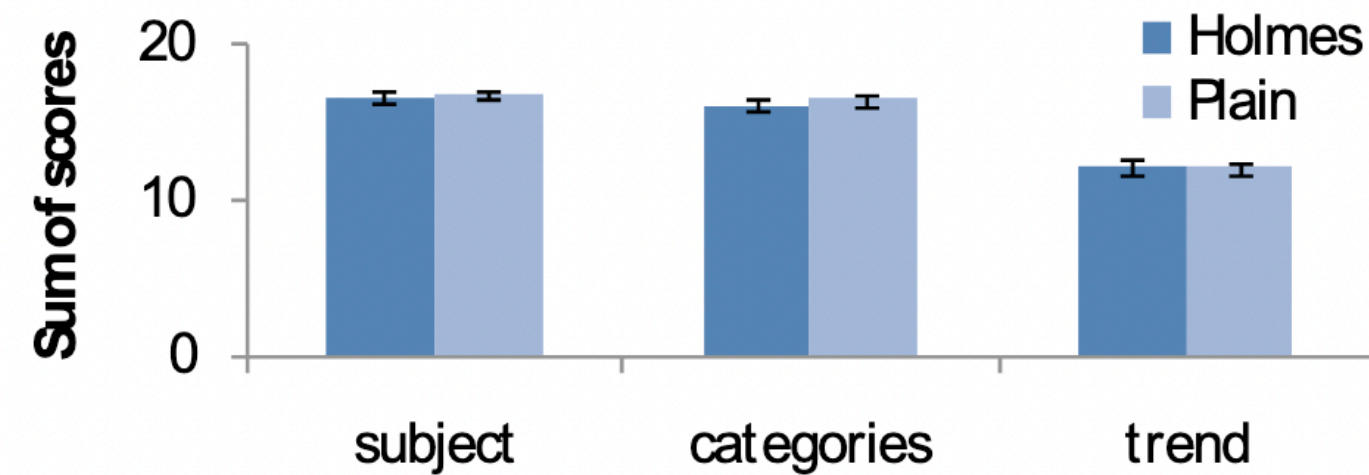


Figure 4. Means \pm SE for description scores.

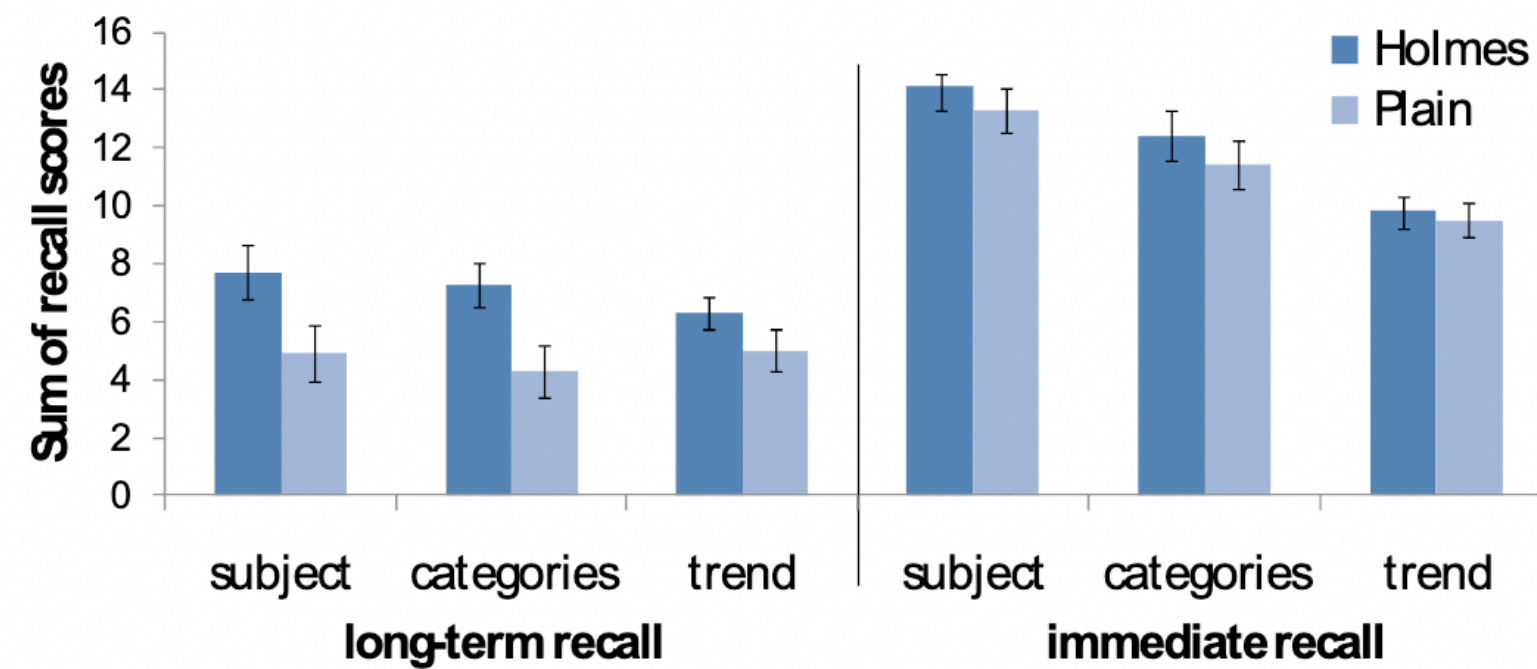


Figure 5. Means \pm SE for recall scores for long-term and immediate recall.

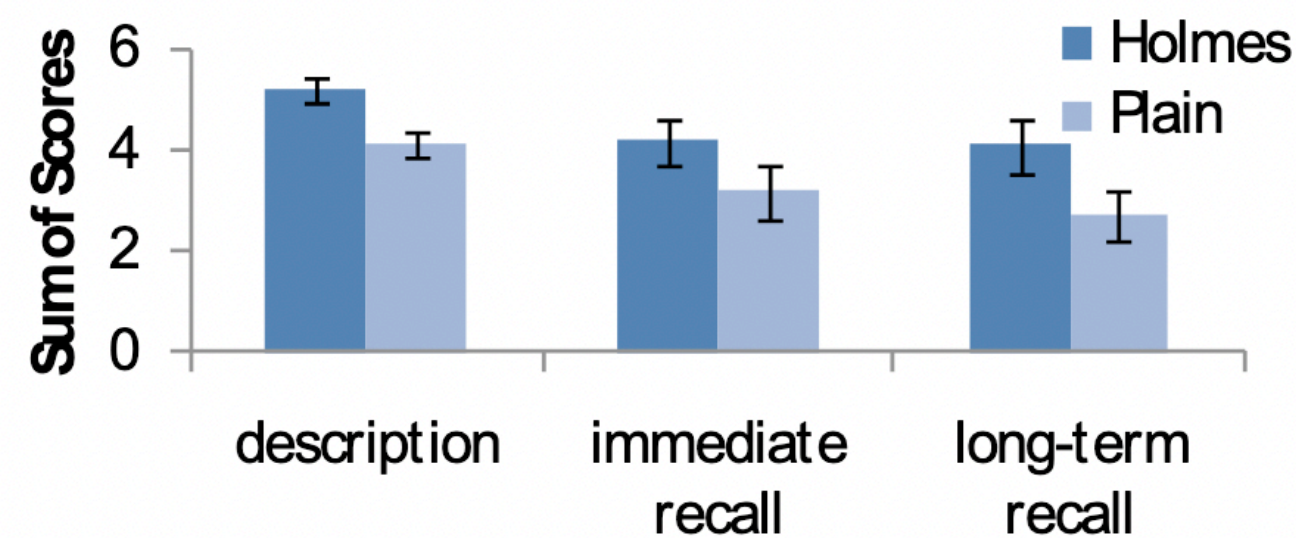
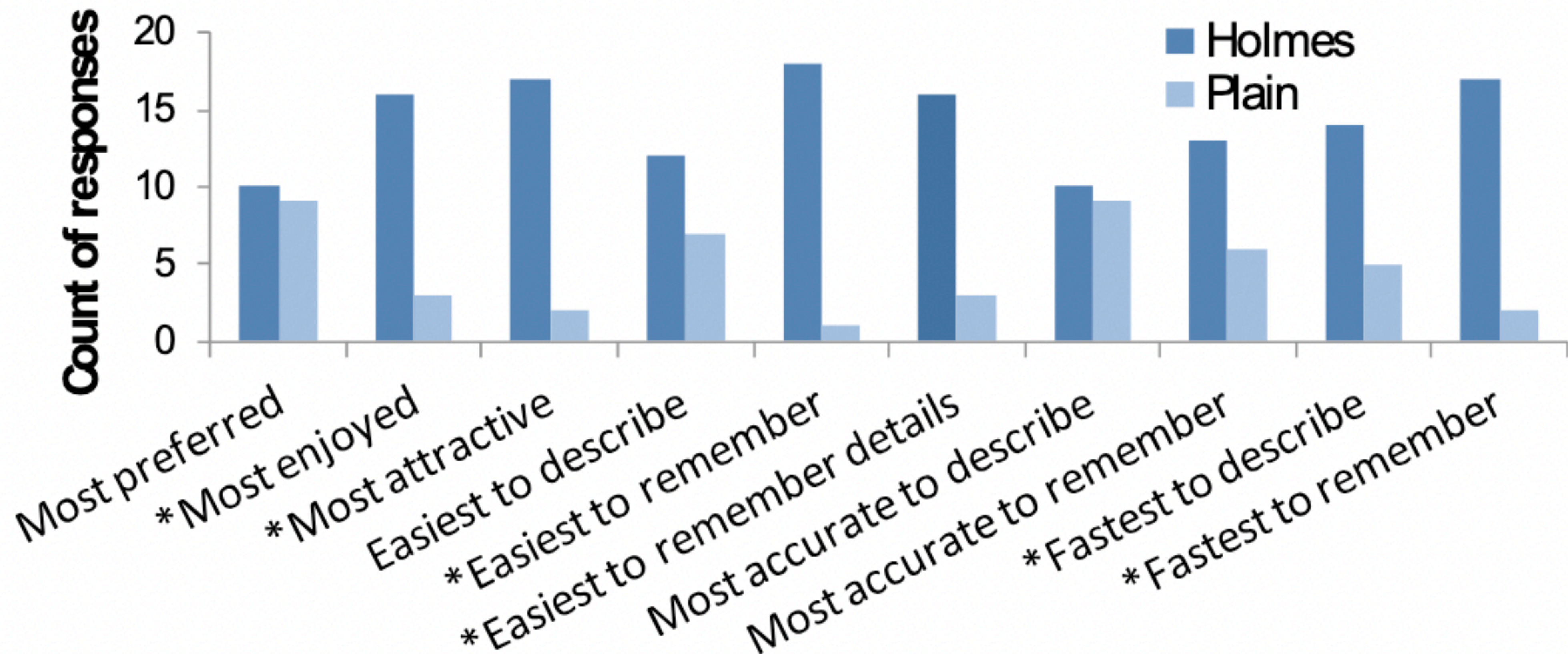


Figure 6. Means \pm SE for sum of value message scores.

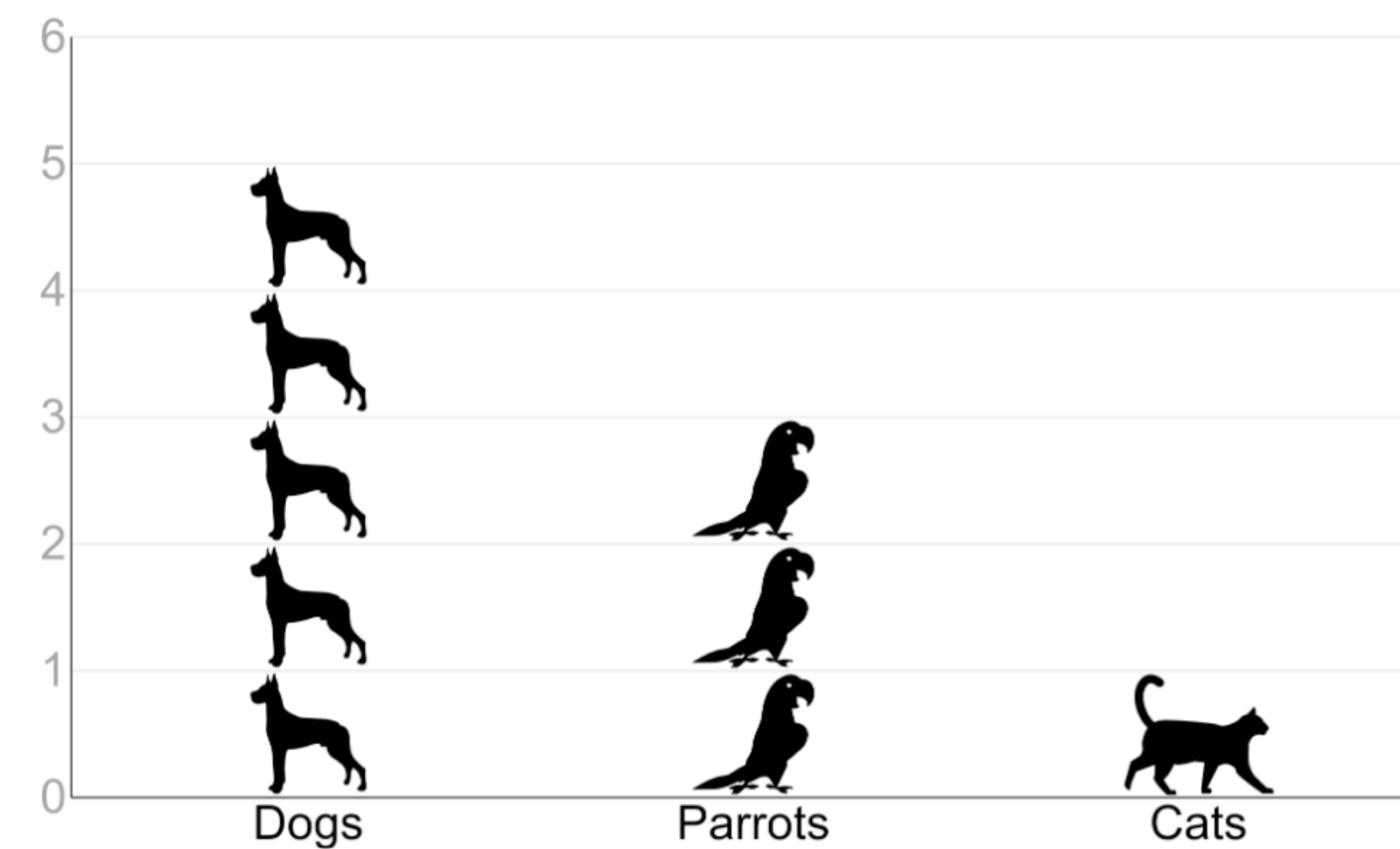
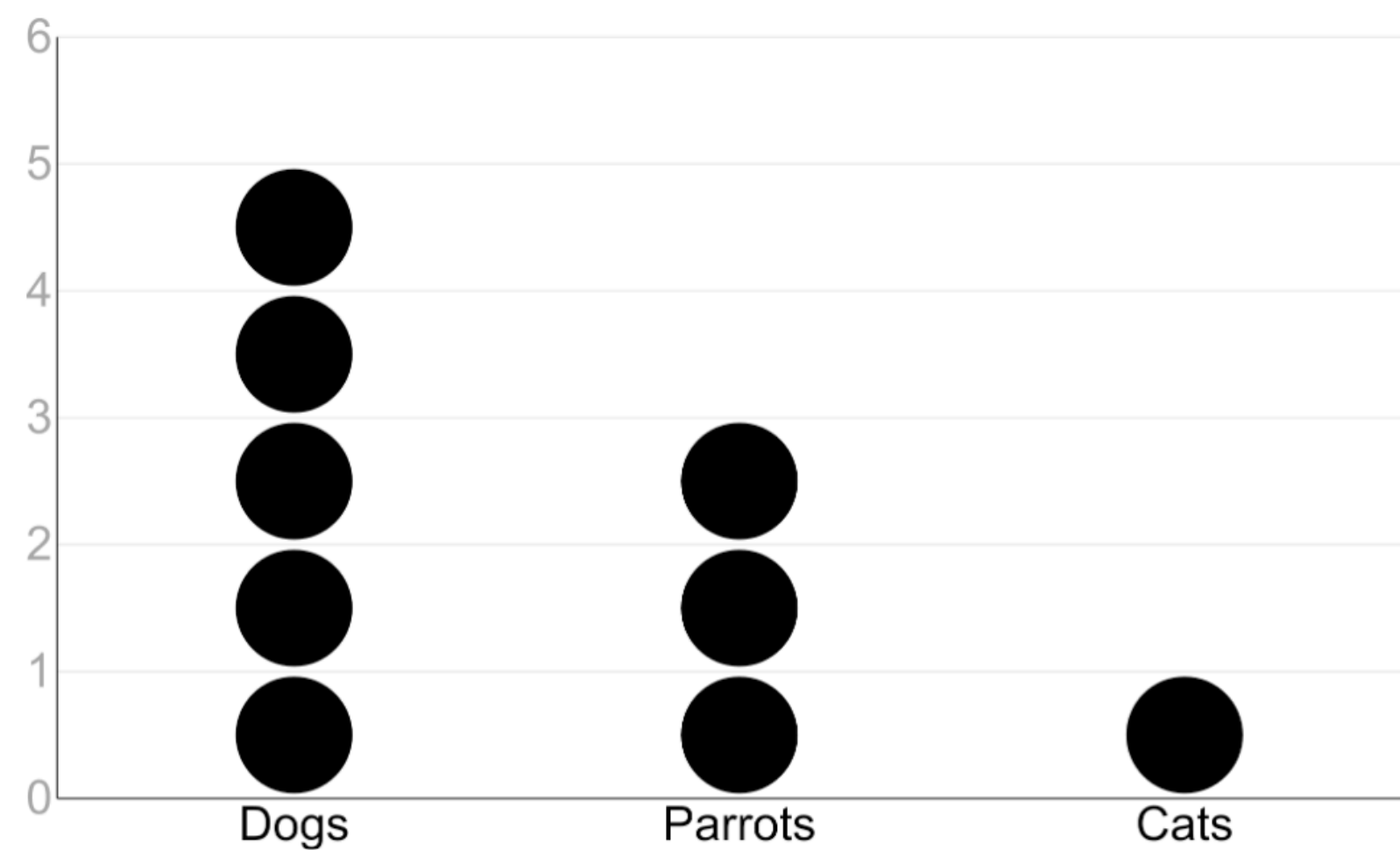
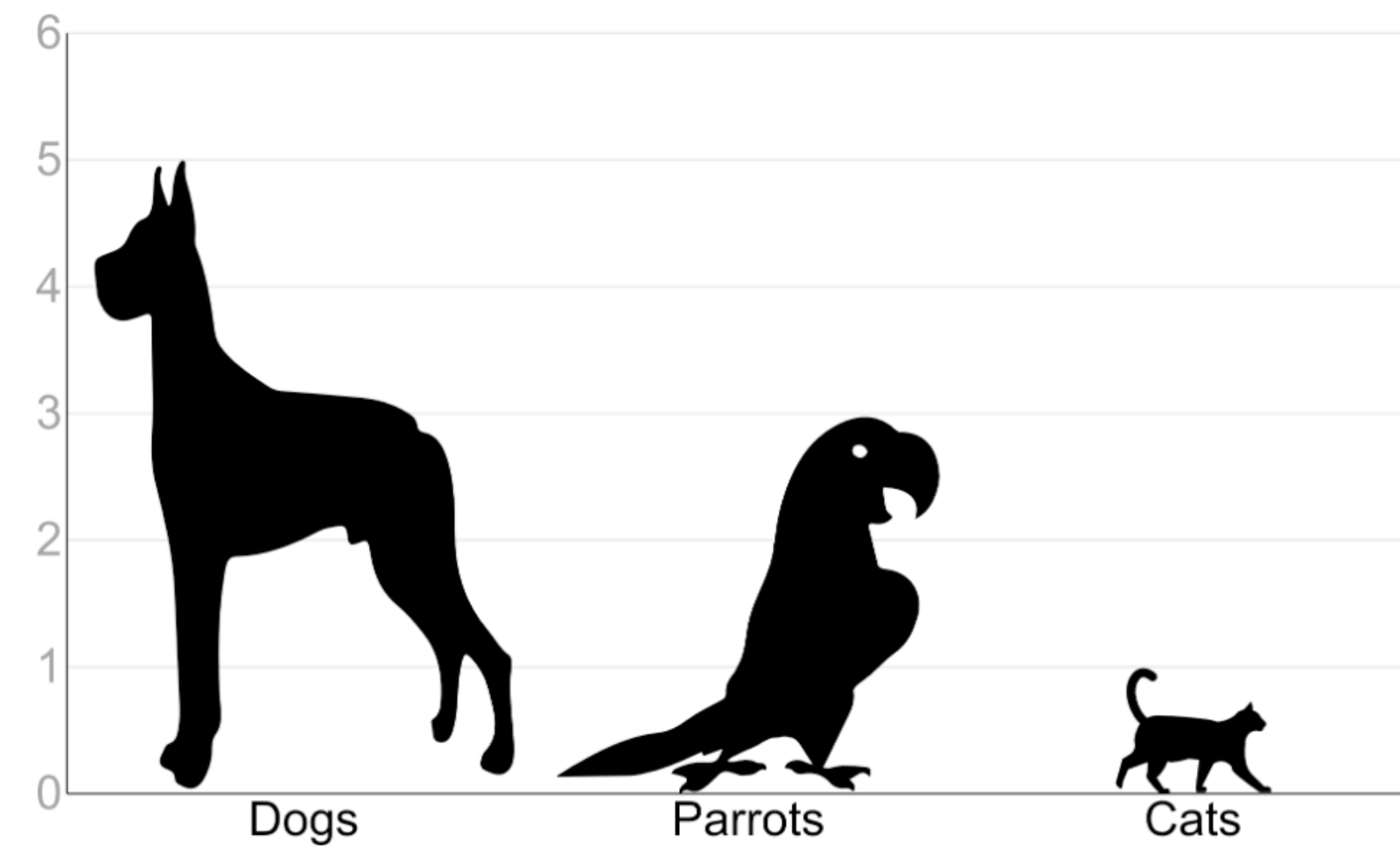
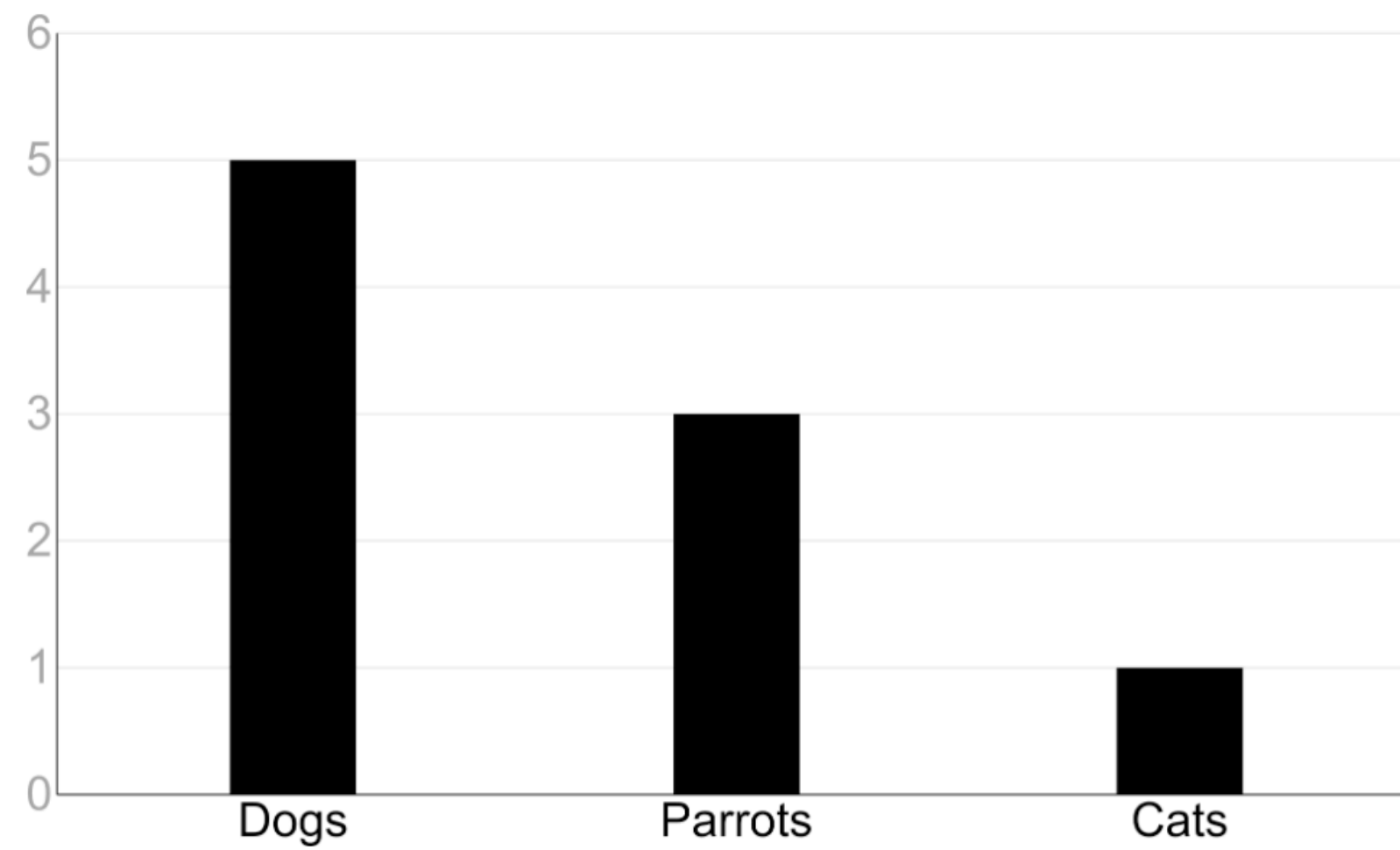
Results showed subject better remembered the information from plots with “Chart junk”!

More results

They also preferred them!

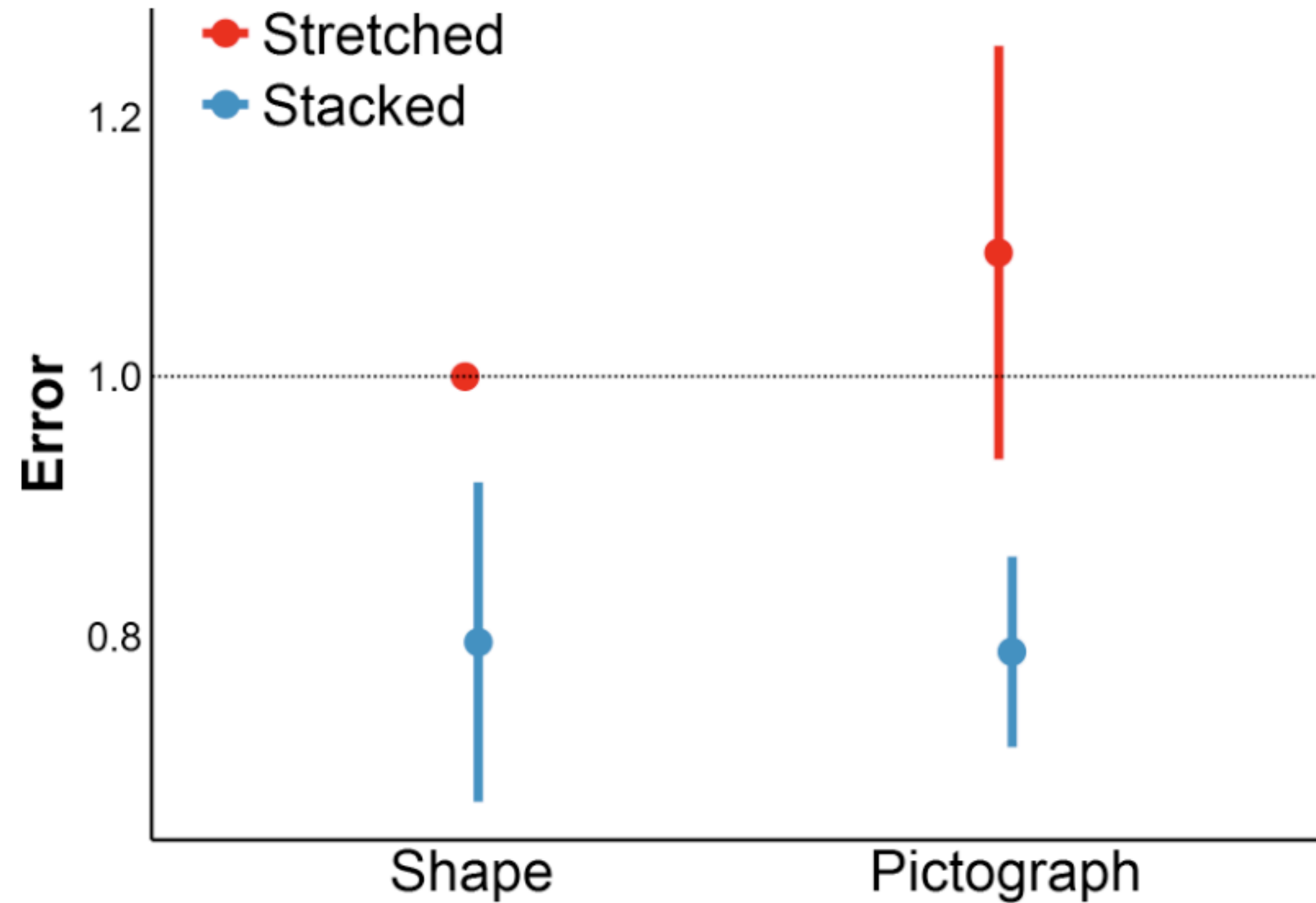


Pictograms

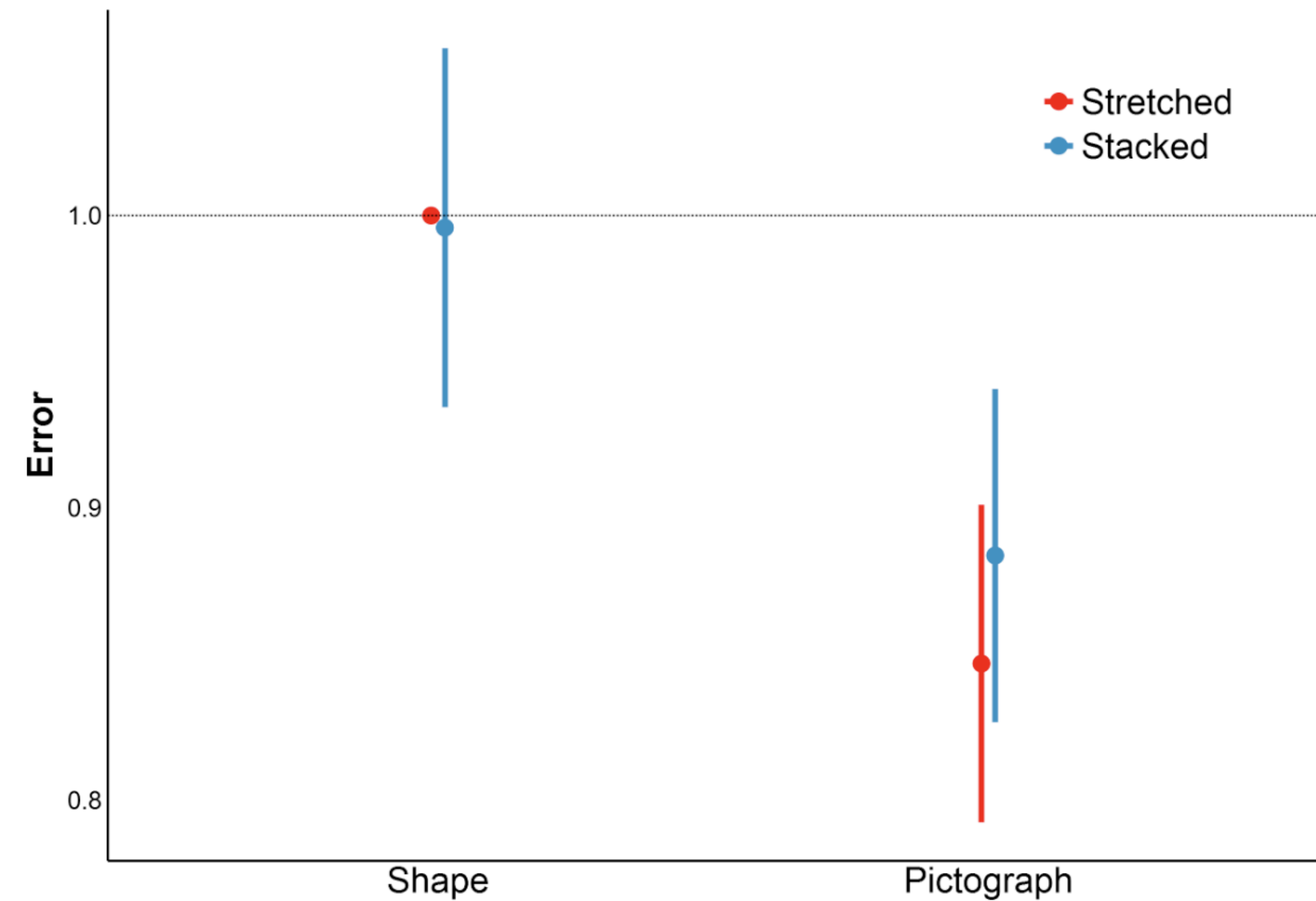


Haroz et. al. used a controlled study comparing pictographs to basic shapes.

Pictograms



Similar error (for stacked version)



Much better recall!

Caveat

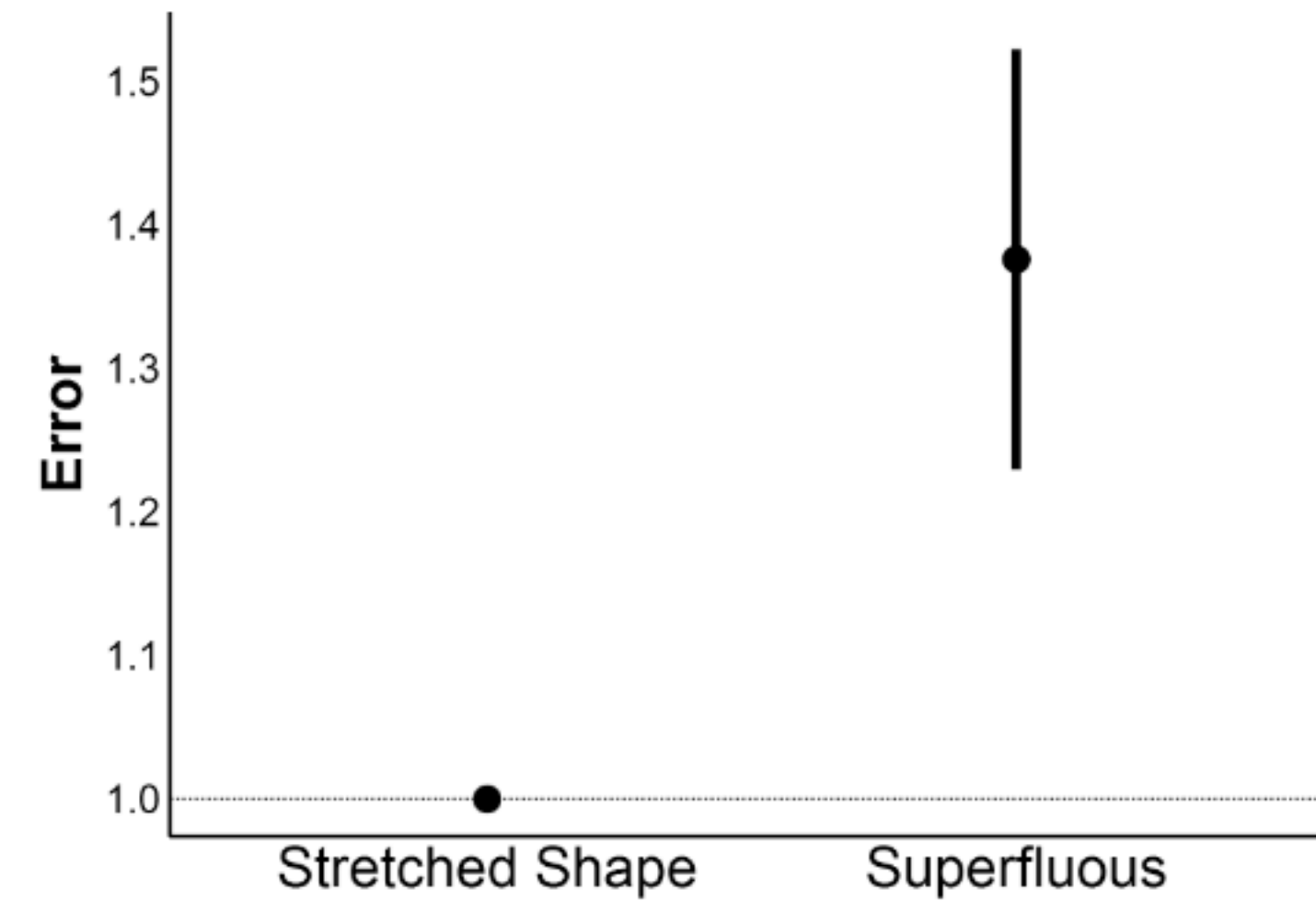
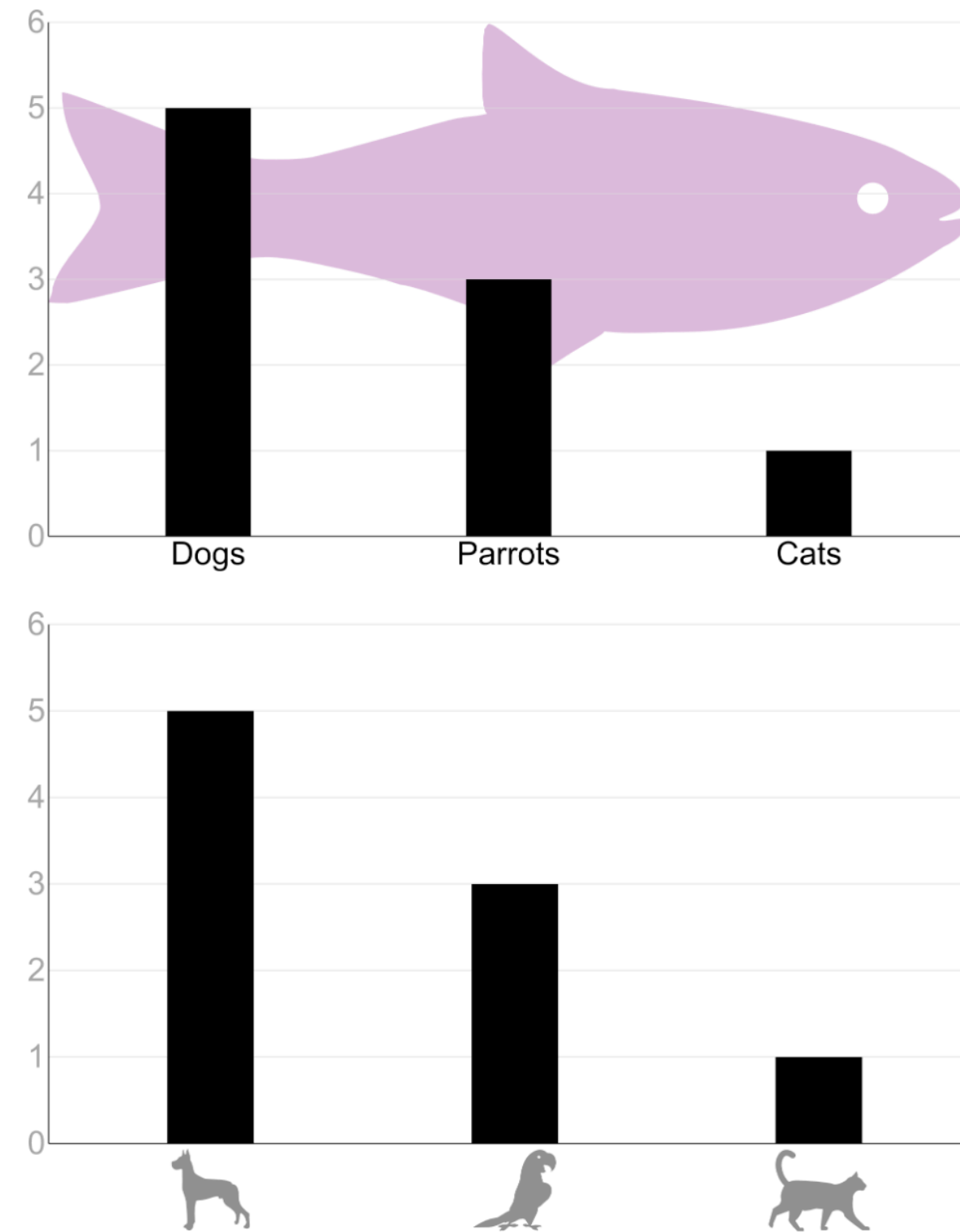
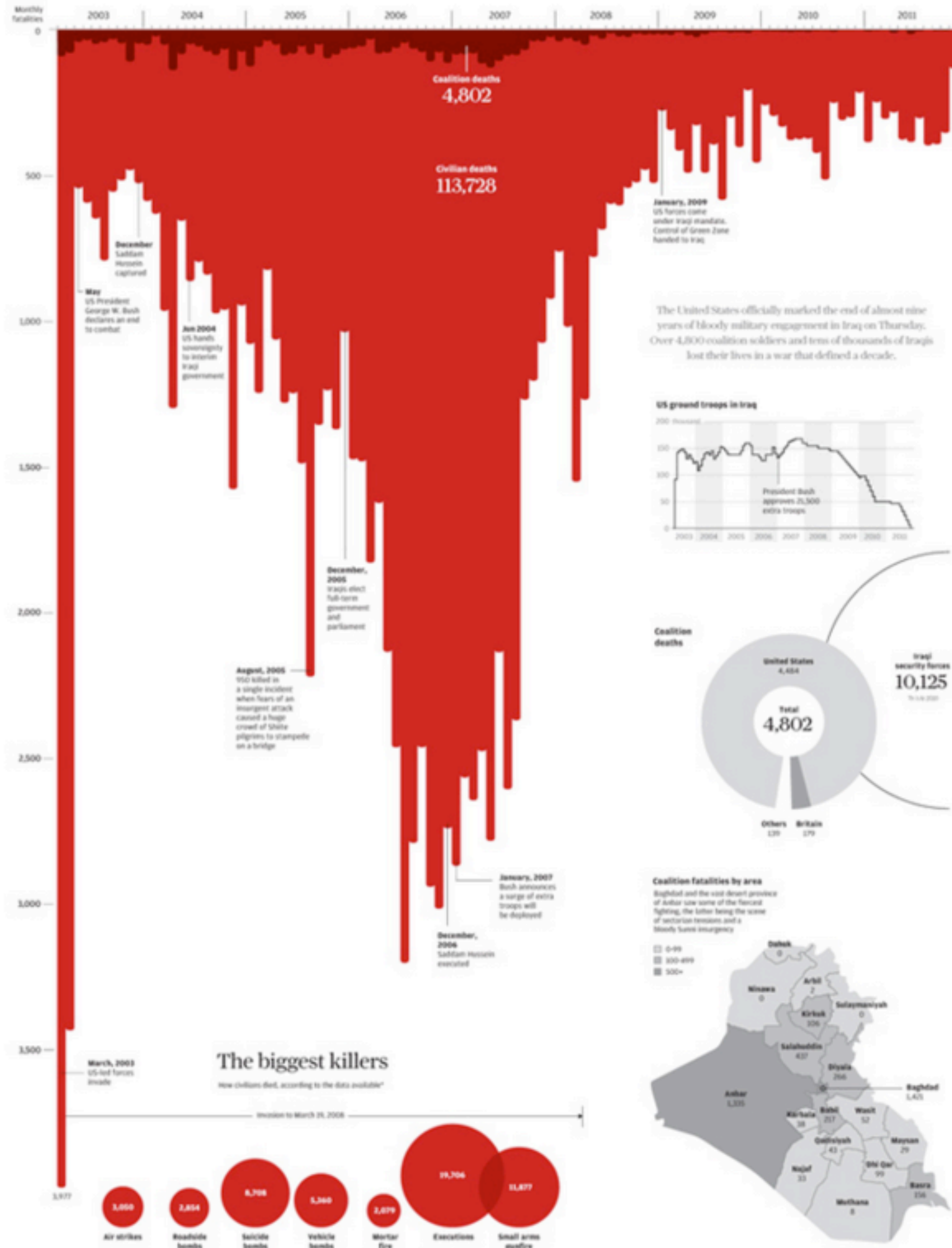


Fig. 7. An otherwise simple bar chart with a superfluous background image yields much higher error levels than an unembellished simple bar chart.

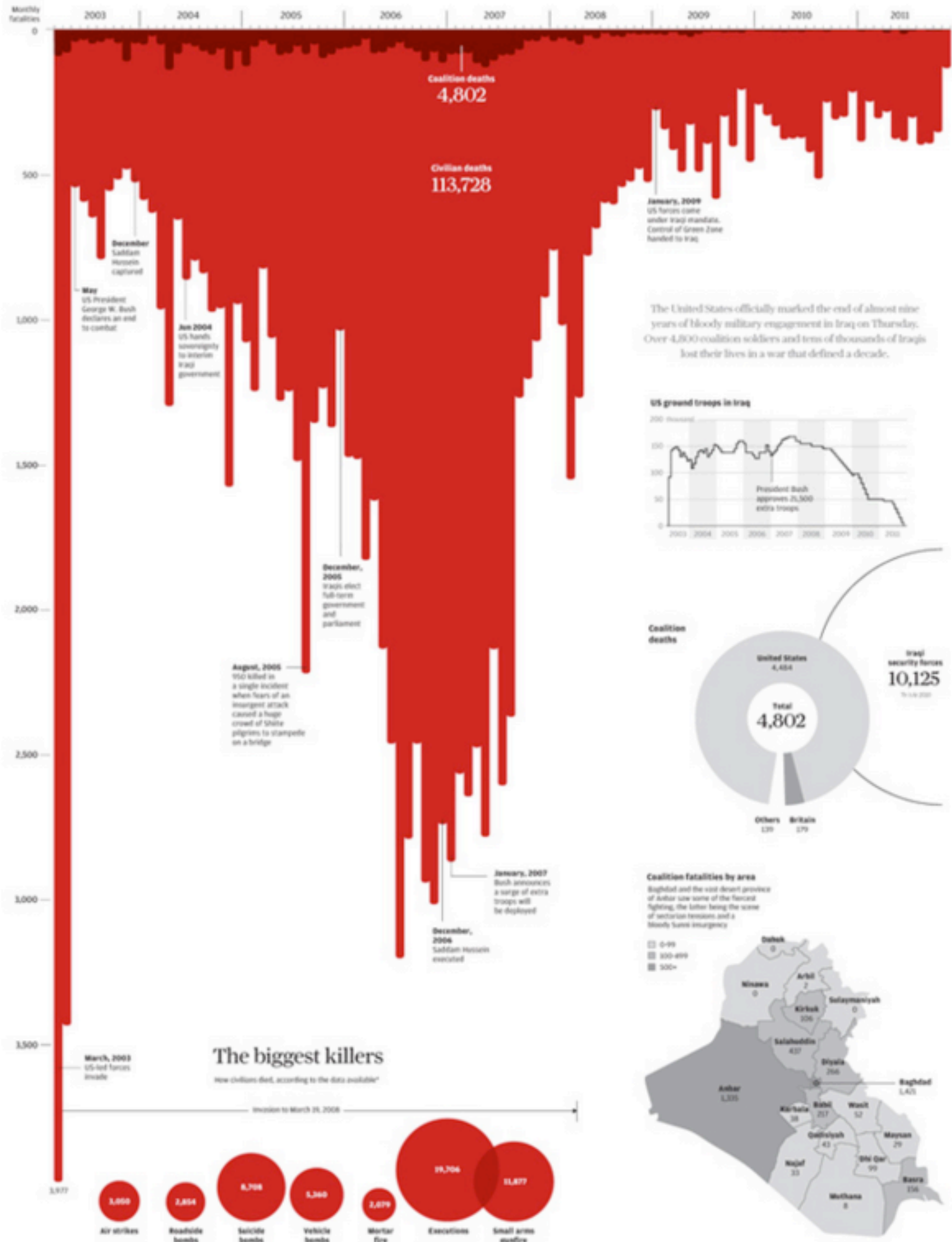
Superfluous pictograms increased error.

**Takeaway: Often a trade-off
between clarity and impact**

Iraq's bloody toll



Iraq's bloody toll



Iraq: Deaths on the decline

