

# Information Visualization

SWE 432, Fall 2019

Design and Implementation of Software for the Web

# Today

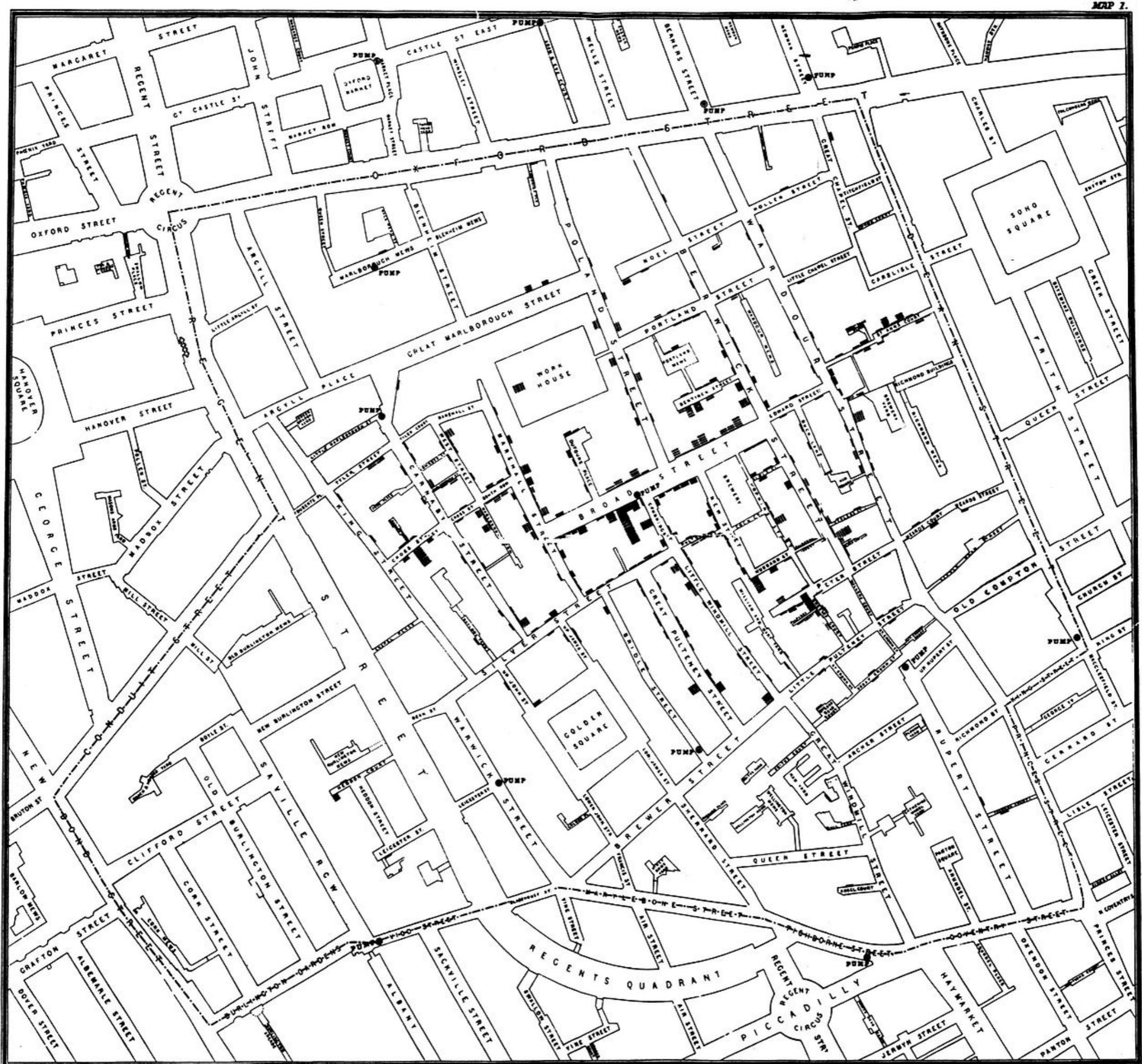
- What is an information visualization?
- Lots of examples of visualizations
- Patterns and anti-patterns for visualization

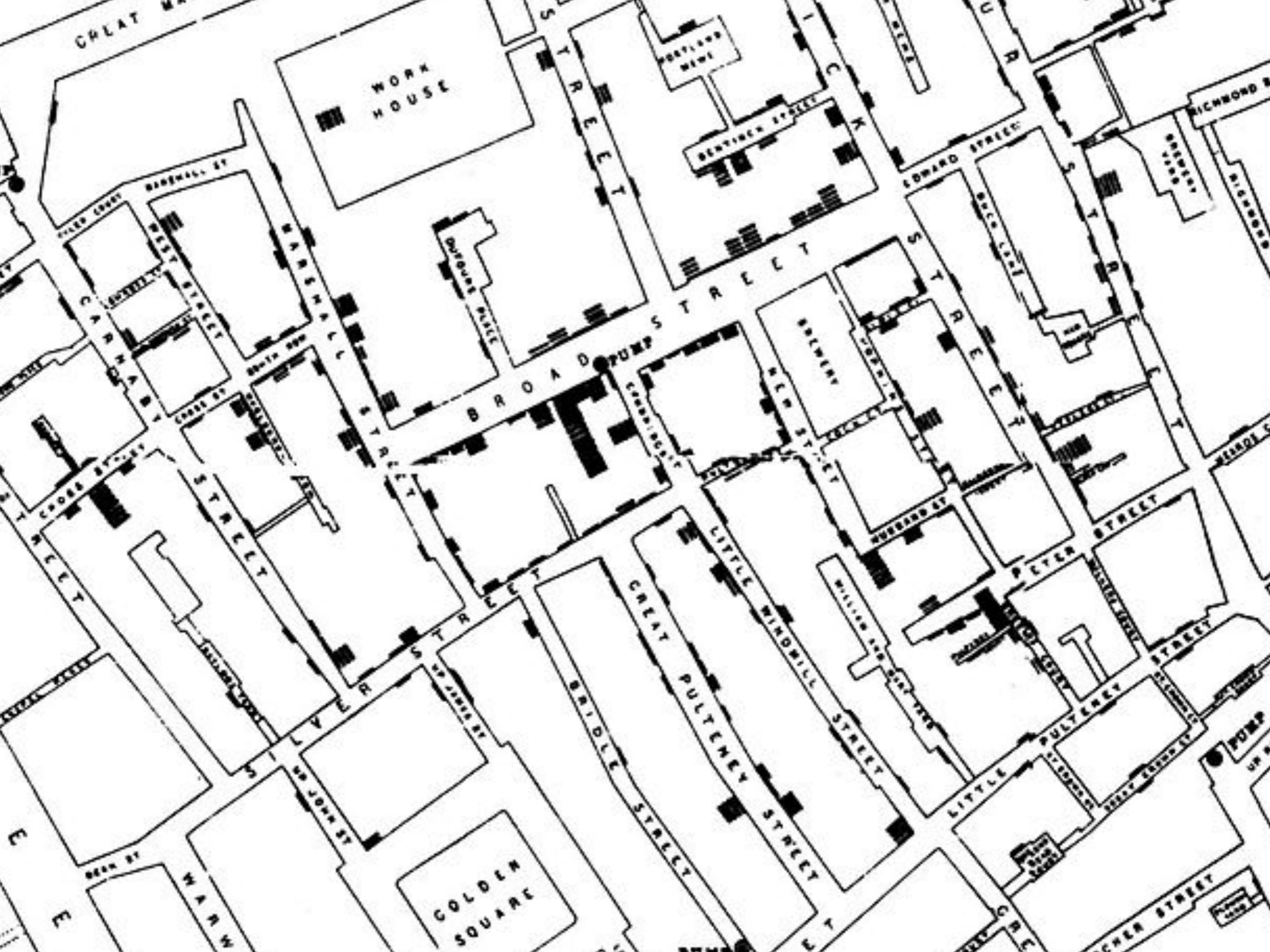
# Information visualization

- Technology has made data pervasive
  - health, finance, commerce, customer, travel, demographics, communications, ...
  - some of it “big”
- Information visualization: the use of interactive visual representations to amplify cognition
  - e.g., discover insights, answer questions

# 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





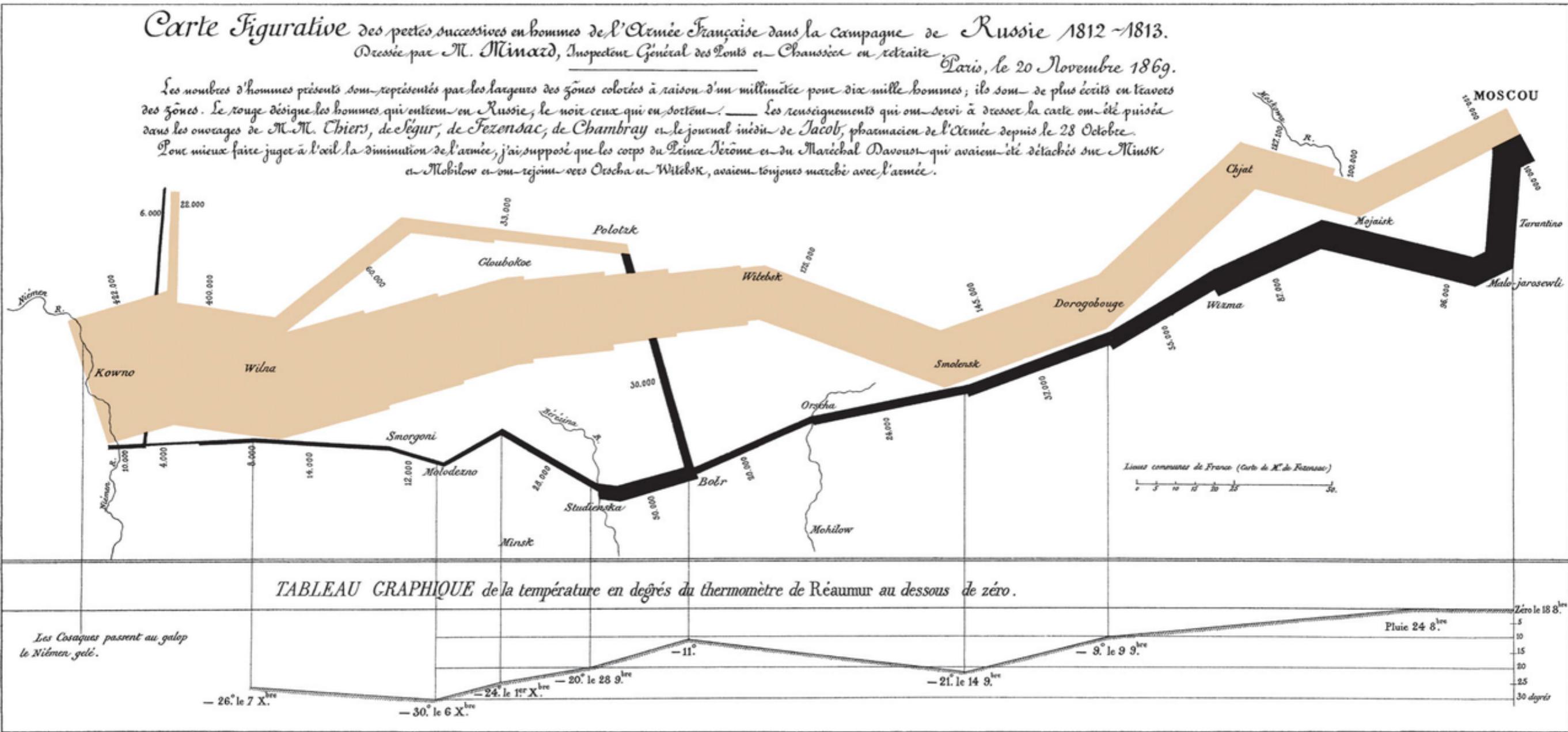
# Investigation and aftermath

- 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

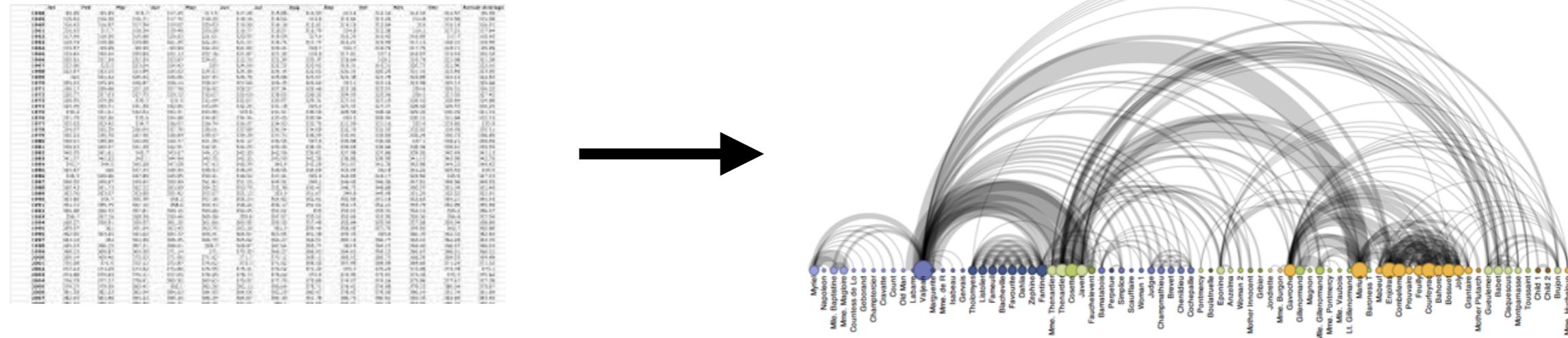
# Methods used by Snow

- 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

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



# What is an information visualization?



- Data → Visual representation
  - Rows in data table → elements in data visualization
    - e.g., historical person → circle in visualization
  - Columns of data → visual variables
    - e.g., relationship to another person → edge in network visualization

# Data transformations

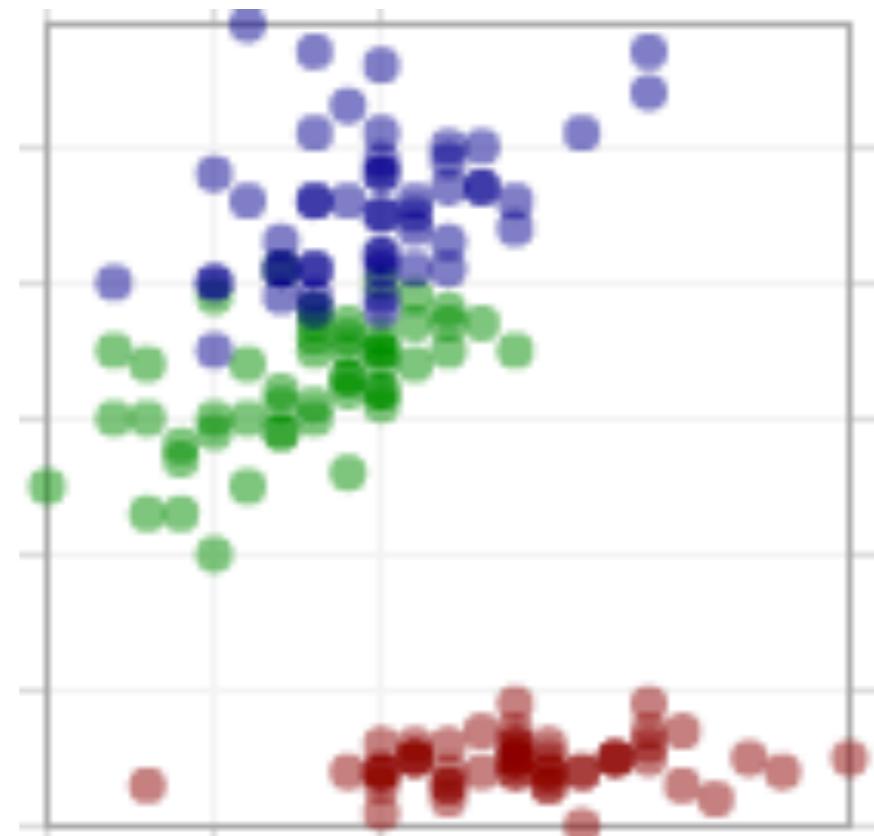
- Classing / binning: Quantitative —> ordinal
  - Maps ranges onto **classes** of variables
  - Can also count # of items in each class w/ histogram
- Sorting: Nominal —> ordinal
  - Add order between items in sets
- Descriptive statistics: mean, average, median, max, min, ...

# Visual structures

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

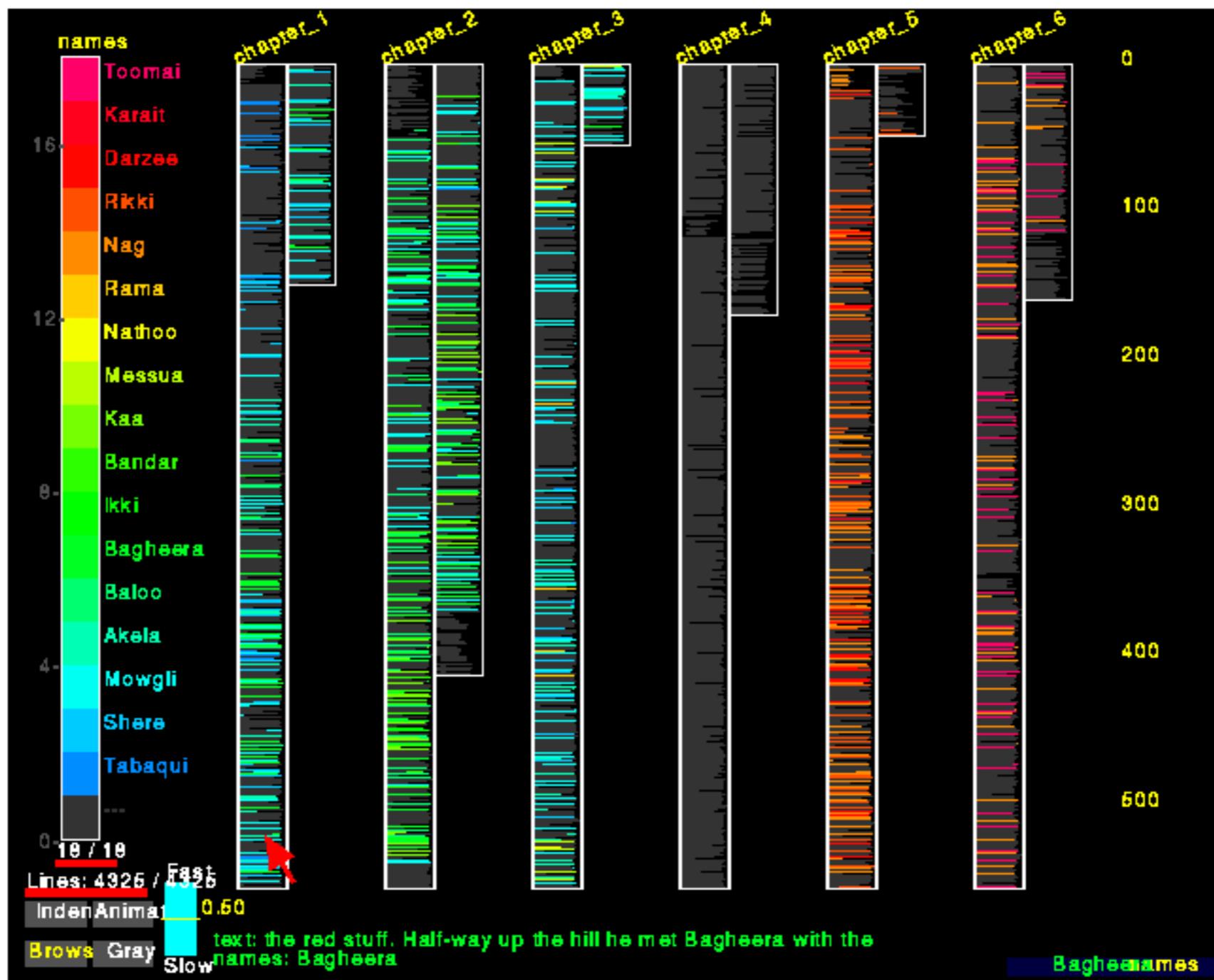
# Spatial substrate

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



# Folding

- continuing an axis by continuing in different space



# Marks

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

# Marks' graphical properties

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

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

# Effectiveness of graphical properties

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

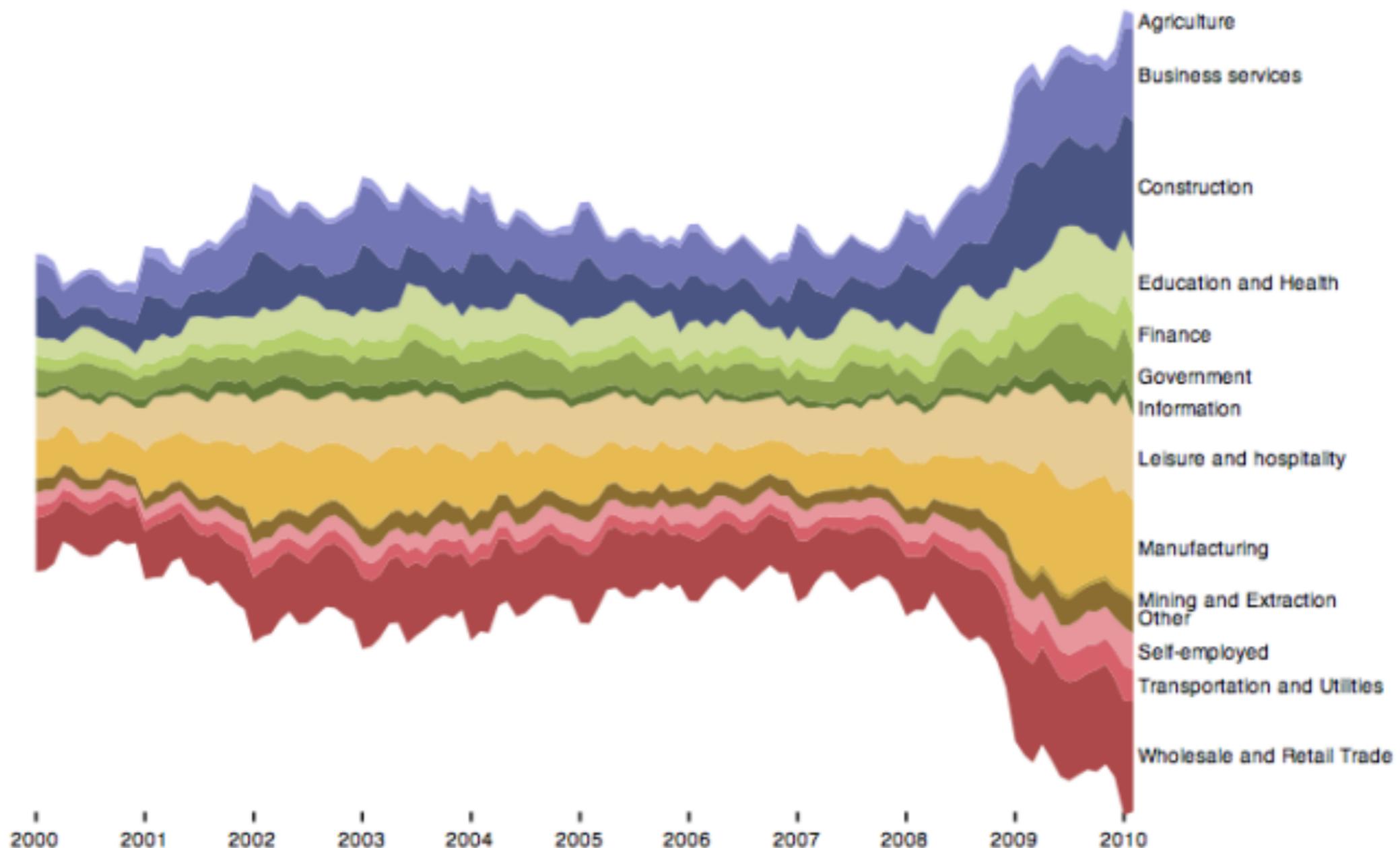
	Spatial	Q	O	N	Object	Q	O	N
Extent	(Position)	●	●	●	Grayscale	○	●	○
	Size	●	●	●				
Differential	Orientation	○	○	●	Color	○	○	●
					Texture	○	○	●
					Shape	○	○	●

Examples of  
visualizations

# Time-series data

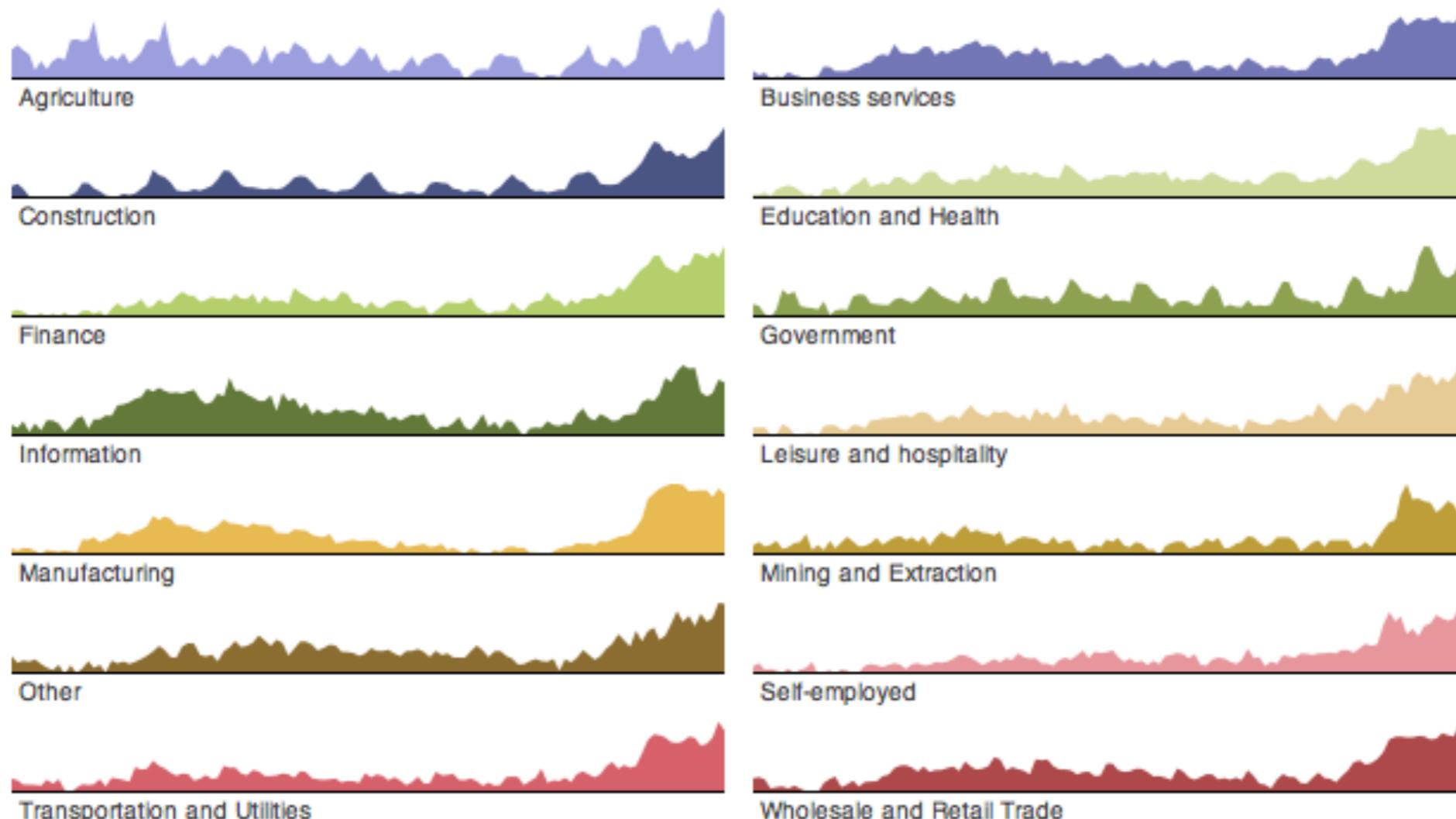
# Stacked graph

- Supports visual summation of multiple components



# Small multiples

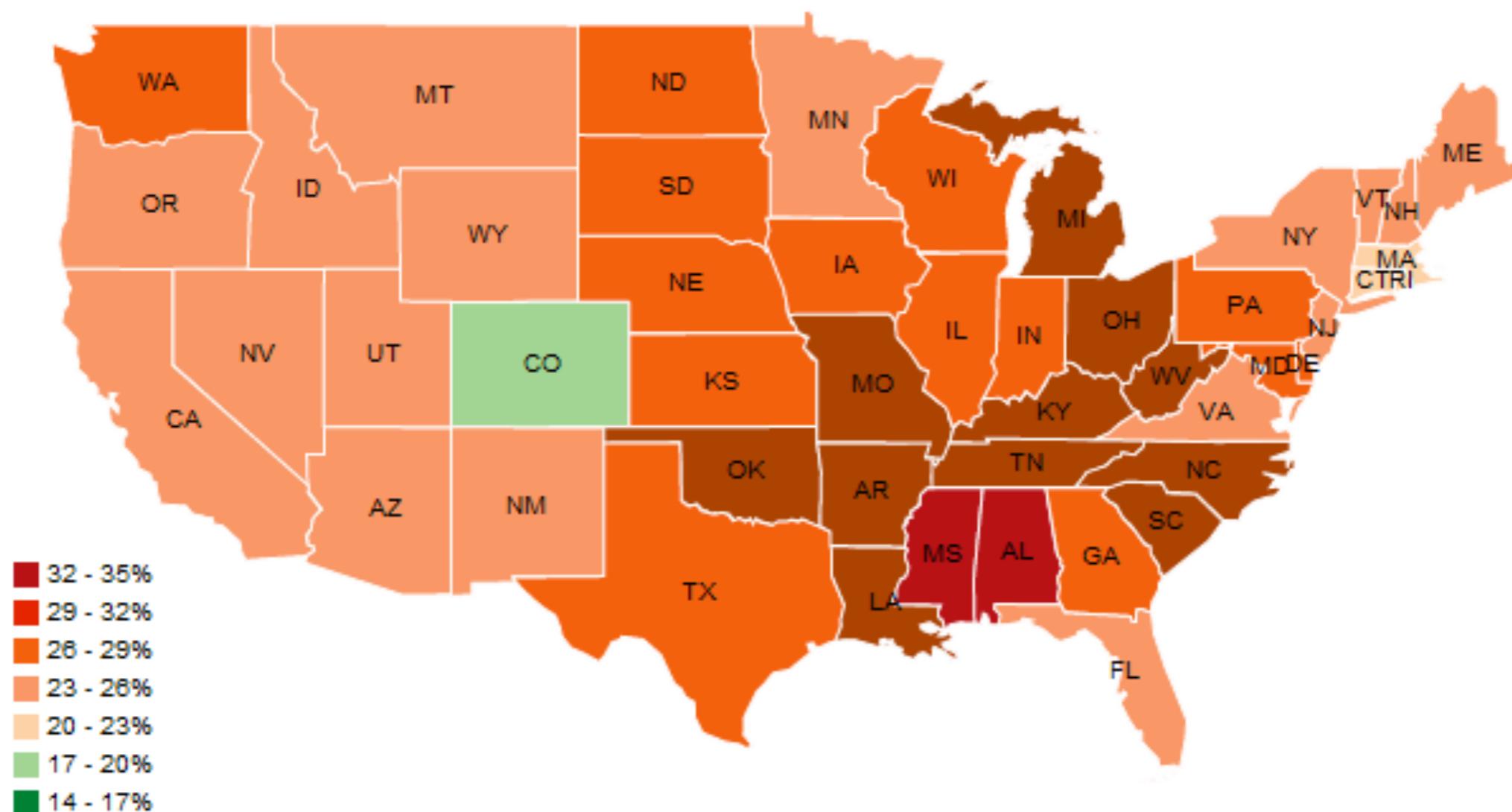
- supports separate comparison of data series
- may have better legibility than placing all in single plot



# Maps

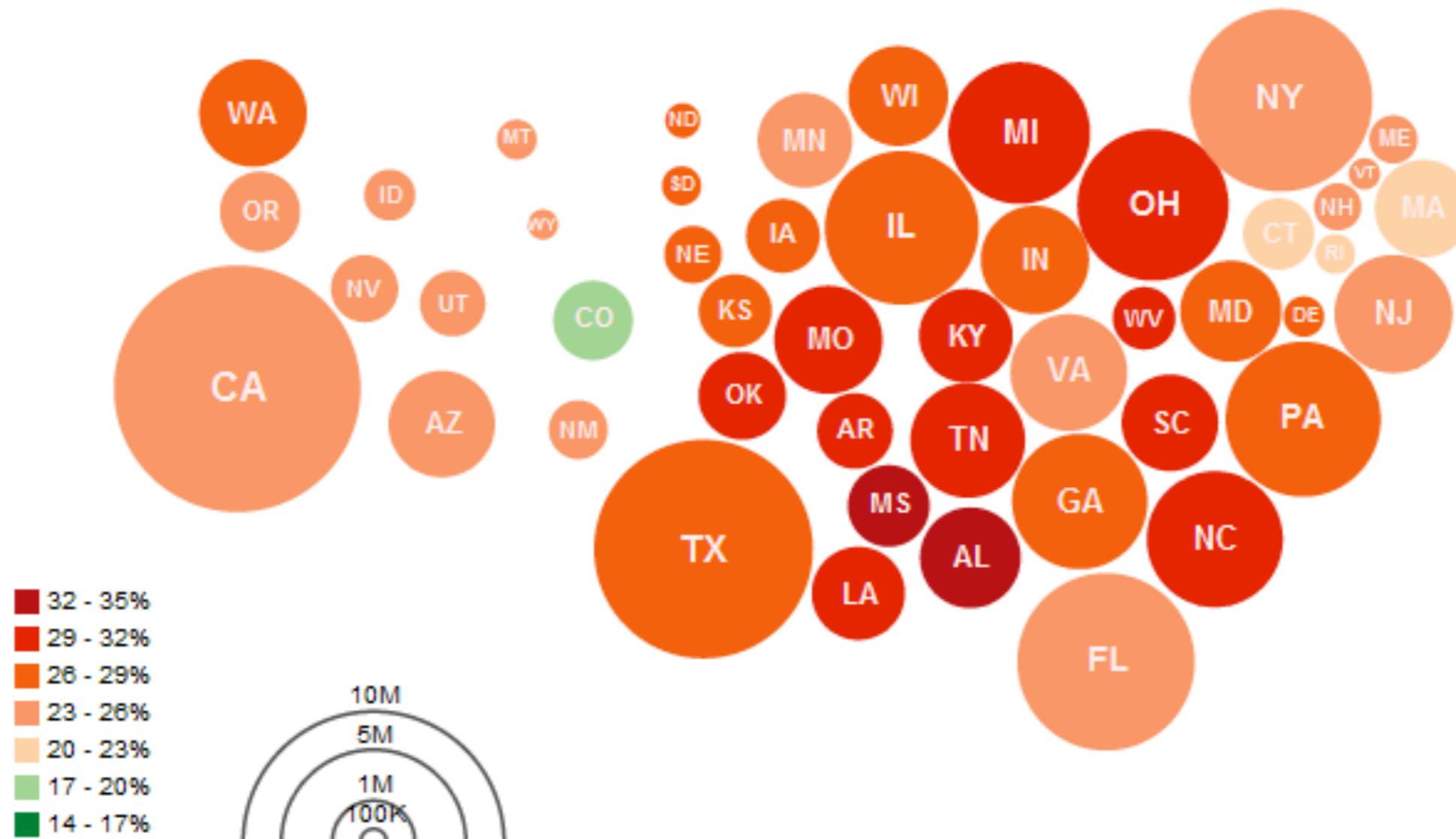
# Choropleth map

- Groups data by area, maps to color



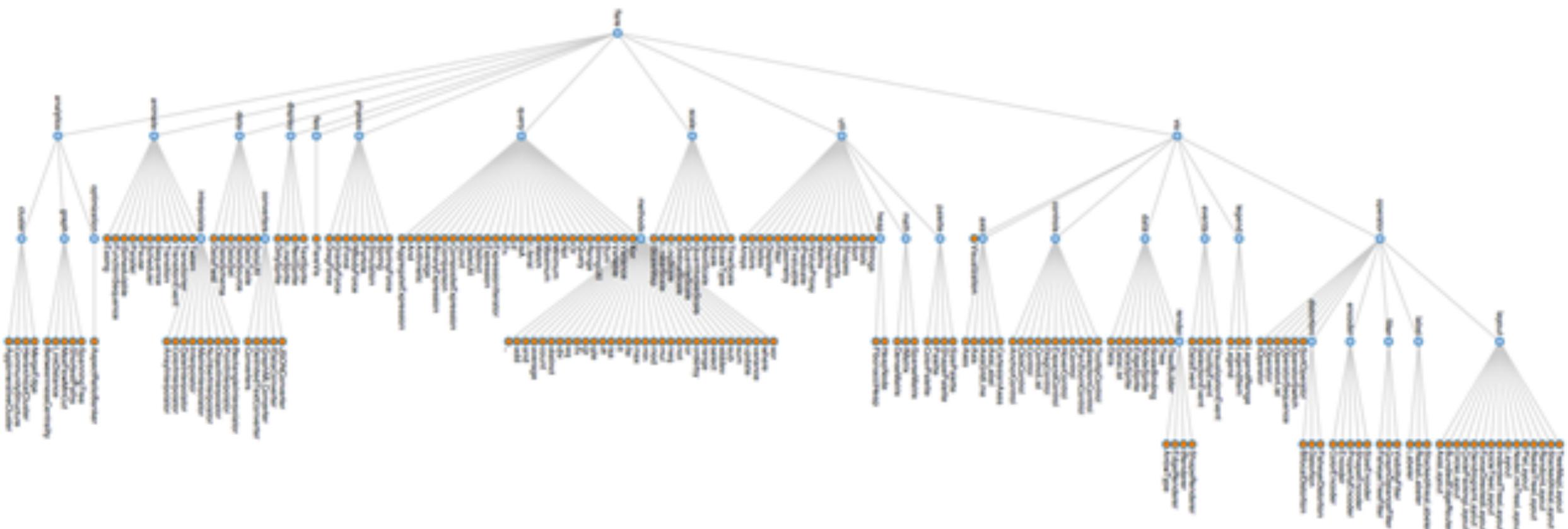
# Cartograms

- Encodes two variables w/ size & color



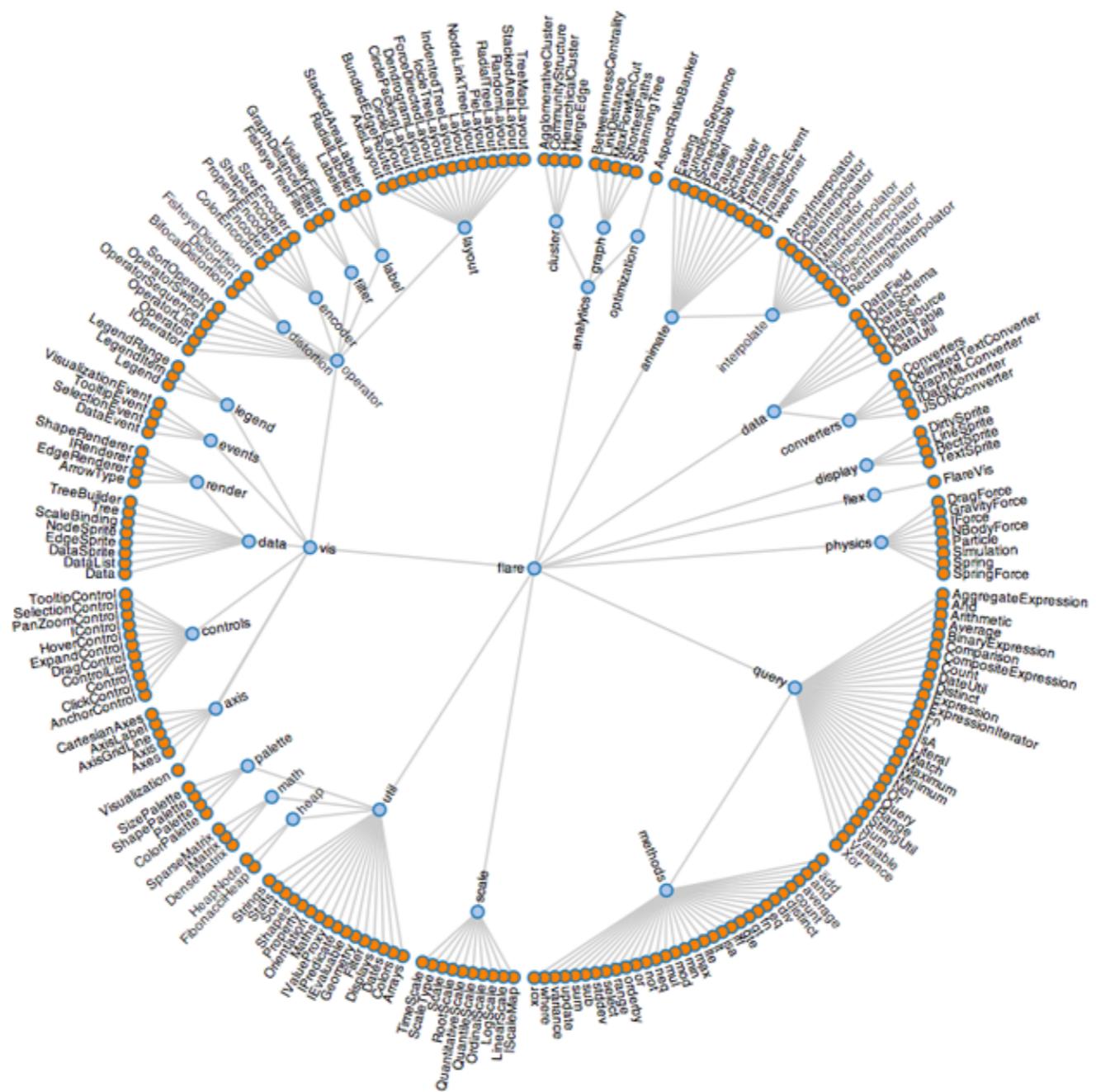
# Hierarchies

# Node link diagram

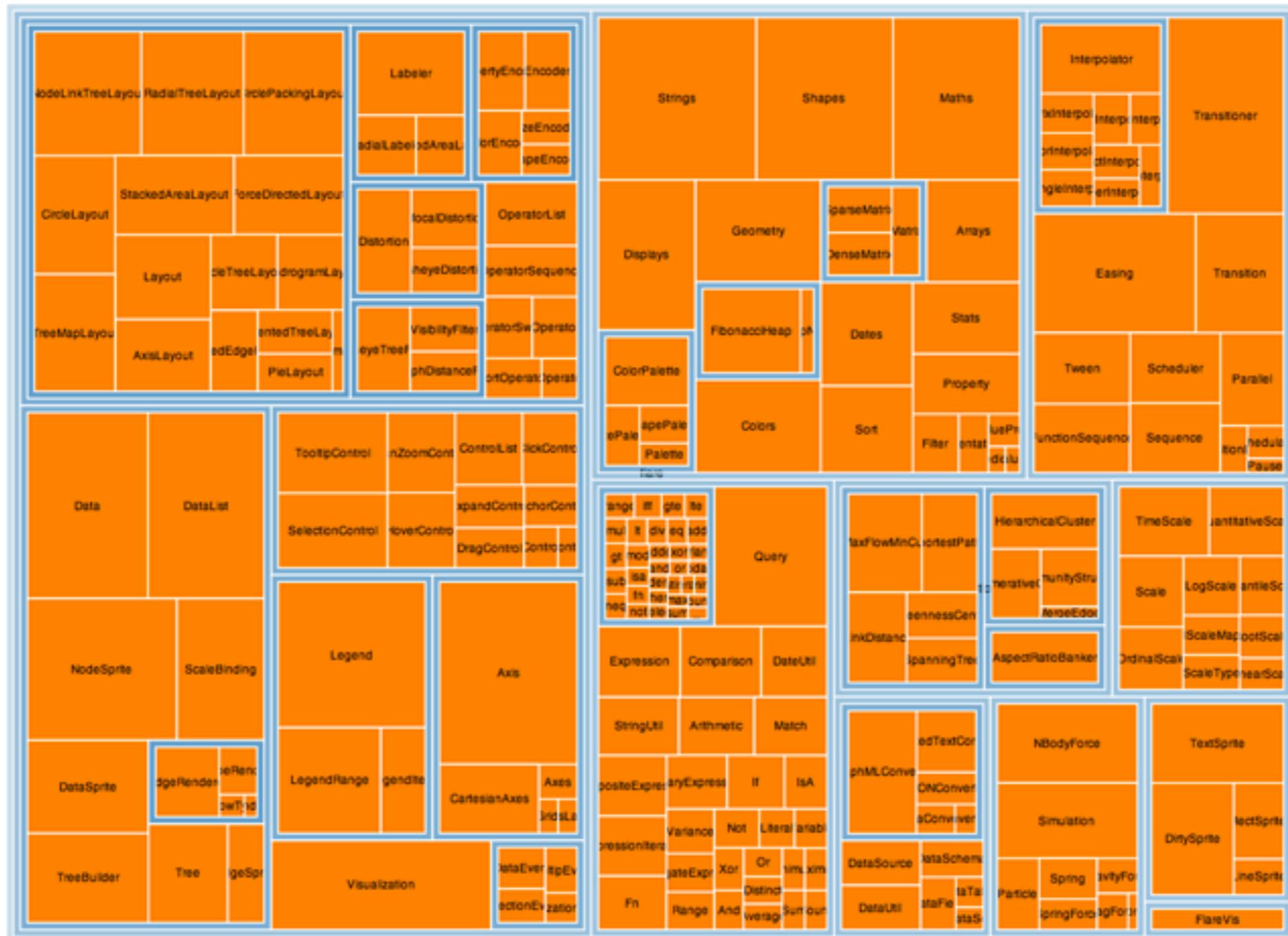


# Dendrogram

- leaf nodes of hierarchy on edges of circle



# Treemaps



# Design considerations

# Tufte's principles of graphical excellence

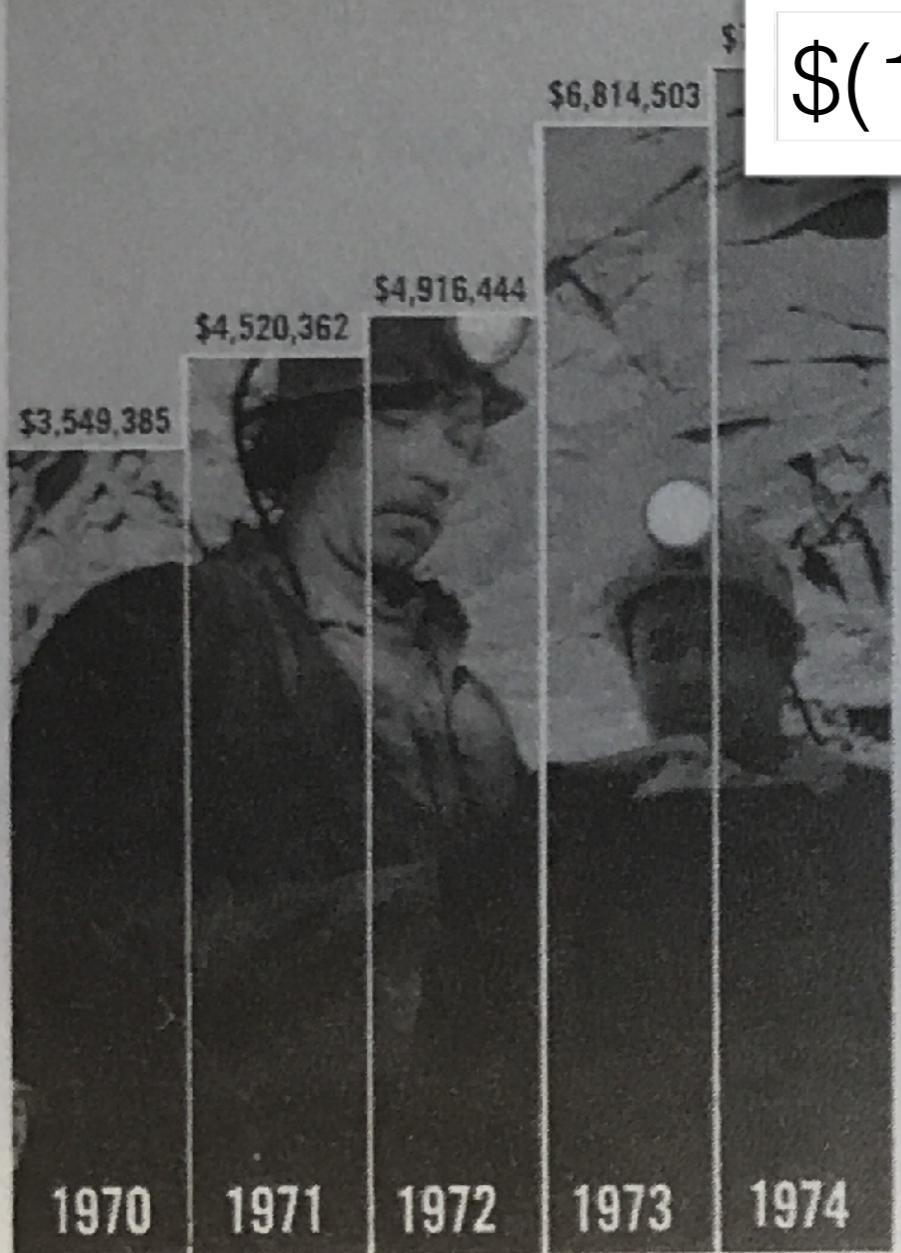
- show the **data**
- induce the viewer to think about the substance rather than the methodology
- avoid distorting what the data have to say
- present **many** numbers in a small space
- make large data sets **coherent**
- encourage the eye to **compare** different pieces of data
- reveal data at several levels of detail, from overview to fine structure
- serve reasonable clear **purpose**: description, exploration, tabulation, decoration

# Distortions in visualizations

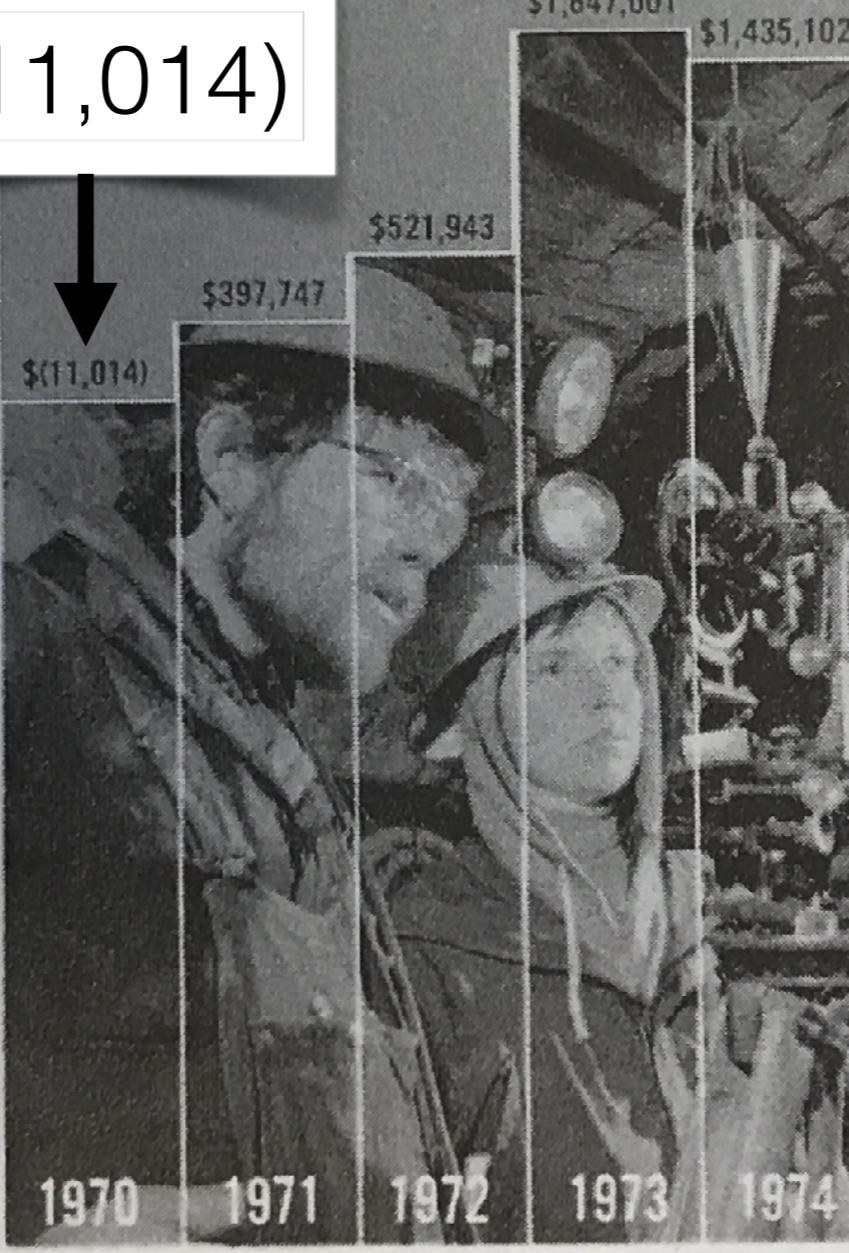
- Visualizations may distort the underlying data, making it harder for reader to understand truth
- Use of **design** variation to try to falsely communicate **data** variation

# Example

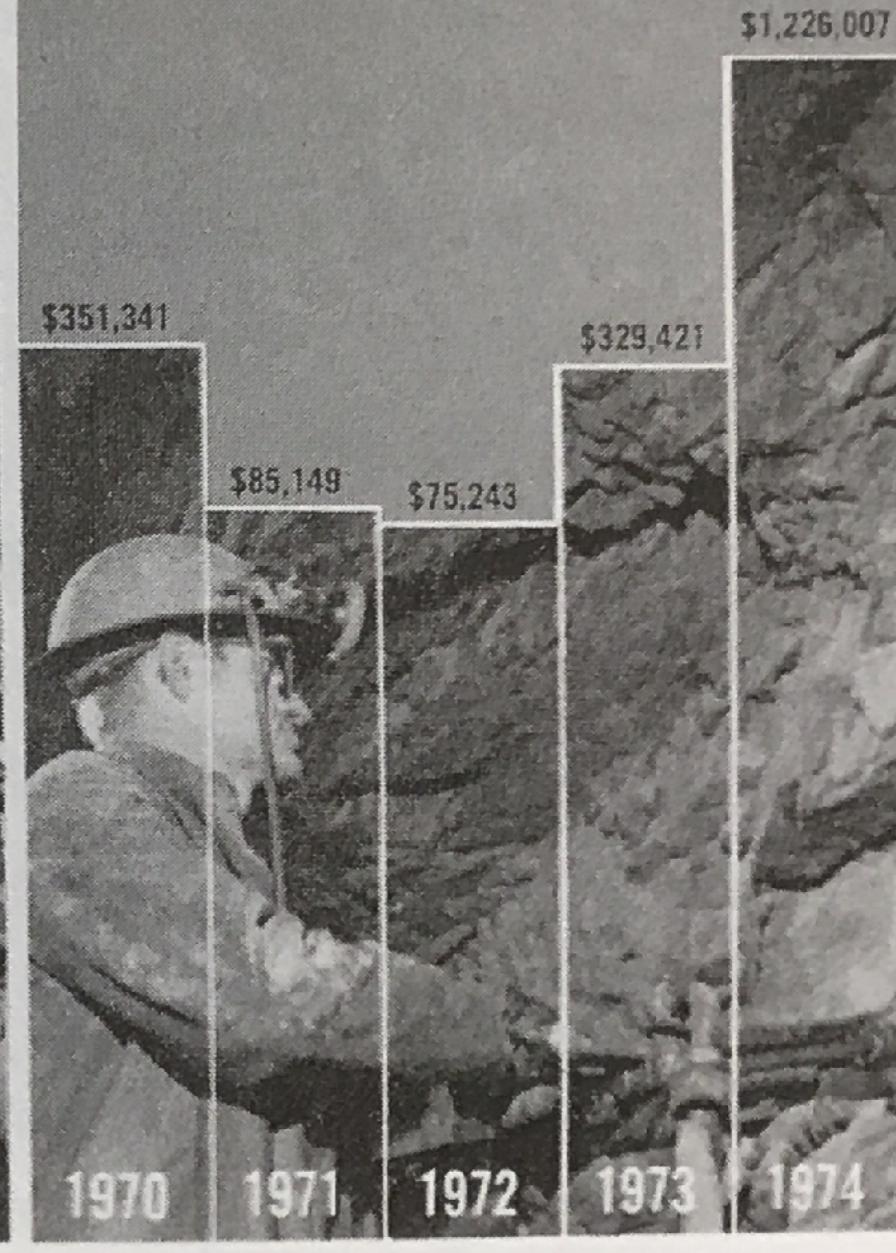
## OPERATING REVENUES



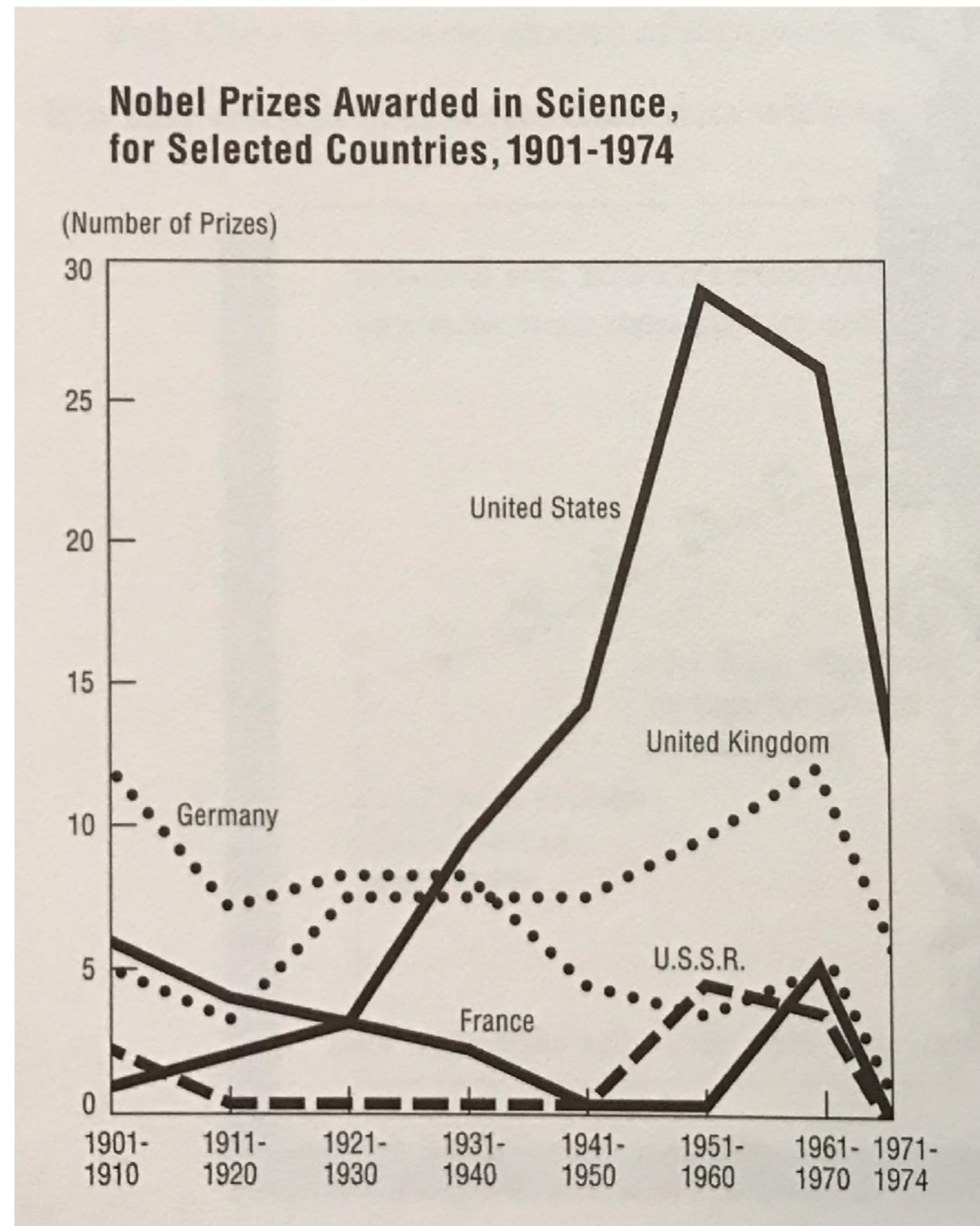
## NET INCOME (LOSS)



## EXPLORATION & DEVELOPMENT EXPENDITURES

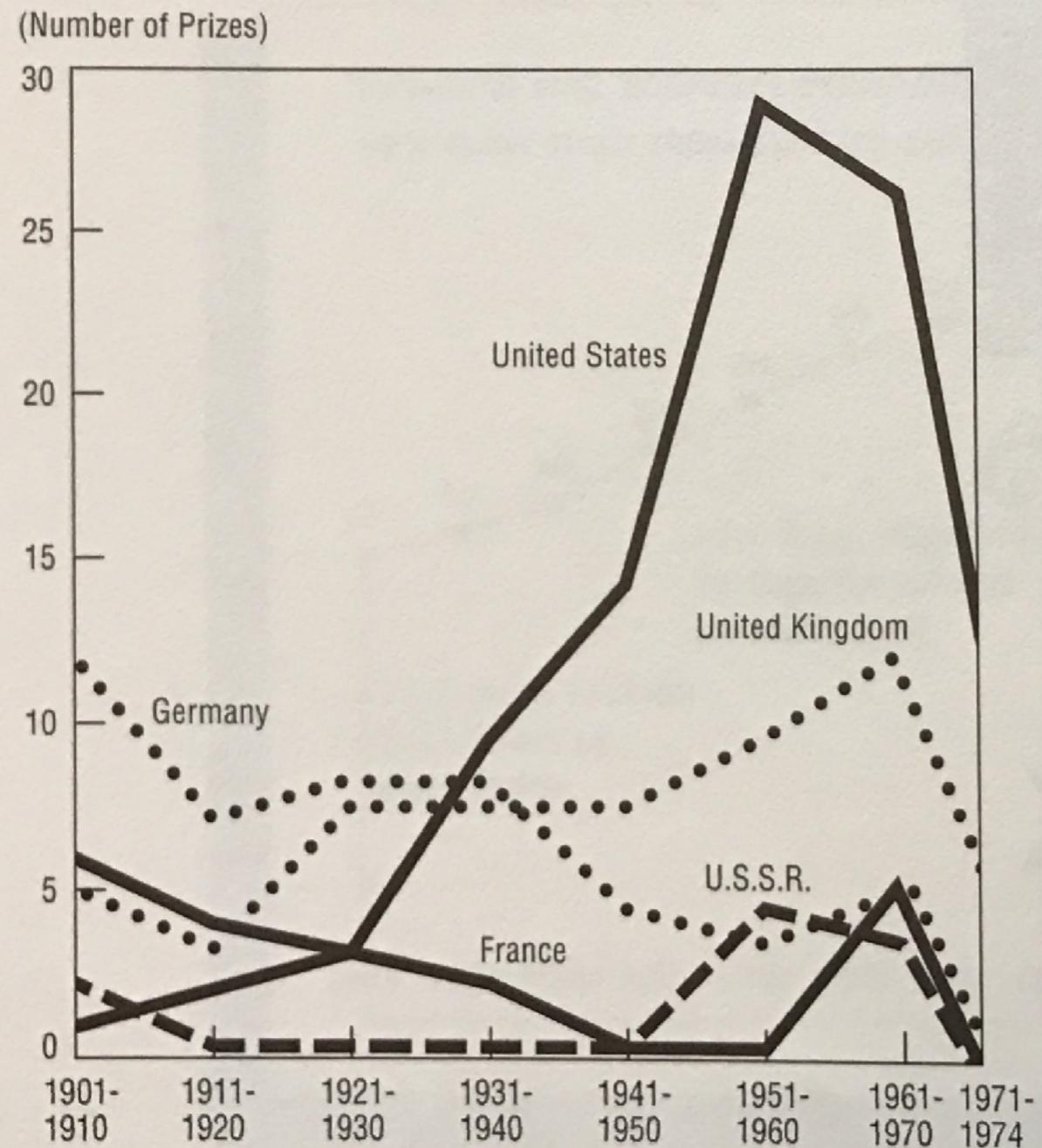


# Example

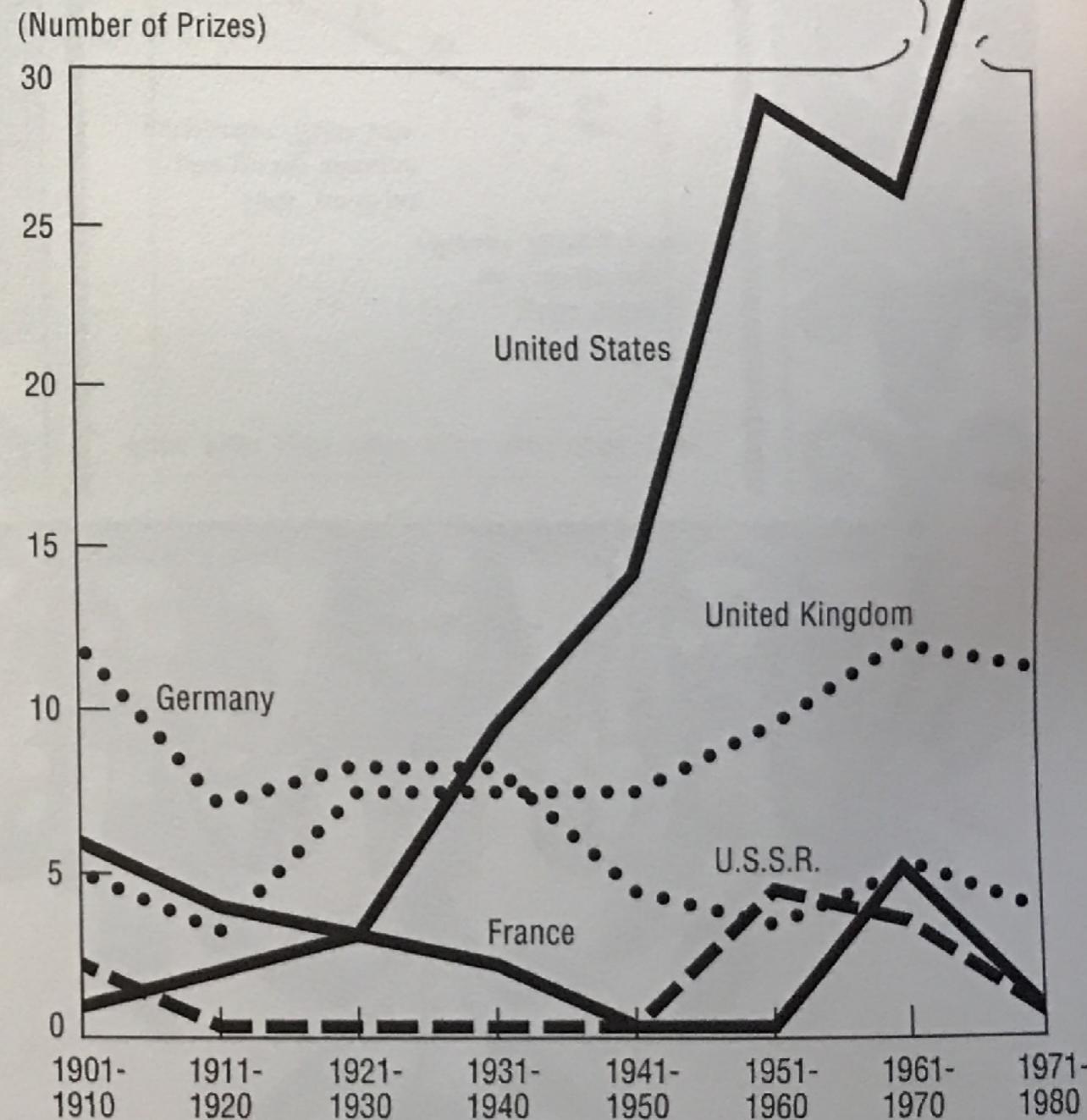


# Example (corrected)

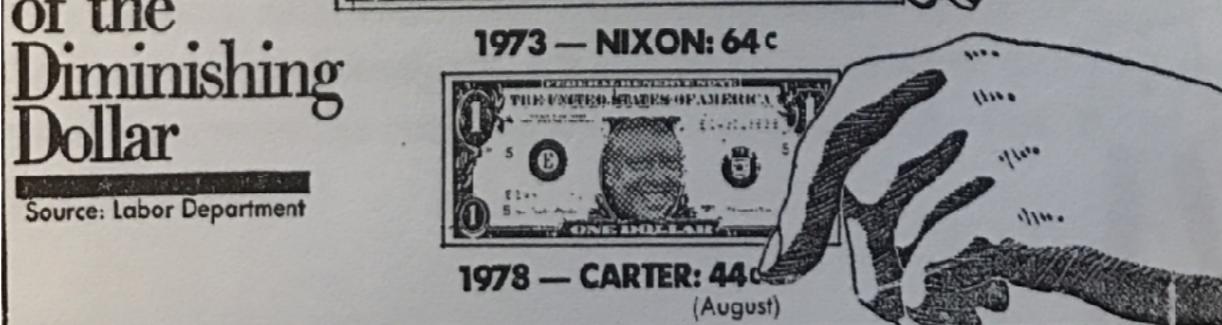
Nobel Prizes Awarded in Science,  
for Selected Countries, 1901-1974



Nobel Prizes Awarded in Science,  
for Selected Countries, 1901-1980



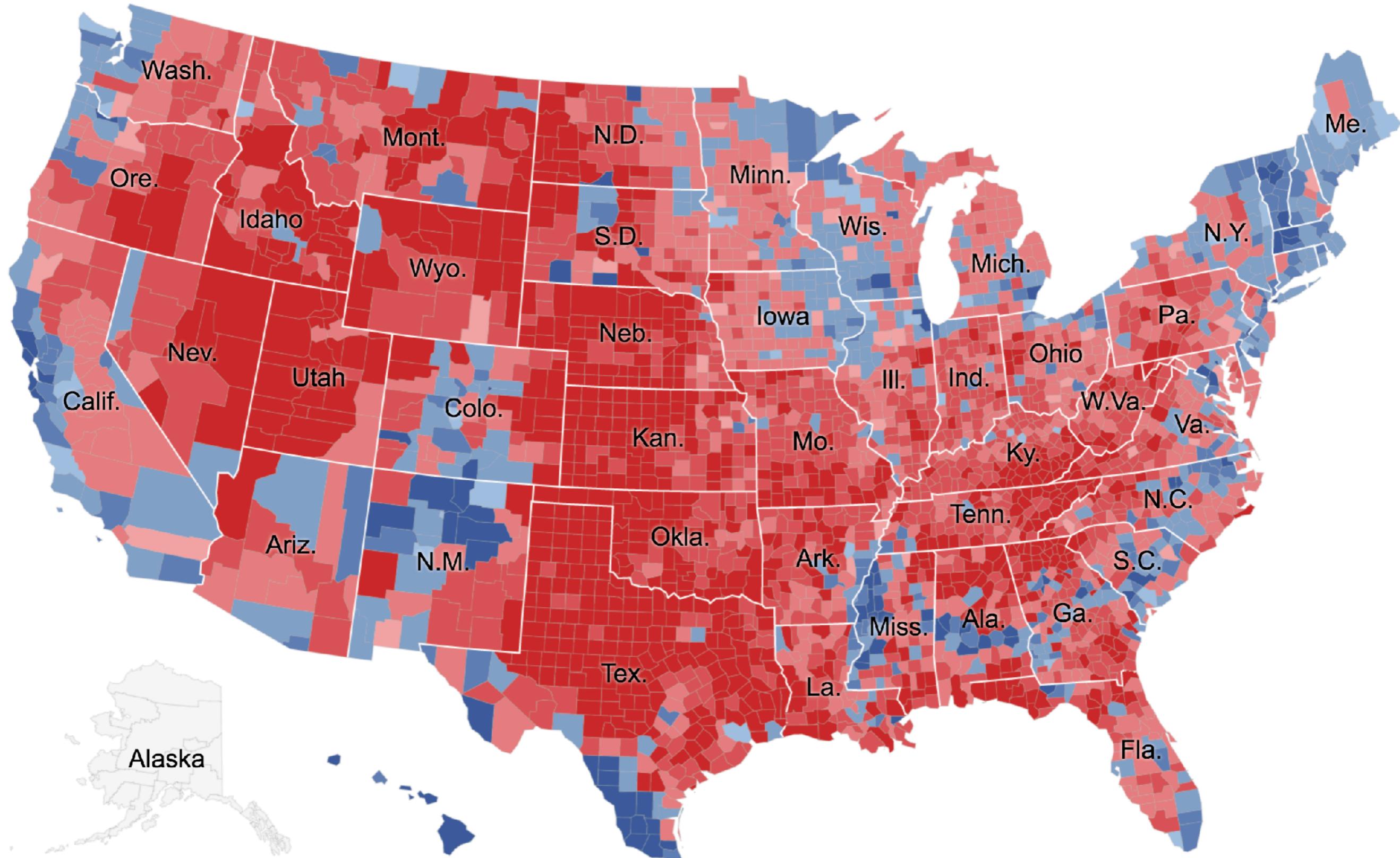
# Example



Purchasing  
Power  
of the  
Diminishing  
Dollar

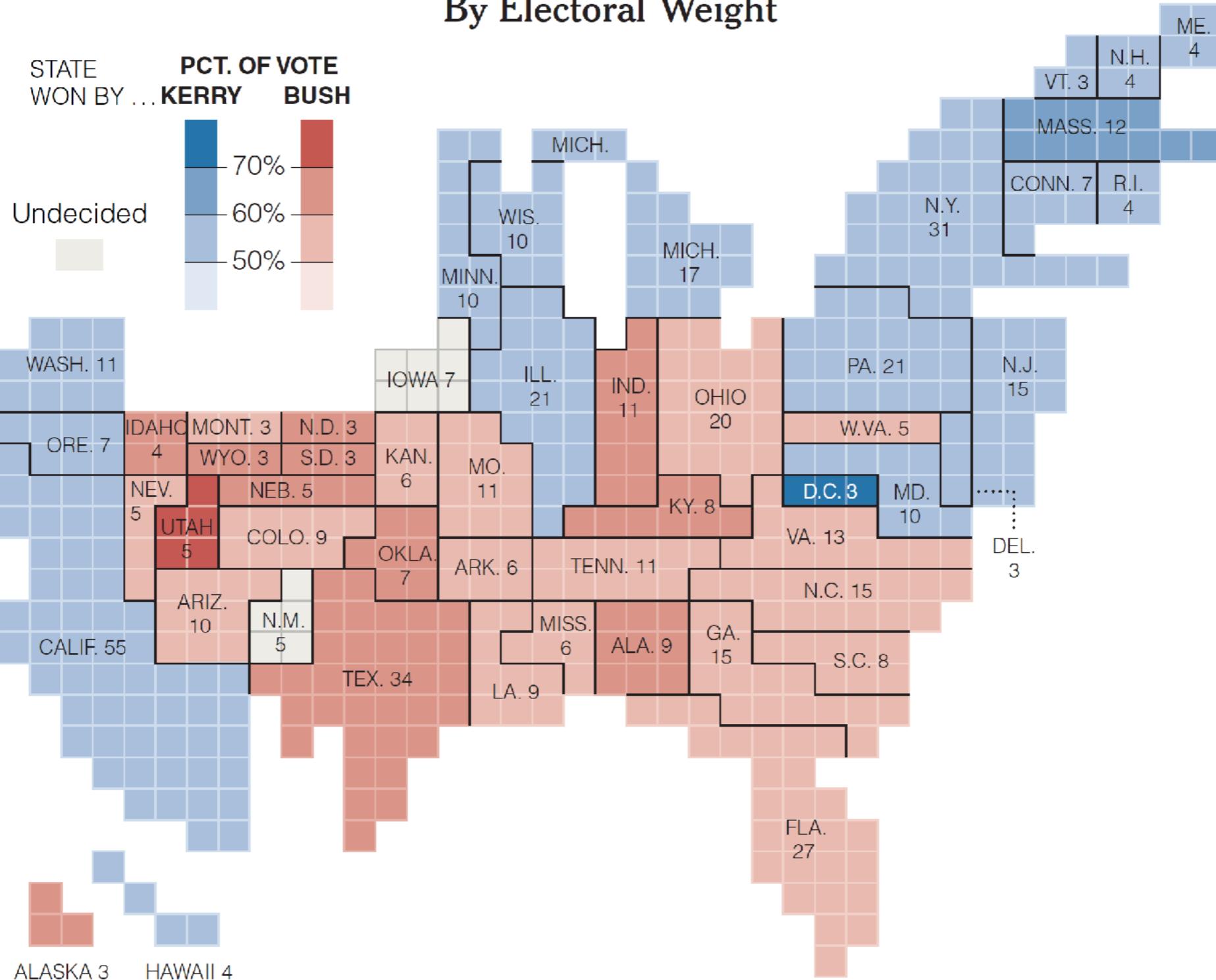
Source: Labor Department

# Traditional Electoral Map



# Weighted Electoral Map

## By Electoral Weight



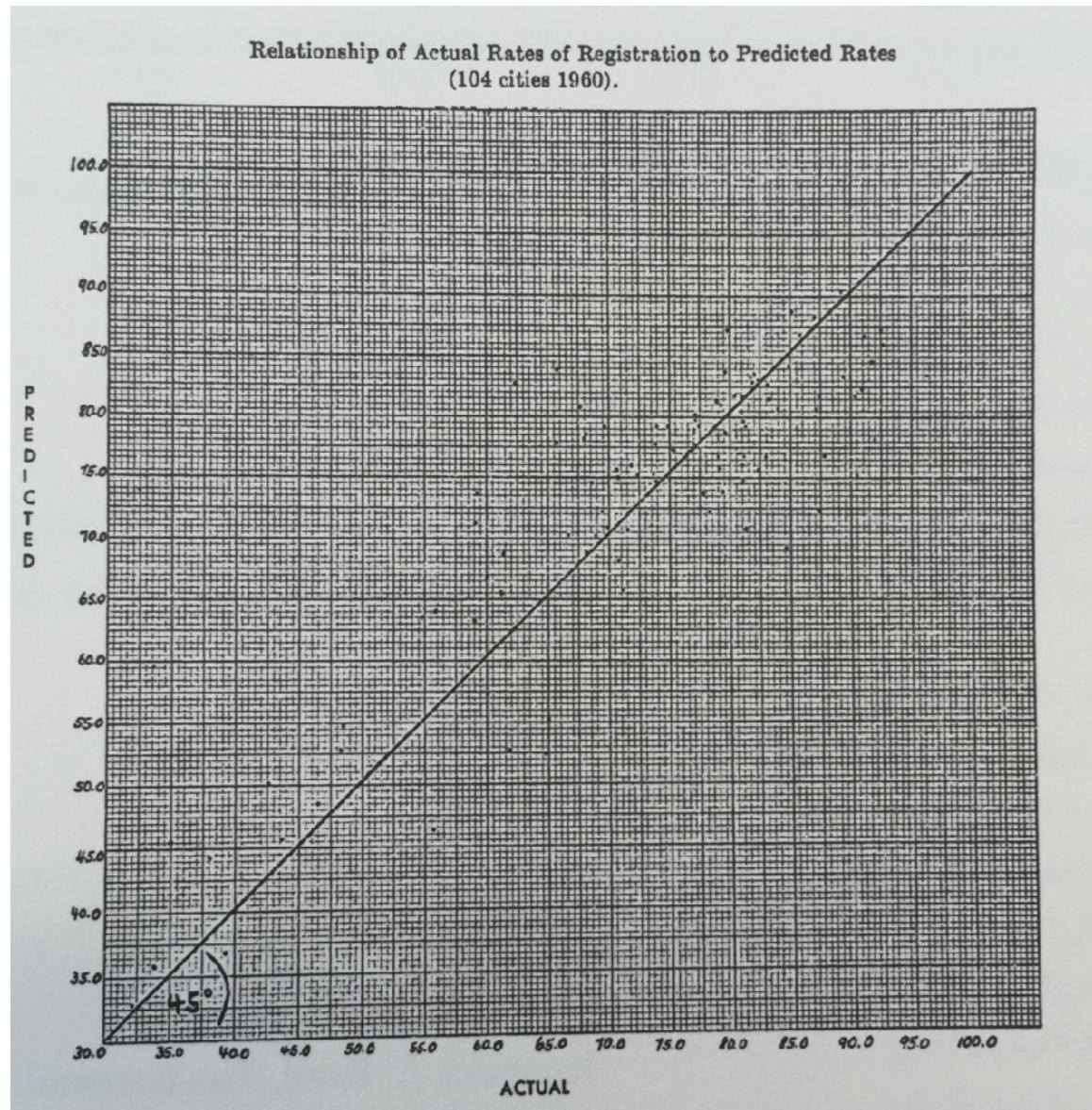
# Data-ink

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

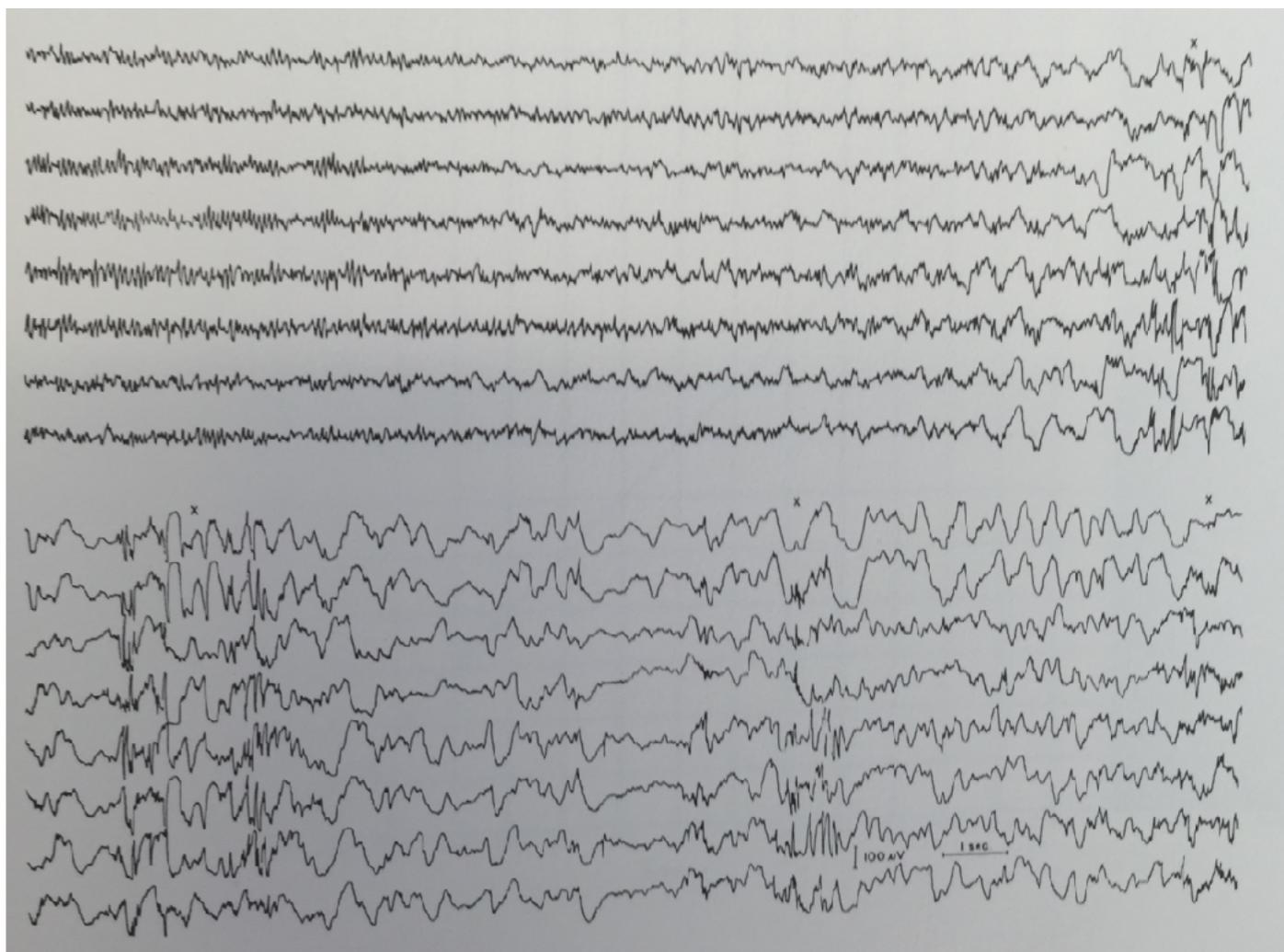
$$\text{Data-ink ratio} = \frac{\text{Data-ink}}{\text{Total ink used to print the graphic}}$$

- = proportion of a graphic's ink devoted to the non-redundant display of data-information
- = 1.0 – proportion of a graphic that can be erased

# Examples of data-ink ratio



$\sim 0$

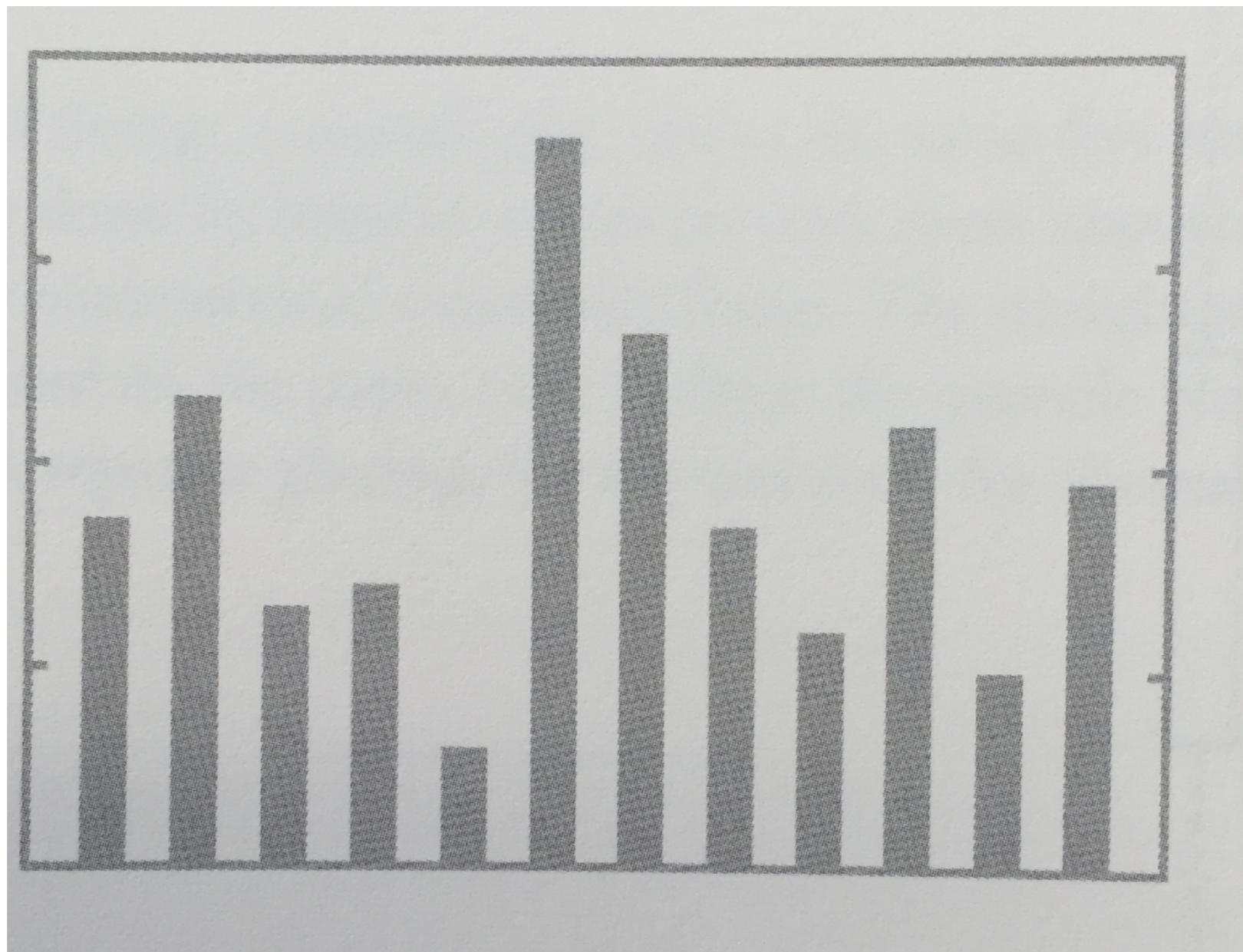


1.0

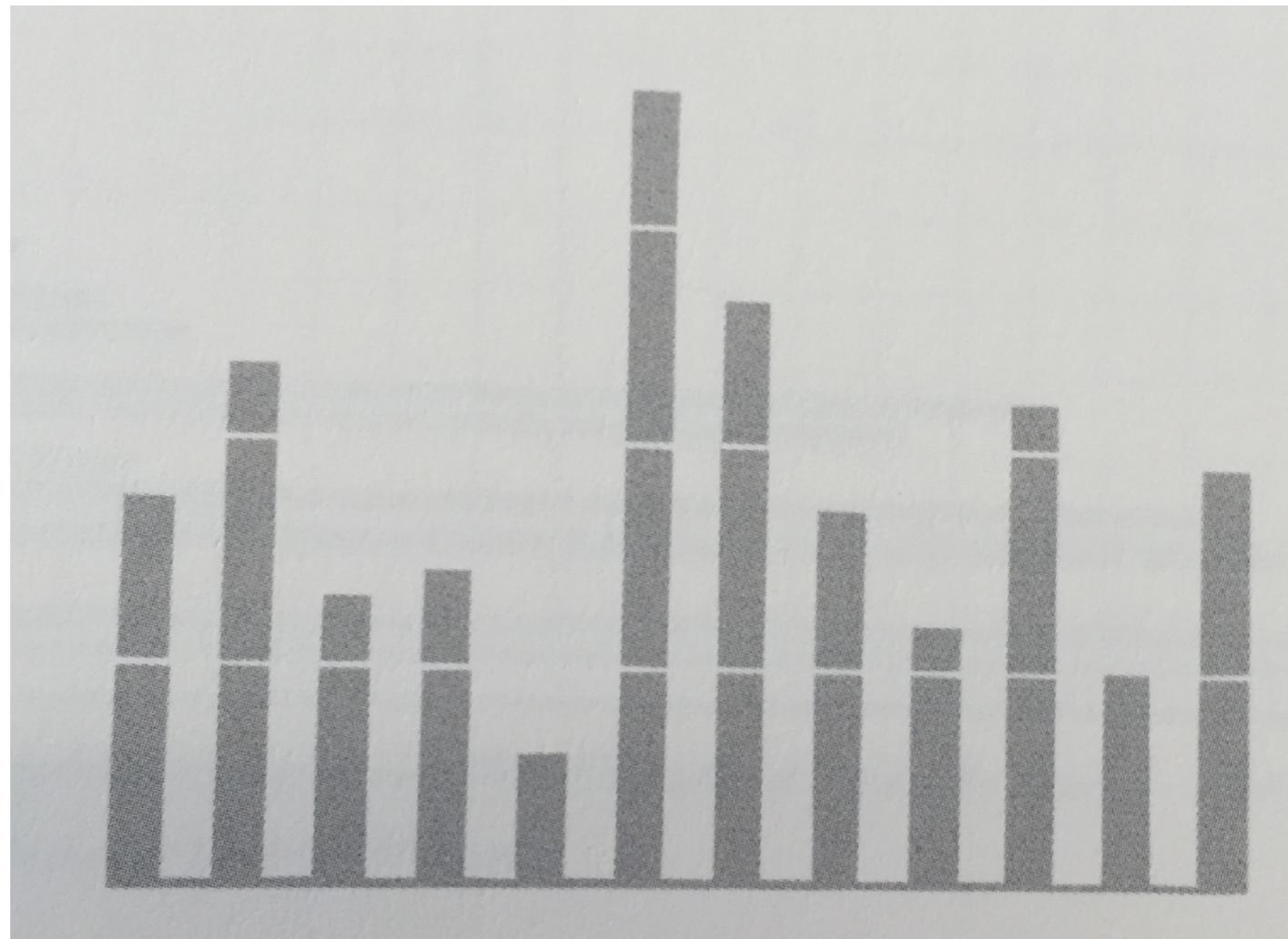
# Design principles for data-ink

- (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

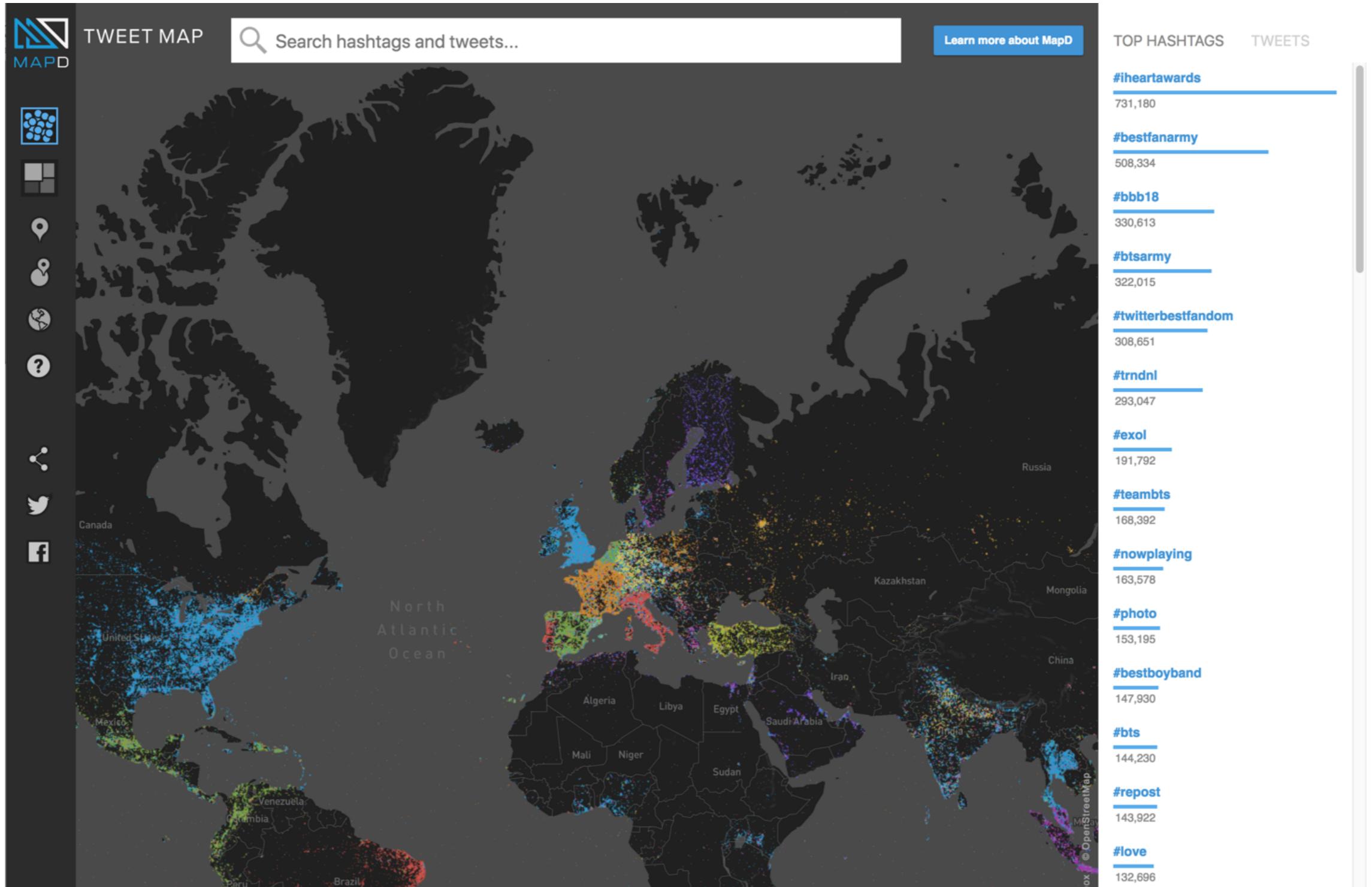
# Interactive visualizations

- 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/>

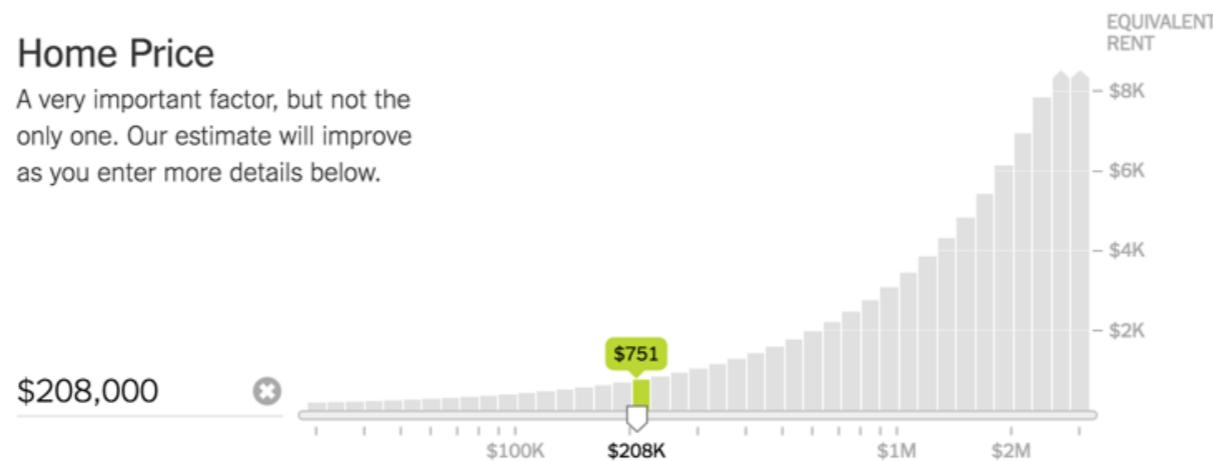
# Is It Better to Rent or Buy?

By MIKE BOSTOCK, SHAN CARTER and ARCHIE TSE

The choice between buying a home and renting one is among the biggest financial decisions that many adults make. But the costs of buying are more varied and complicated than for renting, making it hard to tell which is a better deal. To help you answer this question, our calculator takes the most important costs associated with buying a house and computes the equivalent monthly rent. [RELATED ARTICLE](#)

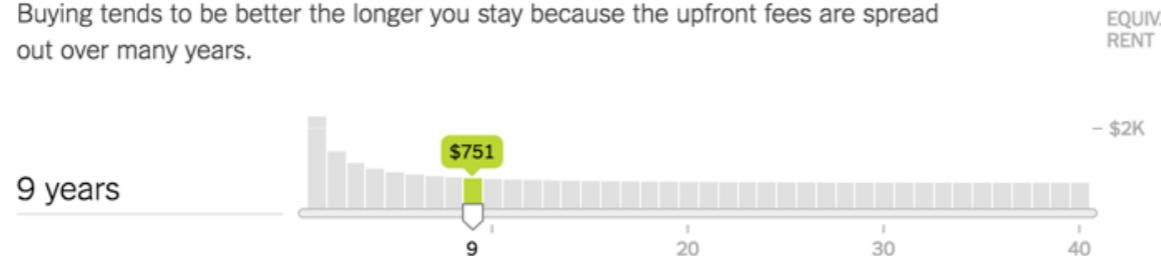
## Home Price

A very important factor, but not the only one. Our estimate will improve as you enter more details below.



## How Long Do You Plan to Stay?

Buying tends to be better the longer you stay because the upfront fees are spread out over many years.



If you can rent a similar home for less than ...

**\$751** PER MONTH

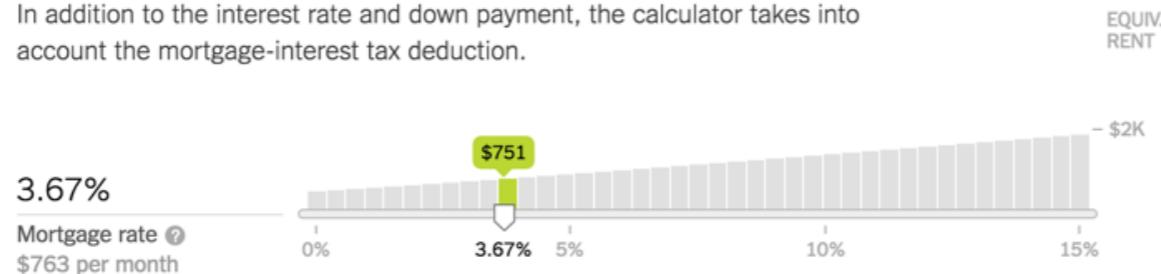
... then renting is better.

	Rent	Buy
Initial costs	\$751	\$49,920
Recurring costs	\$90,937	\$137,912
Opportunity costs	\$13,092	\$37,376
Net proceeds	-\$751	-\$121,180
Total	\$104,029	\$104,029

**How to Read the Charts** Charts that are relatively flat indicate factors that are not particularly important to the outcome. Conversely, the factors that have steep slopes have a large impact.

## What Are Your Mortgage Details?

In addition to the interest rate and down payment, the calculator takes into account the mortgage-interest tax deduction.



[https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html?\\_r=0](https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html?_r=0)

# 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