IBEHS 4C03: Statistical Methods in Biomedical Engineering

Principles of Data Visualization

Carol Bassim , DMD, MHS
Assistant Professor, CLA
Division of Education and Innovation
Department Of Medicine

location: MDCL
phone: (905) 525-9140
email: bassimc@mcmaster.ca
web: http://ibiomed.mcmaster.ca

McMaster University
BRIGHTER WORLD

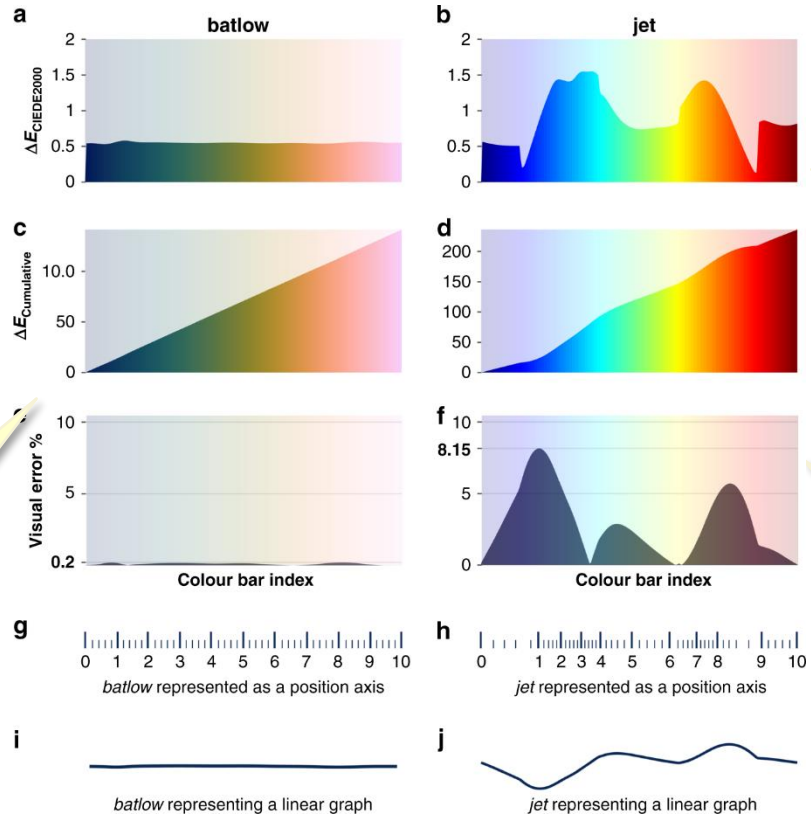
Data Visualization: Inclusion and Differences in Learning 11

McMaster University

BRIGHTER WORLD

Tips on Colour

- Colour can be effective but may not always be possible
- There is **no standard colour progression** – safest is black to white
- Design in black and white and then add colour
- Contrast in colours against one another and against background are important (3:1) ratio



Perceptually uniform is better

Nonlinear perceptual distance

Avoid “RGB rainbow”

It distorts your data

Tips on Colour: useful references

- ColorCET: information on perceptually uniform colour maps for different purposes
- General advice on colours for scientific visualization
- Nature article: discussing good and bad use of colour in scientific publications

Tips on Colour: example

- Position and length can't be used to convey information here
- Use colour
  - Hue = value
  - Saturation = uncertainty

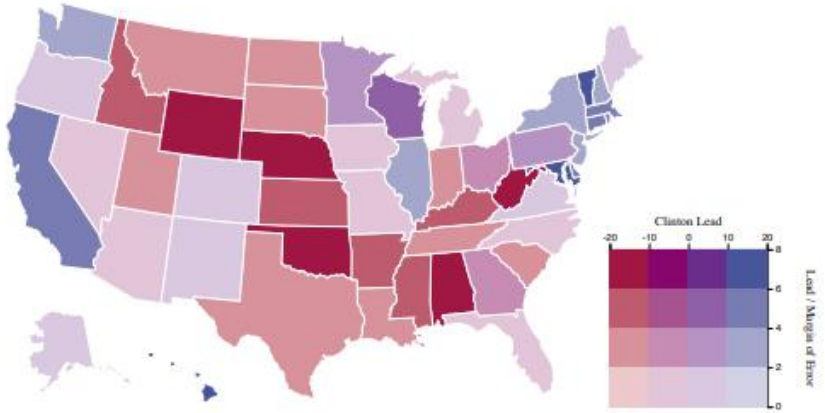


Figure 4: Polling data prior to the 2016 U.S. Presidential Election, encoded in a traditional 2D bivariate map. The redness and blueness show the polling lead for Trump and Clinton, respectively. But many polls had high margins of error, creating uncertainty about the election results.

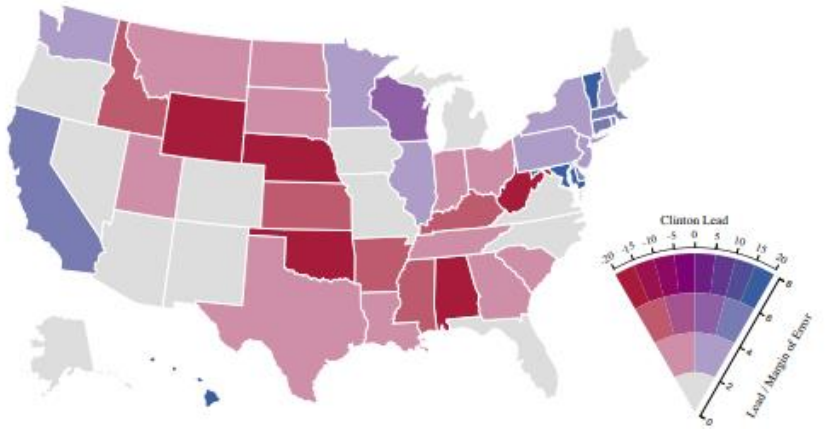


Figure 5: Polling data prior to the 2016 U.S. Presidential Election, encoded with a VSUP.

https://idl.cs.washington.edu/files/2018-UncertaintyPalettes-CHI.pdf

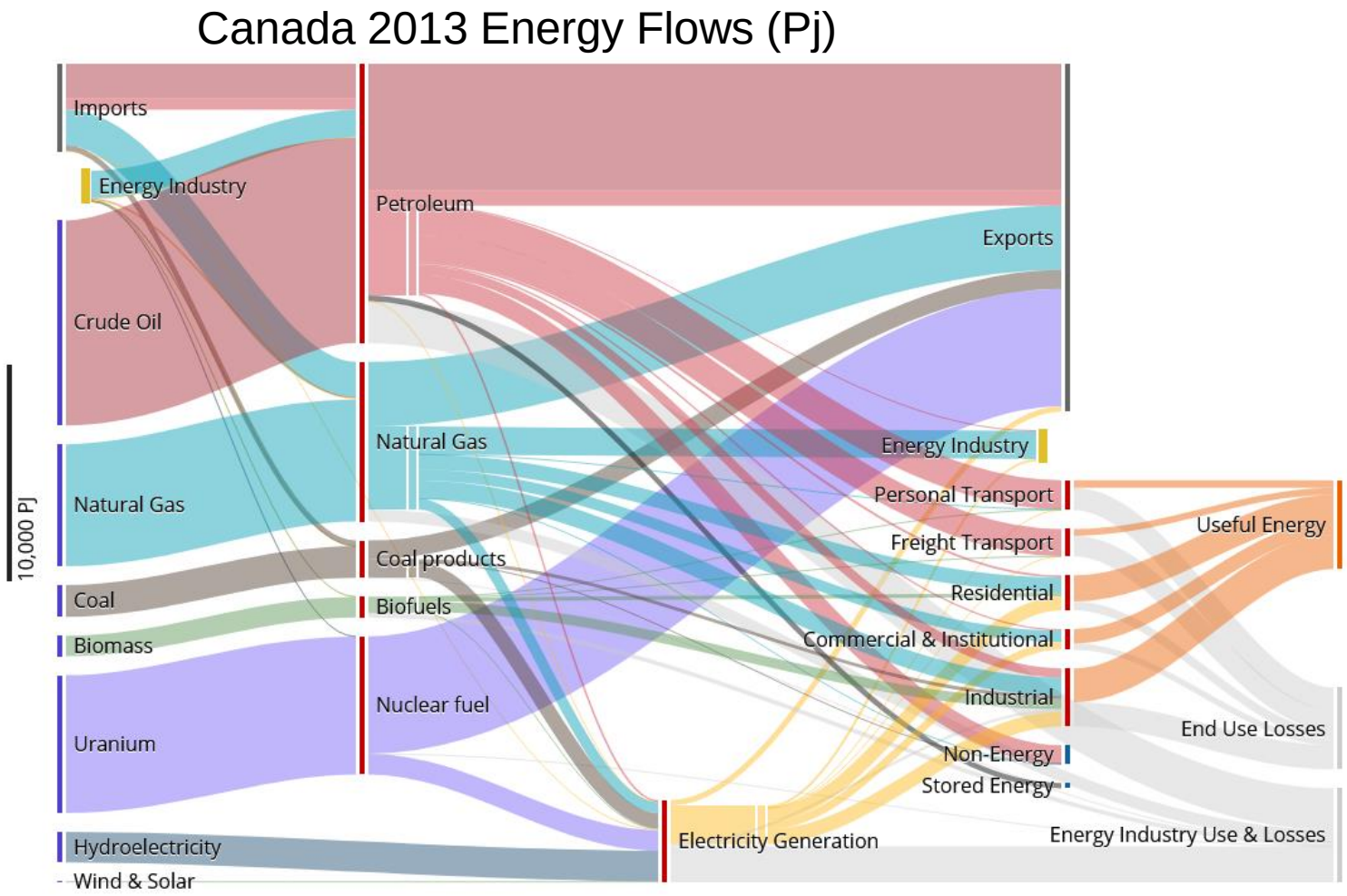
Think about visual challenges

- Make text clear and readable
- Simplify the main message that you are trying to convey
- Think about aging population and visual challenges
- Think through the mobile nature of information sharing
  - Will the visualization be read on a tablet, on a phone, on a watch?

Visualization and Differences in Learning

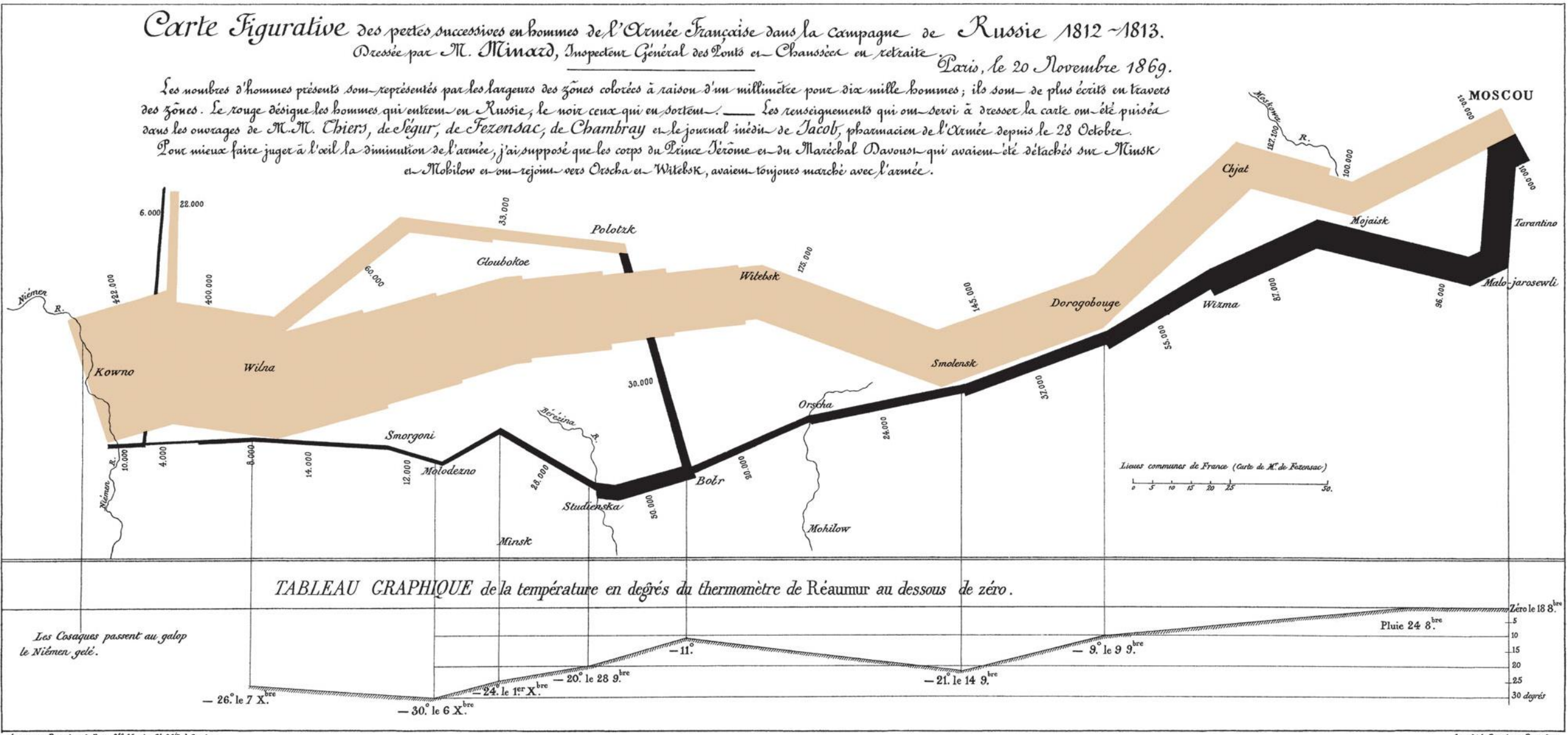
- Difference in thinking and learning styles
  - Visual, text based, descriptive, analytic
- Plots and text go well together: a plot = a paragraph of text
  - Add labels to interesting data points
  - Add equations
  - Add a summary table
- Think of your audience

Other plots: Sankey diagrams



https://www.cesarnet.ca/visualization/sankey-diagrams-canadas-energy-systems

https://www.python-graph-gallery.com/sankey-diagram/



Charles Joseph Minard's map depicting Napoleon's disastrous 1812 march on Russia during the Napoleonic Wars.

Other plots: Useful References

- Data visualization catalogue: descriptions of different visualization types
- Python Graph Gallery: Graph examples and the Python code to generate them
    - They may use different libraries than what I show in this course
- The Matplotlib documentation has a lot of good examples

Principles of Data Visualization 12

McMaster University

BRIGHTER WORLD

Data visualization will always be part of your job

- *Coworker:* Here are the yields from a batch system for the last 3 years (1256 data points), can you please help me:
  - Understand more about the time-trends over the 3 years?
  - Efficiently summarize the yield from all batches run last year?
- *Manager:* I need you to effectively summarize both the number and type of defects on the 17 different aluminum grades we produces that occurred in the last 12 months
- Yourself: 24 different variables being measured vs time (5readings per minute, over 300 minutes) for each batch; how can we visualize these 36,000 data points?

We need good plots to make decisions quickly, correctly, and confidently

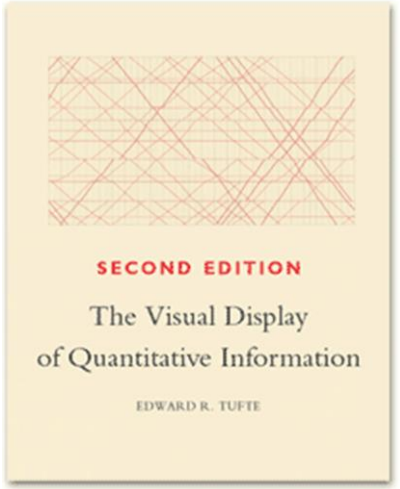
References

1. **Edward Tufte**, Envisioning Information, Graphics Press, 1990. (10th printing in 2005)
2. Edward Tufte, The Visual Display of Quantitative Information, Graphics Press, 2001.
3. Edward Tufte, Visual Explanations: Images and Quantities, Evidence and Narrative, 2nd edition, Graphics Press, 1997.
4. **William Cleveland**, Visualizing Data, and The Elements of Graphing Data, Hobart Press; 2nd edition, 1994.
5. Stephen Few, Show Me the Numbers, and \Now You See It", Analytics Press.
6. Su, It's easy to produce chartjunk using Microsoft Excel 2007 but hard to make good graphs, Computational Statistics and Data Analysis, 52 (10), 4594-4601,2008
7. Semiology of Graphics – Bertin (1967) (latest printing in 2010)
8. Show me the numbers: designing tables and graphs to enlighten – Stephen Few (2004)
9. **Wainer Howard**. How to Display Data Badly. The American Statistician; 38: 137-147 (1984)

Tufte's Principles of Graphical Excellence
The Visual Display of Quantitative Information (2001)

Edward Rolf Tufte was a political scientist and statistician who defined the principles of graphical excellence:

- **Graphical excellence is the well-designed presentation of interesting data—a matter of substance, of statistics, and of design.**
- **Graphical excellence consists of complex ideas communicated with clarity, precision and efficiency.**
- Graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space.
- Graphical excellence is nearly always multivariate.
- And graphical excellence requires telling the truth.



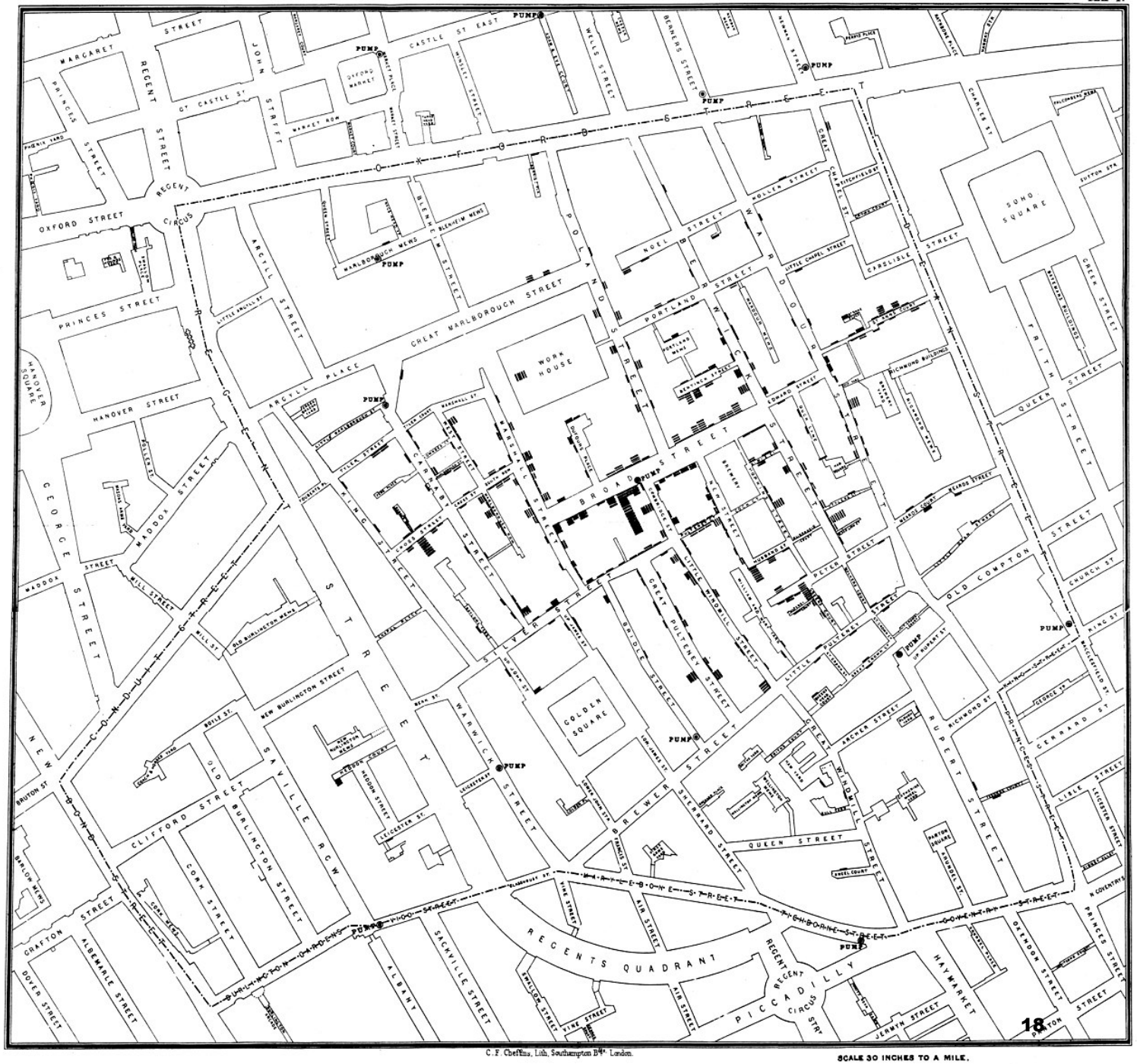
Principles of Graphical Excellence

- Show the data
  - The human eye and brain are excellent at pattern recognition
    - Detection: Sorting signals from noise
    - Estimation: Comparing size, strength, and relative intensities
    - Assembly: Seeing multiple layers of information
      - Encourage the eye to compare different pieces of information
      - Reveal the data at several levels of detail
- We can deal with bad plots, but this takes time and effort
  - Serve a reasonable clear purpose with your graph

Graphical excellence is that which gives to the **viewer** the **greatest number of ideas in the shortest time** with the **least ink in the smallest space**.

- Trust the viewer: Show the data
- Maximize data density (think about the data ink)
- Minimize data junk
- Multivariant

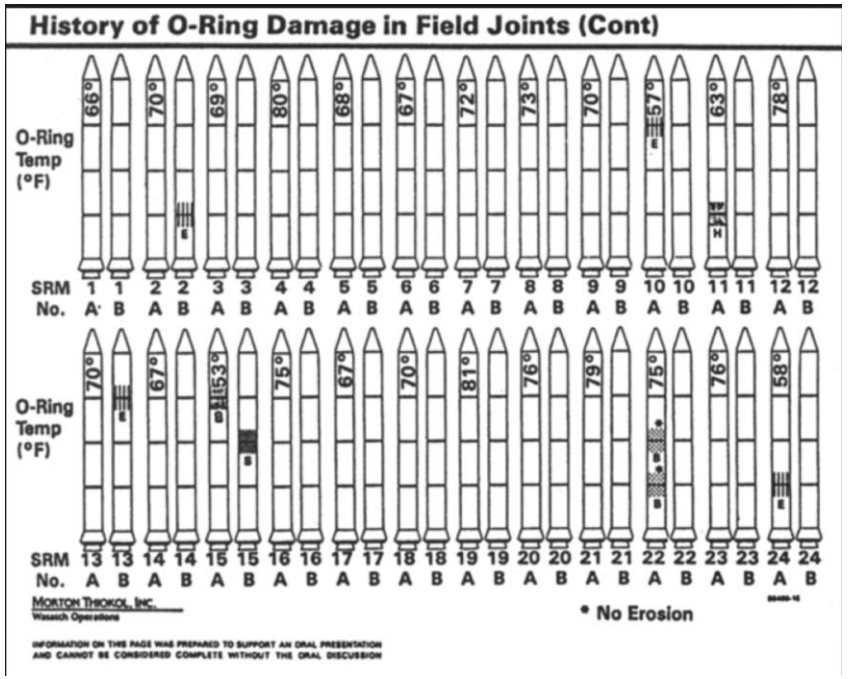
John Snow's Cholera Outbreak Map 1854



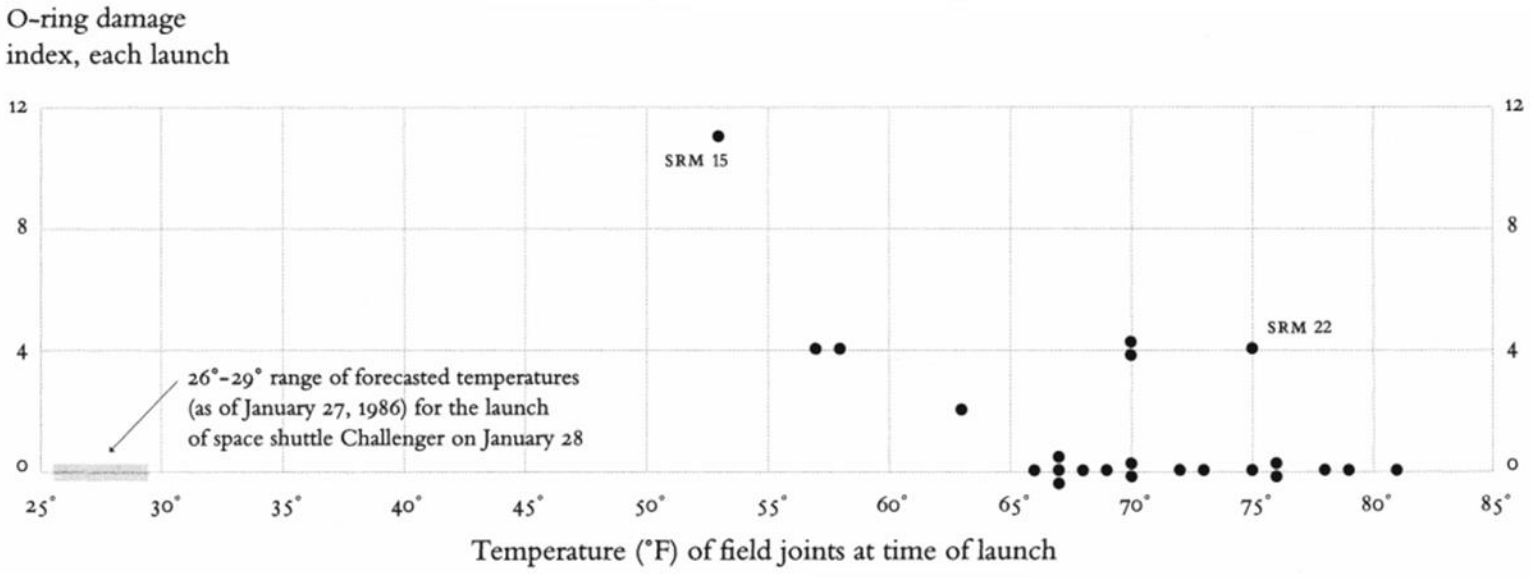
John Snow's Cholera Outbreak Map 1854

Principles of Graphical Excellence

Before the 1986 Challenger launch, NASA engineers had collected data indicating that rubber O-rings failed to seal correctly in cold weather.



Scatterplot showing the temperature by O-ring damage data for the test launches prior to the Challenger launch and explosion.



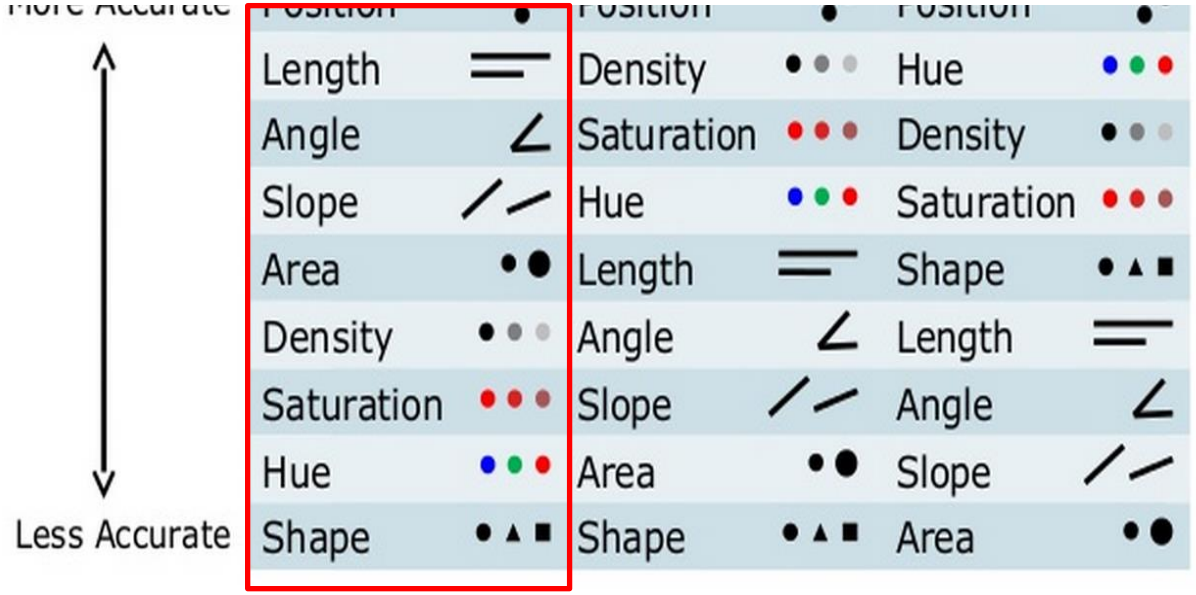
References:

The Space shuttle Challenger explosion blog post, by Vikram Dayal
Robison et al. (2002) Representation and Misrepresentation: Tufte and the Morton Thiokol Engineers on the Challenger, Science and Engineering Ethics, 8, 59-81.
Tufte, Edward R. (1997) “Visual Explanations: Images and Quantities, Evidence and Narrative”, Graphics Press, Cheshire, Connecticut.

And graphical excellence requires telling the truth.

- 6 Principles of Graphical Integrity (adapted)
  - Proportional representation of truth (the lie factor).
  - Clear and detailed labeling.
  - Show data variation, not design variation.
  - In time-series displays of money, deflated and standardized units of monetary measurement are nearly always better than nominal units.
  - The number of information-carrying (variable) dimensions depicted should not exceed the number of dimensions in the data.
  - Graphics must not quote data out of context.

Cleveland (1994) graphical features hierarchy

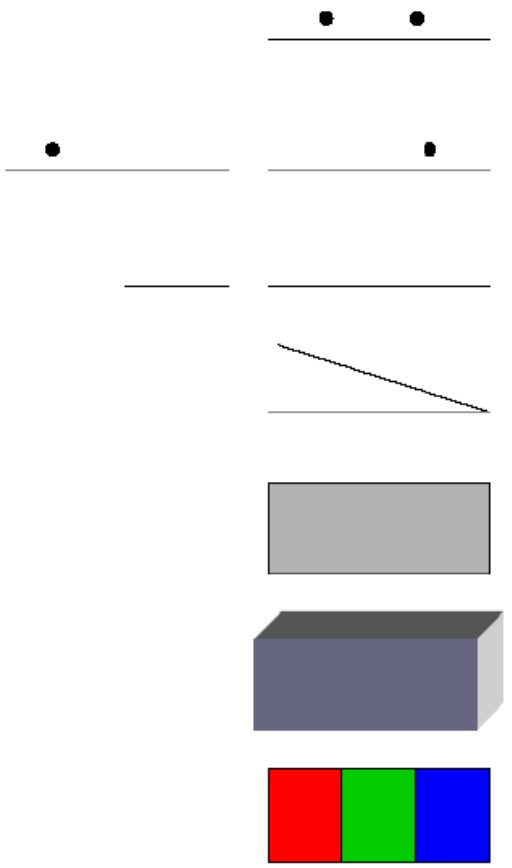


- We live in 3D but work best in 2D
- The human eye is better at comparing some things than others
- Typically, in engineering we deal with quantitative data

Cleveland, William (1994). The elements of graphing data. Summit, New Jersey: Hobart Press. ISBN 0-9634884-1-4.

Cleveland (1994) graphical features hierarchy

Best information transfer



Position along a common scale: *bar chart, time series, scatter plot*

Position along nonaligned axis: *side by side graphs*

Length: *stacked bar charts, waterfall charts*

Angle / Slope: *pie charts*

Area: *pie charts, matrix charts, radar charts*

Volume: *3d charts*

Colour

Worst information transfer

12 Tips on How to Display Data Badly

Adapted from Wainer H. How to Display Data Badly. The American Statistician 1984; 38: 137-147.

- Show as few data points as possible
- Hide what data you do show; minimize the data-ink ratio
- Ignore the visual metaphor altogether
- Only order matters
- Graph data out of context
- Change scales in mid-axis
- Emphasize the trivial; ignore the important
- Jiggle the baseline (don’t make comparisons easy)
- Alphabetize everything
- Make your labels illegible, incomplete, incorrect, and ambiguous.
- More is murkier: use a lot of decimal places and make your graphs three dimensional whenever possible.
- If it has been done well in the past, think of another way to do it

End of Module: Reviewing Learning Objectives

- Characterize and create basic graphs based on datatype.
  - Time series, tables, * scatterplots.

- Data visualization practices
  - Inclusion and learning differences
  - Tufte's Principles of Graphical Excellence
  - Cleveland's Graphical Features Hierarchy
  - Wainer's Tips on How to Display Data Badly