

Data Visualization

Why Data Visualization

- Humans are very visual
- Raw data (tables of numbers/values) are verbose
- Visualization helps communicate relationships between data points
- Visualization summarizes and removes extraneous values and allows us to discern new information

Much of this information comes from the incomplete draft of *Data Visualization* found at

<https://socviz.co/lookatdata.html>.

What Makes a Good Visualization

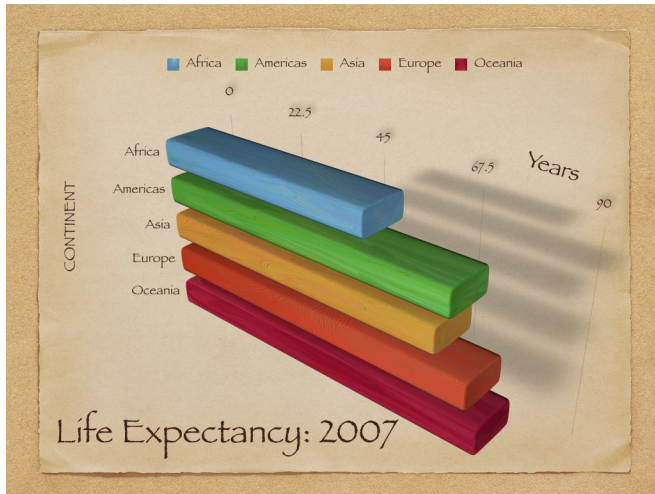
- Display the data
- Tell a story: what is this data saying?
 - Make the story honest: don't misrepresent the data.
 - Keep the story short: show a large amount of data in less space
 - Make the story make sense: make sure your representation is coherent
 - Have a story moral: what is the purpose of the visualization?
- Allow the consumer to see different levels of granularity. How much granularity is enough?

For further discussion, we will assume that you, the visualizer, is acting in good faith, is not trying to misrepresent the data intentionally, and have taken steps to ensure your data is as good as possible.

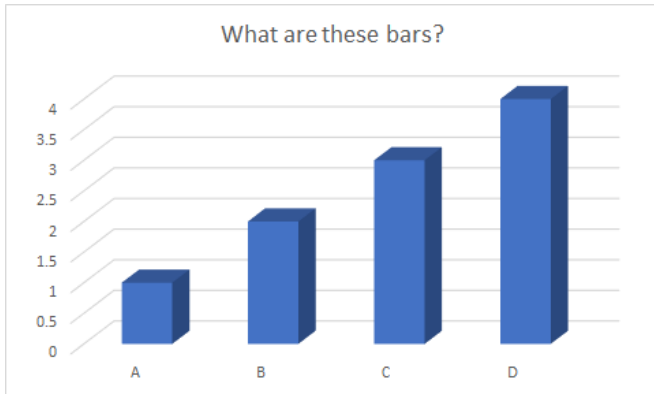
Why Can This Be Hard?

- “Visually appealing“ is subjective (what is good taste?)
- Human perception is complicated
- Data can be bad
- ...

Bad Taste in Visualizations: Chart Junk

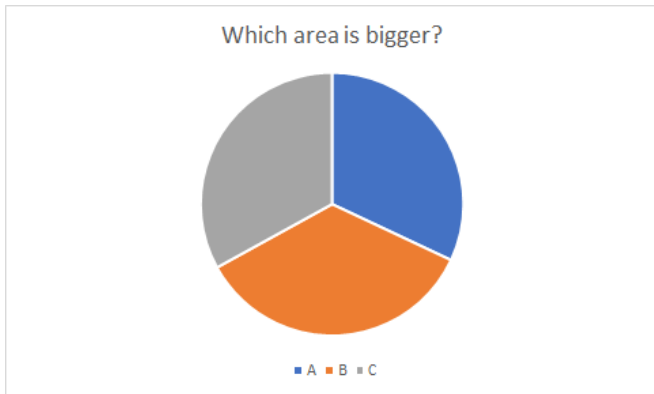


Perception is Hard!



Would you believe the y values are 1, 2, 3, and 4?

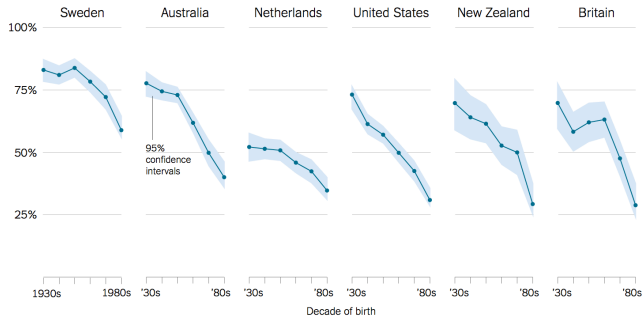
Perception is Hard! Again!



The values are 32, 33, and 35. Can you tell which is which?

Bad Data, Go Sit in the Corner

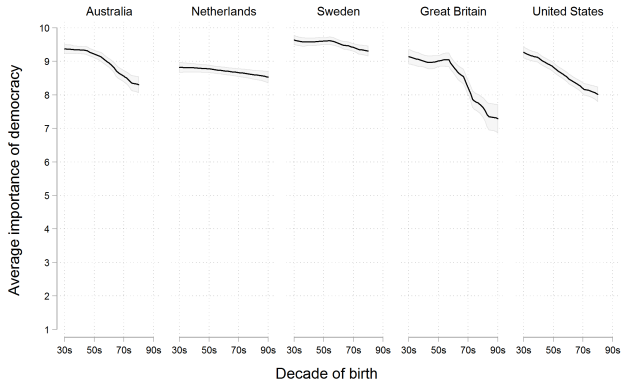
Percentage of people who say it is “essential” to live in a democracy



Source: Yascha Mounk and Roberto Stefan Foa, "The Signs of Democratic Deconsolidation," *Journal of Democracy* | By The New York Times

The original question was a 10 point scale (0 = not important, 10 = very important); how was this summarized here?

Bad Data, Go Sit in the Corner



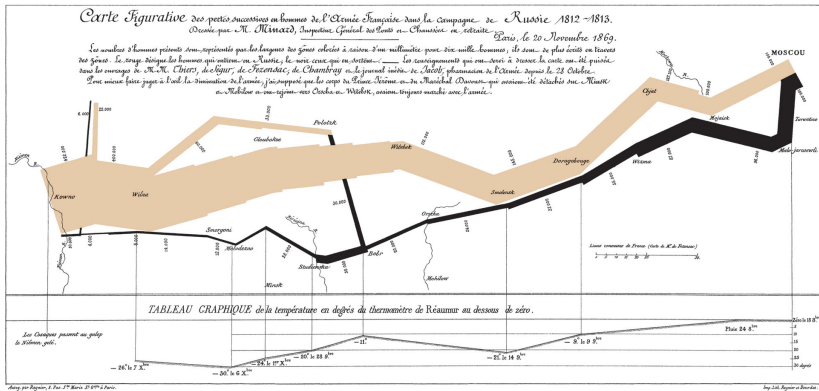
Graph by Erik Voeten, based on WVS 5

What if we look at the average result of the original scale?

A Good Visualization? Can it Be?

- A classic visualization from 1869
- Charles Joseph Minard visualized Napoleon's invasion (and retreat) from Russia
- Visualization shows 6 variables (time, 2D position (as on a map), army size, army direction of motion, and temperature during the retreat)
- This may seem difficult to express in a 2D visualization
- A visualization of this quality is difficult to reproduce with standard tools (this visualization works because it is specialized to the data being shown)

Napoleon's March on Moscow



Perception, Color, and Shapes

- Perception is hard
- There is an entire research world of color theory, cognitive research, and perception
- A good intro: <https://socviz.co/lookatdata.html#perception-and-data-visualization>
- I've highlighted many of the key points here

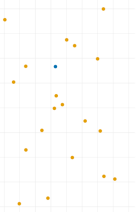
Watch Out for Optical Illusions

- Hermann Grid
- The dress (black and blue? white and gold?)
- The Checkershadow illusion
- I can't see blue LEDs on black backgrounds...
- Color combinations?

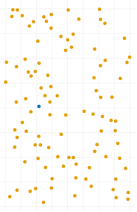
- Color theory is huge
- Accessibility is a big problem (colorblindness)
- Researching and using colors effectively and appropriately for a given problem is vital to a good visualization
- Don't go crazy (avoid chart junk)

Color and Shape

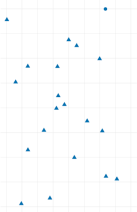
Color Only, N=20



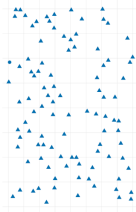
Color Only, N=100



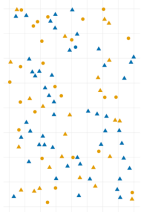
Shape Only, N=20



Shape Only, N=100

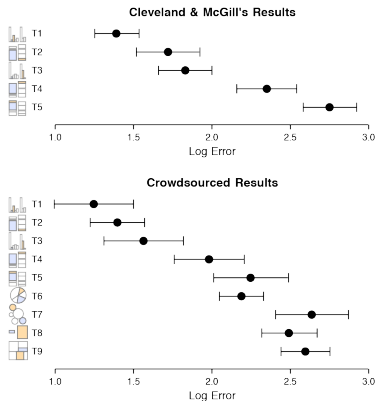


Color & Shape, N=100



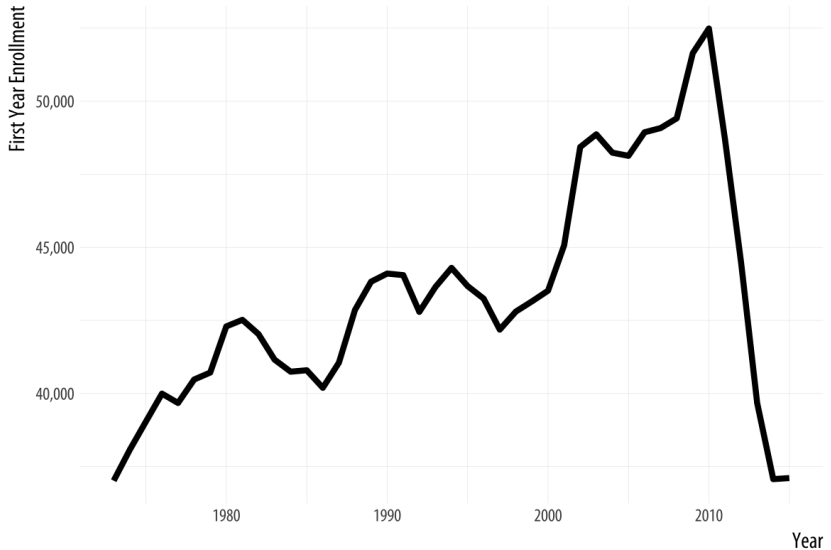
Find the blue circle; differences in color stands out more!

Not All Visualizations Are Created Equal



Notice how the transition from comparing heights, to angles, to areas makes us worse at discerning differences...

People Are Going To Get Mad



Is one of these being dishonest? What are we trying to show?

So It Can't Be Done, Right?

- It can! It's done well every day!
- Rules of thumb can help, but overall, you need to ask yourself:
 - What is this data telling me?
 - How can I clearly and easily convey that to my audience?
 - How can I make it visually pleasing enough to catch my audience's attention?
- Let's talk thumbs and rules

Choosing The Visualization

- What kind of data do you have?
- What kind of tools are available to you?
- Who is your audience?

We are using Python (and matplotlib by default), so our tools are not very limited:

<https://www.machinelearningplus.com/plots/top-50-matplotlib-visualizations-the-master-plots-python/>