get your plot game right

Graphical Integrity
Size

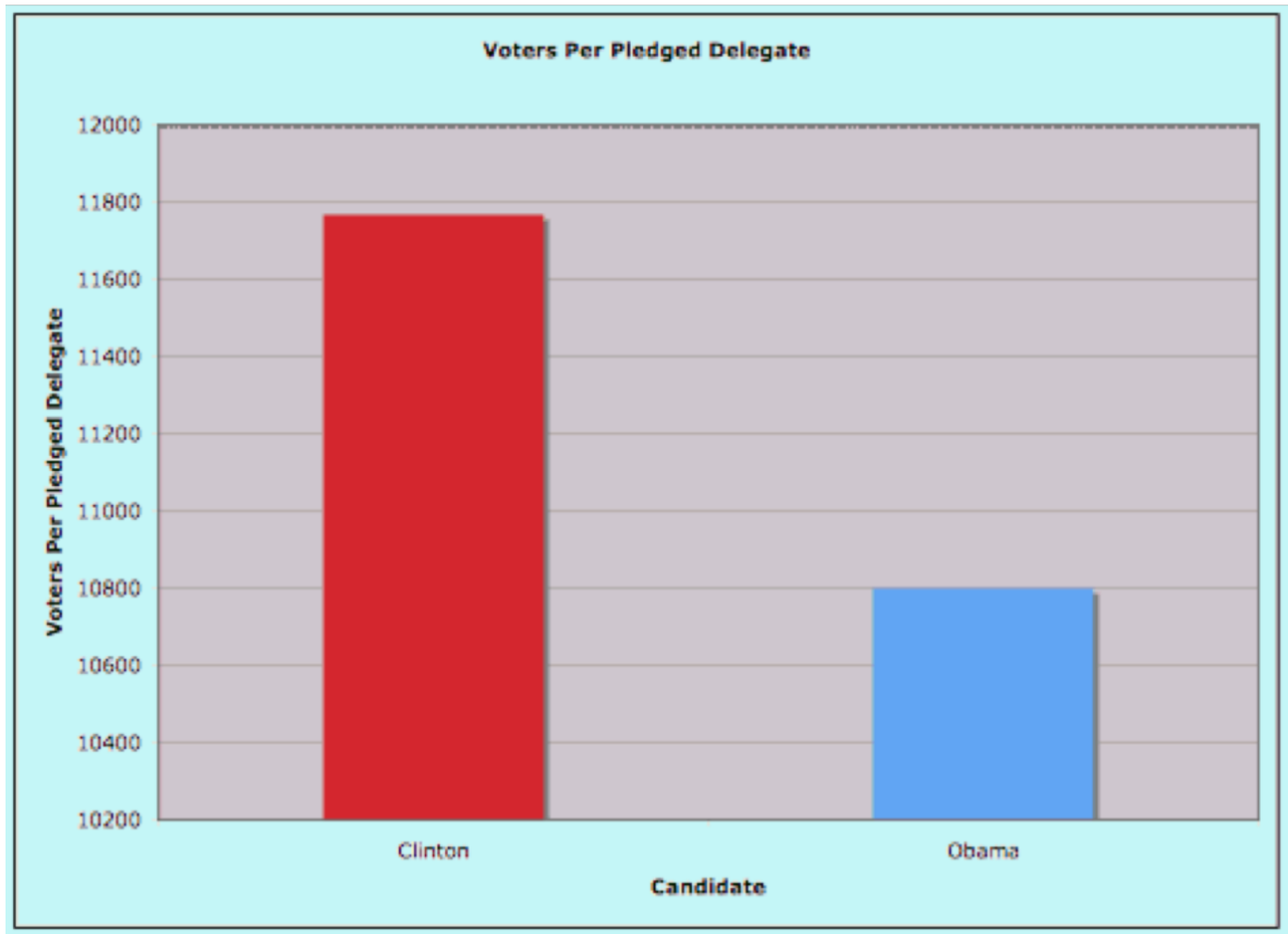
Things in your visuals should be sized in proportion to the data they represent.

- (generally) Start scales at 0 when plotting
- If you’re using circles (like in a bubble chart) be sure to size them by area, not radius/ diameter (since a change in radius makes a change^2 in area)

Examples

A graph from Real Clear Politics in 2008.

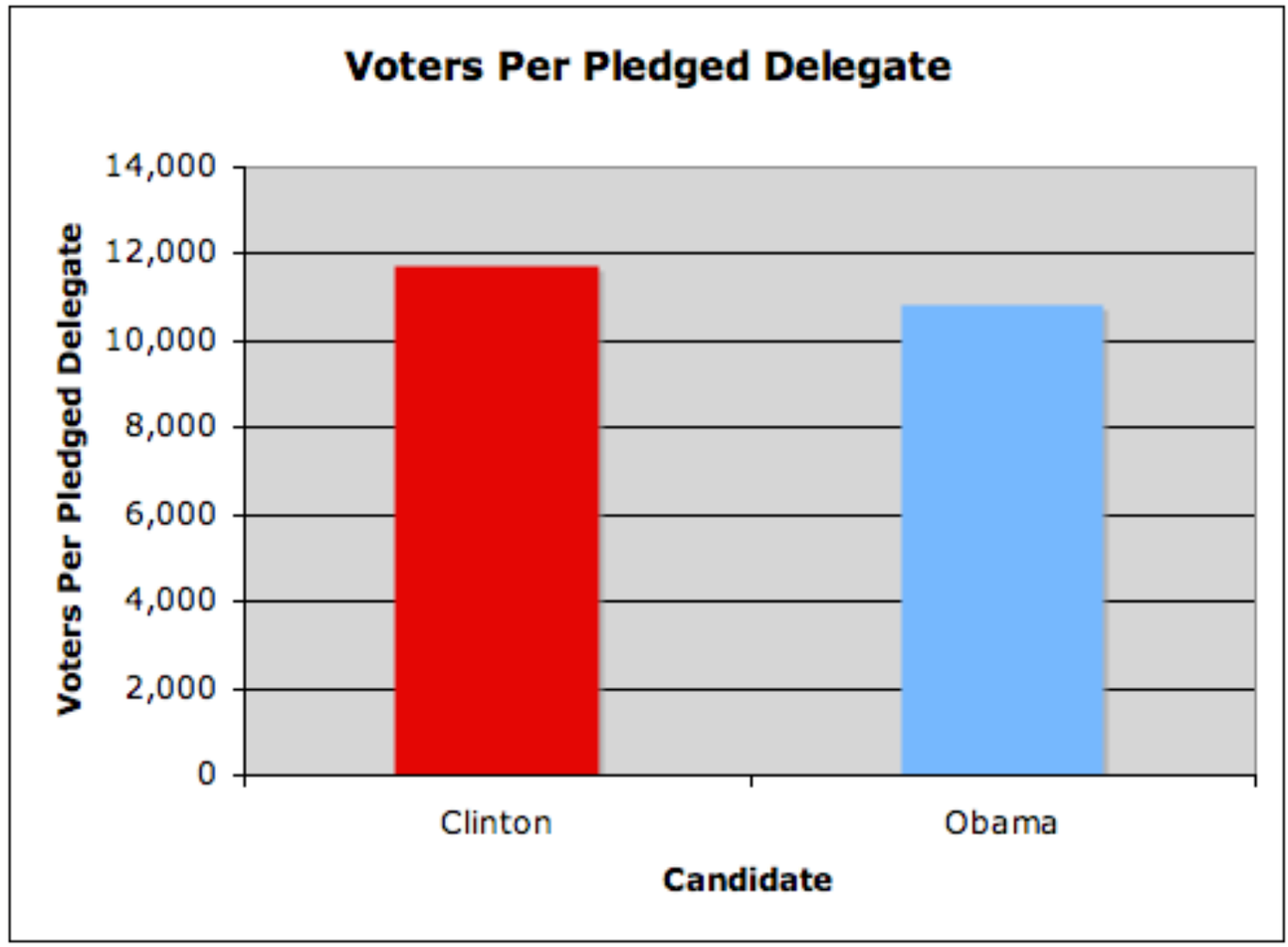
Graph shows a 61.3% difference



Examples

There's really only a 13.2% difference between the two.

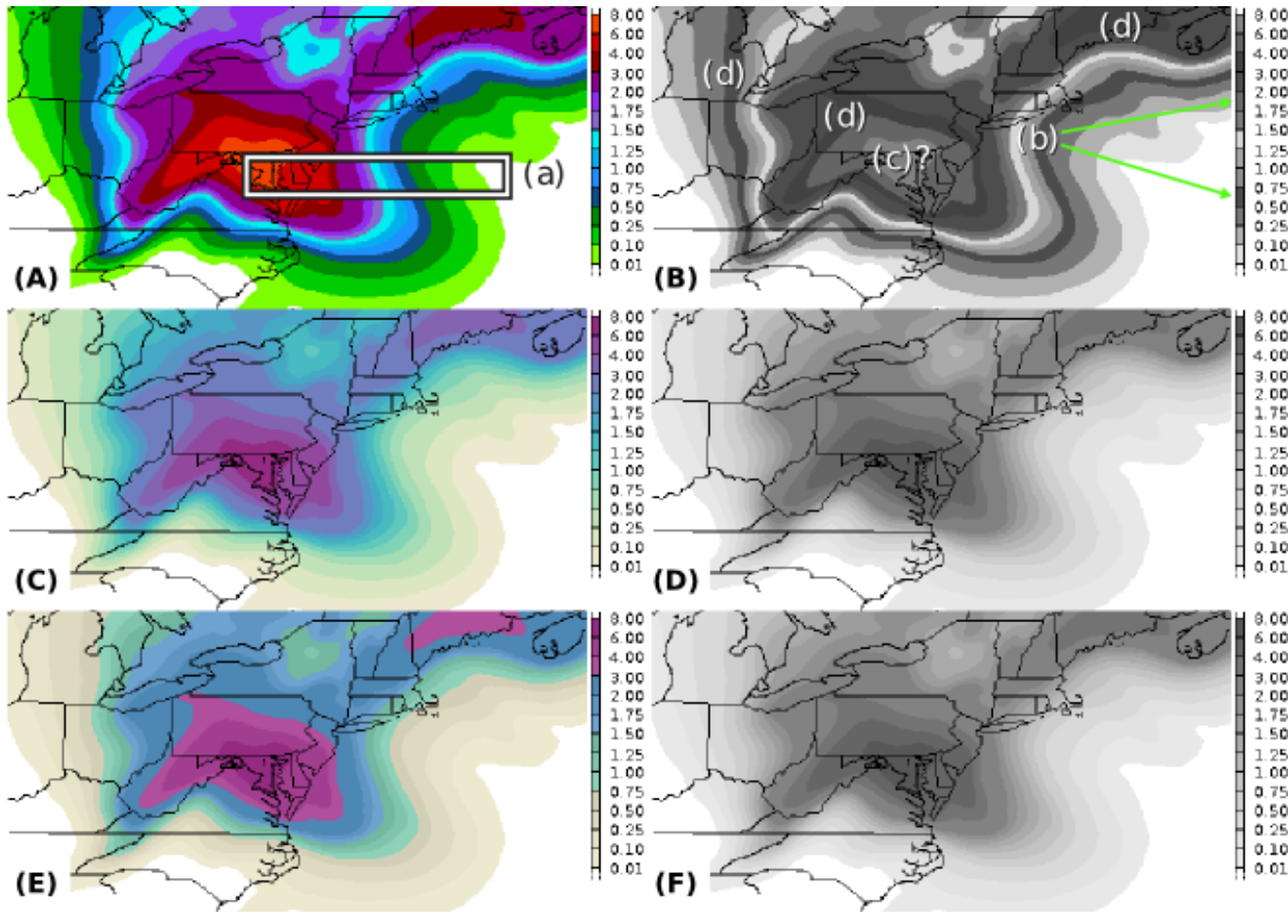
Tufte lie factor of 4.64![1]



[1] graphs courtesy the 20bits blog

Color

Tricky bit about color: the way we perceive color doesn't match the way (most) color systems quantify it.



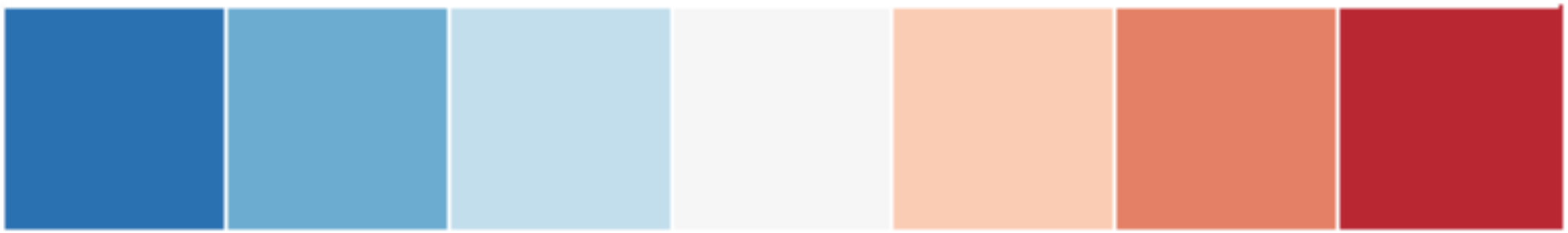
The easy solution:

Use prebuilt palettes like from Colorbrewer.org

Sequential
for data on a 0 → n scale. E.g. unemployment rates, per capita income, etc.

Diverging

for plotting difference with respect to an average or past datapoint. E.g. Change in poverty rate since 2000, etc.



Qualitative (also called ordinal or categorical) for comparing distinct things or categories. E.g. Cat owners vs dog owners, etc.

Seaborn makes using these and other predefined palettes easy

The more complicated solution:

Use a color scale like HCL, where equivalent chroma and lightness values correspond to colors with similar perceptual weight.

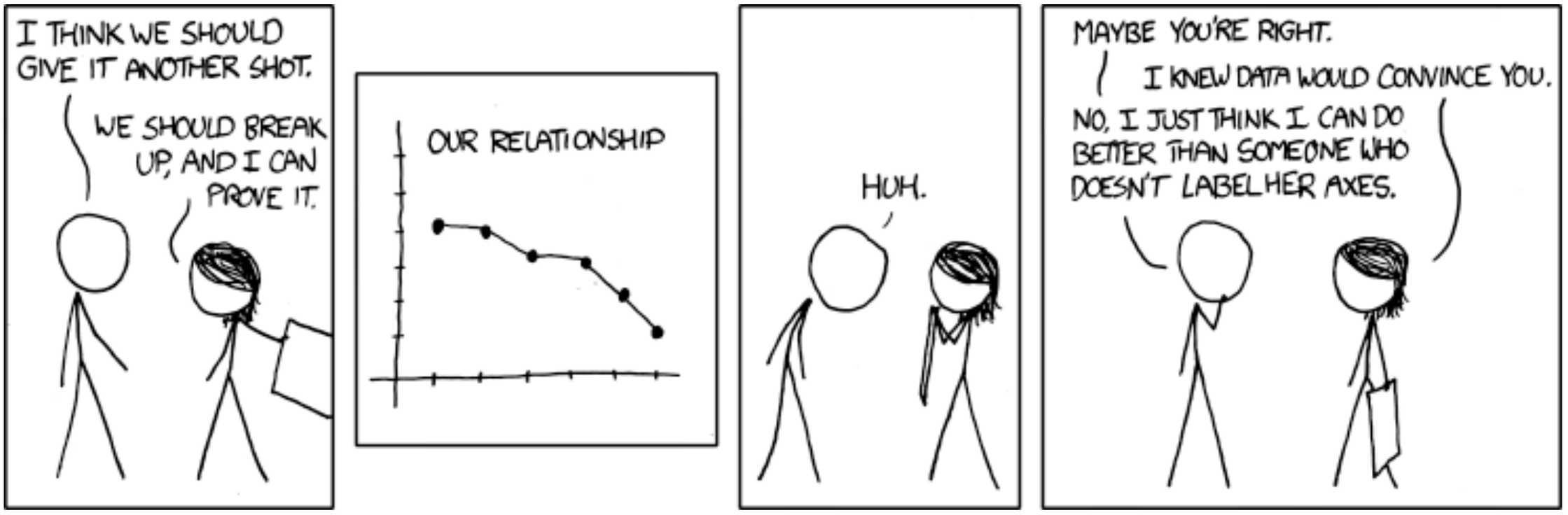
If you want to nerd out on this, here's a great link

Up your data-to-ink ratio

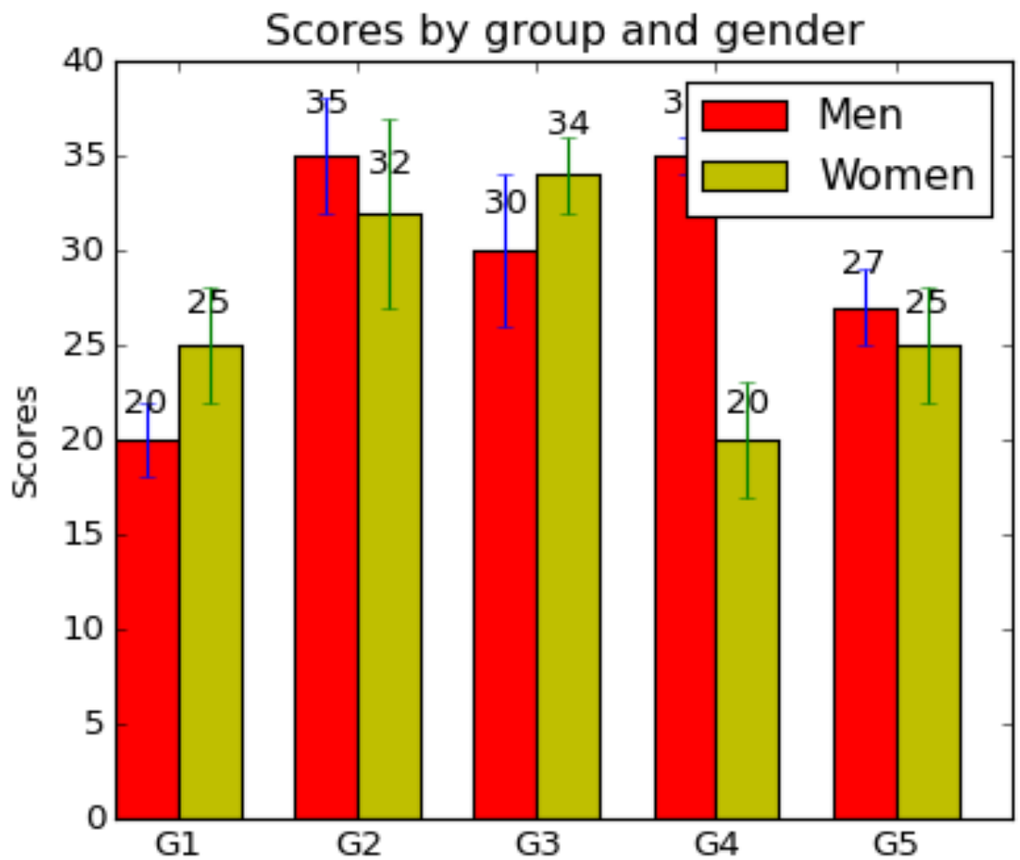
Ways to do this:

- Erase non-data ink where possible
- Erase redundant data ink where possible

Goldilocks labels



The opposite extreme



(actual example plot from matplotlib documentation)

Remove
to improve
(the **data-ink** ratio)

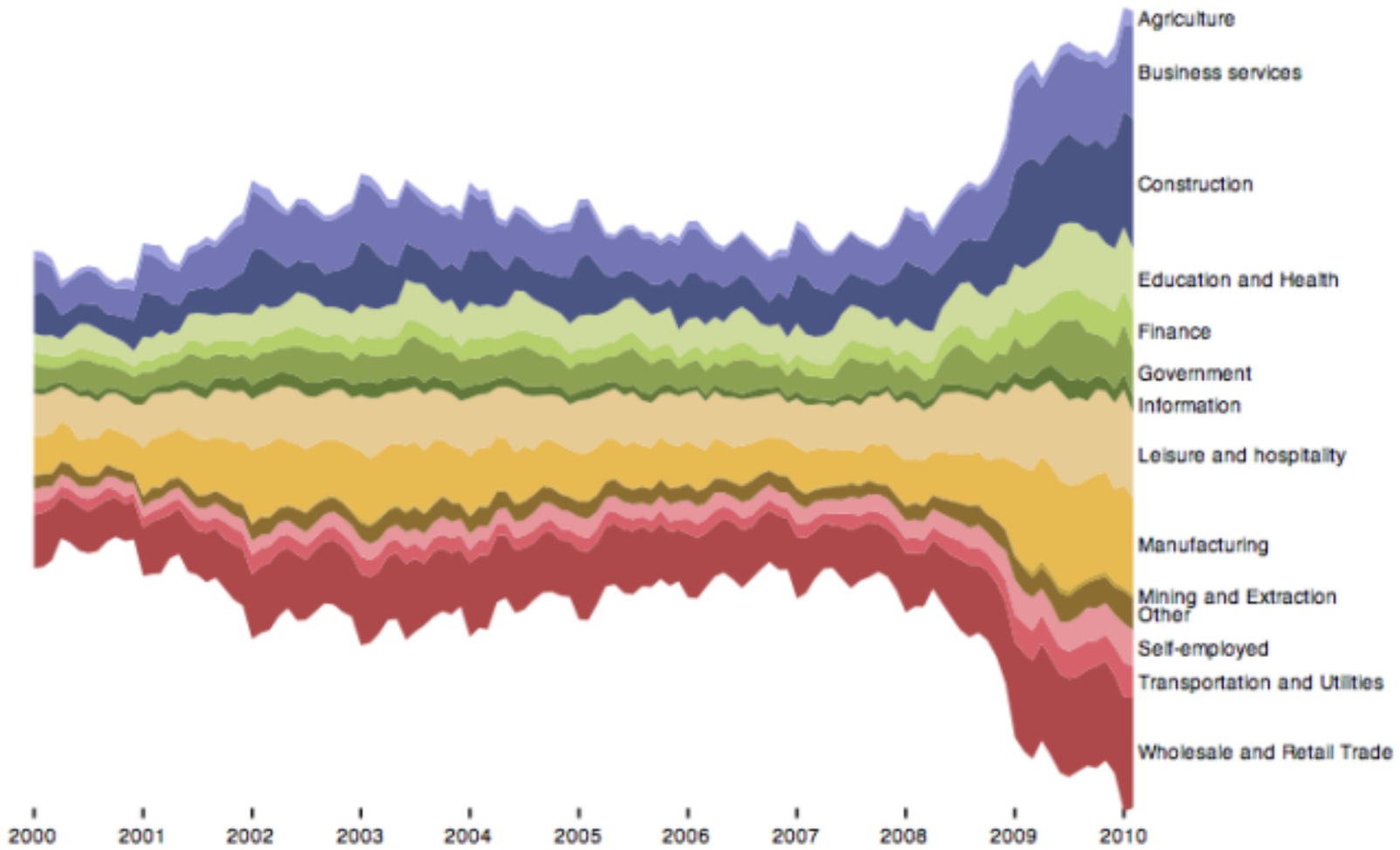
One more caveat...

There is no more reason to expect one graph to "tell all" than to expect one number to do the same.

– John Tukey (inventor of the box plot)

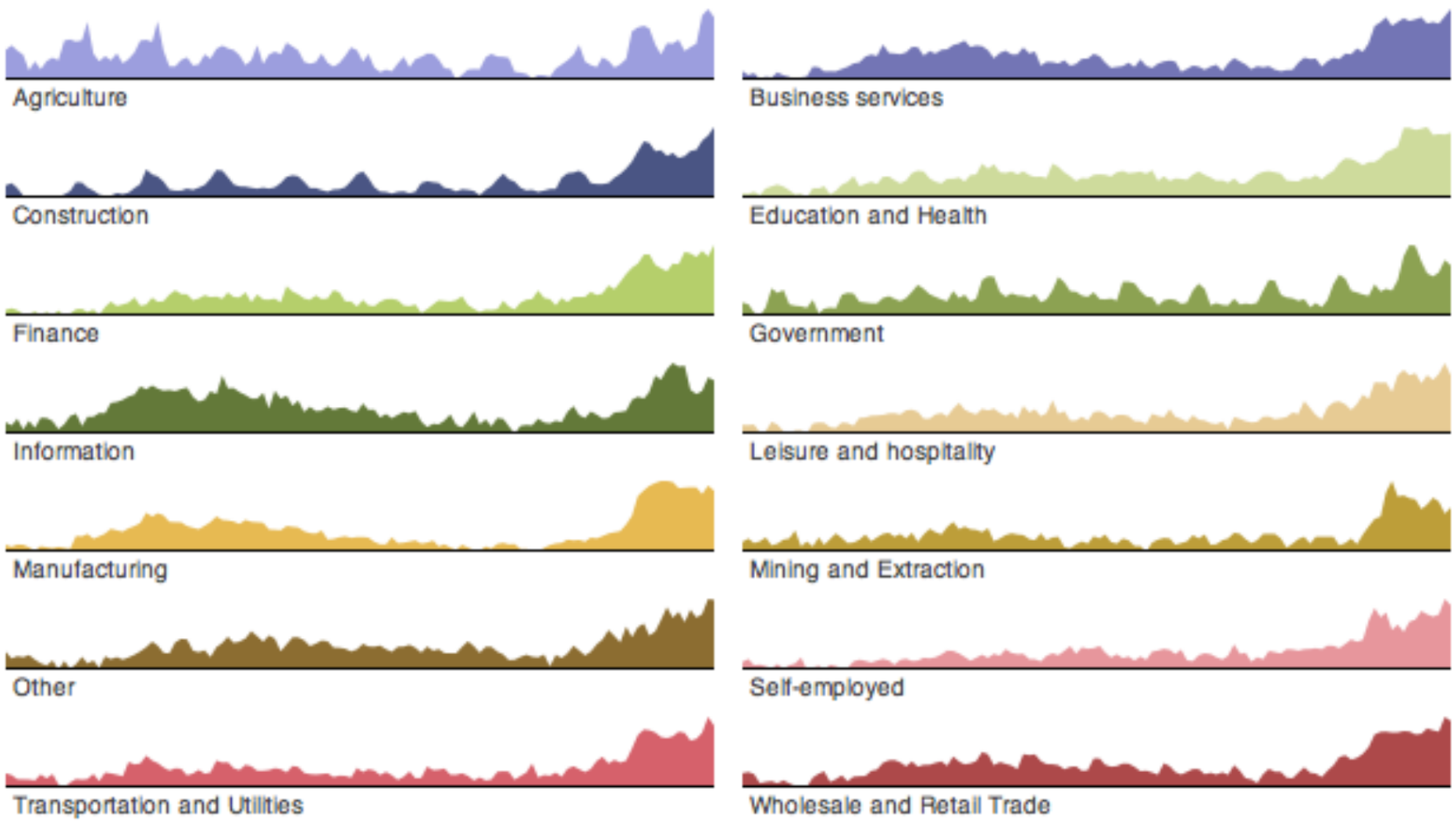
Number of unemployed workers in the United States over the last decade, subdivided by industry.

Streamgraph



Number of unemployed workers in the United States over the last decade, subdivided by industry.

Small multiples



A plot should tell a story.