Visualization

Philipp Eichmann

Original slides by Emanuel Zgraggen
Many slides borrowed from Drucker, Agrawala, Heer, Stasko, etc.
References at the end.

How much data is produced every day?

2.5 exabytes

530,000,000 songs
150,000,000 iPhones
5,000,000 Laptops
250,000 Libraries
90 Years of HD video

[http://www.northeastern.edu/levelblog/2016/05/13/how-much-data-produced-every-day/]

How do we avoid being overwhelmed?

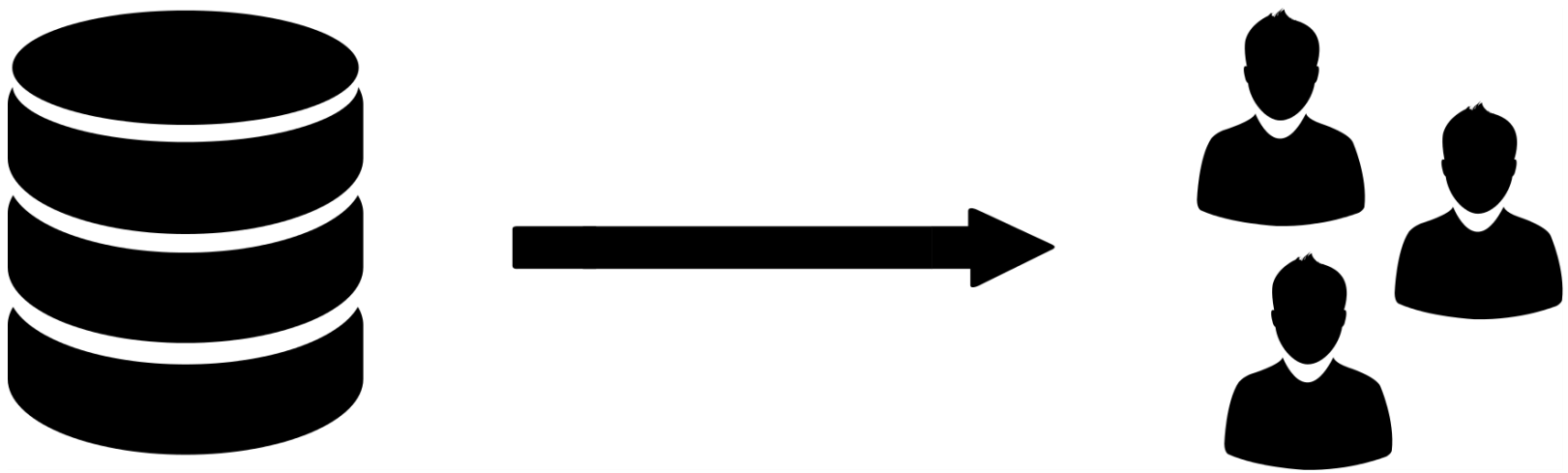
How do we make sense of the data?

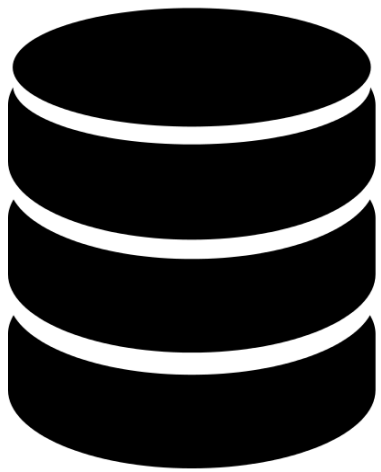
How do we use data in decision-making processes?

Challenge

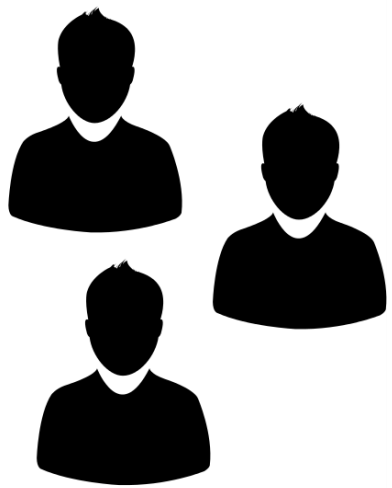
Transform the data into understanding and insight.

Make it useful to people.

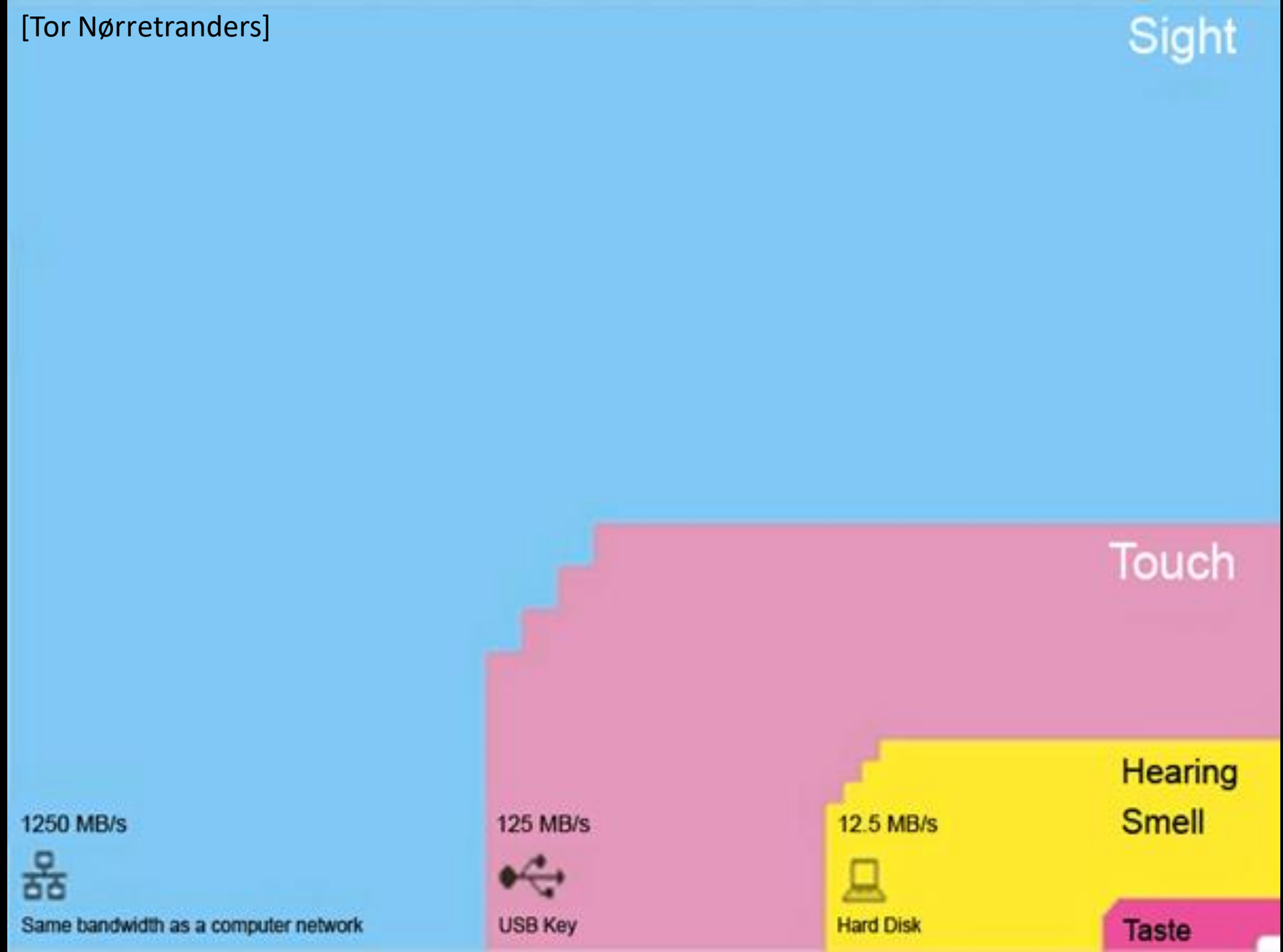


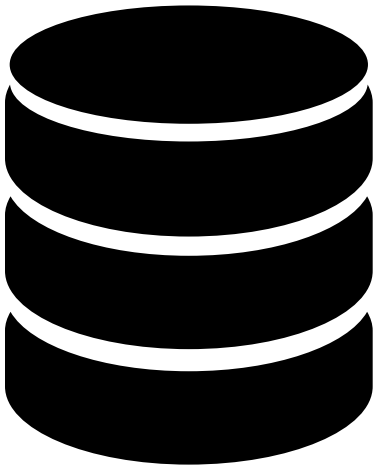


Sight
Hearing
Taste
Smell
Touch



[Tor Nørretranders]





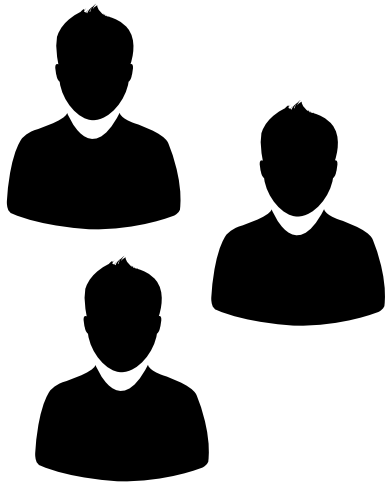
Sight

Hearing

Taste

Smell

Touch



Human Vision

Highest bandwidth sense

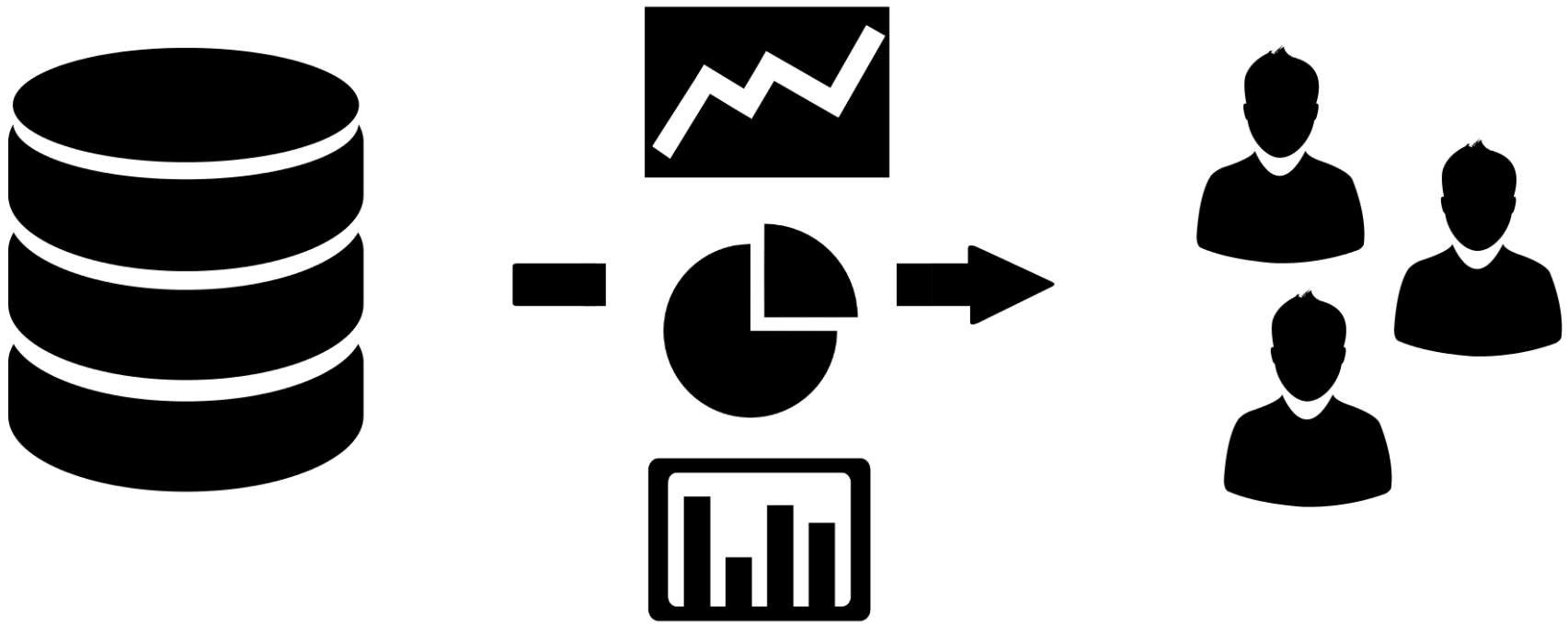
Fast, parallel

Pattern recognition

Pre-attentive

Extends memory and cognitive capacity

People think visually



Why create visualizations?

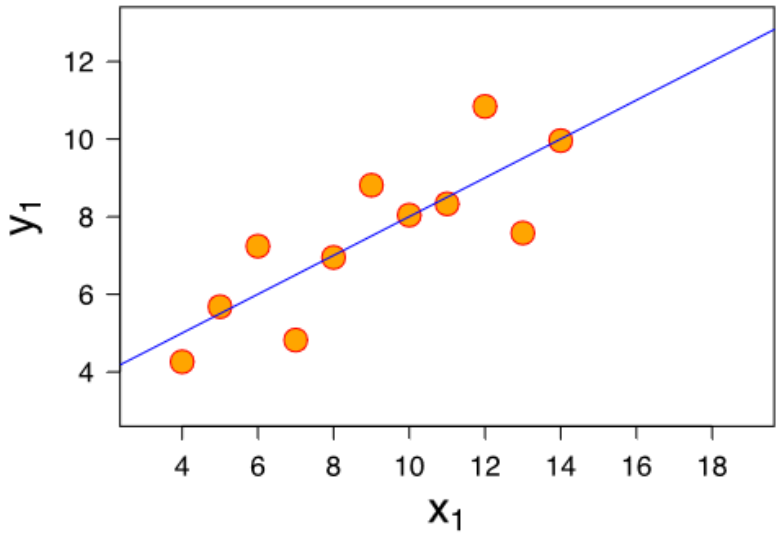
I		II		III		IV	
x	y	x	y	x	y	x	y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.84	12	9.13	12	8.15	8	5.56
7	4.82	7	7.26	7	6.42	8	7.91
5	5.68	5	4.74	5	5.73	8	6.89

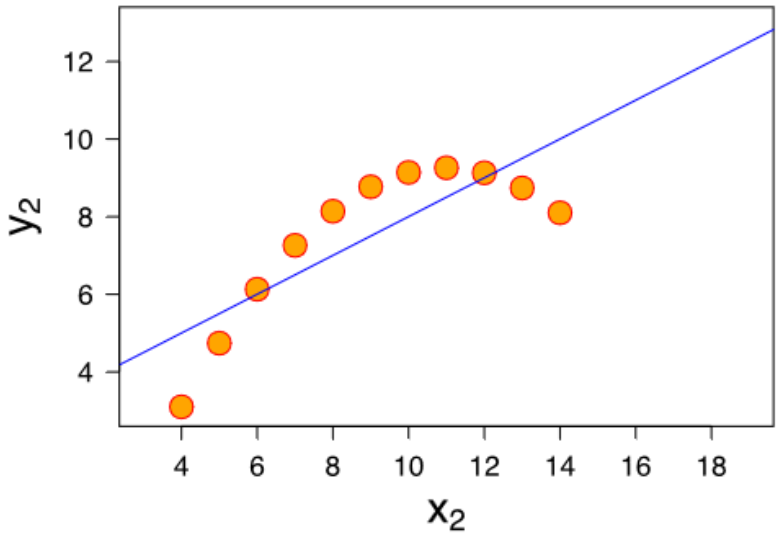
[Anscombe's Quartet]

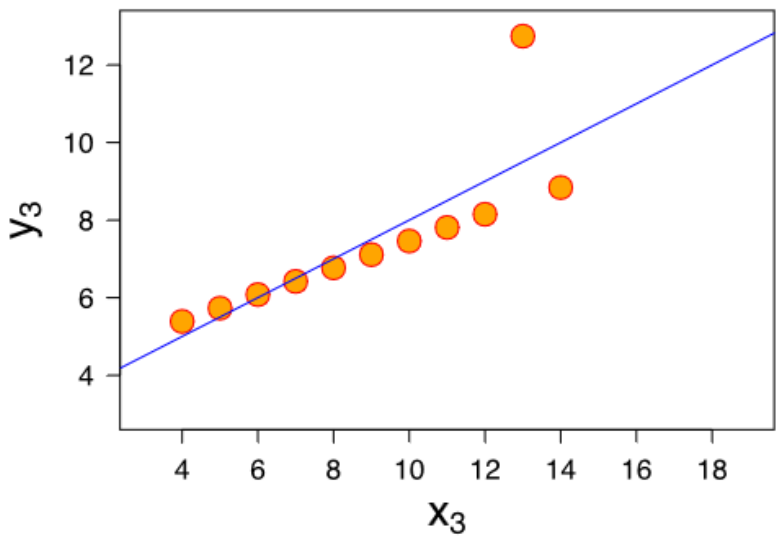
Mean of x in each case	9
Mean of y in each case	7.50
Sample **variance** of x in each case	11
Sample **variance** of y in each case	4.122 or 4.127
Correlation between x and y in each case	0.816
Linear regression line in each case	$y = 3.00 + 0.500x$

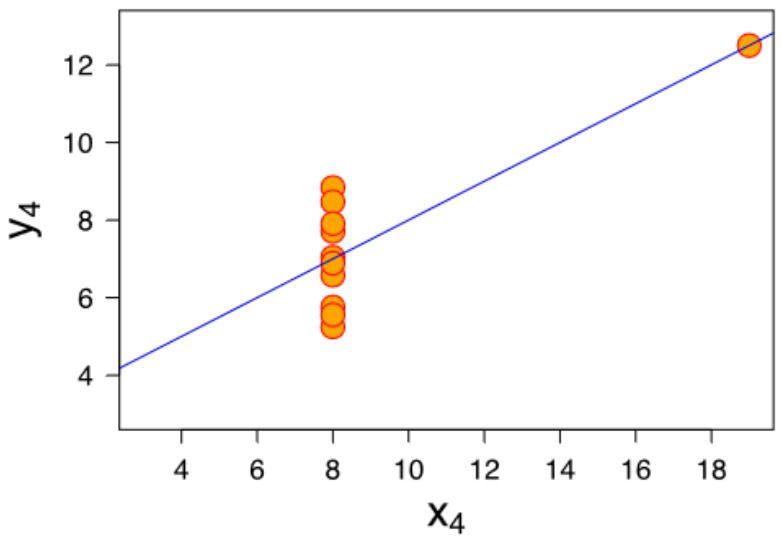
[Anscombe's Quartet]

Find patterns







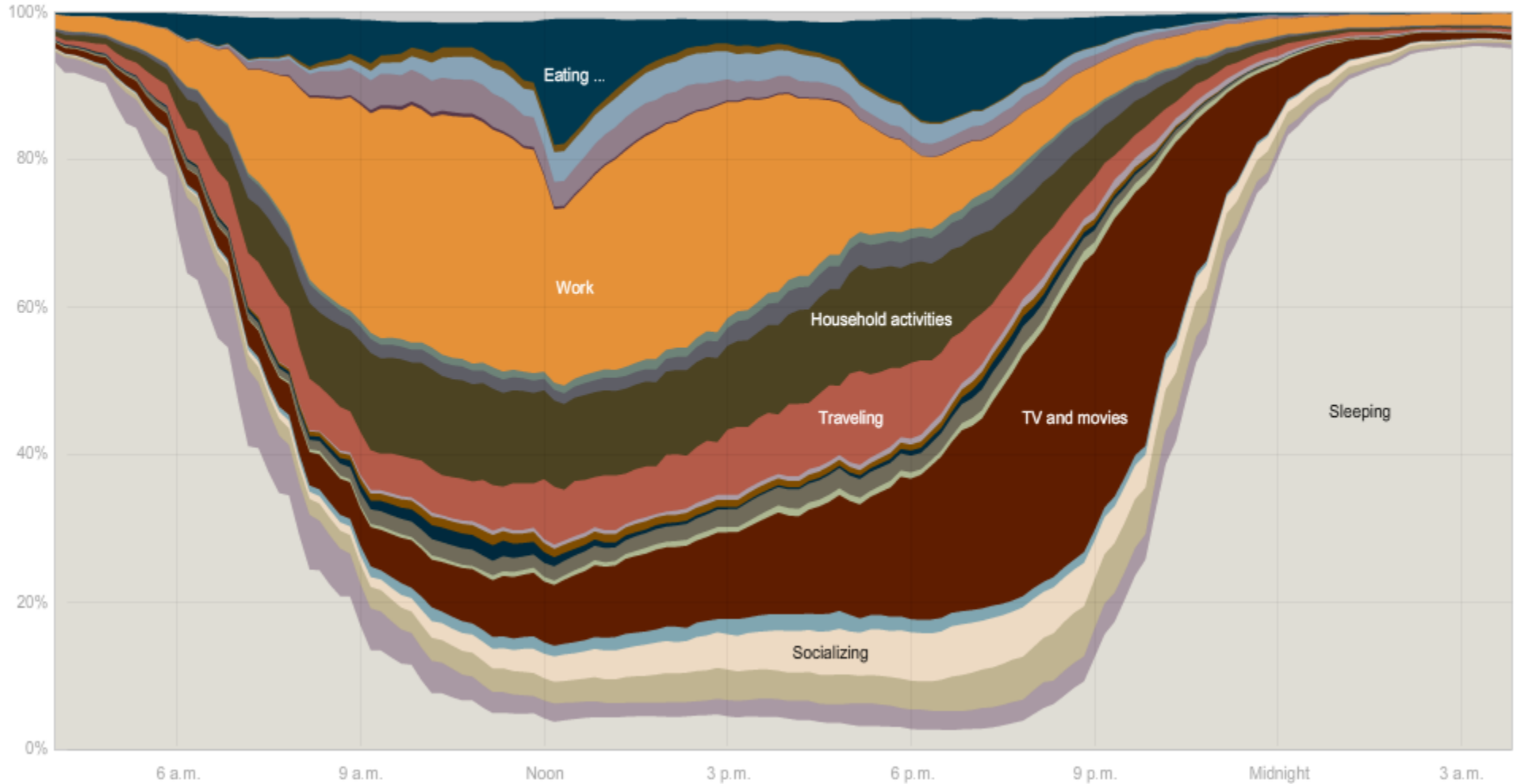


[Anscombe's Quartet]

Everyone

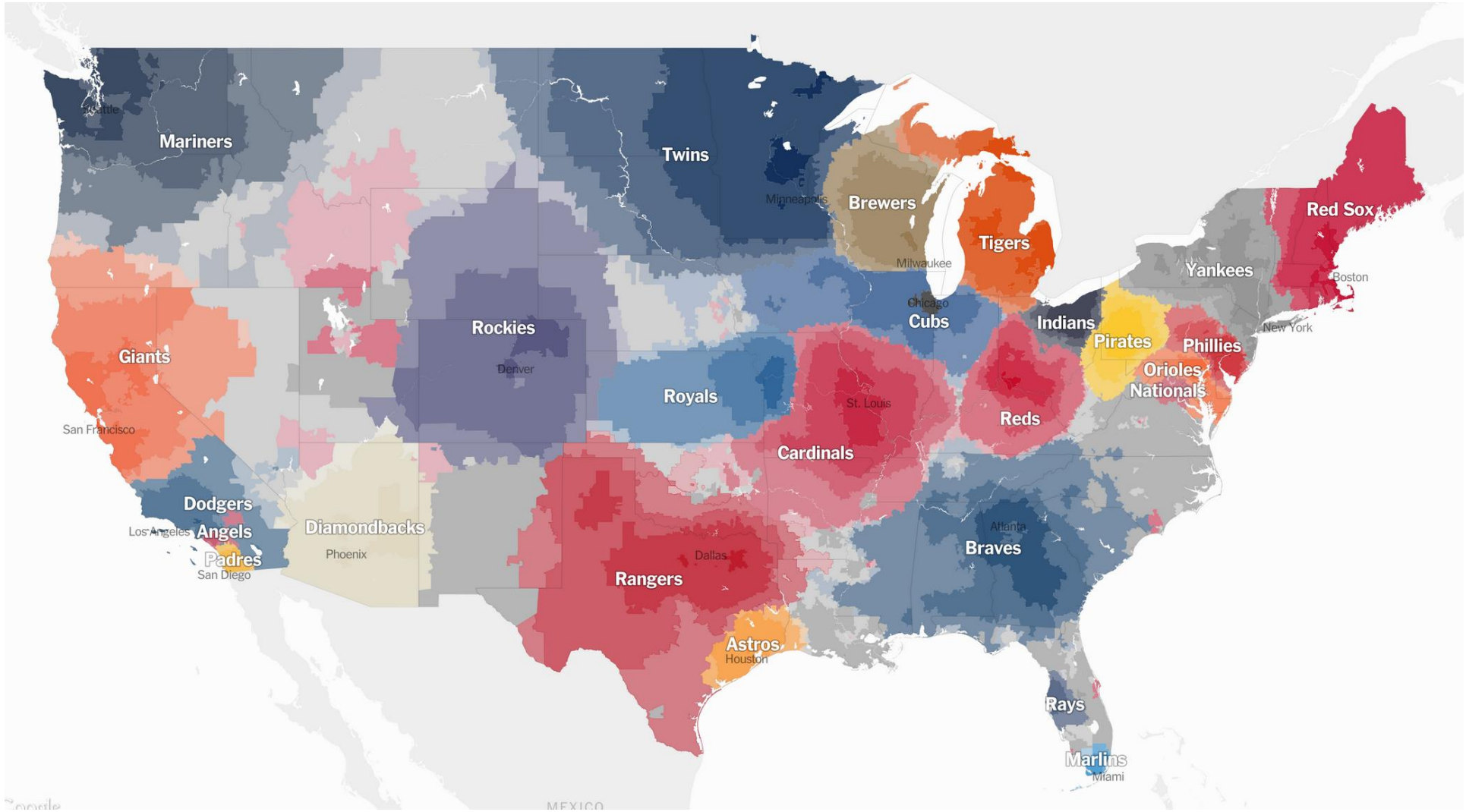
Sleeping, eating, working and watching television take up about two-thirds of the average day.

Everyone	Employed	White	Age 15-24	H.S. grads	No children
Men	Unemployed	Black	Age 25-64	Bachelor's	One child
Women	Not in lab...	Hispanic	Age 65+	Advanced	Two+ children



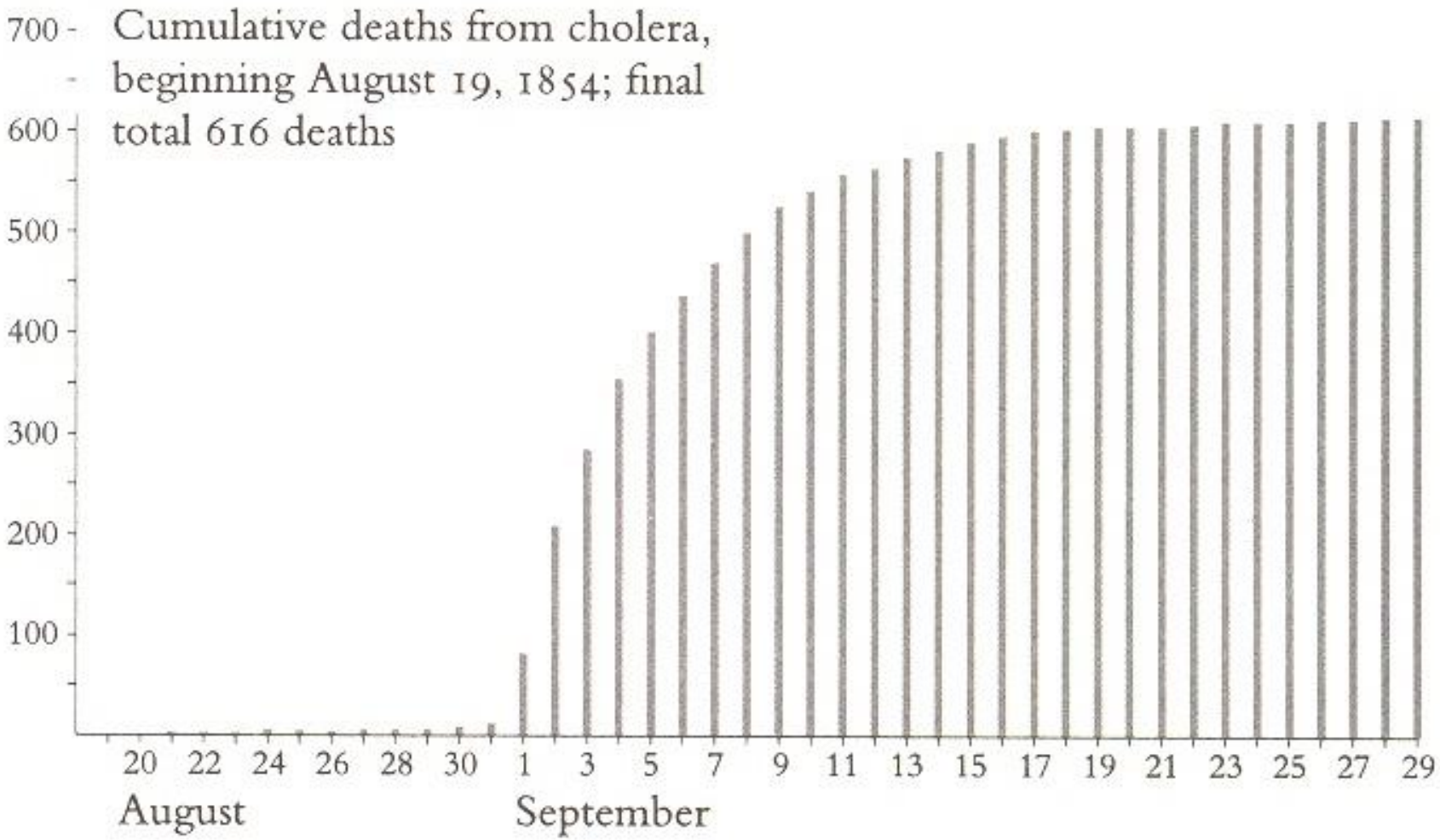
[New York Times]

See data in context



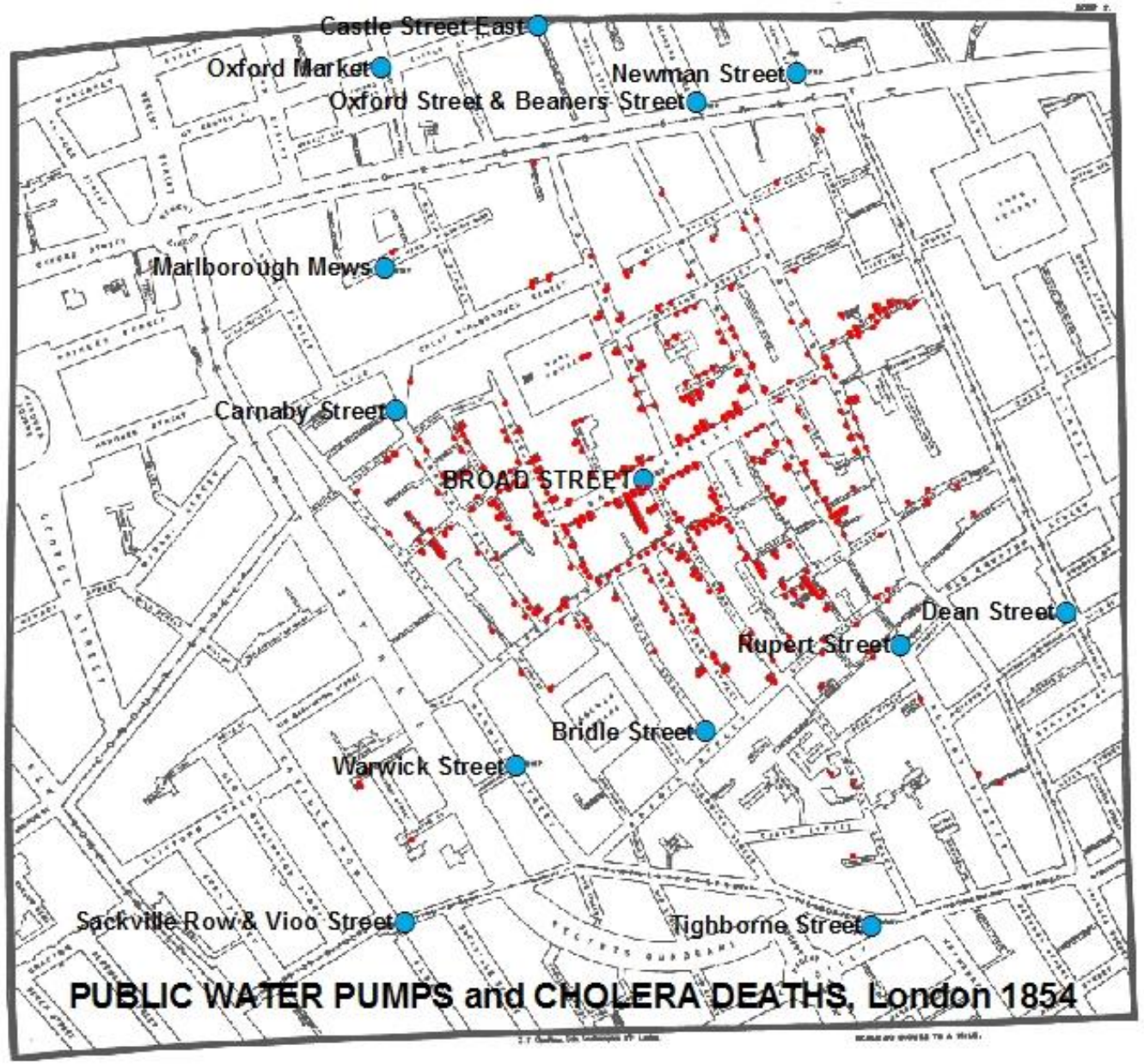
[New York Times]

See data in context



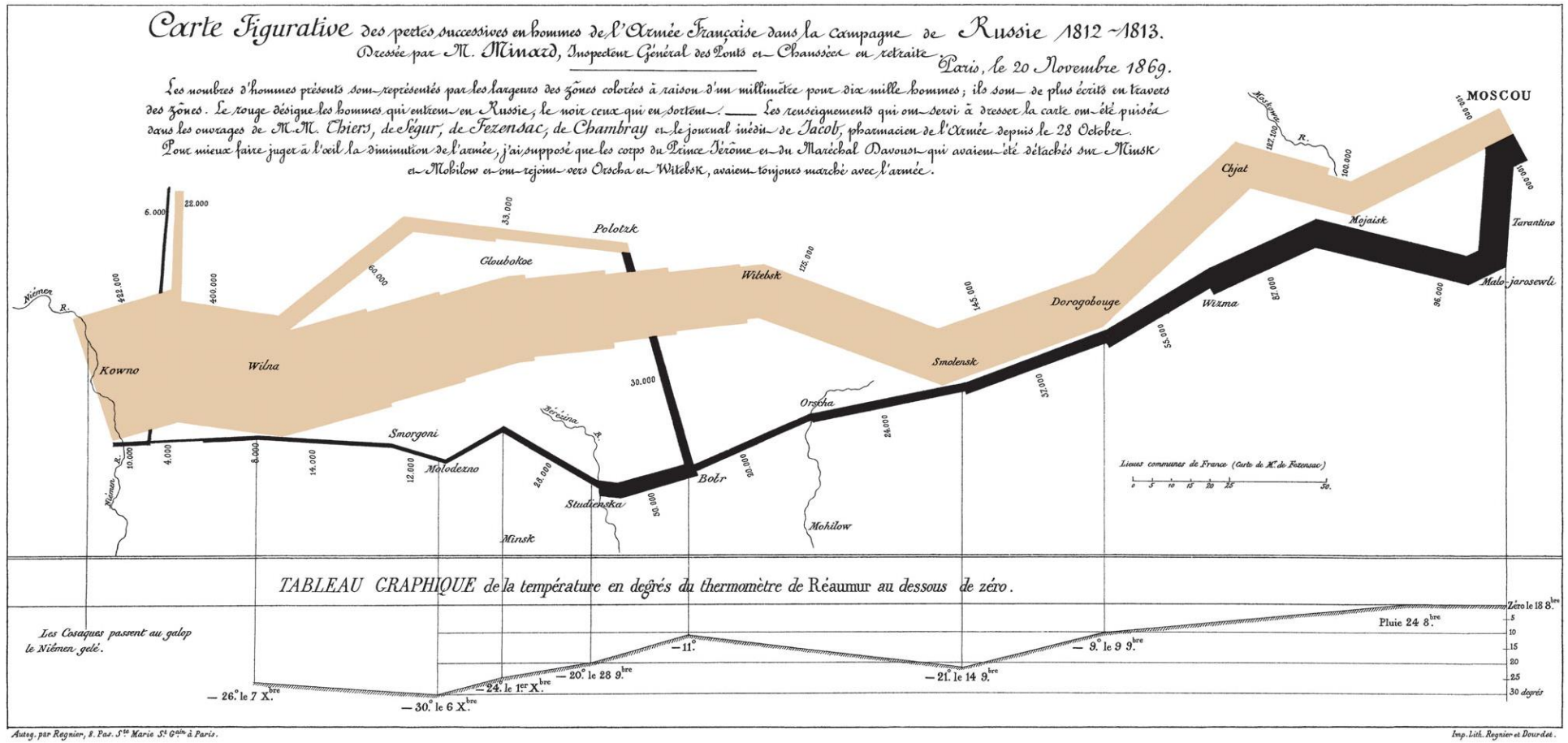
See data in context

See data in context



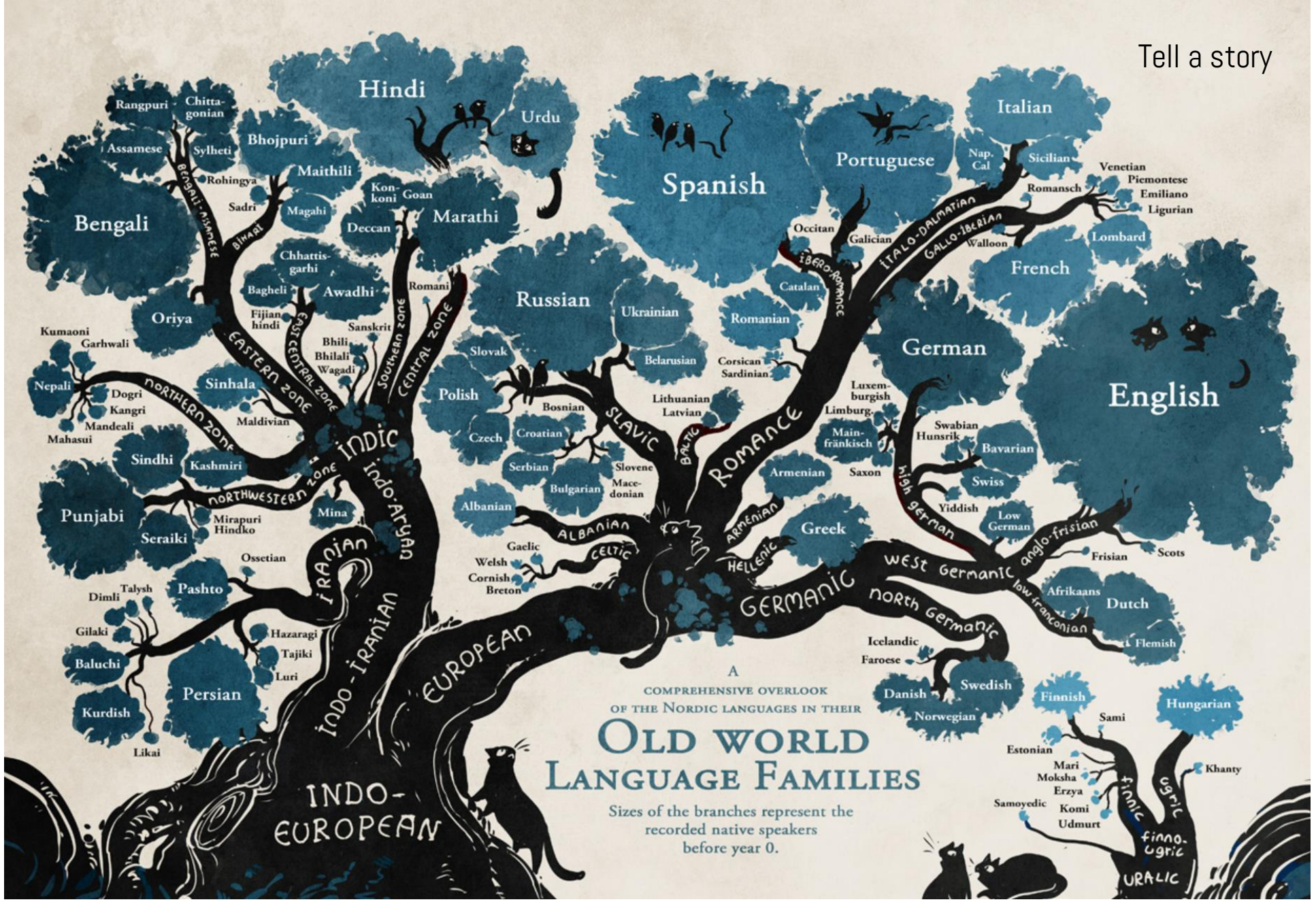
[http://www.udel.edu/johnmack/frec682/cholera/]

Tell a story



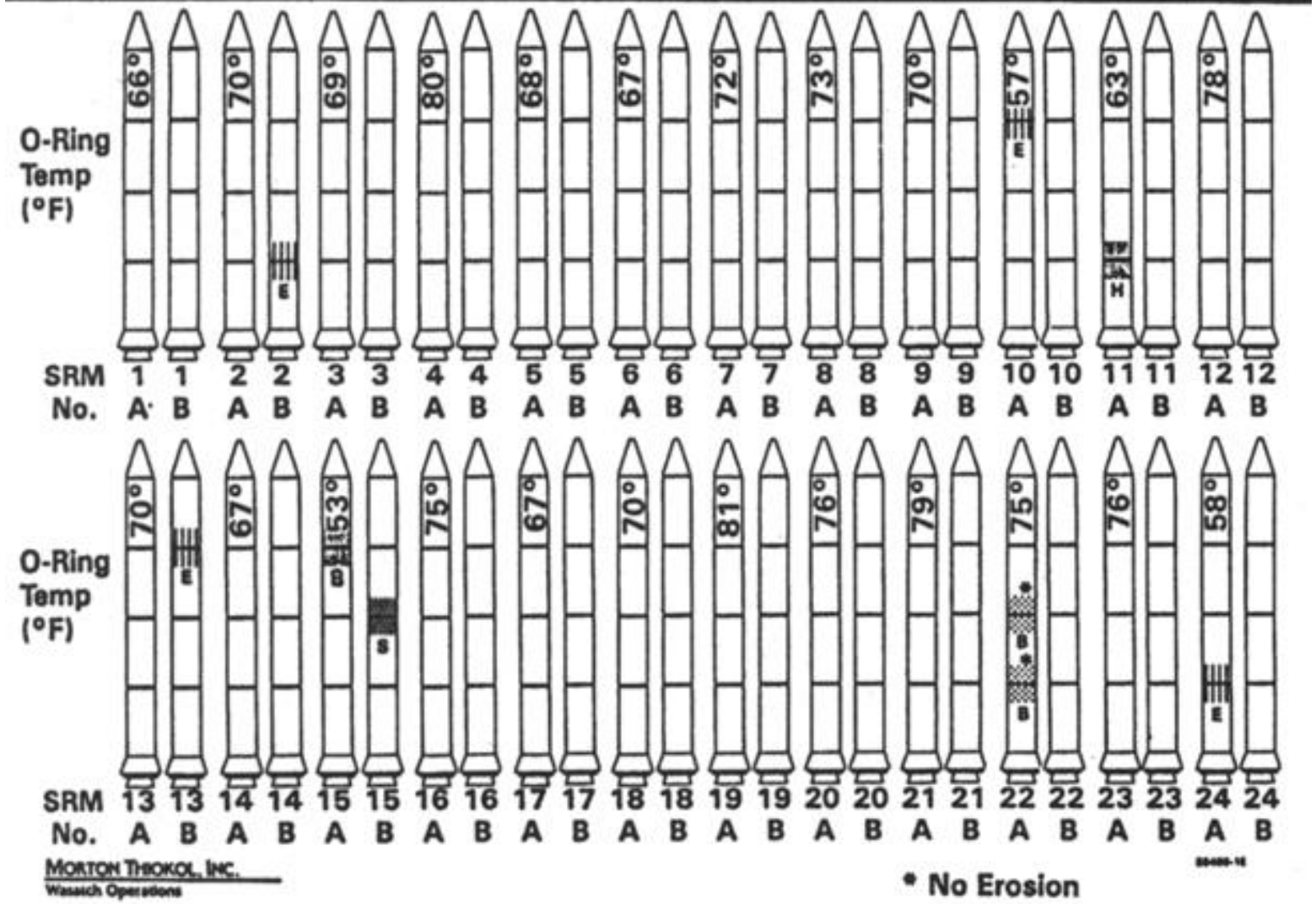
[Minard 1869, Tufte]

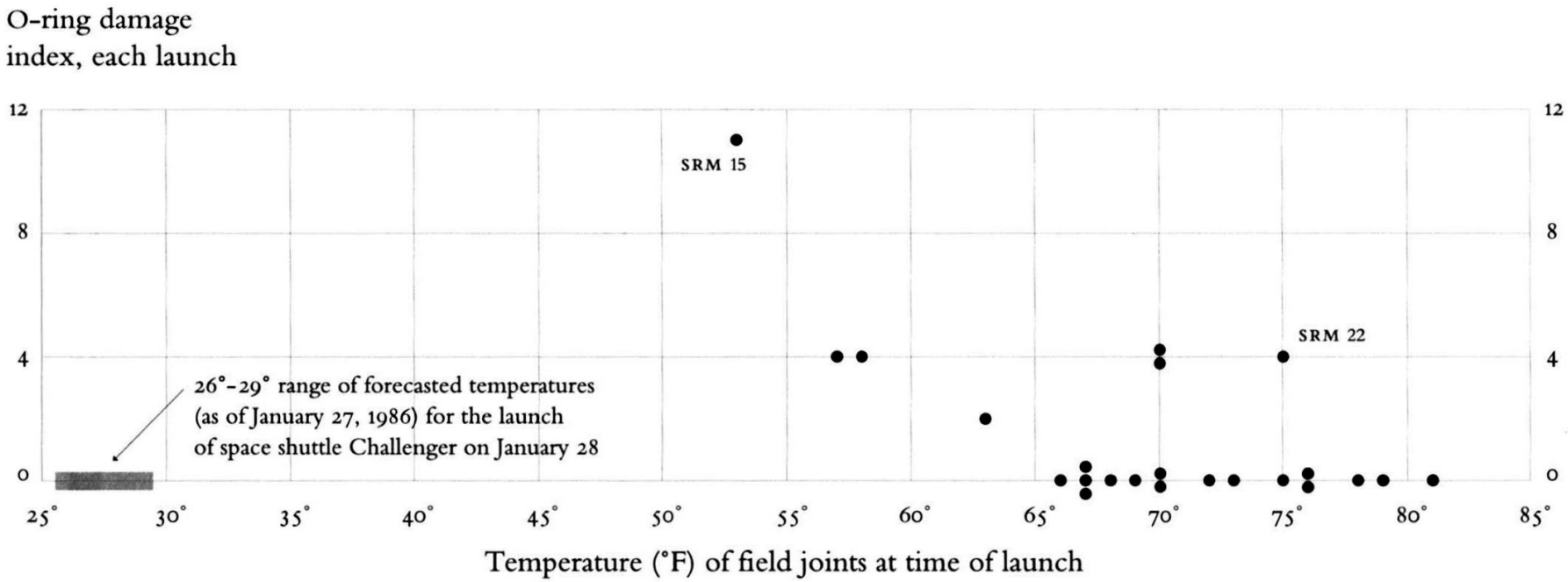
Tell a story

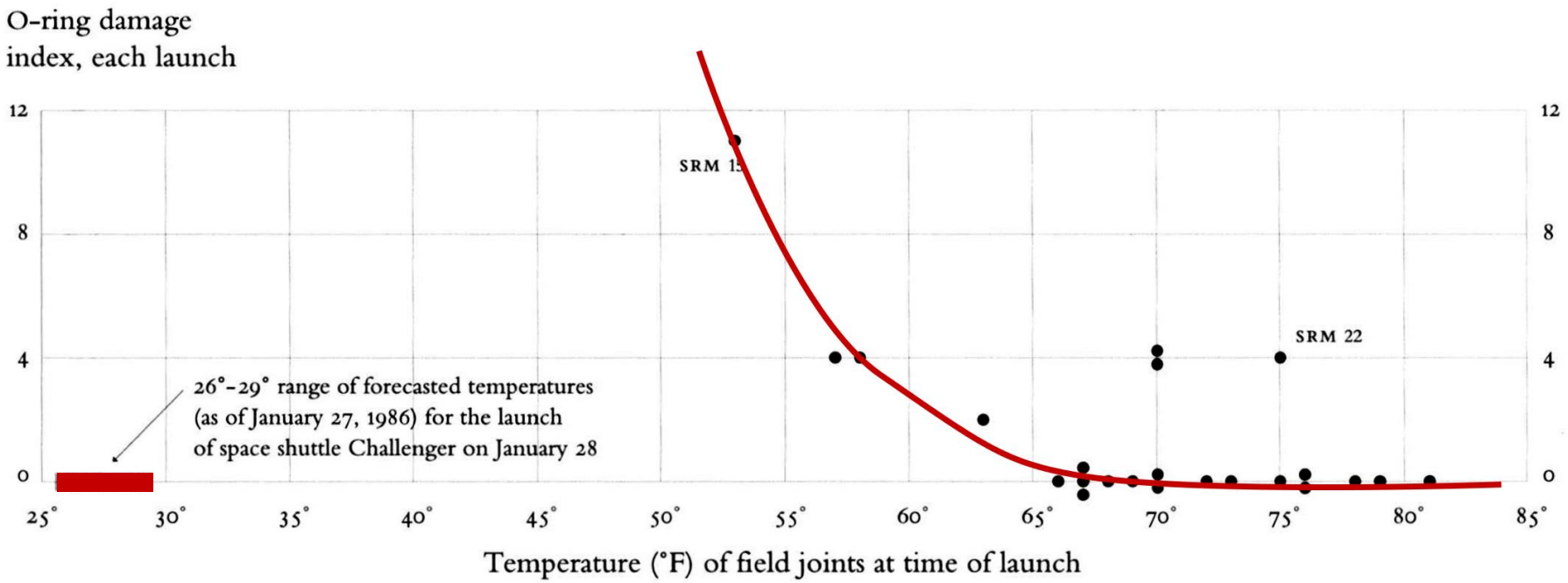


[Minna Sundberg]

History of O-Ring Damage in Field Joints (Cont)

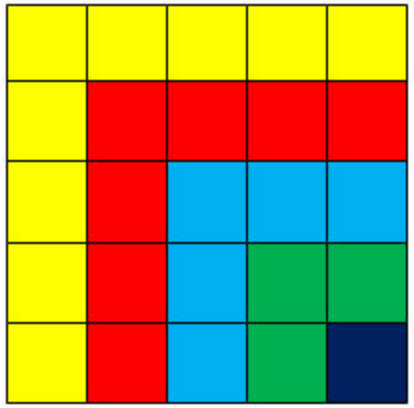



O-ring damage
index, each launch
12
8
4
0
SRM 15
SRM 22
26°–29° range of forecasted temperatures
(as of January 27, 1986) for the launch
of space shuttle Challenger on January 28
25°
30°
35°
40°
45°
50°
55°
60°
65°
70°
75°
80°
85°
Temperature (°F) of field joints at time of launch


O-ring damage
index, each launch
12
8
4
0
SRM 15
SRM 22
26°–29° range of forecasted temperatures
(as of January 27, 1986) for the launch
of space shuttle Challenger on January 28
25°
30°
35°
40°
45°
50°
55°
60°
65°
70°
75°
80°
85°
Temperature (°F) of field joints at time of launch

Sum of first **n** odd numbers:

$1 + 3 + 5 + 7 + 9 = 25$



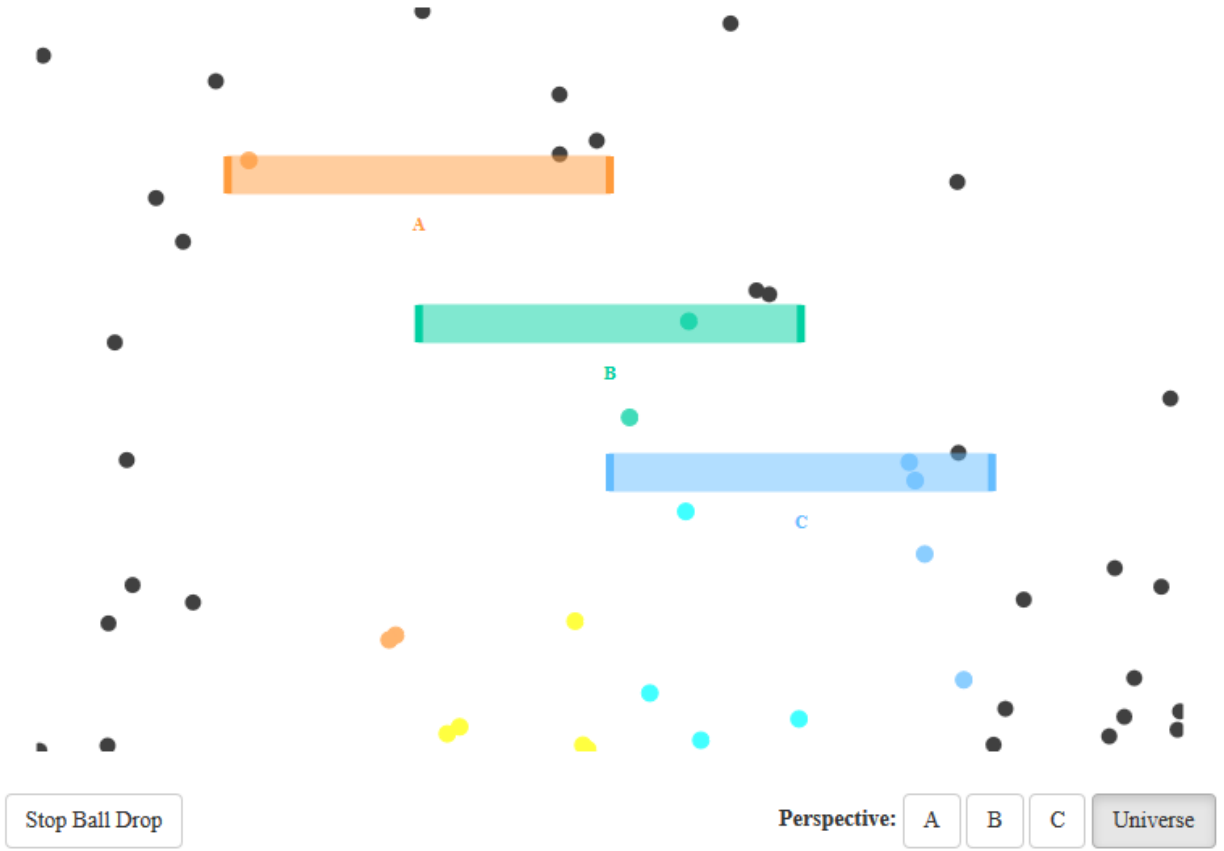
Conditional Probability

Conditional probability is a measure of the probability of an event given another event has occurred. It can be expressed mathematically as the following:

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

To visualize this concept move and rescale the shelfs by drag and drop and toggle the current perspective. What is the probability a ball lands on a shelf? What is the probability given the ball has or will hit another shelf?

This visualization was adapted from Victor Powell's fantastic visualization of conditional probability.

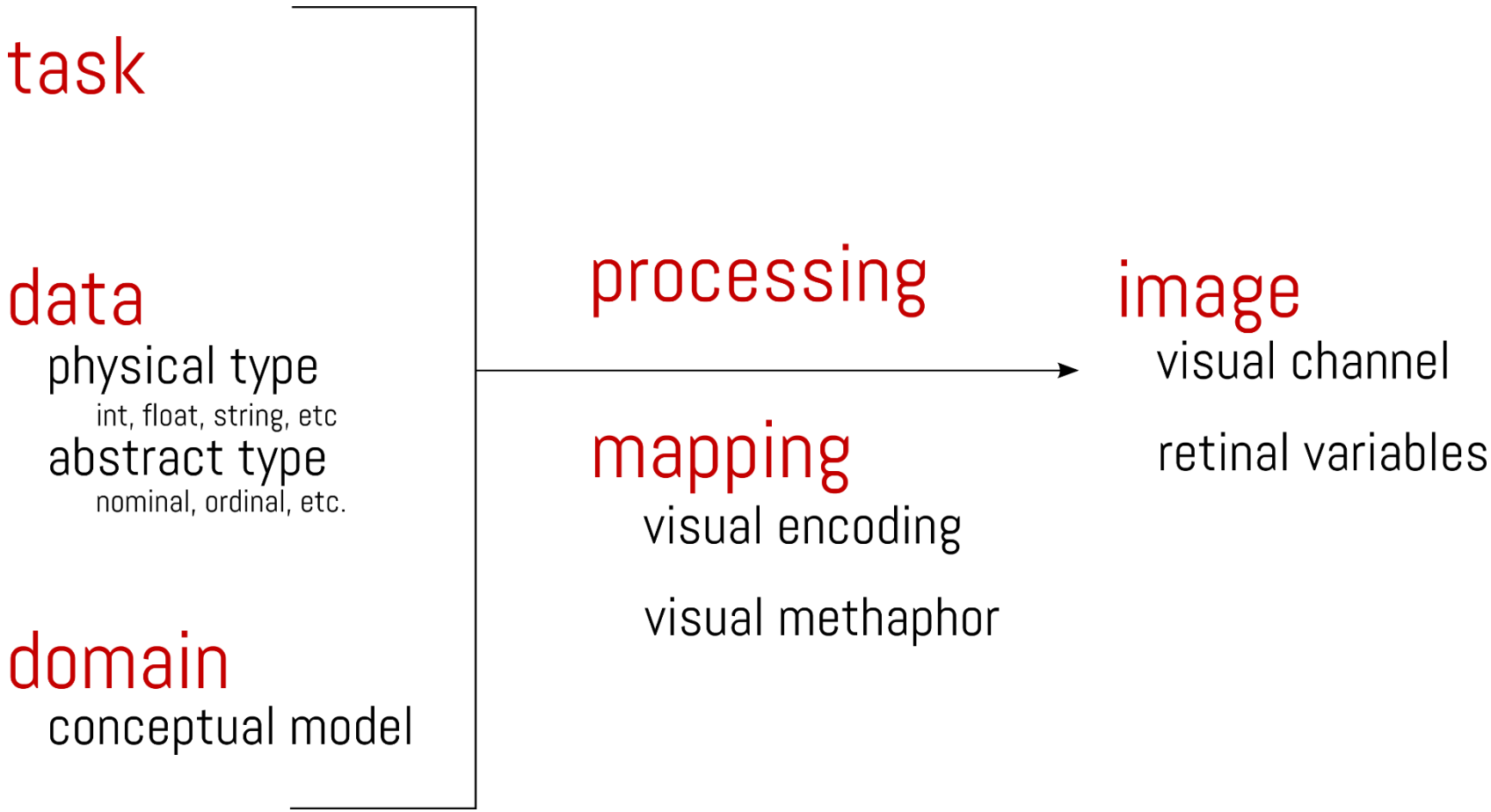


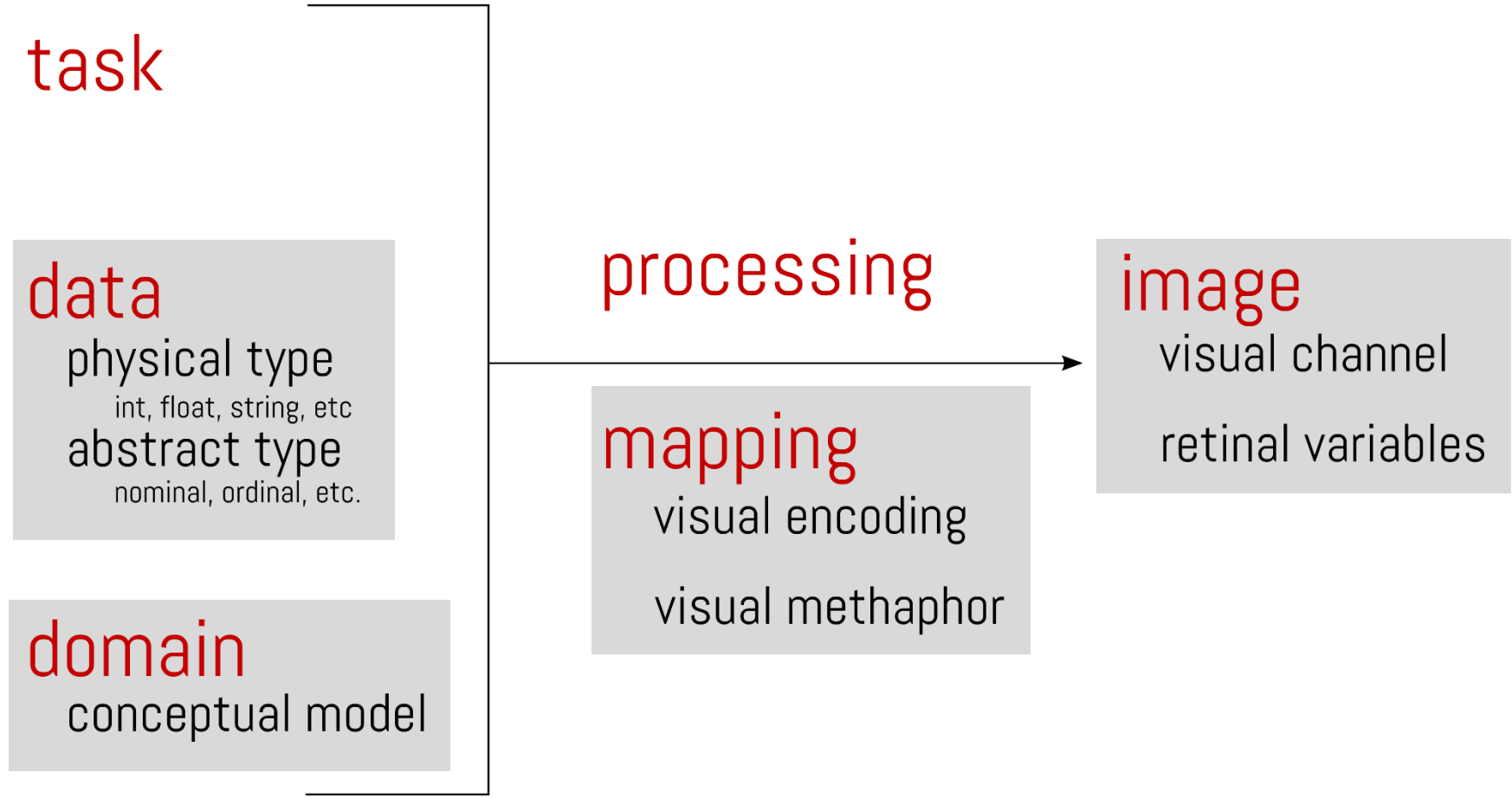
[http://students.brown.edu/seeing-theory/ Daniel Kunin]

Why create visualizations?

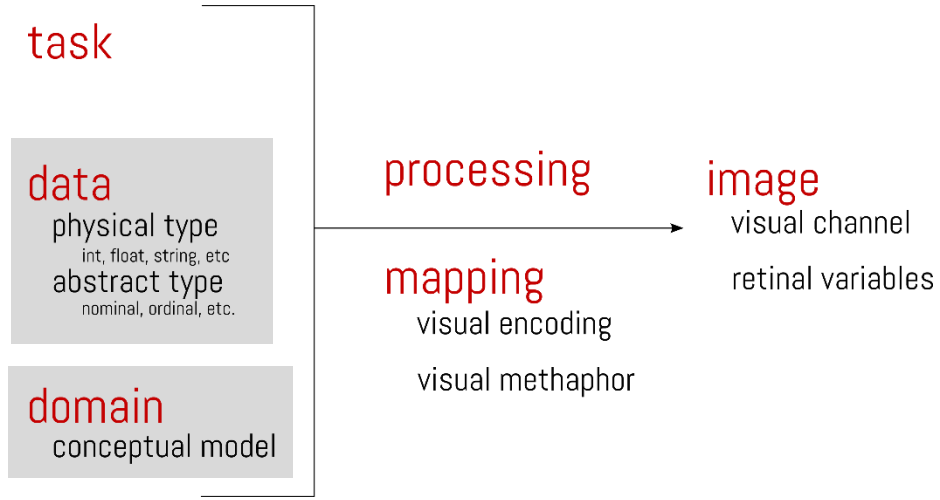
- Answer questions (or discover them)
- Make decisions
- See data in context
- Expand memory
- Support graphical calculation
- Find patterns
- Present argument or tell a story
- Teach

How do we create visualizations?


task
data
physical type
int, float, string, etc
abstract type
nominal, ordinal, etc.
domain
conceptual model
processing
mapping
visual encoding
visual methaphor
image
visual channel
retinal variables


task
data
physical type
int, float, string, etc
abstract type
nominal, ordinal, etc.
domain
conceptual model
processing
mapping
visual encoding
visual methaphor
image
visual channel
retinal variables

Data & Domain


task
data
physical type
int, float, string, etc
abstract type
nominal, ordinal, etc.
domain
conceptual model
processing
mapping
visual encoding
visual methaphor
image
visual channel
retinal variables

Data Model

- How the data is organized
- How are data elements related

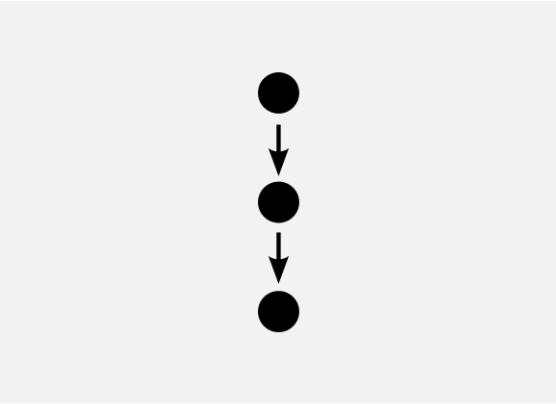
Conceptual Model

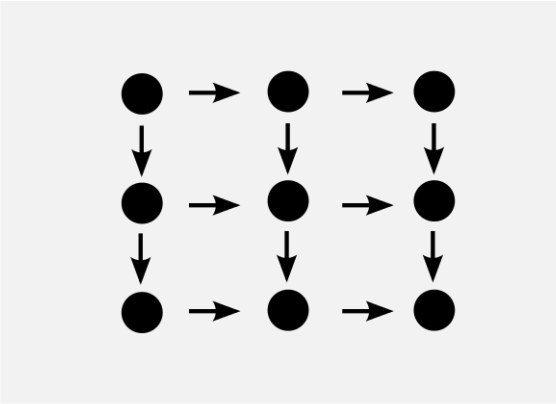
- Mental constructions
- Include semantics and support reasoning

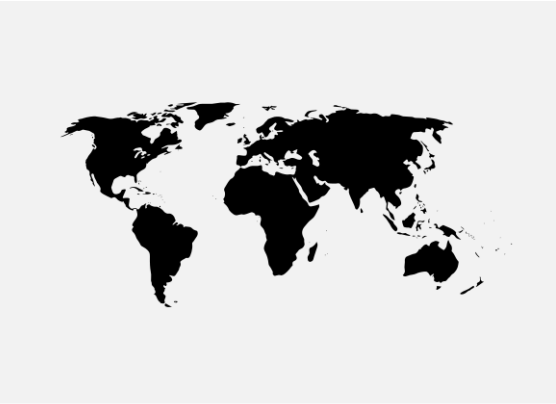
Data vs. Conceptual

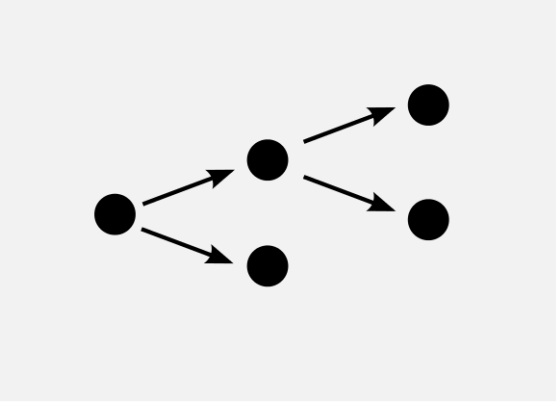
- 1D list of floats vs. Temperature
- 3D list of floats vs. Space

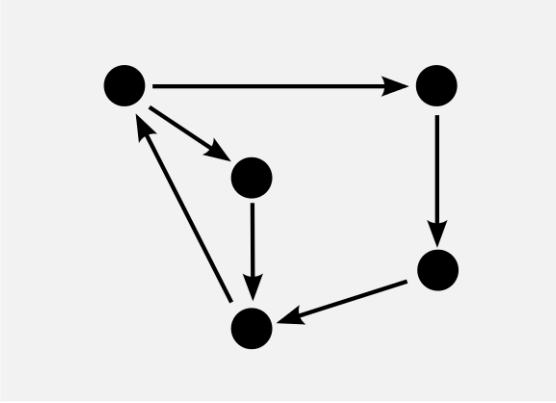
Data Model Taxonomy

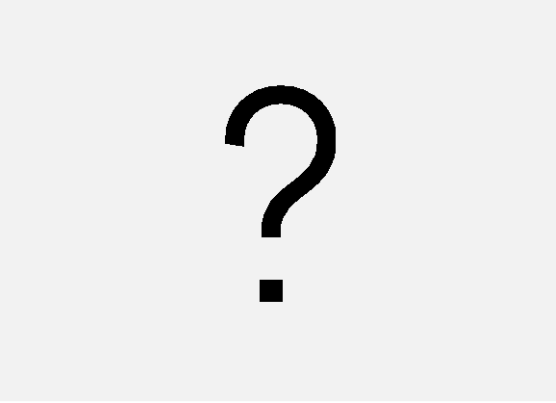












Variables

- **Physical types**
  - Characterized by storage format
  - Characterized by machine operations
  - **Example:**
    - bool, short, int32, float, double, string, …
- **Level of measurement**
  - Describes the relationship among values
    - Nominal
    - Ordinal
    - Quantitative

Nominal, Ordinal and Quantitative

- N – Nominal (labels):
  - Fruits: Apples, Oranges, …
- O – Ordinal
  - Quality of meat: Grade A, AA, AAA
- Q – Ratio (zero fixed)
  - E.g.: Length, Mass, Money, Temperature in Kelvin
  - Counts and amounts
- Q – Interval (Location of zero is arbitrary)
  - E.g.: Date/Time, geographic coordinates
  - Only differences (i.e. Intervals) can be compared

[S. S. Stevens, On the theory of scales of measurements, 1946]

Nominal, Ordinal and Quantitative

- N - Nominal (labels):
  - Operations: =, ≠
- O - Ordinal
  - Operations: =, ≠, <, >, ≤, ≥
- Q - Ratio (zero fixed)
  - Can measure ratios or proportions
  - Operations: =, ≠, <, >, ≤, ≥, -, ÷
- Q - Interval (Location of zero arbitrary)
  - Can measure distances or spans
  - Operations: =, ≠, <, >, ≤, ≥, -

[S. S. Stevens, On the theory of scales of measurements, 1946]

Example

- Data Model
  - 32.50, 54.0, 17.30, …
  - 1D, floats
- Conceptual Model
  - Temperature
- N,O,Q Type
  - Burned vs. Not burned (N)
  - Hot, warm, cold (O)
  - Continuous range of values (Q)

Microsoft Excel - fischer.iris.2.xls

File Edit View Insert Format Tools Data Window Help

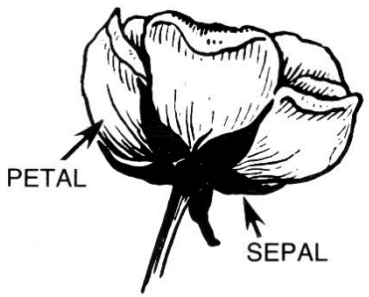
Type a question for help

A1 ID

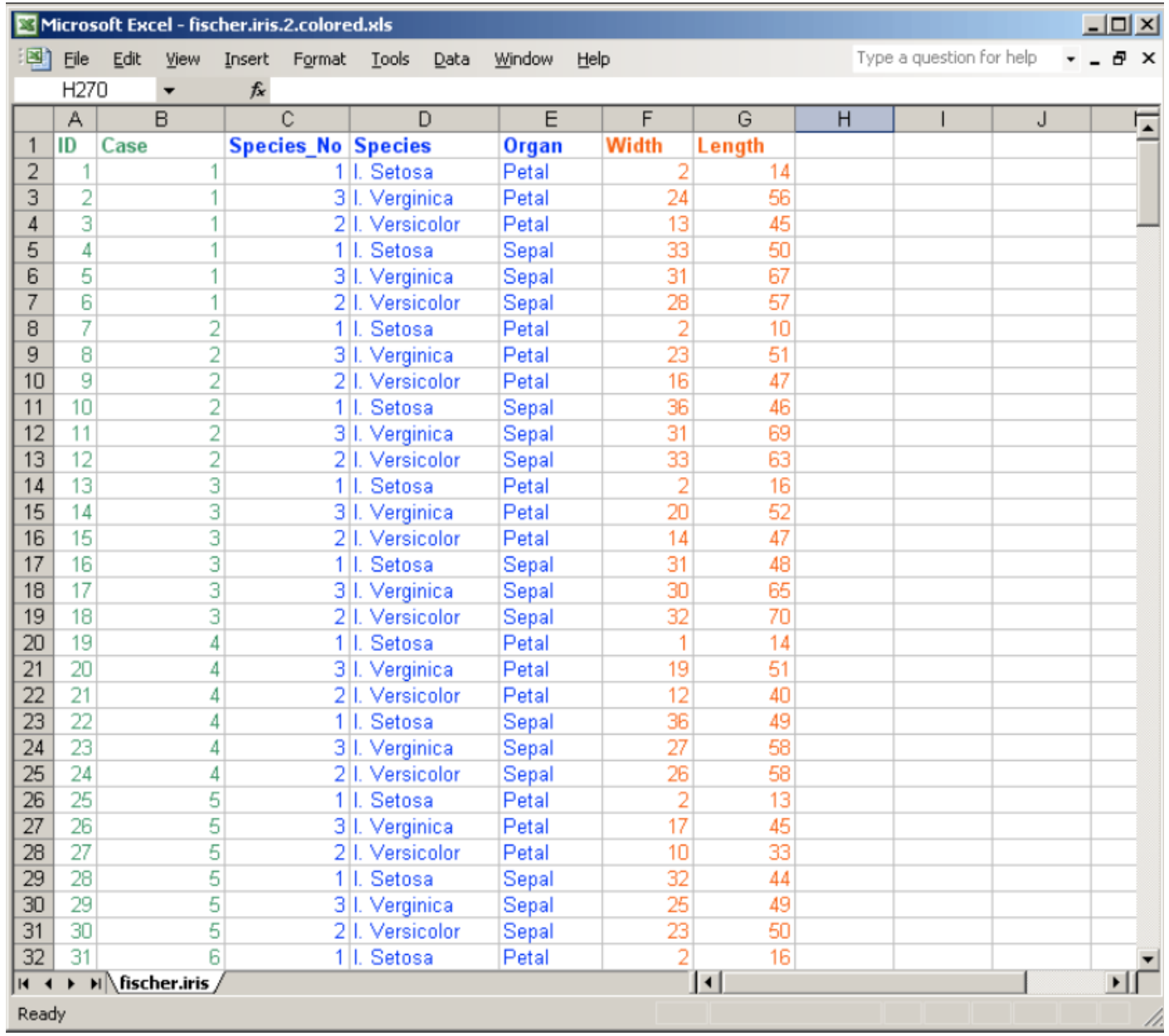
	A	B	C	D	E	F	G	H	I	J
1	**ID**	**Case**	**Species_No**	**Species**	**Organ**	**Width**	**Length**			
2	1	1	1	I. Setosa	Petal	2	14			
3	2	1	3	I. Verginica	Petal	24	56			
4	3	1	2	I. Versicolor	Petal	13	45			
5	4	1	1	I. Setosa	Sepal	33	50			
6	5	1	3	I. Verginica	Sepal	31	67			
7	6	1	2	I. Versicolor	Sepal	28	57			
8	7	2	1	I. Setosa	Petal	2	10			
9	8	2	3	I. Verginica	Petal	23	51			
10	9	2	2	I. Versicolor	Petal	16	47			
11	10	2	1	I. Setosa	Sepal	36	46			
12	11	2	3	I. Verginica	Sepal	31	69			
13	12	2	2	I. Versicolor	Sepal	33	63			
14	13	3	1	I. Setosa	Petal	2	16			
15	14	3	3	I. Verginica	Petal	20	52			
16	15	3	2	I. Versicolor	Petal	14	47			
17	16	3	1	I. Setosa	Sepal	31	48			
18	17	3	3	I. Verginica	Sepal	30	65			
19	18	3	2	I. Versicolor	Sepal	32	70			
20	19	4	1	I. Setosa	Petal	1	14			
21	20	4	3	I. Verginica	Petal	19	51			
22	21	4	2	I. Versicolor	Petal	12	40			
23	22	4	1	I. Setosa	Sepal	36	49			
24	23	4	3	I. Verginica	Sepal	27	58			
25	24	4	2	I. Versicolor	Sepal	26	58			
26	25	5	1	I. Setosa	Petal	2	13			
27	26	5	3	I. Verginica	Petal	17	45			
28	27	5	2	I. Versicolor	Petal	10	33			
29	28	5	1	I. Setosa	Sepal	32	44			
30	29	5	3	I. Verginica	Sepal	25	49			
31	30	5	2	I. Versicolor	Sepal	23	50			
32	31	6	1	I. Setosa	Petal	2	16			

fischer.iris

Ready



[Fisher, 1936] http://en.wikipedia.org/wiki/Iris_flower_data_set



Microsoft Excel - fischer.iris.2.colored.xls

File Edit View Insert Format Tools Data Window Help

Type a question for help

H270

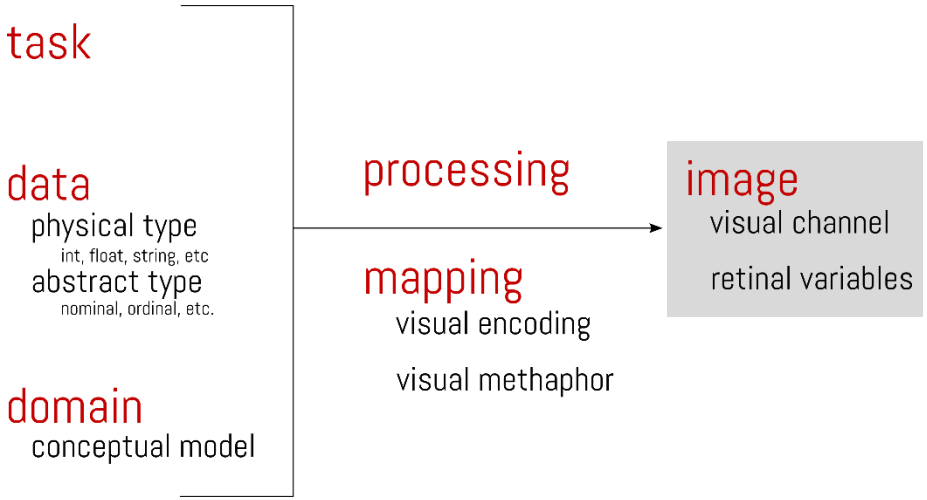
	A	B	C	D	E	F	G	H	I	J
1	ID	Case	Species_No	Species	Organ	Width	Length			
2	1	1	1	I. Setosa	Petal	2	14			
3	2	1	3	I. Verginica	Petal	24	56			
4	3	1	2	I. Versicolor	Petal	13	45			
5	4	1	1	I. Setosa	Sepal	33	50			
6	5	1	3	I. Verginica	Sepal	31	67			
7	6	1	2	I. Versicolor	Sepal	28	57			
8	7	2	1	I. Setosa	Petal	2	10			
9	8	2	3	I. Verginica	Petal	23	51			
10	9	2	2	I. Versicolor	Petal	16	47			
11	10	2	1	I. Setosa	Sepal	36	46			
12	11	2	3	I. Verginica	Sepal	31	69			
13	12	2	2	I. Versicolor	Sepal	33	63			
14	13	3	1	I. Setosa	Petal	2	16			
15	14	3	3	I. Verginica	Petal	20	52			
16	15	3	2	I. Versicolor	Petal	14	47			
17	16	3	1	I. Setosa	Sepal	31	48			
18	17	3	3	I. Verginica	Sepal	30	65			
19	18	3	2	I. Versicolor	Sepal	32	70			
20	19	4	1	I. Setosa	Petal	1	14			
21	20	4	3	I. Verginica	Petal	19	51			
22	21	4	2	I. Versicolor	Petal	12	40			
23	22	4	1	I. Setosa	Sepal	36	49			
24	23	4	3	I. Verginica	Sepal	27	58			
25	24	4	2	I. Versicolor	Sepal	26	58			
26	25	5	1	I. Setosa	Petal	2	13			
27	26	5	3	I. Verginica	Petal	17	45			
28	27	5	2	I. Versicolor	Petal	10	33			
29	28	5	1	I. Setosa	Sepal	32	44			
30	29	5	3	I. Verginica	Sepal	25	49			
31	30	5	2	I. Versicolor	Sepal	23	50			
32	31	6	1	I. Setosa	Petal	2	16			

fischer.iris

Ready

[Fisher, 1936] http://en.wikipedia.org/wiki/Iris_flower_data_set

Image


task
data
physical type
int, float, string, etc
abstract type
nominal, ordinal, etc.
domain
conceptual model
processing
mapping
visual encoding
visual methaphor
image
visual channel
retinal variables

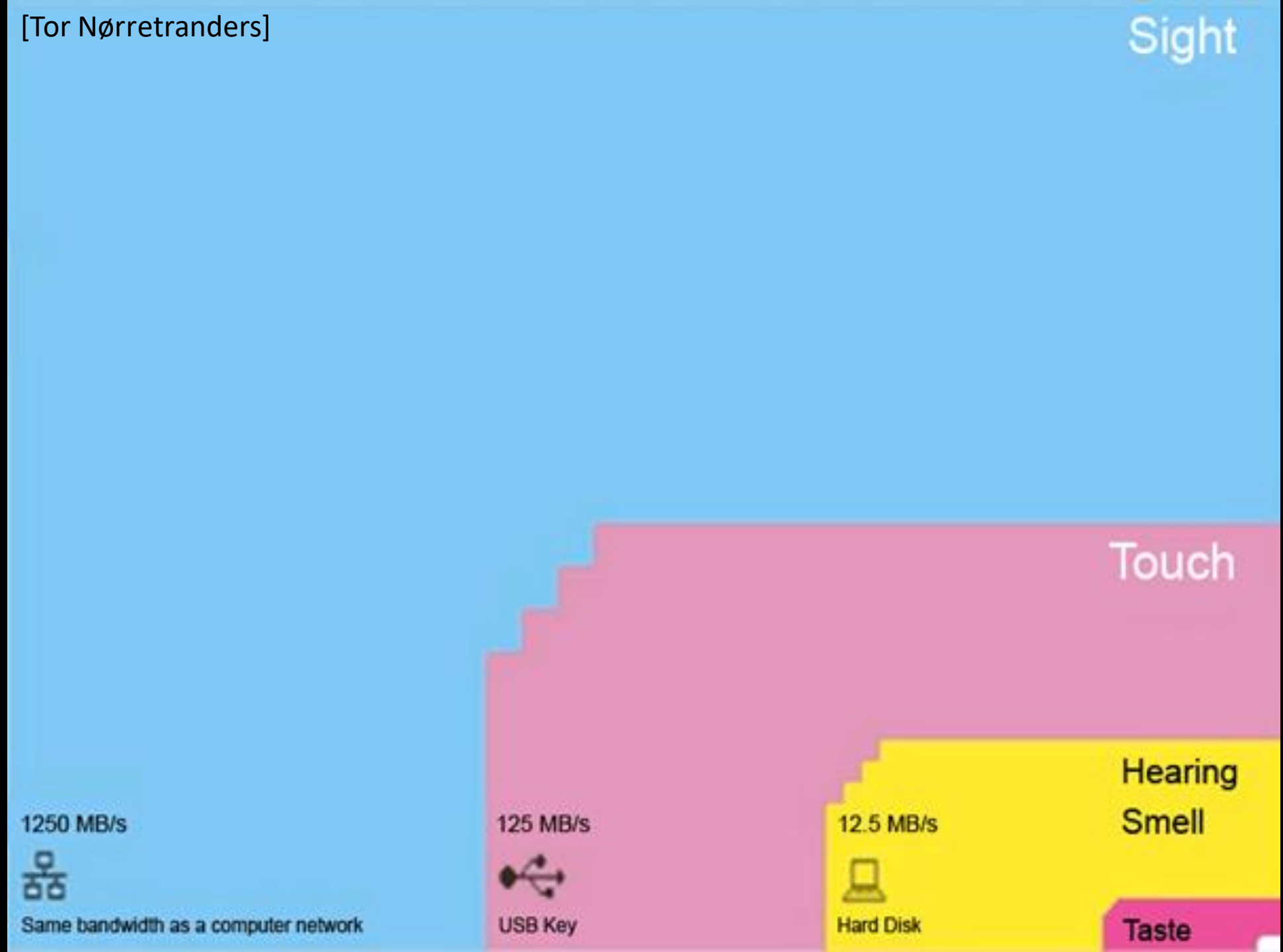
Pre-attentive

unconscious, parallel, fast

Attentive

conscious, serial, slow

[Tor Nørretranders]

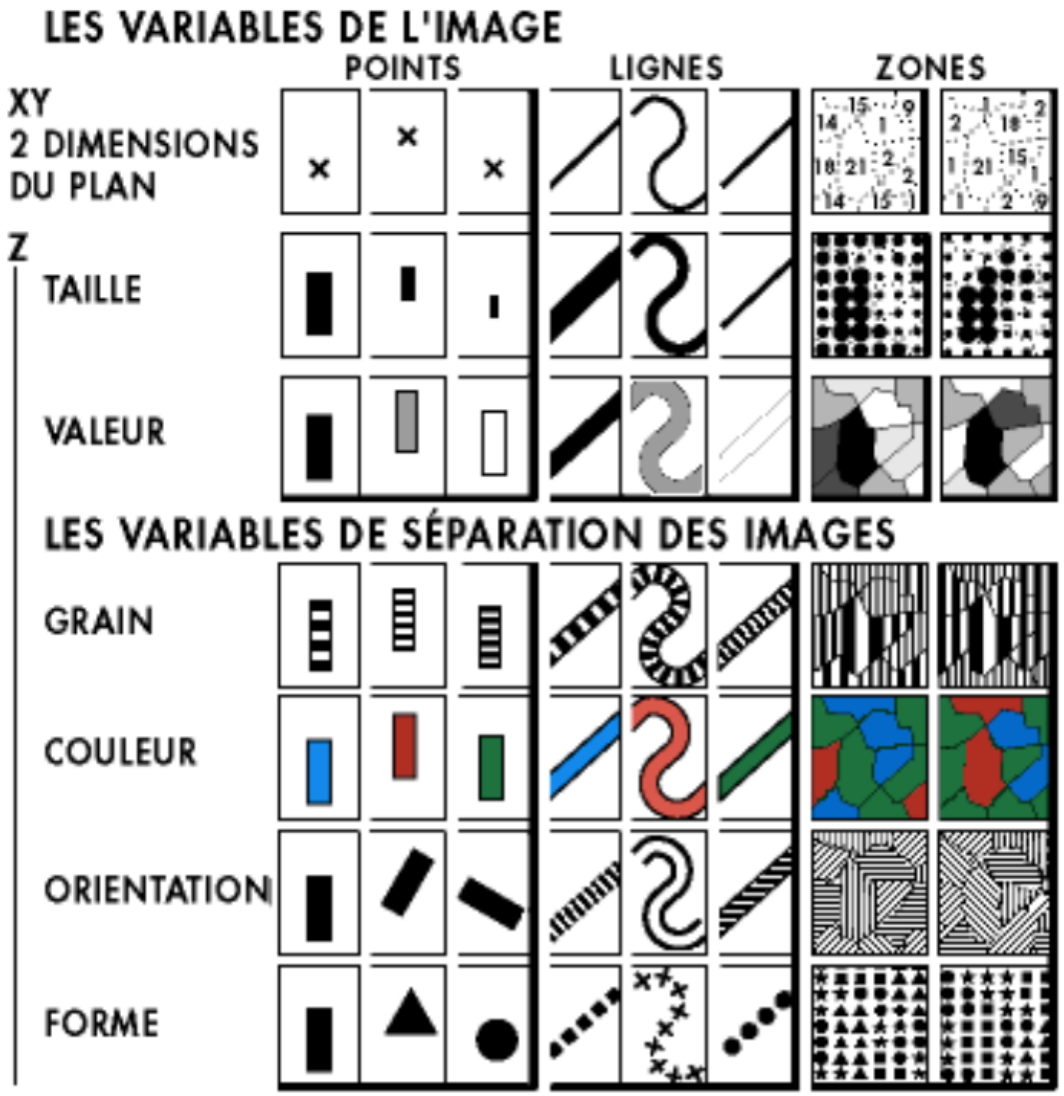


How many 3s?

1281768756138976546984506985604982826762
9809858458224509856458945098450980943585
9091030209905959595772564675050678904567
8845789809821677654876364908560912949686

How many 3s?

1281768756**3**897654698450698560498282676298098584582245098564589450984509809**3**585909103020990595959577256467505067890456788457898098216776548763649085609129496 86



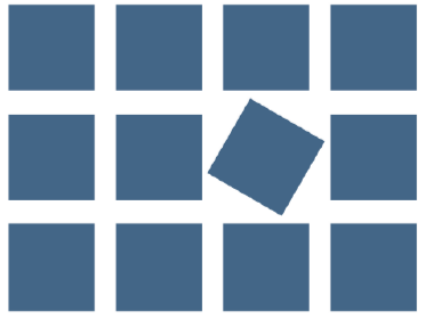
[Bertin, Simonolgy of Graphics, 1983]

Color hue

Color brightness

Position

Orientation

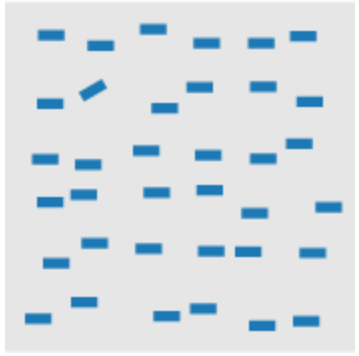


Color saturation

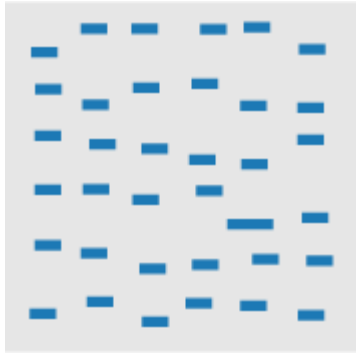
Size

Texture

Shape

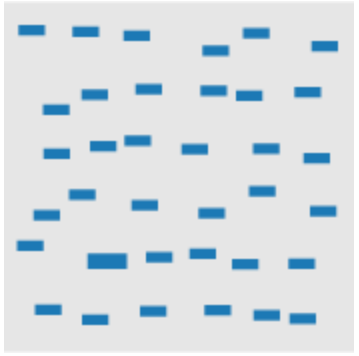


Line orientation

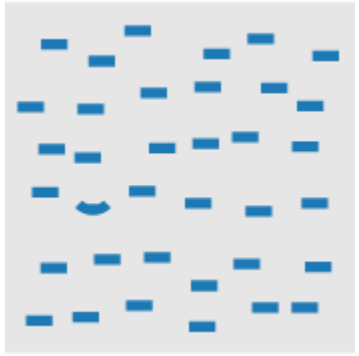


Length, Width

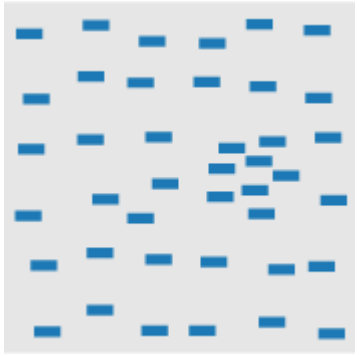
Closure



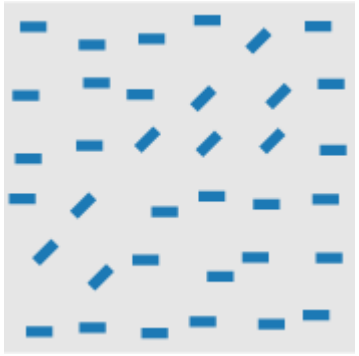
Size



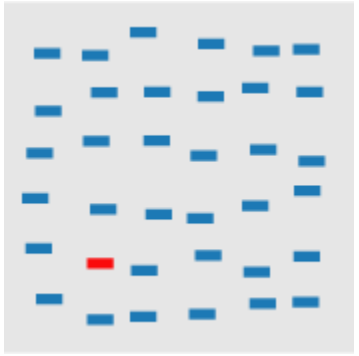
Curvature



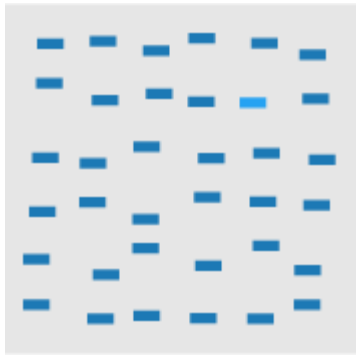
Density



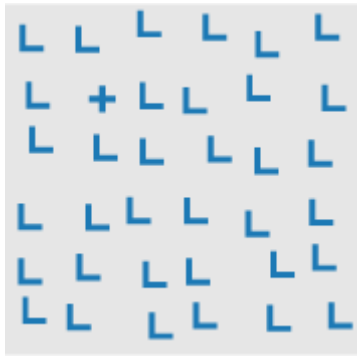
Number, Estimation



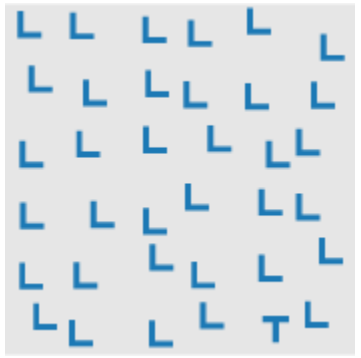
Hue



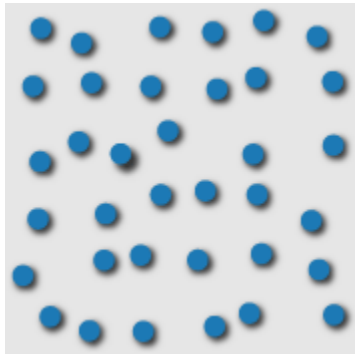
Intensity



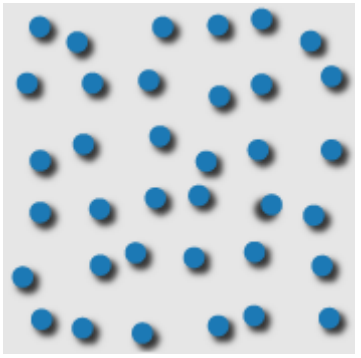
Intersection



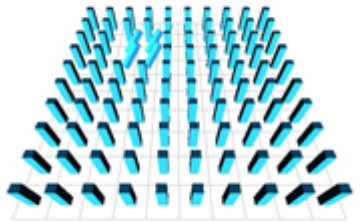
Terminators



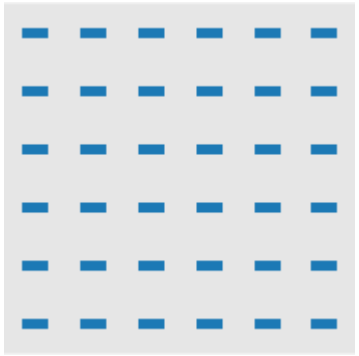
3D depth cues,
stereoscopic depth



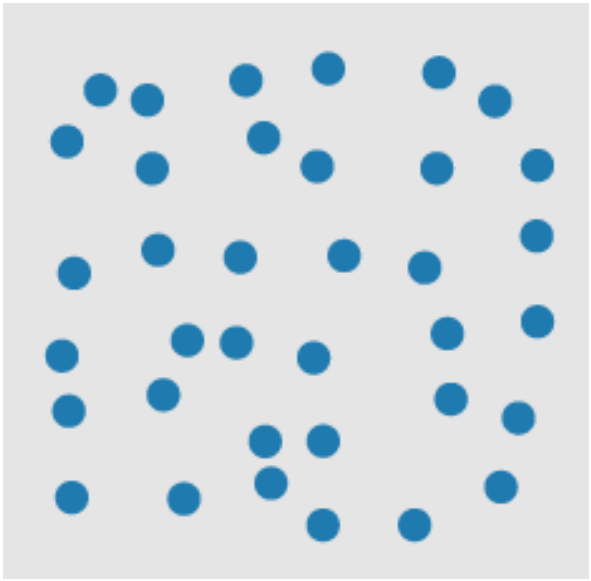
Lighting direction

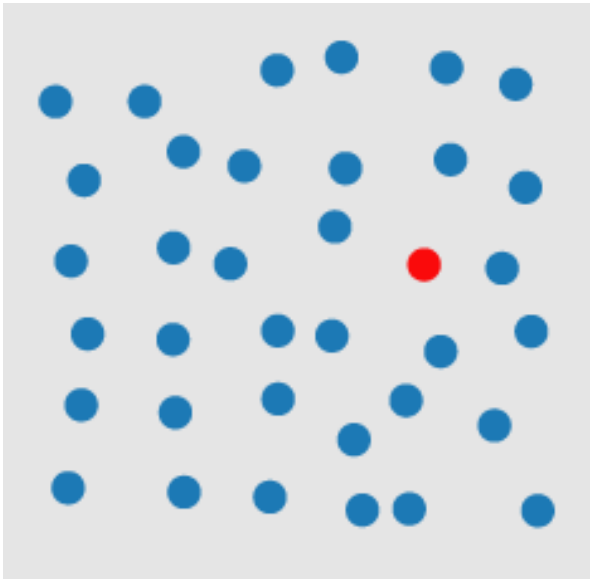


3D orientation

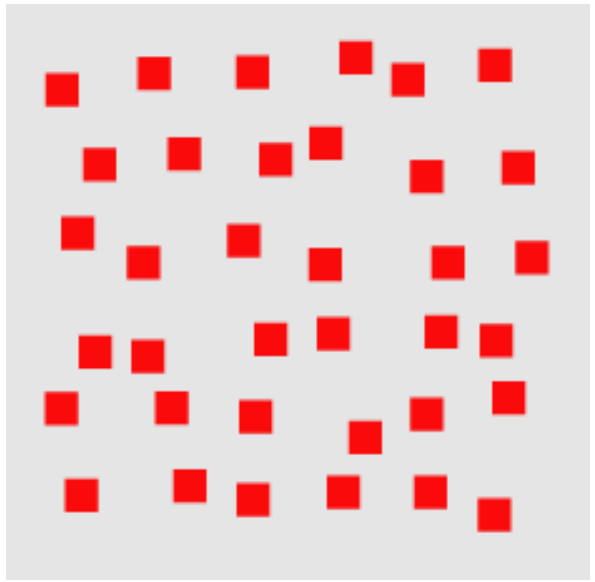


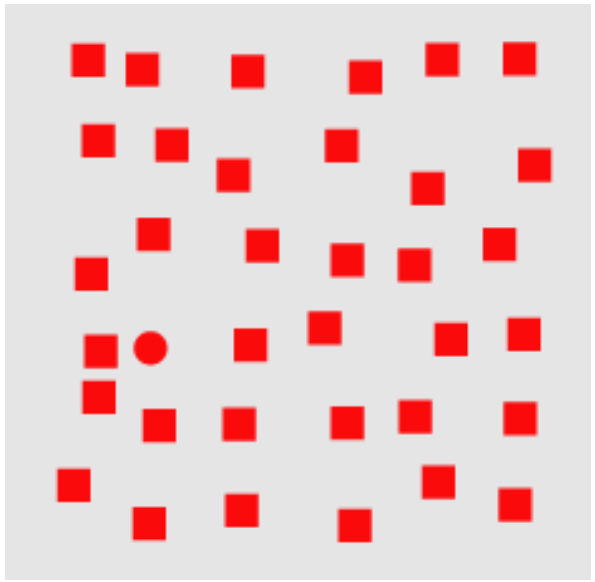
Flicker



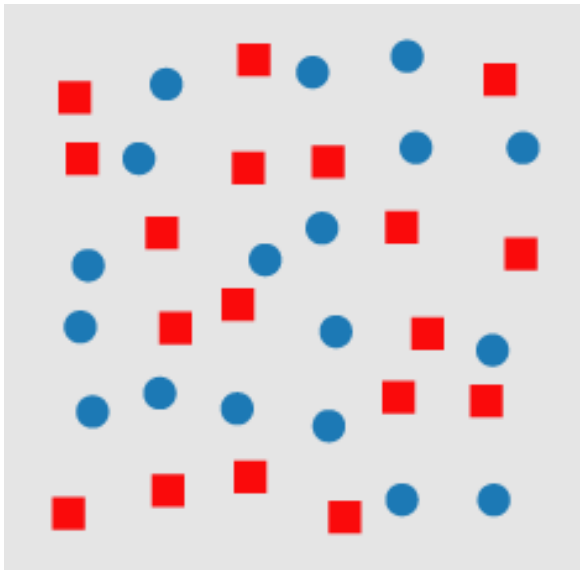


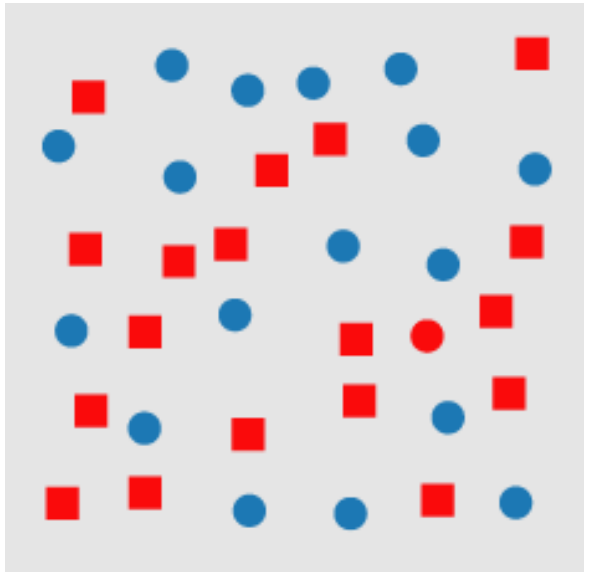
[http://www.csc.ncsu.edu/faculty/healey/PP/index.html]



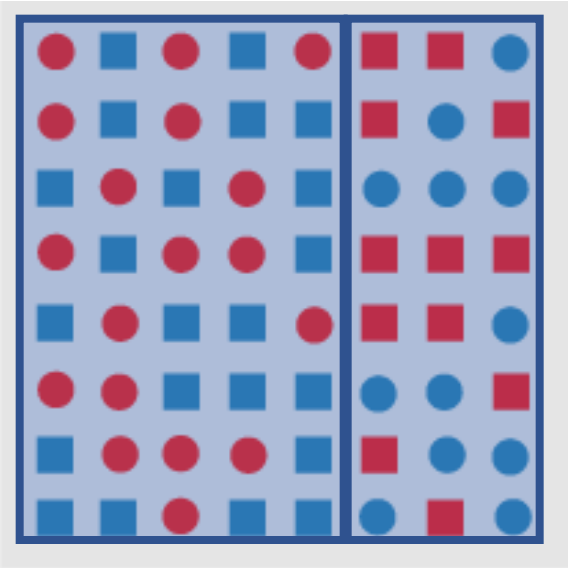


[http://www.csc.ncsu.edu/faculty/healey/PP/index.html]



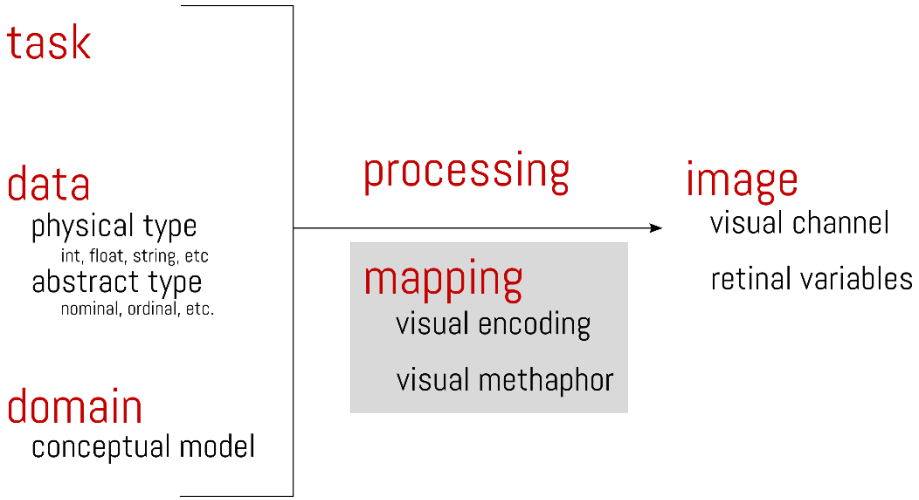


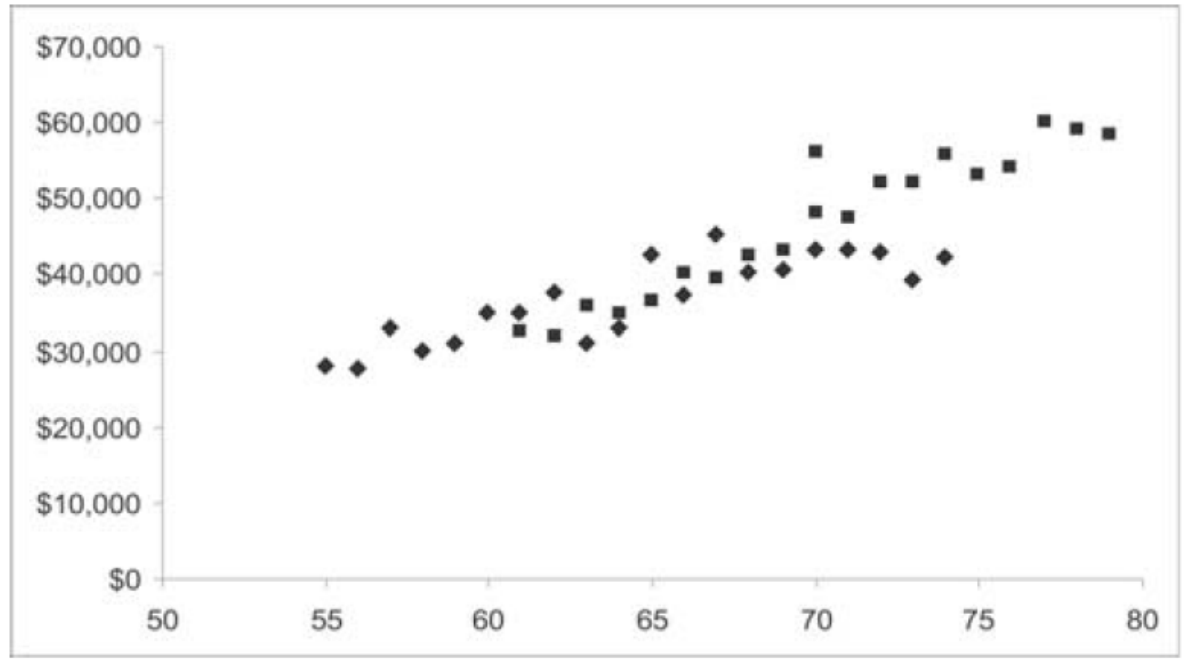
[http://www.csc.ncsu.edu/faculty/healey/PP/index.html]

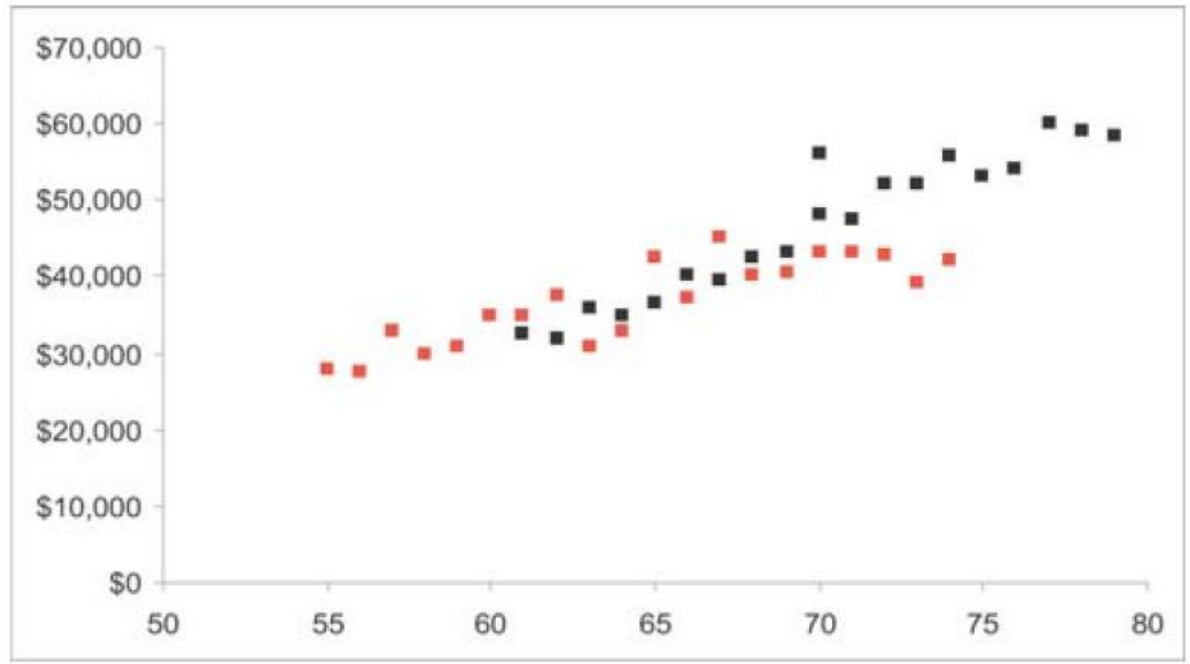


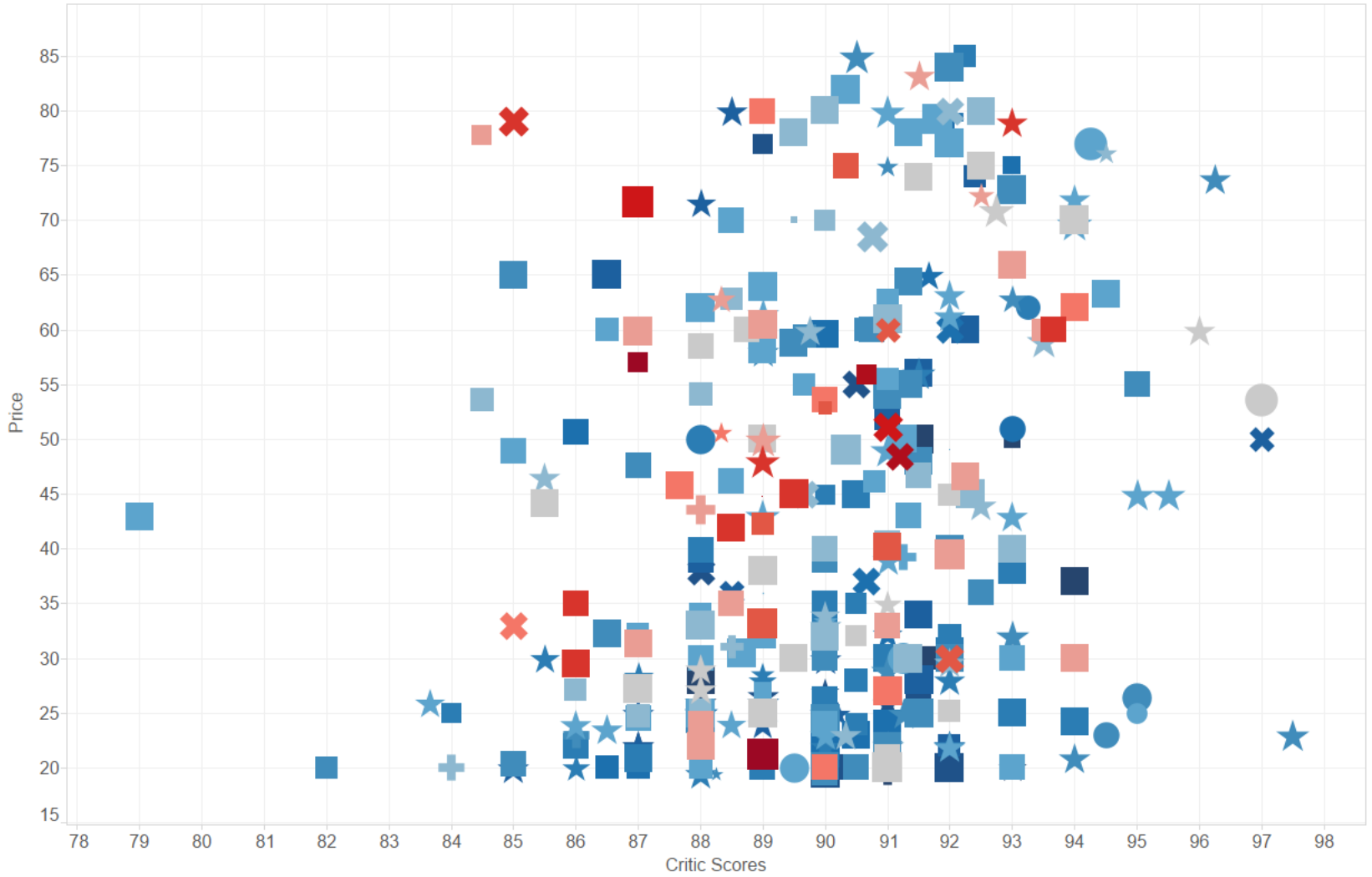
[http://www.csc.ncsu.edu/faculty/healey/PP/index.html]

Mapping


task
data
physical type
int, float, string, etc
abstract type
nominal, ordinal, etc.
domain
conceptual model
processing
mapping
visual encoding
visual methaphor
image
visual channel
retinal variables


$70,000
$60,000
$50,000
$40,000
$30,000
$20,000
$10,000
$0
50
55
60
65
70
75
80

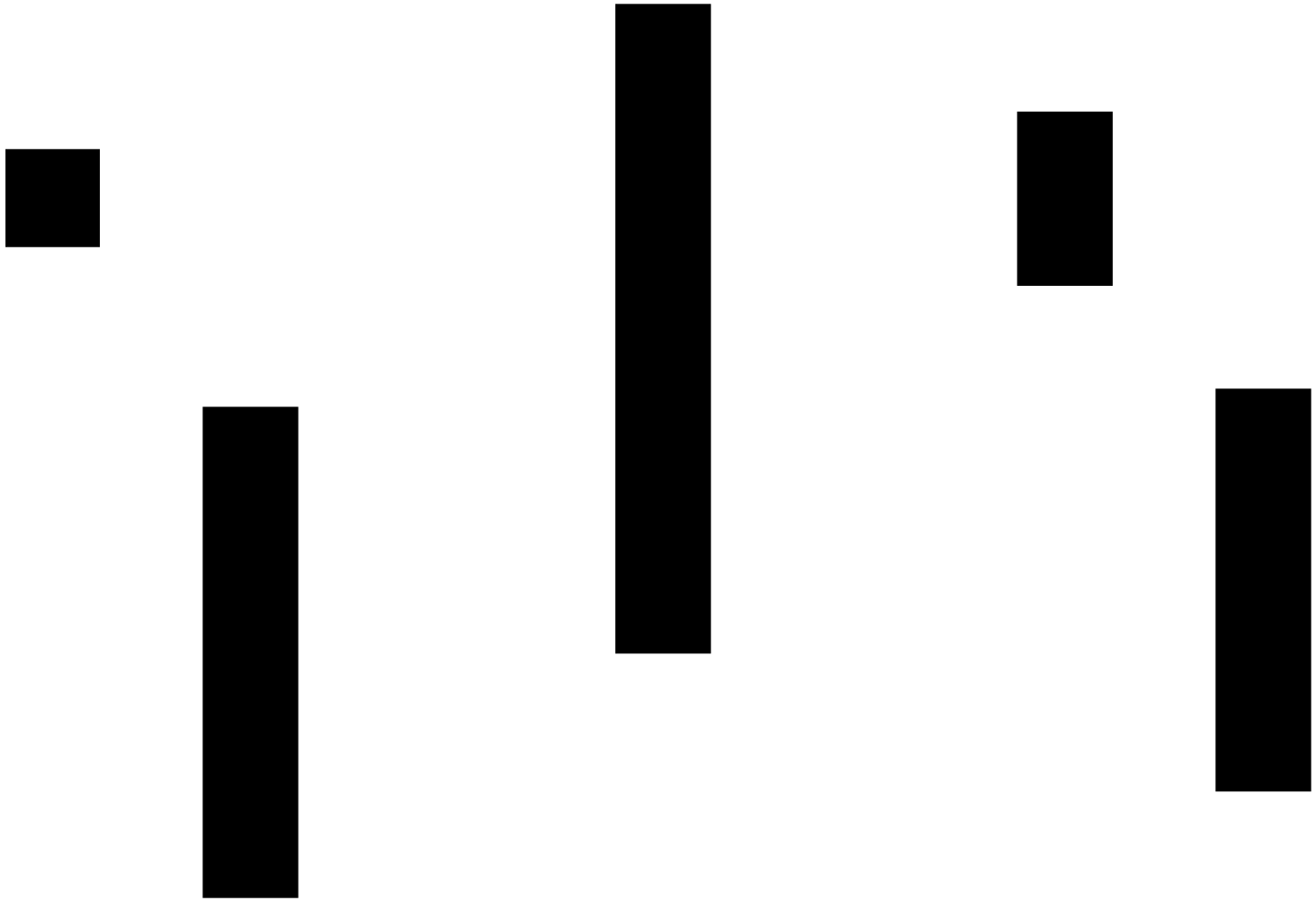

$70,000
$60,000
$50,000
$40,000
$30,000
$20,000
$10,000
$0
50
55
60
65
70
75
80

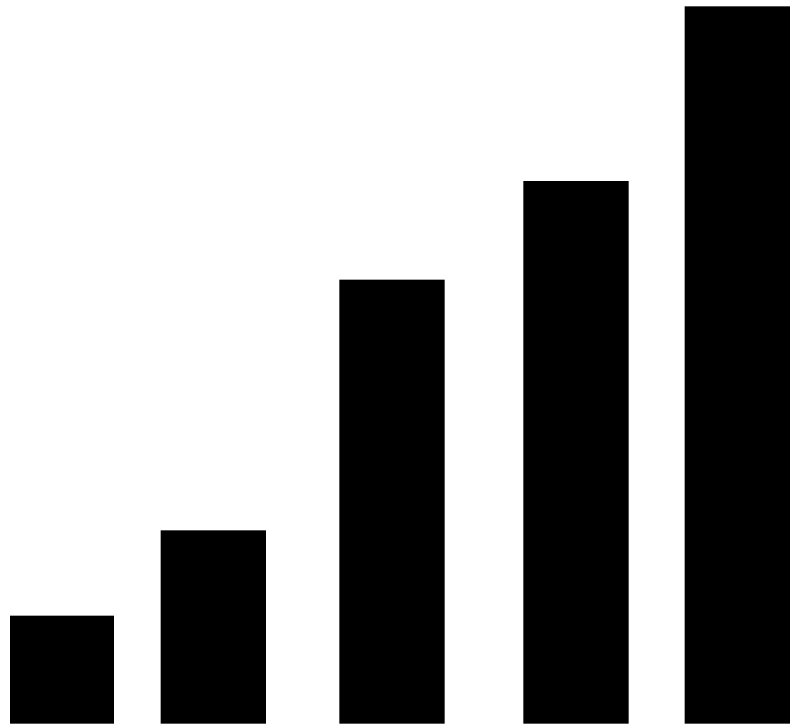


X: Critic Scores, **Y:** Price, **Size:** User Rating, **Color:** Vintage, **Shape:** Type

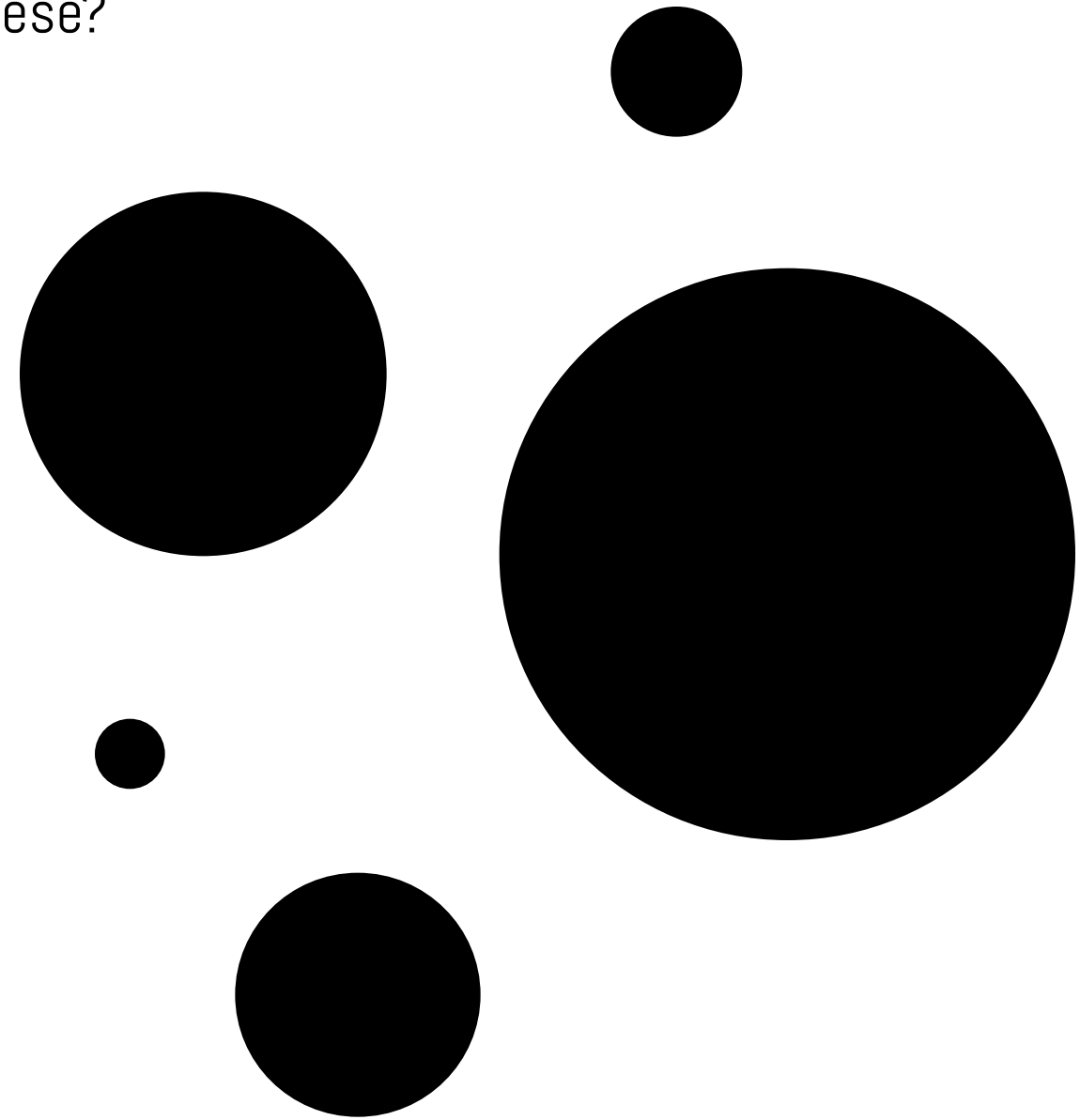
Can you order these?

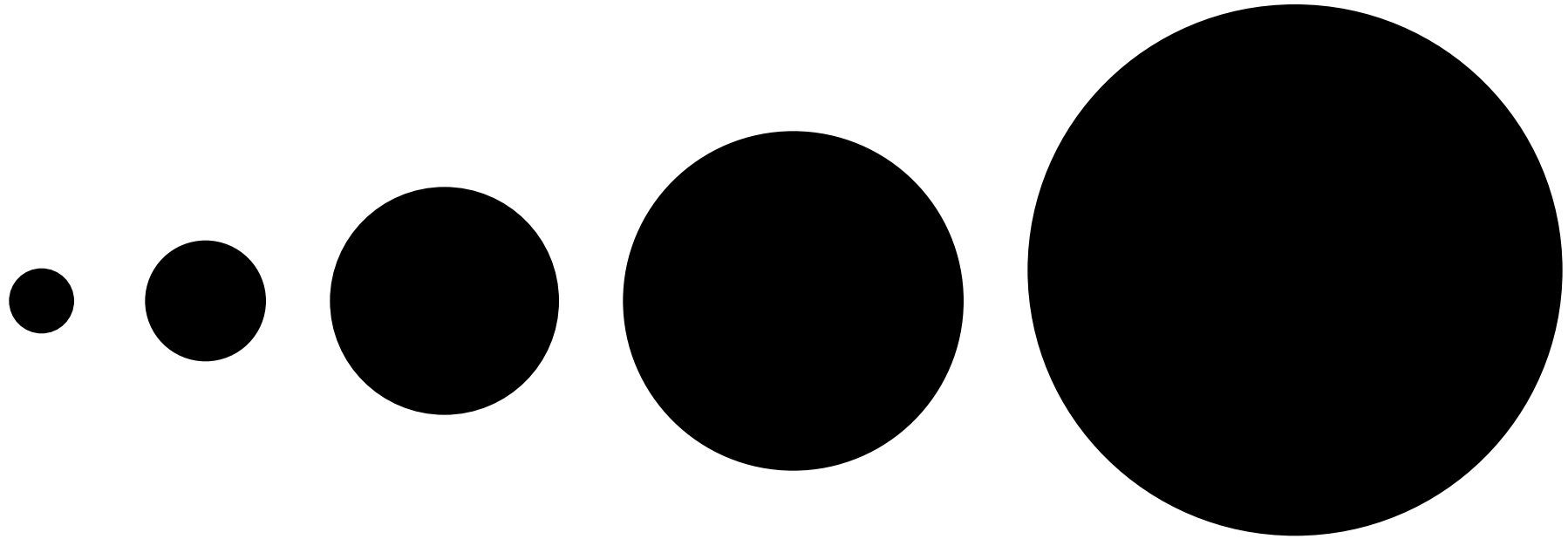
(low -> high)





Can you order these?
(low -> high)



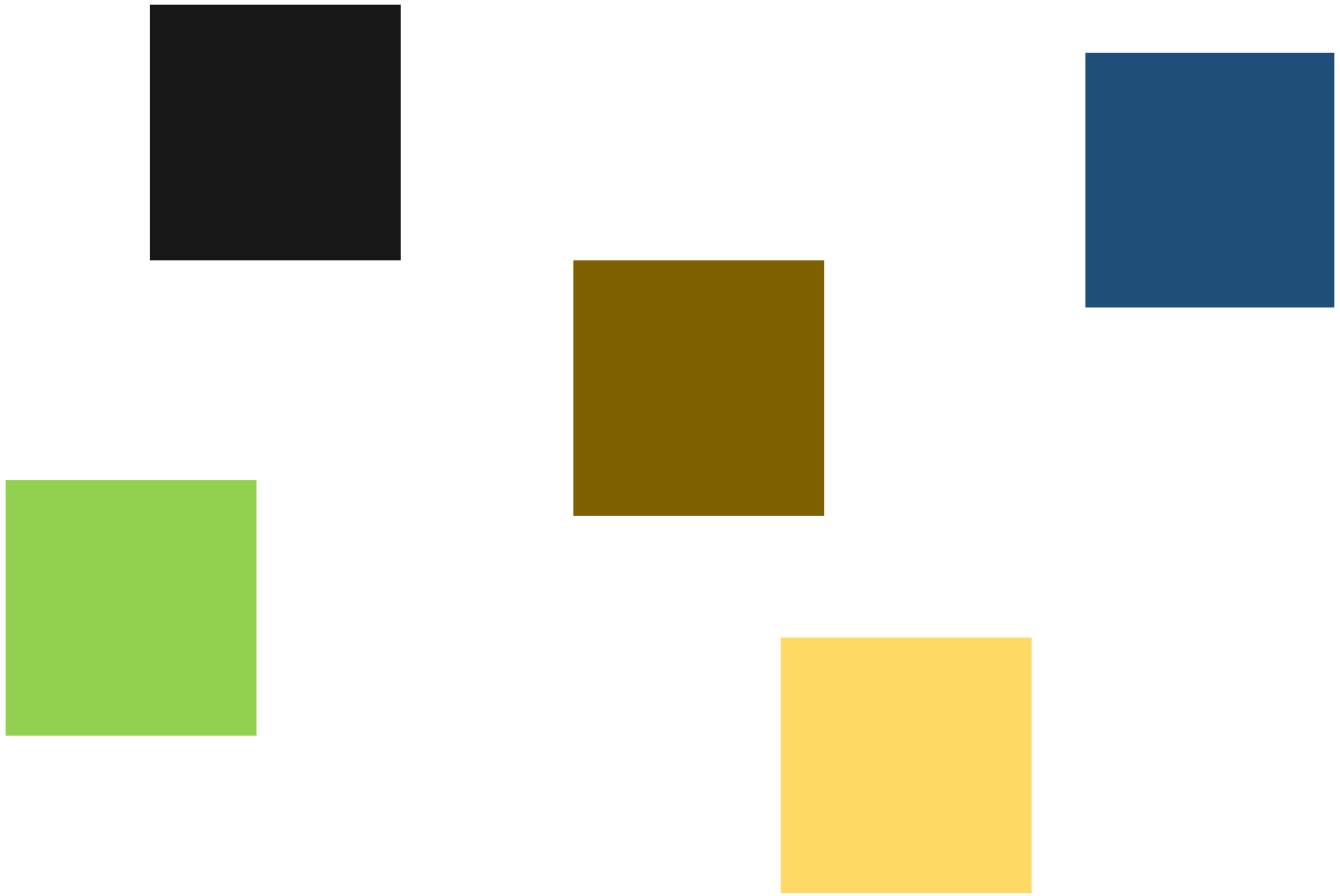


Can you order these?

(low -> high)

Can you order these?

(low -> high)

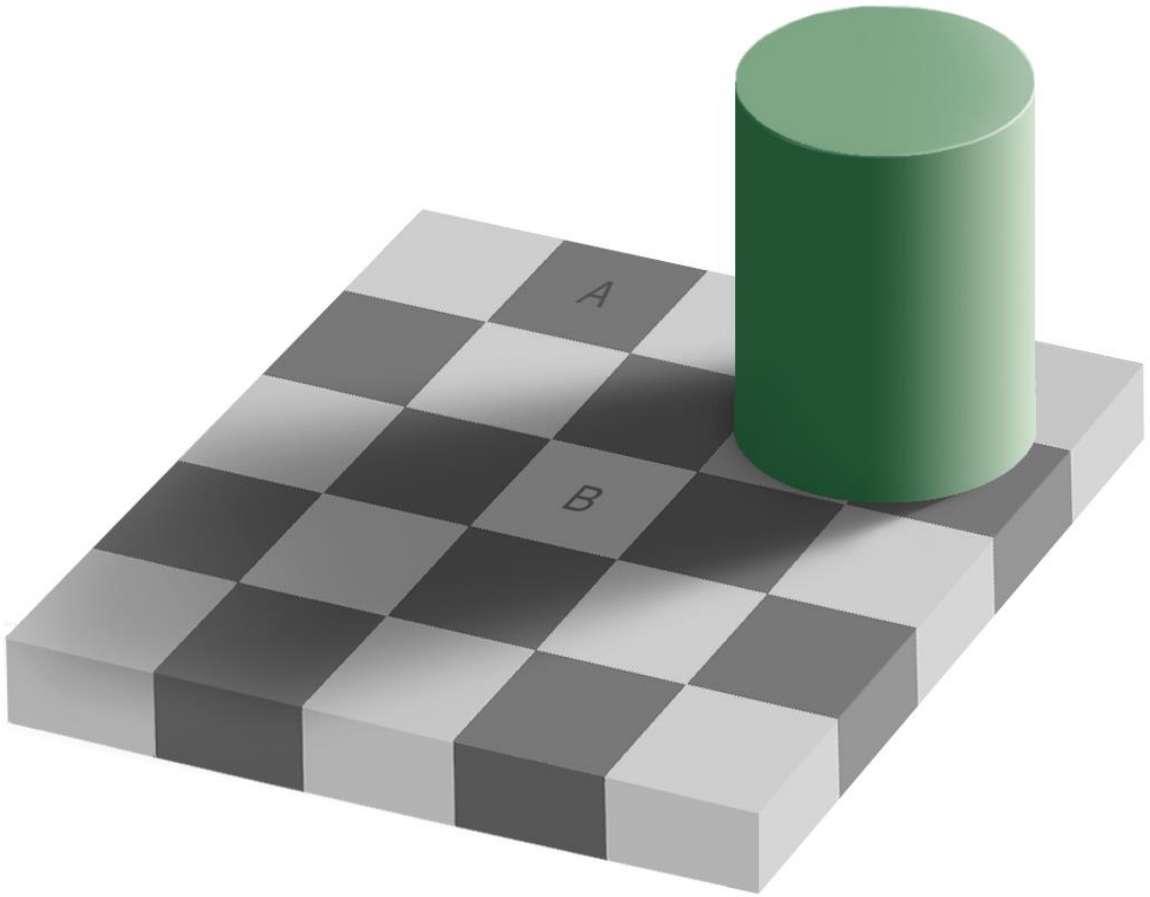


Nominal, Ordinal and Quantitative

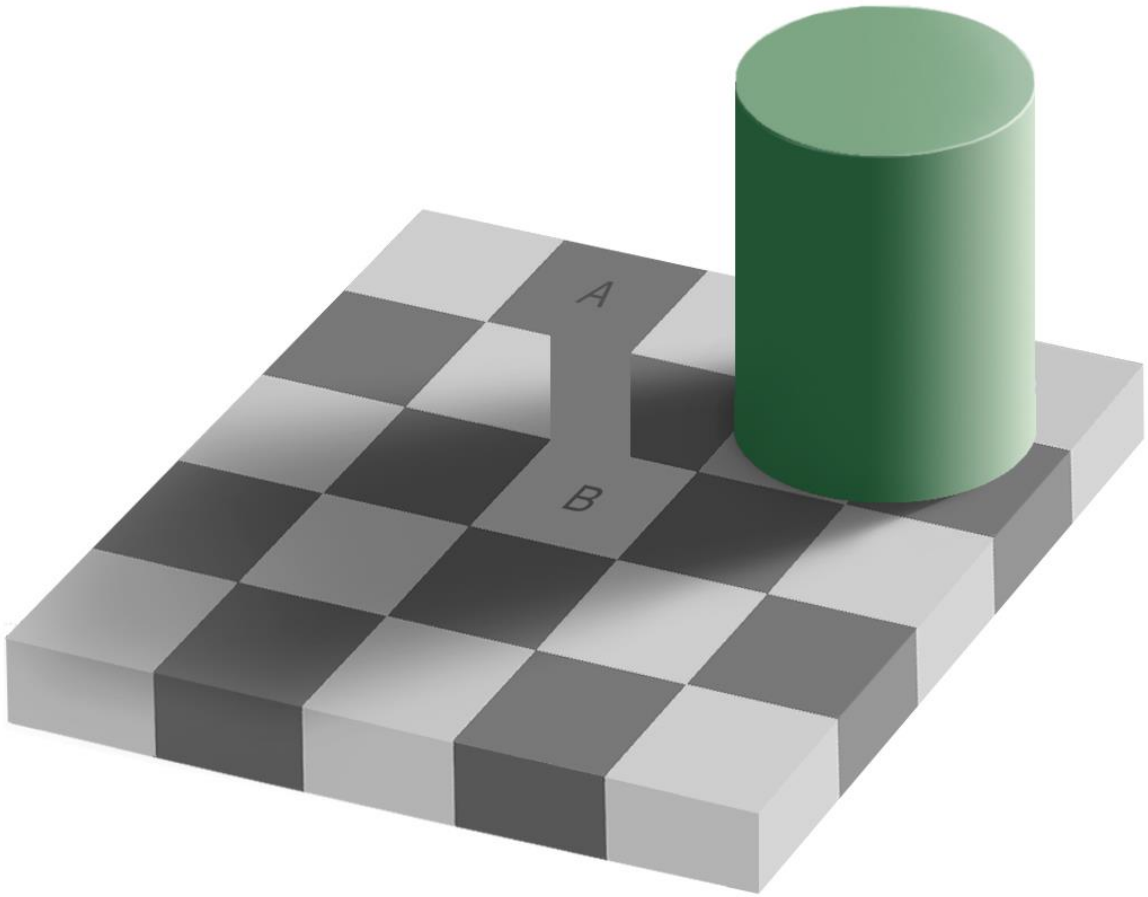
Position	N	O	Q
Size	N	O	Q
Value	N	O	Q
Texture	N	O	
Color	N		
Orientation	N		
Shape	N		

Nominal
Ordered
Quantitative

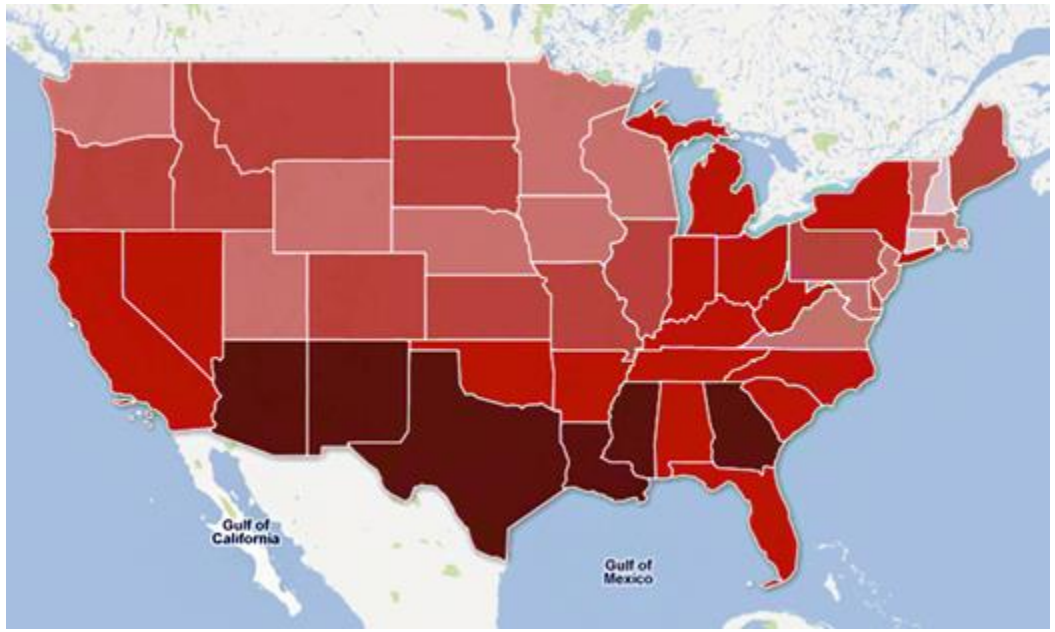
Colors can be misleading



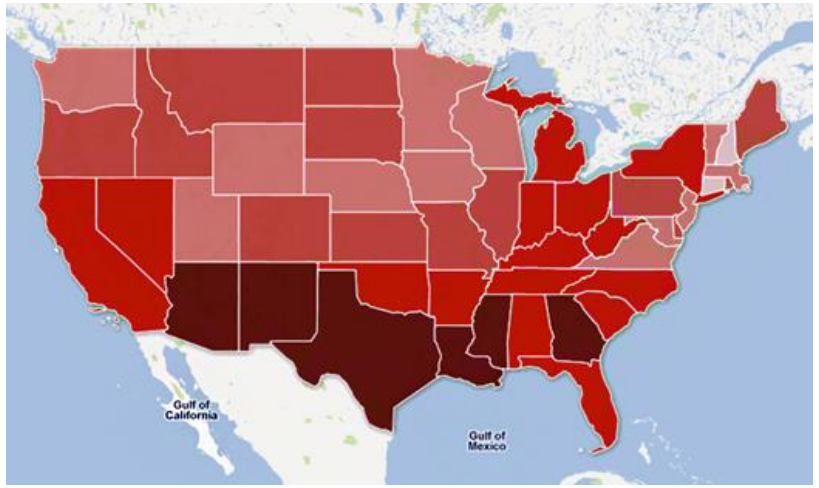
[Edward Adelson, 1995]

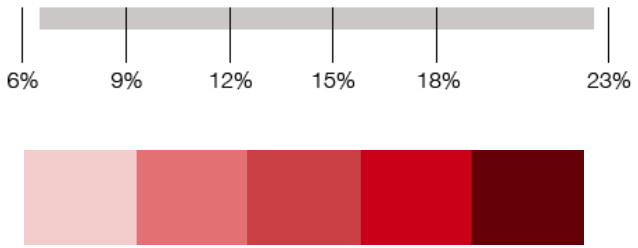


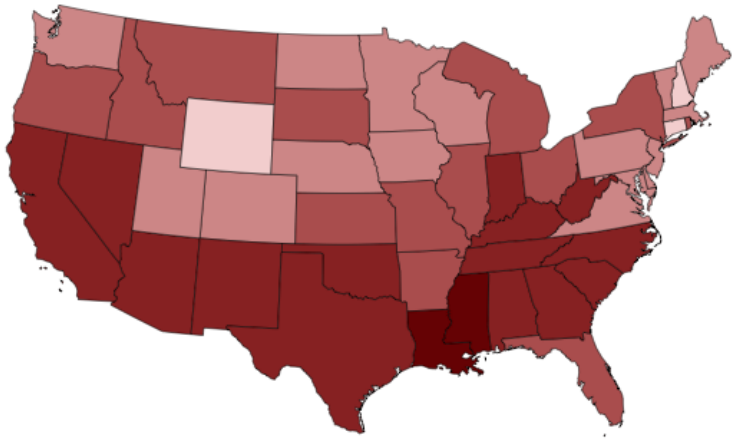
[Edward Adelson, 1995]

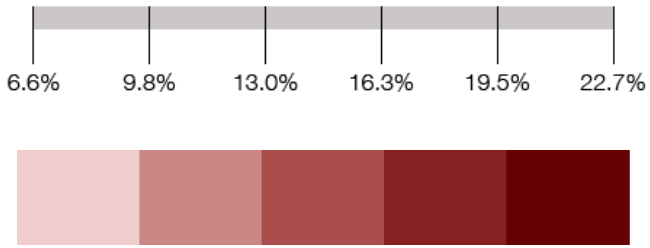


[http://vis4.net/blog/posts/choropleth-maps/]





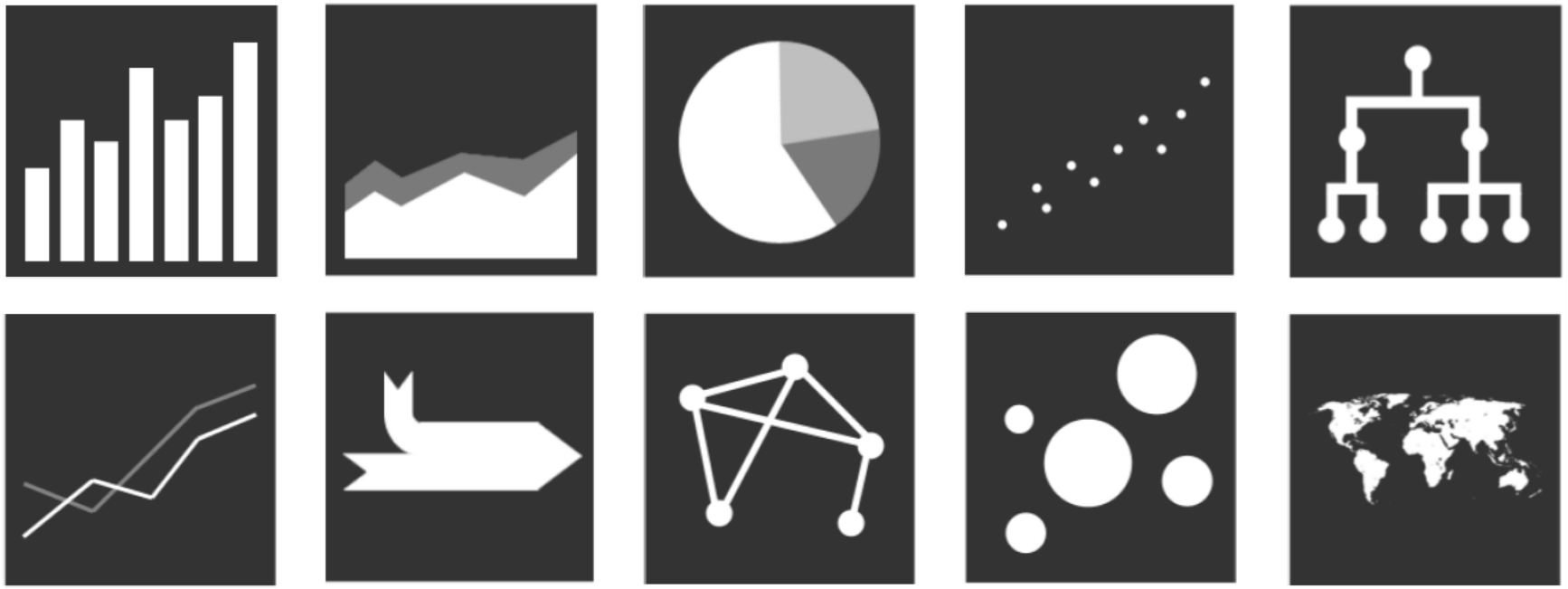




[http://vis4.net/blog/posts/choropleth-maps]

Some Principles & Guidelines

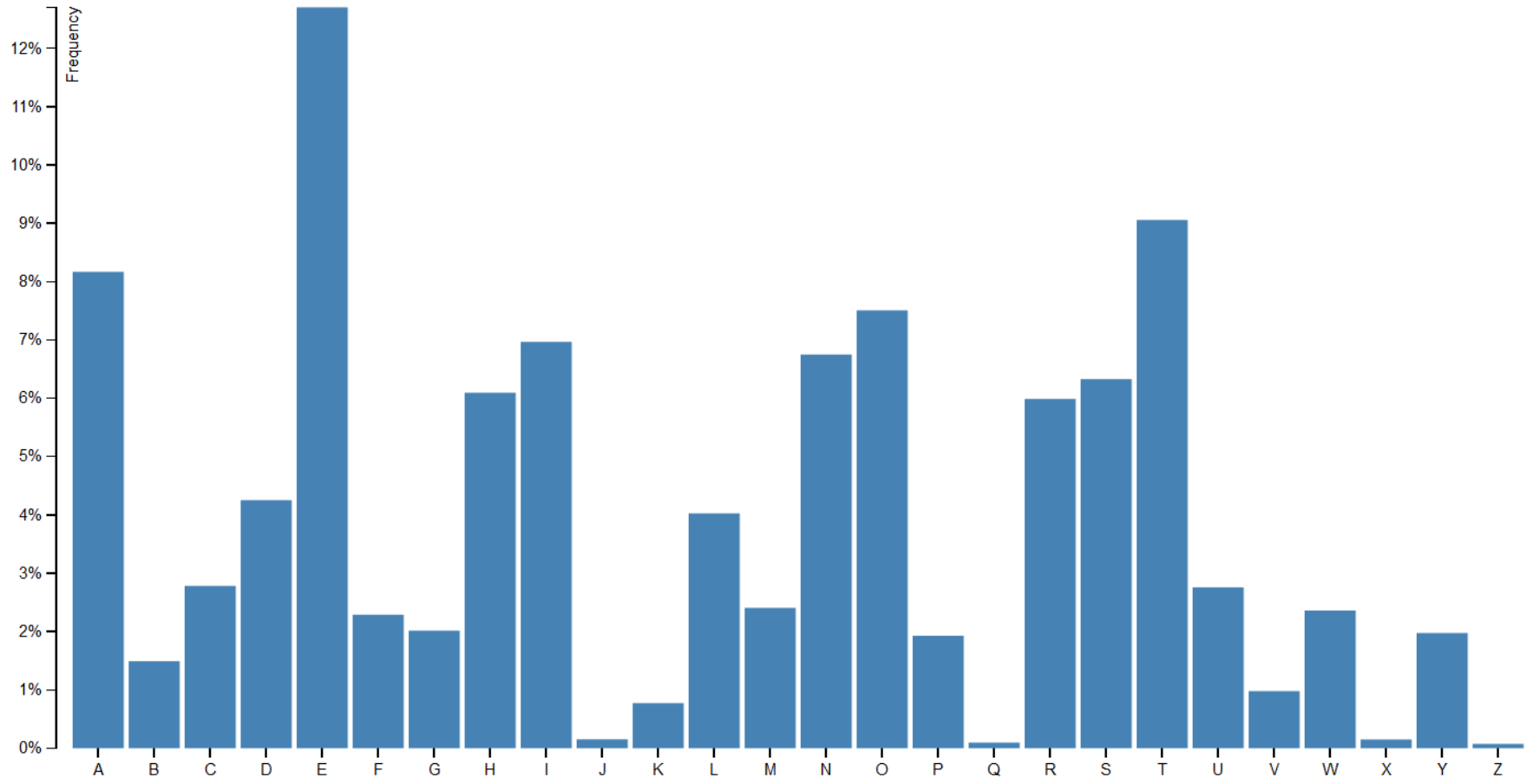
Find and adapt existing visualizations



http://homes.cs.washington.edu/~jheer/files/zoo/

Bar chart

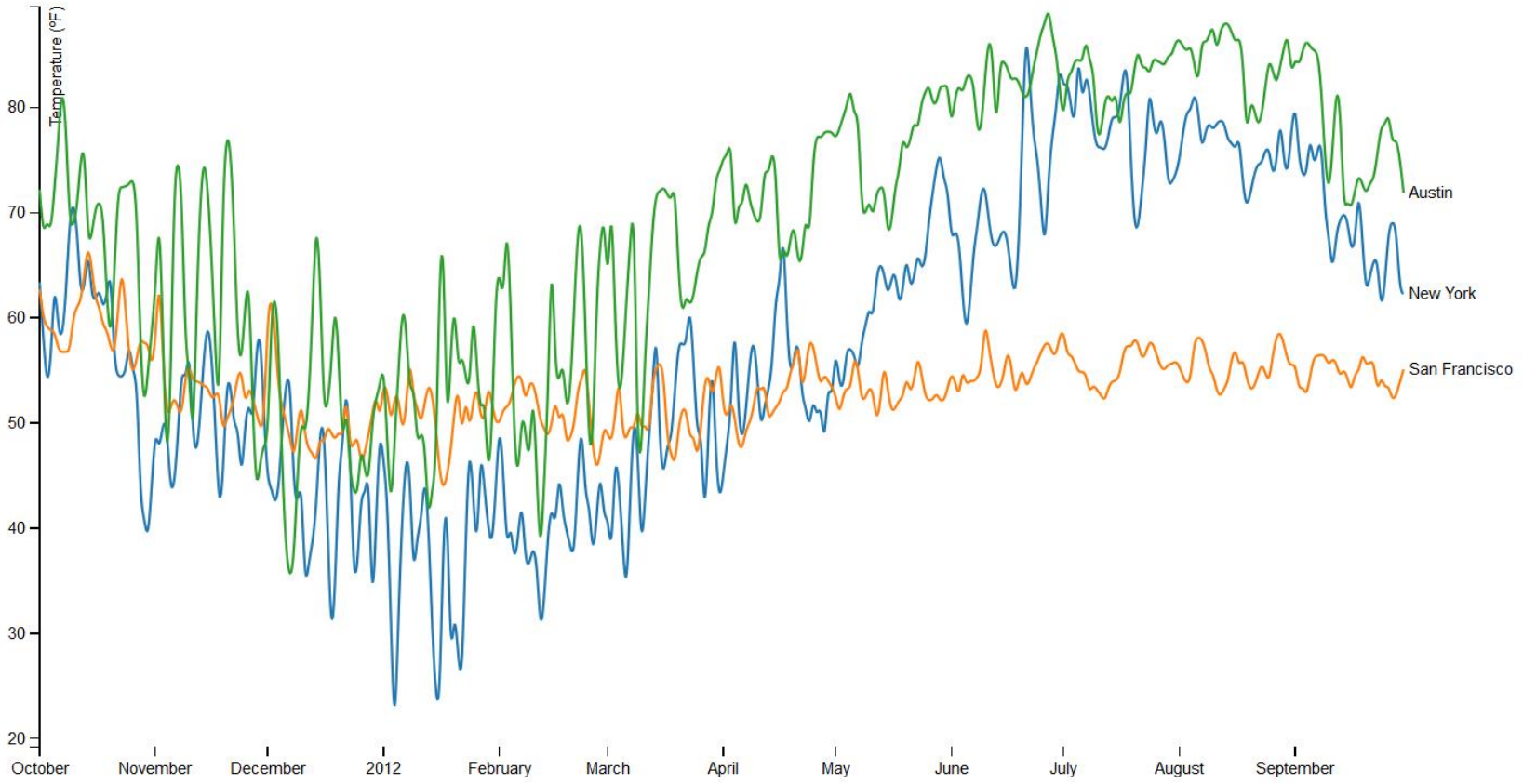
Display different quantities of single-variable data



https://github.com/mbostock/d3/wiki/Gallery

Line chart

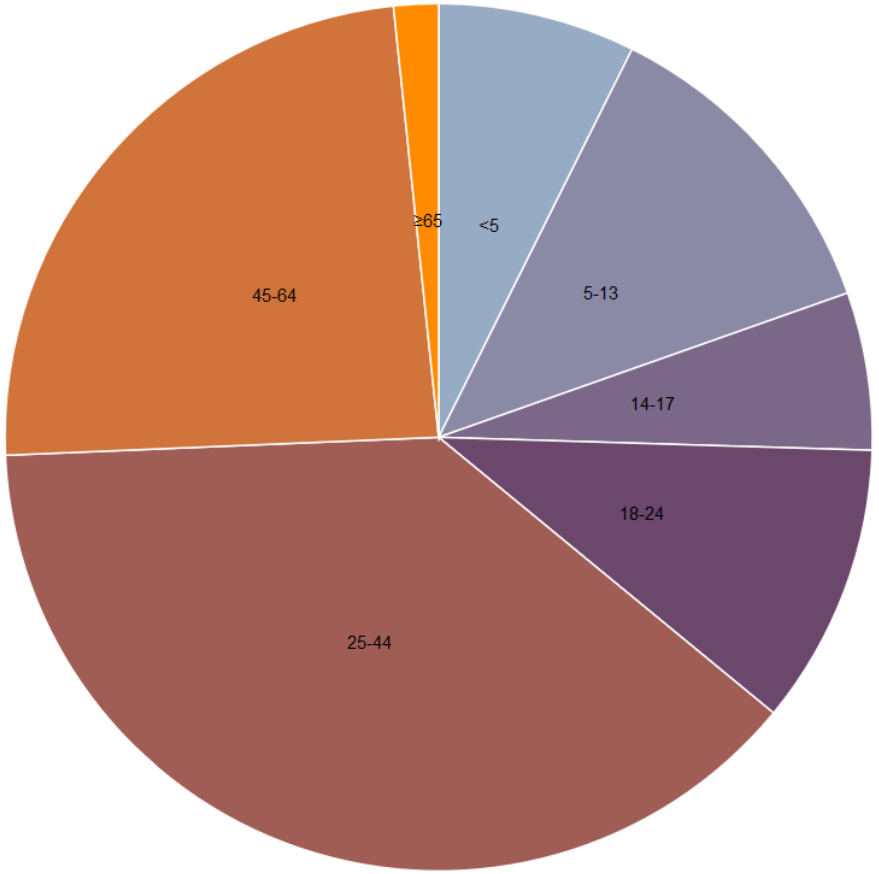
Display how a variable develops over time



https://github.com/mbostock/d3/wiki/Gallery

Pie chart

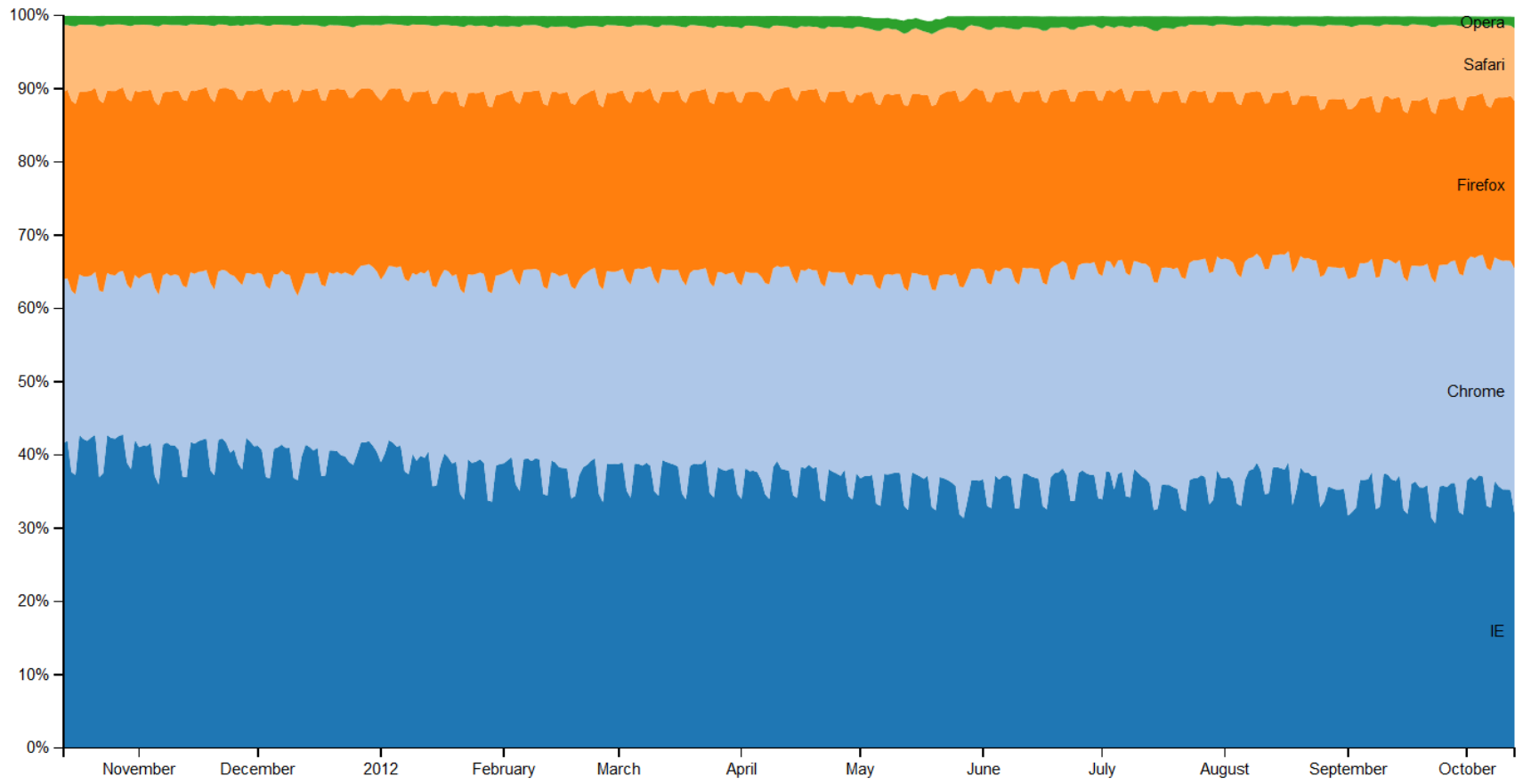
Display distribution of a variable



https://github.com/mbostock/d3/wiki/Gallery

Stacked area chart

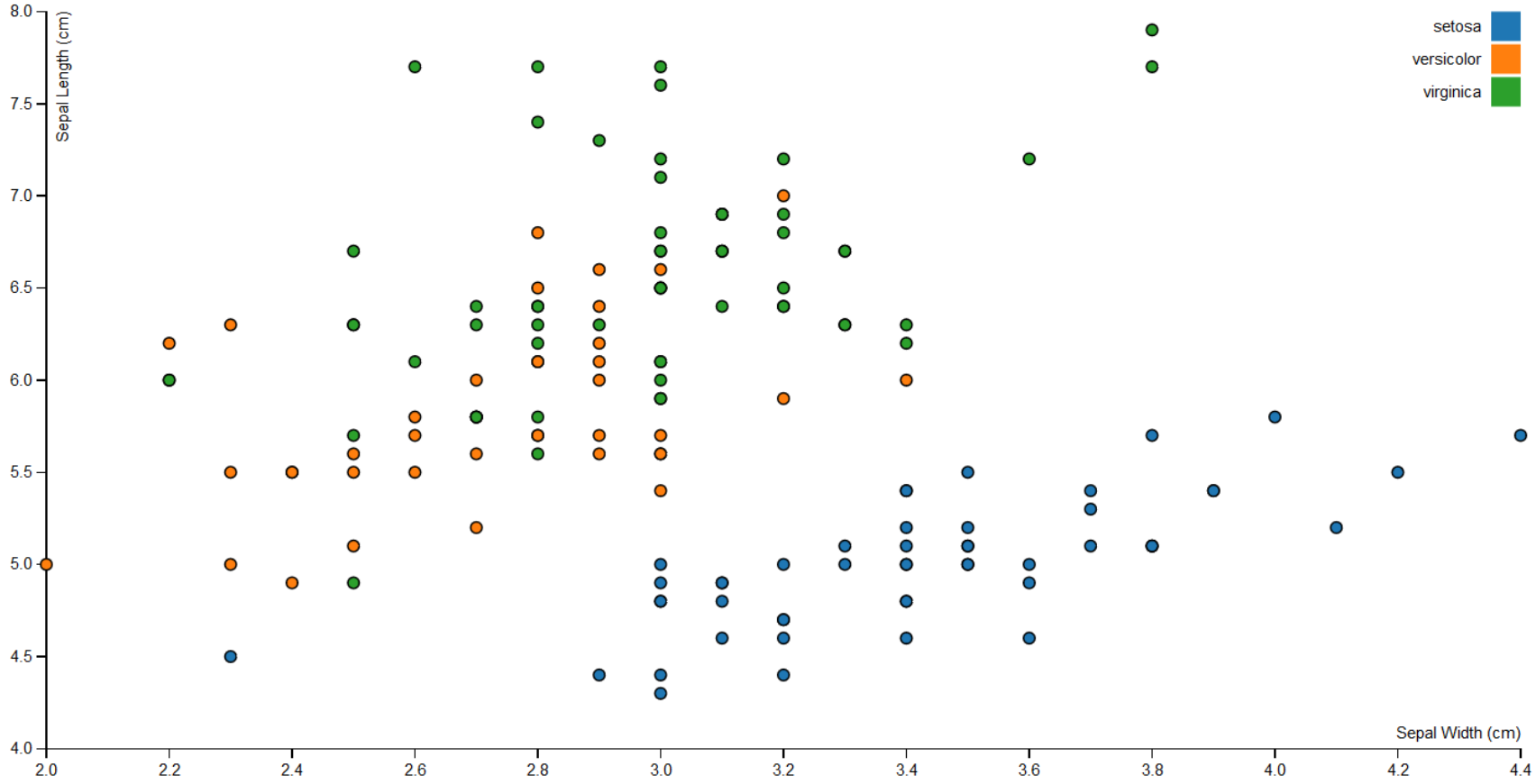
Display total of a variable over time



https://github.com/mbostock/d3/wiki/Gallery

Scatterplot

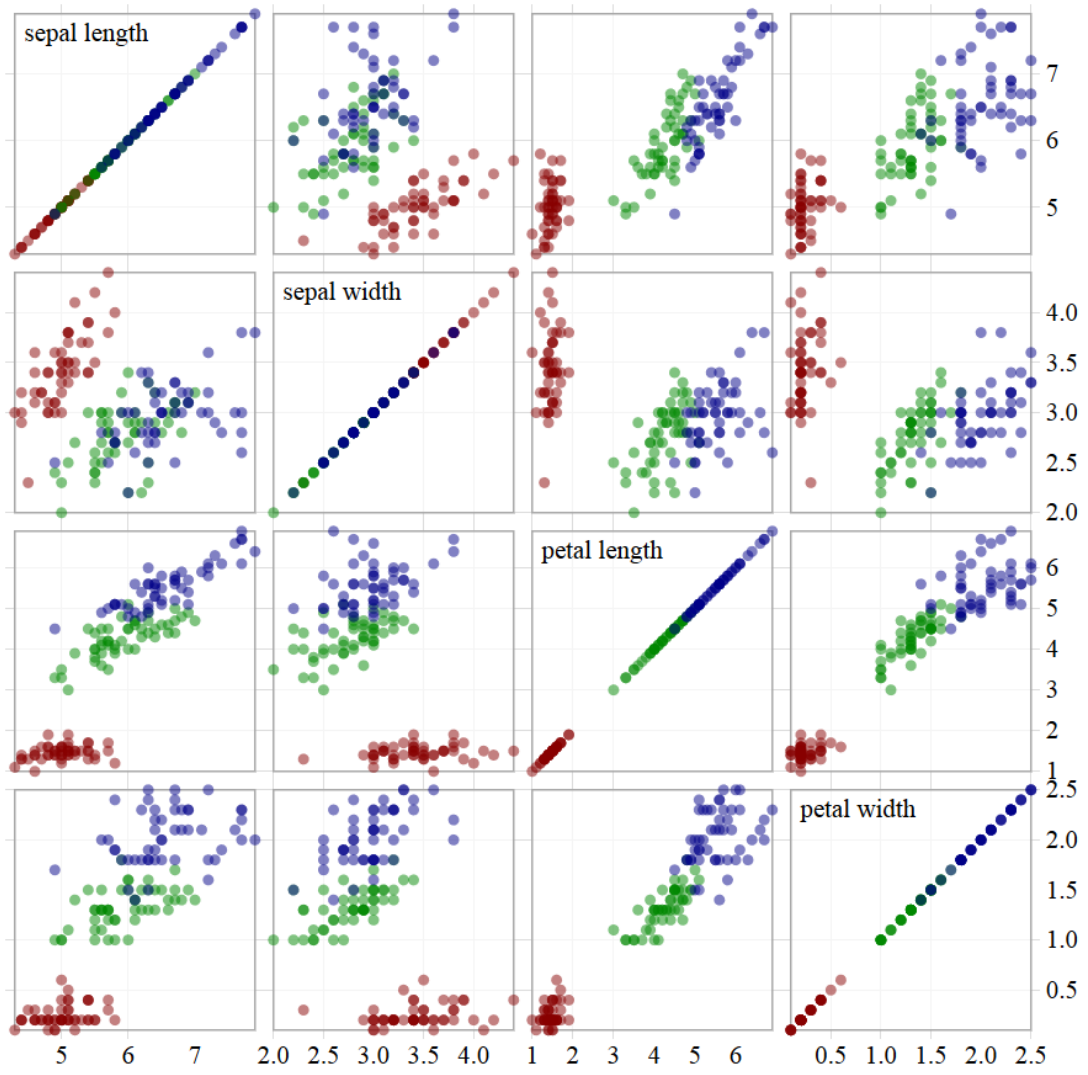
Display relationship between two variables



https://github.com/mbostock/d3/wiki/Gallery

Scatterplot matrix

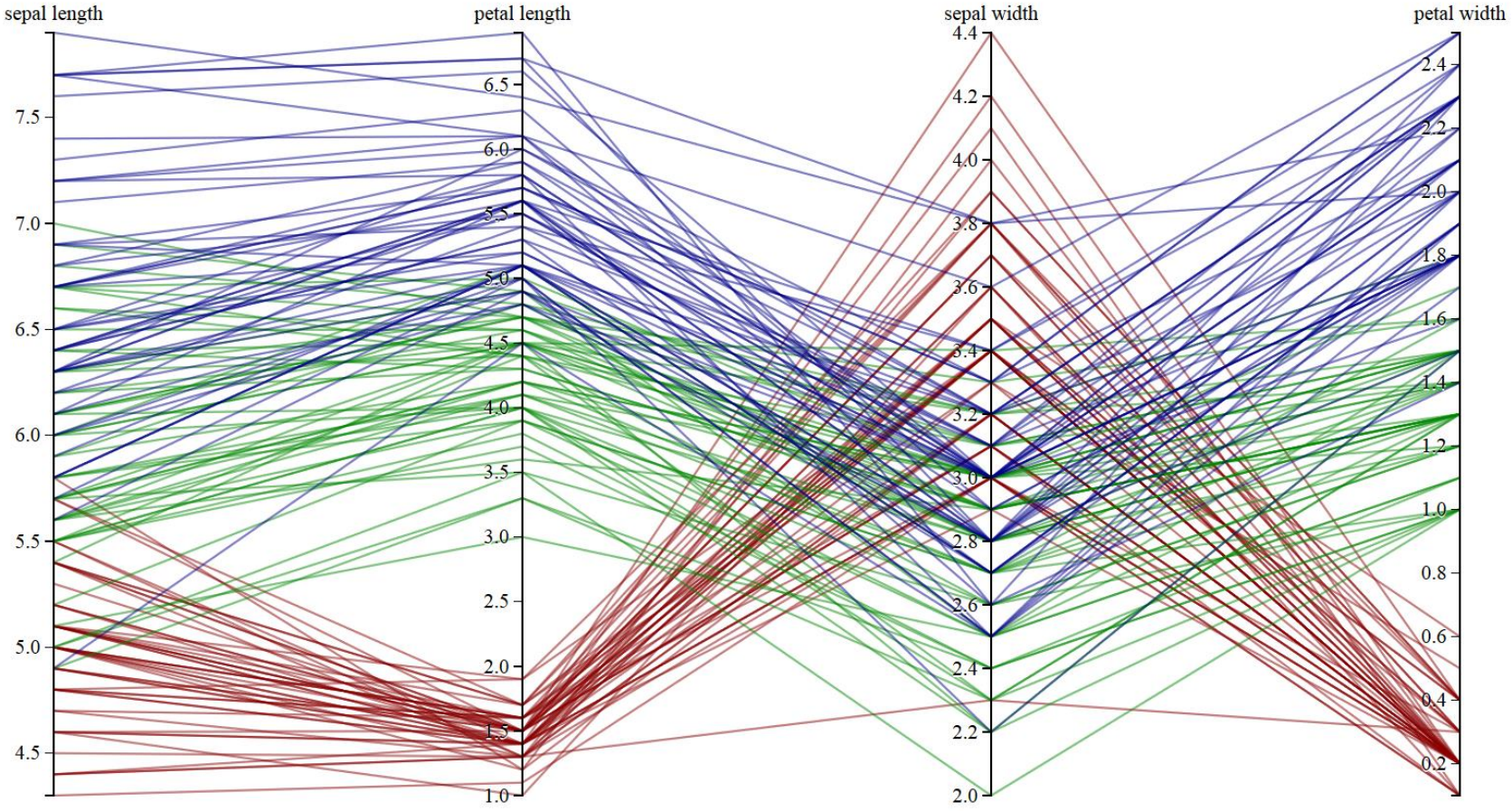
Display relationship between multiple variables



https://github.com/mbostock/d3/wiki/Gallery

Parallel coordinate plot

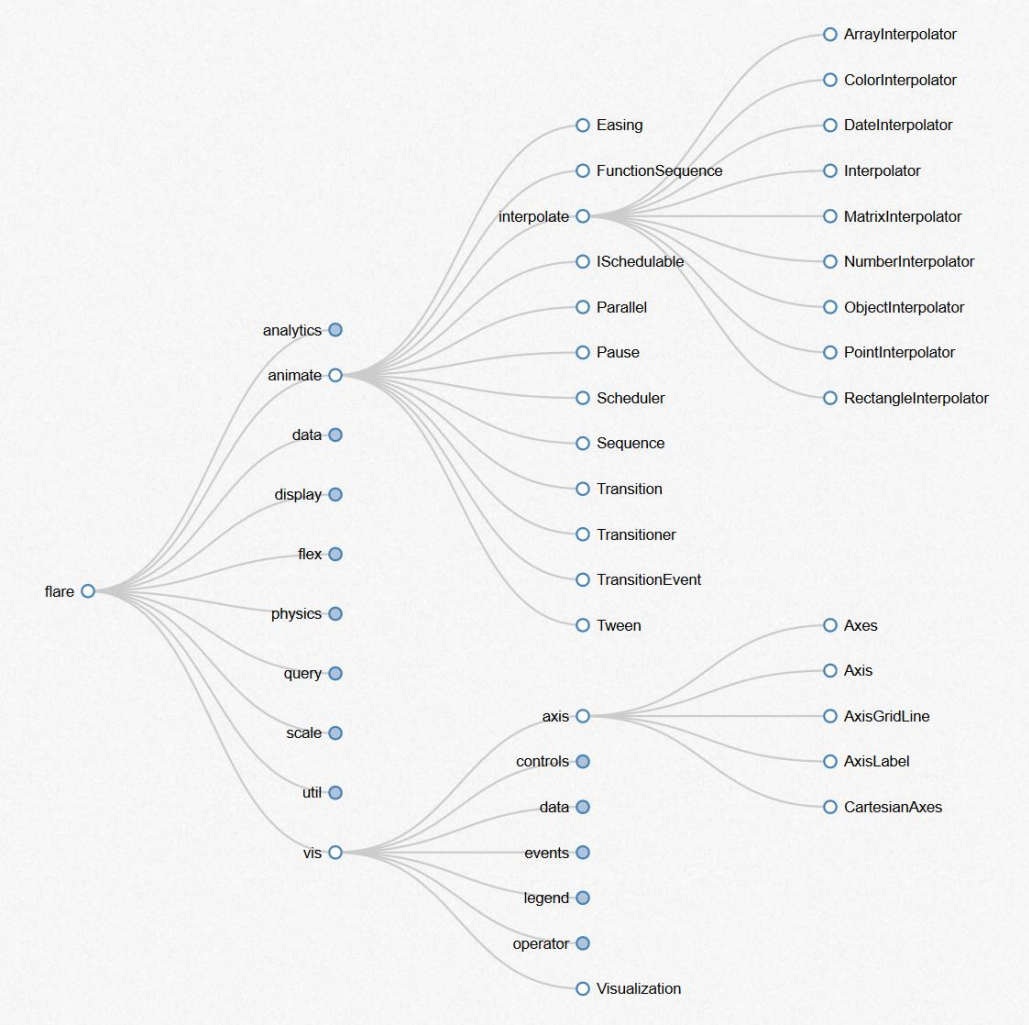
Display relationship between multiple variables



https://github.com/mbostock/d3/wiki/Gallery

Tree

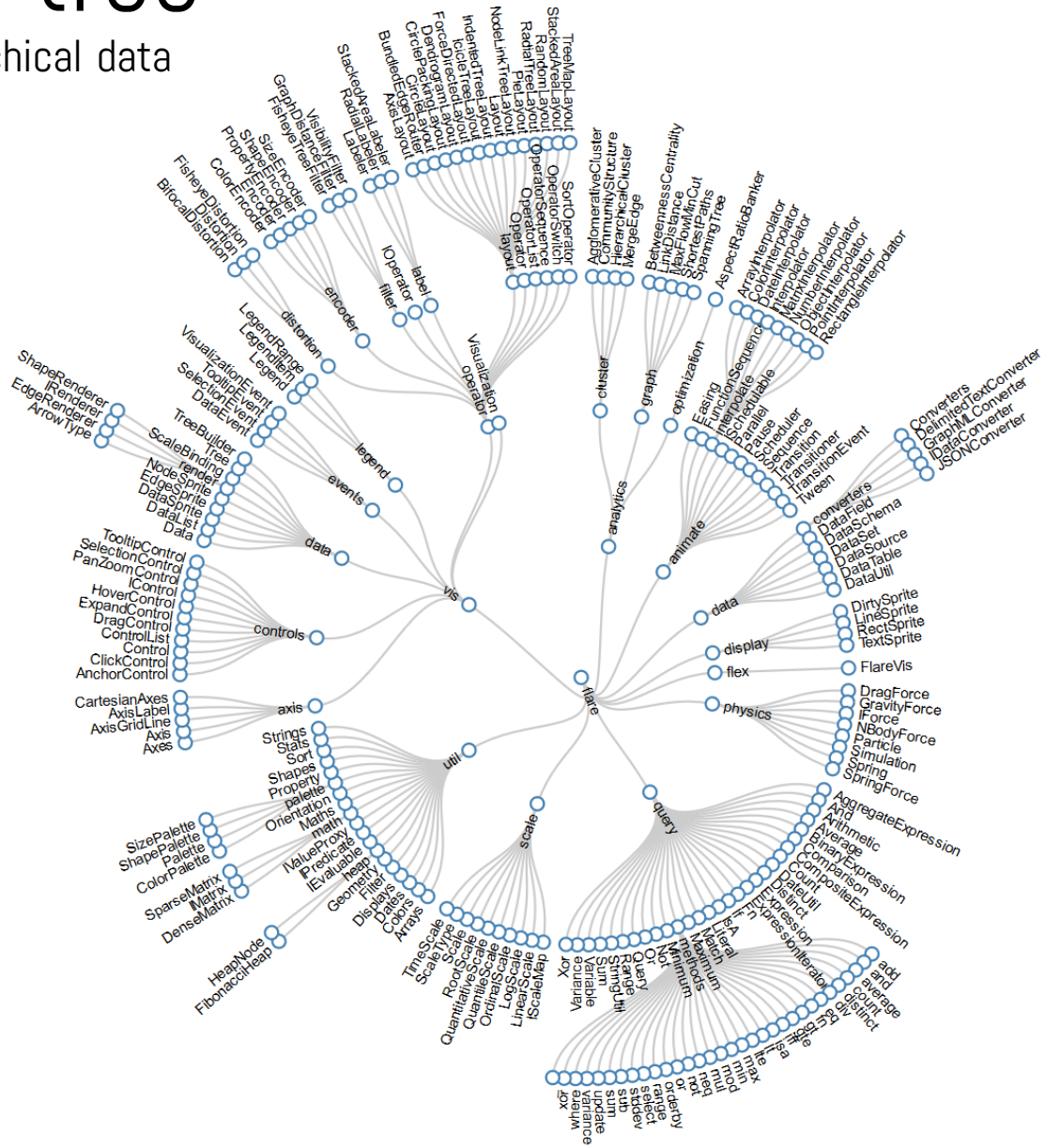
Display hierarchical data



https://github.com/mbostock/d3/wiki/Gallery

Radial tree

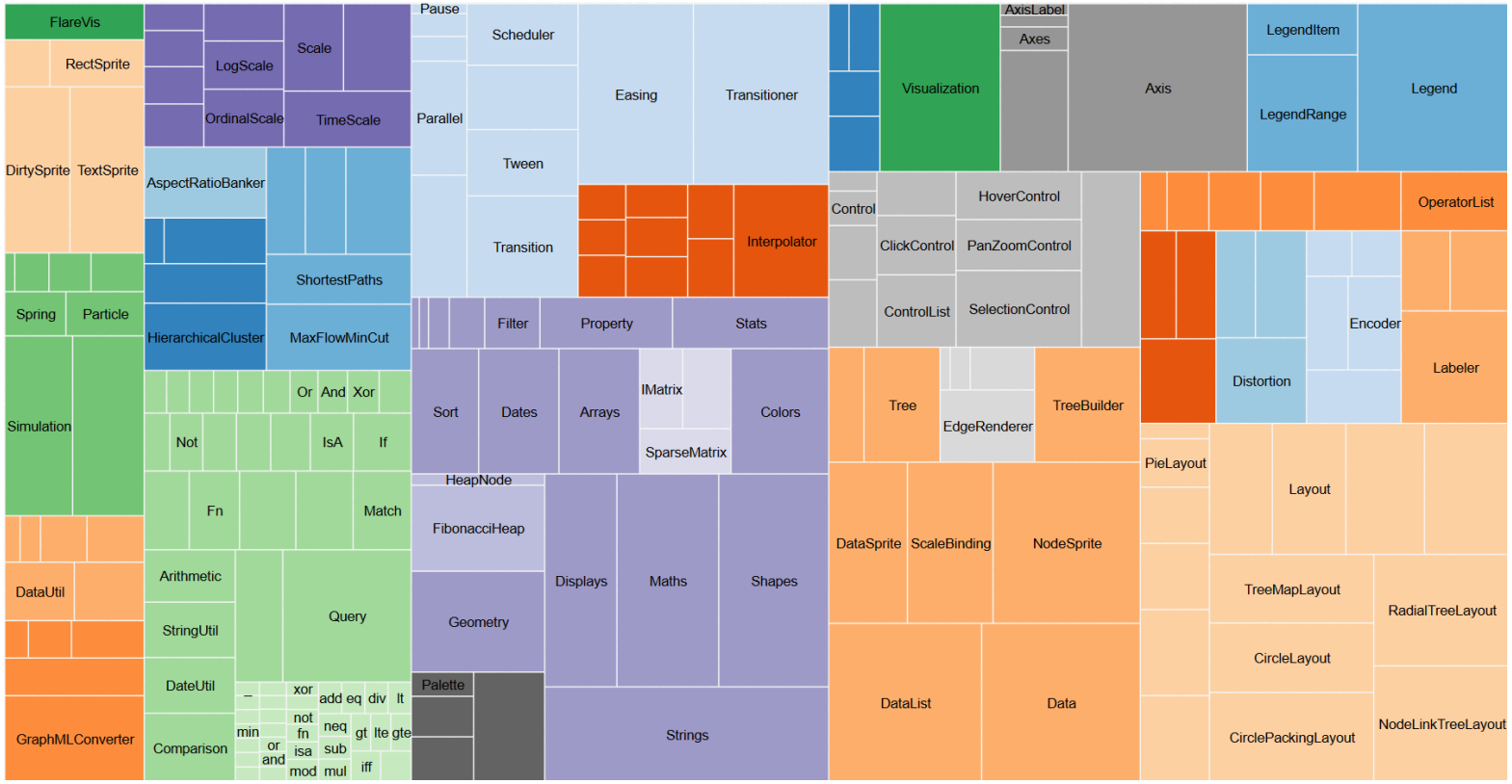
Display hierarchical data



https://github.com/mbostock/d3/wiki/Gallery

Tree map

Display hierarchical data



https://github.com/mbostock/d3/wiki/Gallery

Force-directed graph

Display graphs, networks, relationships

https://github.com/mbostock/d3/wiki/Gallery

Radial graph

Display graphs, networks, relationships

https://github.com/mbostock/d3/wiki/Gallery

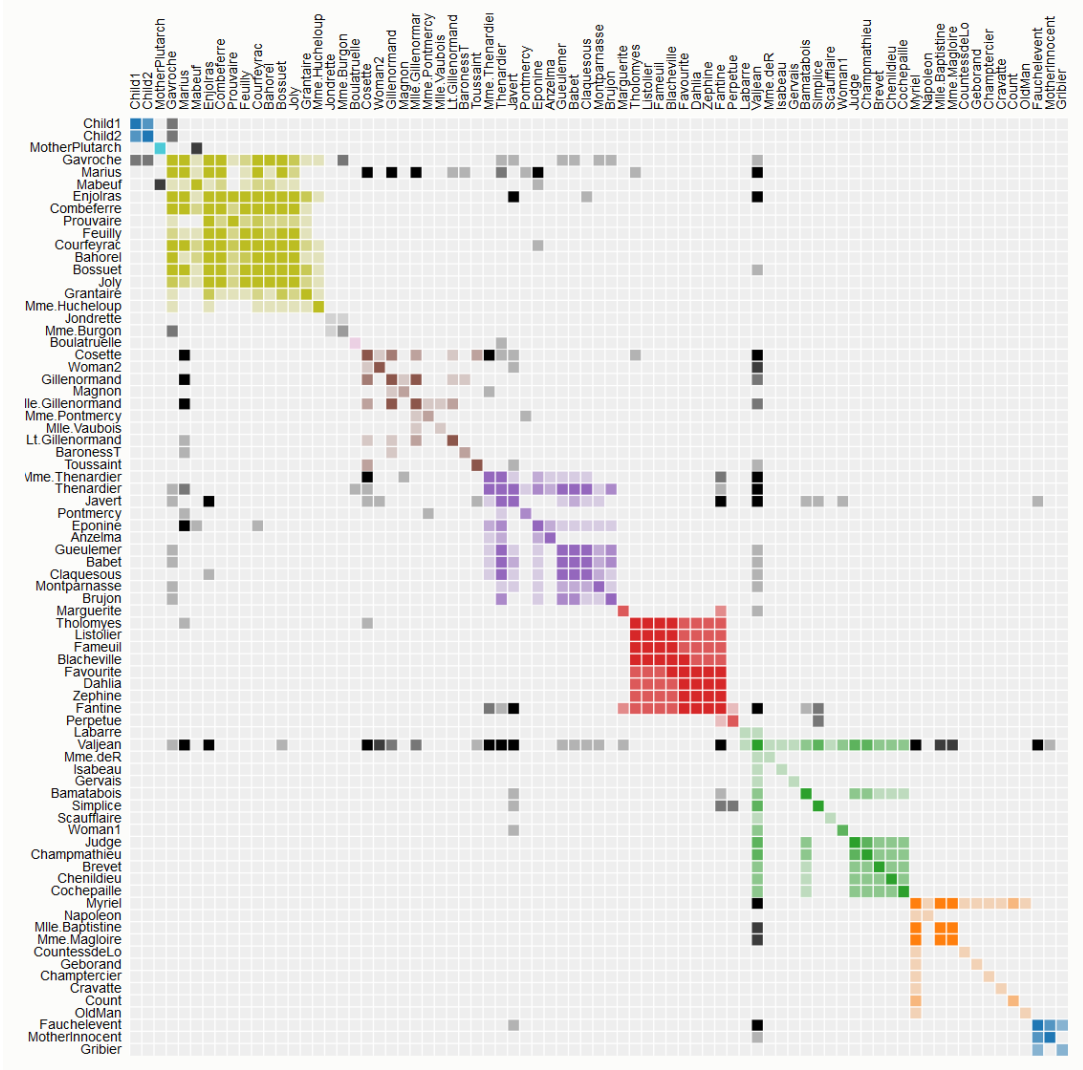
Radial graph

Display graphs, networks, relationships

https://github.com/mbostock/d3/wiki/Gallery

Co-occurrence matrix

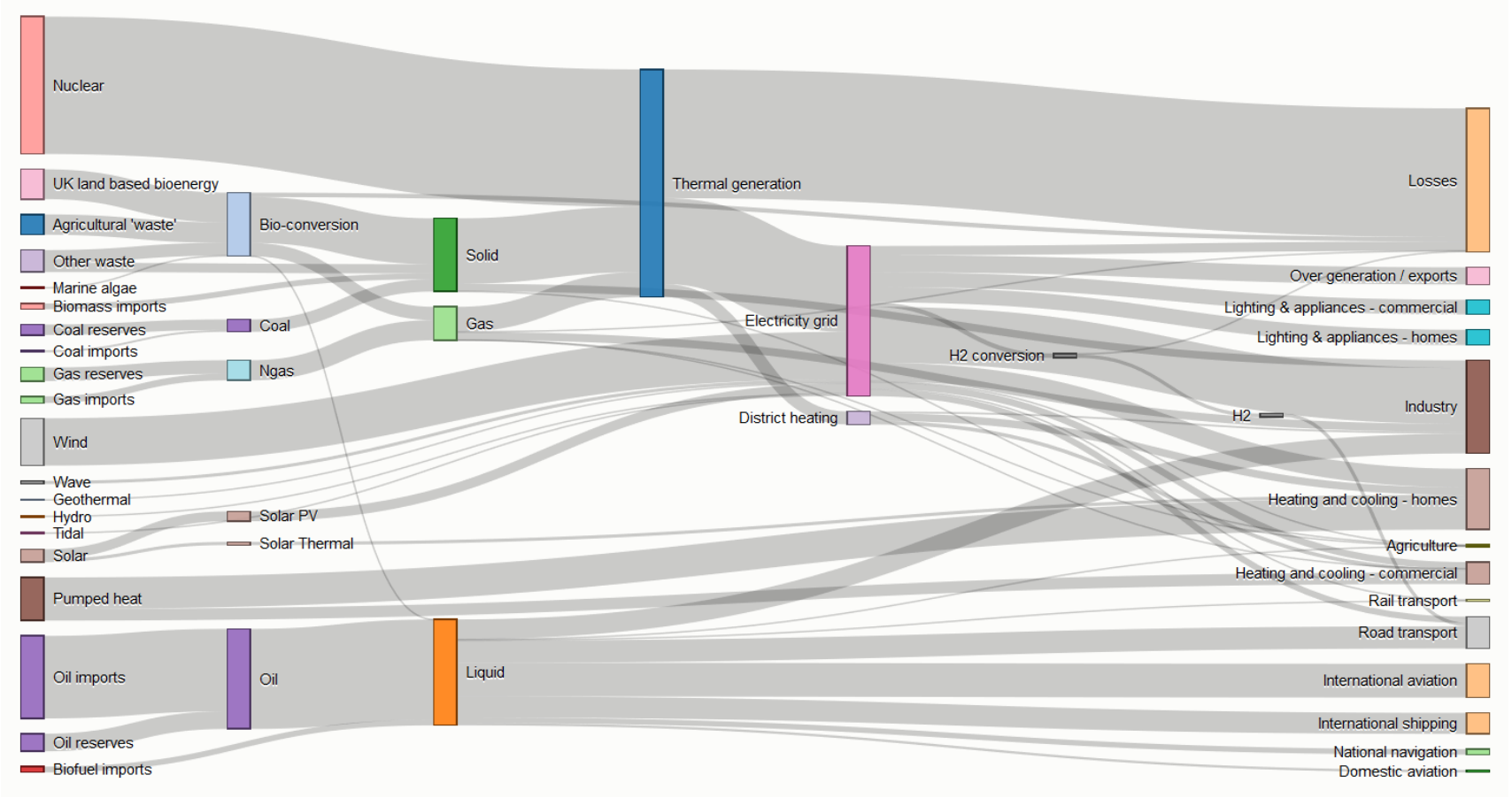
Display graphs, networks, relationships



https://github.com/mbostock/d3/wiki/Gallery

Sankey diagram

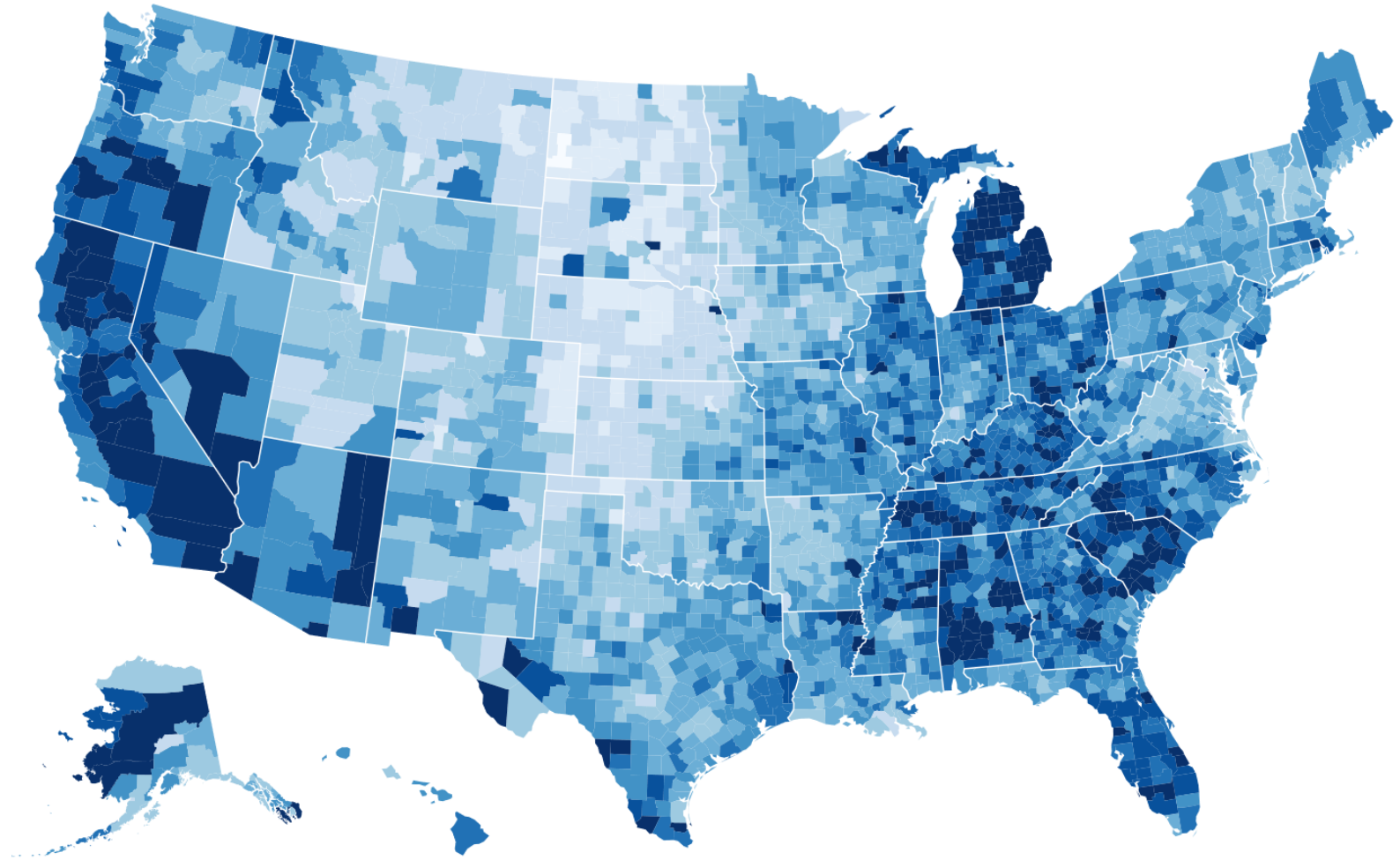
Display flow amongst entities



https://github.com/mbostock/d3/wiki/Gallery

Maps

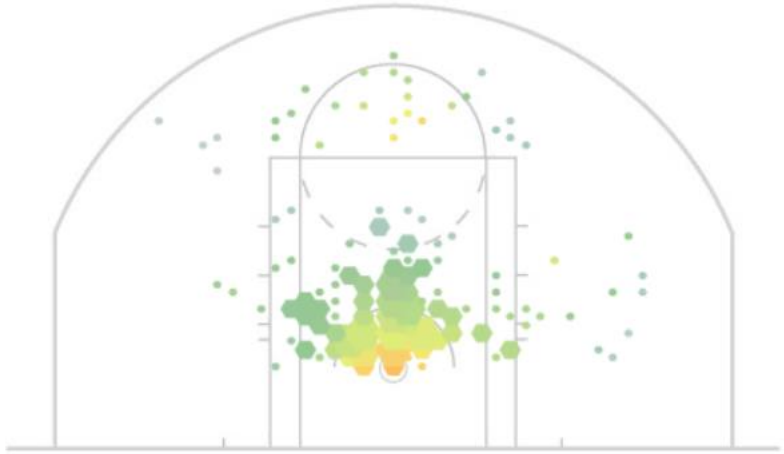
Display a spatial variable



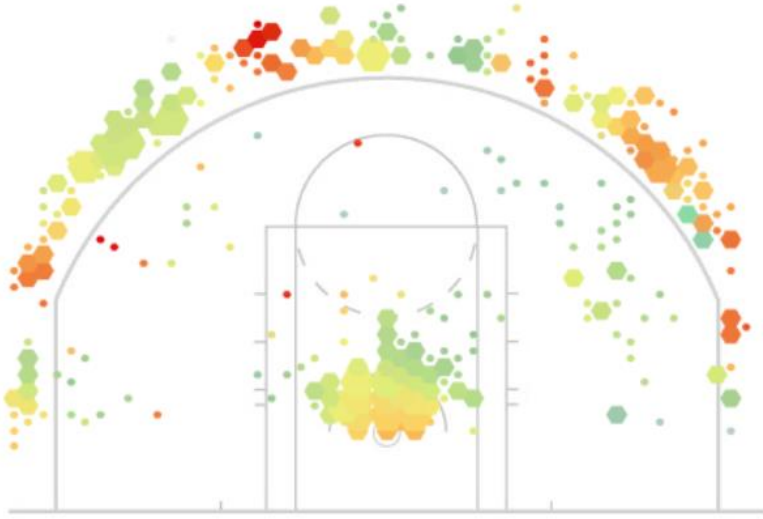
https://github.com/mbostock/d3/wiki/Gallery

Maps

Display a spatial variable



Kendrick Perkins



James Harden

[New York Times]

Expressiveness and Effectiveness

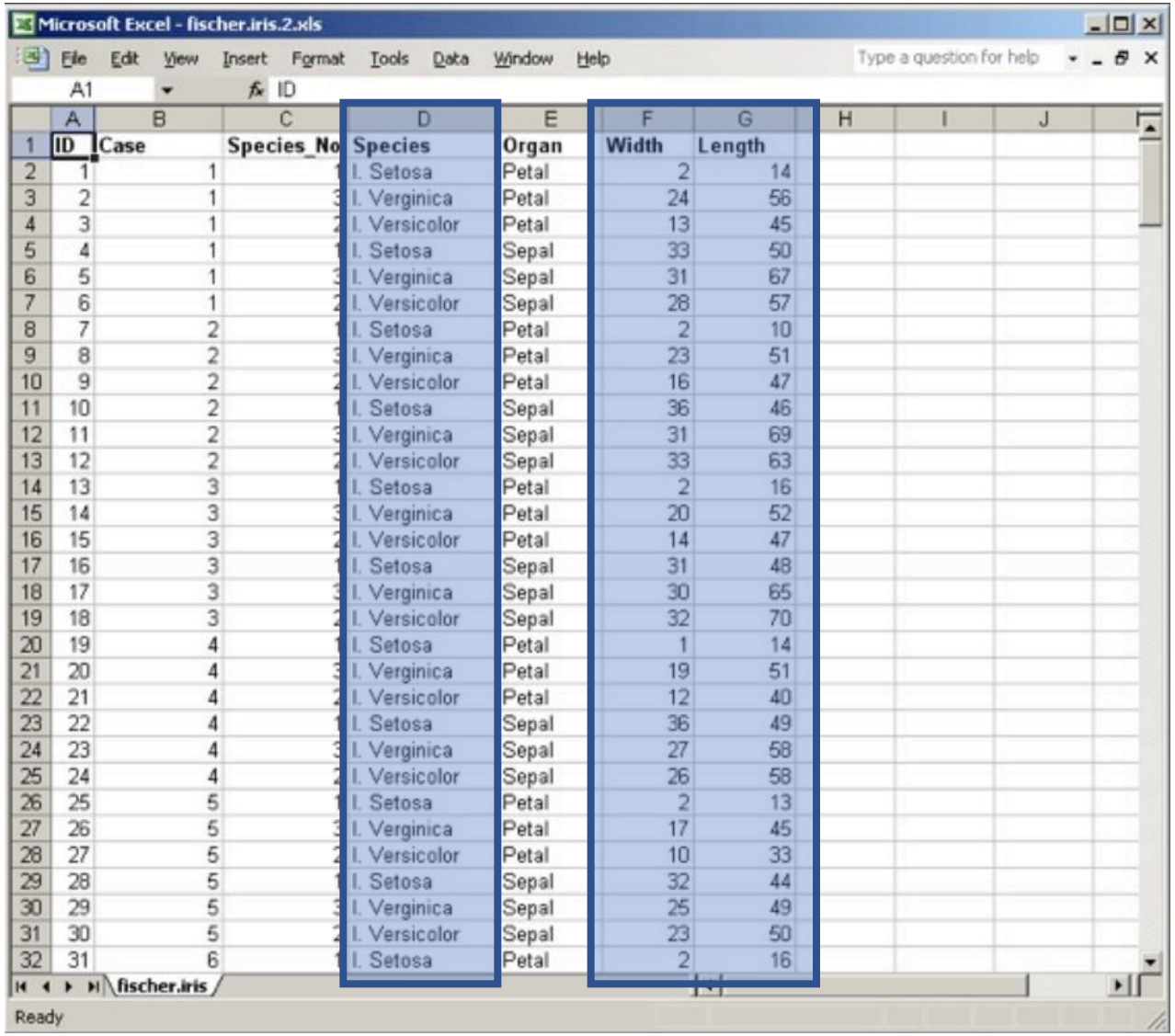
Expressiveness

A set of facts is expressible in a visual language if the sentences (i.e. the visualizations) in the language express all the facts in the set of data, and only the facts in the data.

Effectiveness

A visualization is more effective than another visualization if the information conveyed by one visualization is more readily perceived than the information in the other visualization.

[Mackinlay, APT, 1986]

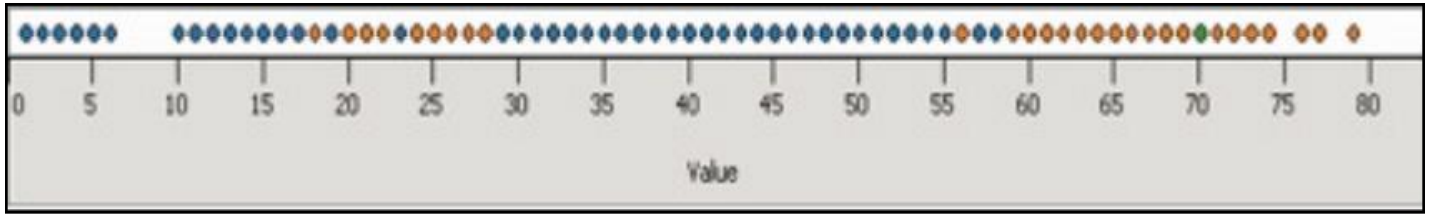


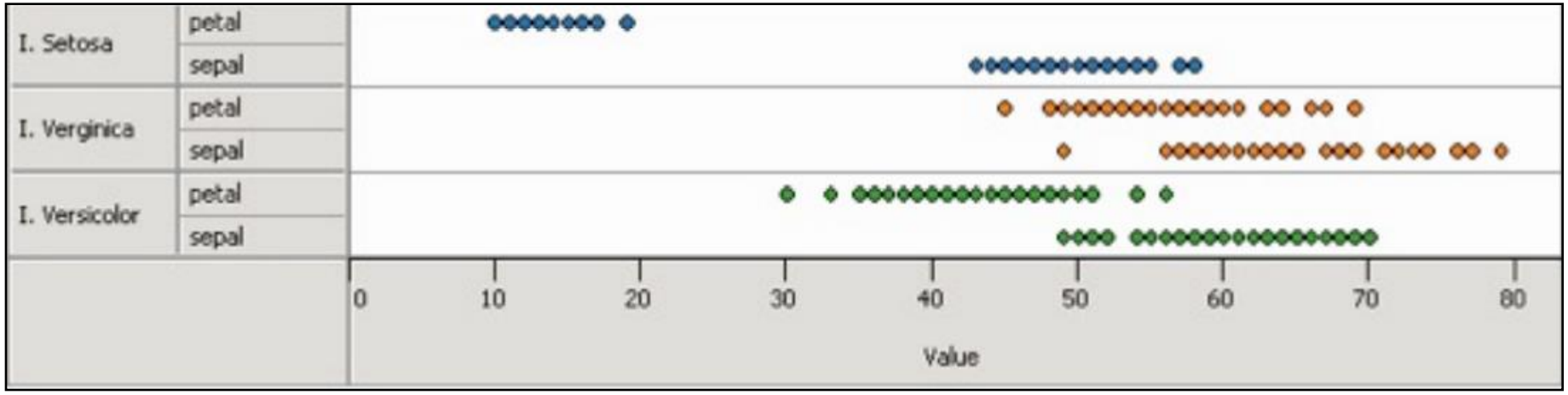
ID	Case	Species_No	Species	Organ	Width	Length
1	1	1	I. Setosa	Petal	2	14
2	1	3	I. Verginica	Petal	24	56
3	1	2	I. Versicolor	Petal	13	45
4	1	1	I. Setosa	Sepal	33	50
5	1	3	I. Verginica	Sepal	31	67
6	1	2	I. Versicolor	Sepal	28	57
7	2	1	I. Setosa	Petal	2	10
8	2	3	I. Verginica	Petal	23	51
9	2	2	I. Versicolor	Petal	16	47
10	2	1	I. Setosa	Sepal	36	46
11	2	3	I. Verginica	Sepal	31	69
12	2	2	I. Versicolor	Sepal	33	63
13	3	1	I. Setosa	Petal	2	16
14	3	3	I. Verginica	Petal	20	52
15	3	2	I. Versicolor	Petal	14	47
16	3	1	I. Setosa	Sepal	31	48
17	3	3	I. Verginica	Sepal	30	65
18	3	2	I. Versicolor	Sepal	32	70
19	4	1	I. Setosa	Petal	1	14
20	4	3	I. Verginica	Petal	19	51
21	4	2	I. Versicolor	Petal	12	40
22	4	1	I. Setosa	Sepal	36	49
23	4	3	I. Verginica	Sepal	27	58
24	4	2	I. Versicolor	Sepal	26	58
25	5	1	I. Setosa	Petal	2	13
26	5	3	I. Verginica	Petal	17	45
27	5	2	I. Versicolor	Petal	10	33
28	5	1	I. Setosa	Sepal	32	44
29	5	3	I. Verginica	Sepal	25	49
30	5	2	I. Versicolor	Sepal	23	50
31	6	1	I. Setosa	Petal	2	16

[Fisher, 1936]

Example 1: Cannot express the facts

A one-to-many (1->N) relation cannot be expressed in a single horizontal dot plot because multiple tuples are mapped to the same position.

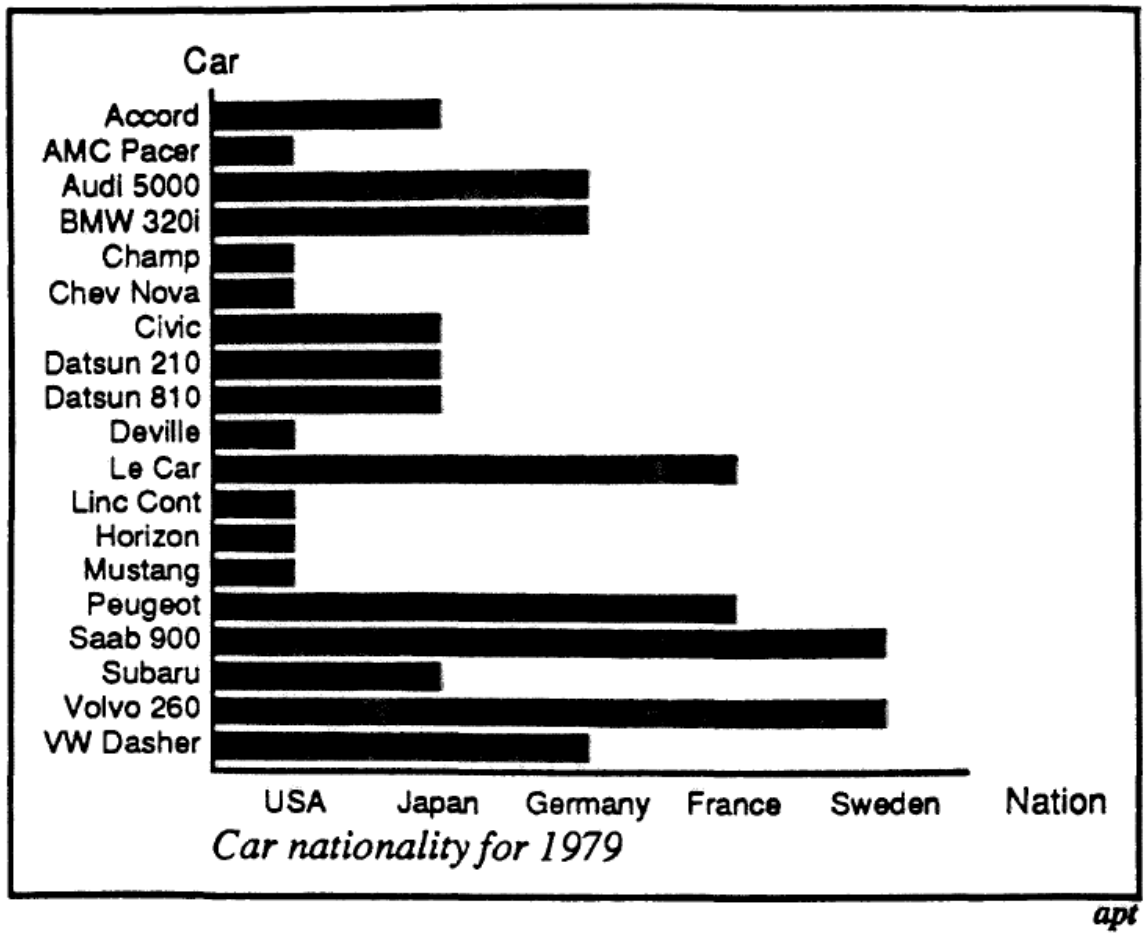


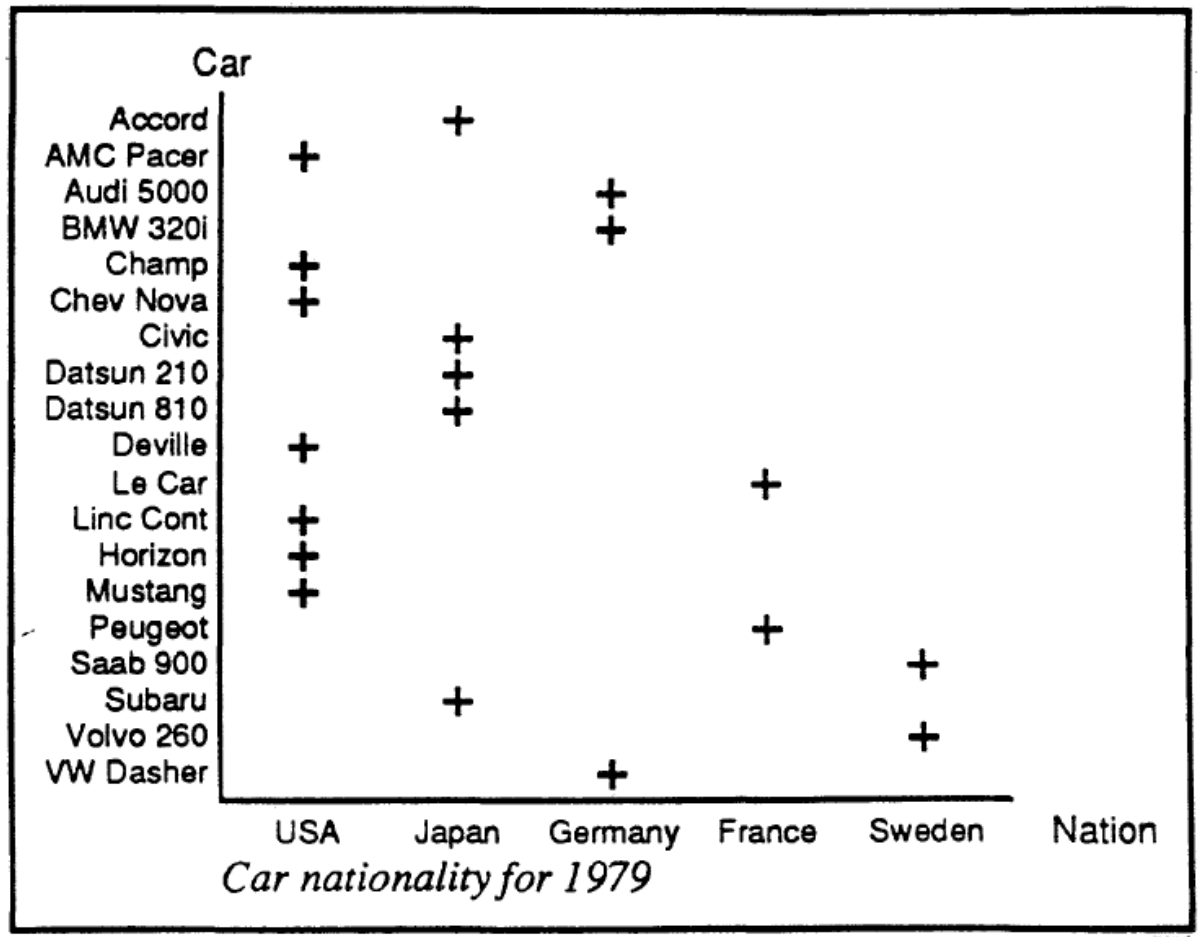


Example 2: Express facts not in the data

A length is interpreted as a quantitative value;

Length of bar says something untrue about Nominal data

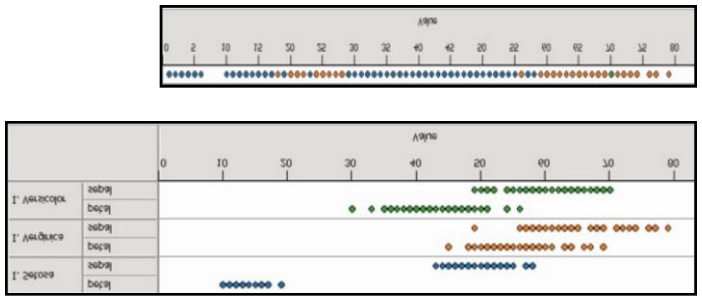


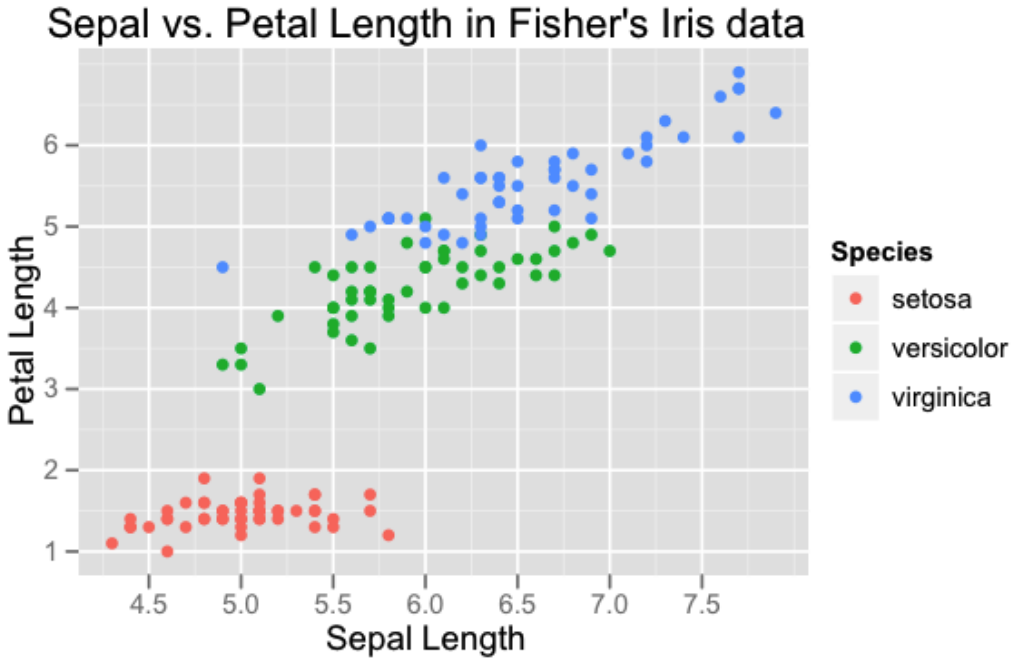


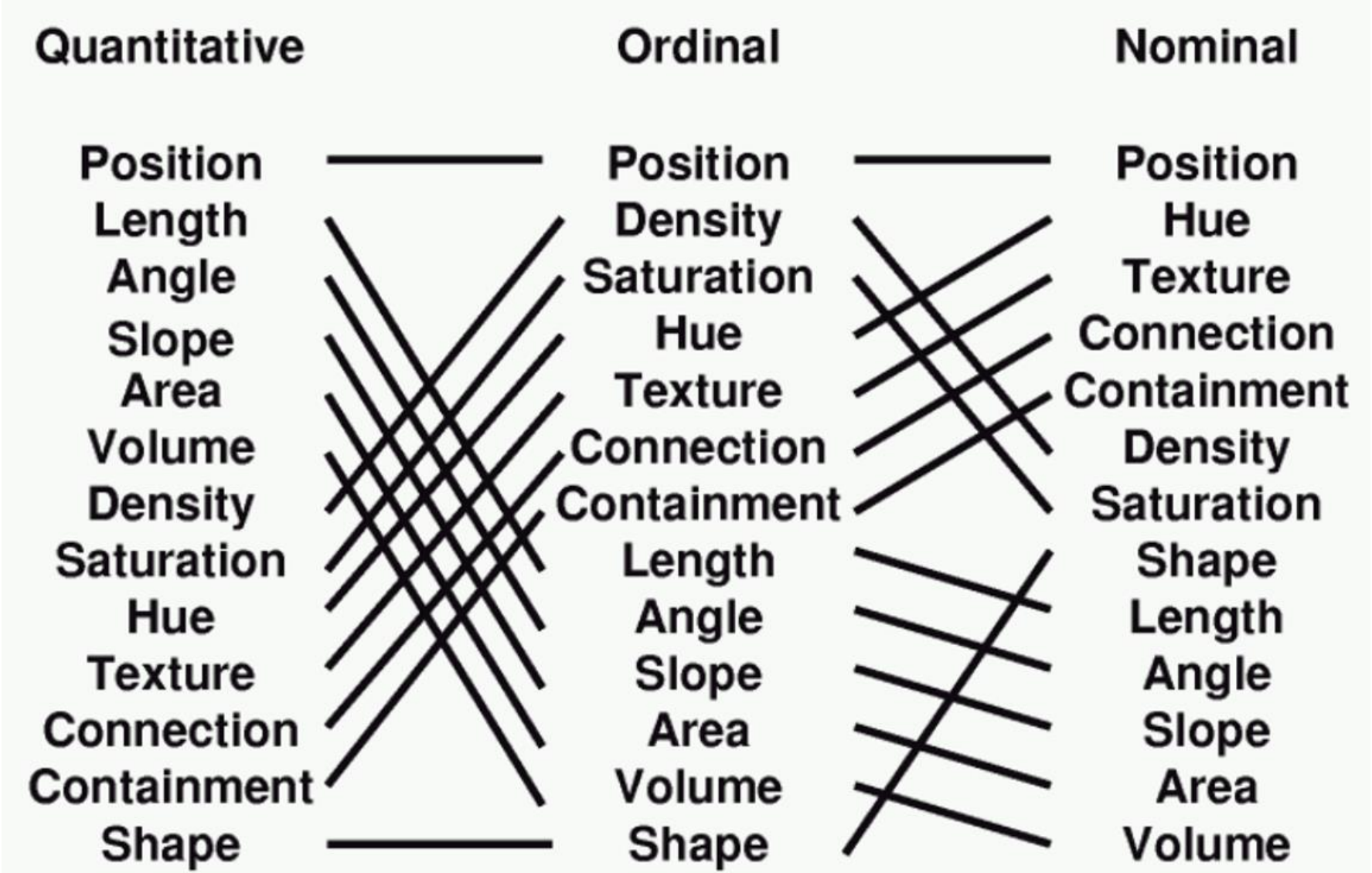
Car nationality for 1979

apt

Example 3: Effectiveness





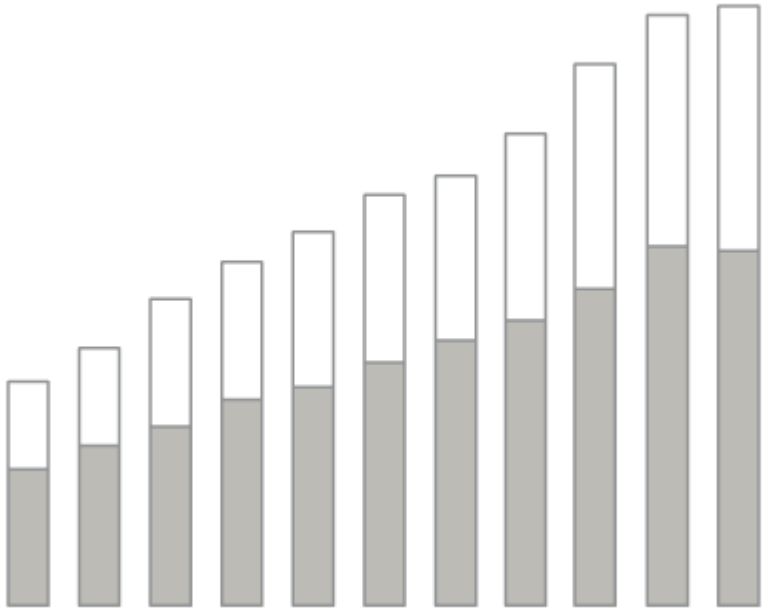

Quantitative
Ordinal
Nominal
Position
Length
Angle
Slope
Area
Volume
Density
Saturation
Hue
Texture
Connection
Containment
Shape
Position
Density
Saturation
Hue
Texture
Connection
Containment
Length
Angle
Slope
Area
Volume
Shape
Position
Hue
Texture
Connection
Containment
Density
Saturation
Shape
Length
Angle
Slope
Area
Volume

[Mackinlay]

Tufte’s Rules

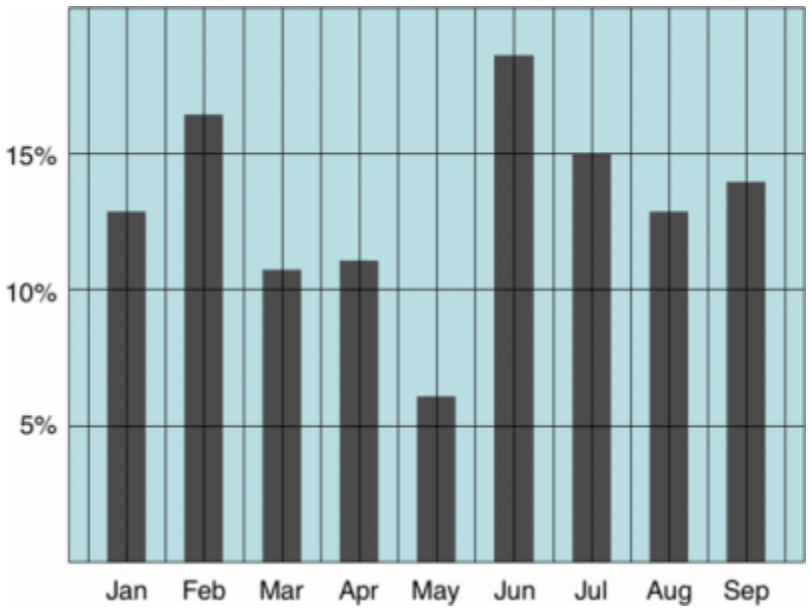
[http://www.sealthreinhold.com/tuftes-rules/rule_three.php]

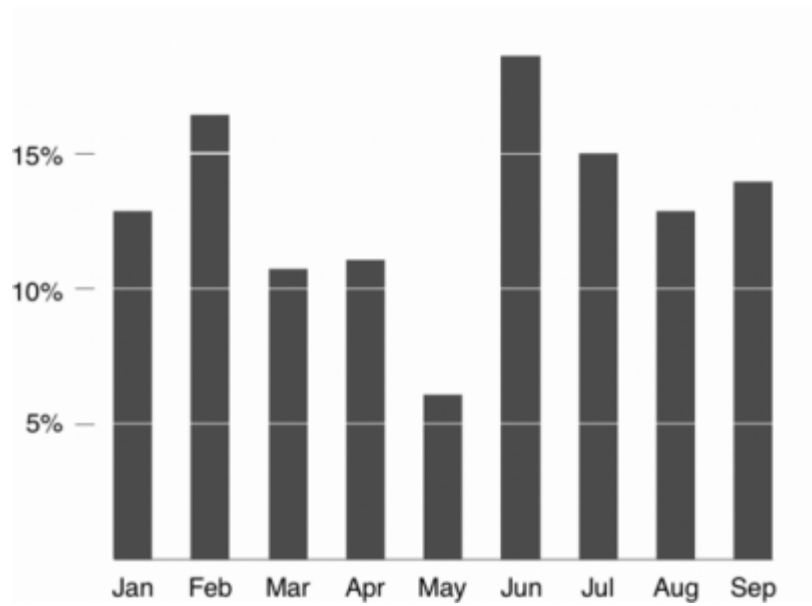
Avoid Chartjunk



Maximize Data-Ink Ratio

$$\text{Data-Ink Ratio} = \frac{\text{Data-Ink}}{\text{Total Ink used}}$$


15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep


15%
10%
5%
Jan
Feb
Mar
Apr
May
Jun
Jul
Aug
Sep

Interaction

(Demos)

Animation

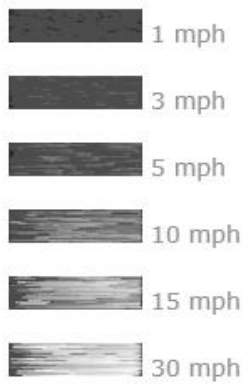
Why Use Animation?

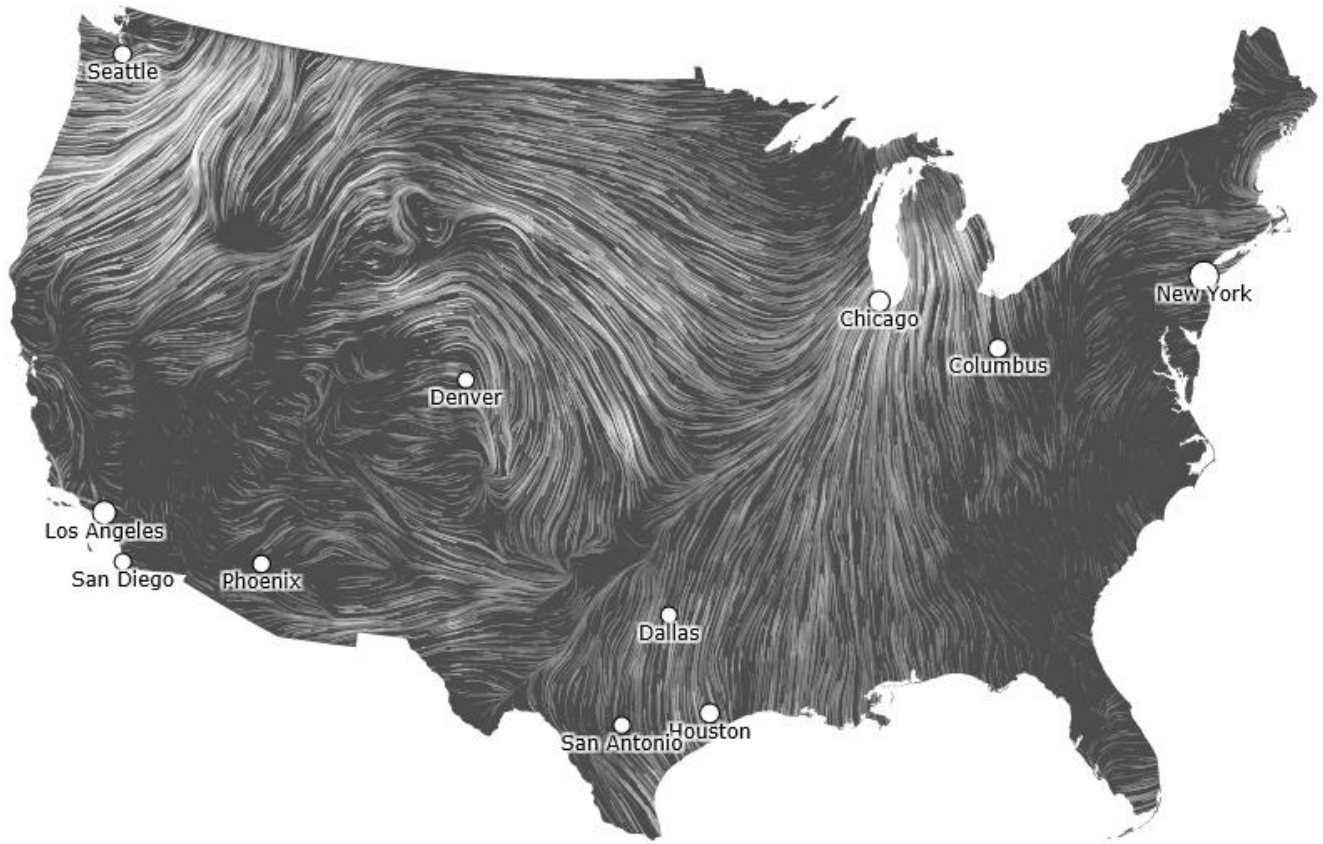
- Visual variable to encode data
- Direct attention
- Understand system dynamics
- Understand state transition (maintain context)
- Increase engagement

wind map

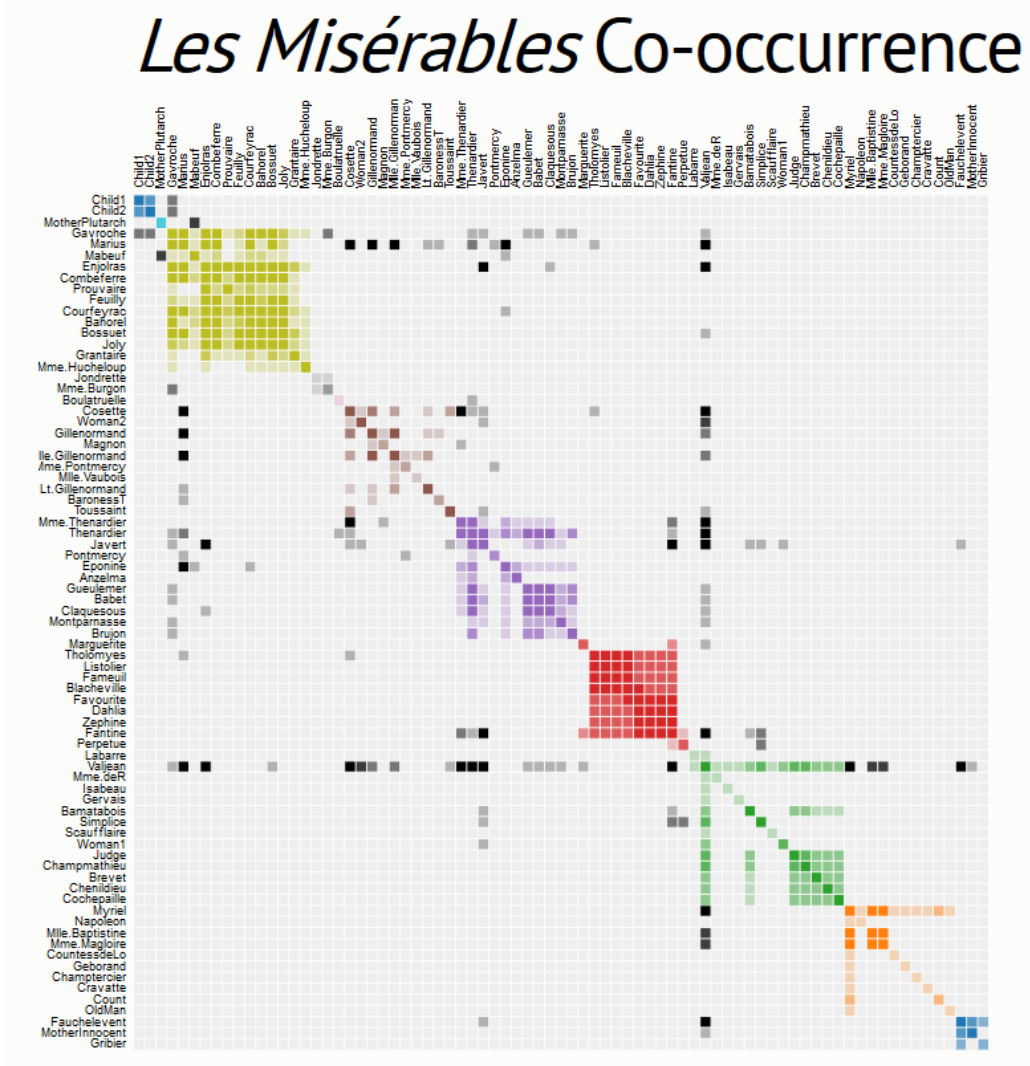
February 18, 201⁴
6:55 pm EST
(time of forecast download)

top speed: **33.7 mph**
average: **10.3 mph**





http://hint.fm/wind/



http://bost.ocks.org/mike/miserables/

(Video)

[Drucker et. al. http://research.microsoft.com/en-us/projects/sanddance/]

(Video)

[Heer and Robertson http://vis.berkeley.edu/papers/animated_transitions/]

Implementation Tips

- D3 is great (https://d3js.org)
  - Well documented, hundreds of examples (e.g. https://github.com/d3/d3/wiki/Gallery)
  - Portable because of open standards (HTML/SVG, CSS)
  - Support for interactions
- For Python, use seaborn (https://seaborn.pydata.org)
- For interactive visualizations with heavy render load, leverage the GPU
  - E.g. using PIXIJS (http://www.pixijs.com)

Courses

- Maneesh Agrawala
  - http://vis.berkeley.edu/courses/cs294-10-fa14/wiki/index.php/Main_Page
- Jeff Heer
  - http://courses.cs.washington.edu/courses/cse512/14wi/
- John Stasko
  - http://www.cc.gatech.edu/~stasko/7450/

Blogs & Websites

- https://www.reddit.com/r/dataisbeautiful/
- http://fivethirtyeight.com/
- http://flowingdata.com/
- http://www.informationisbeautiful.net/
- http://infosthetics.com/
- http://junkcharts.typepad.com/
- http://datavisualization.ch/
- http://eagereyes.org/
- http://blog.okcupid.com/
- https://twitter.com/nytgraphics

Articles & Others

- Tufte: The Visual Display of Quantitative Information
  - http://www.edwardtufte.com/tufte/
- http://www.csc.ncsu.edu/faculty/healey/PP/index.html
- http://www.sealthreinhold.com/tuftes-rules/rule_three.php
- Ted Talk: Beauty of Data Visualization
  - https://www.youtube.com/watch?v=pLqjQ55tz-U
- http://piksels.com/wp-content/uploads/2009/01/visualizingdata.pdf
- http://homes.cs.washington.edu/~jheer/files/zoo/
- http://www.targetprocess.com/articles/visual-encoding.html
- http://en.wikipedia.org/wiki/Misleading_graph