Information Visualization

SWE 632
Fall 2023

© Thomas LaToza

Administrivia

- HW6 due today
- HW7 due next week
- Project presentations in class in 2 weeks

Overview of Information Visualization

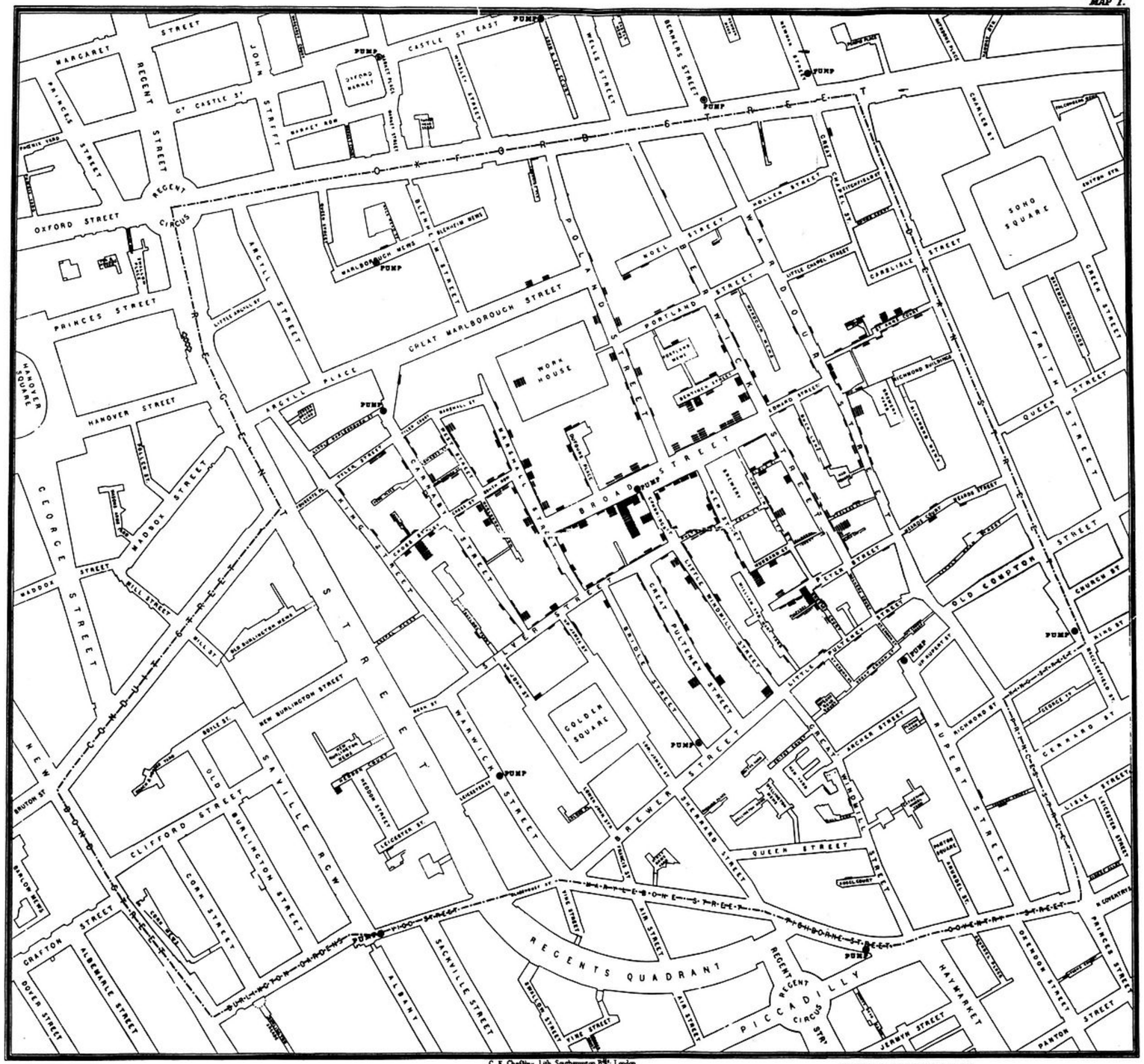
Today

- 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

MAP 1.



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

SCALE 80 INCHES TO A MILE.


WORK HOUSE
PUMP
BROAD STREET
PUMP
MARSHALL
CARNABY
KING STREET
GREAT PULTENEY STREET
LITTLE WINDMILL
BRIDLE STREET
SILVER STREET
GOLDEN SQUARE
PORTLAND MEWS
BENTINCK STREET
EDWARD STREET
RICHMOND BUILDINGS
RICHMOND MEWS
PETER STREET
OLD COMPT
LITTLE PULTENEY
DUFOURS PLACE
TYLER STREET
CROSS STREET
WEST STREET
NEW STREET
BREWERY
HUSBAND ST
CAMBRIDGE ST
LITTLE MARLBOROUGH ST
BURLINGTON STREET
MEARDS STREET
MEARDS COURT
DEAN ST
GREAT MARLBOROUGH PLACE
MARSHALL 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

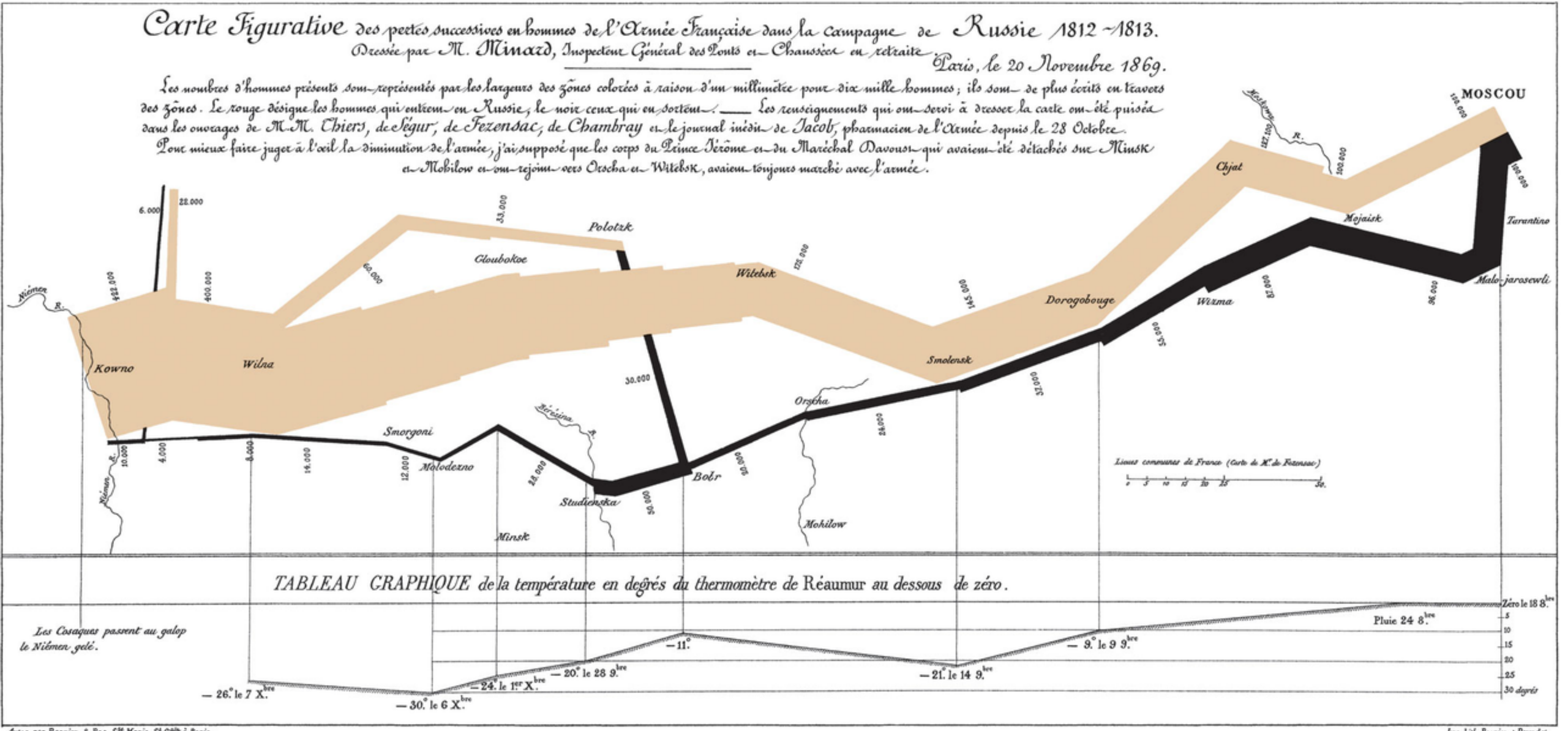
- Placed data in appropriate ***context*** for assessing cause & effect
  - Plotted on map, included well location
  - Reveals proximity as cause
- Made quantitative ***comparisons***
  - Fewer deaths closer to brewery, could investigate cause
- Considered ***alternative*** explanations & contrary cases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

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*

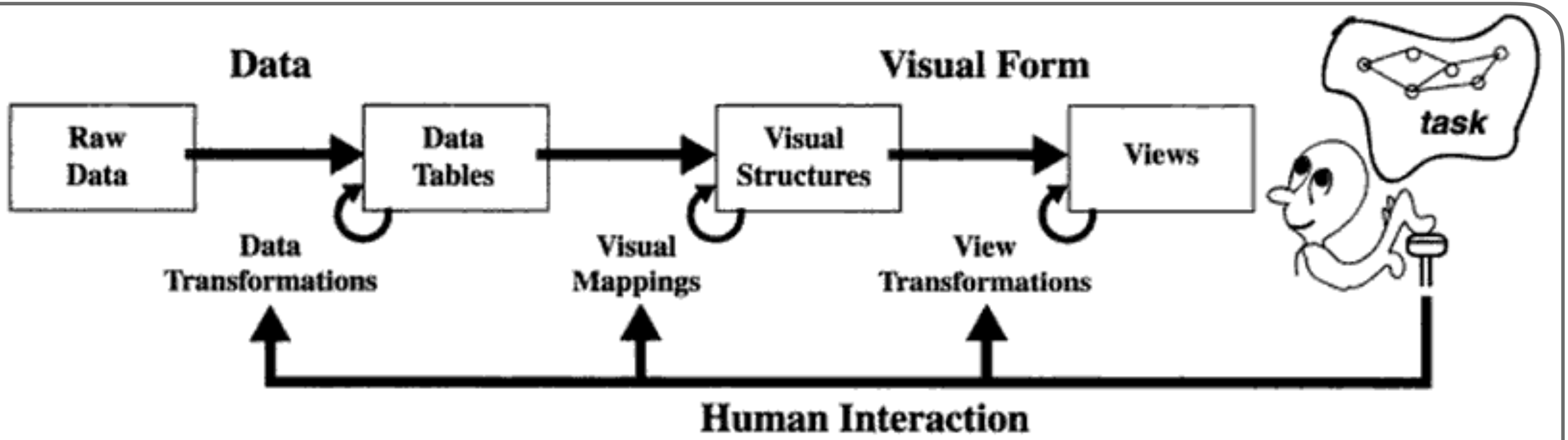
Charles Minard's Map of Napoleon's Russian Campaign of 1812



Drawn in Paris, November 20, 1869.

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 ***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

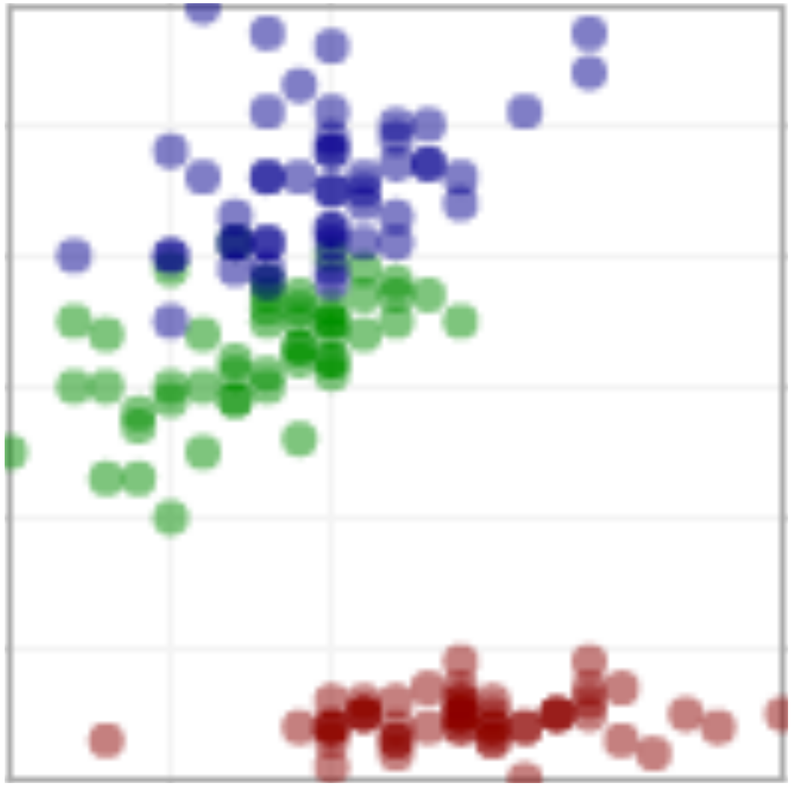
- Classing / binning: Quantitative —> ordinal
  - Maps ranges onto ***classes*** of variables
  - Can also count # of items in each class w/ histogram
- Sorting: Nominal —> ordinal
  - Add order between items in sets
- Descriptive statistics: mean, average, median, max, min, …

Visual Structures

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

Spatial Substrate

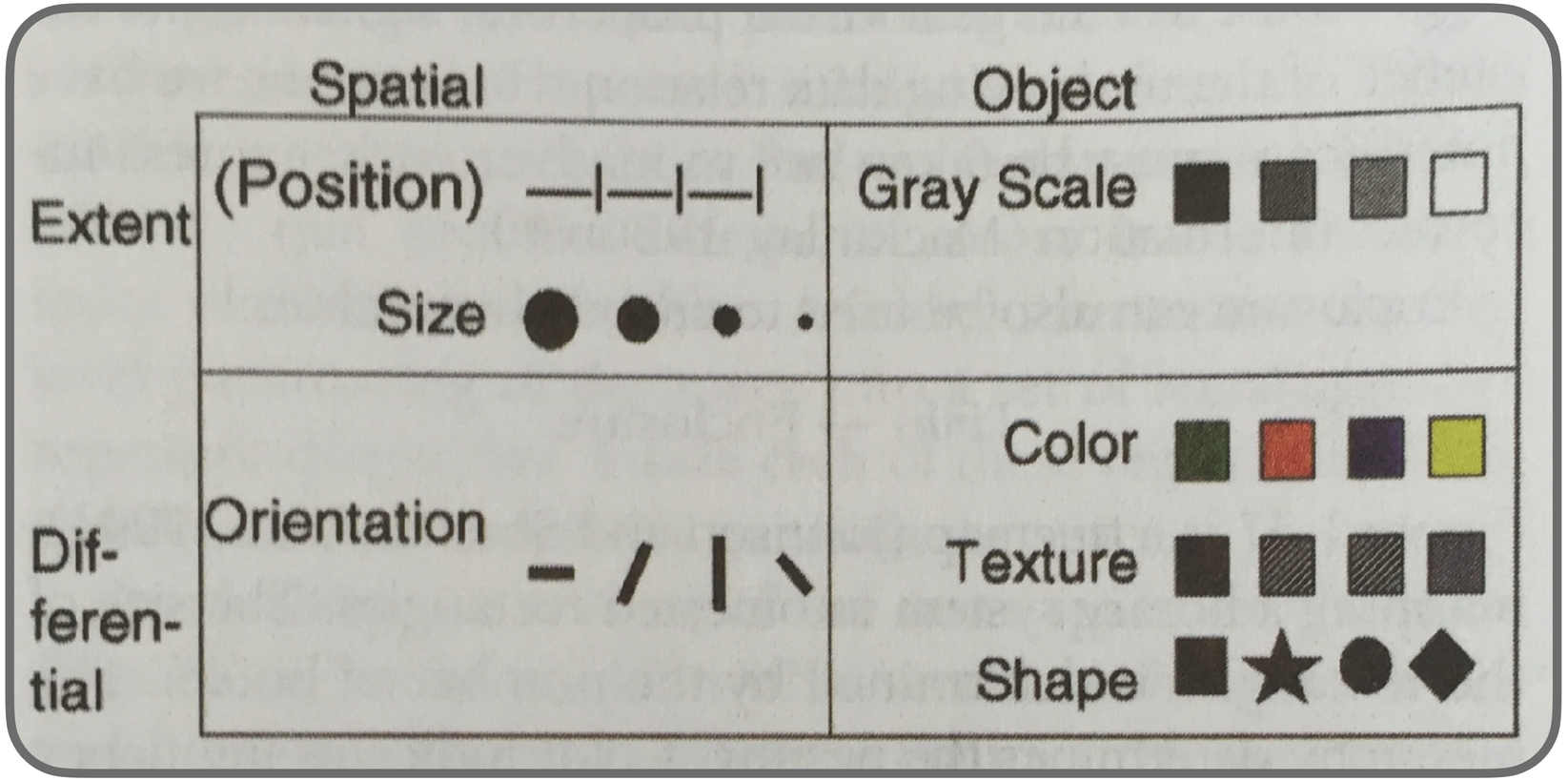
- Axes that divide space
- Types of axes - unstructured, nominal, ordinal, quantitative
- Composition - use of multiple orthogonal axes (e.g., 2D scatterplot, 3D)



Marks

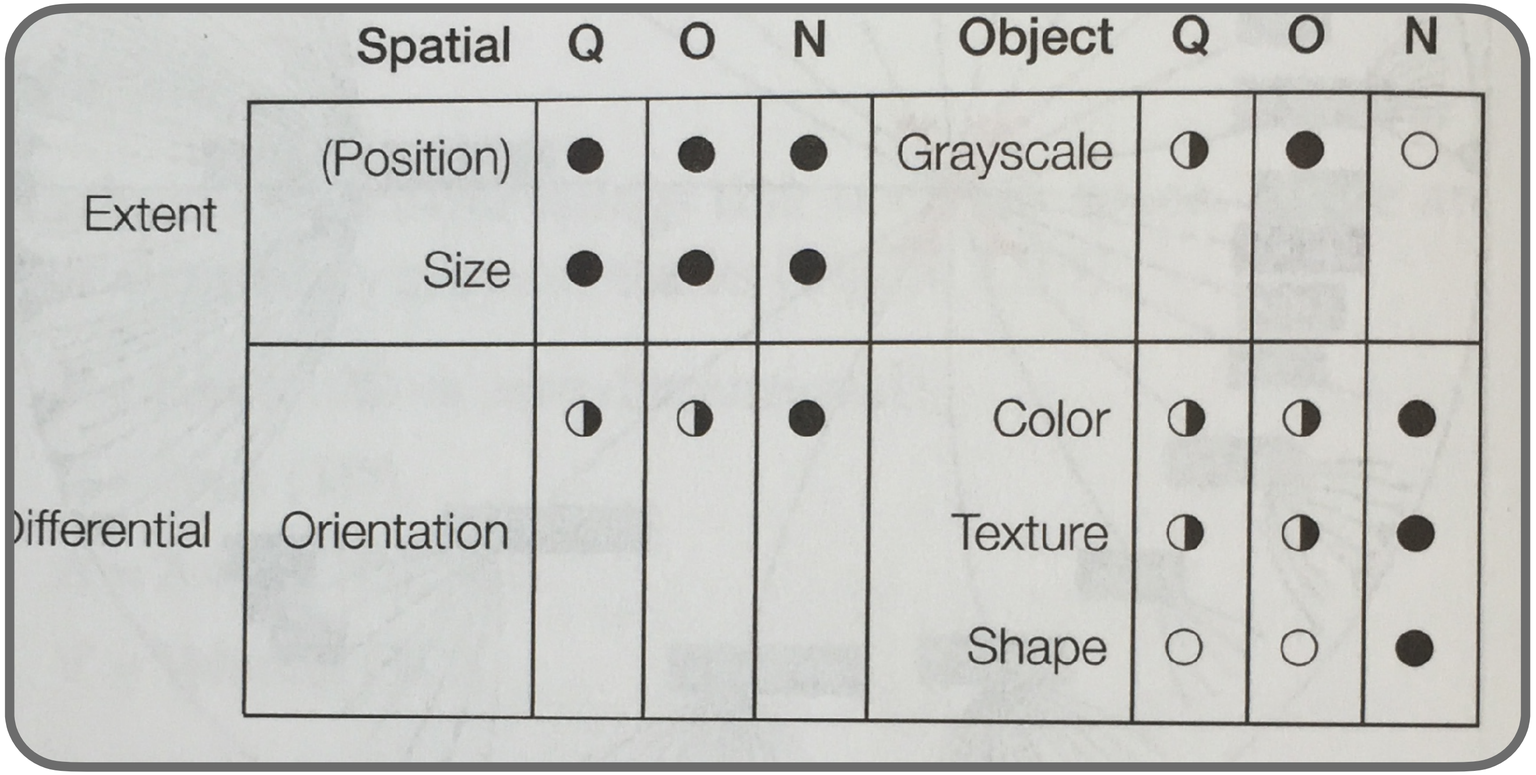
- Points (0D)
- Lines (1D)
- Areas (2D)
- Volumes (3D)

Marks' Graphical Properties



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

Effectiveness of Graphical Properties



	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

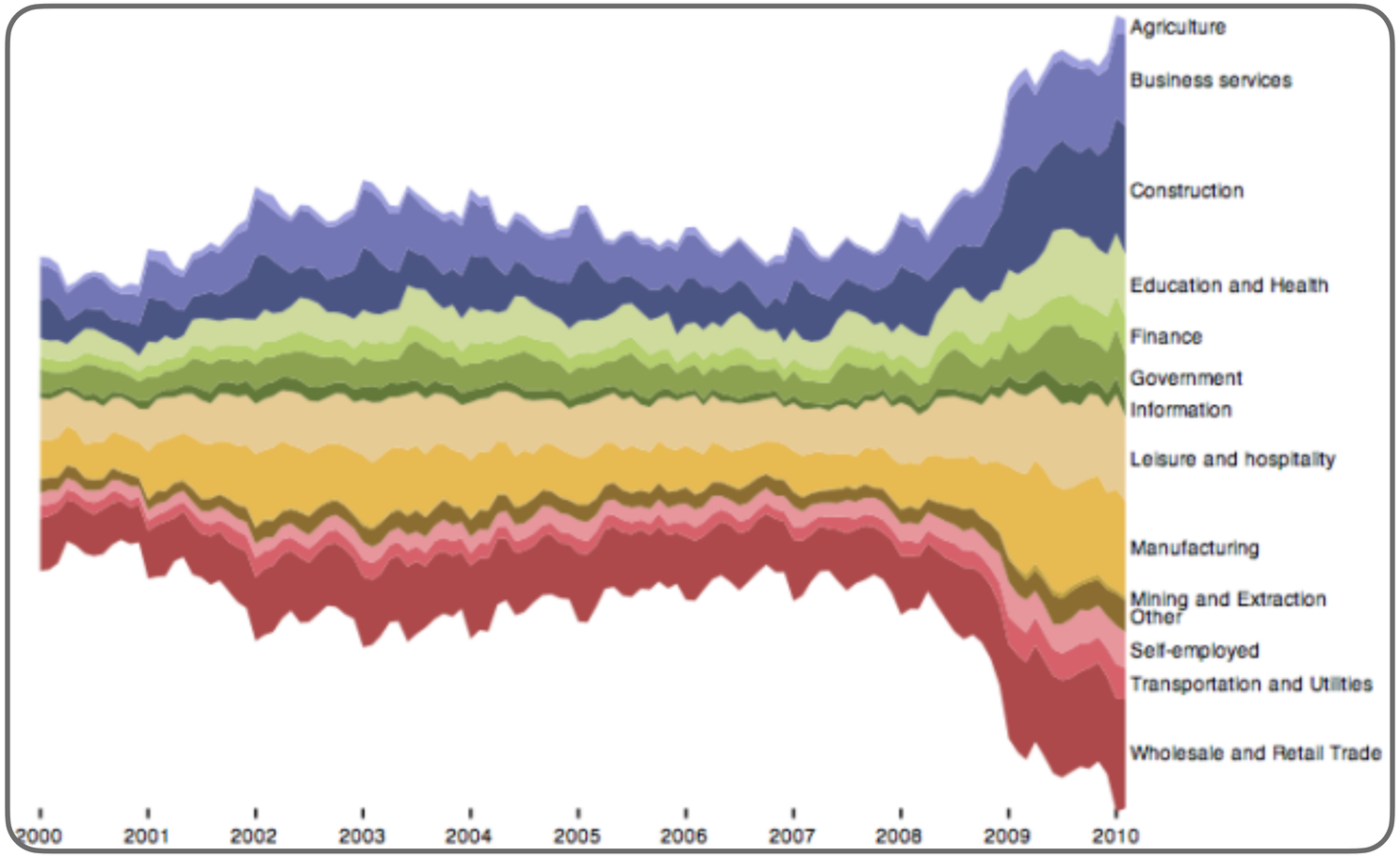
Animation

- 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

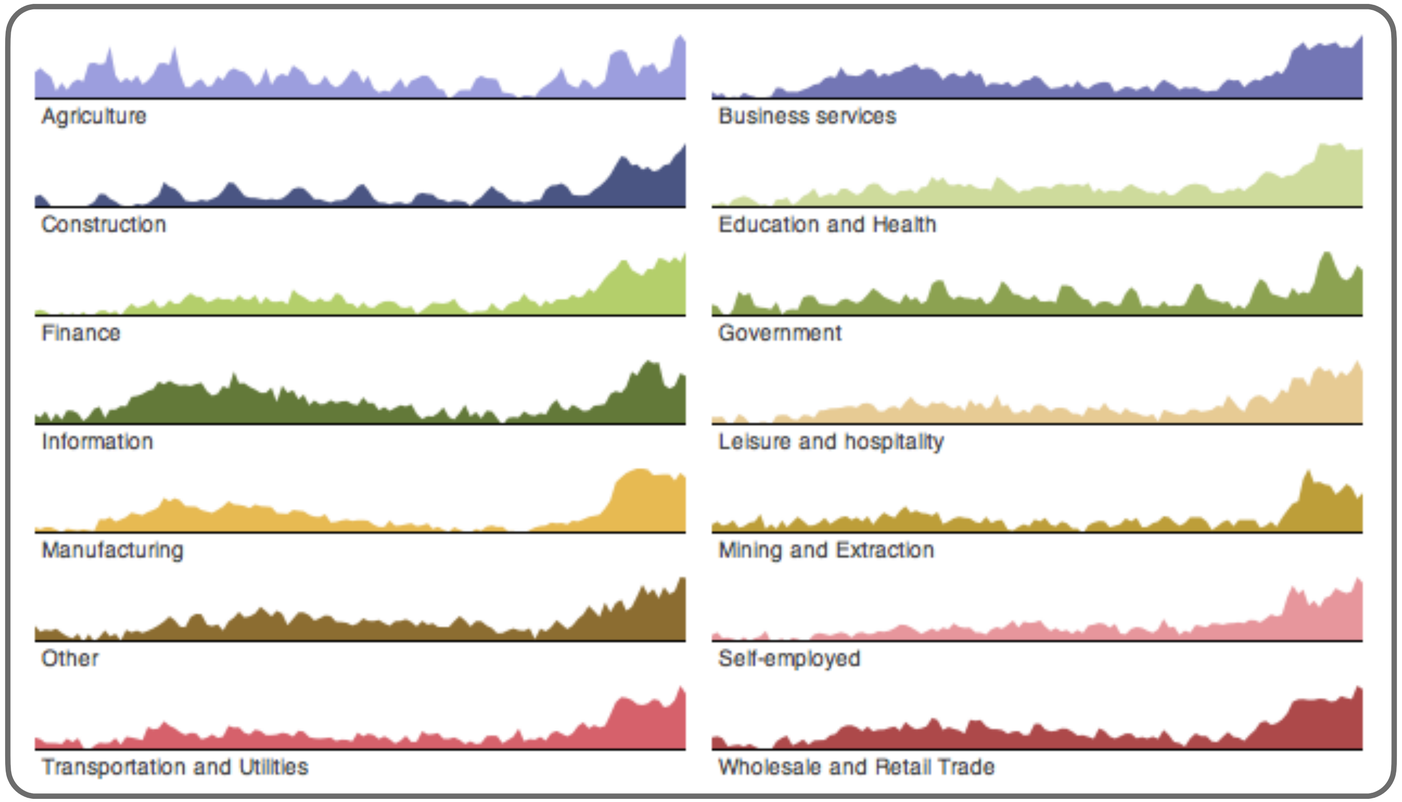
Time-series Data

Stacked Graph



- Supports visual summation of multiple components

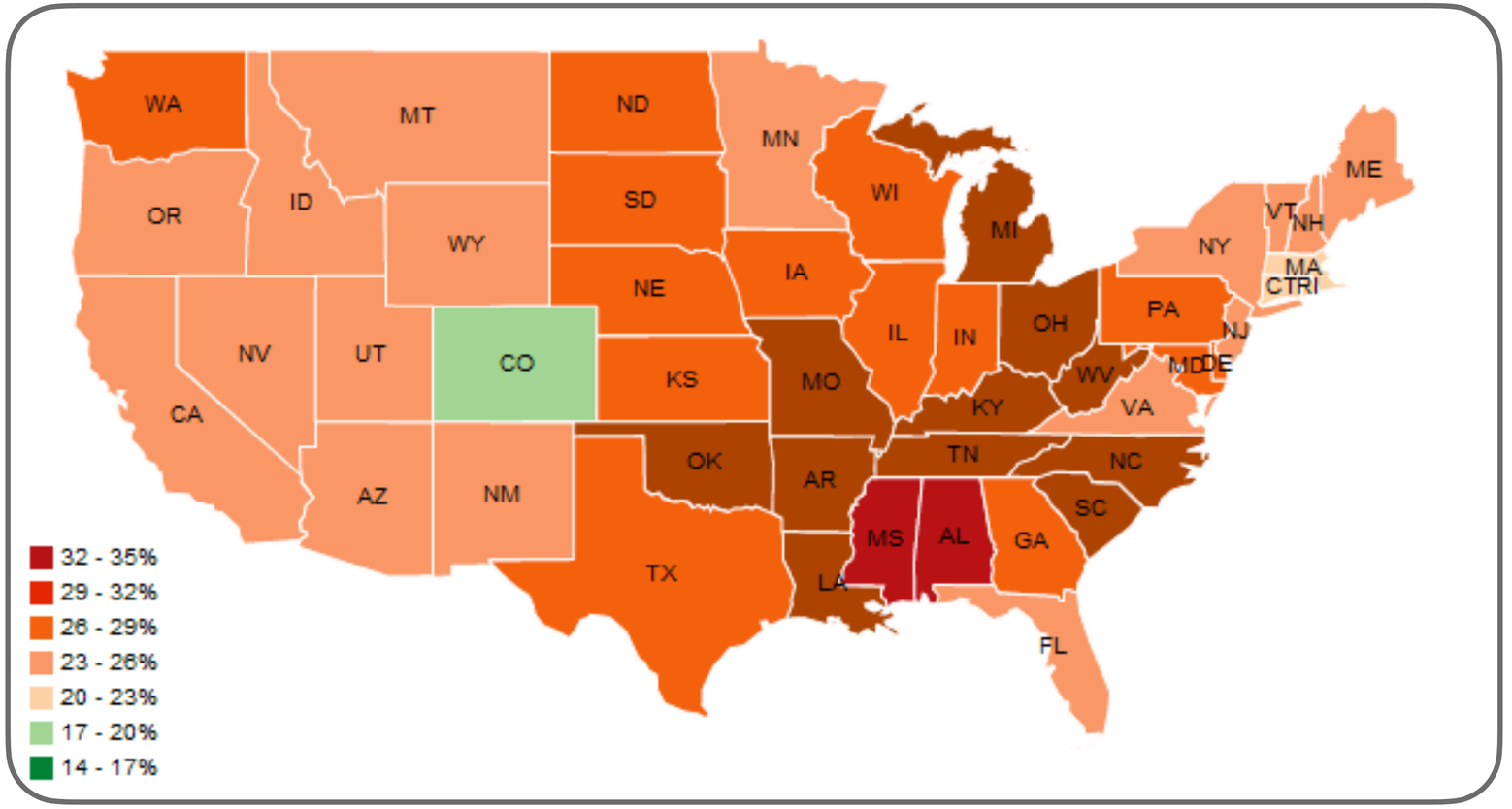
Small Multiples



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

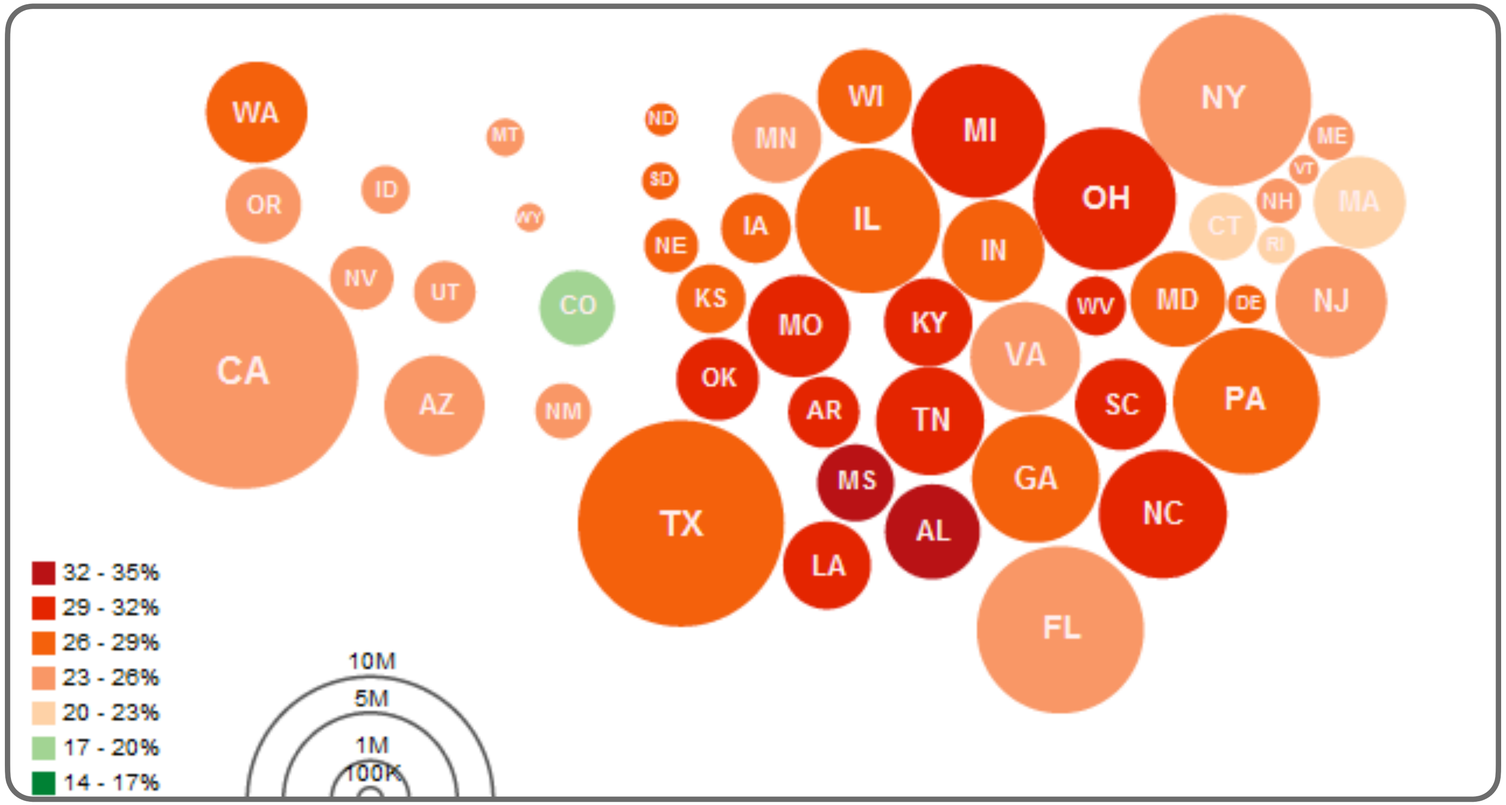
Maps

Choropleth Map



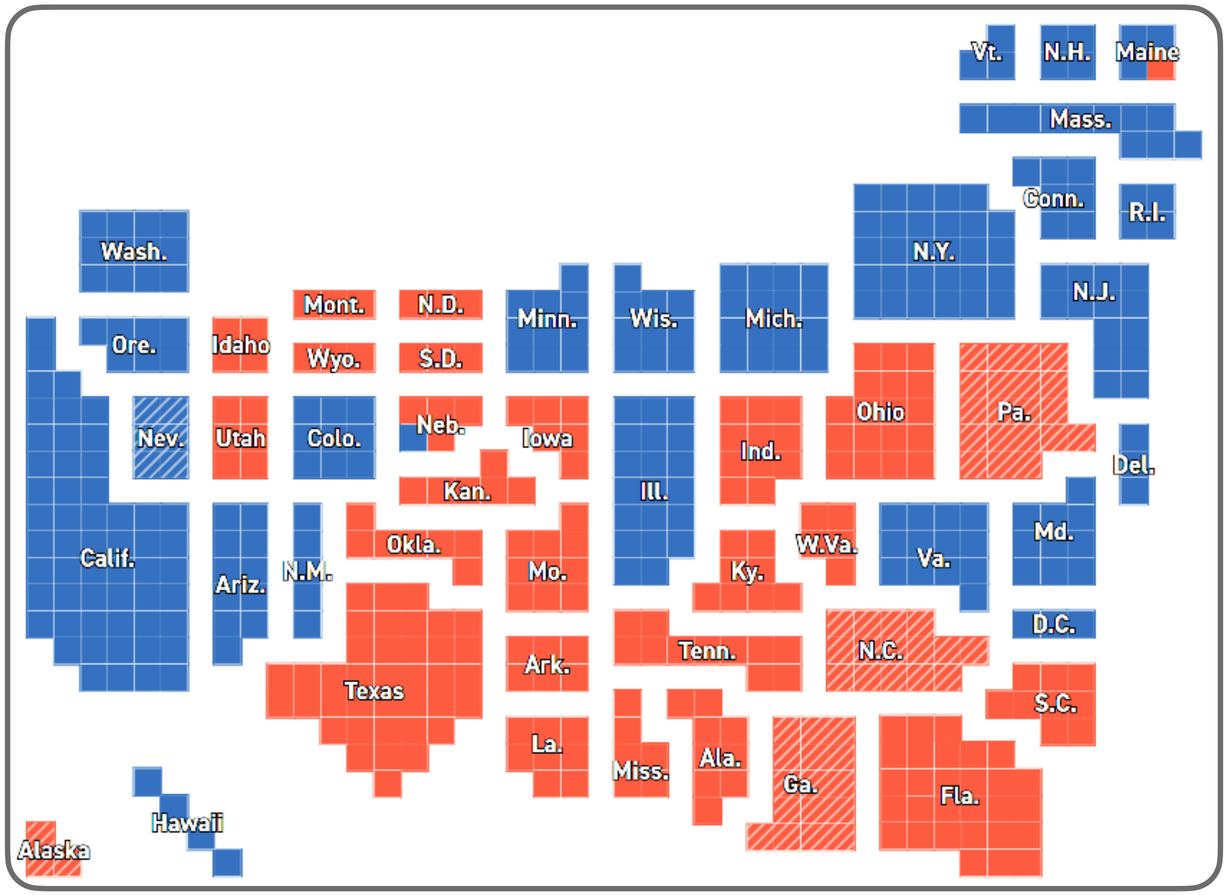
- Groups data by area, maps to color

Cartograms



- Encodes two variables w/ size & color

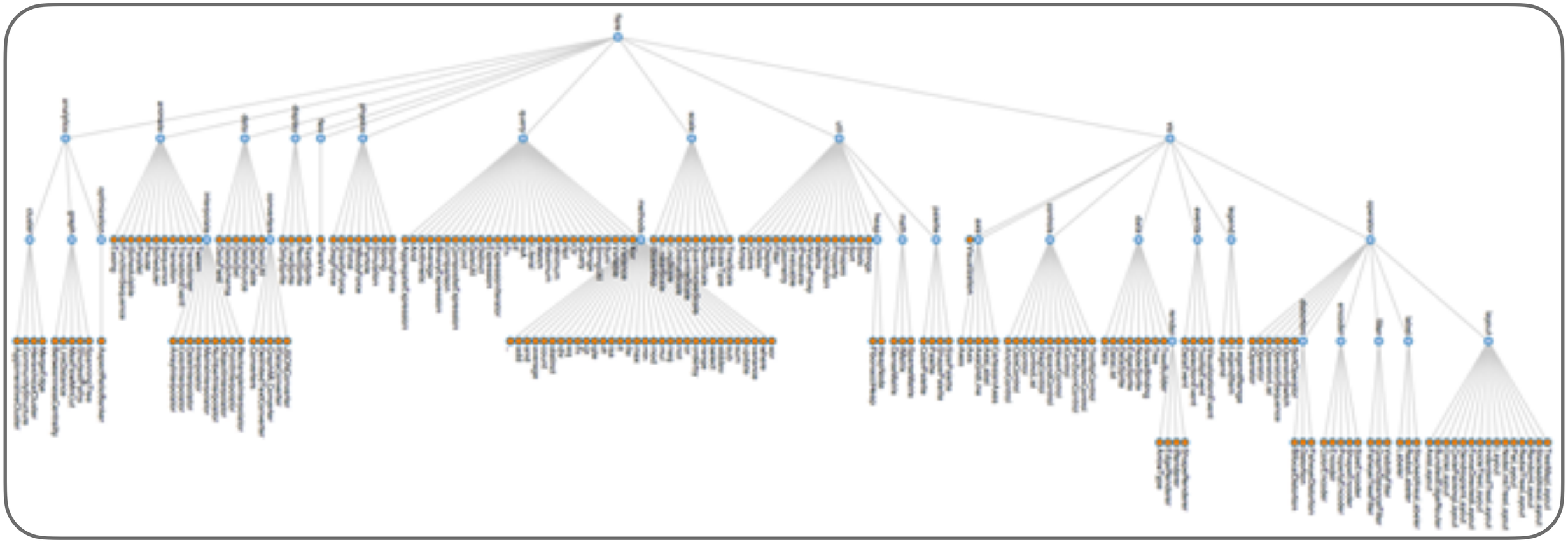
Cartograms



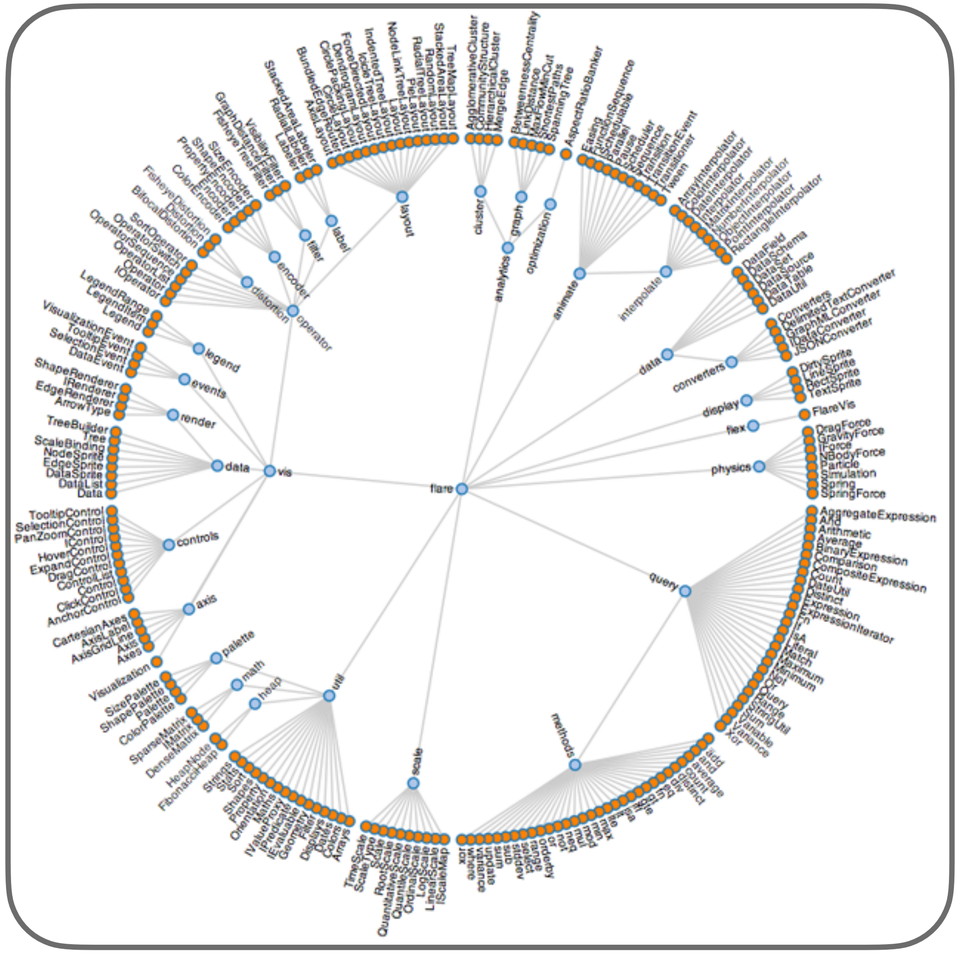
- Encodes two variables w/ size & color

Hierarchies

Node Link Diagram

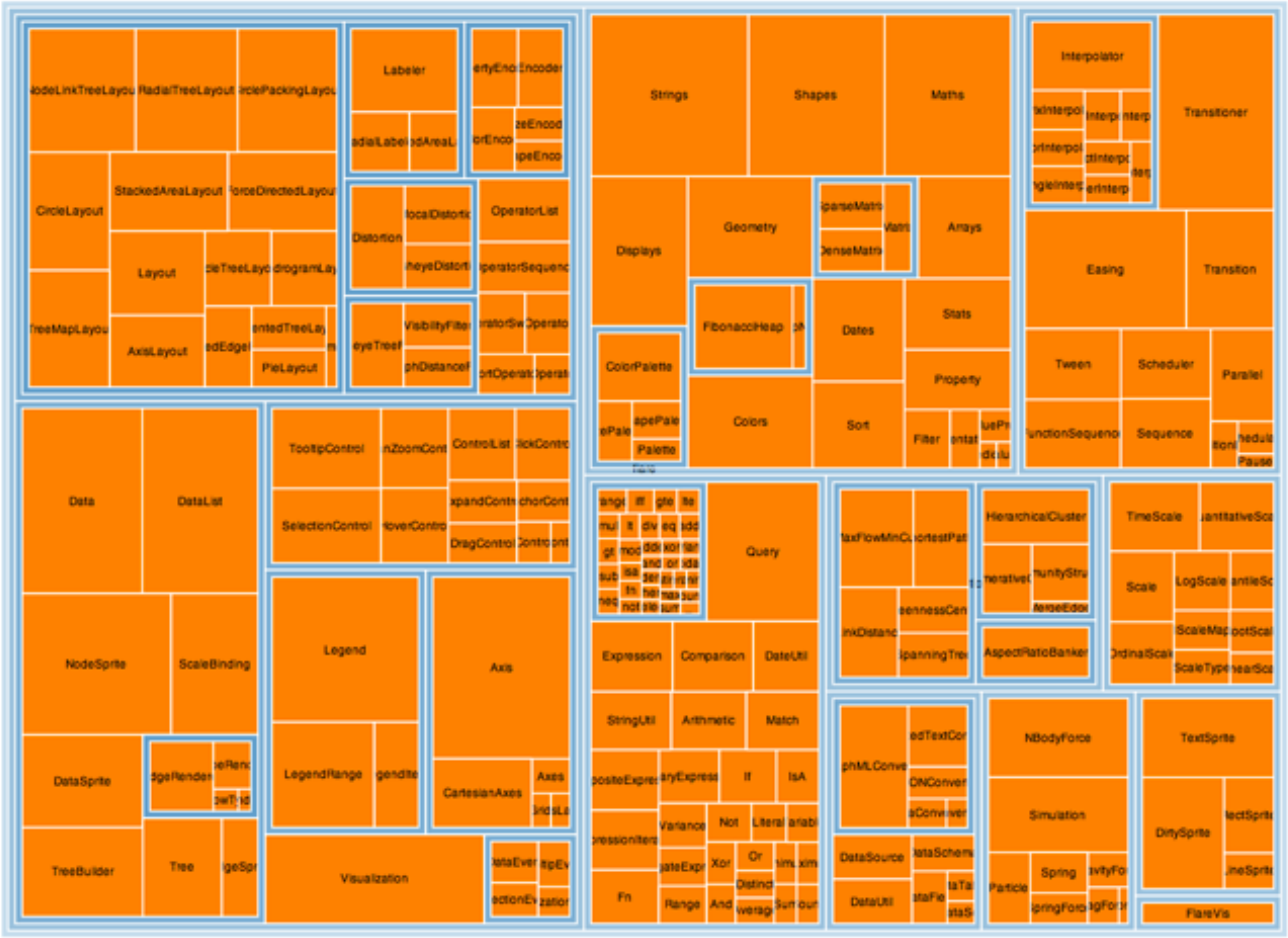


Dendrogram

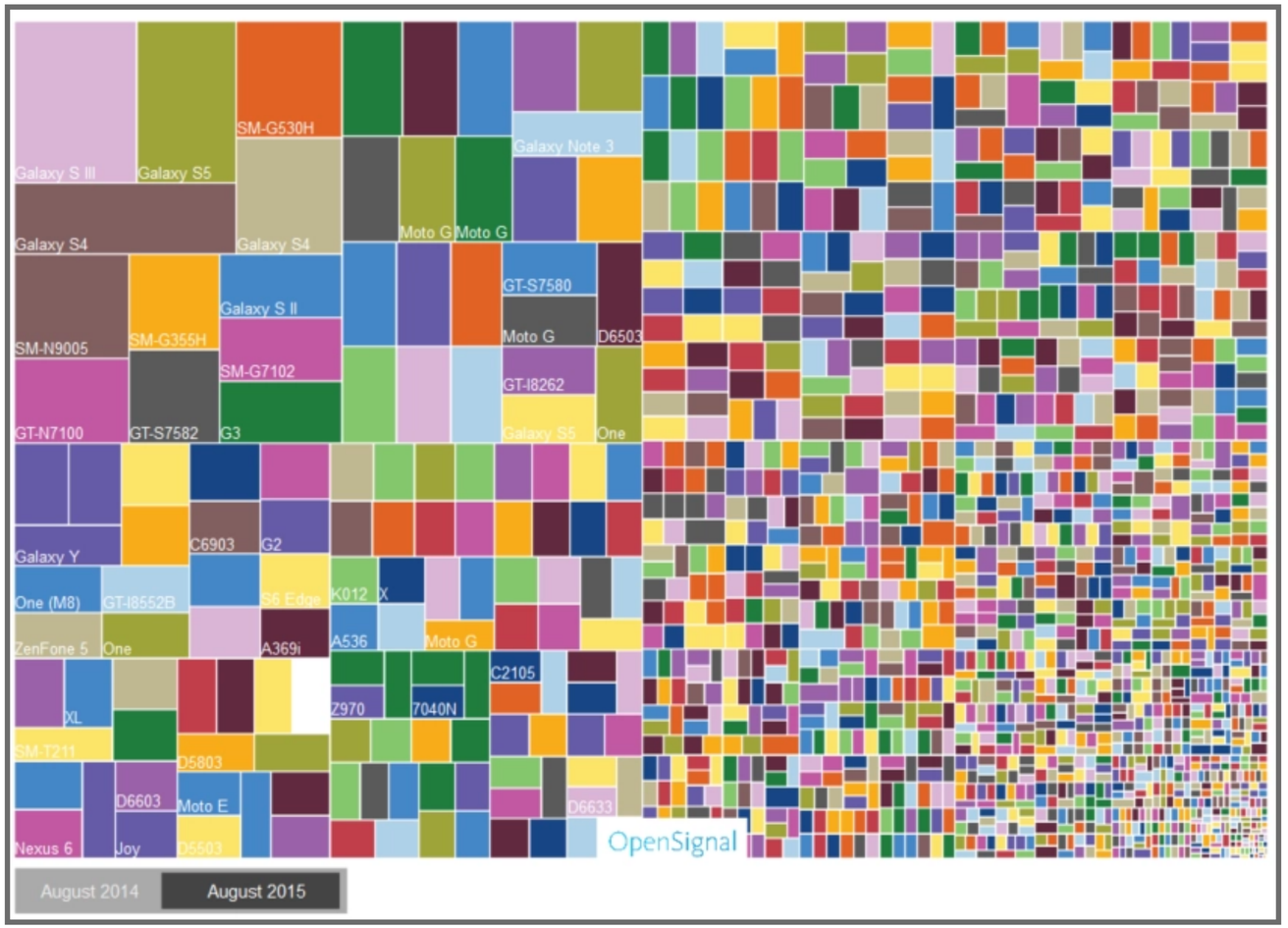


- Leaf nodes of hierarchy on edges of circle

Treemaps

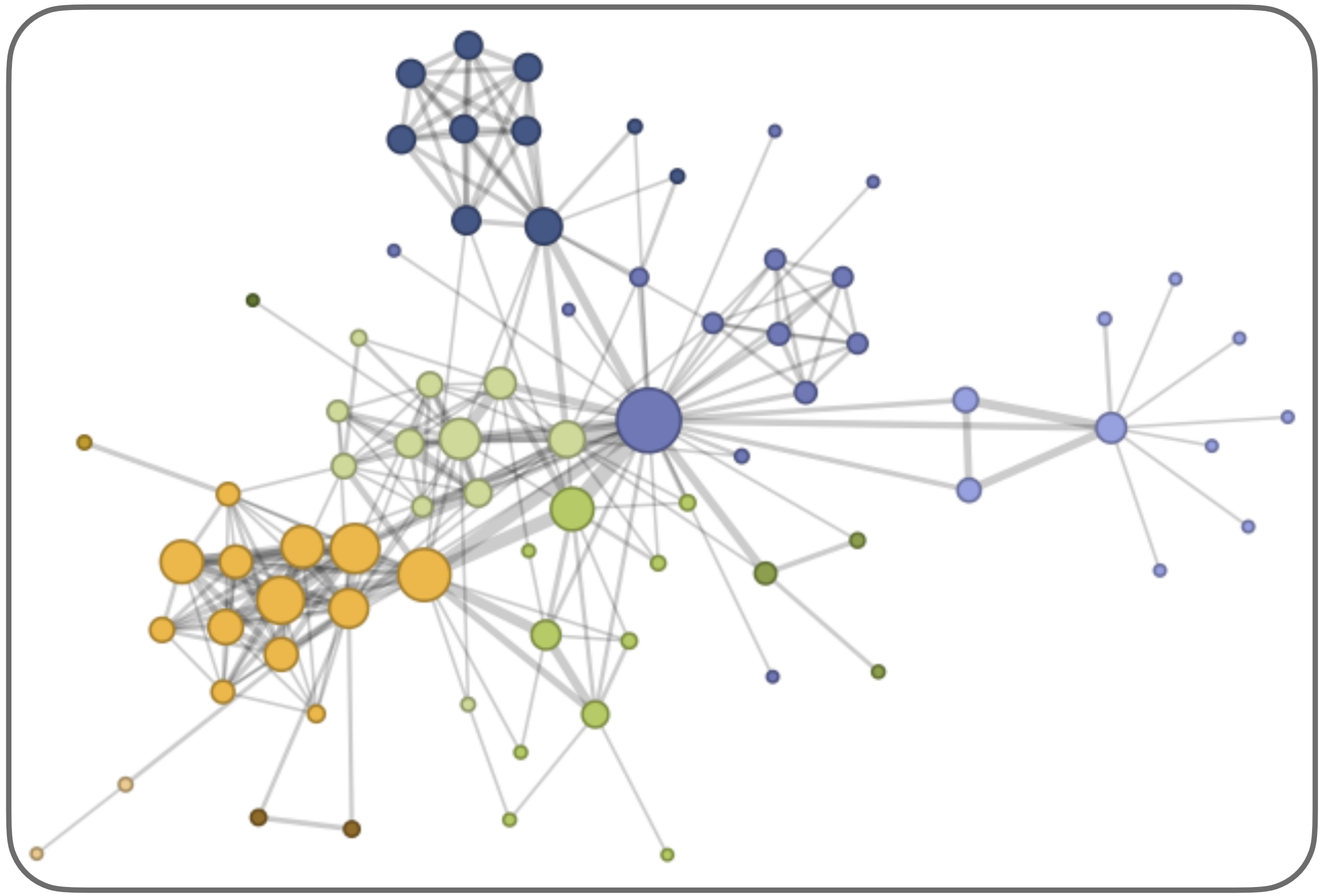

RadialTreeLayout
Labeler
StackedAreaLayout
CircleLayout
Layout
AxisLayout
PieLayout
Distortion
OperatorList
Strings
Shapes
Maths
Displays
Geometry
Arrays
FibonacciHeap
Dates
Stats
ColorPalette
Property
Colors
Sort
Filter
Palette
Interpolator
Transitioner
Easing
Transition
Tween
Scheduler
Parallel
Sequence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

Treemaps


Galaxy S III
Galaxy S5
SM-G530H
Galaxy Note 3
Galaxy S4
Galaxy S4
Moto G
Moto G
GT-S7580
Galaxy S II
SM-N9005
SM-G355H
Moto G
D6503
SM-G7102
GT-I8262
GT-N7100
GT-S7582
G3
Galaxy S5
One
Galaxy Y
C6903
G2
One (M8)
GT-I8552B
S5 Edge
K012
X
ZenFone 5
One
A369i
A536
Moto G
C2105
Z970
7040N
XL
SM-T211
D5803
D6603
Moto E
D6633
Nexus 6
Joy
D5503
OpenSignal
August 2014
August 2015

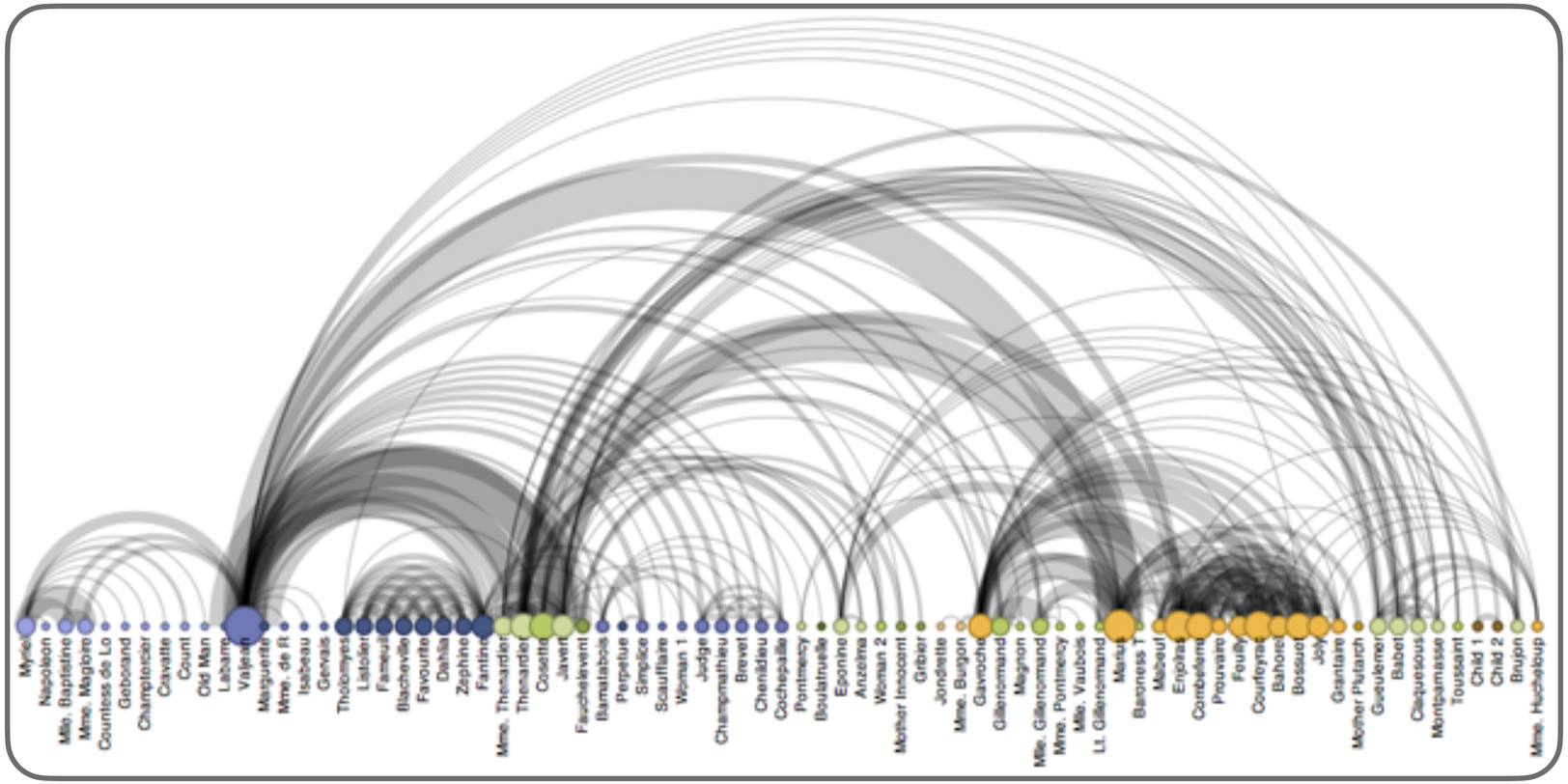
Networks

Force-directed Layout



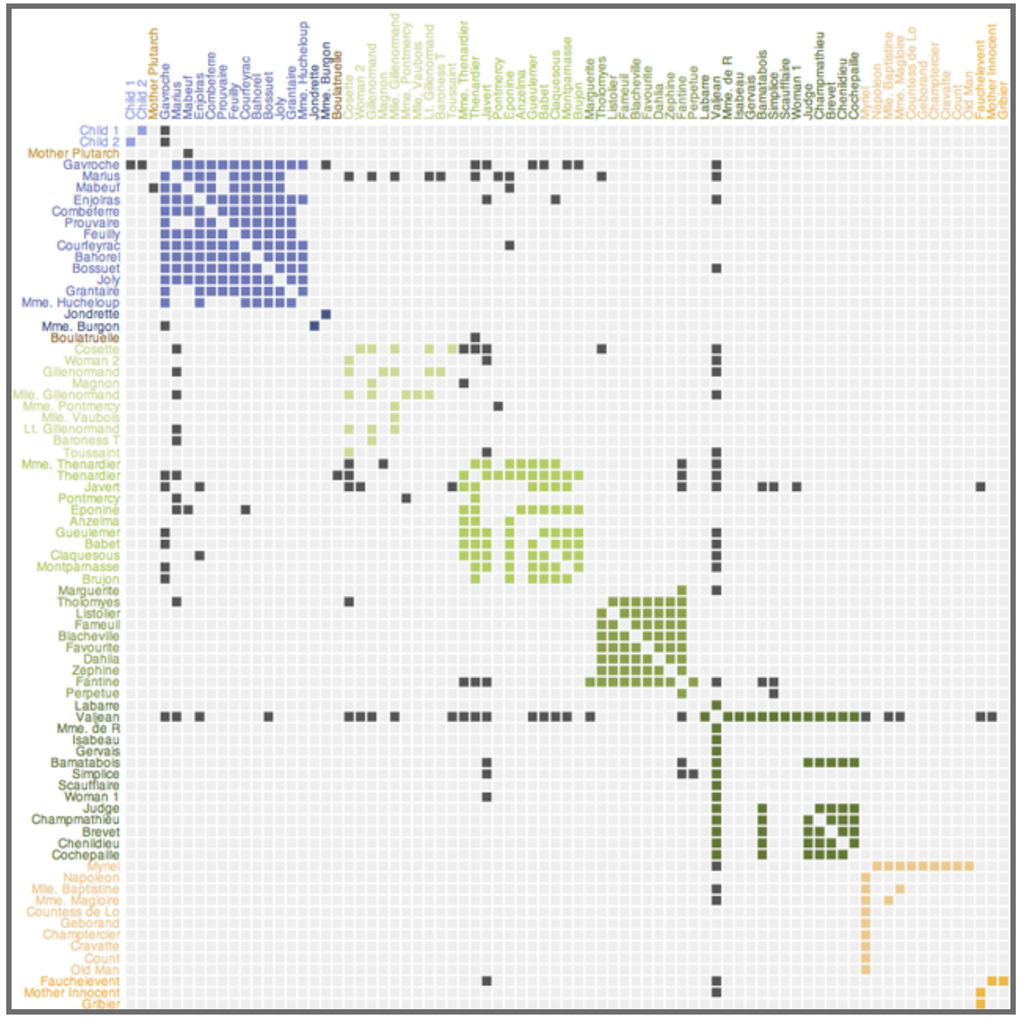
- Edges function as springs, find least energy configuration

Arc Diagram



- Can support identifying cliques & bridges w/ right order

Adjacency Matrix



Design Considerations

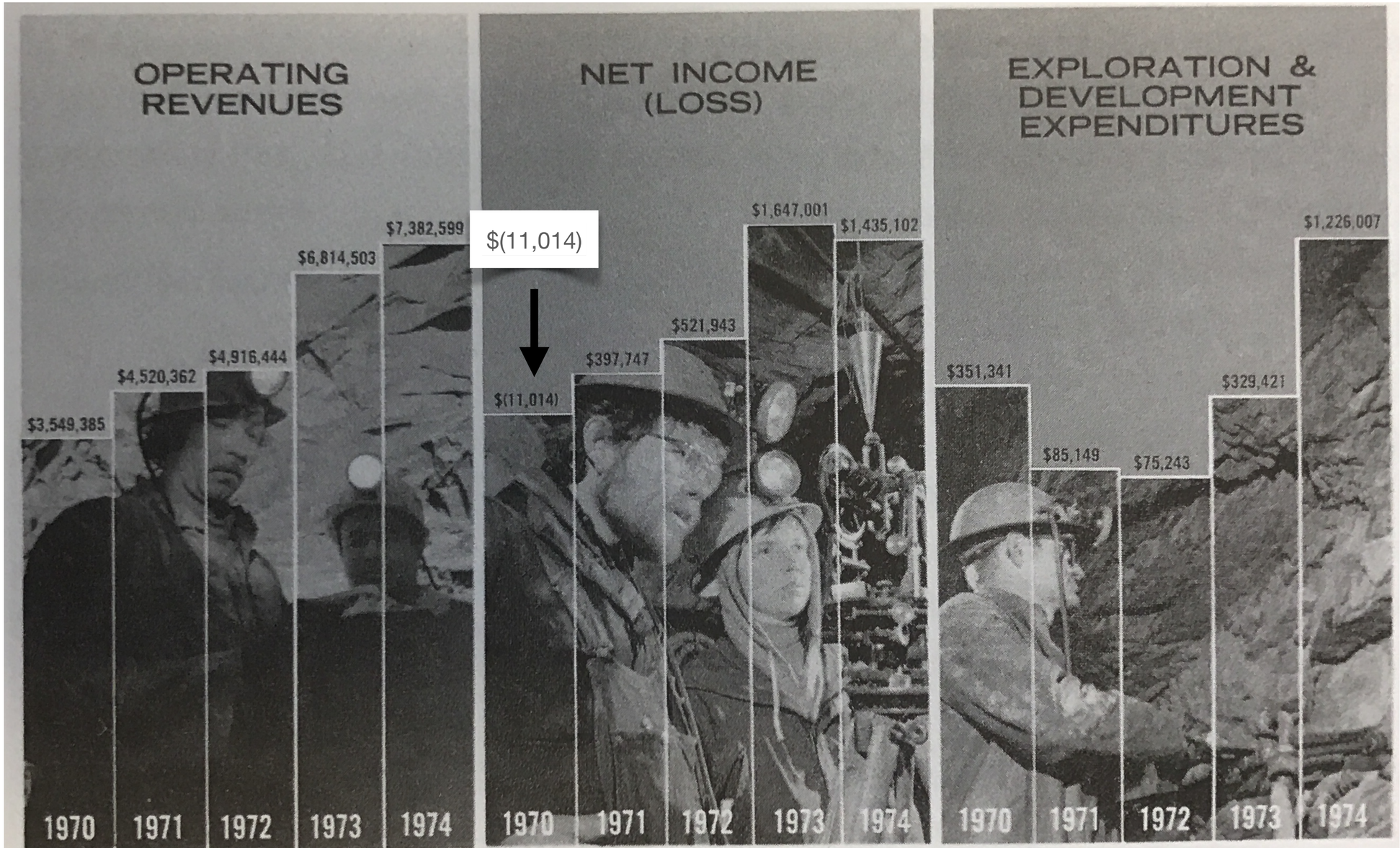
Tufte’s principles of graphical excellence

- 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

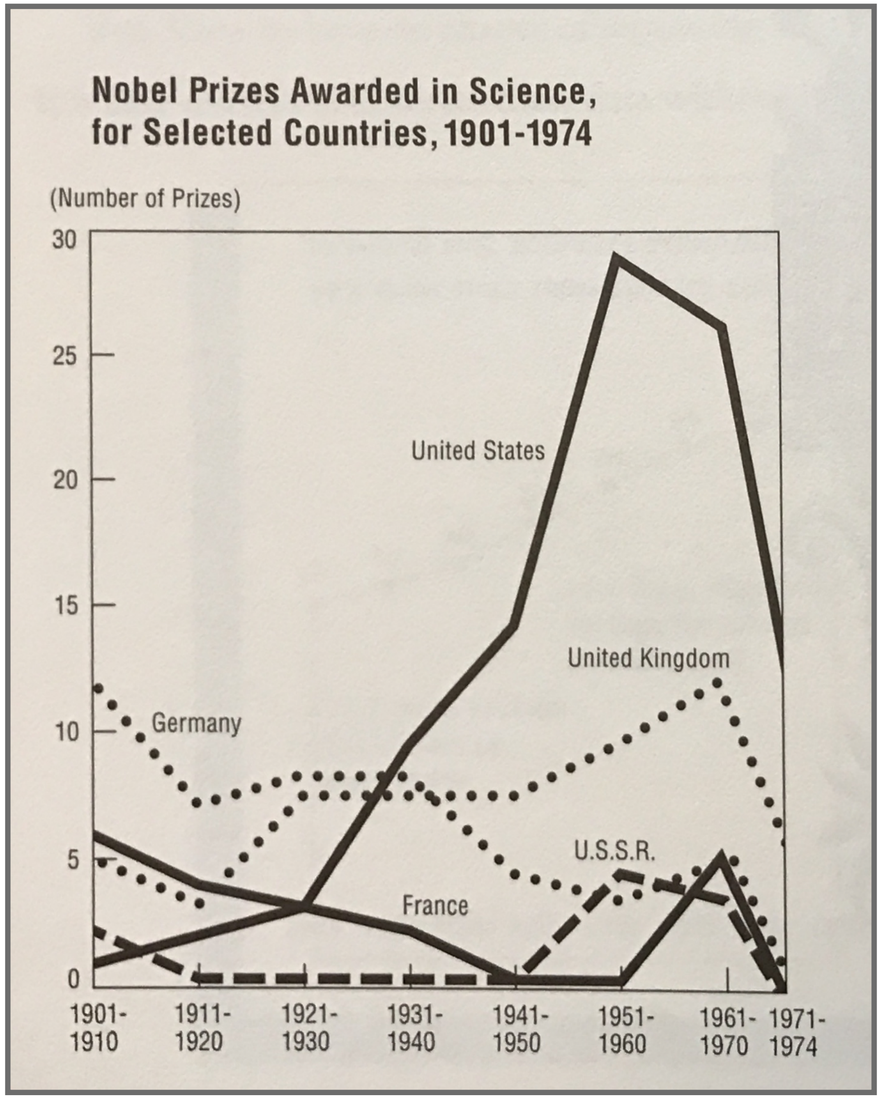
Distortions in Visualizations

- 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

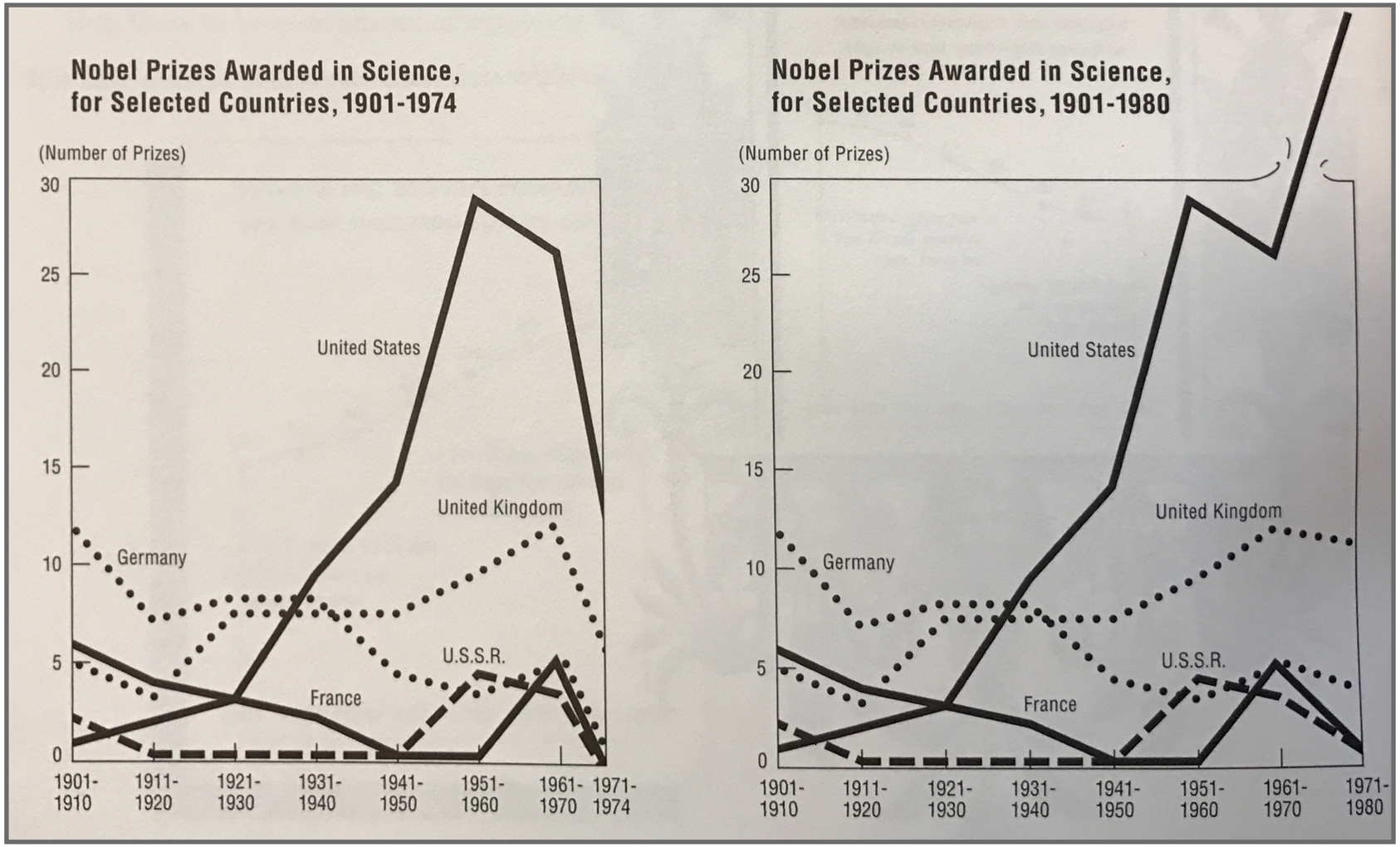
Example


OPERATING REVENUES
$3,549,385
$4,520,362
$4,916,444
$6,814,503
$7,382,599
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

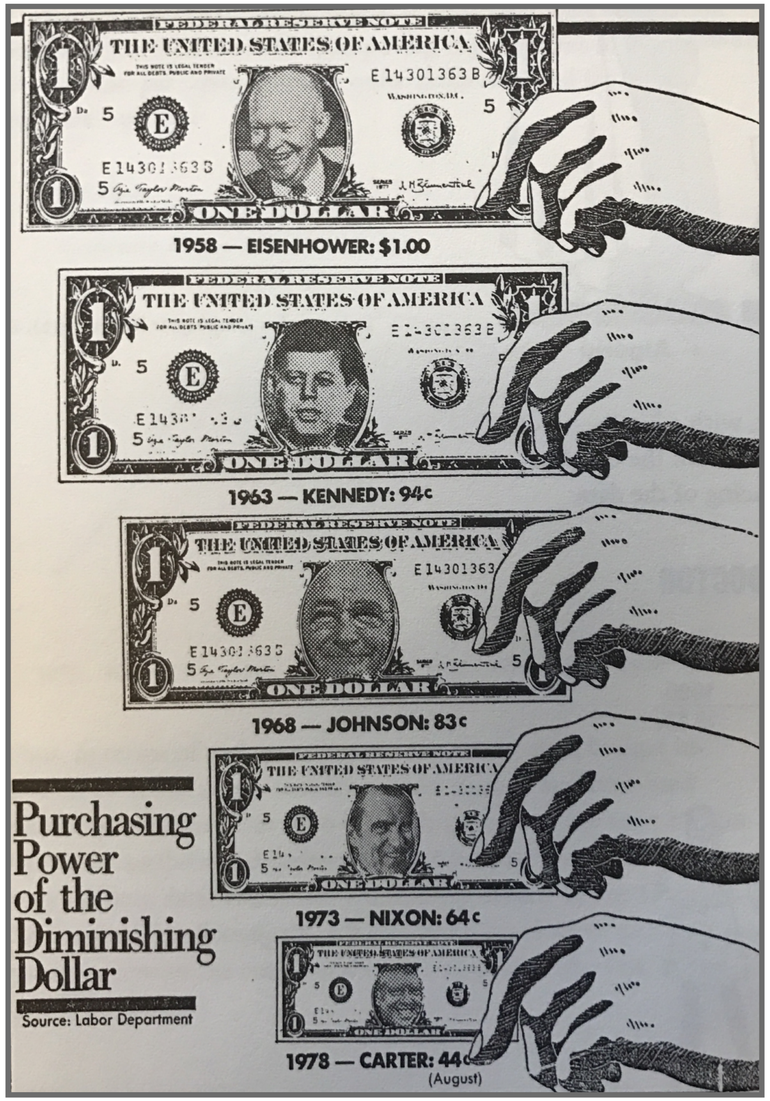
Example


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

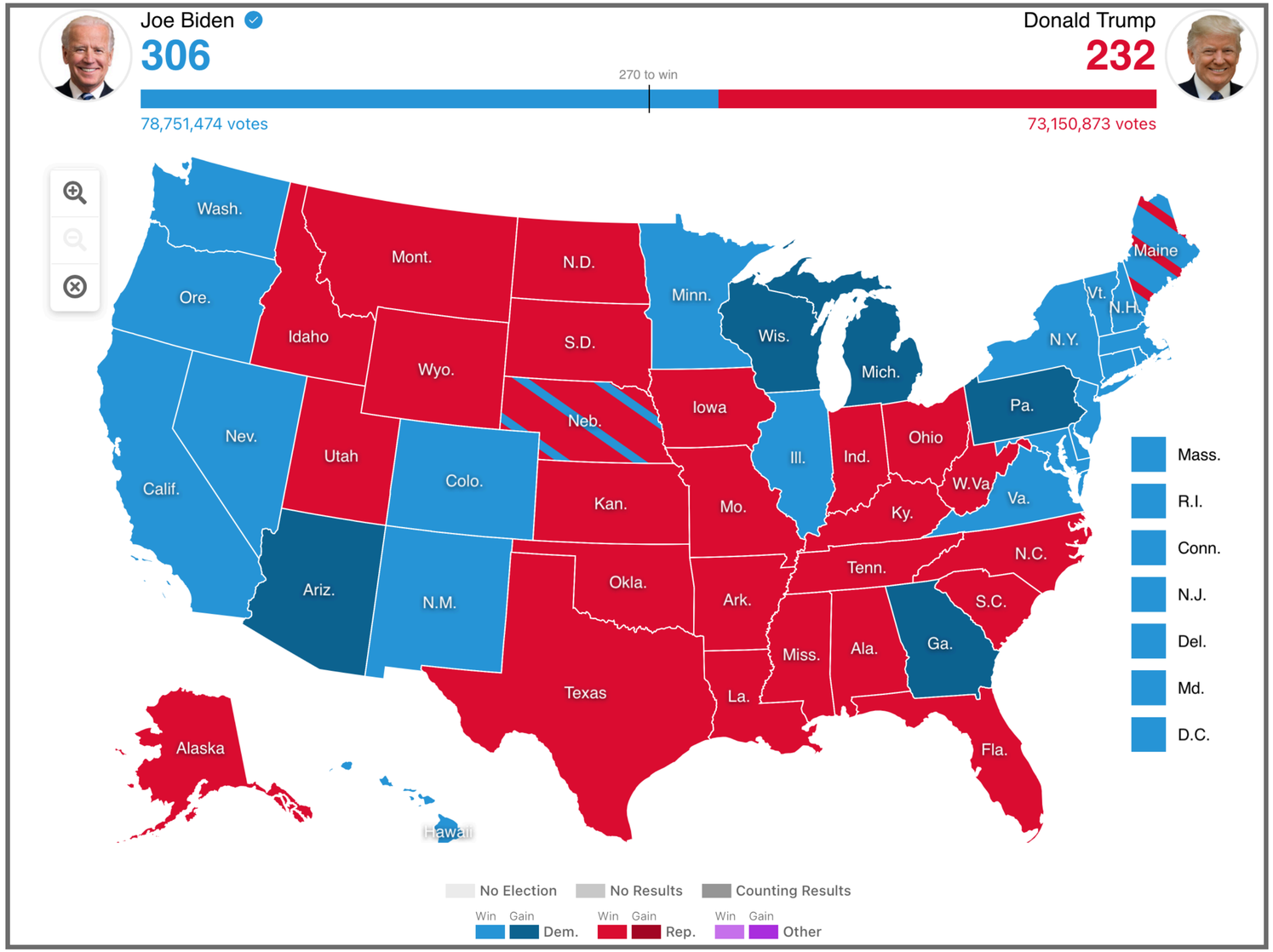
Example (corrected)



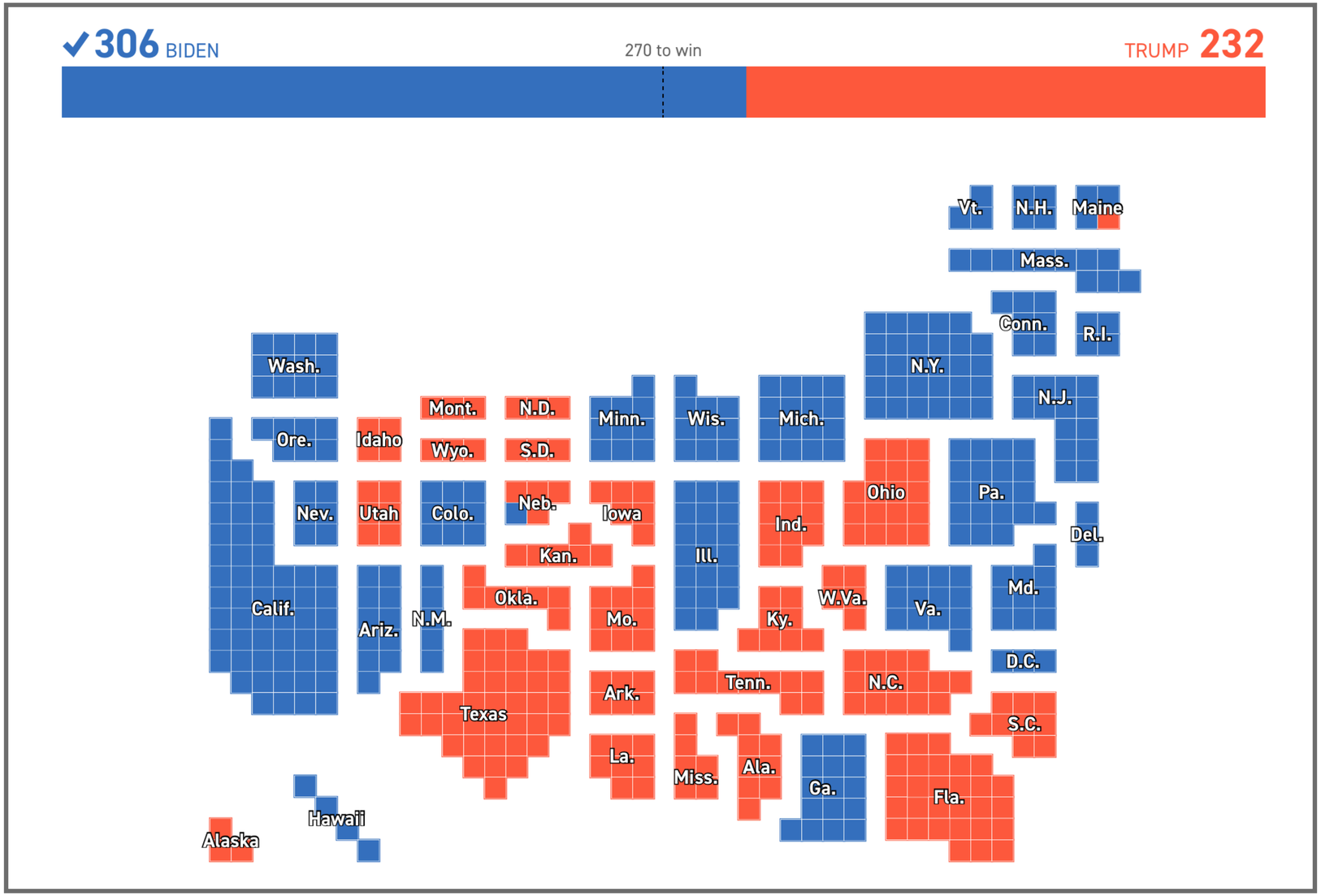
Example


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

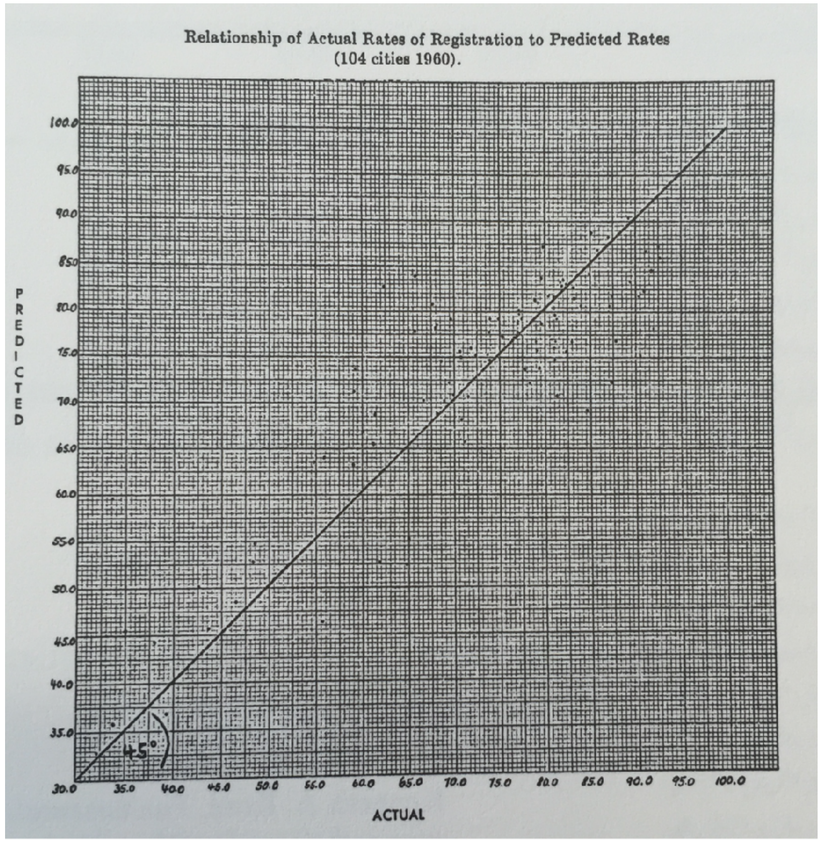
- Data-ink - non-redundant ink encoding data information

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

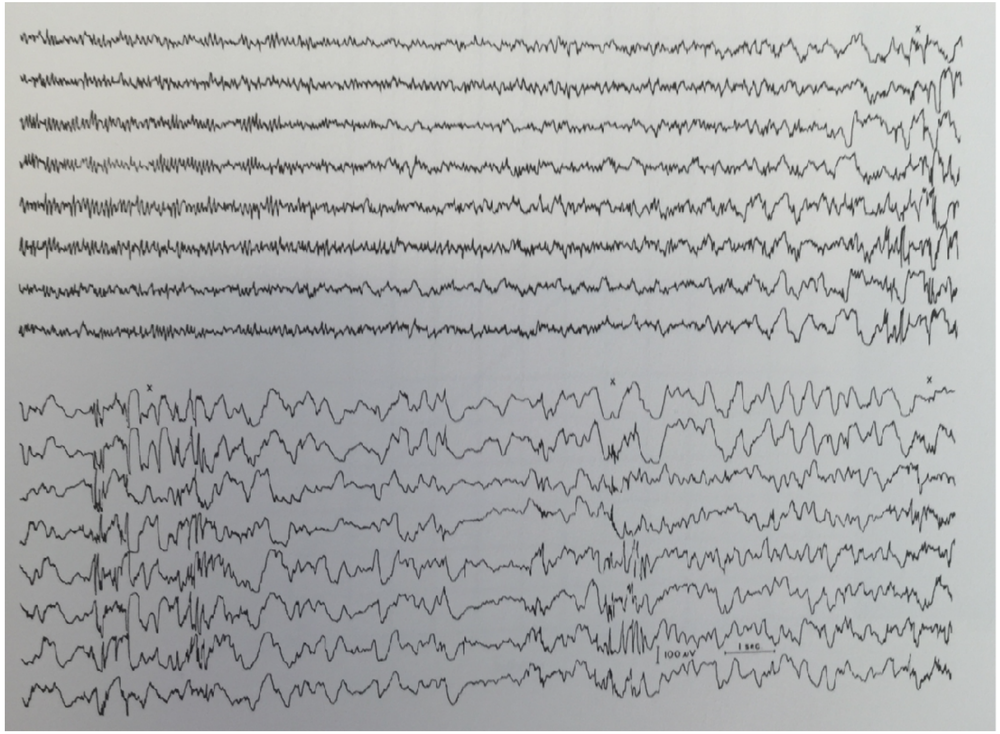
$$= \text{proportion of a graphic's ink devoted to the non-redundant display of data-information}$$

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

Examples of Data-ink Ratio



~0

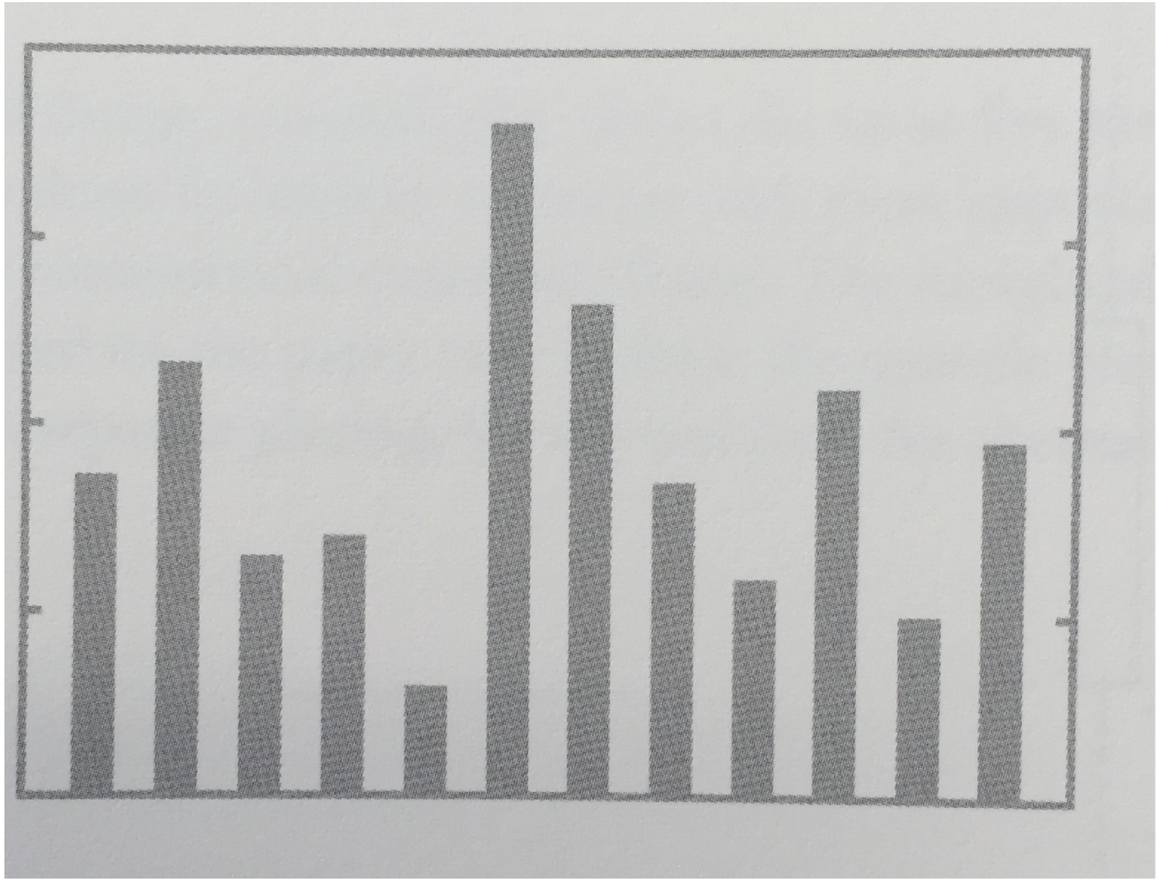


1.0

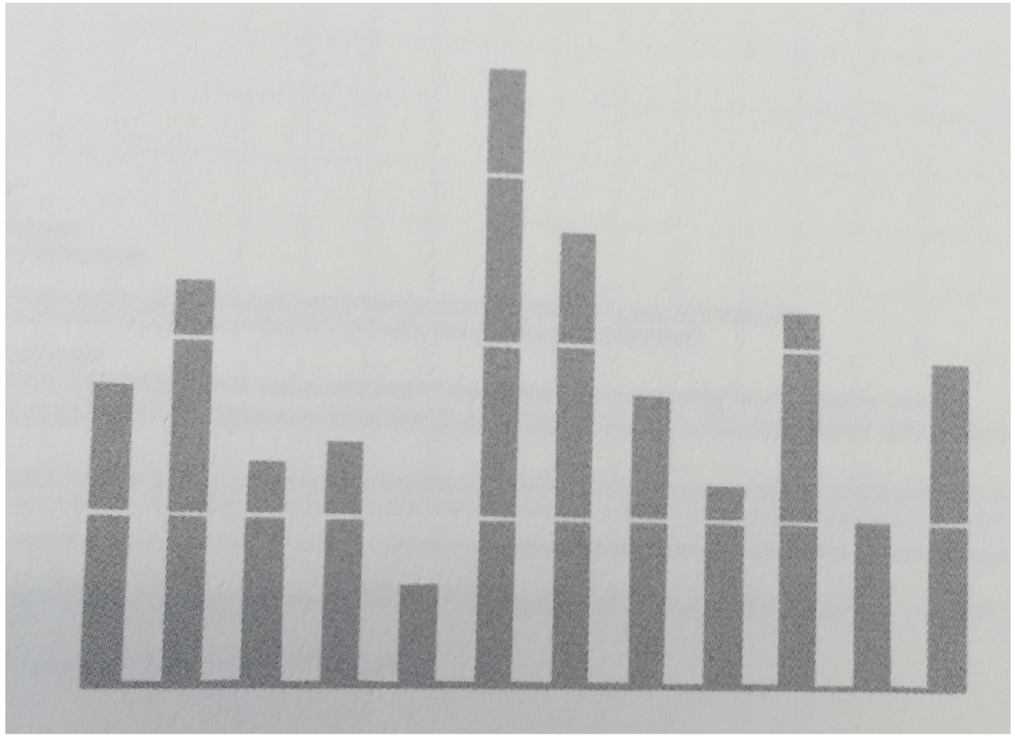
Design Principles for Data-ink

- (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

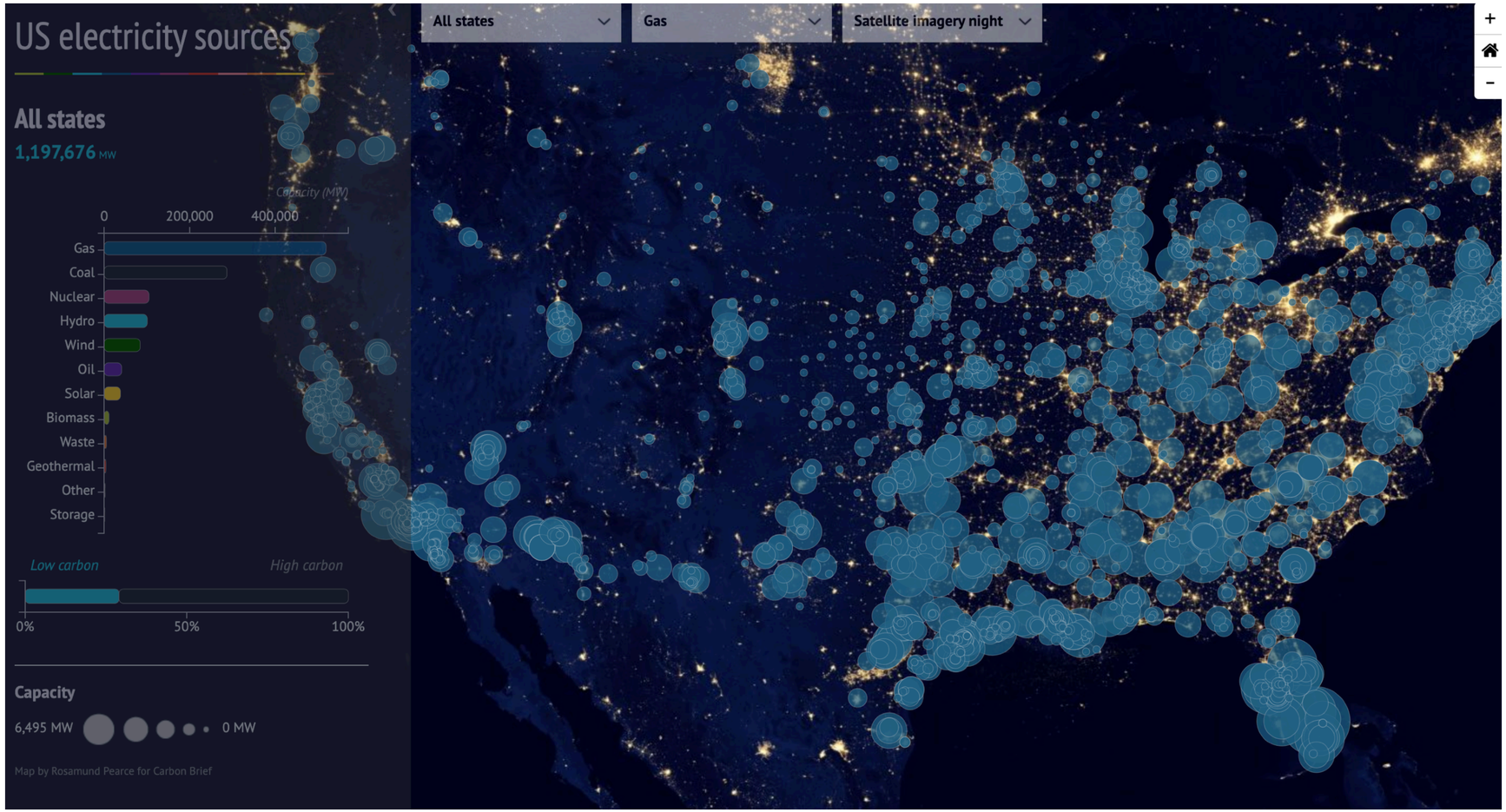
Interactive Visualizations

- 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

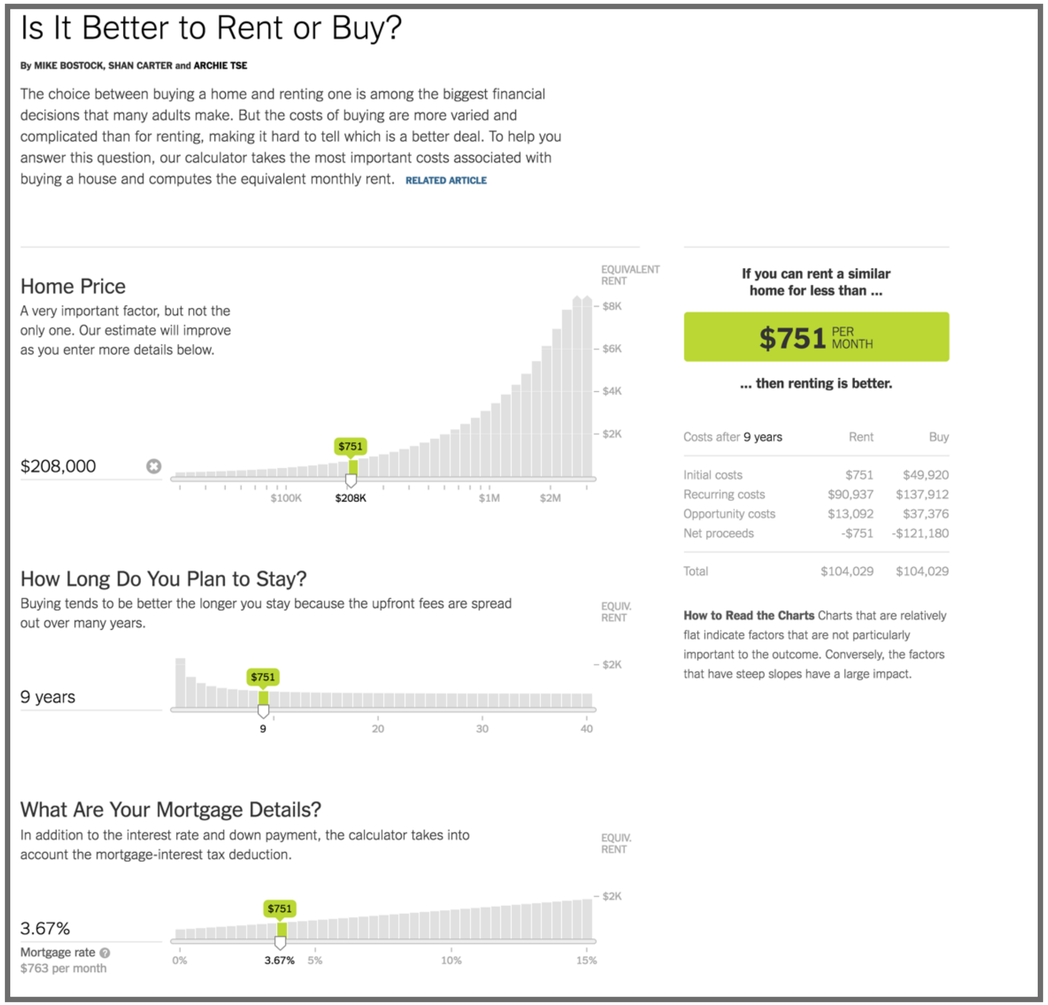
- ***Overview:*** gain an overview of entire collection
- ***Zoom:*** zoom in on items of interest
- ***Filter:*** filter out uninteresting items
- ***Details on Demand:*** select an item or group and get details
- ***Relate:*** view relationships between items
- ***History:*** support undo, replay,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

US Electricity Sources



https://www.carbonbrief.org/mapped-how-the-us-generates-electricity/

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
- Due by 6:25pm today