

Visualization

Emanuel Zgraggen

Many slides borrowed from Drucker, Agrawala, Heer, Stasko, etc. References at the end.

A DAY IN DATA

The exponential growth of data is undisputed, but the numbers behind this explosion – fuelled by internet of things and the use of connected devices – are hard to comprehend, particularly when looked at in the context of one day

500m
tweets are sent every day
Twitter



4PB

of data created by Facebook, including

350m photos

100m hours of video watch time
Facebook Research

294bn

billion emails are sent
Radical Group

320bn

emails to be sent each day by 2021

306bn

emails to be sent each day by 2020



4TB

of data produced by a connected car
Intel

65bn

messages sent over WhatsApp and two billion minutes of voice and video calls made
Facebook



463EB

of data will be created every day by 2025
IIC

95m

photos and videos are shared on Instagram
Instagram Business



3.9bn

people use emails



Searches made a day

5bn

Searches made a day from Google

3.5bn



ACCUMULATED DIGITAL UNIVERSE OF DATA

4.4ZB

44ZB

PwC

2019

2020

DEMYSTIFYING DATA UNITS

From the more familiar 'bit' or 'megabyte', larger units of measurement are more frequently being used to explain the masses of data

Unit	Value	Size
b bit	0 or 1	1/8 of a byte
B byte	8 bits	1 byte
KB kilobyte	1,000 bytes	1,000 bytes
MB megabyte	1,000 ² bytes	1,000,000 bytes
GB gigabyte	1,000 ³ bytes	1,000,000,000 bytes
TB terabyte	1,000 ⁴ bytes	1,000,000,000,000 bytes
PB petabyte	1,000 ⁵ bytes	1,000,000,000,000,000 bytes
EB exabyte	1,000 ⁶ bytes	1,000,000,000,000,000,000 bytes
ZB zettabyte	1,000 ⁷ bytes	1,000,000,000,000,000,000,000 bytes
YB yottabyte	1,000 ⁸ bytes	1,000,000,000,000,000,000,000,000 bytes

*A lowercase "b" is used as an abbreviation for bits, while an uppercase "B" represents bytes.

28PB

to be generated from wearable devices by 2020
Statista



RACONTEUR

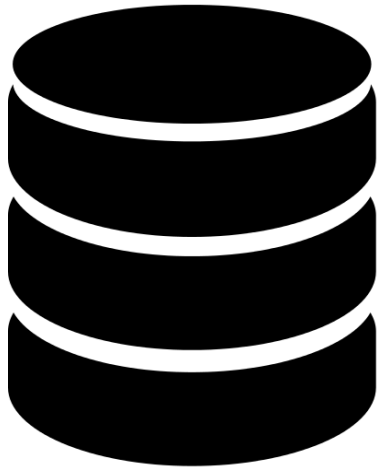
How do we make sense of all this data?

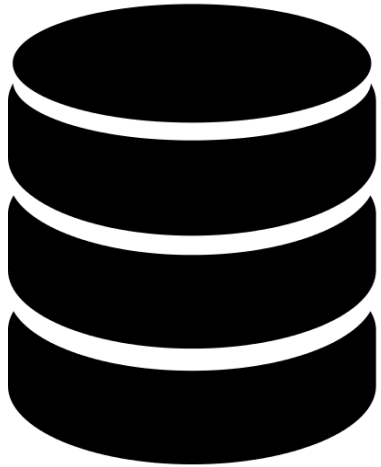
How do we use data in decision-making processes?

How do we avoid being overwhelmed?

Challenge

Transform the data into understanding and insight thus making it useful to people





Sight
Hearing
Taste →
Smell
Touch





Human Vision

Highest bandwidth sense

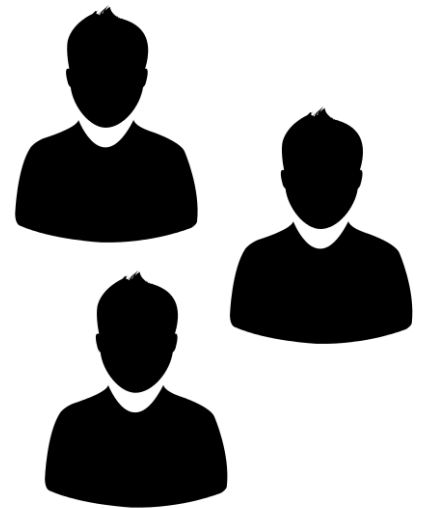
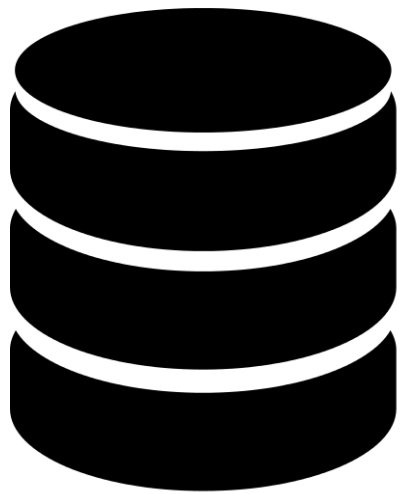
Fast, parallel

Pattern recognition

Pre-attentive

Extends memory and cognitive capacity

People think visually



Why create visualizations?

I

x	y
10	8.04
8	6.95
13	7.58
9	8.81
11	8.33
14	9.96
6	7.24
4	4.26
12	10.84
7	4.82
5	5.68

II

x	y
10	9.14
8	8.14
13	8.74
9	8.77
11	9.26
14	8.1
6	6.13
4	3.1
12	9.13
7	7.26
5	4.74

III

x	y
10	746
8	6.77
13	12.74
9	7.11
11	7.81
14	8.84
6	6.08
4	5.39
12	8.15
7	642
5	5.73

IV

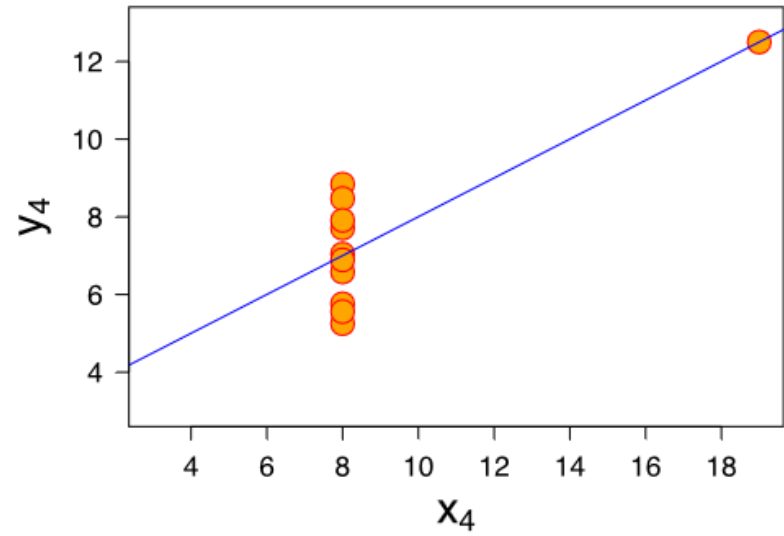
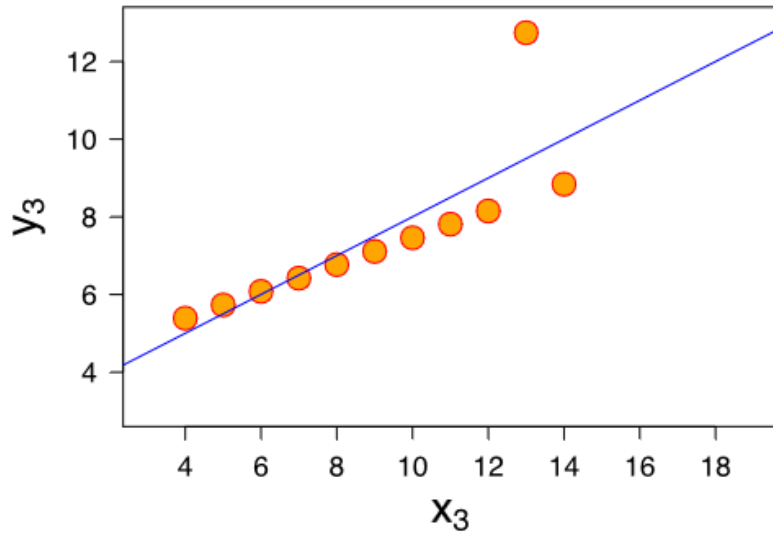
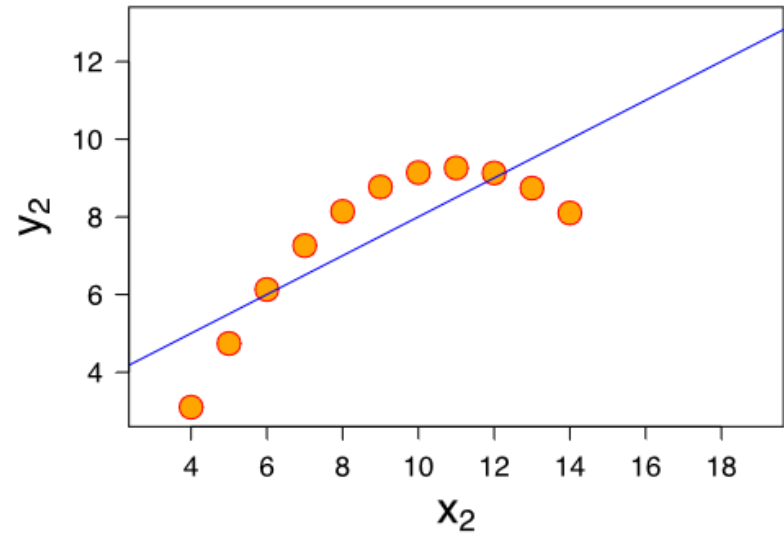
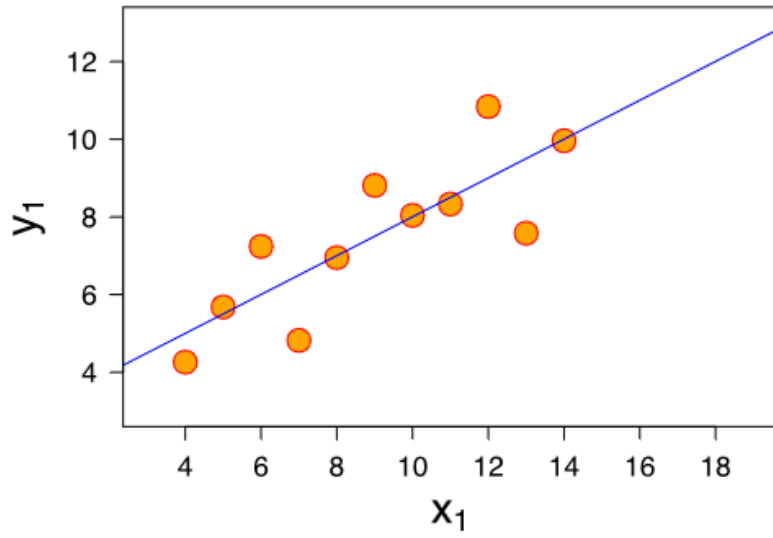
x	y
8	6.58
8	5.76
8	7.71
8	8.84
8	847
8	7.04
8	5.25
19	12.5
8	5.56
8	7.91
8	6.89

[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

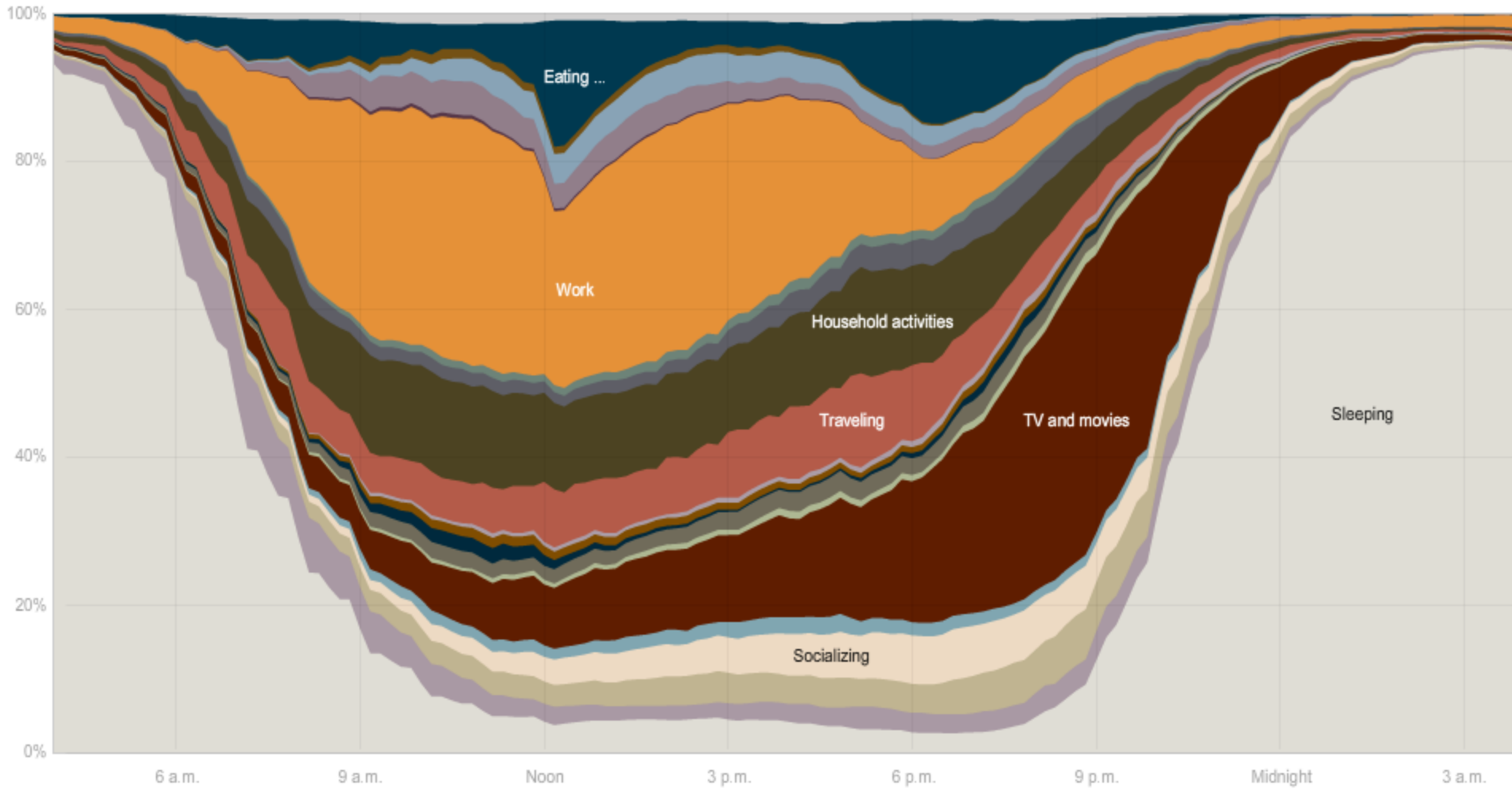


[Anscombe's Quartet]

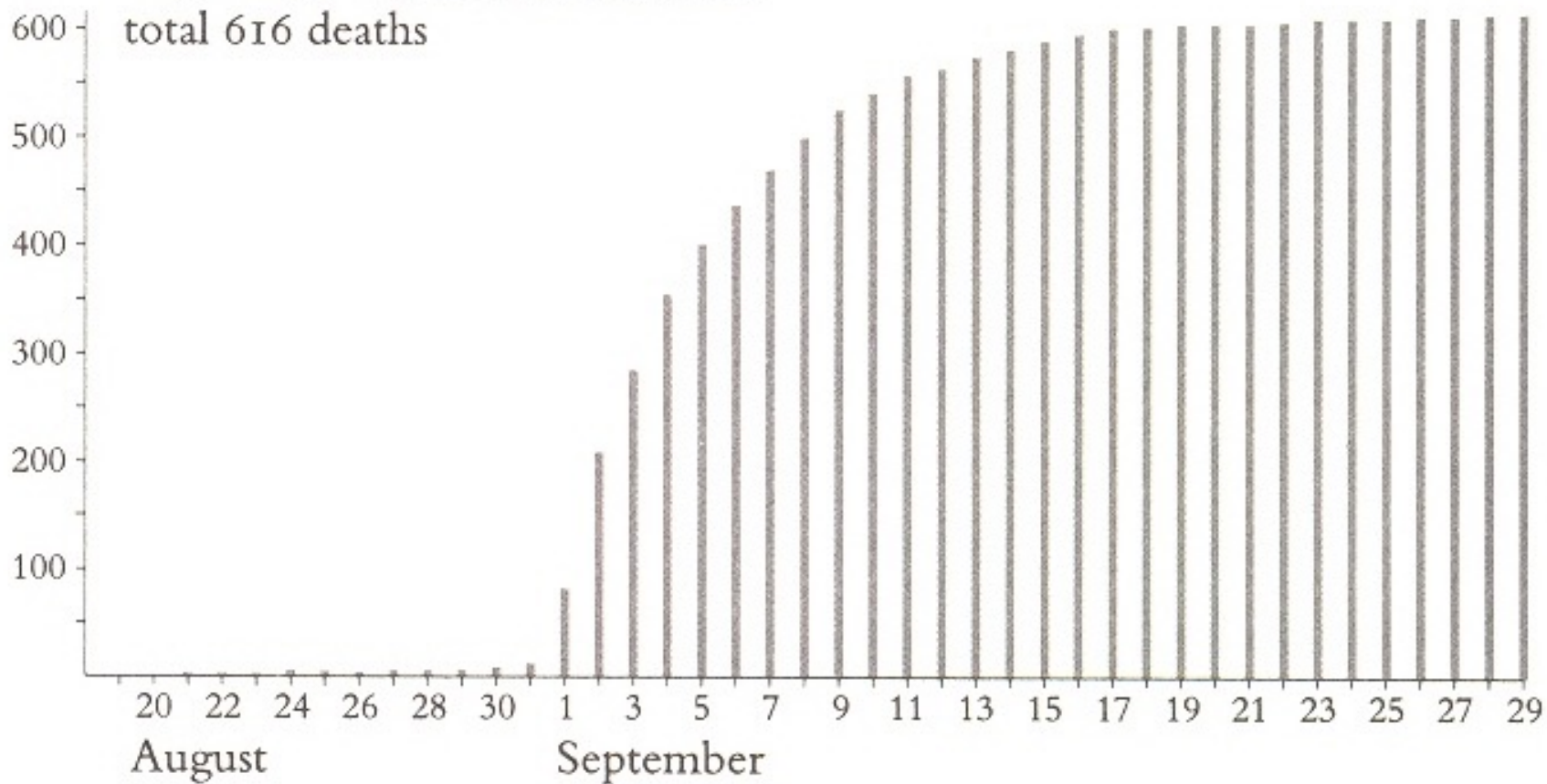
Everyone

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

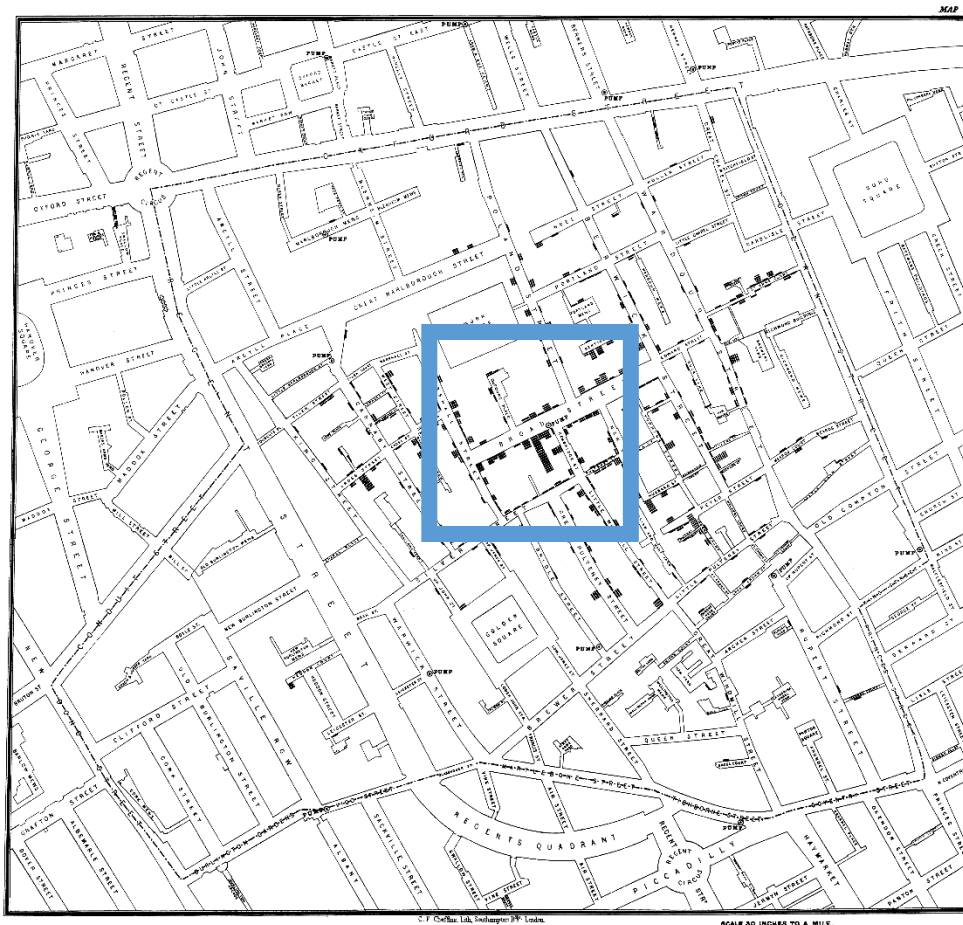
Everyone	Employed	White	Age 15-24	H.S. grads	No children
Men	Unemployed	Black	Age 25-64	Bachelor's	One child
Women	Not in lab...	Hispanic	Age 65+	Advanced	Two+ children

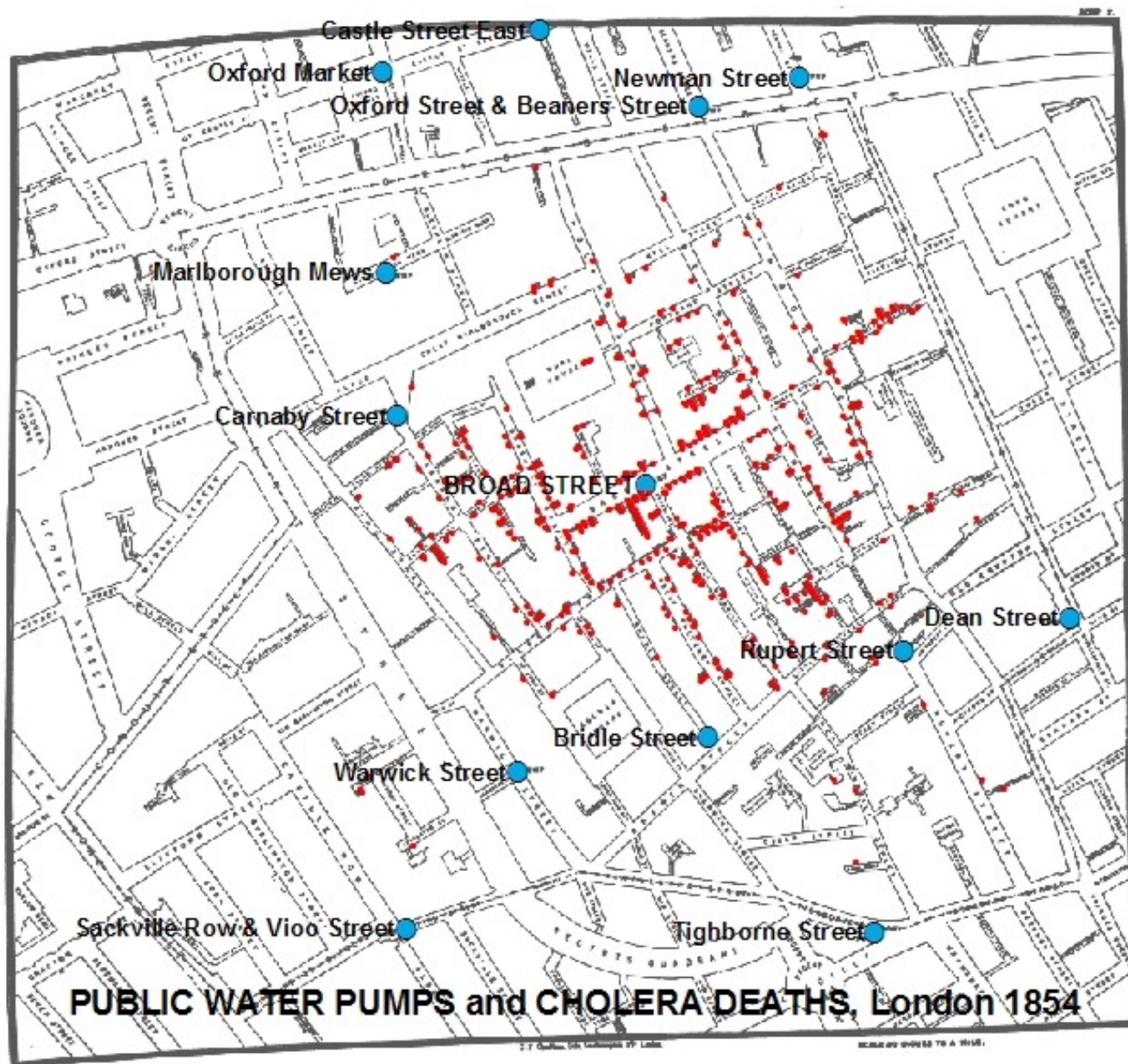


700 - Cumulative deaths from cholera,
beginning August 19, 1854; final
total 616 deaths

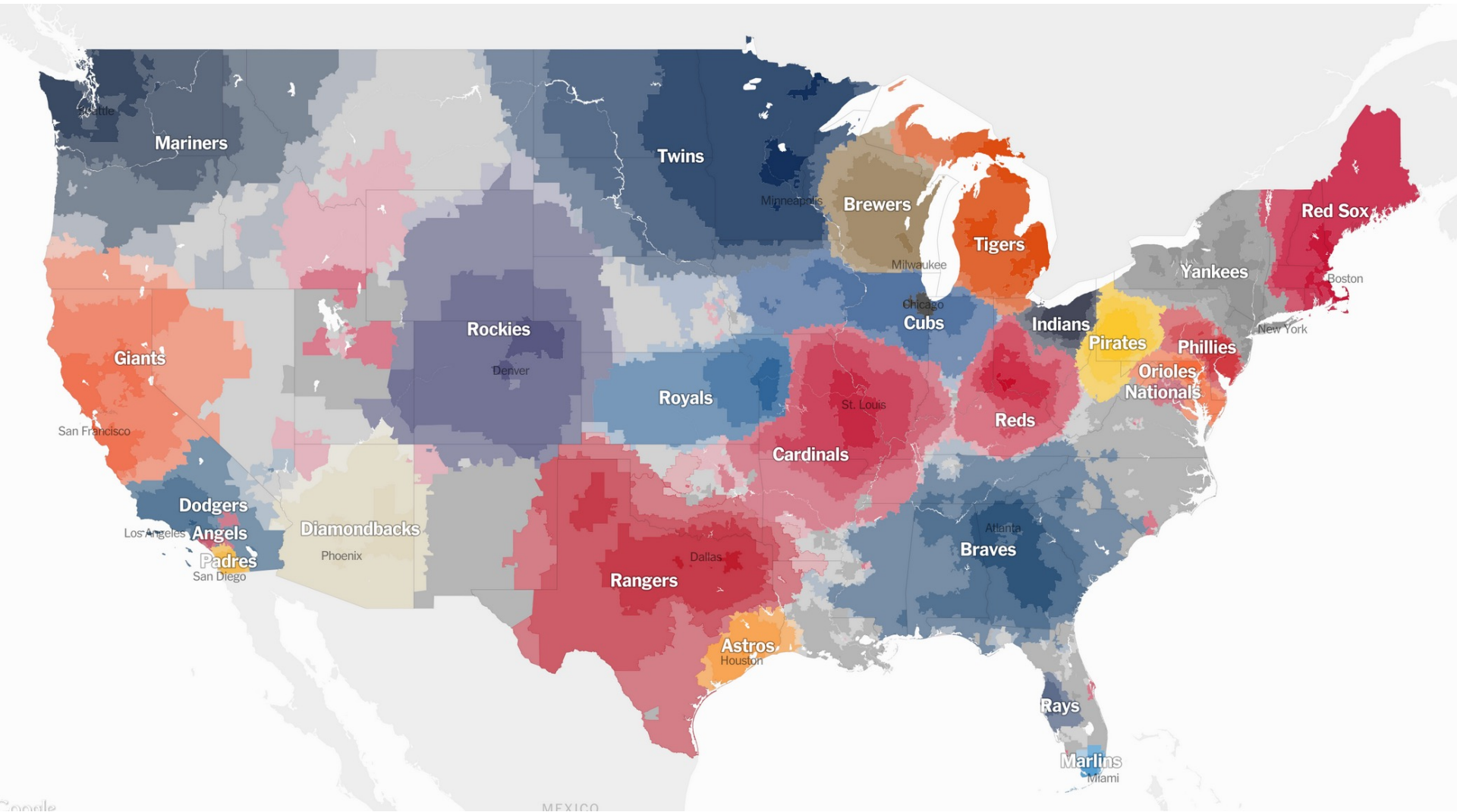


See data in context





See data in context



[New York Times]

Carte Figurative des pertes successives en hommes de l'Armée Française dans la campagne de Russie 1812-1813.

Dressée par M. Minard, Inspecteur Général des Ponts et Chaussées en retraite Paris, le 20 Novembre 1869.

Les nombres d'hommes présents sont représentés par les largeurs des zones colorées à raison d'un millimètre pour dix mille hommes; ils sont de plus écrits en travers des zones. Le rouge désigne les hommes qui ont péri en Russie, le noir ceux qui en sont sortis. Les renseignements qui ont servi à dresser la carte ont été puisés dans les ouvrages de M. M. Chiers, de Légar, de Fezensac, de Chambray et le journal inédit de Jacob, pharmacien de l'Armée depuis le 28 Octobre. Pour mieux faire juger à l'œil la diminution de l'armée, j'ai supposé que les corps du Prince Jérôme et du Maréchal Davout, qui avaient été détachés sur Minsk et Mohilow et qui rejoignirent vers Orscha et Witebsk, avaient toujours marché avec l'armée.

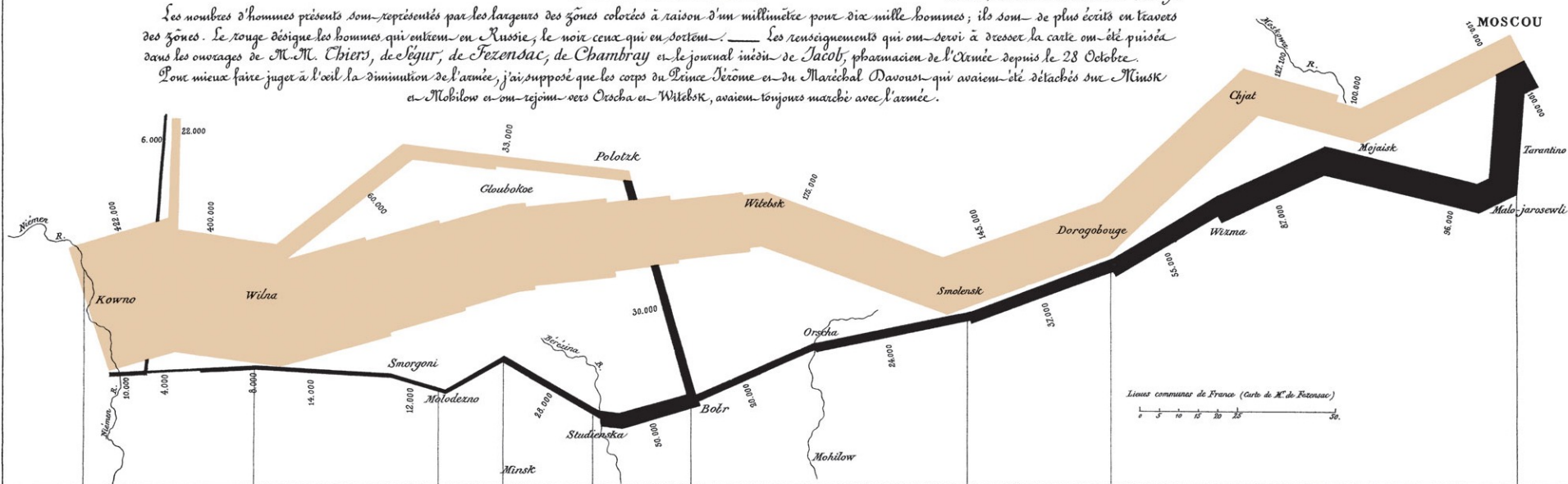
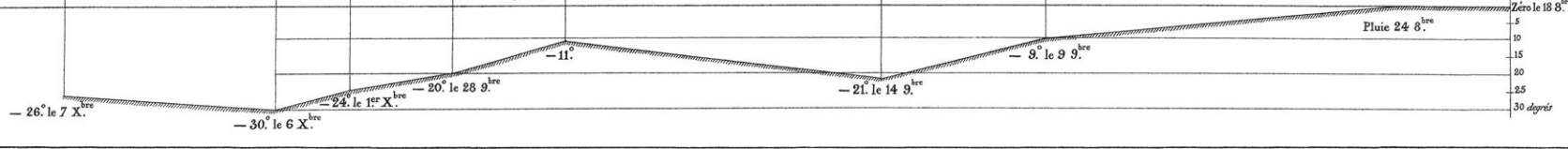


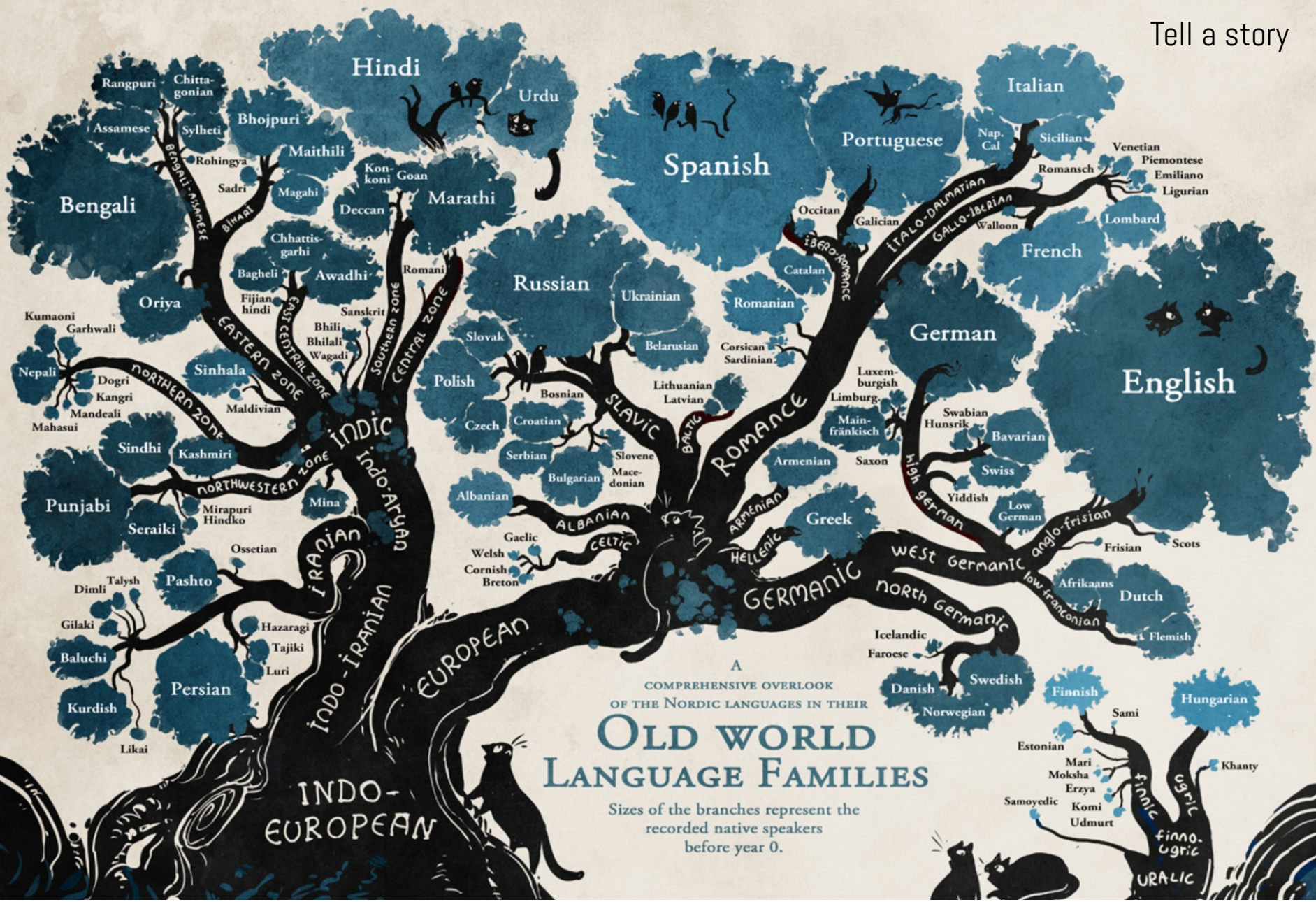
TABLEAU GRAPHIQUE de la température en degrés du thermomètre de Réaumur au dessous de zéro.

Les Cosaques passent au galop le Niéme gelé.



Autog. par Ragnier, 8, Rue S^{te} Marie S^o 6² à Paris.

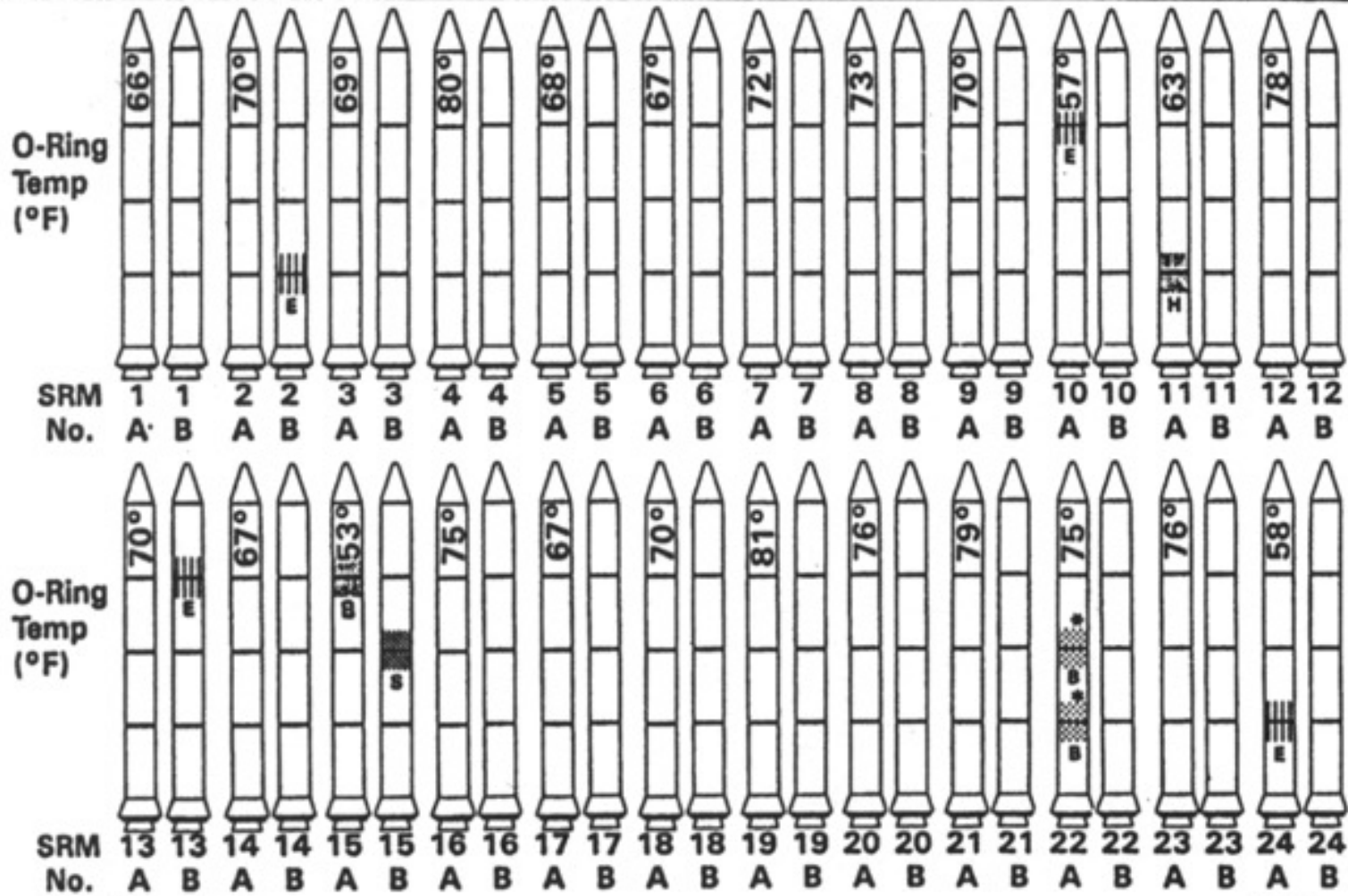
Imp. Lith. Ragnier et Douvdat.



A
 COMPREHENSIVE OVERLOOK
 OF THE NORDIC LANGUAGES IN THEIR
**OLD WORLD
 LANGUAGE FAMILIES**

Sizes of the branches represent the recorded native speakers before year 0.

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



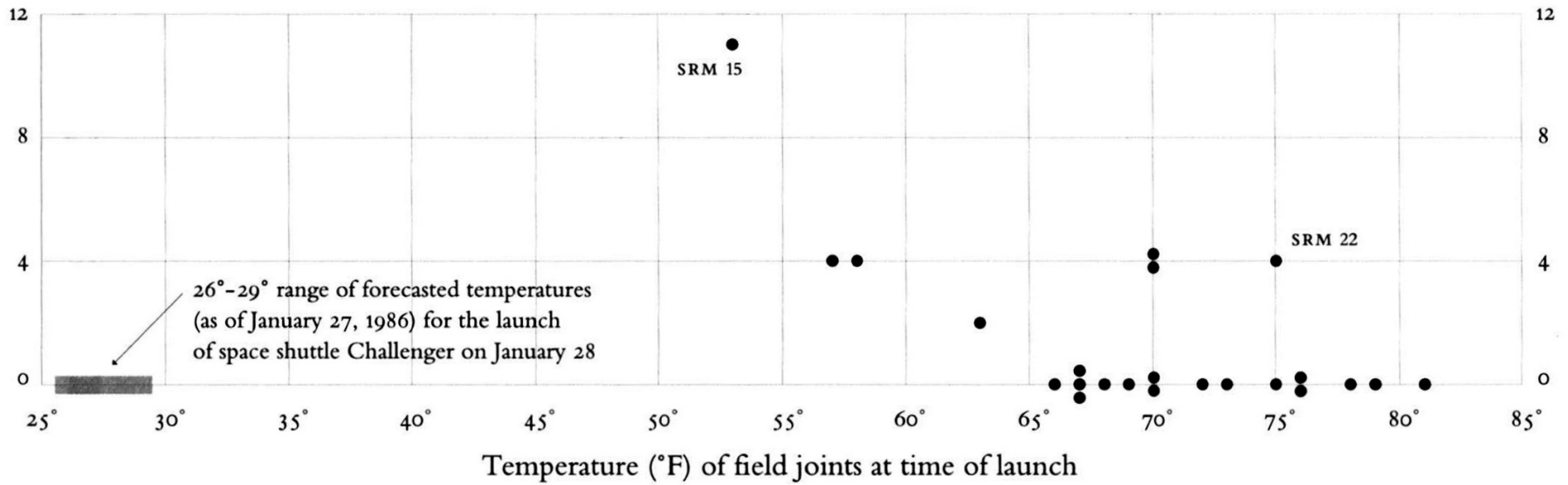
MORTON THIOKOL, INC.
Wasatch Operations

* No Erosion

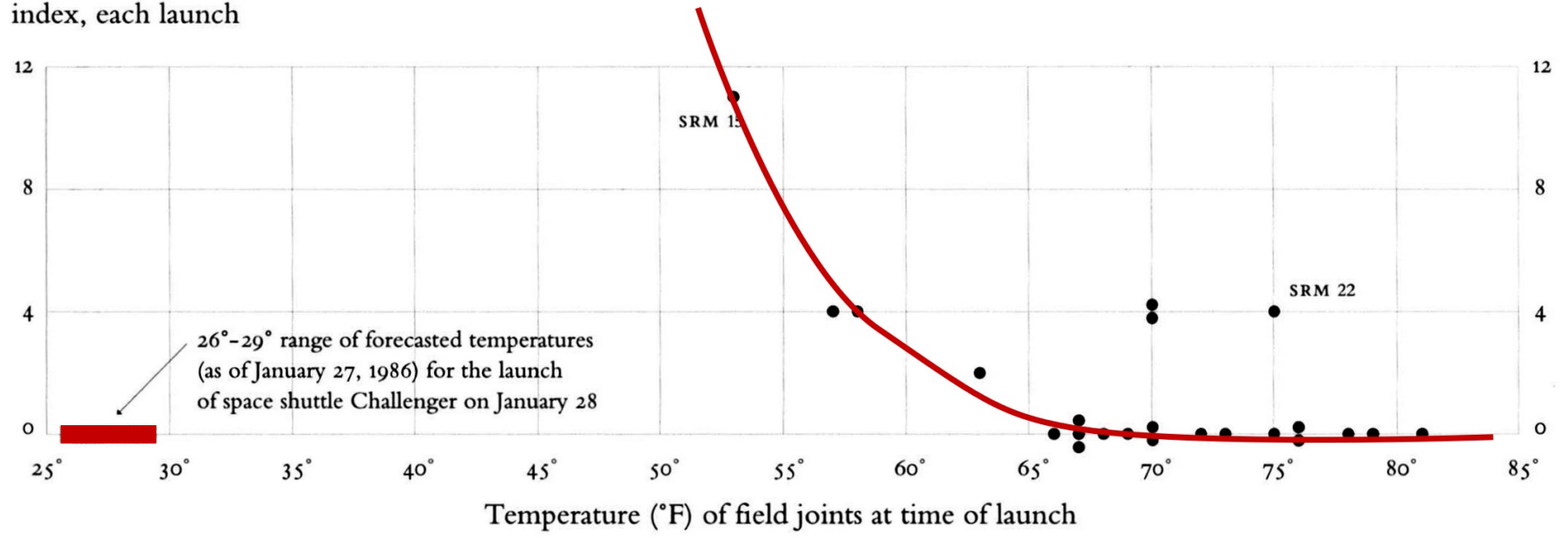
50400-14

INFORMATION ON THIS PAGE WAS PREPARED TO SUPPORT AN ORAL PRESENTATION AND CANNOT BE CONSIDERED COMPLETE WITHOUT THE ORAL DISCUSSION

O-ring damage index, each launch

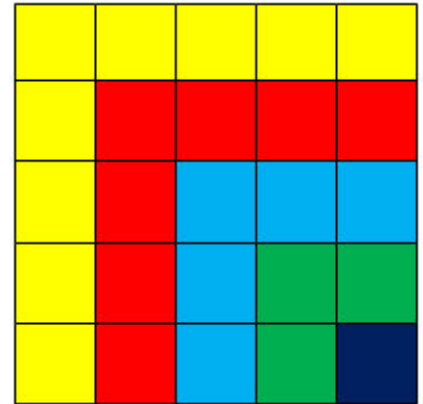


O-ring damage index, each launch



Sum of odd numbers:

$$1 + 3 + 5 + 7 + 9 = 5^2$$



Expectation

The expectation of a random variable is a number that attempts to capture the center of that random variable's distribution. It can be interpreted as the long-run average of many independent samples from the given distribution. More precisely, it is defined as the probability-weighted sum of all possible values in the random variable's support,

$$E[X] = \sum_{x \in \mathcal{X}} xP(x)$$

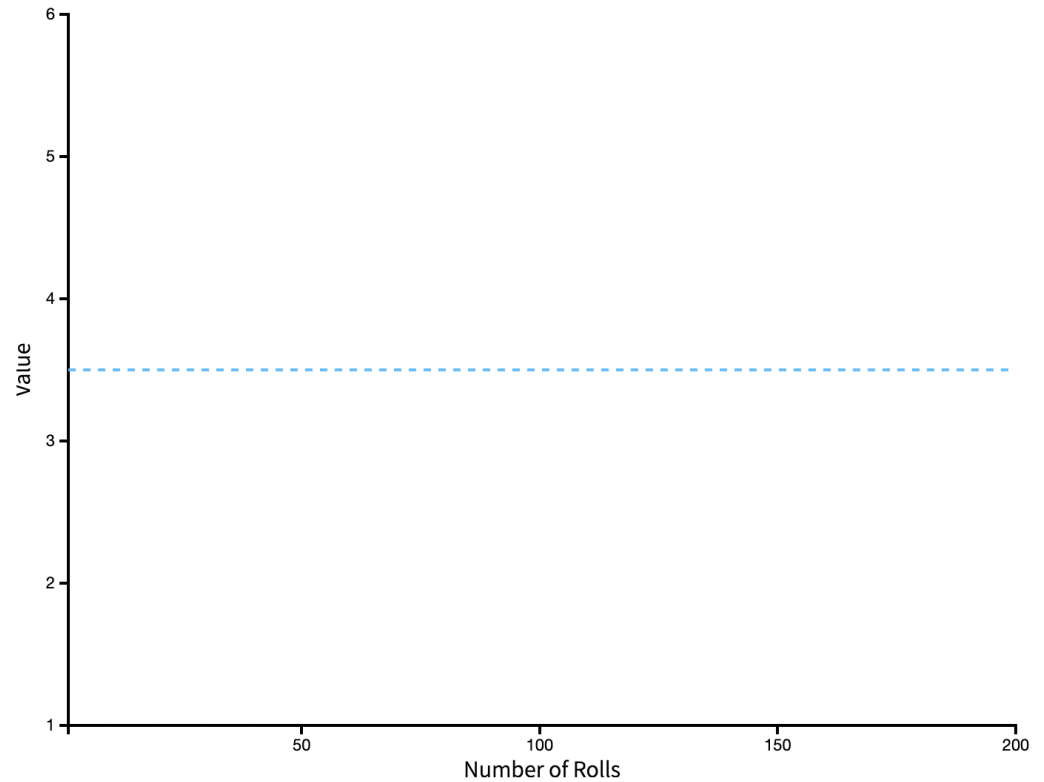
Consider the probabilistic experiment of rolling a fair die and watch as the running sample mean converges to the expectation of 3.5.



Roll the Die

Roll 100 times

Change the distribution of the different faces of the die (thus making the die biased or "unfair") by adjusting the **blue** bars below and observe how this changes the expectation.



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

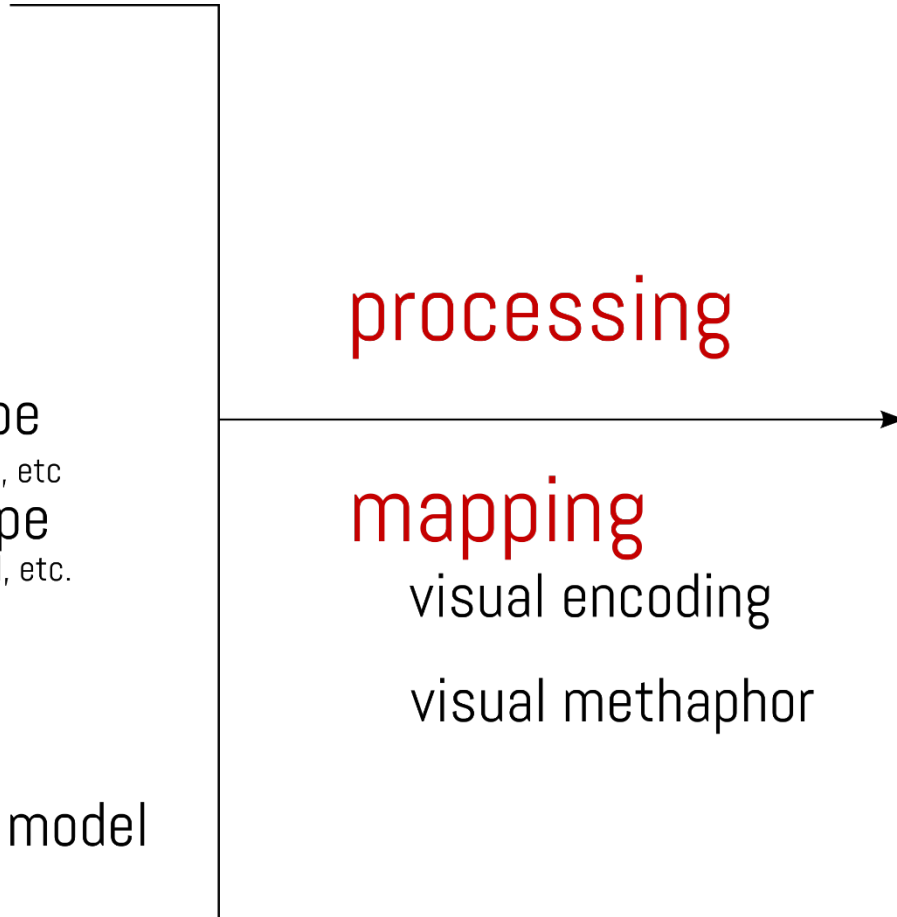
processing

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

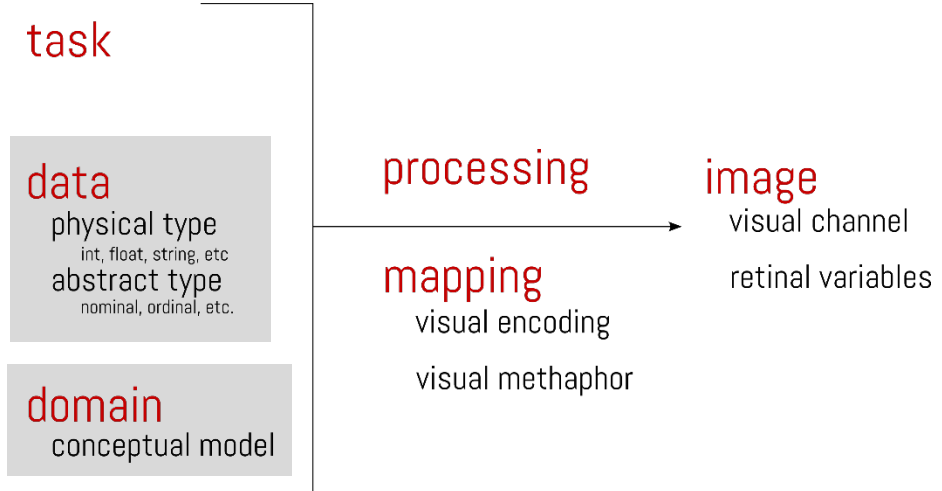
mapping

visual encoding
visual methaphor

image

visual channel
retinal variables

Data & Domain



Data Model

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

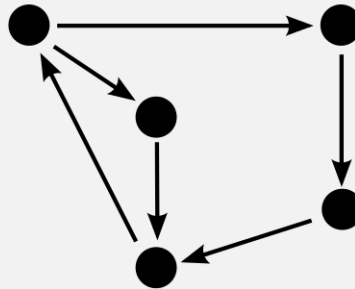
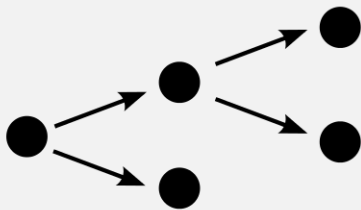
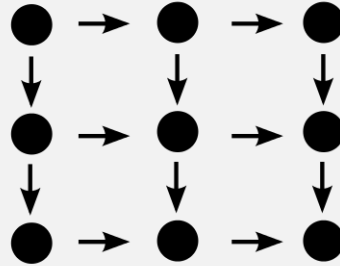
Conceptual Model

- 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 – Interval (Location of zero is arbitrary)
 - Dates: Mar. 14, 1933
 - Lat: 26.1, Long: -110.0
 - Only differences (i.e. Intervals) can be compared
- Q – Ratio (zero fixed)
 - Physical measurements: Length, Mass
 - Counts and amounts

Nominal, Ordinal and Quantitative

- N - Nominal (labels):
 - Operations: =, \neq
- O - Ordinal
 - Operations: =, \neq , $<$, $>$, \leq , \geq
- Q - Interval (Location of zero arbitrary)
 - Operations: =, \neq , $<$, $>$, \leq , \geq , -
 - Can measure distances or spans
- Q - Ratio (zero fixed)
 - Operations: =, \neq , $<$, $>$, \leq , \geq , -, \div
 - Can measure ratios or proportions

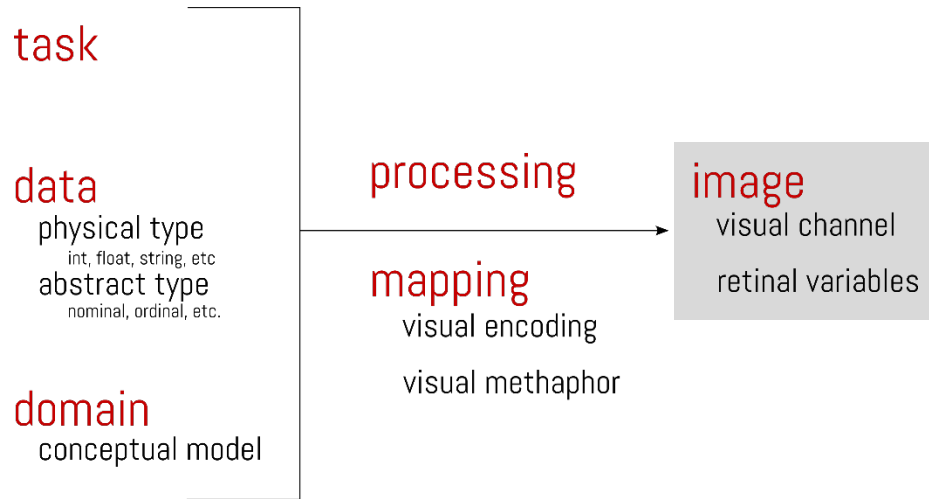
Example

- 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)

Width	Length	Species
1.2	5.1	setosa
1.4	4.9	versicolor
0.8	4.7	virginica
1.7	4.6	setosa
4.9	4.3	virginica

Q, Q, N

Image



Pre-attentive

unconscious, parallel, fast

Attentive

conscious, serial, slow



How many 3's?

1281768756138976546984506985604982826762
9809858458224509856458945098450980943585
9091030209905959595772564675050678904567
8845789809821677654876364908560912949686

How many 3's?

12817687561**3**8976546984506985604982826762
980985845822450985645894509845098094**3**585
90910**3**0209905959595772564675050678904567
8845789809821677654876**3**64908560912949686

LES VARIABLES DE L'IMAGE

	POINTS			LIGNES			ZONES	
XY 2 DIMENSIONS DU PLAN								
Z TAILLE								
VALEUR								

LES VARIABLES DE SÉPARATION DES IMAGES

GRAIN								
COULEUR								
ORIENTATION								
FORME								

Color hue



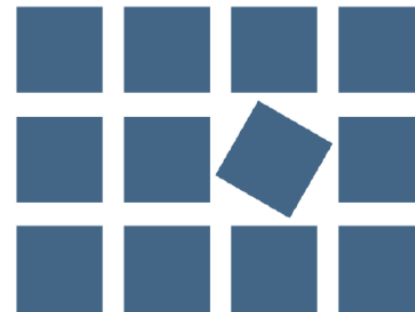
Color brightness



Position



Orientation



Color saturation



Size

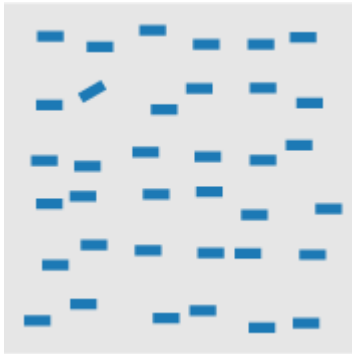


Texture

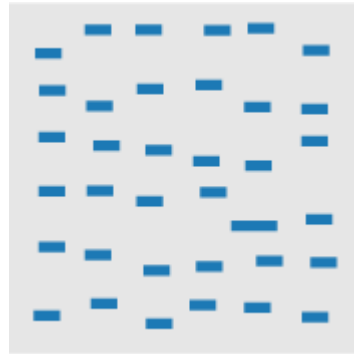


Shape

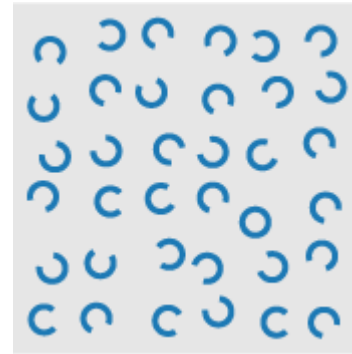




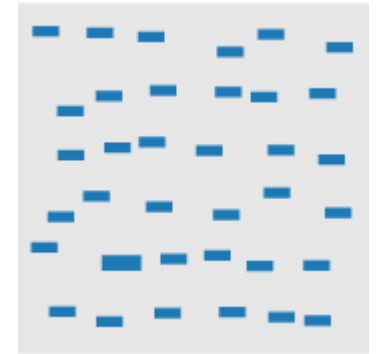
Line orientation



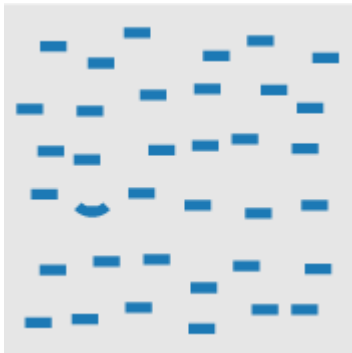
Length, width



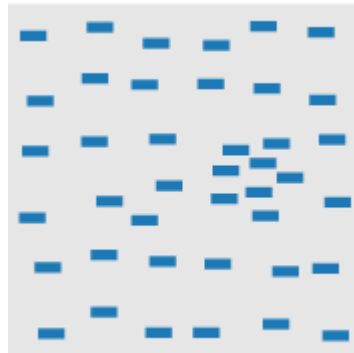
closure



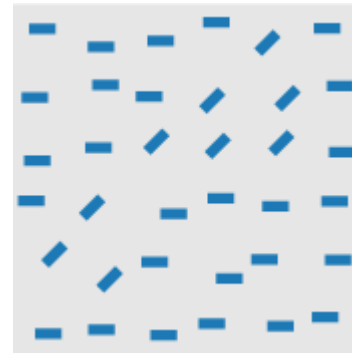
size



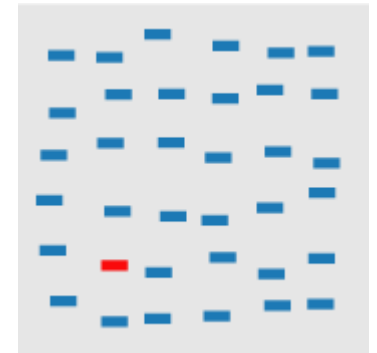
curvature



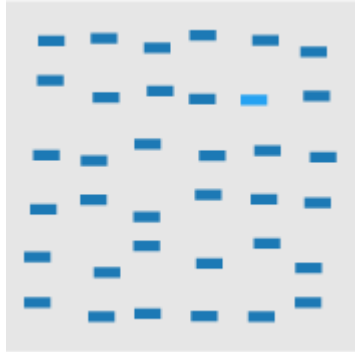
density



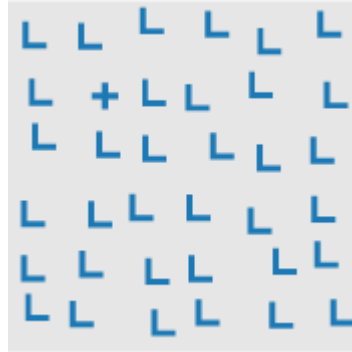
number, estimation



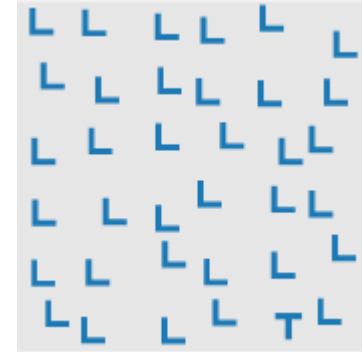
hue



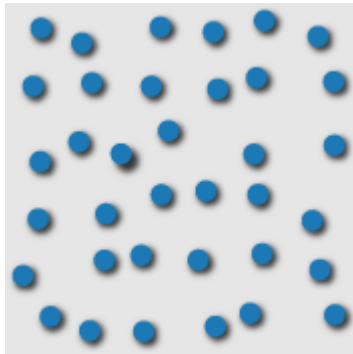
intensity



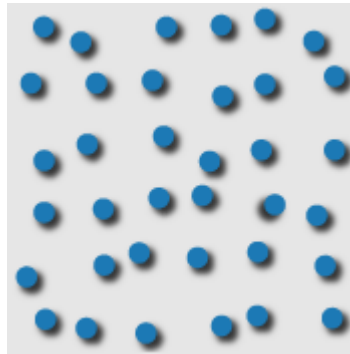
intersection



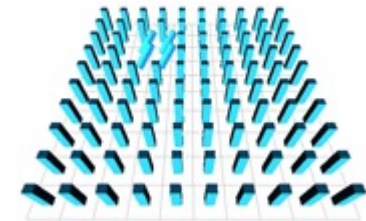
terminators



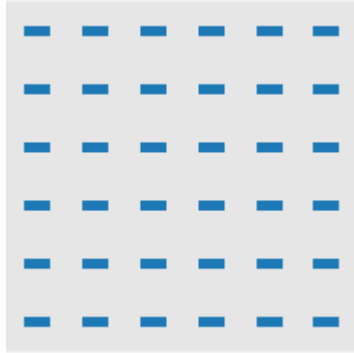
3D depth cues,
stereoscopic depth



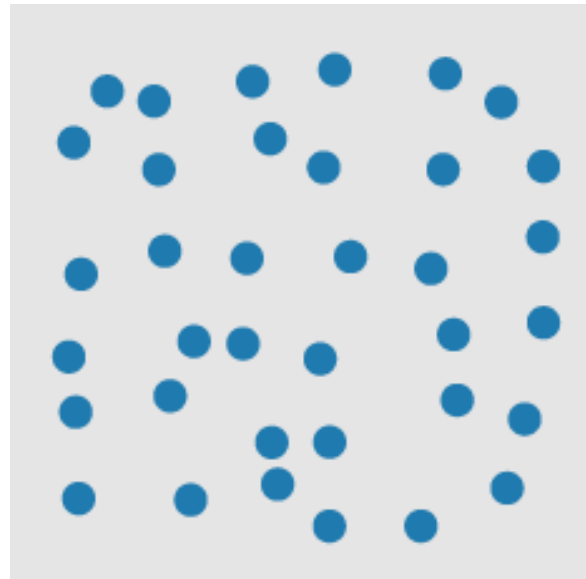
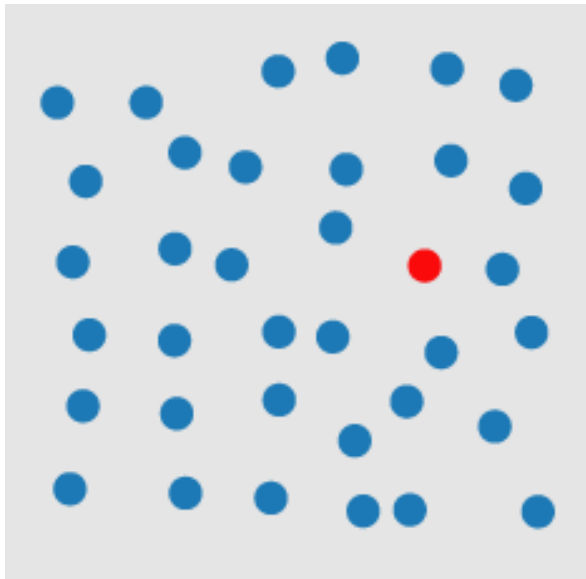
Lighting direction

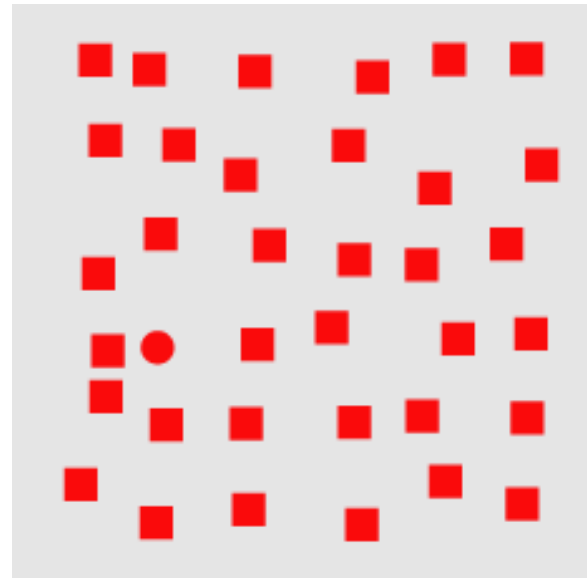
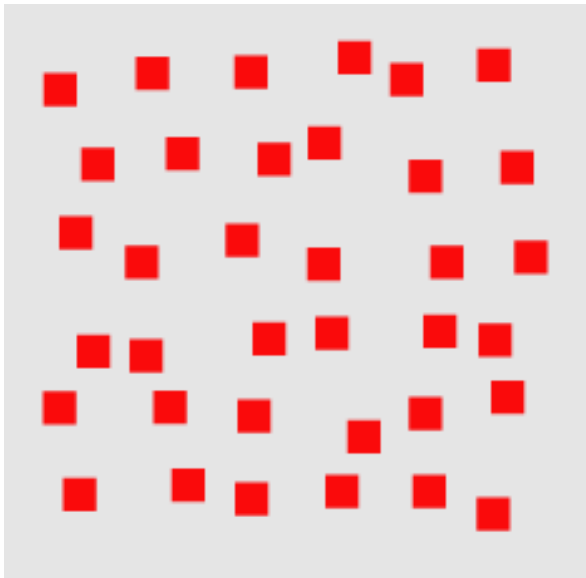


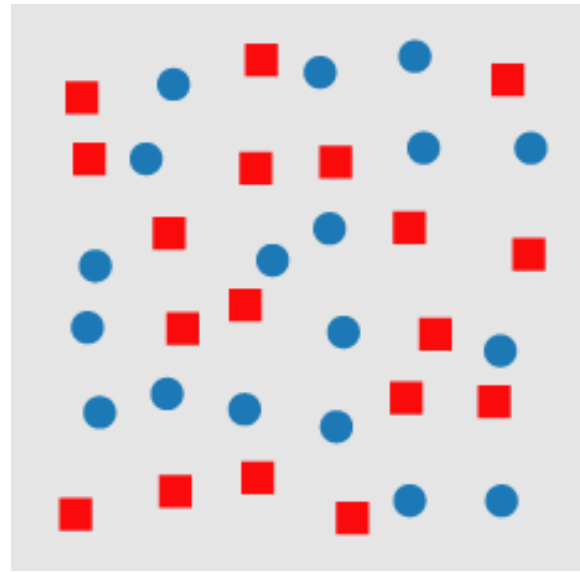
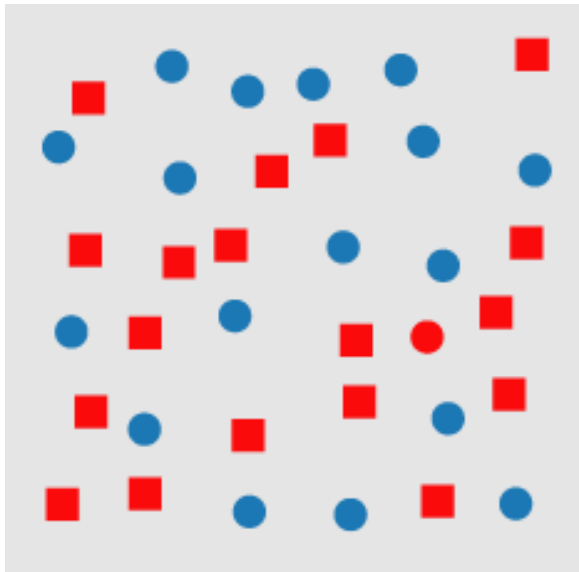
3D orientation

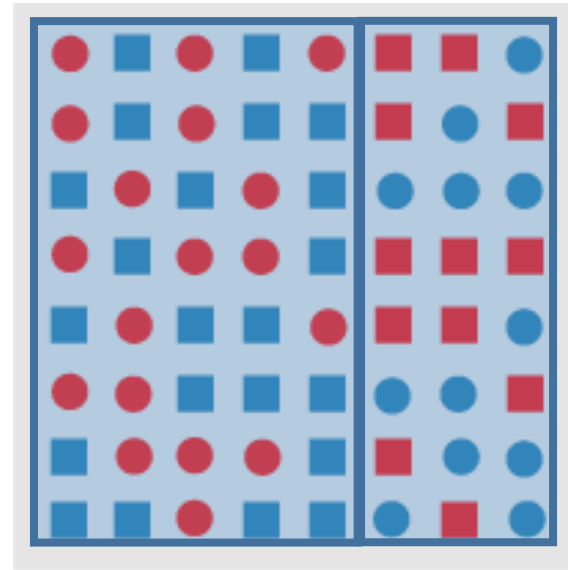


Flicker

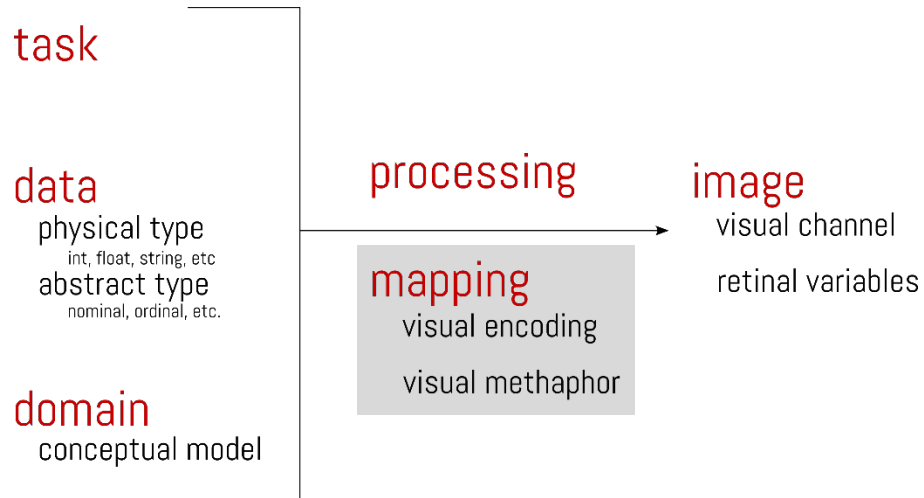


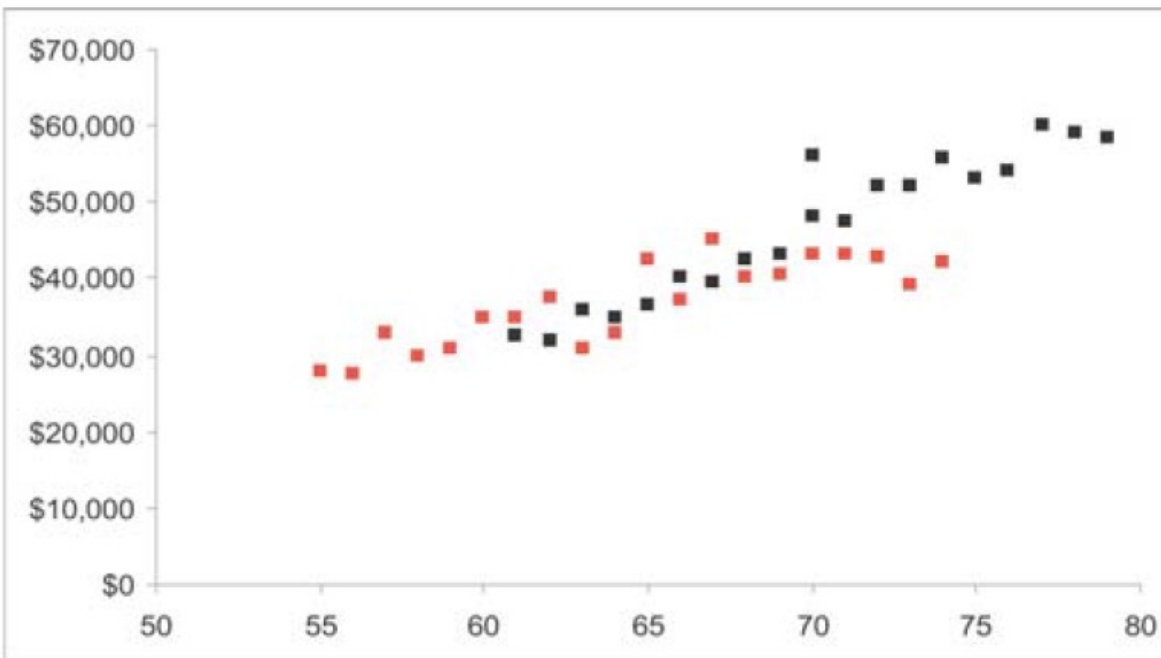
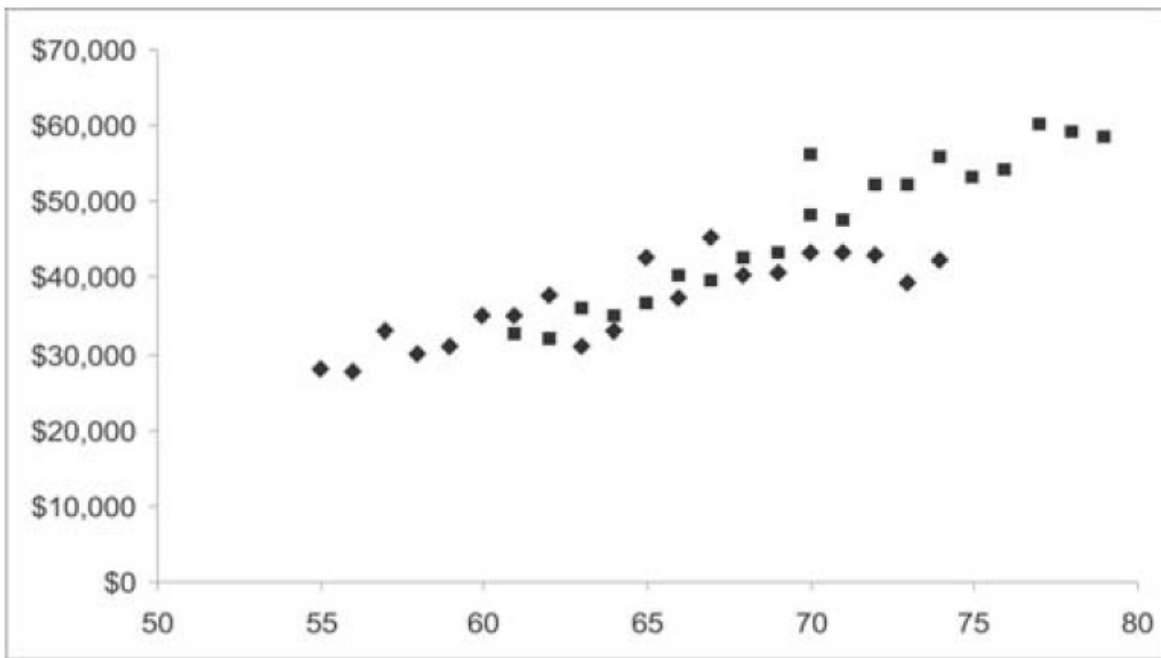


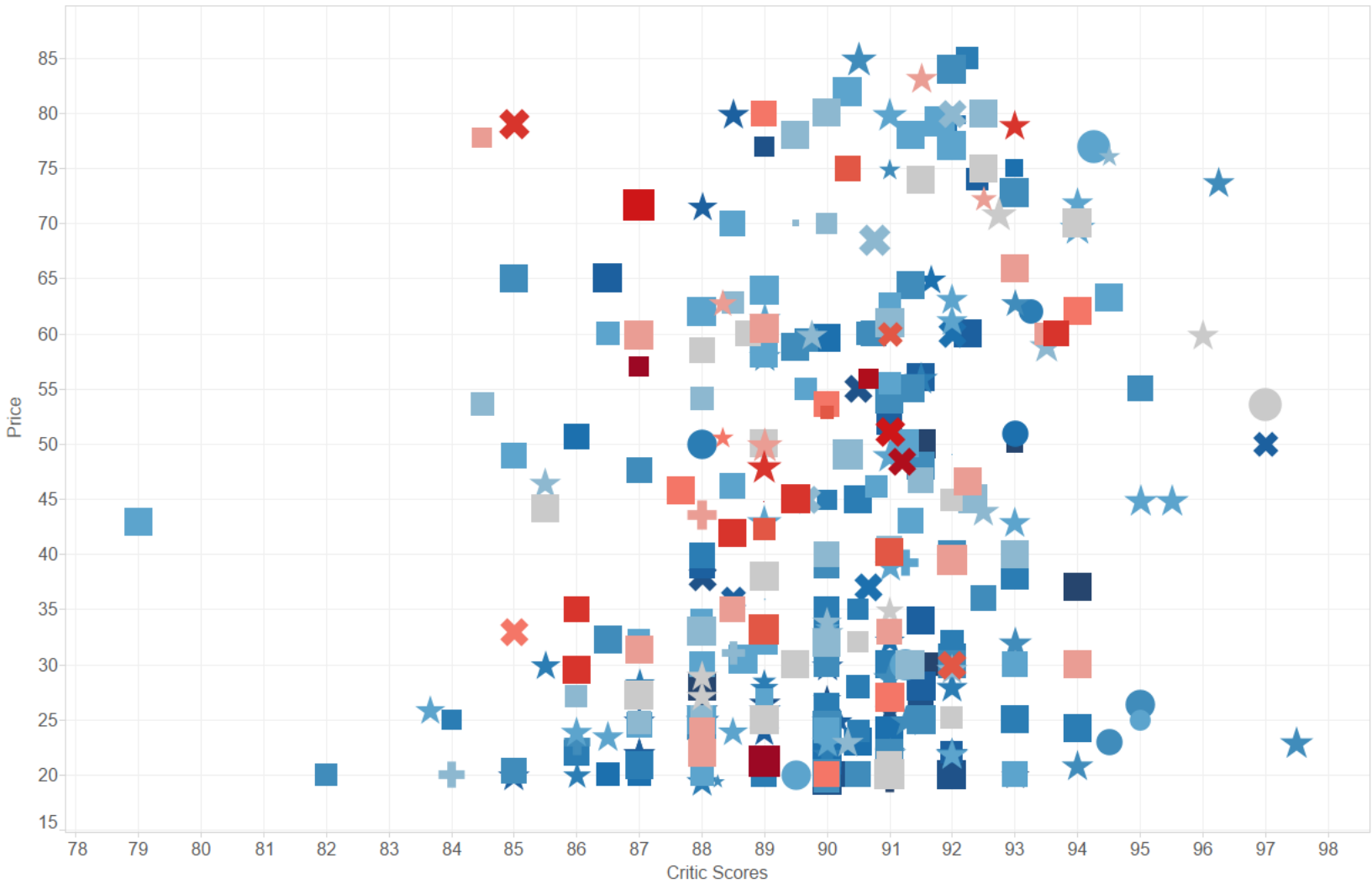




Mapping



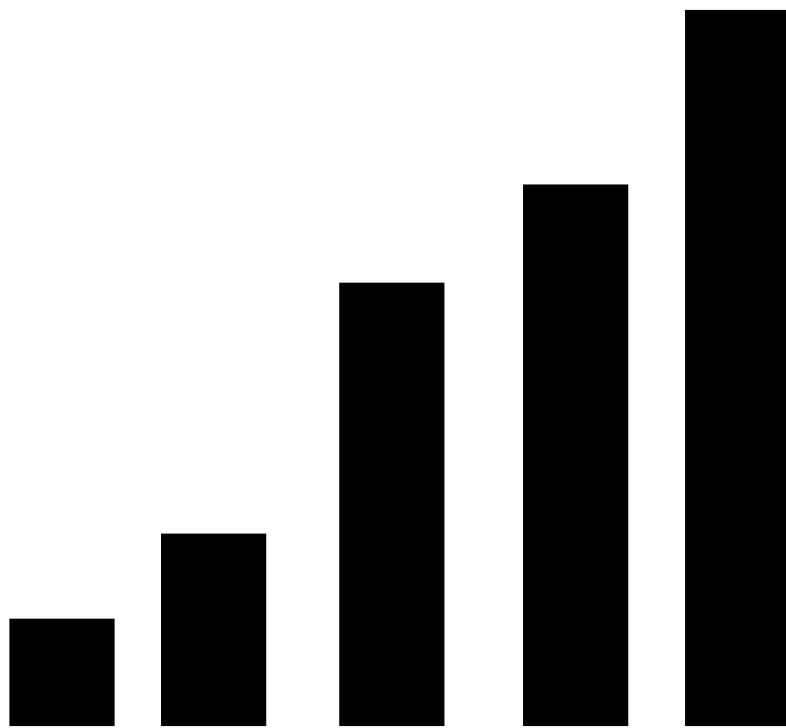




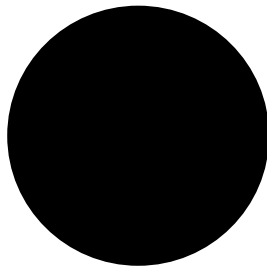
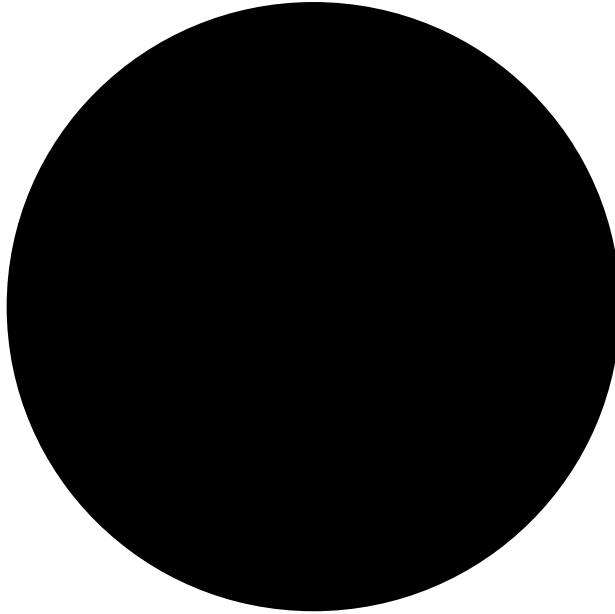
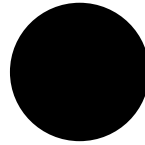
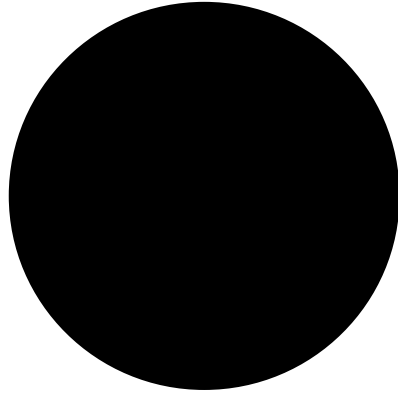
X: Critic Scores, Y: Price, Size: User Rating, Color: Vintage, Shape: Type

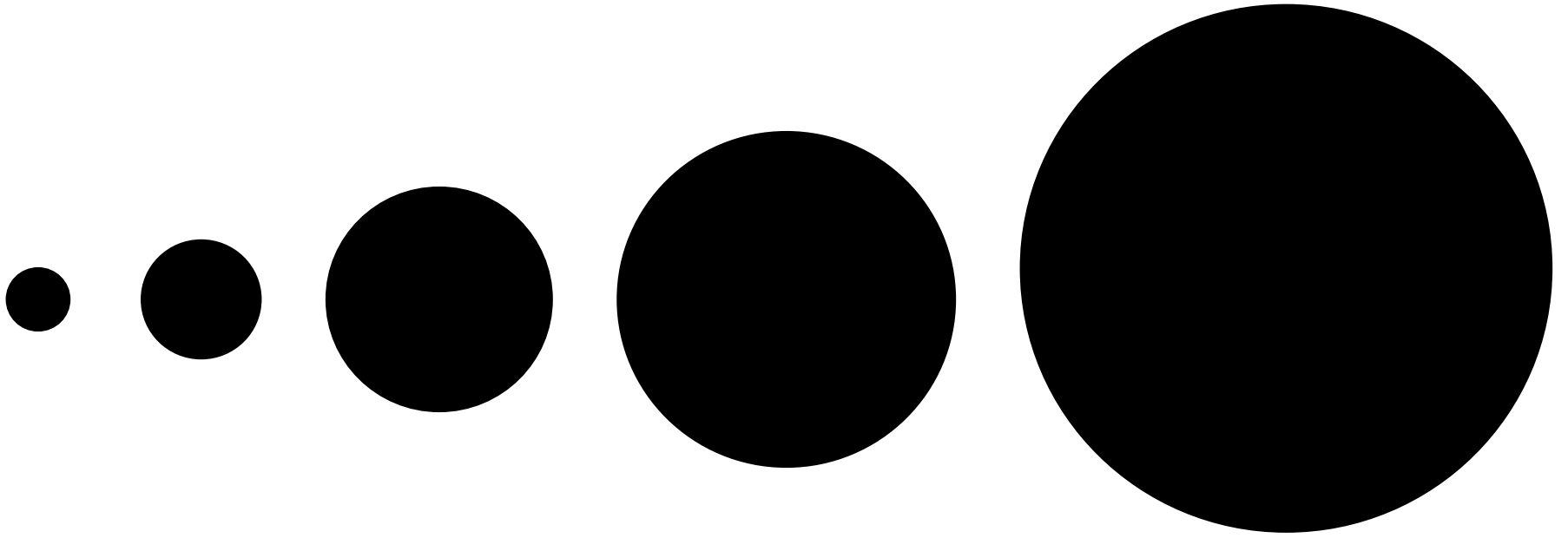
Can you order these?
(low -> high)



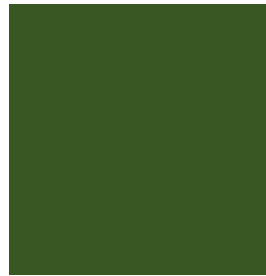
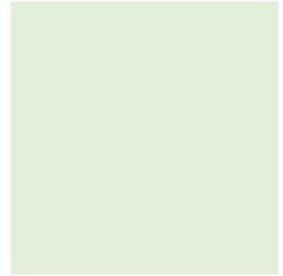


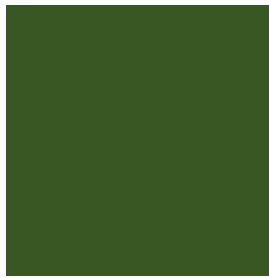
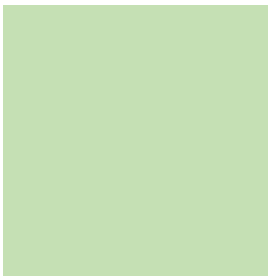
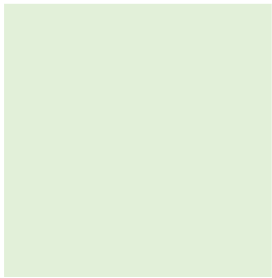
Can you order these?
(low -> high)



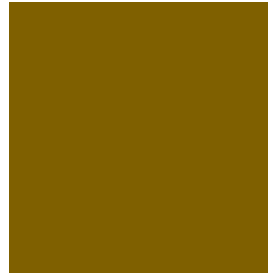
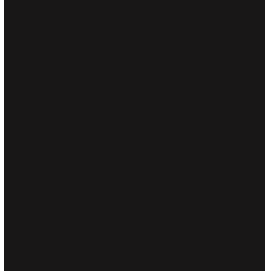


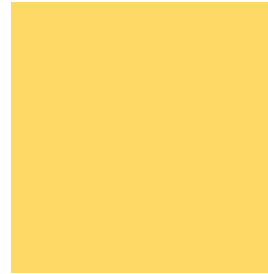
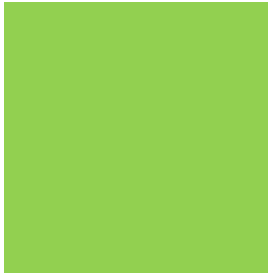
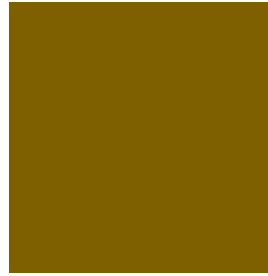
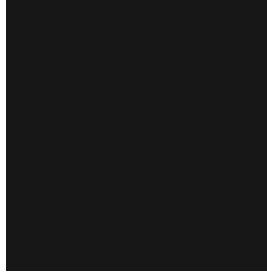
Can you order these?
(low -> high)





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		
---	--	--

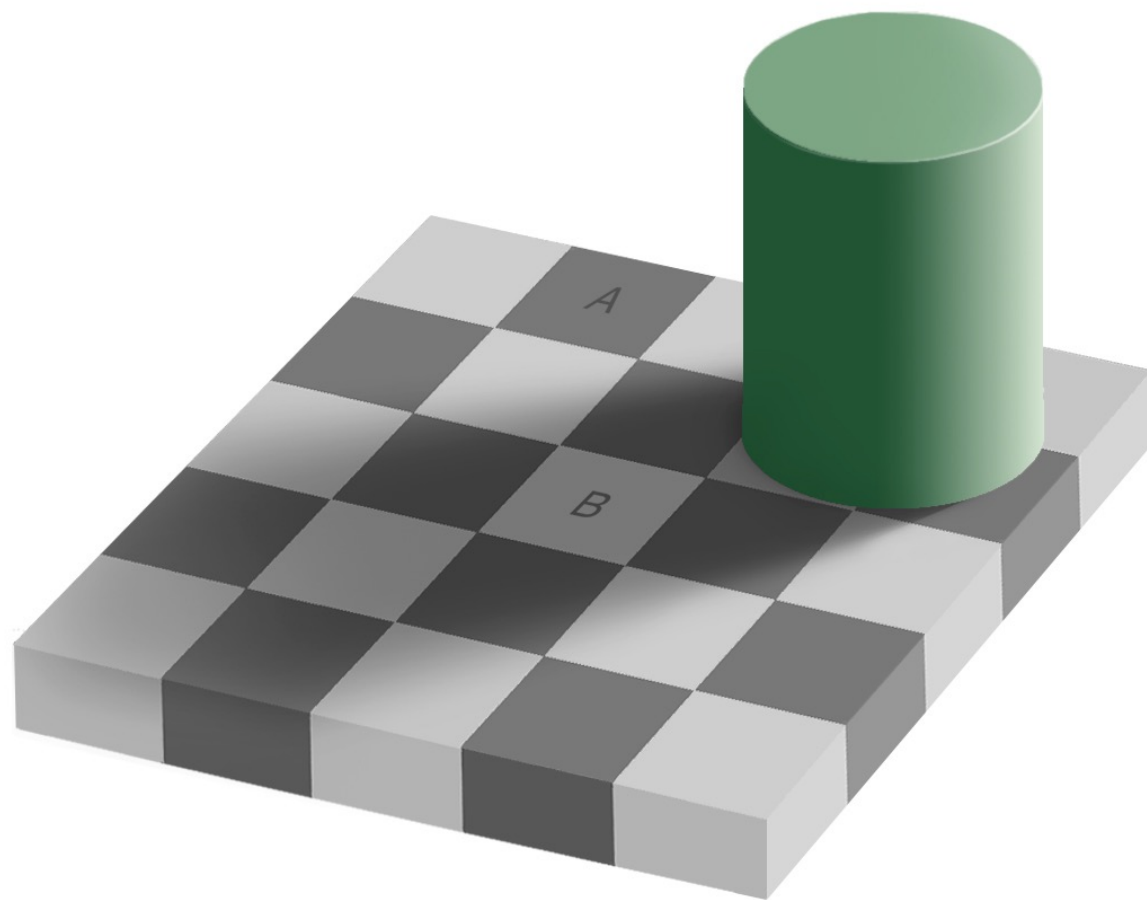
Shape

N		
---	--	--

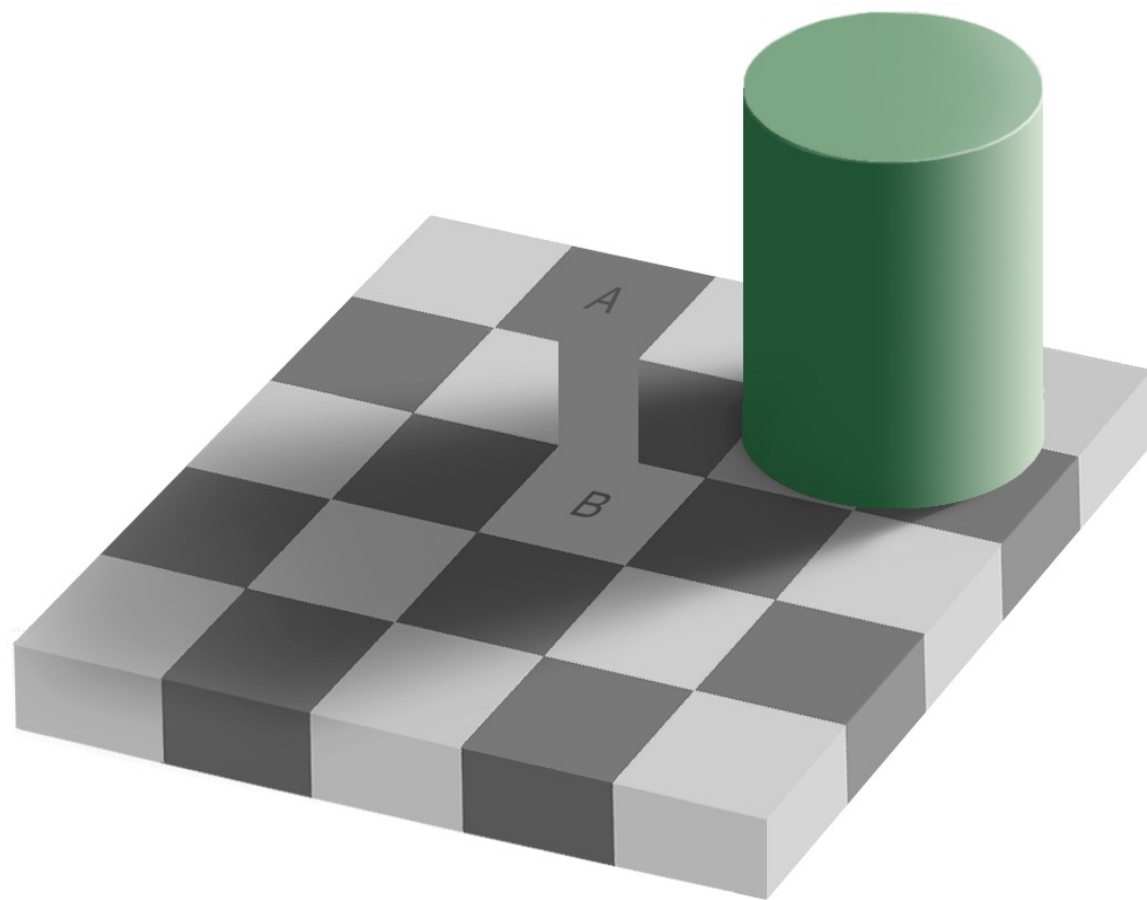
Nominal

Ordered

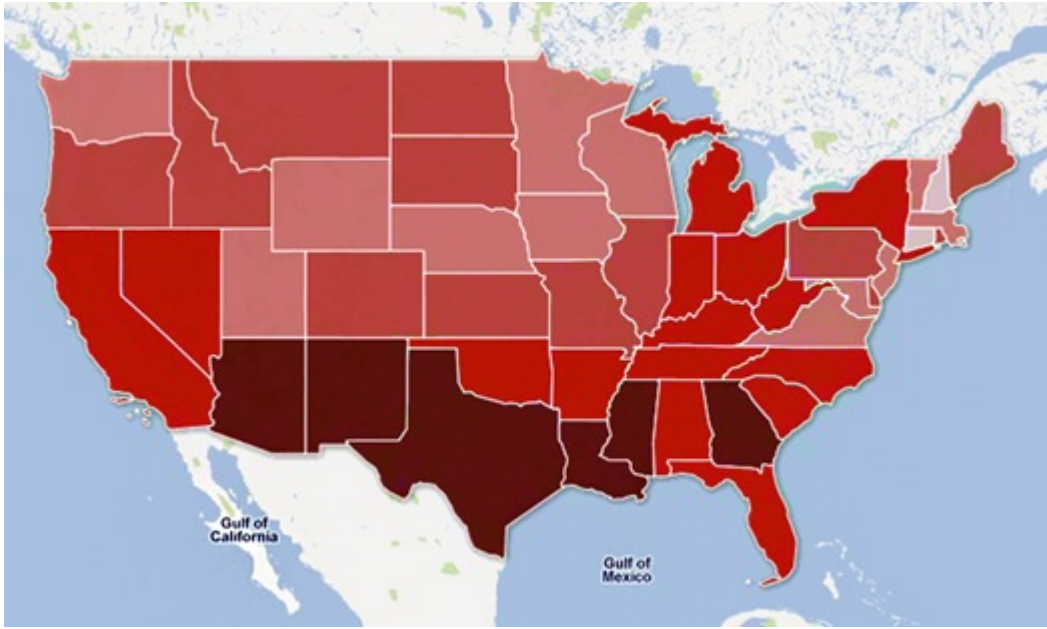
Quantitative

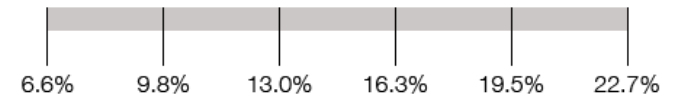
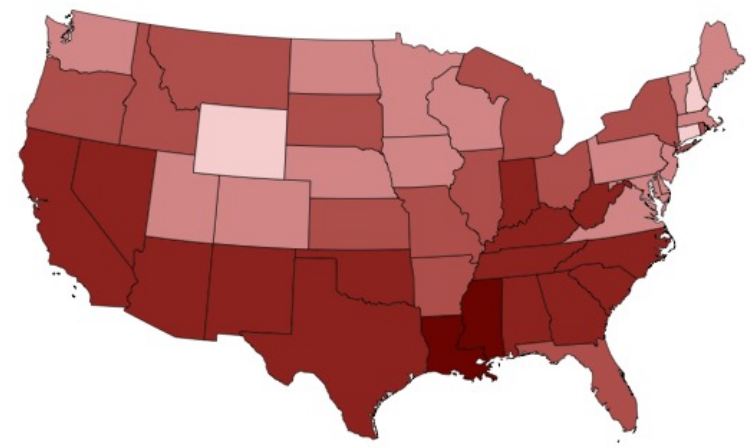
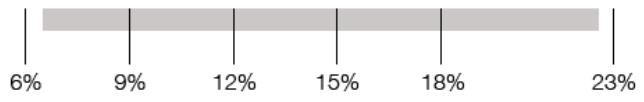
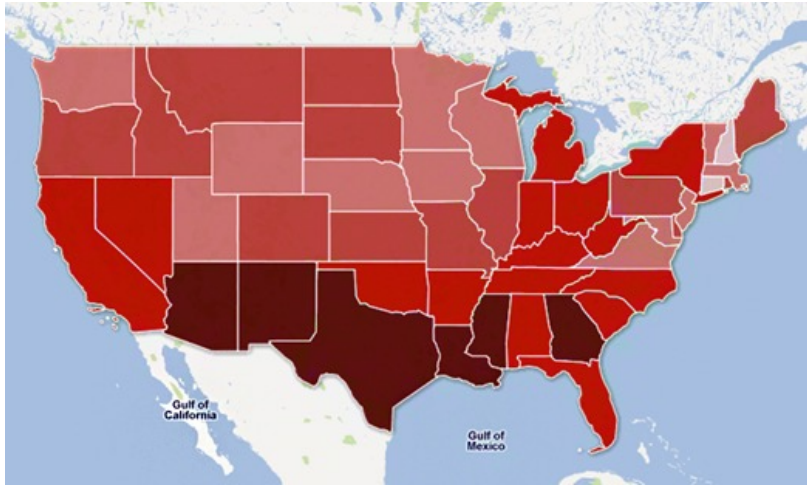


[Edward Adelson, 1995]

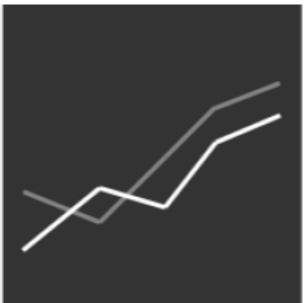
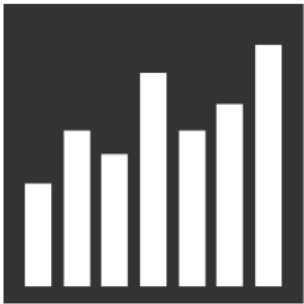


[Edward Adelson, 1995]



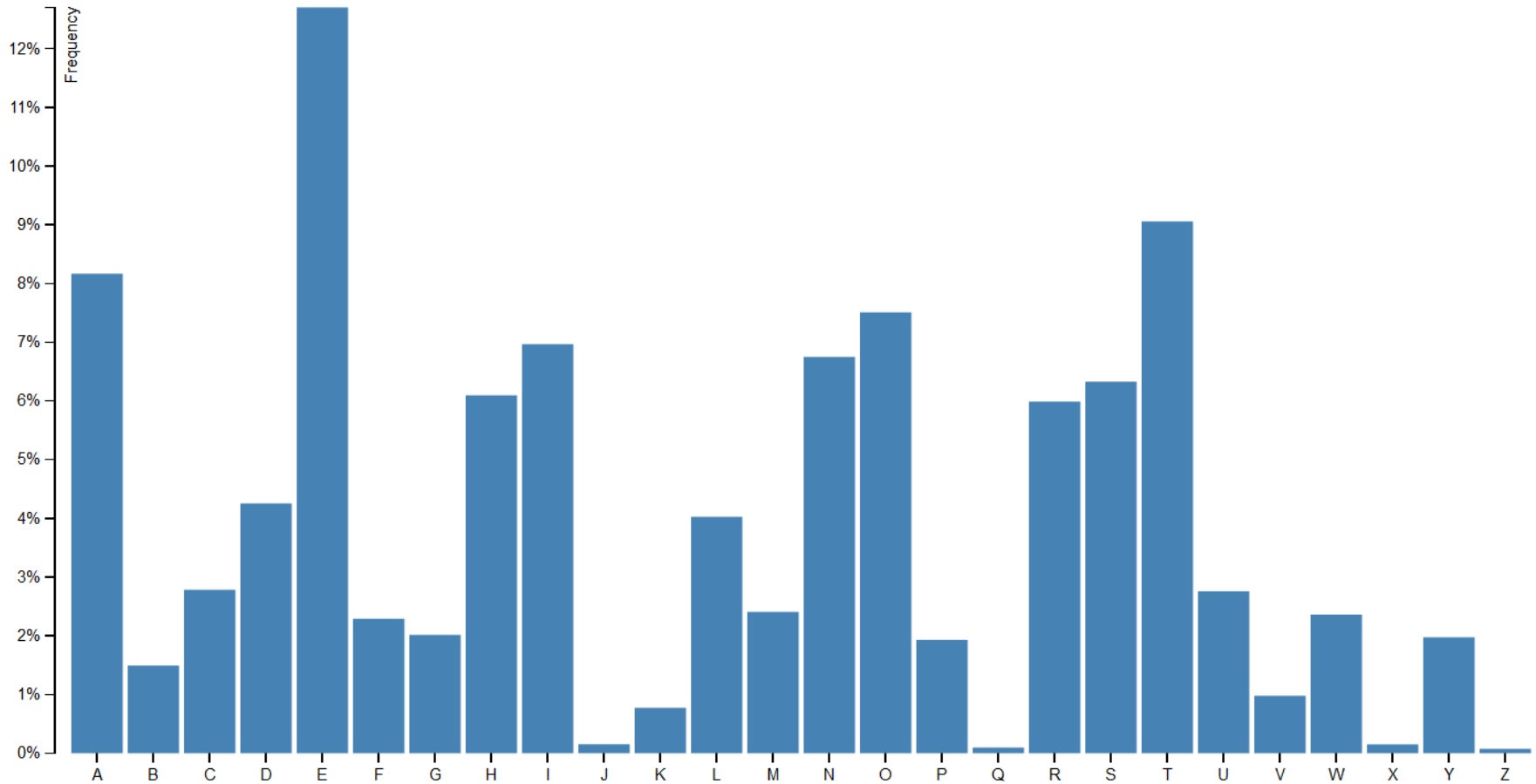


Some Principles & Guidelines



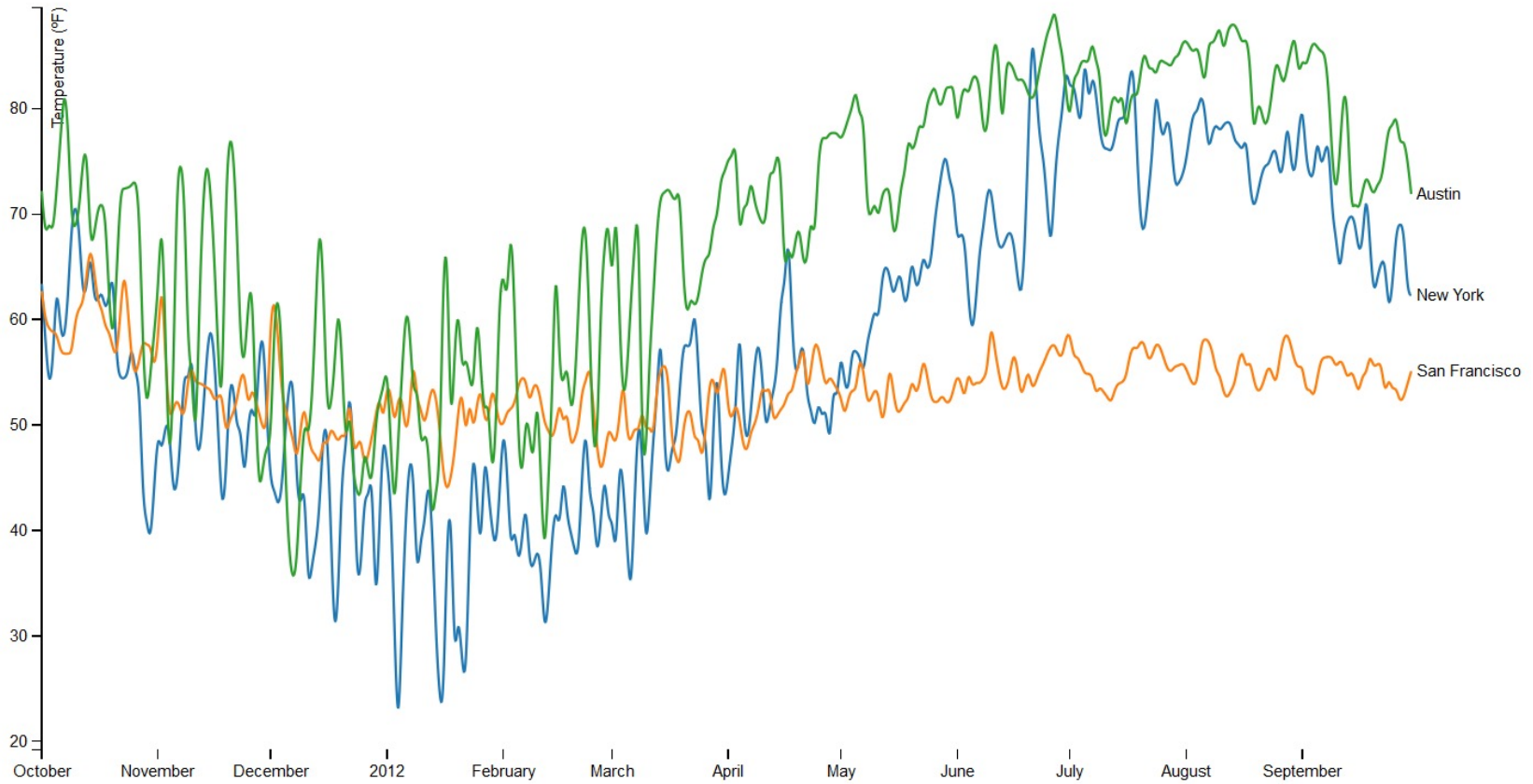
Bar chart

Display different quantities of single-variable data



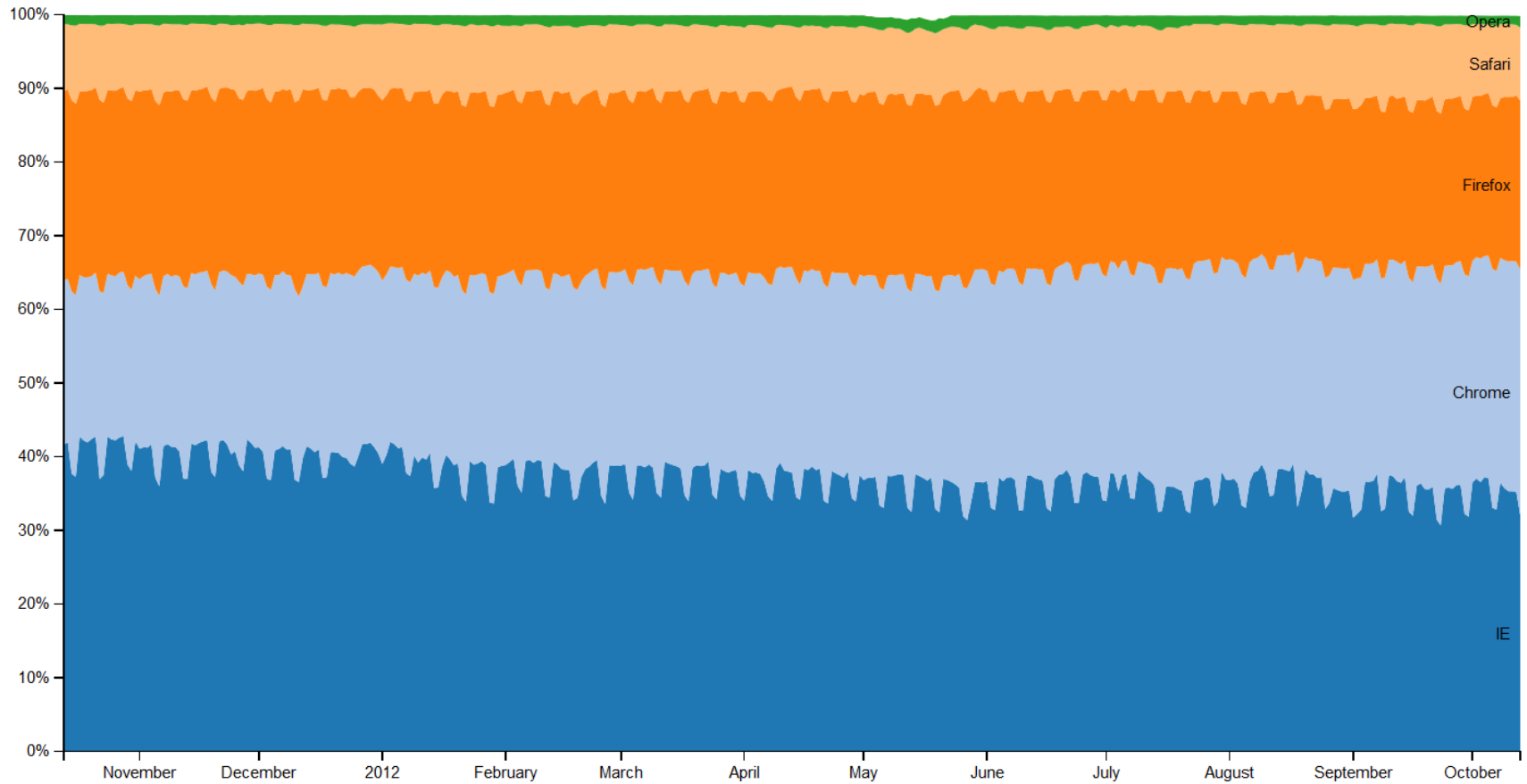
Line chart

Display how a variable develops over time



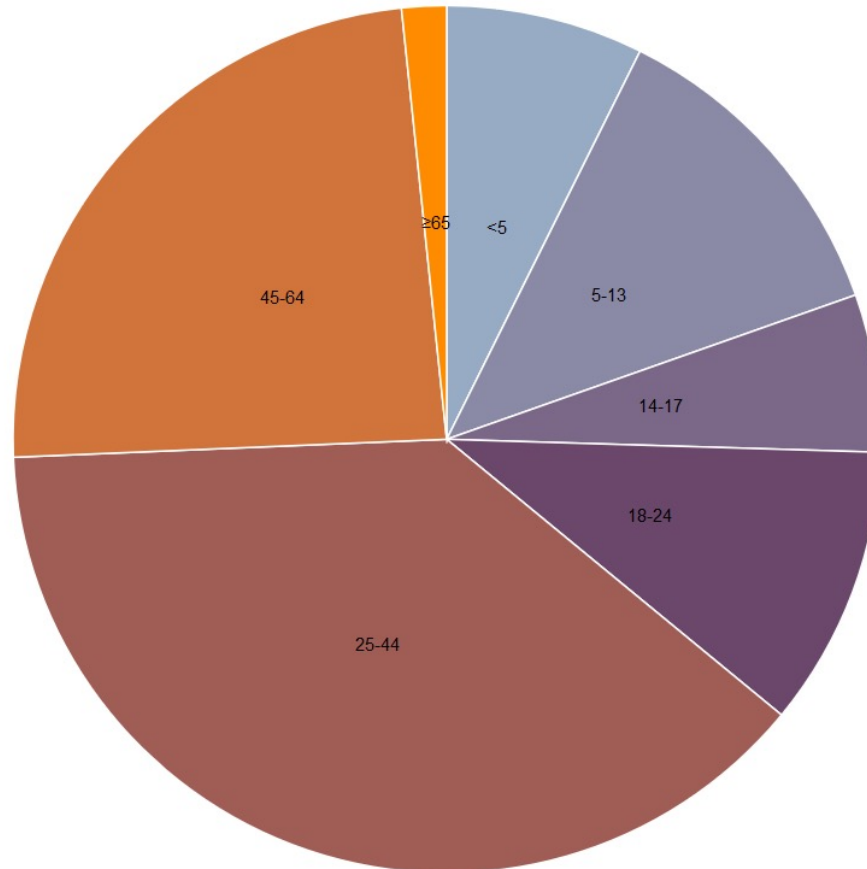
Stacked area chart

Display total of a variable over time



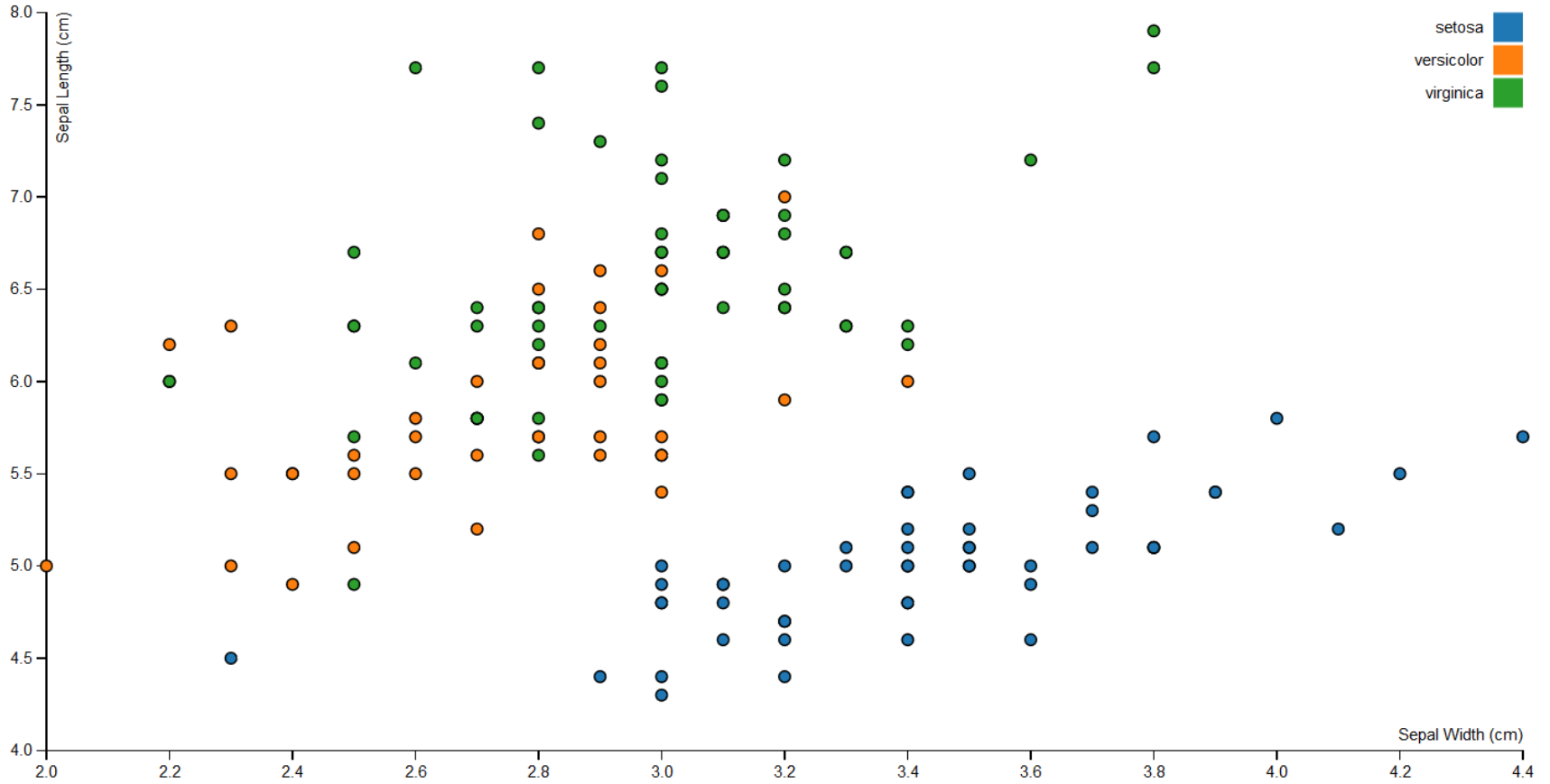
Pie chart

Display distribution of a variable



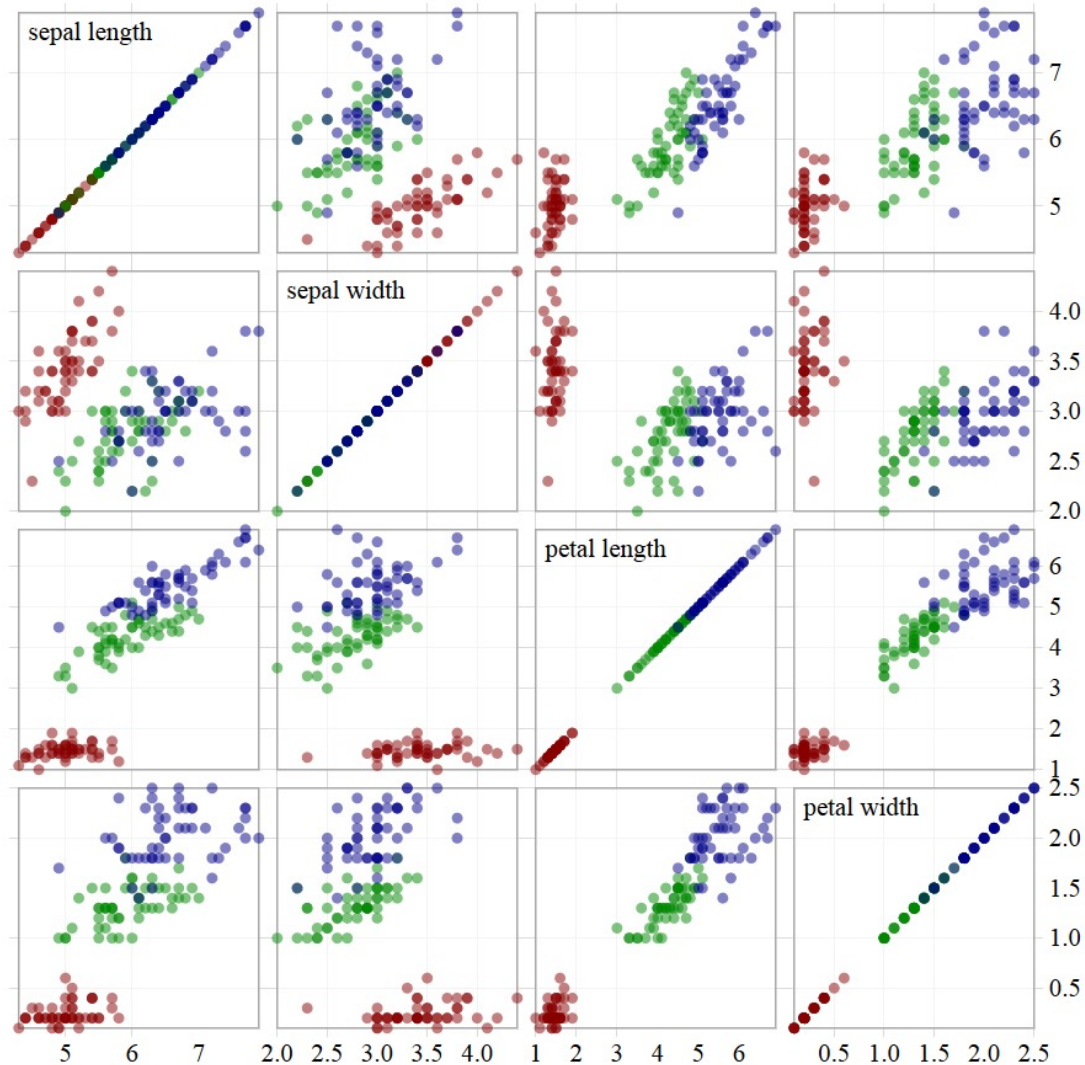
Scatterplot

Display relationship between two variables



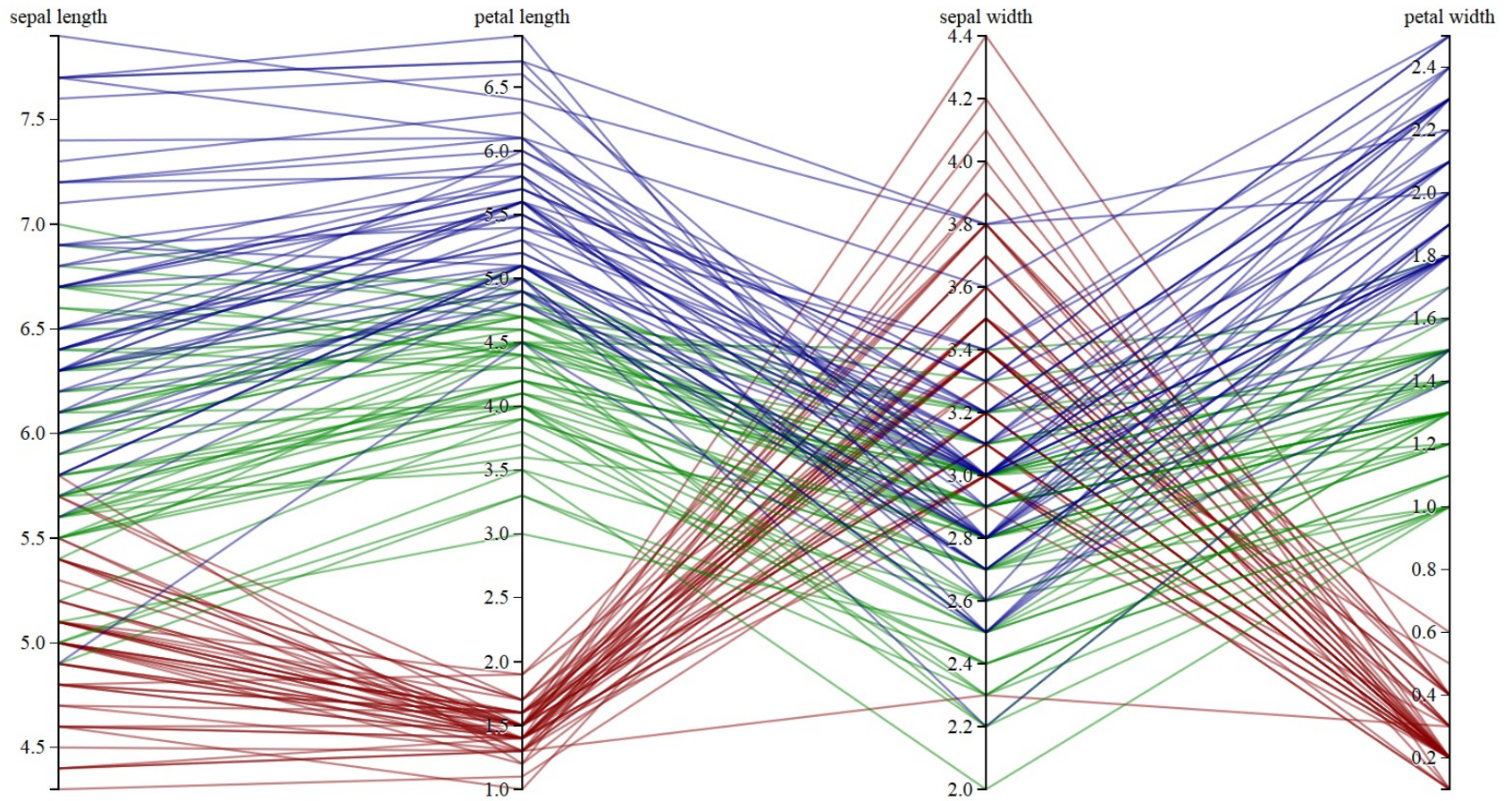
Scatterplot matrix

Display relationship between multiple variables



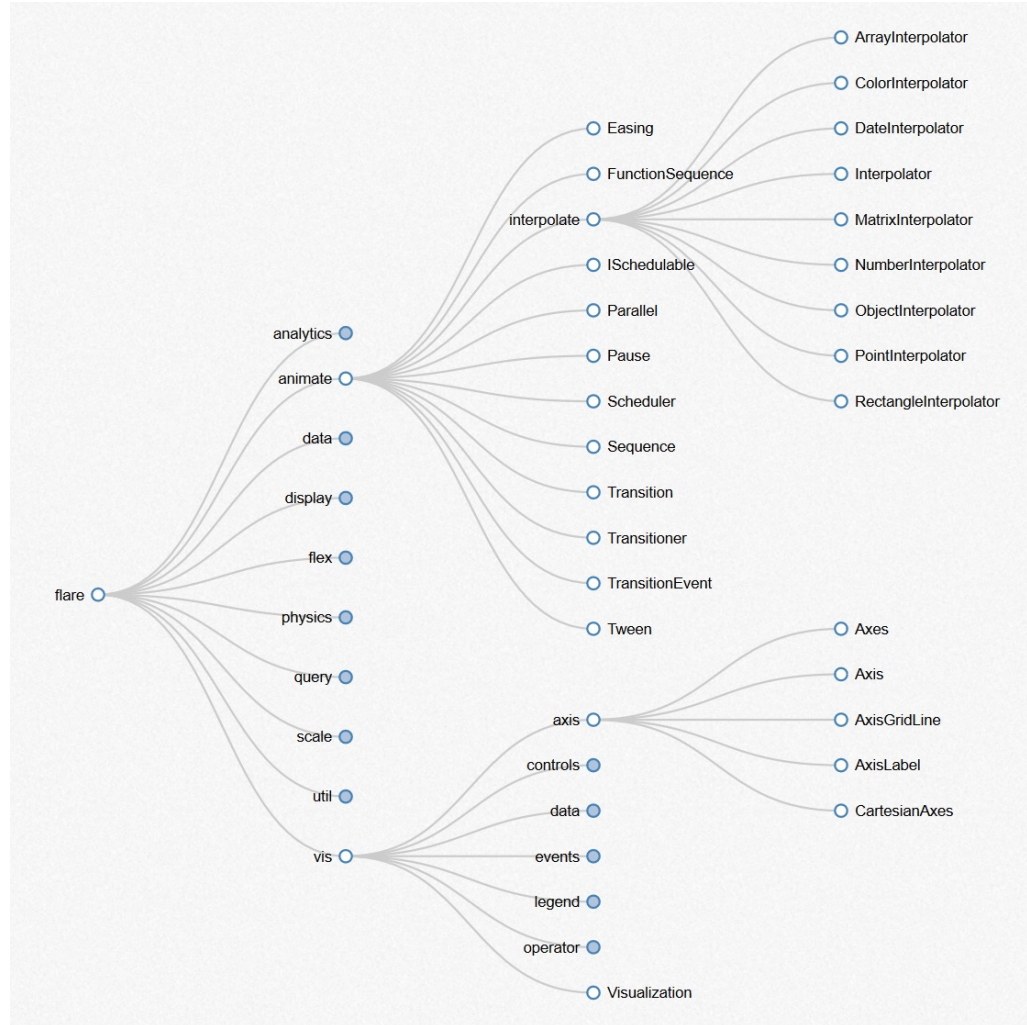
Parallel coordinate plot

Display relationship between multiple variables



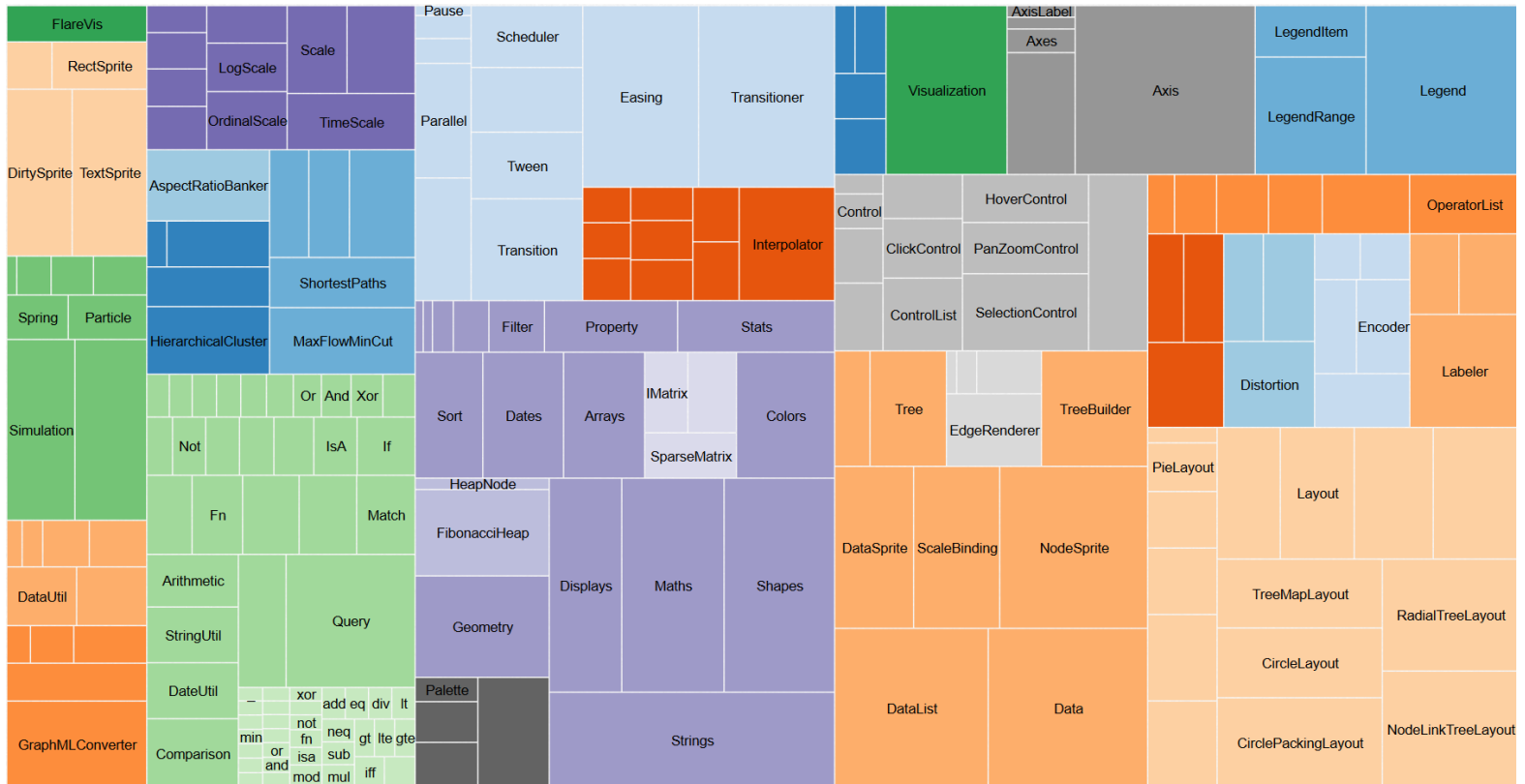
Tree

Display hierarchical data



Tree map

Display hierarchical data



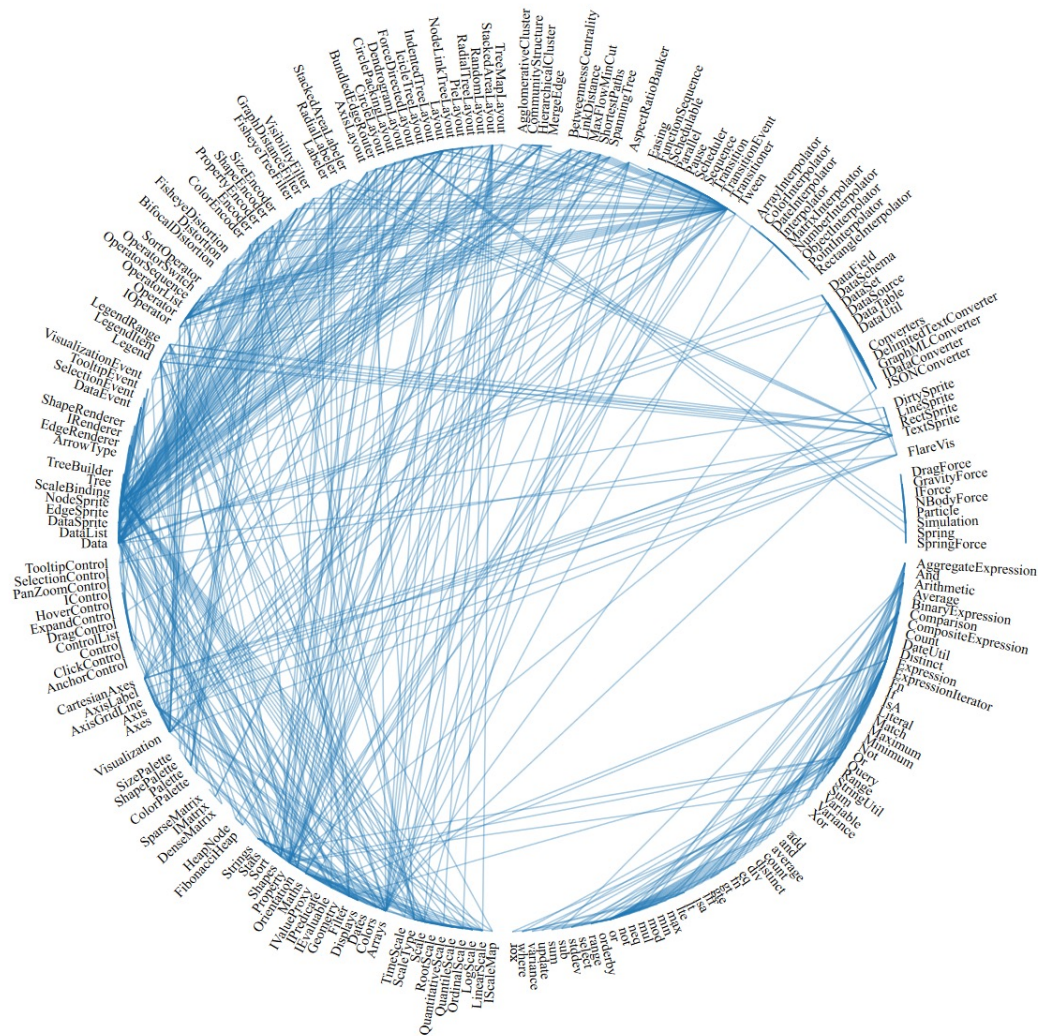
Force-directed graph

Display graphs, networks, relationships



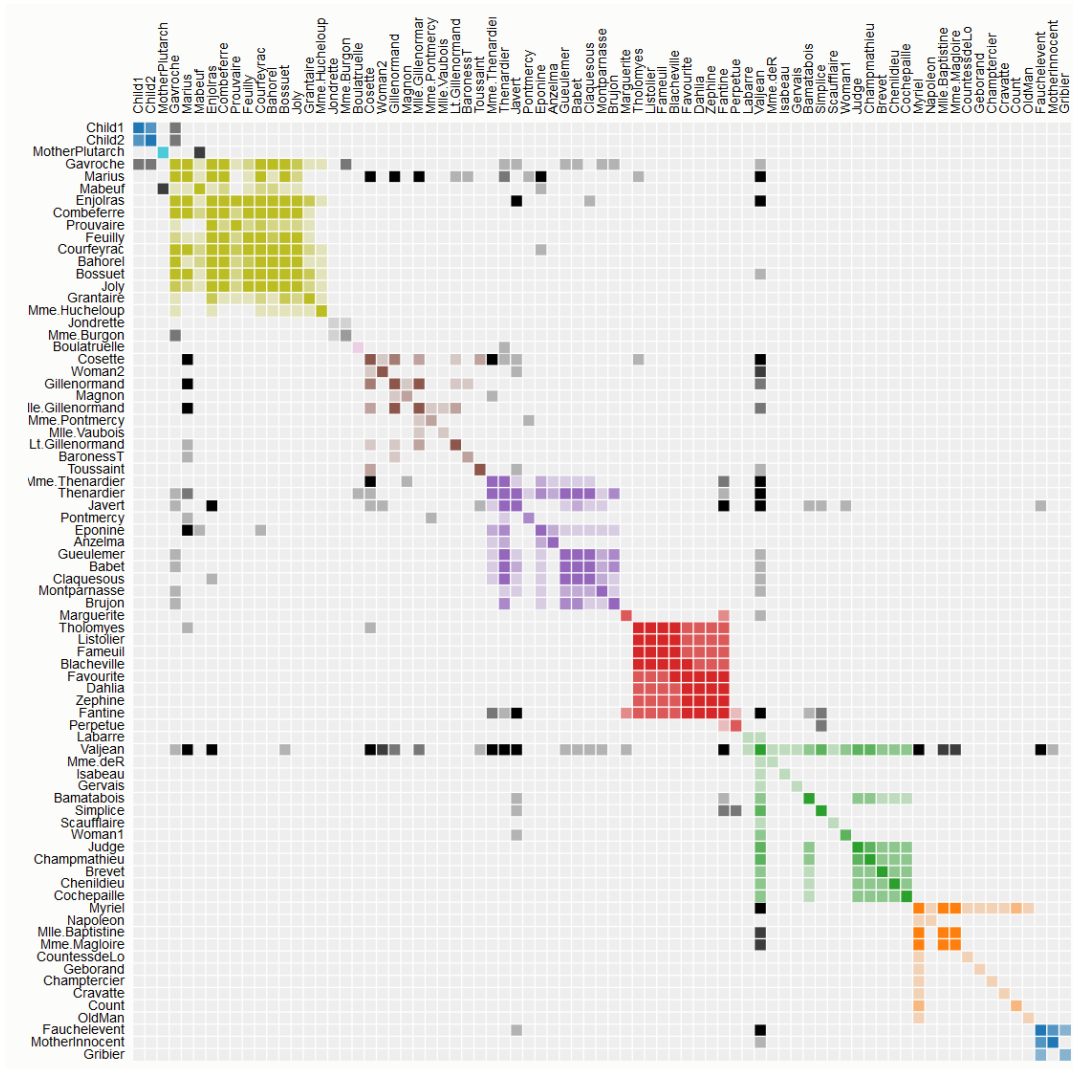
Radial graph

Display graphs, networks, relationships



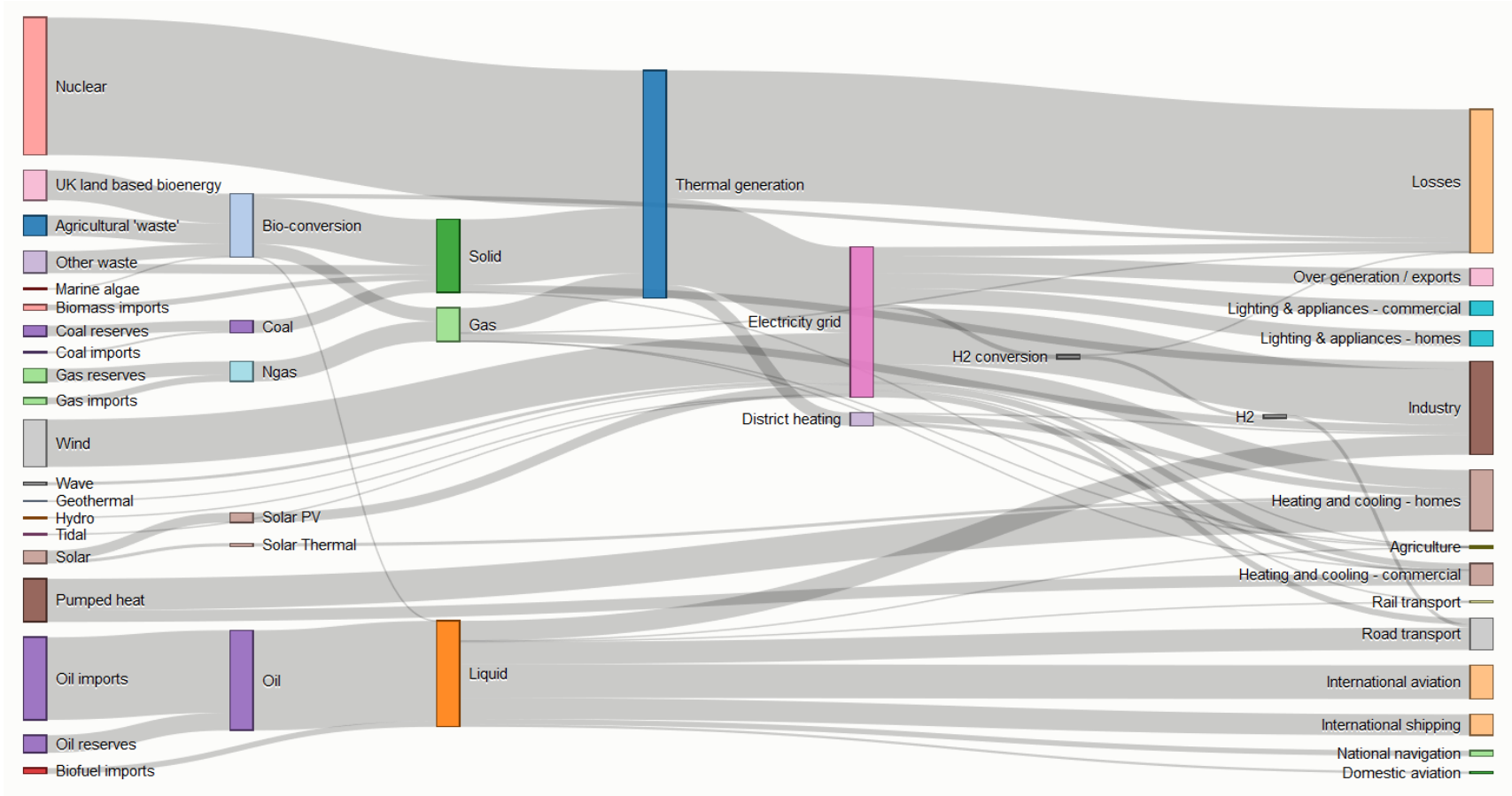
Co-occurrence matrix

Display graphs, networks, relationships



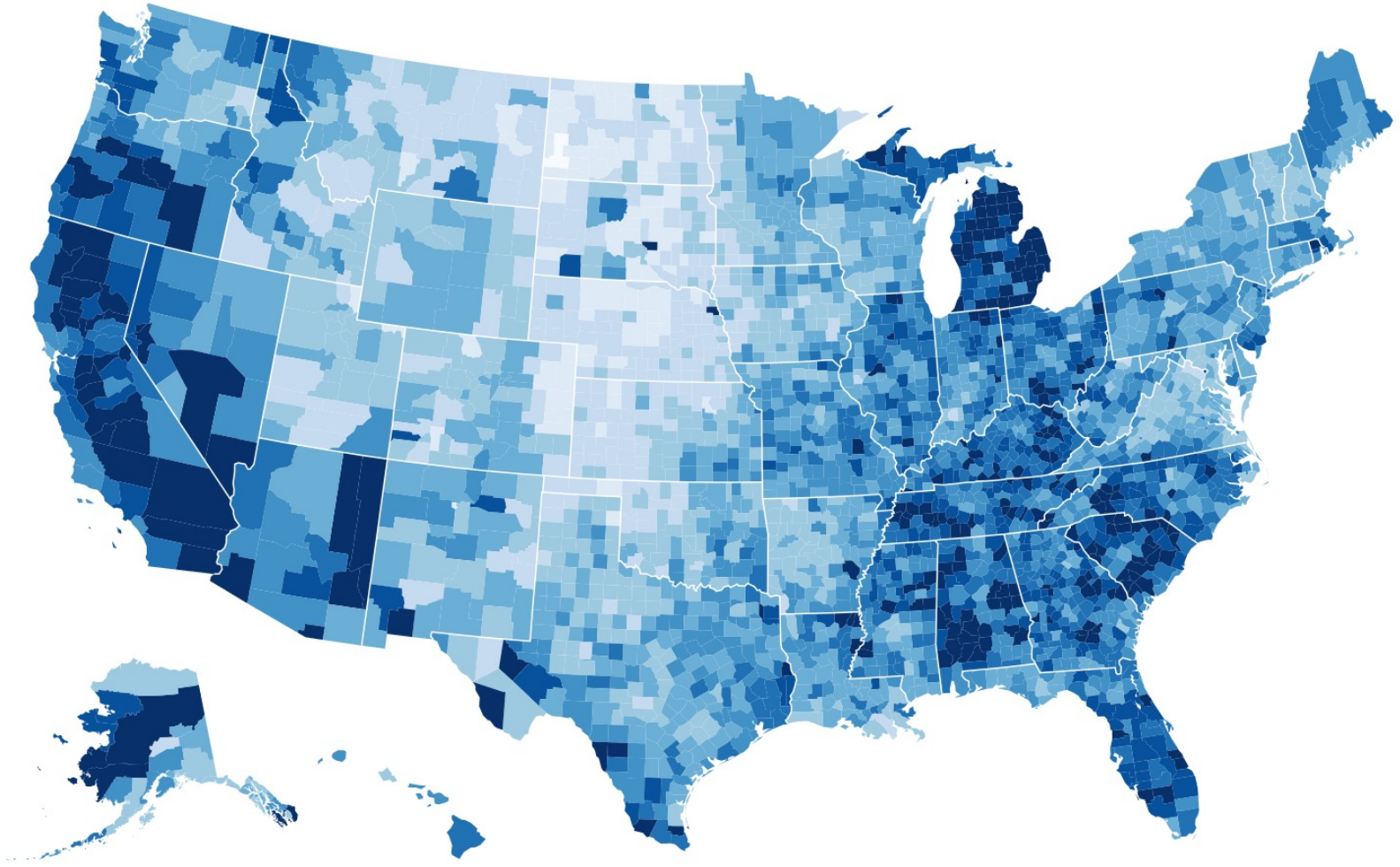
Sankey diagram

Display flow amongst entities



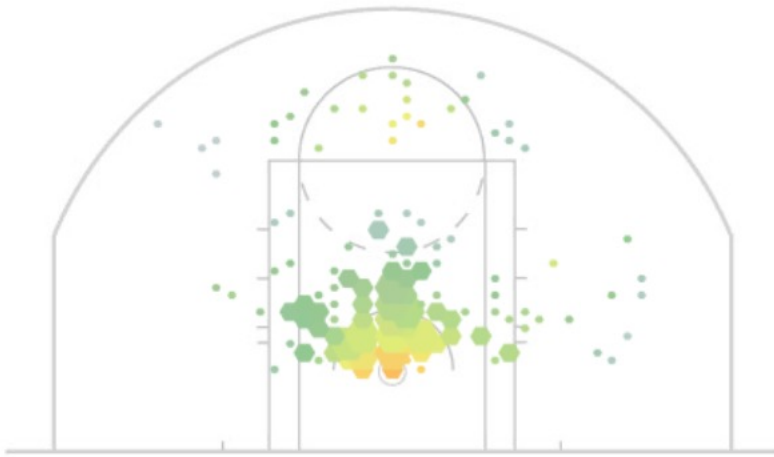
Maps

Display a spatial variable

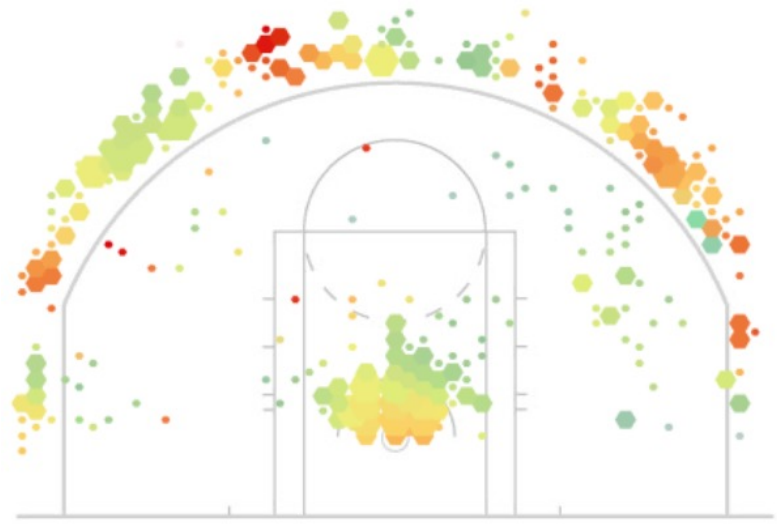


Maps

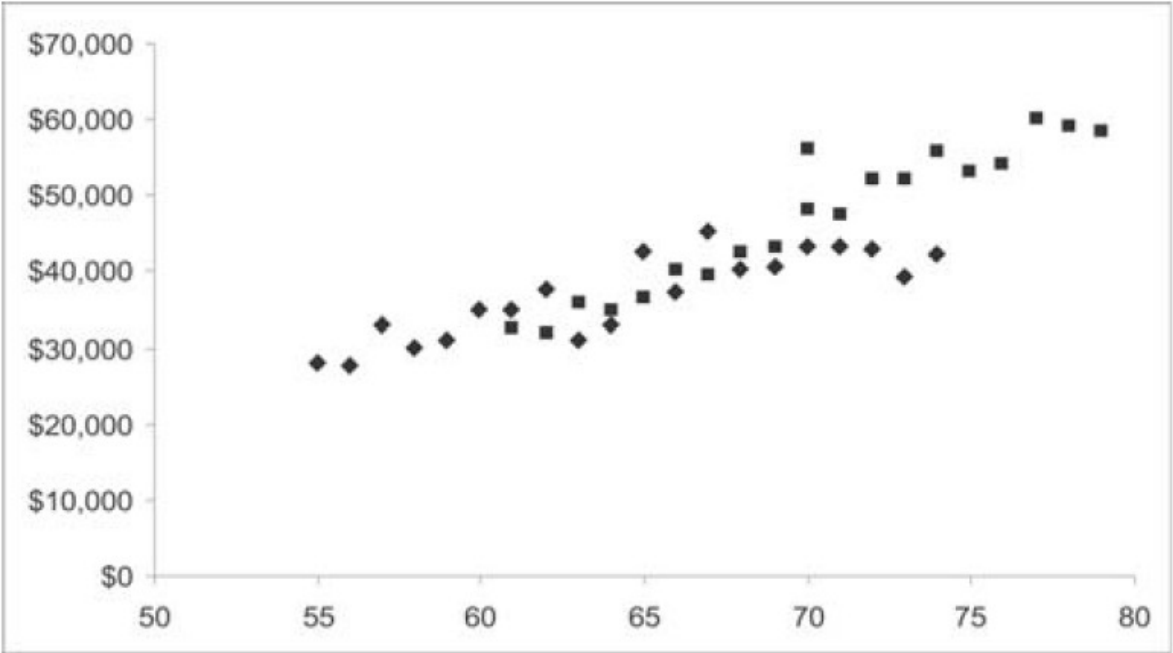
Display a spatial variable



Kendrick Perkins



James Harden



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.

Microsoft Excel - fischer.iris.2.xls

File Edit View Insert Format Tools Data Window Help

Type a question for help

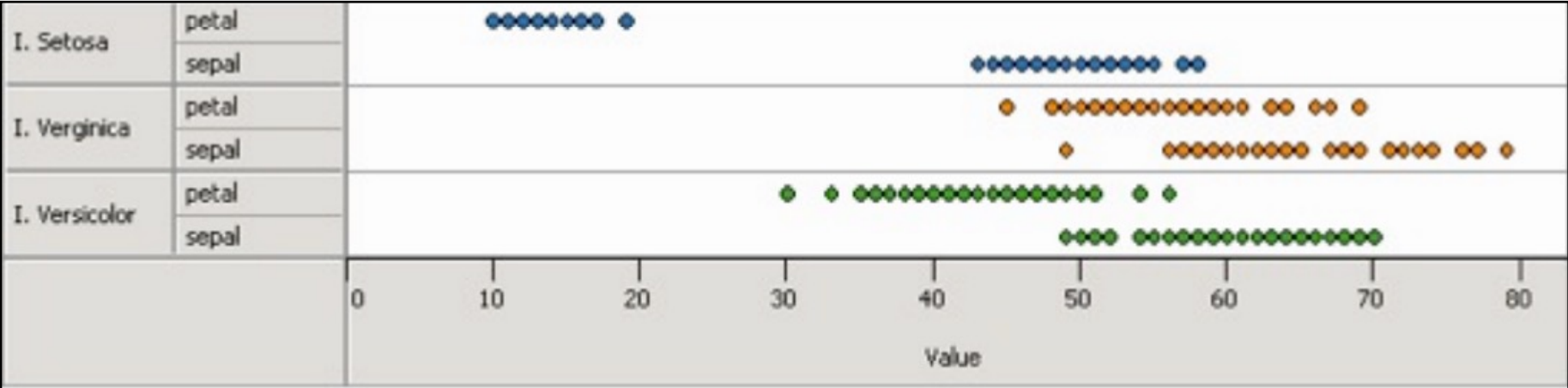
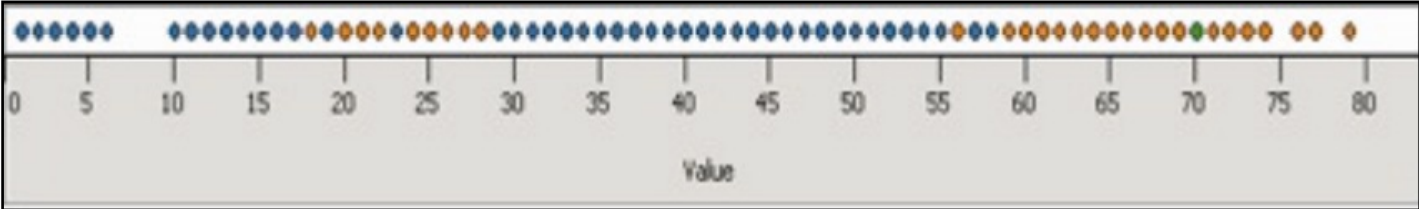
	A	B	C	D	E	F	G	H	I	J
1	ID	Case	Species_No	Species	Organ	Width	Length			
2	1	1	1	I. Setosa	Petal	2	14			
3	2	1	3	I. Verginica	Petal	24	56			
4	3	1	2	I. Versicolor	Petal	13	45			
5	4	1	1	I. Setosa	Sepal	33	50			
6	5	1	3	I. Verginica	Sepal	31	67			
7	6	1	2	I. Versicolor	Sepal	28	57			
8	7	2	1	I. Setosa	Petal	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			

Ready

[Fisher, 1936]

Example 1: Cannot express the facts

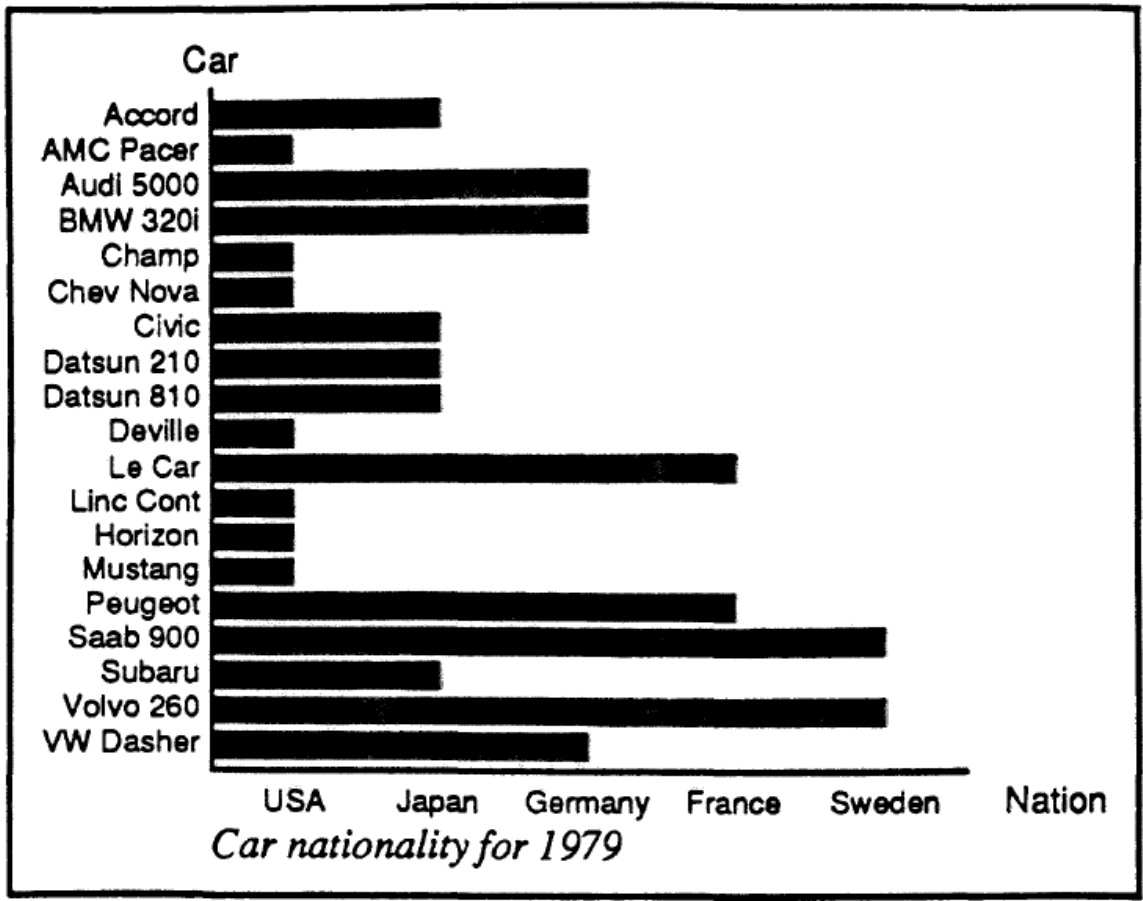
A one-to-many (1->N) relation cannot be expressed in a single horizontal dot plot because multiple tuples are mapped to the same position.

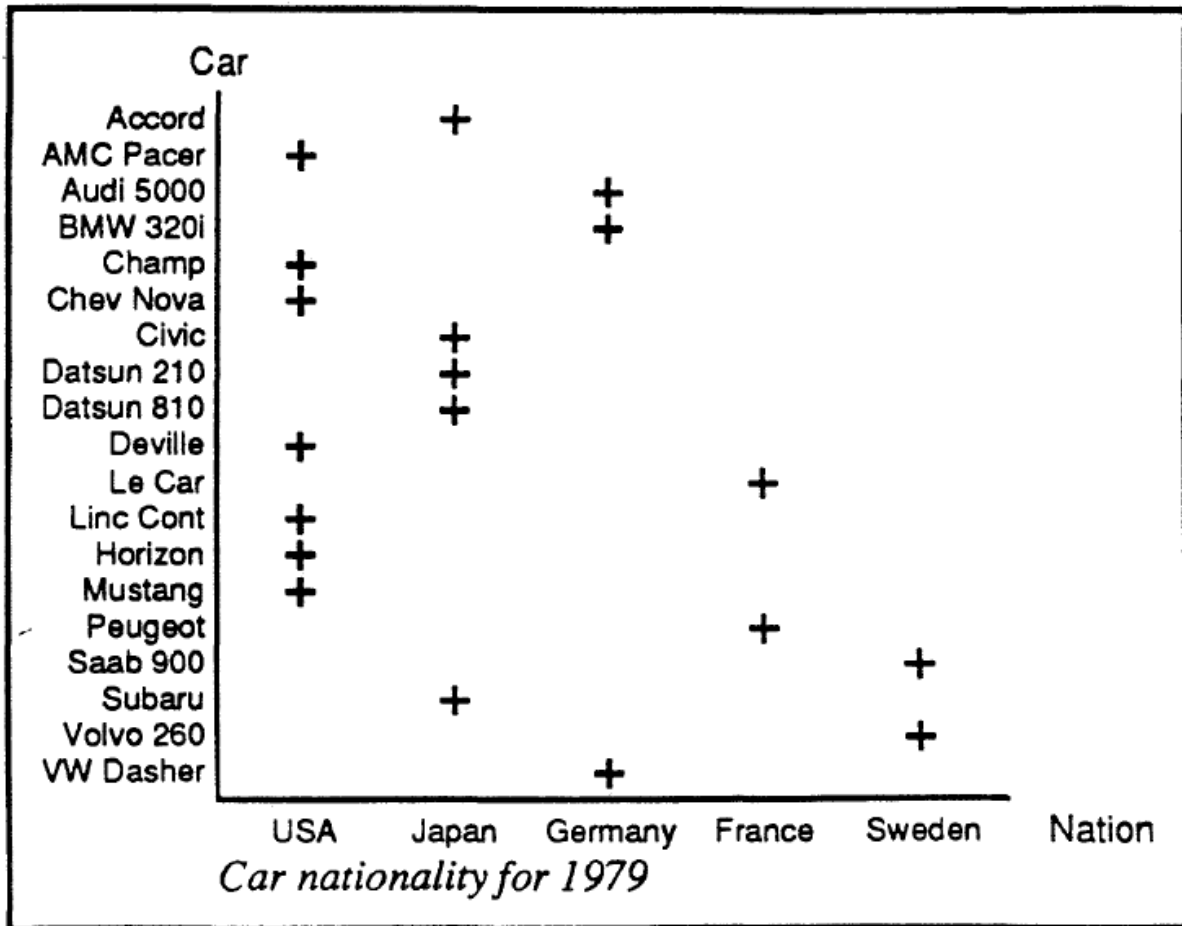


Example 2: Express facts not in the data

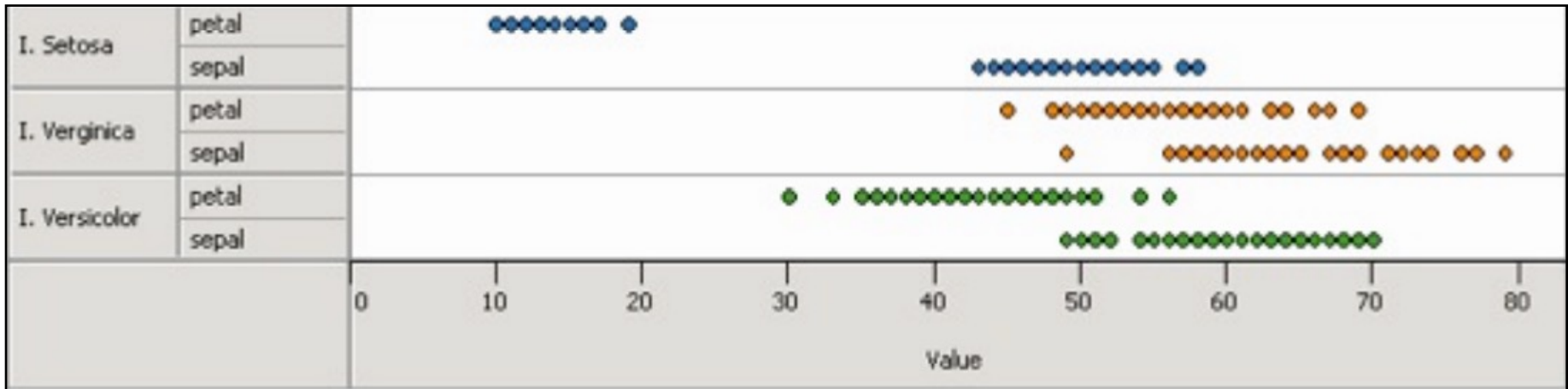
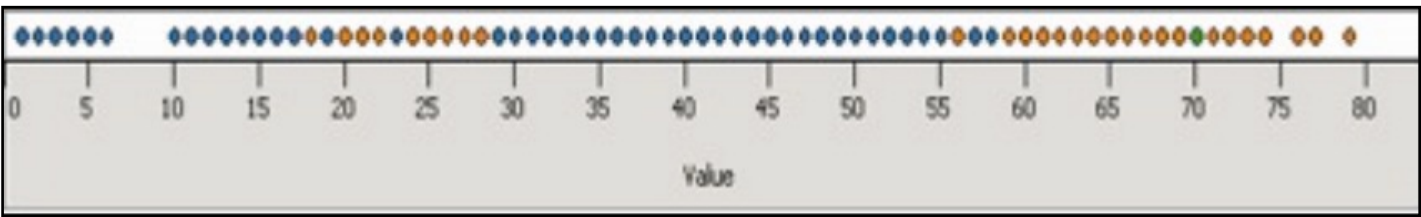
A length is interpreted as a quantitative value;

Length of bar says something untrue about Nominal data

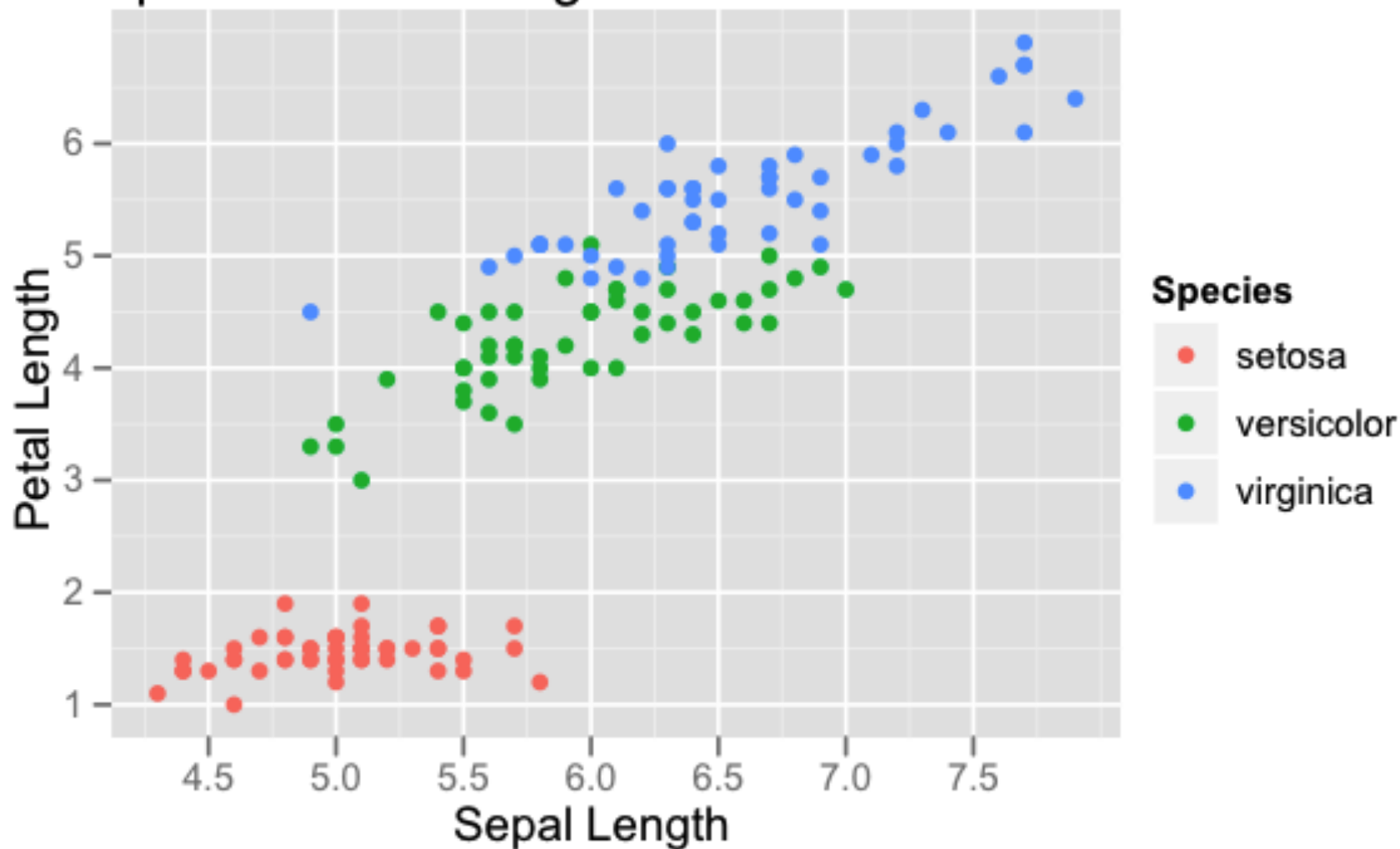




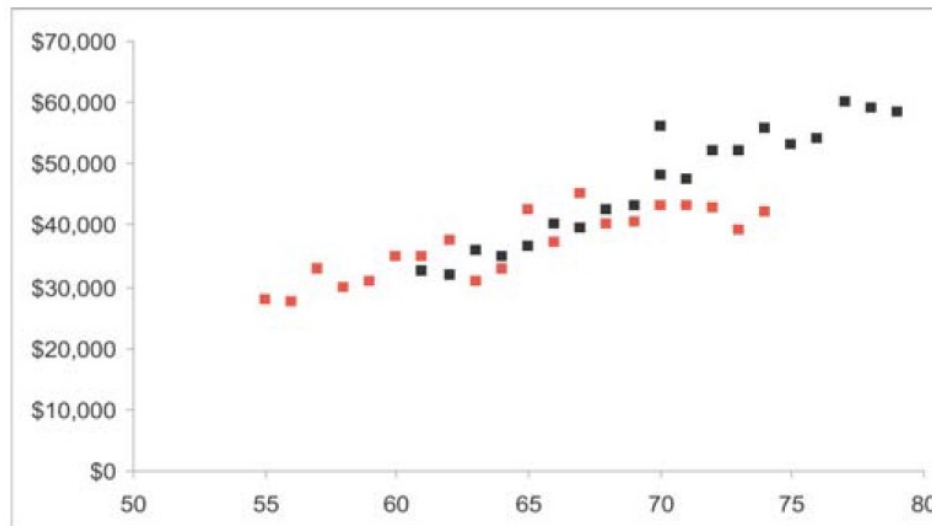
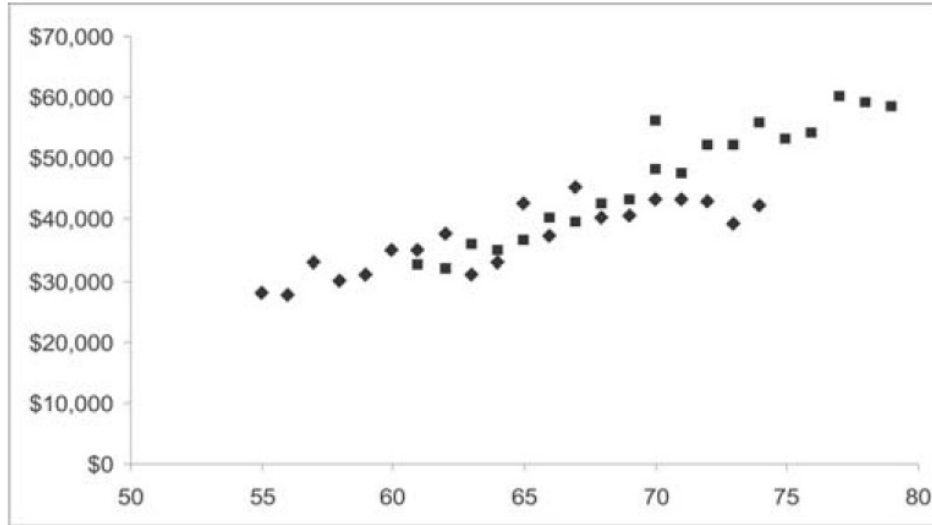
Example 4: Effectiveness



Sepal vs. Petal Length in Fisher's Iris data



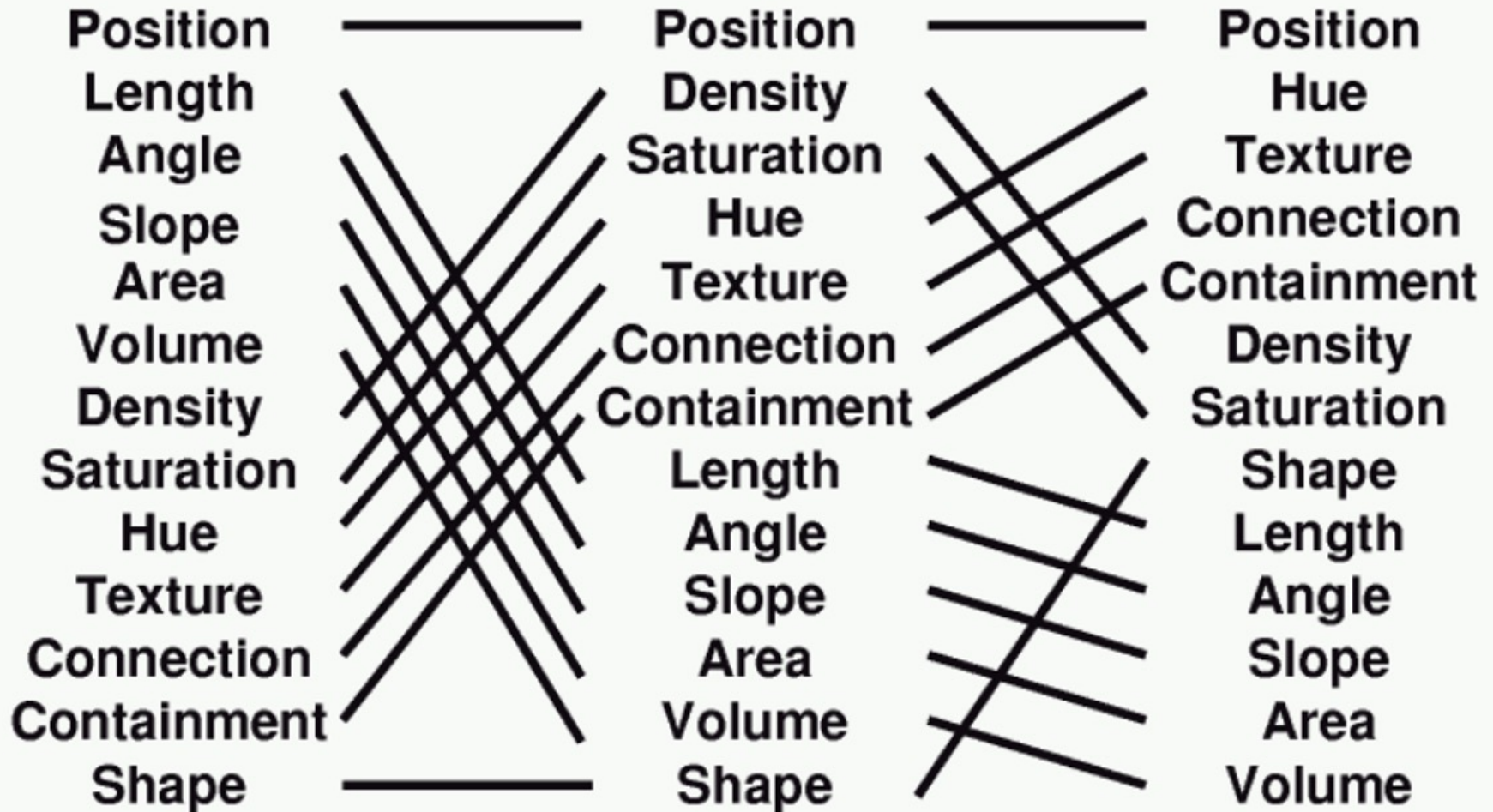
Example 3: Effectiveness



Quantitative

Ordinal

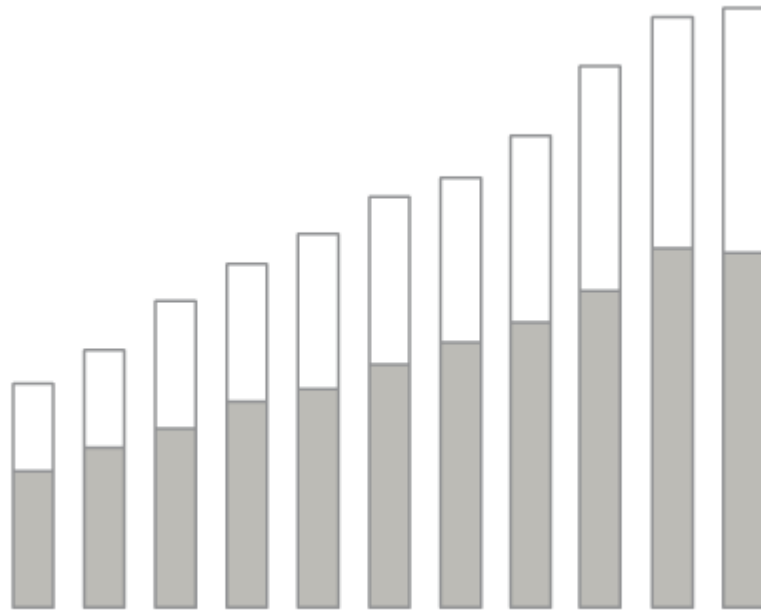
Nominal



Tufte's Rules

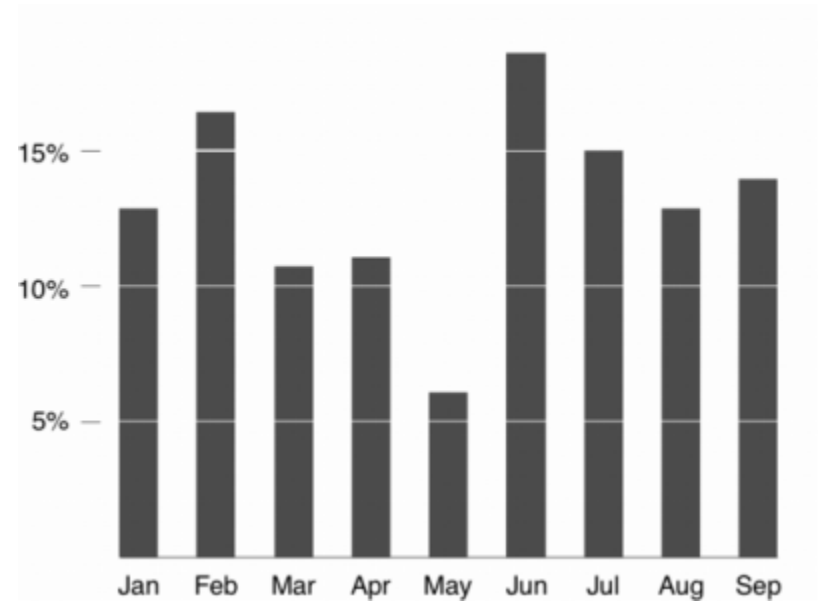
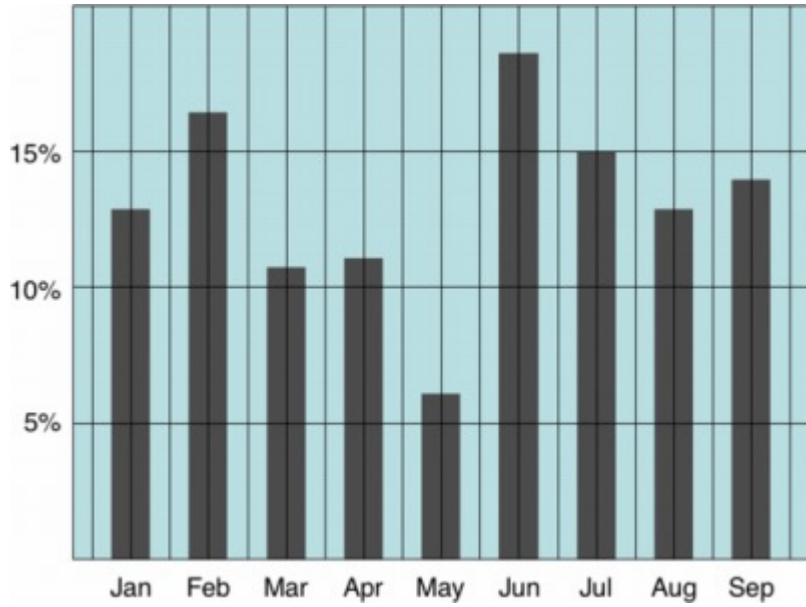
[http://www.sealthreinhold.com/tuftes-rules/rule_three.php]

Avoid Chartjunk



Maximize Data-Ink Ratio

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



Interaction

COVID-19 Dashboard by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU)

Last Updated at (M/D/YYYY)
3/28/2022, 8:20 AM

Total Cases
481,026,349

Total Deaths
6,124,193

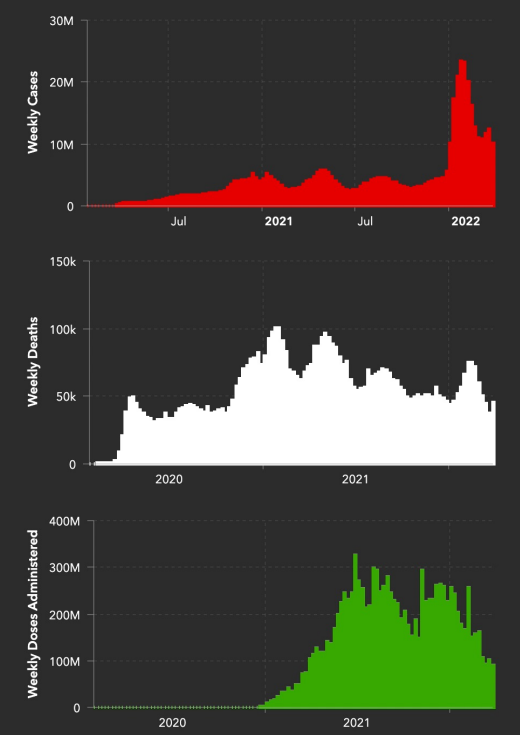
Total Vaccine Doses Administered
10,894,100,192

28-Day Cases
45,169,179

28-Day Deaths
178,916


28-Day Vaccine Doses Administered
366,101,901

- Cases** | Deaths by Country/Region/Sovereignty
- Korea, South**
28-Day: **8,868,598** | 7,128
Totals: **12,003,054** | 15,186
 - Vietnam**
28-Day: **5,690,468** | 2,162
Totals: **9,011,473** | 42,306
 - Germany**
28-Day: **4,712,847** | 4,886
Totals: **19,492,672** | 127,599
 - France**
28-Day: **2,354,756** | 3,556
Totals: **25,216,913** | 142,706
 - United Kingdom**
28-Day: **1,910,367** | 3,249
Totals: **20,848,913** | 165,046
 - Italy**
28-Day: **1,600,165** | 4,222
Totals: **14,364,723** | 158,782
 - Netherlands**
28-Day: **1,515,445** | 406
Totals: **7,929,975** | 22,515
 - Russia**
28-Day: **1,448,686** | 16,087
Totals: **17,525,184** | 360,347



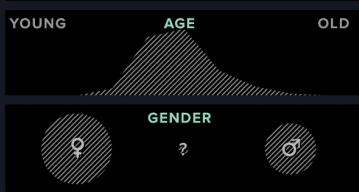
DEMOGRAPHICS

CITY



YOUNG AGE OLD

GENDER



POSE

UP

LEFT RIGHT

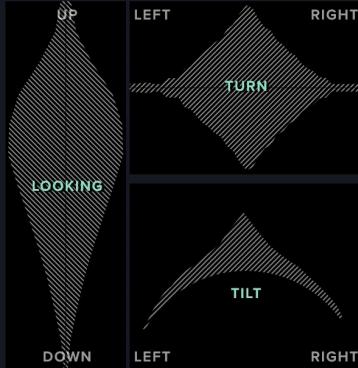
TURN

LOOKING

DOWN

LEFT RIGHT

TILT



FEATURES

EYES

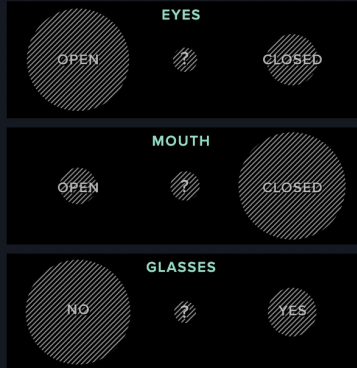
OPEN ? CLOSED

MOUTH

OPEN ? CLOSED

GLASSES

NO ? YES

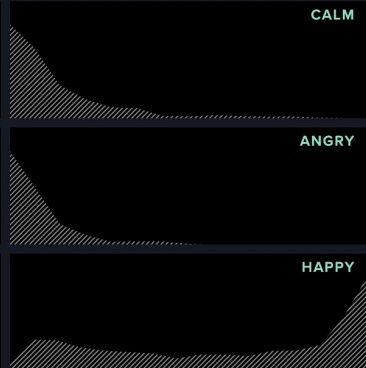


MOOD

CALM

ANGRY

HAPPY



3840 of 3840 selfies.



Animation

Why Use Animation?

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

Expectation

The expectation of a random variable is a number that attempts to capture the center of that random variable's distribution. It can be interpreted as the long-run average of many independent samples from the given distribution. More precisely, it is defined as the probability-weighted sum of all possible values in the random variable's support,

$$E[X] = \sum_{x \in \mathcal{X}} xP(x)$$

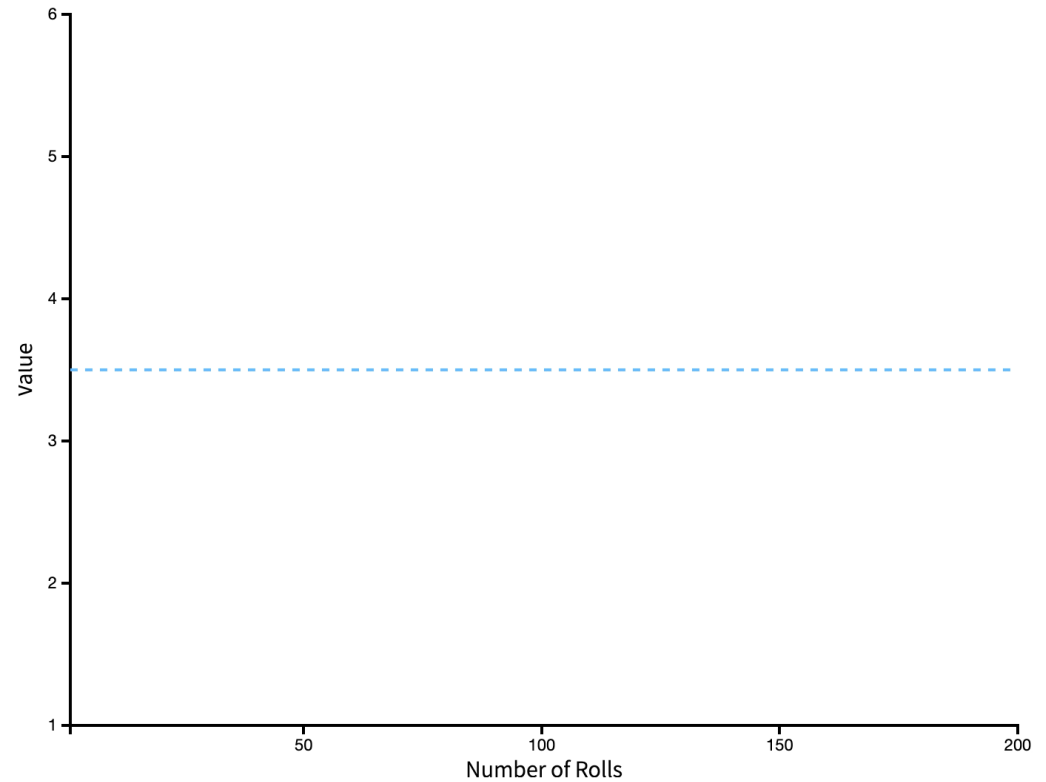
Consider the probabilistic experiment of rolling a fair die and watch as the running sample mean converges to the expectation of 3.5.



Roll the Die

Roll 100 times

Change the distribution of the different faces of the die (thus making the die biased or "unfair") by adjusting the **blue** bars below and observe how this changes the expectation.



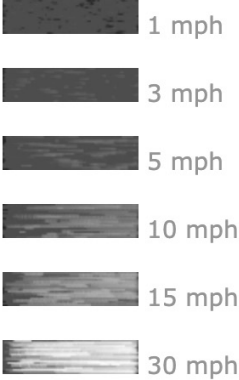
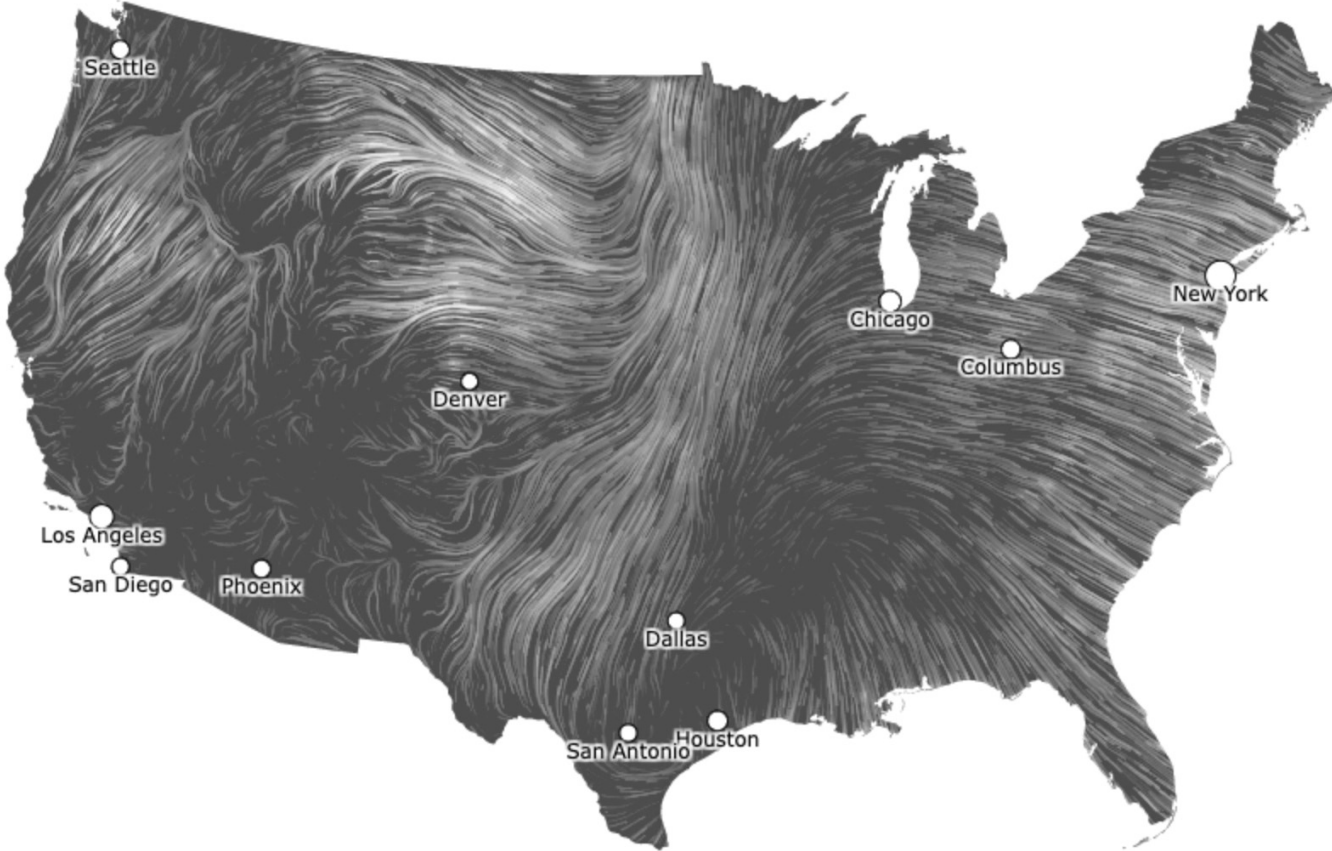
wind map

March 12, 2022

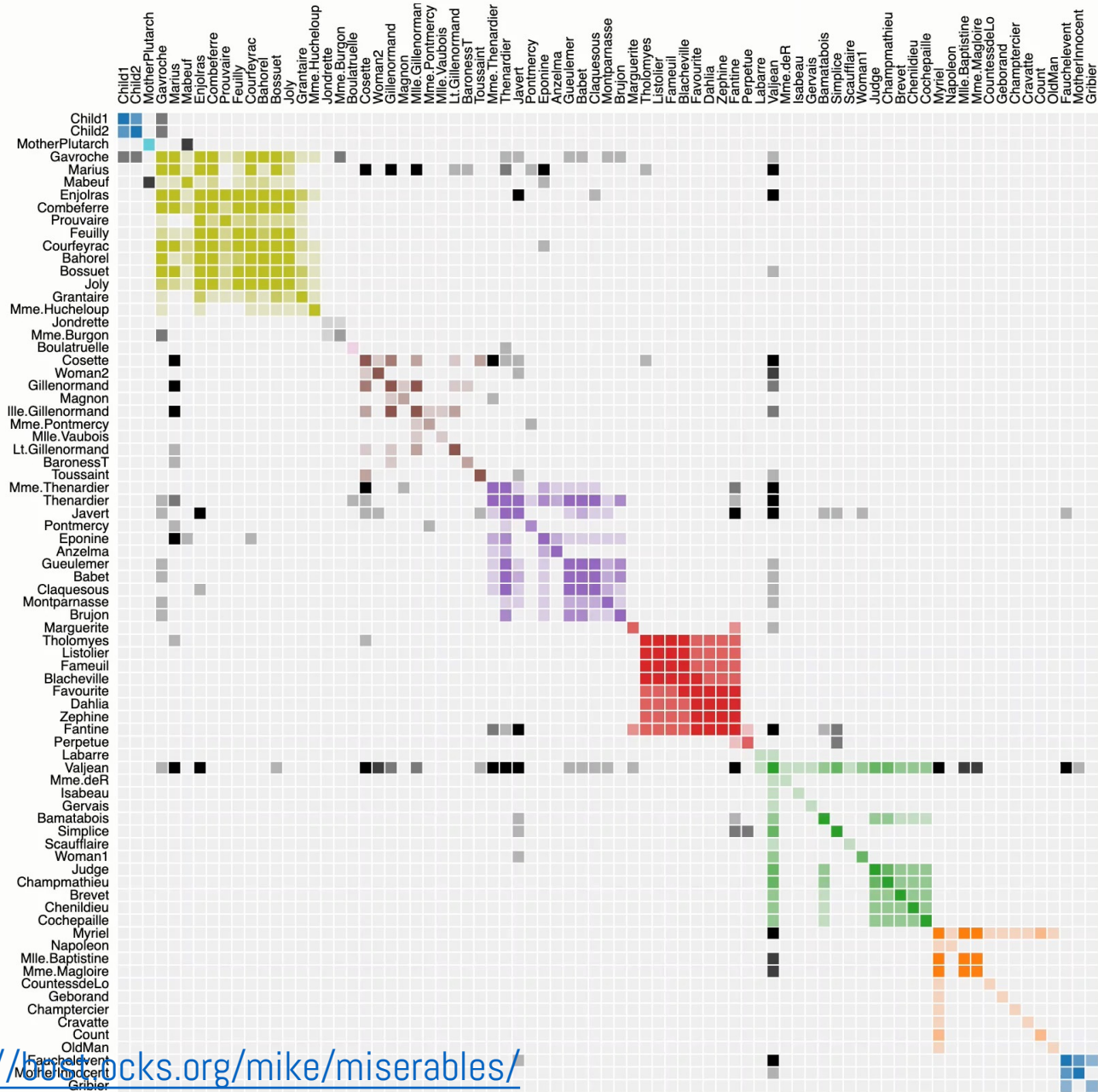
5:12 pm EST

(time of forecast download)

top speed: **50.5 mph**
average: **14.4 mph**



Les Misérables Co-occurrence



Courses

- Arvind Satyanarayan
 - <http://vis.csail.mit.edu/classes/6.894/>
- 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/>

Tools

- Vega ecosystem
 - <https://vega.github.io/vega/>
 - <https://vega.github.io/vega-lite/>
 - <https://altair-viz.github.io/>
- Tableau
 - <https://public.tableau.com/en-us/s/>
- Einblick
 - <https://www.einblick.ai/>
 - Fell free to ping me (ez@einblick.ai) with questions, feedback, etc.

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