Information Visualization

SWE 632
Fall 2015

GEORGE MASON UNIVERSITY

Administrivia

- HW 4 due today
- HW 5 due next week
- Midterms returned in-class today

Comments on midterm

- Answer the question given
- Does not need to be 2 pages
- Bootstrap can be customized to be more distinctive (no one lost points here)

Course grade

- In-class and online discussion participation: 5%
- Tech talk: 10%
- HWs and project presentation: 40%
- Mid-term exam: 20%
- Final exam: 25%

Information Visualization

Graphics is the visual means of resolving logical problems.

-Bertin (1977)

Information visualization

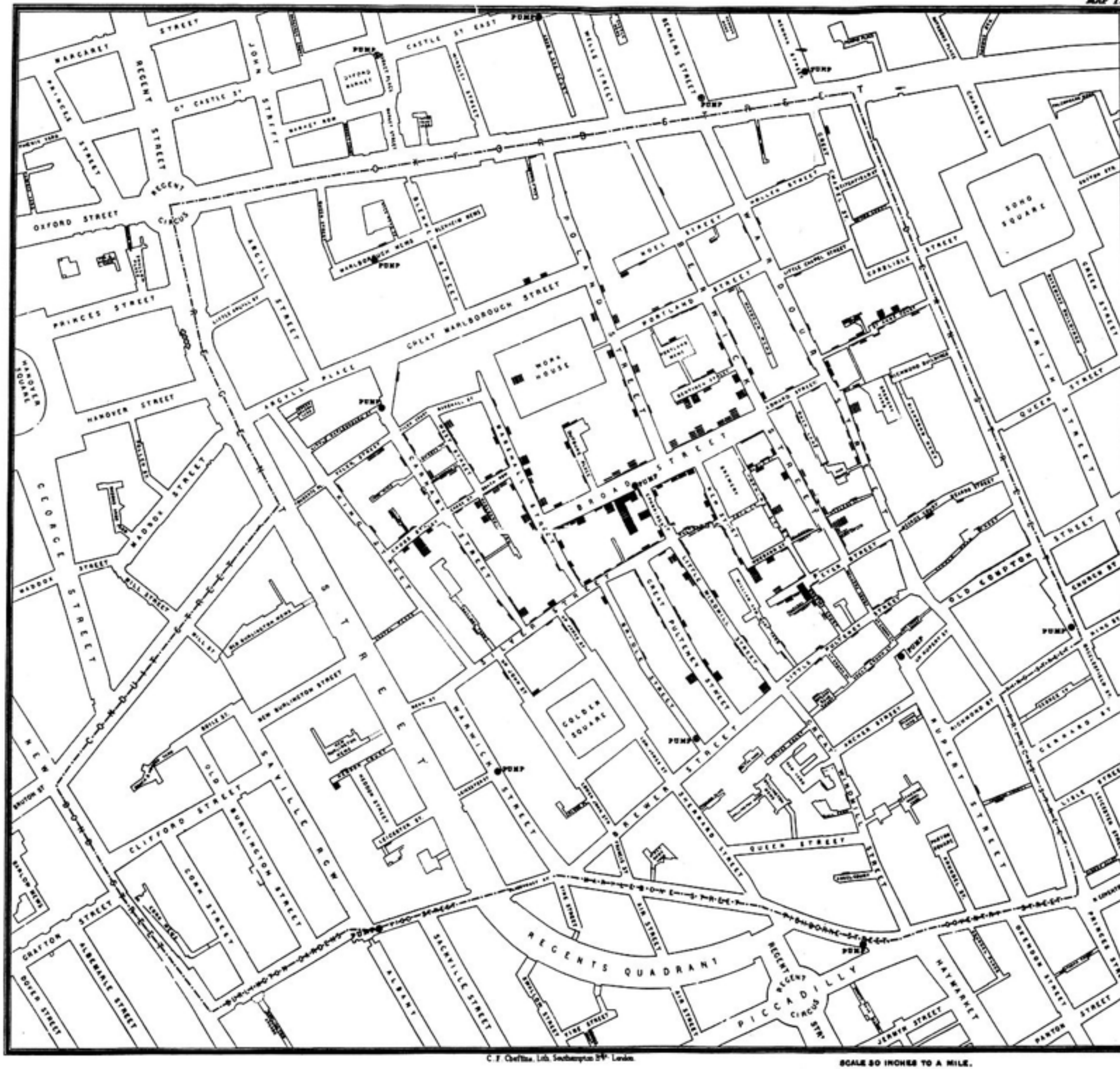
- Technology has made data **pervasive**
  - health, finance, commerce, customer, travel, demographics, communications, …
  - some of it "**big**"
- Information visualization: the use of interactive visual representations to amplify cognition
  - e.g., discover insights, answer questions

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

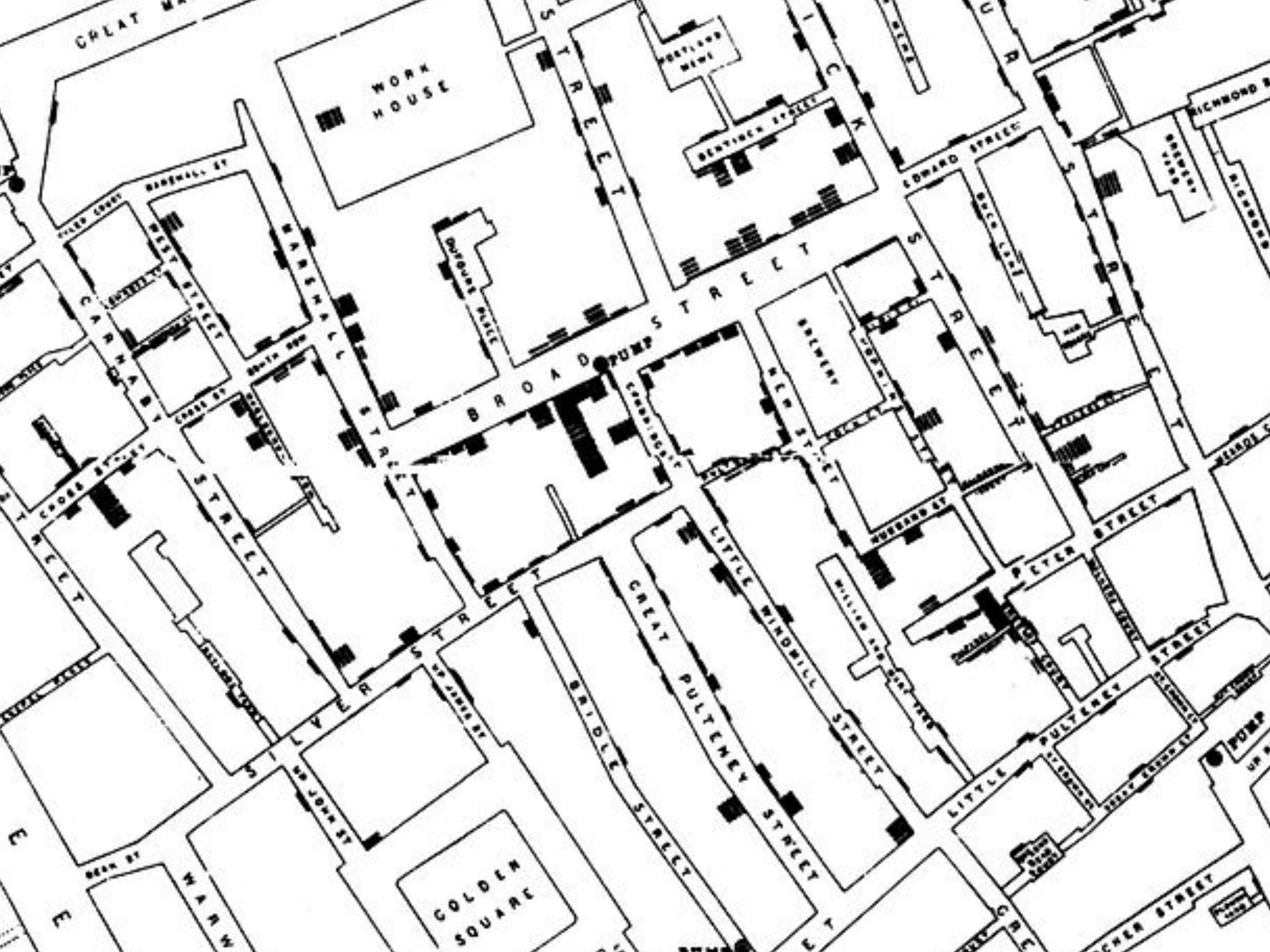
John Snow

- Set out to investigate **cause**
- Suspected it might be due to water from community **pump**
- Tested water —> no obvious impurities
- What more evidence could there be?
    - Could list of 83 deaths, plotted on map



C. F. Cheffins, Lith, Southampton Bldgs. London.

SCALE 30 INCHES TO A MILE.


WORK HOUSE
BROAD STREET
PUMP
GOLDEN SQUARE
SILVER STREET
MARSHALL STREET
CARNABY STREET
WEST STREET
CROSS STREET
PORTLAND MEWS
BENTINCK STREET
EDWARD STREET
DUFOURS PLACE
BREWERY
NEW STREET
CAMBRIDGE STREET
HUBBARD ST
PETER STREET
GREAT PULTENEY STREET
LITTLE WINDMILL STREET
BRIDLE STREET
LITTLE PULTENEY STREET
WILLIAM AND MARY YARD
SOUTH ROW
BERWICK STREET
RICHMOND ST

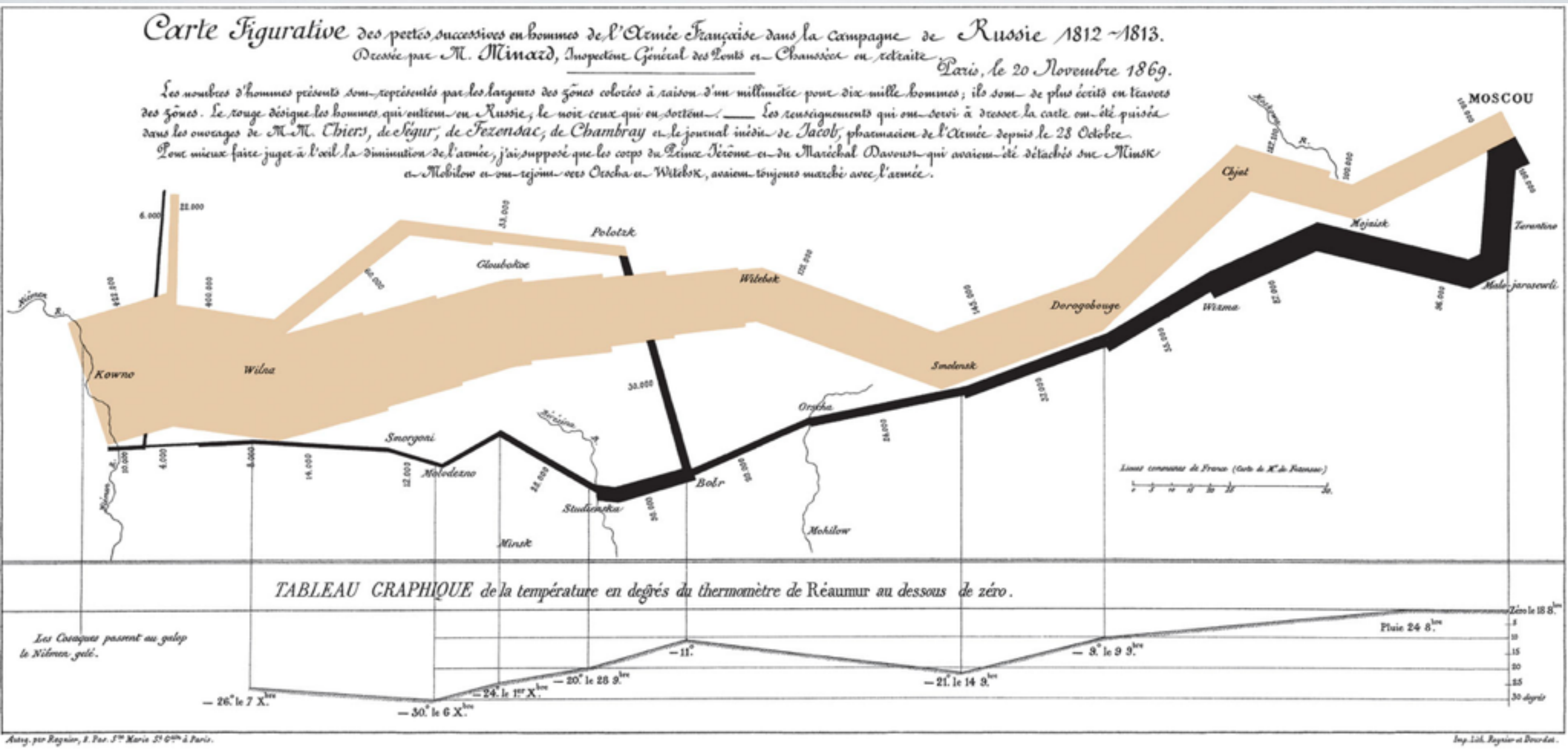
Investigation and aftermath

- 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

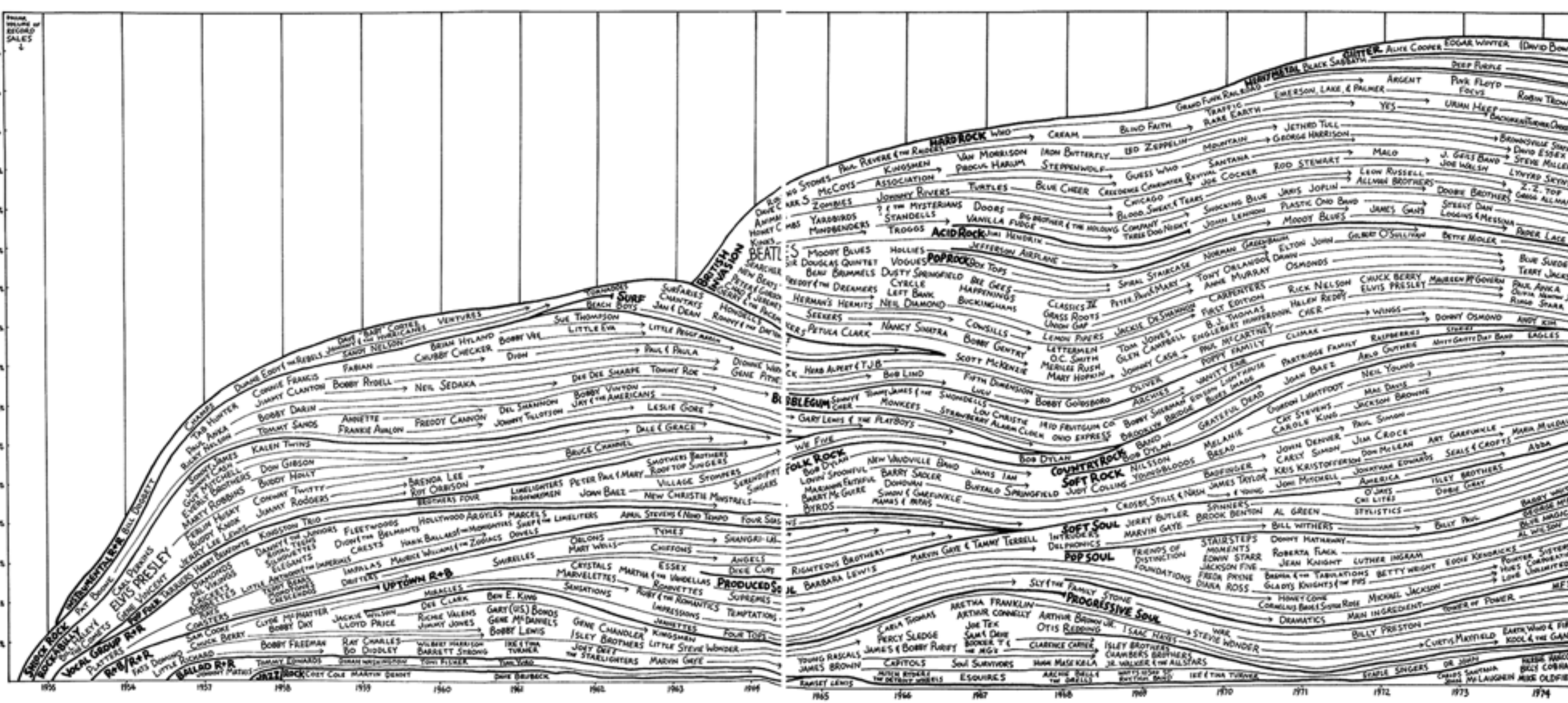
Methods used by Snow

- 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



Chapel & Garofalo, Rock 'N Roll is Here to Pay: The History and Politics of the Music Industry

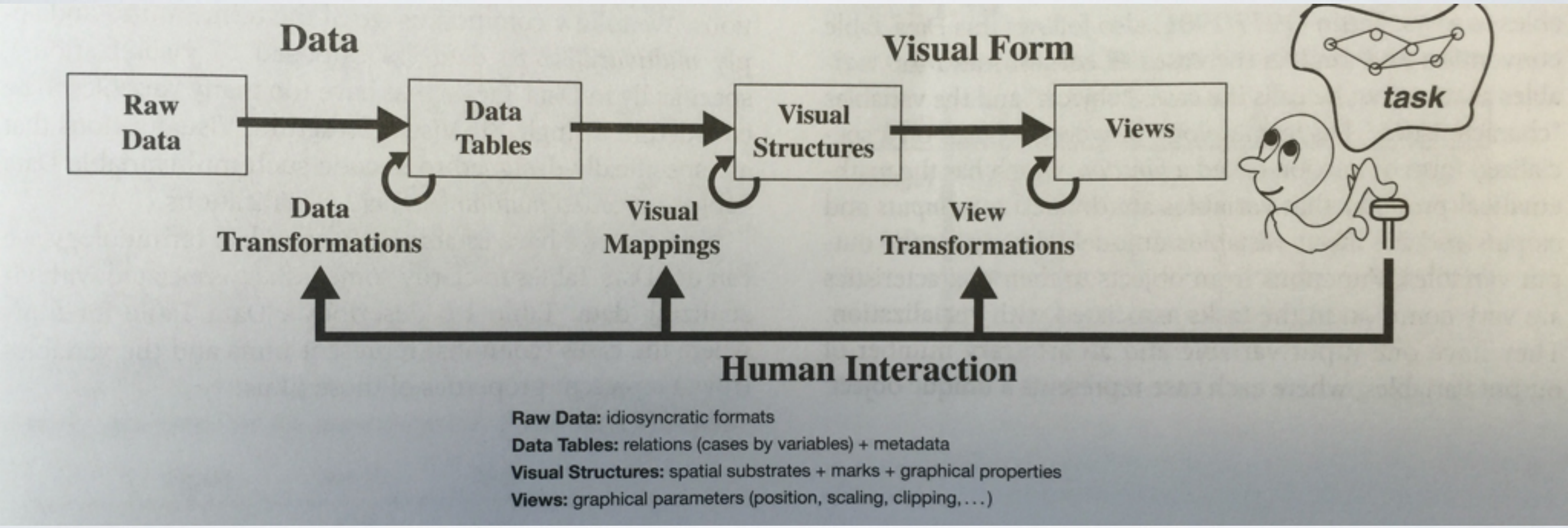


How information visualization amplifies cognition.

Increased Resources	
High-bandwidth hierarchical interaction	The human moving gaze system partitions limited channel capacity so that it combines high spatial resolution and wide aperture in sensing visual environments (Resnikoff, 1987).
Parallel perceptual processing	Some attributes of visualizations can be processed in parallel compared to text, which is aerial.
Offload work from cognitive to perceptual system	Some cognitive inferences done symbolically can be recoded into inferences done with simple perceptual operations (Larkin and Simon, 1987).
Expanded working memory	Visualizations can expand the working memory available for solving a problem (Norman, 1993).
Expanded storage of information	Visualizations can be used to store massive amounts of information in a quickly accessible form (e.g., maps).
Reduced Search	
Locality of processing	Visualizations group information used together, reducing search (Larkin and Simon, 1987).
High data density	Visualizations can often represent a large amount of data in a small space (Tufte, 1983).
Spatially indexed addressing	By grouping data about an object, visualizations can avoid symbolic labels (Larkin and Simon, 1987).
Enhanced Recognition of Patterns	
Recognition instead of recall	Recognizing information generated by a visualization is easier than recalling that information by the user.
Abstraction and aggregation	Visualizations simplify and organize information, supplying higher centers with aggregated forms of information through abstraction and selective omission (Card, Robertson, and Mackinlay, 1991; Resnikoff, 1987).
Visual schemata for organization	Visually organizing data by structural relationships (e.g., by time) enhances patterns.
Value, relationship, trend	Visualizations can be constructed to enhance patterns at all three levels (Bertin, 1977/1981).
Perceptual Inference	
Visual representations make some problems obvious	Visualizations can support a large number of perceptual inferences that are extremely easy for humans (Larkin and Simon, 1987).
Graphical computations	Visualizations can enable complex specialized graphical computations (Hutchins, 1996).
Perceptual Monitoring	Visualizations can allow for the monitoring of a large number of potential events if the display is organized so that these stand out by appearance or motion.
Manipulable Medium	Unlike static diagrams, visualizations can allow exploration of a space of parameter values and can amplify user operations.

Mapping data to visual form

Designing an information visualization



Types of variables

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

Data transformations

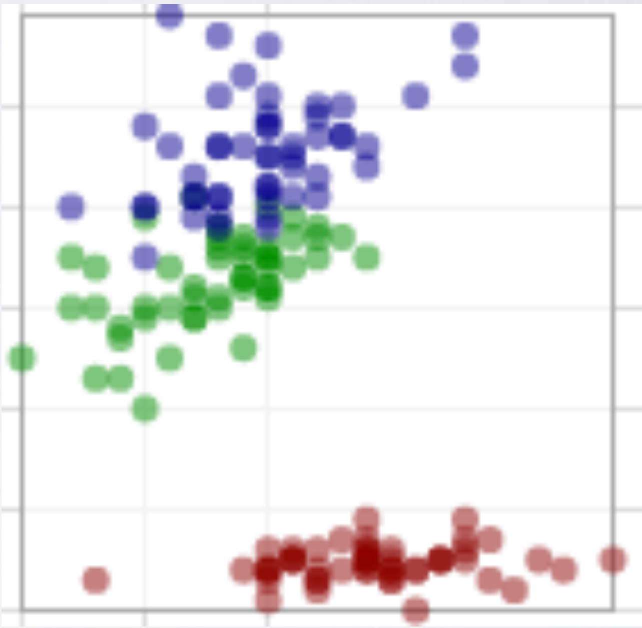
- Classing / binning: Quantitative —> ordinal
  - Maps ranges onto **classes** of variables
  - Can also count # of items in each class w/ histogram
- Sorting: Nominal —> ordinal
  - Add order between items in sets
- Descriptive statistics: mean, average, median, max, min, …

Visual structures

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

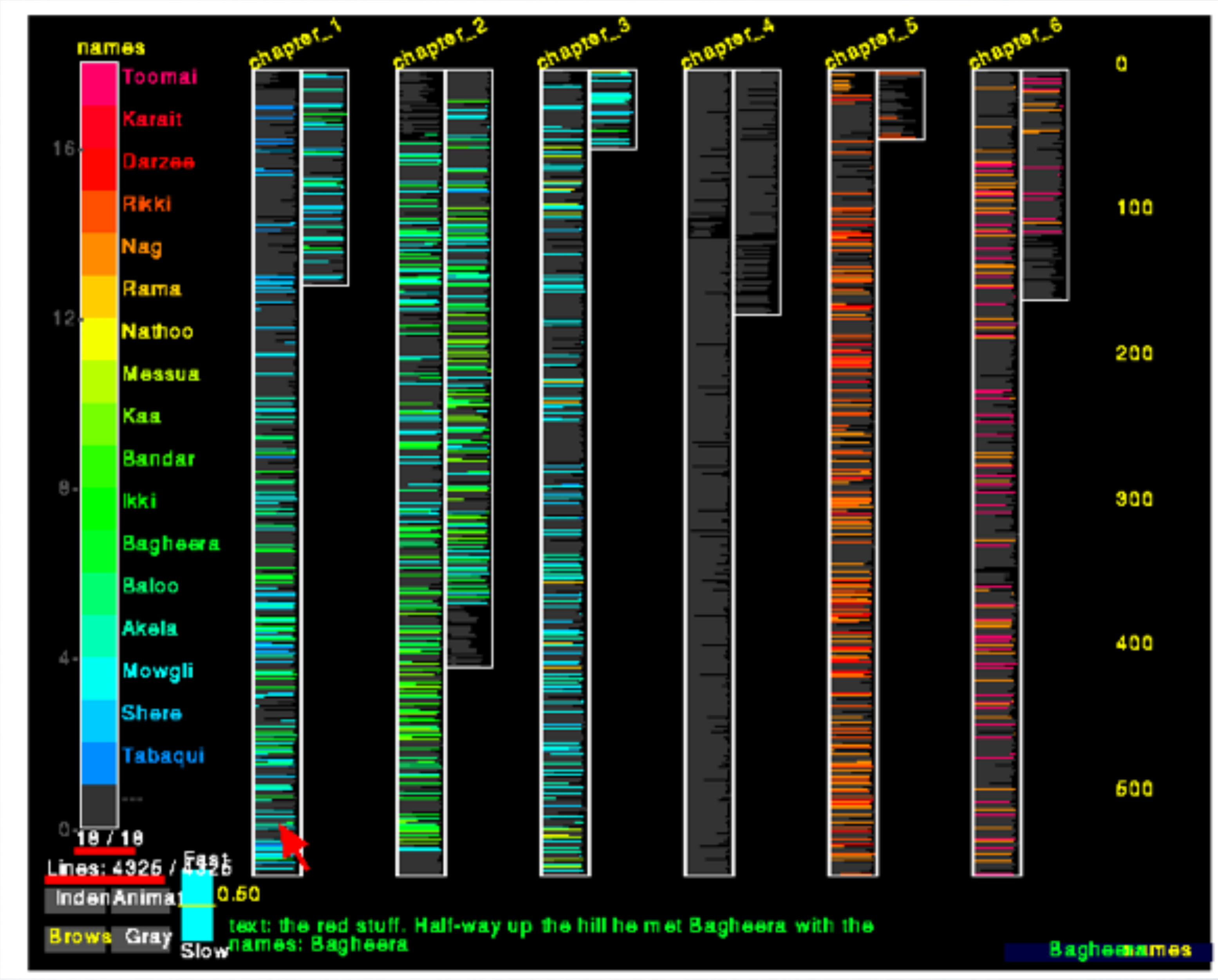
Spatial substrate

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



Folding

- continuing an axis by continuing in different space

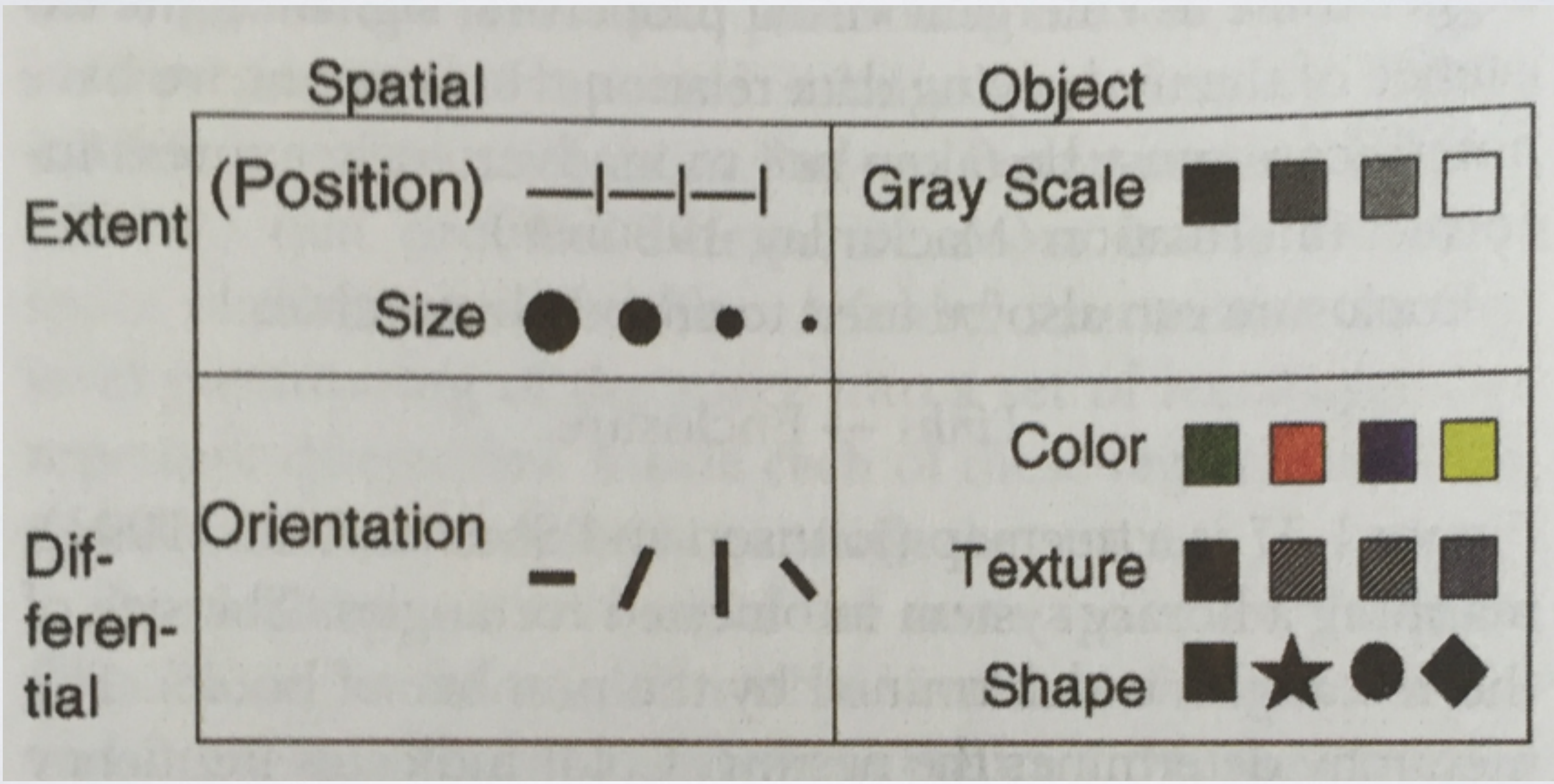


Marks

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

Marks' graphical properties

a.k.a. Bertin's retinal properties



Effectiveness of graphical properties

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

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

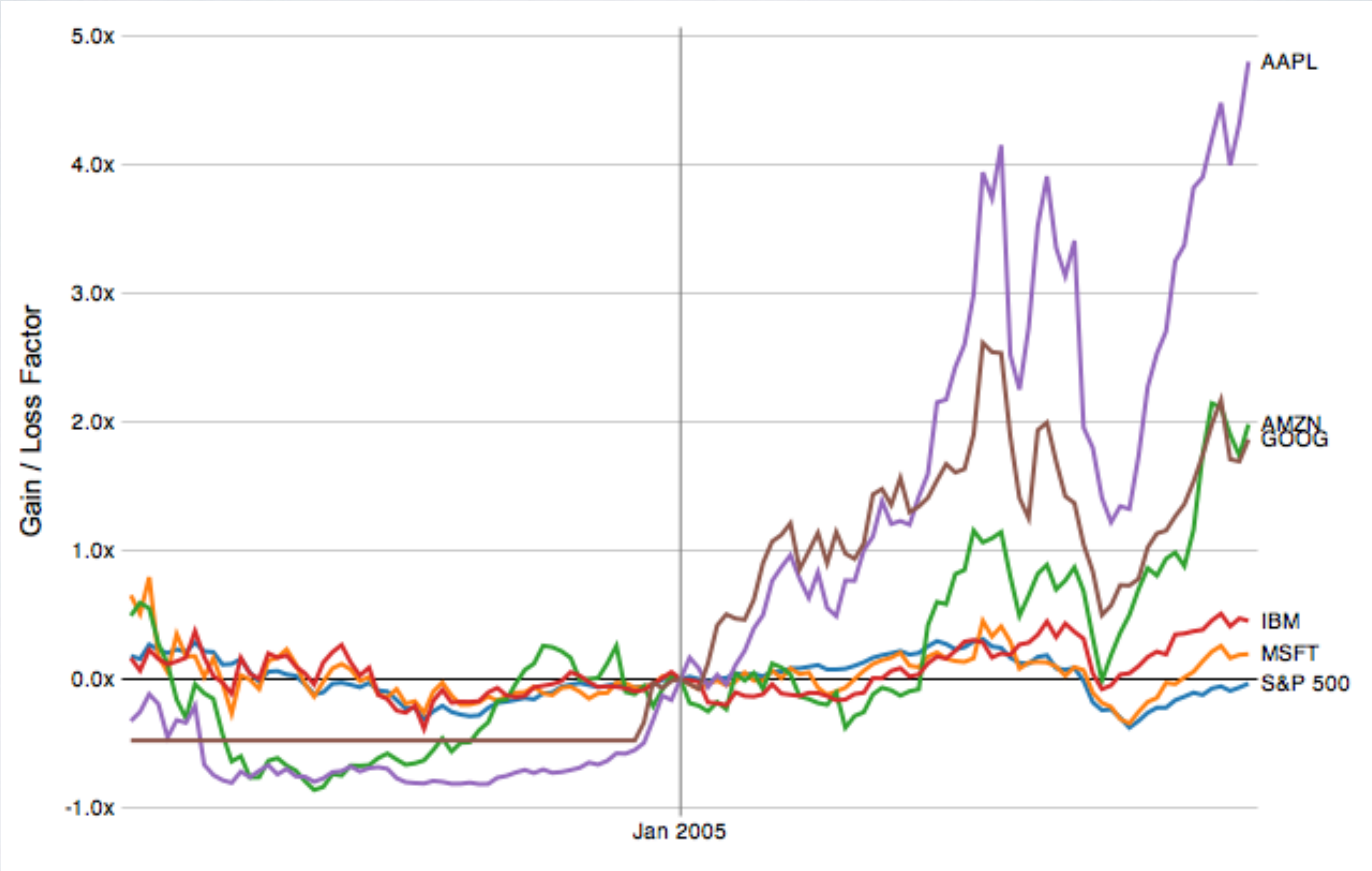
Animation

- 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

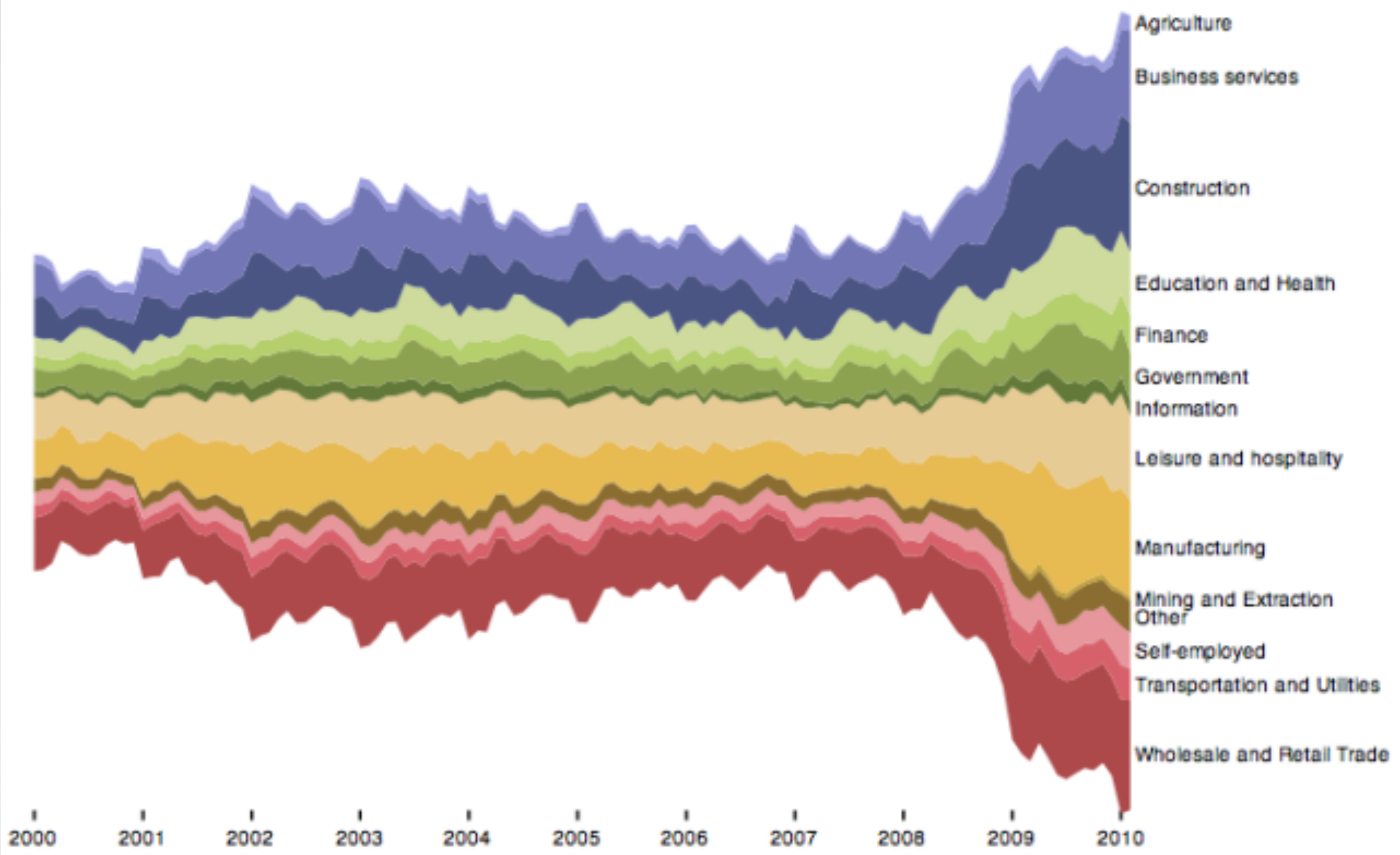
Time-series data

Index chart



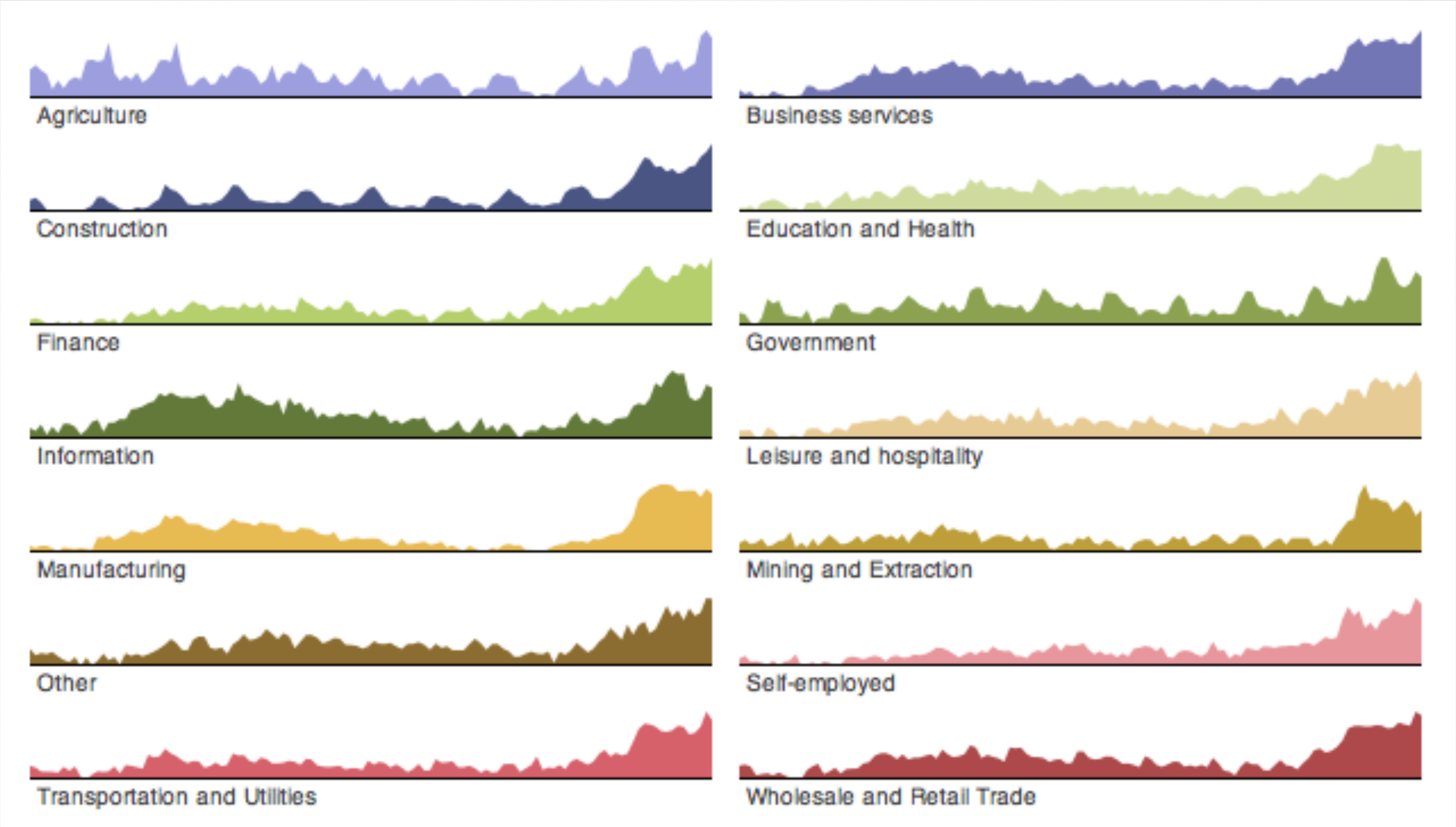
- Depicts % change relative to baseline point

Stacked graph



- Supports visual summation of multiple components

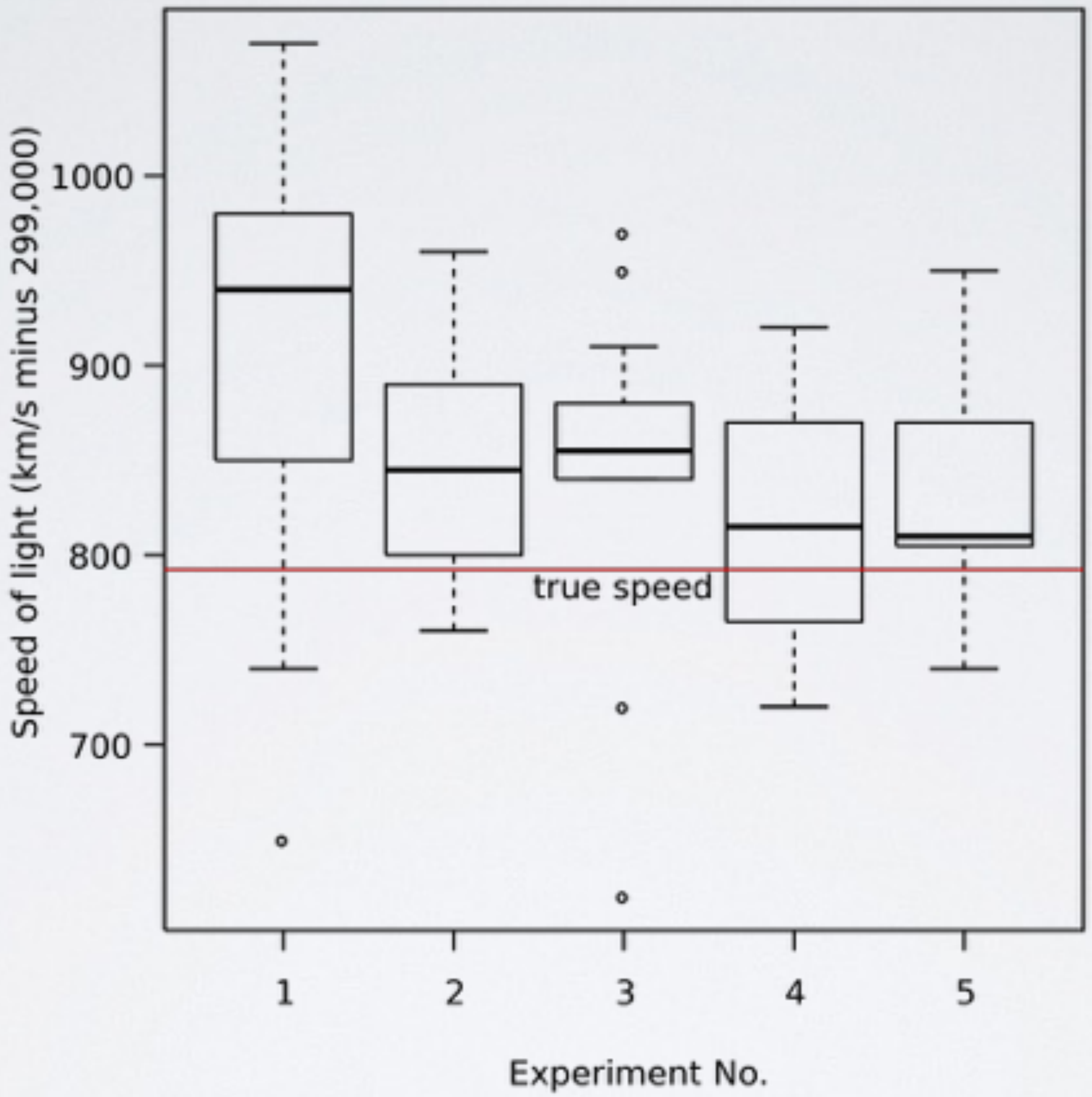
Small multiples



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

Statistical distributions

Box plot



- shows distribution with median, quantiles, min / max
- outliers: 1.5 x interquartile range (height of box)

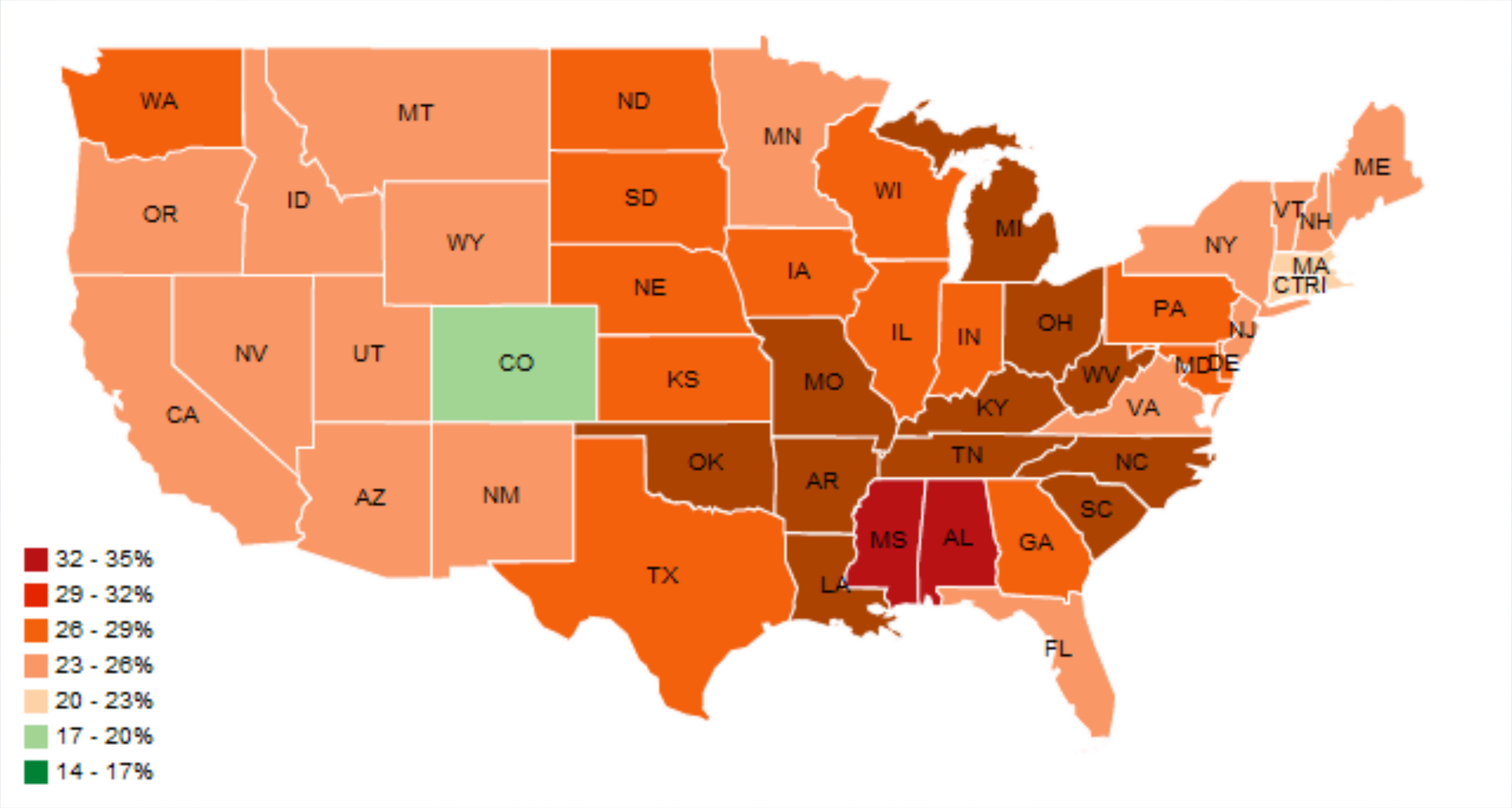
Stem and left plots

0	1 1 1 2 2 2 2 3 3 3 3 3 3 4 4 4 4 4 4 4 4 4 5 6 7 8 8 8 8 8 8 9
1	0 0 0 0 1 1 1 1 2 2 3 3 3 3 4 4 4 4 5 5 6 7 7 8 9 9 9 9 9
2	0 0 1 1 1 5 7 8 9
3	0 0 1 2 3 3 3 4 6 6 8 8
4	0 0 1 1 1 1 3 3 4 5 5 5 6 7 8 9
5	0 2 3 5 6 7 7 7 9
6	1 2 6 7 8 9 9 9
7	0 0 0 1 6 7 9
8	0 0 1 2 3 4 4 4 4 4 4 4 5 6 7 7 7 9
9	1 3 3 5 7 8 8 8 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9
10	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

- bins numbers by first digit, stacks remaining digits
- more detail focused alternative to histogram

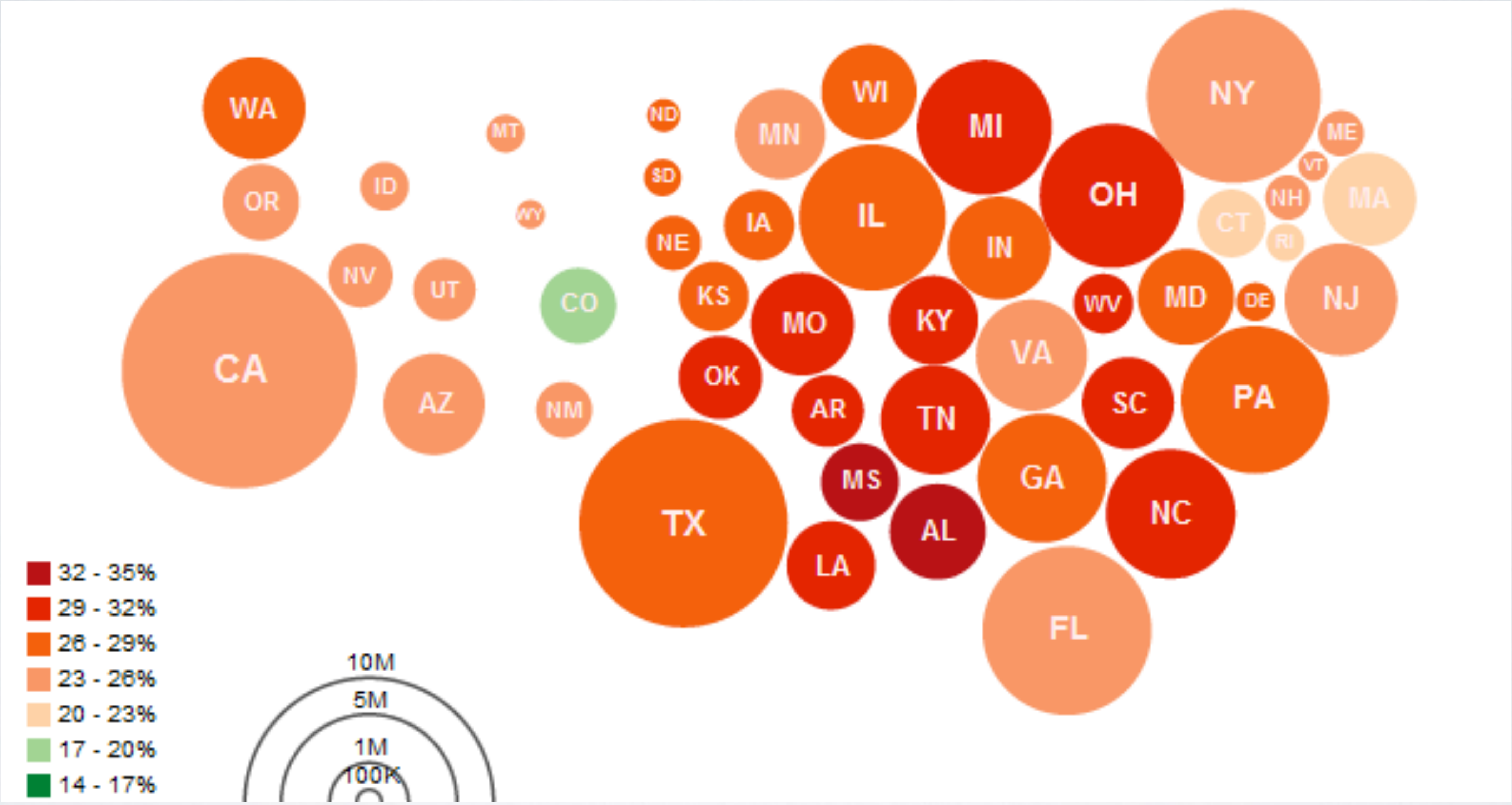
Maps

Choropleth map



- Groups data by area, maps to color

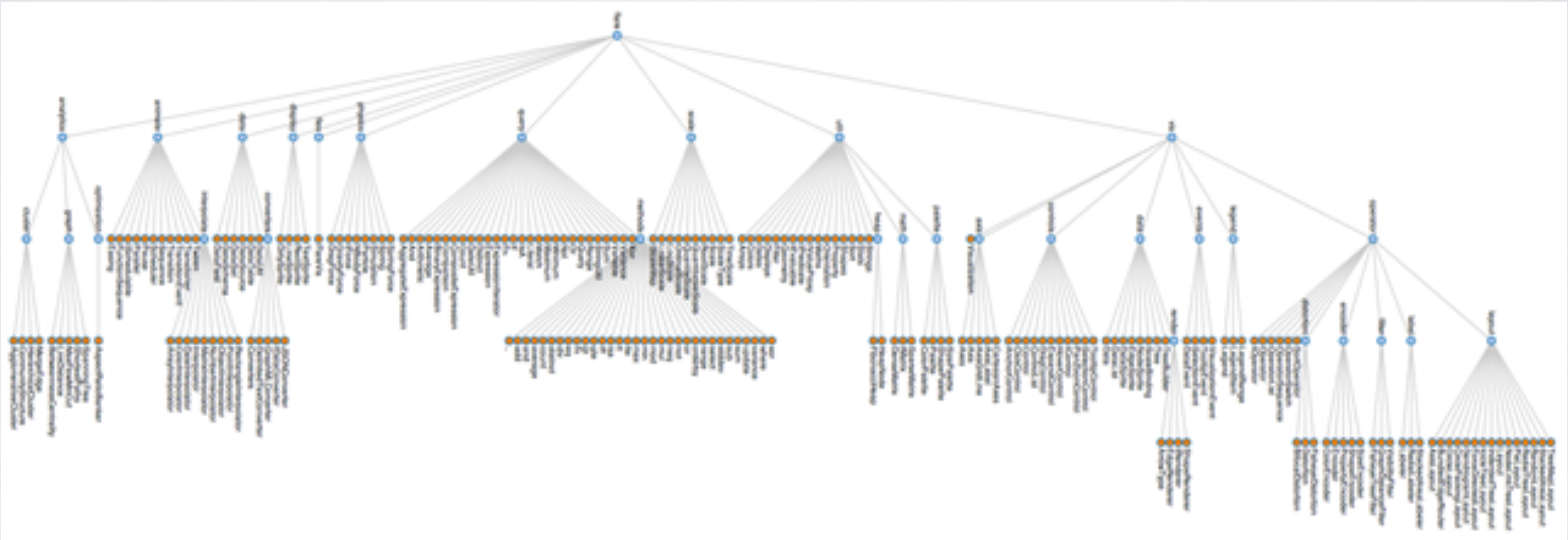
Cartograms



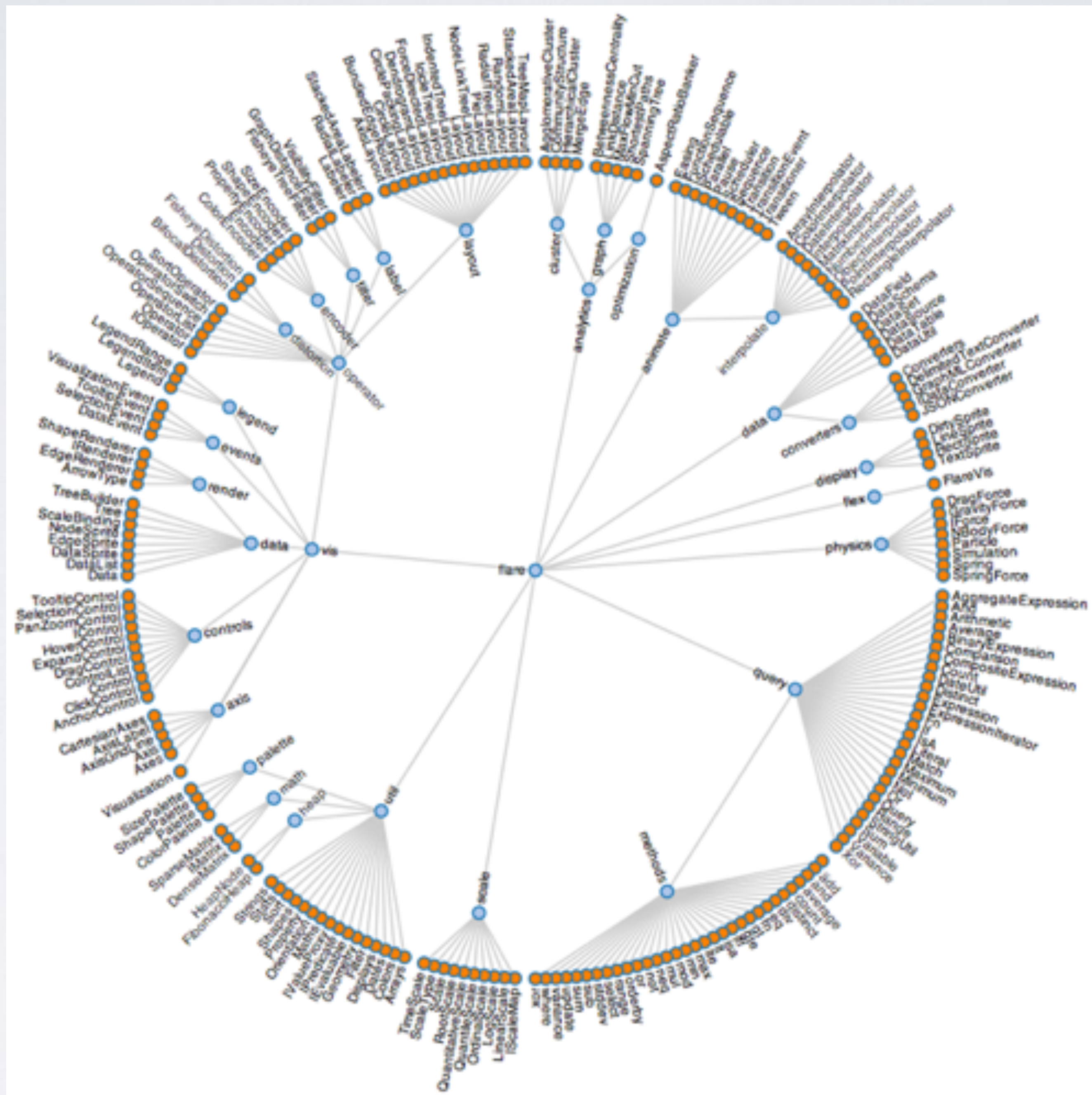
- Encodes two variables w/ size & color

Hierarchies

Node link diagram

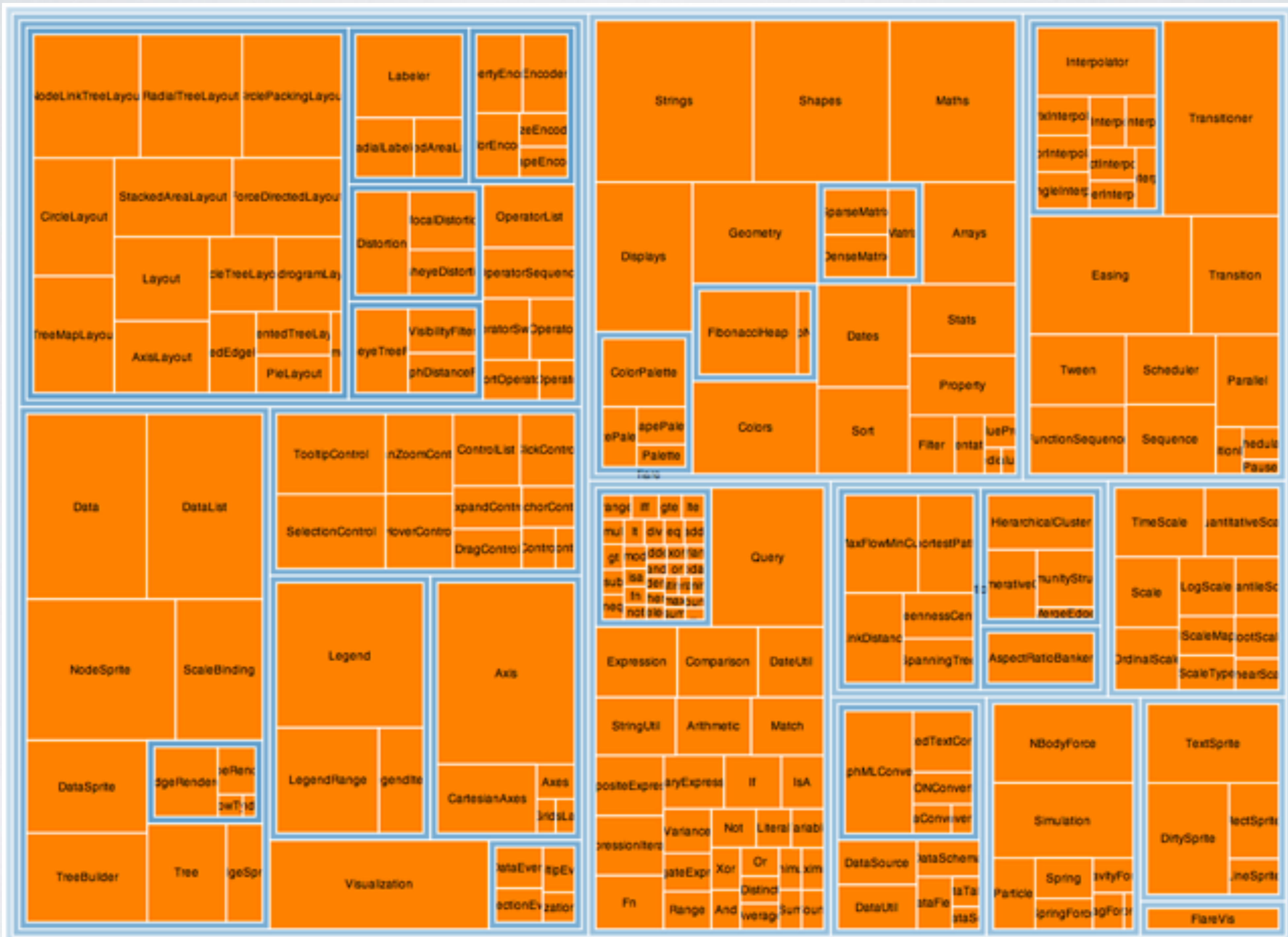


Dendrogram



- leaf nodes of hierarchy on edges of circle

Treemaps



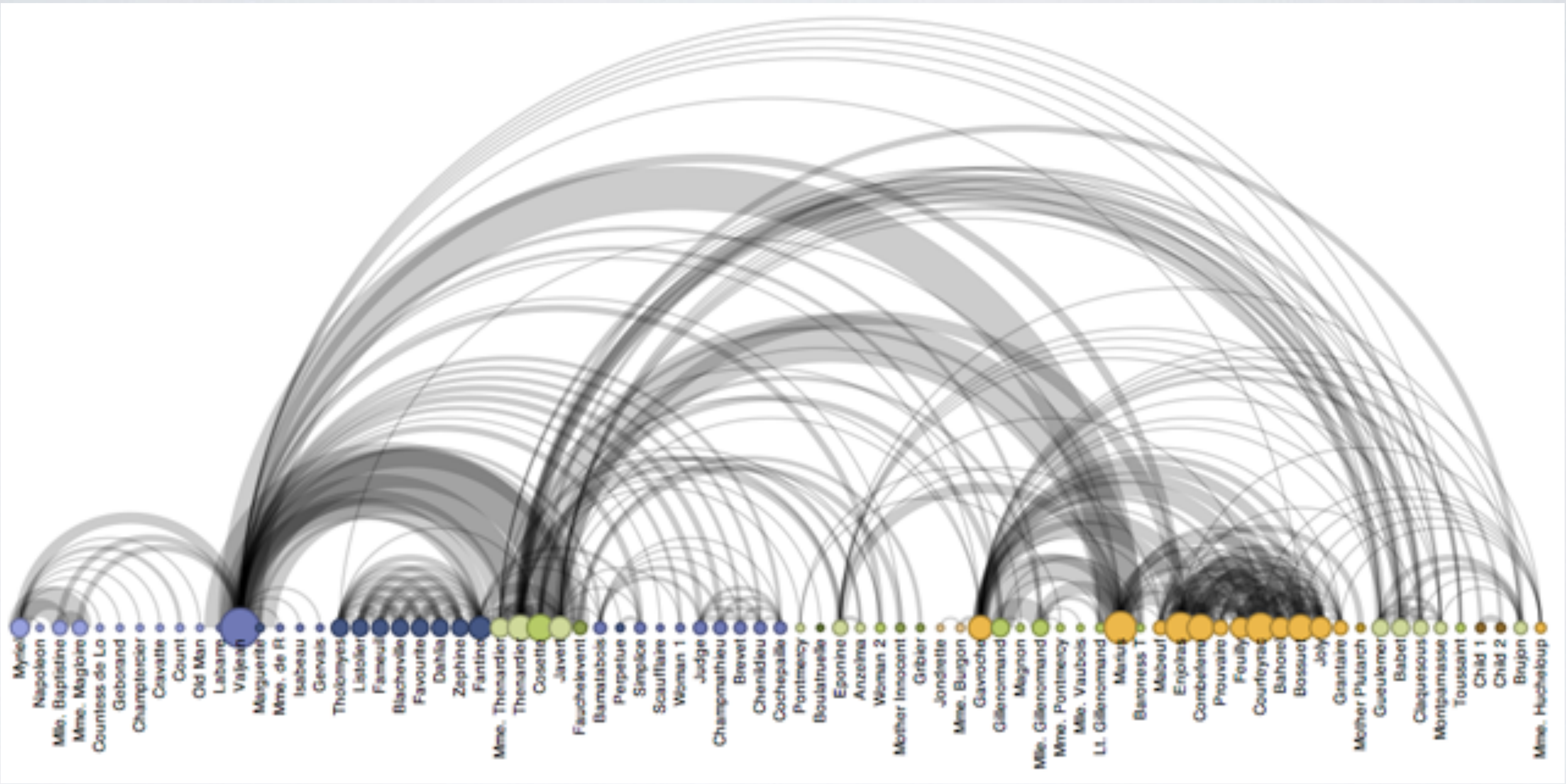
RadialTreeLayout
Labeler
CircleLayout
StackedAreaLayout
OperatorList
Layout
Distortion
AxisLayout
PieLayout
VisibilityFilter
Strings
Shapes
Maths
Interpolator
Transitioner
Displays
Geometry
Arrays
Easing
Transition
FibonacciHeap
Dates
Stats
ColorPalette
Property
Tween
Scheduler
Parallel
Colors
Sort
Filter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

Networks

Force-directed layout

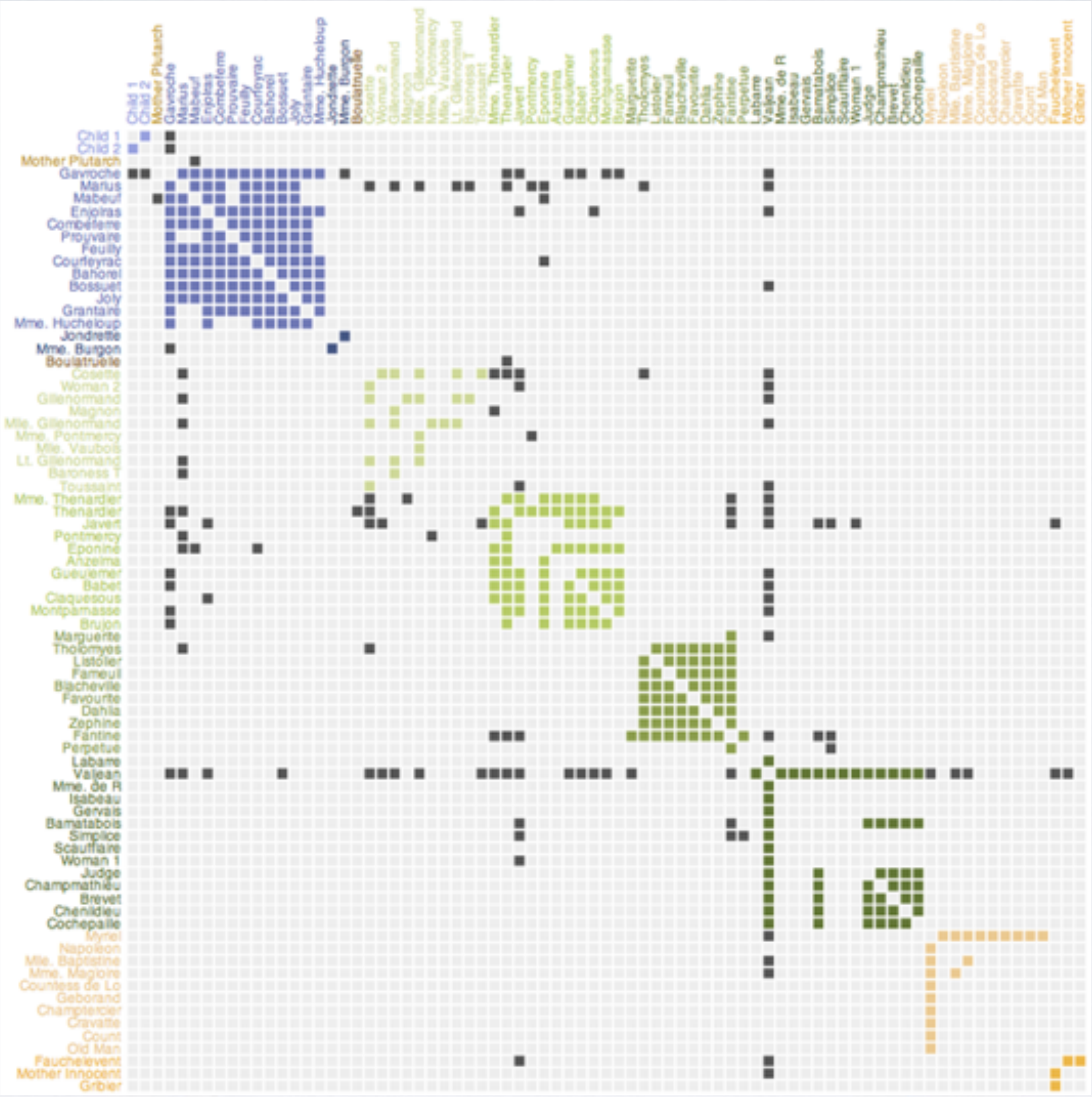
- edges function as springs, find least energy configuration

Arc diagram



- can support identifying cliques & bridges w/ right order

Adjacency matrix



Design considerations

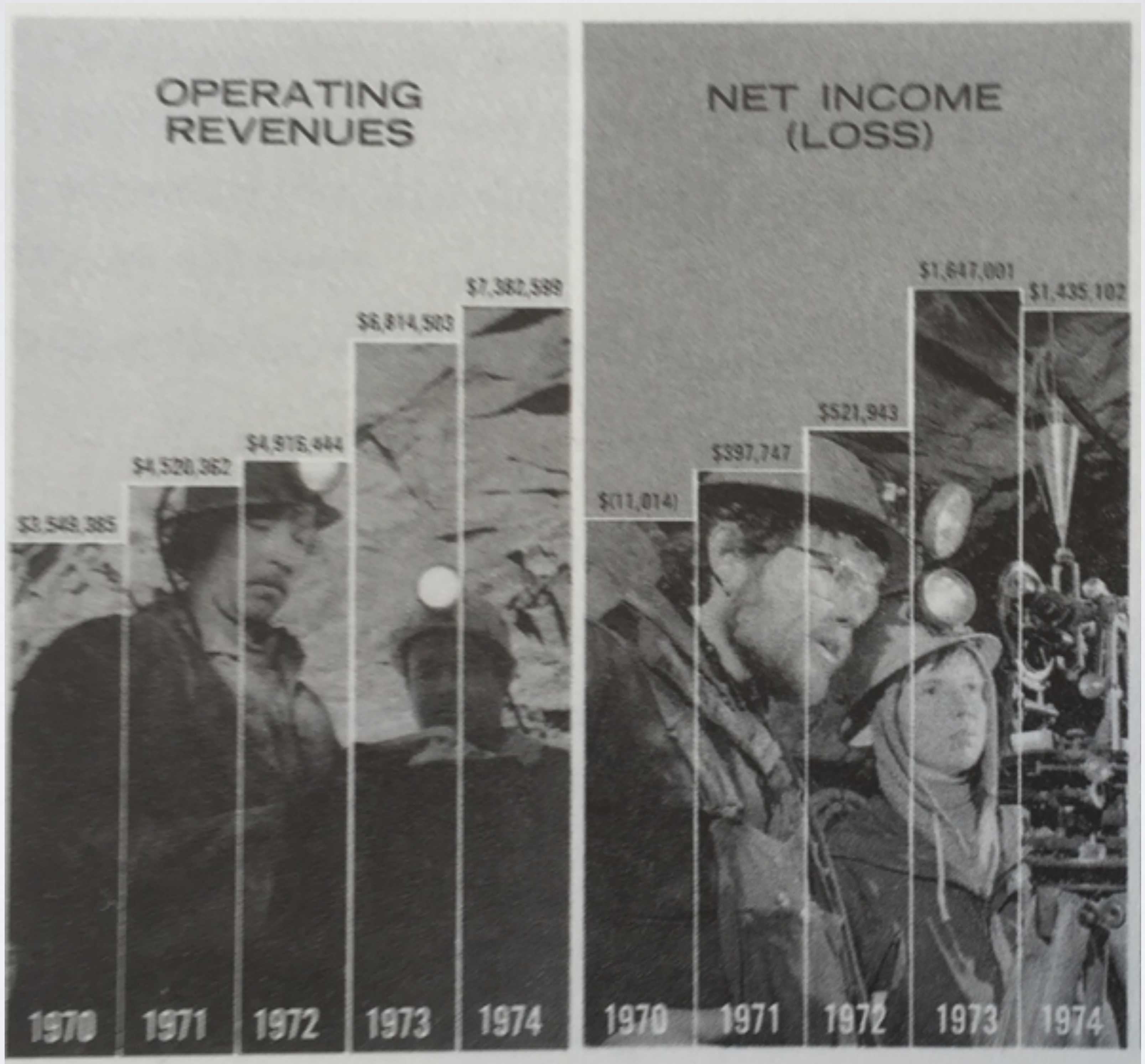
Tufte's principles of graphical excellence

- 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

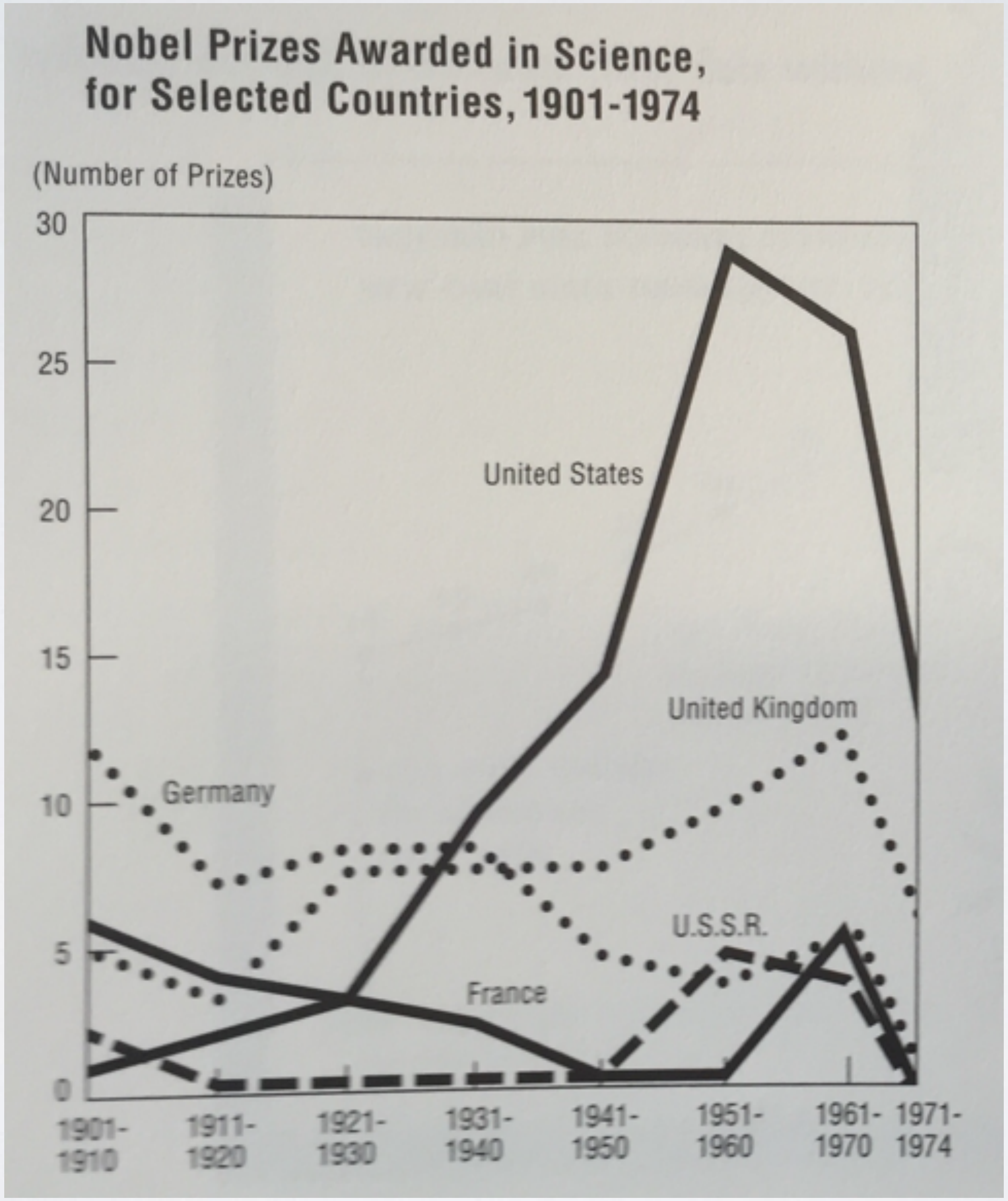
Distortions in visualizations

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

Example

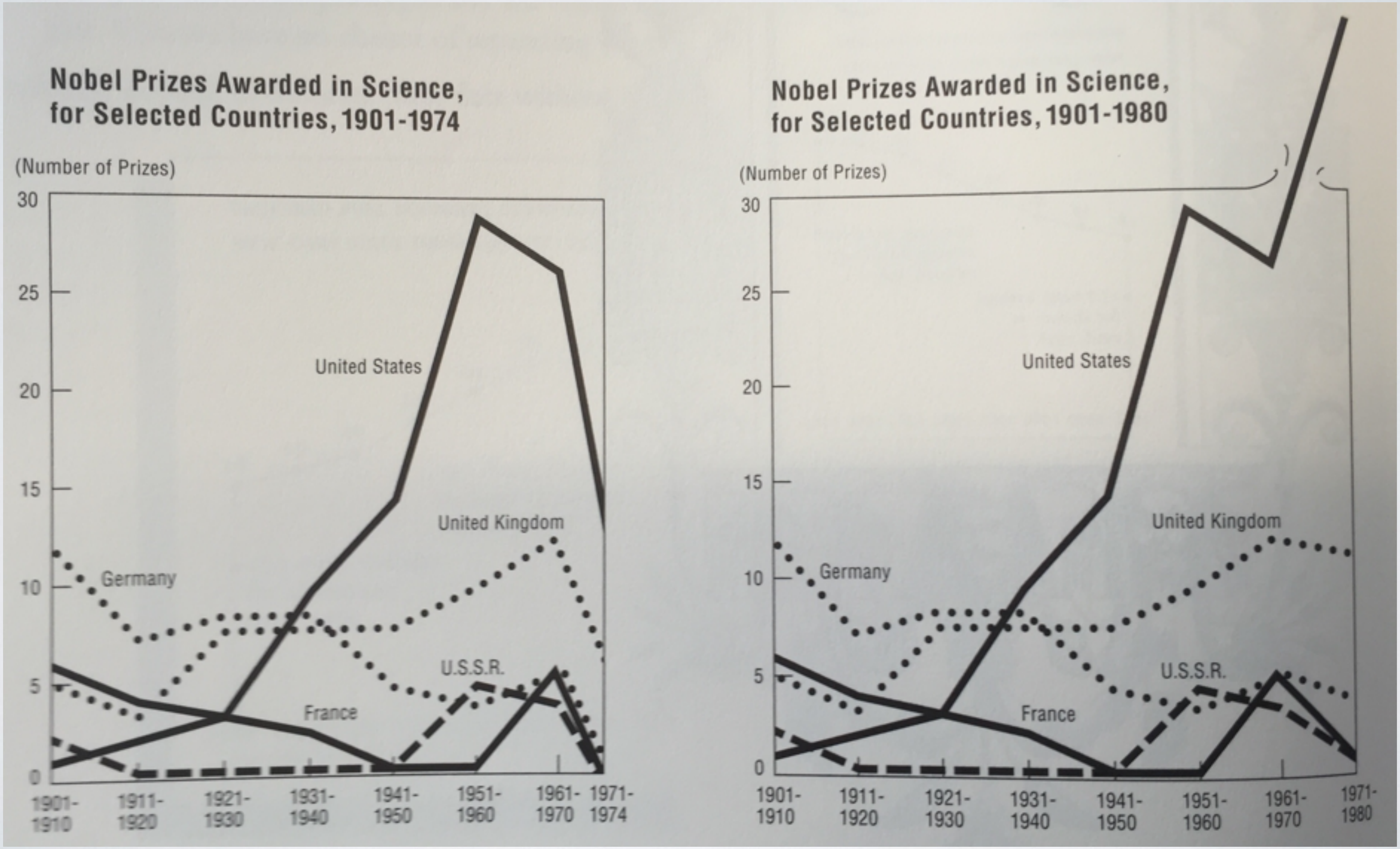

OPERATING REVENUES
$3,549,385
$4,520,362
$4,976,444
$6,814,503
$7,382,599
1970
1971
1972
1973
1974
NET INCOME (LOSS)
$(11,014)
$397,747
$521,943
$1,647,001
$1,435,102
1970
1971
1972
1973
1974

Example

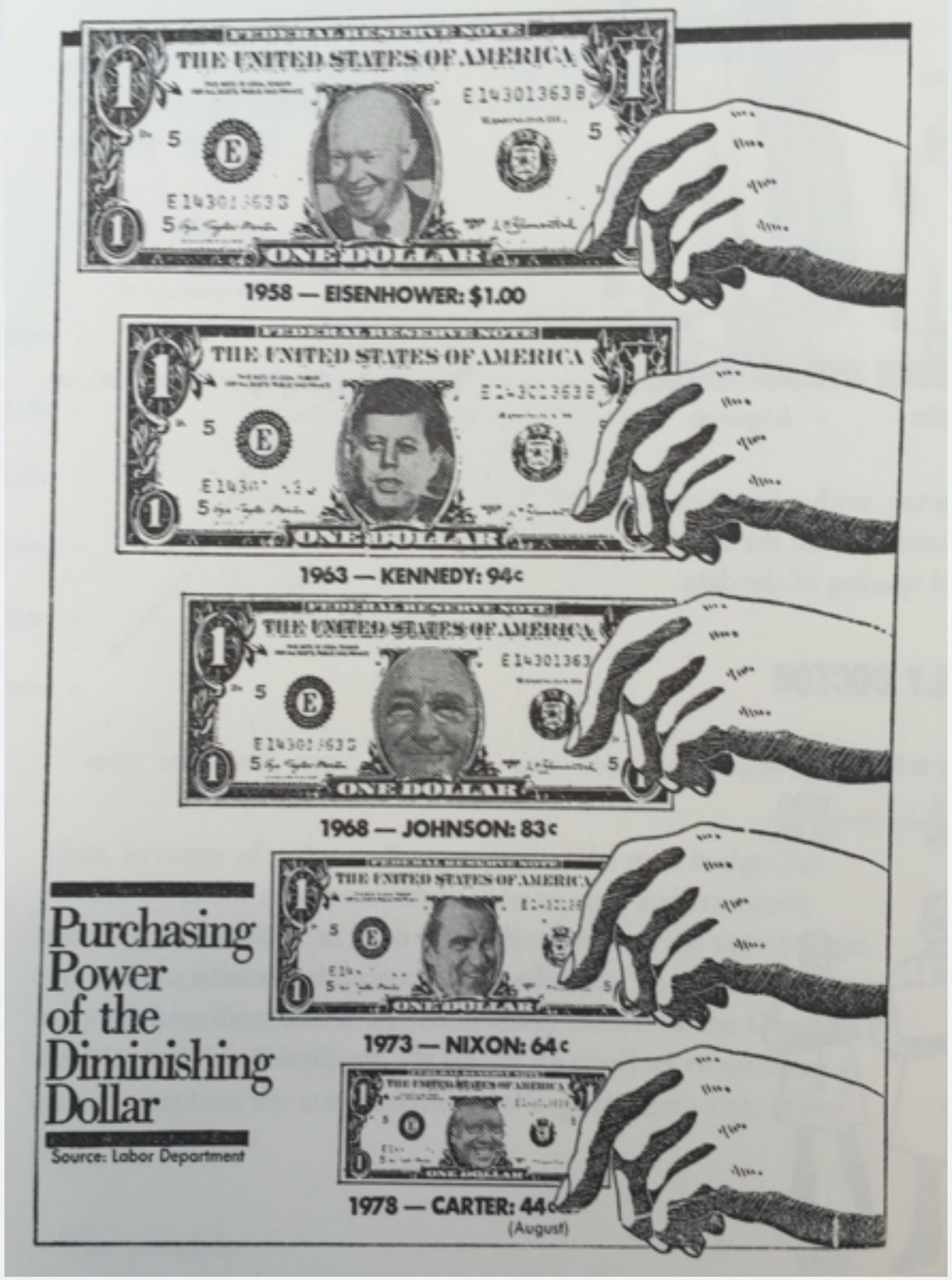


Nobel Prizes Awarded in Science, for Selected Countries, 1901-1974
(Number of Prizes)
30
25
20
15
10
5
0
United States
United Kingdom
Germany
U.S.S.R.
France
1901-1910
1911-1920
1921-1930
1931-1940
1941-1950
1951-1960
1961-1970
1971-1974

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

Data-ink

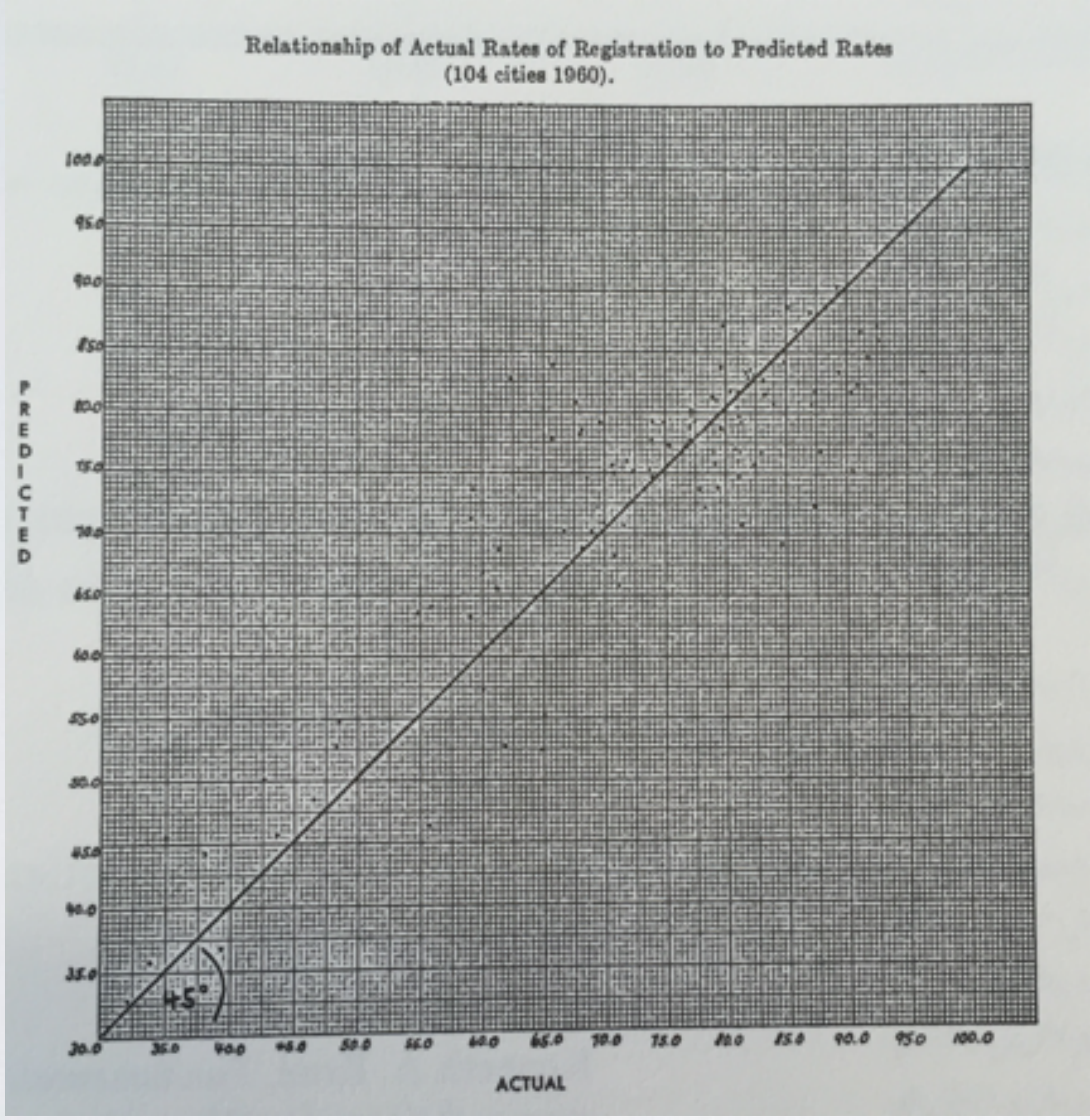
- Data-ink - non-redundant ink encoding data information

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

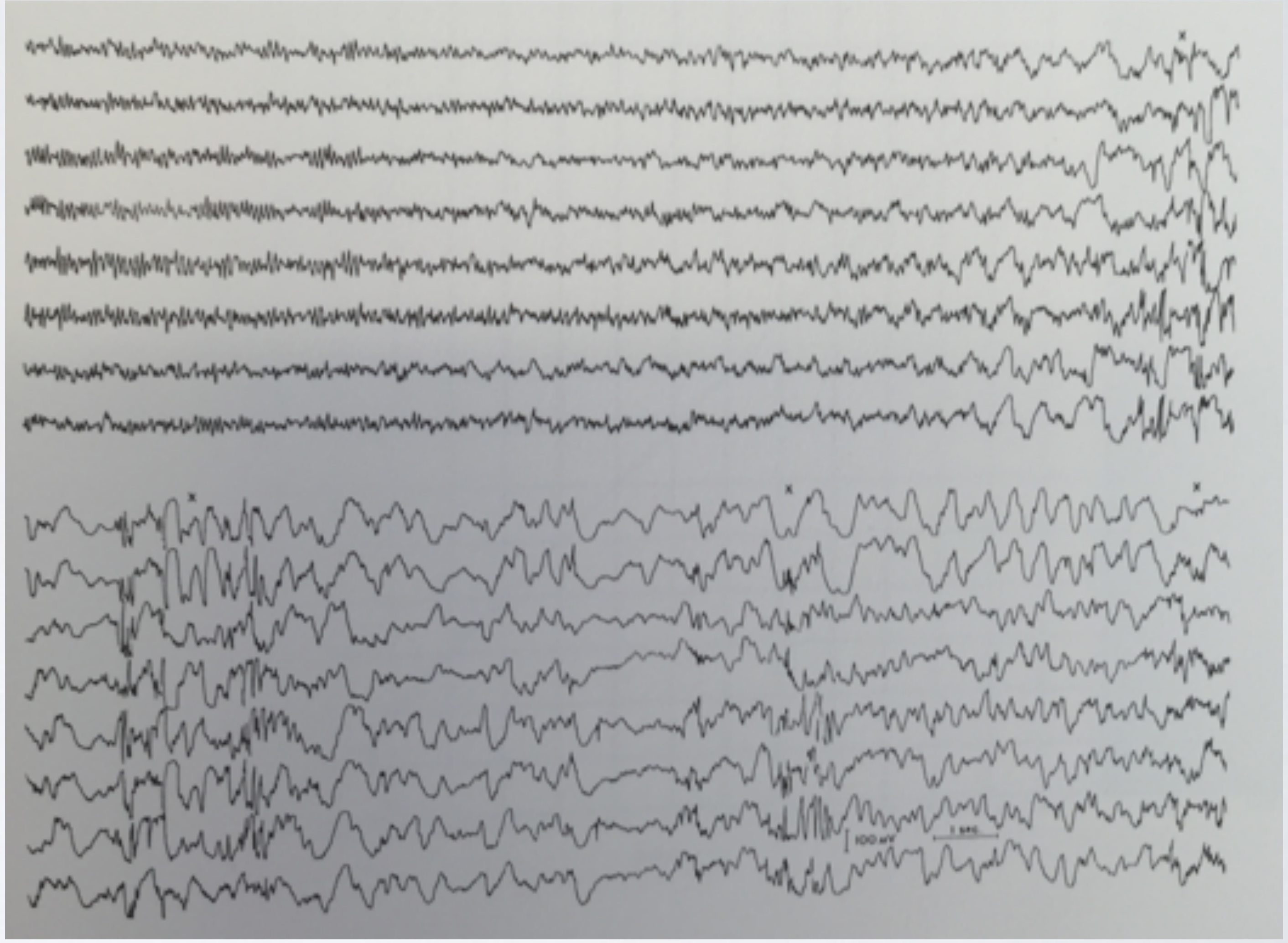
= proportion of a graphic's ink devoted to the non-redundant display of data-information

= 1.0 – proportion of a graphic that can be erased without loss of data-information.

Data-ink ratio



~0

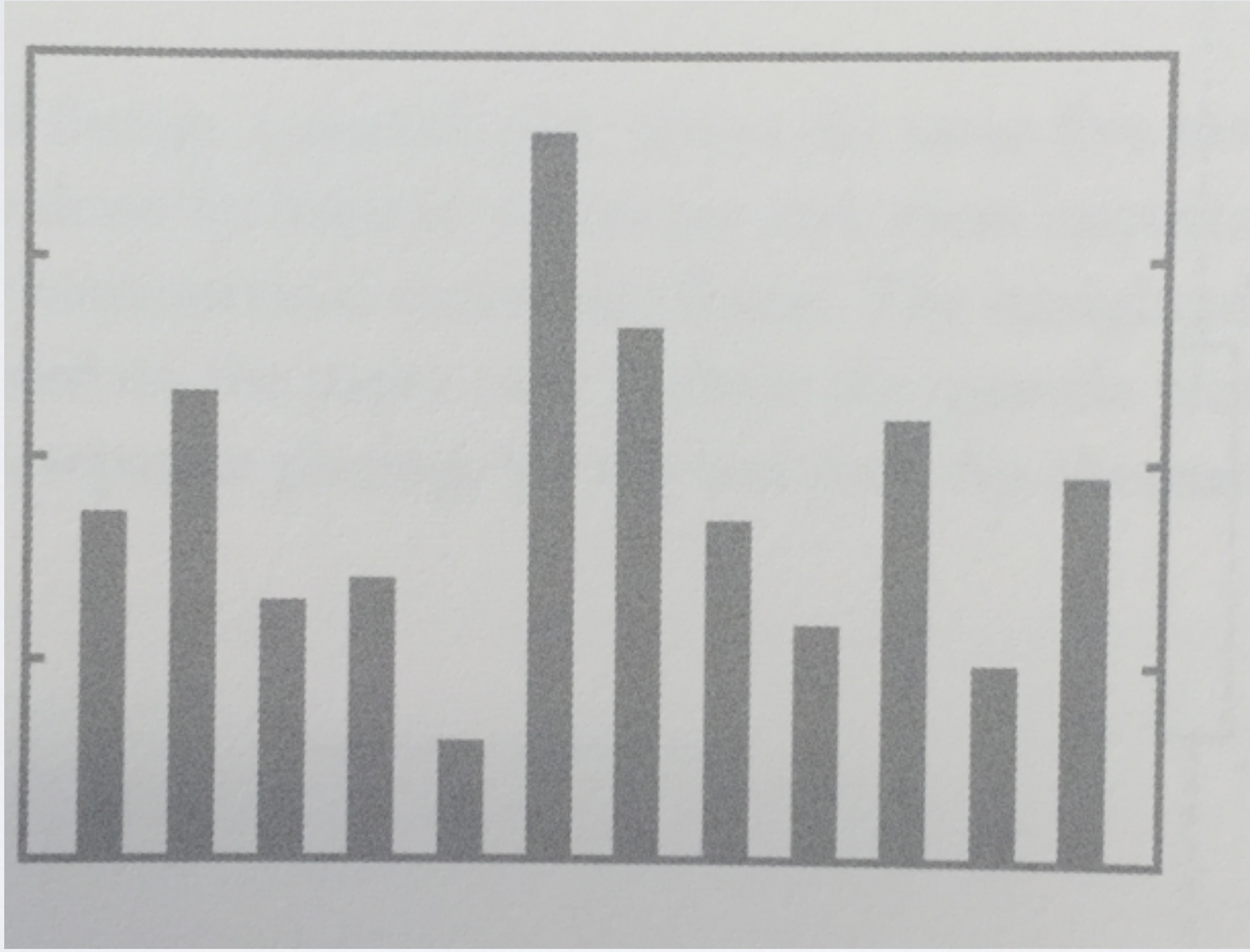


1.0

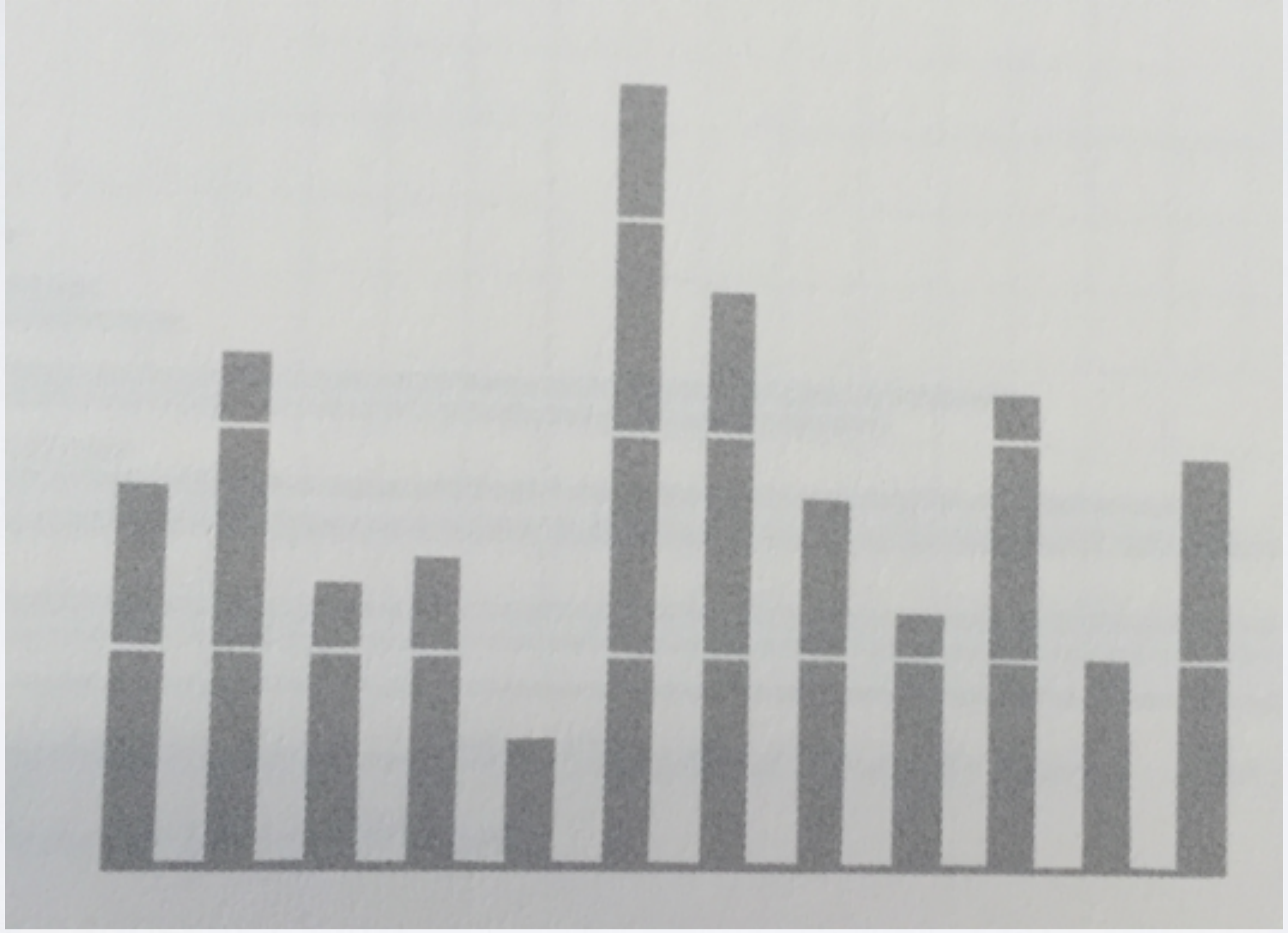
Design principles for data-ink

- (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**

Shneiderman's 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
  - Select a set of data to visualize and three 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