Information Visualization

SWE 432, Fall 2017

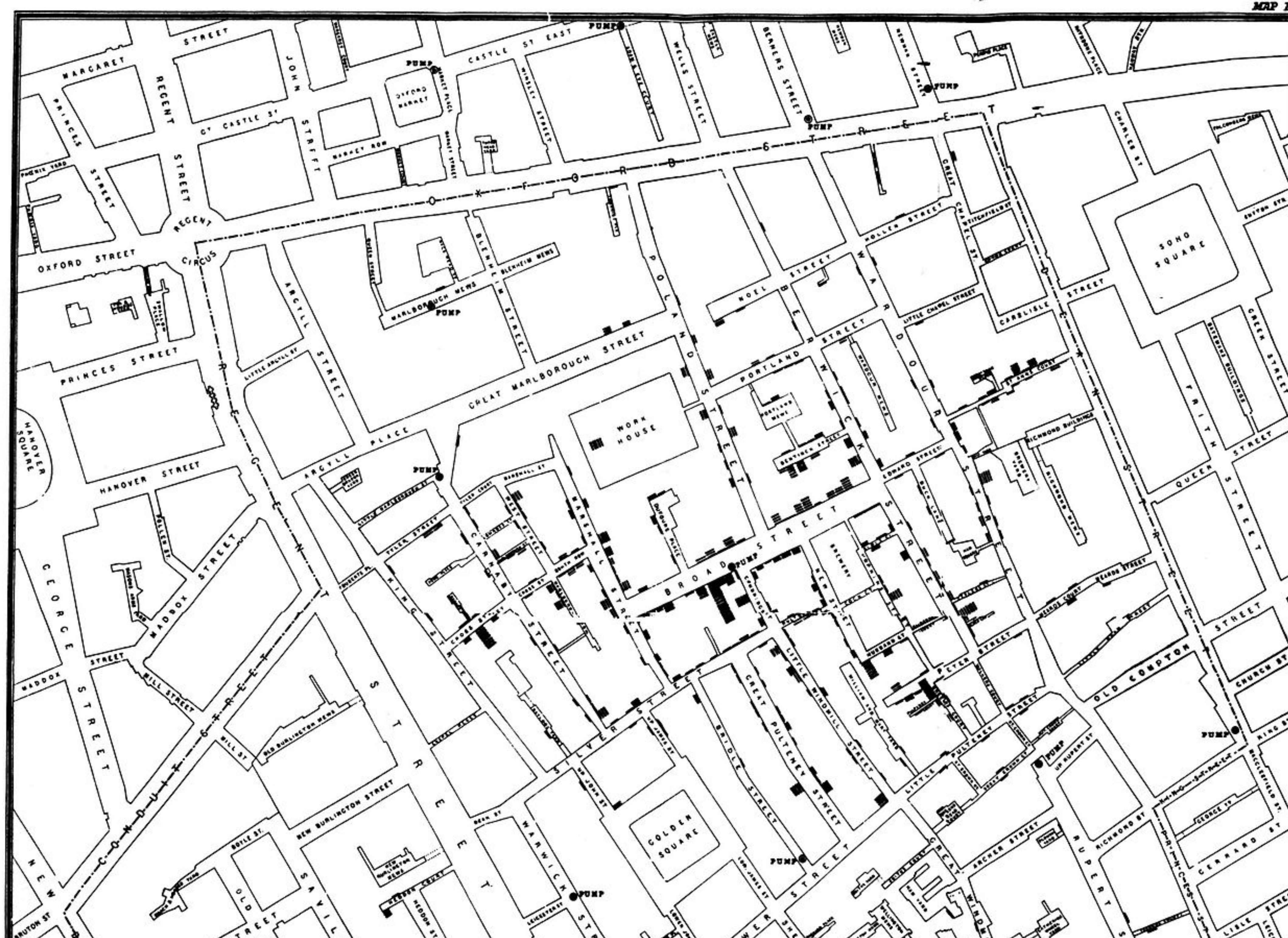
Design and Implementation of Software for the Web

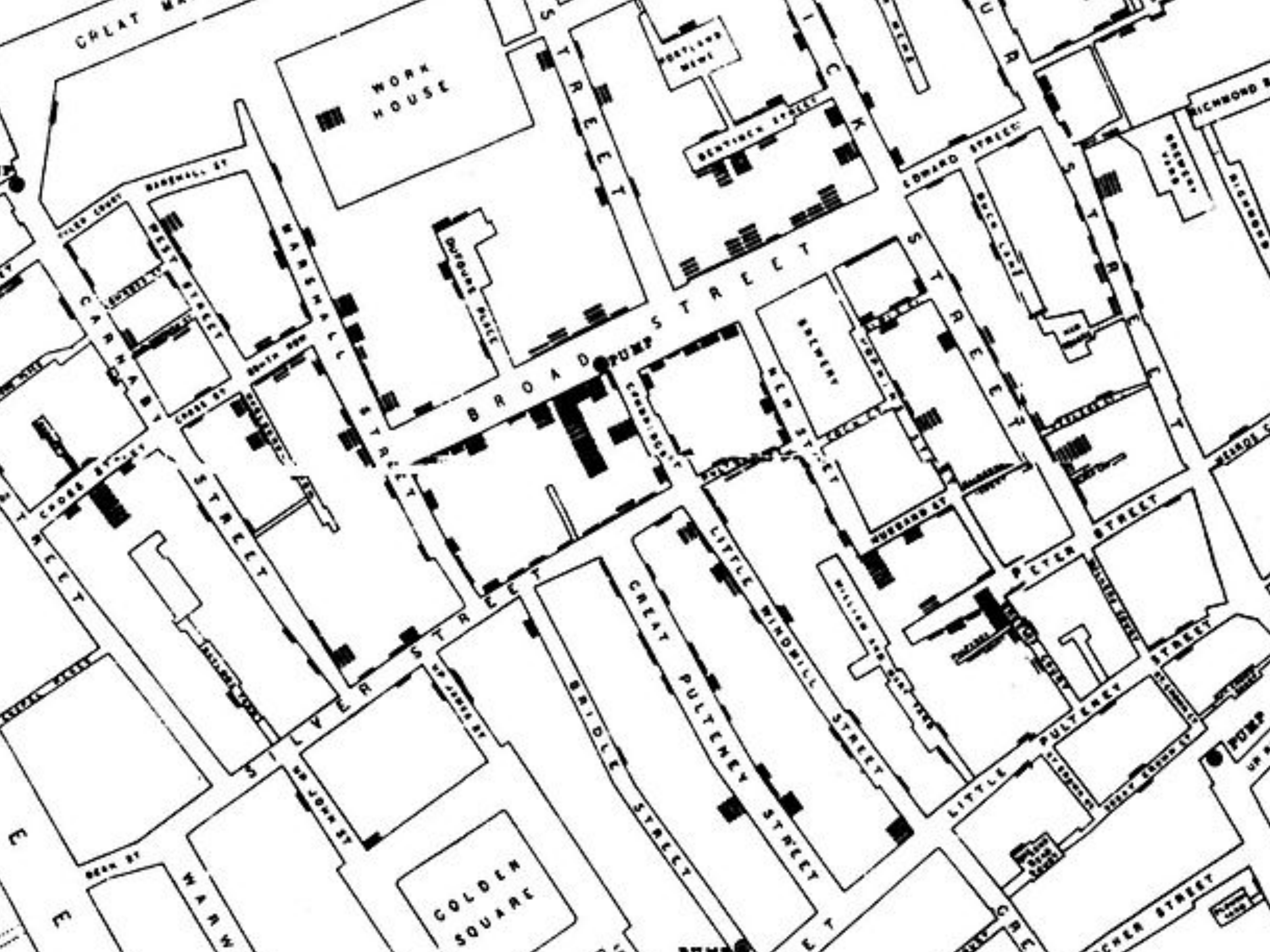
Today

- What types of information visualization are there?
  - Which one should you choose?
- What does usability mean for 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


MAP 1.
OXFORD STREET
REGENT STREET
REGENT CIRCUS
MARGARET STREET
PRINCES STREET
GT CASTLE ST
CASTLE ST EAST
JOHN STREET
OXFORD MARKET
WELLS STREET
BERNERS STREET
NEWMAN STREET
CHARLES ST
SOHO SQUARE
GREEK STREET
FRITH STREET
DEAN STREET
WARDOUR STREET
BERWICK STREET
POLAND STREET
NOEL STREET
PORTLAND STREET
GREAT MARLBOROUGH STREET
MARLBOROUGH MEWS
BLENHEIM STREET
ARGYLL STREET
ARGYLL PLACE
WORK HOUSE
BROAD STREET
MARSHALL STREET
CARNABY STREET
KING STREET
SILVER STREET
CAMBRIDGE STREET
GREAT PULTENEY STREET
LITTLE WINDMILL STREET
BRIDLE STREET
GOLDEN SQUARE
WARWICK STREET
BREWER STREET
QUEEN STREET
OLD COMPTON STREET
CHURCH ST
RUPERT STREET
GERRARD STREET
LISLE STREET
PRINCES STREET
COVENTRY STREET
PICCADILLY
REGENT CIRCUS
REGENTS QUADRANT
HAYMARKET
JERMYN STREET
PANTON STREET
SACKVILLE STREET
ALBANY
VIGO STREET
BURLINGTON GARDENS
OLD BURLINGTON STREET
NEW BURLINGTON STREET
SAVILLE ROW
CORK STREET
CLIFFORD STREET
GRAFTON STREET
ALBEMARLE STREET
DOVER STREET
NEW BOND STREET
CONDUIT STREET
MADDOX STREET
GEORGE STREET
HANOVER SQUARE
HANOVER STREET
HANOVER STREET
MILL STREET
BRUTON ST
SWALLOW STREET
AIR STREET
VINE STREET
GLASSHOUSE ST
TICHBORNE STREET
MARYLEBONE STREET
HARDWICK STREET
RICHMOND BUILDINGS
QUEEN STREET
PUMP
C. F. Cheffins, Lith, Southampton Bgs. London.
SCALE 30 INCHES TO A MILE.


WORK HOUSE
PORTLAND MEWS
BENTINCK STREET
EDWARD STREET
RICHMOND ST
STREET
MARSHALL ST
DUFOURS PLACE
BROAD STREET
PUMP
CARNABY
WEST STREET
CROSS ST
CROSS STREET
STREET
SILVER STREET
BREWERY
NEW STREET
HUBBARD ST
PETER STREET
GREAT PULTENEY STREET
LITTLE WINDMILL STREET
BRIDLE STREET
LITTLE PULTENEY STREET
GOLDEN SQUARE
PUMP

Investigation and aftermath

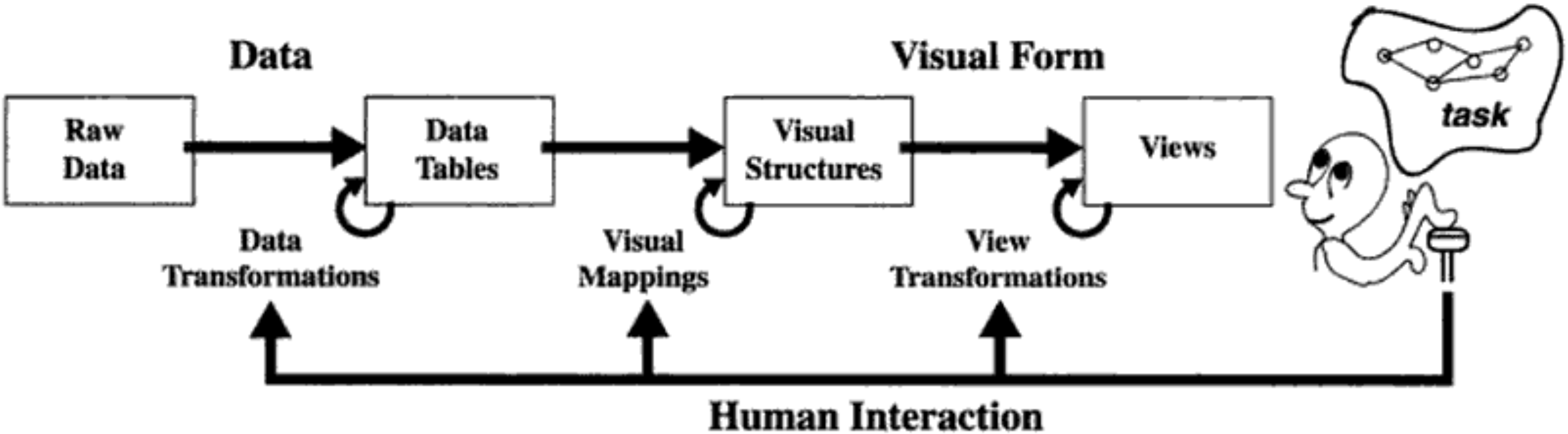
- Based on **visualization**, did case by case investigation
- Found that **61 / 83** positive identified as using well water from Broad Street pump
- Board ordered pump-handle to be removed from well
- Epidemic soon **ended**
- Solved centuries old question of how cholera spread

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

Mapping data to visual form

Designing an information visualization



Raw Data: idiosyncratic formats
Data Tables: relations (cases by variables) + metadata
Visual Structures: spatial substrates + marks + graphical properties
Views: graphical parameters (position, scaling, clipping, ...)

Types of raw data

- Nominal - unordered set
- Ordinal - ordered set
- Quantitative - numeric range

Data transformations

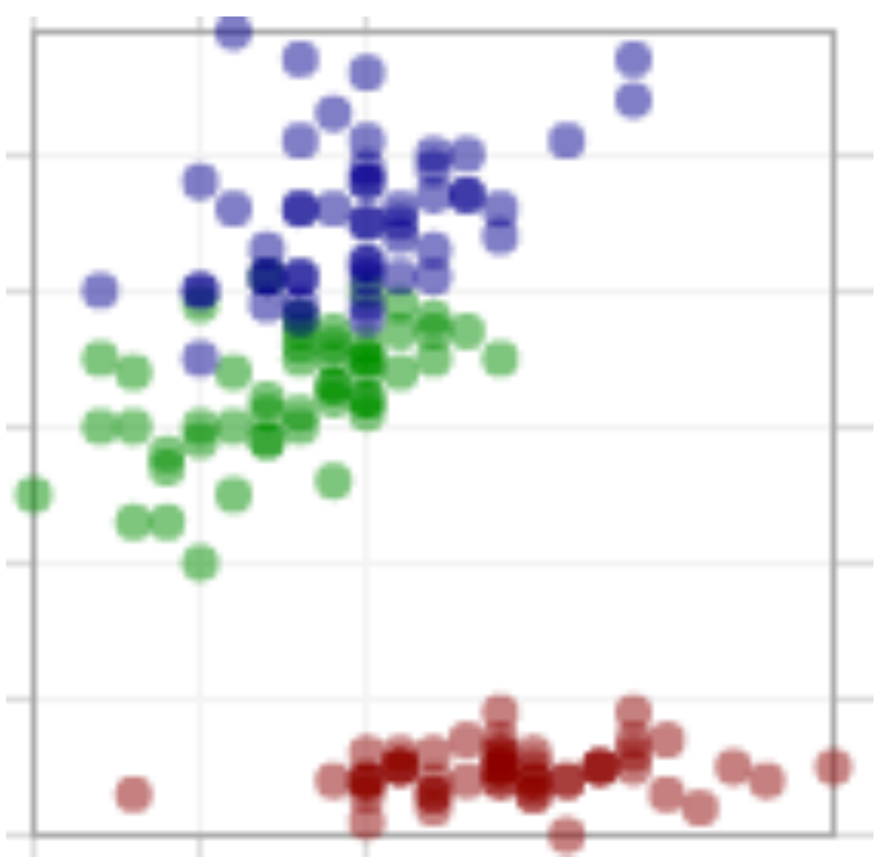
- Classing / binning: Quantitative —> ordinal
  - Maps ranges onto **classes** of variables
  - Can also count # of items in each class w/ histogram
- Sorting: Nominal —> ordinal
  - Add order between items in sets
- Descriptive statistics: mean, average, median, max, min, …

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)

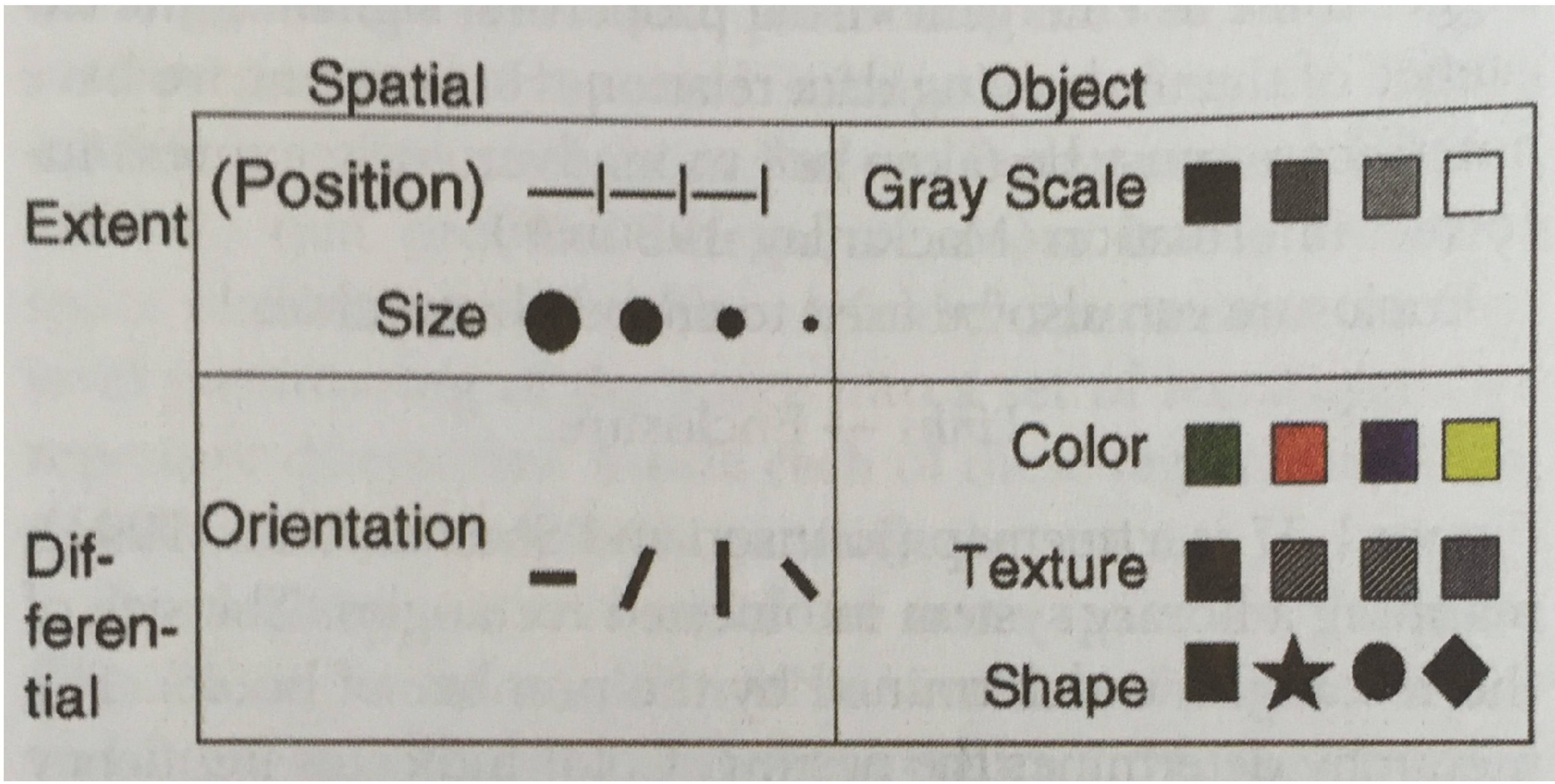


Marks

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

Marks' graphical properties

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



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	○	○	●

Animation

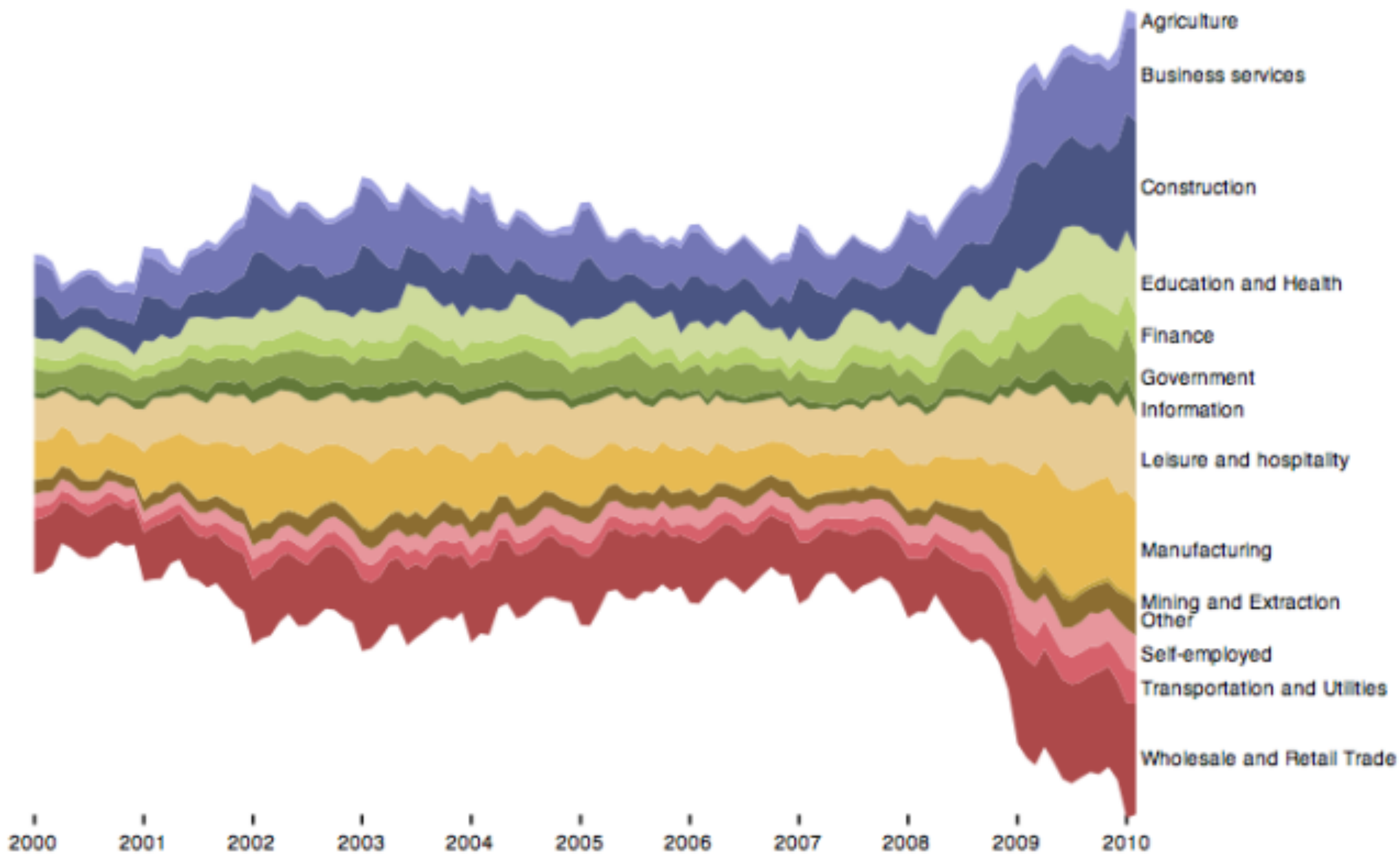
- Visualization can change over time
- Could be used to encode data as a function of time
  - But often not effective as makes direct comparisons hard
- Can be more effective to animate transition from before to after as user configures visualization

Examples of visualizations

Time-series data

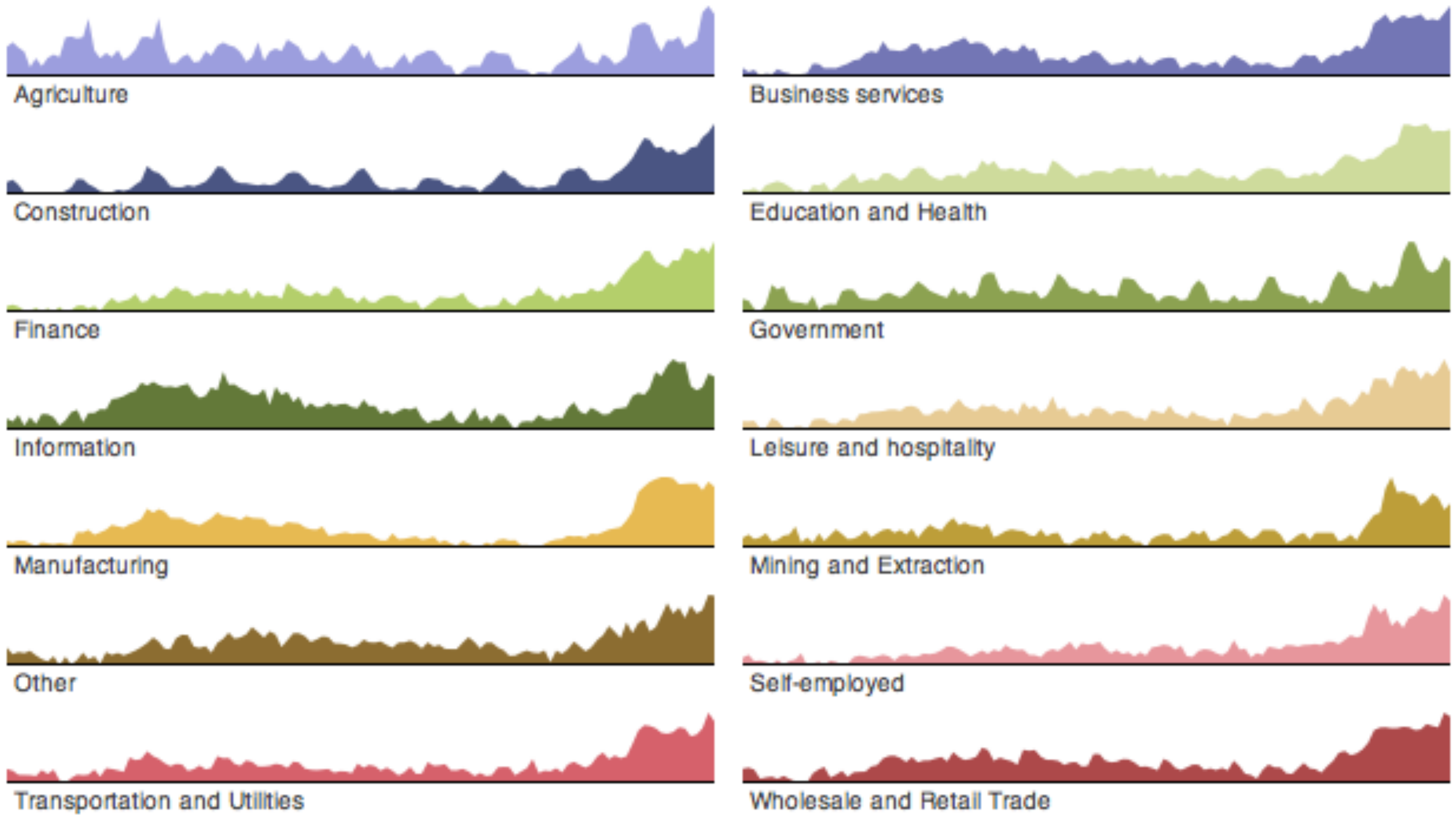
Stacked graph

- Supports visual summation of multiple components



Small multiples

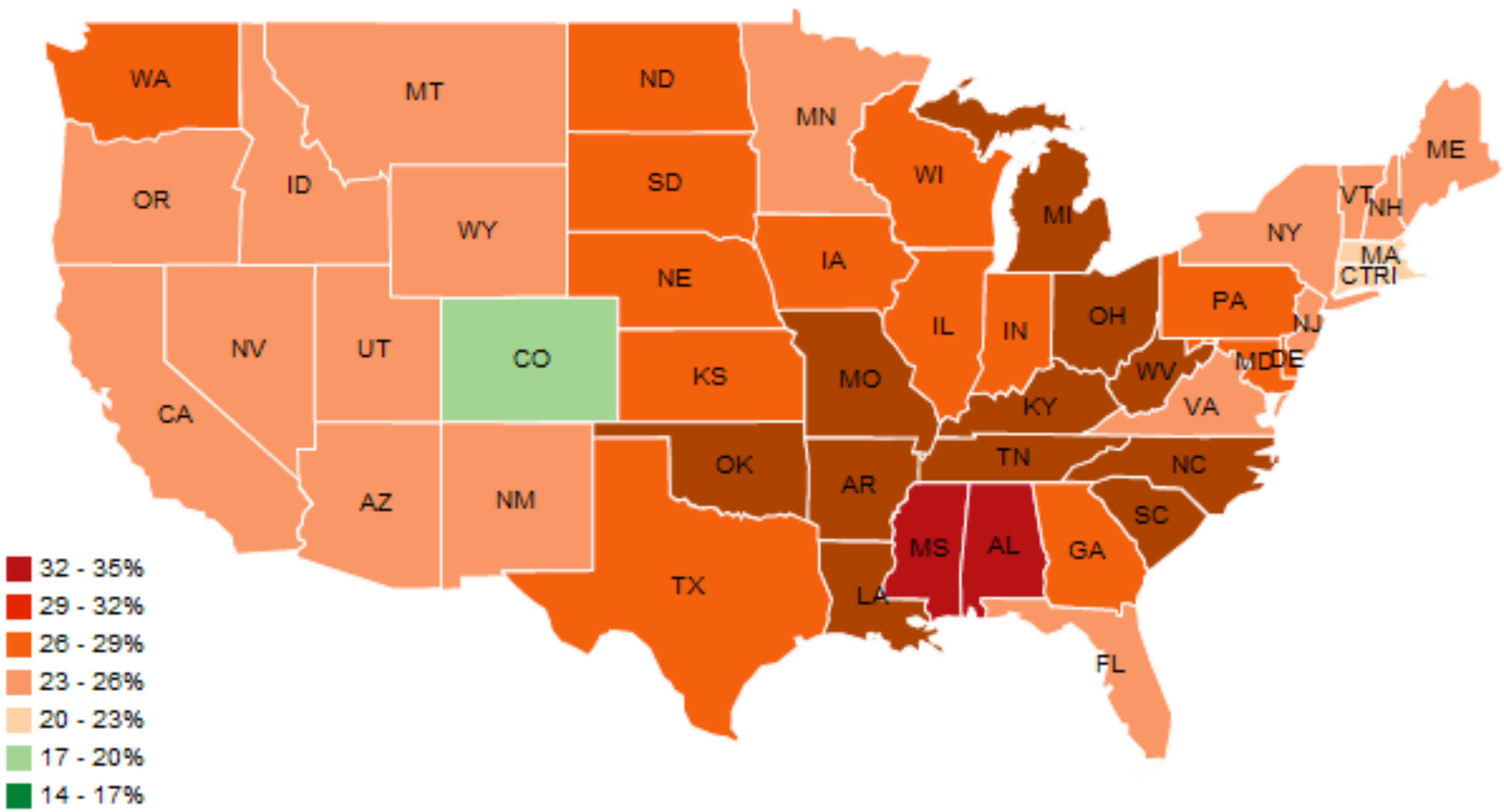
- supports separate comparison of data series
- may have better legibility than placing all in single plot



Maps

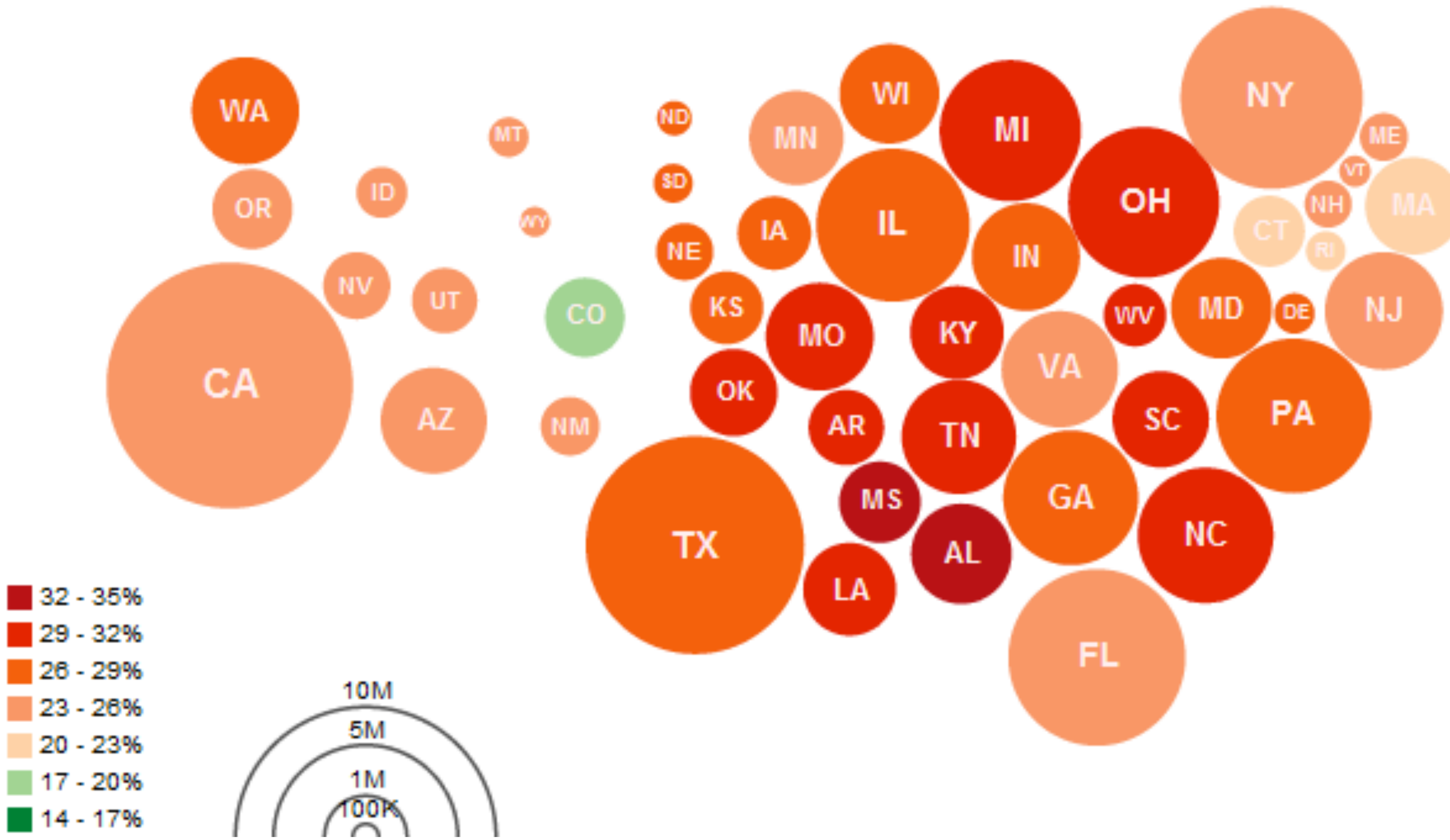
Choropleth map

- Groups data by area, maps to color



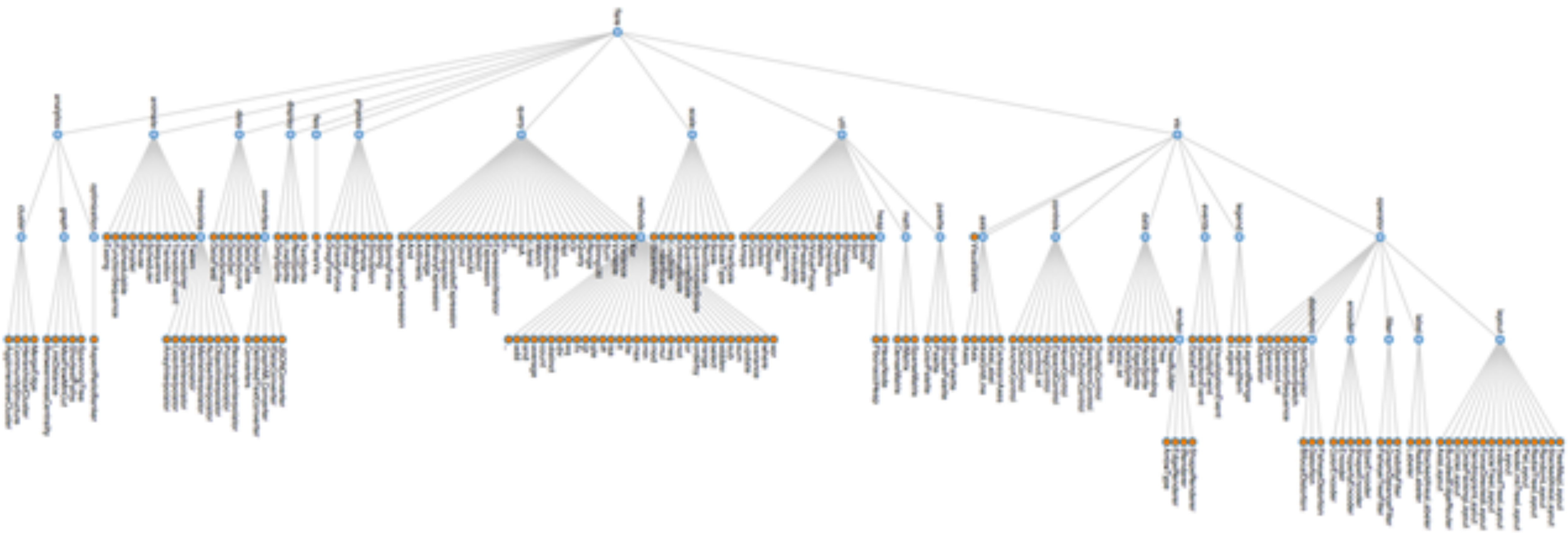
Cartograms

- Encodes two variables w/ size & color



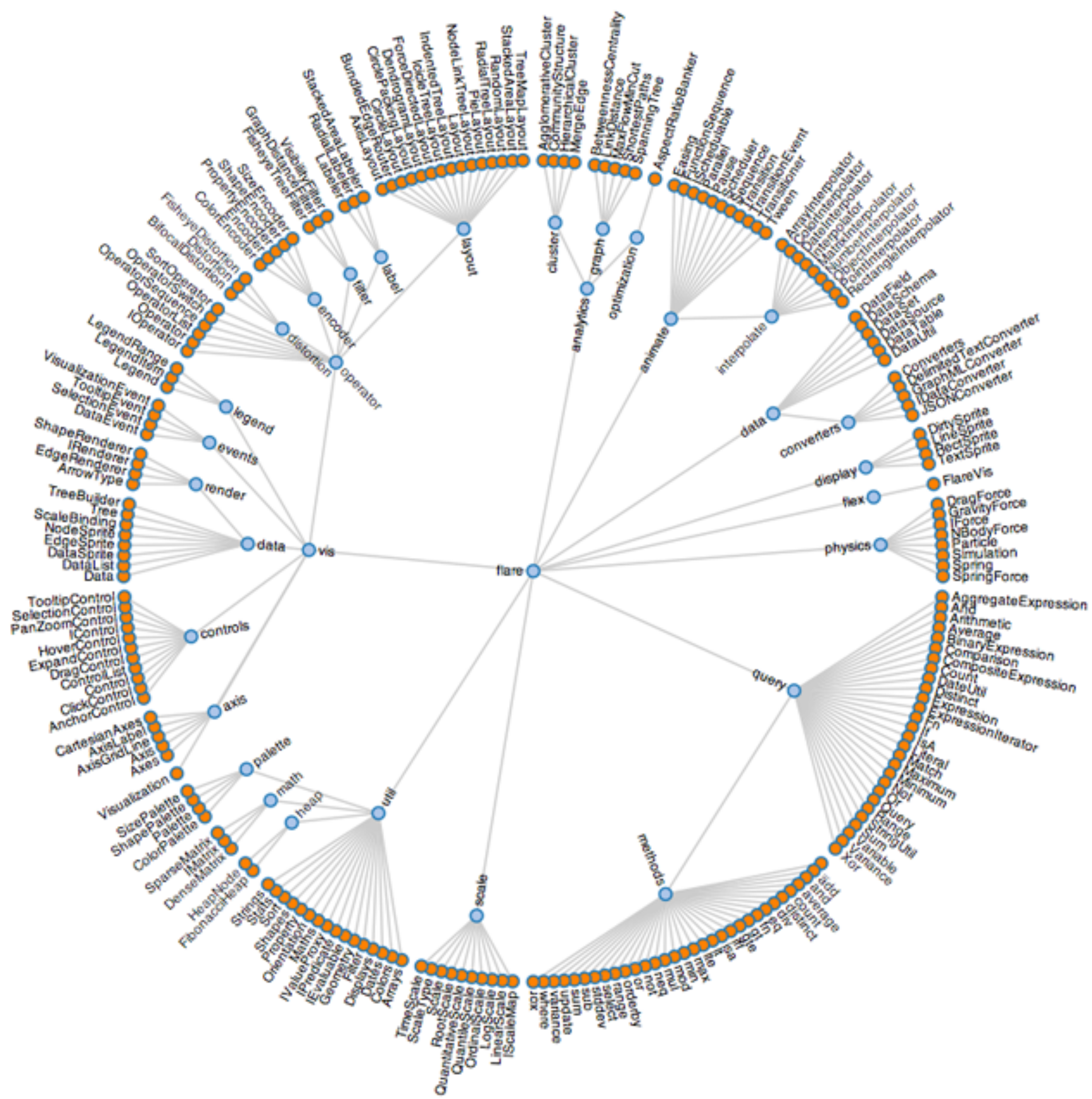
Hierarchies

Node link diagram

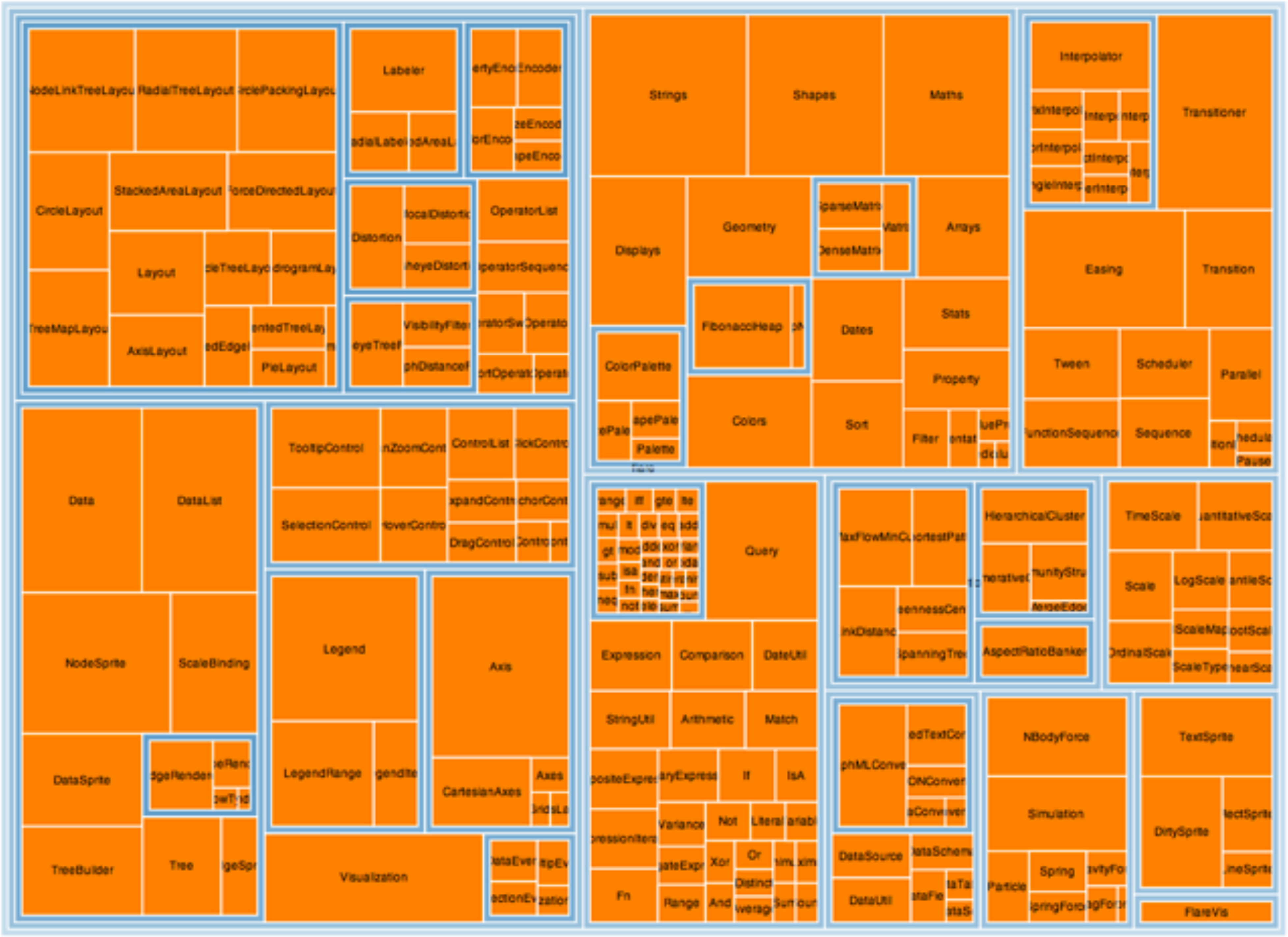


Dendrogram

- leaf nodes of hierarchy on edges of circle



Treemaps


Labeler
RadialTreeLayout
CircleLayout
StackedAreaLayout
Layout
AxisLayout
PieLayout
Distortion
OperatorList
VisibilityFilter
Strings
Shapes
Maths
Interpolator
Transitioner
Displays
Geometry
Arrays
Easing
Transition
FibonacciHeap
Dates
Stats
ColorPalette
Property
Palette
Colors
Sort
Filter
Tween
Scheduler
Parallel
Sequence
Pause
Data
DataList
TooltipControl
ControlList
SelectionControl
DragControl
NodeSprite
ScaleBinding
Legend
Axis
DataSprite
LegendRange
CartesianAxes
Axes
TreeBuilder
Tree
Visualization
Query
Expression
Comparison
DateUtil
StringUtil
Arithmetic
Match
If
IsA
Variance
Not
Xor
Or
Distinct
Fn
Range
And
HierarchicalCluster
AspectRatioBanker
TimeScale
Scale
LogScale
DataSource
DataUtil
NBodyForce
Simulation
Particle
Spring
TextSprite
DirtySprite
FlareVis

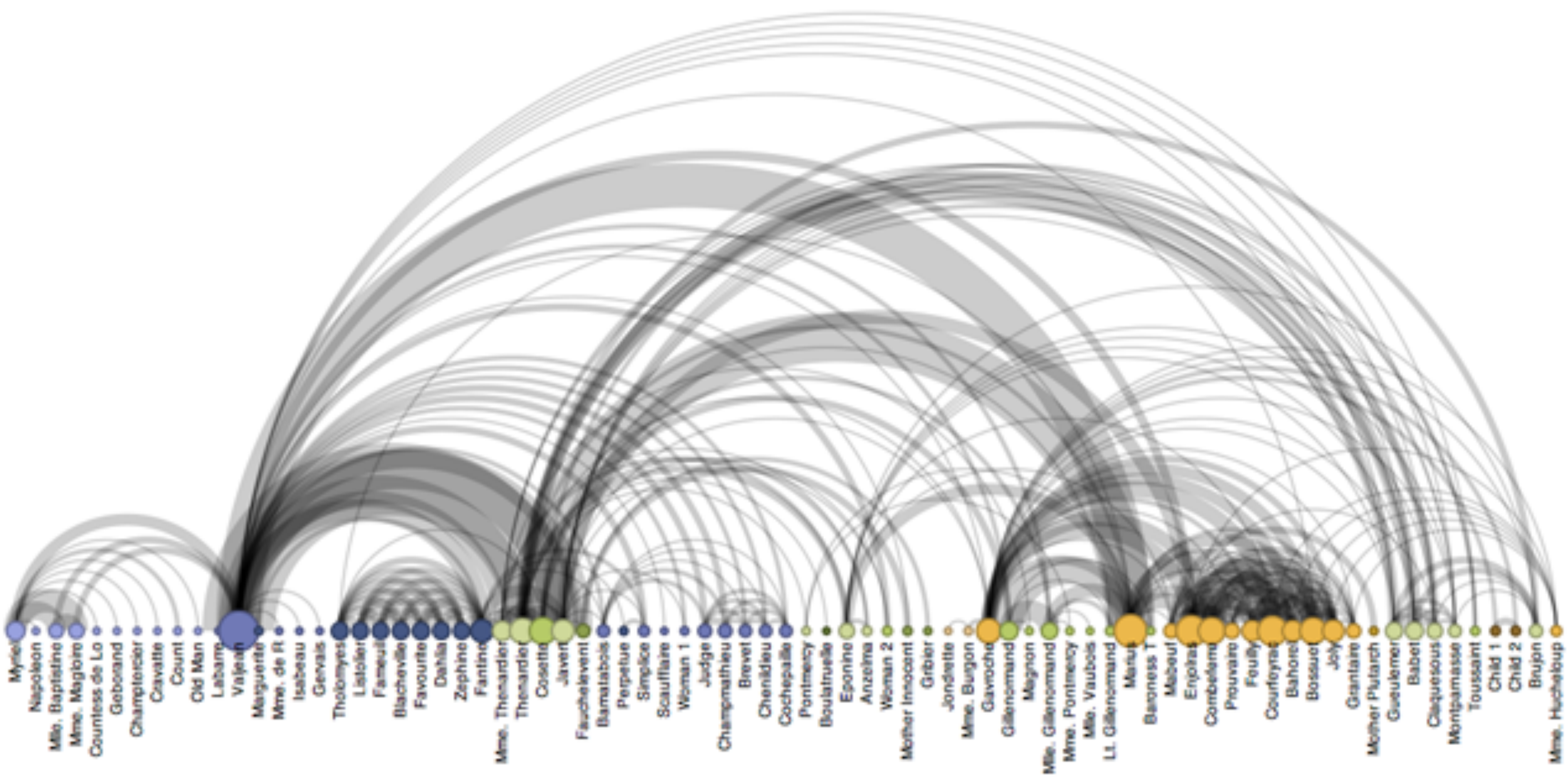
Networks

Force-directed layout

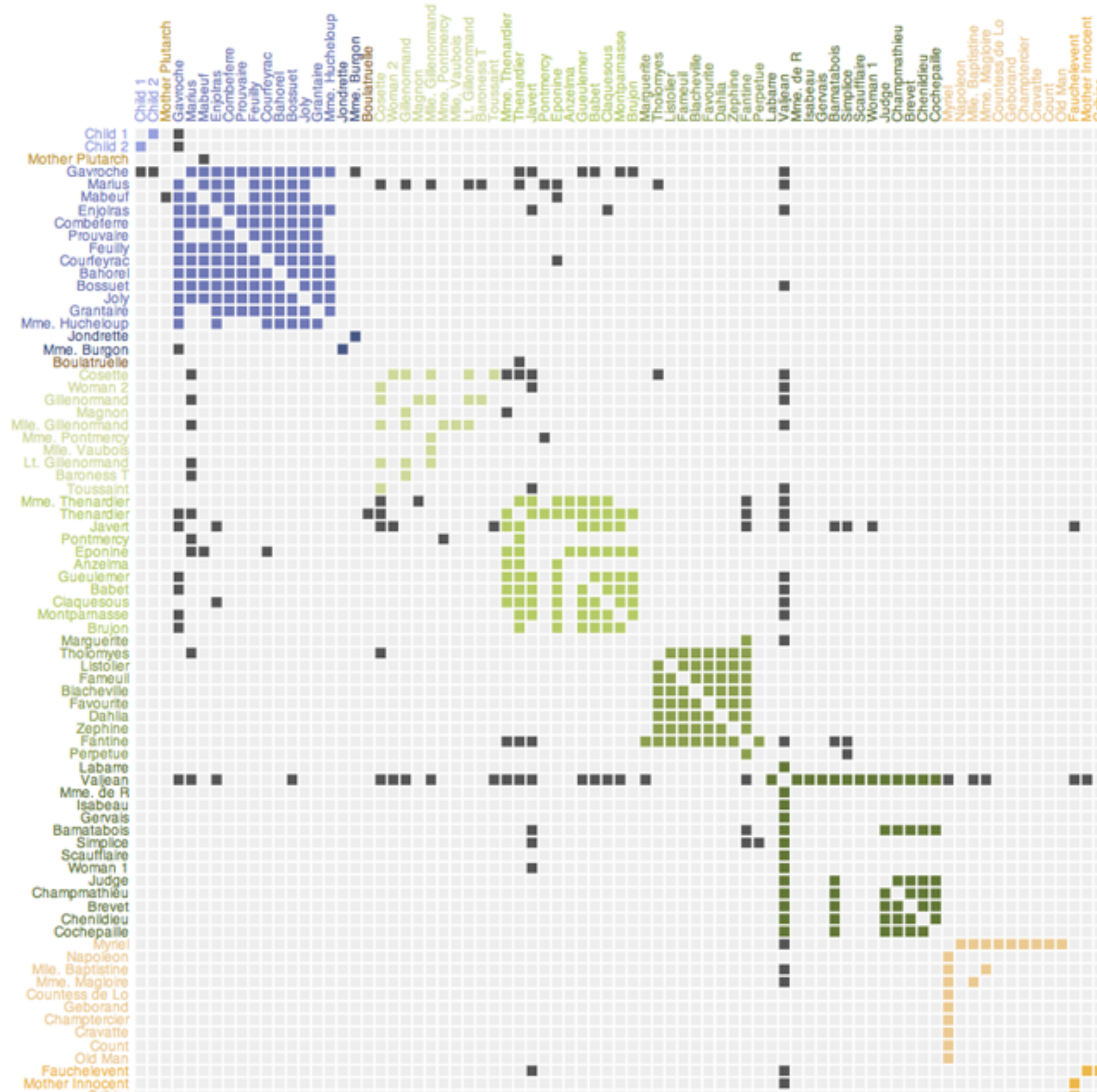
- edges function as springs, find least energy configuration

Arc diagram

- can support identifying cliques & bridges w/ right order



Adjacency matrix



Design considerations

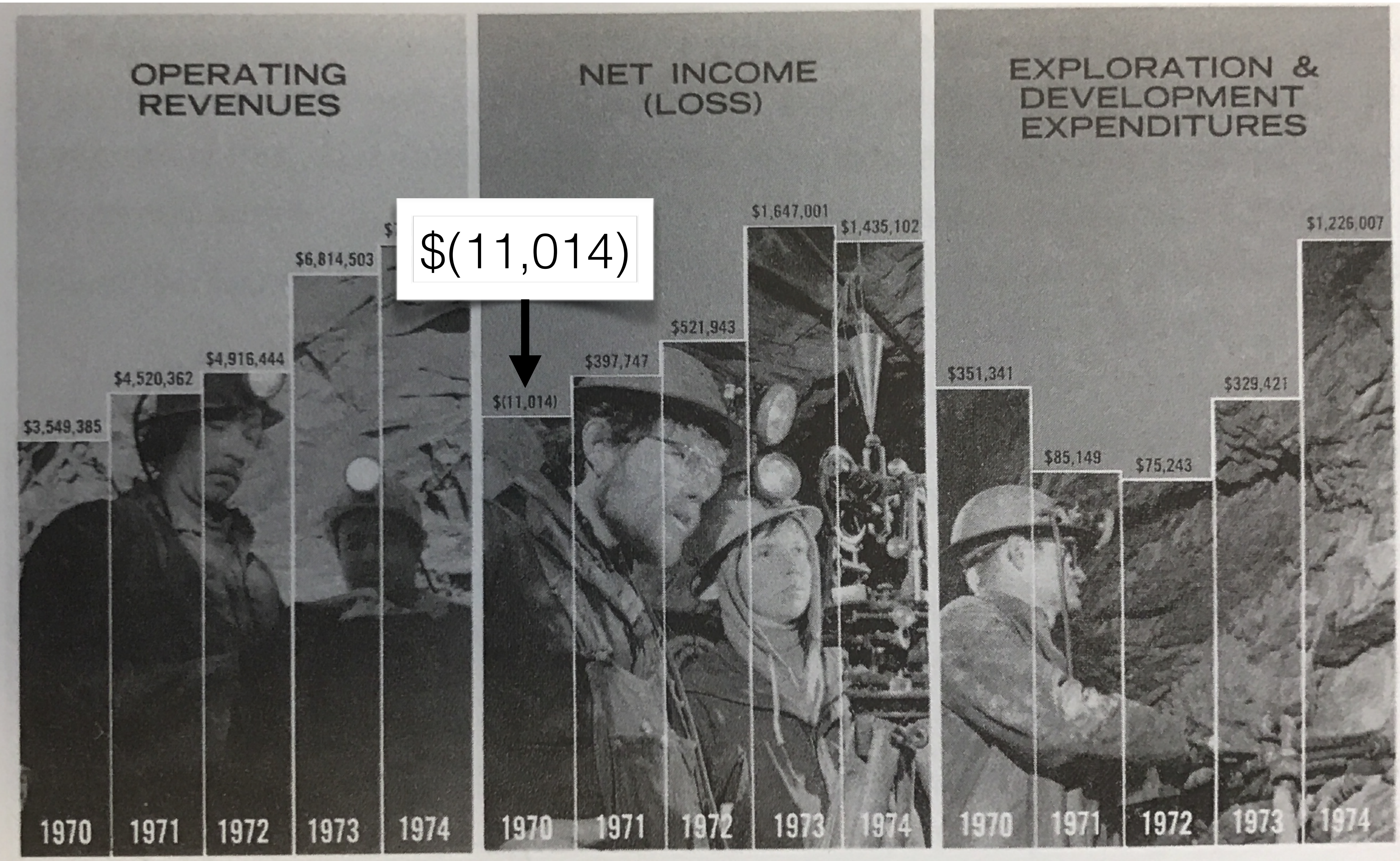
Tufte's principles of graphical excellence

- 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

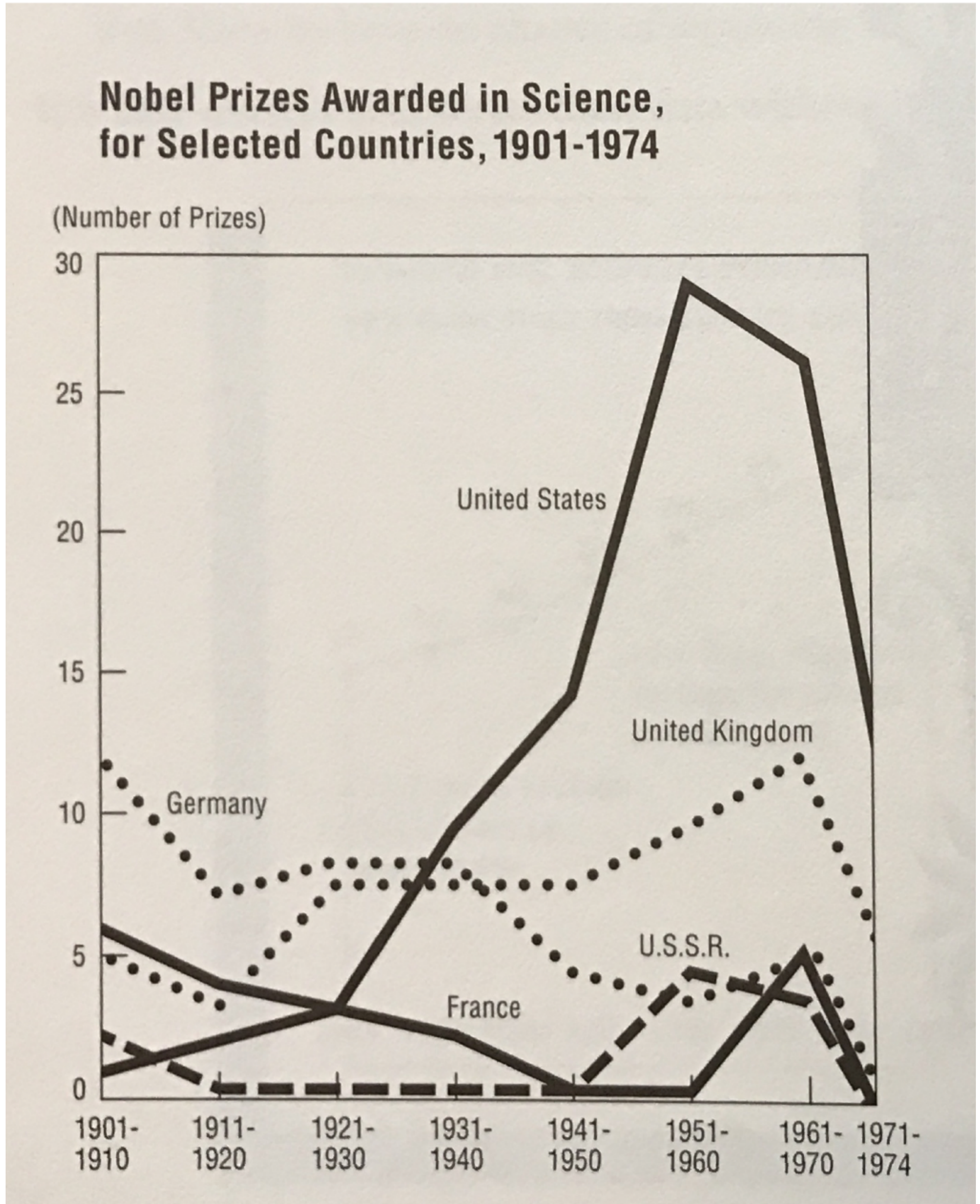
- Visualizations may distort the underlying data, making it harder for reader to understand truth
- Use of **design** variation to try to falsely communicate **data** variation

Example



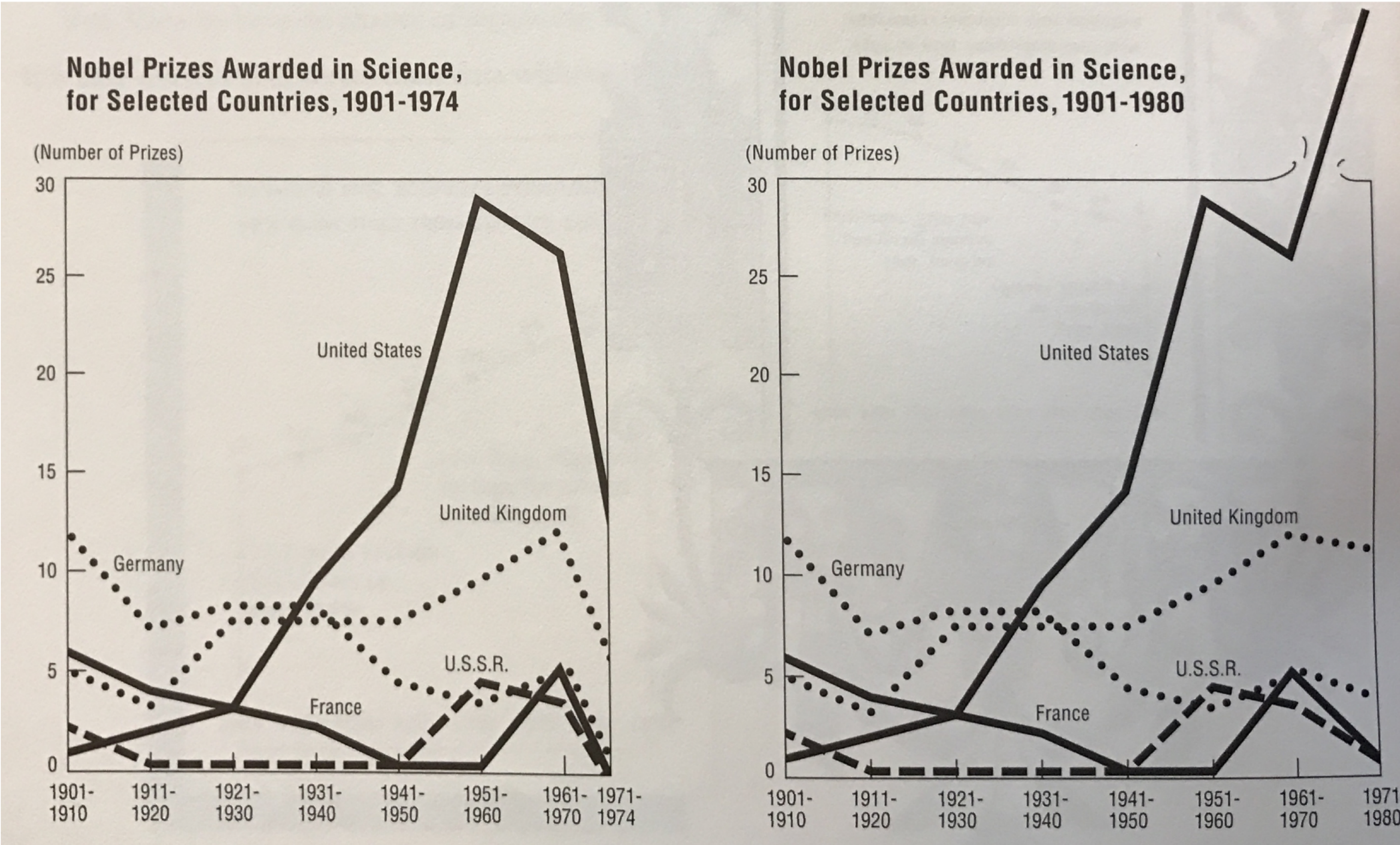
OPERATING REVENUES
$3,549,385
$4,520,362
$4,916,444
$6,814,503
1970
1971
1972
1973
1974
NET INCOME (LOSS)
$(11,014)
$(11,014)
$397,747
$521,943
$1,647,001
$1,435,102
1970
1971
1972
1973
1974
EXPLORATION & DEVELOPMENT EXPENDITURES
$351,341
$85,149
$75,243
$329,421
$1,226,007
1970
1971
1972
1973
1974

Example

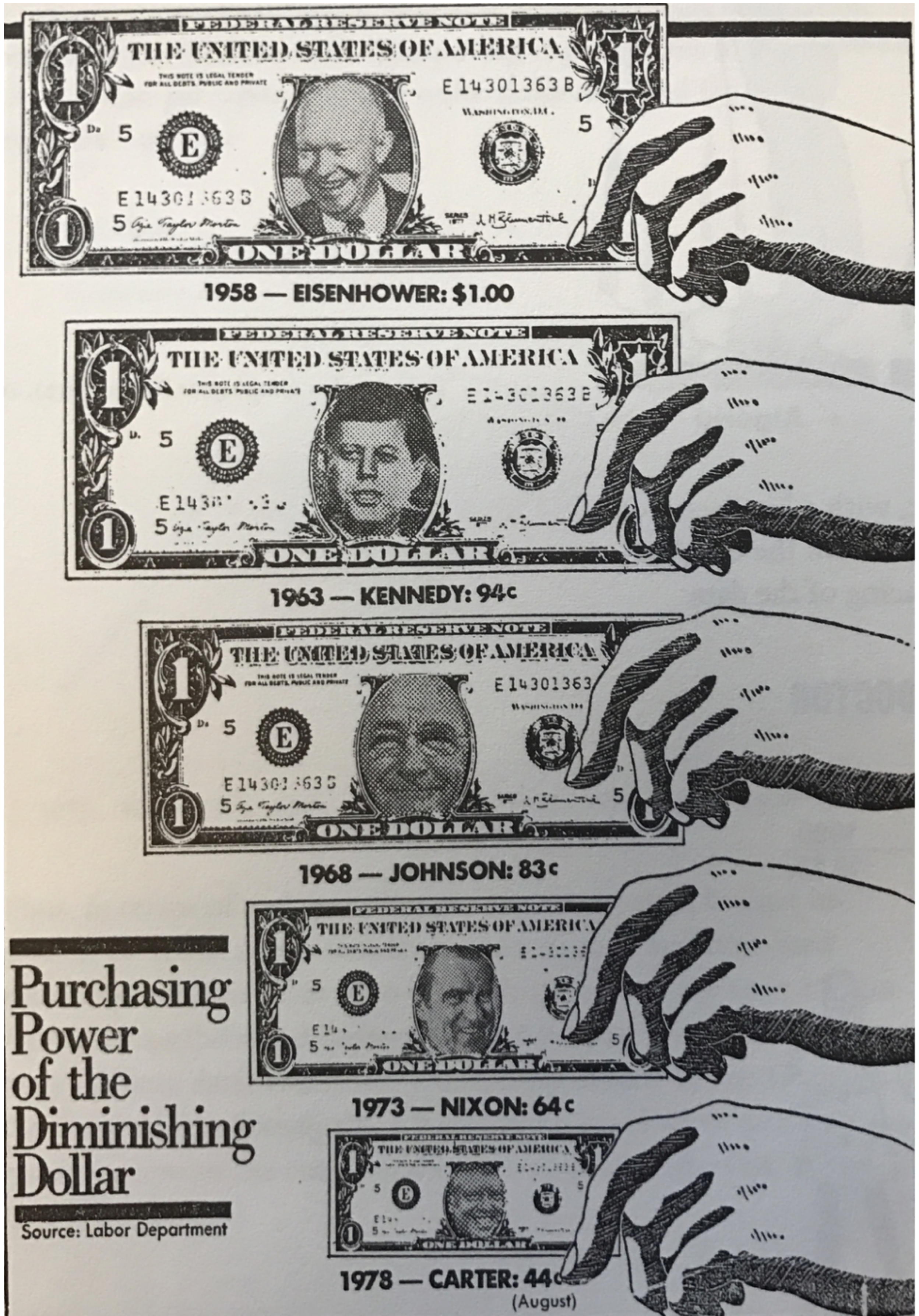


Nobel Prizes Awarded in Science, for Selected Countries, 1901-1974
(Number of Prizes)
30
25
20
15
10
5
0
United States
United Kingdom
Germany
U.S.S.R.
France
1901-1910
1911-1920
1921-1930
1931-1940
1941-1950
1951-1960
1961-1970
1971-1974

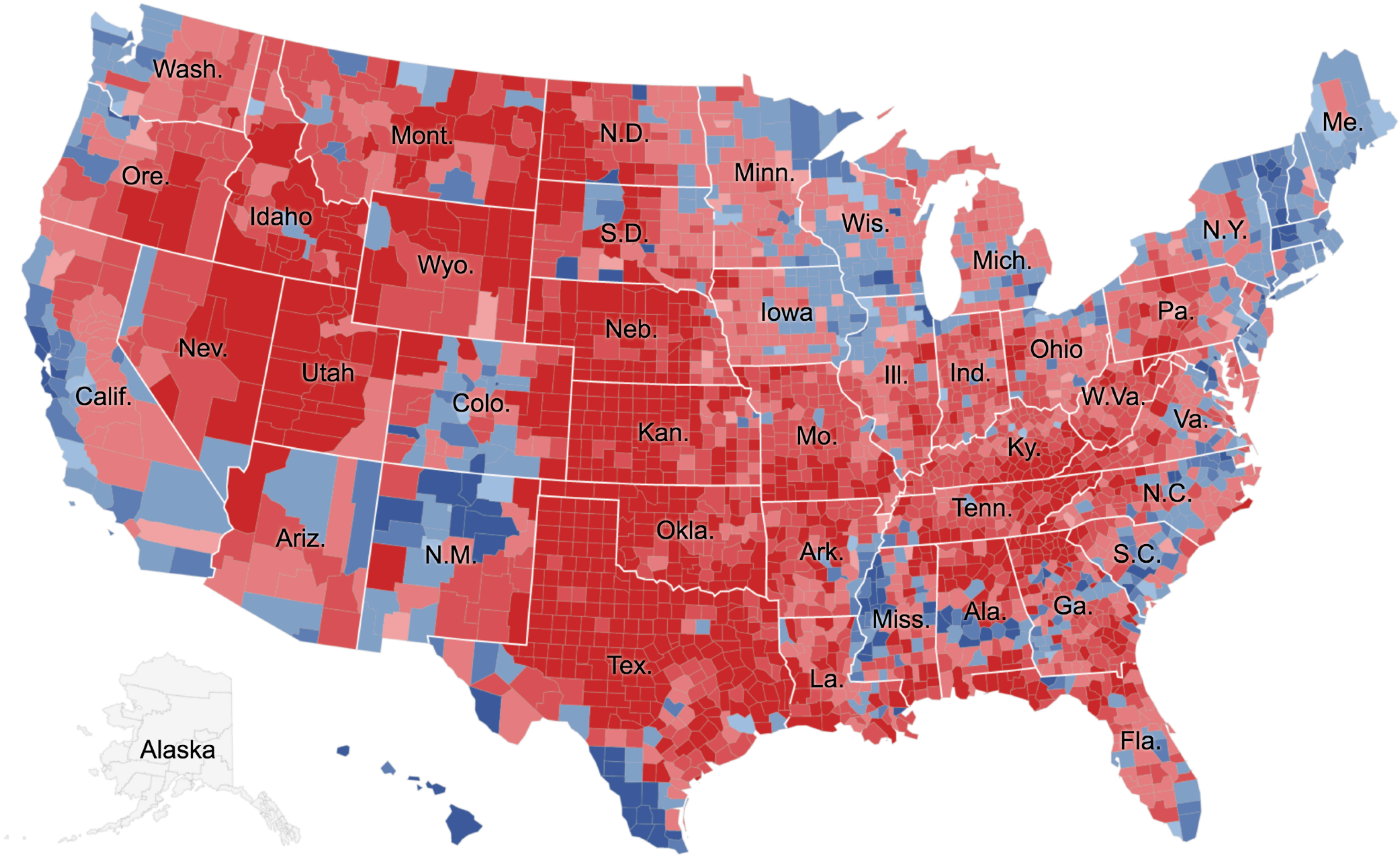
Example (corrected)



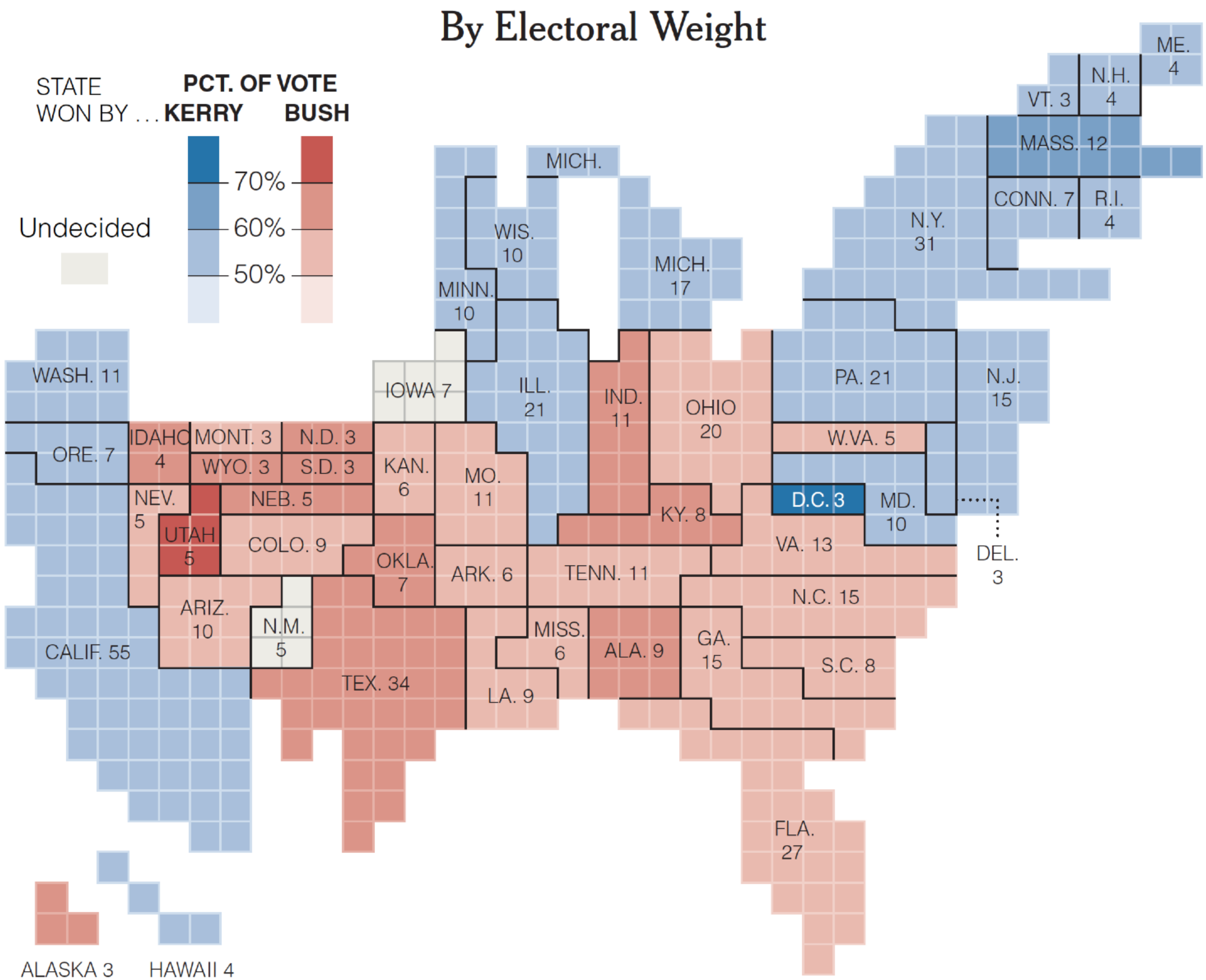
Example


1958 — EISENHOWER: $1.00
1963 — KENNEDY: 94¢
1968 — JOHNSON: 83¢
1973 — NIXON: 64¢
1978 — CARTER: 44¢
(August)
Purchasing Power of the Diminishing Dollar
Source: Labor Department

Traditional Electoral Map



Weighted Electoral Map



Data-ink

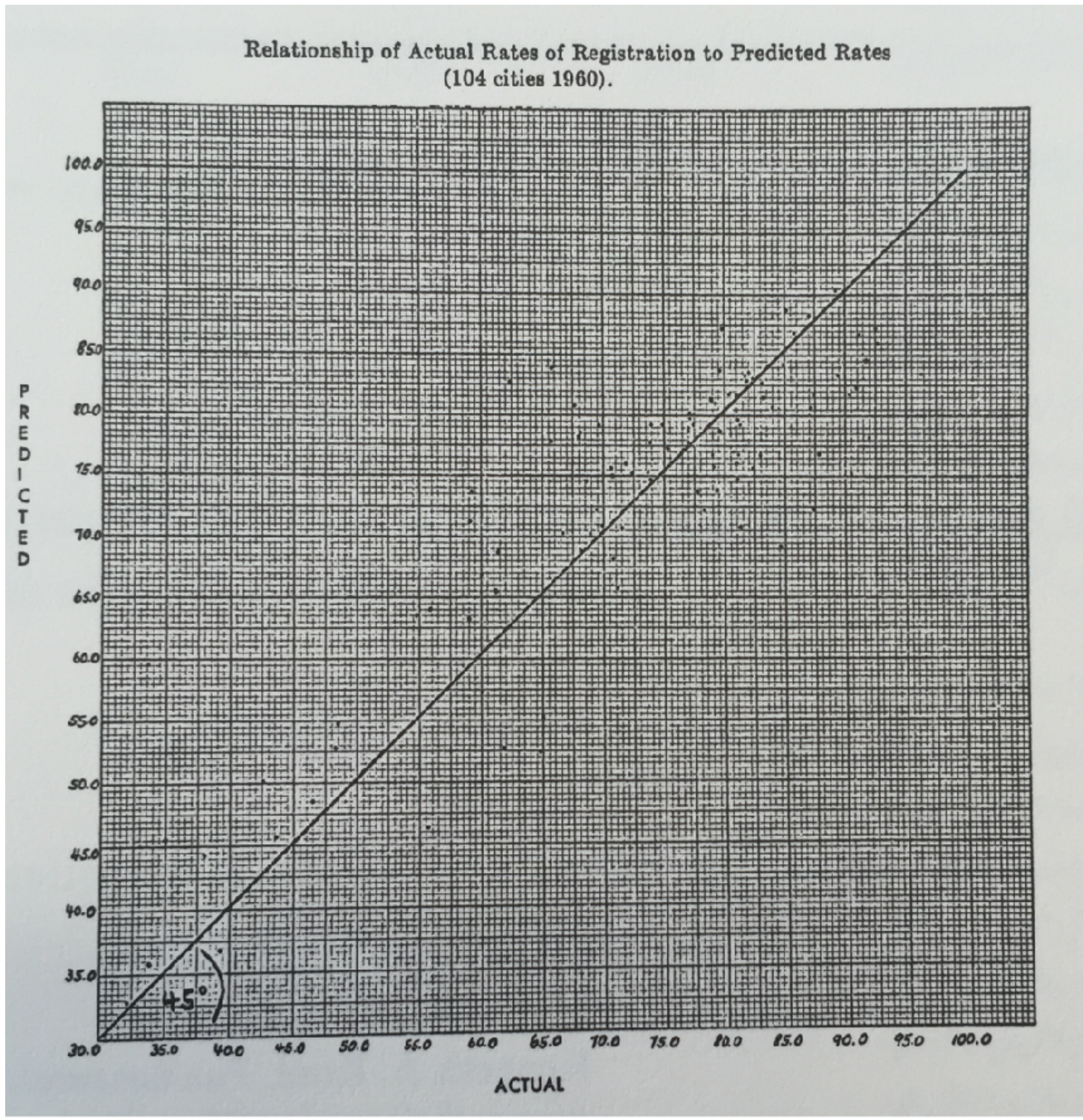
- Data-ink - non-redundant ink encoding data information

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

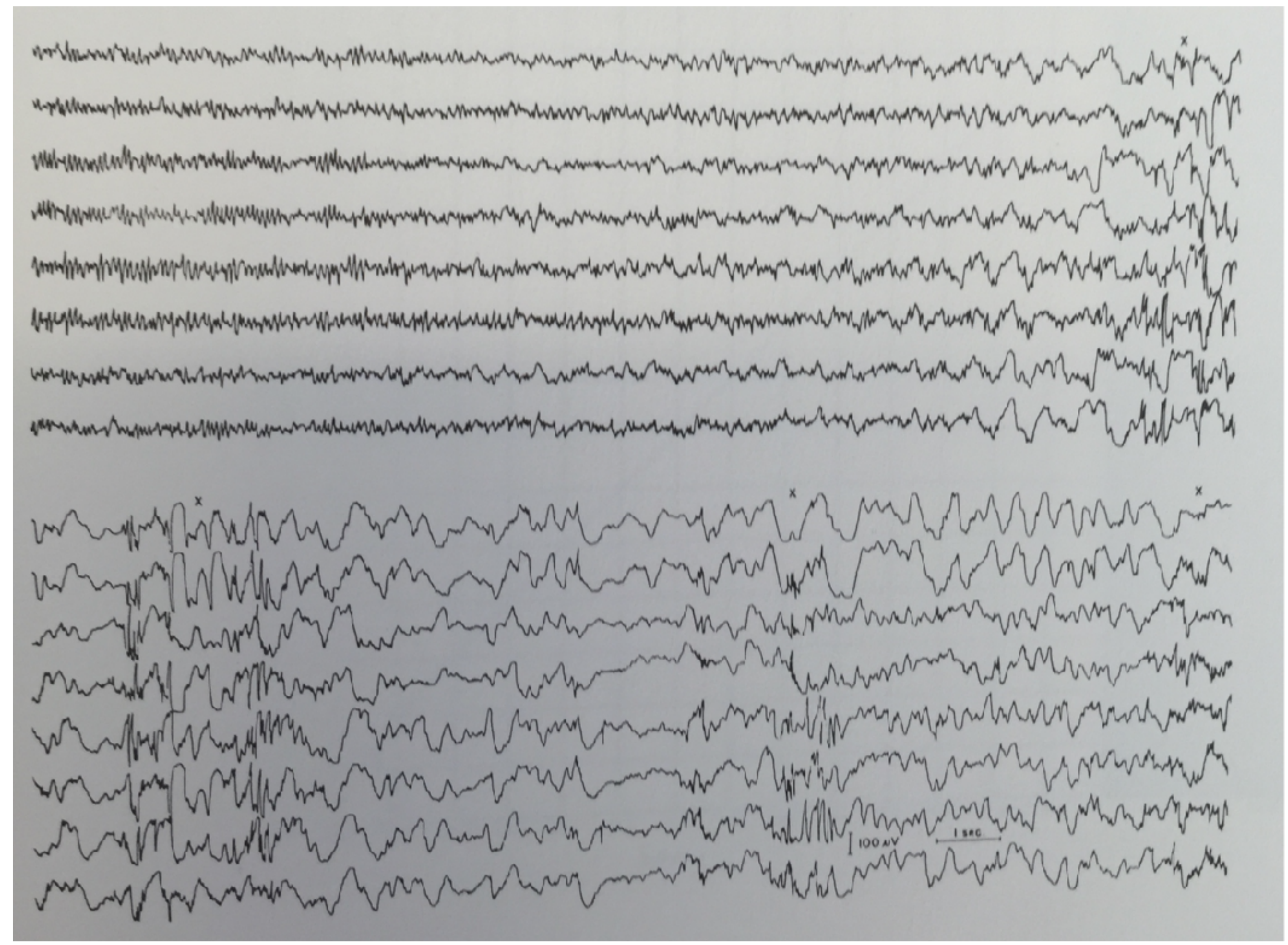
$$= \text{proportion of a graphic's ink devoted to the non-redundant display of data-information}$$

$$= 1.0 - \text{proportion of a graphic that can be erased}$$

Examples of data-ink ratio



~0

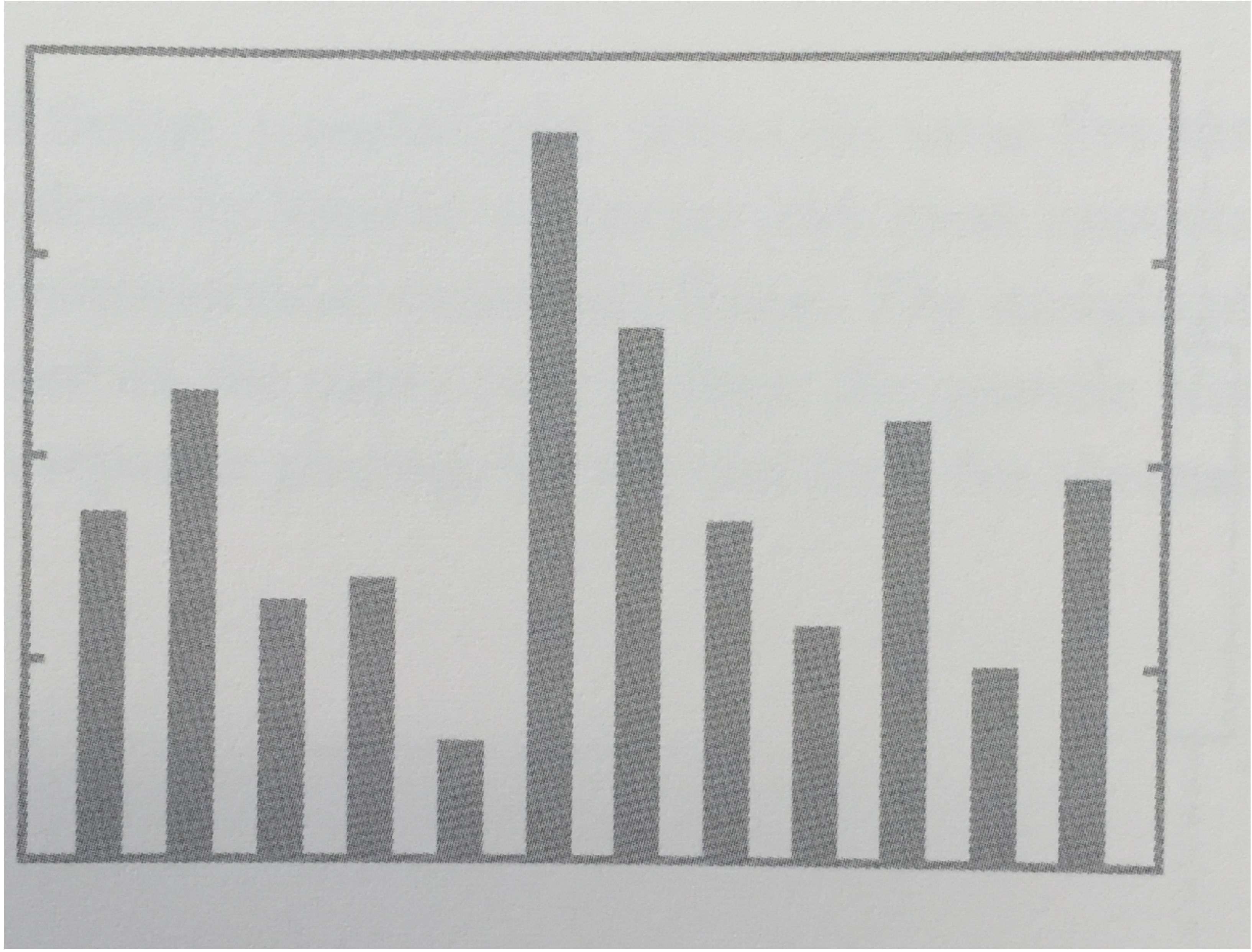


1.0

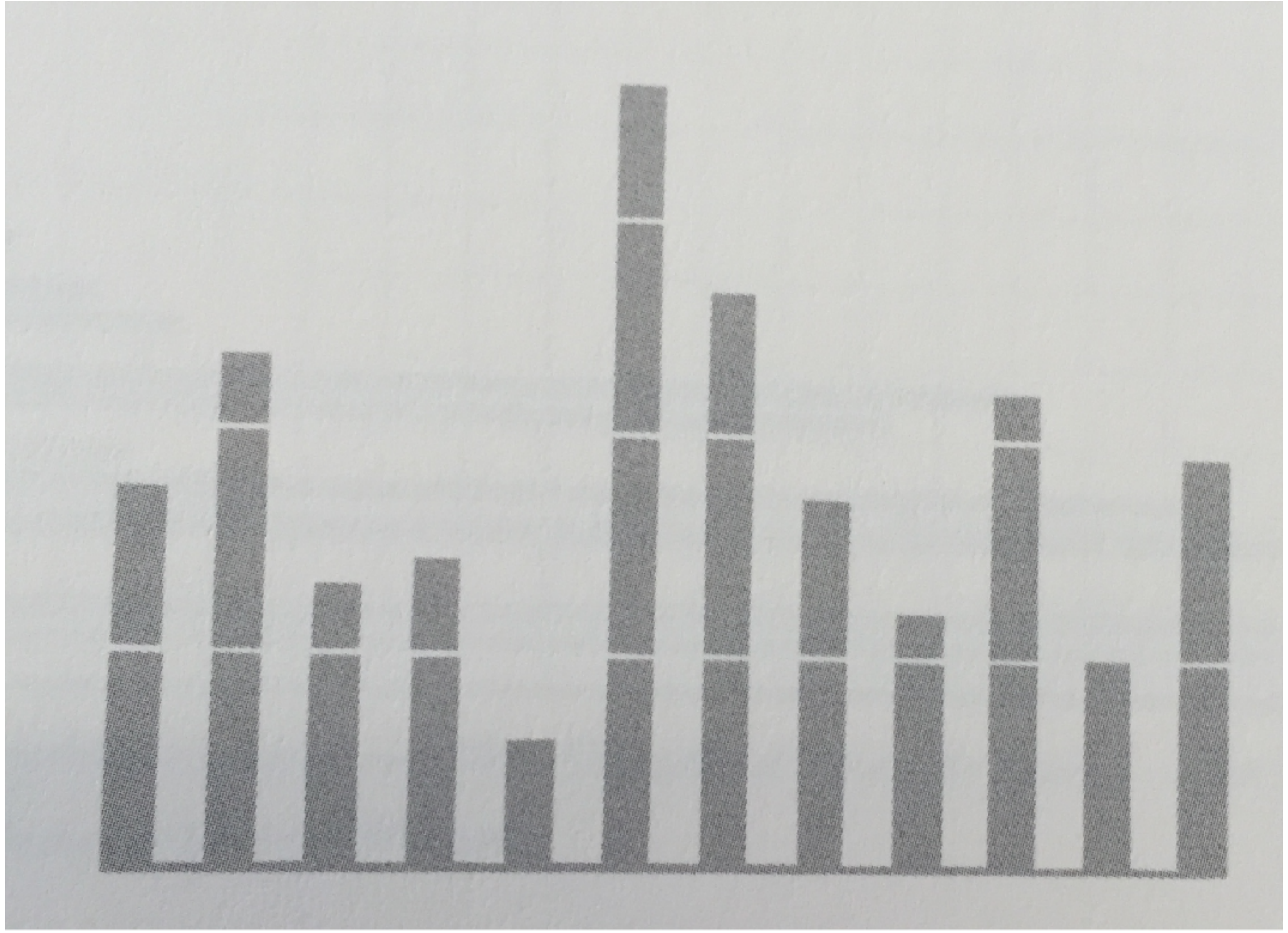
Design principles for data-ink

- (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

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