Data Exploration & Visualization

Module 5

Design Principles

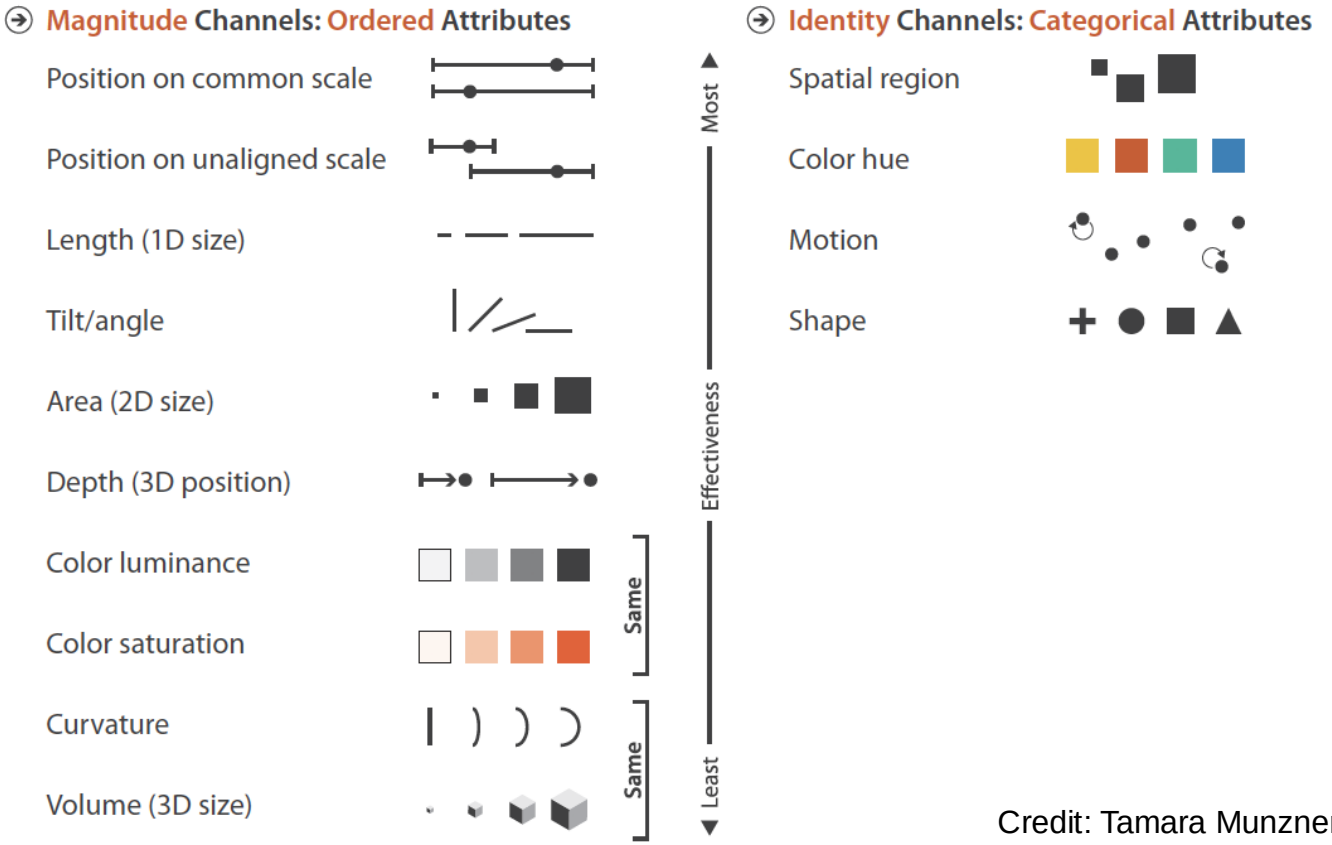
Dr. ZENG Wei

DSAA 5024

The Hong Kong University of Science and Technology (Guangzhou)

Effectiveness

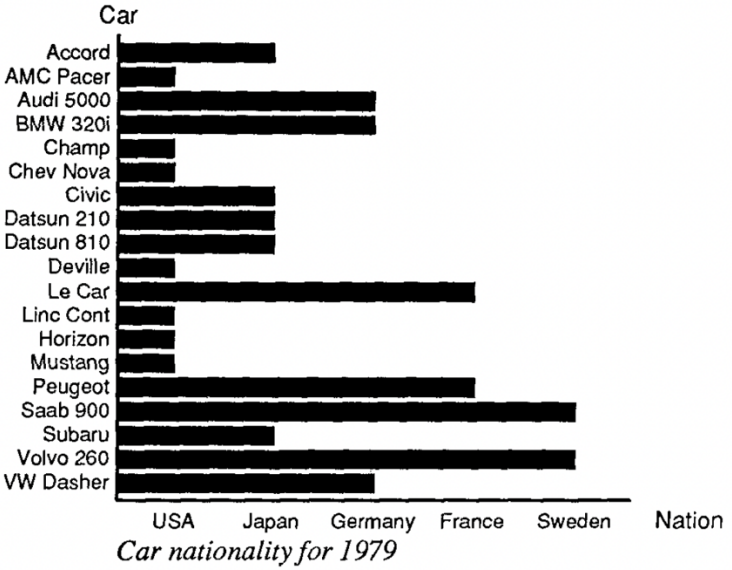
- Effectiveness rules: effective selection of visual channels (i.e., soft constraints)

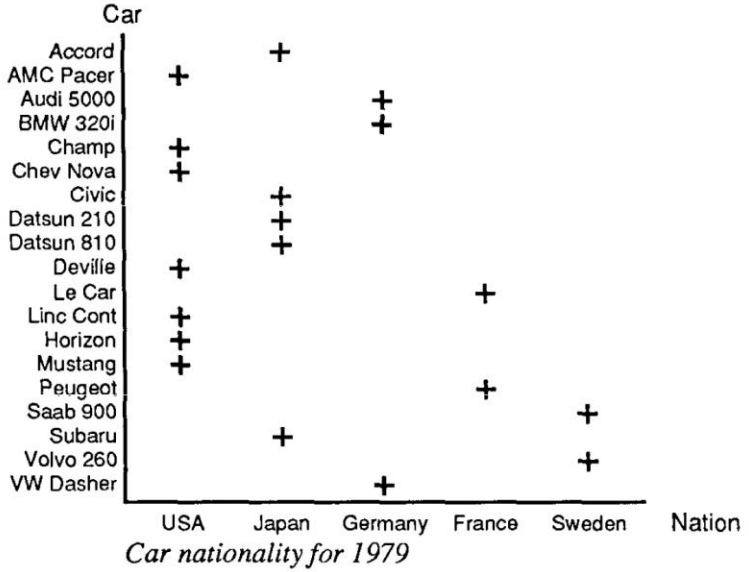


Credit: Tamara Munzner

Expressiveness

- Expressiveness rules: correct usage of visual channels (i.e., hard constraints).





J. Mackinlay, "Automating the design of graphical presentations of relational information," *ACM Trans. Graph.,* vol. 5, no. 2, pp. 110-141, 1986.

Data Exploration & Visualization

Module 5: Design Principles

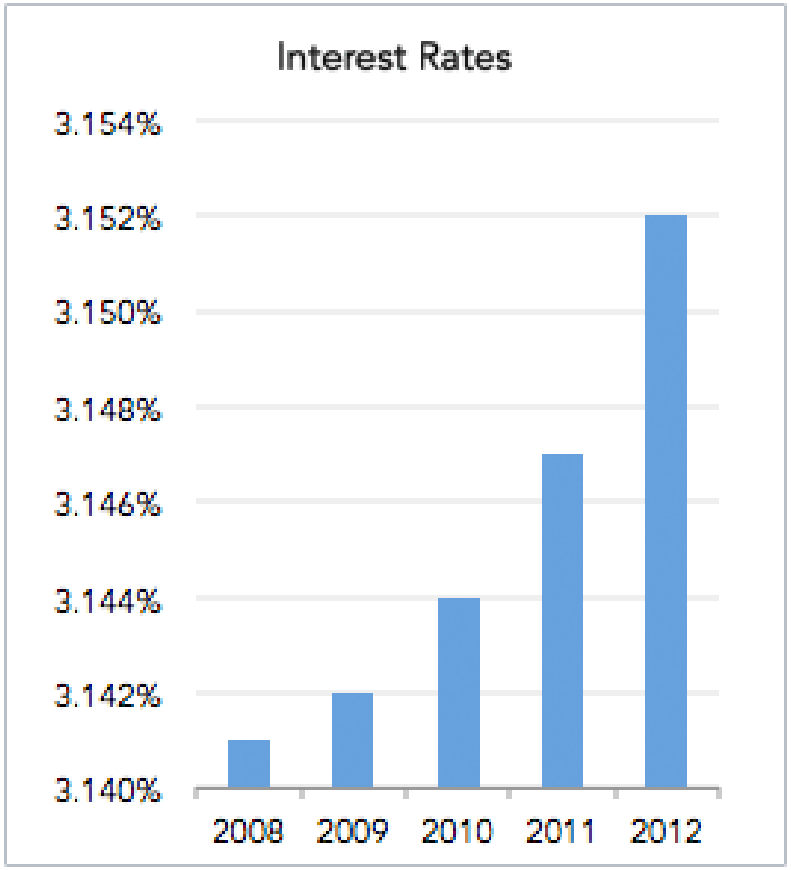
- Integrity principles
  - Not to lie with data visualization
- Tufte’s rules
- Chart-junk debate
- Nested model

Integrity principles

- **Graphical integrity**: visual representations of data must tell the truth
- Not to lie with data visualization
  - Common lies in data visualization
    - Truncated Y-Axis
    - Cumulative graphs
    - Ignoring conventions
    - Inconsistent scales
    - Size & volume encoding

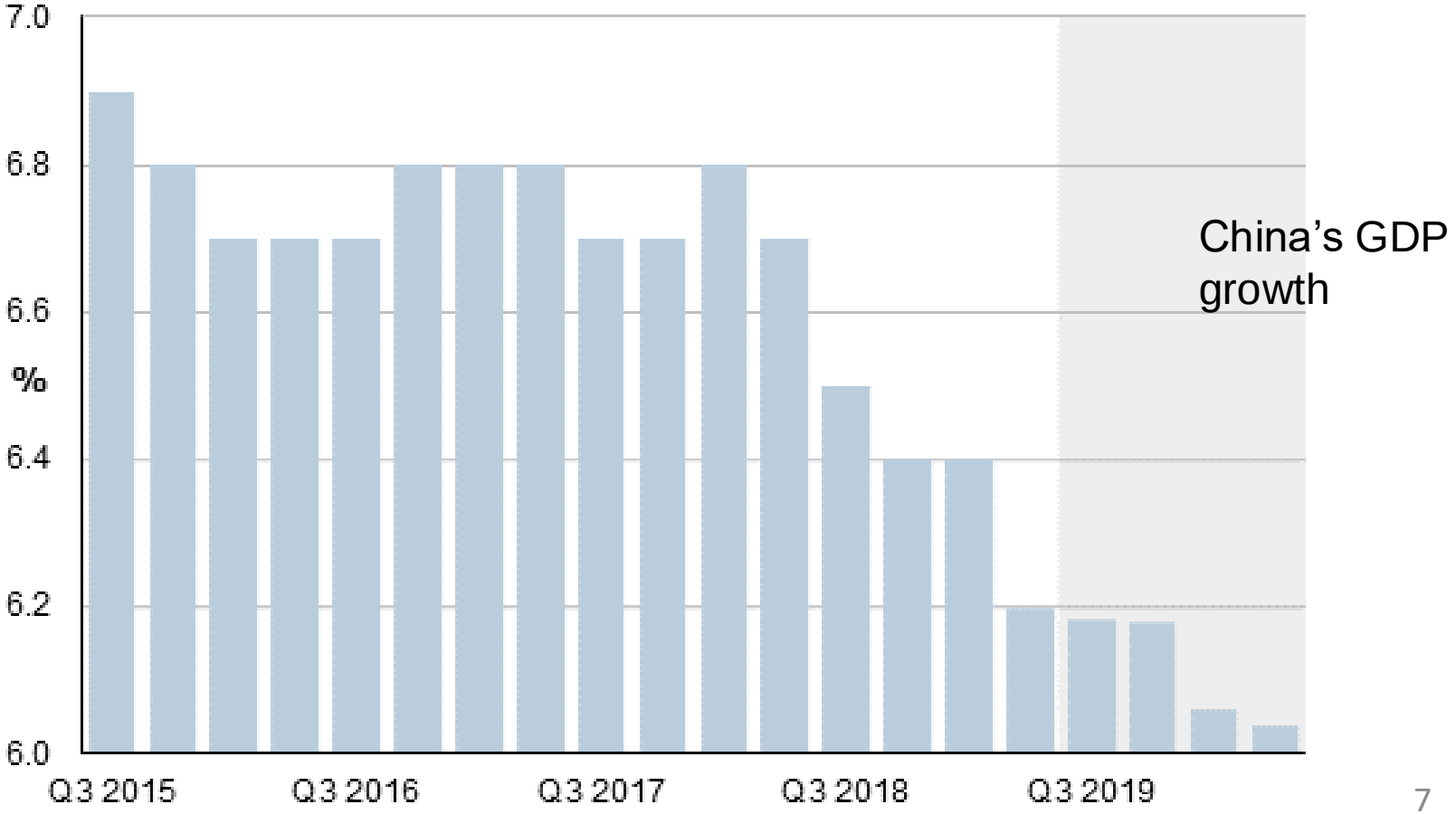
Truncated Y-Axis

- Same data, different Y-Axis



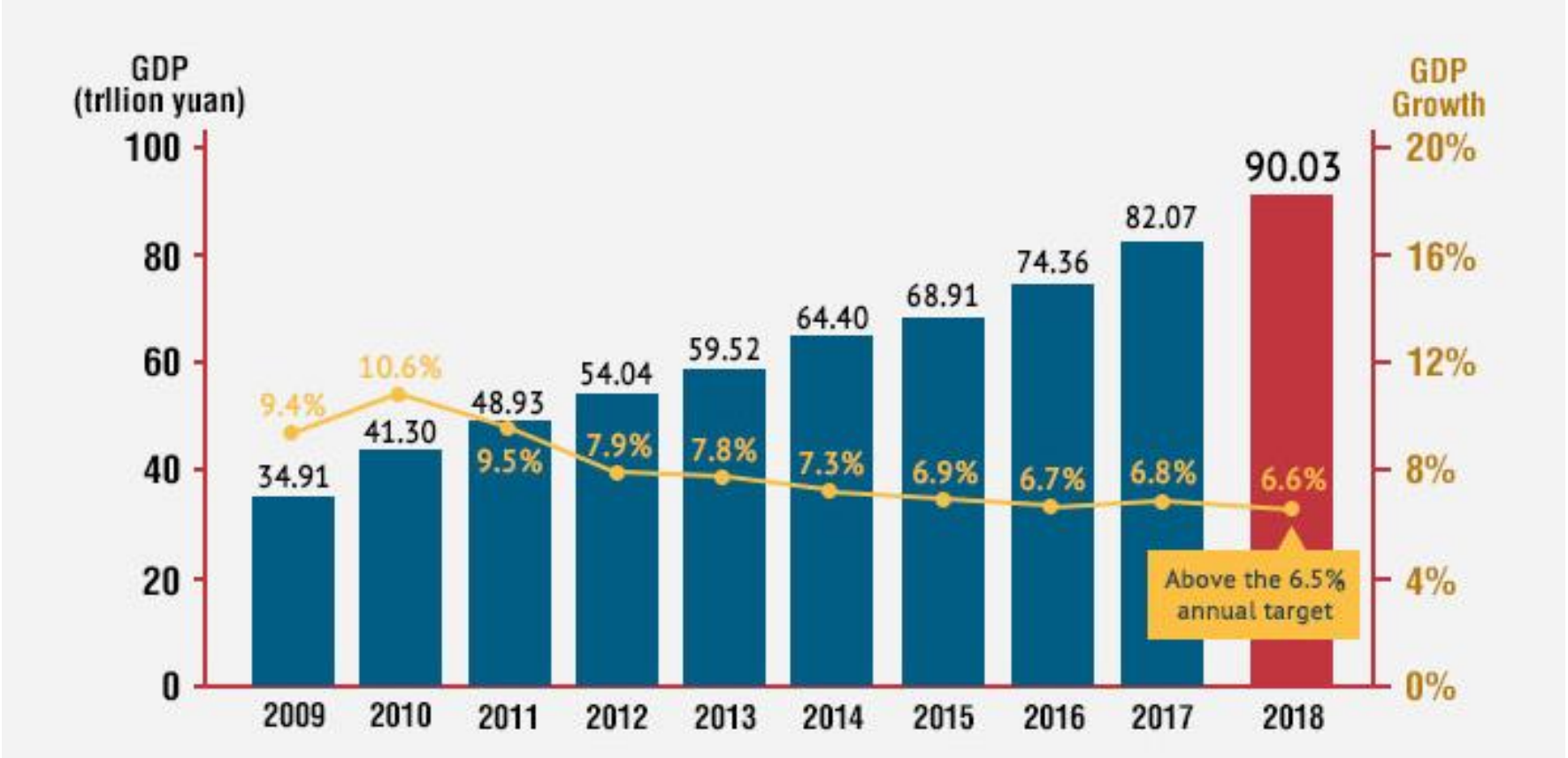
Truncated Y-Axis

- China economy crash?



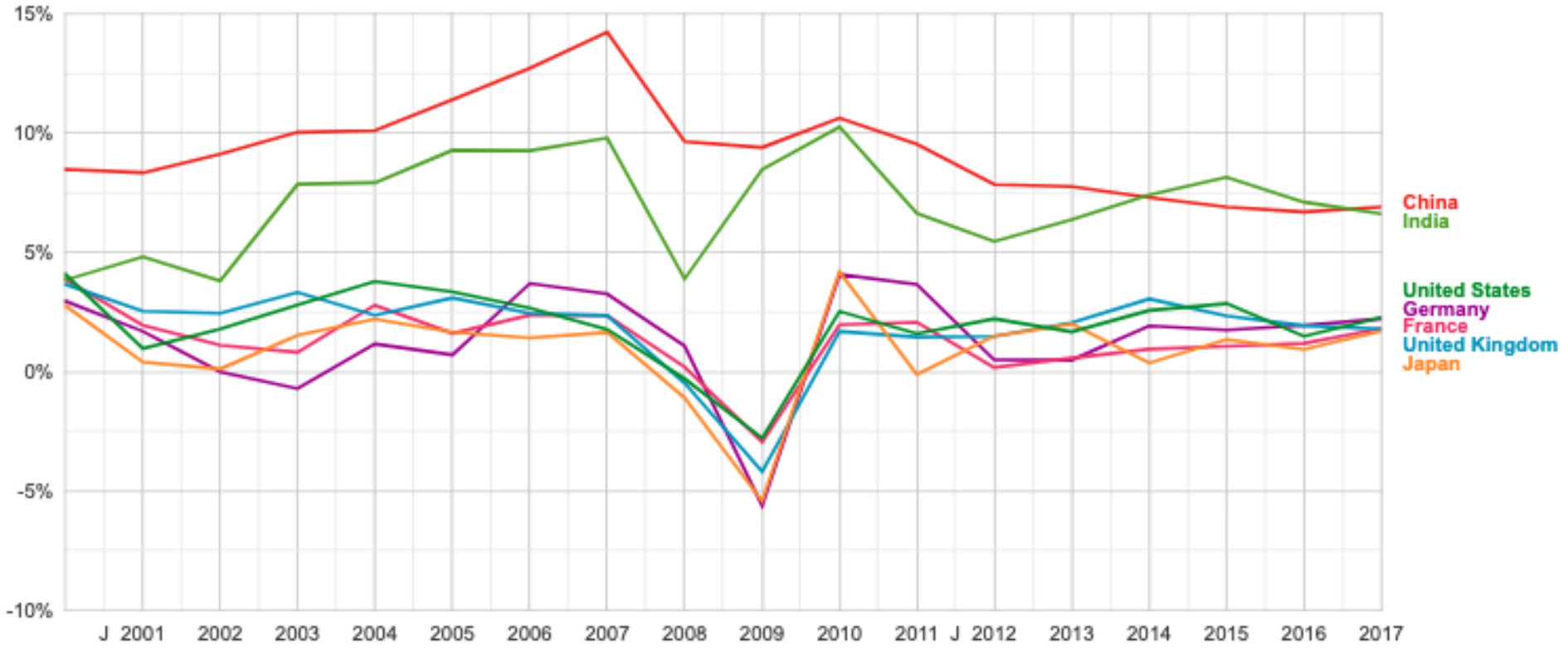
Truncated Y-Axis

- China economy crash?
  - Shown in entire scale



Truncated Y-Axis

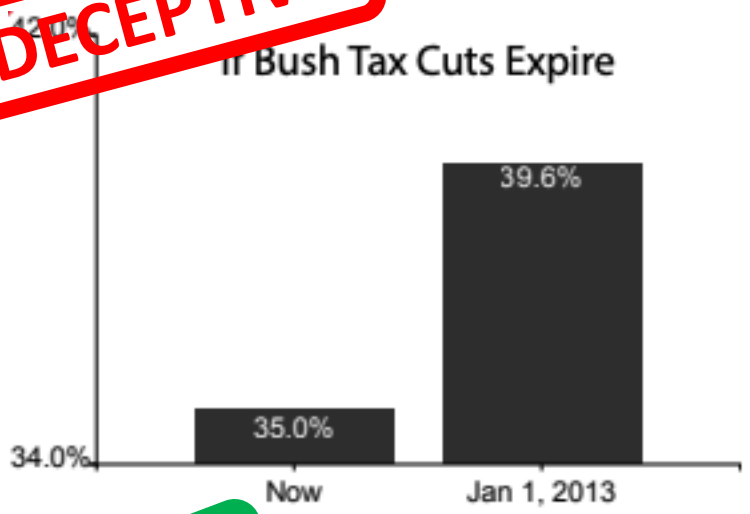
- China economy crash?
  - Shown in context



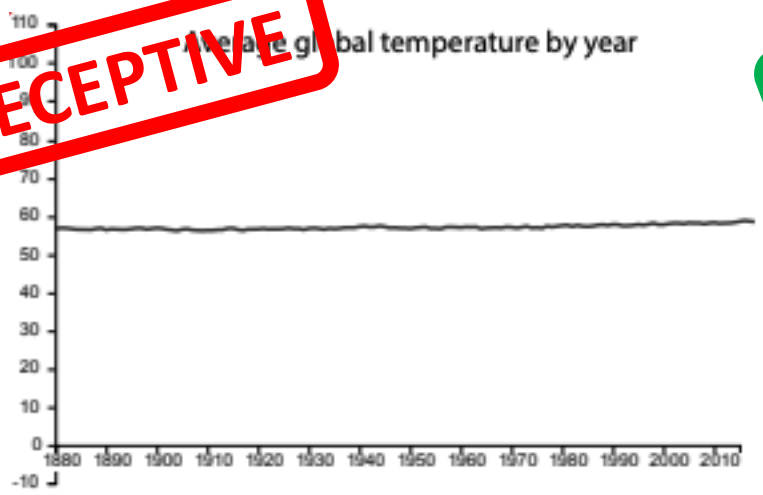
Truncating the Y-Axis

HONEST

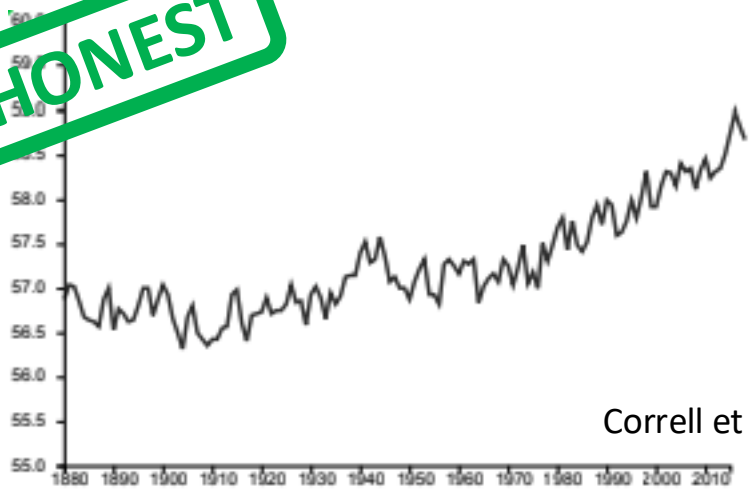
DECEPTIVE



DECEPTIVE



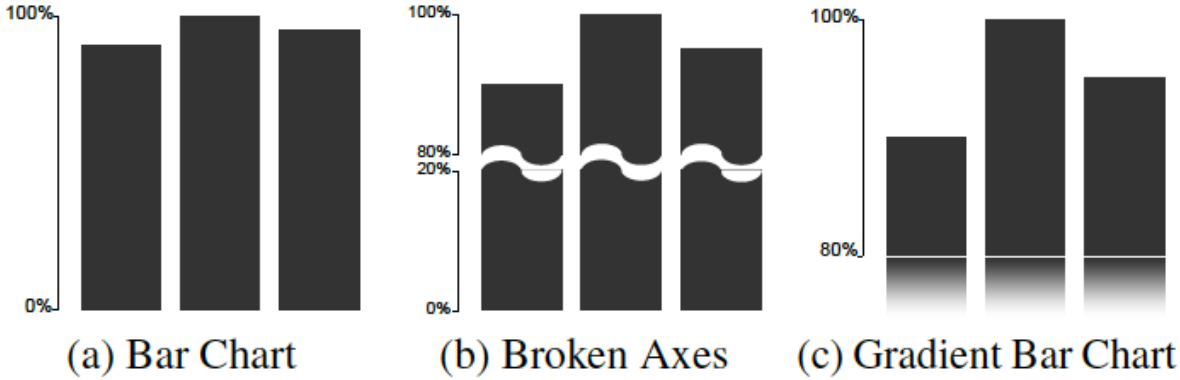
HONEST



Correll et al., 2020

Truncating the Y-Axis

- Y-axis truncation can be beneficial and harmful
  - Depending on the communicative and analytic intent.
  - A consistent and significant impact on the perceived importance of effect sizes.



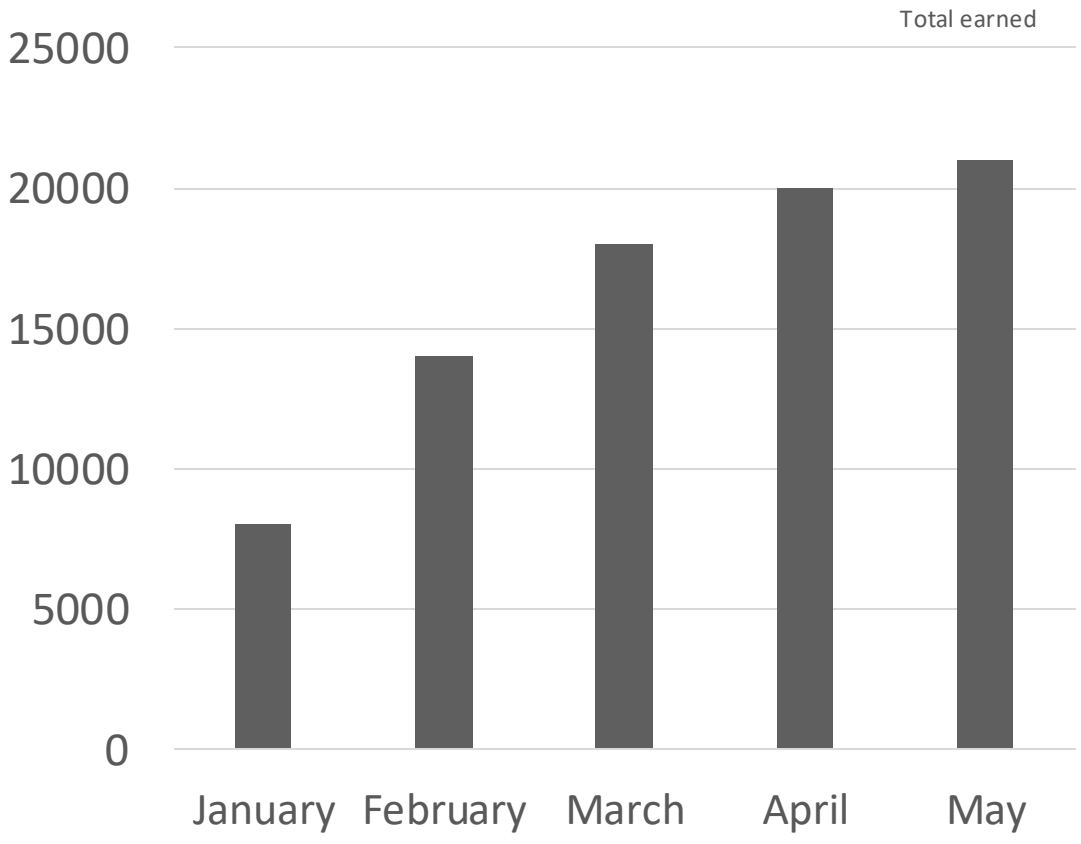
(a) Bar Chart (b) Broken Axes (c) Gradient Bar Chart

  - Visual indicators of broken or truncated axes are not significantly helpful.

Correll et al., 2020

Cumulative graphs

- A cumulative graph is a graph plotted from a cumulative frequency table.



Cumulative graphs

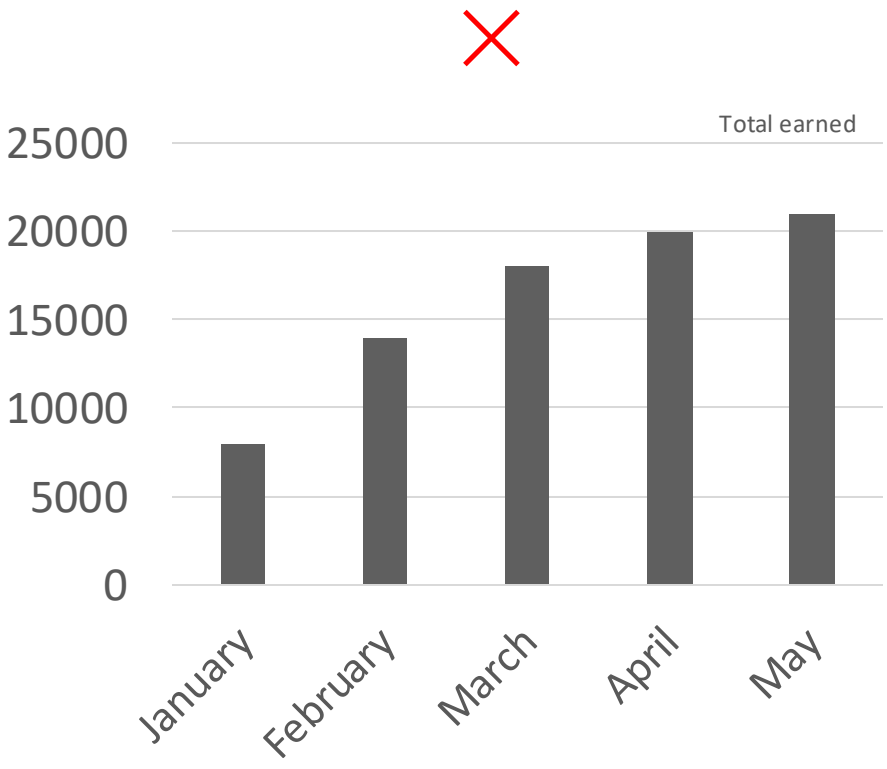
- A cumulative graph is a graph plotted from a cumulative frequency table.

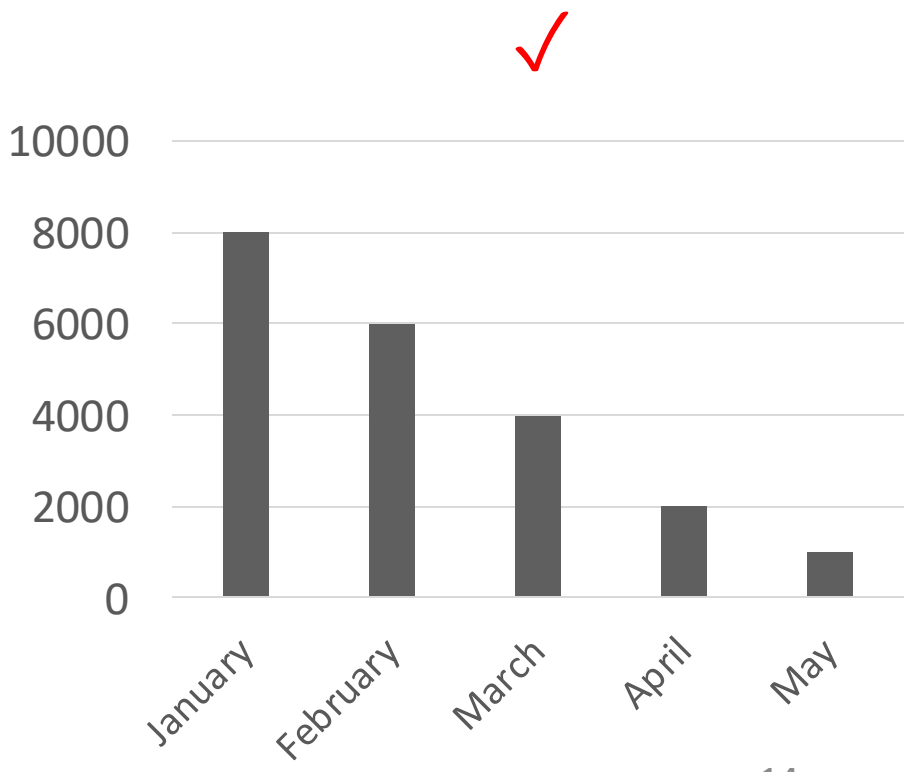
Month	Earned
January	8,000
February	6,000
March	4,000
April	2,000
May	1,000

Month	Total Earned
January	8,000
February	14,000 (8,000 + 6,000)
March	18,000 (14,000 + 4,000)
April	20,000 (18,000 + 2,000)
May	21,000 (20,000 + 1,000)

Cumulative graphs

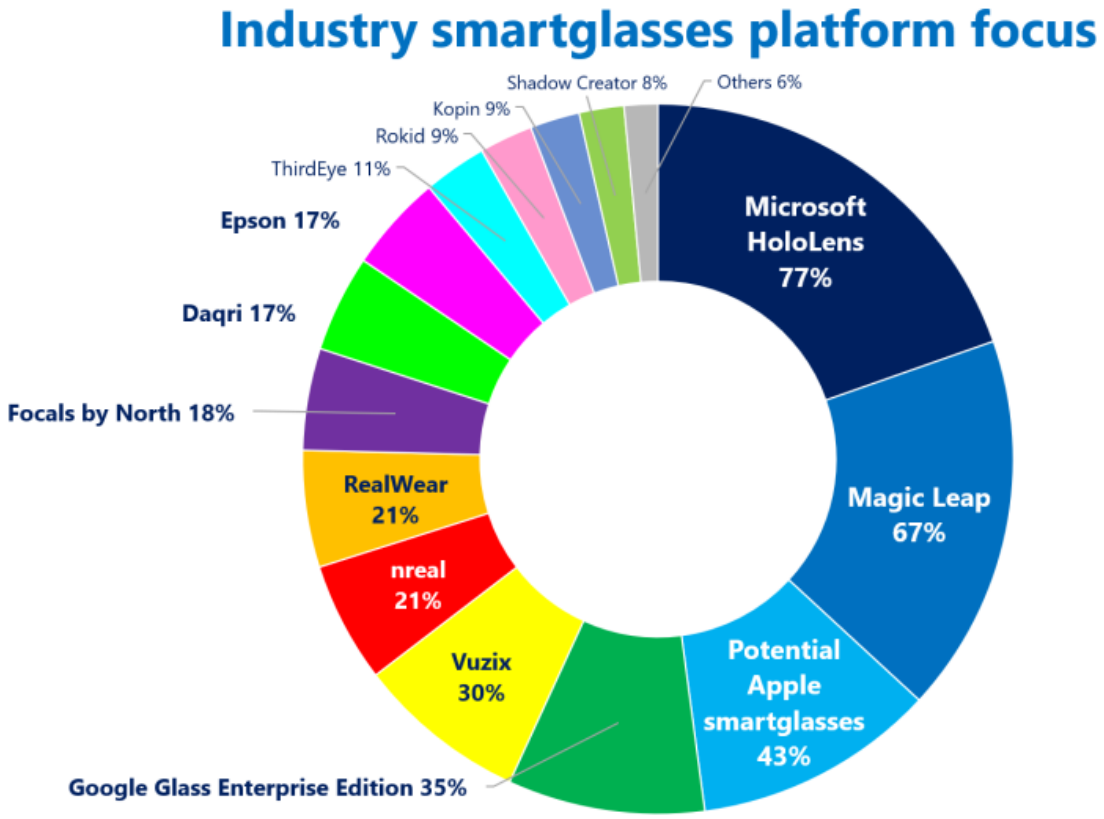
- A cumulative graph is a graph plotted from a cumulative frequency table.





Ignoring conventions

- Standard practices
  - pie charts represent parts of a whole



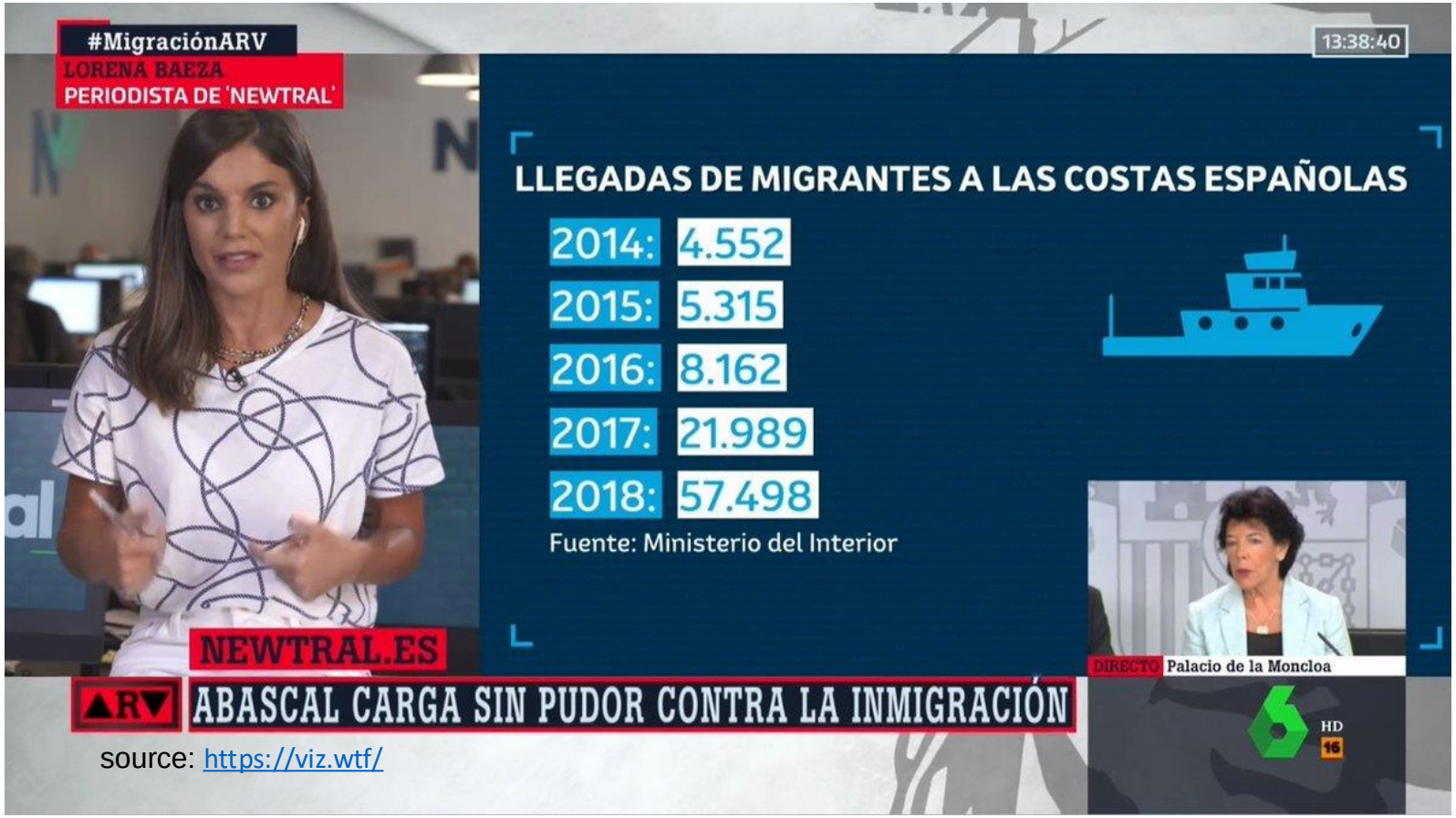
Digi-Capital

Notes: Percentages not cumulative, "Others" include DreamGlass, EverySight, Glassear, IMMY, Iristick, Lenovo, Letinar, Lumus, Optinvent, Realmax, Snap Spectacles, Tilt Five

awe

Ignoring conventions

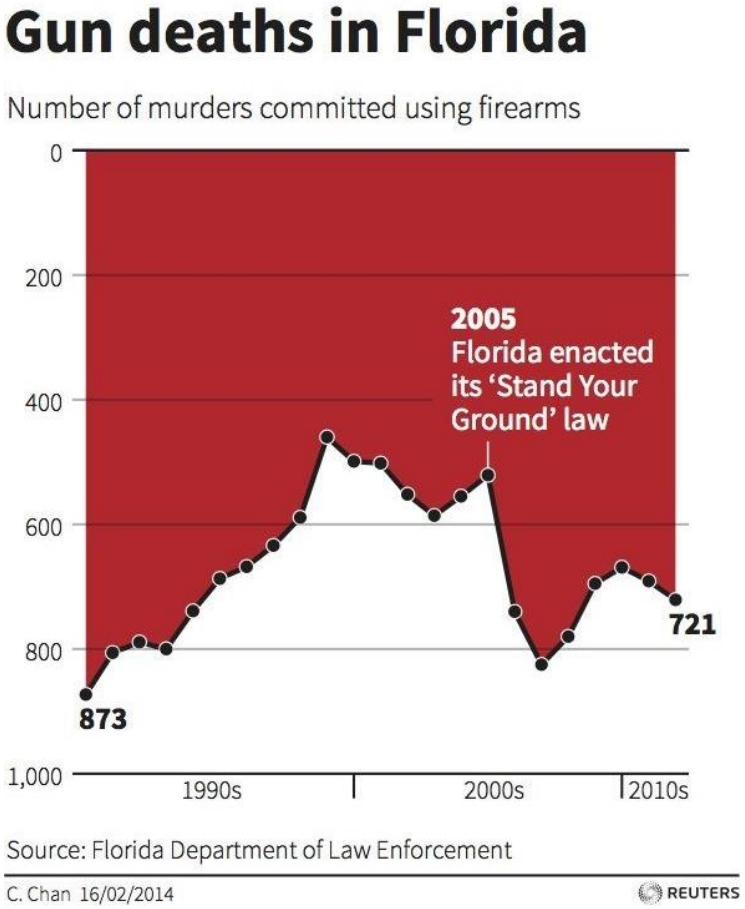
- Standard practices
  - Longer bars indicate larger numbers



source: https://viz.wtf/

Ignoring conventions

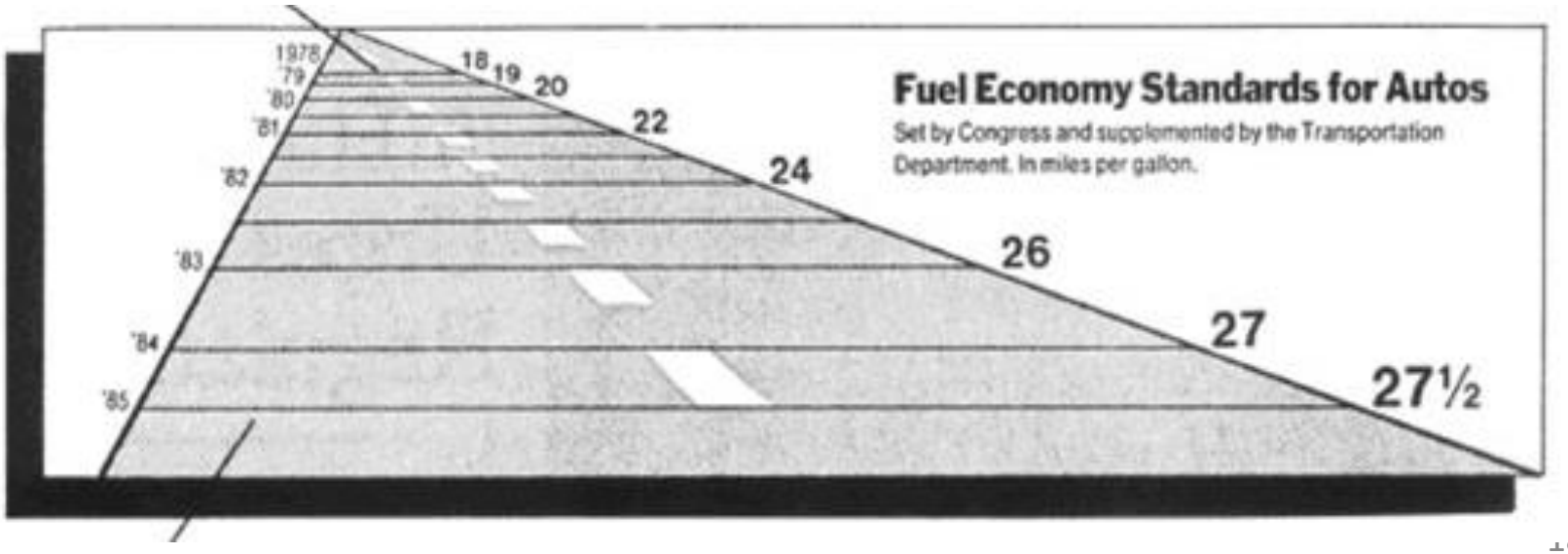
- Standard practices
  - Y-values increase as we move up the page



Inconsistent scales

Lie factor = size of effect shown in the graphic
/ size of effect in the data
= {(5.3 – 0.6) / 0.6} / {(27.5-18) / 18}
= 783% / 53% = 14.8

0.6 inches

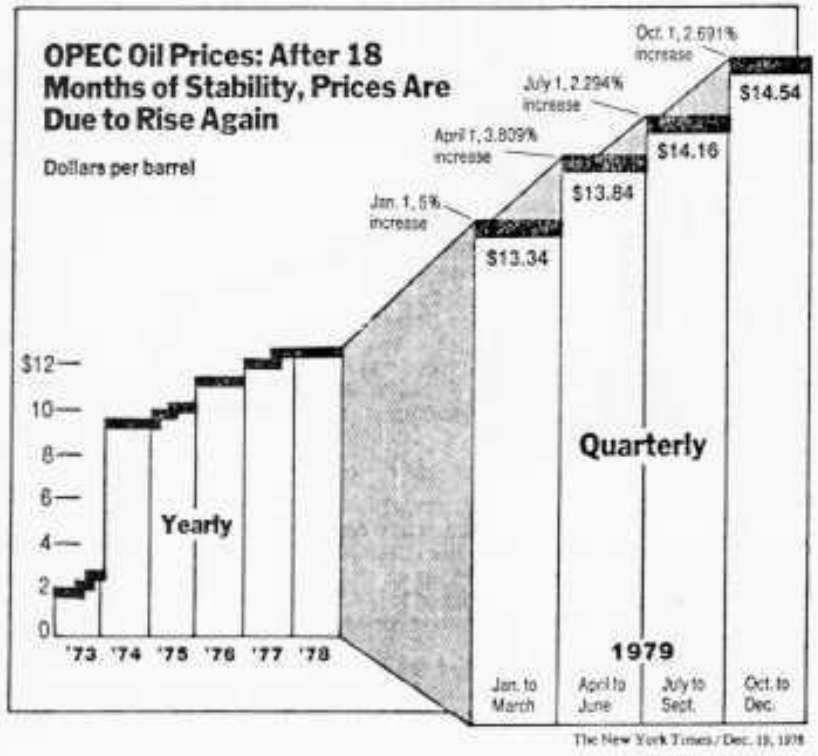


5.3 inches

Inconsistent scales

Lie factor = size of effect shown in the graphic / size of effect in the data

Design variation corrupts this display:



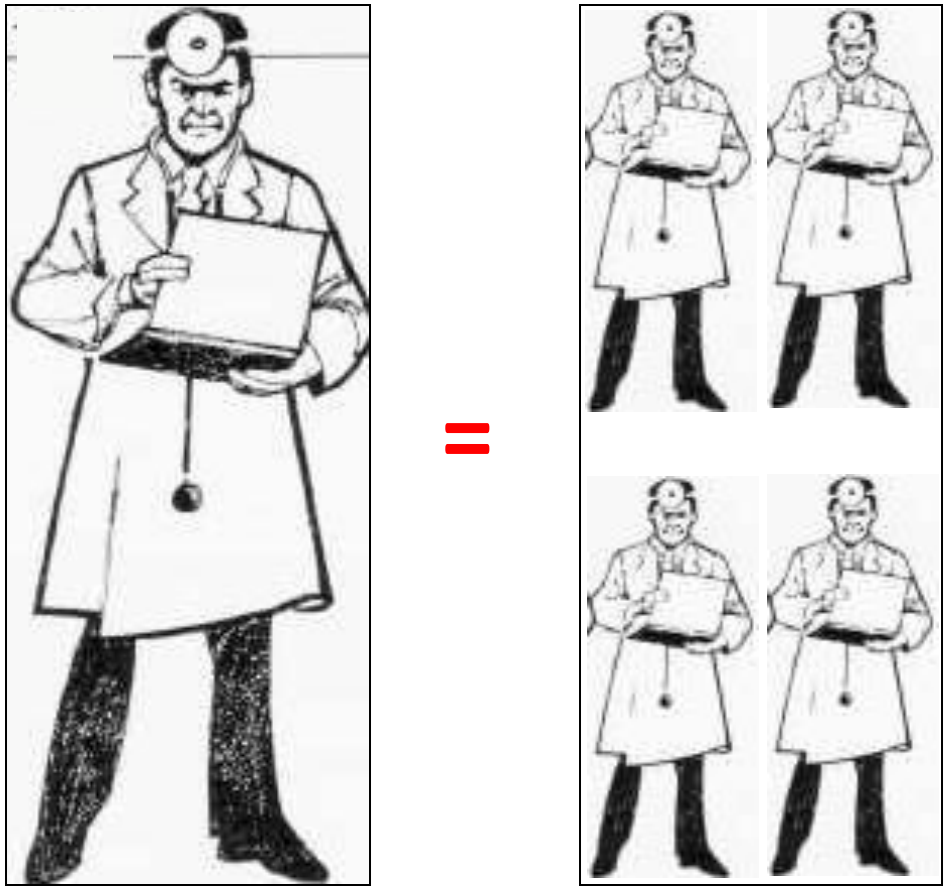
New York Times, December 19, 1978, p. D-7.

Area encoding

Area used for linear value

Los Angeles Times, August 5, 1979, p. 3.

Area encoding



Area used for linear value

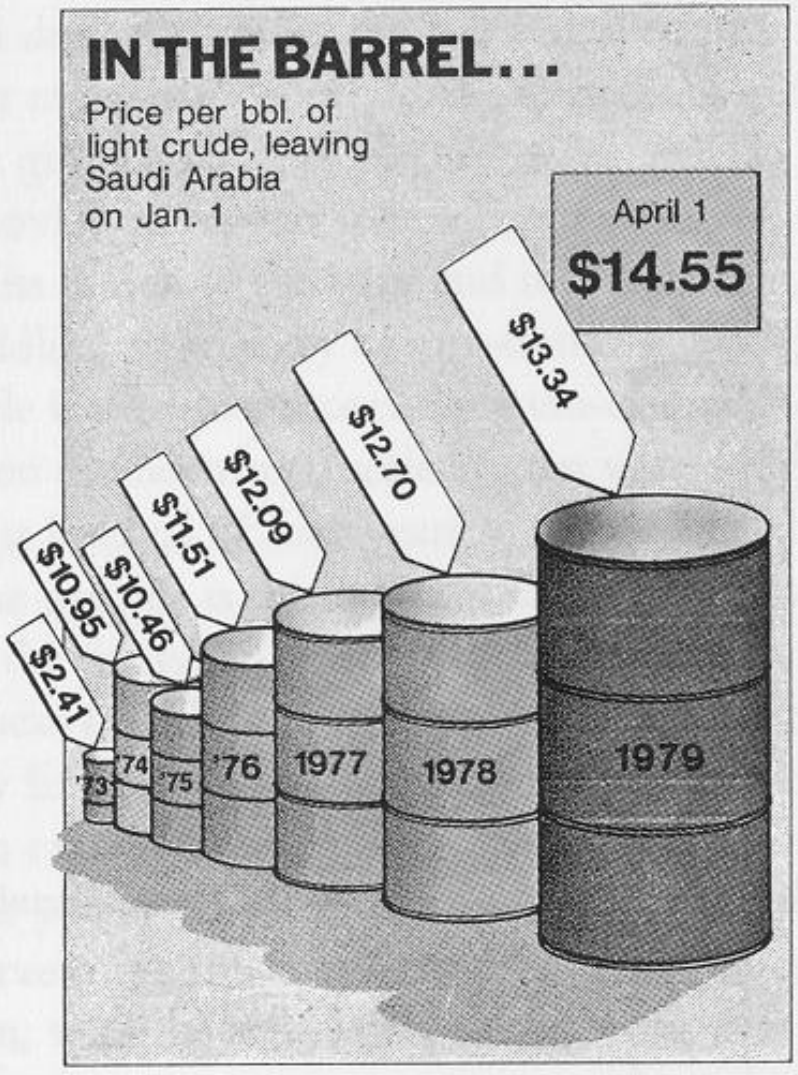
Area encoding

Encoding	Effect
Height: value	Area = value^2
Width: value	

Problem:
Using 2 dimensions to represent 1 dimension.

Encoding	Effect
Area: value	Height = value$^{0.5}$
	Width = value$^{0.5}$

Volume encoding



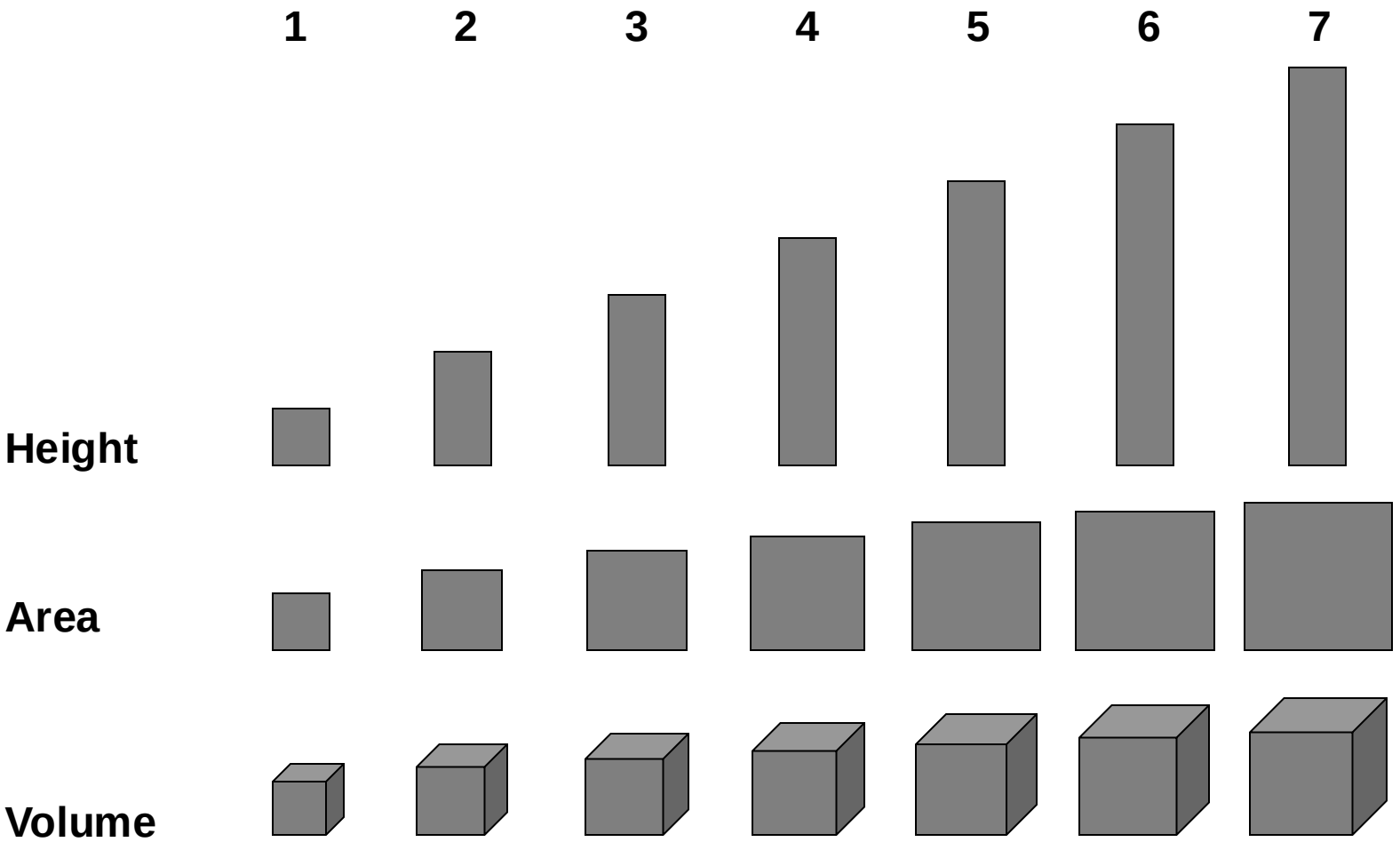
Time, April 9, 1979, p. 57.

Height?
Diameter?
Surface area?
Volume?

73 – 79 data difference = 5.5x
73 – 79 volume difference = 270x

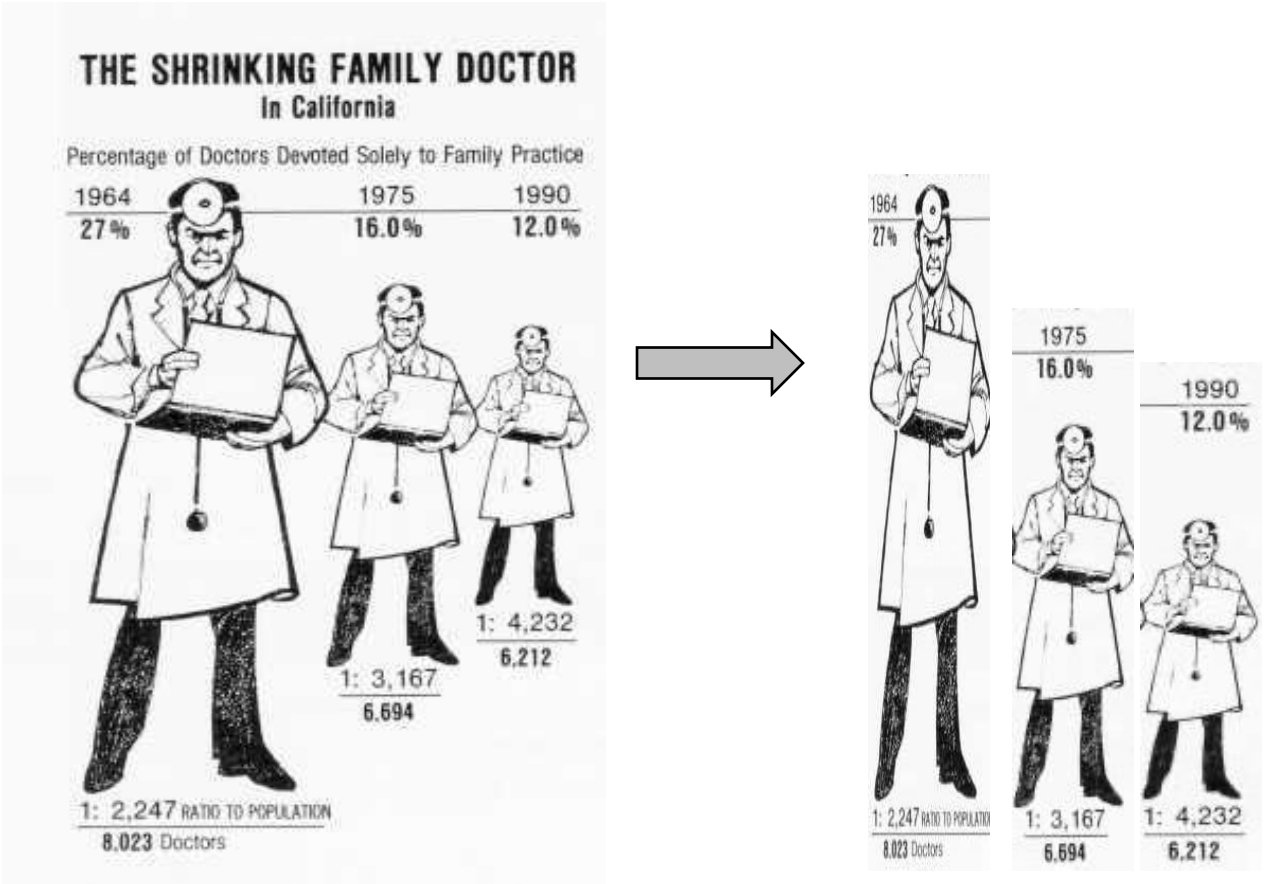
Lie factor?

Height *vs.* Area *vs.* Volume



Solution: Just use height

- The number of information carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.



How not to lie with data visualization

- Show entire scale
- Show data in context
- Not use cumulative graphs when unnecessary
- Follow conventions
- Consistent, linear scale
  - Log scale for log data
- Avoid size/volume encoding
  - Use height OR width
  - Don’t use both for same data attribute

Data Exploration & Visualization

Module 5: Design Principles

- Integrity principles
  - Not to lie with data visualization
- Tufte’s rules
- Chart-junk debate
- Nested model

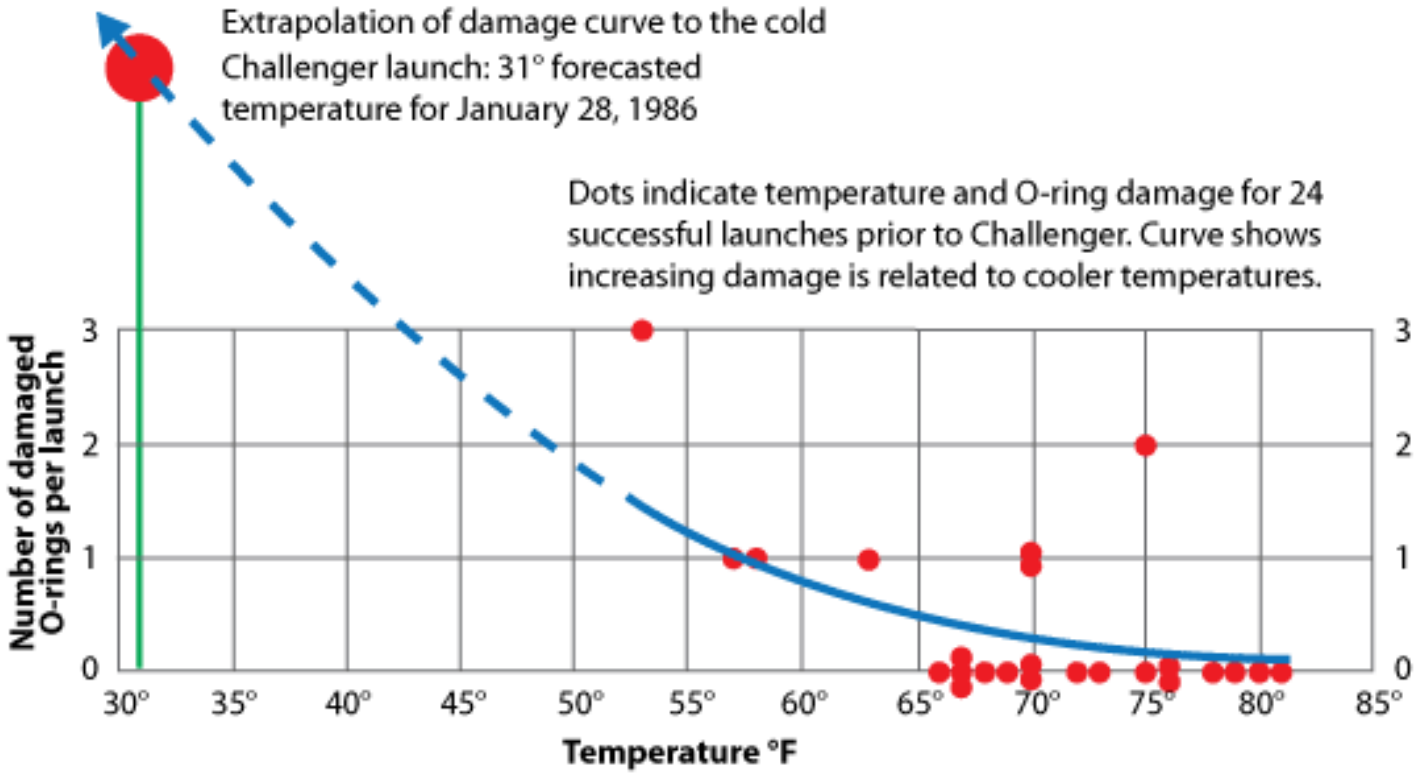
Design principles

- Visualizing with clarity and precision
  - Tufte's rules (https://www.edwardtufte.com/)
    - A minimalist on data visualization.
    - 5 books on data visualization.

  - Chart-junk debate
    - Is the 'minimalist' approach to visualizing information always correct?

Design principles

- Visualizing with clarity and precision
  - Edward Tufte's figure on the 1986 Challenger Space Shuttle launch decision



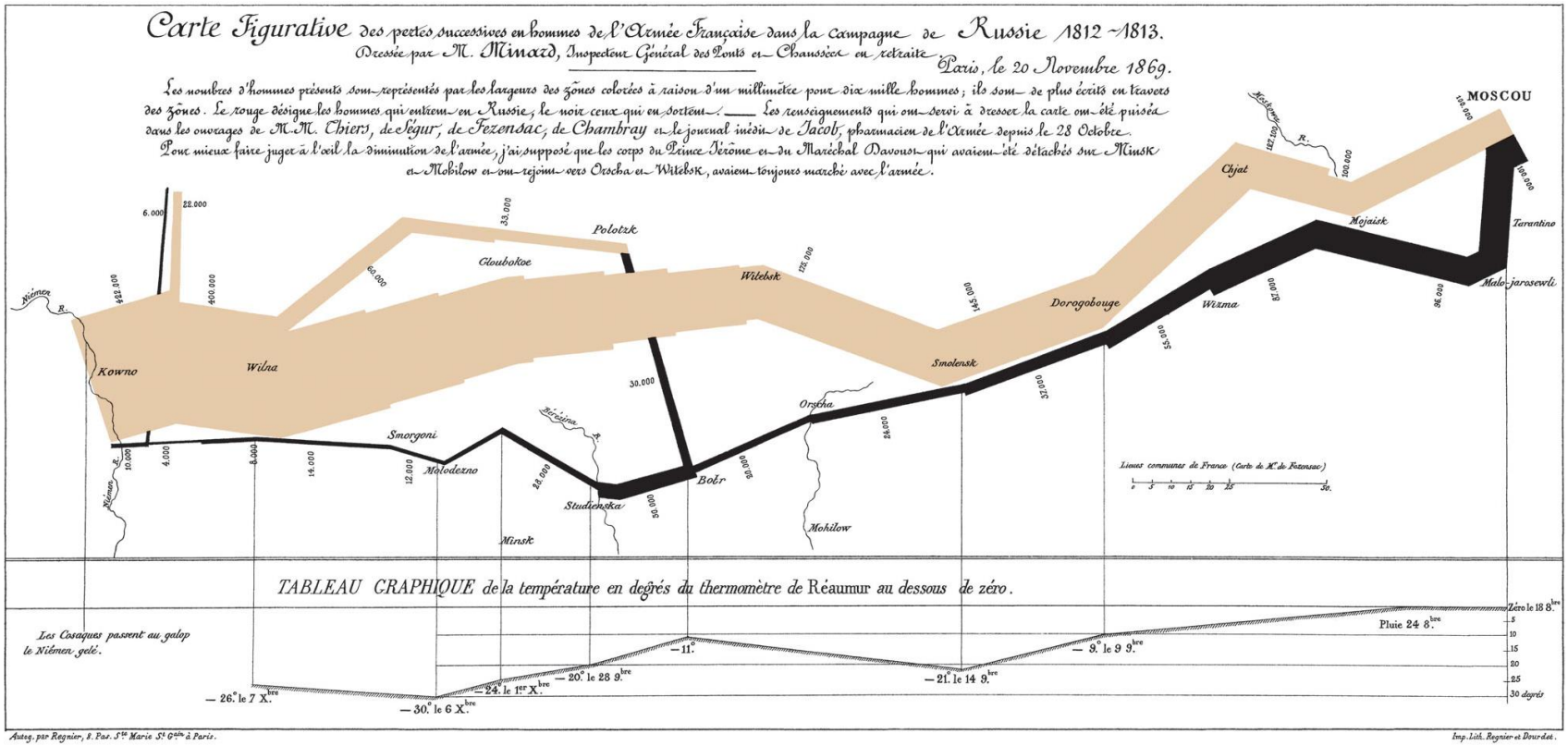
Tufte's rules

- Use graphics
- Let the data speak
- Use labels
- Avoid chartjunk
- Utilize data-ink ratio
- Utilize micro/macro
- Use small multiples

http://www.sealthreinhold.com/school/tuftes-rules/rule_one.php

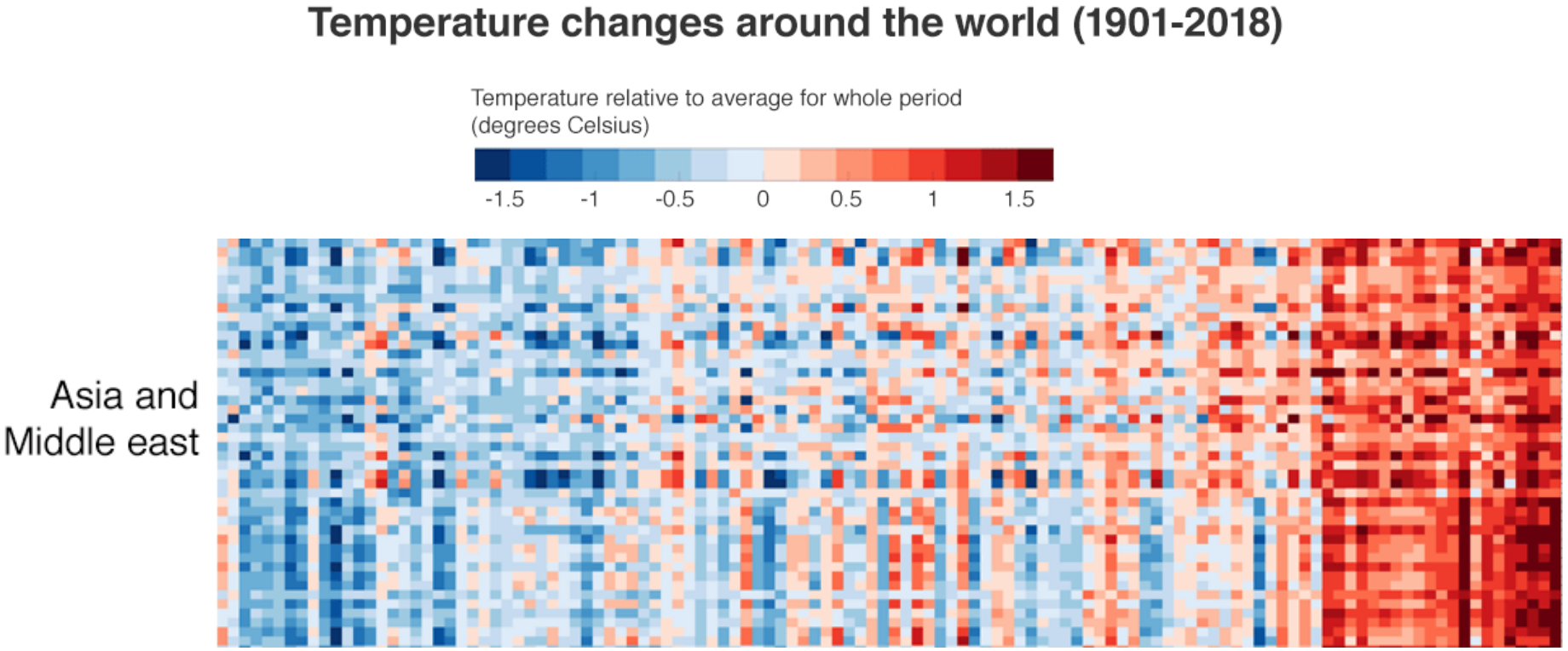
Use graphics

- A picture is worth a thousand words
- Consider using pictures/icons/glyphs in place of words



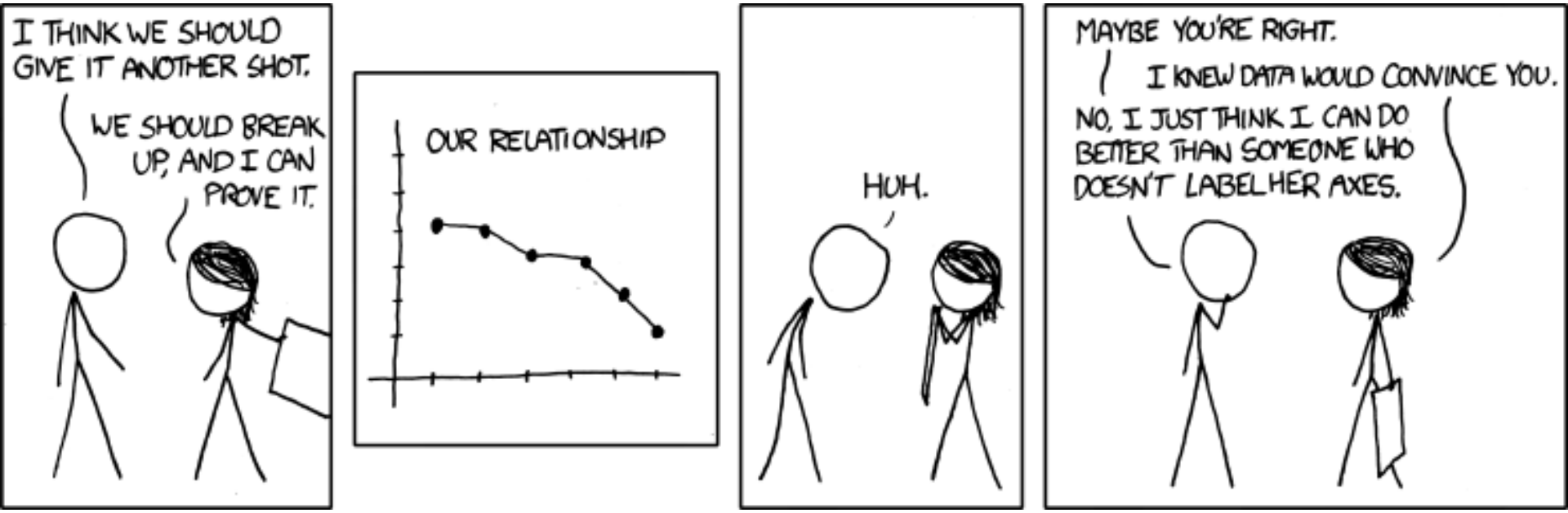
Let the data speak

- Summaries and aggregations only when necessary
- Rely on the deductive, inductive and abductive reasoning of the viewer



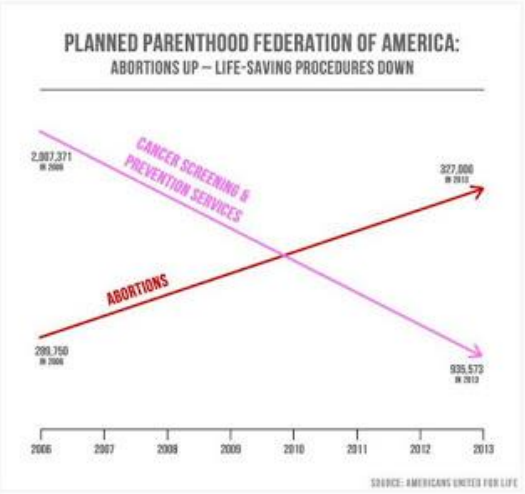
Use labels

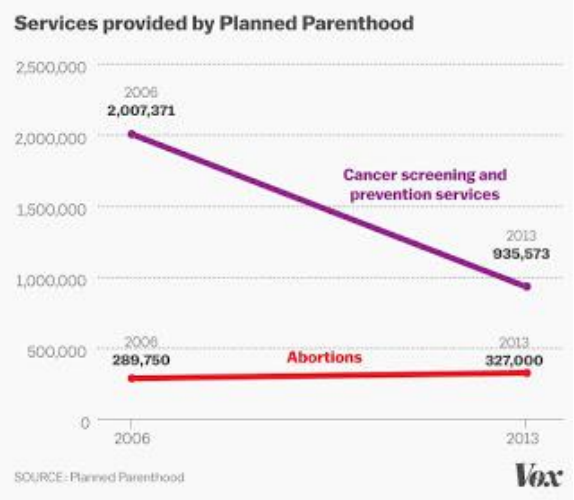
- Label your axes
- Pictures still need words
- Label should stand out from data

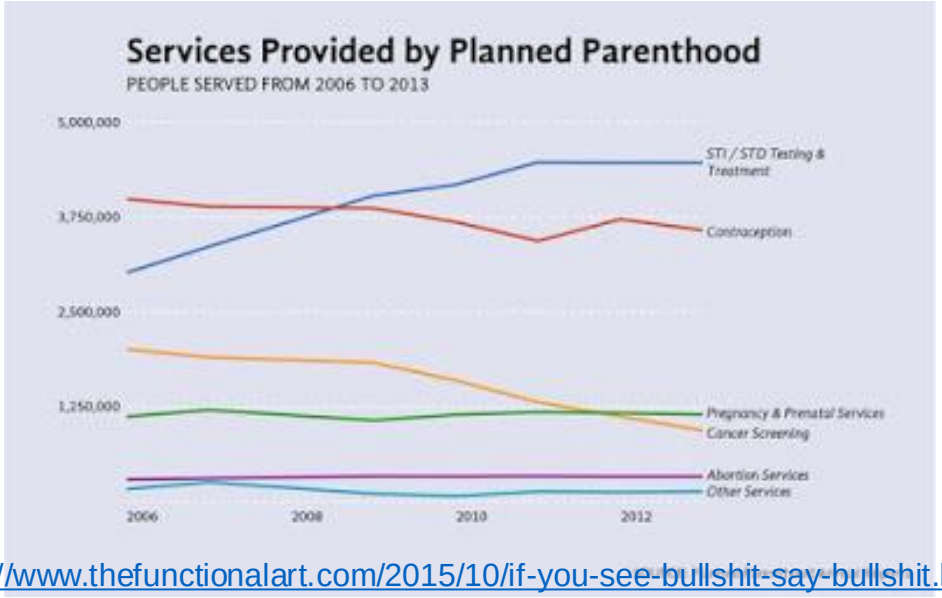


Use labels

- Labelled axis is critical
- Dual axes controversial
  - acceptable if commensurate
  - beware, very easy to mislead!







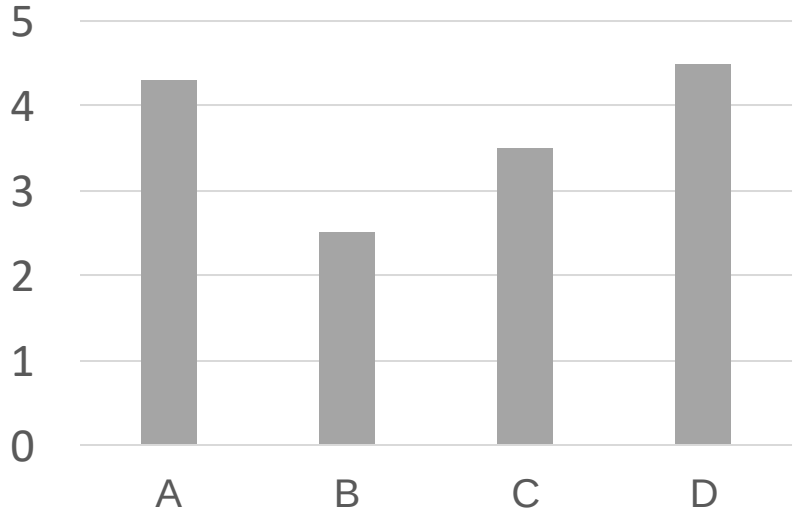
http://www.thefunctionalart.com/2015/10/if-you-see-bullshit-say-bullshit.html

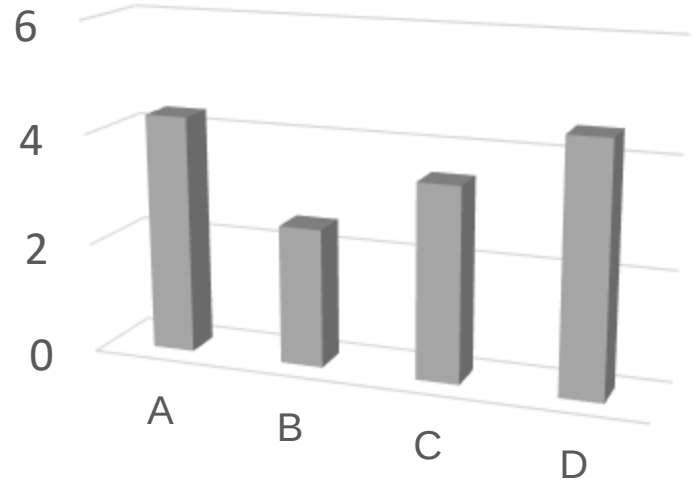
Avoid chartjunk

- The interior decoration of graphics generates a lot of ink that does not tell the viewer anything new...
- It is all non-data-ink or **redundant data-ink**, and it is often **chartjunk**
  - 3D
  - Distracting patterns
  - Overbearing colors
  - Unnecessary grids and outlines

Avoid chartjunk

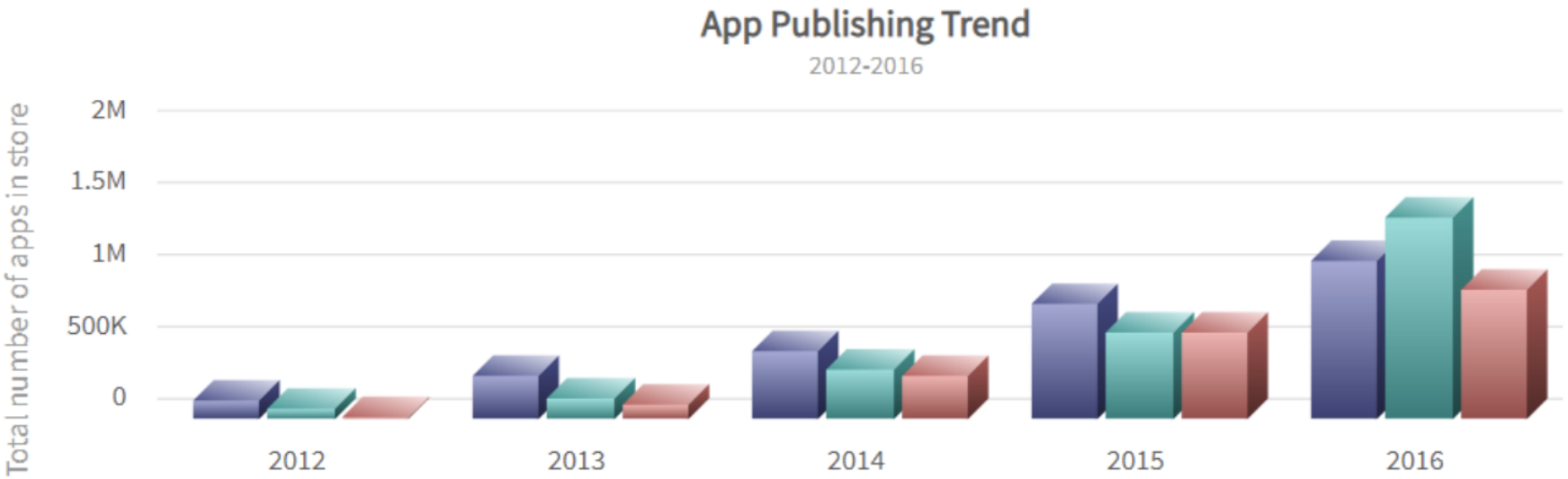
- Pretty ≠ effective
- Sometimes 3D can make a 2D boring chart more engaging
- 3D can often lead to erroneous interpretations





3D visualization

- 3D plots are quite popular, in particular in business presentations but also among academics.
- They are also almost always inappropriately used.
- **Don't use 3D for abstract data**

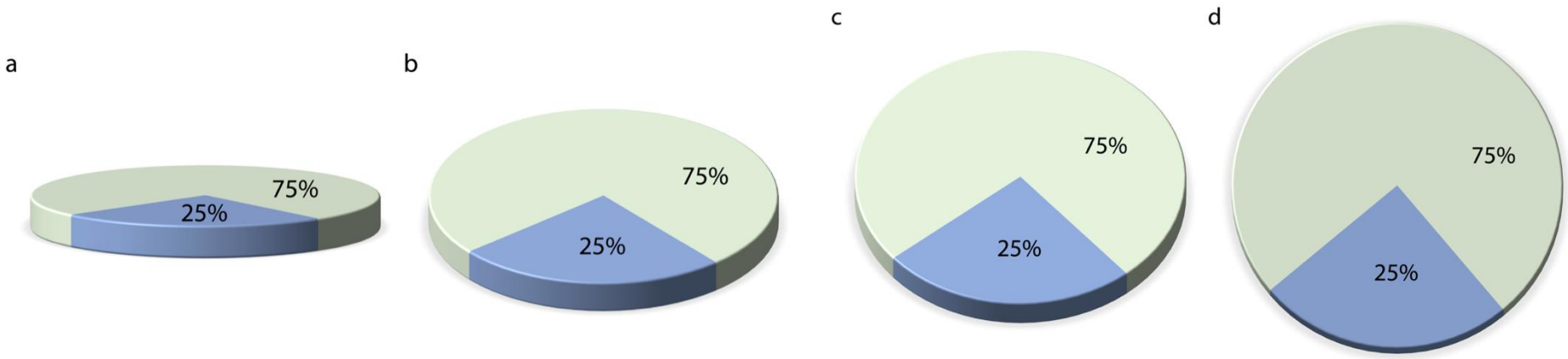


3D Data Visualization Software Tools: Best and Free NO. 1 FusionCharts

http://intellspot.com/3d-data-visualization-tools/

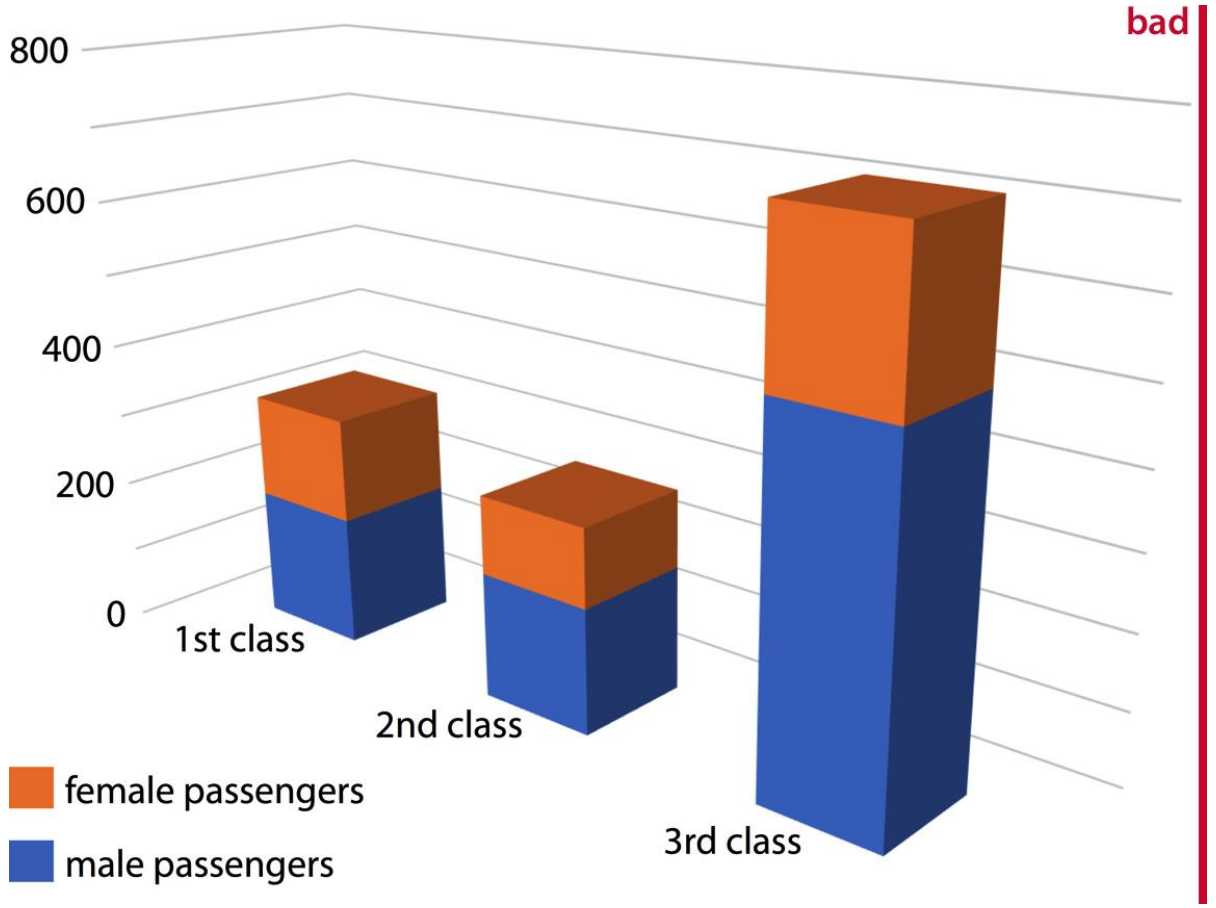
Avoid gratuitous 3D

- Examples:
  - pie charts turned into disks rotated in space
  - bar plots turned into columns
  - line plots turned into bands.
- None of these cases does the third dimension convey any actual data. 3D is used simply to decorate and adorn the plot.
  - distorts the data.



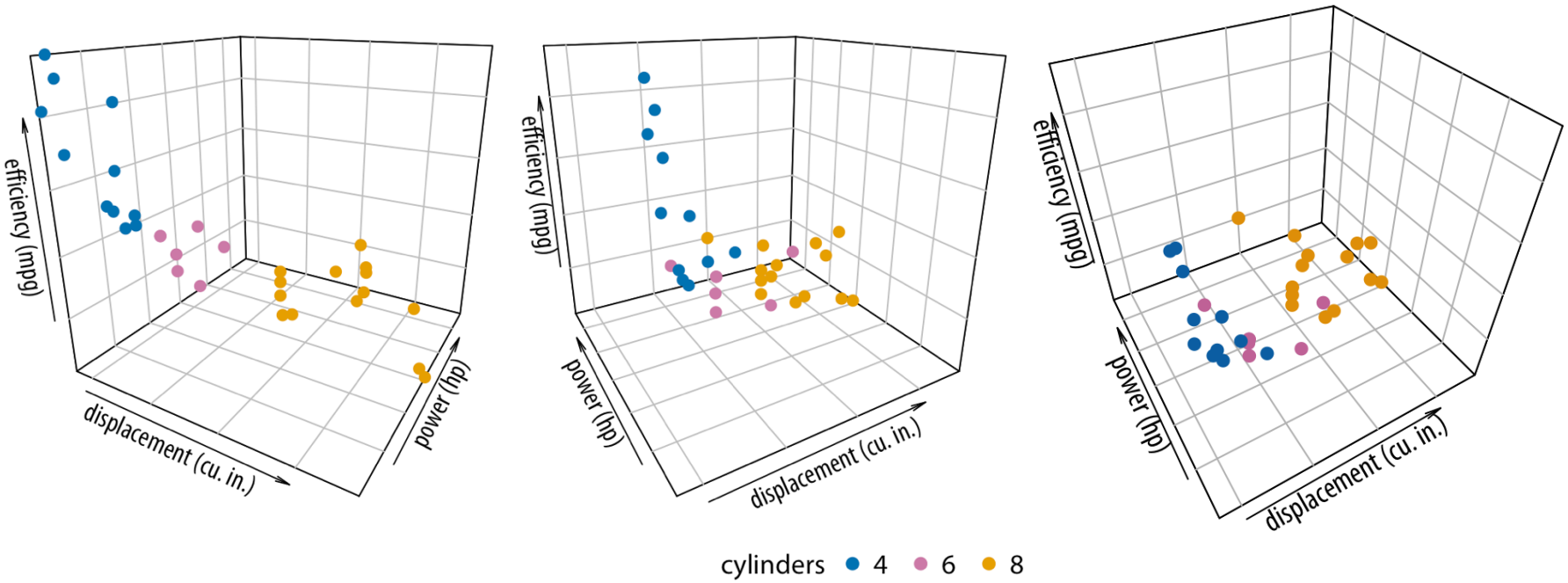
Avoid gratuitous 3D

- The breakdown of Titanic passengers by class and gender using 3D bars.



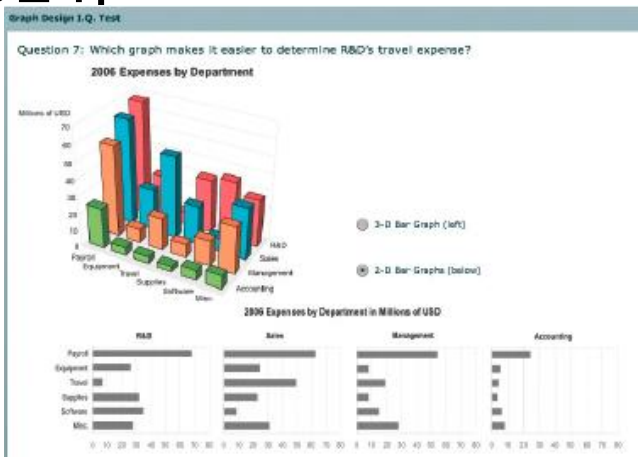
Avoid gratuitous 3D

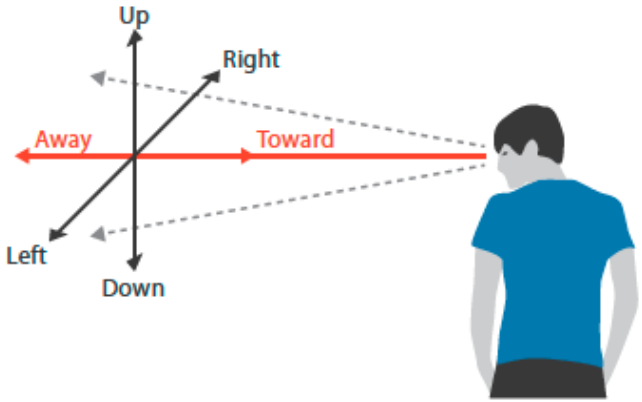
- Visualizations using three position scales (x, y, and z) to represent data - the third dimension serves an actual purpose.
  - plot displacement along the *x* axis
  - power along the *y* axis
  - fuel efficiency along the *z* axis

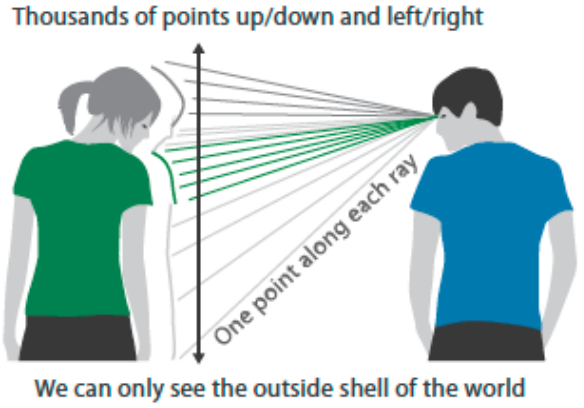


Avoid gratuitous 3D

- No unjustified 3D [Munzner 2014]
  - disparity of depth
  - occlusion
  - perspective distortion
  - interaction complexity
  - text legibility

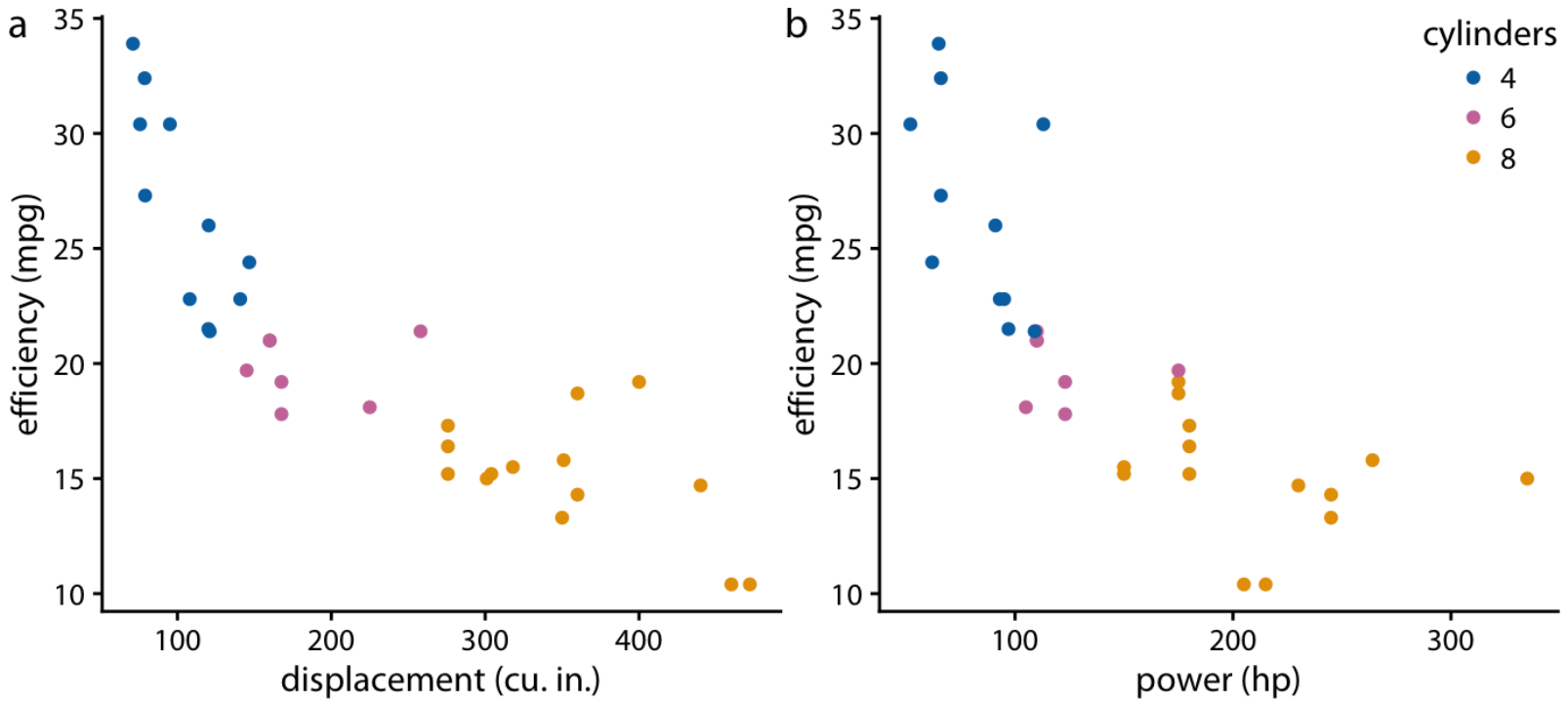






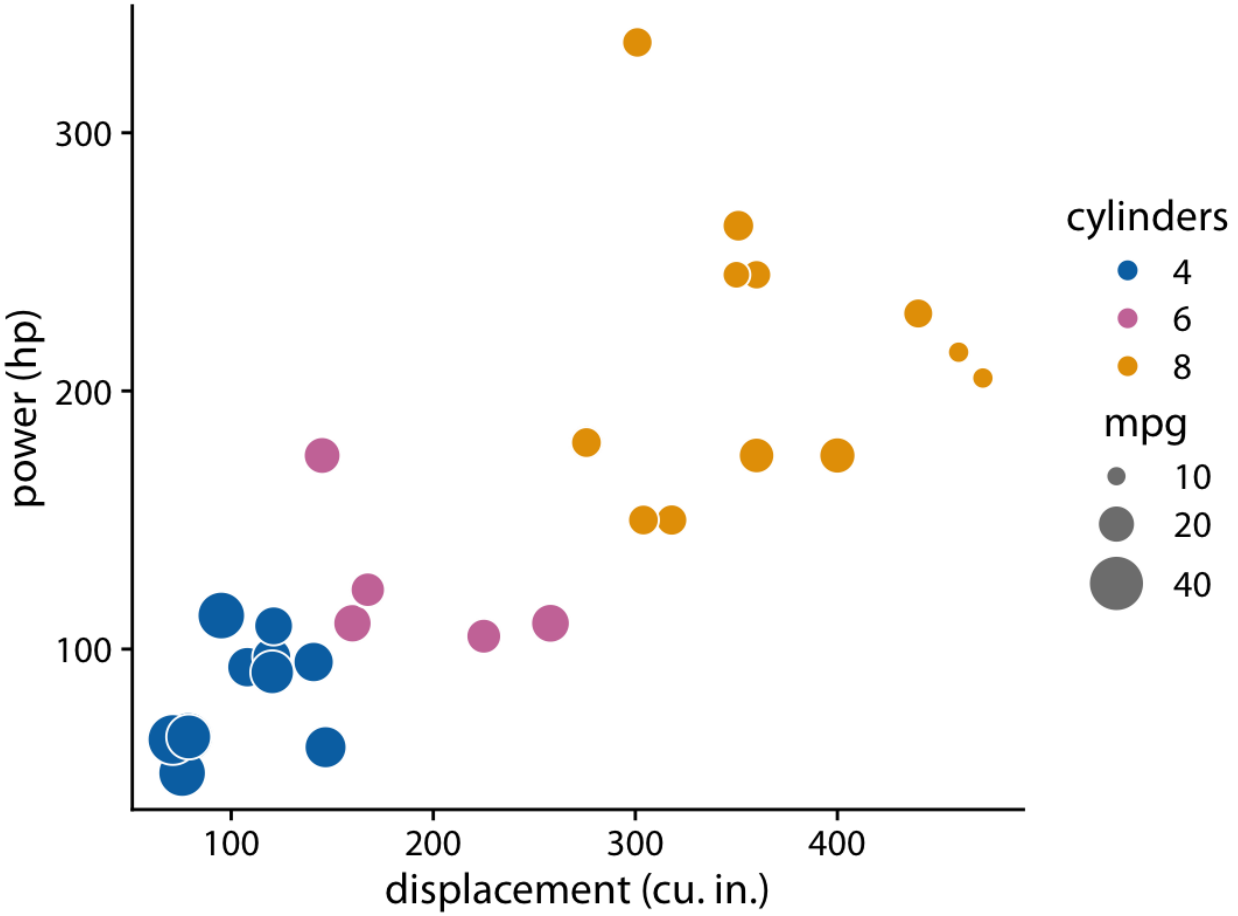
Alternative design

- Better design: turn into regular 2D figures.
  - Plot fuel efficiency twice, one against displacement (left), and one against power (right).



Alternative design

- Better design: turn into regular 2D figures.
  - X-axis: displacement , Y-axis: power, Dot size: fuel efficiency



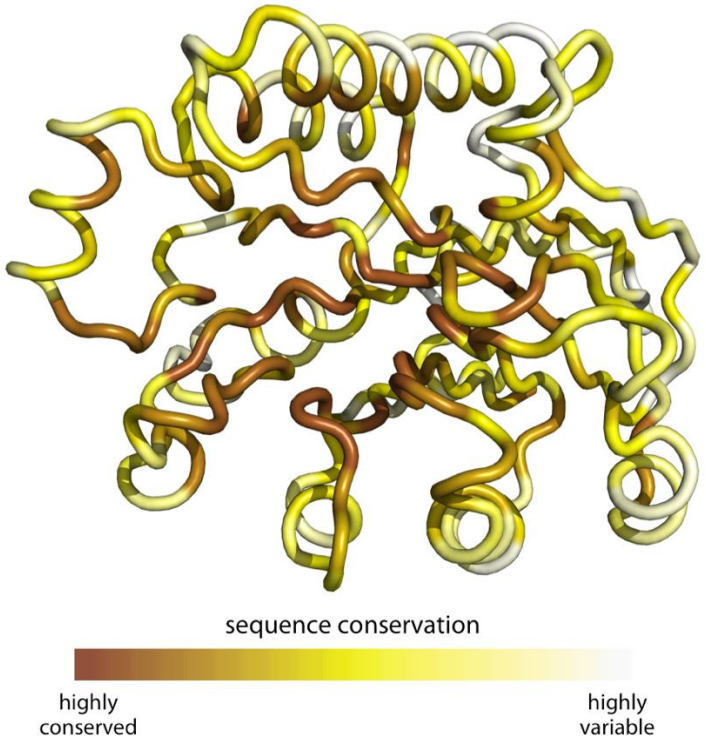
Appropriate use of 3D visualizations

- 3D is appropriate when the user's task involves *shape understanding* of inherently three-dimensional structures
- The issues are of lesser concern
  - The visualization is interactive.
  - It is shown in a VR or augmented reality environment where it can be inspected from multiple angles.
  - 3D visualization slowly rotate can discern where in 3D space different graphical elements reside
    - The human brain is very good at reconstructing a 3D scene from a series of images taken from different angles.

Appropriate use of 3D visualizations

- It makes sense to use 3D visualizations when we want to show actual 3D objects and/or data mapped onto them.
  - Spatial data
  - Preferably in interactive version or rotating

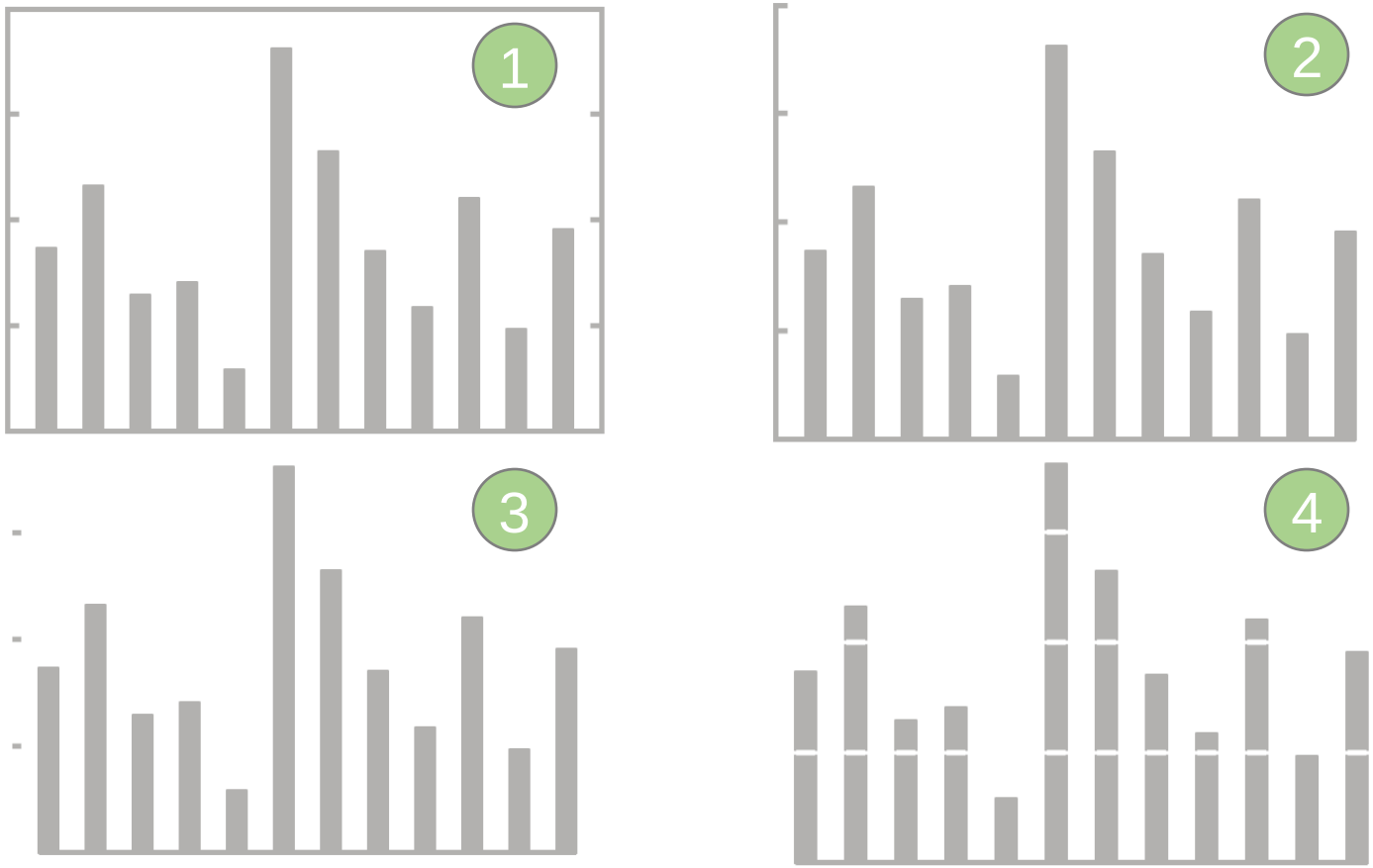
topographic relief of a mountainous island



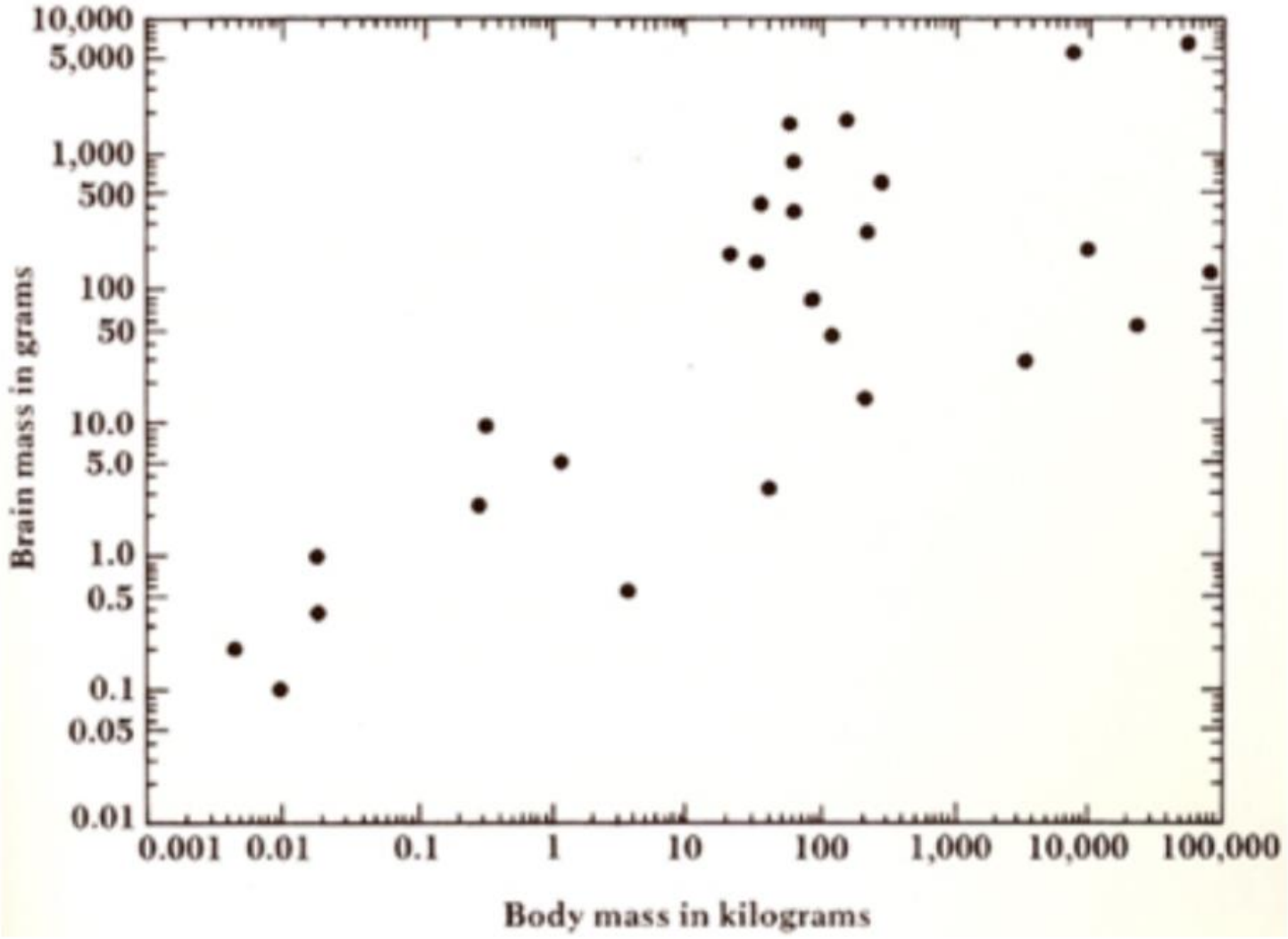
Patterns of evolutionary variation in a protein

Utilize data-ink ratio

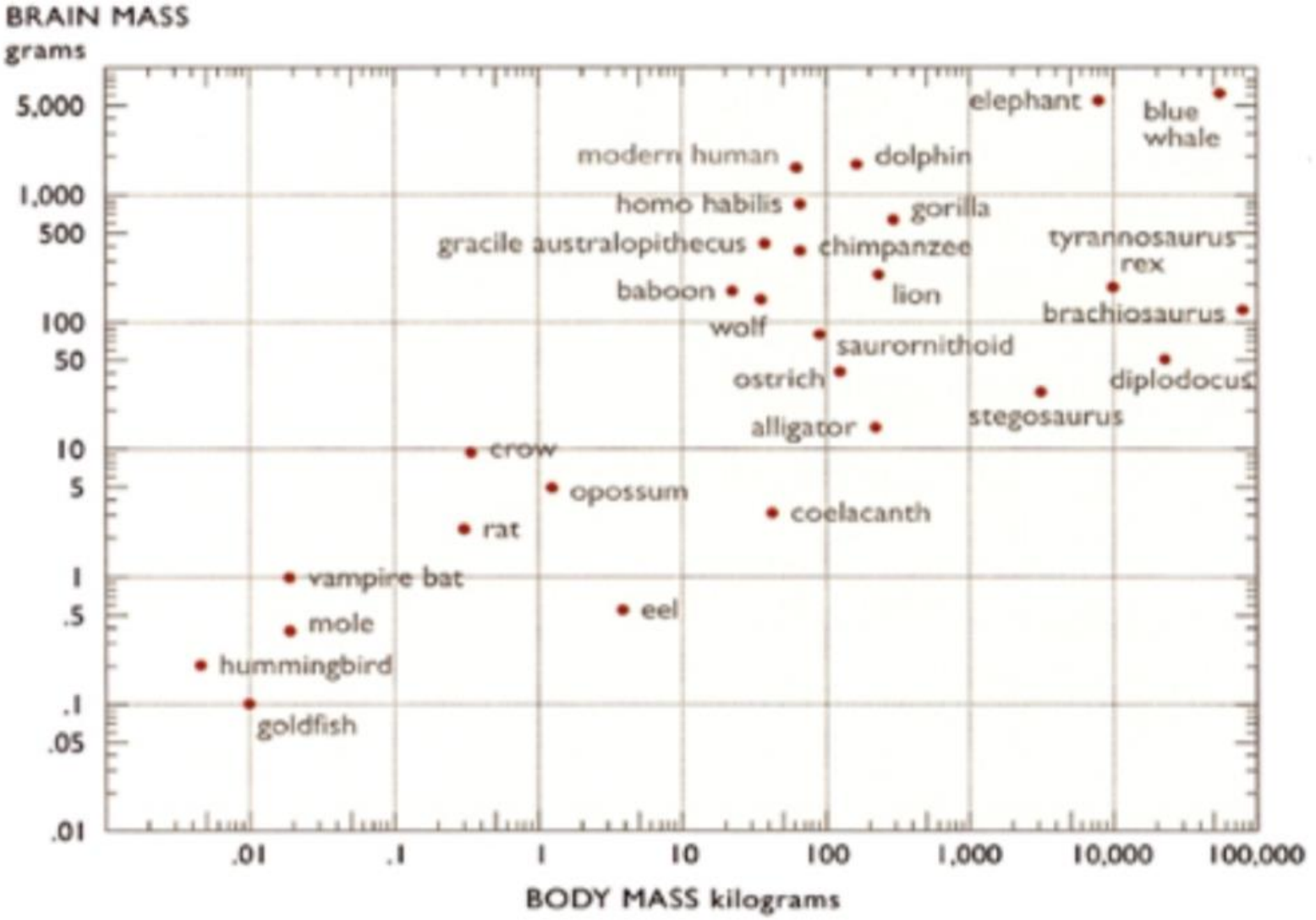
- Data-ink ratio = Data / ink
  - Don’t waste ink on elements of the visualization not associated with the data



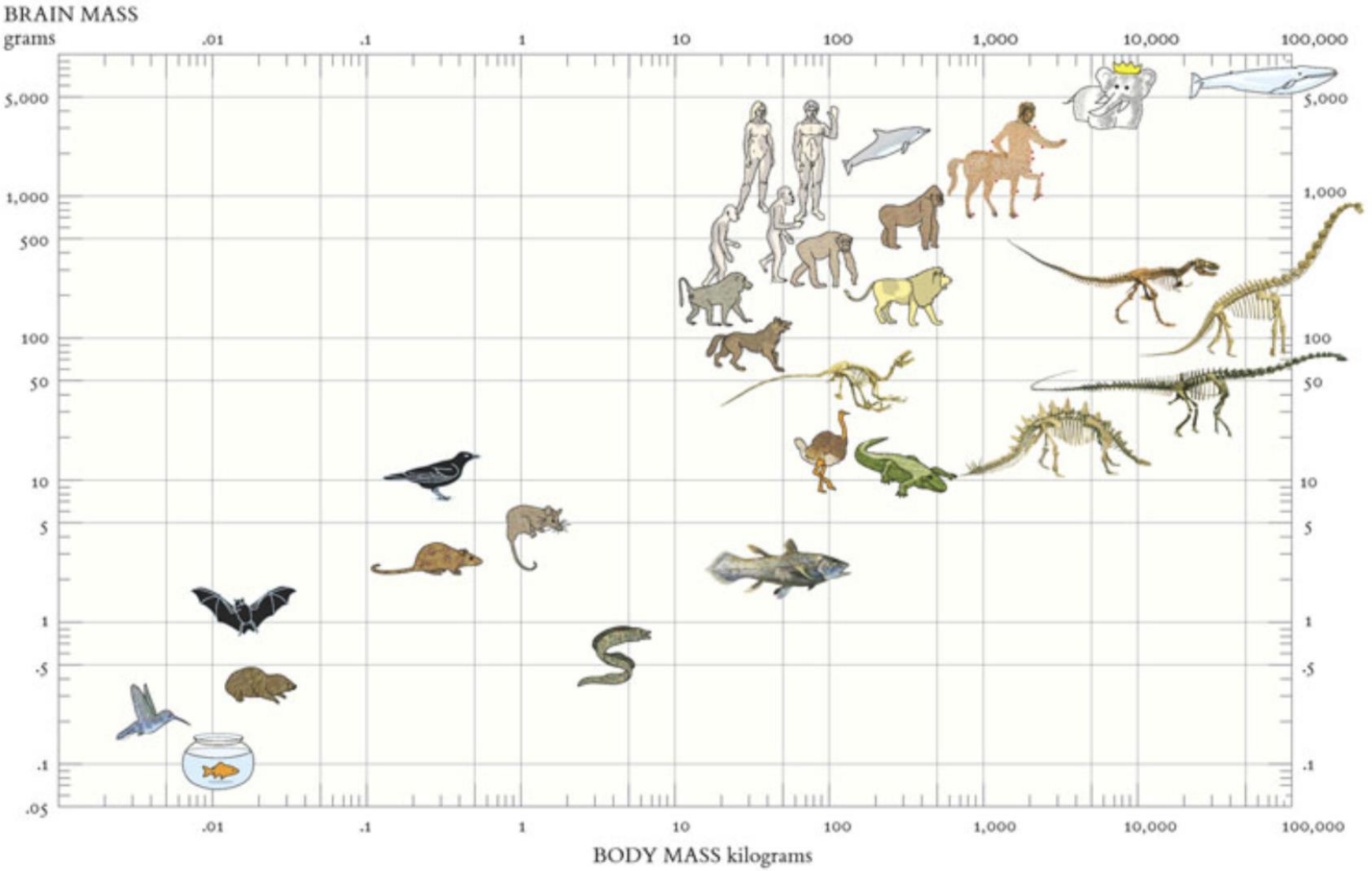
Utilize data-ink ratio



Utilize data-ink ratio

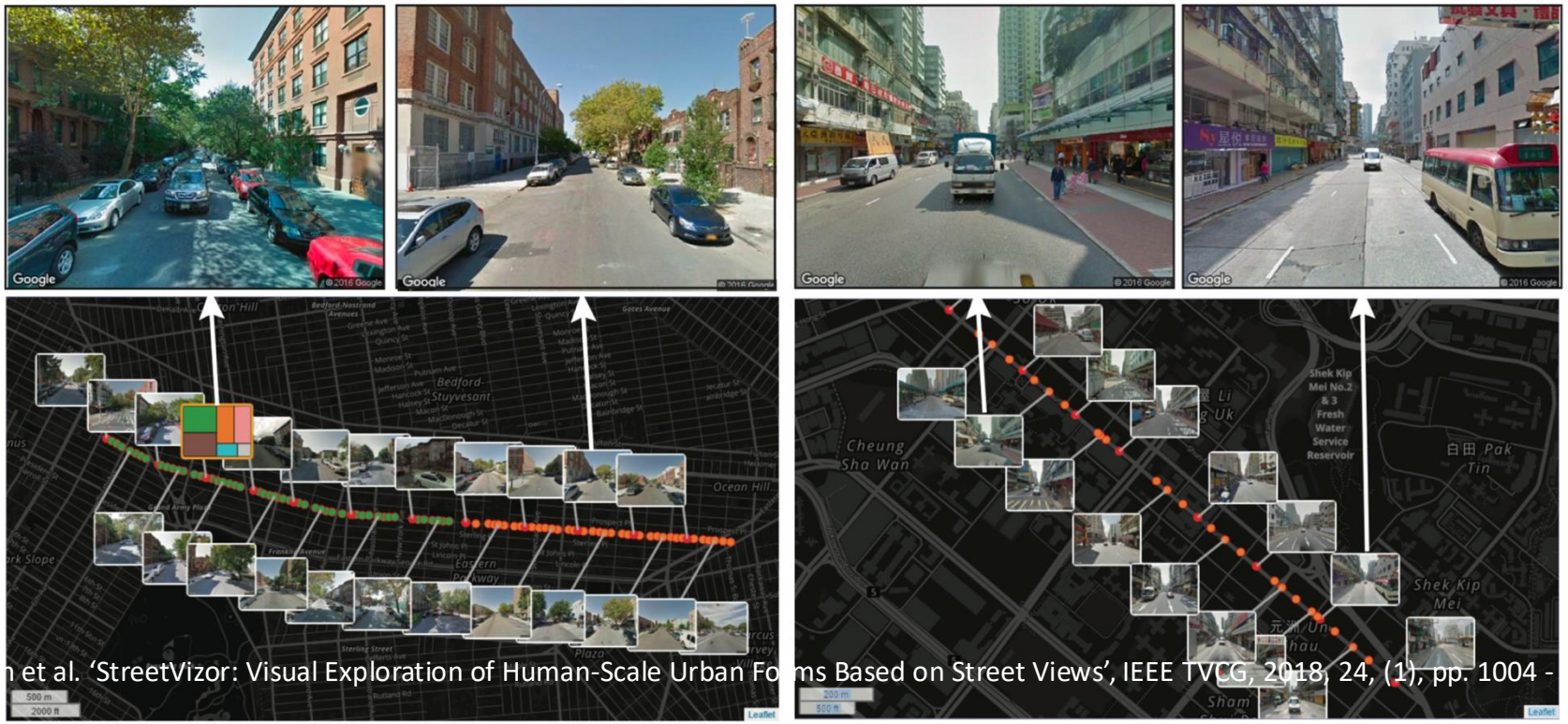


Utilize data-ink ratio



Utilize micro/macro

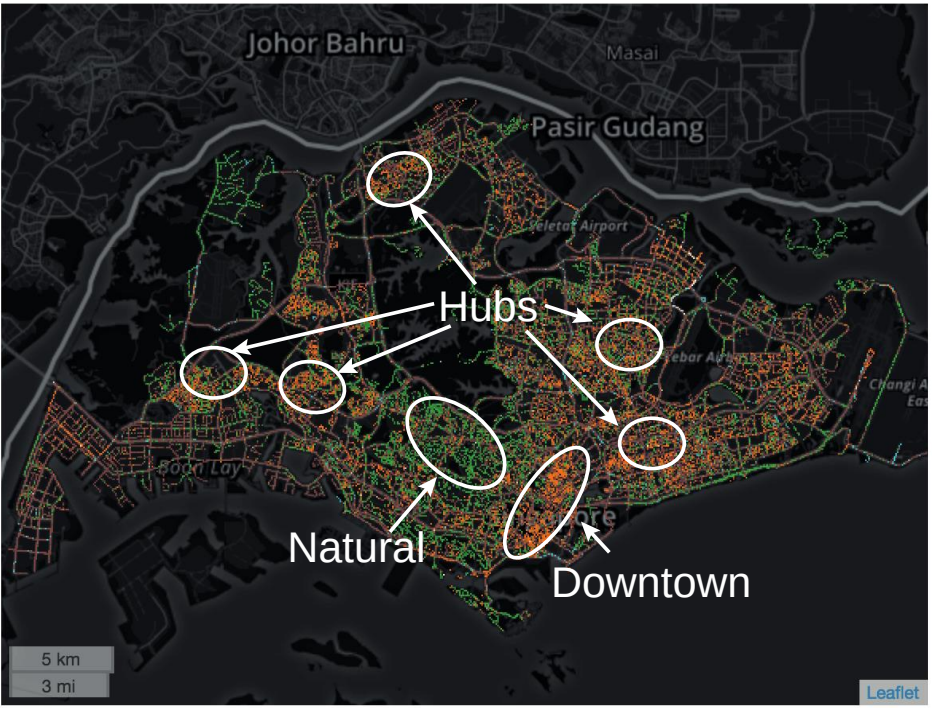
- "overview first, zoom and filter, then details-on-demand" – Shneiderman's Mantra
  - Create zoomable interfaces when possible
  - To clarify, add details: the clarity of the macro is determined by the quality and quantity of the micro

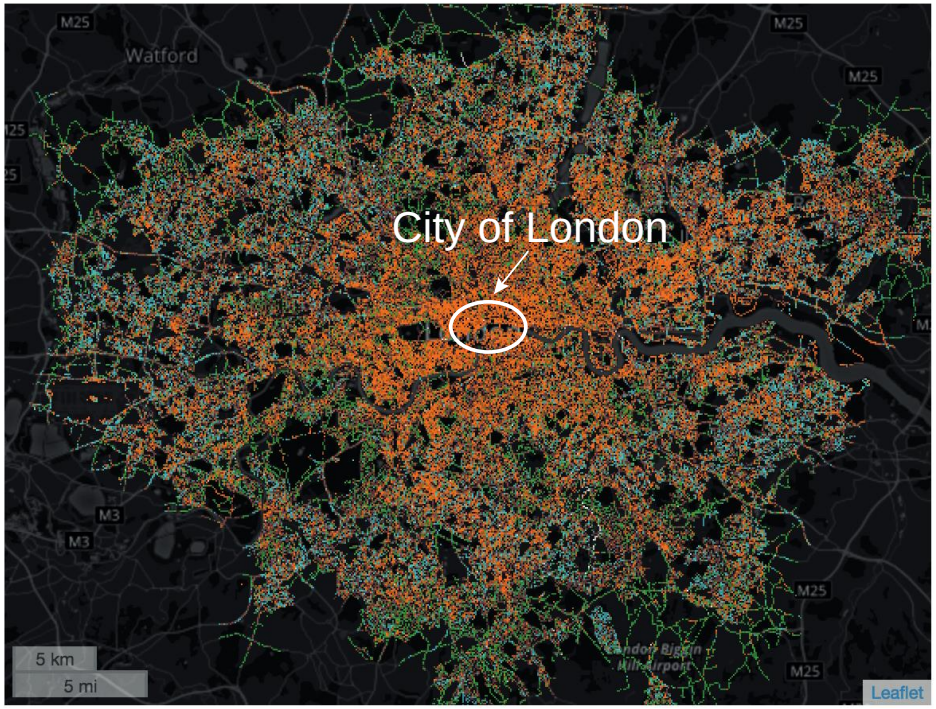


et al. 'StreetVizor: Visual Exploration of Human-Scale Urban Forms Based on Street Views', IEEE TVCG, 2018, 24, (1), pp. 1004 -

Use small multiples

- Maintain a consistent design
  - Consistent appearance puts emphasis on data, not the visual design
  - Changes in design can distract from irregularities in the data





Shen et al. 'StreetVizor: Visual Exploration of Human-Scale Urban Forms Based on Street Views', IEEE TVCG, 2018, 24, (1), pp. 1004 - 1013

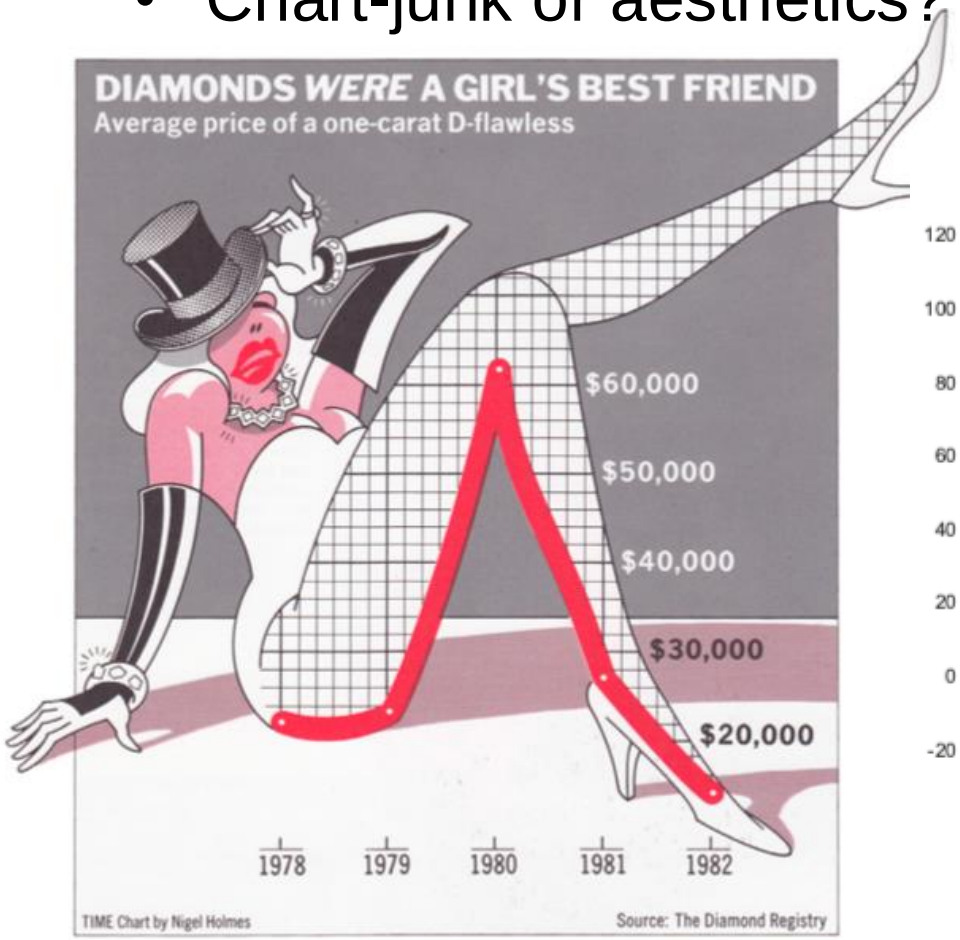
Data Exploration & Visualization

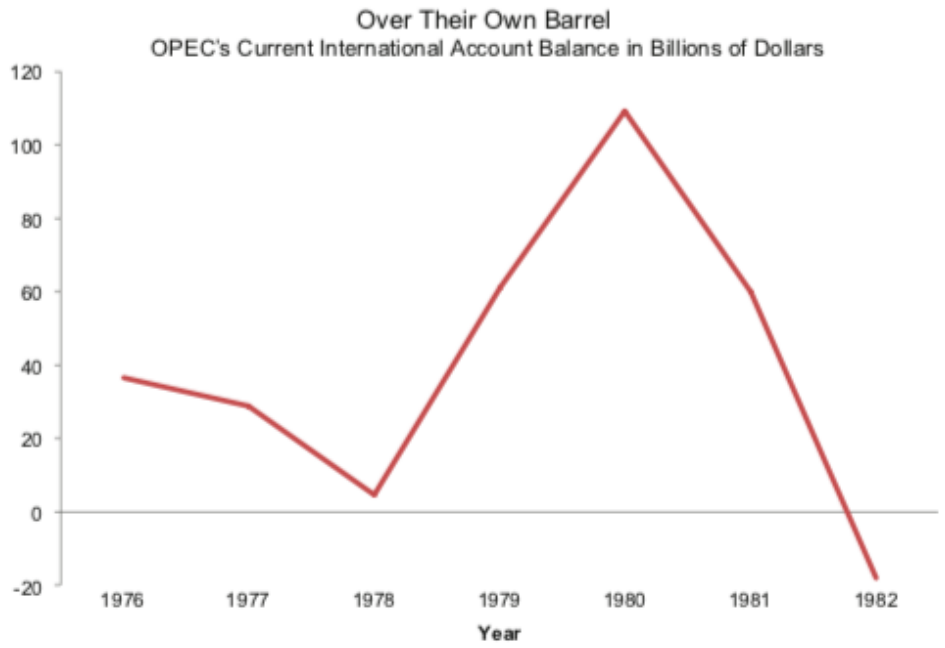
Module 5: Design Principles

- Integrity principles
  - Not to lie with data visualization
- Tufte’s rules
- Chart-junk debate
- Nested model

Chart-junk debate

- Chart-junk or aesthetics?

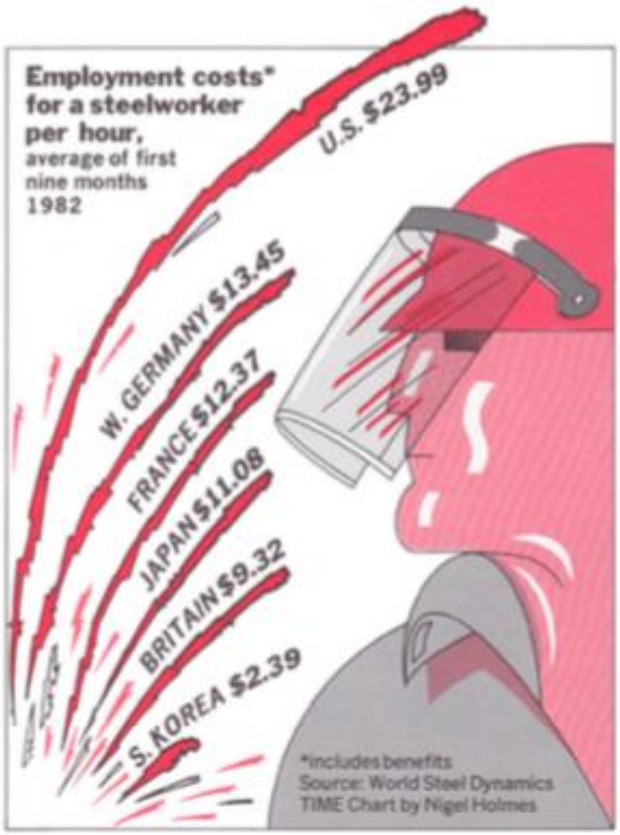


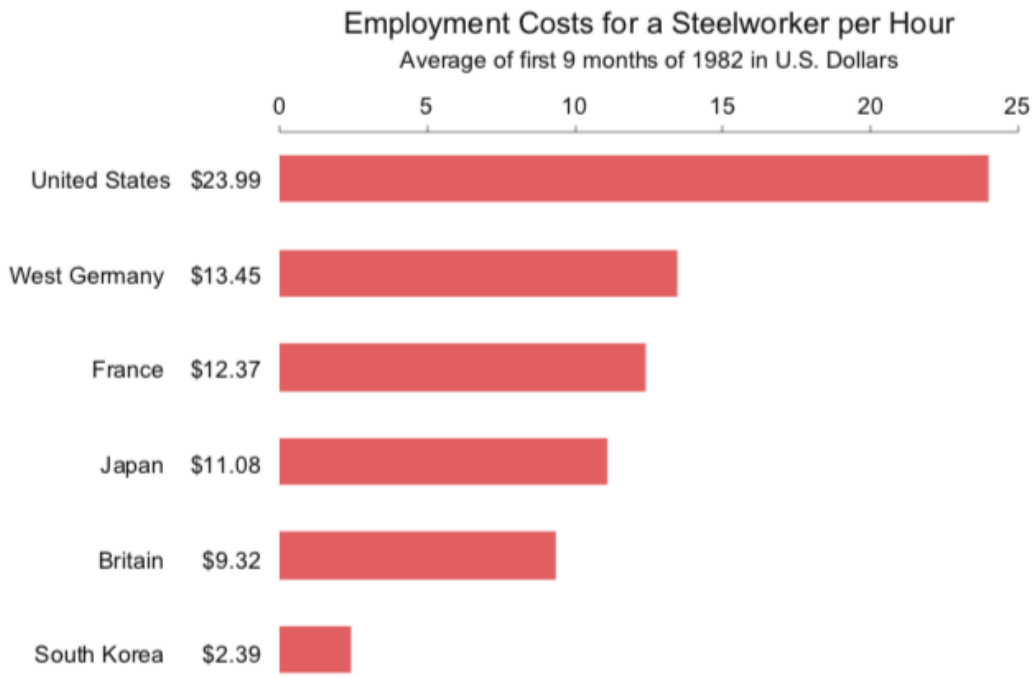


S. Bateman et al. "Useful junk?: the effects of visual embellishment on comprehension and memorability of charts," in ACM CHI, Atlanta, Georgia, USA, 2010, pp. 2573-2582, 1753716: ACM.

Chart-junk debate

- Chart-junk or aesthetics?





S. Bateman et al. “Useful junk?: the effects of visual embellishment on comprehension and memorability of charts,” in ACM CHI, Atlanta, Georgia, USA, 2010, pp. 2573-2582, 1753716: ACM.

Chart-junk debate

- People's accuracy in describing the embellished charts was no worse than for plain charts.
- Recall after a two-to-three-week gap was significantly better.
- Although we are cautious about recommending that all charts be produced in this style, our results question some of the premises of the minimalist approach to chart design.

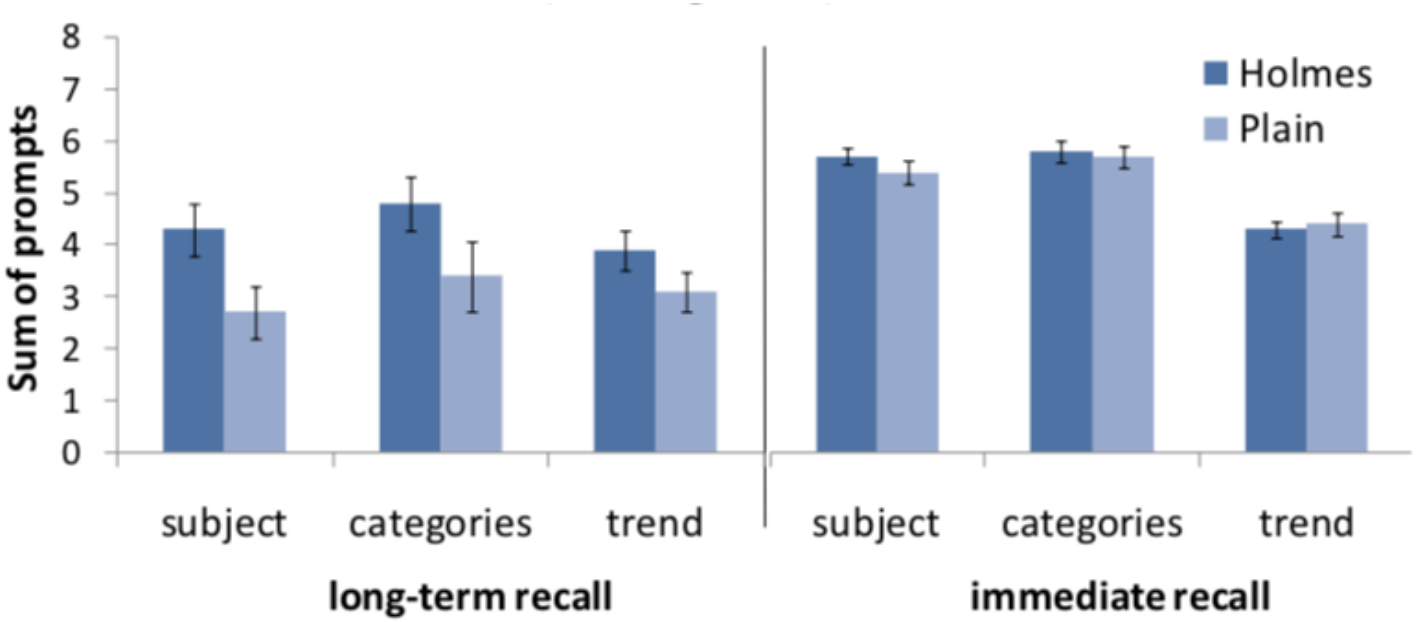
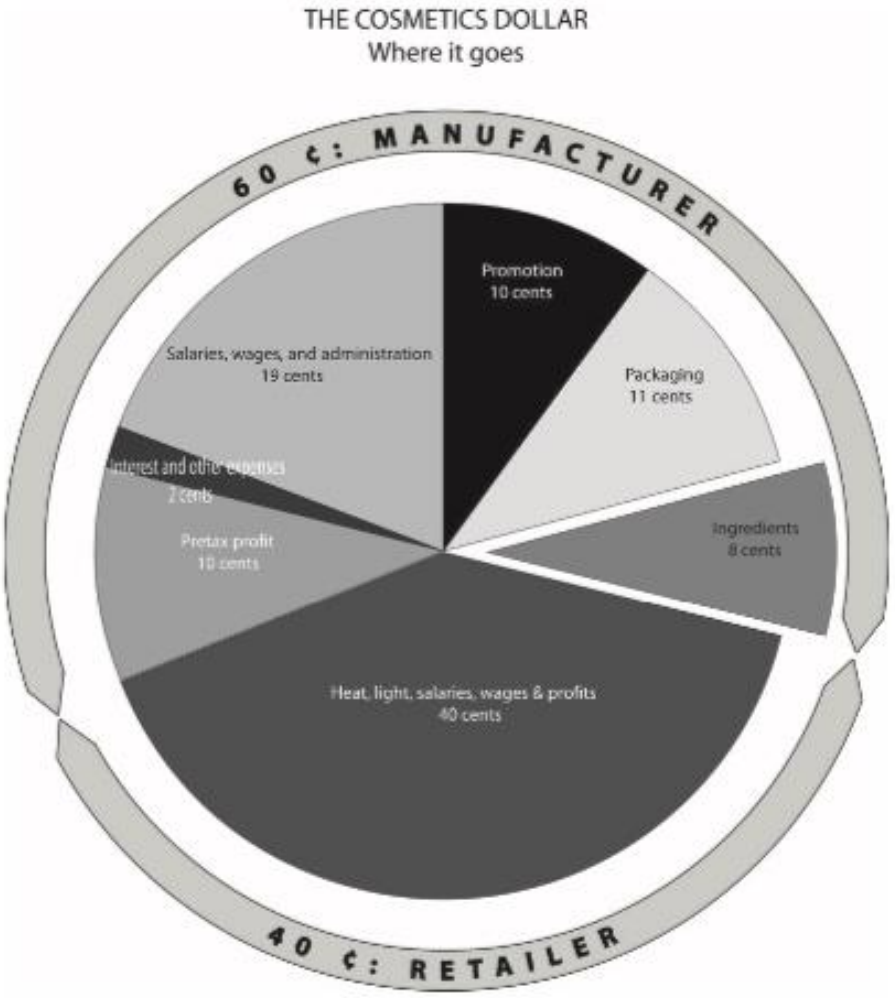


Chart-junk debate

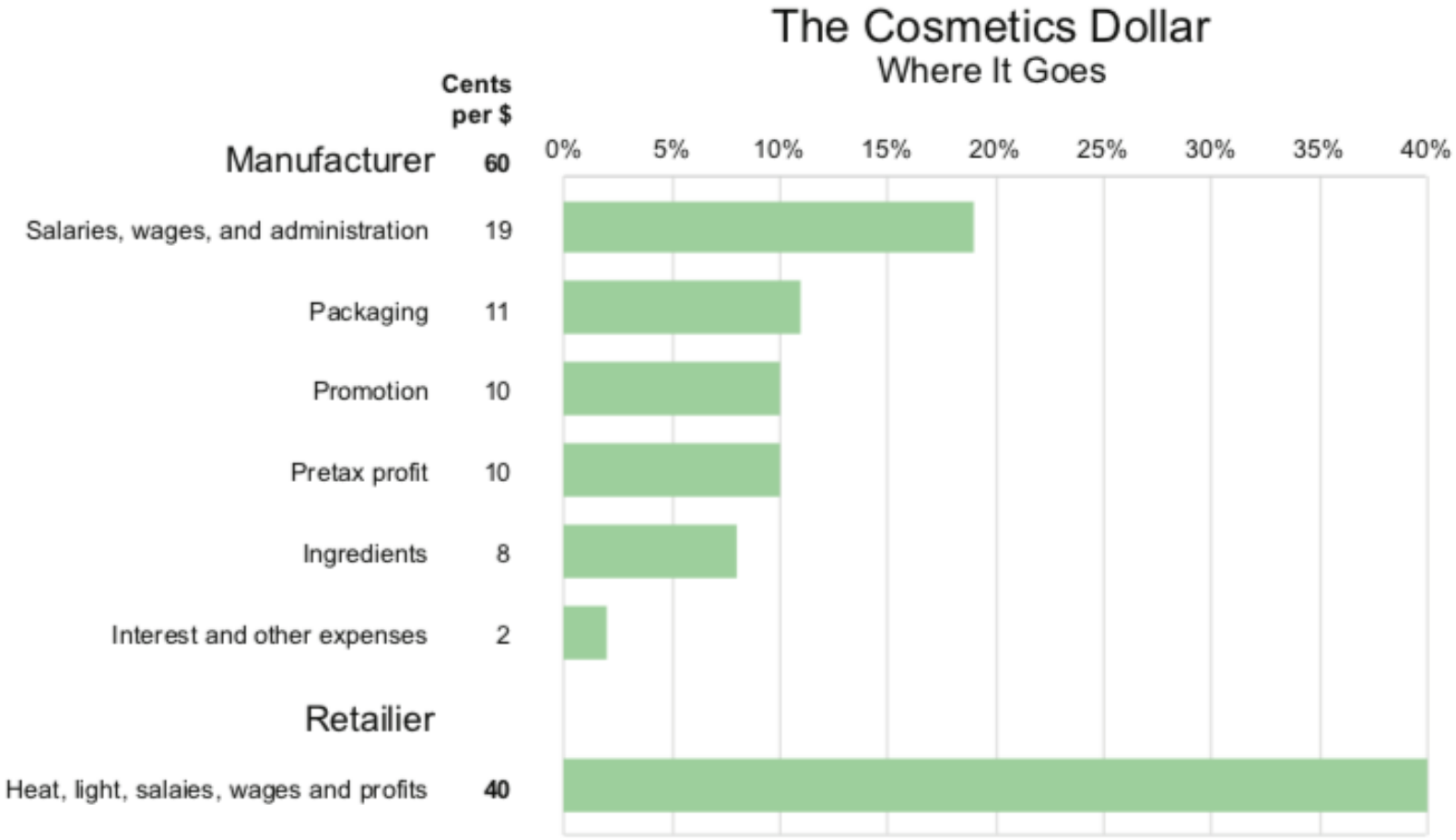
- Unproperly designed chart.



S. Few. (2011) The Chartjunk Debate - A Close Examination of Recent Findings. Visual Business Intelligence Newsletter.

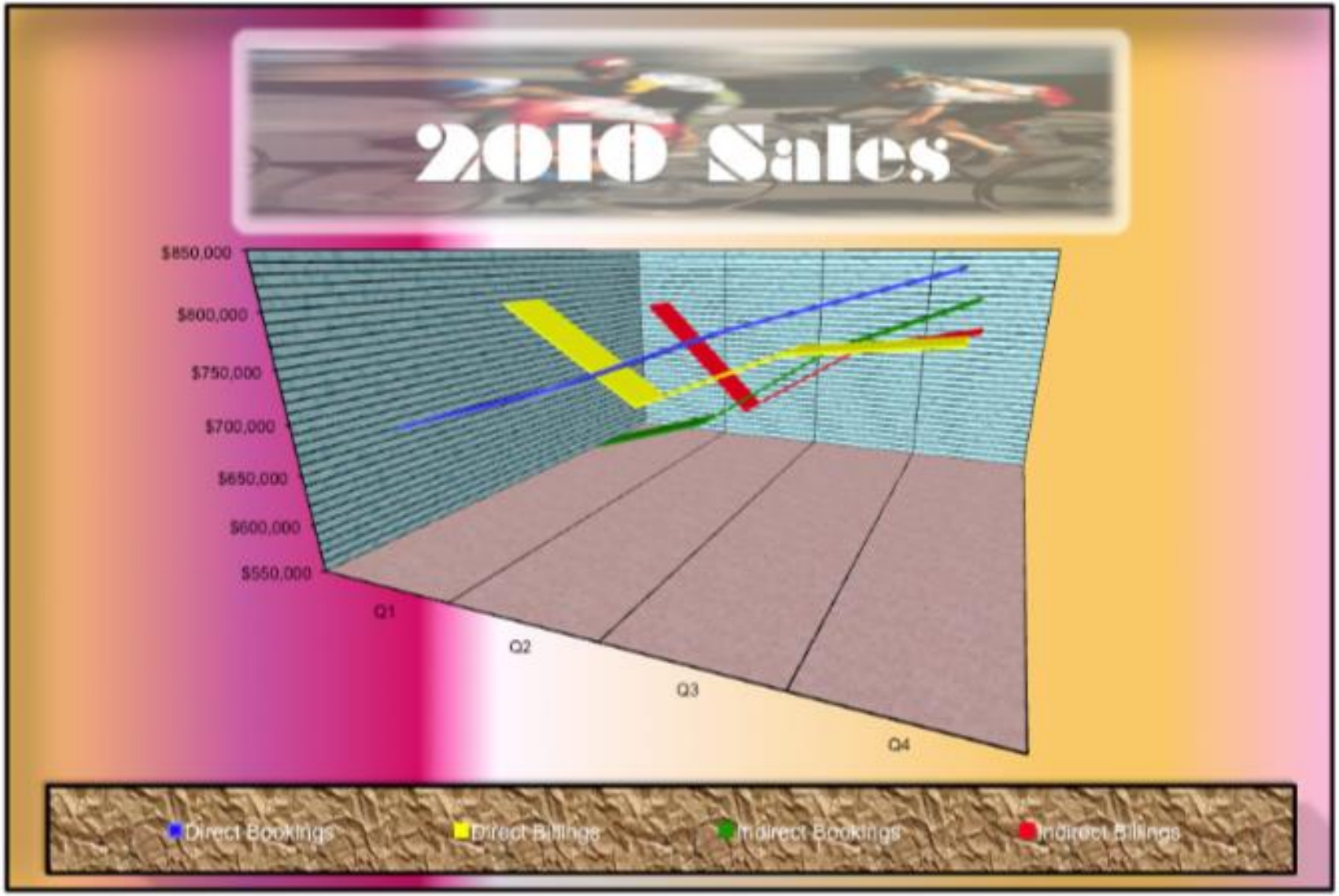
Chart-junk debate

- Simpler, more understandable, and more visually pleasing.



S. Few. (2011) The Chartjunk Debate - A Close Examination of Recent Findings. Visual Business Intelligence Newsletter.

Chart-junk debate



Summary

1. Present all the data that is needed for the audience to see and understand what's meaningful.
2. Present nothing that isn't needed.
3. Represent data accurately.
4. Represent data in a way that is easy for the eyes to perceive and the brain to interpret.
5. Provide appropriate context for interpreting the meaning of the data.

Self readings

- ‘HOW CHARTS LIE’ - Alberto Cairo http://albertocairo.com/
- S. Bateman et al. “Useful junk?: the effects of visual embellishment on comprehension and memorability of charts,” in ACM CHI, 2010, pp. 2573-2582.
- S. Few. (2011) The Chartjunk Debate - A Close Examination of Recent Findings. Visual Business Intelligence Newsletter.

2022 CSIG-VIS
International Lecture Series 13

March 24, 2022 9:00-10:30
Beijing time (UTC/GMT+08:00)
https://live.bilibili.com/24003948

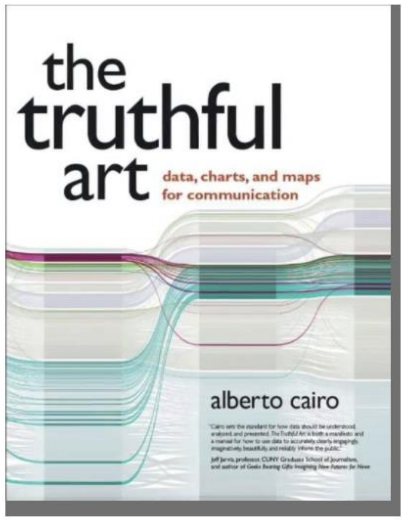
Prof. Alberto Cairo
University of Miami

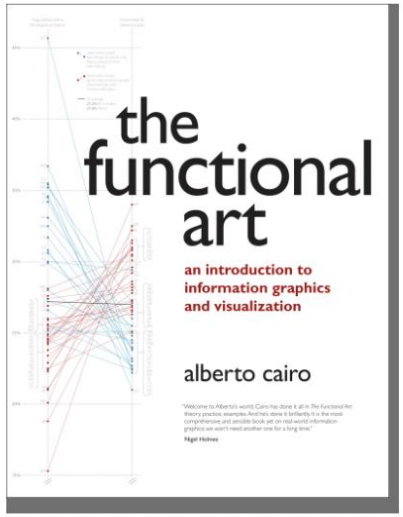
Visualization for the General Public

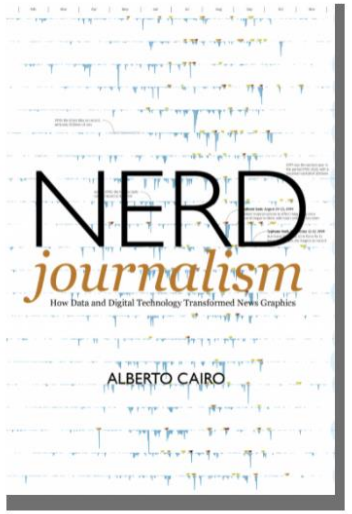
Academic visualization has traditionally been focused on designing and studying visuals to be used by experts in different domains to aid their peers. But what happens when those experts want to present visualizations to a more general public? Confusion and misunderstanding may ensue. This talk provides some tips on how to make our graphics more appealing but also more understandable.

Alberto Cairo is the Knight Chair in Visual Journalism at the School of Communication of the University of Miami (UM), He's also the director of the visualization program at UM' s Institute for Data Science and Computing (iDSC). Cairo is the author of several books about visualization, such as The Truthful Art (2016), and How Charts Lie (2019). He has led visualization teams in media organizations in Spain, Brazil, and the United States, and it's a regular consultant for several governmental organizations and companies such as Google.

Contact email:csig_vis_forum@163.com
China Society of Image and Graphics, Technical Committee on Visualization and Visual Analysis







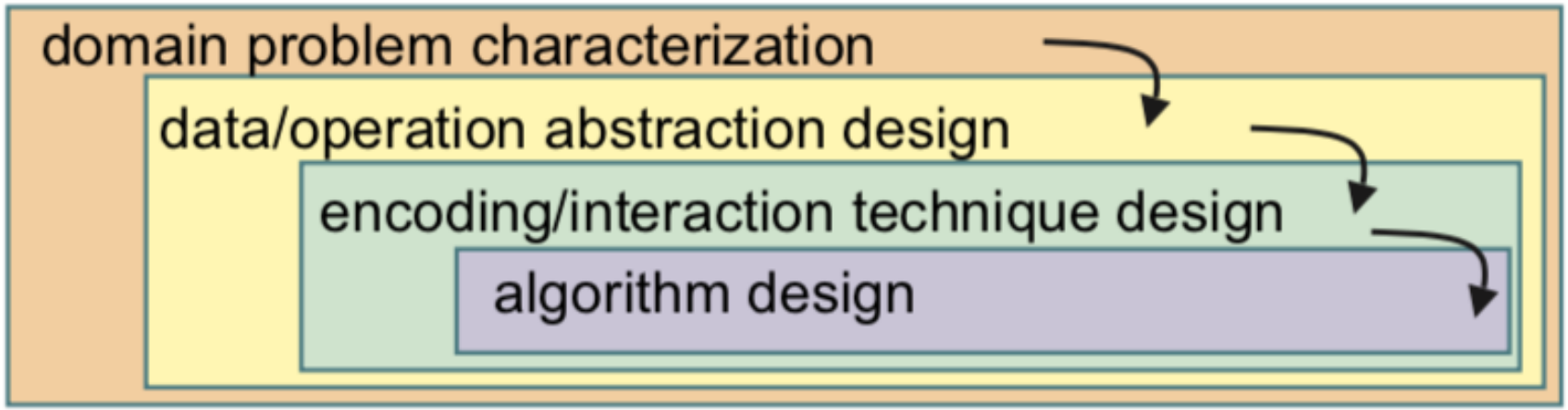
Data Exploration & Visualization

Module 5: Design Principles

- Integrity principles
  - Not to lie with data visualization
- Tufte’s rules
- Chart-junk debate
- Nested model

Nested model

- Nested model
  1. Domain problem and data characterization
  2. Operation and data type abstraction
  3. Visual encoding and interaction design
  4. Algorithm design



T. Munzner, "A Nested Model for Visualization Design and Validation," IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 921-928, 2009.

Nested model

- Domain problem characterization
  - Domain: physics, biology, transportation, etc.
  - Problems: a detailed set of domain questions
    - cure disease ✗
    - investigate microarray data showing gene expression levels and the network of gene interactions ✓
- Data characterization
  - General features of objects in the target class
    - Data type errors, nulls
    - Quintiles, max, min, mean
    - Create meta-data as needed
    - …

T. Munzner, "A Nested Model for Visualization Design and Validation," IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 921-928, 2009.

Nested model

- Abstraction: from **specific domain** to **abstract** and **generic description** in computer science
- Operation Abstraction
  - **High-level:** expose uncertainty, confirm hypotheses, concretize relationships, formulate cause and effect…
  - **Low-level:** retrieve value, filter, find extremum, sort, determine range, find anomalies, cluster, correlate…
- Data Abstraction
  - Tabular, node-link graph or tree, spatial

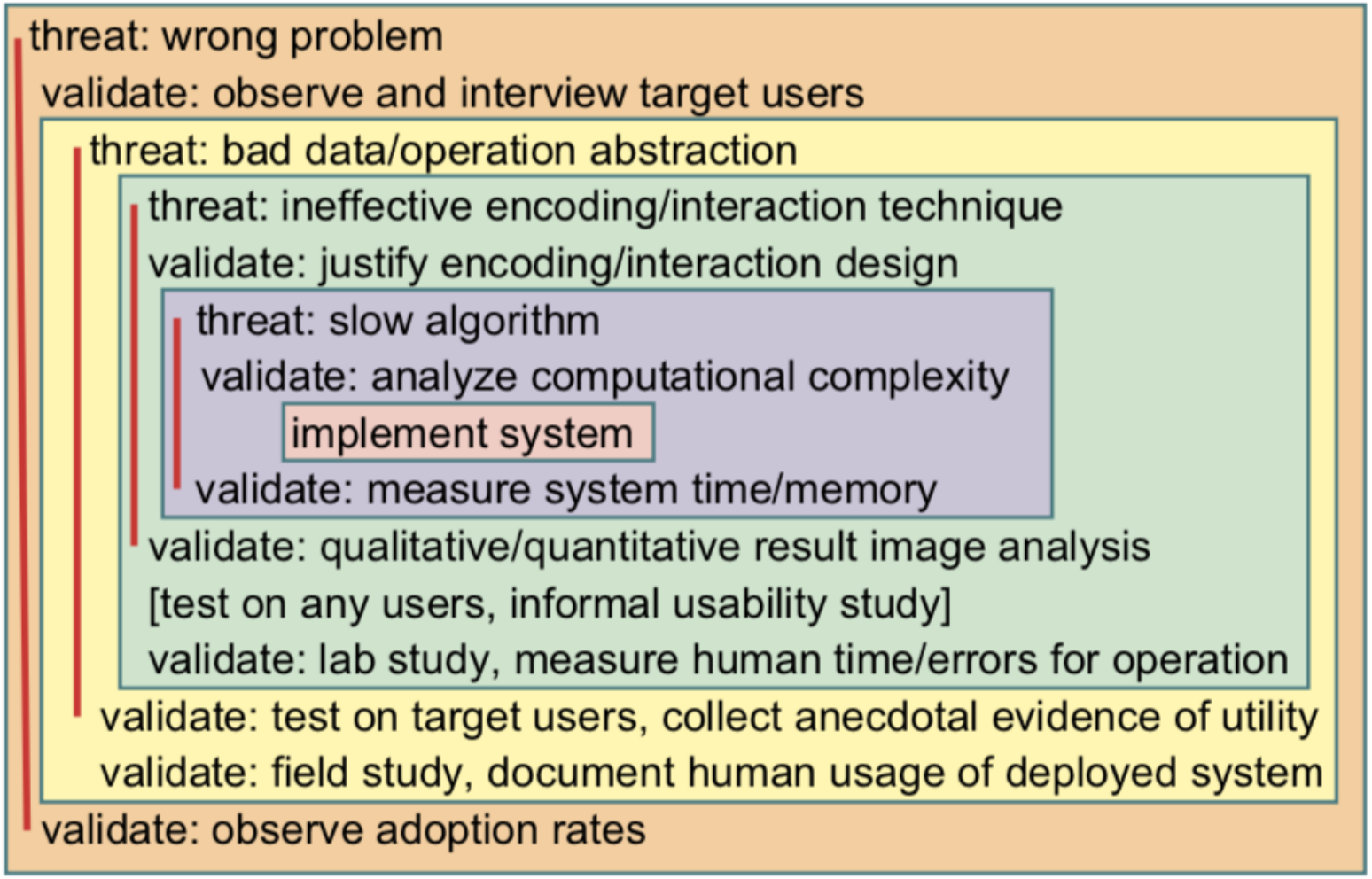
T. Munzner, "A Nested Model for Visualization Design and Validation," IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 921-928, 2009.

Nested model

- Visual Encoding and Interaction Design
  - Visualization techniques
    - Bar, line, circle, area, text…
  - Interaction techniques
    - Linking & brushing
    - Filtering
    - …
- Algorithm Design
  - to carry out the visual encoding and interaction designs automatically

T. Munzner, "A Nested Model for Visualization Design and Validation," IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 921-928, 2009.

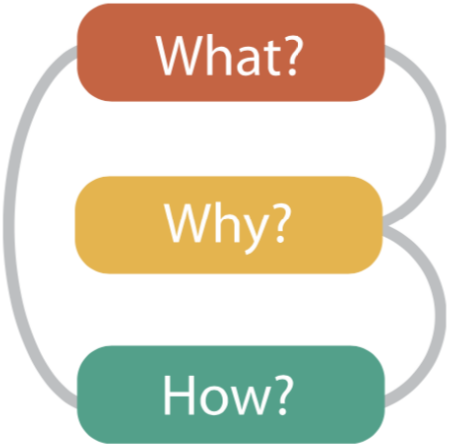
Nested model



T. Munzner, "A Nested Model for Visualization Design and Validation," IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, pp. 921-928, 2009.

What-why-how

- **what** is shown?
  - **data** abstraction
- **why** is the user looking at it?
  - **task** abstraction
- **how** is it shown?
  - **visual encoding and interaction**

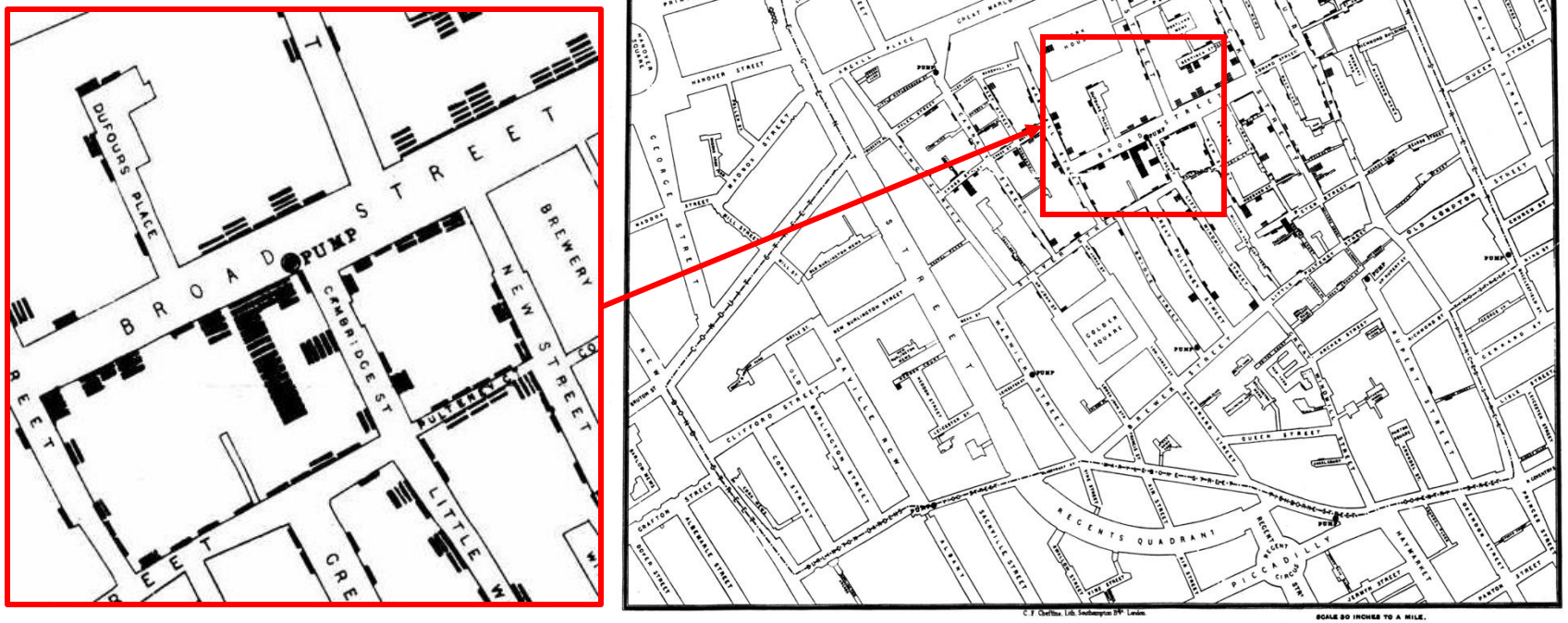


Credit: Tamara Munzner

- Abstract vocabulary avoids domain-specific terms

What-why-how

- Map of the 1854 London cholera outbreak
  - John Snow



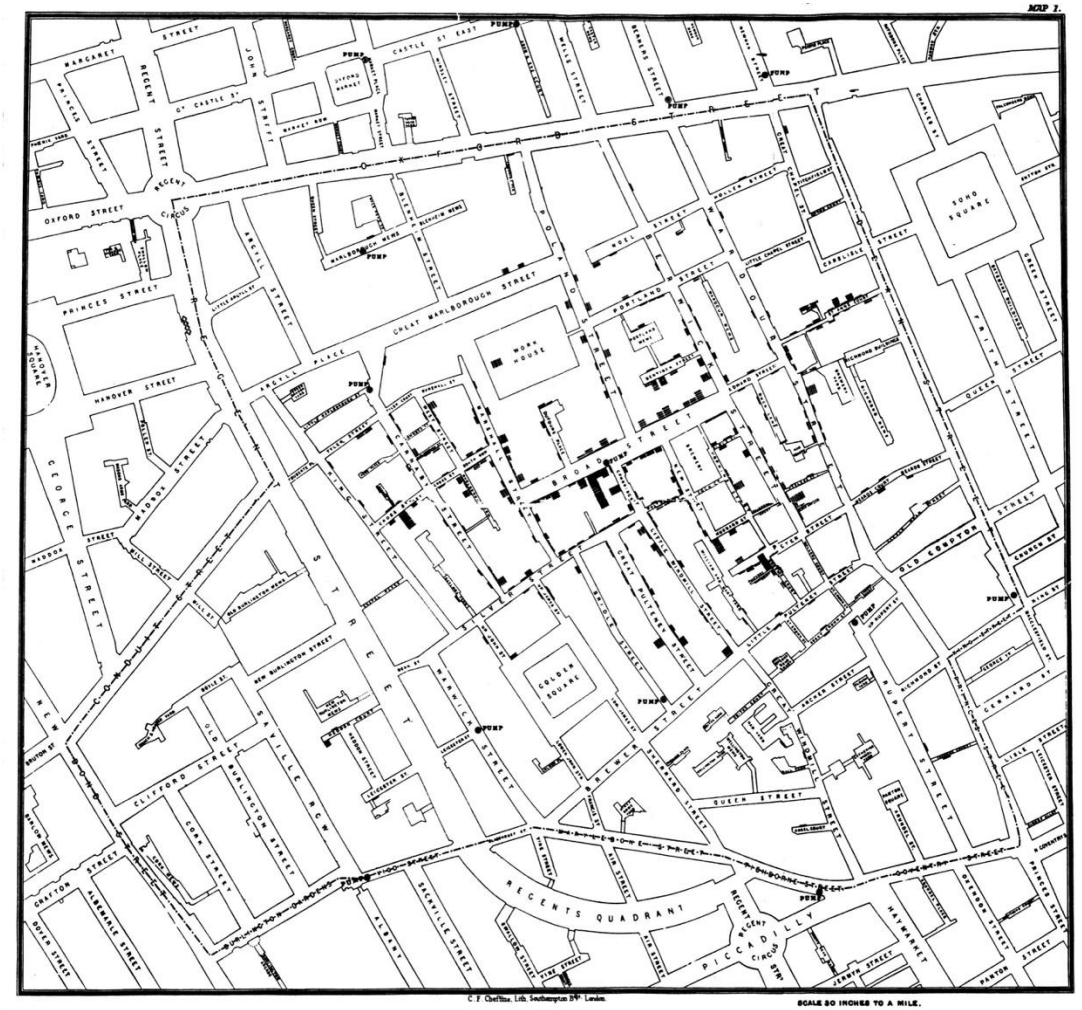
What-why-how

- Snow mapped 13 public wells and all known cholera deaths.
- Noted the spatial clustering of cases around one particular water pump on the southwest corner of the intersection of Broad Street and Cambridge Street.

- He examined water samples from various wells under a microscope, and confirmed the presence of an unknown bacterium in the Broad Street samples.

What-why-how

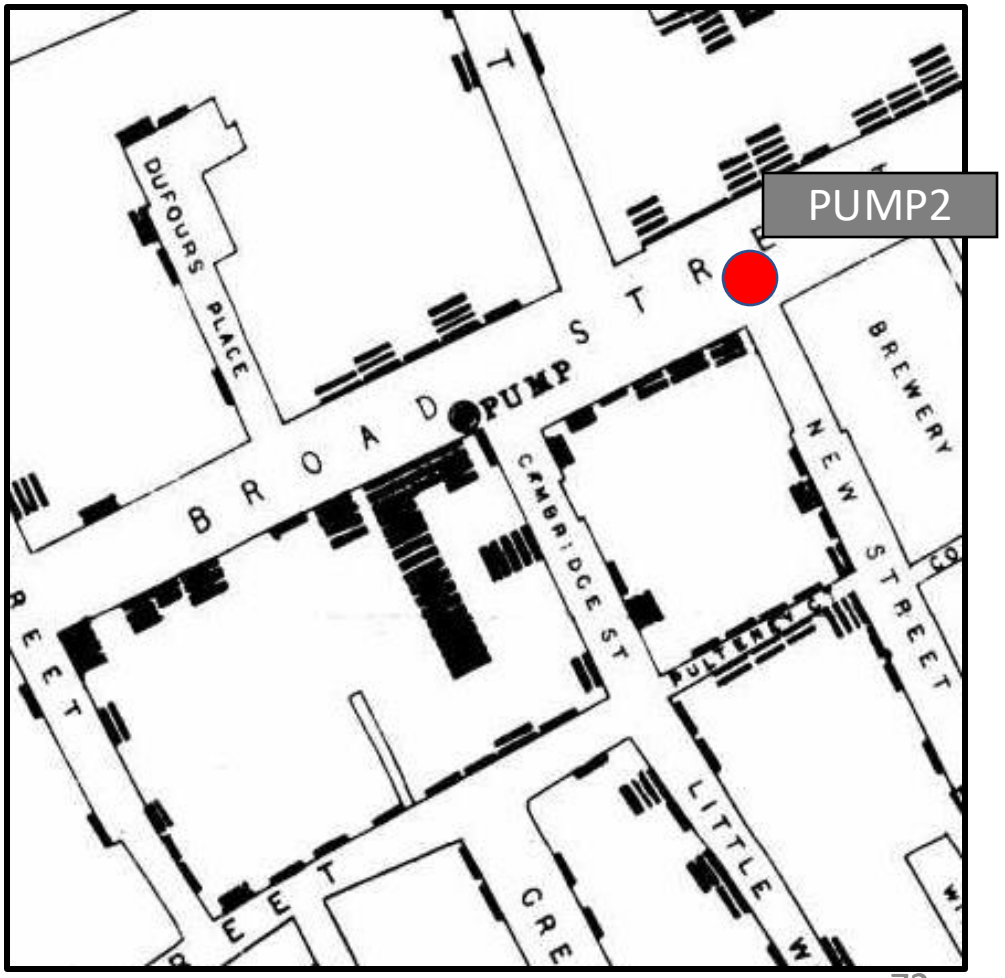
- **What**?
  - Cholera deaths
  - Public wells
- **Why?**
  - Cluster
  - Correlate
- **How?**
  - Wells → dots on the map
  - Deaths → stacked bars



Design options

- Visualization designs are task- and data-dependent.

- What-if scenarios
  - Another pump nearby?

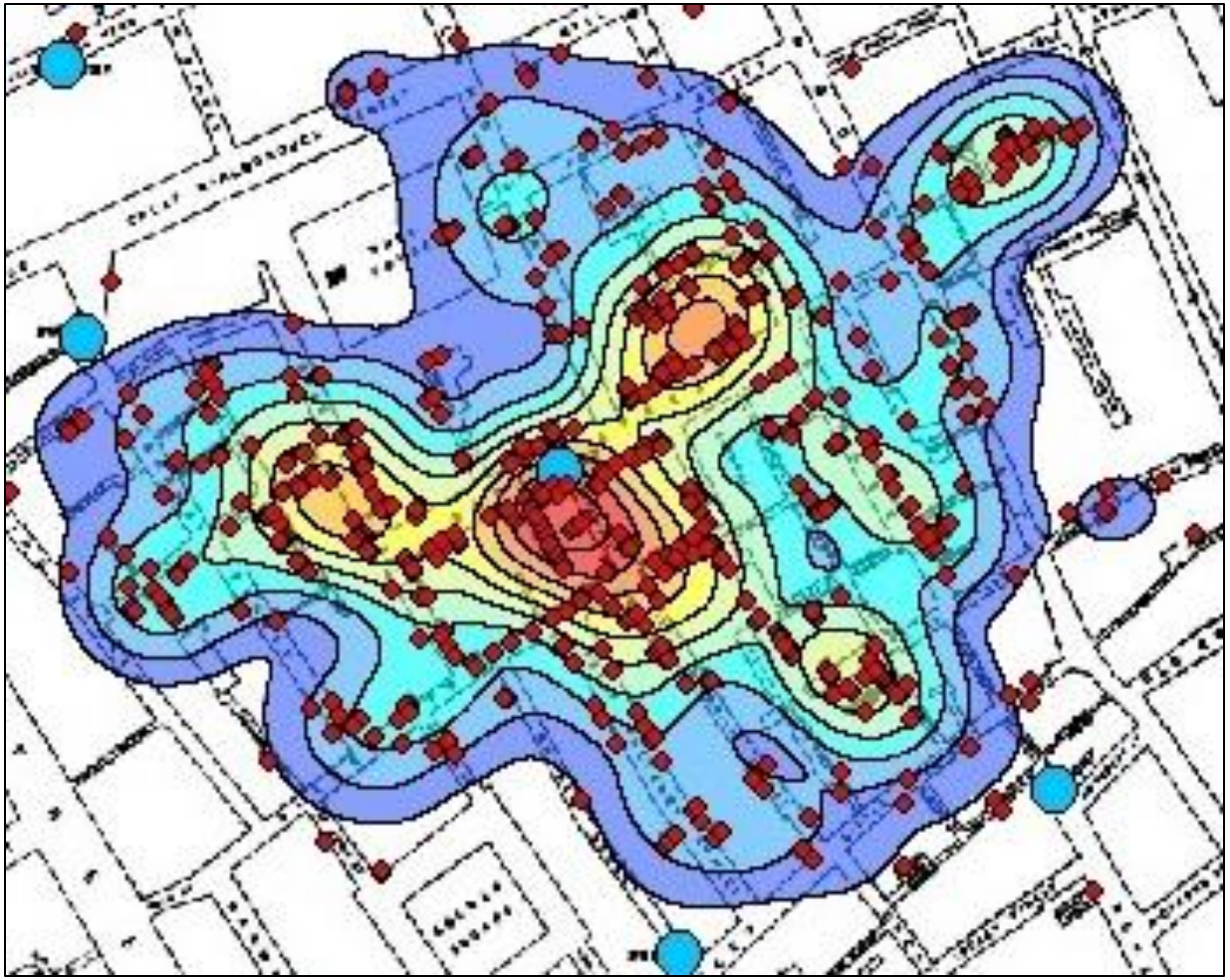


Design options

- Map + circle charts

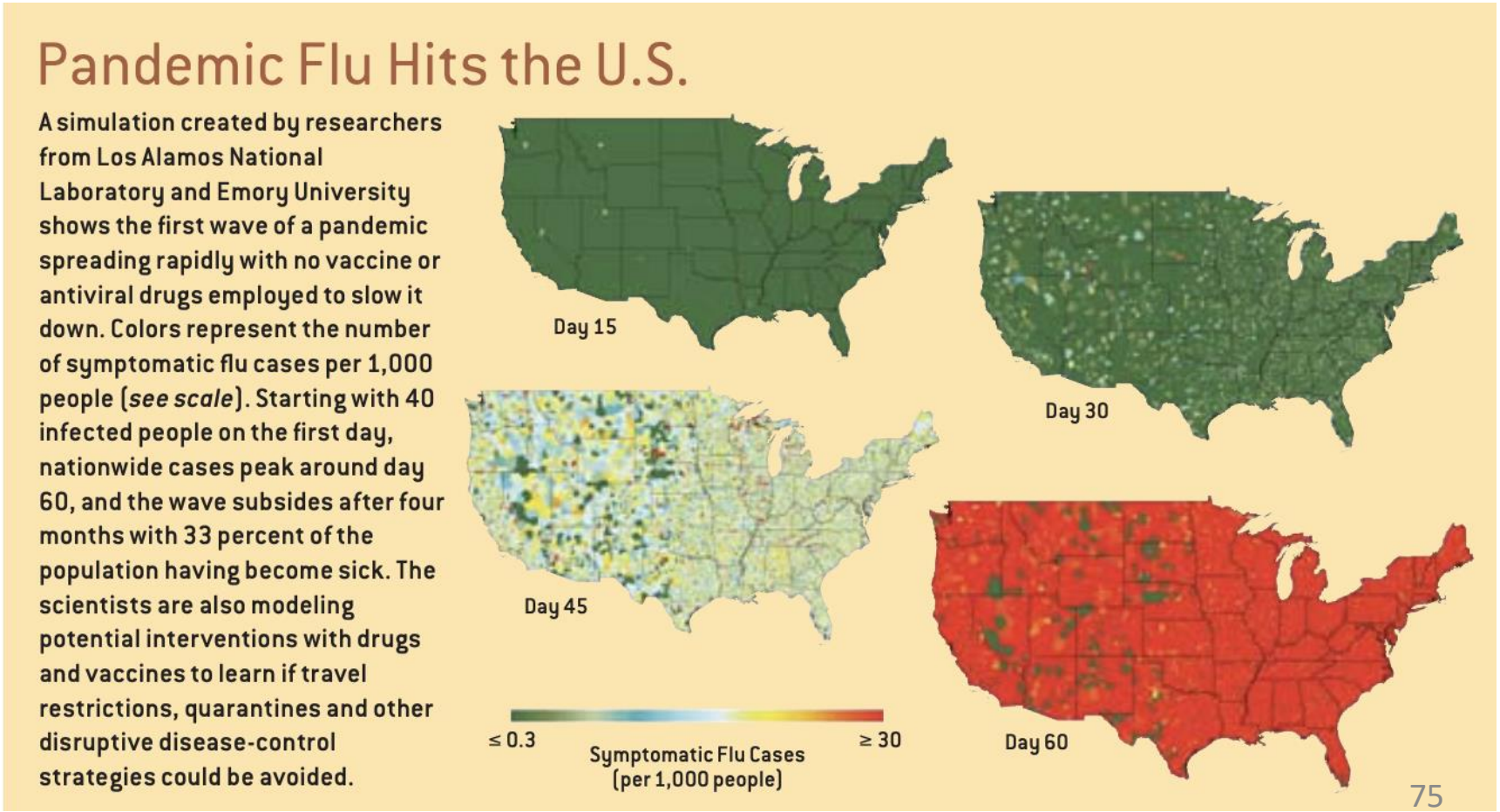
Design options

- Density map



In-class exercise: Design critique

- Critique the design: what works, what doesn't



Source: Preparing for a Pandemic *Scientific American*, 293(5). November, 2005, p. 50