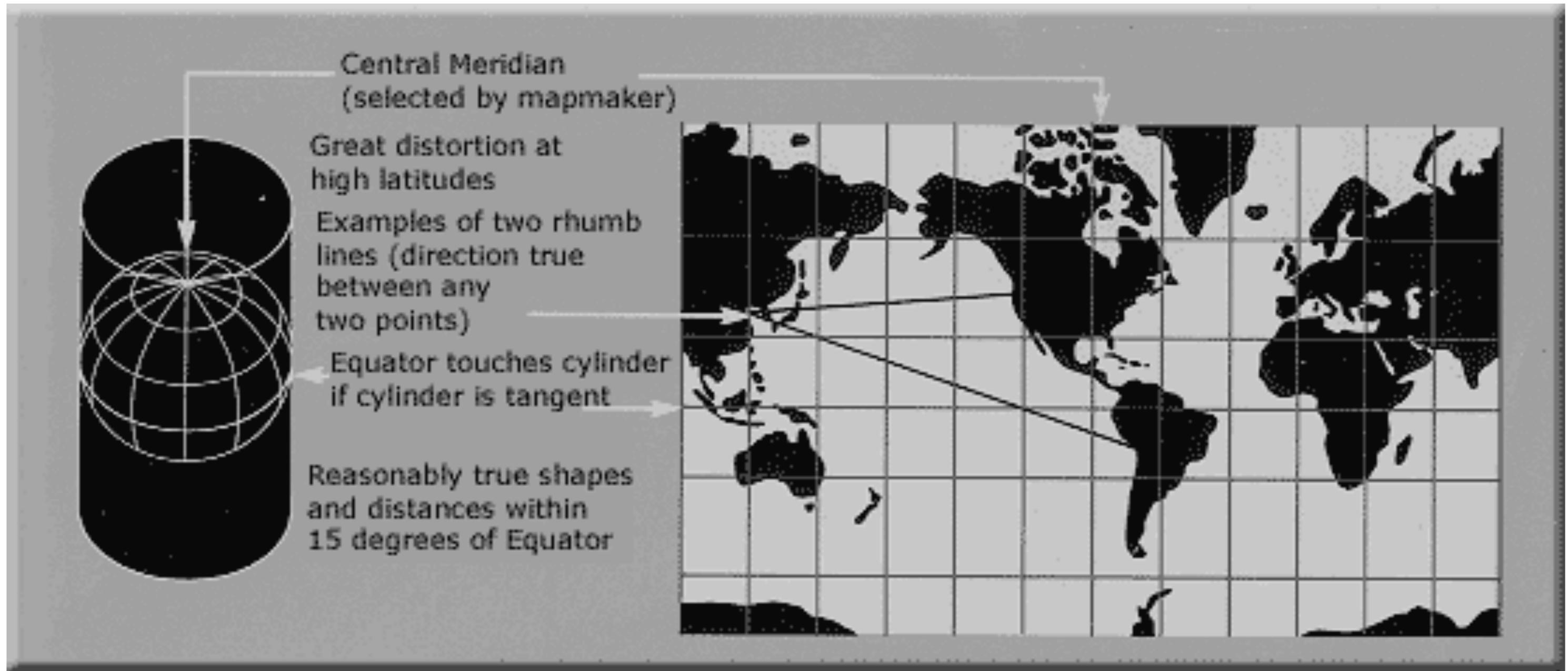


Data Visualization (CSCI 627/490)

Design

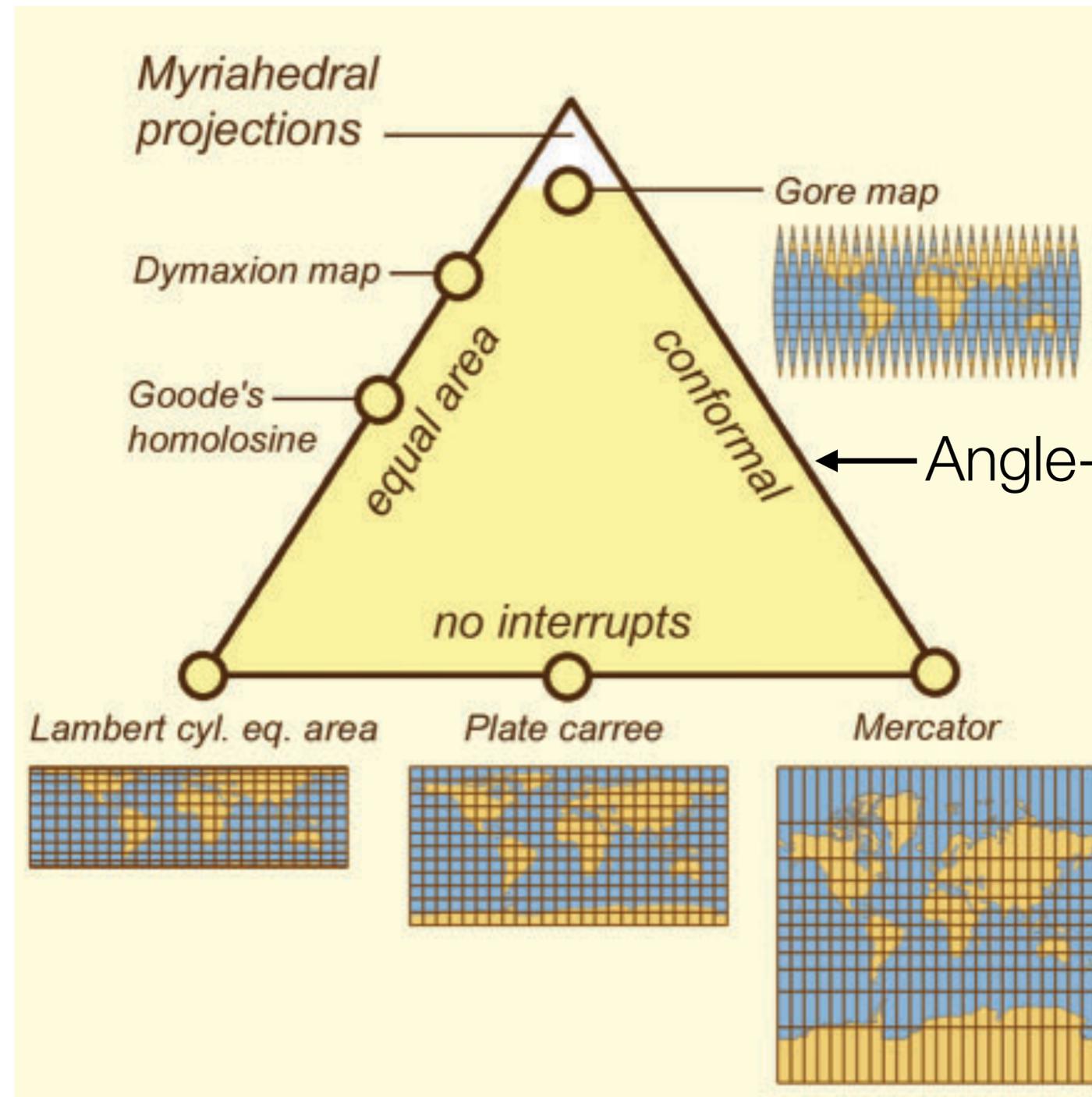
Dr. David Koop

Geographic Data: 3D to 2D: Projection



[USGS Map Projections]

Projection Classification



← Angle-preserving

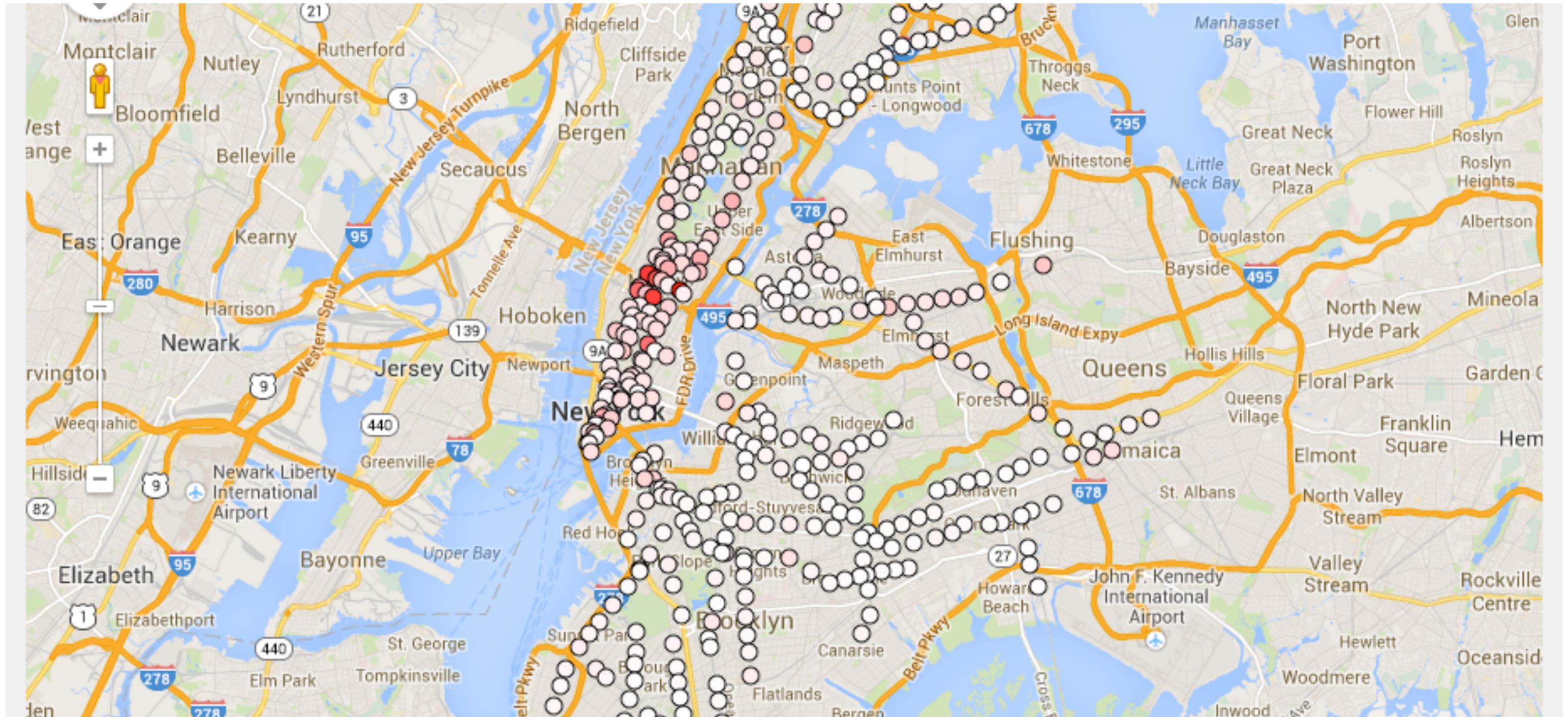
[J. van Wijk, 2008]

Search Tasks

| | Target known | Target unknown |
|------------------|---|--|
| Location known |  <i>Lookup</i> |  <i>Browse</i> |
| Location unknown |  <i>Locate</i> |  <i>Explore</i> |

[Munzner (ill. Maguire), 2014]

Adding Discrete Data to a Map: Quantitative Points



Adding Continuous Data to a Map: Isolines



[USGS via Wikipedia]

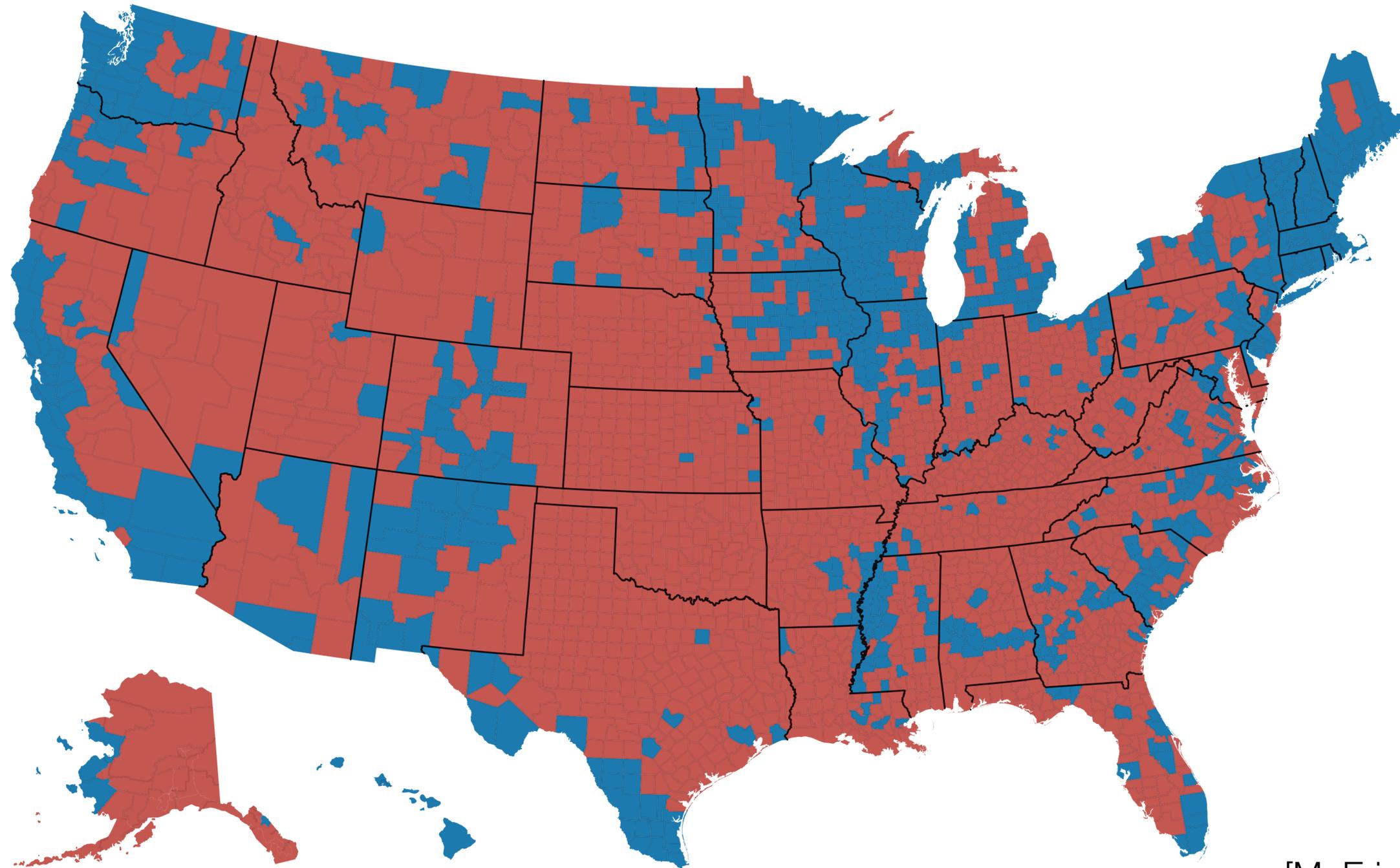
Assignment 4

- To be announced soon
- Colormaps, geospatial vis, networks

Project

- Next Step: Design
 - Given dataset (what) and tasks (why), work on the how
 - Don't do this the other way around: do not start with "I want to make a streamgraph" and then decide what tasks could work with that
 - This includes interactive design

Choropleth (Two Hues)



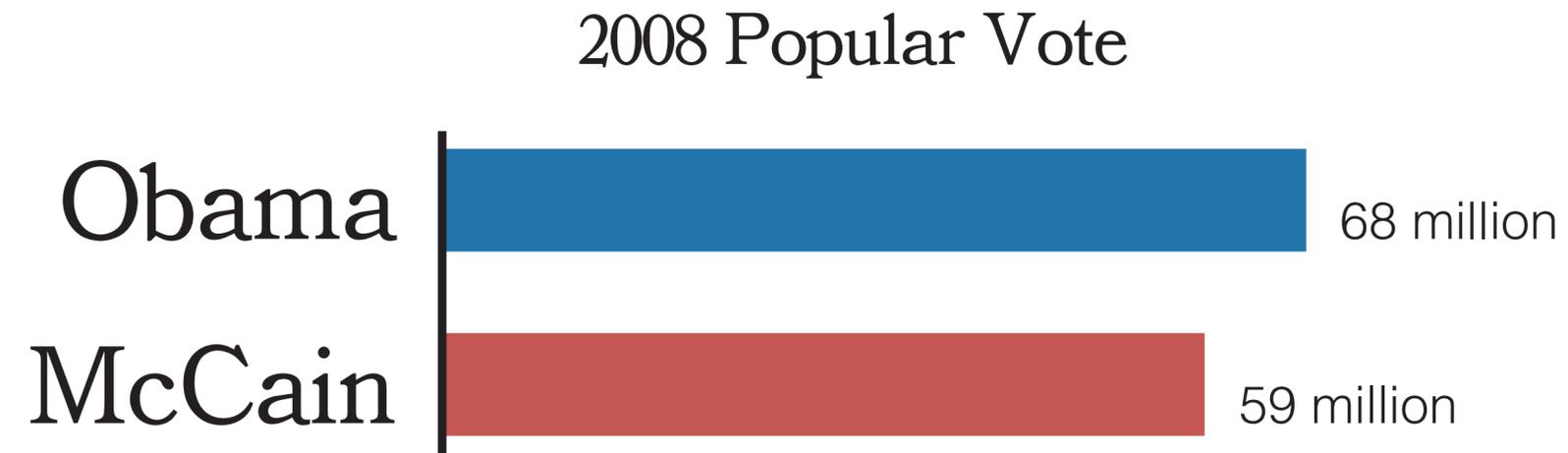
[M. Ericson, New York Times]

Choropleth Map

- Data: geographic geometry data & one quantitative attribute per region
- Tasks: trends, patterns, comparisons
- How: area marks from given geometry, color hue/saturation/luminance
- Scalability: thousands of regions

- Design choices:
 - Colormap
 - Region boundaries (level of summarization)

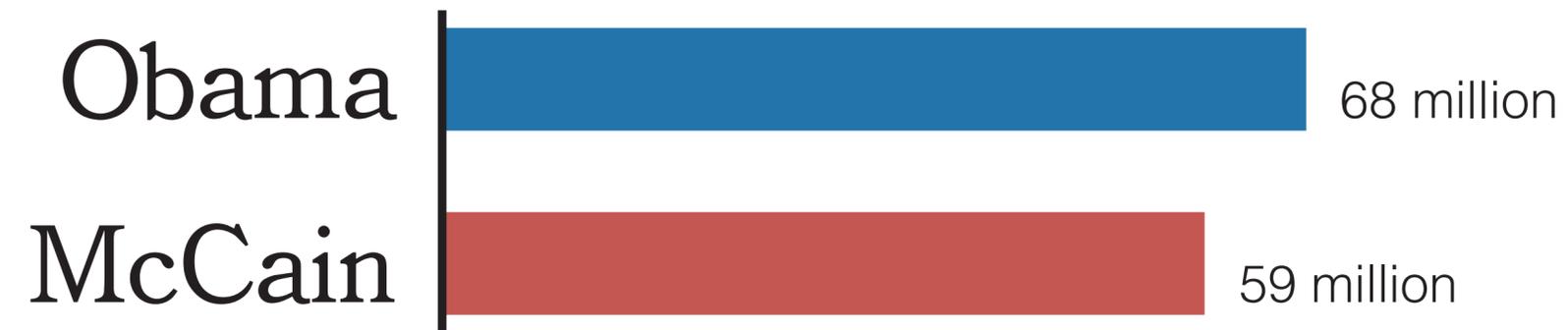
Problem?



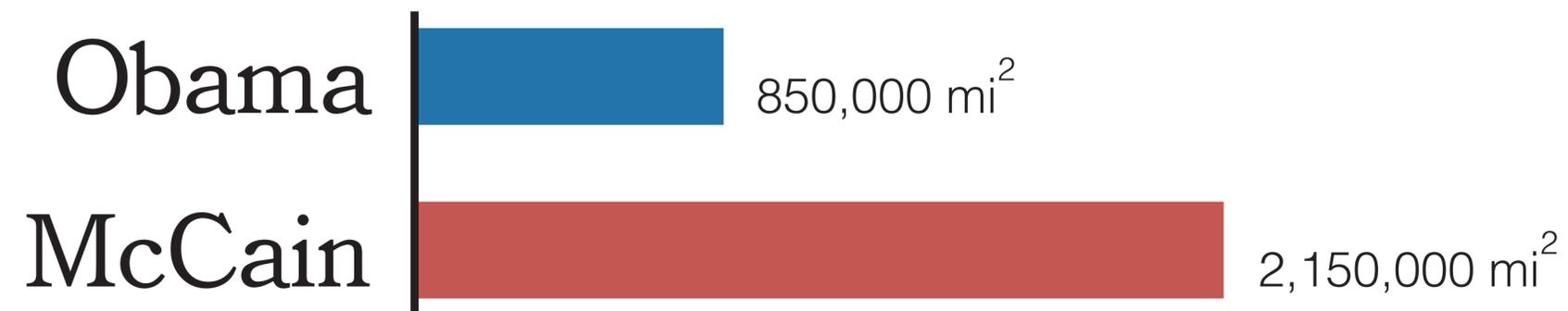
[M. Ericson, New York Times]

Problem?

2008 Popular Vote



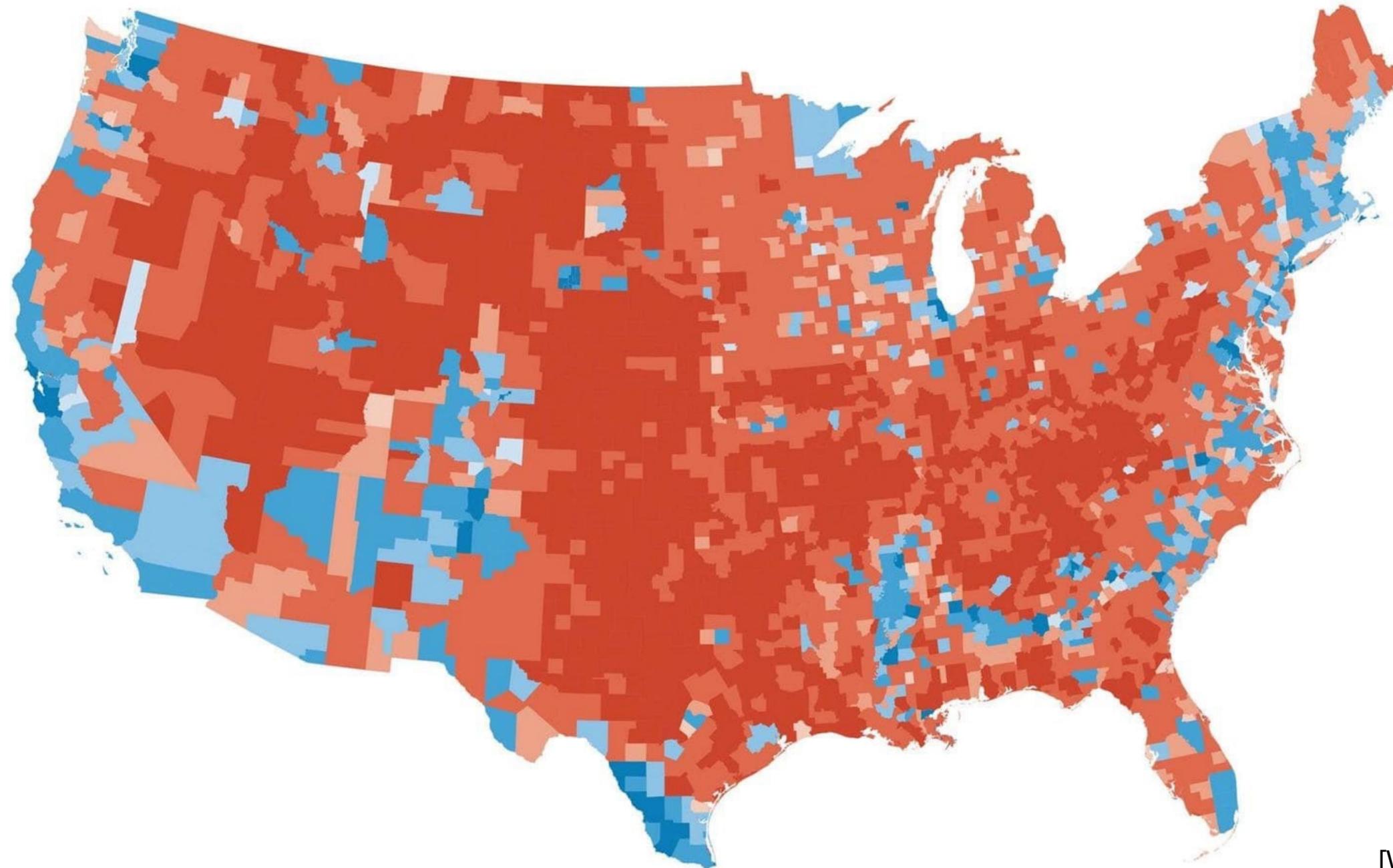
Amount of red and blue shown on map



[M. Ericson, New York Times]

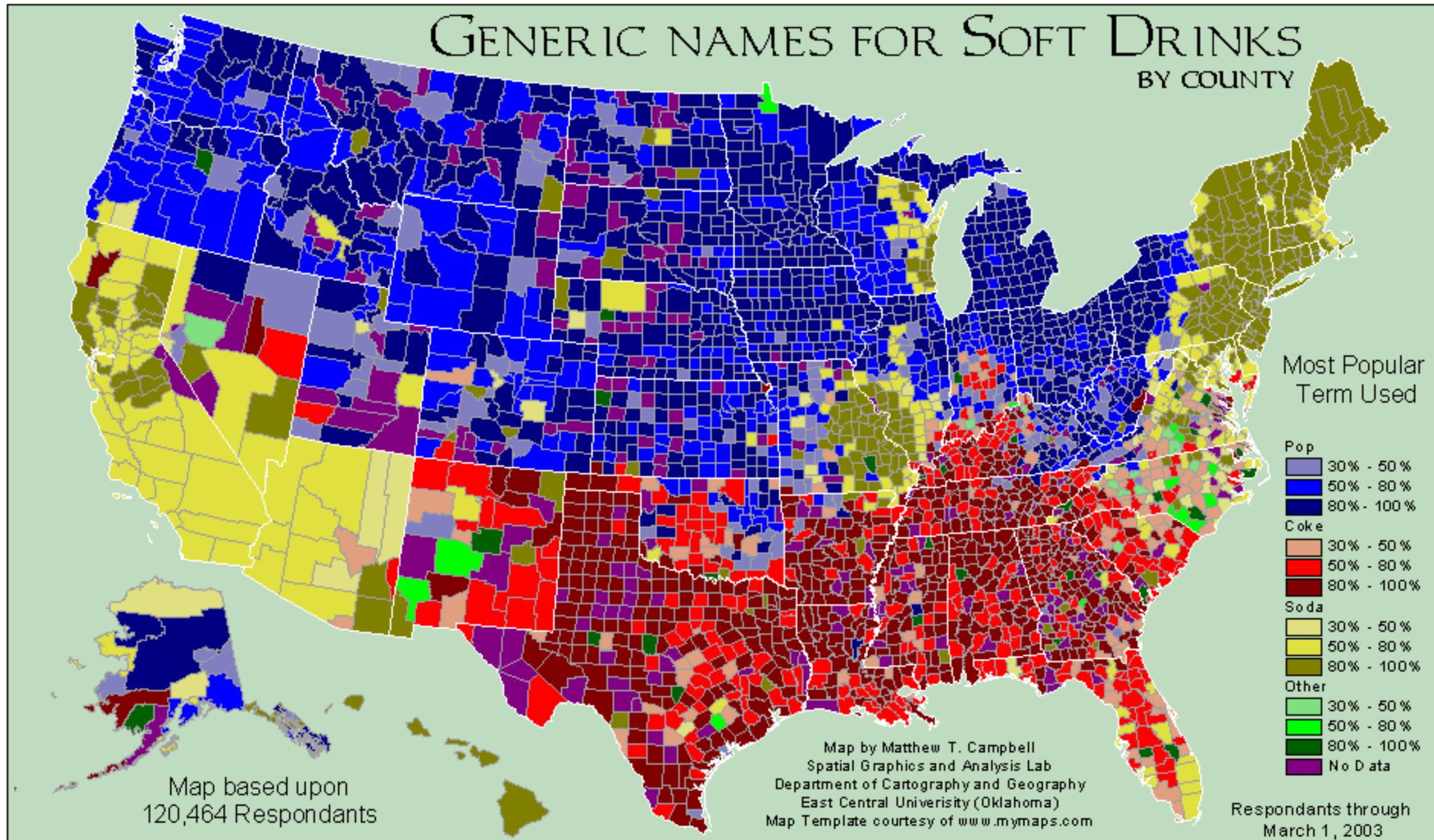
Adding Saturation

Clinton +50-100 +15-50 +2.1-15 +0-2.1 Trump +0-2.1 +2.1-15 +15-50 +50-100



[Washington Post, 2018]

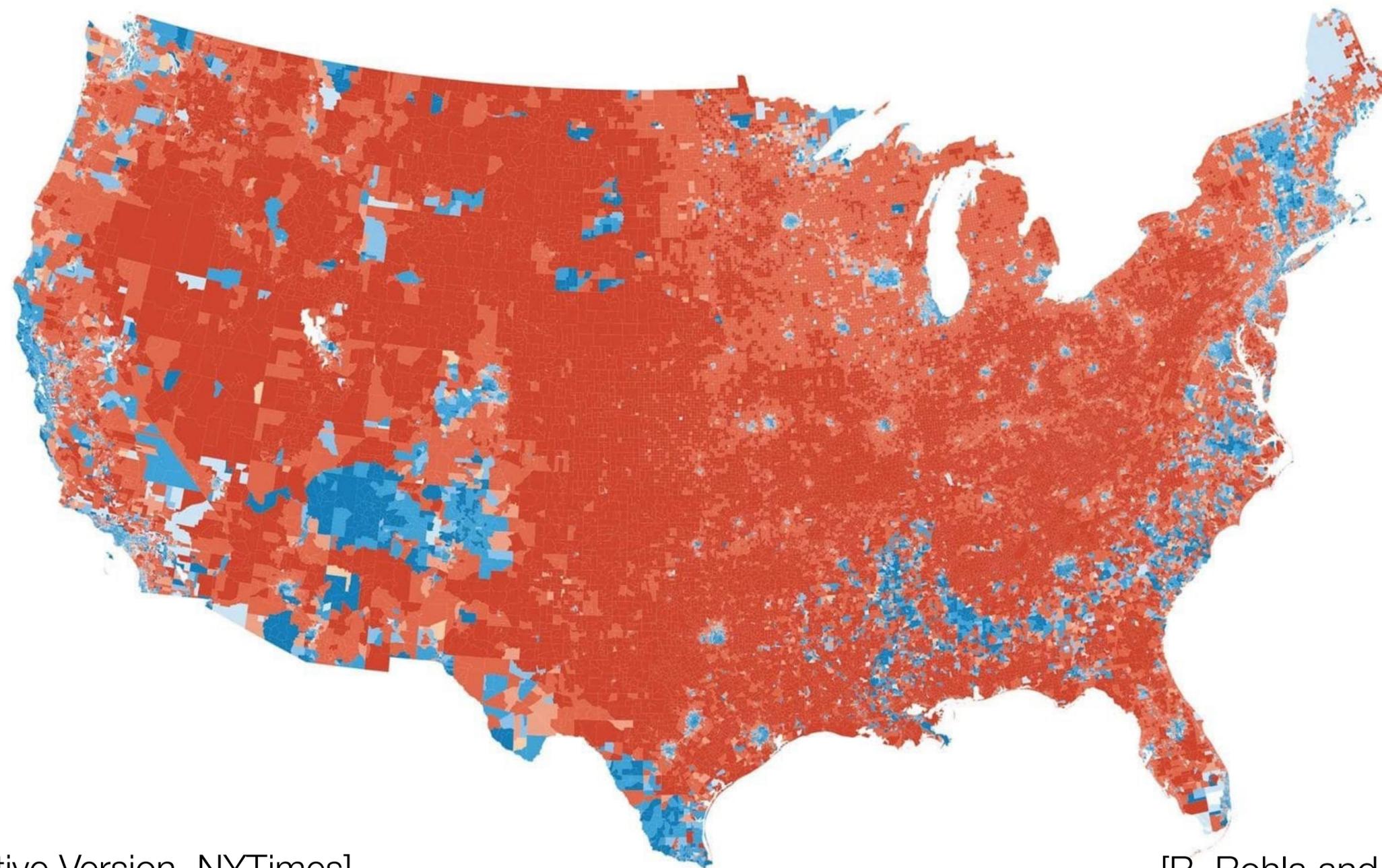
Area Marks and Color Hue & Saturation



[popvssoda.com]

Aggregation: 2016 Election by Precinct

Clinton +50-100 +15-50 +2.1-15 +0-2.1 Trump +0-2.1 +2.1-15 +15-50 +50-100

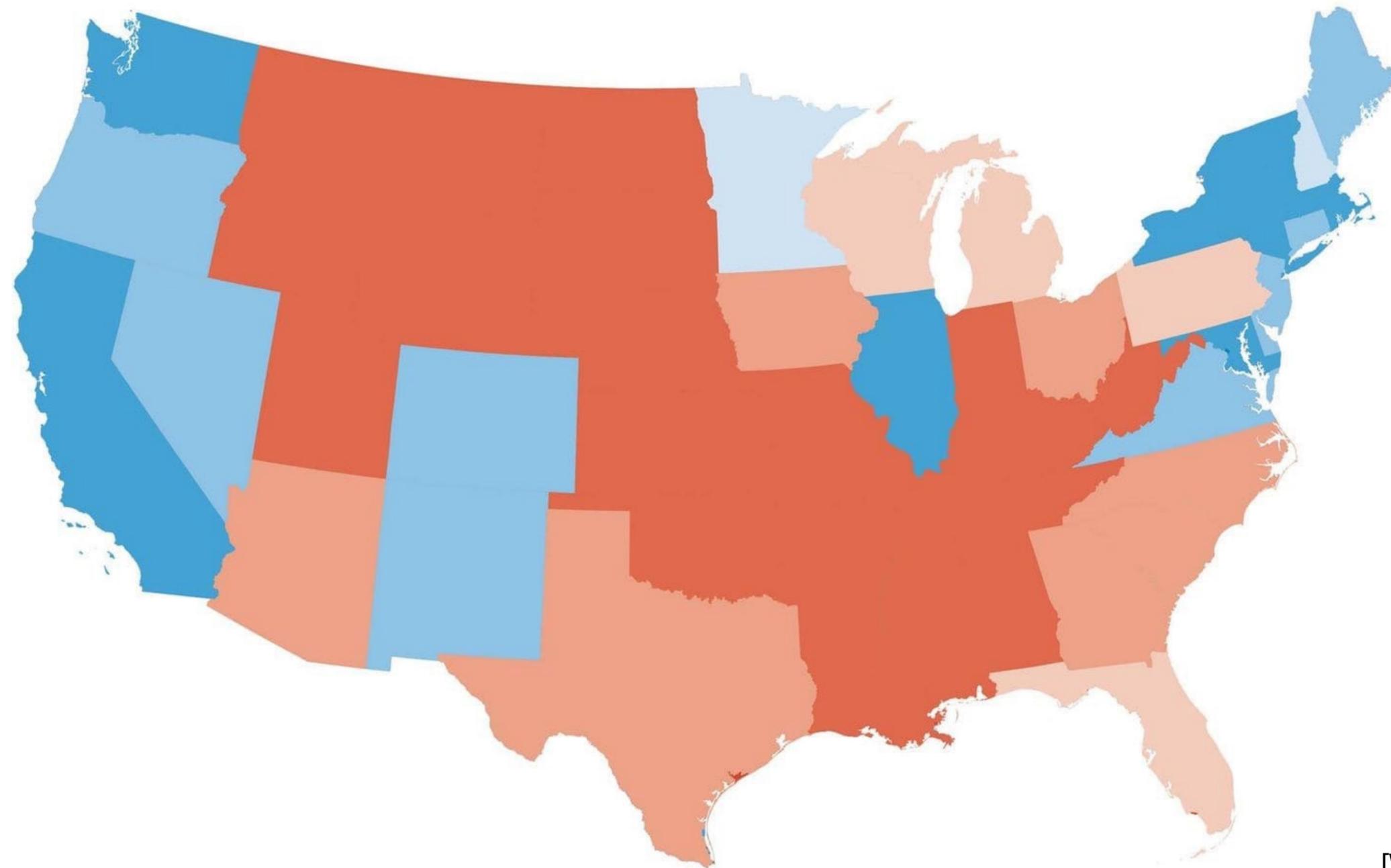


[[Interactive Version](#), NYTimes]

[R. Rohla and [Washington Post](#), 2018]

Aggregation: 2016 Election by State

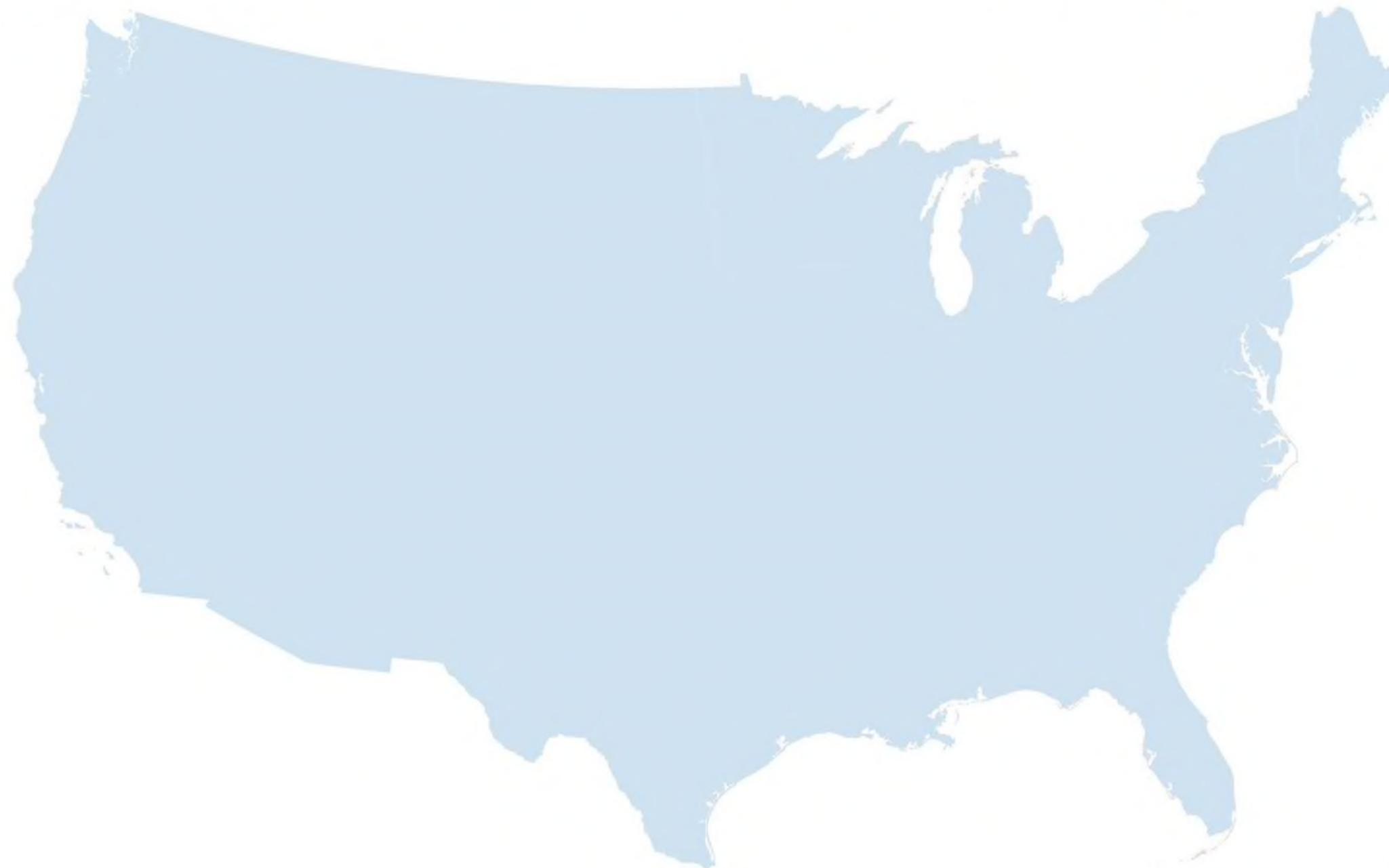
Clinton +50-100 +15-50 +2.1-15 +0-2.1 Trump +0-2.1 +2.1-15 +15-50 +50-100



[Washington Post, 2018]

Aggregation: 2016 Election by Country

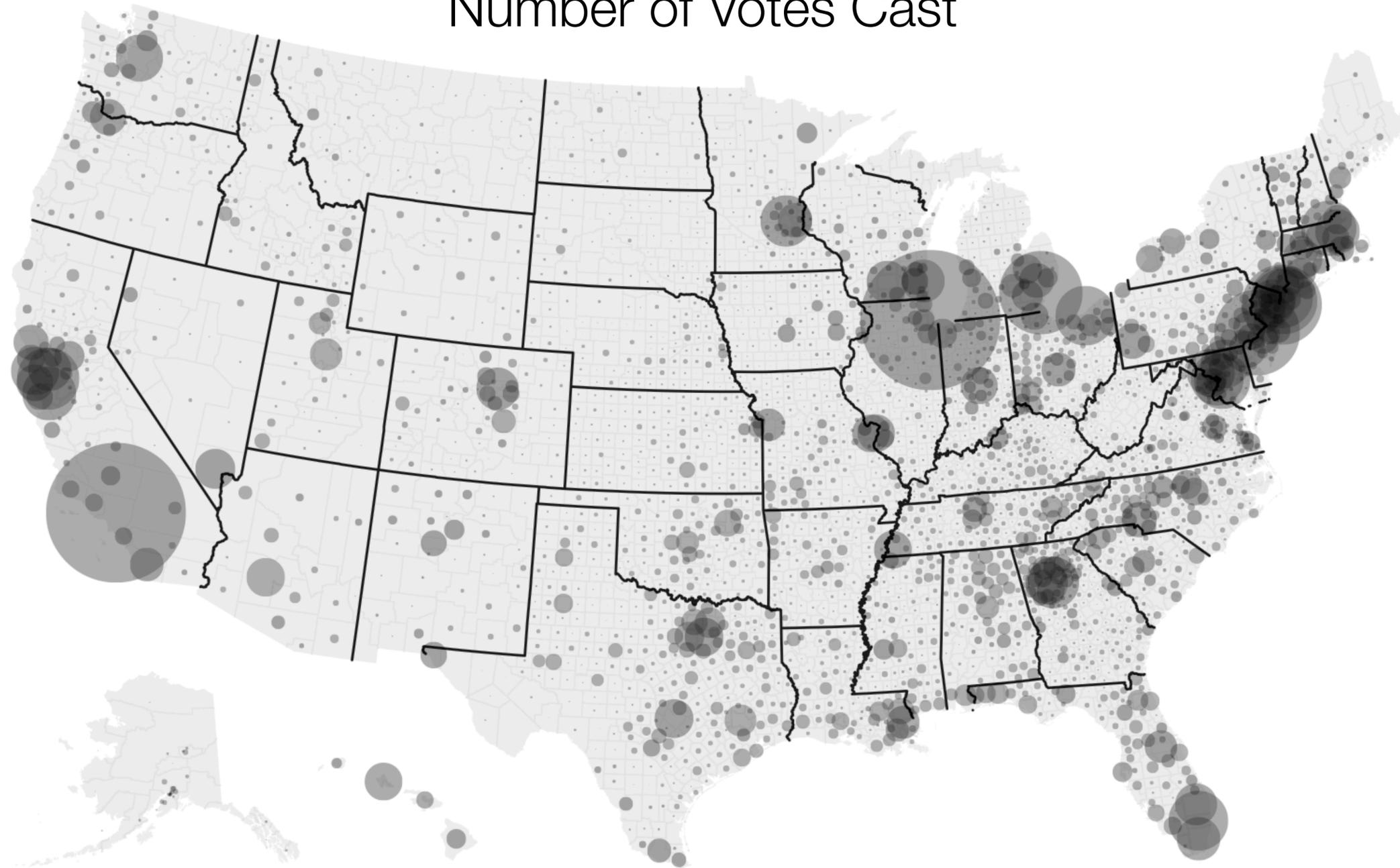
Clinton +50-100 +15-50 +2.1-15 +0-2.1 Trump +0-2.1 +2.1-15 +15-50 +50-100



[Washington Post, 2018]

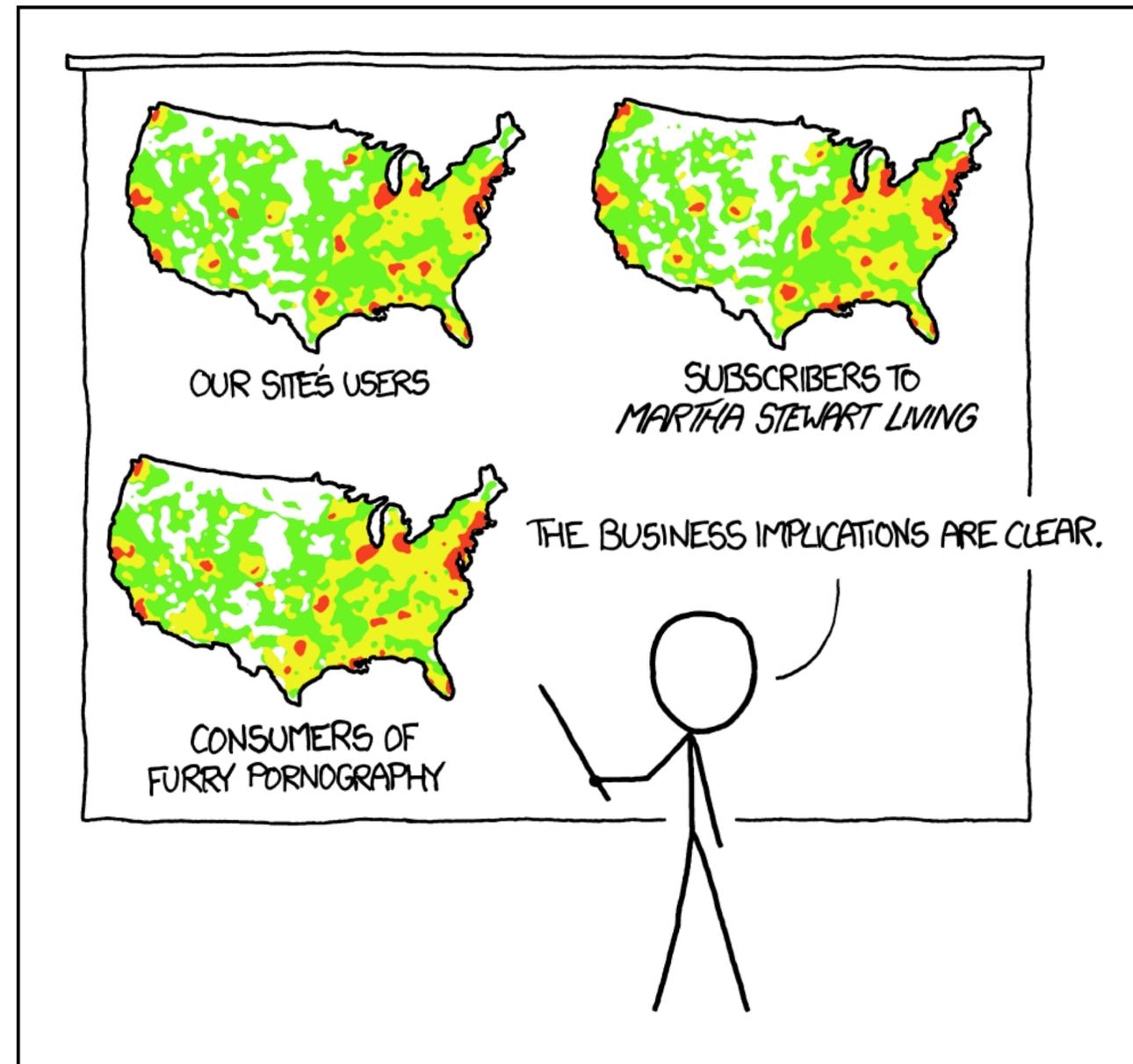
Maps: What trends do you see?

Number of Votes Cast



[Desaturated by D. Koop, M. Ericson, New York Times]

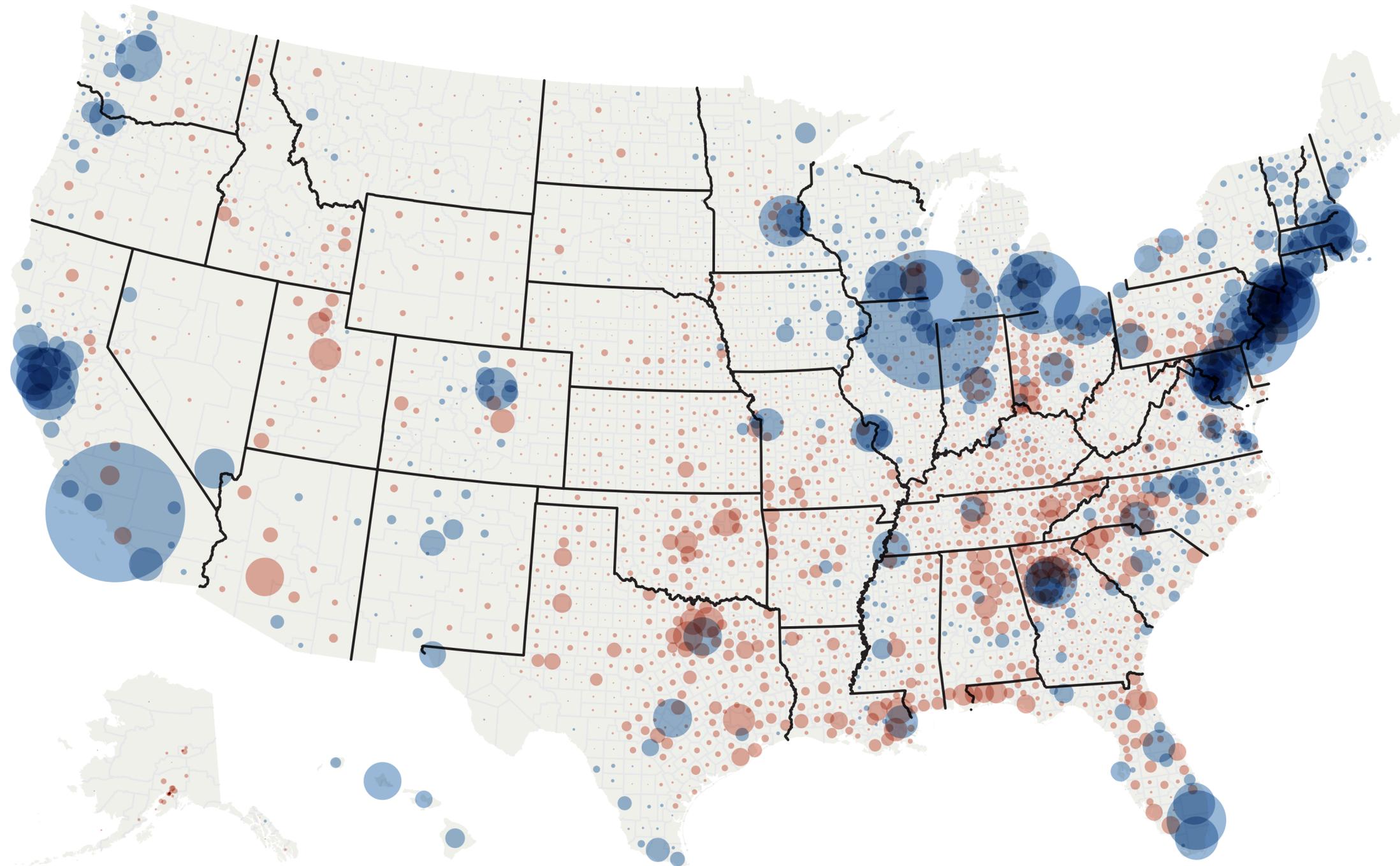
Don't Just Create Population Maps!



PET PEEVE #208:
GEOGRAPHIC PROFILE MAPS WHICH ARE
BASICALLY JUST POPULATION MAPS

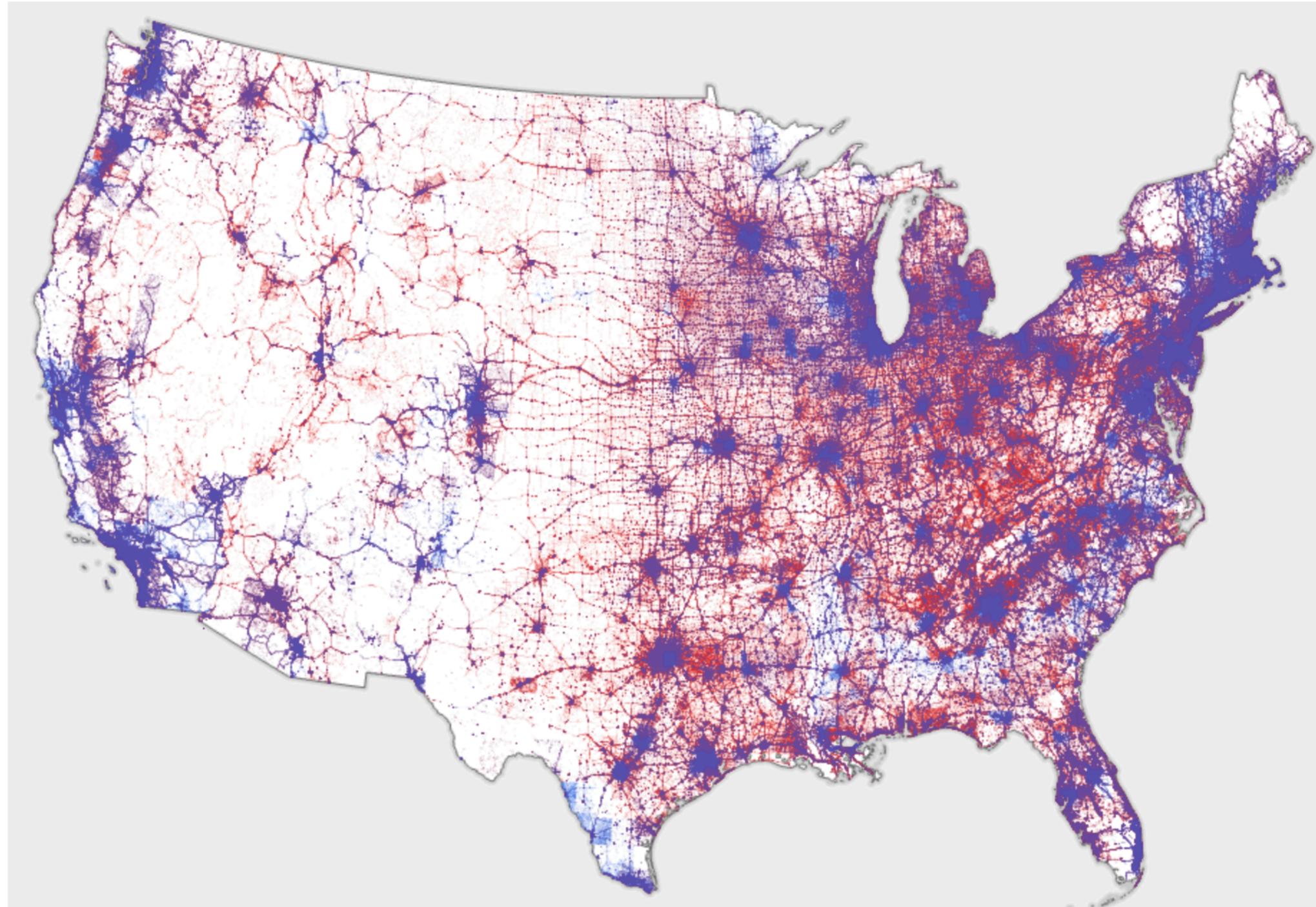
[xkcd]

Size Encoding



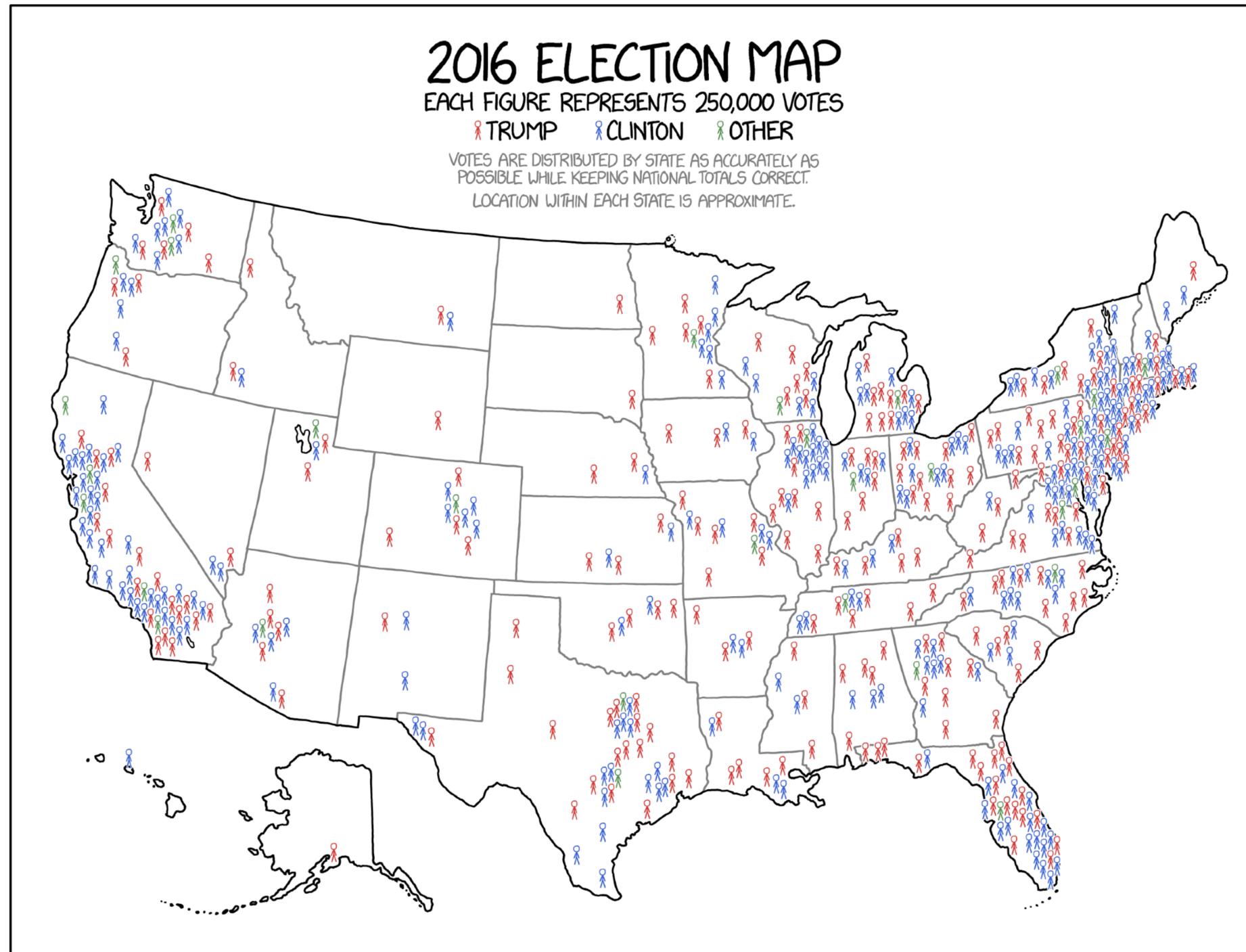
[M. Ericson, New York Times]

Dasymetric Dot Density



[K. Field]

Glyphs: xkcd's Map



[xkcd]

Cartograms

US Presidential Election 2016

Results mapped at county level showing the candidate with the largest vote share in each area

Overall result:

Trump

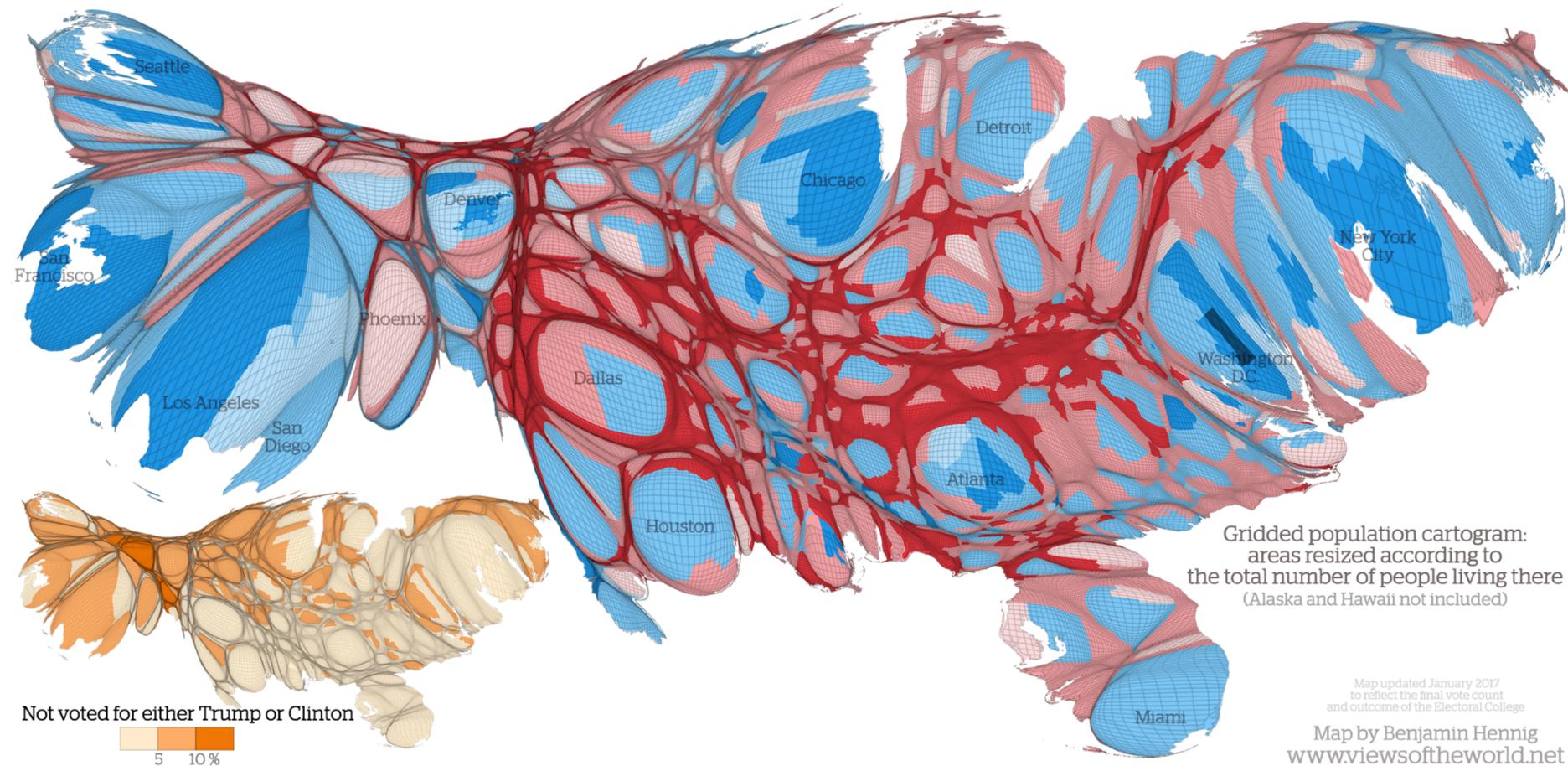
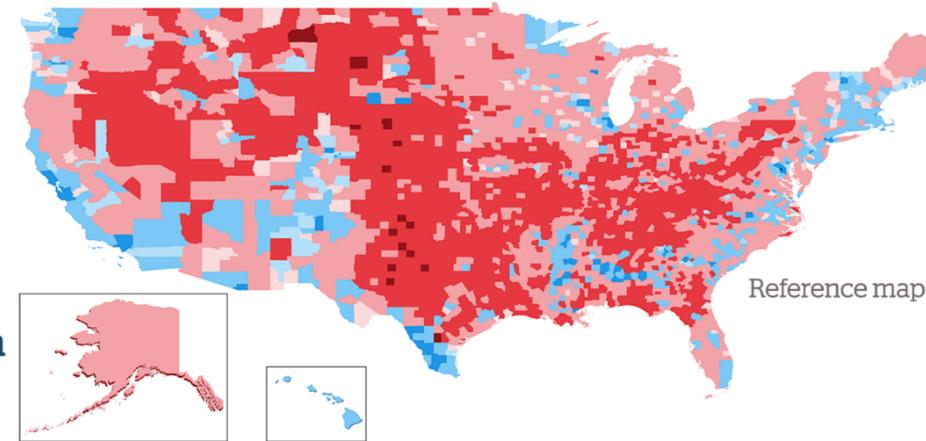
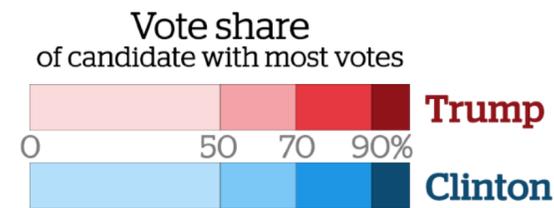
62,979,636 votes (46.1%)
306 electoral votes
(received 304 in the Electoral College)

Clinton

65,844,610 votes (48.2%)
232 electoral votes
(received 227 in the Electoral College)

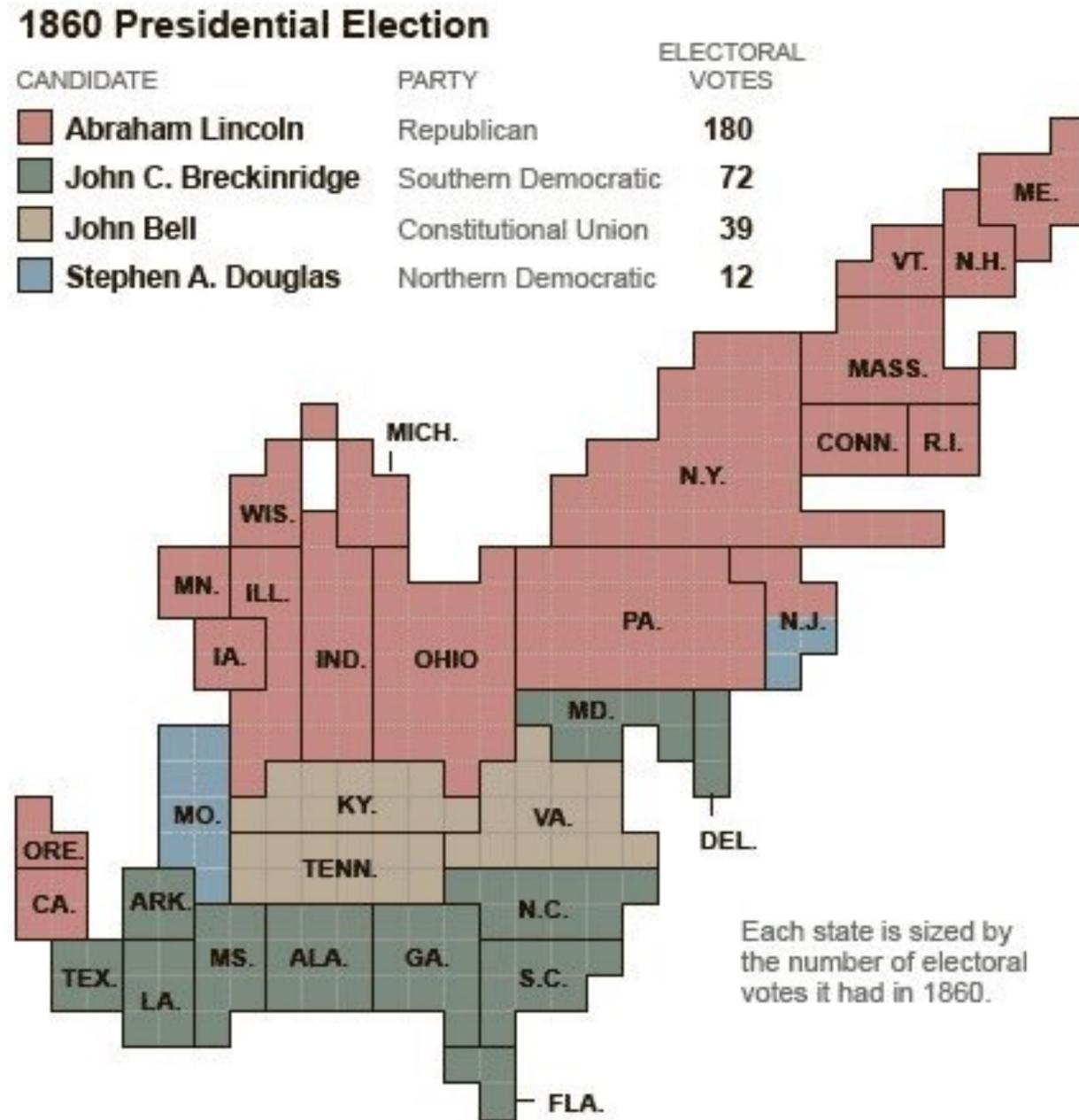
Other candidates

7,804,213 votes (5.7%)



[B. Hennig]

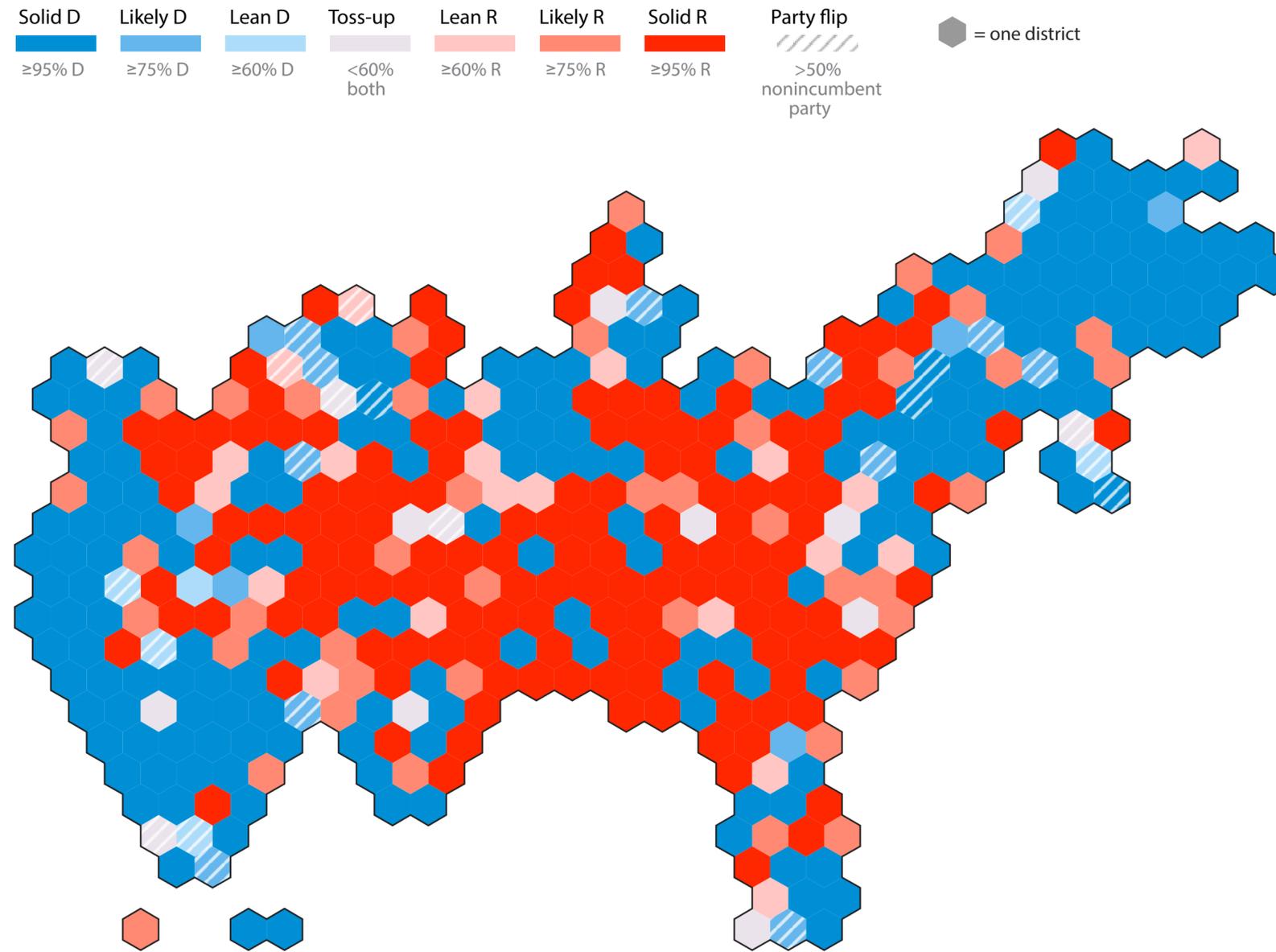
Cartograms



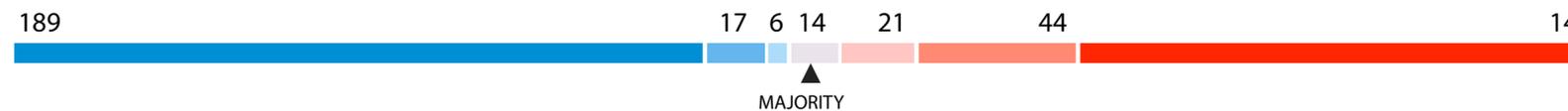
- Data: geographic geometry data & **two** quantitative attributes (one part-of-whole)
- Derived data: new geometry derived from the part-of-whole attribute
- Tasks: trends, comparisons, part-of-whole
- How: area marks from derived geometry, color hue/saturation/luminance
- Scalability: thousands of regions
- Design choices:
 - Colormap
 - Geometric deformation

[New York Times]

Hexagonal Cartogram

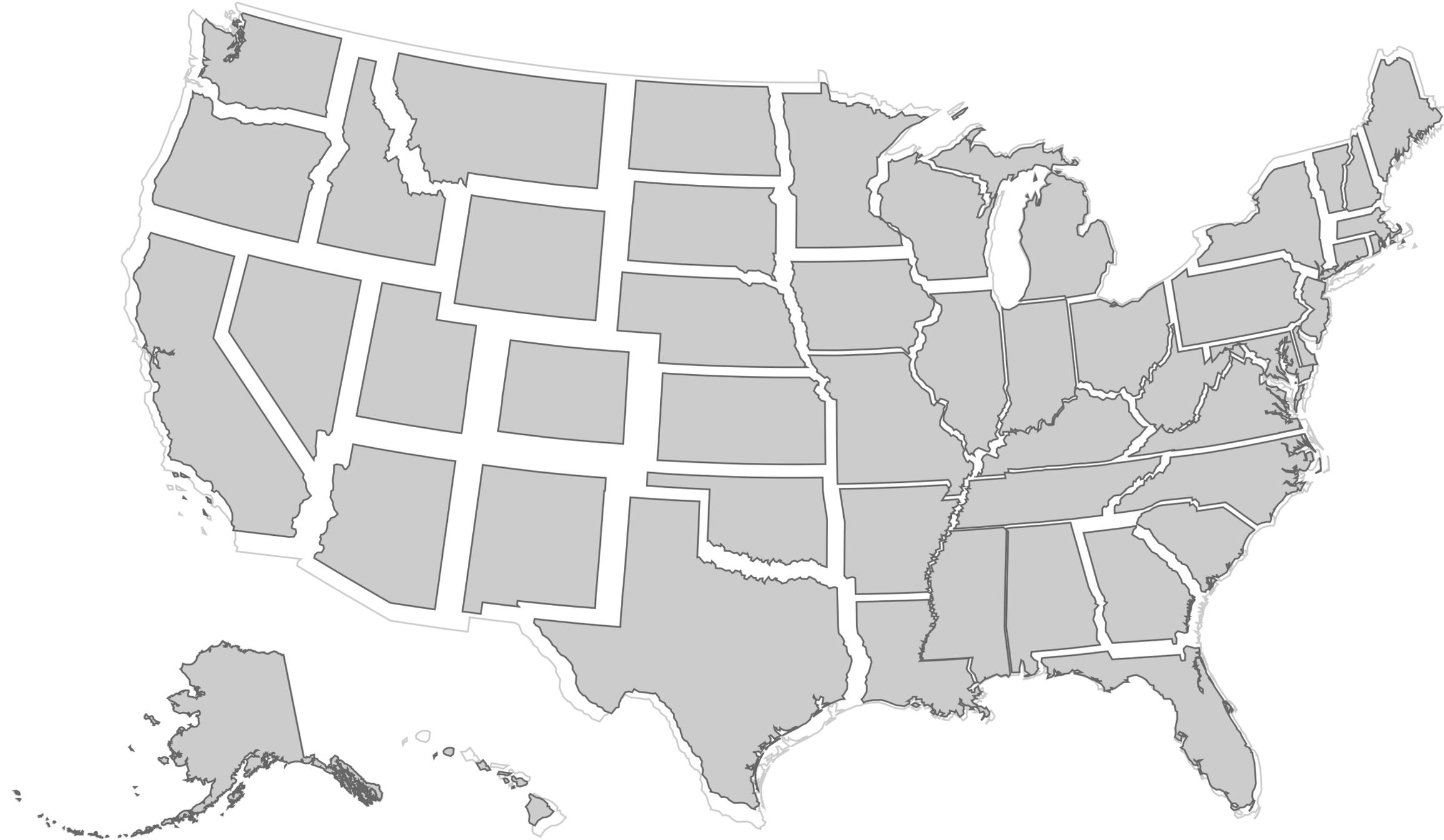


District totals by category



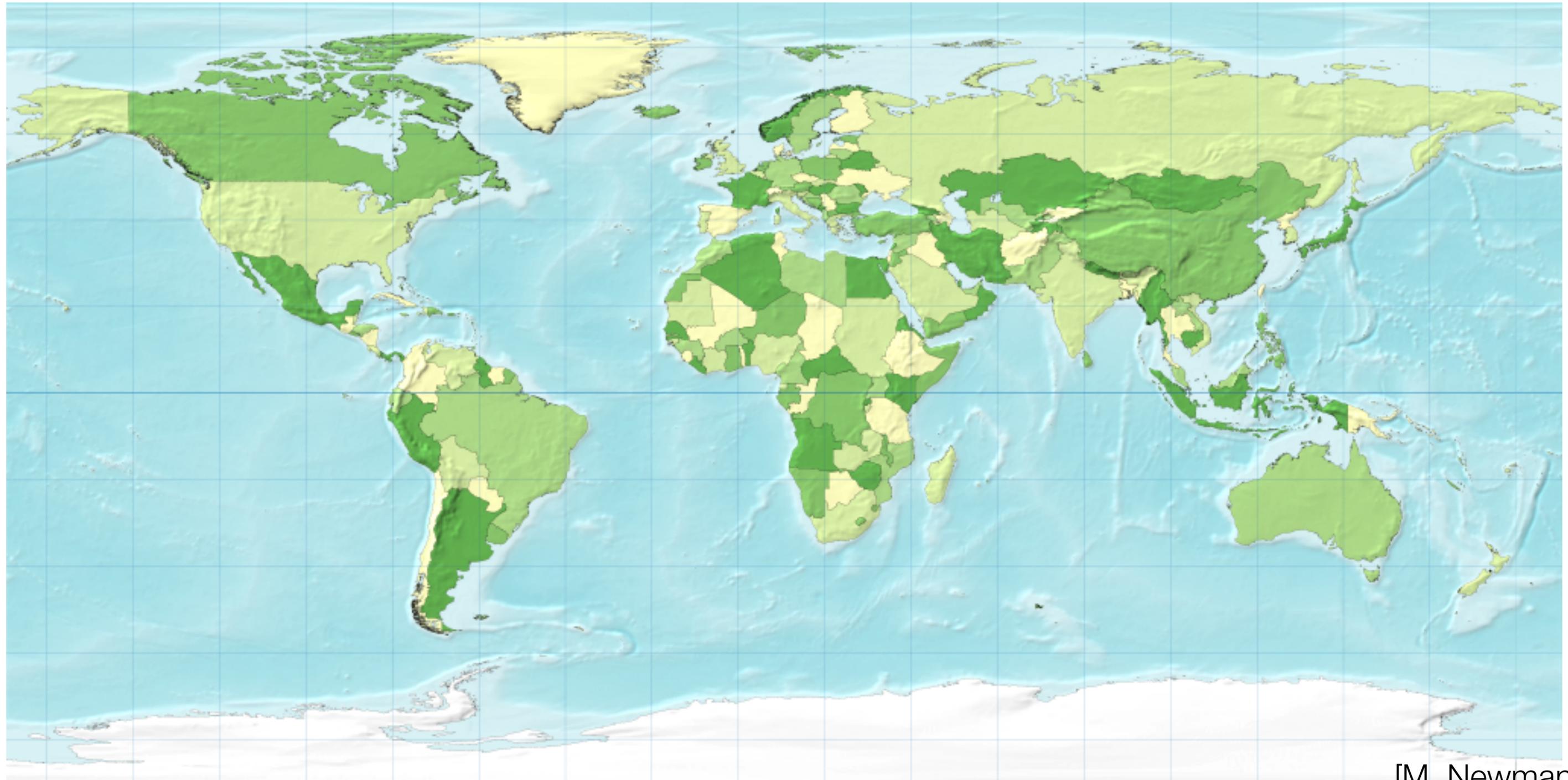
[FiveThirtyEight, 2018]

Non-Contiguous Cartogram



[M. Bostock, 2012]

World Cartograms



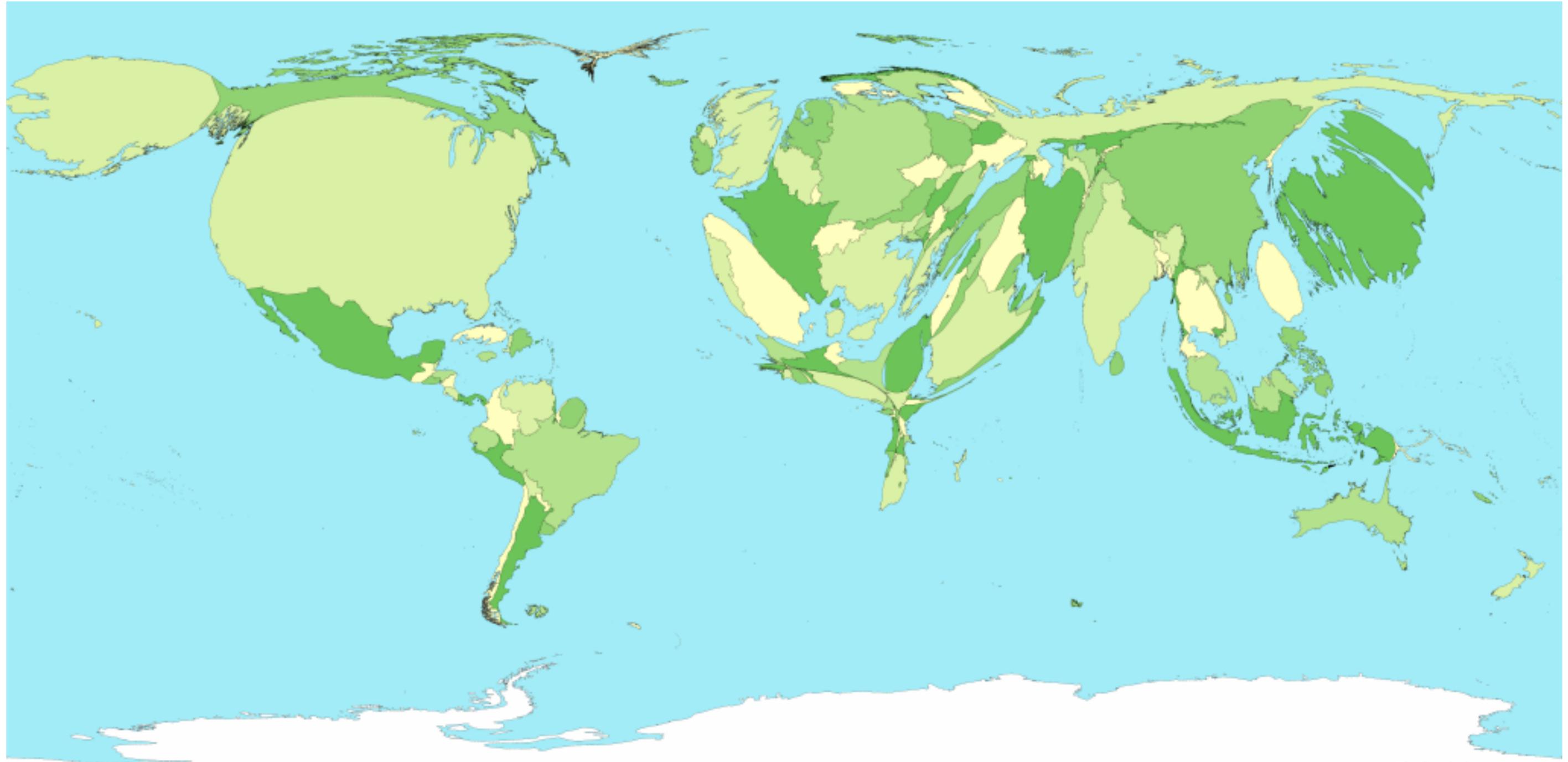
[M. Newman, 2009]

World Population



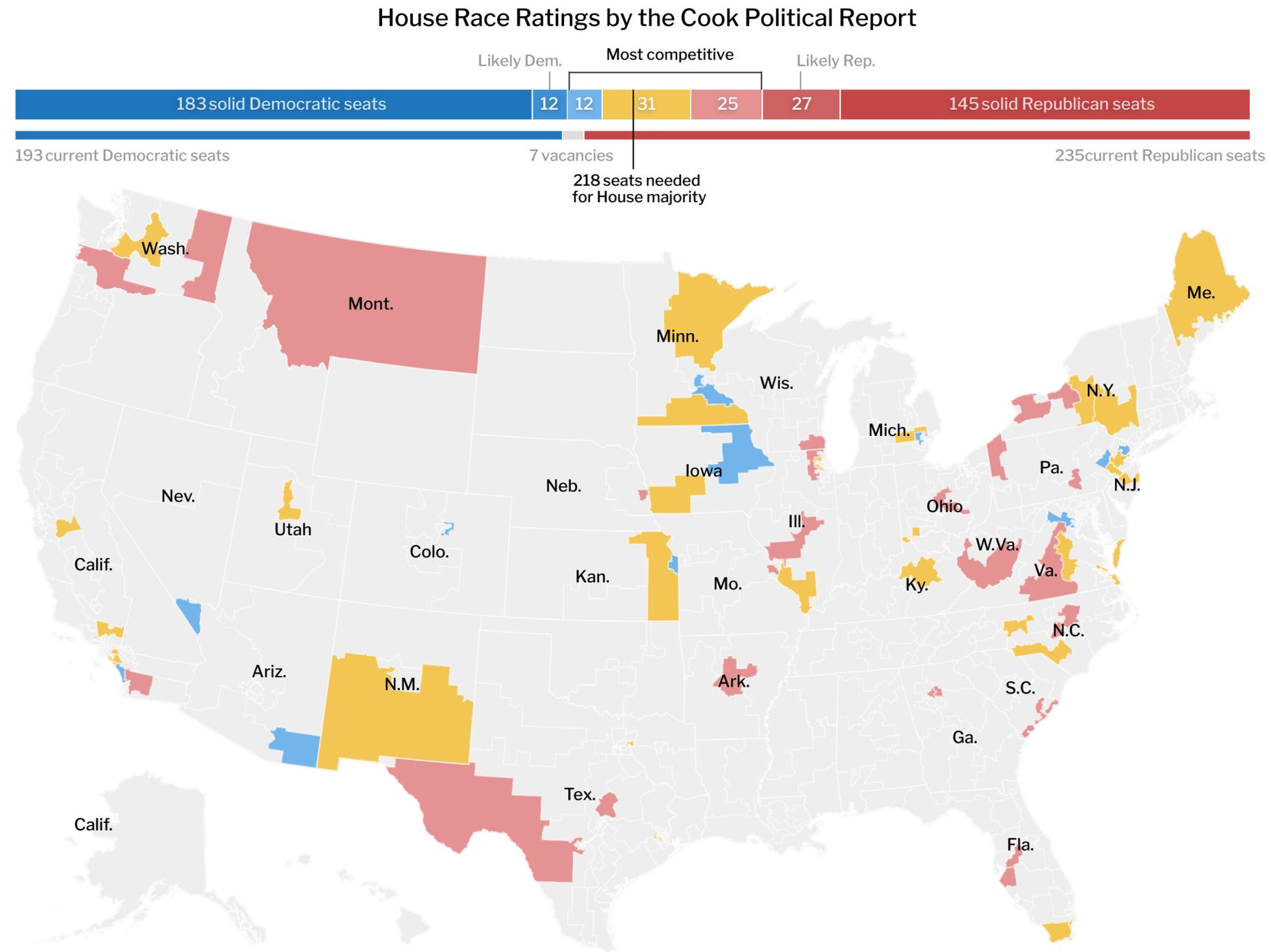
[M. Newman, 2009]

World Energy Consumption



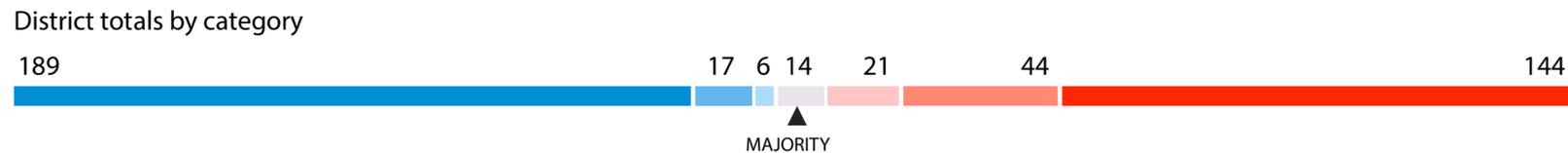
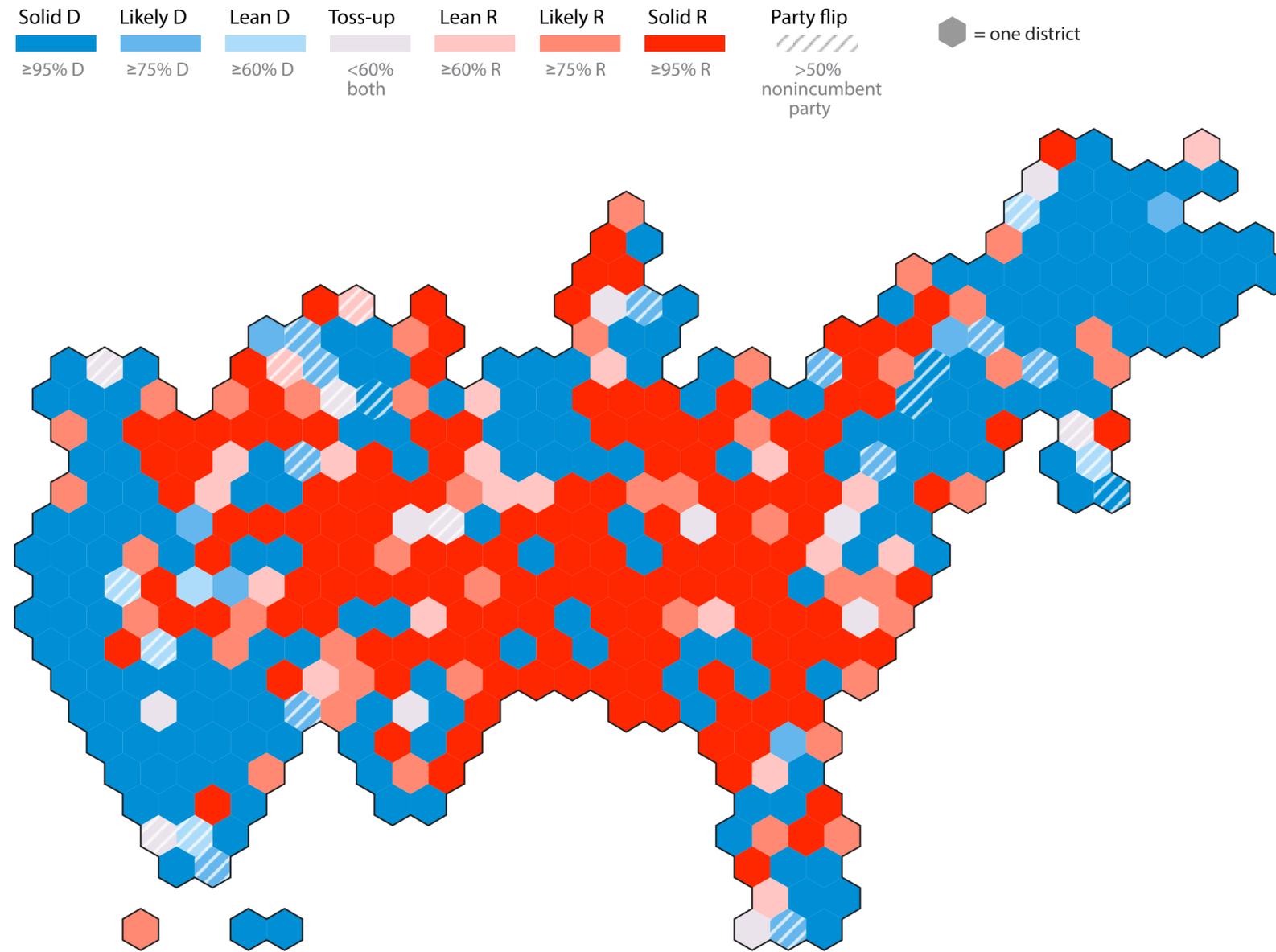
[M. Newman, 2009]

House Races: Map?



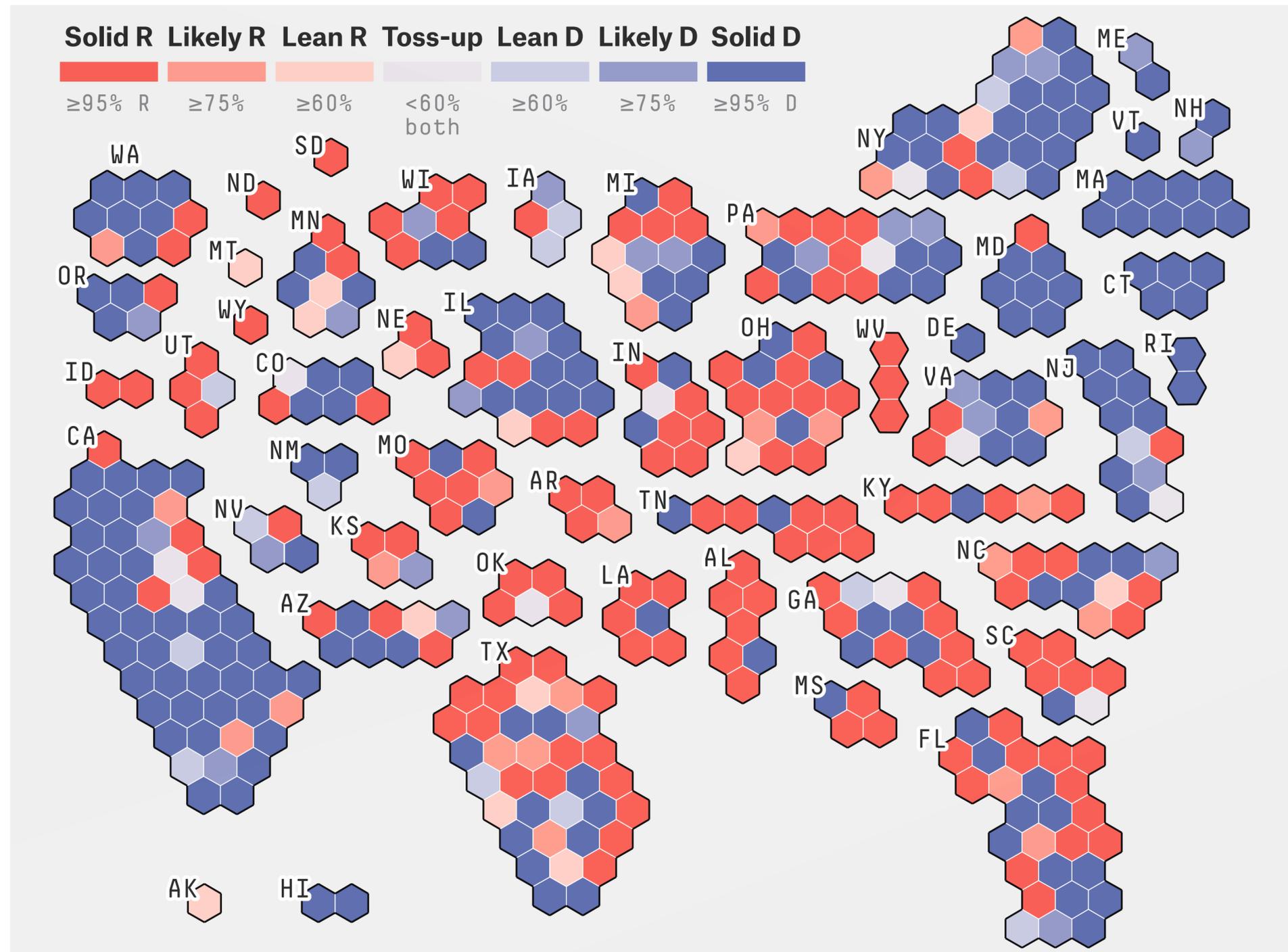
[New York Times, 2018]

House Races: Cartogram?



[FiveThirtyEight, 2018]

House Races: Non-Contiguous "Cartogram"



[FiveThirtyEight, 2020]

Maps Aren't Always Best: Close House Races

12 Lean Democratic

- AZ-02 Open (McSally)
- CA-49 Open (Issa)
- CO-06 Coffman
- IA-01 Blum
- KS-03 Yoder
- MI-11 Open (Trott)
- MN-02 Lewis
- MN-03 Paulsen
- NV-03 Open (Rosen)
- NJ-11 Open (Frelinghuysen)
- PA-07 Vacant (formerly Dent)
- VA-10 Comstock

31 Tossups

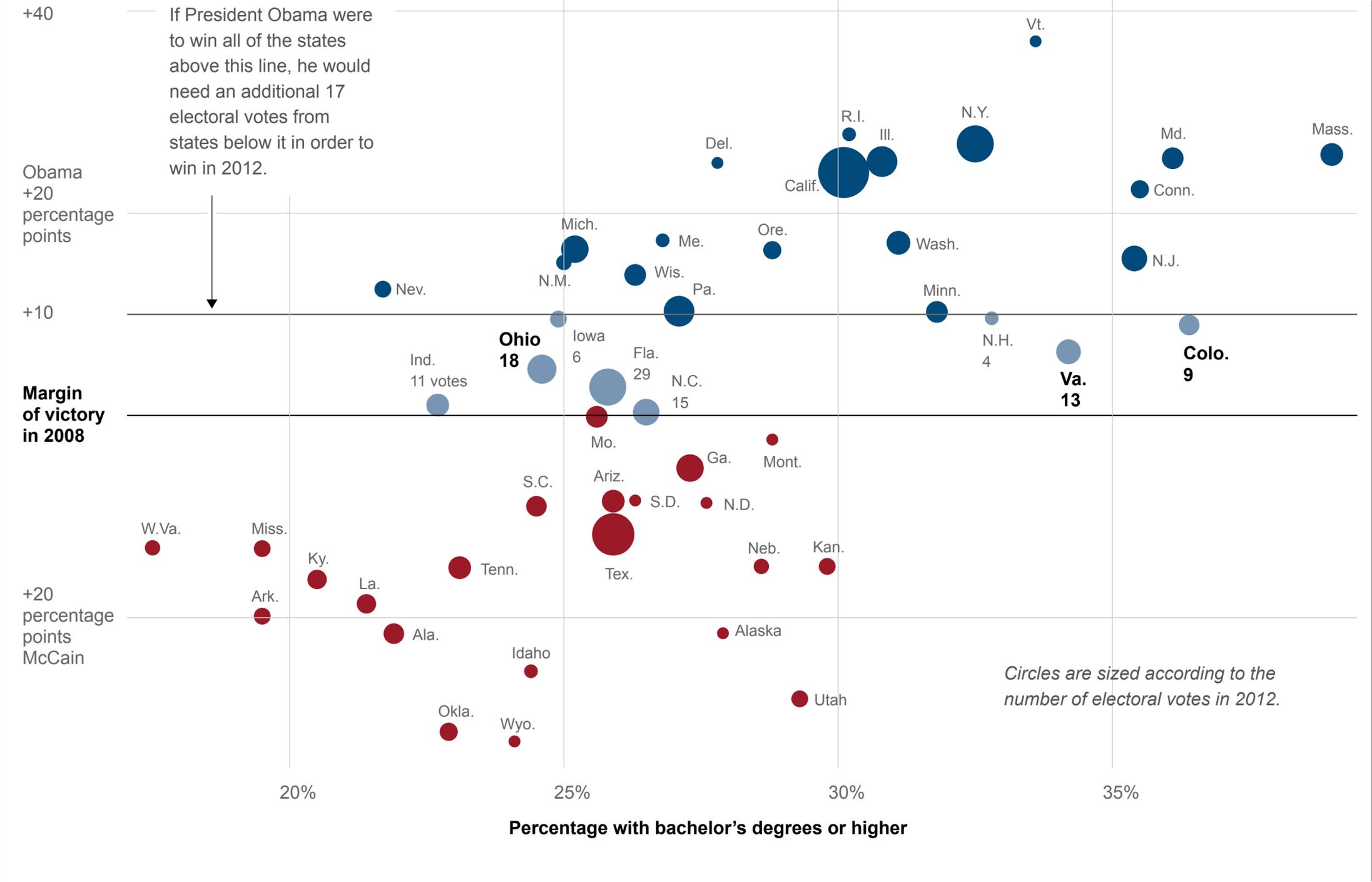
- CA-10 Denham
- CA-25 Knight
- CA-39 Open (Royce)
- CA-45 Walters
- CA-48 Rohrabacher
- FL-26 Curbelo
- FL-27 Open (Ros-Lehtinen)
- IL-06 Roskam
- IL-12 Bost
- IA-03 Young
- KS-02 Open (Jenkins)
- KY-06 Barr

25 Lean Republican

- AR-02 Hill
- CA-50 Hunter
- FL-15 Open (Ross)
- FL-16 Buchanan
- GA-06 Handel
- GA-07 Woodall
- IL-13 Davis
- IL-14 Hultgren
- MO-02 Wagner
- MT-AL Gianforte
- NE-02 Bacon
- NY-24 Katko

[New York Times, 2018]

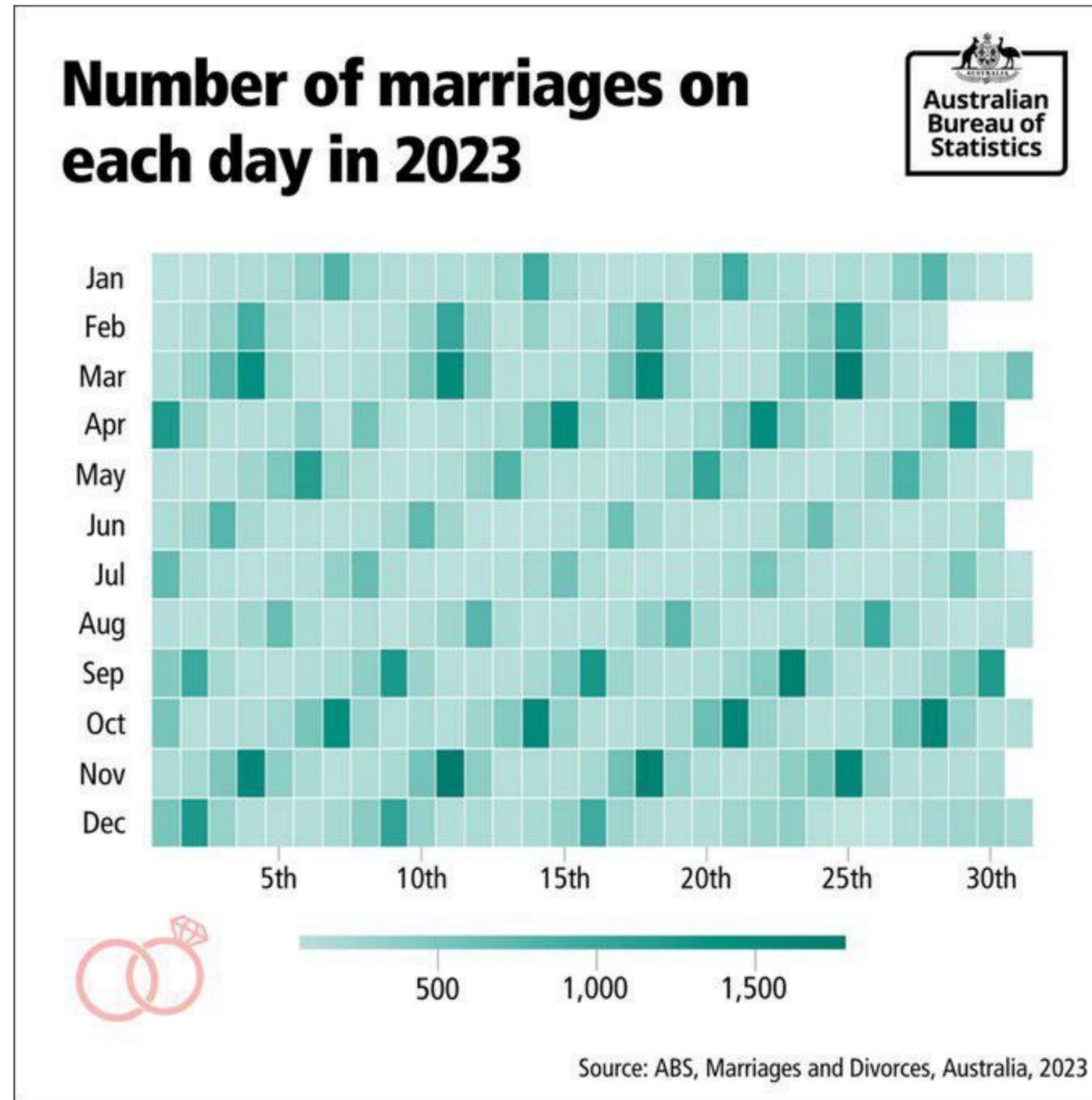
Maps Aren't Always Best: Obama Targets



[NYTimes]

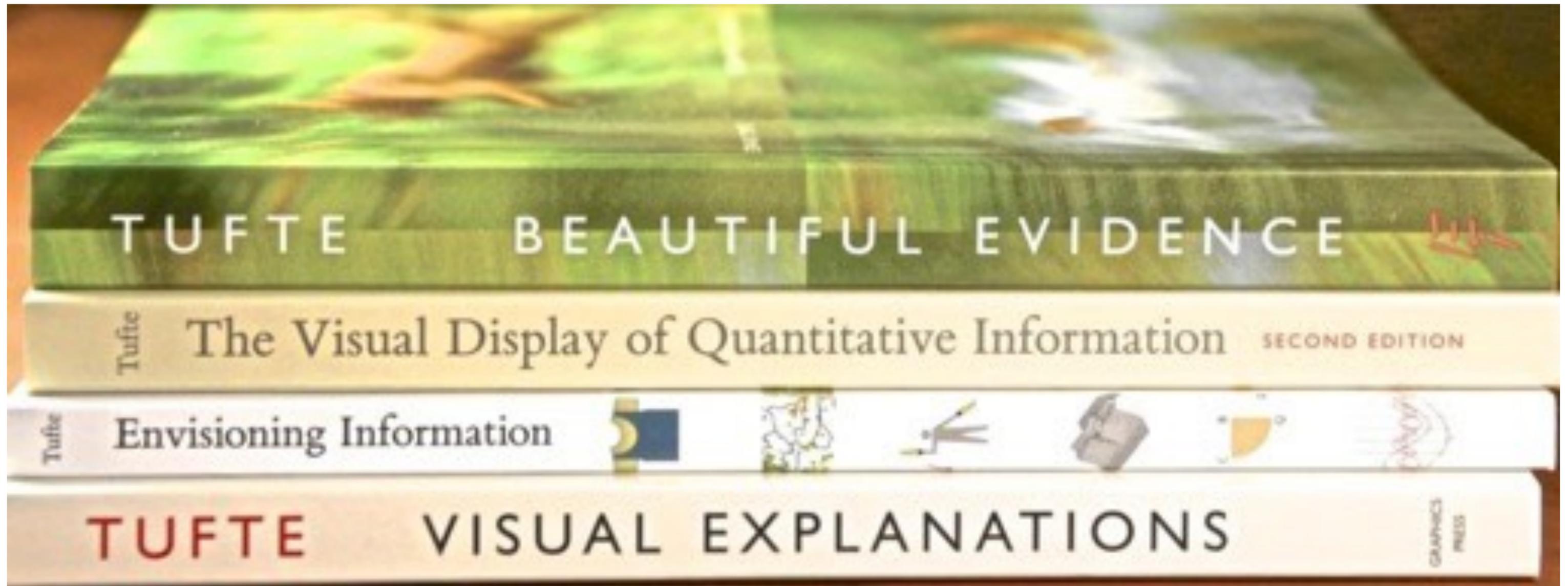


Guidelines for Visualization Design



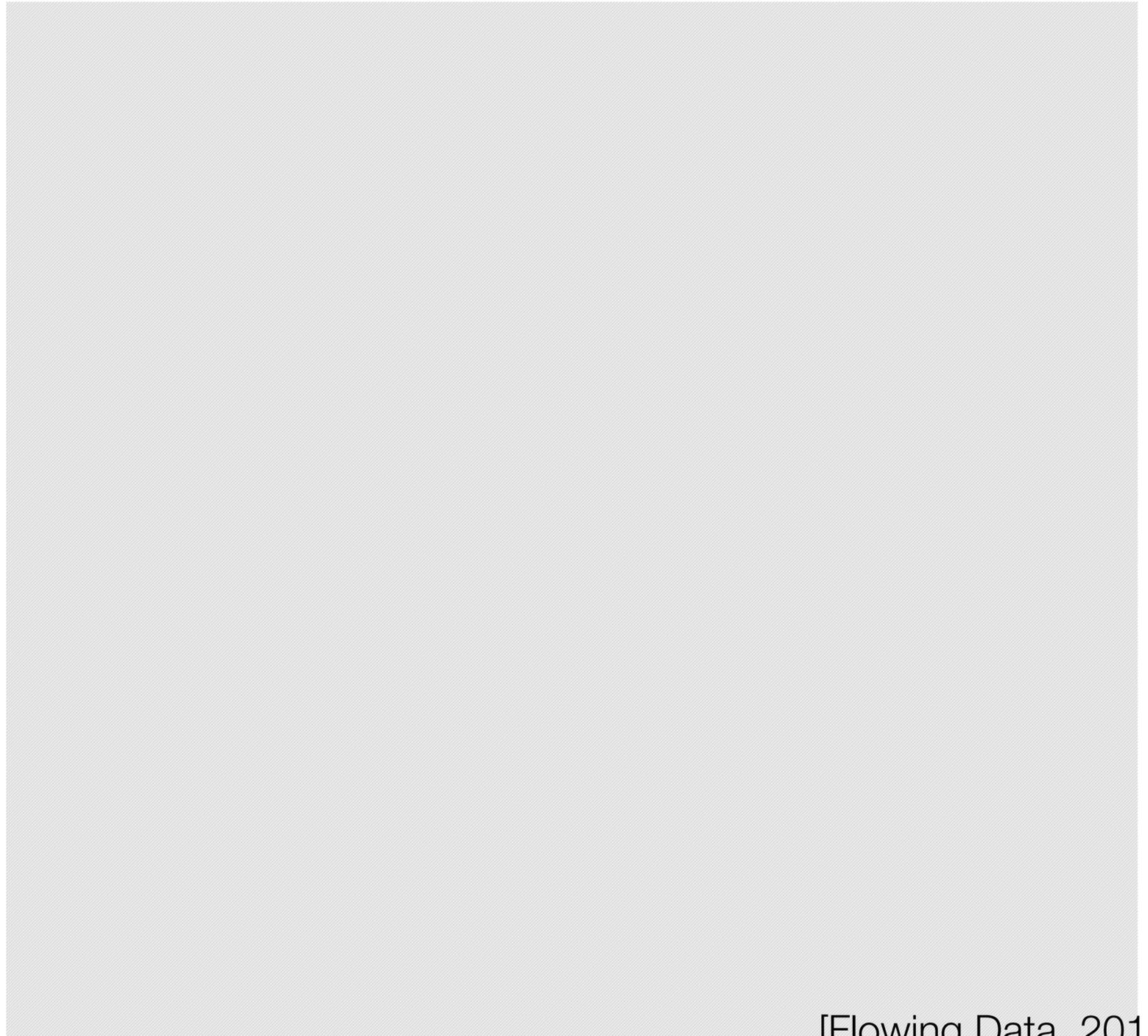
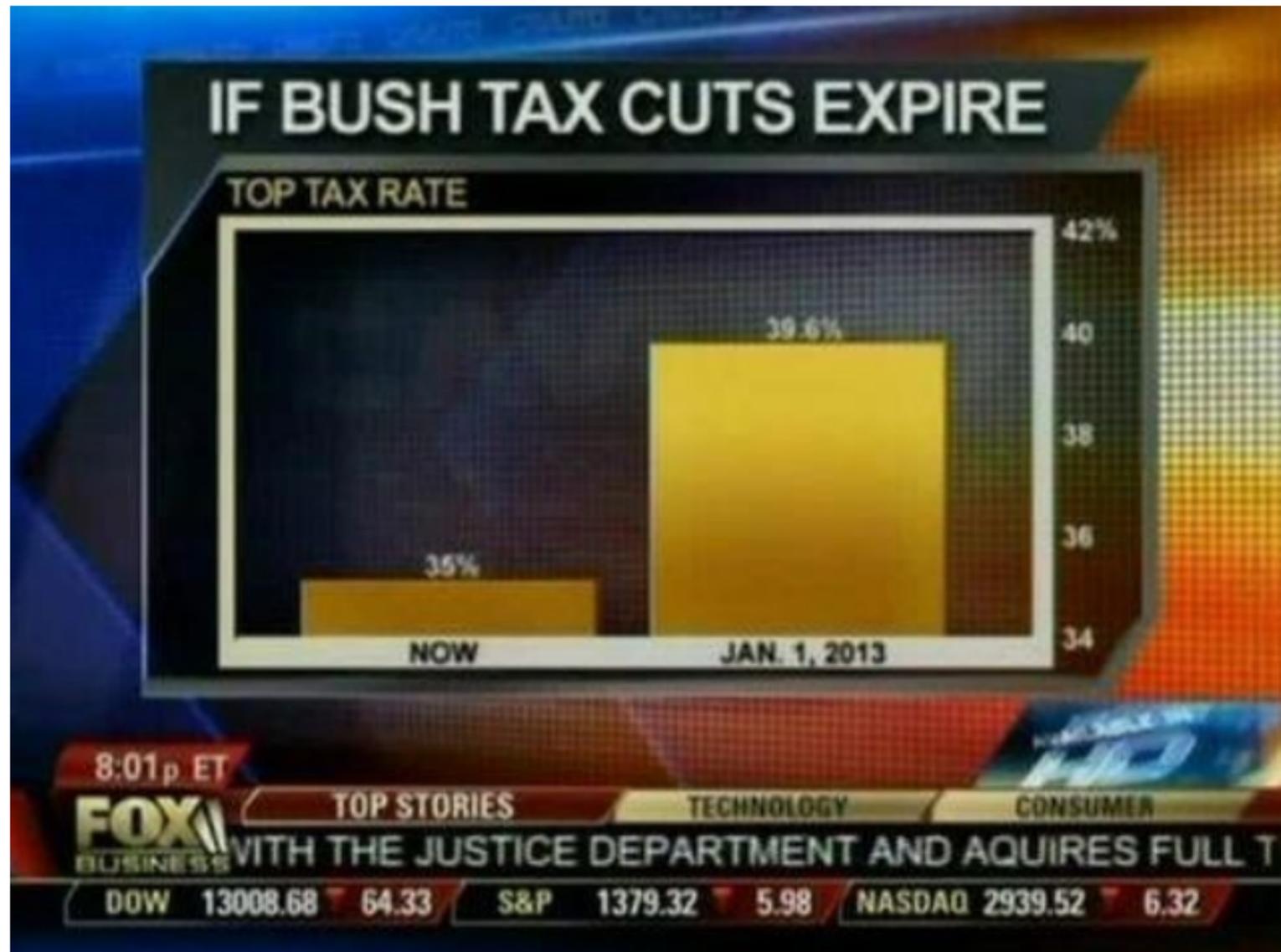
[Reddit]

Tufte: "The da Vinci of Data" – NYTimes



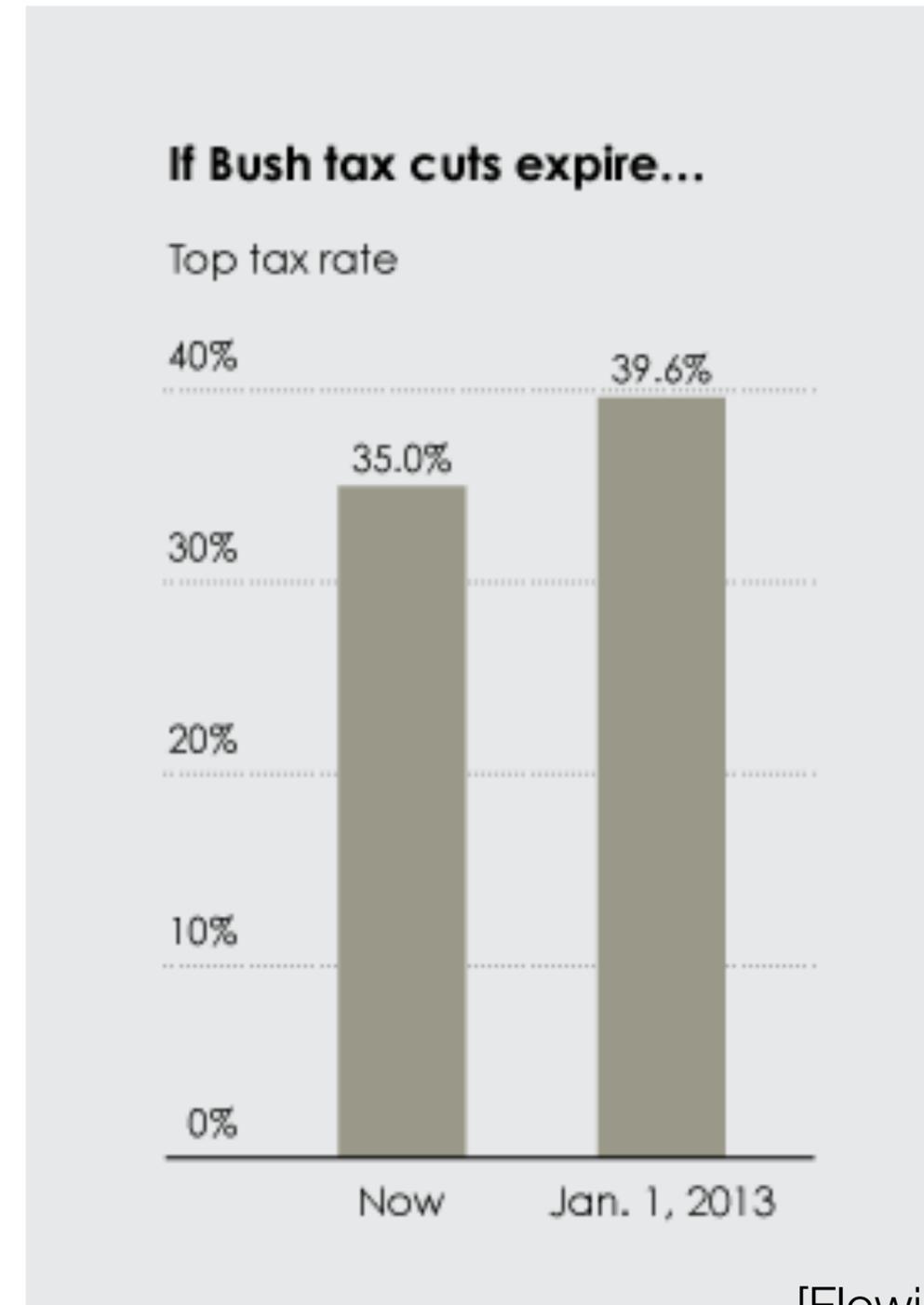
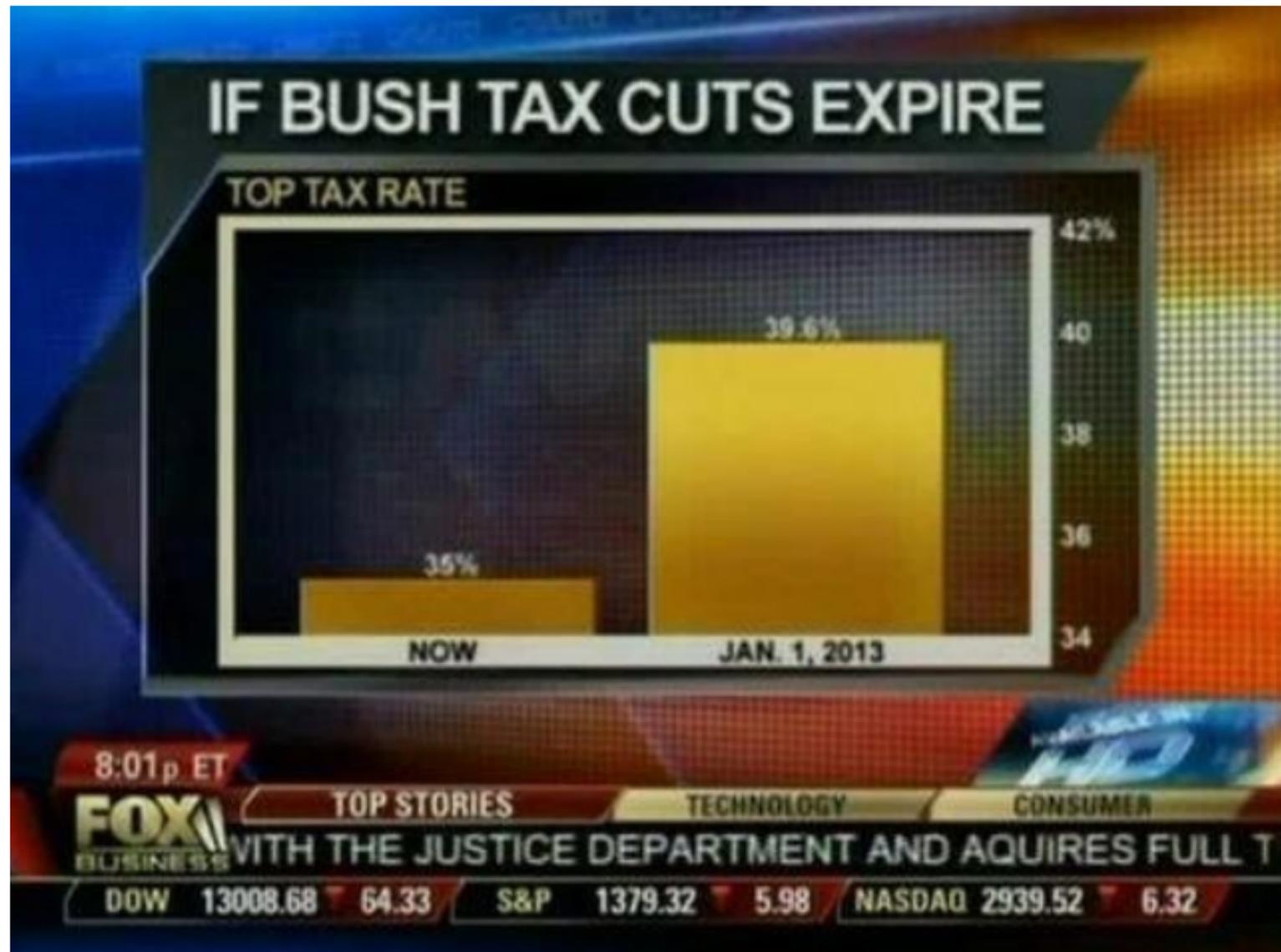
[<https://www.edwardtufte.com/tufte/>, 2017]

Bad: Data magnitude \neq Mark magnitude



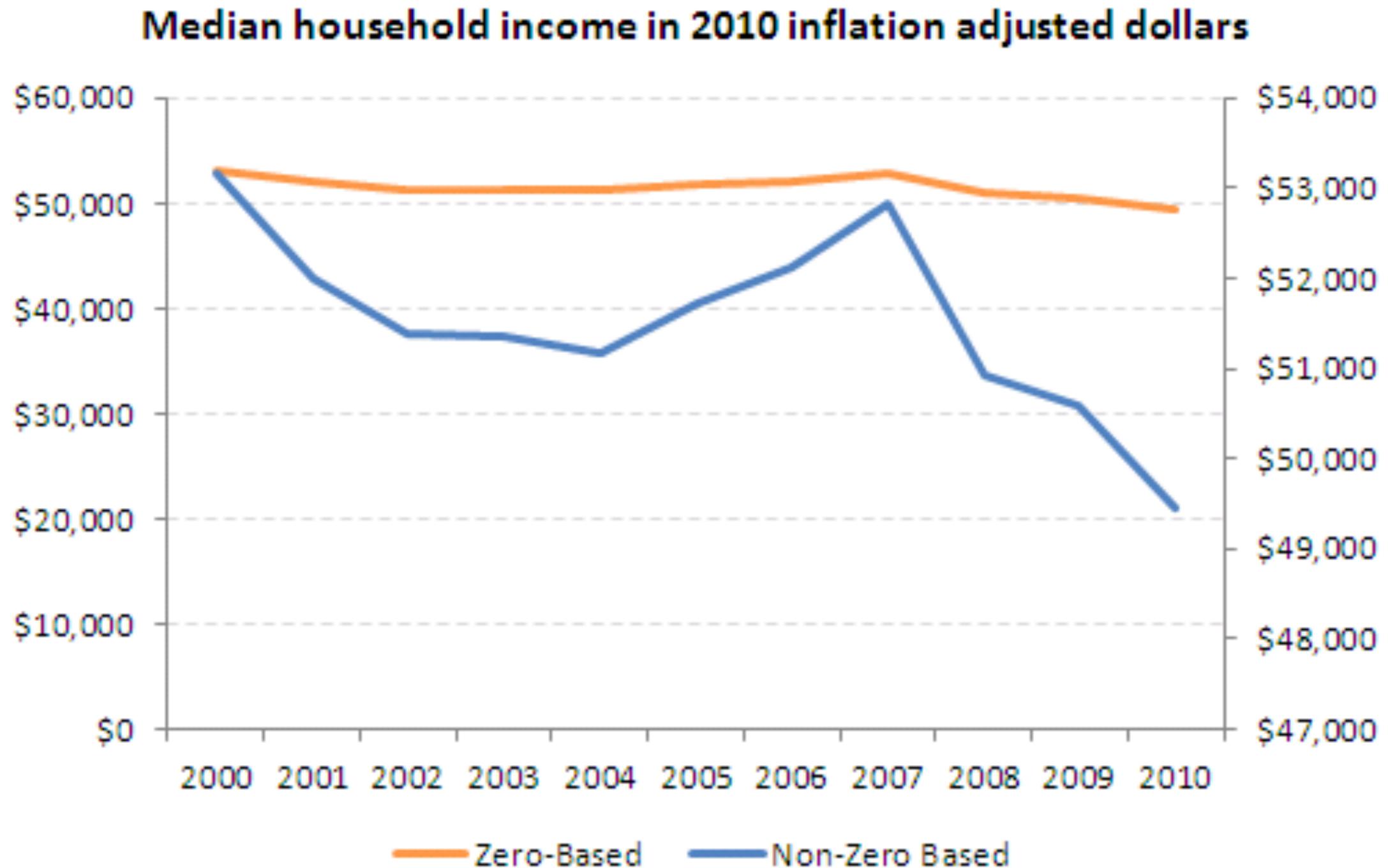
[Flowing Data, 2012]

Good: Data magnitude \Leftrightarrow Mark magnitude



[Flowing Data, 2012]

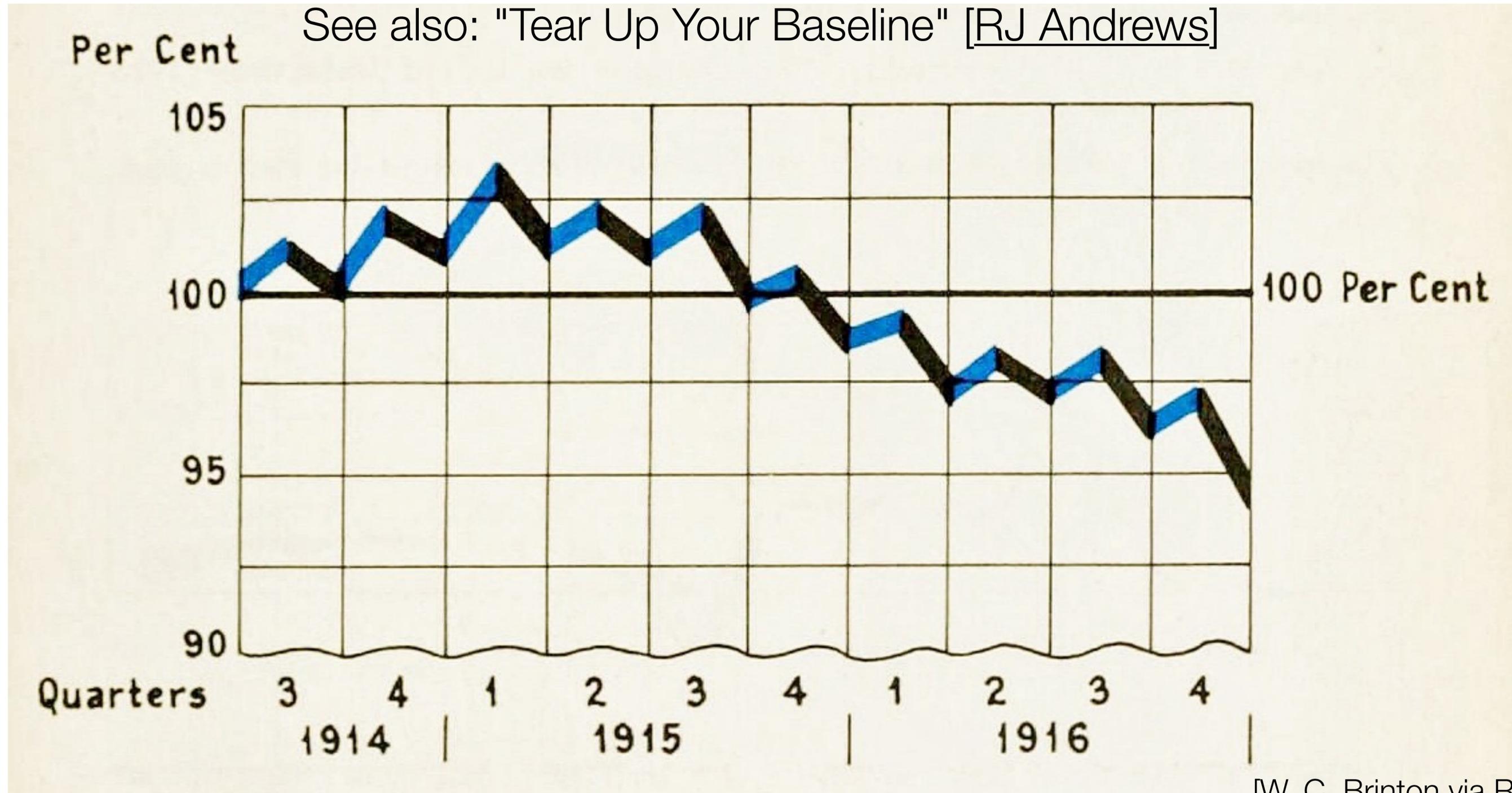
Starting Scales at Zero?



[A. Kirebel, VizWiz]

Wavy baselines for non-zero starts

See also: "Tear Up Your Baseline" [[RJ Andrews](#)]



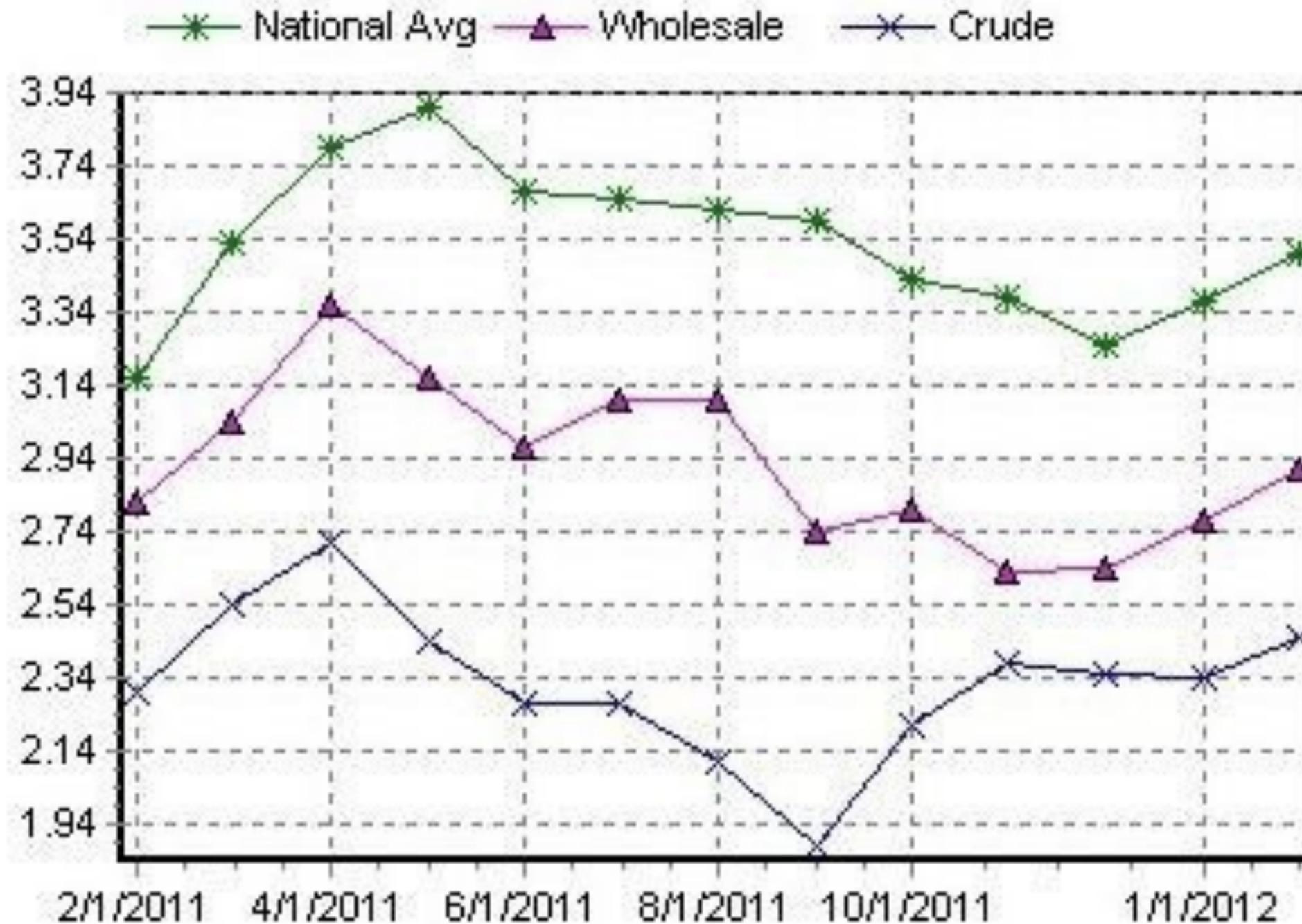
[W. C. Brinton via [RJ Andrews](#)]

Cherry-picking data



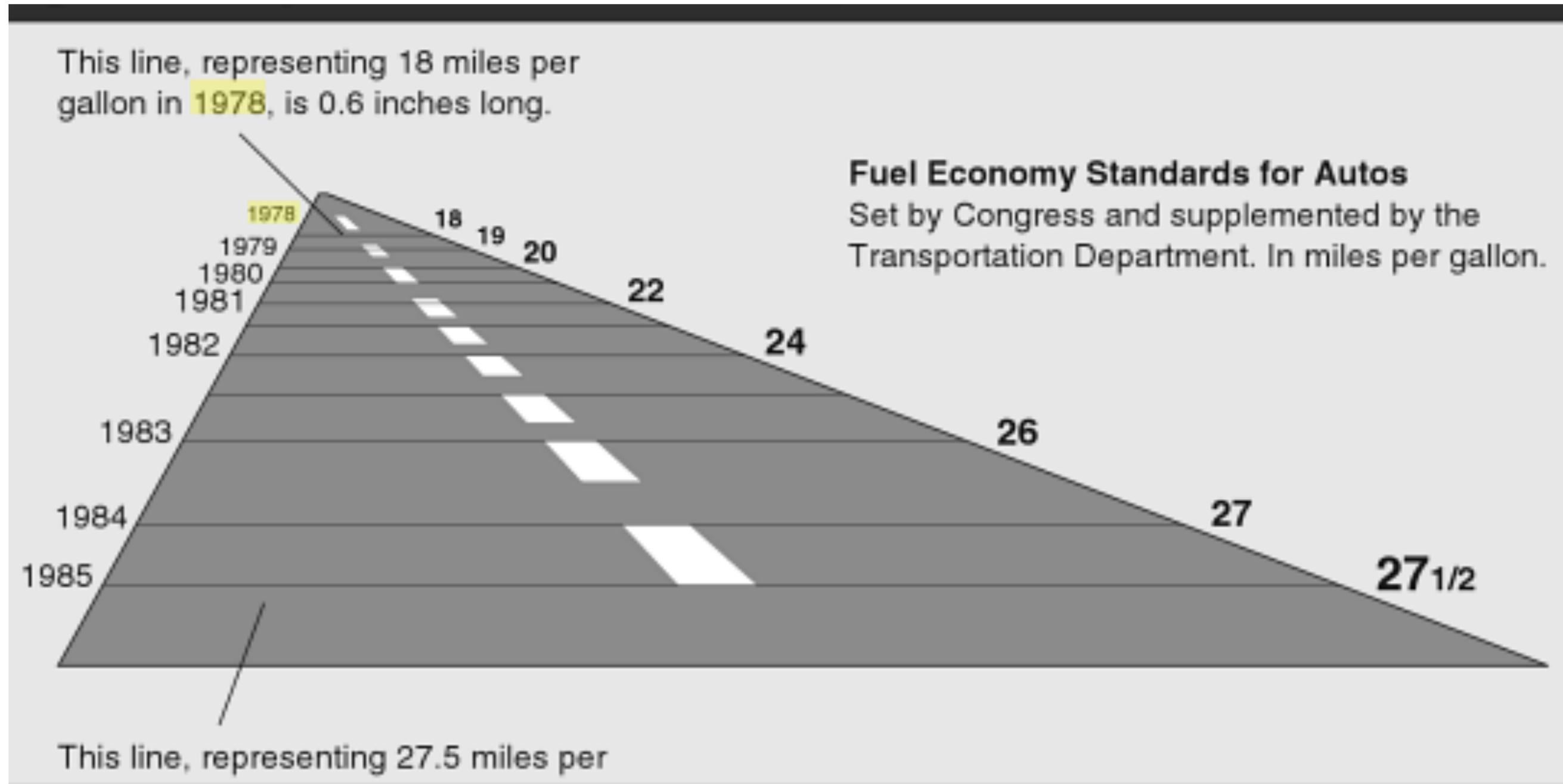
[Fox News via [Media Matters](#), 2012]

Show **all** the data



[AAA via [Media Matters](#), 2012]

Tufte's Lie Factor

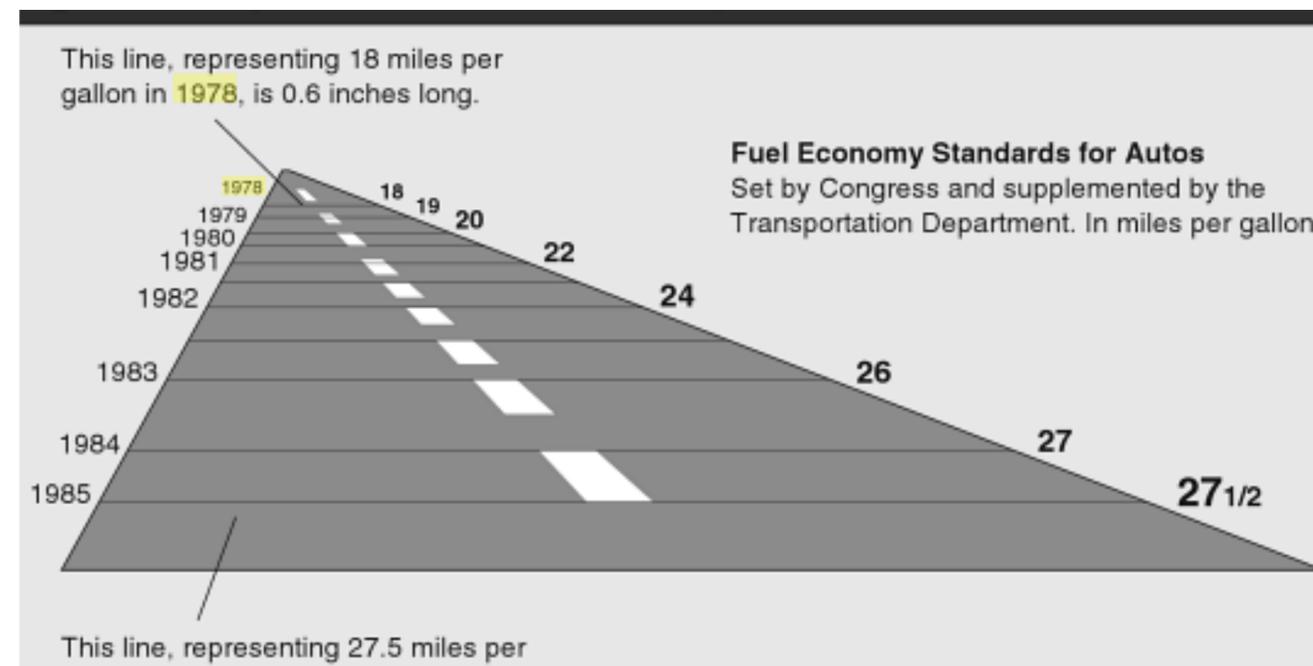


[NYTimes via Tufte, 1991]

Tufte's Lie Factor

- Size of effect = (2nd value - 1st value) / (1st value)
- Lie factor = (size of effect in graphic) / (size of effect in data)
- In the graphic:

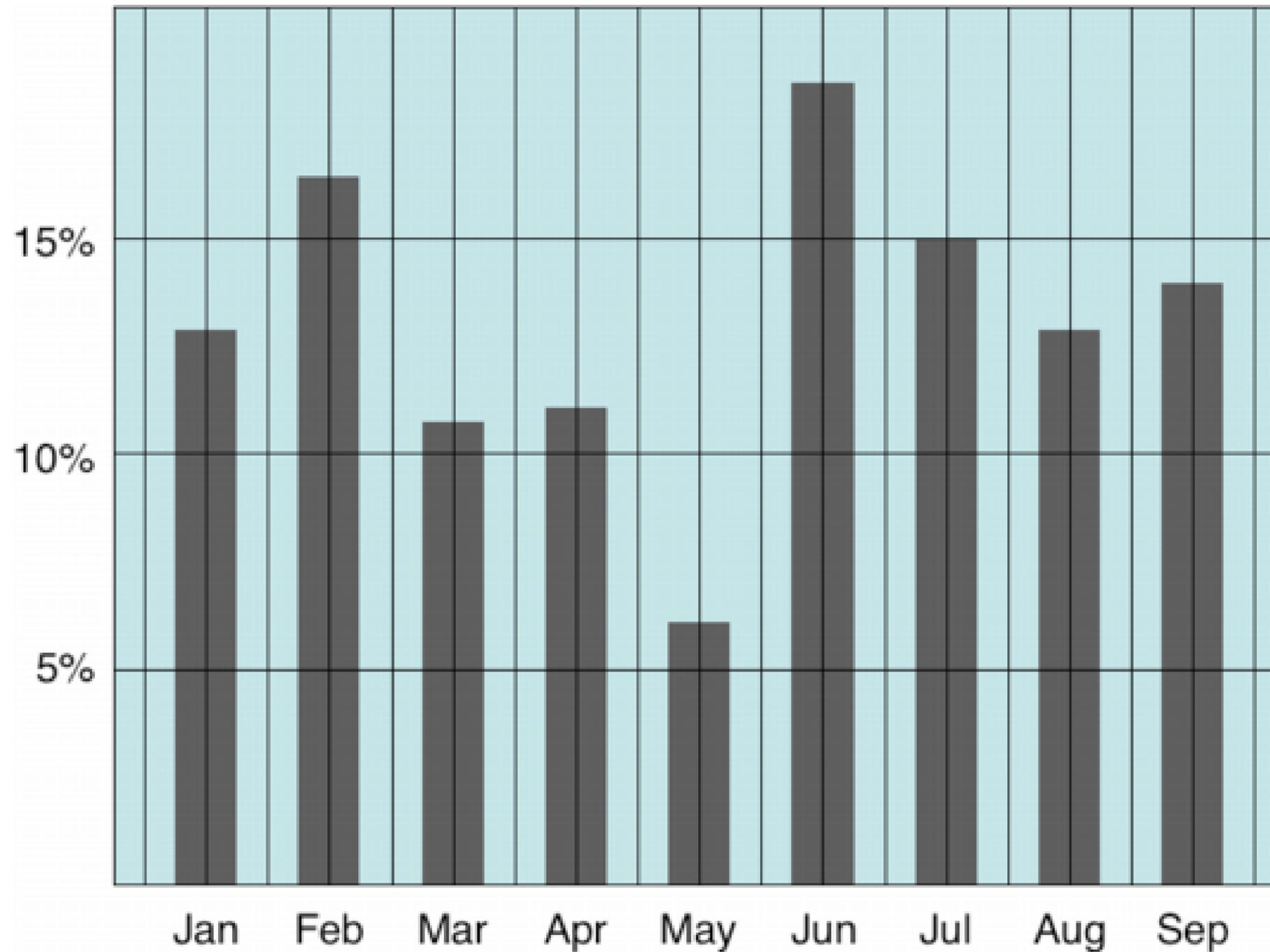
$$\text{Lie Factor} = \frac{\frac{5.3 - 0.6}{0.6}}{\frac{27.5 - 18}{18}} = 14.8$$



(Some of) Tufte's Integrity Principles

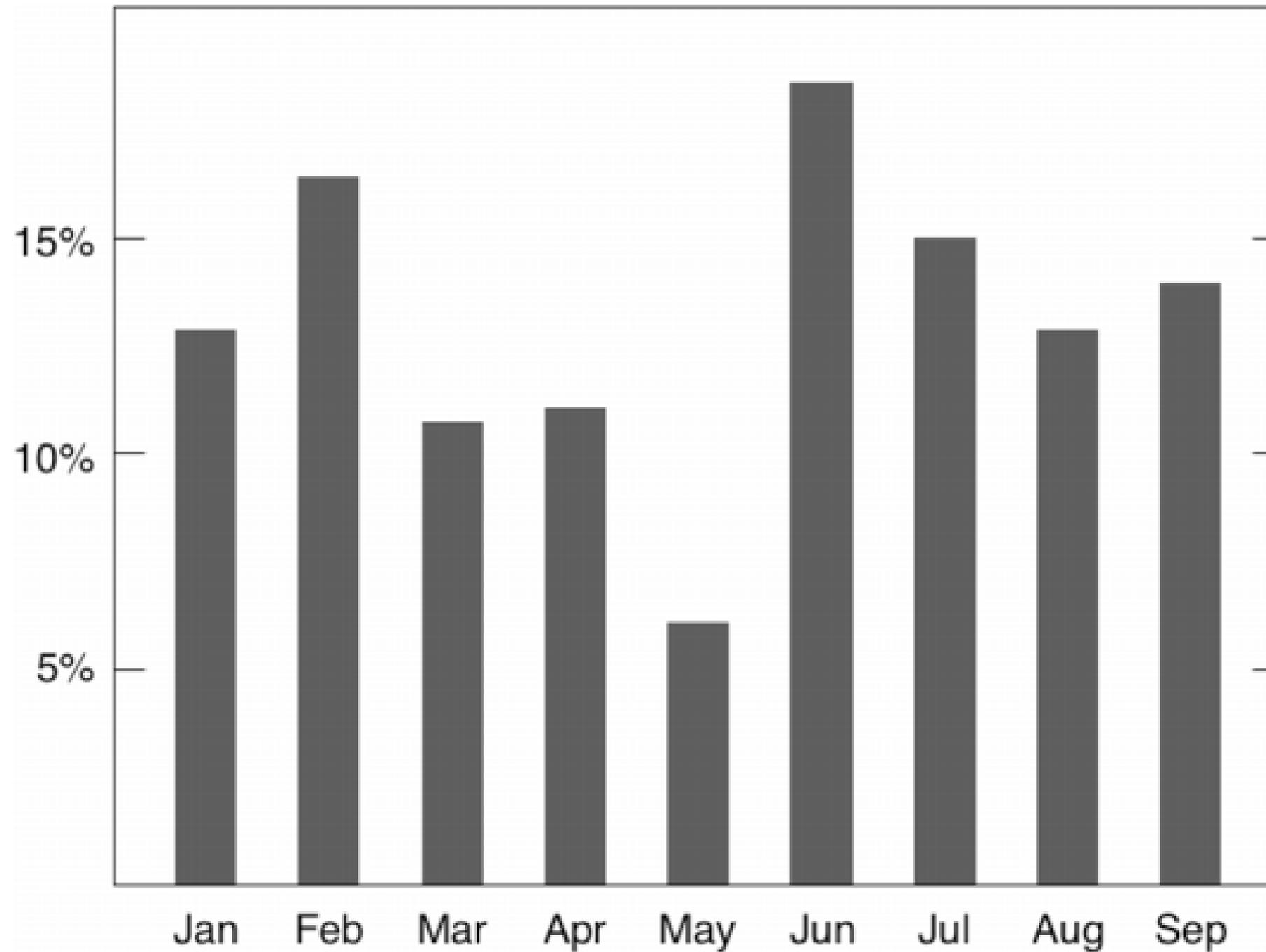
- Show data variation, not design variation
- Clear, detailed, and thorough labeling and appropriate scales
- Size of the graphic effect should be directly proportional to the numerical quantities ("lie factor")

Avoid Chartjunk



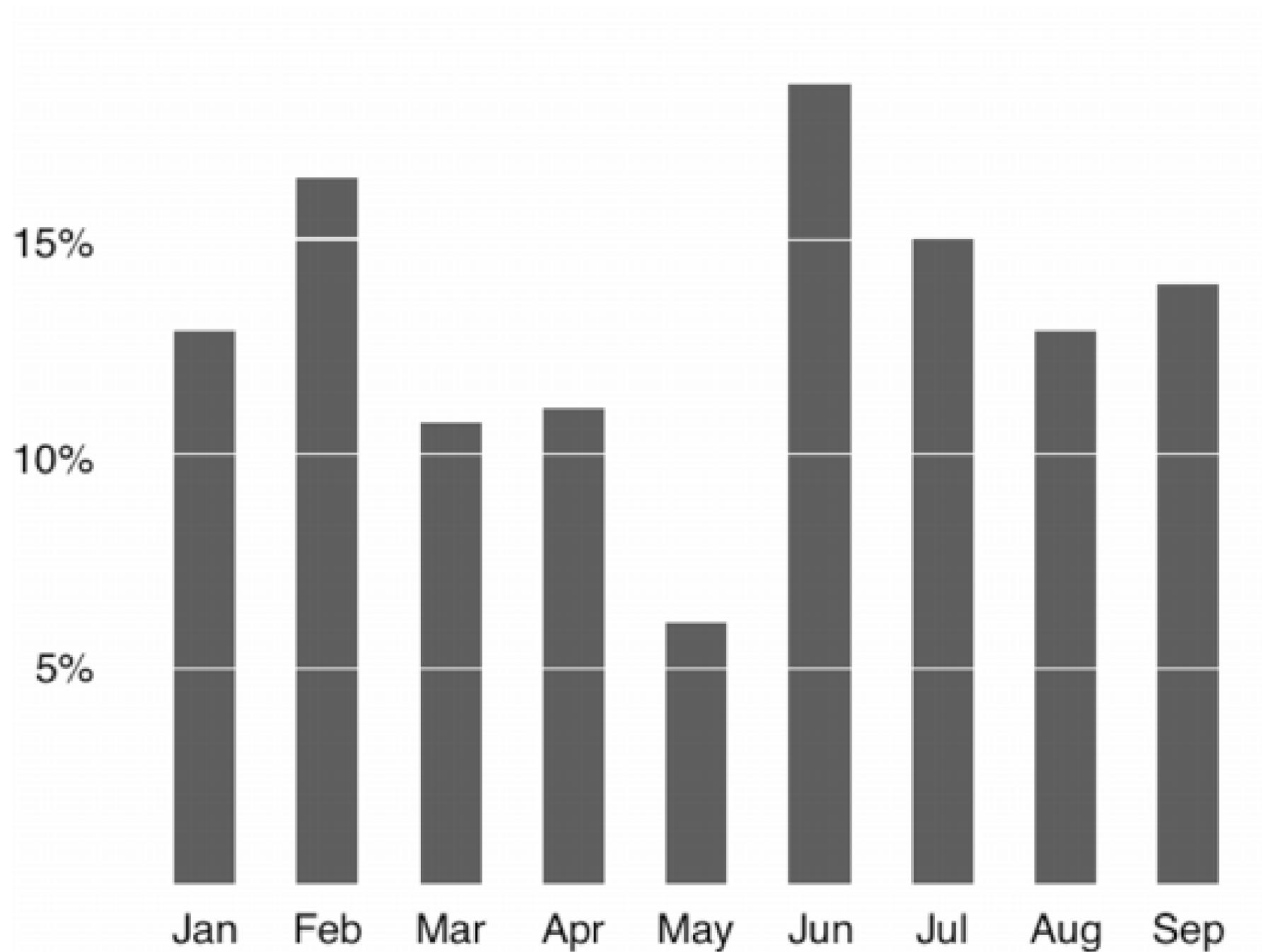
[T. Brey via [A. Lex](#)]

Avoid Chartjunk



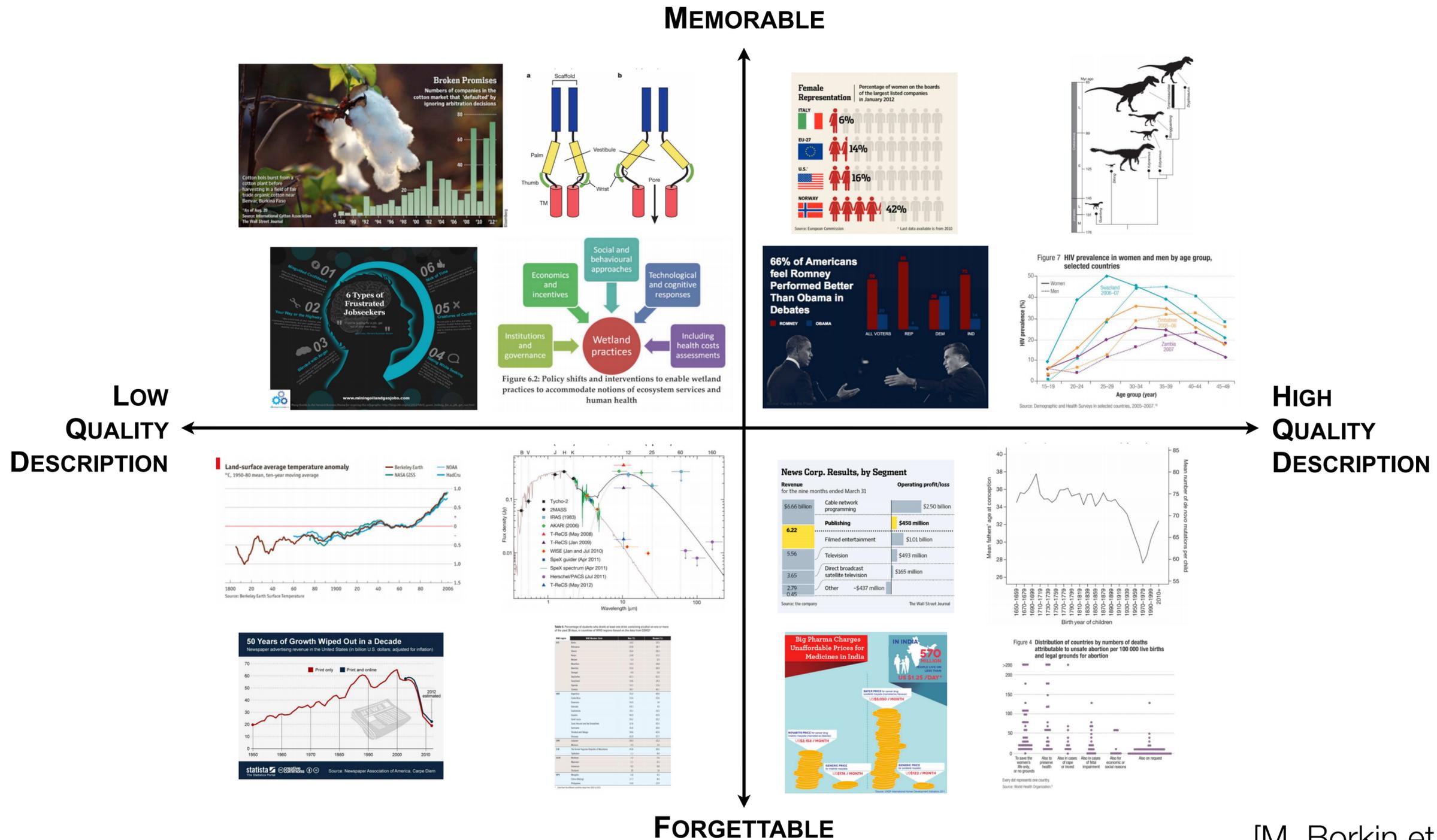
[T. Brey via [A. Lex](#)]

Avoid Chartjunk



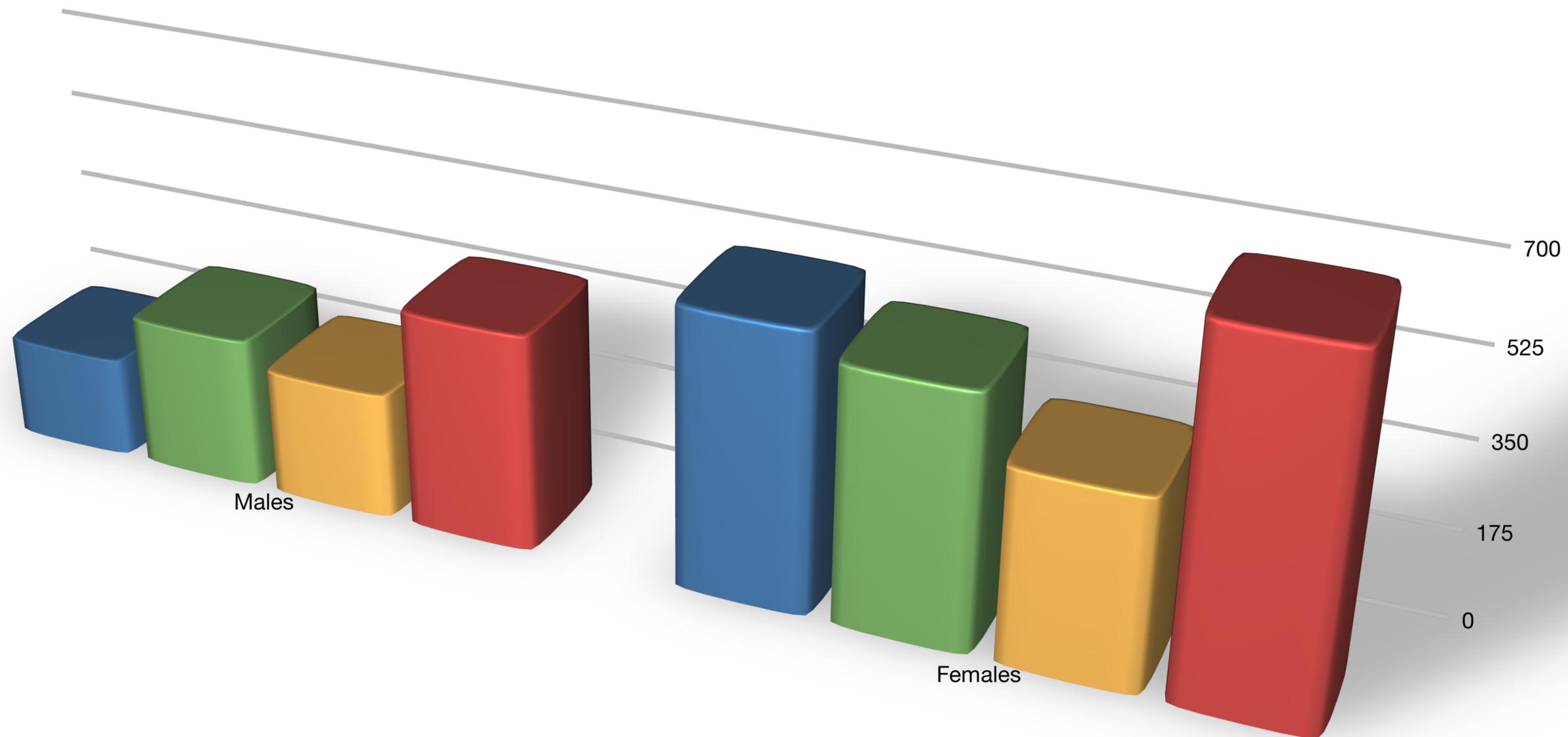
[T. Brey via [A. Lex](#)]

Avoid Chartjunk?



[M. Borkin et al., InfoVis 2015]

Data-to-Ink Ratio (Also Unjustified 3D)



■ 0-\$24,999

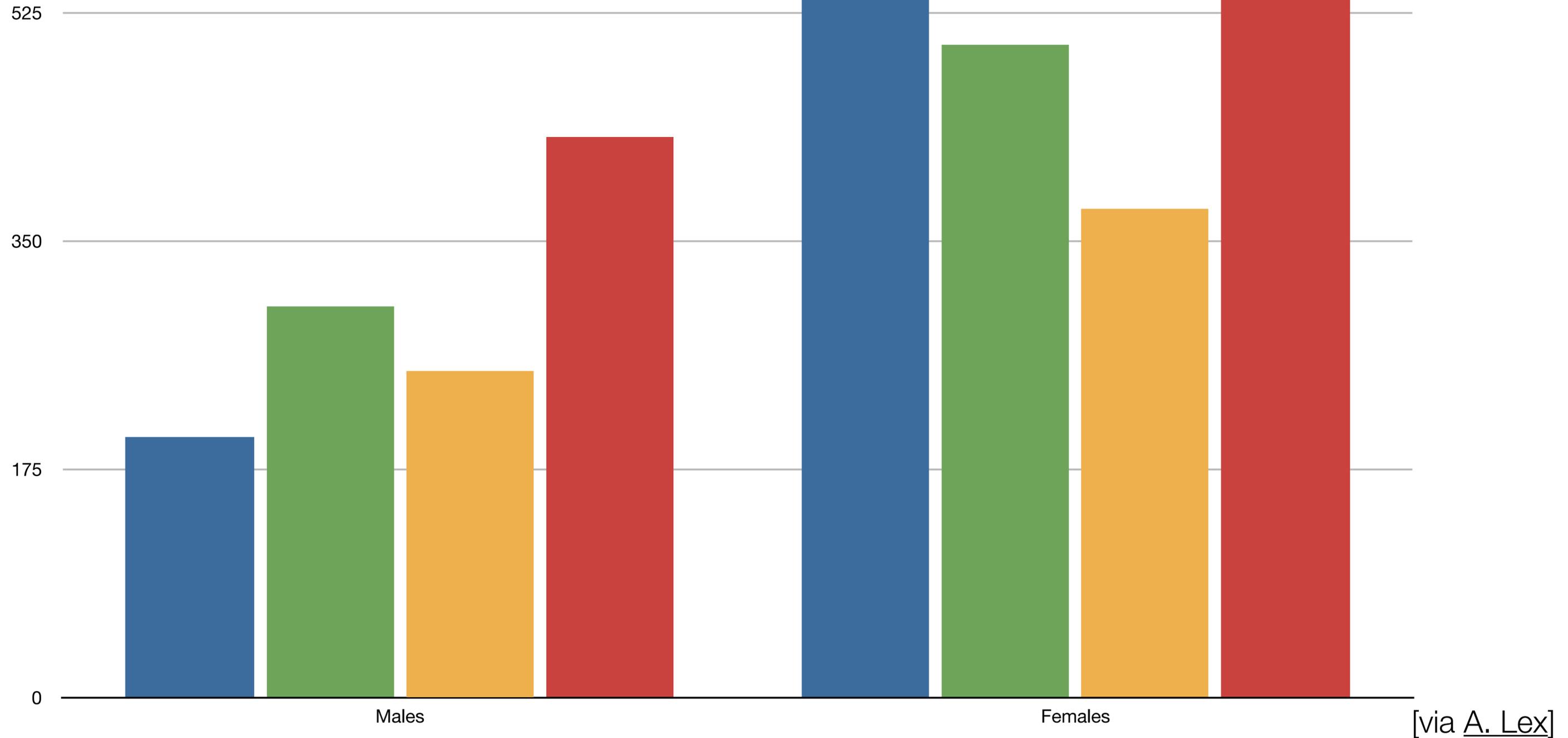
■ \$25,000+

■ 0-\$24,999

■ \$25,000+

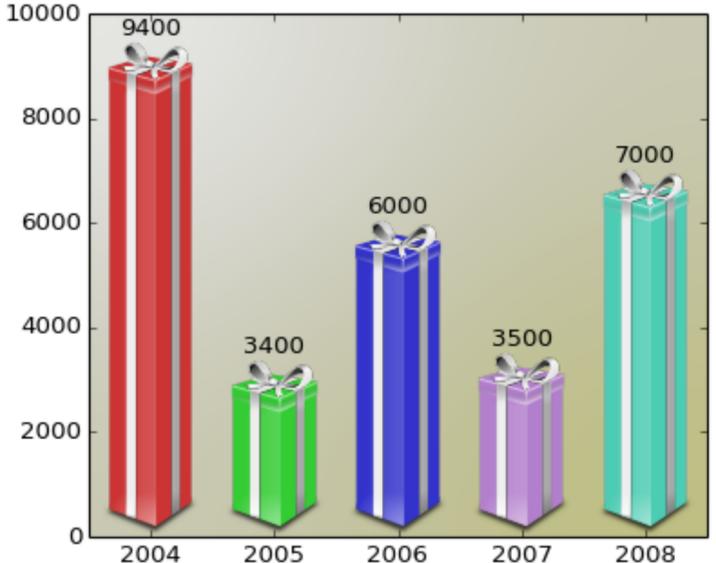
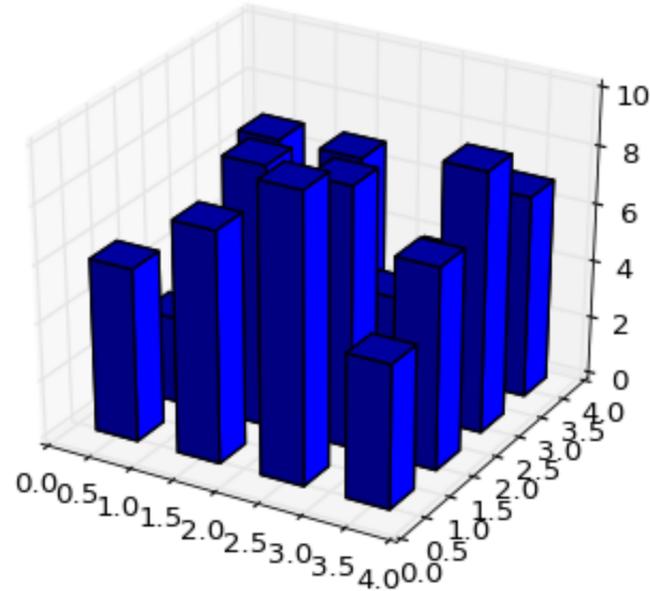
[via [A. Lex](#)]

Maximize Data-to-Ink Ratio

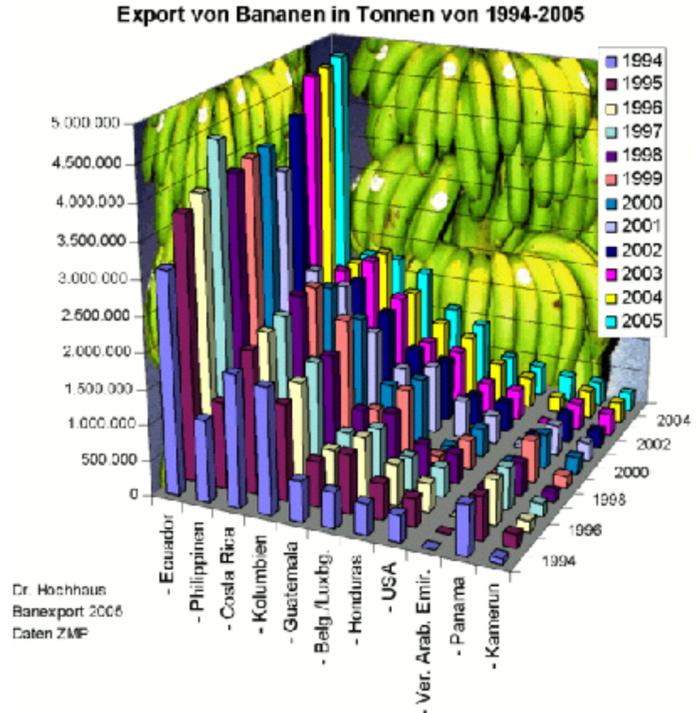


[via A. Lex]

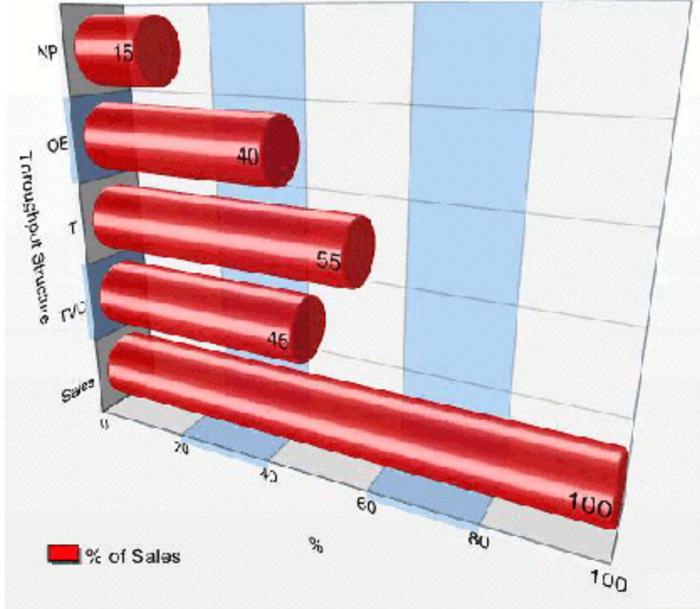
No Unjustified 3D



matplotlib gallery



Dr. Hochhaus
Banexport 2006
Daten ZMF

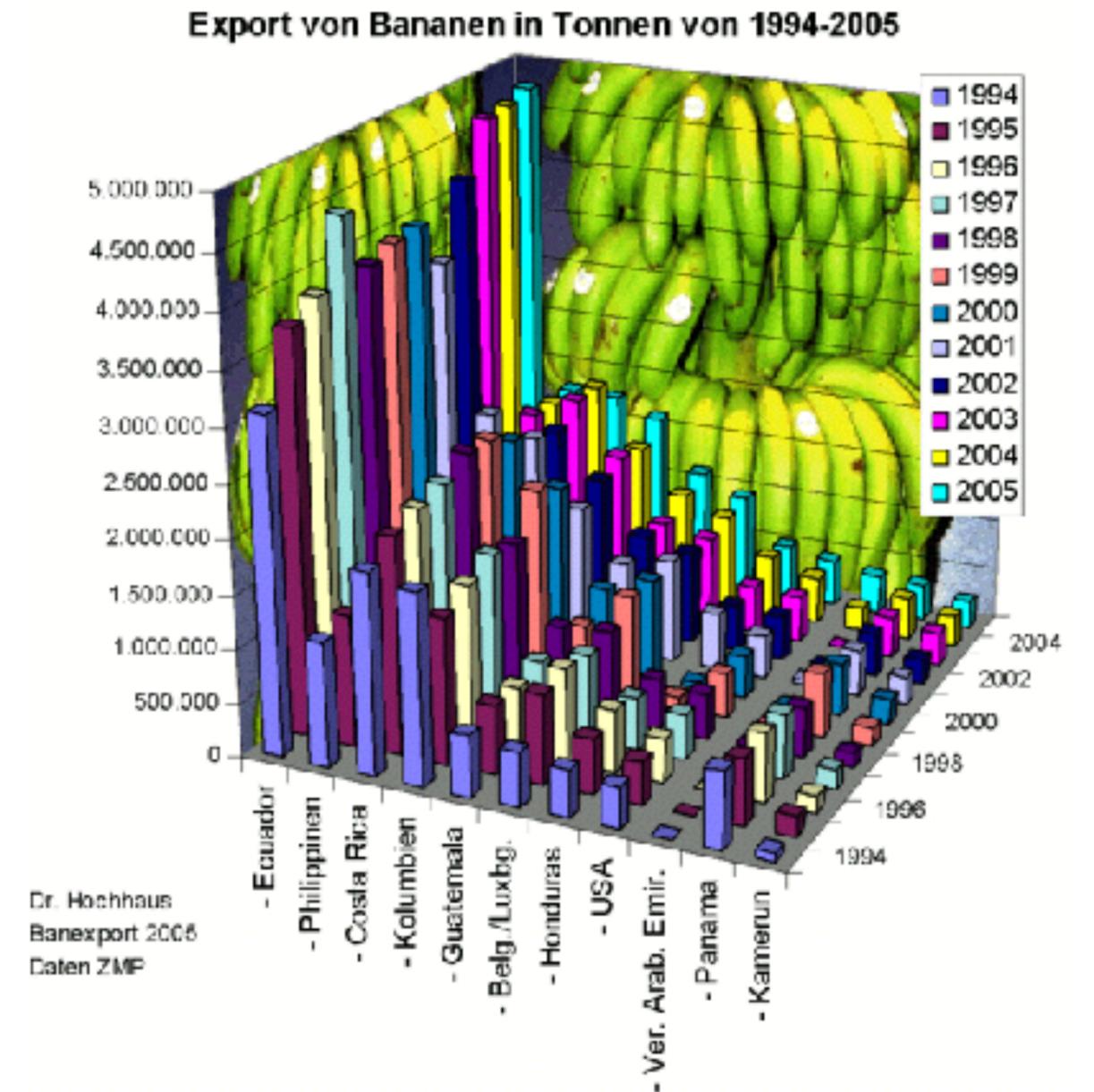


Excel Charts Blog

[via A. Lex]

No Unjustified 3D

- Occlusion hides information
- Perspective distortion dangers
- Tilted text isn't legible
- Can **help** with shape perception



[via A. Lex]

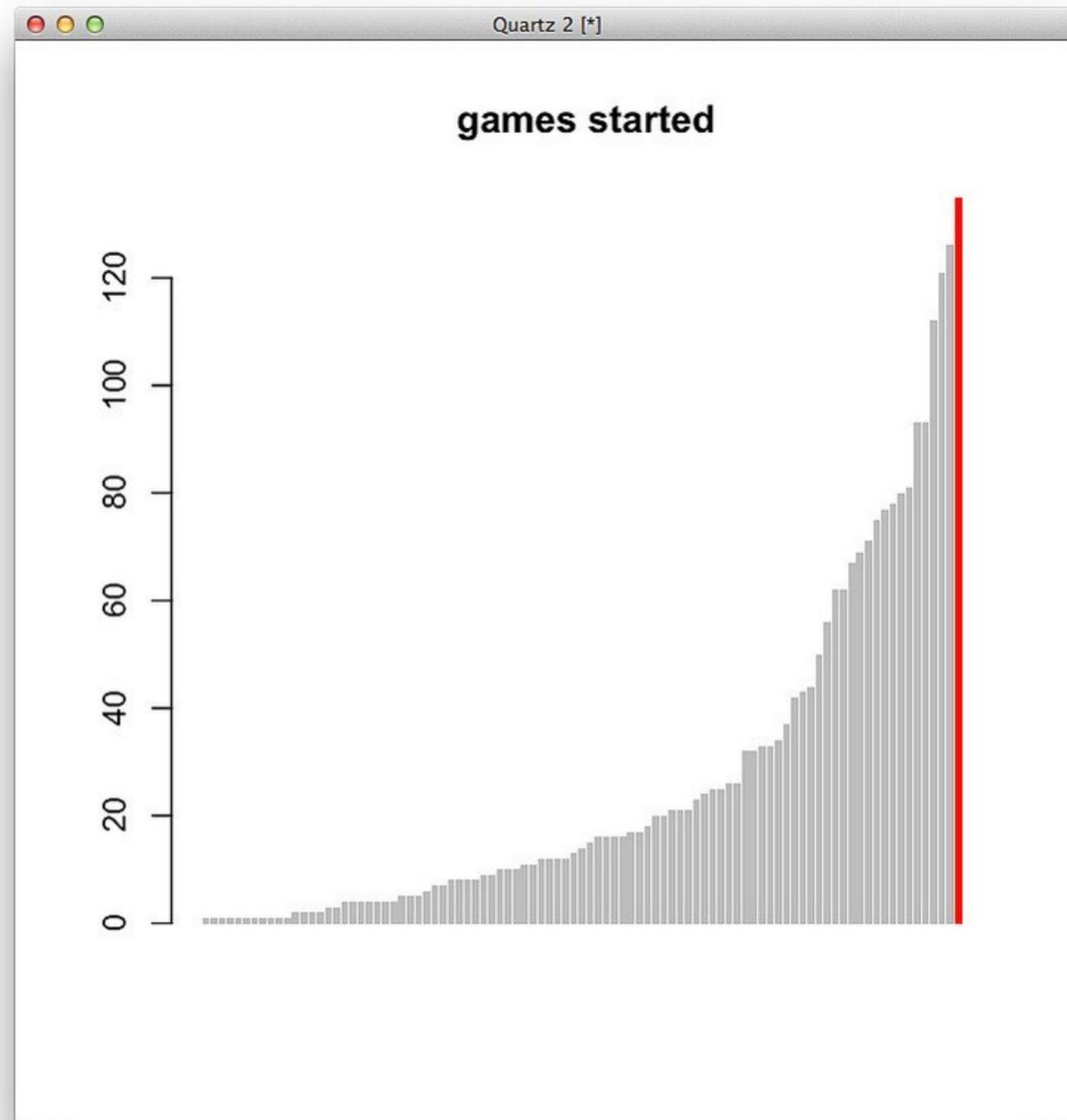
Eyes Beat Memory

- Reduce cognitive load (using up working memory)
- Animation versus side-by-side views
- Change blindness

“Computer-based visualization systems provide visual representations of datasets **designed** to help people carry out tasks more effectively.”

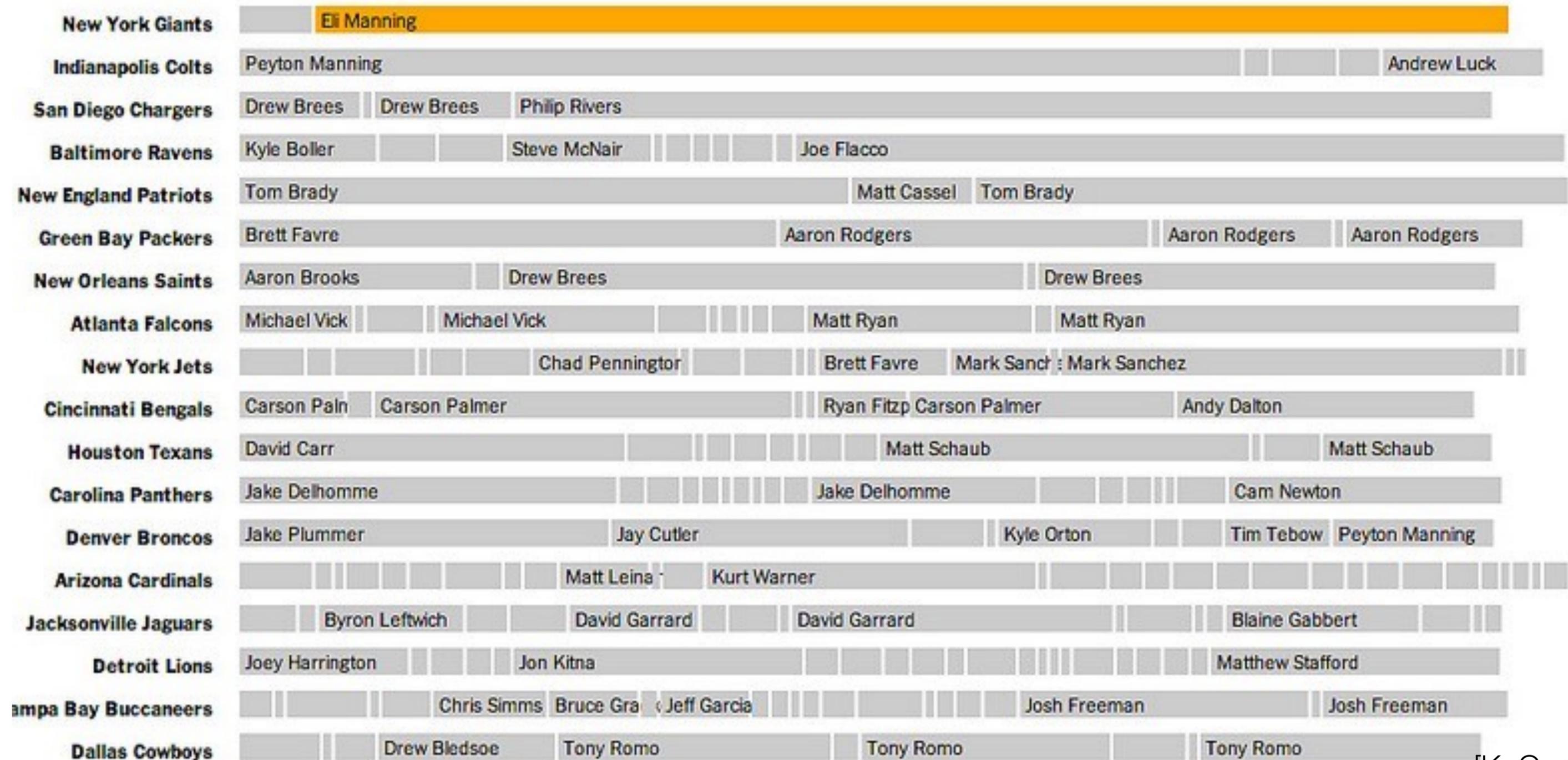
— T. Munzner

Design Iteration



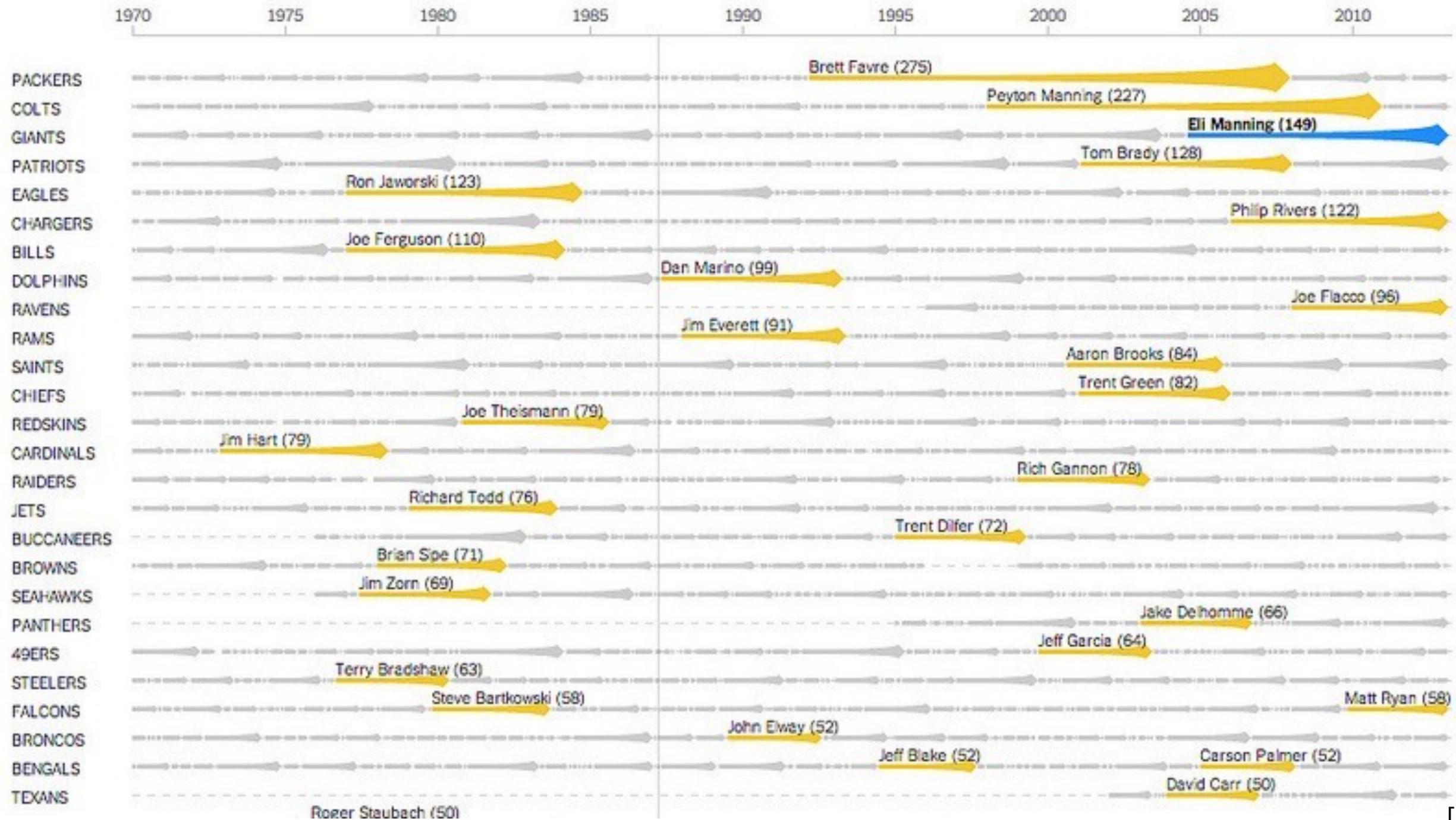
[K. Quealy, 2013]

Design Iteration



[K. Quealy, 2013]

Design Iteration

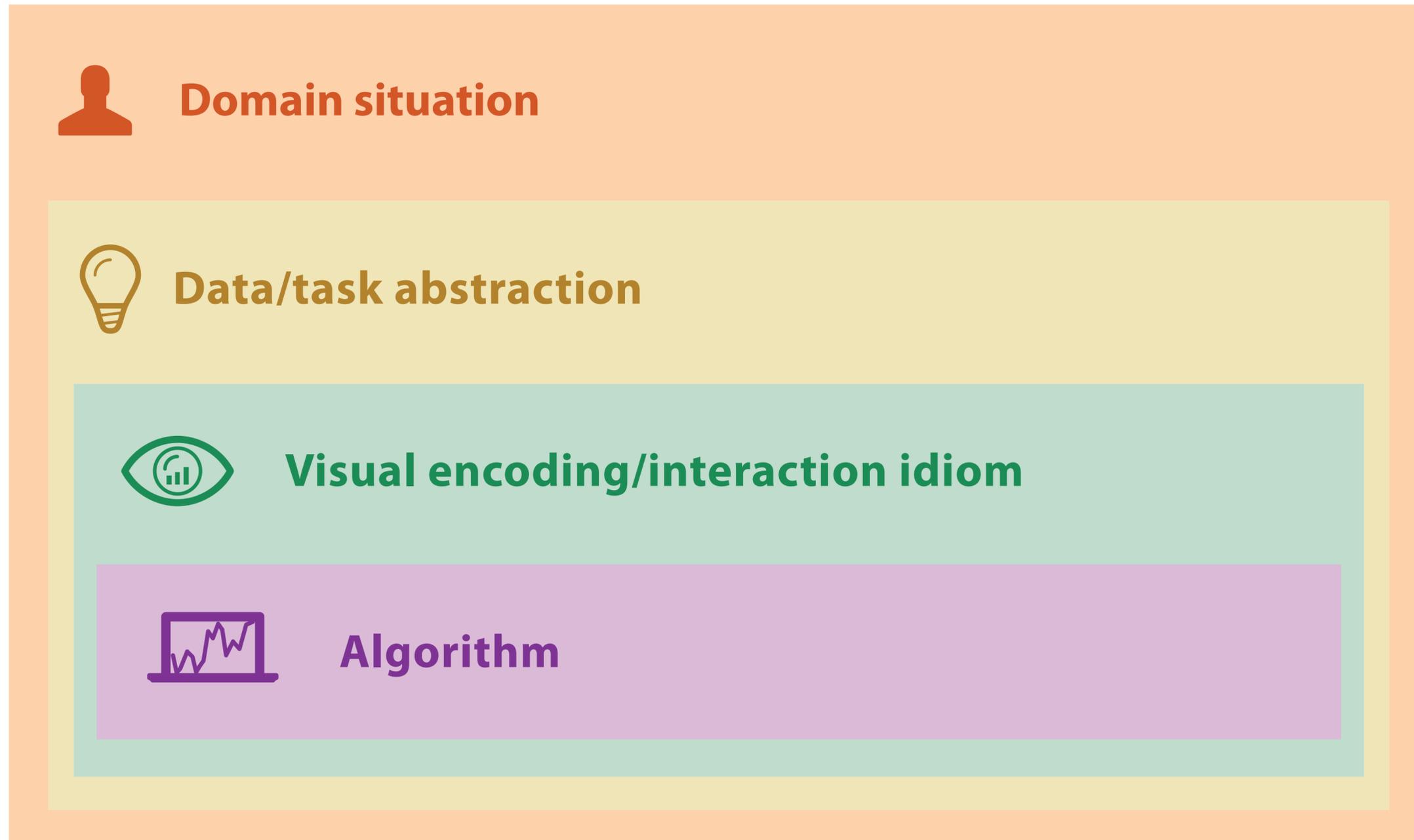


[K. Quealy, 2013]

Design

- Unlike a math problem, there are many different approaches for the visualization of some data
- Need to have some way to discuss how to determine whether a visualization is doing what we want
- Validation: Understand why a design is effective
 - What problems can be effective
 - Do this at different levels

Four Nested Levels of Design



[Munzner, 2014]

Potential problems at each level

 **Domain situation**
You misunderstood their needs

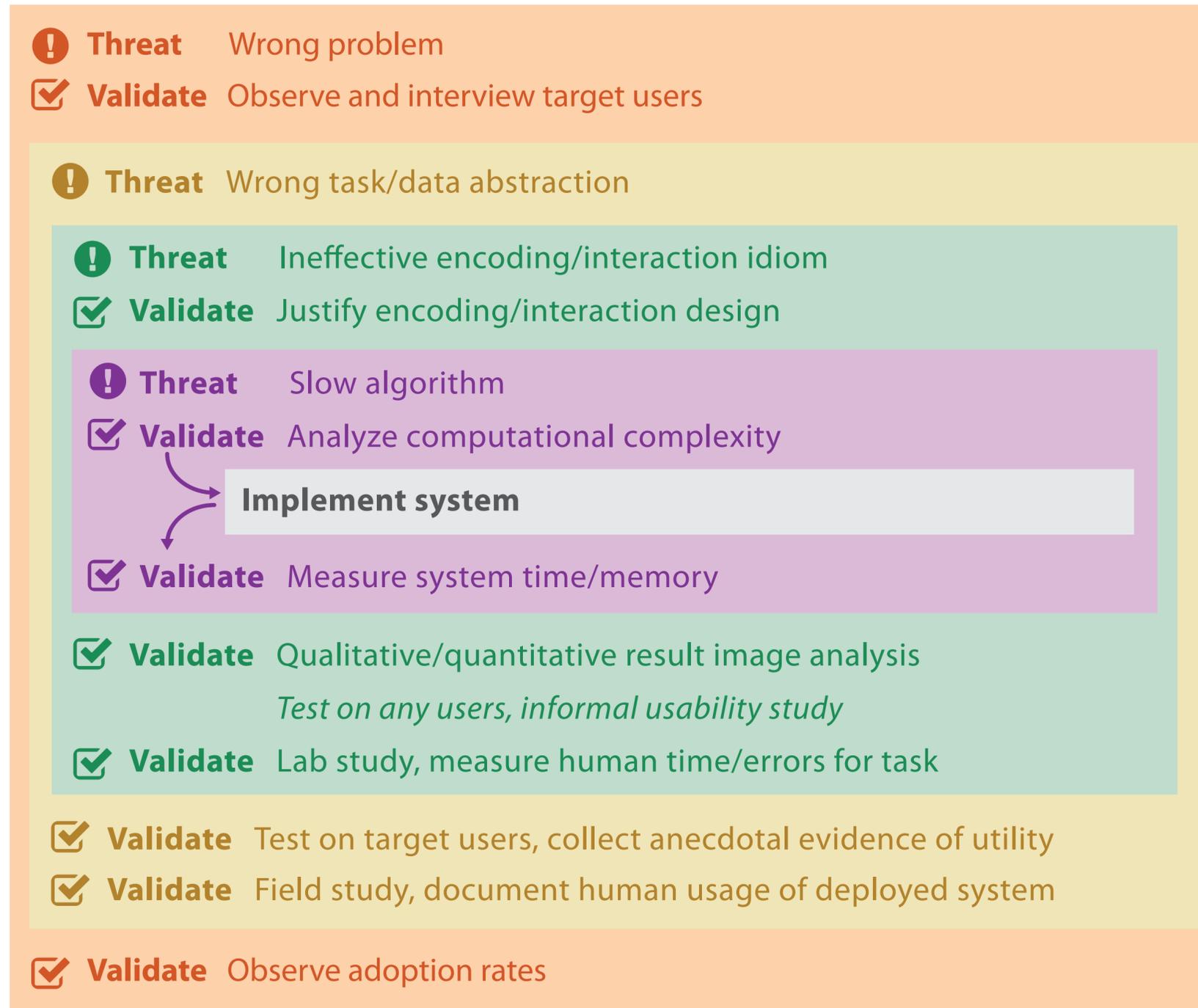
 **Data/task abstraction**
You're showing them the wrong thing

 **Visual encoding/interaction idiom**
The way you show it doesn't work

 **Algorithm**
Your code is too slow

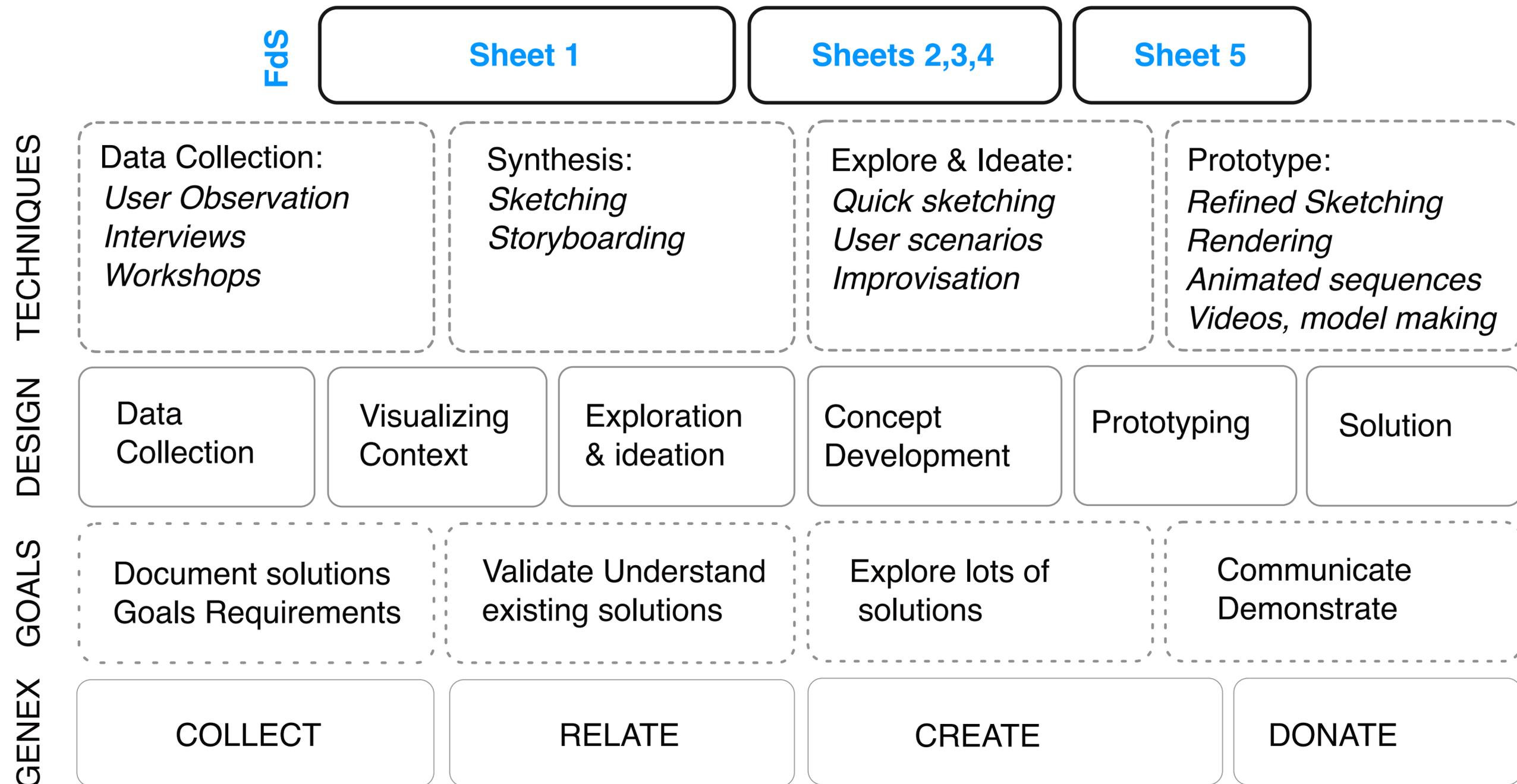
[Munzner, 2014]

Validation at each level



[Munzner, 2014]

Five Design-Sheet Methodology



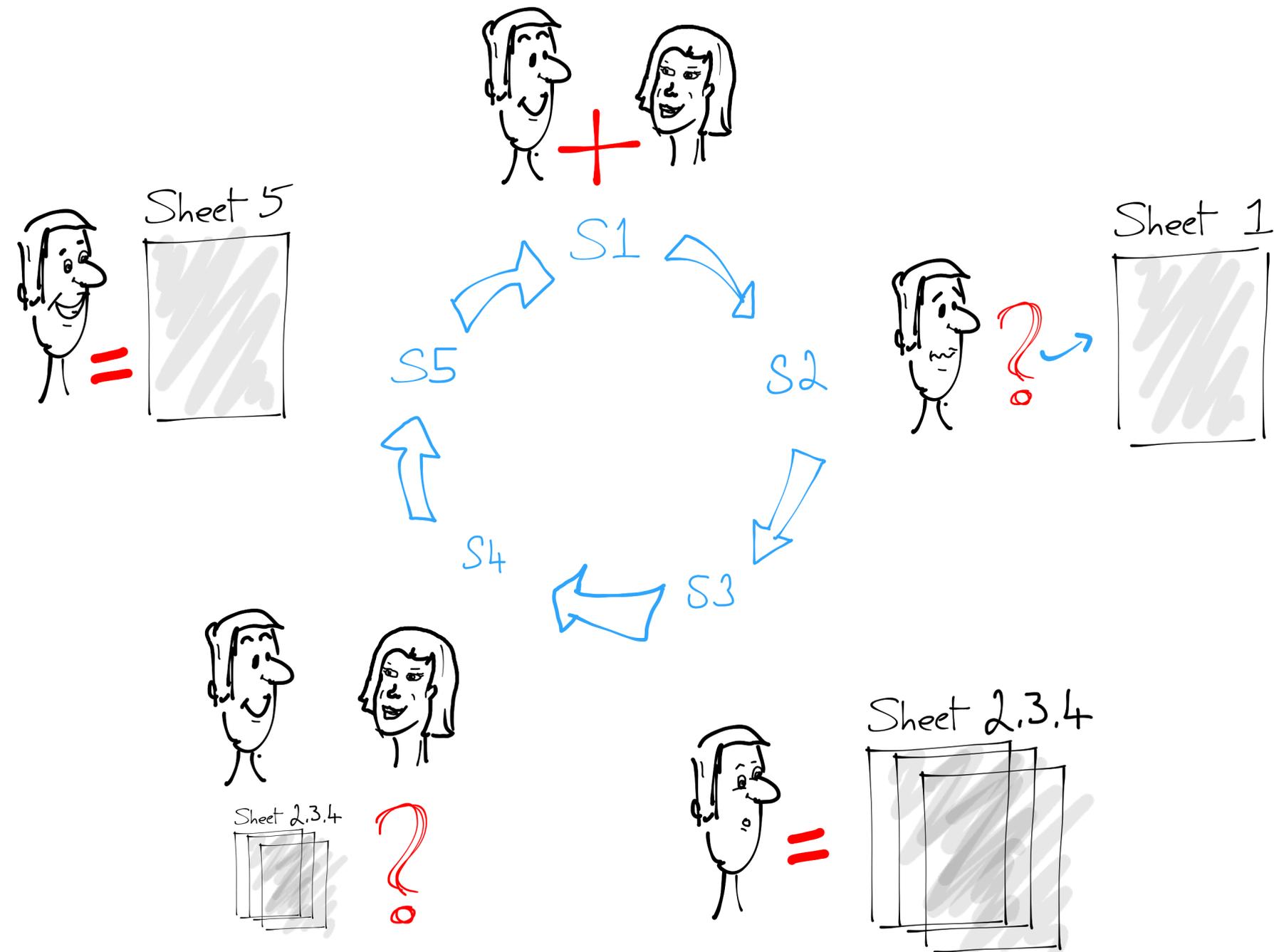
[J. Roberts et al., 2016]

Five Stages

1. Meet with client and consider task; or contemplate task on own.
2. Ideate and sketch small ideas.
3. Sketch and plan three alternative designs.
4. Consider solutions with client; or deliberate on own.
5. Generate realization sheet, and implement prototype. Discuss with client and re-iterate if necessary.

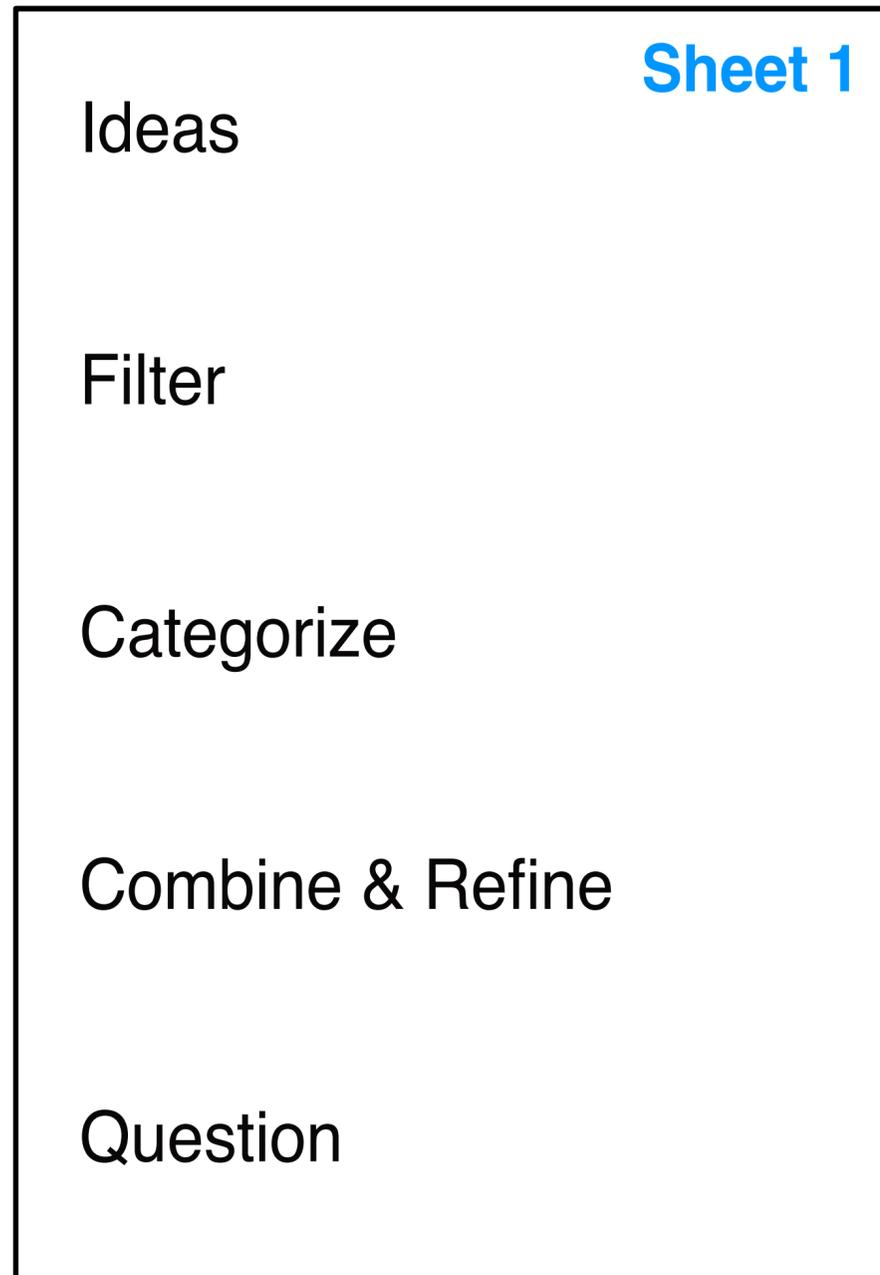
[J. Roberts et al., 2016]

Five Stages

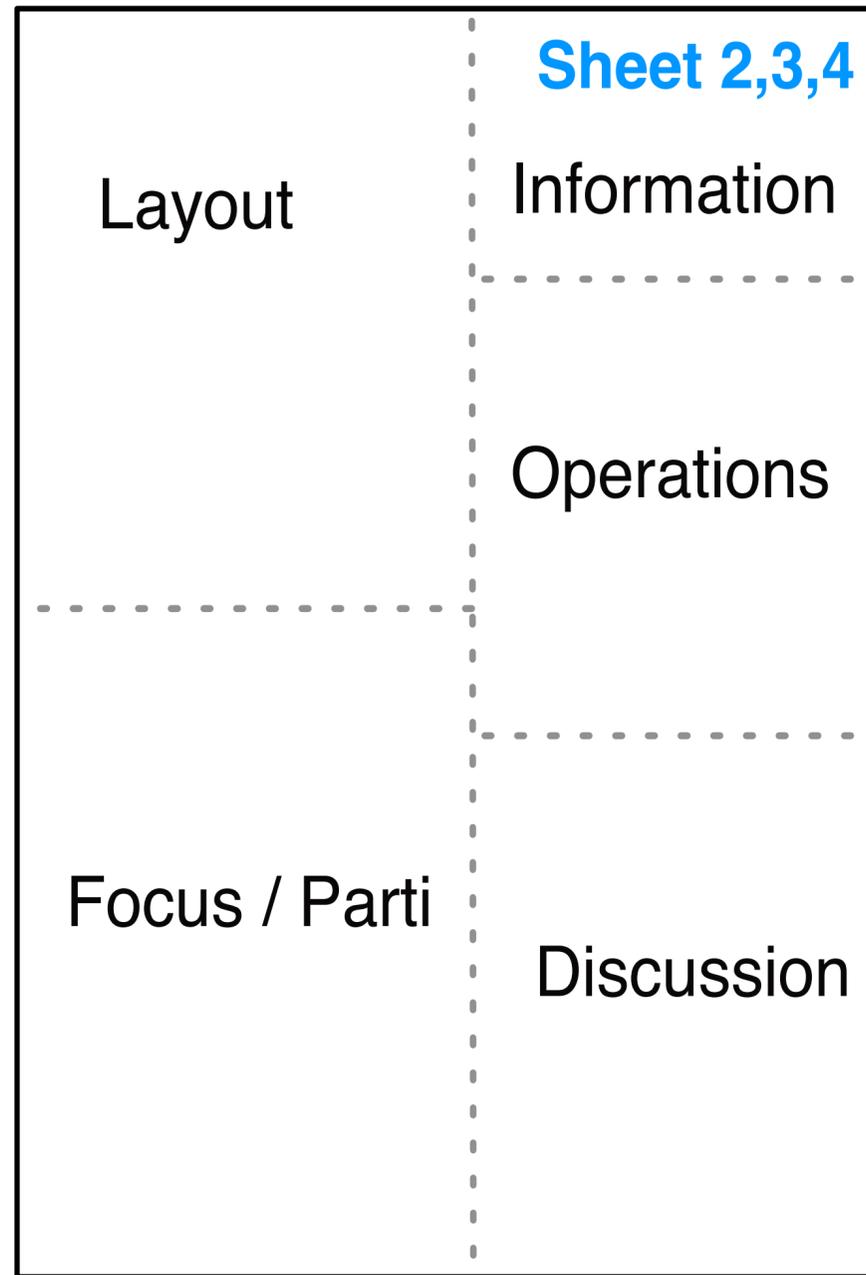


[J. Roberts et al., 2016]

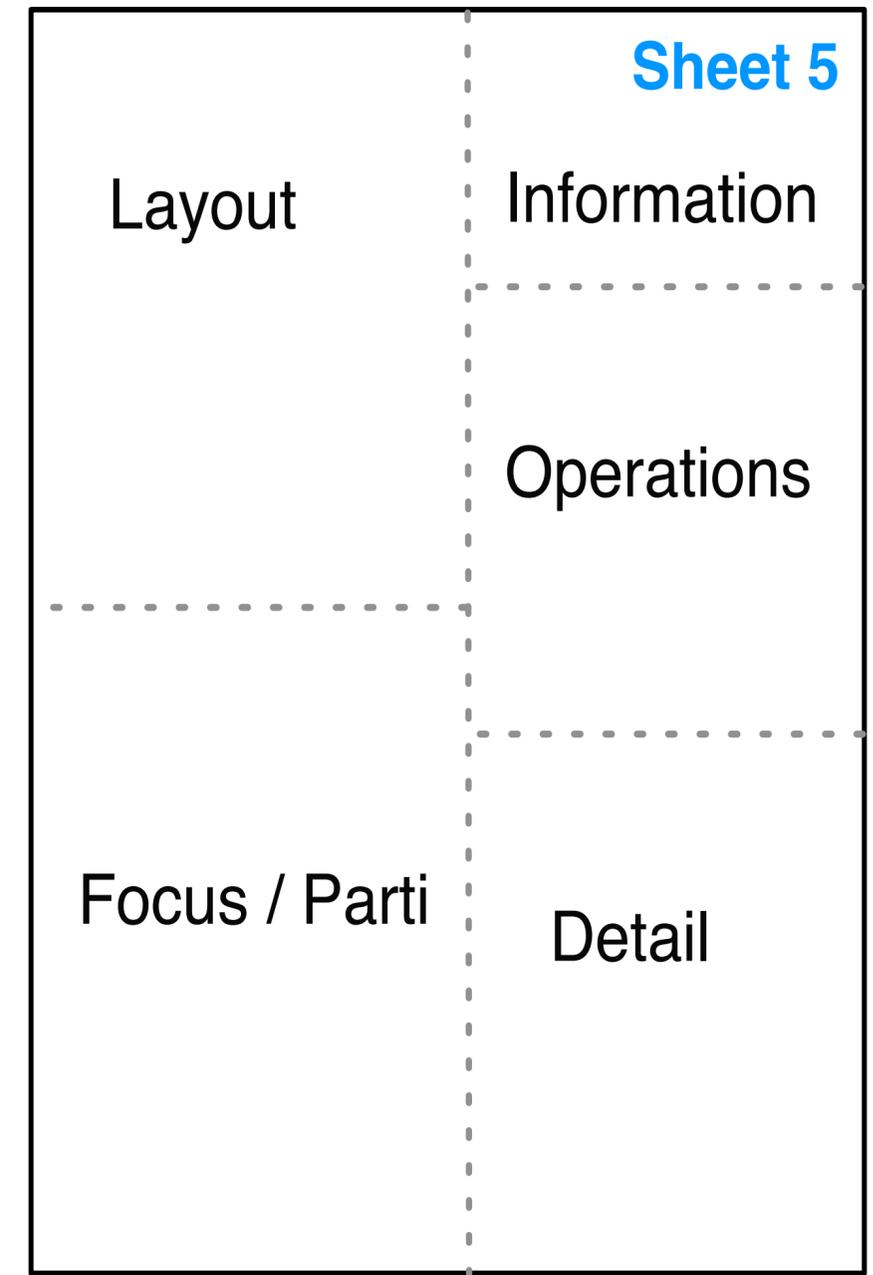
The Five Sheets



Ideation



Alternative Designs



Realization

[J. Roberts et al., 2016]

Sheets 2-4

TITLE: RELATIONSHIPS BETWEEN OSA & POST-HE EMPLOYABILITY
AUTHOR:
DATE: 16/11/2012
SHEET: 1 - FDS2
TASK: GLYPH REPRESENTATION OF THE DATA ON A MAP

OPERATIONS

- USER CAN SELECT A YEAR AND DISPLAY THAT DATA ON THE MAP
- USER CAN SELECT/EXCLUDE UNIVERSITIES FROM THE DATA
- DATA PRE-LOADED
- HOVER-OVER GLYPHS TO SHOW POP-UP WITH ADDITIONAL INFO ABOUT THE UNIVERSITY.
- GRADIENTS USED TO ADD AN ADDITIONAL VISUAL CUE OF DIFFERENCES BETWEEN UNIVERSITIES.

DISCUSSION

- IS THE MAP USEFUL?
 - QUADRANT MAP INSTANT?
 - DOES LOCATION DATA HAVE TO BE ACCURATE?
- AS WITH LAYOUT TOTAL SHOULD BE WITH THIS TOTALS GRAPHIC. THE TOTALS SHOULD BE MOVED AWAY FROM THE INTERACTIVE ELEMENTS.
- COULD MORE DATA BE GAINED BY CHANGING ON THE GLYPH?
- COLOUR... WHAT PALETTE WOULD MAKE AN APPROPRIATE MAP?
 - POS/NEG COLOURS (+/-)
 - COLOUR BLEND RANGE
 - DISAGILETY DATA SHOULD BE ACCESSIBLE TO ALL
 - DO ANY COLOURS MAKE EYEBALANCING EASIER?

OPERATIONS

- HOVERING OVER EACH AVERAGE UNIVERSITY BAR DRAWS THE YEAR BY YEAR BREAKDOWN.
- CLICKING ON AN INDIVIDUAL YEAR WOULD SET ALL BACKGROUND "MAIN" BARS TO THE FIGURES OF THAT YEAR ALONE FOR COMPARISON.
- CHECKING ON THE MINUS SYMBOL WOULD MINIMISE THAT UNE WITH THE REMAINING UNES EXPANDING TO FILL THE SPACE.
- DATA PRE-LOADED
- ABILITY TO CHANGE DISPLAY PALETTE COLOURS FOR DIFFERENT USERS PREFERENCES?

DISCUSSION

- DOES "MINIMISING" UNES ADD ANY USEFUL INTERACTIONS?
- DURING THE HOVER-OVER BREAKDOWN OF DATA SHOULD THE YEAR BE DISPLAYED IN TEXT FOR EASE OF INTERPRETATION?
- SHOW ALL DATA COULD BE VIEWED BUT IT MAY CLUTTER THE DATA.
- VISUALISATION ASSUMES THE ONLY CORRELATIONS WILL BE NUMERICAL.
- COULD OTHER DATA BE DISPLAYED ON THE BARS? PERCENTAGES OR ORDERED CLASSIFICATION BY TYPE?

TITLE: RELATIONSHIPS BETWEEN OSA & POST-HE EMPLOYABILITY
AUTHOR:
DATE: 16/11/2012
SHEET: 2 - FDS3
TASK: BAR-CHART REPRESENTATION OF THE OSA EMPLOYABILITY & LEAVERS DATA

OPERATIONS

- HOVERING OVER THE GRAPH BRUSHED THAT YEAR'S DATA AND DISPLAYS IT NUMERICALLY TOP RIGHT.
- UNIVERSITIES CAN BE EXCLUDED FROM THE VISUALISATION
- DATA-SET PRE-LOADED
- POSSIBLE? - CLICK ON THE FOCUSED / BRUSHED UNI-YEAR AND A NEW DRAWING IS LOADED WITH ADDITIONAL DATA NOT ABOUT THE UNI.
- CLICKABLE TAB ON LINE GRAPH (OSAs) ALLOWS LINE GRAPH TO BE DRAGGED OVER EMPLOYABILITY CHART FOR DIRECT COMPARISON OR SCID AWAY IF REQUIRED TO AVOID CLUTTER

DISCUSSION

- THIS VISUALISATION TREATS EACH INDIVIDUAL YEAR AS A SEPARATE DATA POINT GROUPED BY YEAR.
- EACH DATA-SET IS PLOTTED WITH ITS OWN SCALE RELEVANT TO THAT DATA. IMPORTANT TO CHOOSE RANGE AS WE ARE WORKING IN PERCENTAGE OF PERCENTAGES.
- DATA CAN BE SEEN IN ORIGINAL NUMERICAL FORMAT
- VISUALISATION ASSUMES CORRELATIONS WILL ONLY EXIST NUMERICALLY NOT BY LOCATION OR YEAR.
- THIS VISUALISATION MAY LIMIT THE EMPLOYMENT PROPERTIES THAT COULD BECOME APPARENT DUE TO ITS 2D NATURE

TITLE: RELATIONSHIPS BETWEEN OSA & POST-HE EMPLOYABILITY
AUTHOR:
DATE: 16/11/2012
SHEET: 3 - FDS4
TASK: PARALLEL COORDINATE PLOT REPRESENTATION OF THE DATA WITH EACH YEAR AS A SEPARATE POINT ALONG THE X-AXIS

OPERATIONS

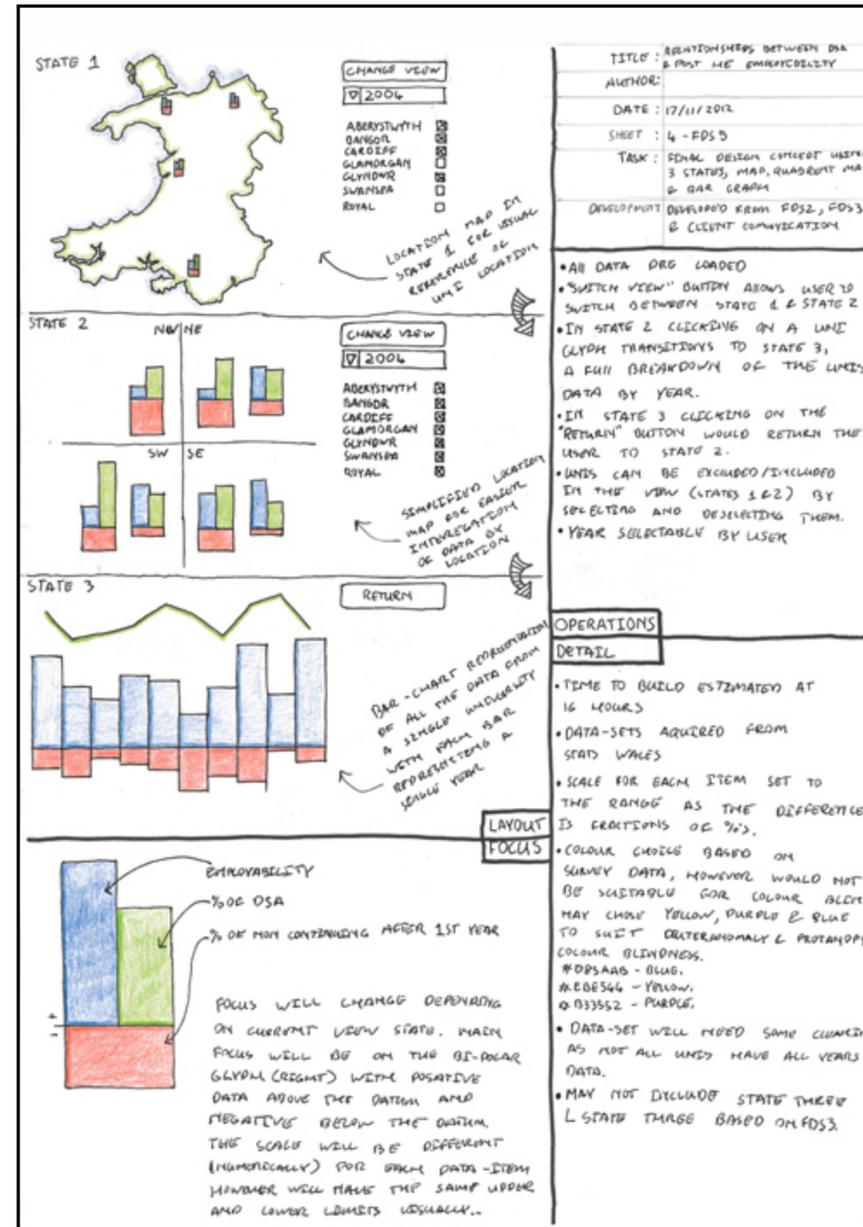
- HOVERING OVER THE LINE GRAPH (OSAs) ALLOWS LINE GRAPH TO BE DRAGGED OVER EMPLOYABILITY CHART FOR DIRECT COMPARISON OR SCID AWAY IF REQUIRED TO AVOID CLUTTER

DISCUSSION

- THIS VISUALISATION TREATS EACH INDIVIDUAL YEAR AS A SEPARATE DATA POINT GROUPED BY YEAR.
- EACH DATA-SET IS PLOTTED WITH ITS OWN SCALE RELEVANT TO THAT DATA. IMPORTANT TO CHOOSE RANGE AS WE ARE WORKING IN PERCENTAGE OF PERCENTAGES.
- DATA CAN BE SEEN IN ORIGINAL NUMERICAL FORMAT
- VISUALISATION ASSUMES CORRELATIONS WILL ONLY EXIST NUMERICALLY NOT BY LOCATION OR YEAR.
- THIS VISUALISATION MAY LIMIT THE EMPLOYMENT PROPERTIES THAT COULD BECOME APPARENT DUE TO ITS 2D NATURE

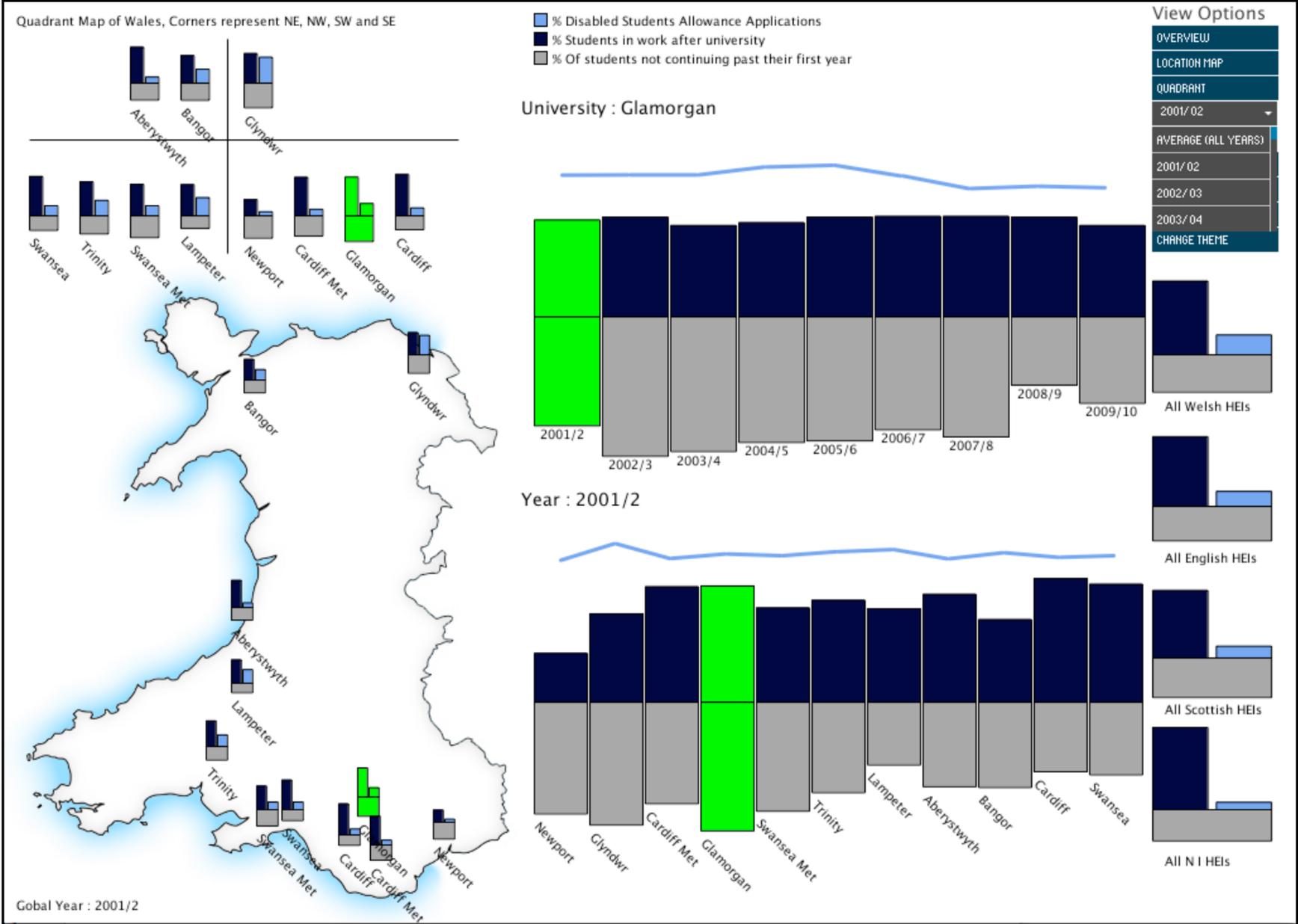
[J. Roberts et al., 2016]

Sheet 5



[J. Roberts et al., 2016]

Prototype



[J. Roberts et al., 2016]