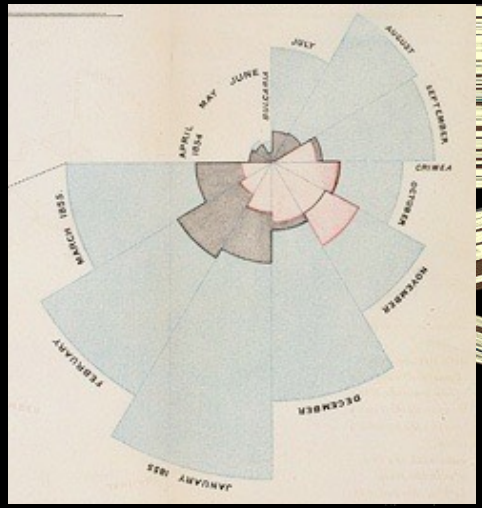
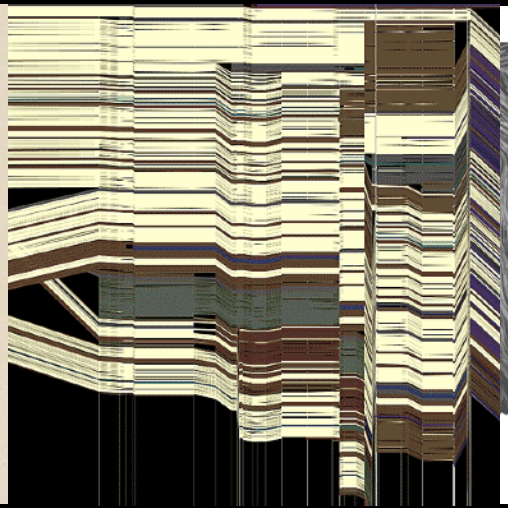
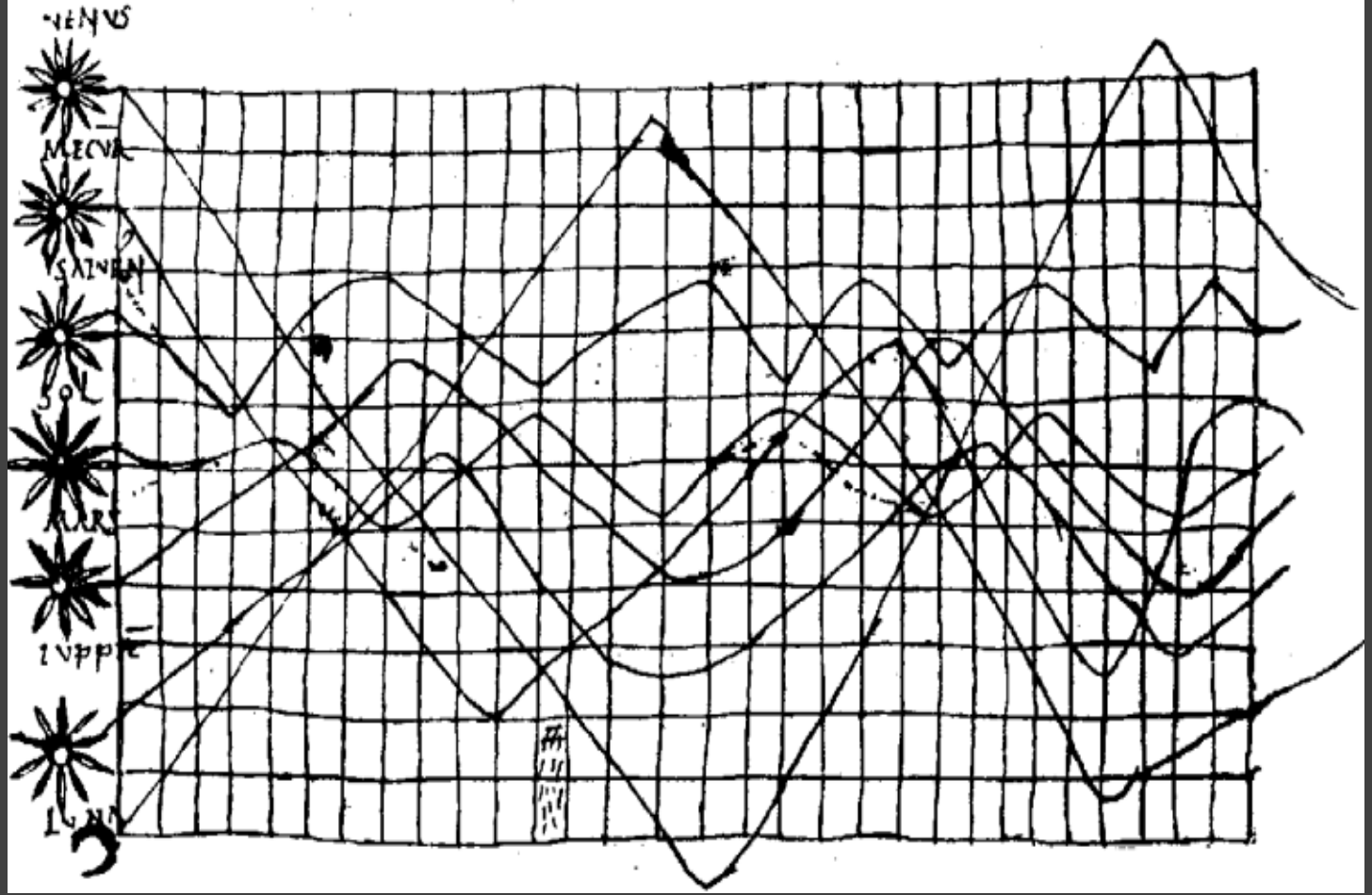
CSE P 590A - Data Visualization

Introduction

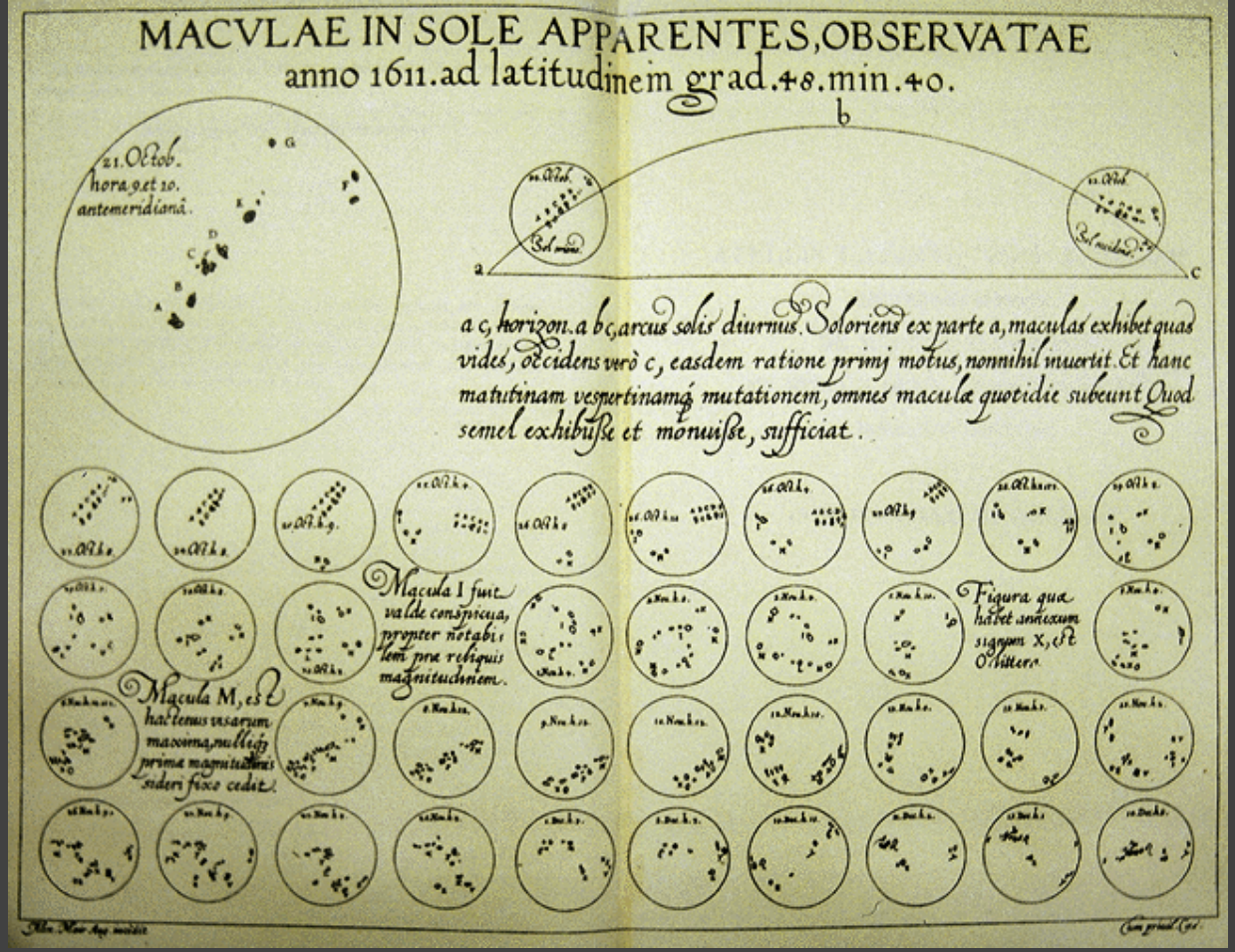




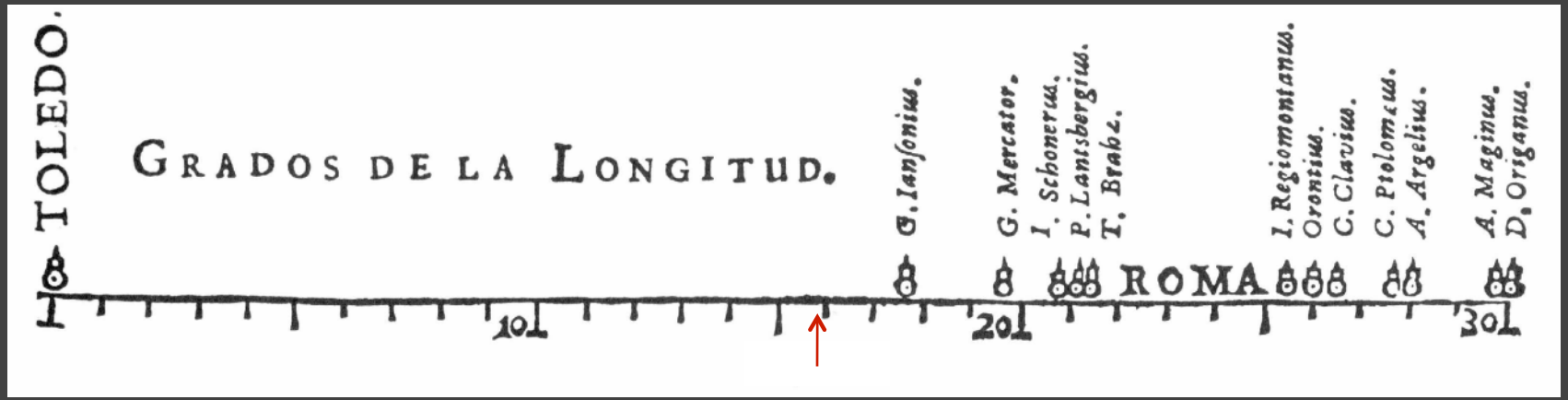
Jeffrey Heer University of Washington



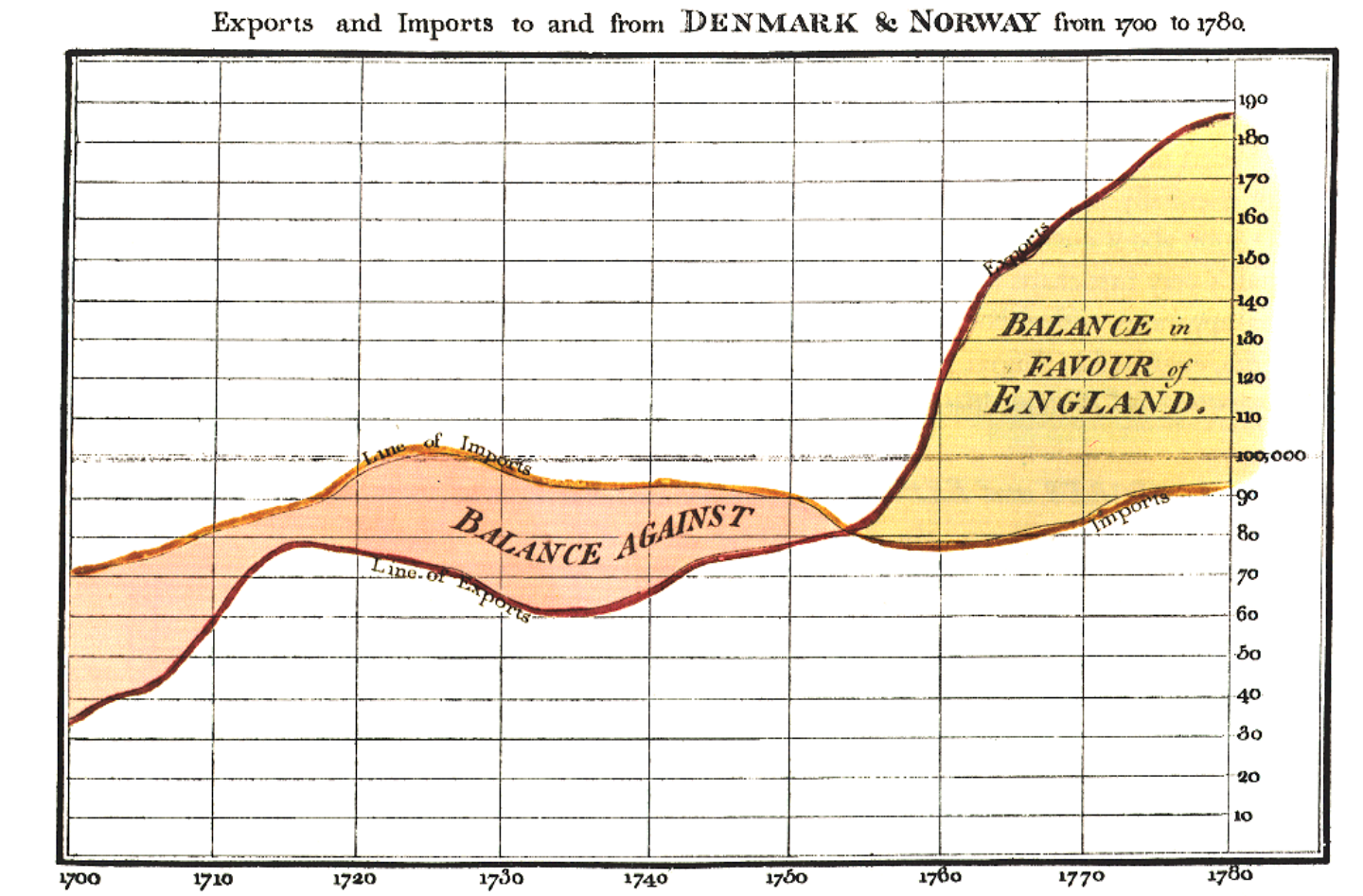
~950 AD Position of Sun, Moon and Planets



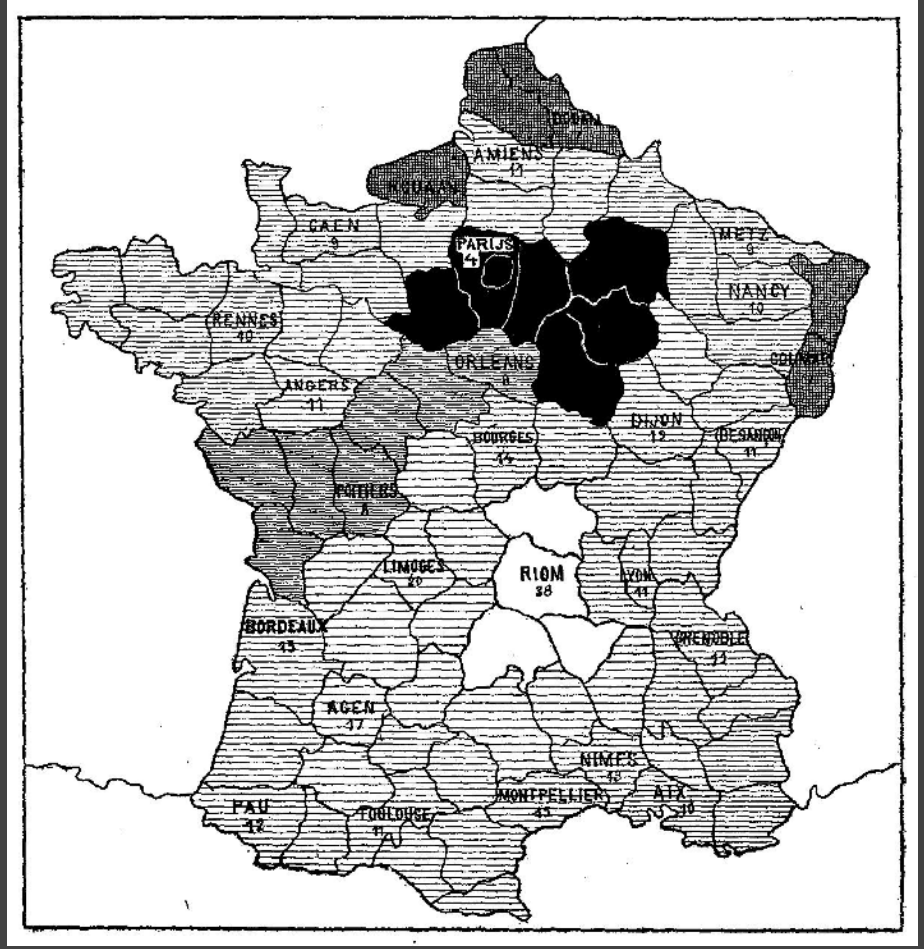
Sunspots over time, Scheiner 1626



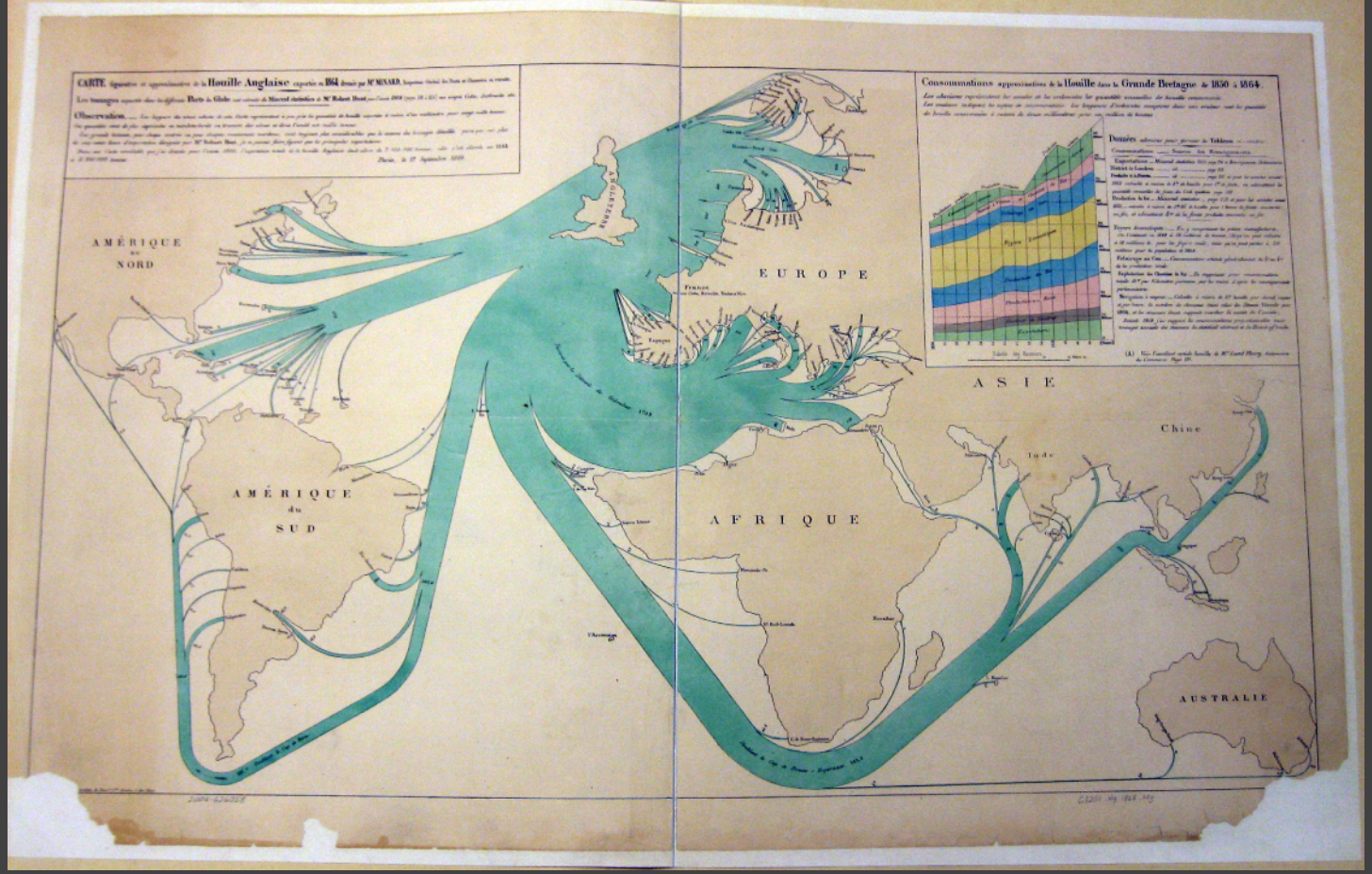
Longitudinal distance between Toledo and Rome, van Langren 1644



The Commercial and Political Atlas, William Playfair 1786



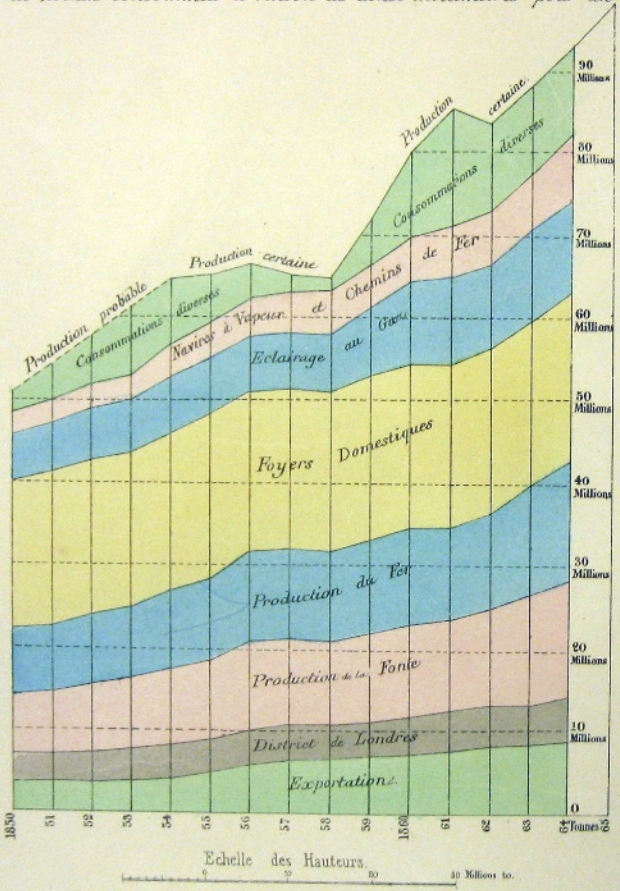
1826(?) Illiteracy in France, Pierre Charles Dupin



1864 British Coal Exports, Charles Minard

Consommations approximatives de la Houille dans la Grande Bretagne de 1850 à 1864.

Les abscisses représentent les années et les ordonnées les quantités annuelles de houille consommée.
Les couleurs indiquent les espèces de consommations. Les longueurs d'ordonnées comprises dans une couleur sont les quantités de houille consommées à raison de deux millimètres pour un million de tonnes.



Données admises pour former le Tableau ci-contre.

Consommations. —— Sources des Renseignements.

Exportations. — *Mineral statistics 1865 page 214 et Renseignements Parlementaires.*

District de Londres. —— *id.* —— *page 213*

Produits de la Fonte. —— *id* —— *page 215 et pour les années avant 1855 calculée à raison de 3 to de houille pour 1 to de fonte, en admettant les quantités annuelles de fonte du* Coal question *page 192.*

Production du fer — *Mineral statistics* — *page 215 et pour les années avant 1855* — *calculée à raison de 3 to 35 de houille pour 1 tonne de fonte convertie en fer; et admettant $\frac{9}{10}$ de la fonte produite convertis en fer.*

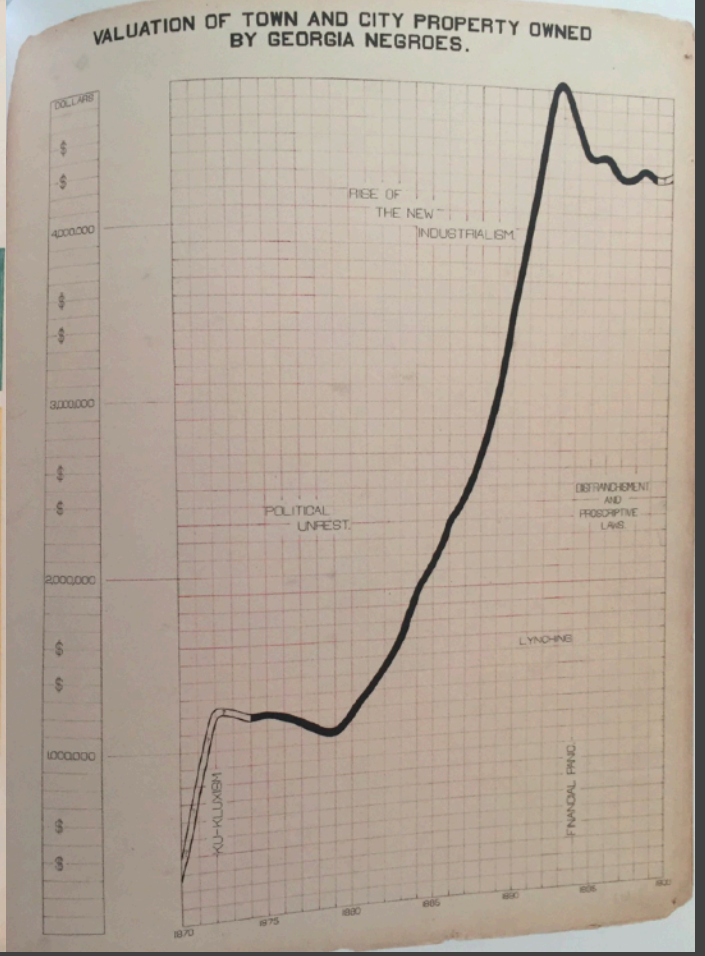
Foyers domestiques. — *En y comprenant les petites manufactures. On l'estimait en 1848 à 19 millions de tonnes, (A) qu'on peut réduire à 18 millions to. pour les foyers seuls, mais qu'on peut porter à 20 millions pour la population de 1864.*

Eclairage au Gaz. — *Consommation estimée généralement du $\frac{1}{9}$e au $\frac{1}{8}$e de la production totale.*

Exploitation des Chemins de Fer. — *En supposant pour consommation totale 10 k par Kilomètre parcouru par les trains d'après les renseignements parlementaires.*

Navigation à vapeur. — *Calculée à raison de 5 k houille par cheval vapeur et par heure, le nombre de chevaux étant celui du Steam Vessels pour 1864, et les steamers étant supposés marcher la moitié de l'année; Avant 1864 j'ai supposé les consommations proportionnelles aux tonnages annuels des steamers du statistical abstract et du Board of trade.*

(A) *Voir l'excellent article houille de M. Lamé Fleury, Dictionnaire du Commerce Page 111.*



1900 Visualizing Black America , W. E. B. DuBois et al.

Entering the 1900s…

Rise of **formal statistical methods** in the physical and social sciences

Little innovation in graphical methods

A period of **application and popularization**

Graphical methods enter textbooks, curricula, and **mainstream use**

Four major influences act on data analysis today:

1. The formal theories of statistics.
2. Accelerating developments in computers and display devices.
3. The challenge, in many fields, of more and larger bodies of data.
4. The emphasis on quantification in a wider variety of disciplines.

Data Analysis & Statistics, Tukey & Wilk 1966

While some of the influences of statistical theory on data analysis have been helpful, others have not.

Data Analysis & Statistics, Tukey & Wilk 1966

Exposure, the effective laying open of the data to **display the unanticipated**, is to us a major portion of data analysis…

It is not clear how the **informality** and **flexibility** appropriate to the **exploratory character** of exposure can be fitted into any of the structures of formal statistics so far proposed.

Data Analysis & Statistics, Tukey & Wilk 1966

Accordingly, both approaches and techniques need to be structured so as to **facilitate human involvement and intervention**.

Some implications for effective analysis are: (1) it is essential to have convenience of **interaction of people and intermediate results** and (2) at all stages of data analysis, the outputs need to be **matched to the capabilities of the people who use it and want it.**

Data Analysis & Statistics, Tukey & Wilk 1966

Set A		Set B		Set C		Set D	
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	9.11	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

Summary Statistics

$u_X = 9.0$ $\sigma_X = 3.317$

$u_Y = 7.5$ $\sigma_Y = 2.03$

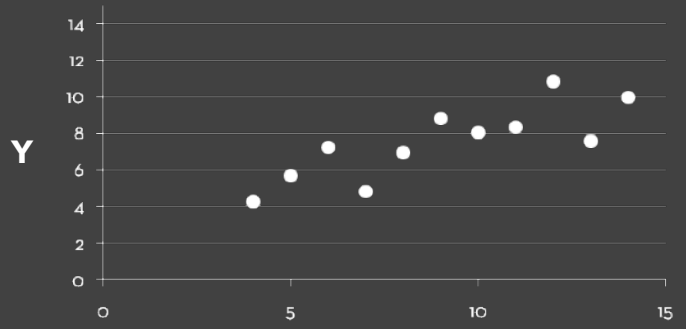
Linear Regression

$Y = 3 + 0.5\ X$

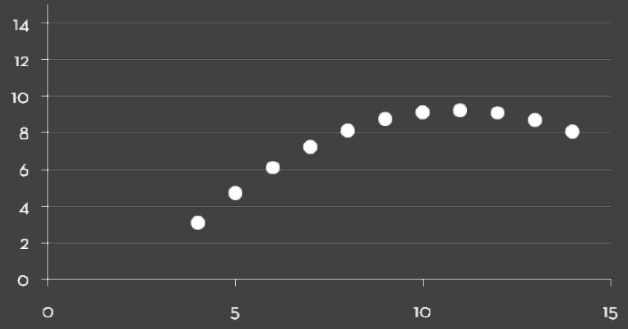
$R^2 = 0.67$

[Anscombe 1973]

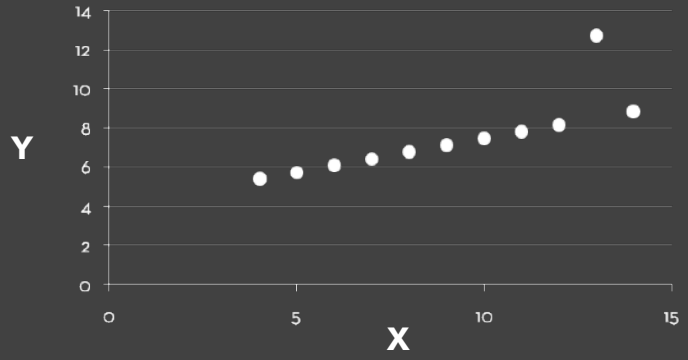
Set A



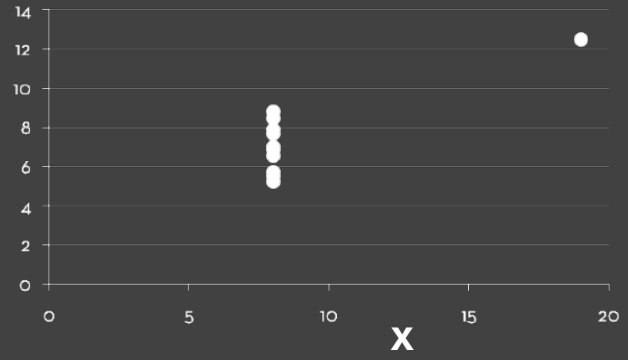
Set B



Set C



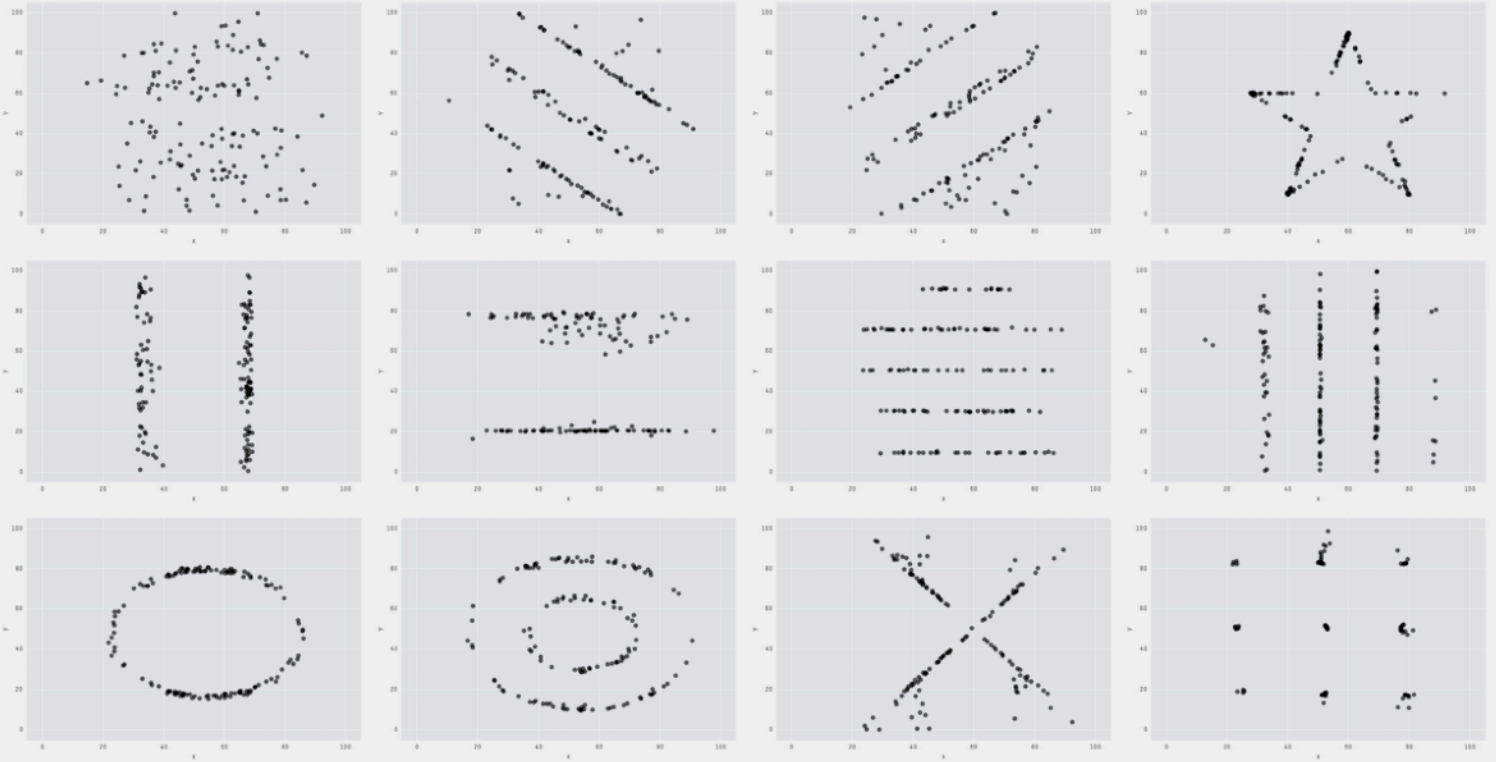
Set D



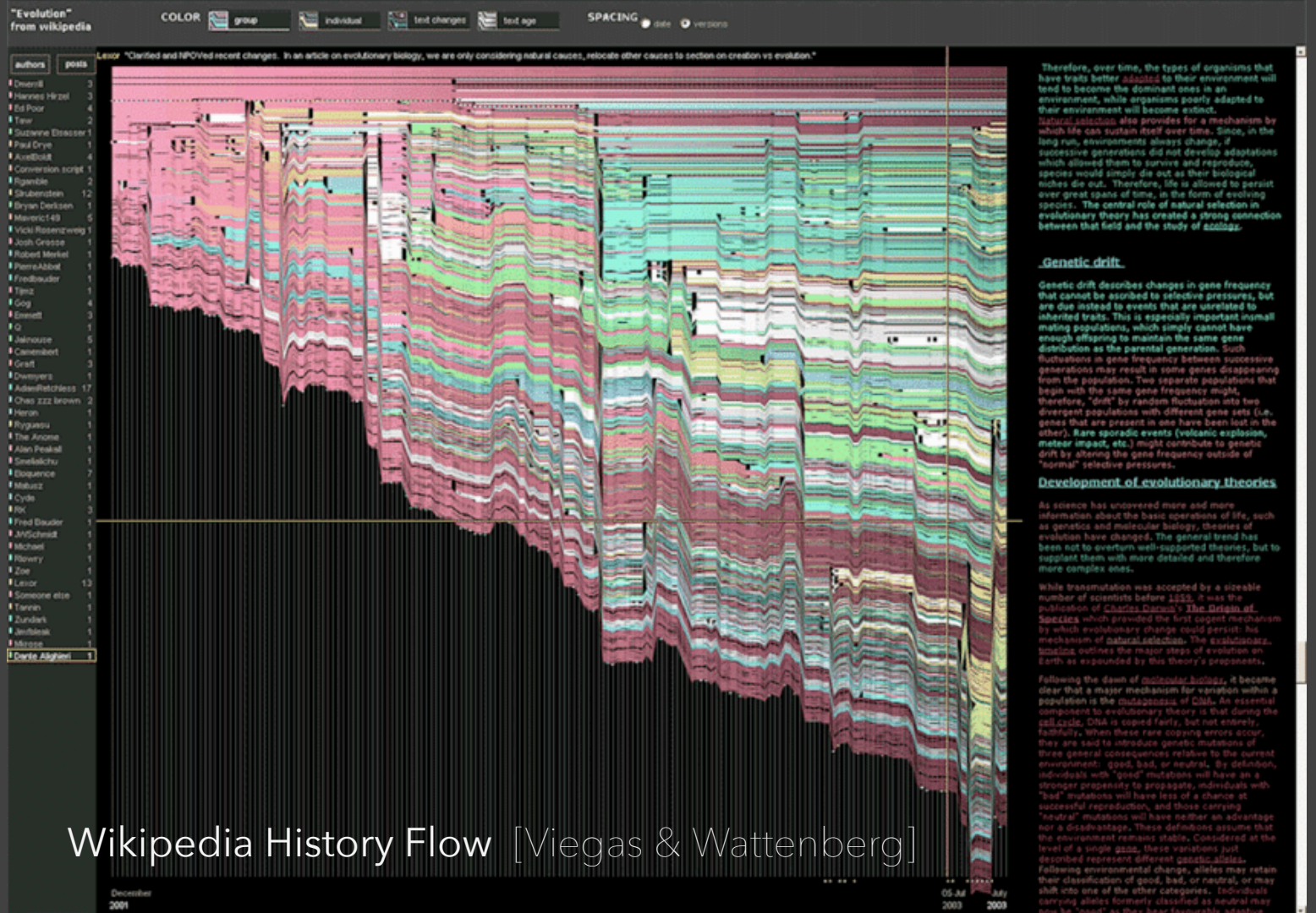
[Anscombe 1973]

X Mean: 54.26
Y Mean: 47.83
X SD : 16.76
Y SD : 26.93
Corr. : -0.06

[Matejka 2017]



Example:
Wikipedia Edits

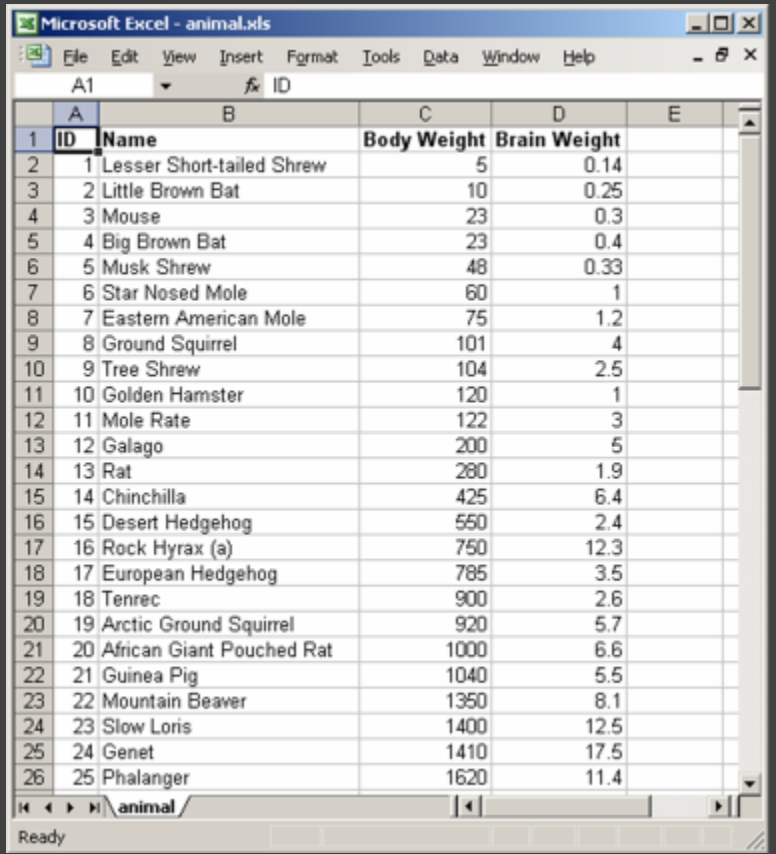


Wikipedia History Flow [Viegas & Wattenberg]

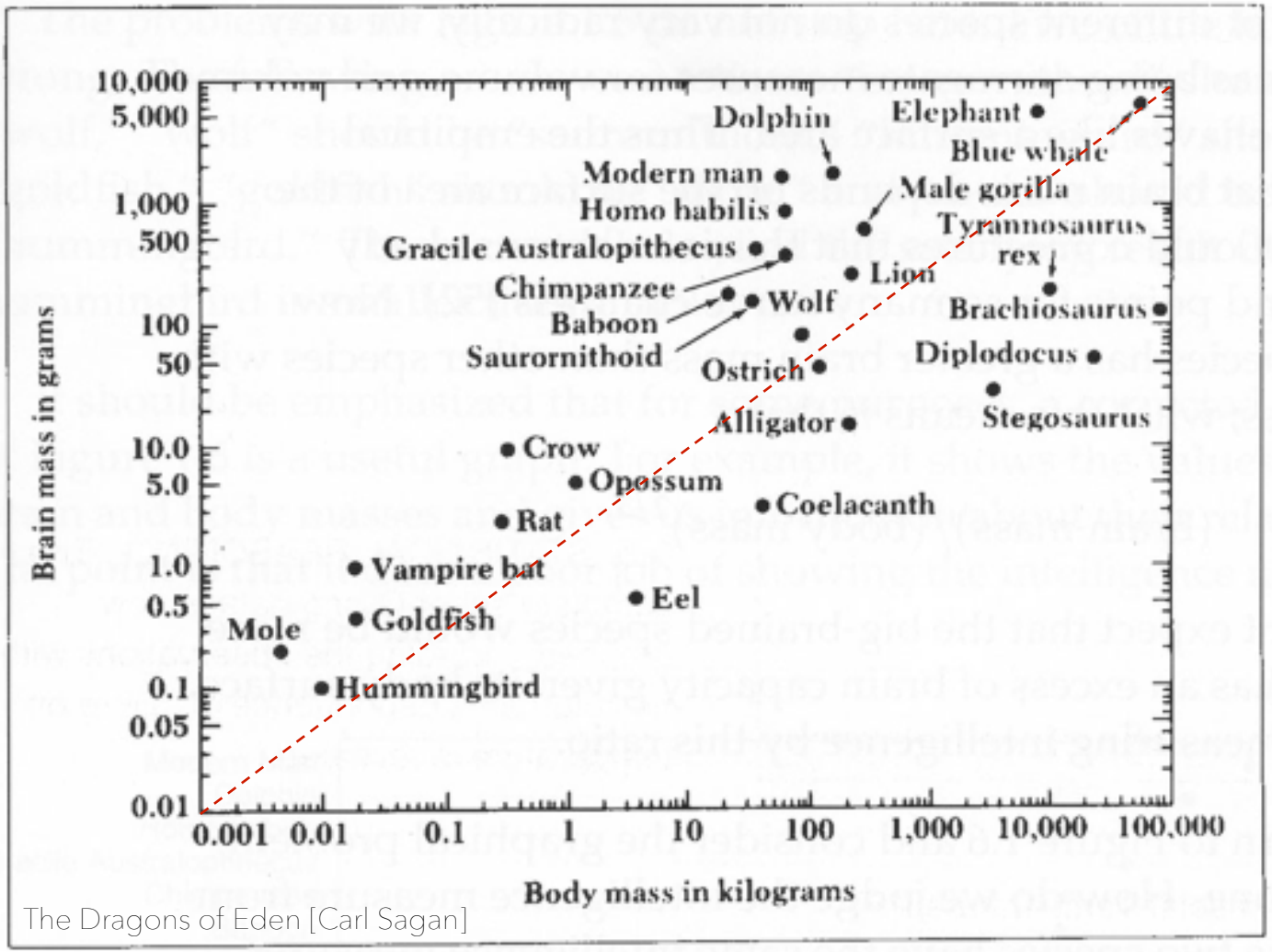
Wikipedia History Flow [Viegas & Wattenberg]

Example:
Animal Brains

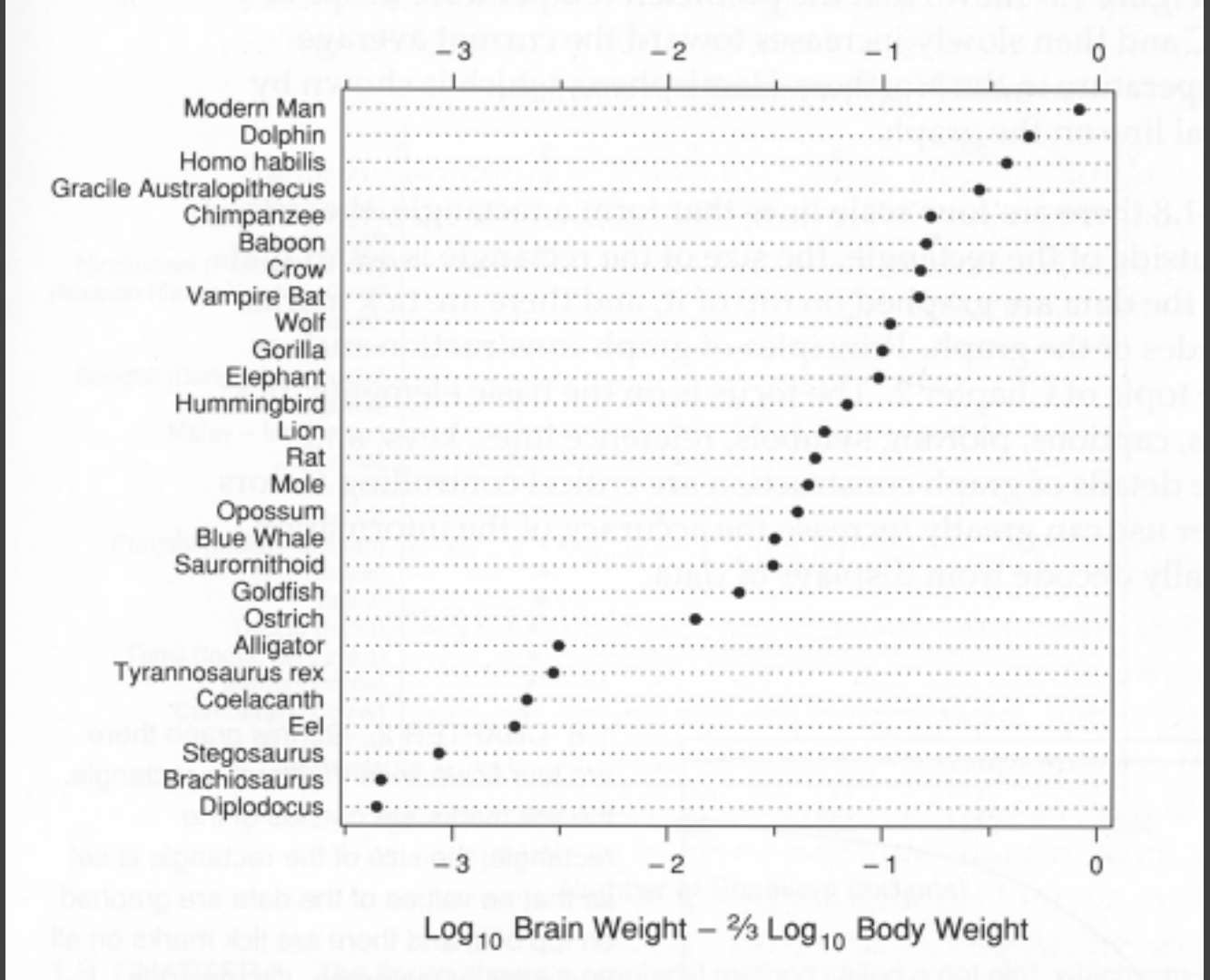
Which animals are the “smartest”?



ID	Name	Body Weight	Brain Weight
1	Lesser Short-tailed Shrew	5	0.14
2	Little Brown Bat	10	0.25
3	Mouse	23	0.3
4	Big Brown Bat	23	0.4
5	Musk Shrew	48	0.33
6	Star Nosed Mole	60	1
7	Eastern American Mole	75	1.2
8	Ground Squirrel	101	4
9	Tree Shrew	104	2.5
10	Golden Hamster	120	1
11	Mole Rate	122	3
12	Galago	200	5
13	Rat	280	1.9
14	Chinchilla	425	6.4
15	Desert Hedgehog	550	2.4
16	Rock Hyrax (a)	750	12.3
17	European Hedgehog	785	3.5
18	Tenrec	900	2.6
19	Arctic Ground Squirrel	920	5.7
20	African Giant Pouched Rat	1000	6.6
21	Guinea Pig	1040	5.5
22	Mountain Beaver	1350	8.1
23	Slow Loris	1400	12.5
24	Genet	1410	17.5
25	Phalanger	1620	11.4

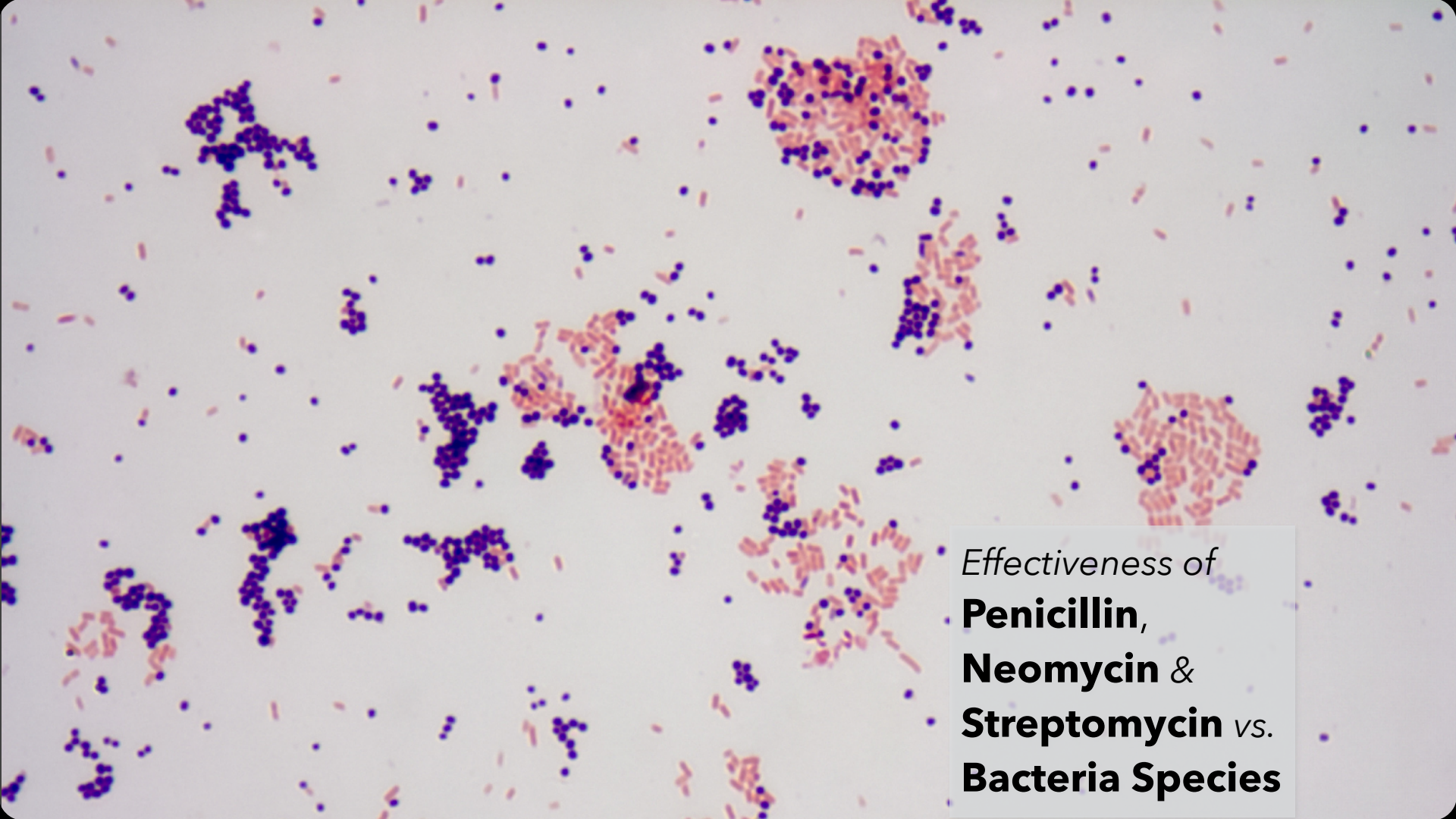


The Dragons of Eden [Carl Sagan]



The Elements of Graphing Data [Cleveland]

Example:
Antibiotic Effectiveness

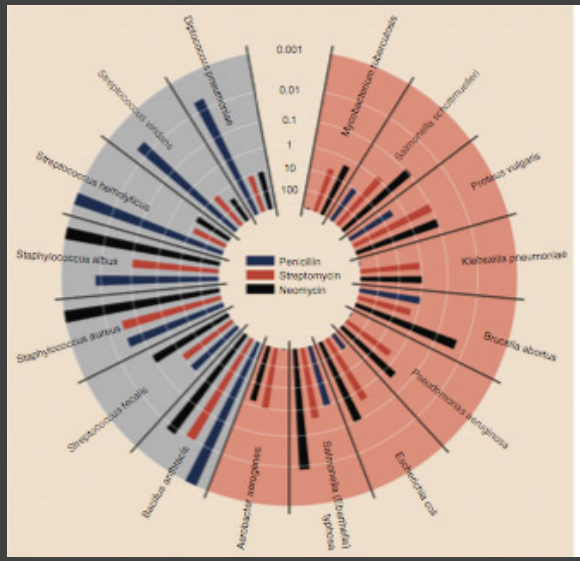

Effectiveness of
Penicillin,
Neomycin *&*
Streptomycin *vs.*
Bacteria Species

What questions might we ask?

Table 1: Burtin's data.

Bacteria	Antibiotic			Gram Staining
	Penicillin	Streptomycin	Neomycin	
Aerobacter *aerogenes*	870	1	1.6	negative
Brucella *abortus*	1	2	0.02	negative
Brucella *anthracis*	0.001	0.01	0.007	positive
Diplococcus *pneumoniae*	0.005	11	10	positive
Escherichia *coli*	100	0.4	0.1	negative
Klebsiella *pneumoniae*	850	1.2	1	negative
Mycobacterium *tuberculosis*	800	5	2	negative
Proteus *vulgaris*	3	0.1	0.1	negative
Pseudomonas *aeruginosa*	850	2	0.4	negative
Salmonella (Eberthella) *typhosa*	1	0.4	0.008	negative
Salmonella *schottmuelleri*	10	0.8	0.09	negative
Staphylococcus *albus*	0.007	0.1	0.001	positive
Staphylococcus *aureus*	0.03	0.03	0.001	positive
Streptococcus *fecalis*	1	1	0.1	positive
Streptococcus *hemolyticus*	0.001	14	10	positive
Streptococcus *viridans*	0.005	10	40	positive

Which antibiotic is most effective?



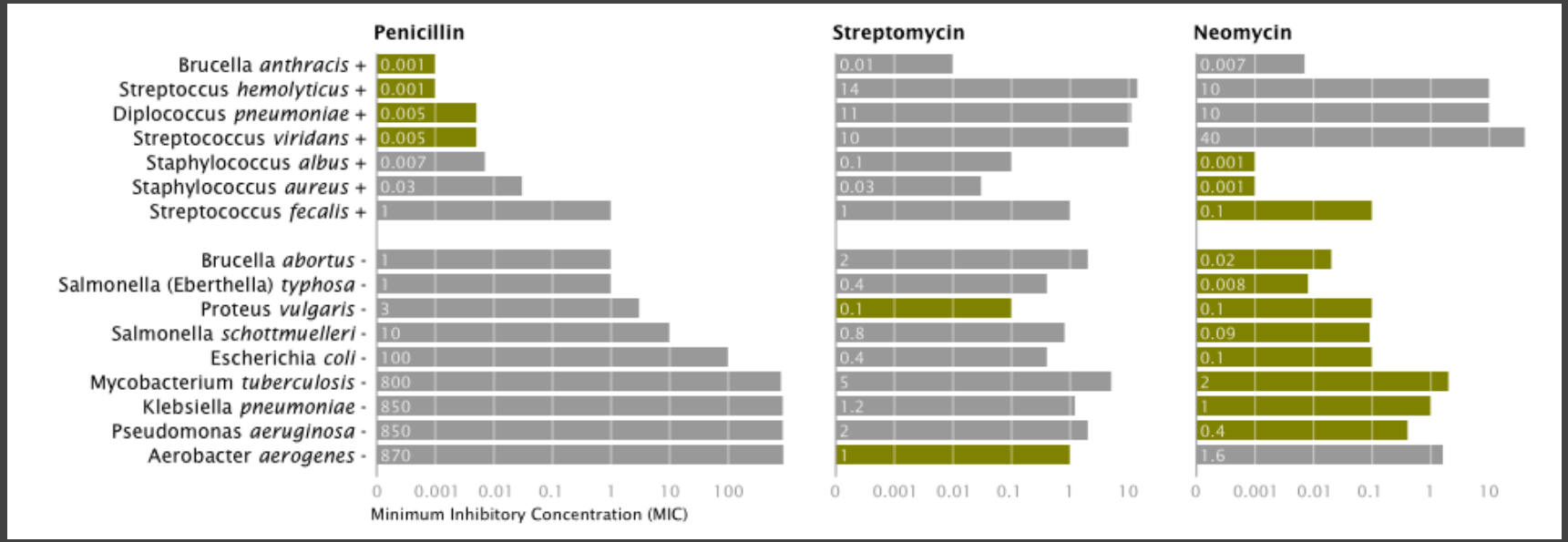
Bacteria	Penicillin	Antibiotic Streptomycin	Neomycin	Gram stain
Aerobacter aerogenes	870	1	1.6	–
Brucella abortus	1	2	0.02	–
Bacillus anthracis	0.001	0.01	0.007	+
Diplococcus pneumoniae	0.005	11	10	+
Escherichia coli	100	0.4	0.1	–
Klebsiella pneumoniae	850	1.2	1	–
Mycobacterium tuberculosis	800	5	2	–
Proteus vulgaris	3	0.1	0.1	–
Pseudomonas aeruginosa	850	2	0.4	–
Salmonella (Eberthella) typhosa	1	0.4	0.008	–
Salmonella schottmuelleri	10	0.8	0.09	–
Staphylococcus albus	0.007	0.1	0.001	+
Staphylococcus aureus	0.03	0.03	0.001	+
Streptococcus fecalis	1	1	0.1	+
Streptococcus hemolyticus	0.001	14	10	+
Streptococcus viridans	0.005	10	40	+

Radius: 1 / log(MIC)

Bar Color: Antibiotic

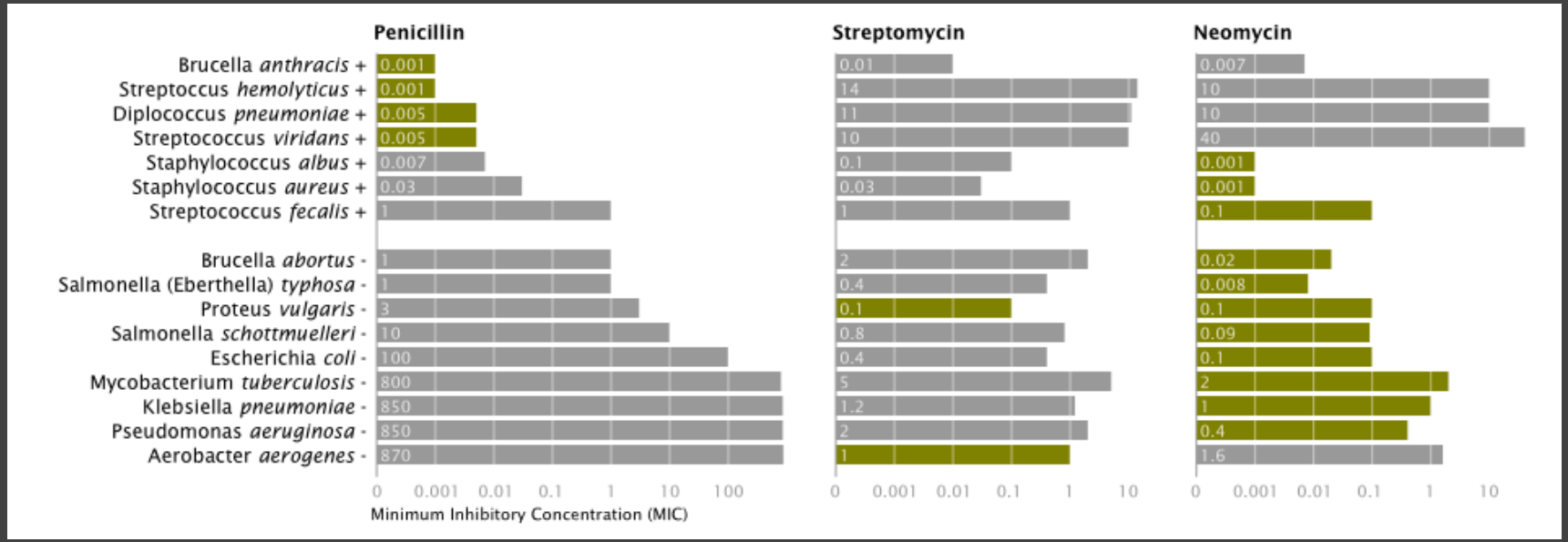
Background Color: Gram Staining

Which antibiotic is most effective?



Mike Bostock

Which antibiotic is most effective?



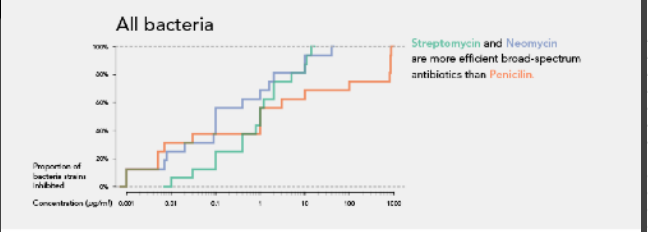
X-axis: Antibiotic | log(MIC)

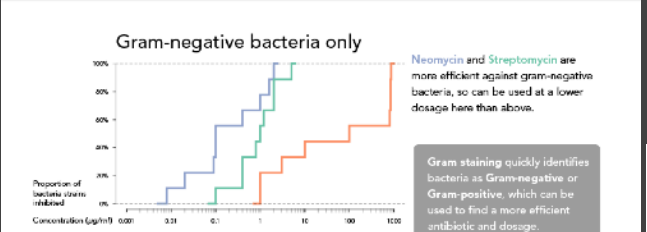
Y-axis: Gram-Staining | Species

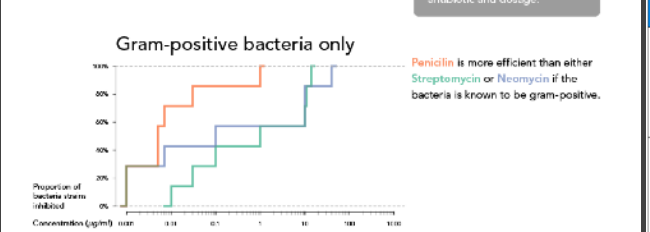
Color: Most-Effective?

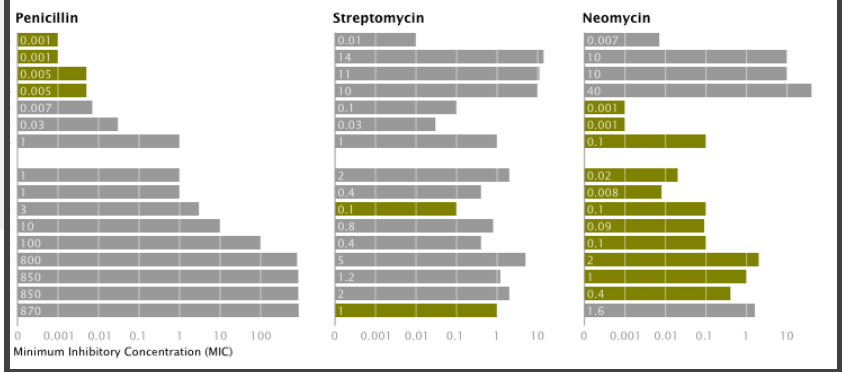
penicillin
streptomycin
neomycin
E. aerogenes
B. abortus
E. coli
K. pneumoniae
M. tuberculosis
P. vulgaris
P. aeruginosa
S. typhi
S. schottmuelleri
B. anthracis
S. pneumoniae
S. albus
S. aureus
E. faecalis
S. pyogenes
S. viridans
Gram positive
Gram negative
0.1 (µg/ml)
1
10
100
500
minimum inhibitory concentration of antibiotics
bowen li
cs448b

Bowen Li

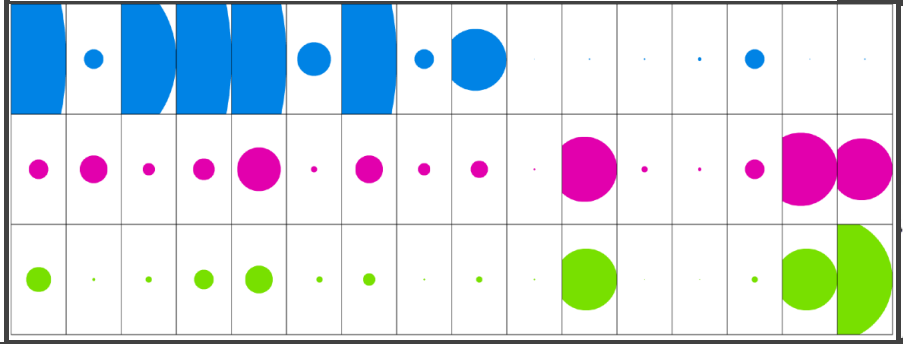

All bacteria
Streptomycin and Neomycin are more efficient broad-spectrum antibiotics than Penicillin.
Proportion of bacteria strains inhibited
Concentration (μg/ml)

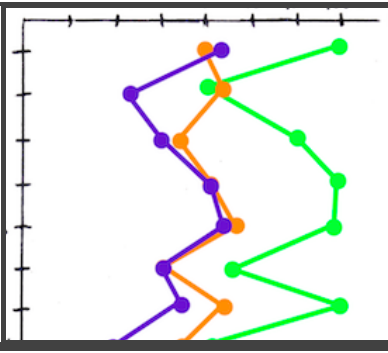

Gram-negative bacteria only
Neomycin and Streptomycin are more efficient against gram-negative bacteria, so can be used at a lower dosage here than above.
Gram staining quickly identifies bacteria as Gram-negative or Gram-positive, which can be used to find a more efficient antibiotic and dosage.
Proportion of bacteria strains inhibited
Concentration (μg/ml)

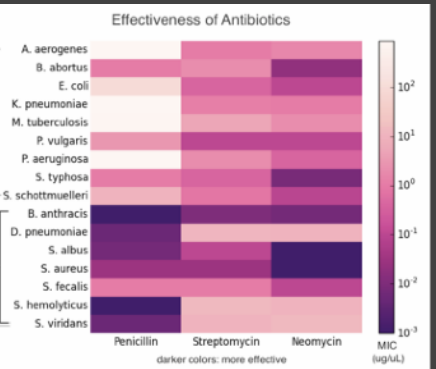

Gram-positive bacteria only
Penicillin is more efficient than either Streptomycin or Neomycin if the bacteria is known to be gram-positive.
Proportion of bacteria strains inhibited
Concentration (μg/ml)

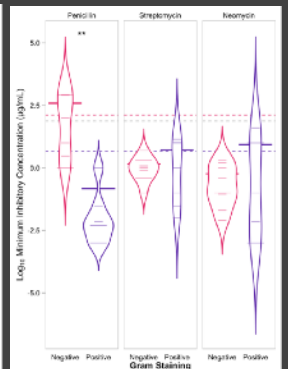

Penicillin
Streptomycin
Neomycin
Minimum Inhibitory Concentration (MIC)

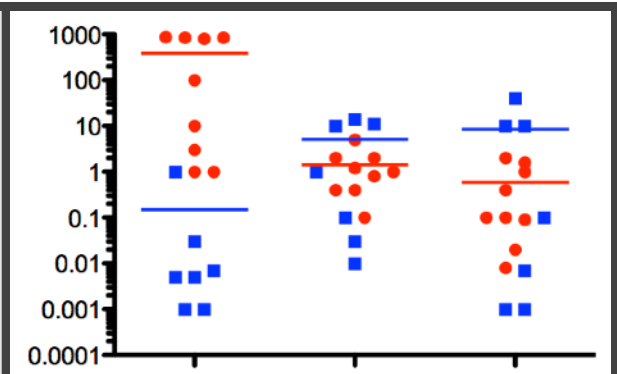

Gram-Positive
Diplococcus pneumoniae
Streptococcus viridans
Staphylococcus albus
Staphylococcus hemolyticus
Staphylococcus aureus
Streptococcus fecalis

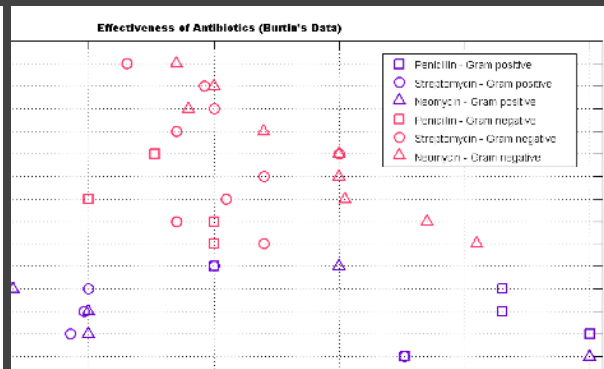


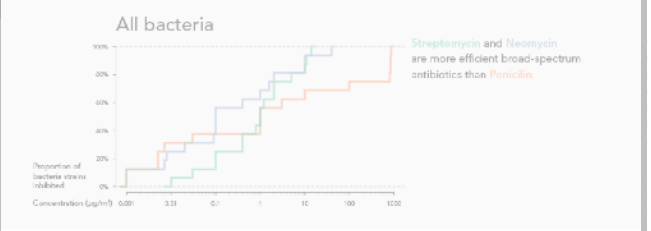


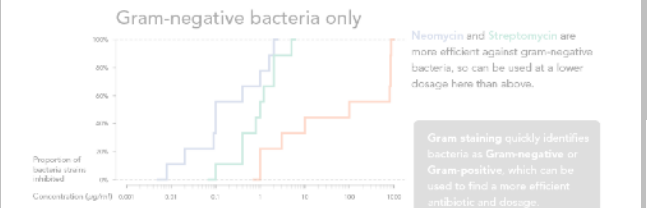

Effectiveness of Antibiotics
A. aerogenes
B. abortus
E. coli
K. pneumoniae
M. tuberculosis
P. vulgaris
P. aeruginosa
S. typhosa
S. schottmuelleri
B. anthracis
D. pneumoniae
S. albus
S. aureus
S. fecalis
S. hemolyticus
S. viridans
Penicillin
Streptomycin
Neomycin
MIC (ug/uL)
darker colors: more effective

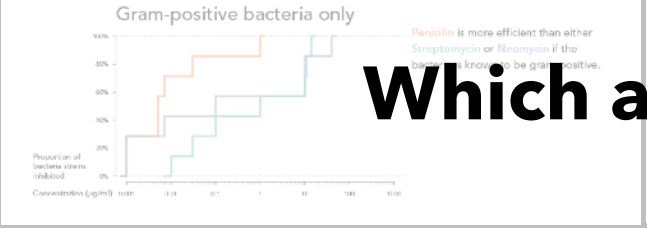

Penicillin
Streptomycin
Neomycin
Log10 Minimum Inhibitory Concentration (ug/mL)
Negative
Positive
Gram Staining

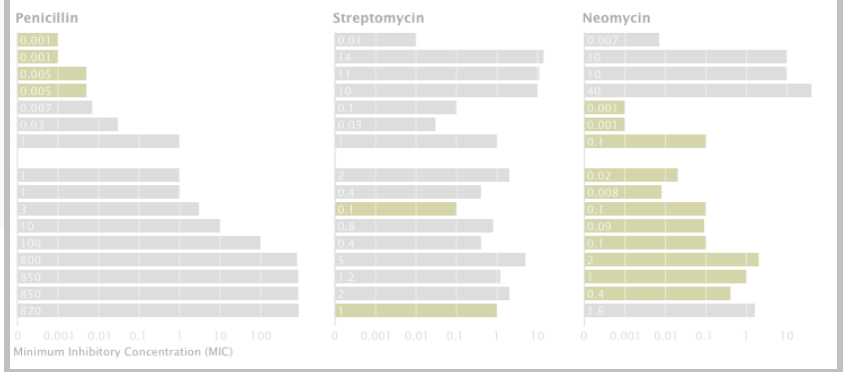

1000
100
10
1
0.1
0.01
0.001
0.0001

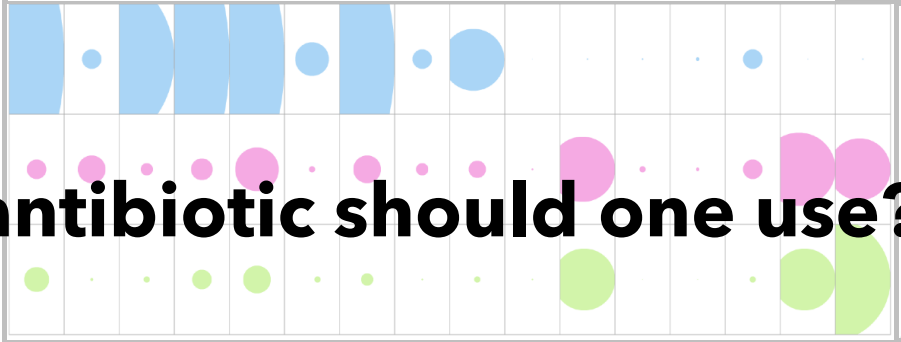

Effectiveness of Antibiotics (Burtis's Data)
Penicillin - Gram positive
Streptomycin - Gram positive
Neomycin - Gram positive
Penicillin - Gram negative
Streptomycin - Gram negative
Neomycin - Gram negative

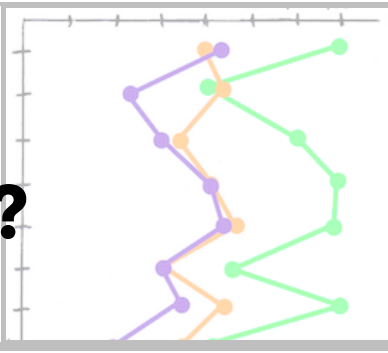




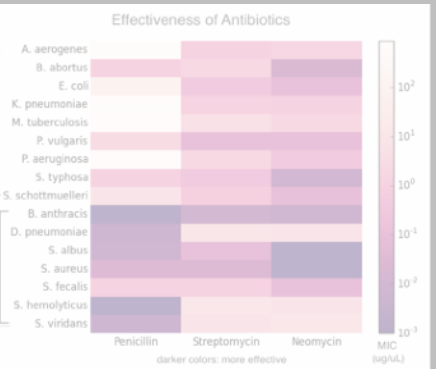


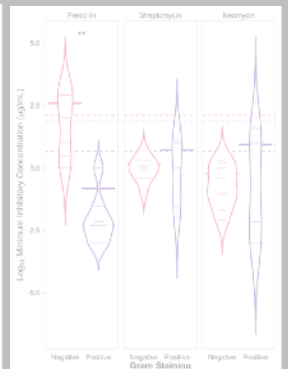


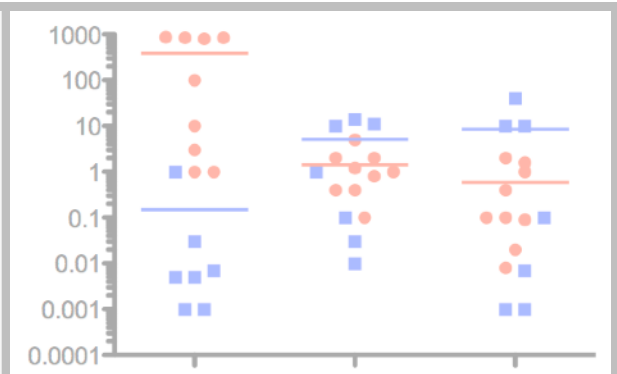


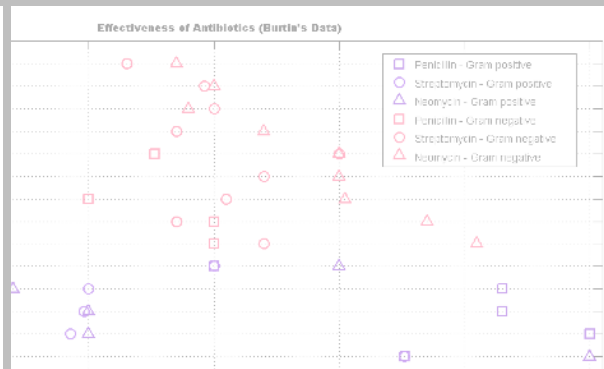


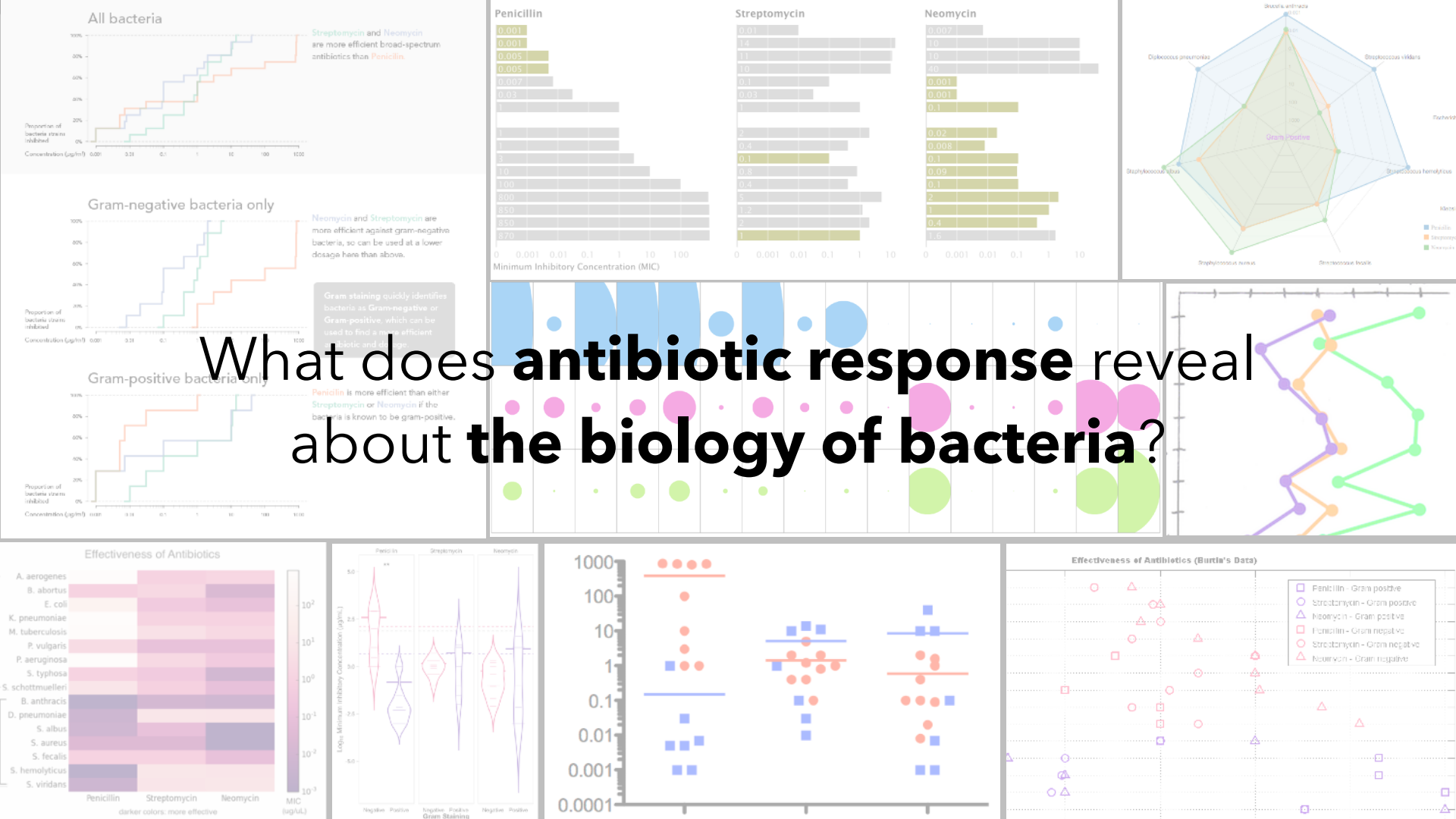
Which antibiotic should one use?





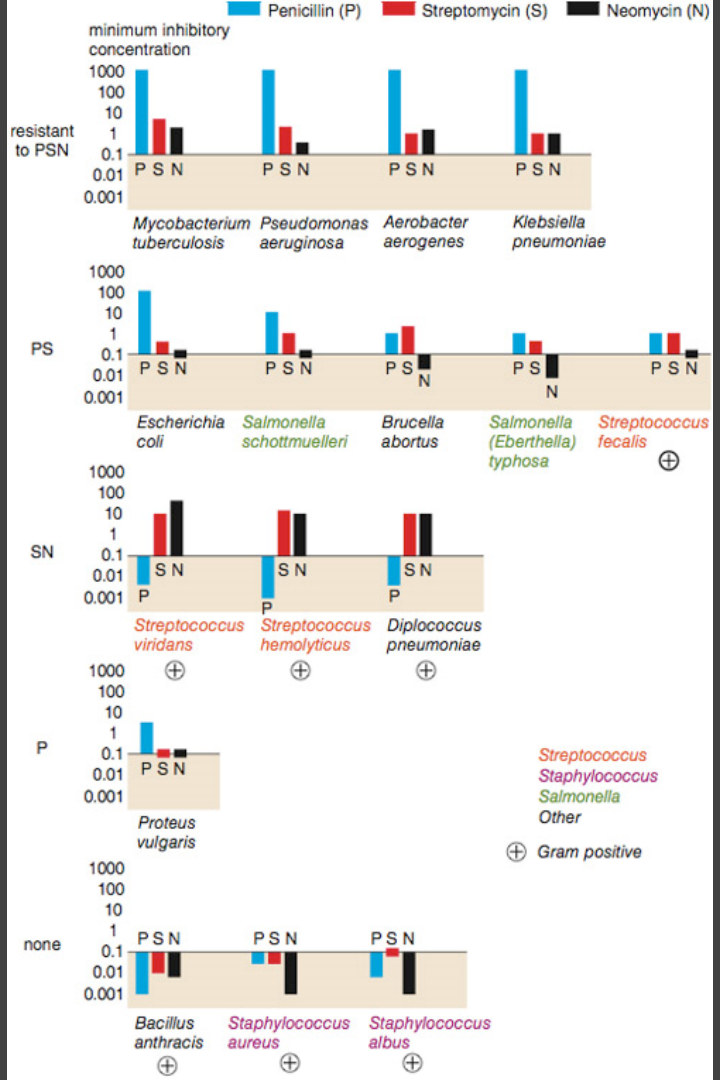






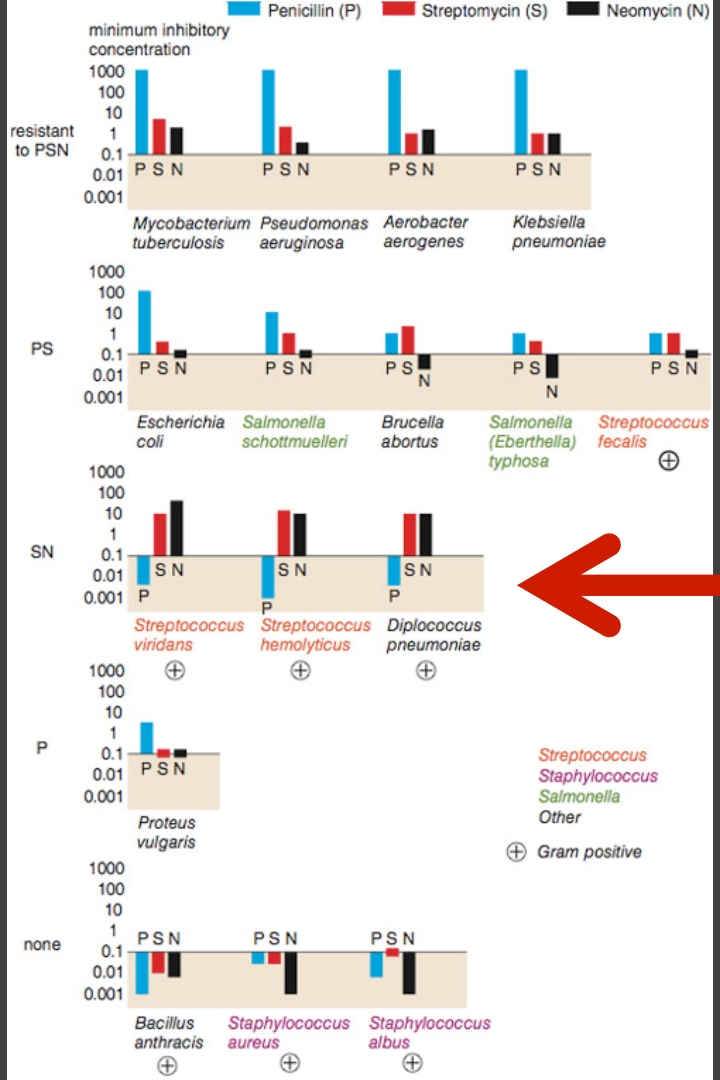
What does **antibiotic response** reveal about **the biology of bacteria**?

Do the bacteria group by antibiotic resistance?



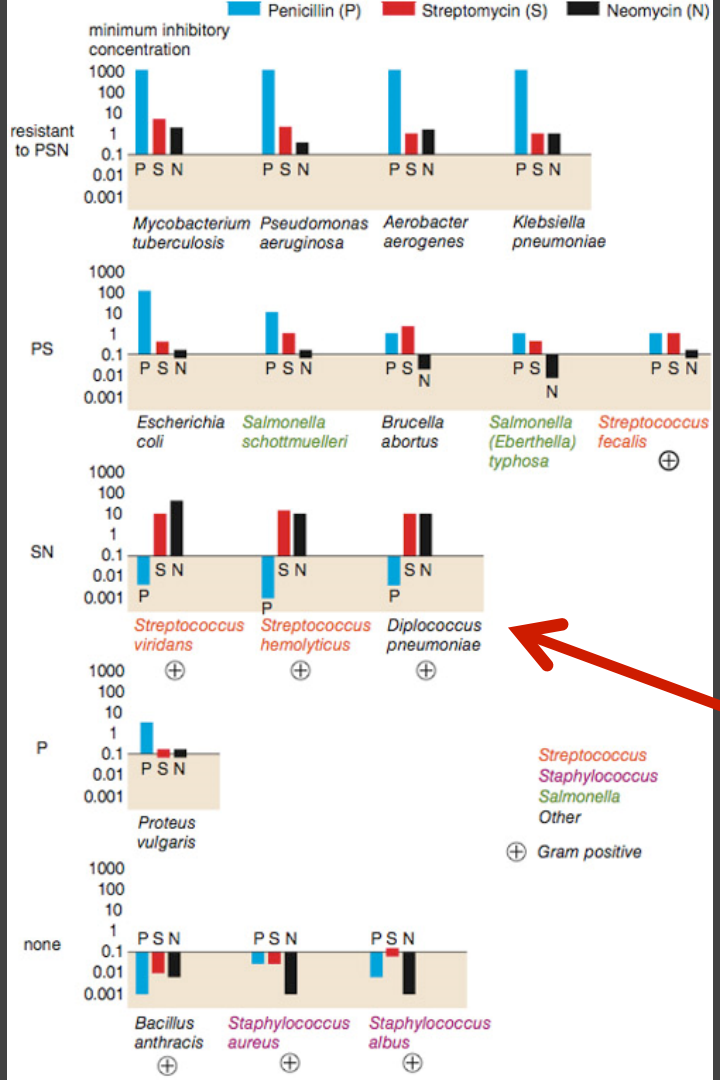
Do the bacteria group by antibiotic resistance?

Wainer & Lysen
American Scientist, 2009



Do the bacteria group by antibiotic resistance?

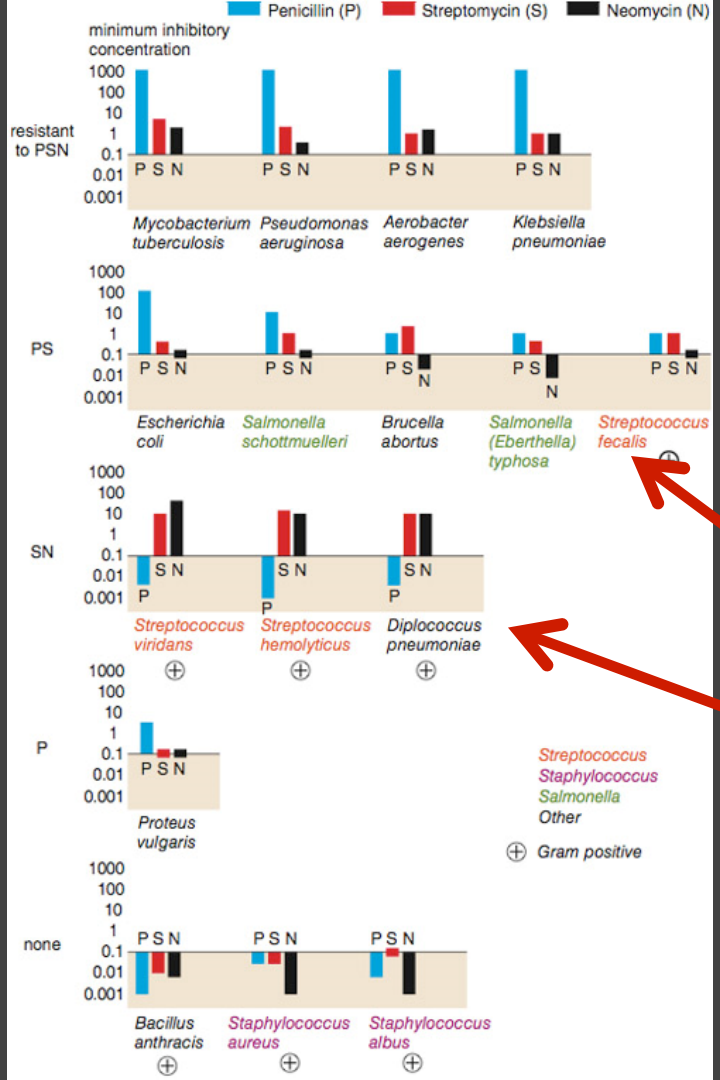
Wainer & Lysen
American Scientist, 2009



Do the bacteria group by antibiotic resistance?

Really a streptococcus!
(realized ~20 yrs later)

Wainer & Lysen
American Scientist, 2009

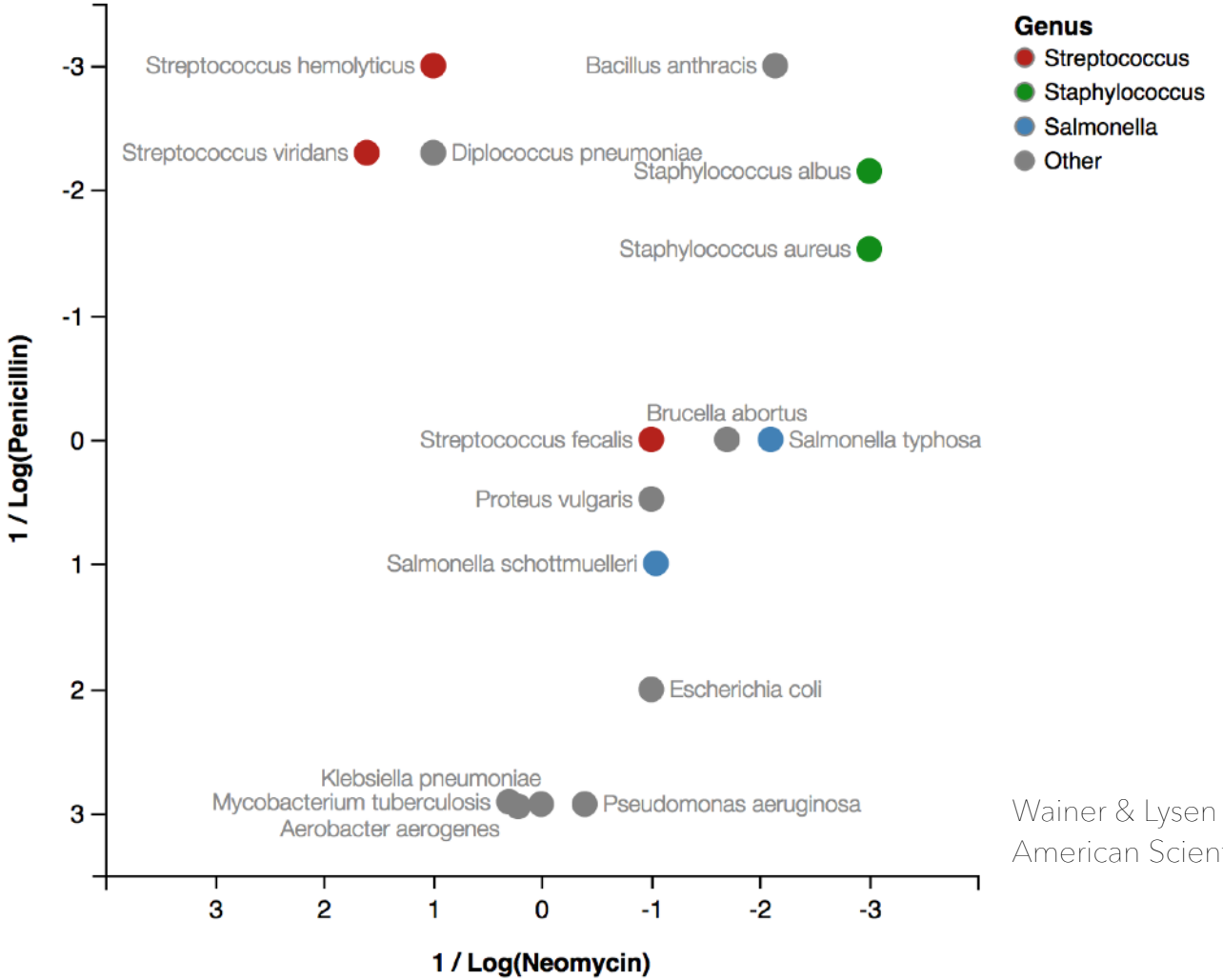


Do the bacteria group by antibiotic resistance?

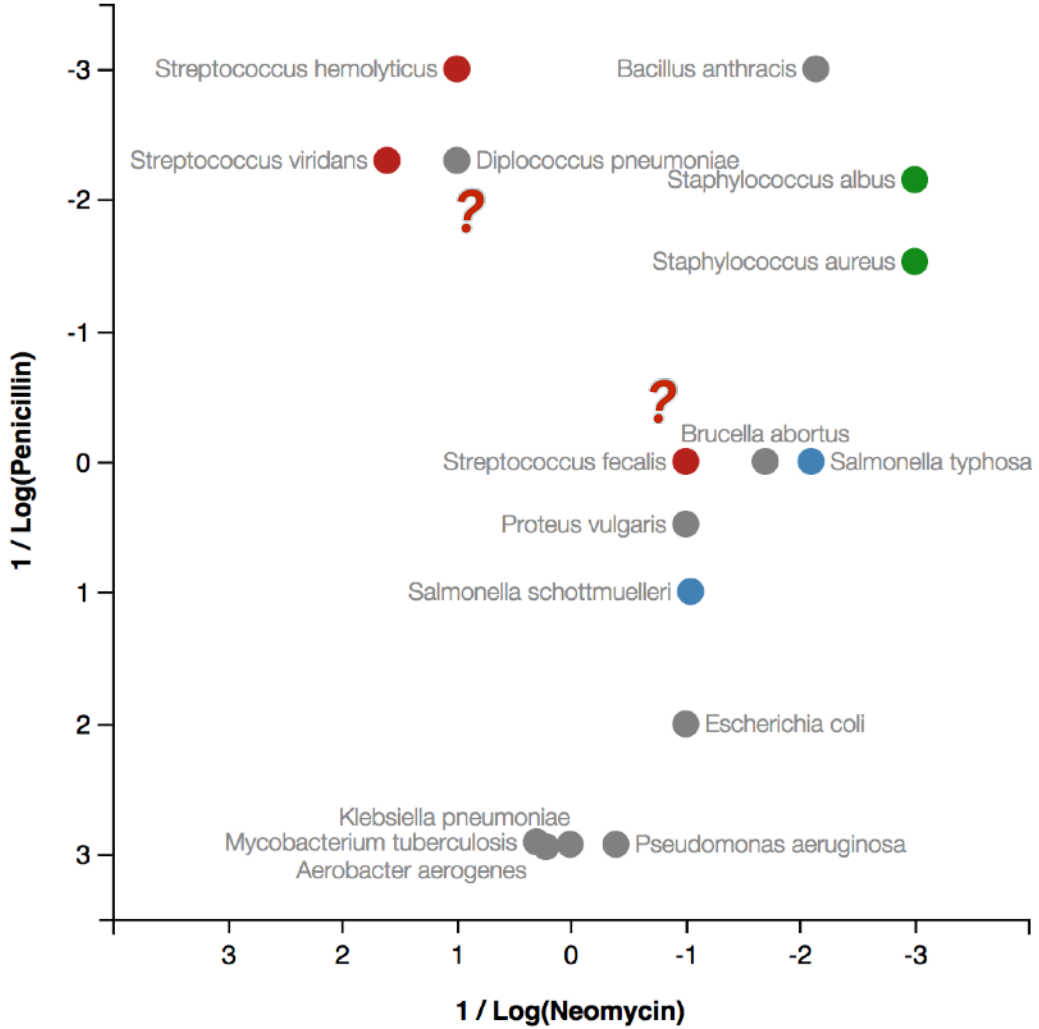
Not a streptococcus!
(realized ~30 yrs later)

Really a streptococcus!
(realized ~20 yrs later)

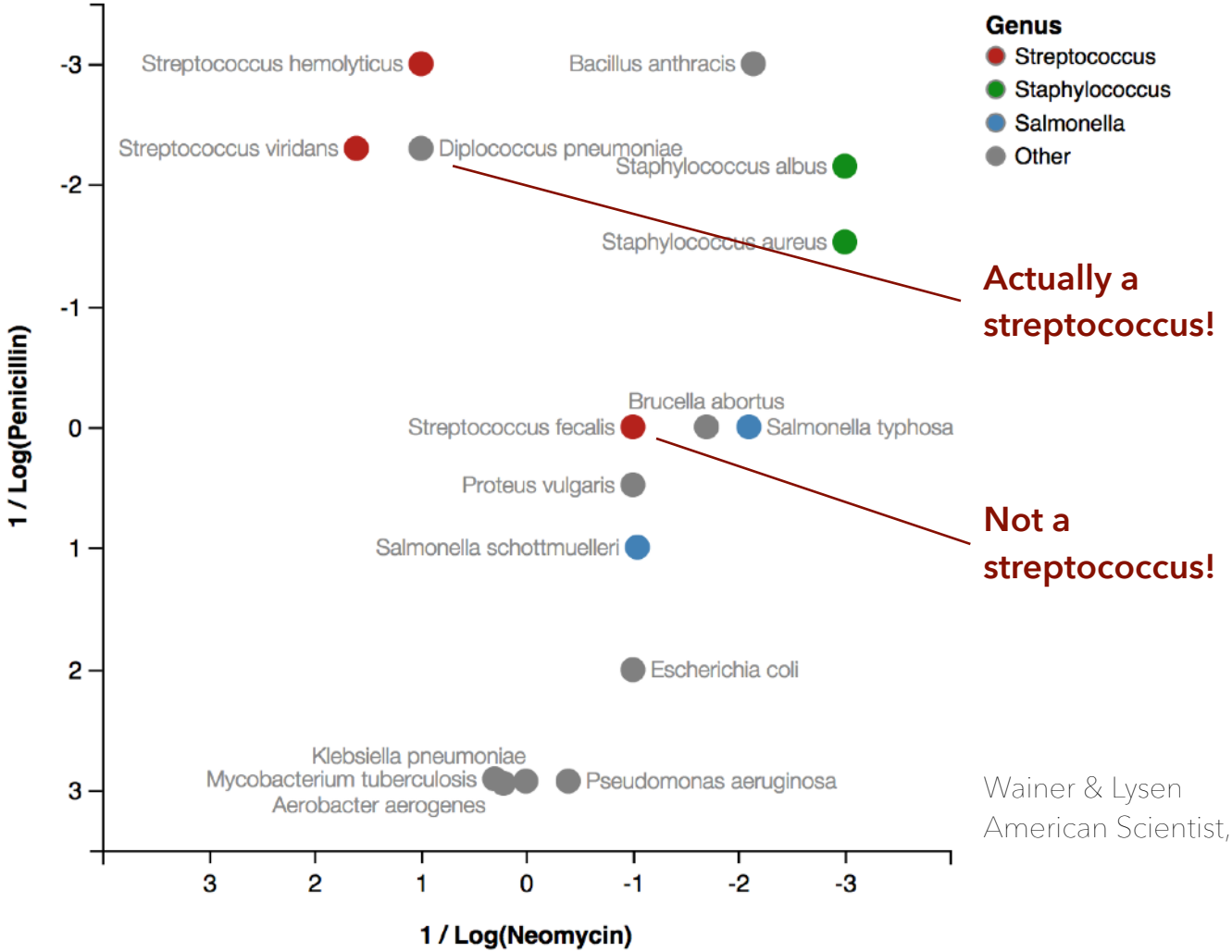
Wainer & Lysen
American Scientist, 2009



Wainer & Lysen
American Scientist, 2009



Wainer & Lysen
American Scientist, 2009



Wainer & Lysen
American Scientist, 2009

Lesson: Iterative Exploration

Exploratory Process

1 Construct graphics to address questions
2 Inspect “answer” and assess new questions
3 Repeat…

Transform data appropriately (e.g., invert, log)

Formulate clear analysis questions & goals

Don’t trust your data!

What is Visualization?

"Transformation of the symbolic into the geometric"
[McCormick et al. 1987]

"... finding the artificial memory that best supports our natural means of perception." [Bertin 1967]

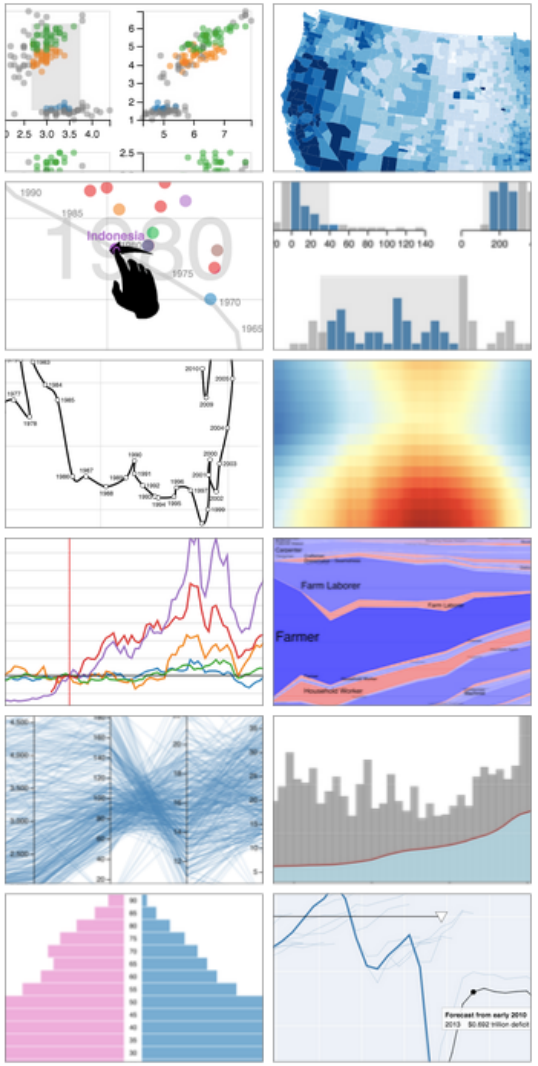
"The use of computer-generated, interactive, visual representations of data to amplify cognition."
[Card, Mackinlay, & Shneiderman 1999]

Exploration Tasks

Data Exploration Tasks

Profile: learn the shape and structure of the data, assess data quality, check modeling assumptions

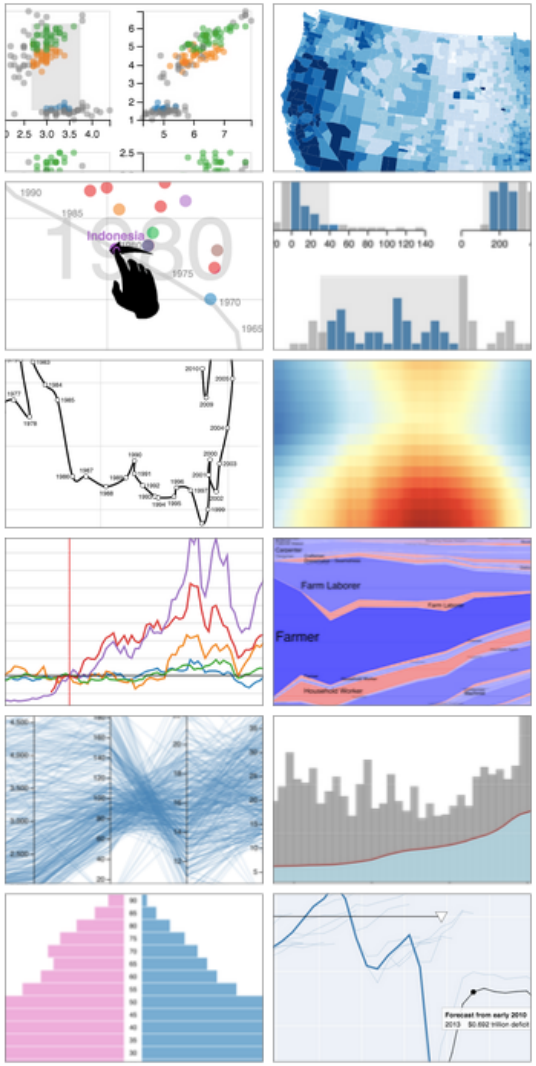
GOAL: Is the data actionable? What can we ask?



Data Exploration Tasks

Profile: learn the shape and structure of the data, assess data quality, check modeling assumptions
GOAL: Is the data actionable? What can we ask?

Search: identify specific data points or relations of interest to form an evidentiary chain
GOAL: Fact-finding, isolate important points/connections

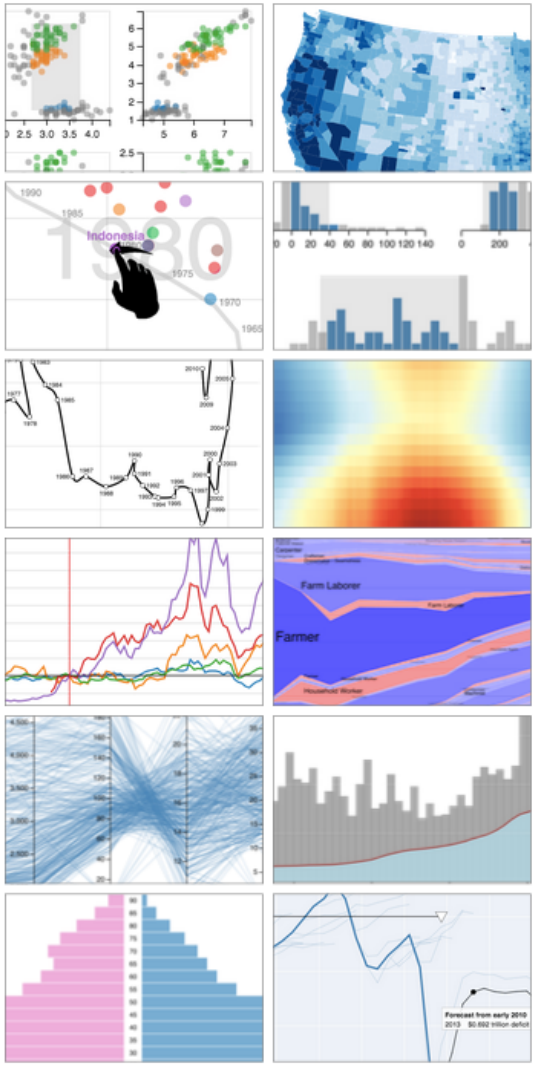


Data Exploration Tasks

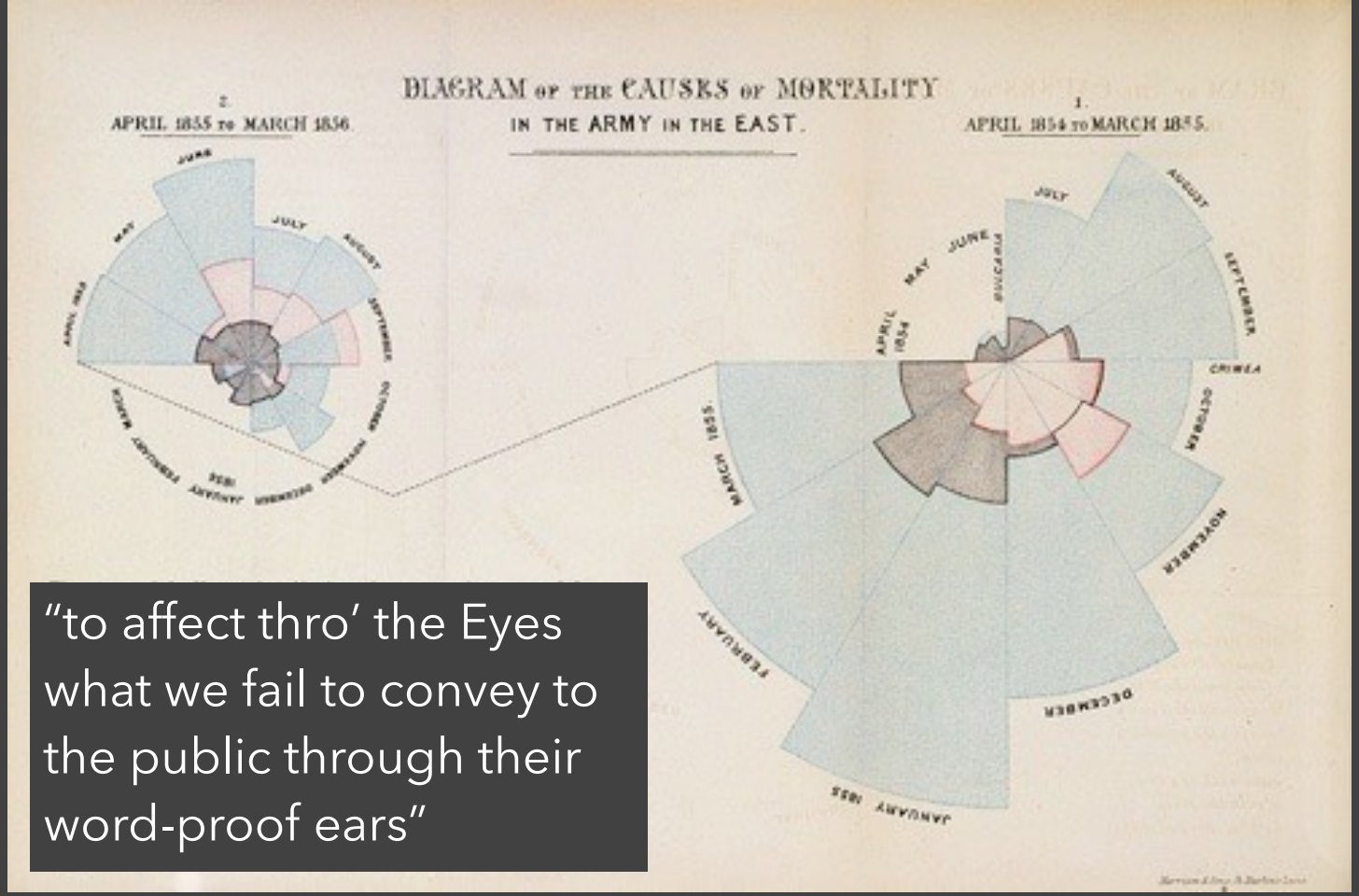
Profile: learn the shape and structure of the data, assess data quality, check modeling assumptions
GOAL: Is the data actionable? What can we ask?

Search: identify specific data points or relations of interest to form an evidentiary chain
GOAL: Fact-finding, isolate important points/connections

Infer: generalize from observed patterns, ascribe observations to specific factors or causes
GOAL: Inform modeling and decision making



Communication Tasks

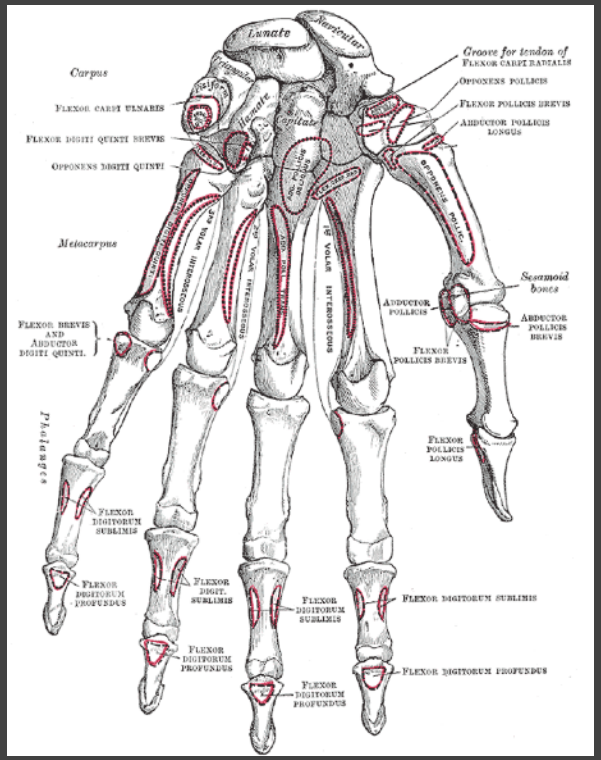


"to affect thro' the Eyes what we fail to convey to the public through their word-proof ears"

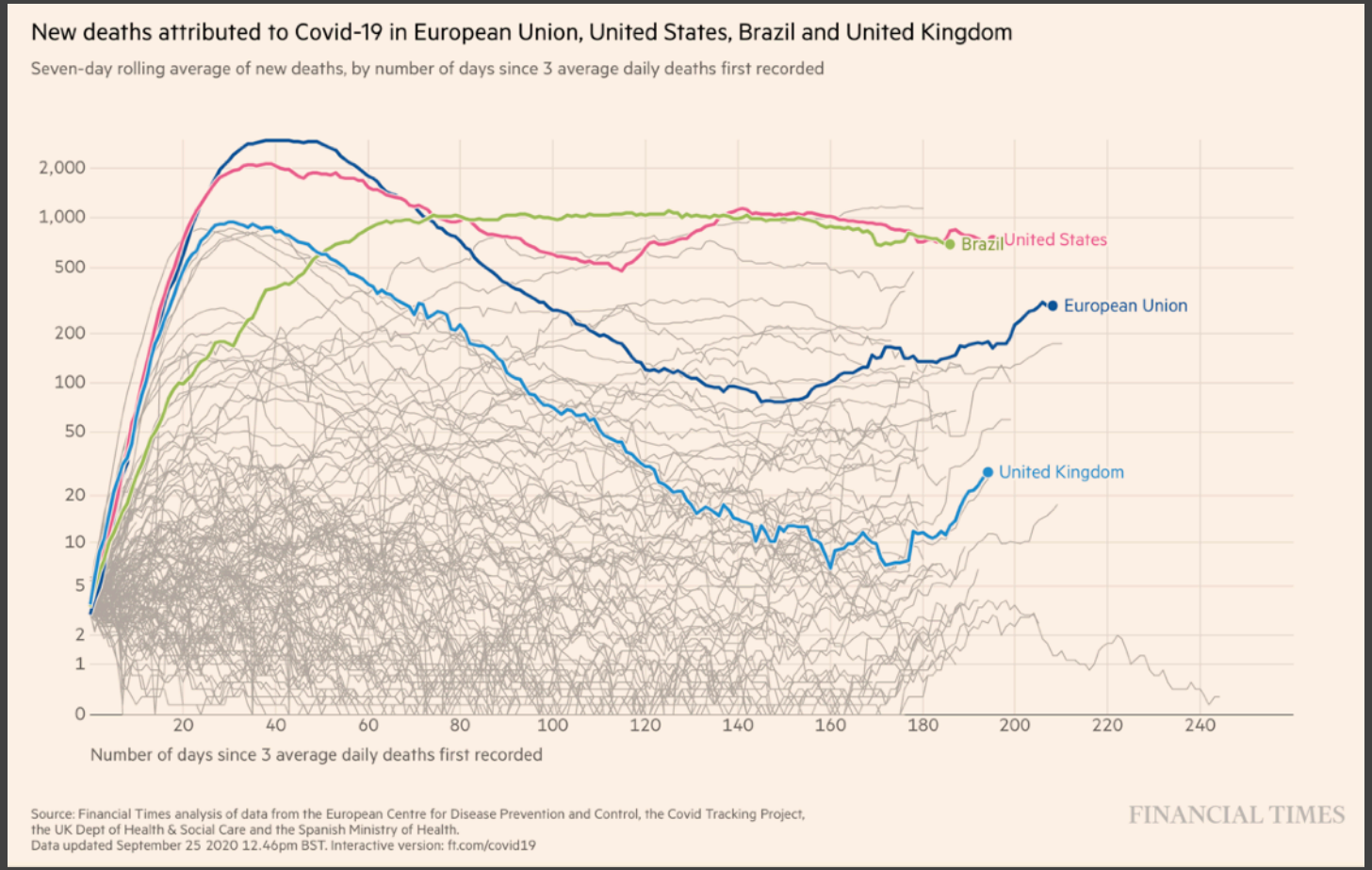
1856 "Coxcomb" of Crimean War Deaths, Florence Nightingale

Communicate, Inform, Inspire

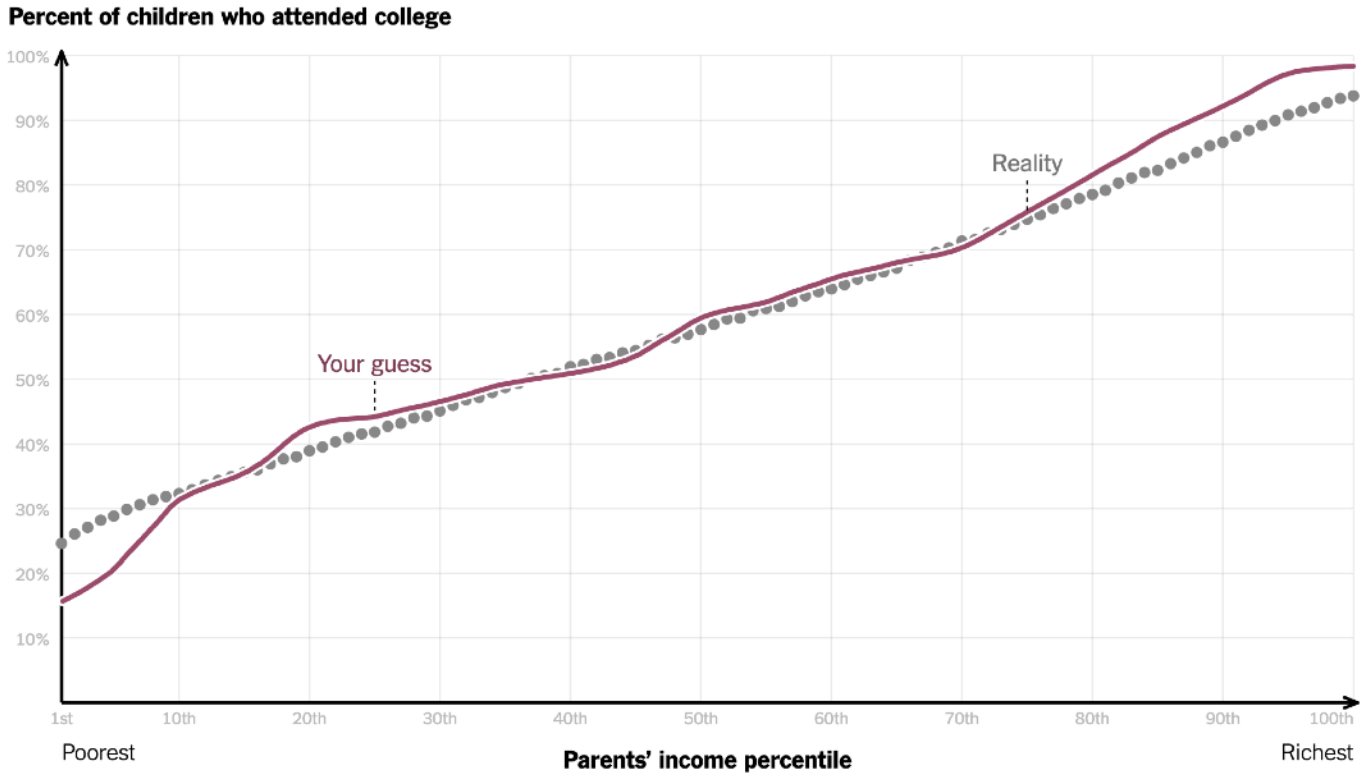
Visualizing Black America, Du Bois et al. 1900



Bones in hand, Gray's Anatomy 1918 ed.

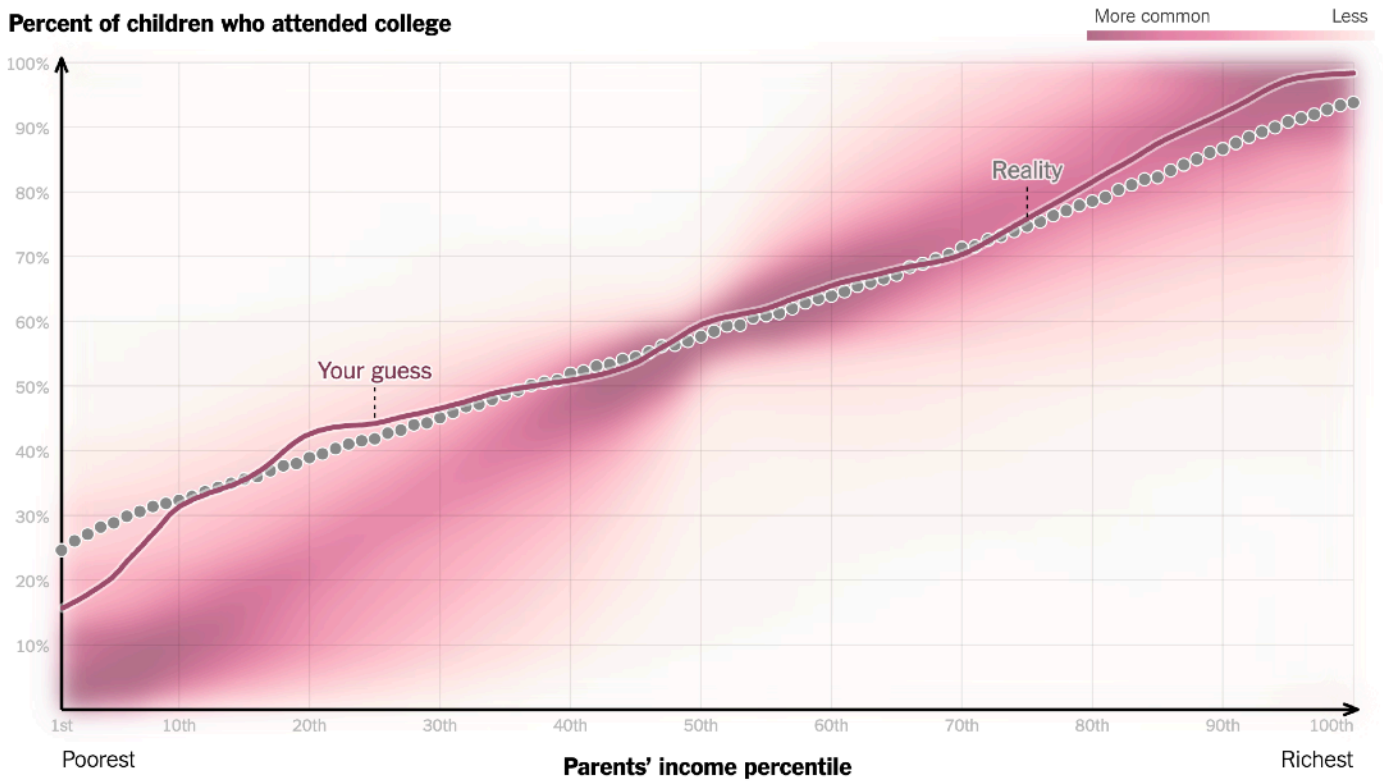


Coronavirus Tracked John Burn-Murdoch & Financial Times



You Draw It: How Family Income Predicts Children's College Chances

[New York Times, May 28, 2015]



You Draw It: How Family Income Predicts Children's College Chances

[New York Times, May 28, 2015]

Course Overview

You should expect to:

1. *Evaluate and critique* visualization designs
2. *Learn* visualization techniques & theory
3. *Implement* interactive data visualizations
4. *Develop* a substantial visualization project

Goals of Visualization Research

1 **Understand** how visualizations convey information

What do people perceive / comprehend?
How do visualizations inform mental models?

2 **Develop principles and techniques** for creating effective visualizations and supporting analysis

Leverage perception & augment cognition
Improve ties between visualization & mental model

Course Overview

W1: Introduction to Visualization (Vega-Lite)

W2: Visual Encoding

W3: Data Transformation

W4: Mapping and Cartography

W5: Interaction Techniques

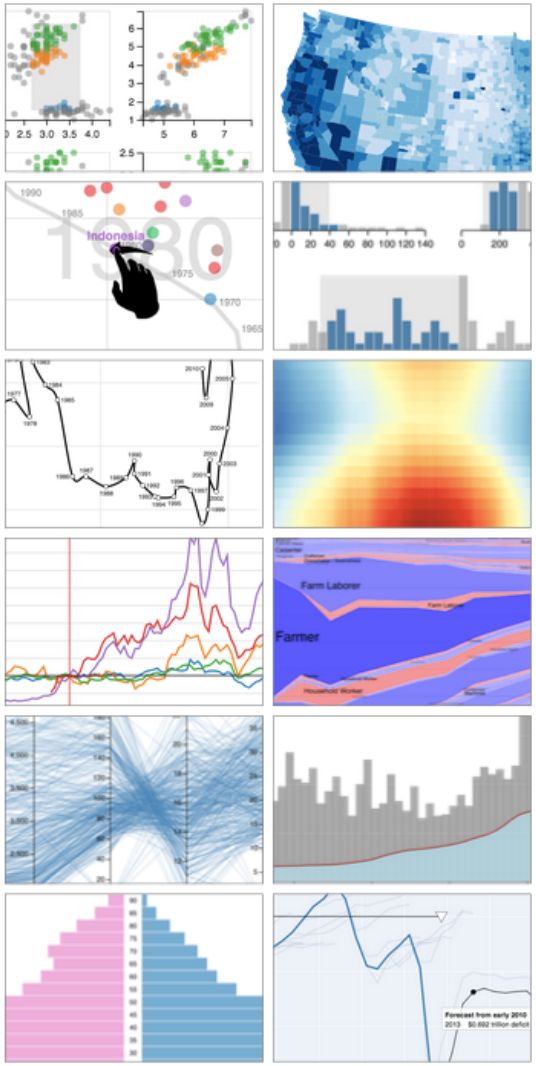
W6: Visualization on the Web (D3.js)

W7: Perception & Color

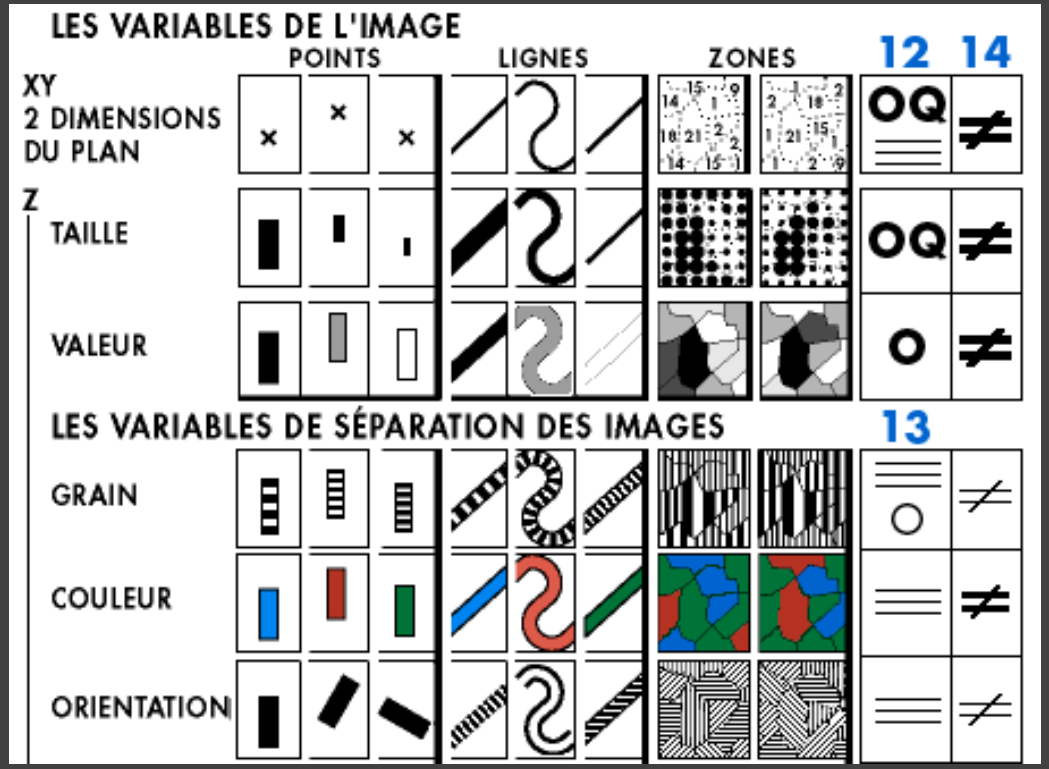
W8: Networks

W9: Uncertainty

W10: Conclusion and Final Project Showcase

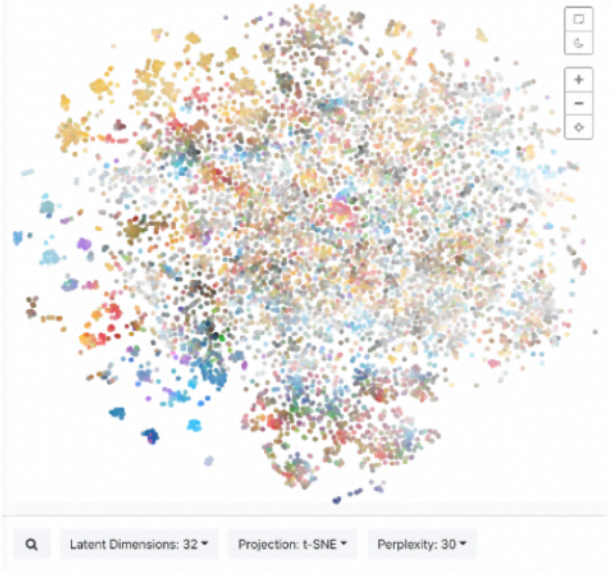


W2: Visual Encoding

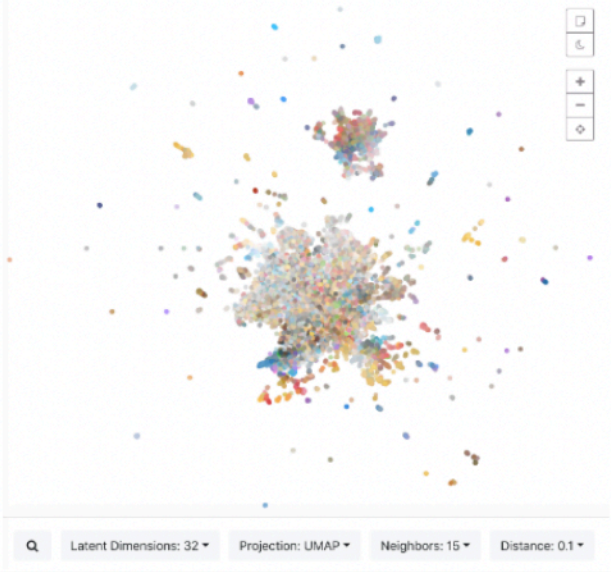


Sémiologie Graphique [Bertin 67]

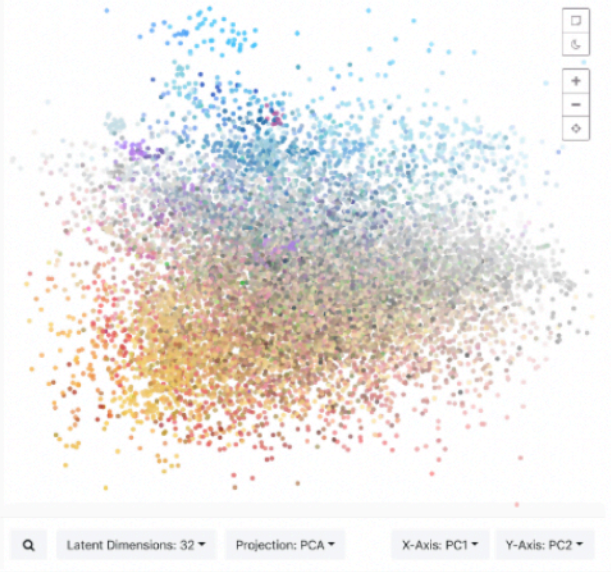
W3: Data Transformation



t-SNE

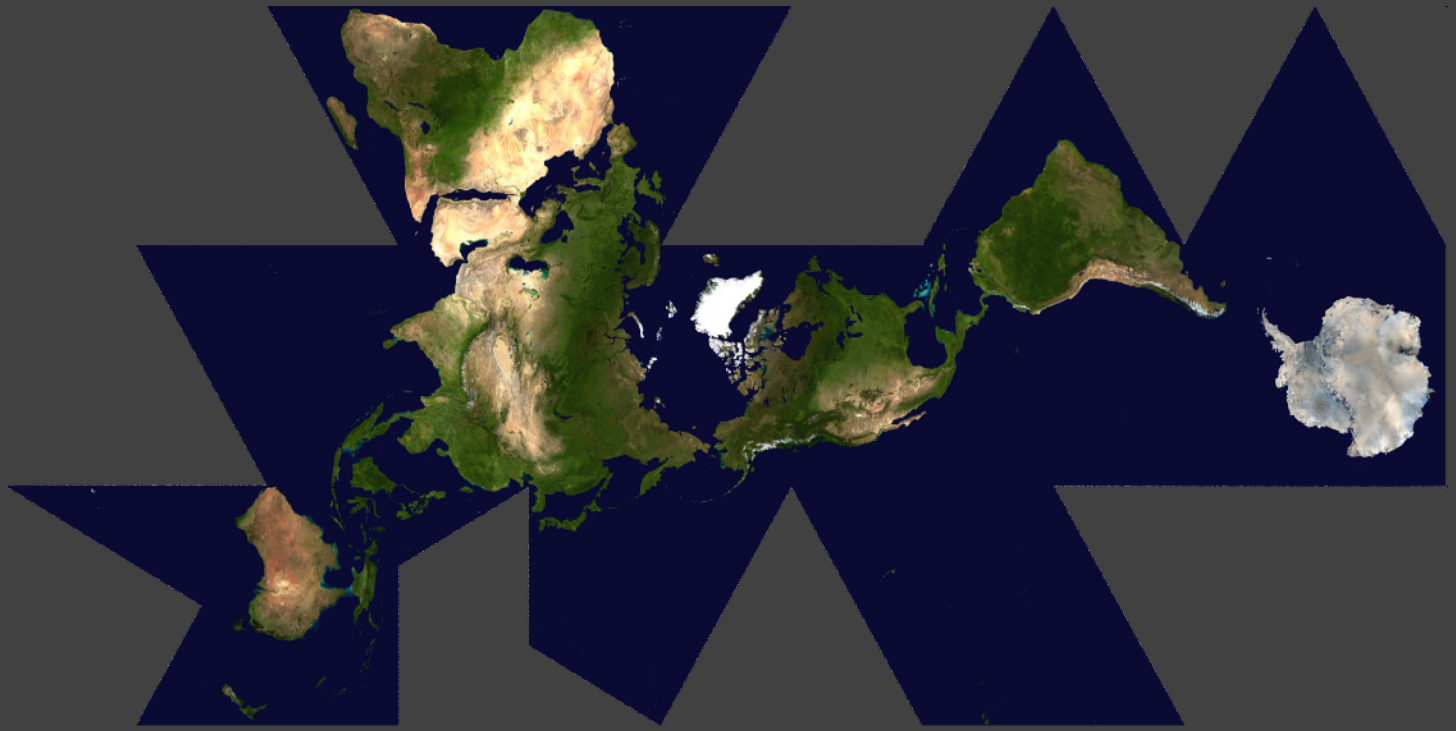


UMAP



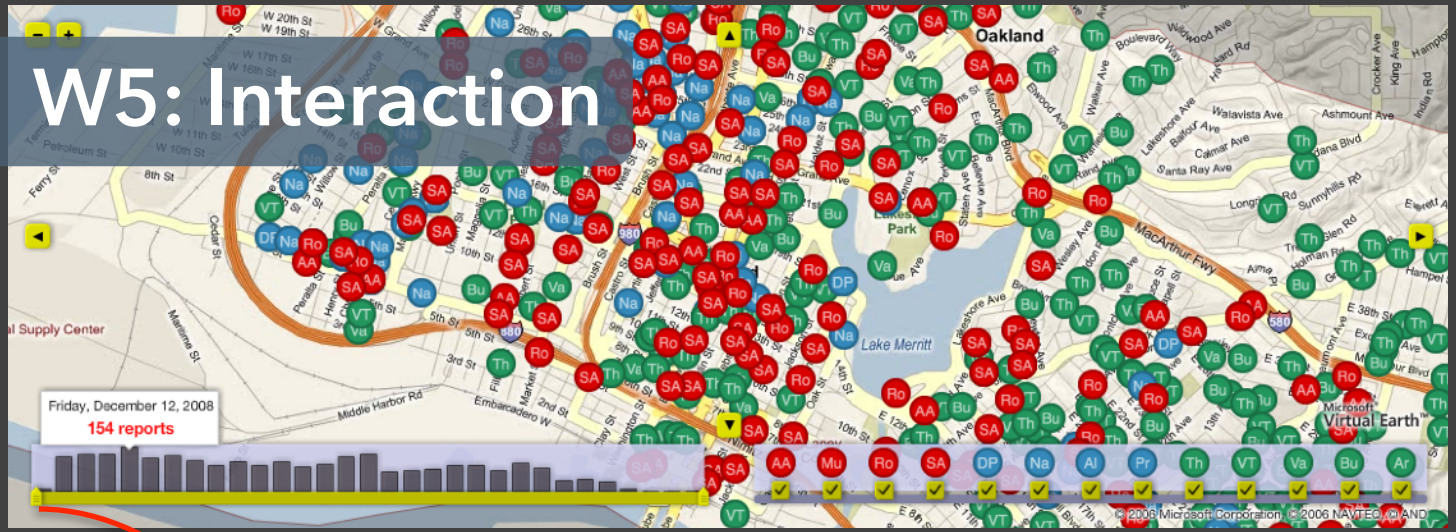
PCA

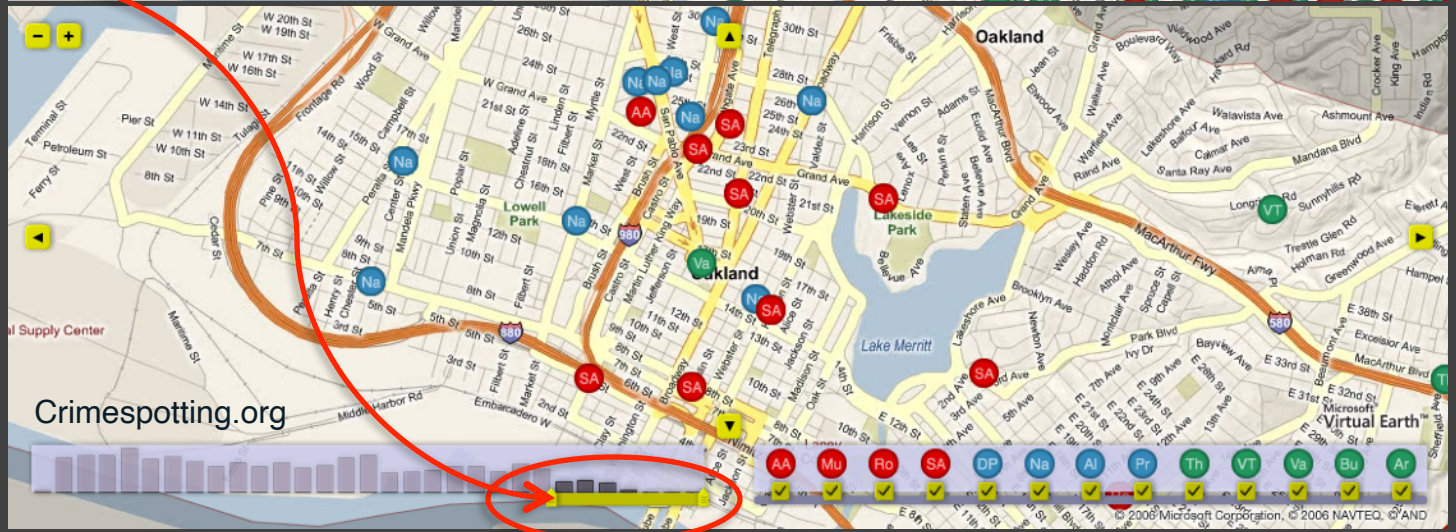
W4: Mapping & Cartography



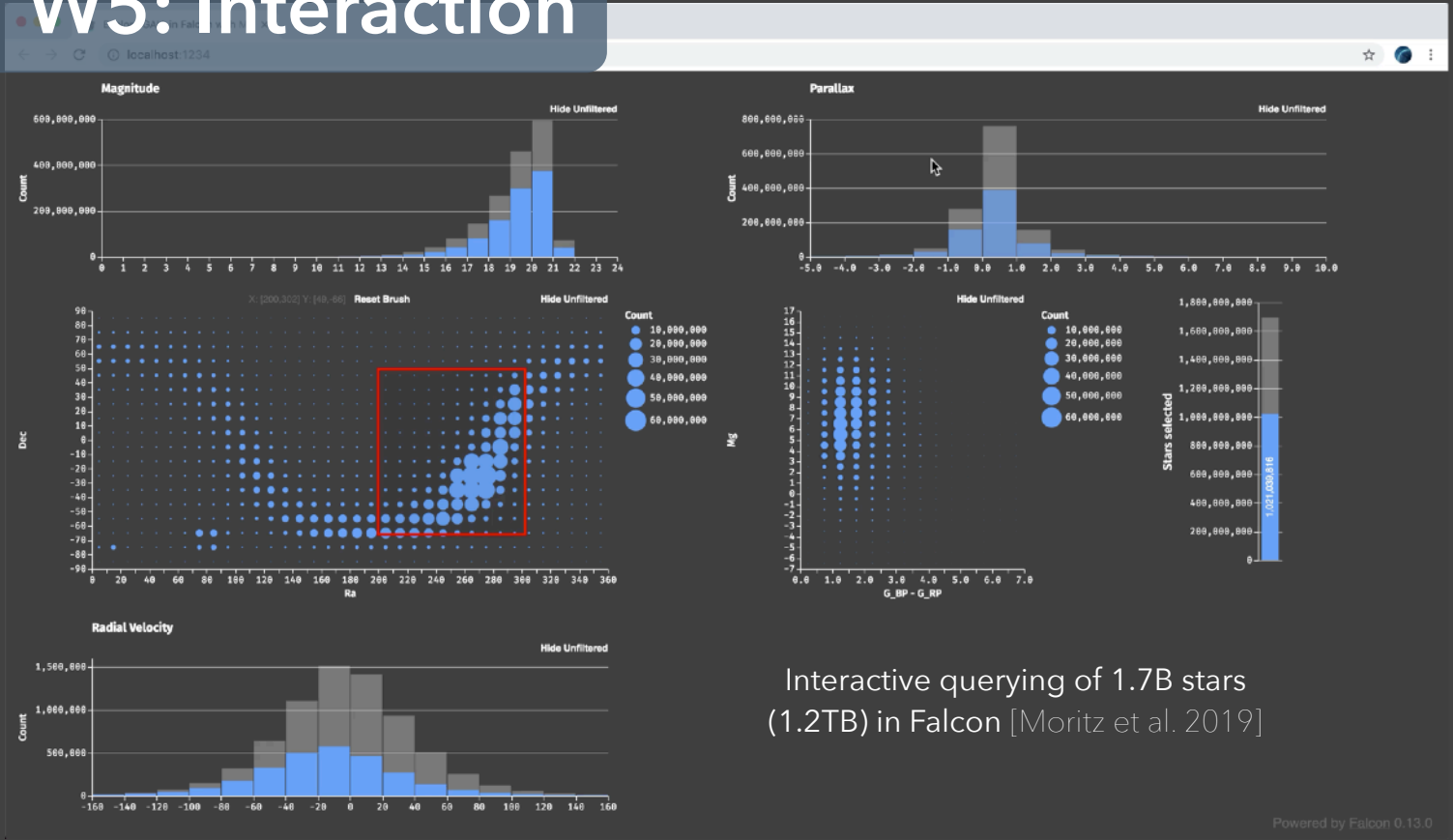
Dymaxion Maps [Fuller 46]

W5: Interaction



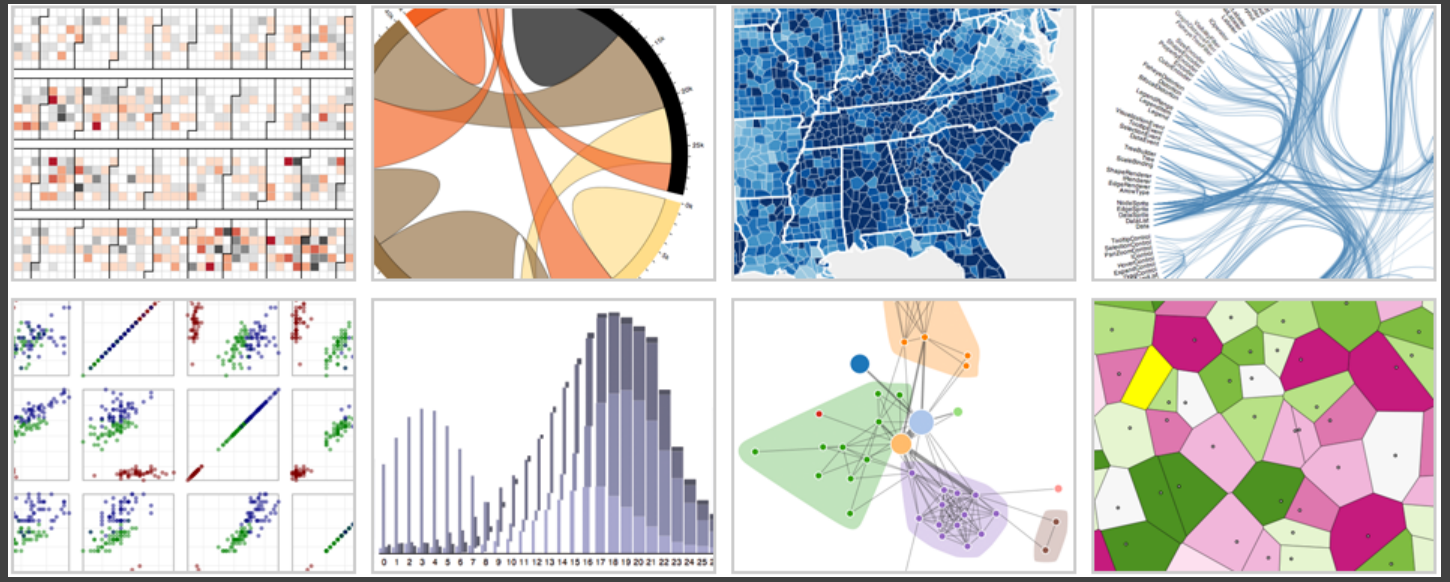


W5: Interaction



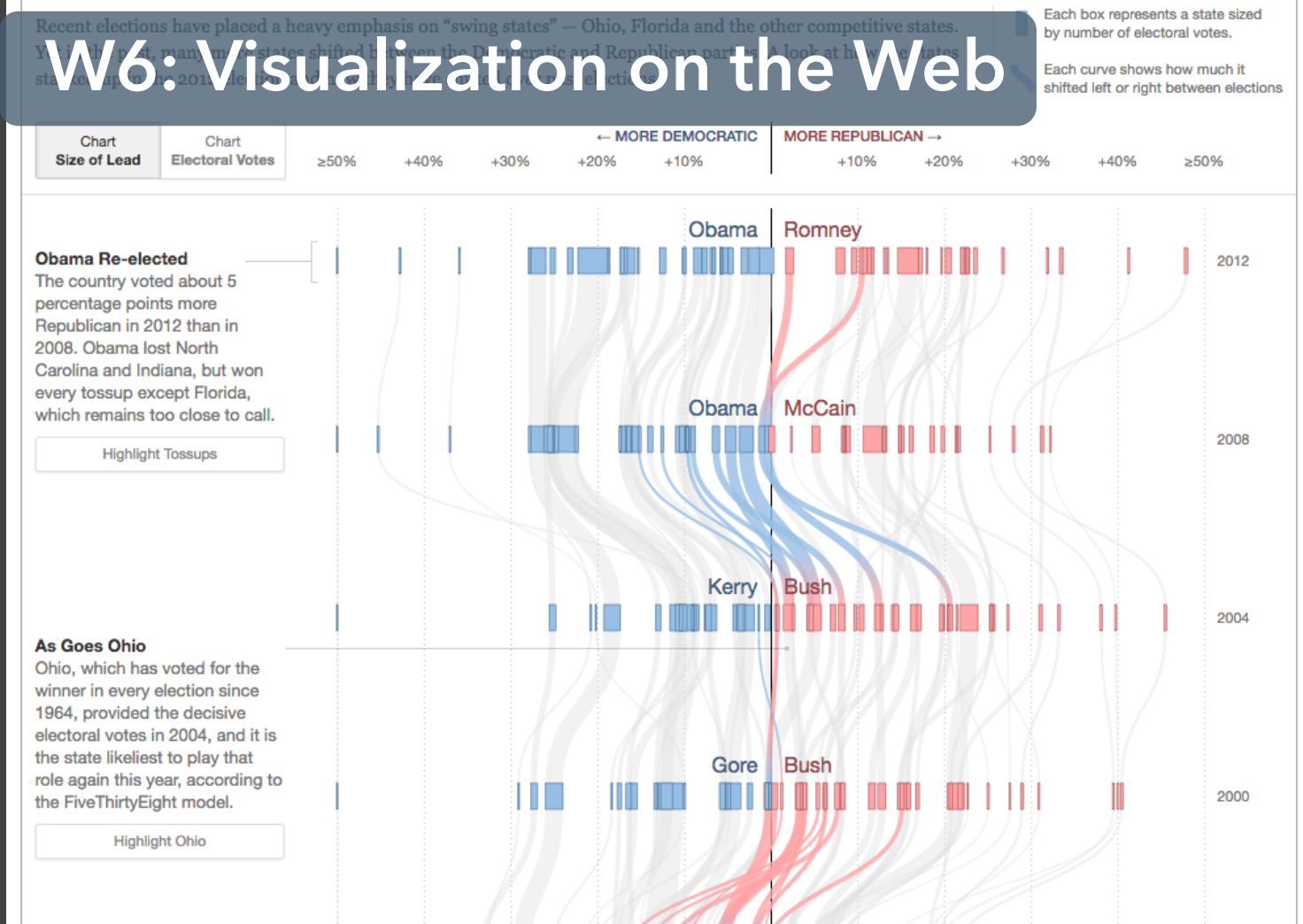
Interactive querying of 1.7B stars (1.2TB) in Falcon [Moritz et al. 2019]

W6: Visualization on the Web

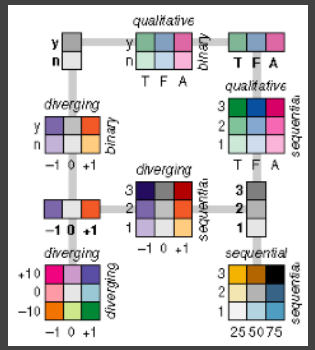


D3: Data-Driven Documents

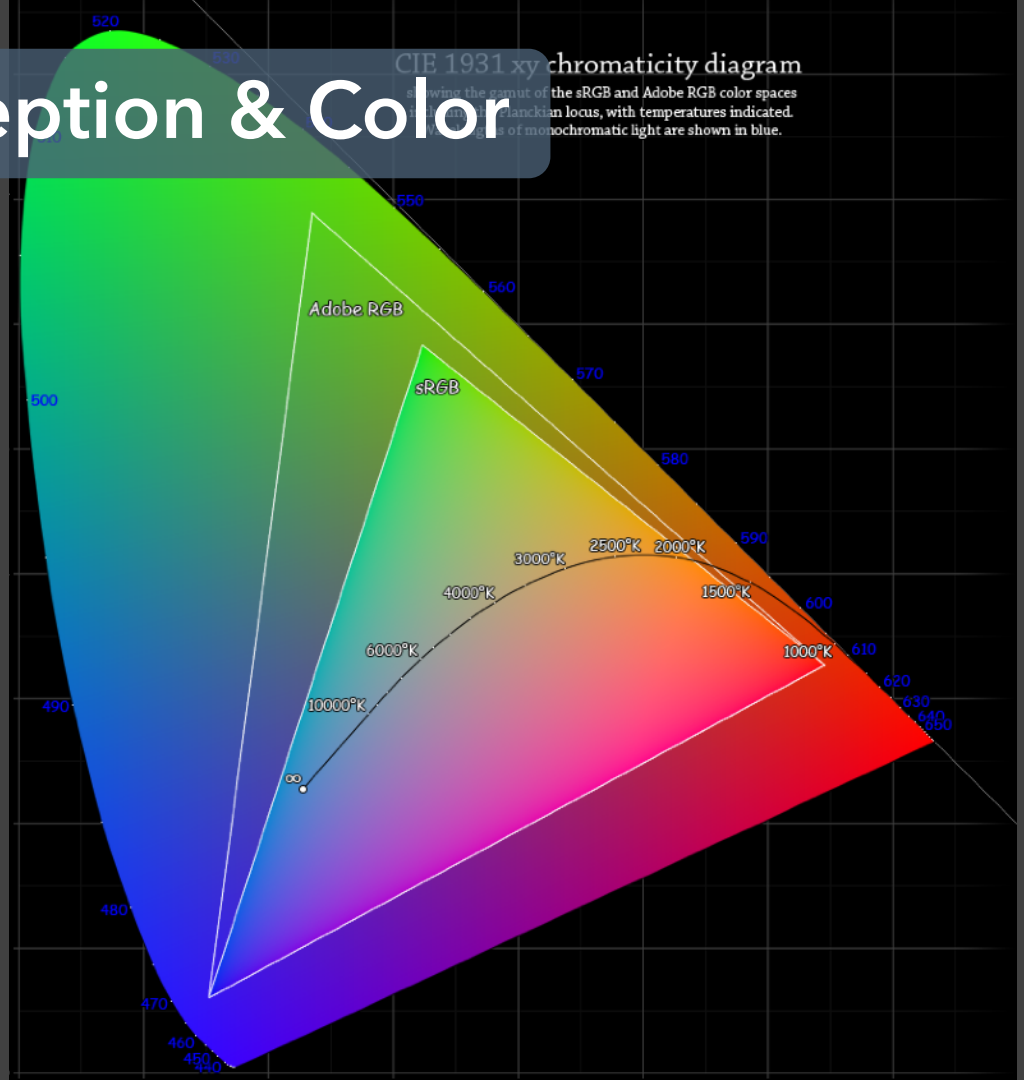
W6: Visualization on the Web



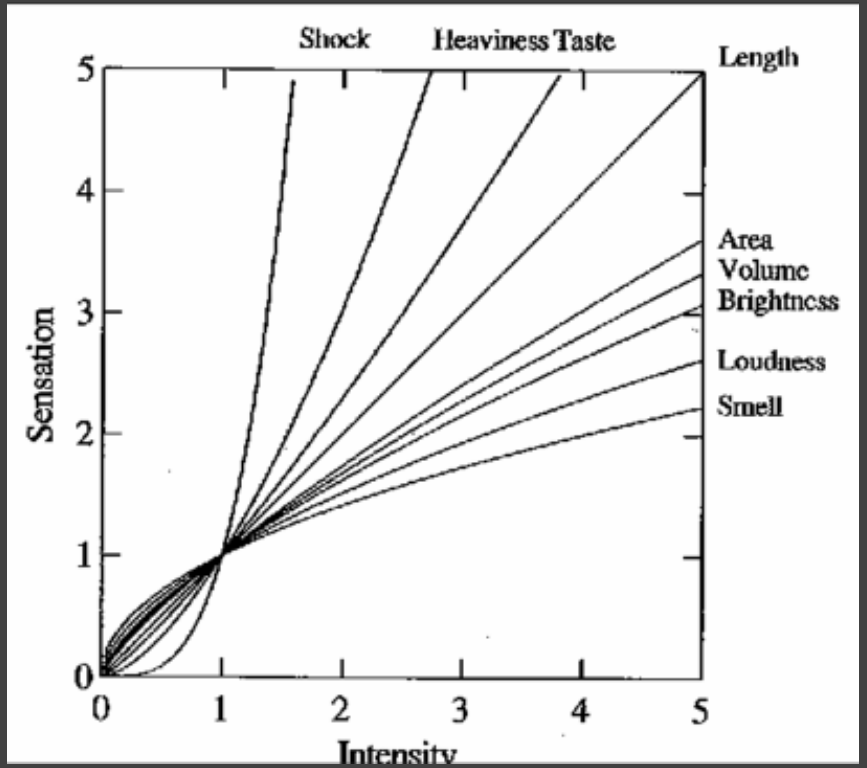
W7: Perception & Color



Color Brewer

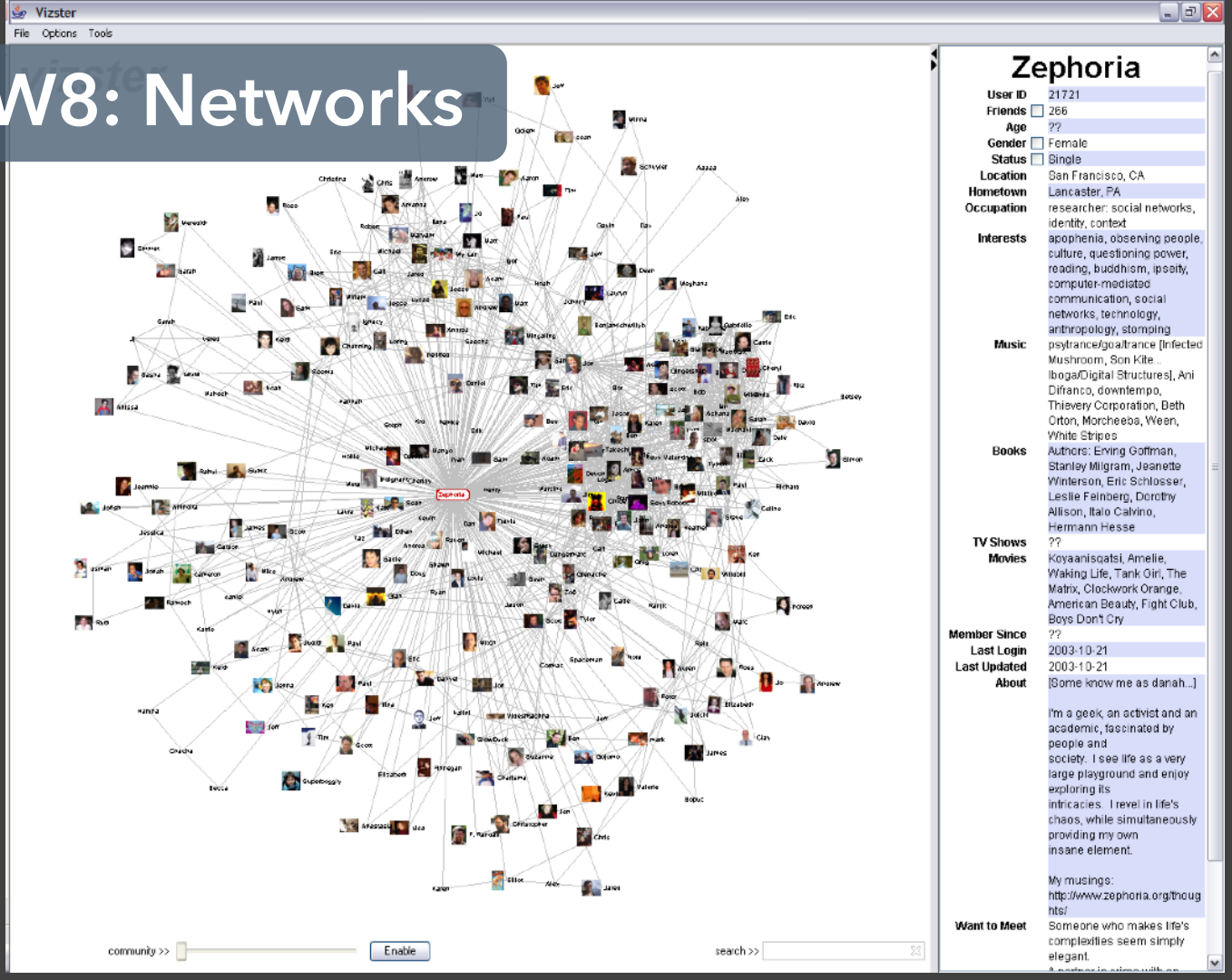


W7: Perception & Color

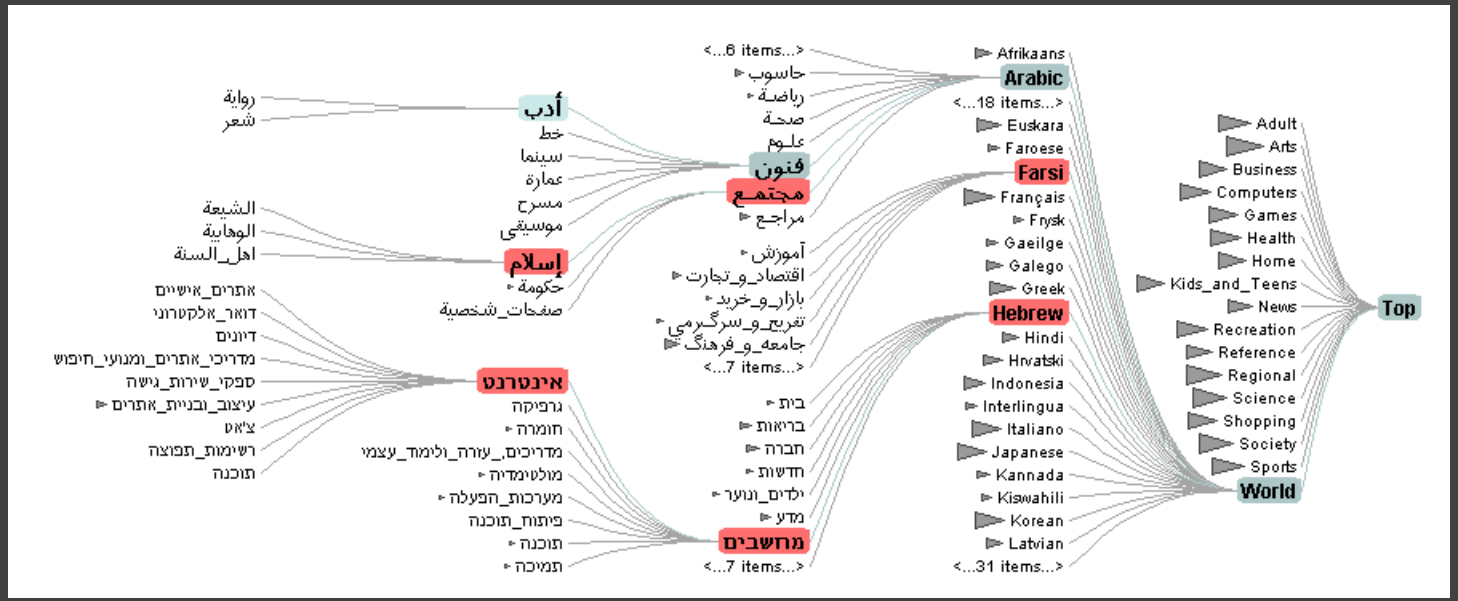


The psychophysics of sensory function [Stevens 61]

W8: Networks

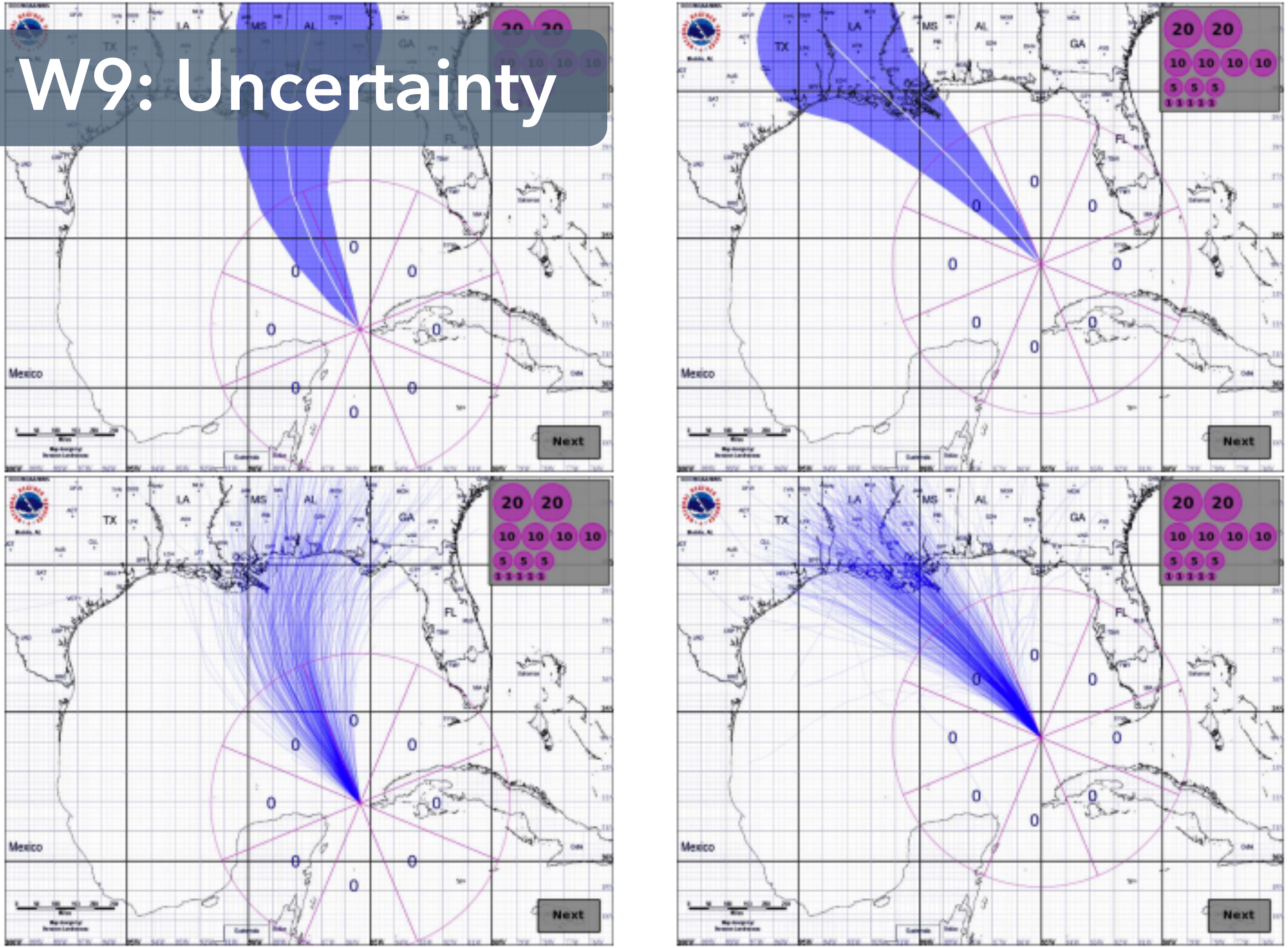

Vizster
File
Options
Tools
community >>
Enable
search >>
Zephoria
User ID 21721
Friends 266
Age ??
Gender Female
Status Single
Location San Francisco, CA
Hometown Lancaster, PA
Occupation researcher: social networks, identity, context
Interests apophenia, observing people, culture, questioning power, reading, buddhism, ipseity, computer-mediated communication, social networks, technology, anthropology, stomping
Music psytrance/goa/trance [Infected Mushroom, Son Kite... Iboga/Digital Structures], Ani Difranco, downtempo, Thievery Corporation, Beth Orton, Morcheeba, Ween, White Stripes
Books Authors: Erving Goffman, Stanley Milgram, Jeanette Winterson, Eric Schlosser, Leslie Feinberg, Dorothy Allison, Italo Calvino, Hermann Hesse
TV Shows ??
Movies Koyaanisqatsi, Amelie, Waking Life, Tank Girl, The Matrix, Clockwork Orange, American Beauty, Fight Club, Boys Don't Cry
Member Since ??
Last Login 2003-10-21
Last Updated 2003-10-21
About [Some know me as danah...]
I'm a geek, an activist and an academic, fascinated by people and society. I see life as a very large playground and enjoy exploring its intricacies. I revel in life's chaos, while simultaneously providing my own insane element.
My musings: http://www.zephoria.org/thoughts/
Want to Meet Someone who makes life's complexities seem simply elegant.

W8: Networks



Degree-Of-Interest Trees [Heer & Card 04]

W9: Uncertainty



Course Overview

W1: Introduction to Visualization (Vega-Lite)

W2: Visual Encoding

W3: Data Transformation

W4: Mapping and Cartography

W5: Interaction Techniques

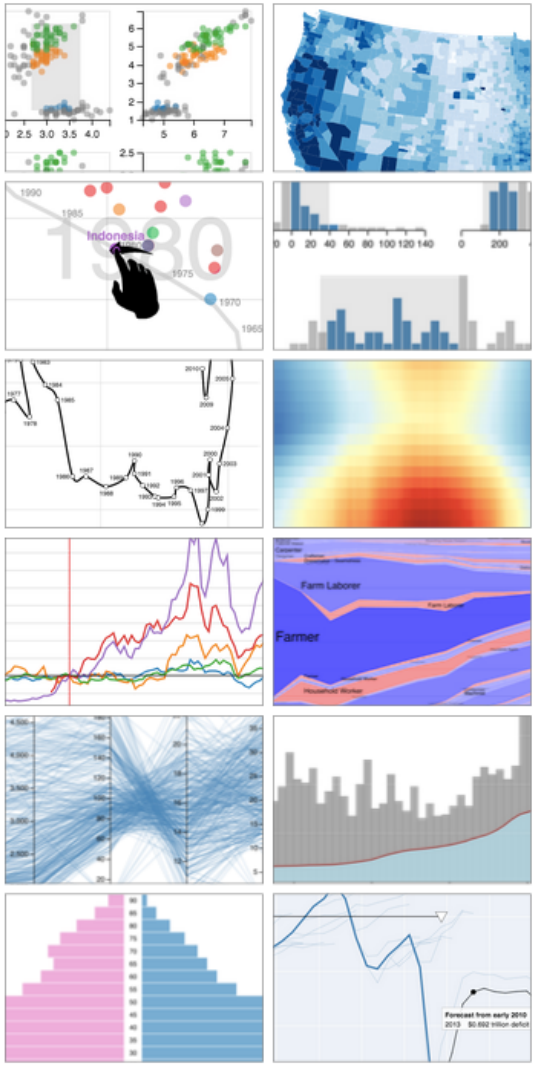
W6: Visualization on the Web (D3.js)

W7: Perception & Color

W8: Networks

W9: Uncertainty

W10: Conclusion and Final Project Showcase



Assignments and Scoring

Each assignment is due the following Tue by 9am.

W1: Expository Visualization (10%)

W2: Deceptive Visualization (10%)

W3: Peer Review (5%)

W4: Journey Map (10%)

W5: Interactive Visualization, Part 1

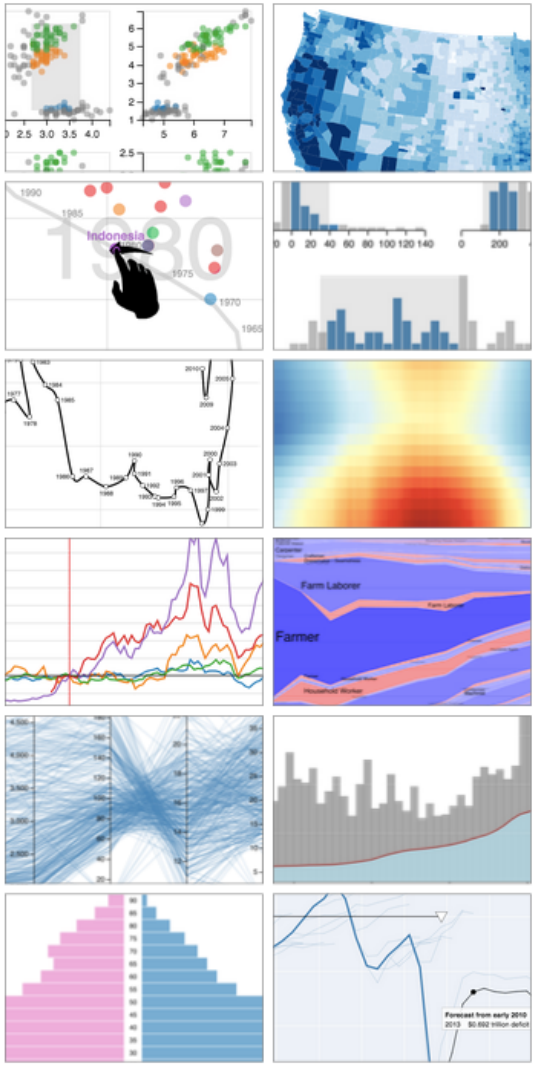
W6: Interactive Visualization, Part 2 (20%)

W7: Peer Review (5%), Final Project Proposal

W8: Final Project Milestone

W9: Final Project Deliverable (30%)

CP: Course Participation (10%)



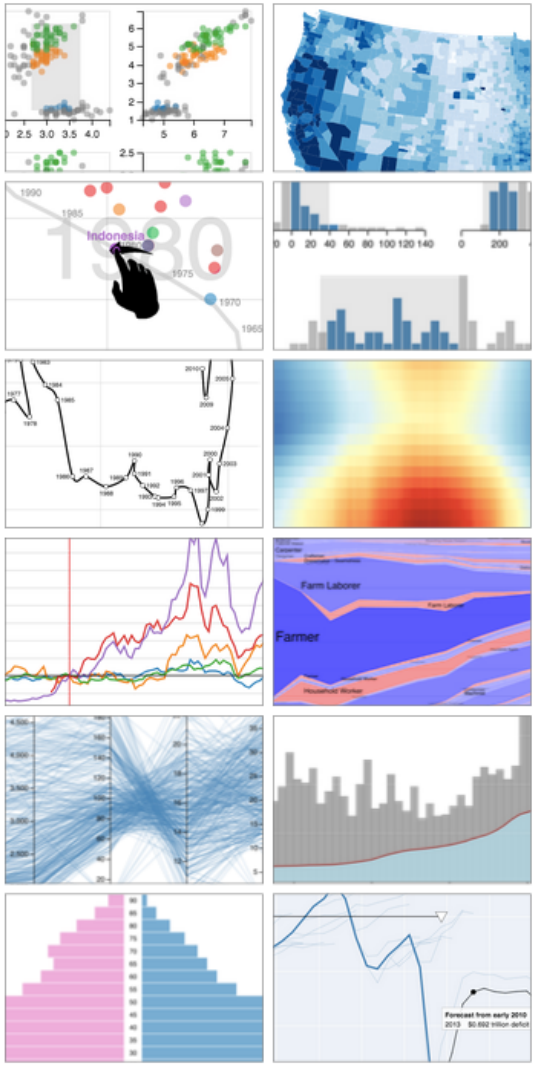
Course Participation

Lecture Attendance

Please attend lectures in person! That said, we know that illness, travel, etc. can prevent attendance. If you can't attend class, please review the recordings online.

Weekly Exercises

We have in-class exercises each week. Complete them even if you can't attend in person. We use "best-effort" grading, so it's OK if you don't complete everything during class time. Focus on assignments, not exercises, between sessions. You also get one exercise "pass".



Course Staff

Lead Instructor

Jeffrey Heer - Professor, CSE
Office Hours: Before/After Class, Gates G10

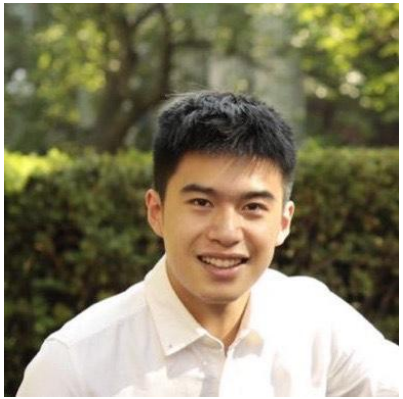
Teaching Assistants

Will Wang - Ph.D. Student, CSE
Office Hours: Thu 5-6pm, Zoom

Parum Misri - B.S. Student, CSE

Will (Huichen) Wang

wwill@cs.washington.edu



Second-year CS PhD student

Research:

- HCI, Visualization, Applications and Limitations of GenAI for Vis & Data Science

Fun:

- Ping Pong
- Movies
- Chess and Go

Undergrad TA: Parum Misri

- Class of 2025
  - BS in Computer Science
  - BS in Economics
- Interests:
  - Data science
  - Digital design
  - Sketching
  - Piano

Warm-Up Design Activity

Visual Encoding Exercise

5 17

How many visualizations can you think of for conveying these two numbers? Feel free to invent tasks or contexts. **Sketch as many as you can!**

Don’t stress over quality, go for quantity.

Time: ~5 minutes

Visual Encoding Exercise

5 17

Take a photo or screenshot of your visualizations, and post it to the shared thread on Ed.

Visual Encoding Exercise

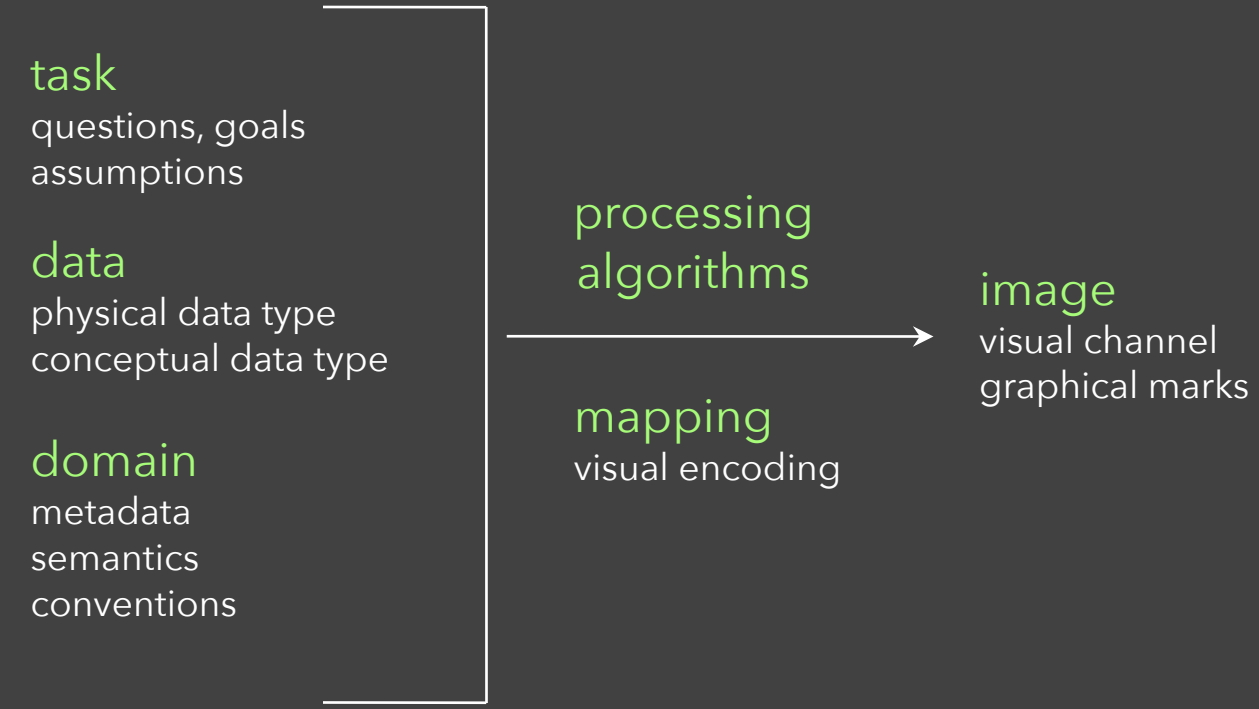
5 17

Share your designs with students near you. Introduce yourselves! Then compare your designs. How many ideas are the same? How many are different?

What do you find highly effective? Highly creative?

Visual Encoding Design

The Big Picture



Data Models

Represent data as a *table* (relation)

Each *row* (tuple) represents a record

Each *column* (field) represents a typed variable

Physical Type: integer, float, date, boolean, string…

Conceptual Type: temperatures, dollars, products…

For visualization it is helpful to classify fields according to the type of comparisons we wish to make:

Nominal (N), *Ordinal* (O), and *Quantitative* (Q) types

Nominal, Ordinal & Quantitative

Nominal, Ordinal & Quantitative

N - Nominal (labels or categories)

- Fruits: apples, oranges, …

Nominal, Ordinal & Quantitative

N - Nominal (labels or categories)

- Fruits: apples, oranges, …

O - Ordered

- Quality of meat: Grade A, AA, AAA

Nominal, Ordinal & Quantitative

N - Nominal (labels or categories)

- Fruits: apples, oranges, …

O - Ordered

- Quality of meat: Grade A, AA, AAA

Q - Interval (location of zero arbitrary)

- Dates: Jan, 19, 2006; Location: (LAT 33.98, LON -118.45)
- Only differences (i.e., intervals) may be compared

Nominal, Ordinal & Quantitative

N - Nominal (labels or categories)

- Fruits: apples, oranges, …

O - Ordered

- Quality of meat: Grade A, AA, AAA

Q - Interval (location of zero arbitrary)

- Dates: Jan, 19, 2006; Location: (LAT 33.98, LON -118.45)
- Only differences (i.e., intervals) may be compared

Q - Ratio (zero fixed)

- Physical measurement: Length, Mass, Time duration, …
- Counts and amounts

Nominal, Ordinal & Quantitative

N - Nominal (labels or categories)

- Operations: =, ≠

O - Ordered

- Operations: =, ≠, <, >

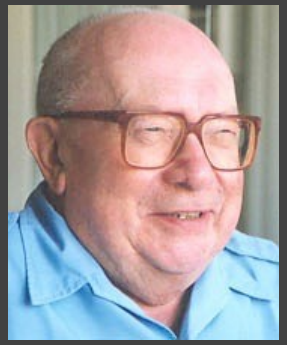
Q - Interval (location of zero arbitrary)

- Operations: =, ≠, <, >, -
- Can measure distances or spans

Q - Ratio (zero fixed)

- Operations: =, ≠, <, >, -, %
- Can measure ratios or proportions

Visual Language is a Sign System



Jacques Bertin

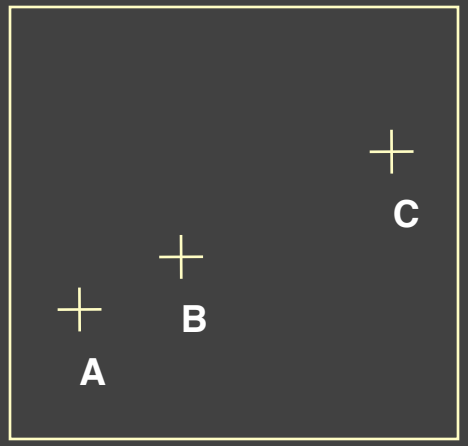
Images perceived as a set of signs

Sender encodes information in signs

Receiver decodes information from signs

Sémiologie Graphique, 1967

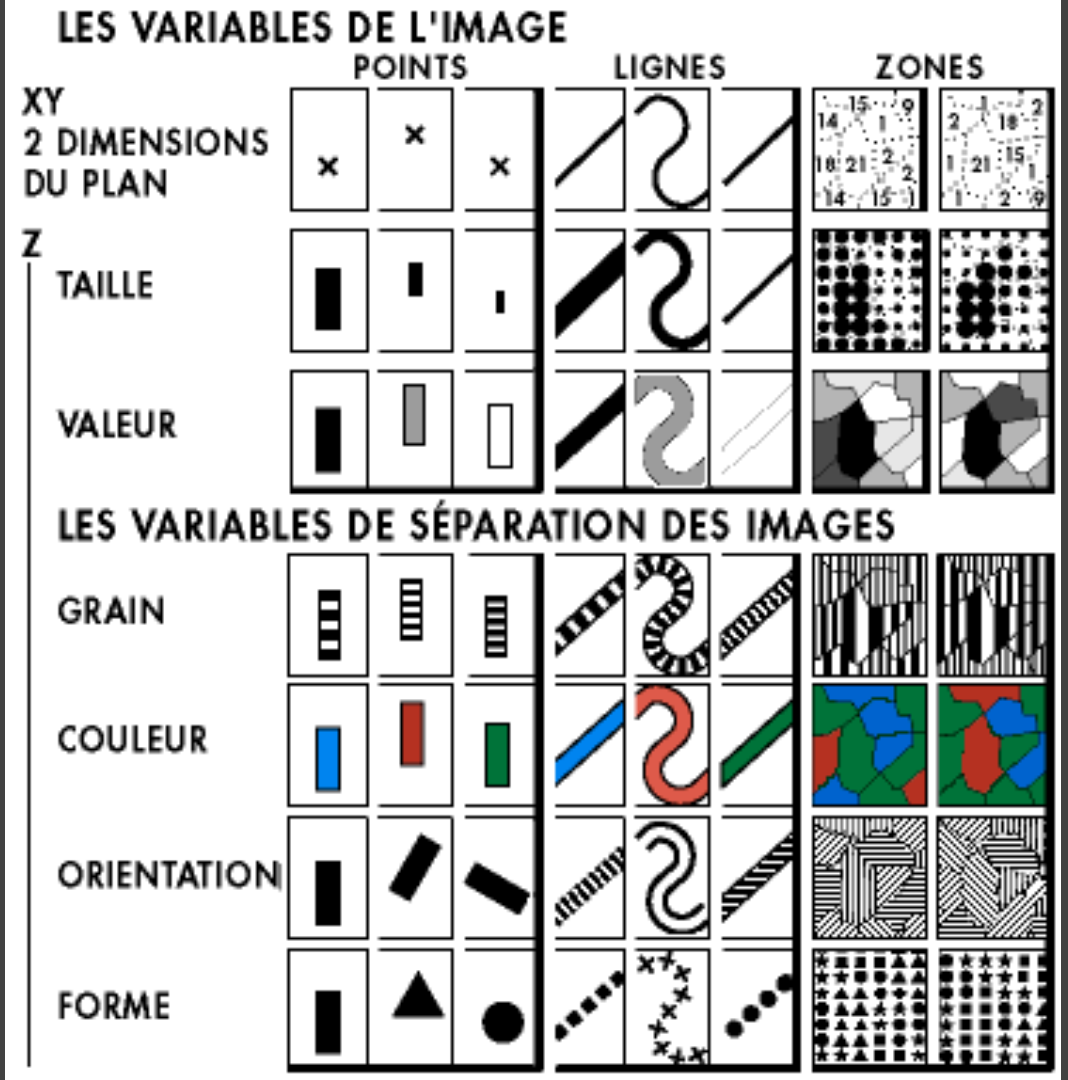
Bertin's Semiology of Graphics



1. A, B, C are distinguishable
2. B is between A and C.
3. BC is twice as long as AB.

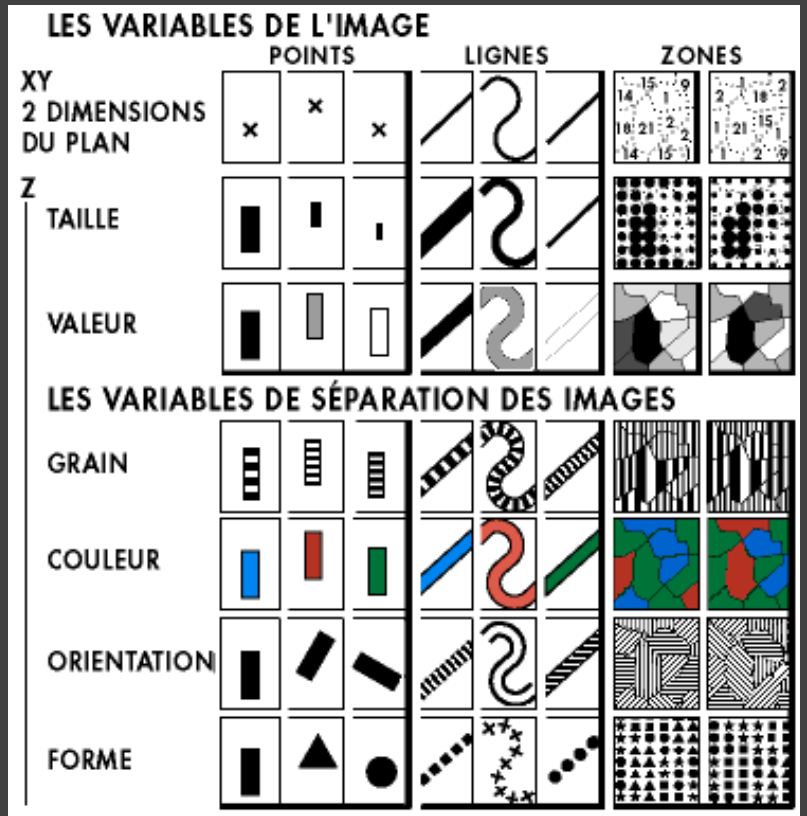
∴ Encode quantitative variables

"Resemblance, order and proportional are the three signfields in graphics." - Bertin


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

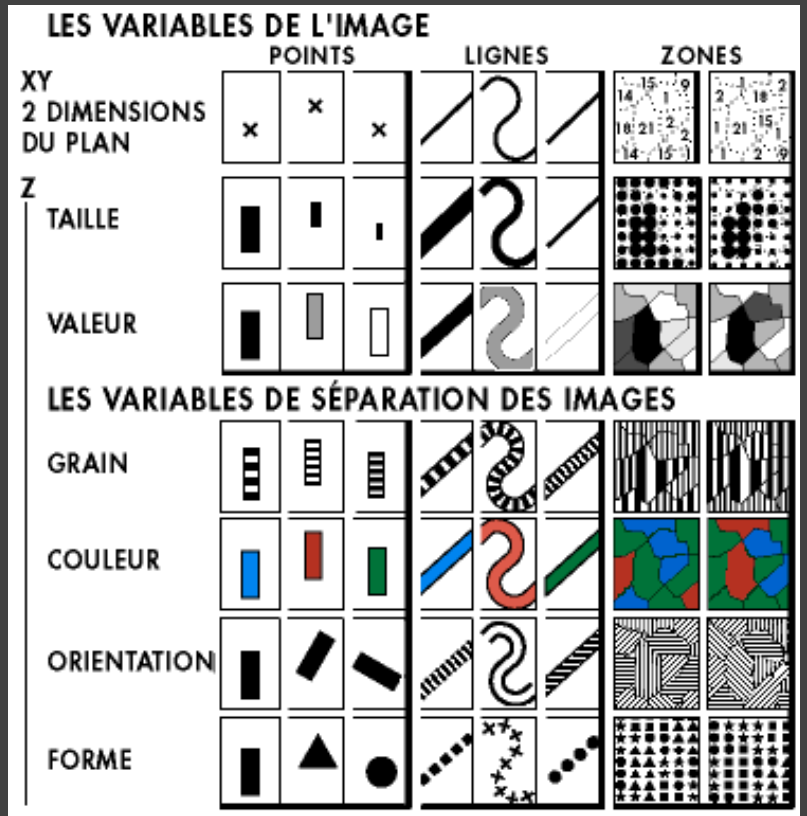
Visual Encoding Channels

Position (x 2)
Size
Value
Texture
Color
Orientation
Shape

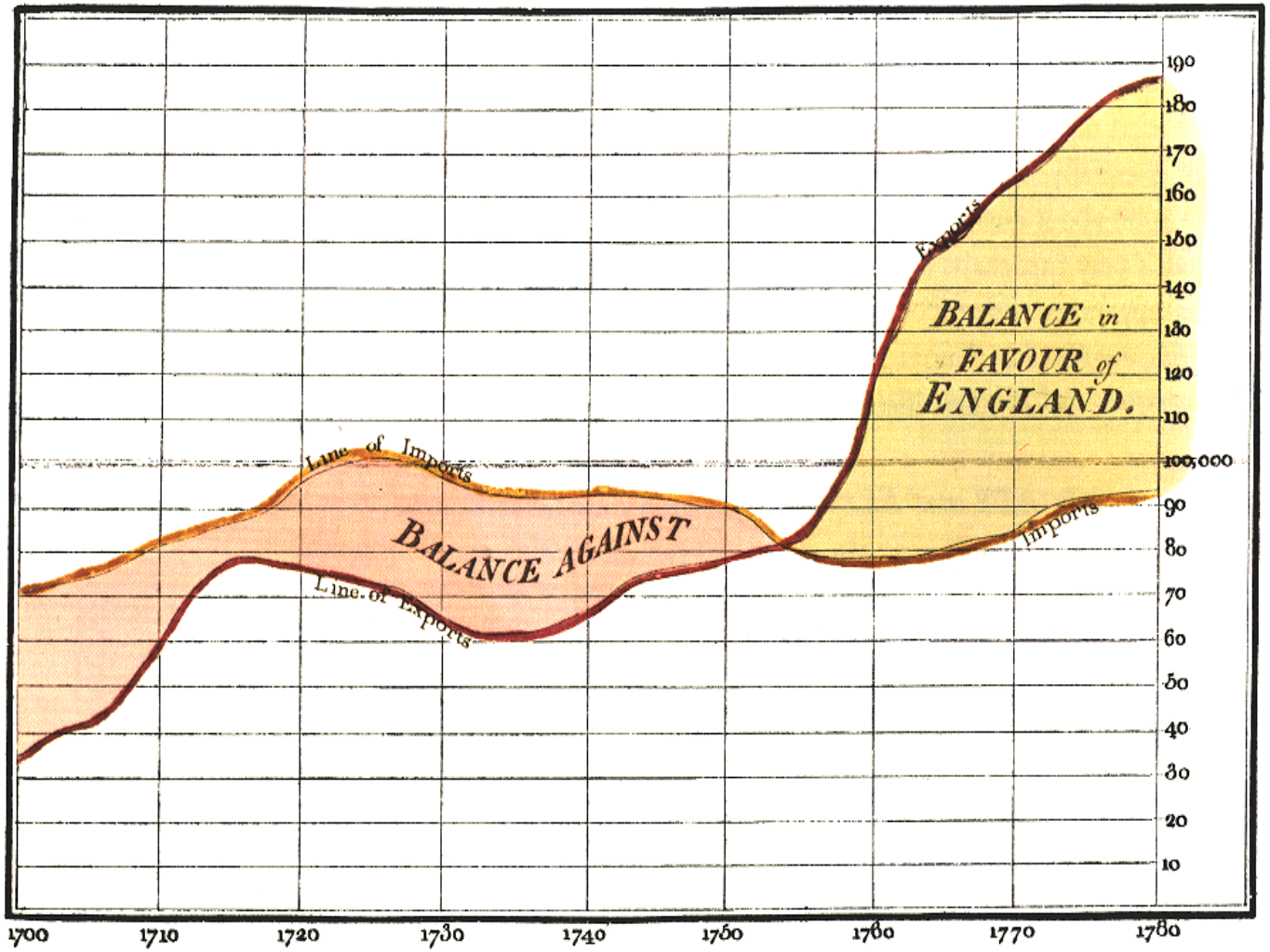


Visual Encoding Channels

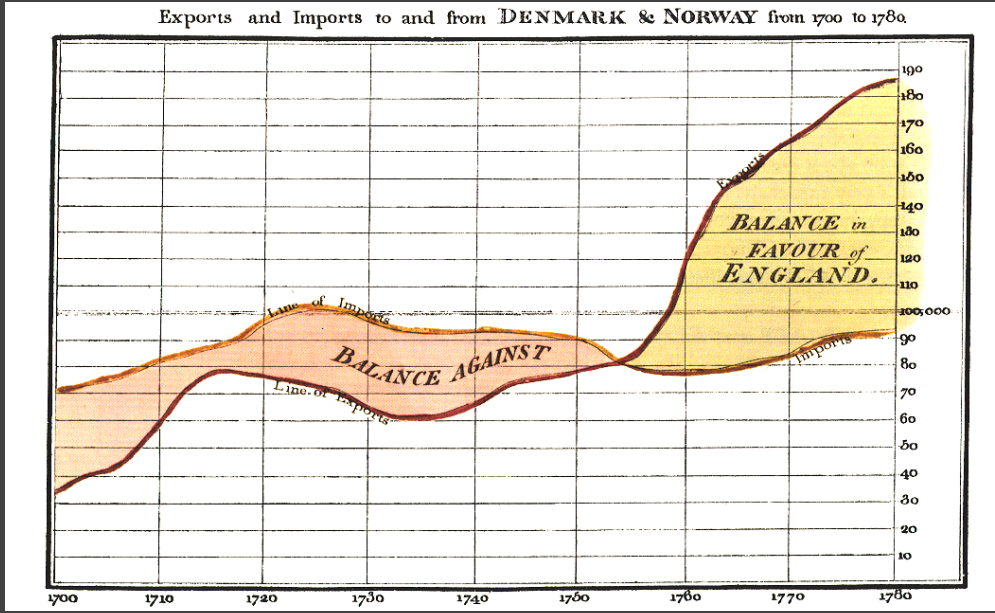
Position
Length
Area
Volume
Value
Texture
Color
Orientation
Shape
Transparency
Blur / Focus ...



Exports and Imports to and from DENMARK & NORWAY from 1700 to 1780.



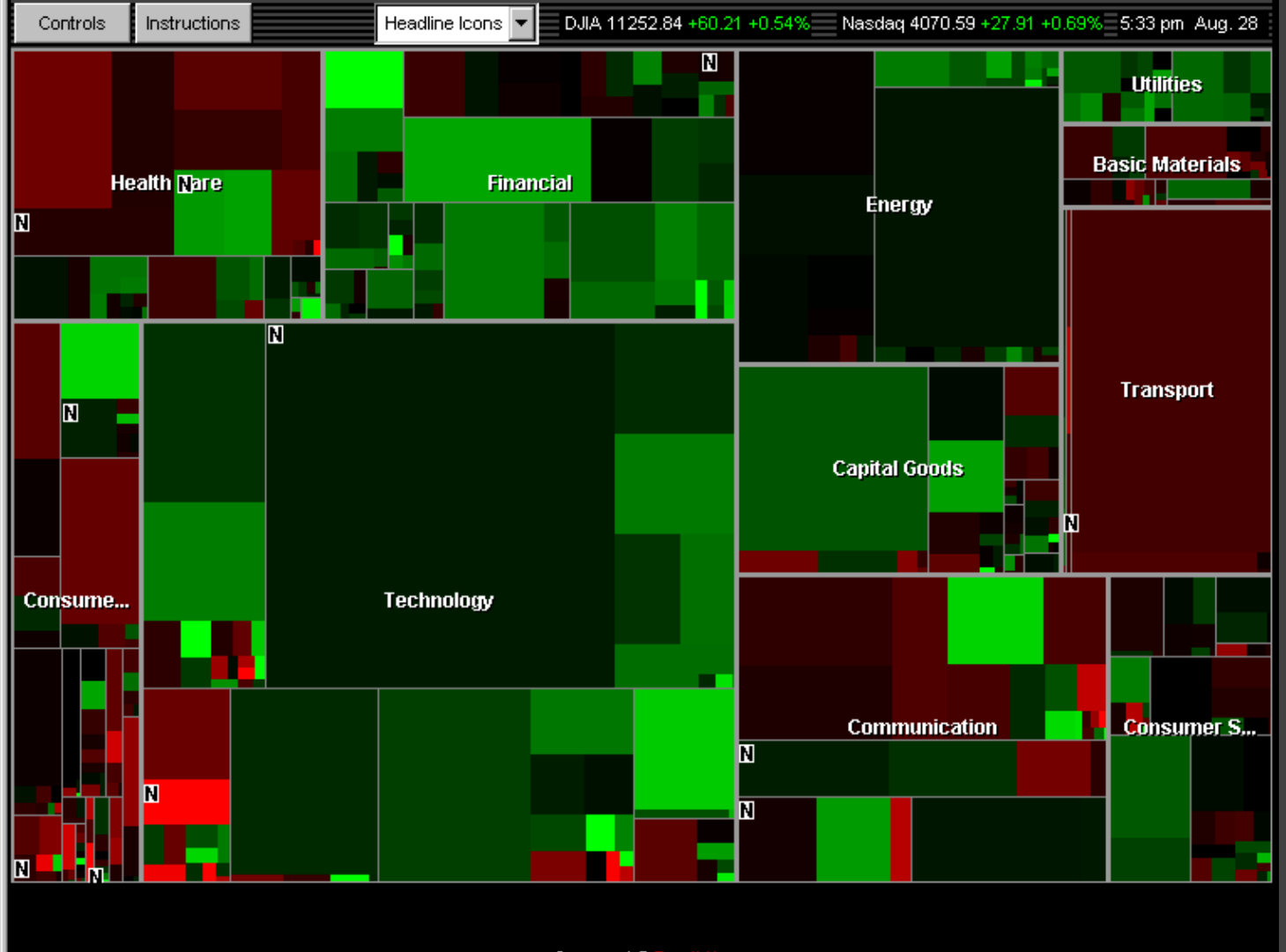
William Playfair, 1786



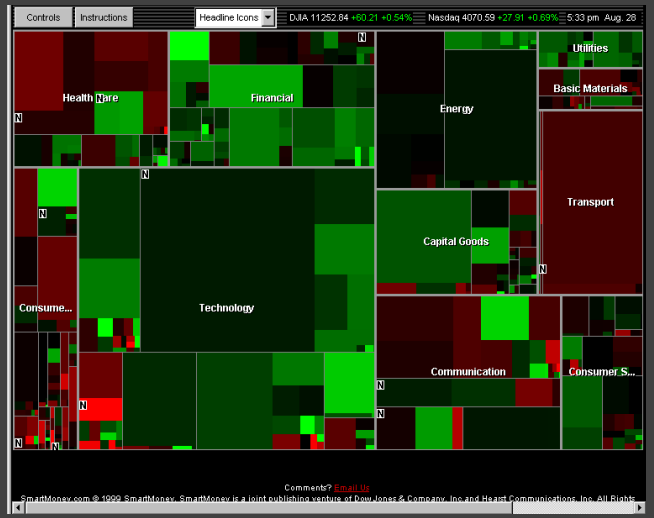
X-axis: year (Q)

Y-axis: currency (Q)

Color: imports/exports (N, O)


Controls
Instructions
Headline Icons
DJIA 11252.84 +60.21 +0.54%
Nasdaq 4070.59 +27.91 +0.69%
5:33 pm Aug. 28
Health Care
Financial
Energy
Utilities
Basic Materials
Transport
Capital Goods
Consume...
Technology
Communication
Consumer S...
N

Wattenberg's Map of the Market



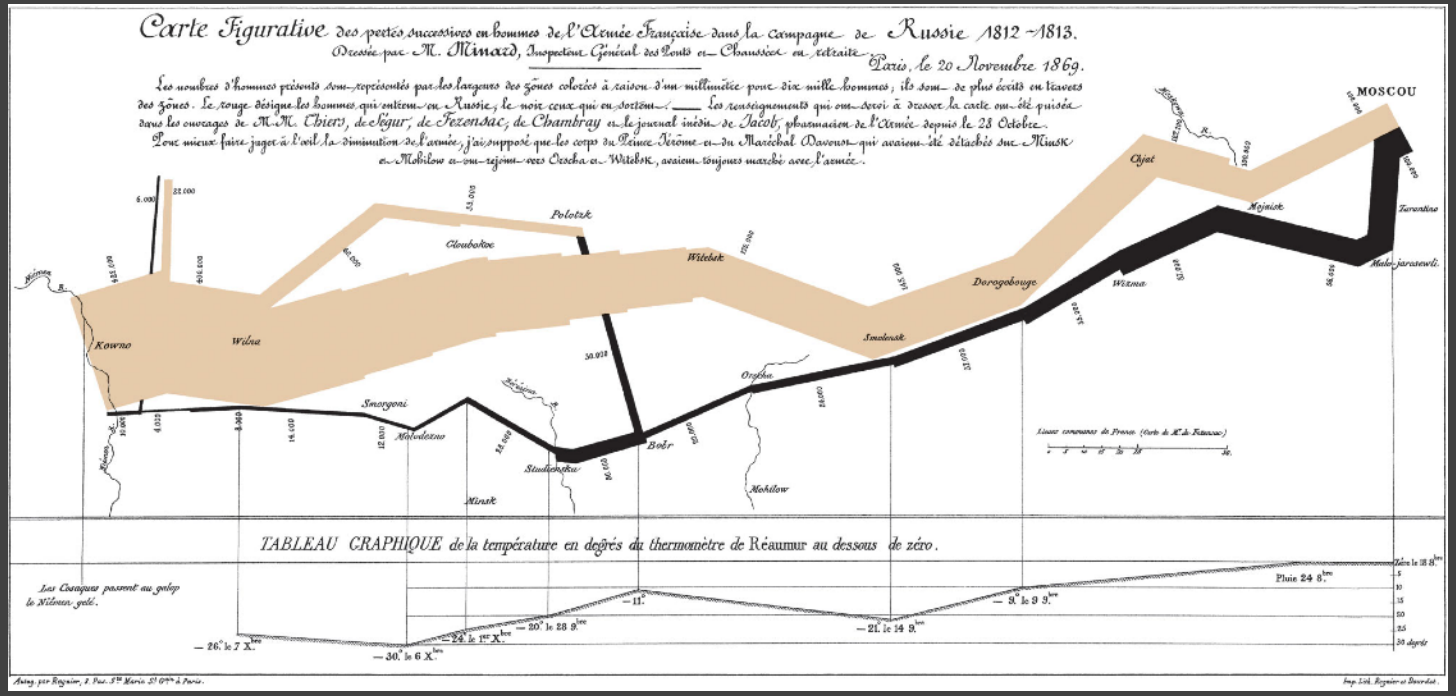
Rectangle Area: market cap (Q)

Rectangle Position: market sector (N), market cap (Q)

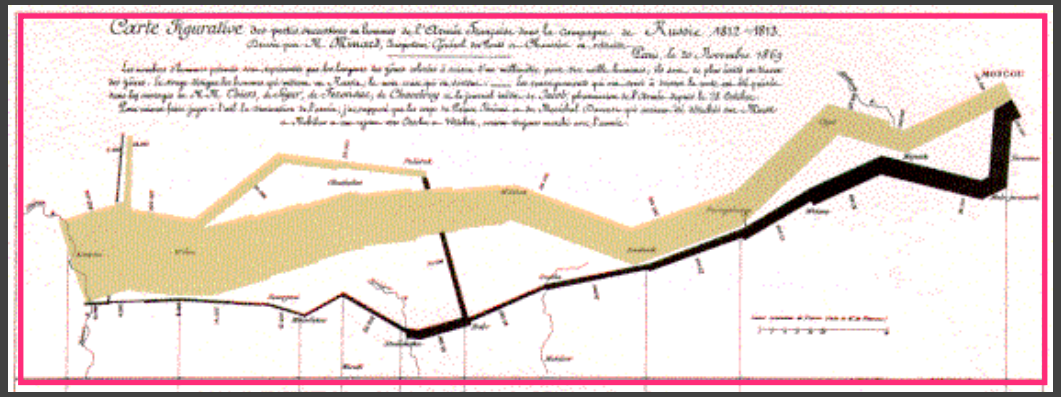
Color Hue: loss vs. gain (N, O)

Color Value: magnitude of loss or gain (Q)

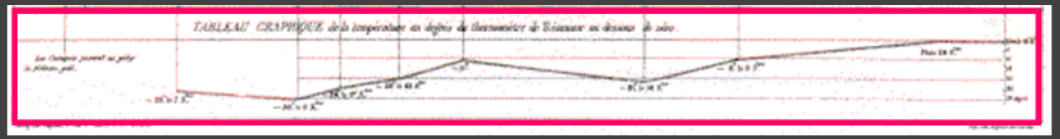
Minard 1869: Napoleon's March



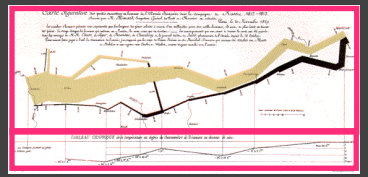
Single-Axis Composition



+



=



Mark Composition

Y-axis: temperature (Q)

\+ **X-axis**: longitude (Q) / time (O)

=

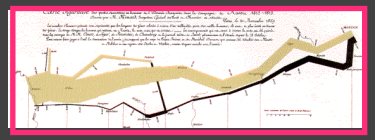
Temp over space/time (Q x Q)

Mark Composition

Y-axis: latitude (Q)

\+ **X-axis**: longitude (Q)

\+ **Width**: army size (Q)

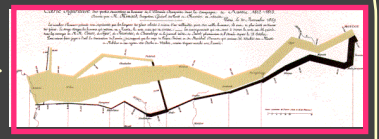


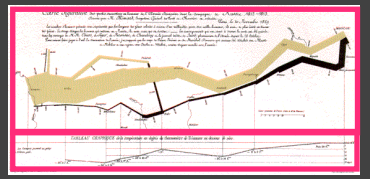
= Army position (Q x Q) and army size (Q)

latitude (Q)

longitude (Q)

army size (Q)

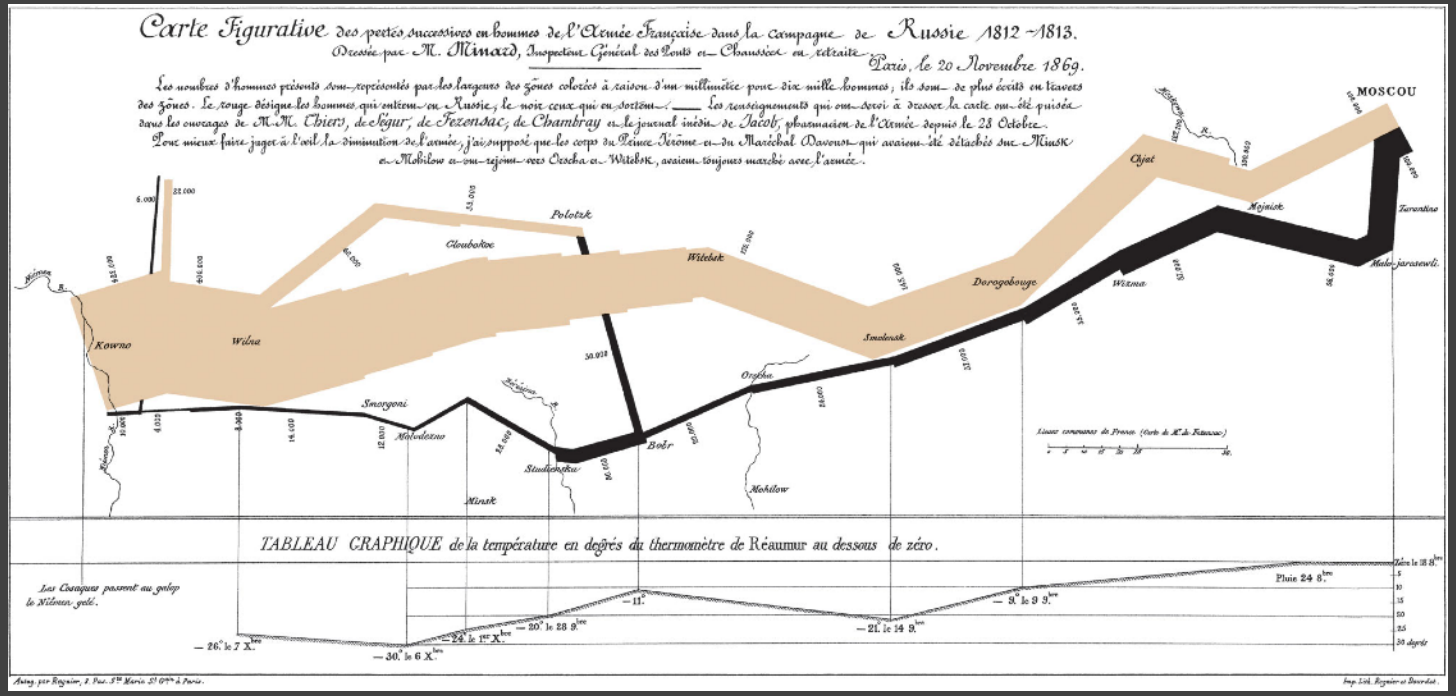




temperature (Q)

longitude (Q) / time (O)

Minard 1869: Napoleon's March



Depicts at least 5 quantitative variables. Any others?

Visualization Tools

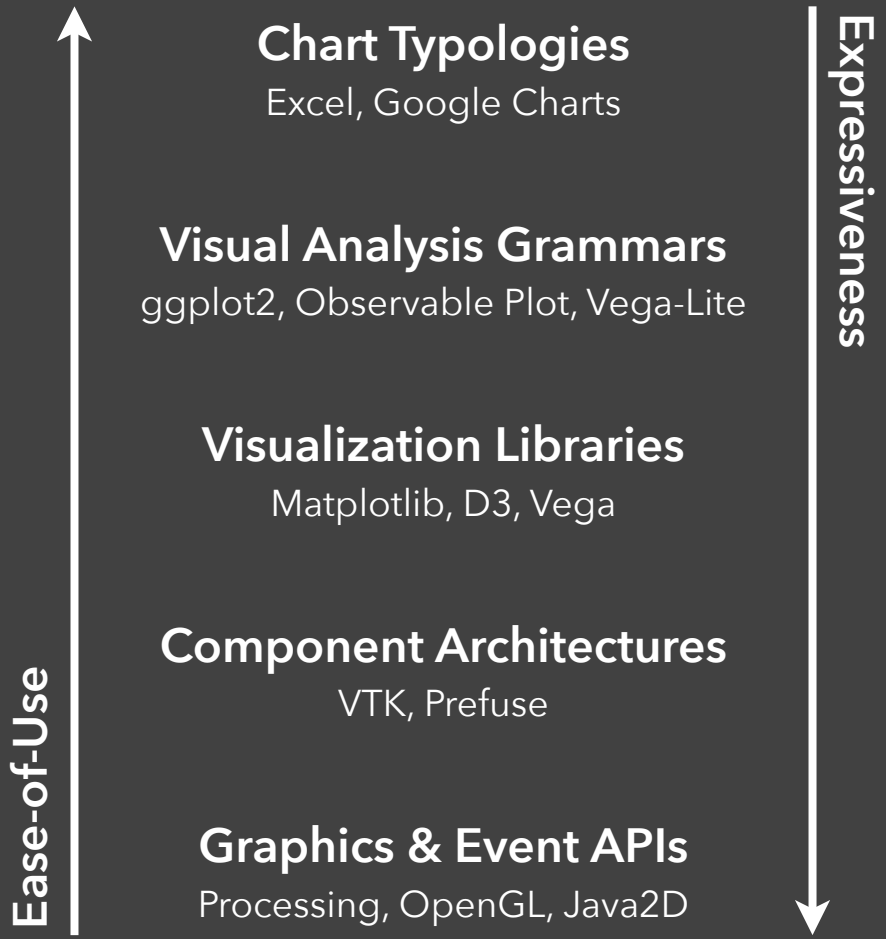


Chart Typologies
Excel, Google Charts
Visual Analysis Grammars
ggplot2, Observable Plot, Vega-Lite
Visualization Libraries
Matplotlib, D3, Vega
Component Architectures
VTK, Prefuse
Graphics & Event APIs
Processing, OpenGL, Java2D
Ease-of-Use
Expressiveness

Ease-of-Use

Chart Typologies
Excel, Google Charts

Visual Analysis Grammars
ggplot2, Observable Plot, Vega-Lite

Visualization Libraries
Matplotlib, D3, Vega

Component Architectures
VTK, Prefuse

Graphics & Event APIs
Processing, OpenGL, Java2D

Expressiveness

Offer **fine-grained control** for composing interactive graphics.

But require **verbose** specifications and technical expertise.

Ease-of-Use

Expressiveness

Chart Typologies
Excel, Google Charts

Visual Analysis Grammars
ggplot2, Observable Plot, Vega-Lite

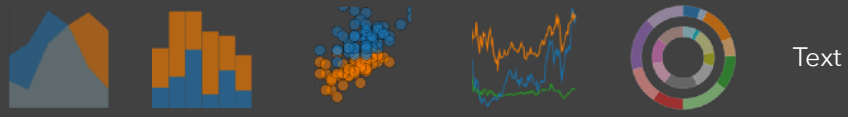
Visualization Libraries
Matplotlib, D3, Vega

Component Architectures
VTK, Prefuse

Graphics & Event APIs
Processing, OpenGL, Java2D

Grammar Building Blocks

Data	Input data to visualize
Transforms	Filter, aggregate, stats, layout
Scales	Map data values to visual values
Guides	Axes & legends to visualize scales
Marks	Data-representative graphics



Ease-of-Use

Expressiveness

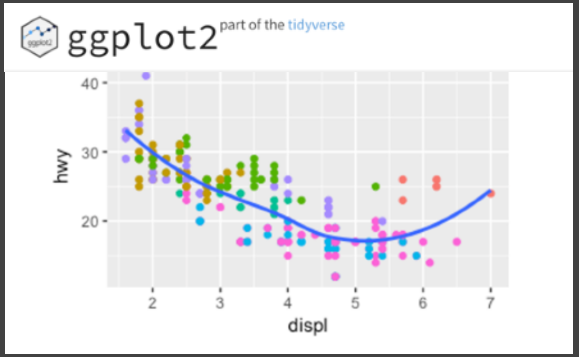
Chart Typologies
Excel, Google Charts

Visual Analysis Grammars
ggplot2, Observable Plot, Vega-Lite

Visualization Libraries
Matplotlib, D3, Vega

Component Architectures
VTK, Prefuse

Graphics & Event APIs
Processing, OpenGL, Java2D



Facilitate **rapid exploration** with **concise** specifications by omitting low-level details.

Infer **sensible defaults** and customize by overriding defaults.

Ease-of-Use

Chart Typologies
Excel, Google Charts

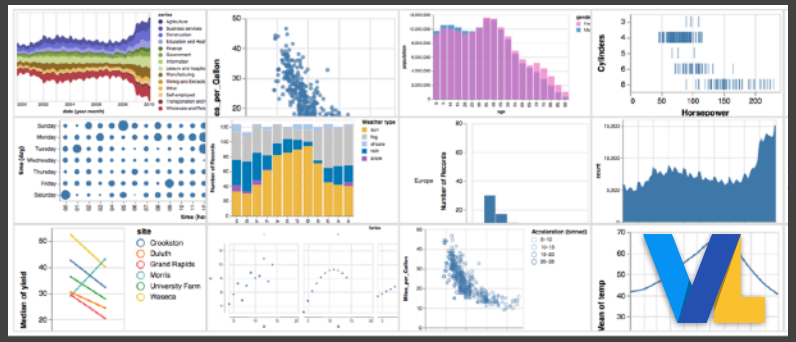
Visual Analysis Grammars
ggplot2, Observable Plot, Vega-Lite

Visualization Libraries
Matplotlib, D3, Vega

Component Architectures
VTK, Prefuse

Graphics & Event APIs
Processing, OpenGL, Java2D

Expressiveness



Facilitate **rapid exploration**
with **concise** specifications
by omitting low-level details.

Infer **sensible defaults** and
customize by overriding defaults.

A Dive Into Vega-Lite

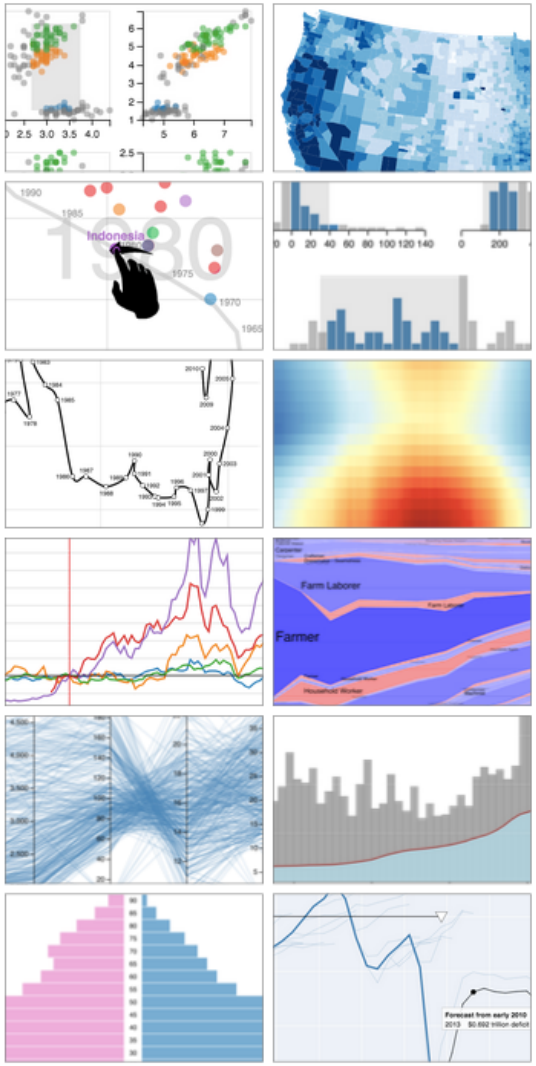
Vega-Lite is a *Visual Analysis Grammar*

Similar in spirit to how SQL provides a language for expressing database queries, Vega-Lite is a high-level language for describing visualizations.

Vega-Lite compiles specifications to interactive, web-based visualizations.

Developed at UW, initially by Dominik Moritz, Kanit Wongsuphasawat, Arvind Satyaranarayan, and Jeffrey Heer. It is now a popular open source project with many contributors.

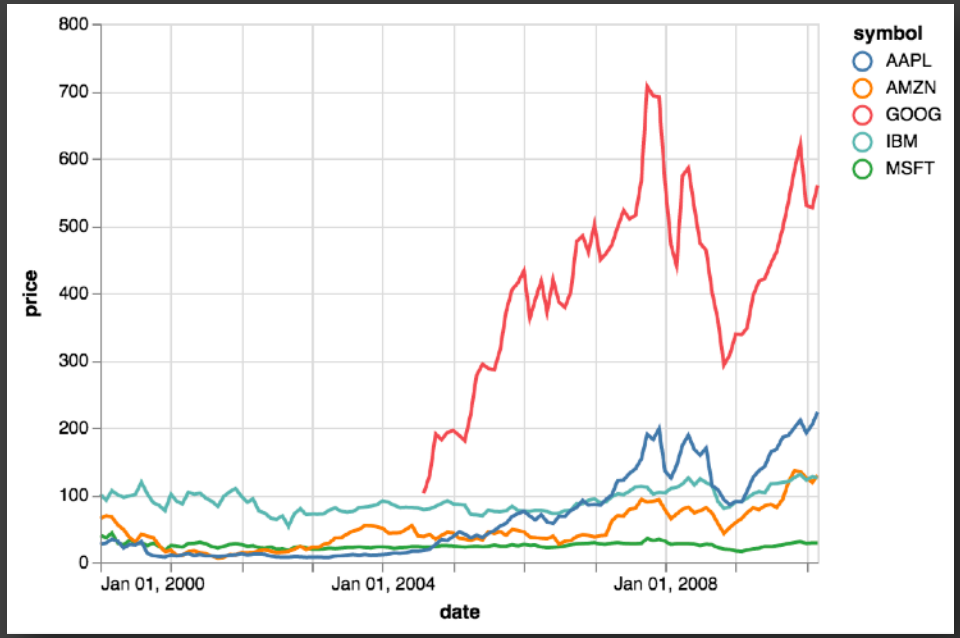
Uptake in Python and Jupyter via the Altair API.



Vega-Lite is a *Visual Analysis Grammar*

Vega Altair Python API

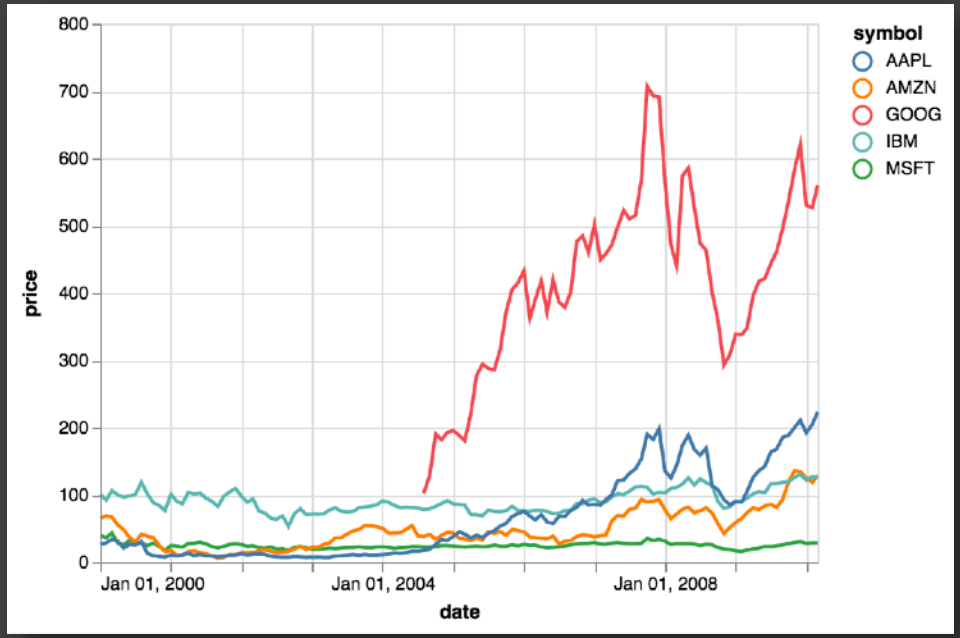
```
alt.Chart('stocks.csv')
  .mark_line()
  .encode(
    x='date',
    y='price',
    color='symbol'
  )
```



Vega-Lite is a *Visual Analysis Grammar*

Vega-Lite JavaScript API

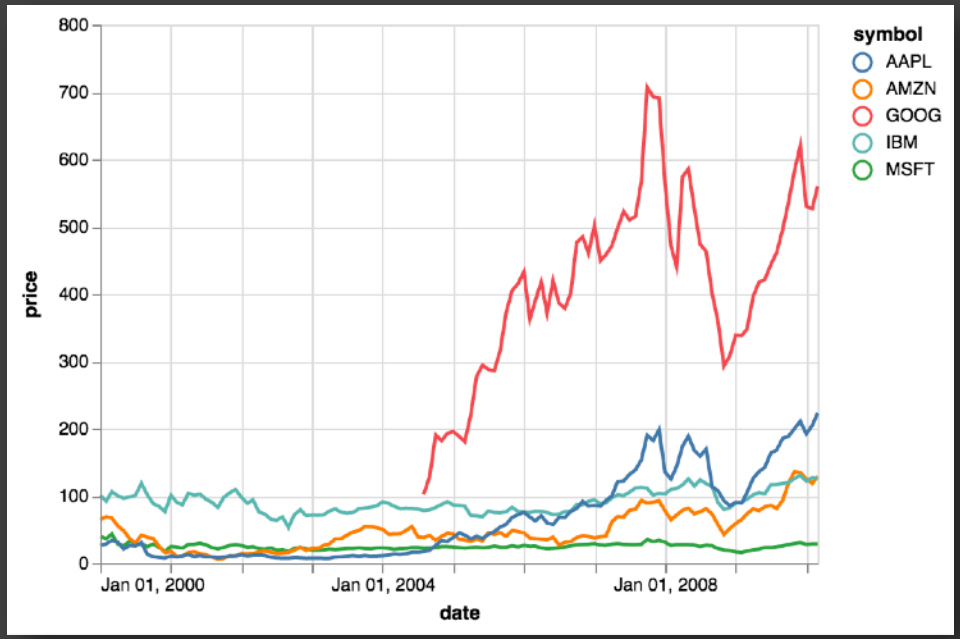
```
vl.data('stocks.csv')
  .markLine()
  .encode(
    vl.x().fieldT('date'),
    vl.y().fieldQ('price'),
    vl.color().fieldN('symbol')
  )
```



Vega-Lite is a *Visual Analysis Grammar*

Vega-Lite JSON Specification

```
{
  data: {url: "stocks.csv"},
  mark: "line",
  encoding: {
    x: {
      type: "temporal",
      field: "date"
    },
    y: {
      type: "quantitative",
      field: "price"
    },
    color: {
      type: "nominal",
      field: "symbol"
    }
  }
}
```



Specifying Single Views

Abstract Data  Visual Representation

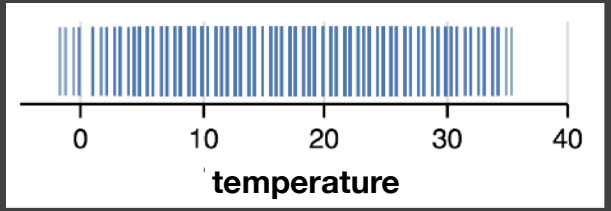
Specifying Single Views

Abstract Data → Visual Representation

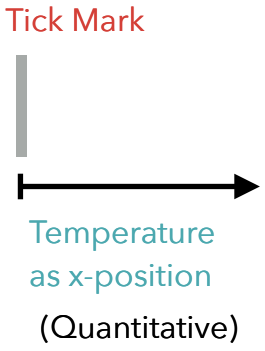
Weather Data for Seattle

date	temperature	precipitation	weather
1/1	10.6	10.9	"rain"
1/2	11.7	0.8	"drizzle"
1/3	12.2	10.2	"rain"
…	…	…	…

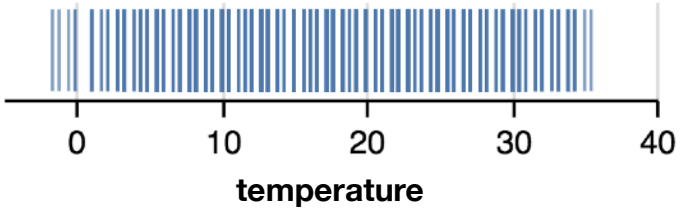
Strip Plot of Temperature



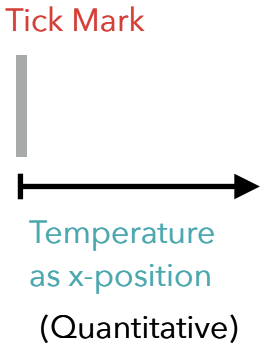
Strip Plot = (Tick with x=field)



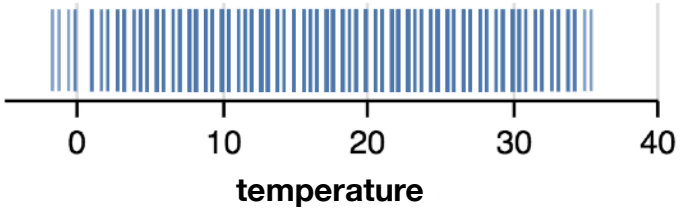

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      field: "temperature",
      type: "quantitative"
    }
  }
}
```



Strip Plot = (Tick with x=field)



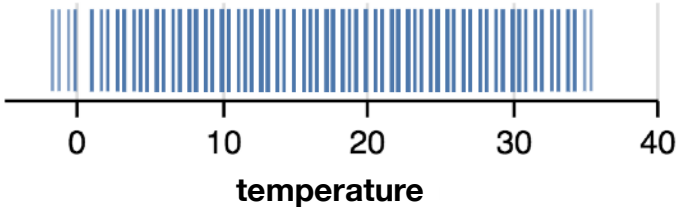

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      field: "temperature",
      type: "quantitative"
    }
  }
}
```



Vega-Lite is portable JSON specification

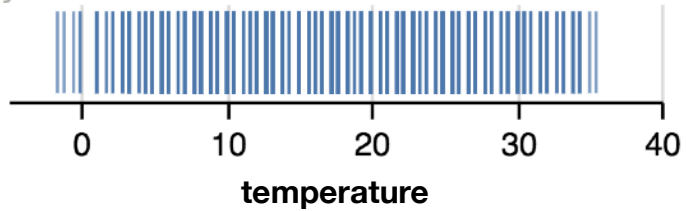
Strip Plot = (Tick with x=field)

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      field: "temperature",
      type: "quantitative"
    }
  }
}
```



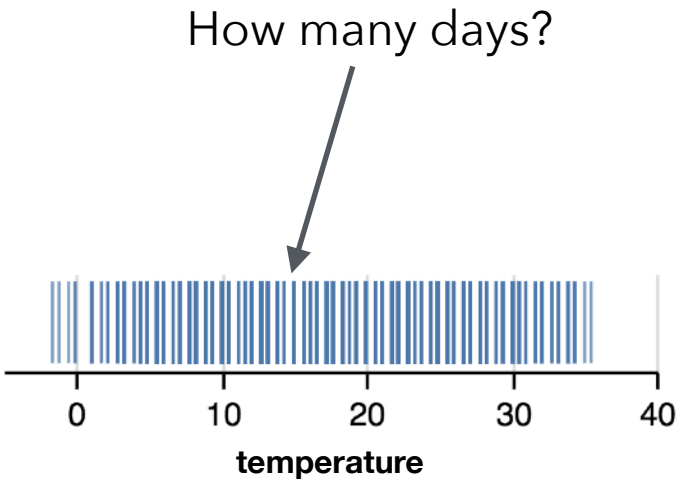
Strip Plot: Default Scales and Axes

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      field: "temperature",
      type: "quantitative",
      scale: {type: "linear", domain: [-10, 40], ...}
      axis: {title: "temperature", grid: true, ...}
    }
  }
}
```



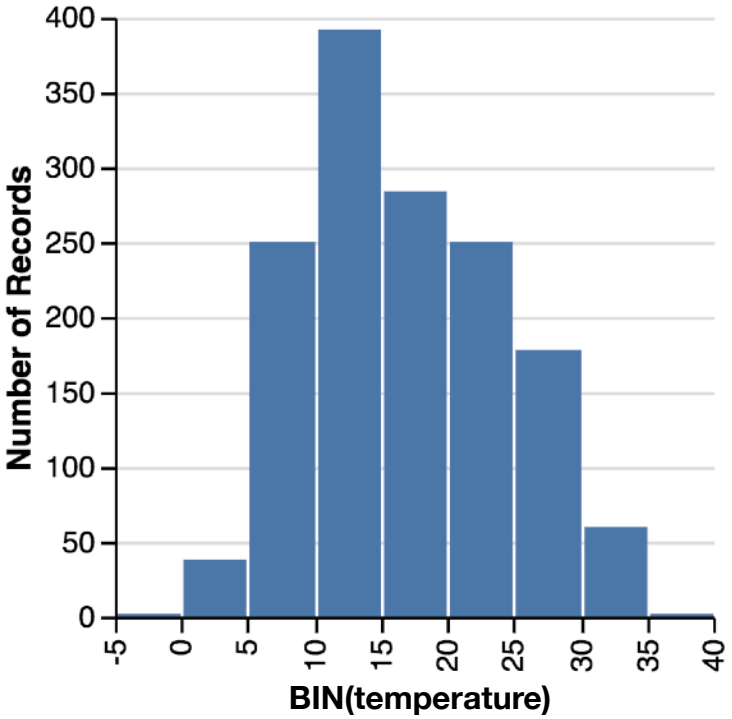
Strip Plot

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      field: "temperature",
      type: "quantitative"
    }
  }
}
```



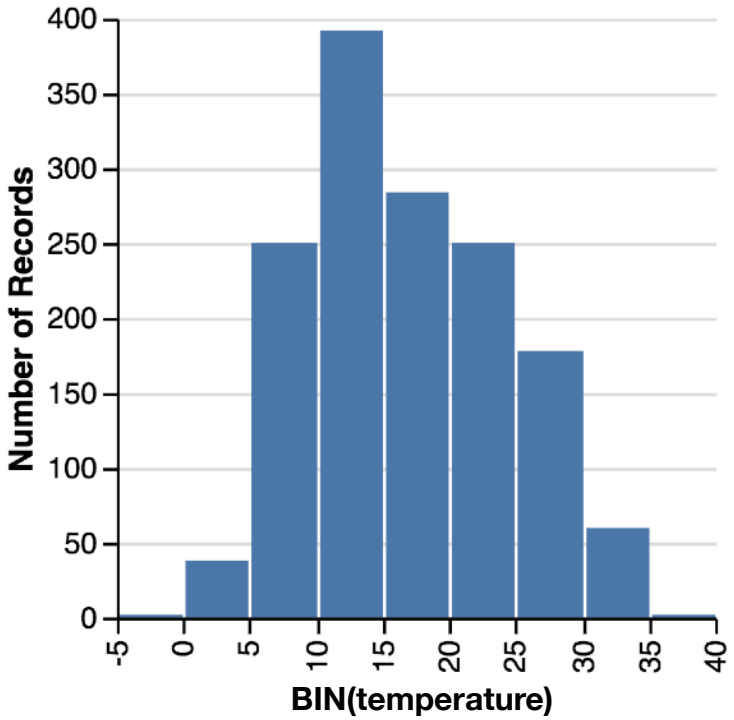
Histogram

Goal



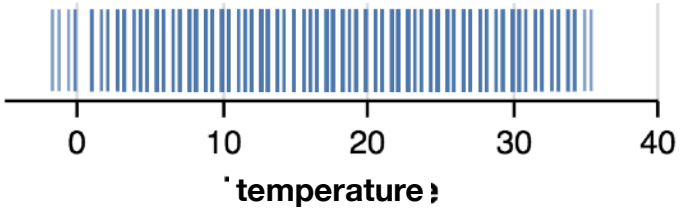
Histogram = (Bar with x=binned field, y=count)

Goal



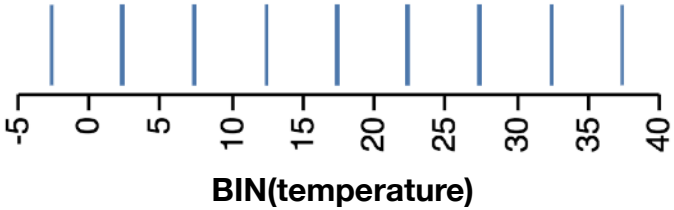
Histogram = (Bar with x=binned field, y=count)

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      field: "temperature",
      type: "quantitative"
    }
  }
}
```



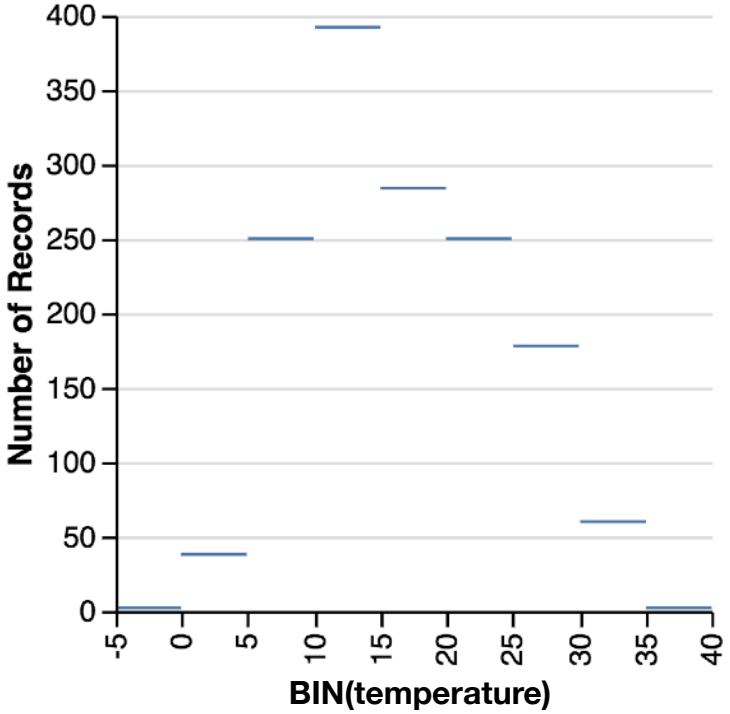
Histogram = (Bar with x=binned field, y=count)

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    }
  }
}
```



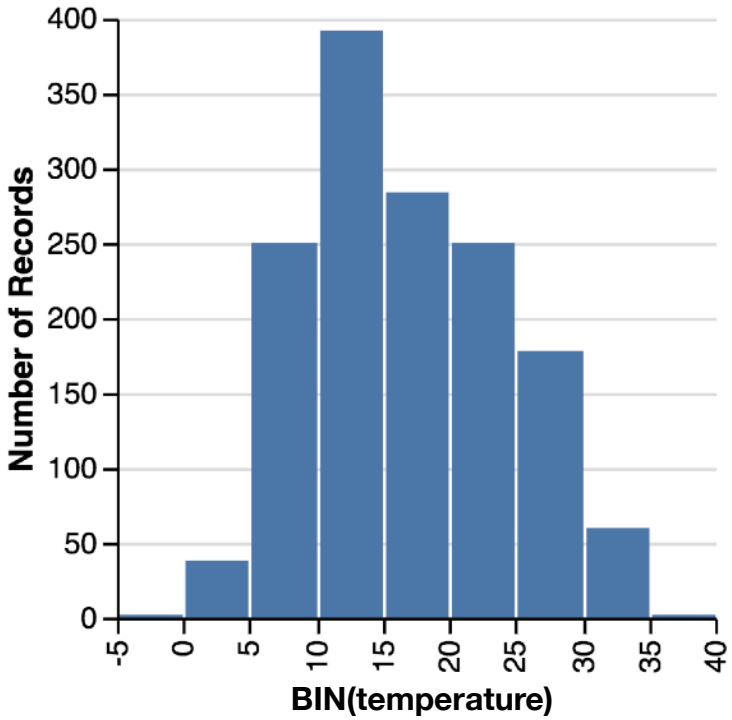
Histogram = (Bar with x=binned field, y=count)

```
{
  data: {url: "weather-seattle.json"},
  mark: "tick",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
    y: {
      aggregate: "count",
      type: "quantitative"
    }
  }
}
```



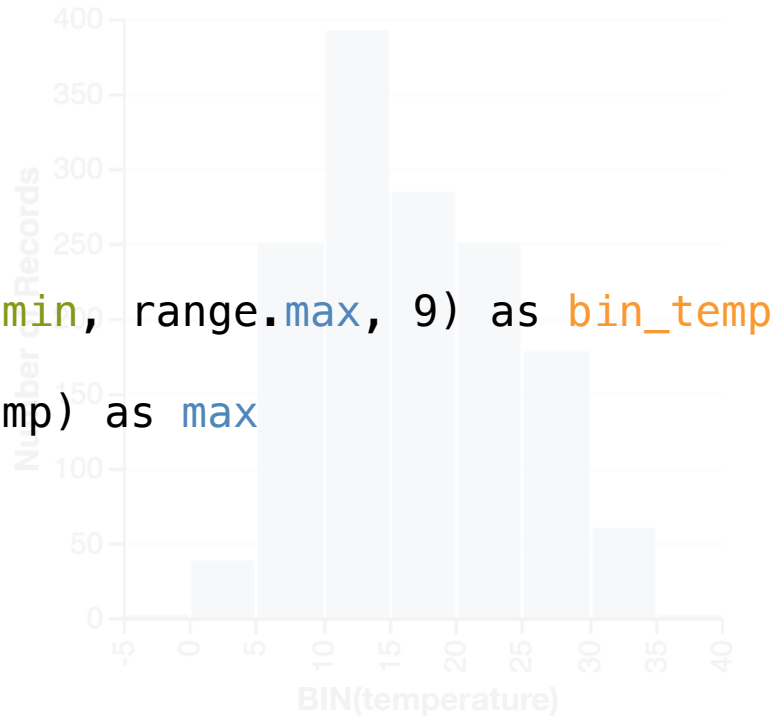
Histogram = (Bar with x=binned field, y=count)

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
    y: {
      aggregate: "count",
      type: "quantitative"
    }
  }
}
```



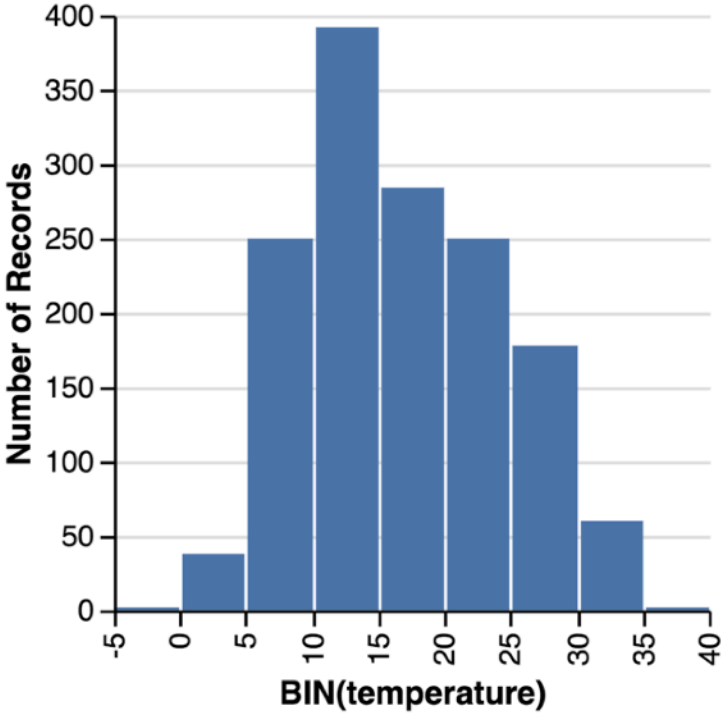
Histogram = (Bar with x=binned field, y=count)

```
SELECT bin_temp, count(*)
FROM (
  SELECT floor(weather.temp, range.min, range.max, 9) as bin_temp
  FROM weather, (
    SELECT min(temp) as min, max(temp) as max
  ) range
)
```



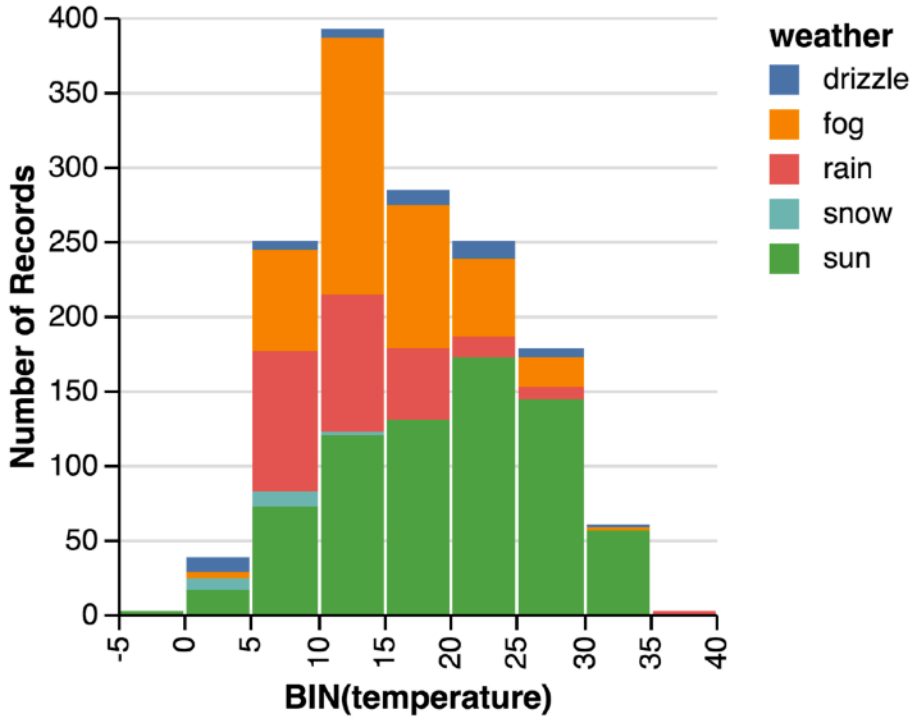
Histogram

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
    y: {
      aggregate: "count",
      type: "quantitative"
    }
  }
}
```



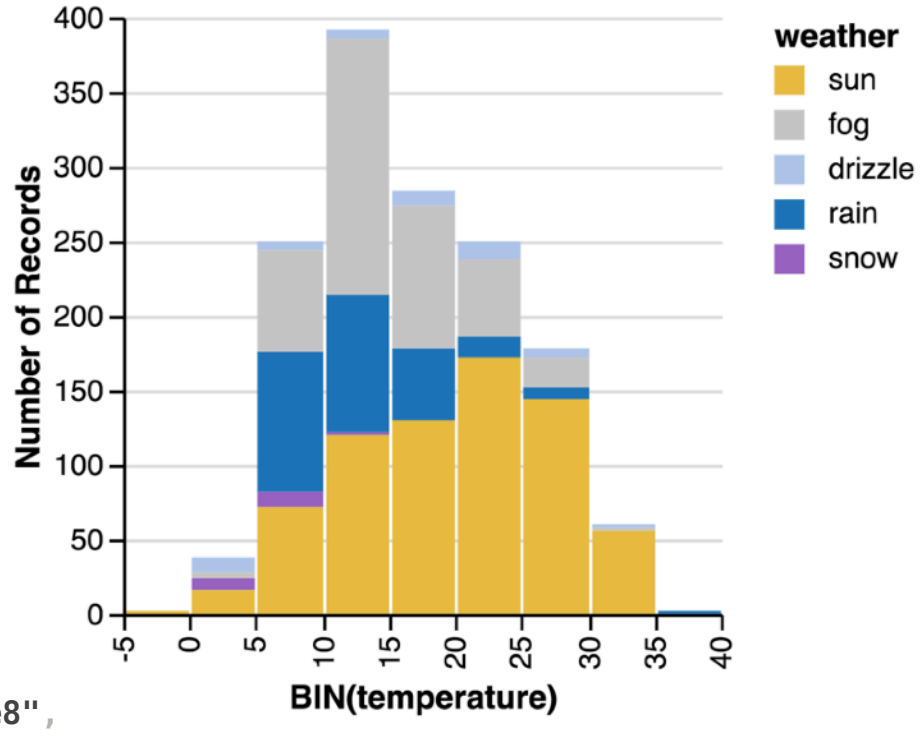
Histogram + Color

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
    y: {
      aggregate: "count",
      type: "quantitative"
    },
    color: {
      field: "weather",
      type: "nominal"
    }
  }
}
```



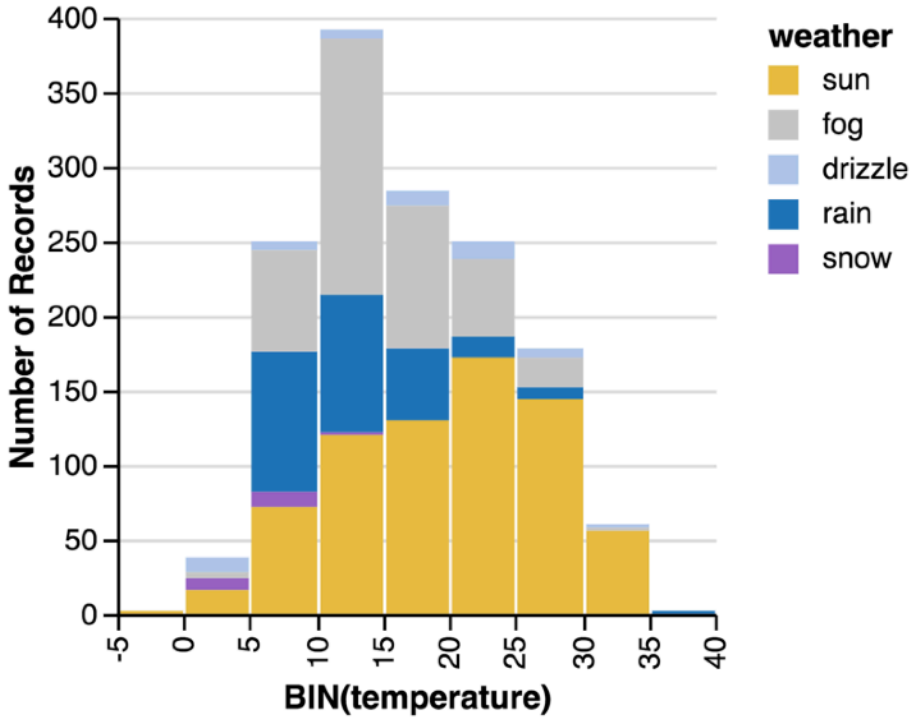
Histogram + Color

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
    y: {
      aggregate: "count",
      type: "quantitative"
    },
    color: {
      field: "weather",
      type: "nominal",
      "scale": {
        "domain": ["sun","fog","drizzle",
                   "rain","snow"],
        "range": ["#e7ba52","#c7c7c7","#aec7e8",
                  "#1f77b4","#9467bd"]
      }
    }
```



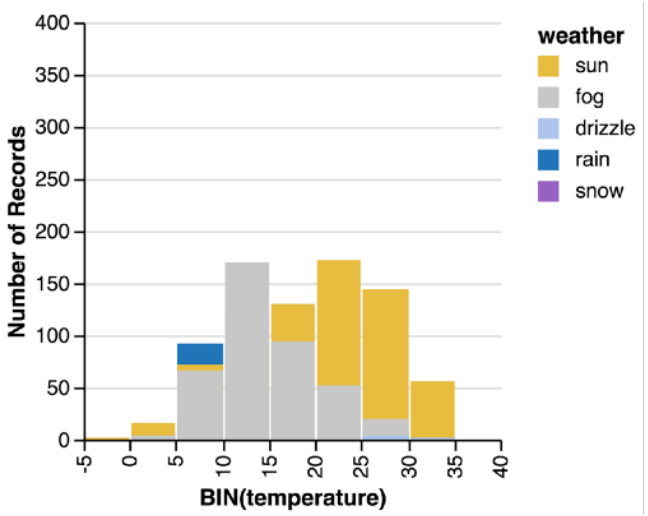
Histogram + Color = Stacked Histogram

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
    y: {
      aggregate: "count",
      type: "quantitative"
    },
    color: {
      field: "weather",
      type: "nominal",
      ...
      }
    }
  }
}
```

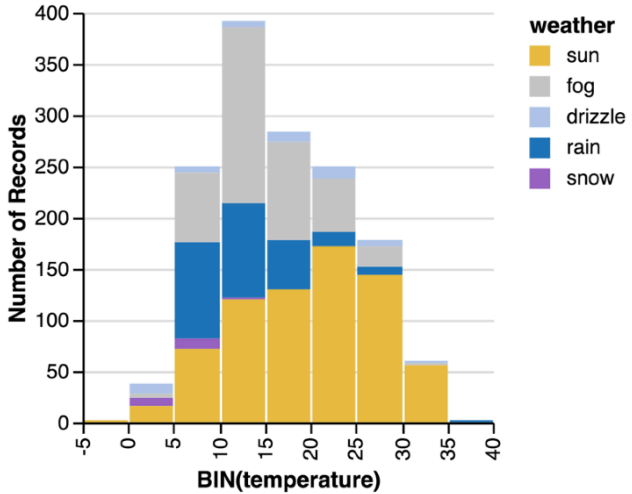


Stacked Histogram: Sensible Defaults

no stack



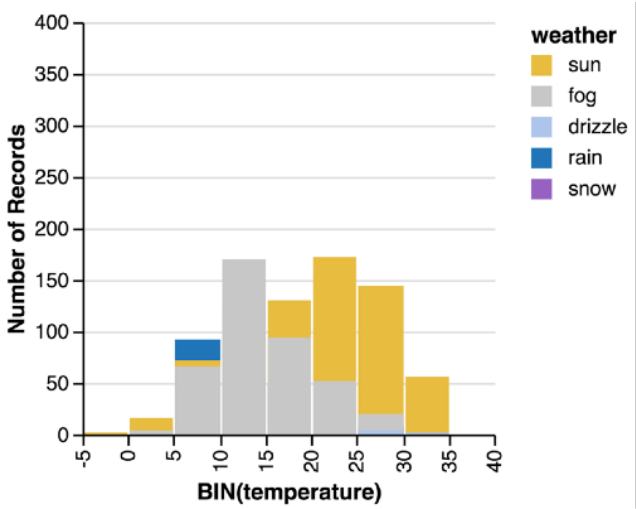
stack (default)



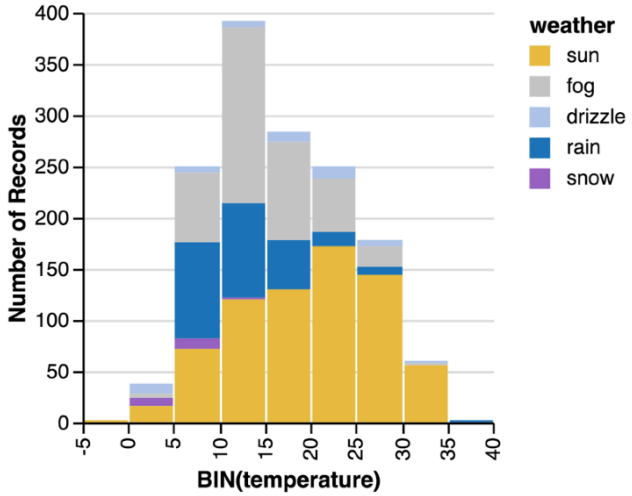
Stacked Histogram: Sensible Defaults

Channel (color) + Mark (bar) automatically enables stacking: a layout transform.

no stack



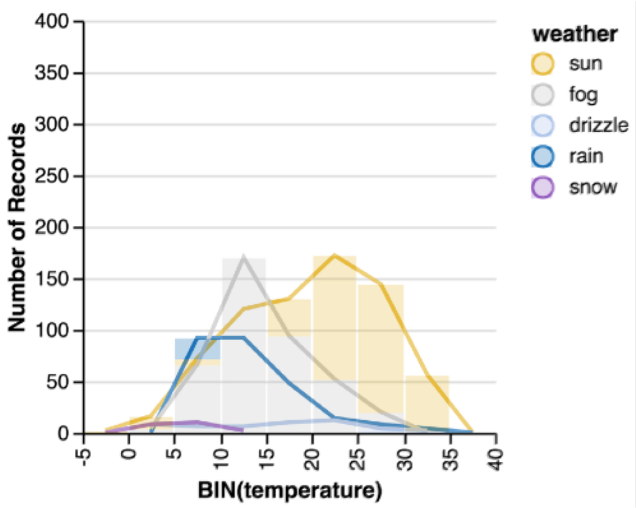
stack (default)



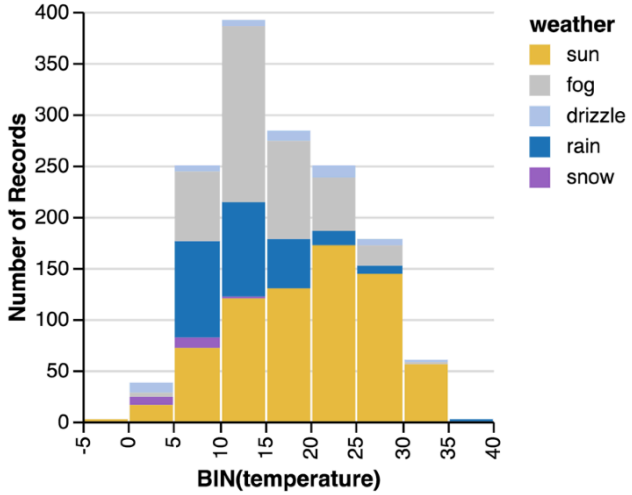
Stacked Histogram: Sensible Defaults

Channel (color) + Mark (bar) automatically enables stacking: a layout transform.

no stack → overlap

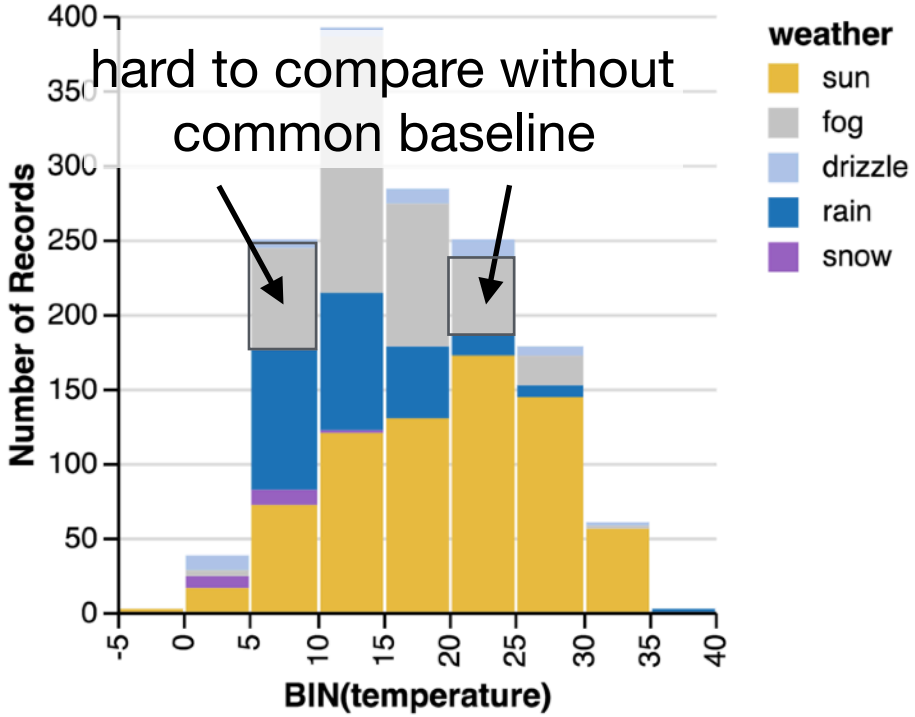


stack (default)



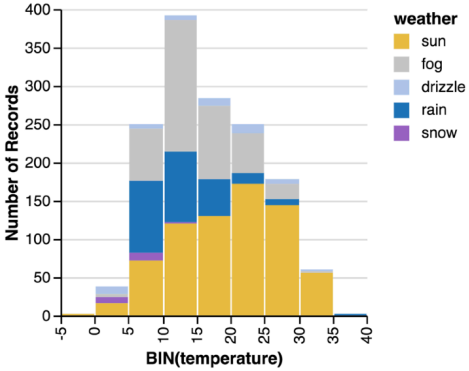
Histogram + Color = Stacked Histogram

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {
      bin: true,
      field: "temperature",
      type: "quantitative"
    },
    y: {
      aggregate: "count",
      type: "quantitative"
    },
    color: {
      field: "weather",
      type: "nominal"
    }
  }
}
```



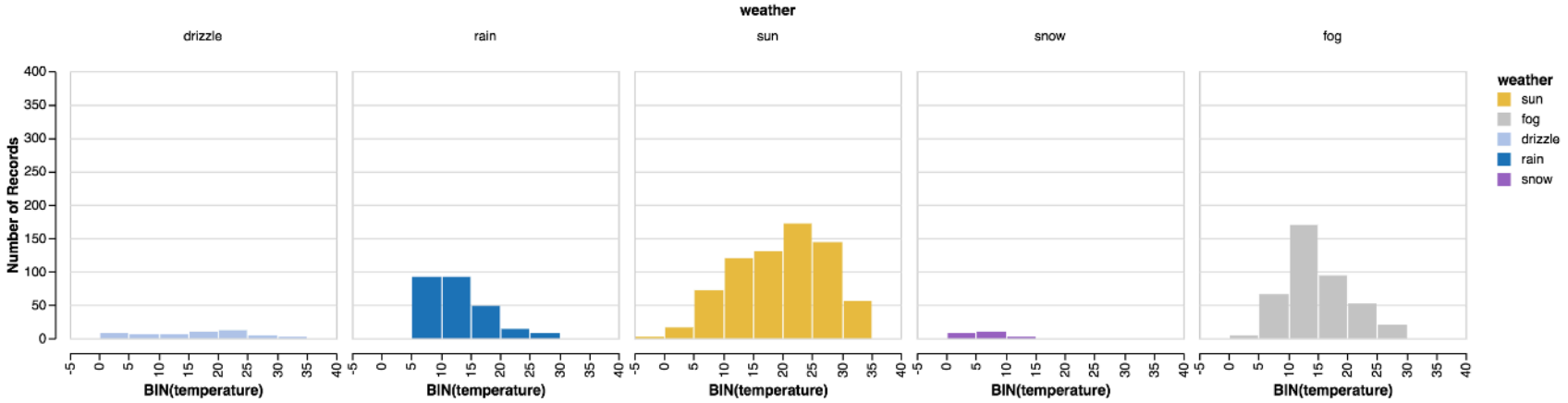
Histogram + Color = Stacked Histogram

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {bin: true, field: "temperature", type: "quantitative"},
    y: {aggregate: "count",              type: "quantitative"},
    color: {       field: "weather",     type: "nominal"}
  }
}
```



Histogram + Column = Trellis Histogram

```
{
  data: {url: "weather-seattle.json"},
  mark: "bar",
  encoding: {
    x: {bin: true, field: "temperature", type: "quantitative"},
    y: {aggregate: "count",              type: "quantitative"},
    column: {      field: "weather",     type: "nominal"}
  }
}
```



Ease-of-Use

Chart Typologies
Excel, Google Charts

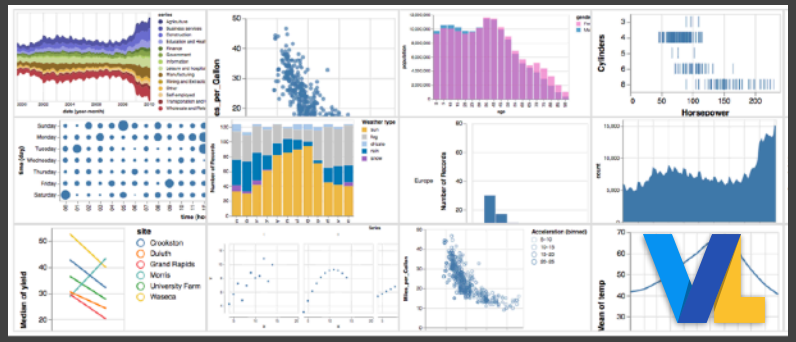
Visual Analysis Grammars
ggplot2, Observable Plot, Vega-Lite

Visualization Libraries
Matplotlib, D3, Vega

Component Architectures
VTK, Prefuse

Graphics & Event APIs
Processing, OpenGL, Java2D

Expressiveness



Facilitate **rapid exploration** with **concise** specifications by omitting low-level details.

Infer **sensible defaults** and customize by overriding defaults.

With **native** support for ***interaction!***

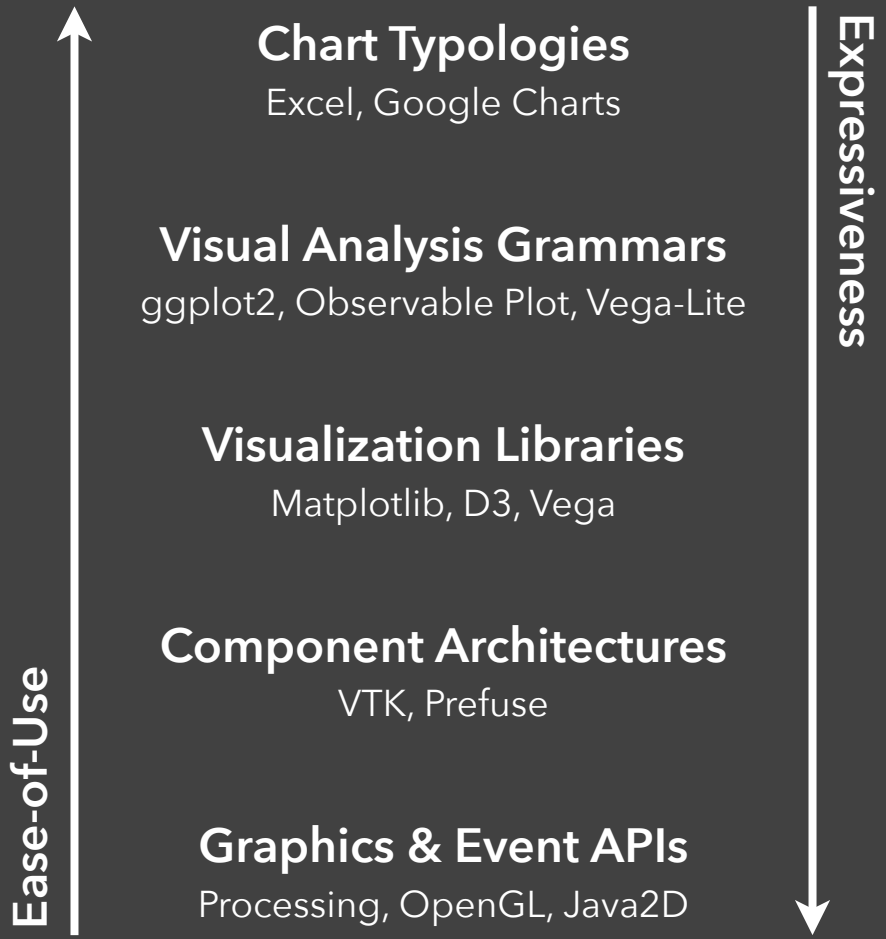


Chart Typologies
Excel, Google Charts
Visual Analysis Grammars
ggplot2, Observable Plot, Vega-Lite
Visualization Libraries
Matplotlib, D3, Vega
Component Architectures
VTK, Prefuse
Graphics & Event APIs
Processing, OpenGL, Java2D
Ease-of-Use
Expressiveness

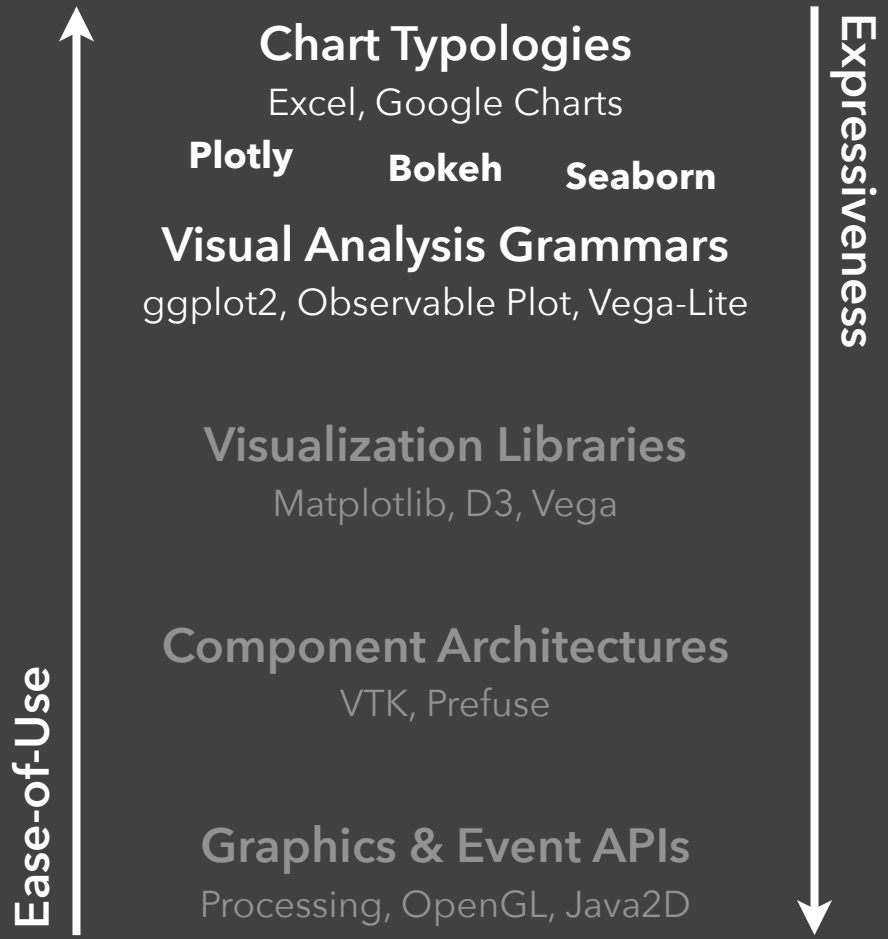
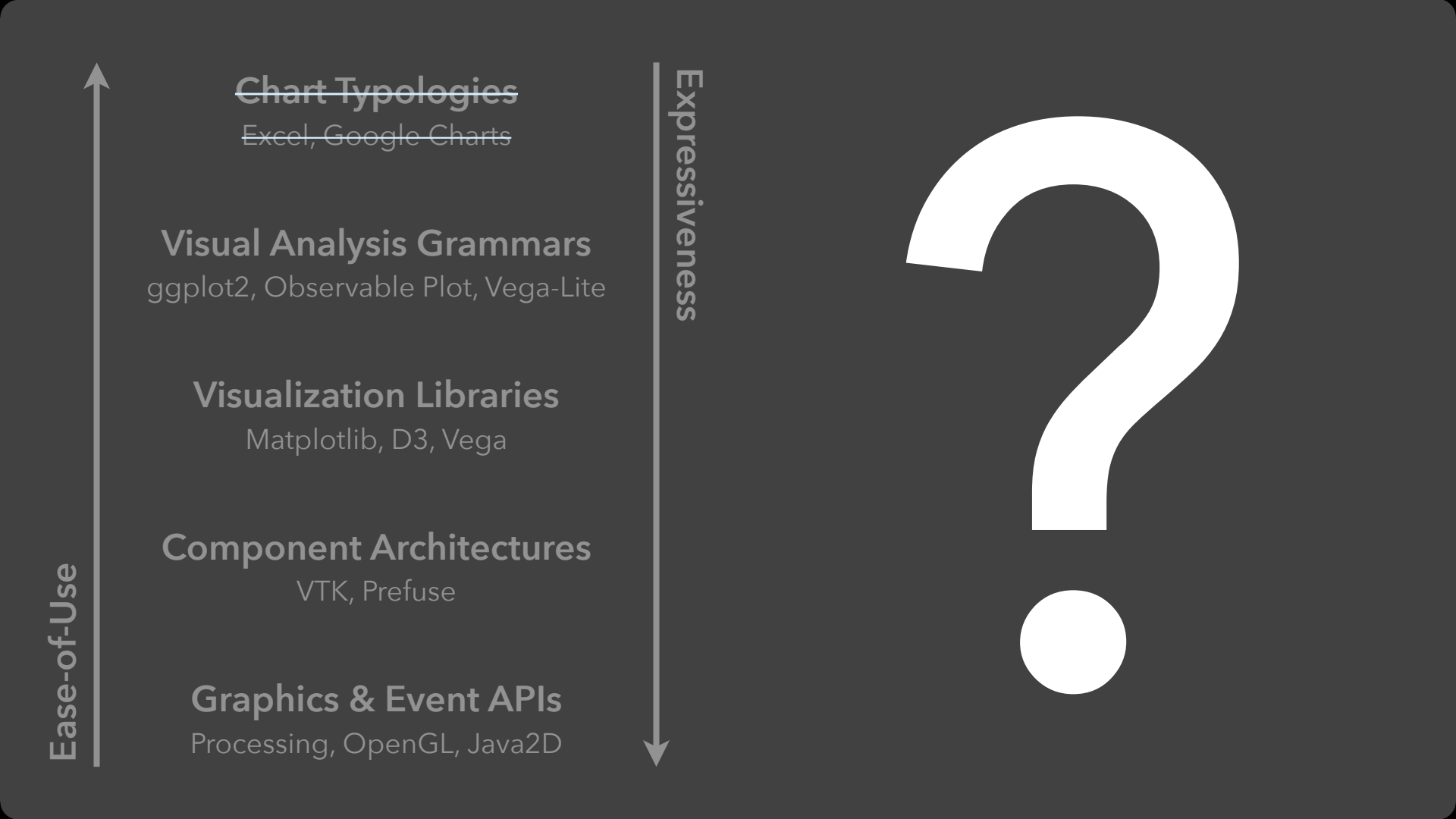


Chart Typologies
Excel, Google Charts
Plotly
Bokeh
Seaborn
Visual Analysis Grammars
ggplot2, Observable Plot, Vega-Lite
Visualization Libraries
Matplotlib, D3, Vega
Component Architectures
VTK, Prefuse
Graphics & Event APIs
Processing, OpenGL, Java2D
Ease-of-Use
Expressiveness


~~Chart Typologies~~
~~Excel, Google Charts~~
Visual Analysis Grammars
ggplot2, Observable Plot, Vega-Lite
Visualization Libraries
Matplotlib, D3, Vega
Component Architectures
VTK, Prefuse
Graphics & Event APIs
Processing, OpenGL, Java2D
Ease-of-Use
Expressiveness
?

Ease-of-Use

Expressiveness

Visualization Guidance

Tableau, Voyager, Lux, LLMs, ...

Visual Analysis Grammars

ggplot2, Observable Plot, Vega-Lite

Visualization Libraries

Matplotlib, D3, Vega

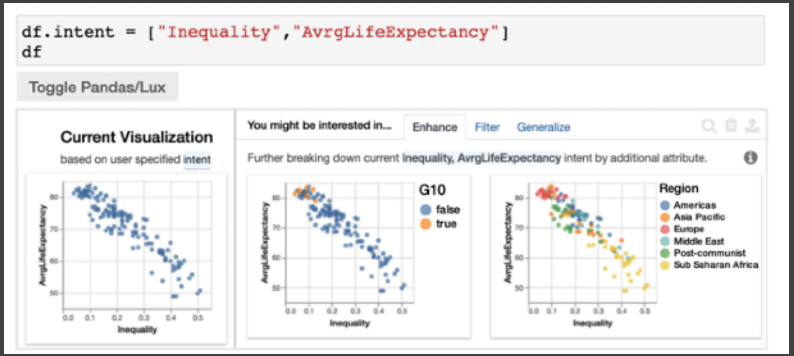
Component Architectures

VTK, Prefuse

Graphics & Event APIs

Processing, OpenGL, Java2D

Voyager, Wongsuphasawat et al. 2017



Lux, Lee et al. 2021

W1: Exercise & Assignment

W1: Hours of Sunshine

The climate of a place can have a tremendous impact on people's lived experience. You will examine average monthly climate measurements for six major U.S. cities, roughly covering the edges of the continental United States.

Our in-class ***exercise*** is to get hands-on experience creating and publishing visualizations. You will create and revise a line chart of average monthly sunshine hours for six U.S. cities.

The ***assignment*** is to then design your own graphic.

W1: Hours of Sunshine

Complete the exercises in the course website, as provided in your personal GitLab repository.

Submit your exercise results by committing and pushing them to your GitLab repo.

You may collaborate in groups of 1-3 people. Each person should update their own repo.

The course staff is here to help! Don’t hesitate to ask us questions spanning design, tech, or more.

W1: Expository Visualization

Using the given climate data set…

Pick a **guiding question**, use it to title your vis.

Design a **static visualization** for that question.

You are free to **use any web-based tool**, so long as you publish the result to your GitLab repo.

Deliverables via Gradescope

Image of your visualization (PNG or JPG format)

Short description + design rationale (≤ 4 paragraphs)

Due by **9:00 am, Tue January 14**.