

Information Visualization

SWE 632

Fall 2021



Administrivia

- HW6 due today
- Last HW (HW7) due next week
- Research Study out next week
- Final project presentation in 2 weeks (11/30)

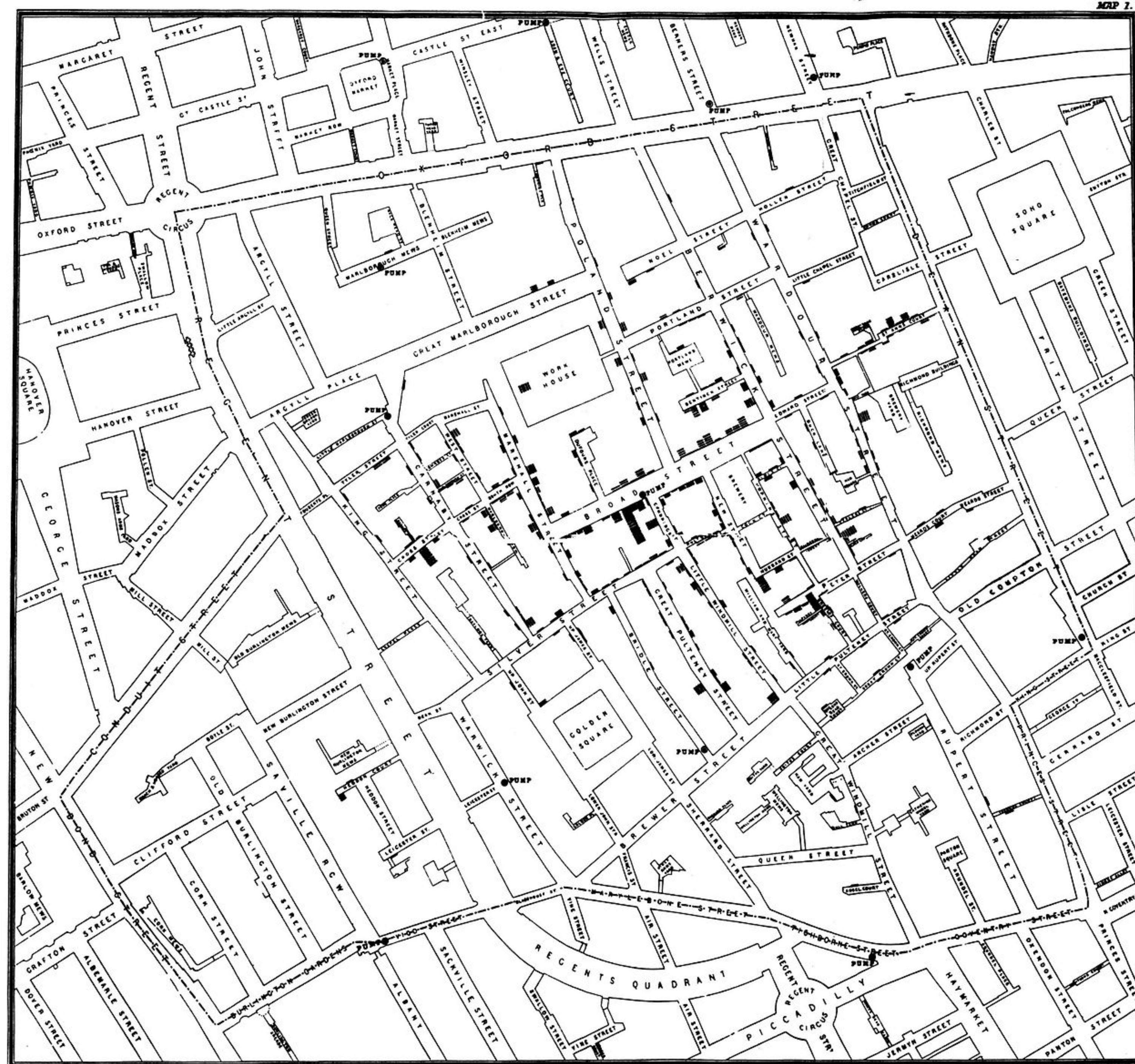
Overview of Information Visualization

Today

- What types of information visualization are there?
 - Which one should you choose?
- What principles and guidelines inform the design of information visualizations?
- How can interactivity be used to design better information visualizations?

Cholera Epidemic in London, 1854

- >500 fatal attacks of cholera in 10 days
 - Concentrated in Broad Street area of London
 - Many died in a few hours
- Dominant theory of disease: caused by noxious odors
- Afflicted streets deserted by >75% inhabitants



Investigation and Aftermath

- Based on visualization, did case by case investigation
- Found that 61 / 83 positive identified as using well water from Broad Street pump
- Board ordered pump-handle to be removed from well
- Epidemic soon ended
- Solved centuries old question of how cholera spread

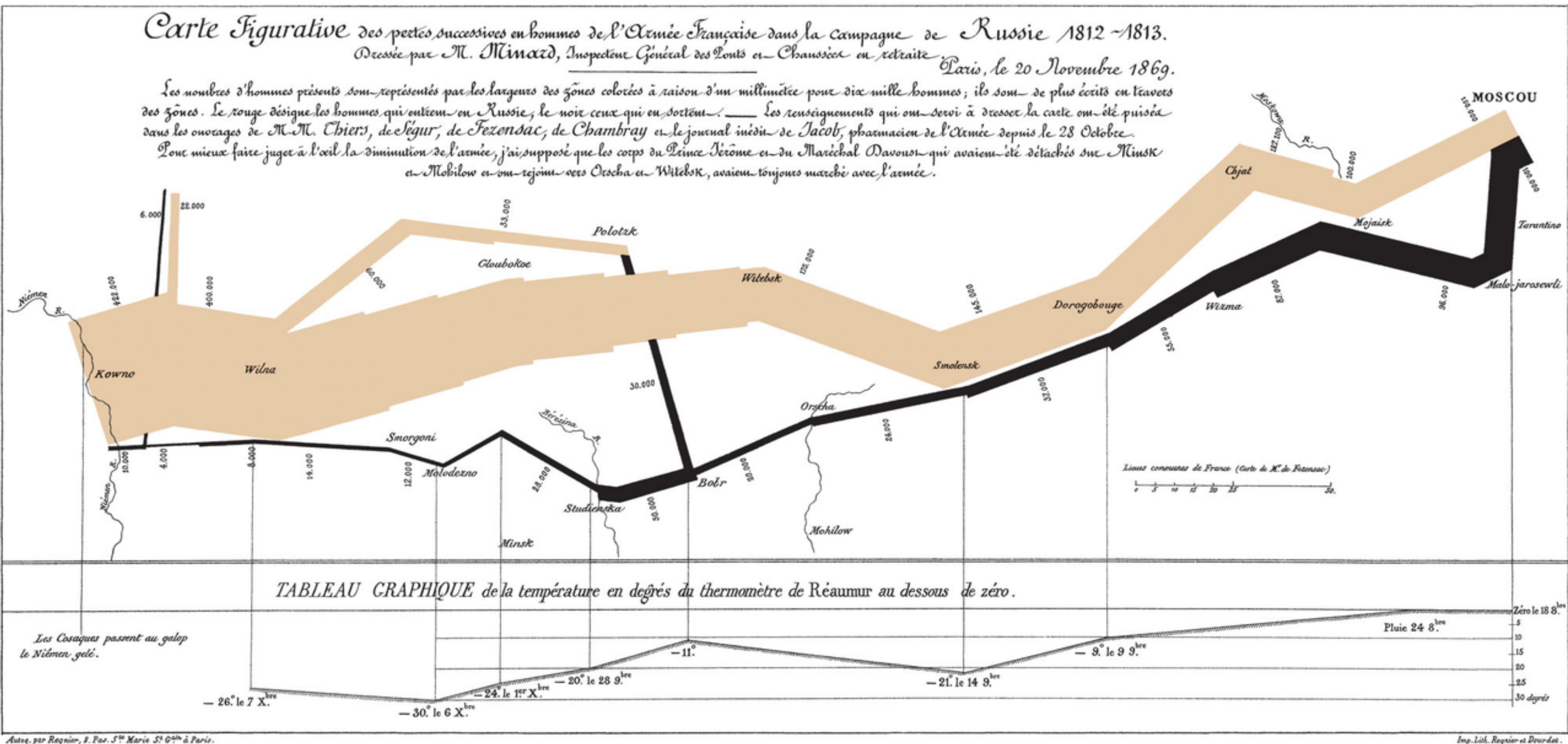
Methods used by Snow

- Placed data in appropriate context for assessing cause & effect
 - Plotted on map, included well location
 - Reveals proximity as cause
- Made quantitative comparisons
 - Fewer deaths closer to brewery, could investigate cause
- Considered alternative explanations & contrary cases
 - Investigated cases not close to pump, often found connection to pump
- Assessment of possible errors in numbers

Amplifying Cognition

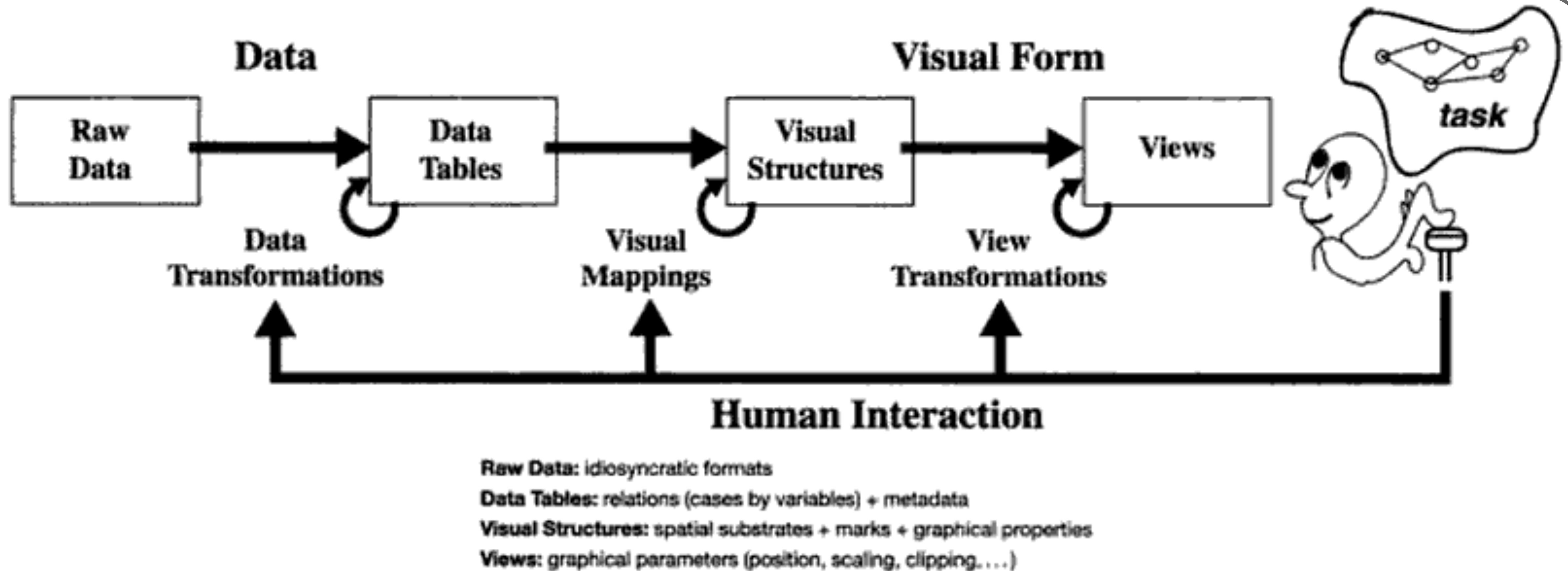
- Information Visualization can amplify cognition by:
 1. *Increasing the memory and processing resources available to users*
 2. *Reducing the search for information*
 3. *Using visual representations to enhance the detection of patterns*
 4. *Enabling perceptual inference*
 5. *Using perceptual attention mechanisms for monitoring*
 6. *Encoding Information in a manipulable medium*

Charles Minard's Map of Napoleon's Russian Campaign of 1812



Mapping Data to Visual Form

Designing an Information Visualization



Types of Raw Data

- Nominal - unordered set without a quantitative value
 - Gender: male, female
 - Hair color: brown, black, blonde, gray, orange, ...
- Ordinal - ordered set, with no meaning assigned to differences
 - How do you feel today: very unhappy, unhappy, ok, happy, very happy
 - Undefined how much better happy is than ok
- Quantitative - numeric value
 - Height, weight, distance, ...

Data Transformations

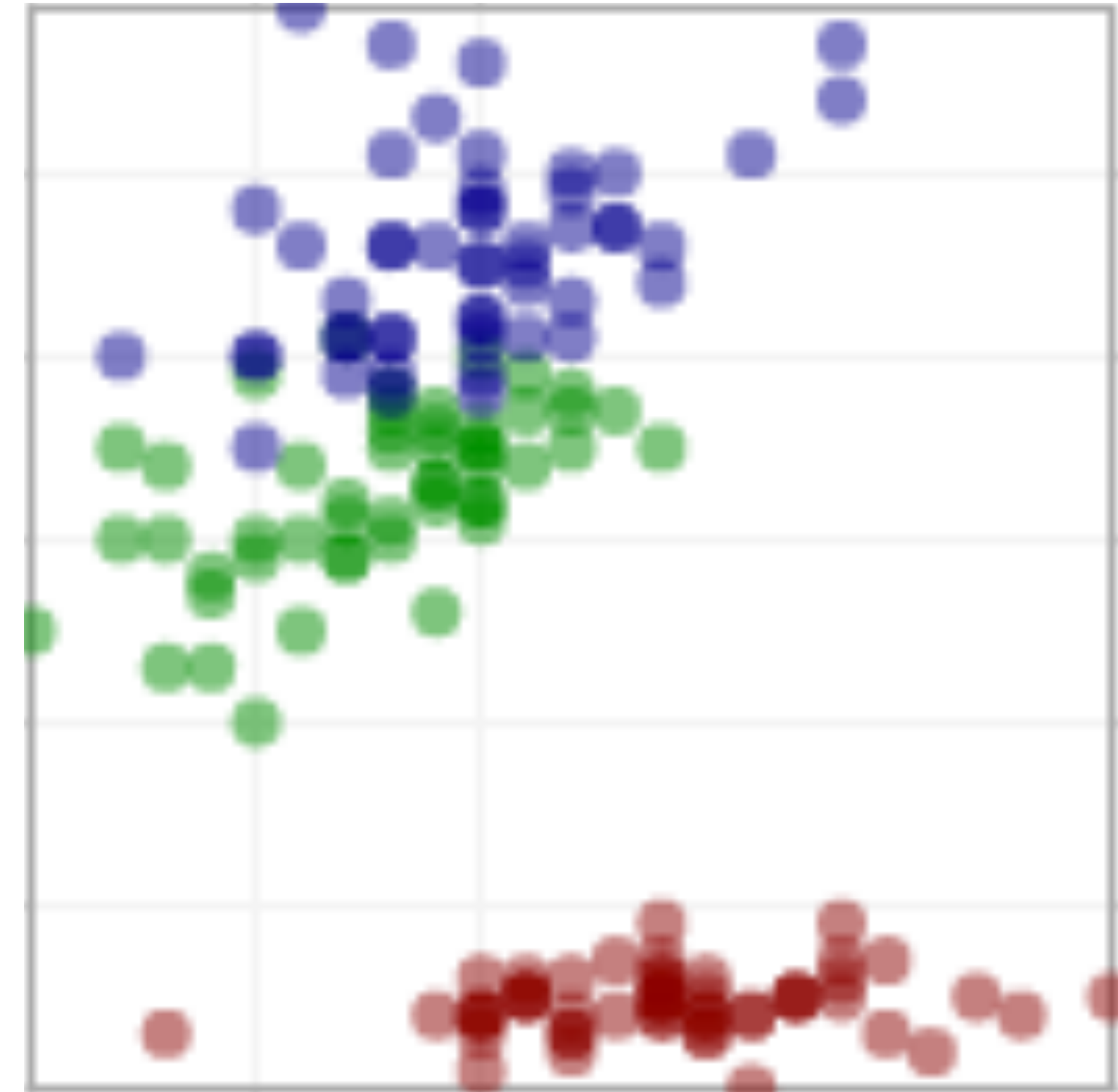
- Classing / binning: Quantitative $\rightarrow$ ordinal
 - Maps ranges onto classes of variables
 - Can also count # of items in each class w/ histogram
- Sorting: Nominal $\rightarrow$ ordinal
 - Add order between items in sets
- Descriptive statistics: mean, average, median, max, min, ...

Visual Structures

- 3 components
 - spatial substrate
 - marks
 - marks' graphical properties

Spatial Substrate

- Axes that divide space
- Types of axes - unstructured, nominal, ordinal, quantitative
- Composition - use of multiple orthogonal axes (e.g., 2D scatterplot, 3D)



Marks

- Points (0D)
- Lines (1D)
- Areas (2D)
- Volumes (3D)

Marks' Graphical Properties

Spatial		Object	
Extent	(Position) — — —	Gray Scale ■ ■ ■ □	
	Size ● ● ● ●		
Dif-feren-tial	Orientation — / \	Color ■ ■ ■ ■	
		Texture ■ ■ ■ ■	
		Shape ■ ★ ● ◆	

- Quantitative (Q), Ordinal (O), Nominal (N)
- Filled circle - good; open circle - bad

Effectiveness of Graphical Properties

		Spatial			Object			
		Q	O	N	Q	O	N	
Extent	(Position)	●	●	●	Grayscale	◐	●	○
	Size	●	●	●				
Differential	Orientation	◐	◐	●	Color	◐	◐	●
					Texture	◐	◐	●
					Shape	○	○	●

- Quantitative (Q), Ordinal (O), Nominal (N)
- Filled circle - good; open circle - bad

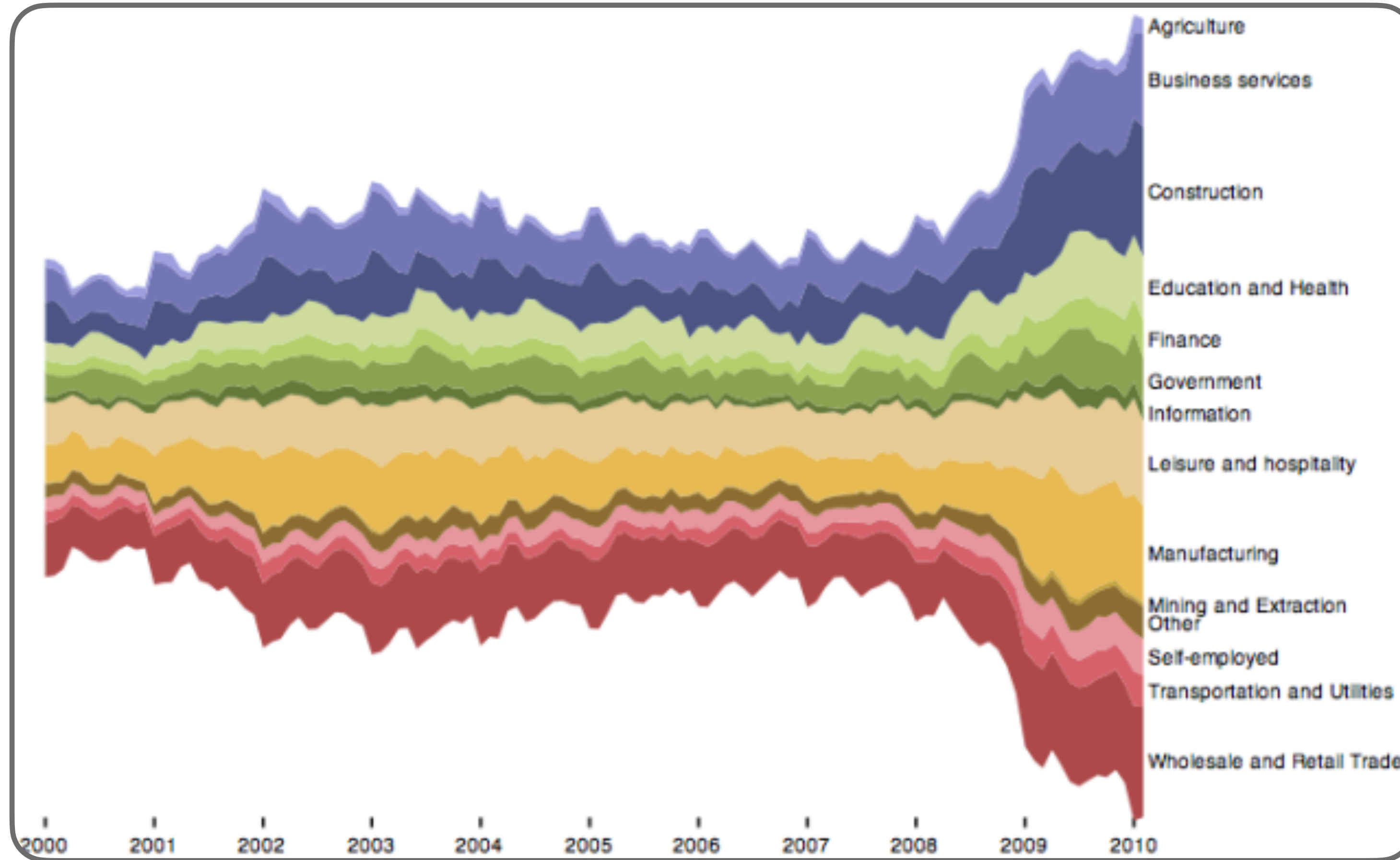
Animation

- Visualization can change over time
- Could be used to encode data as a function of time
 - But often not effective as makes direct comparisons hard
- Can be more effective to animate transition from before to after as user configures visualization

Examples of Visualizations

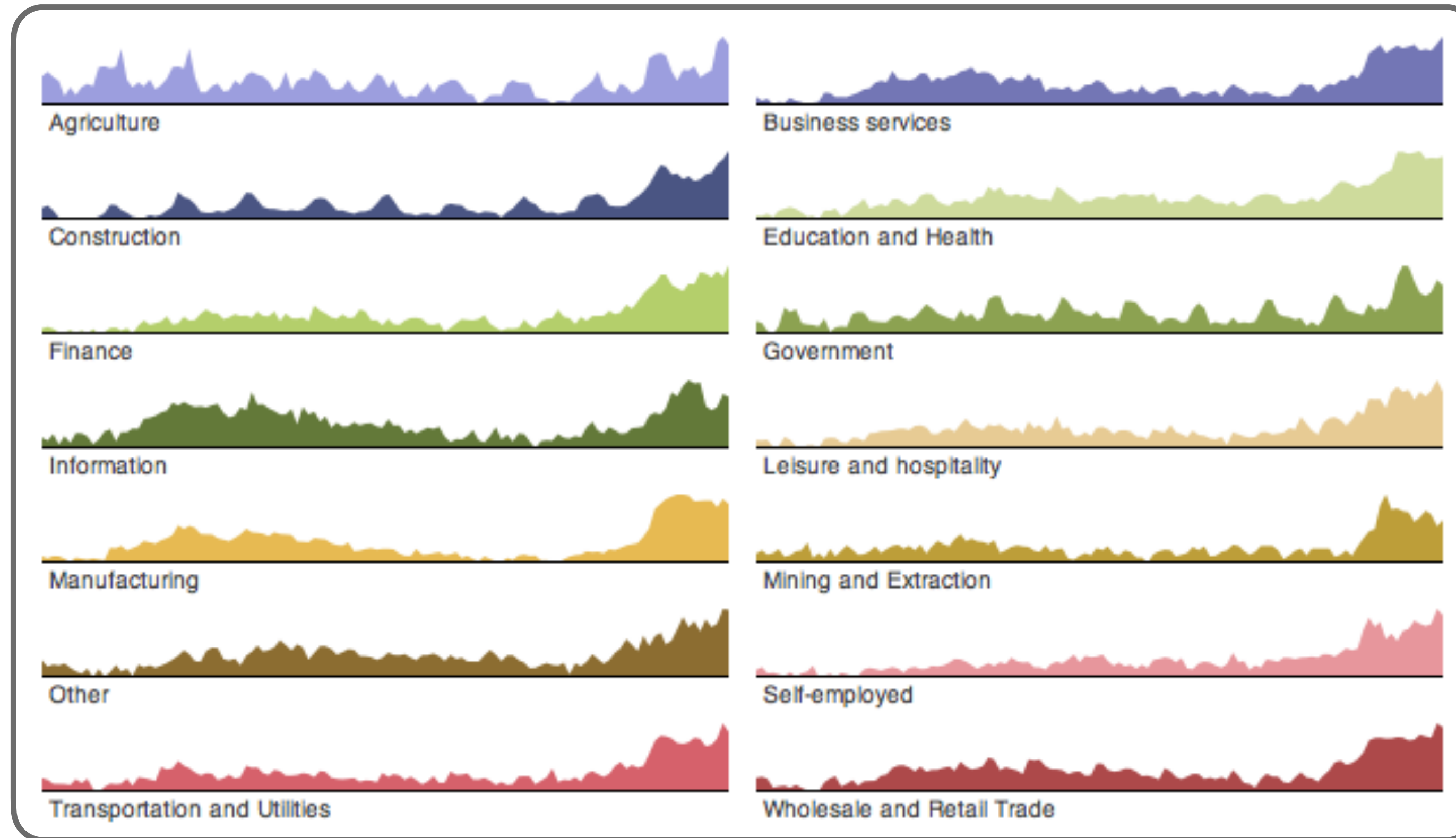
Time-series Data

Stacked Graph



- Supports visual summation of multiple components

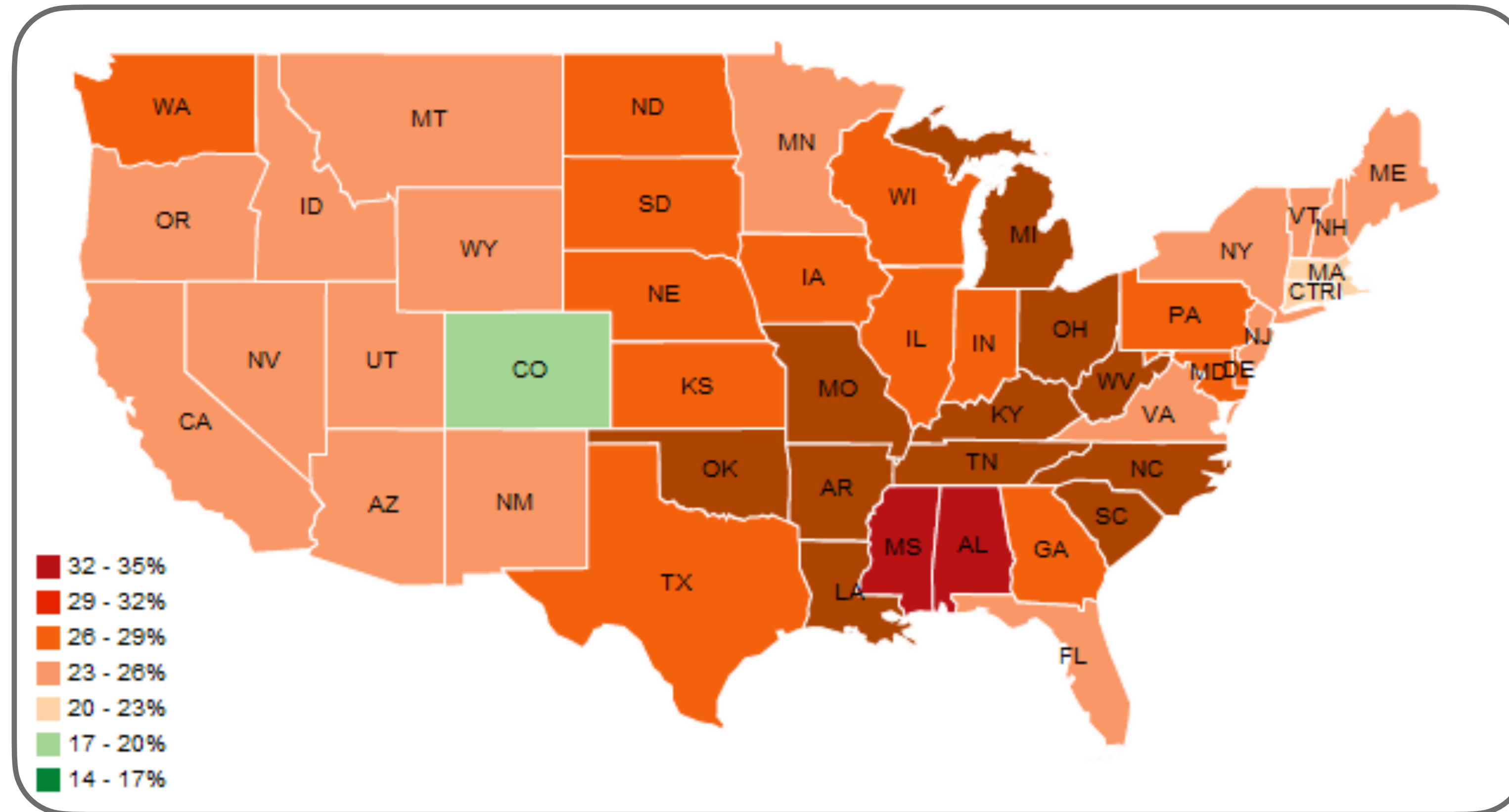
Small Multiples



- Supports separate comparison of data series
- May have better legibility than placing all in single plot

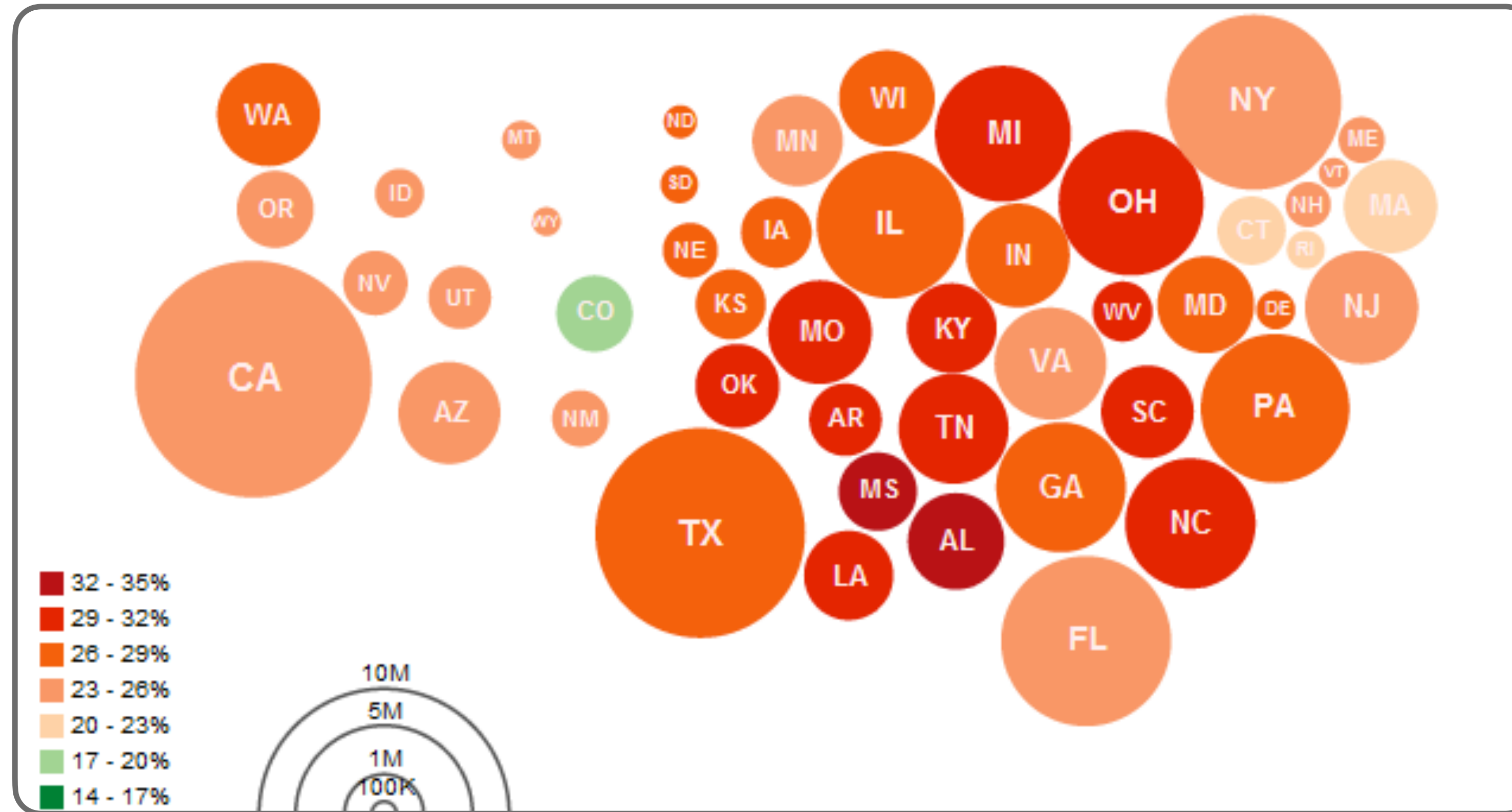
Maps

Choropleth Map



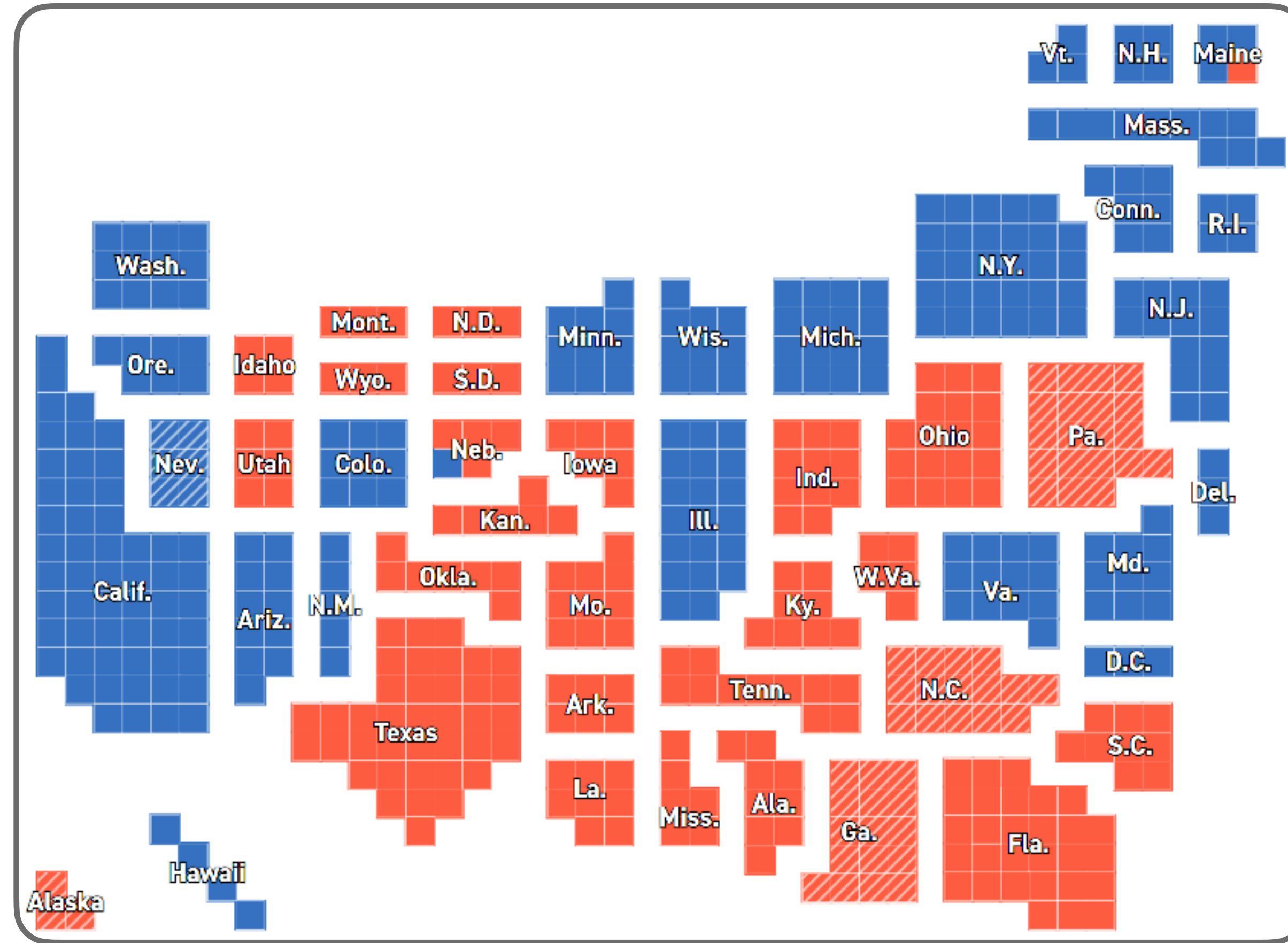
- Groups data by area, maps to color

Cartograms



- Encodes two variables w/ size & color

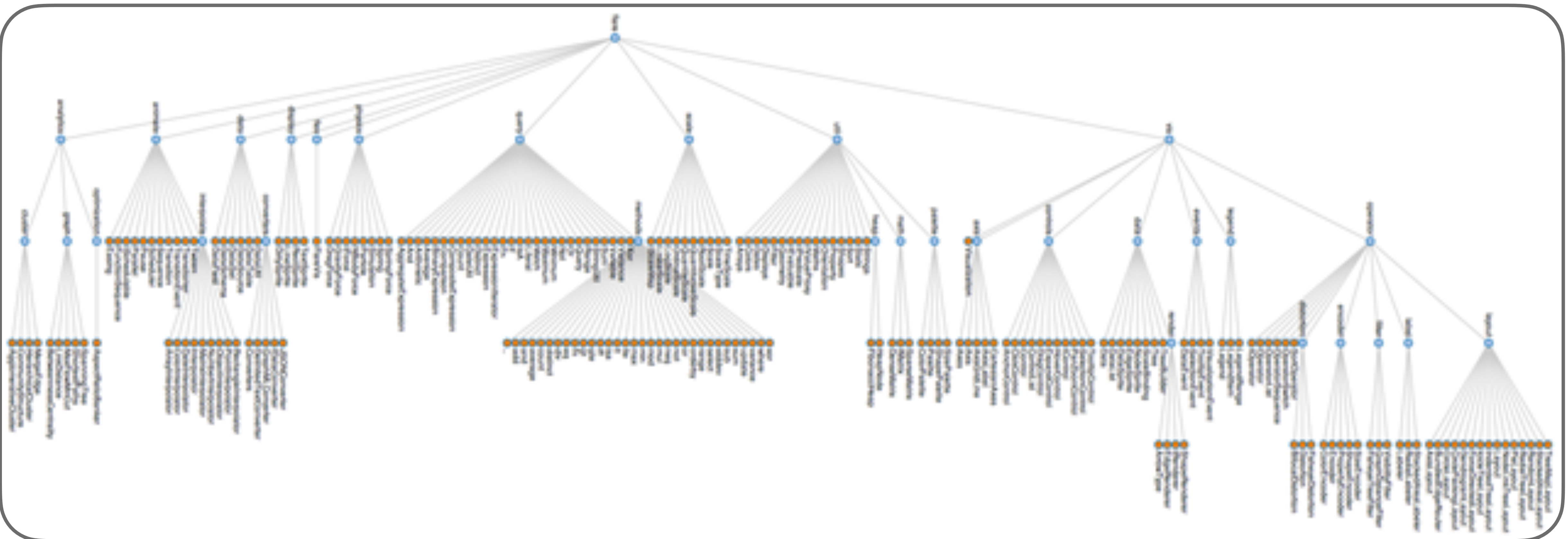
Cartograms



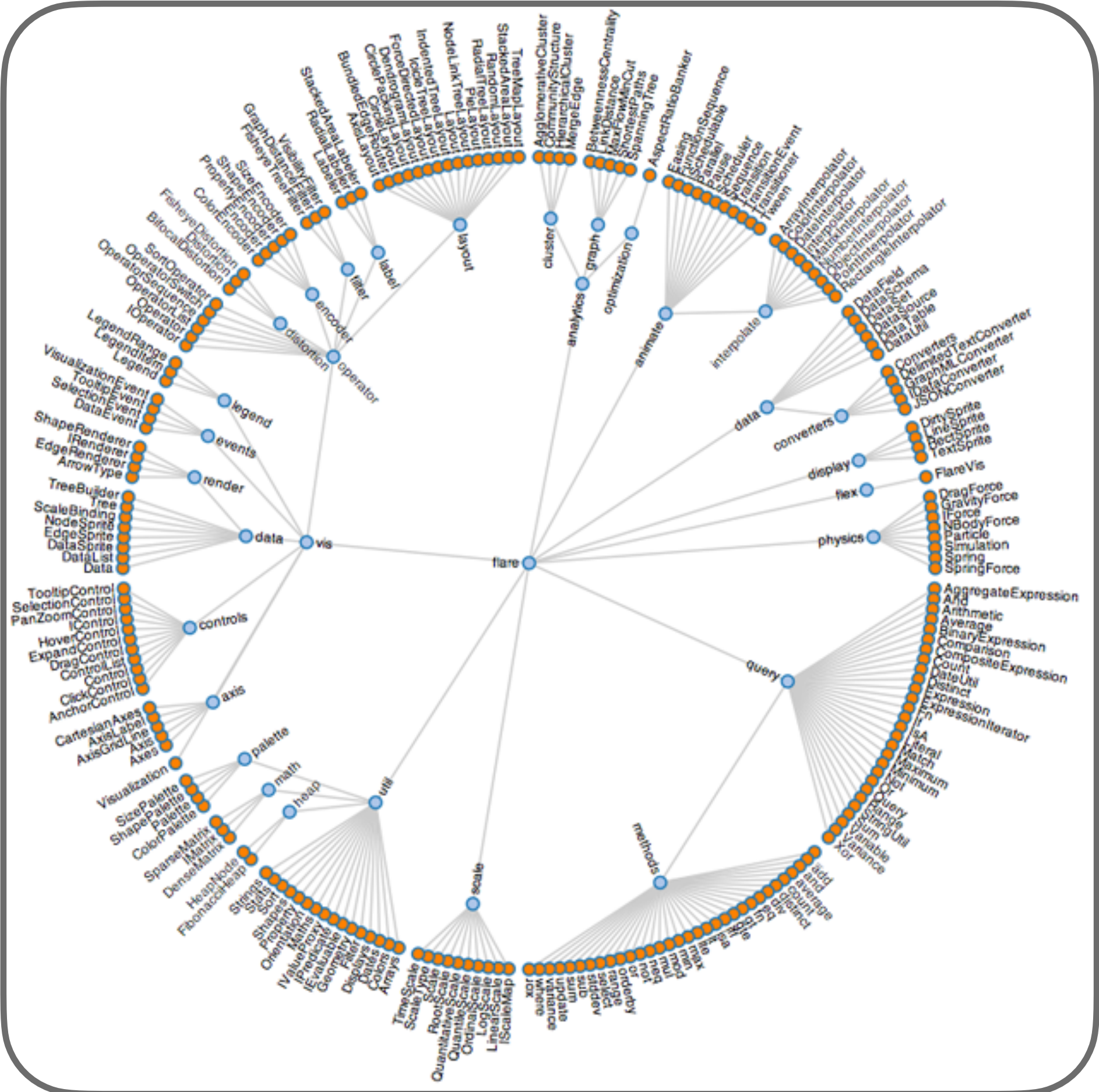
- Encodes two variables w/ size & color

Hierarchies

Node Link Diagram

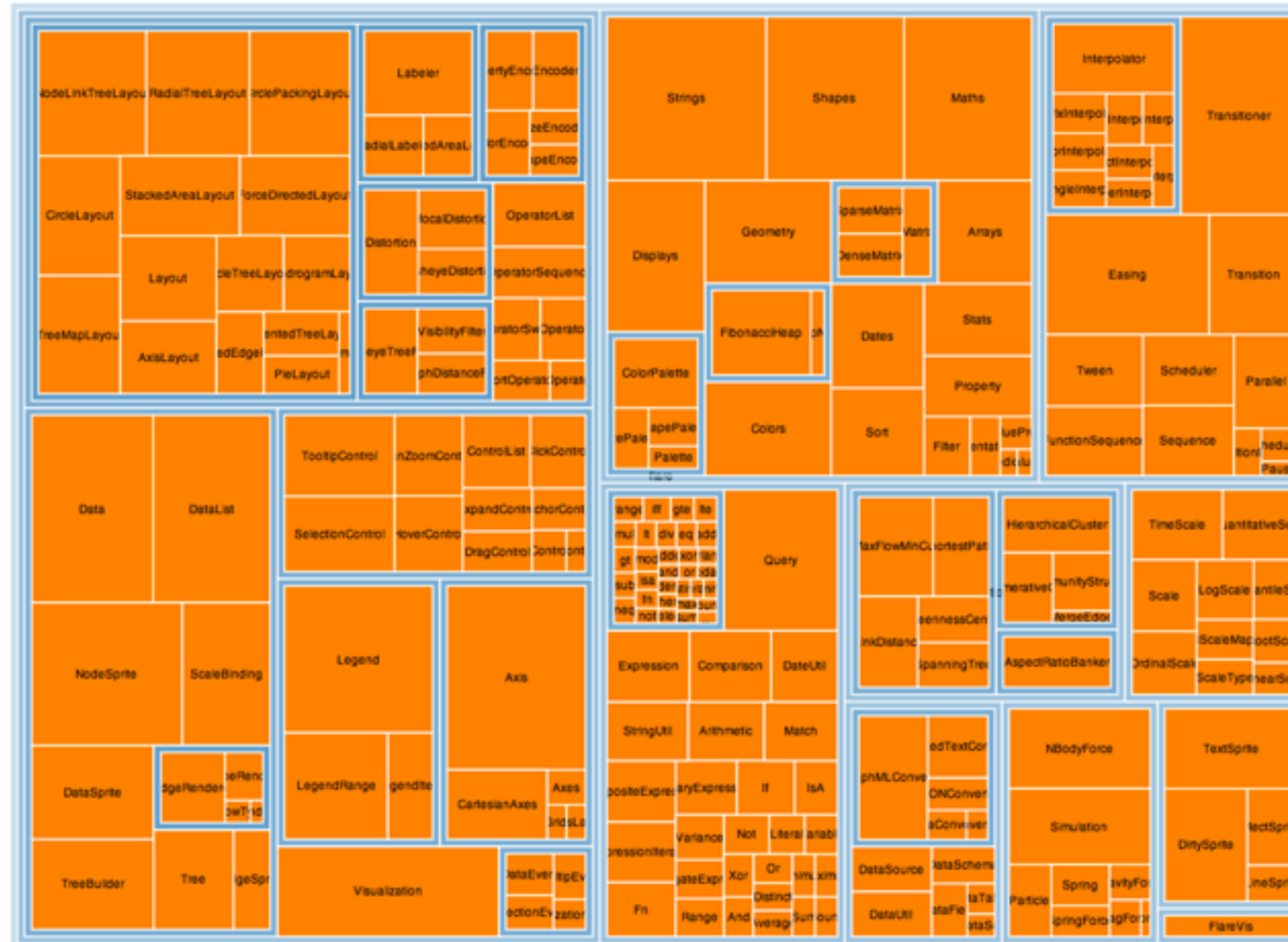


Dendrogram

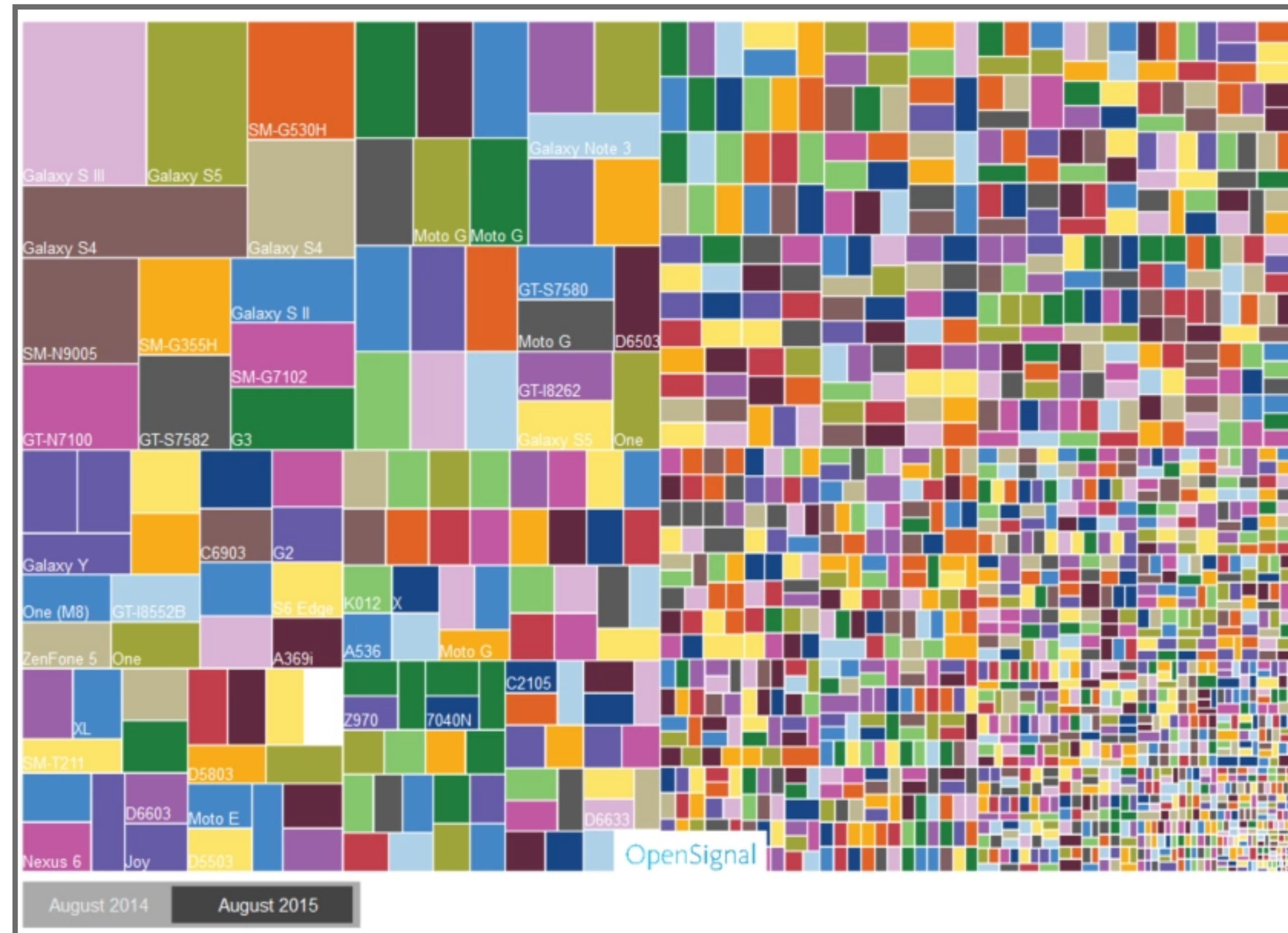


- Leaf nodes of hierarchy on edges of circle

Treemaps

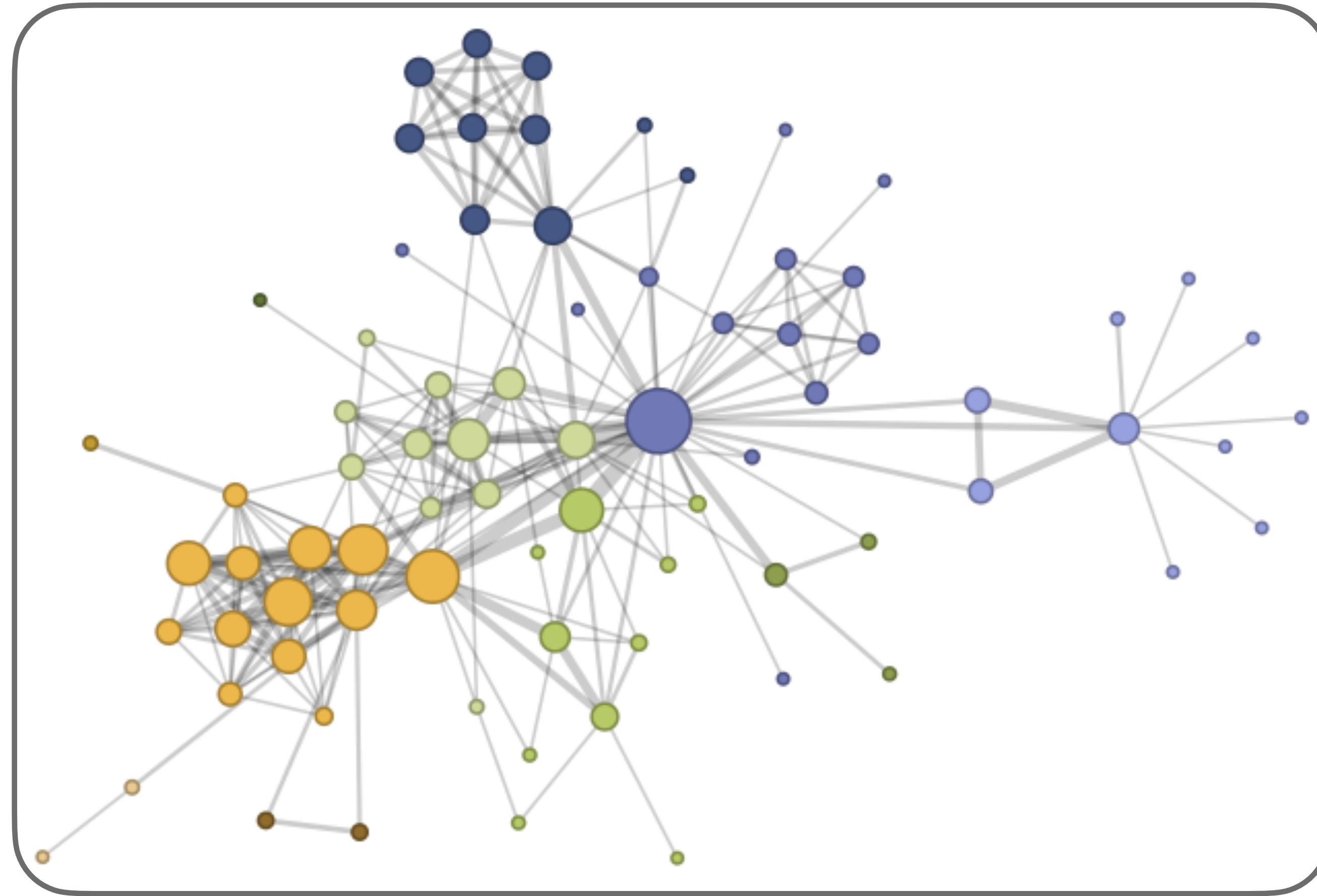


Treemaps



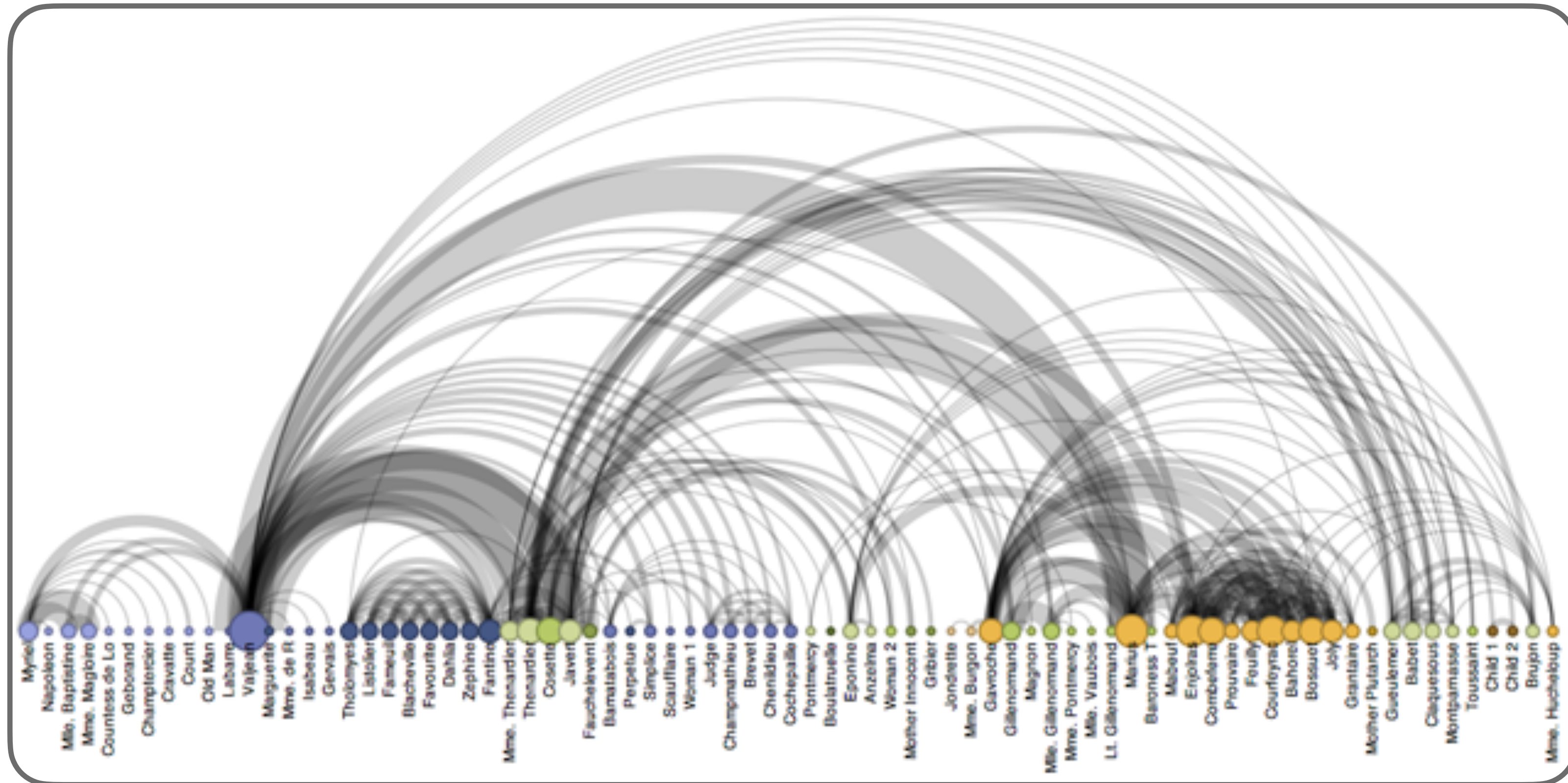
Networks

Force-directed Layout



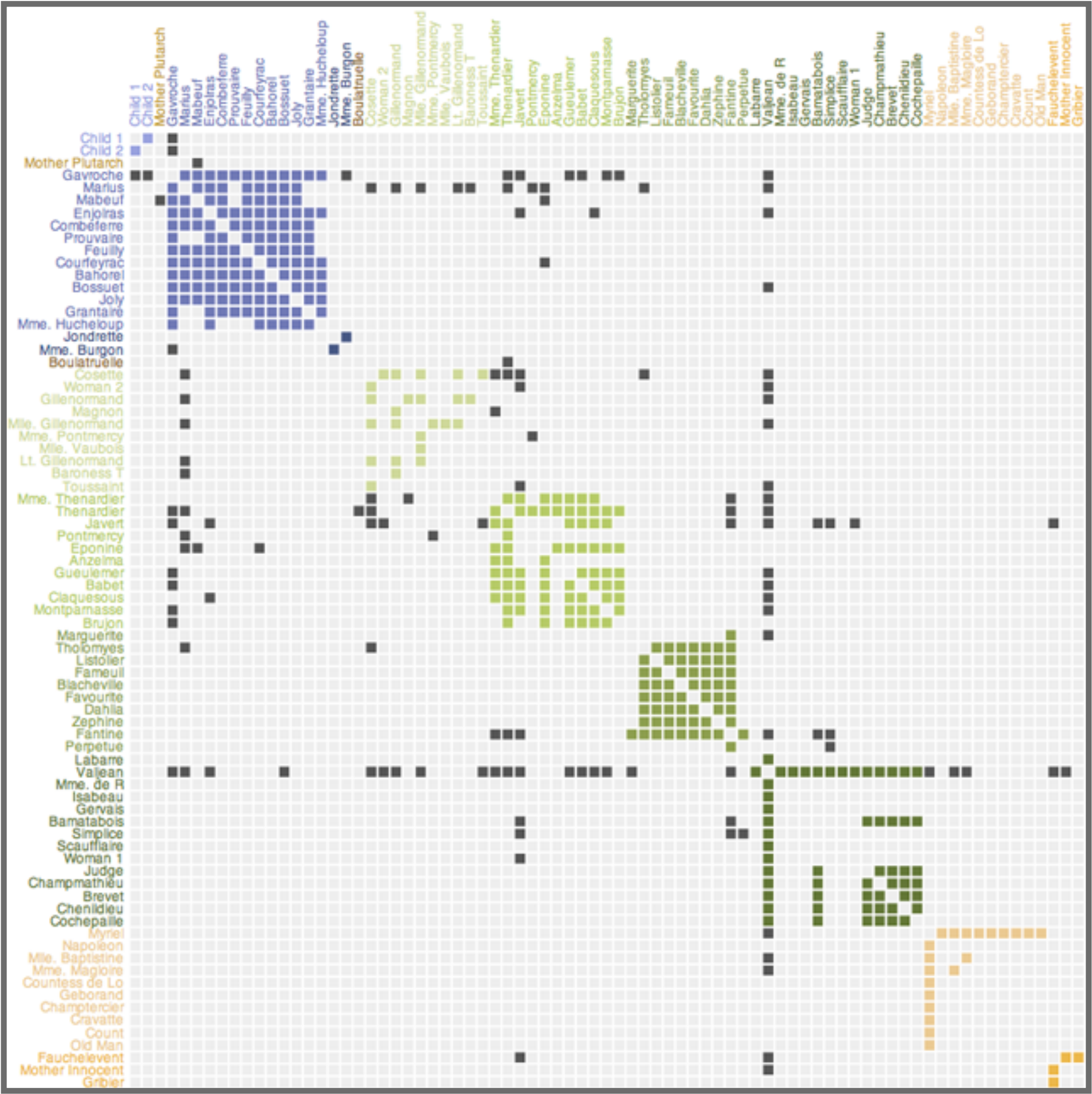
- Edges function as springs, find least energy configuration

Arc Diagram



- Can support identifying cliques & bridges w/ right order

Adjacency Matrix



Design Considerations

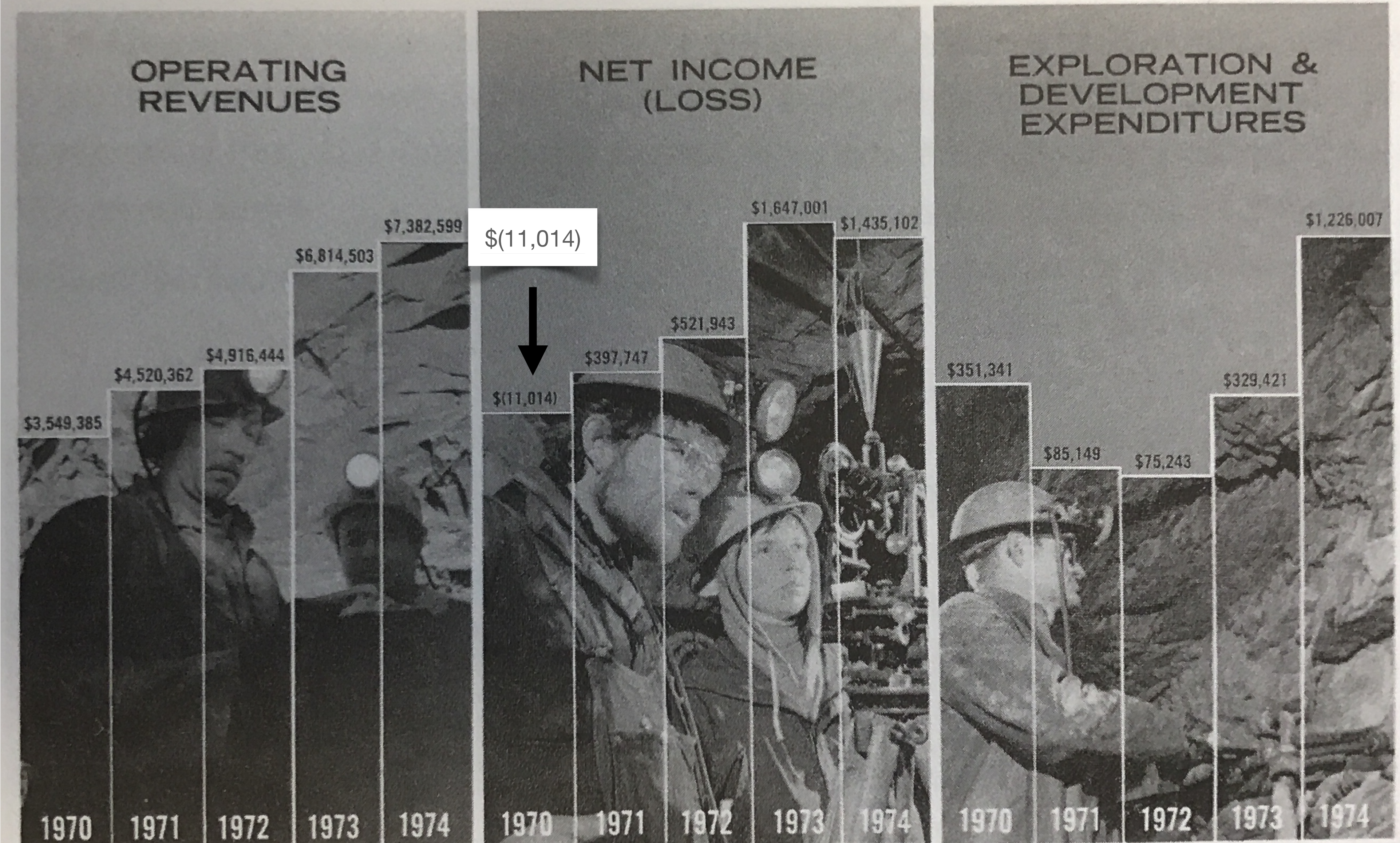
Tufte's principles of graphical excellence

- Show the data
- Induce the viewer to think about the substance rather than the methodology
- Avoid distorting what the data have to say
- Present many numbers in a small space
- Make large data sets coherent
- Encourage the eye to compare different pieces of data
- Reveal data at several levels of detail, from overview to fine structure
- Serve reasonable clear purpose: description, exploration, tabulation, decoration

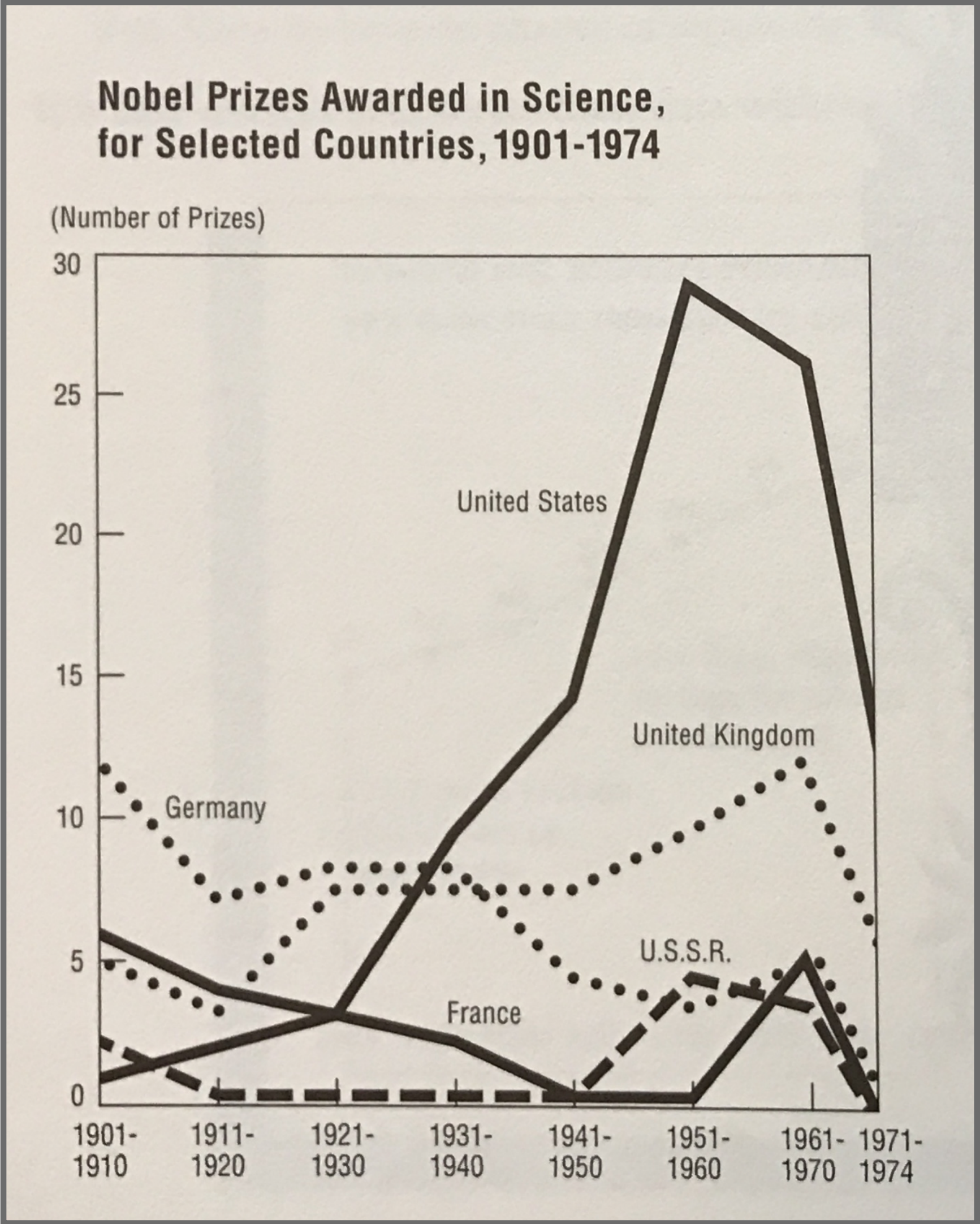
Distortions in Visualizations

- Visualizations may distort the underlying data, making it harder for reader to understand truth
- Use of *design* variation to try to falsely communicate *data* variation

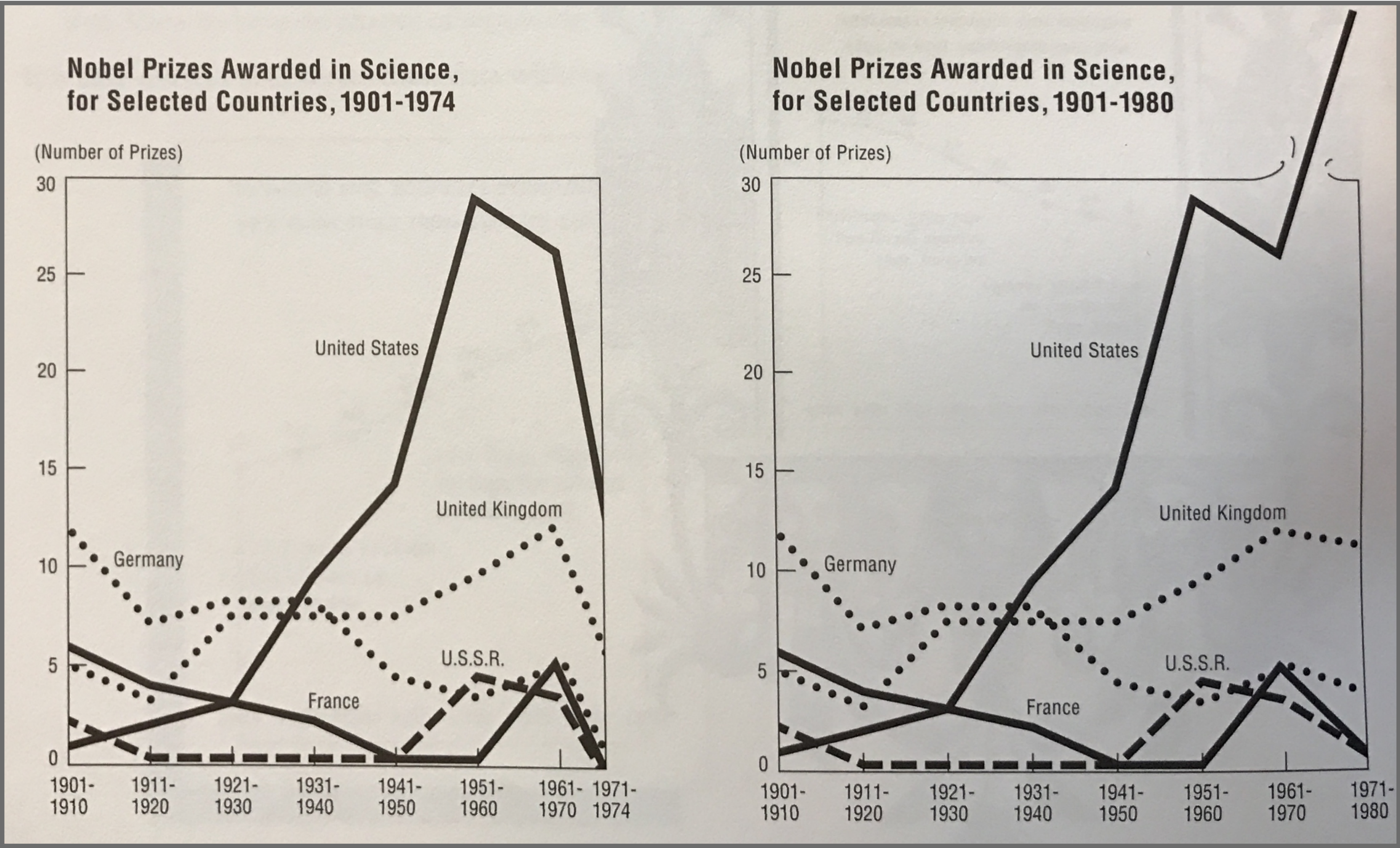
Example



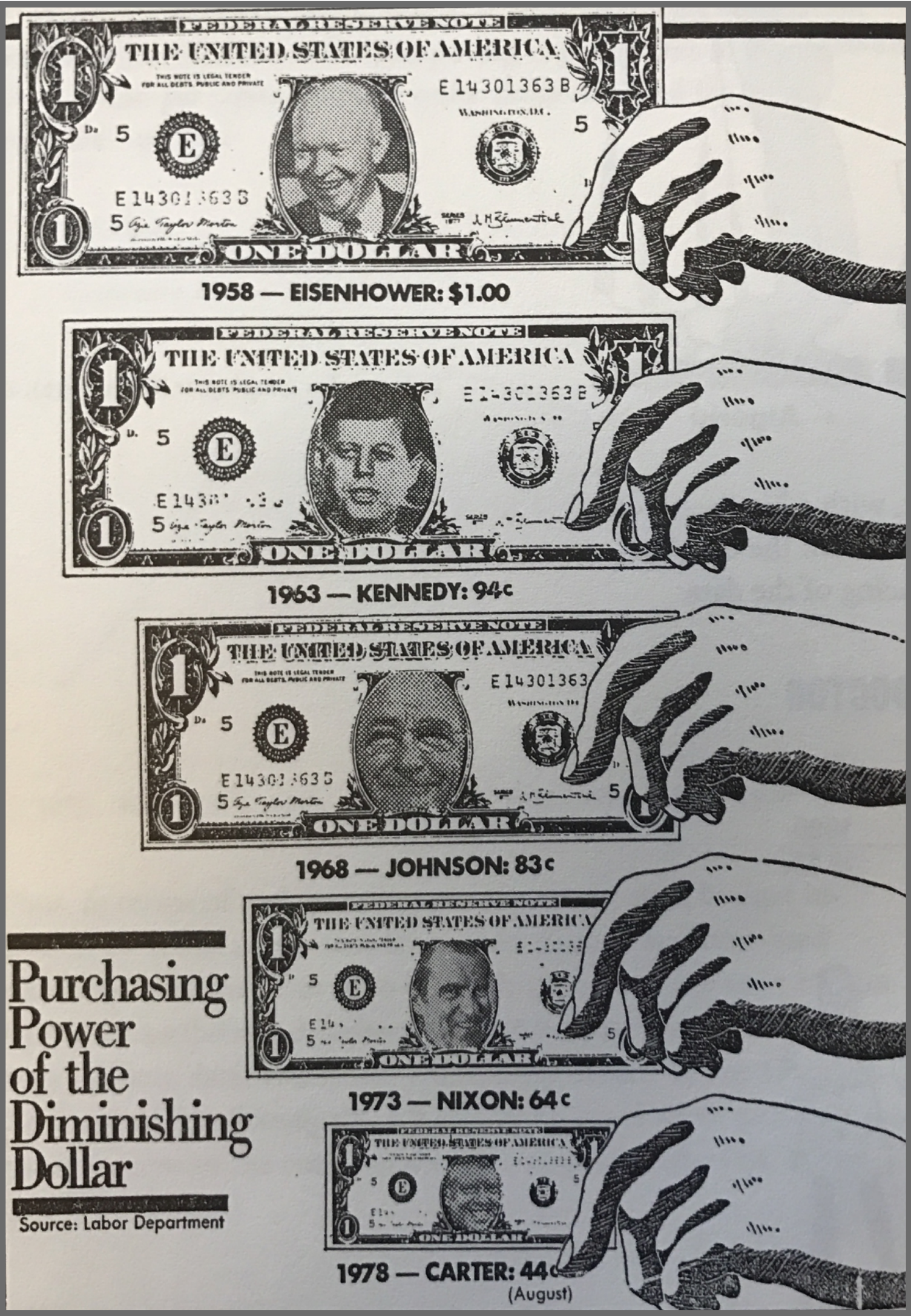
Example



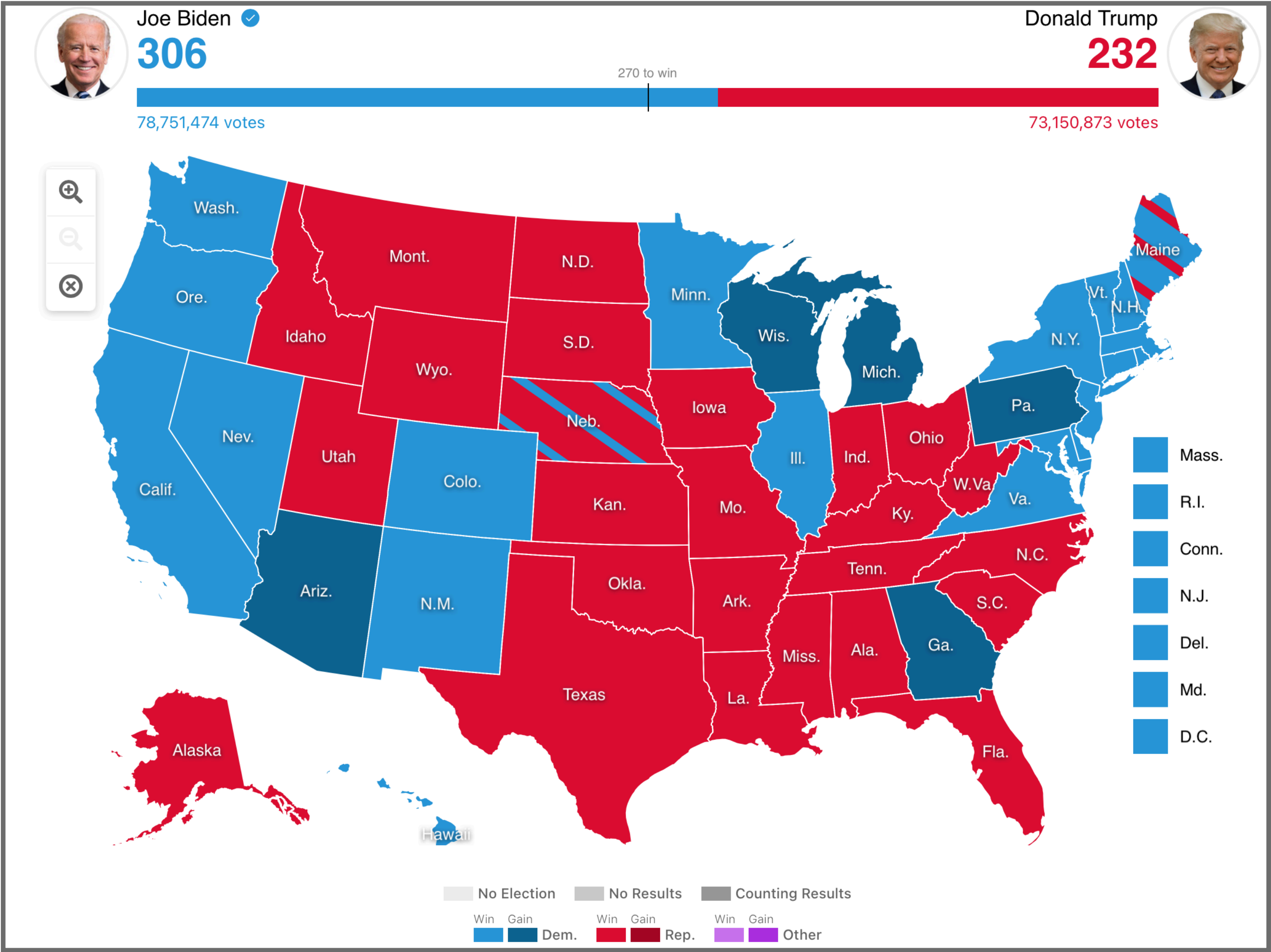
Example (corrected)



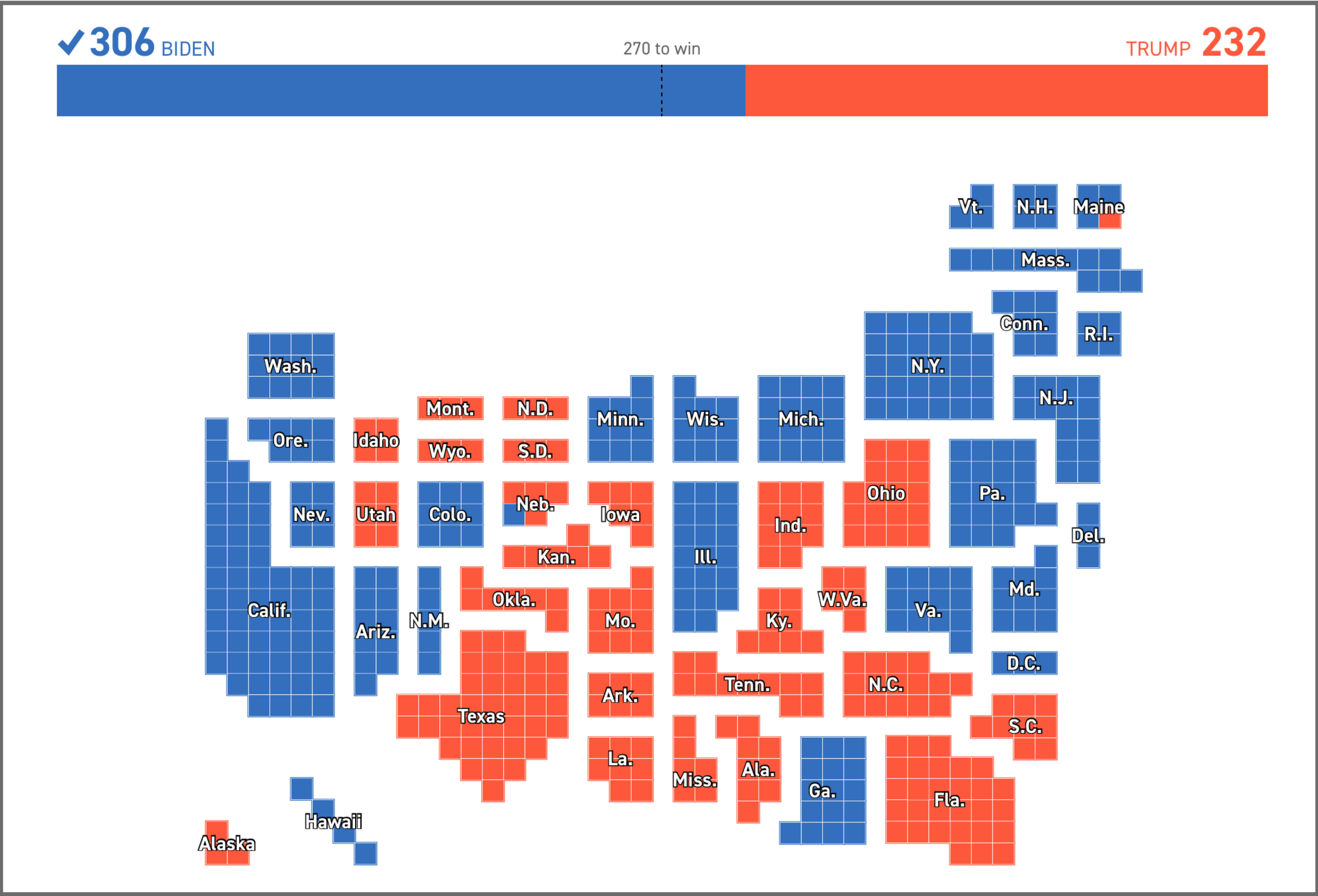
Example



Traditional Electoral Map



Weighted Electoral Map

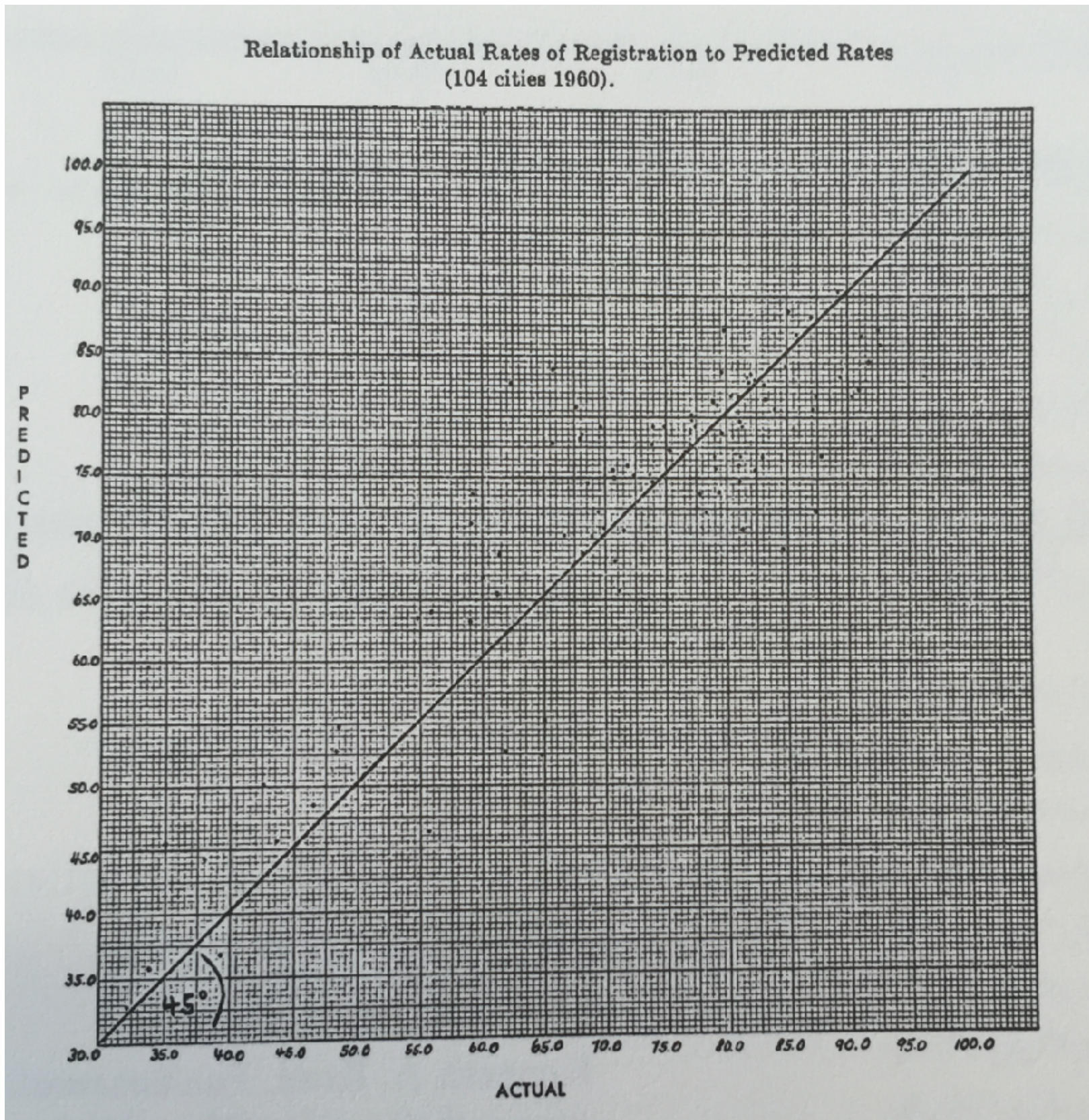


Data-ink

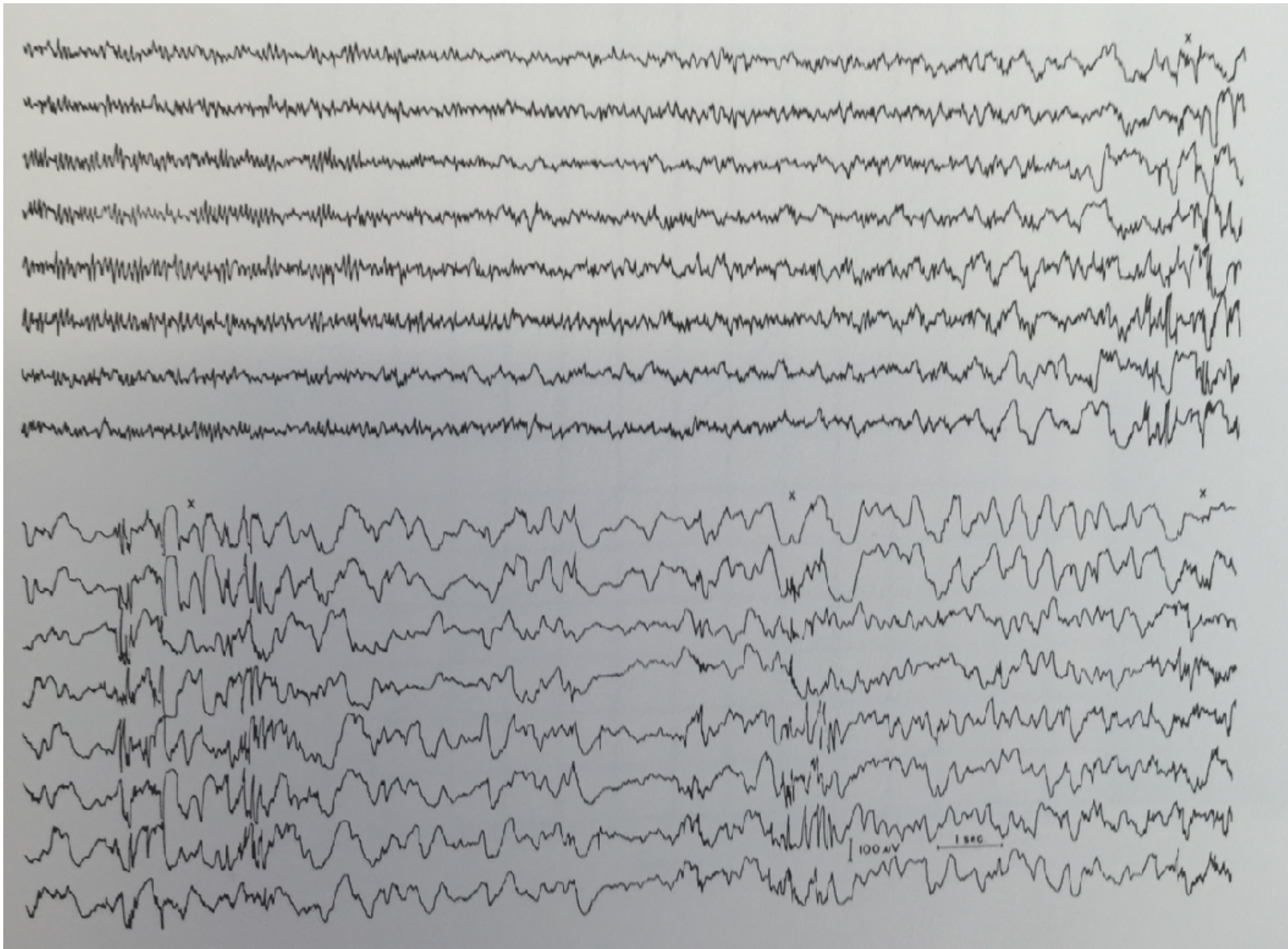
- Data-ink - non-redundant ink encoding data information

$$\begin{aligned}\text{Data-ink ratio} &= \frac{\text{Data-ink}}{\text{Total ink used to print the graphic}} \\ &= \text{proportion of a graphic's ink devoted to the} \\ &\quad \text{non-redundant display of data-information} \\ &= 1.0 - \text{proportion of a graphic that can be erased}\end{aligned}$$

Examples of Data-ink Ratio



~0

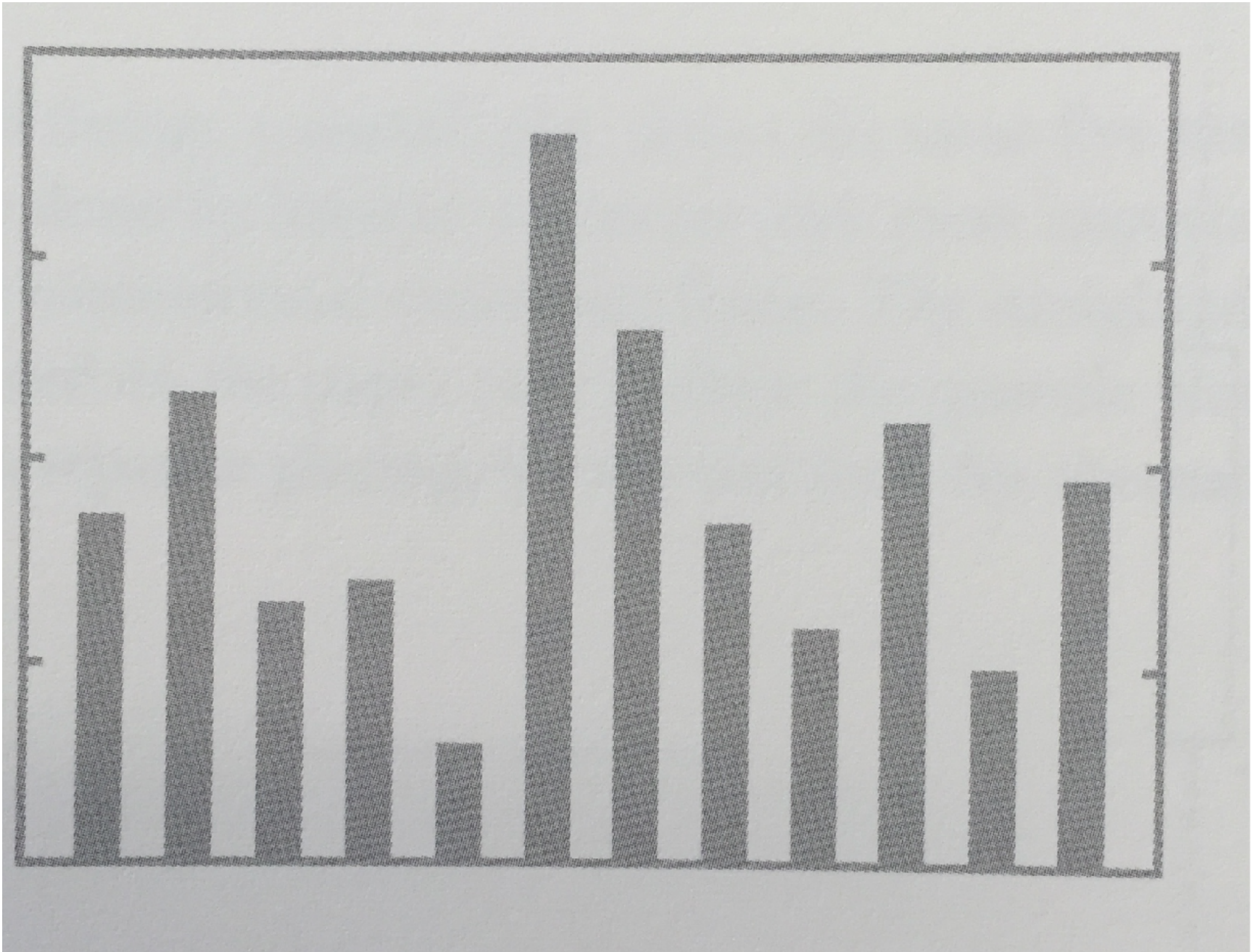


1.0

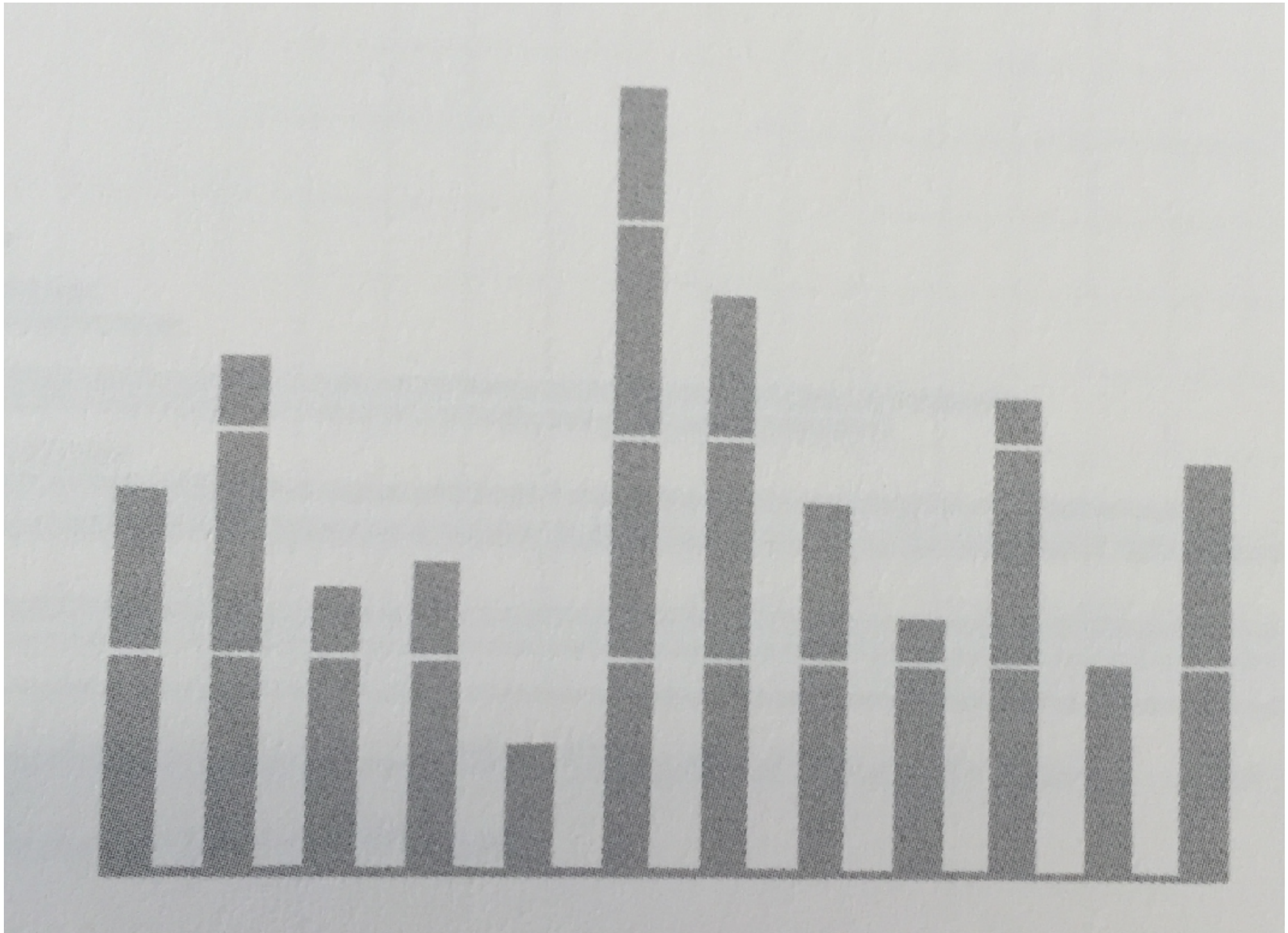
Design Principles for Data-ink

- (a.k.a. aesthetics & minimalism / elegance & simplicity)
- *Above all else show the data*
 - Erase non-data-ink, within reason
 - Often not valuable and distracting
 - Redundancy not usually useful

Example



Example (revised)



Interacting with Visualizations

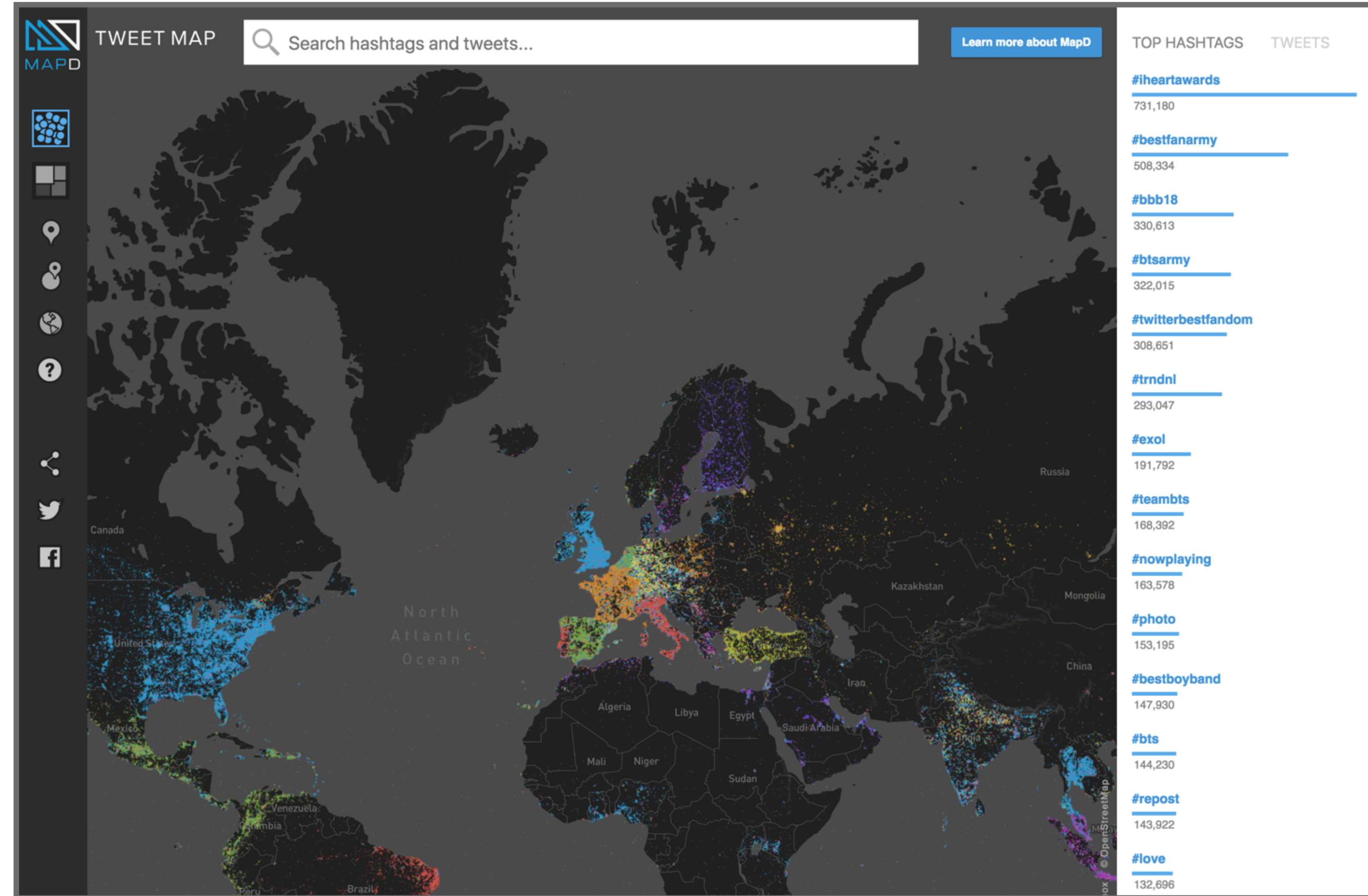
Interactive Visualizations

- Users often use iterative process of making sense of the data
 - Answers lead to new questions
- Interactivity helps user constantly change display of information to answer new questions
- Should offer visualization that offers best view of data moment to moment as desired view changes

Information Visualization Tasks

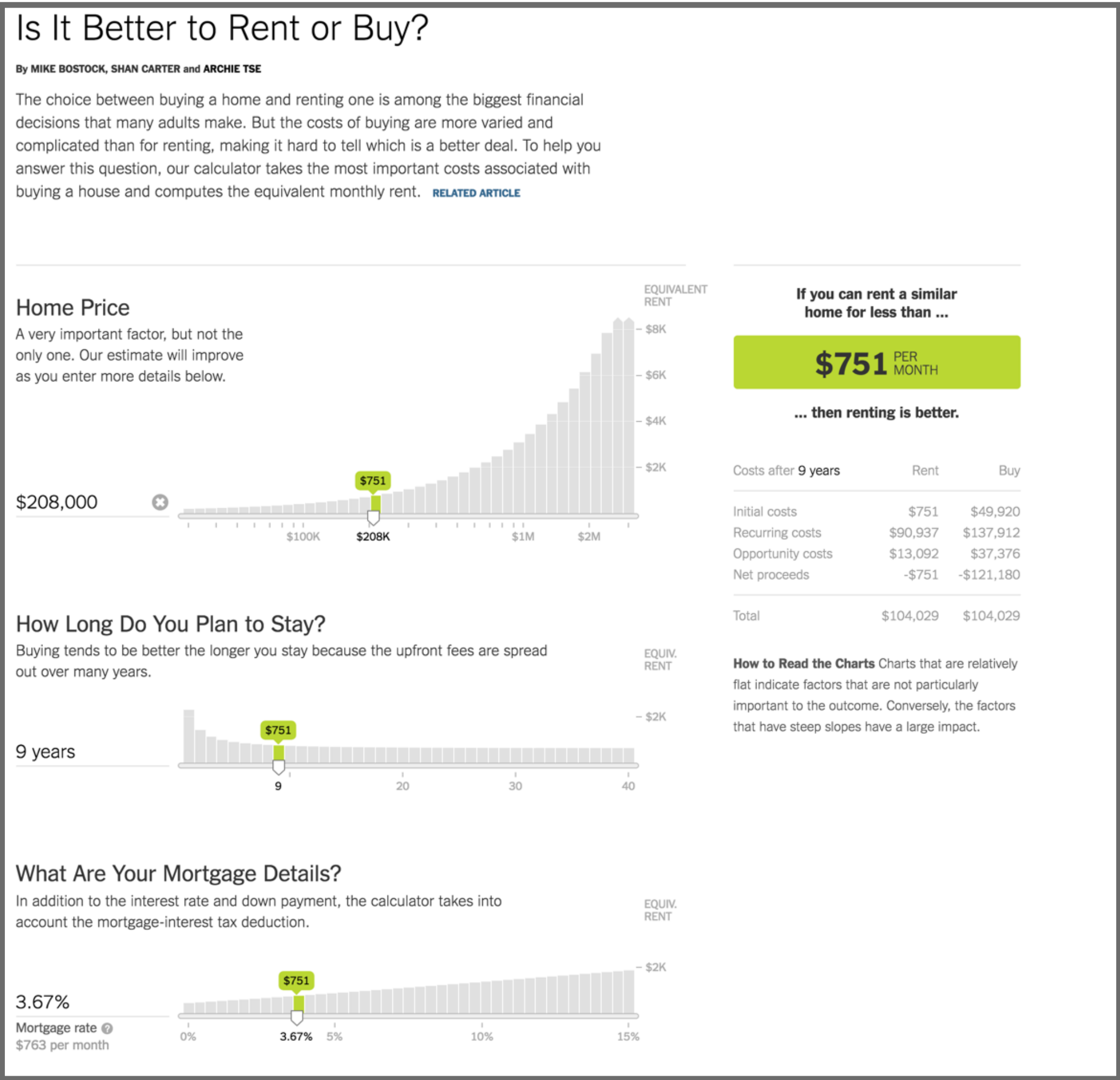
- Overview: gain an overview of entire collection
- Zoom: zoom in on items of interest
- Filter: filter out uninteresting items
- Details on Demand: select an item or group and get details
- Relate: view relationships between items
- History: support undo, replay, progressive refinement
- Extract: allow extraction of sub-collections through queries

Global Tweet Map



<https://www.mapd.com/demos/tweetmap/>

Renting vs. Buying Utility



https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html?_r=0

10 Minute Break

In-Class Activity

Design an Information Visualization

- In groups of 2 or 3
 - Select a set of data to visualize and two or more representative questions to answer using this data
 - Design an *interactive* information visualization
 - Create sketches showing the design of the information visualization
 - Should have multiple views of data, interactions to configure and move between views
- Deliverables: 2+ questions you support, sketches with annotations explaining how users would use visualization to answer questions