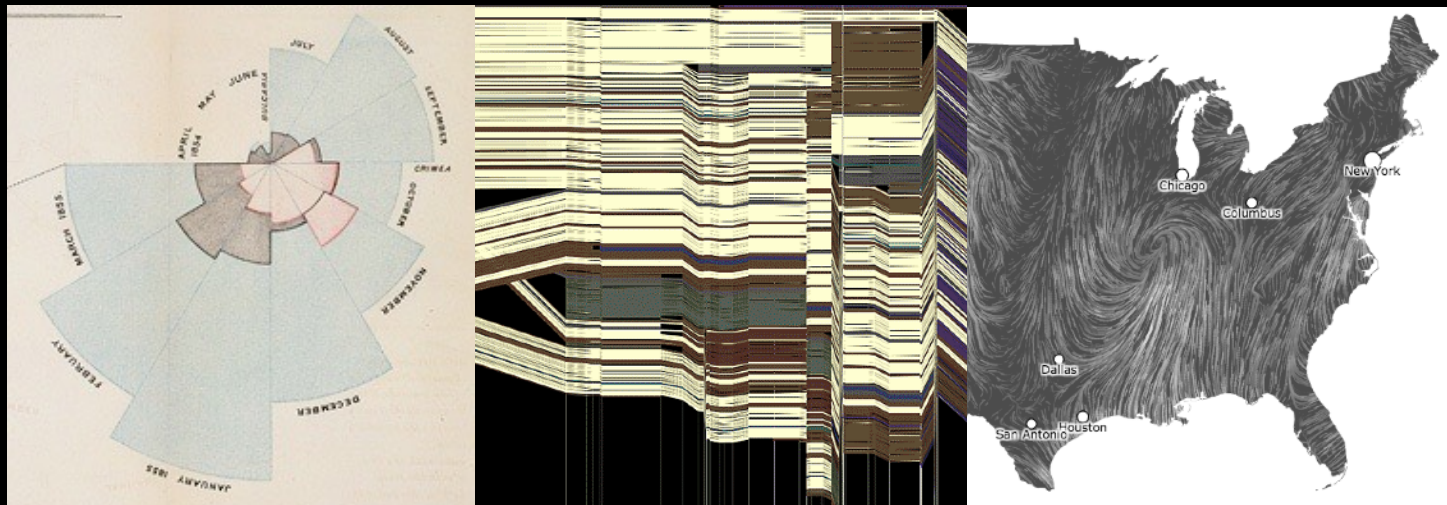
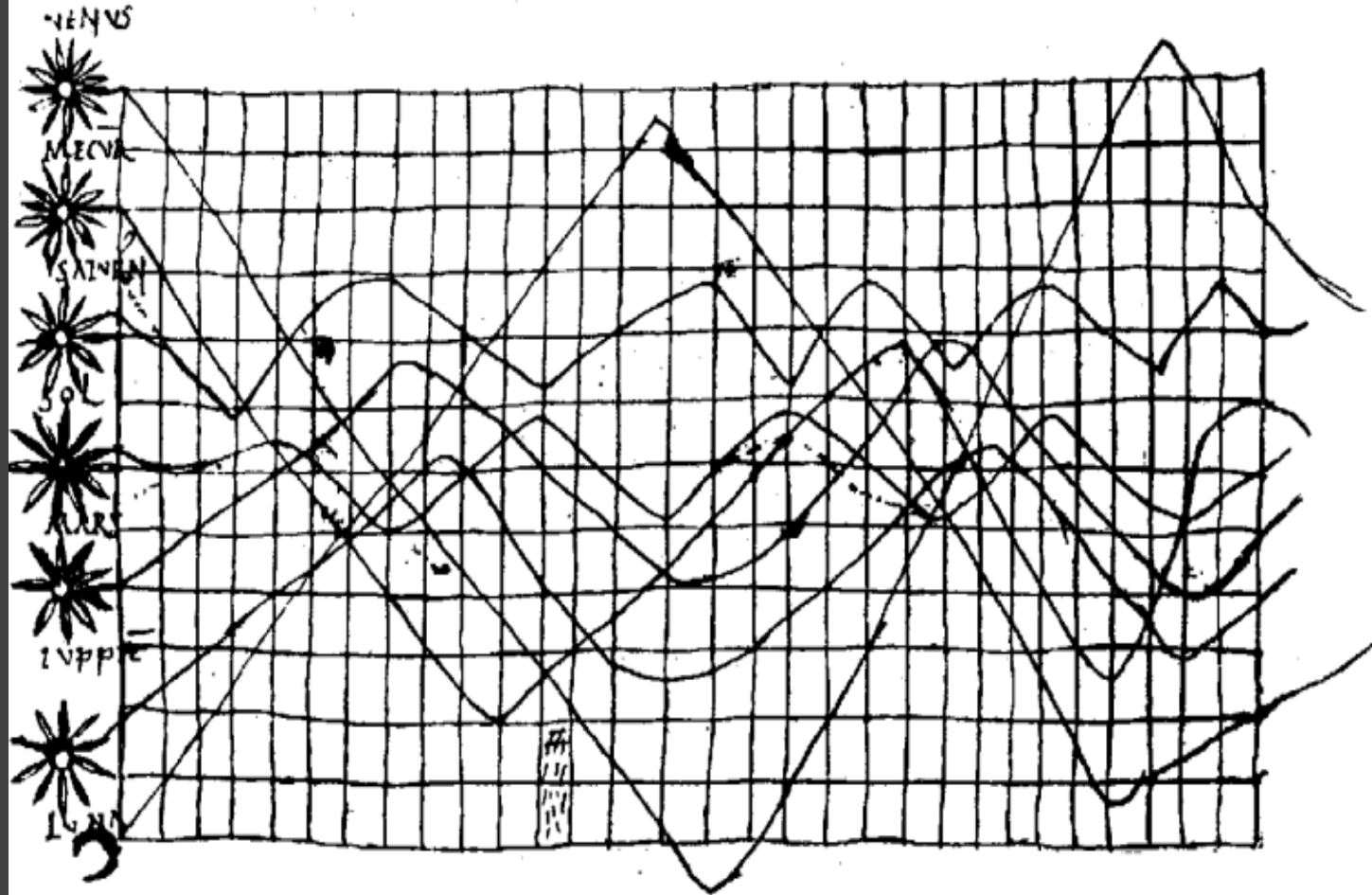


CSE 442 - Data Visualization

Introduction



Jeffrey Heer University of Washington



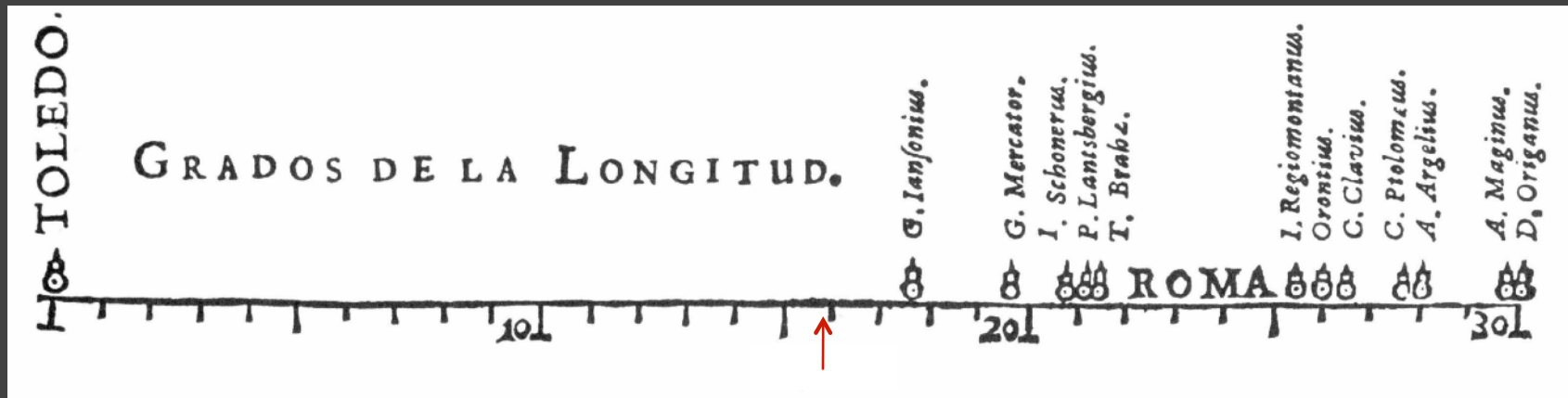
~950 AD Position of Sun, Moon and Planets

MACVLAE IN SOLE APPARENTES, OBSERVATAE
 anno 1611. ad latitudinem grad. 48. min. 40.



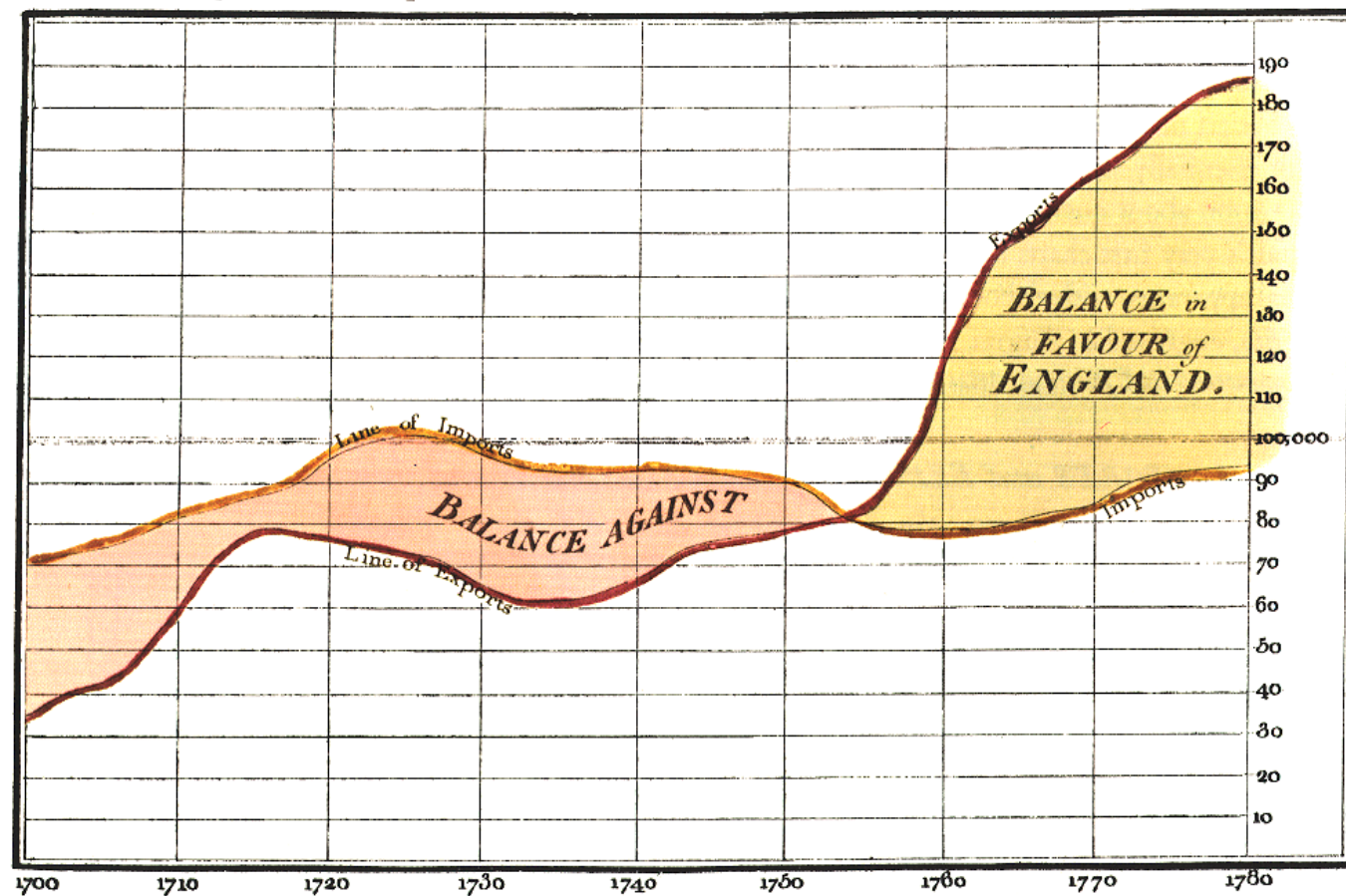
a c, horizon. a b c, arcus solis diurnus. Sol oriens ex parte a, maculas exhibet quas vides, occidens vero c, easdem ratione primj motus, nonnihil inuertit. Et hanc matutinam vespertinamq; mutationem, omnes macula quotidie subeunt. Quod semel exhibuisse et monuisse, sufficiat.





Longitudinal distance between Toledo and Rome, van Langren 1644

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

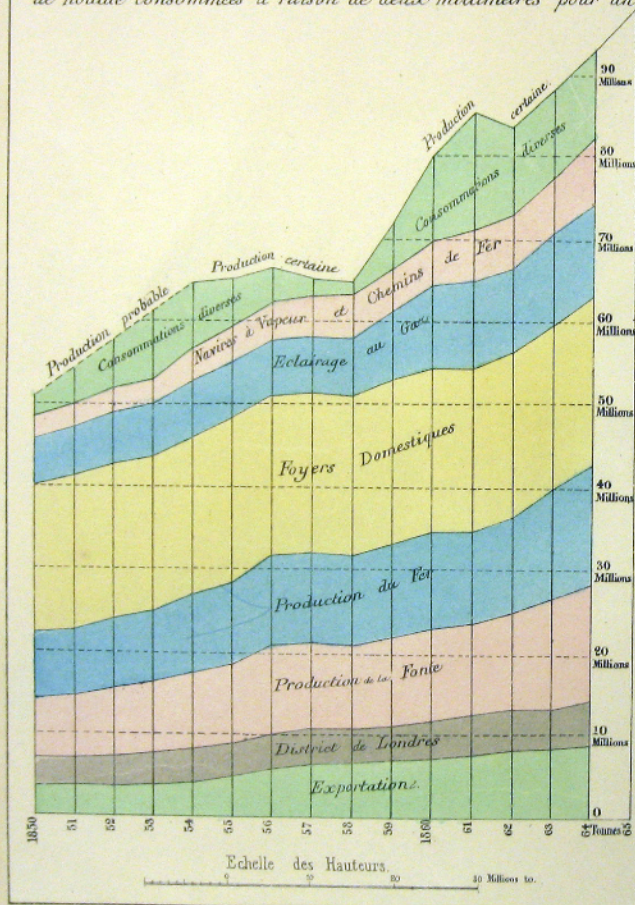


The Commercial and Political Atlas, William Playfair 1786

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^{es} de houille pour 1^{re} 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^{es} 35 de houille pour 1 tonne de fonte convertie en fer, et admettant $\frac{1}{10}$ de la fonte produite convertie 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 $\frac{1}{3}$ au $\frac{1}{2}$ de la production totale.

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

Navigation à vapeur. — Calculée à raison de 5^e 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.^r Lamé Fleury, Dictionnaire du Commerce Page III.

Negro business men in the United States.

Nègres Américains dans les affaires.

Done by Atlanta University.

Estimated capital
Capital évalué

\$ 8,784,637
45,516,254 FRANCS.

General merchandise stores
Magazins de provisions et
d'objets divers

Grocers
Épiciers

Bankers
Banquiers

Undertakers
Entrepreneurs de pompes
funèbres

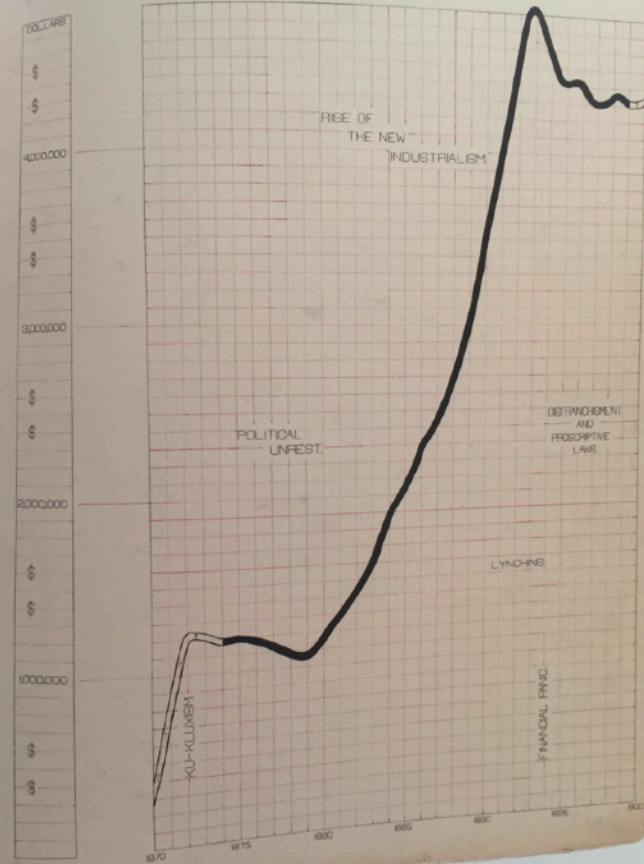
Building contractors
Entrepreneurs de bâtiments

Druggists
Pharmaciens

Publishers
Éditeurs

Building and loan associations
Institutions financières co-opératives

VALUATION OF TOWN AND CITY PROPERTY OWNED BY GEORGIA NEGROES.



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**

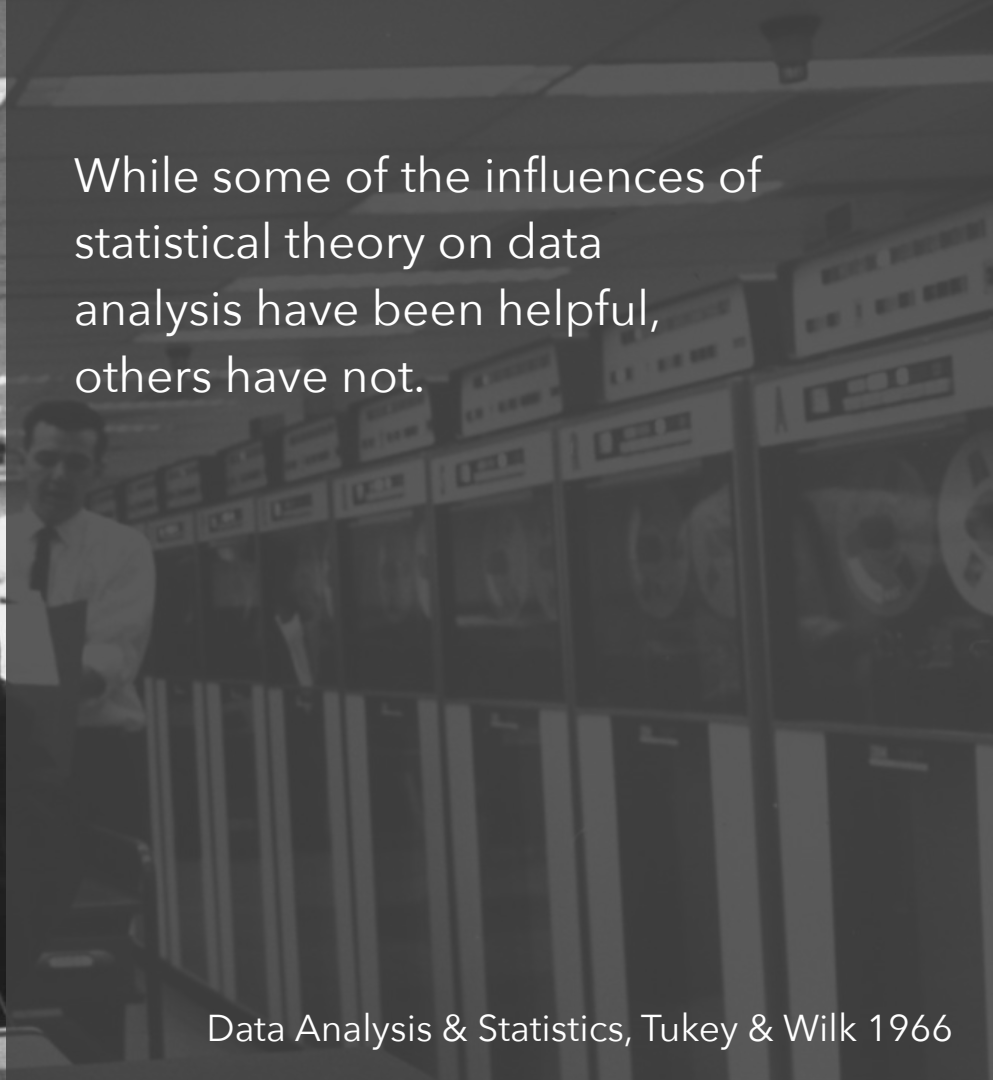


Four major influences act on data analysis today:

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.



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.



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.**

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

$$u_X = 9.0 \quad \sigma_X = 3.317$$

$$u_Y = 7.5 \quad \sigma_Y = 2.03$$

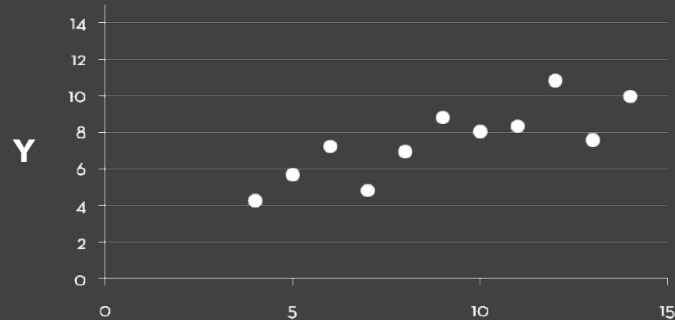
Linear Regression

$$Y = 3 + 0.5 X$$

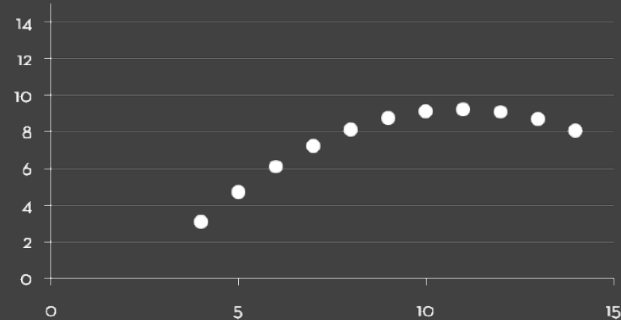
$$R^2 = 0.67$$

[Anscombe 1973]

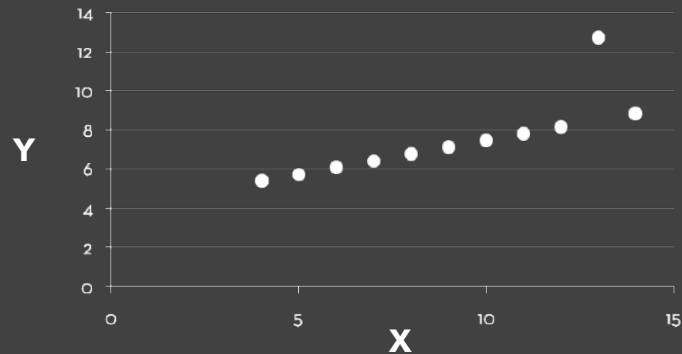
Set A



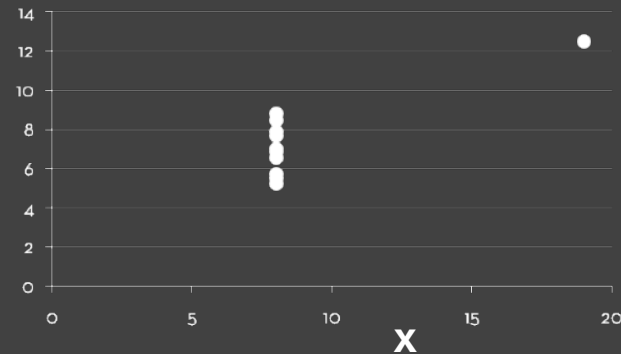
Set B



Set C



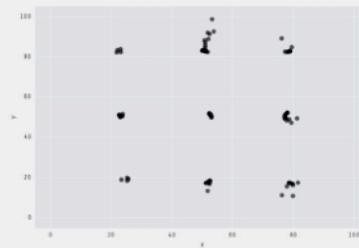
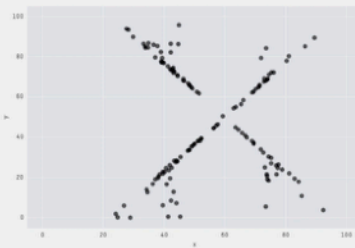
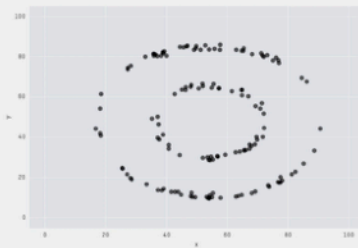
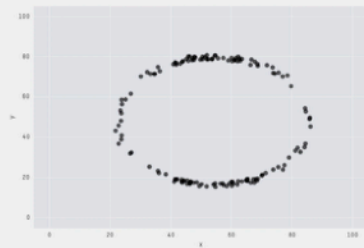
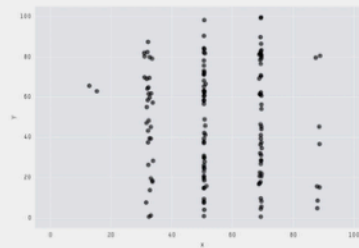
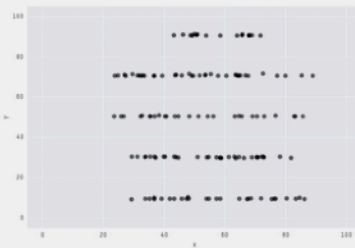
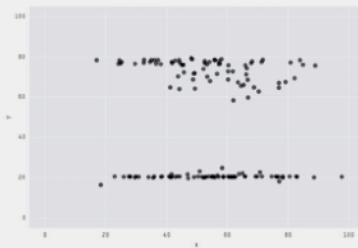
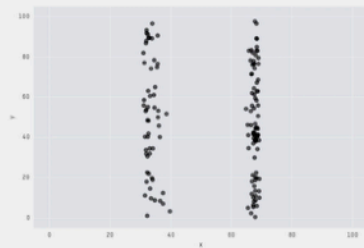
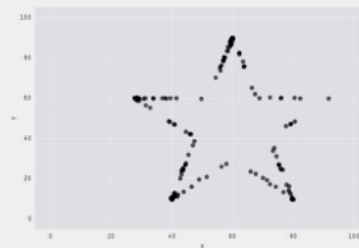
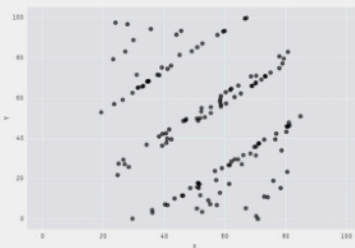
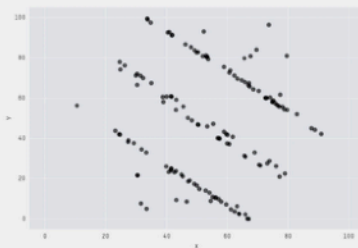
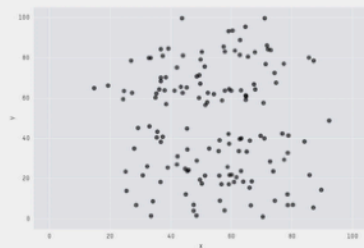
Set D





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

[Matejka 2017]



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]

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

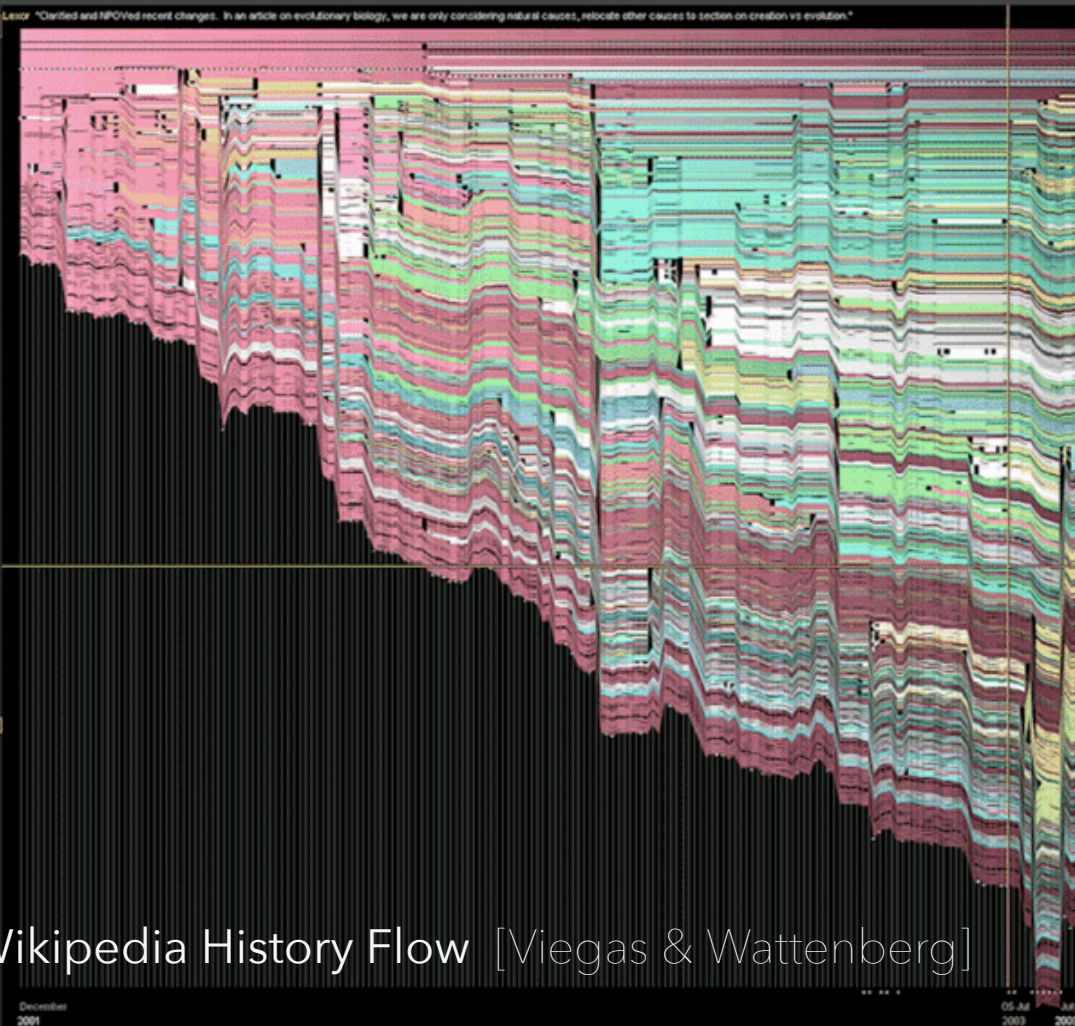
Improve ties between visualization & mental model

Example: Wikipedia Edits

authors posts

Donnell	3
Hannes Hirtel	3
Ed Hour	4
Tier	4
Susanne Brasser	1
Paul Dye	1
Alex Bick	4
Conversion script	1
Rigable	2
Strutenstein	12
Bryan Derksen	1
Maverick49	5
Nicki Rosenzweig	1
Josh Grison	1
Robert Merkel	1
PierreAbdel	1
FredBauder	1
Tjess	1
Gog	4
Emmett	3
Q	1
Jehouse	5
Canemart	1
Oriff	3
Deweyers	1
Adewolfechless	17
Chaz ZZZ brown	2
Heron	1
Ryguess	1
The Anome	1
Alan Peakal	1
Serebichu	1
Boquence	7
Helmut	1
Cyle	1
Ric	3
Fred Bauder	1
ArSchmidt	1
Michael	1
Flawery	1
Zoe	1
Levor	13
Someone else	1
Tanen	1
Sunder	1
Julienne	1
Hirose	1
Corte Alghieri	1

Levor "Clarified and NPOVed recent changes. In an article on evolutionary biology, we are only considering natural causes, relocate other causes to section on creation vs evolution."



Therefore, over time, the types of organisms that have traits better **adapted** to their environment will tend to become the **dominant** ones in an environment, while organisms poorly adapted to their environment will become **extinct**.

Natural selection also provides for a **mechanism** by which life can sustain itself over time. Since, in the long run, environments always change, if successive generations did not develop adaptations which allowed them to survive and reproduce, species would simply die out as their biological niches die out. Therefore, life is allowed to persist over great spans of time, in the form of evolving species. The central role of natural selection in evolutionary theory has created a strong connection between that field and the study of **ecology**.

Genetic drift

Genetic drift describes changes in gene frequency that cannot be ascribed to selective pressures, but are due instead to events that are unrelated to inherited traits. This is especially important in small mating populations, which simply cannot have enough offspring to maintain the same gene distribution as the parental generation. Such fluctuations in gene frequency between successive generations may result in some genes disappearing from the population. Two separate populations that begin with the same gene frequency might, therefore, "drift" by random fluctuation into two divergent populations with different gene sets (i.e. genes that are present in one have been lost in the other). Rare sporadic events (volcanic explosion, meteor impact, etc.) might contribute to genetic drift by altering the gene frequency outside of "normal" selective pressures.

Development of evolutionary theories

As science has uncovered more and more information about the basic operations of life, such as genetics and molecular biology, theories of evolution have changed. The general trend has been not to overturn well-supported theories, but to supplant them with more detailed and therefore more complex ones.

While transmutation was accepted by a sizeable number of scientists before 1859, it was the publication of Charles Darwin's *The Origin of Species* which provided the first cogent mechanism by which evolutionary change could persist, his mechanism of **natural selection**. The evolutionary **timeline** outlines the major steps of evolution on Earth as expounded by this theory's proponents.

Following the dawn of **molecular biology**, it became clear that a major mechanism for variation within a population is the **misapposness** of **DNA**. An essential component to evolutionary theory is that during the **cell cycle**, DNA is copied fairly, but not entirely, faithfully. When these rare copying errors occur, they are said to introduce genetic mutations of three general consequences relative to the current environment: good, bad, or neutral. By definition, individuals with "good" mutations will have an a stronger propensity to propagate, individuals with "bad" mutations will have less of a chance at successful reproduction, and those carrying "neutral" mutations will have neither an advantage nor a disadvantage. These definitions assume that the environment remains stable. Considered at the level of a single gene, these variations just described represent different genetic **alleles**. Following environmental change, alleles may retain their classification of good, bad, or neutral, or may shift into one of the other categories. Individuals carrying alleles formerly classified as neutral may now be "cousins" to those having formerly neutral

Wikipedia History Flow [Viegas & Wattenberg]

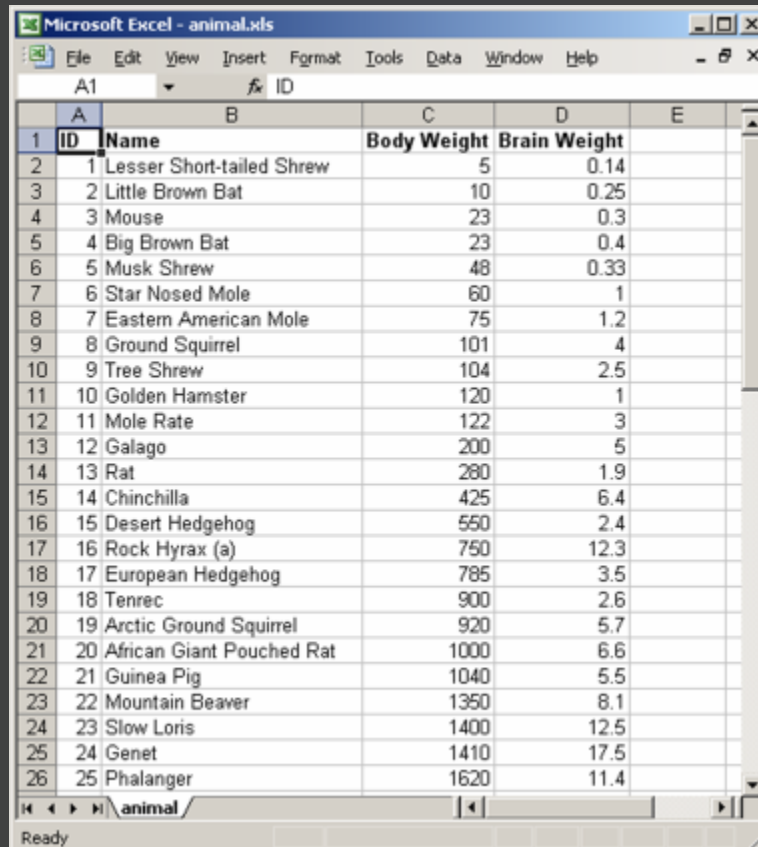


Edit War...

Wikipedia History Flow [Miegas & Wattenberg]

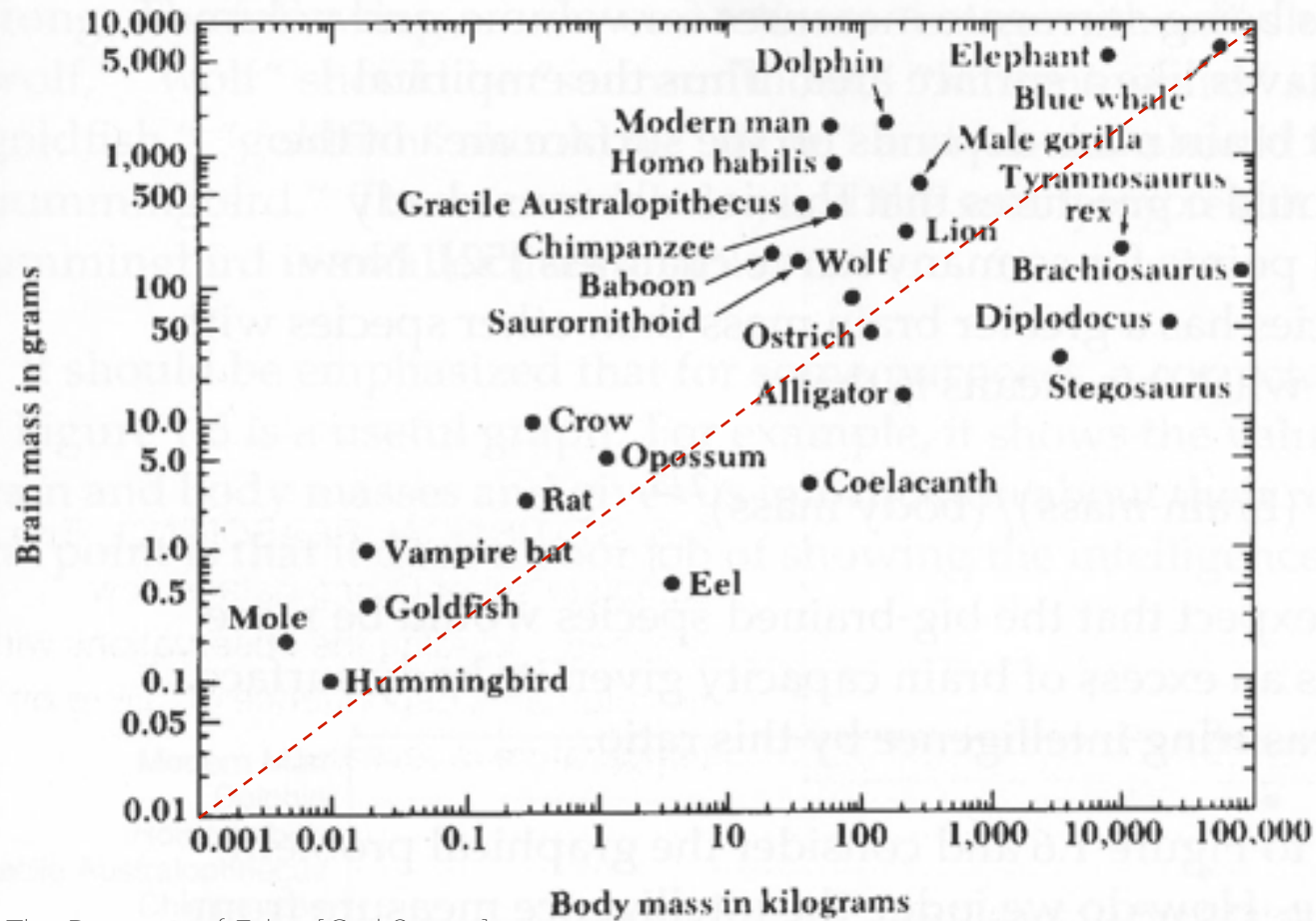
Example: Animal Brains

Which animals are the “smartest”?

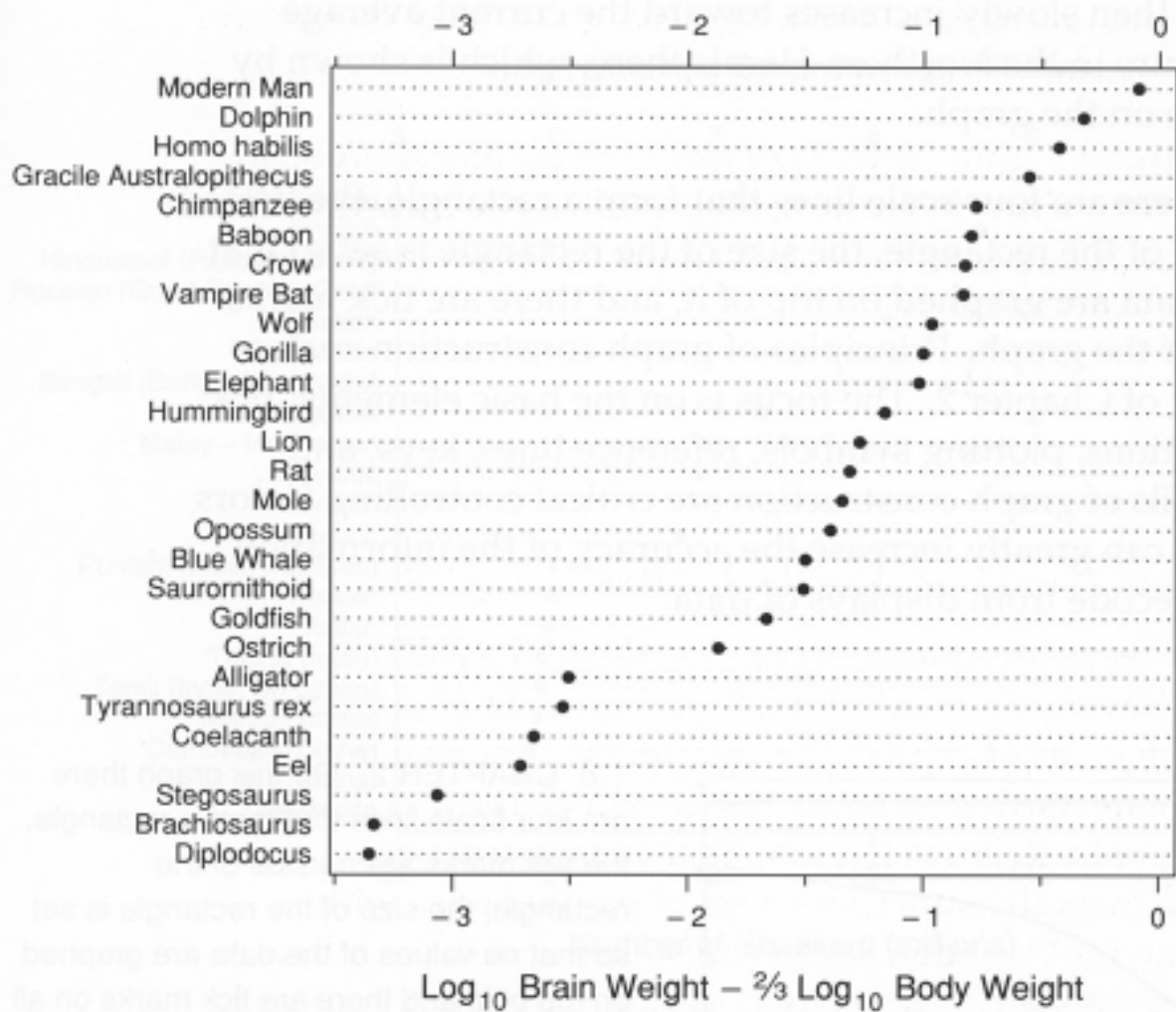


A screenshot of a Microsoft Excel spreadsheet titled "animal.xls". The spreadsheet contains a table with four columns: ID, Name, Body Weight, and Brain Weight. The data is organized into 26 rows, each representing a different animal species. The "ID" column contains sequential numbers from 1 to 25. The "Name" column lists the species names. The "Body Weight" column shows numerical values, and the "Brain Weight" column shows decimal values. The status bar at the bottom indicates the file is "Ready".

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

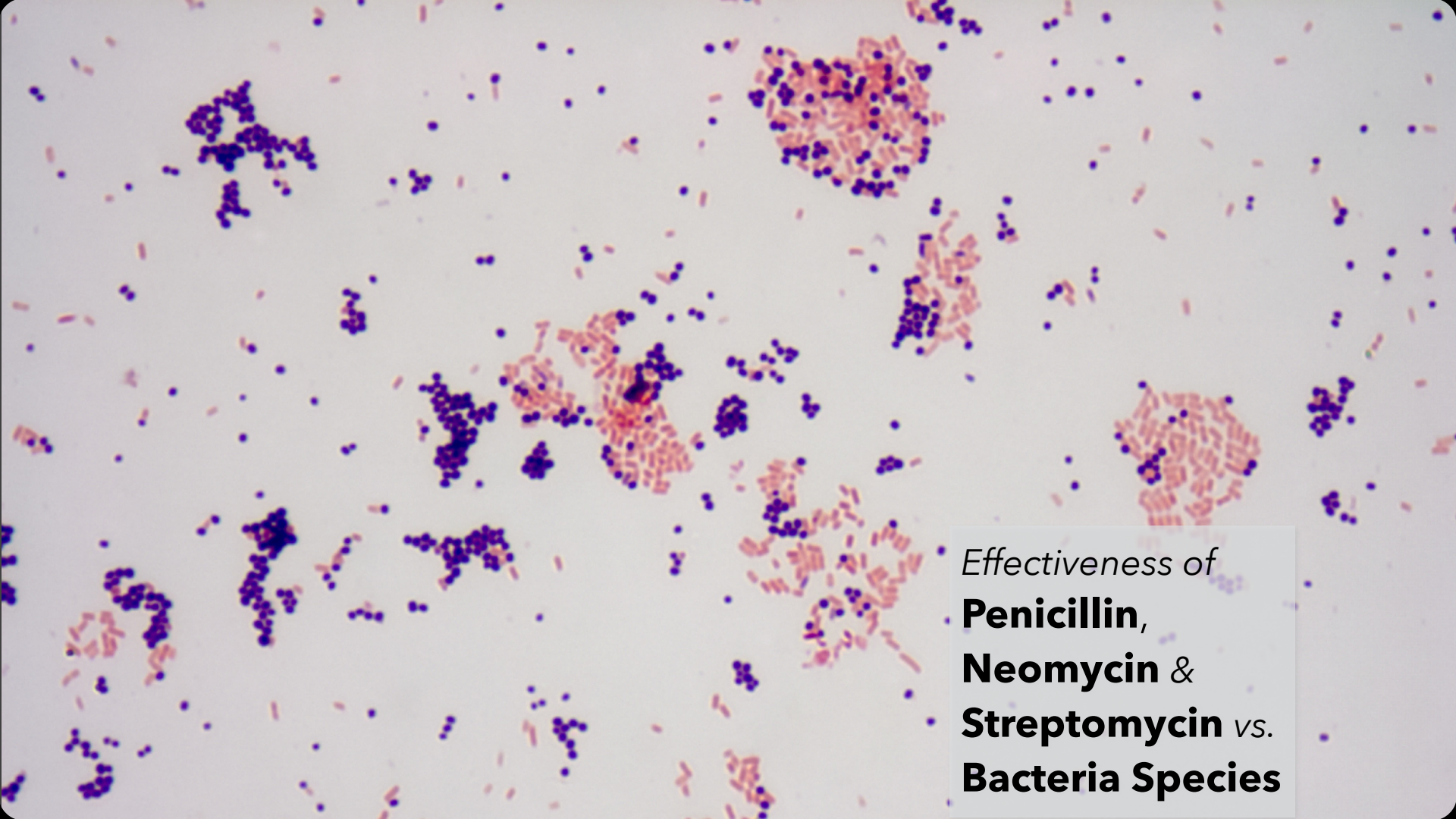


The Dragons of Eden [Carl Sagan]



Example:

Antibiotic Effectiveness

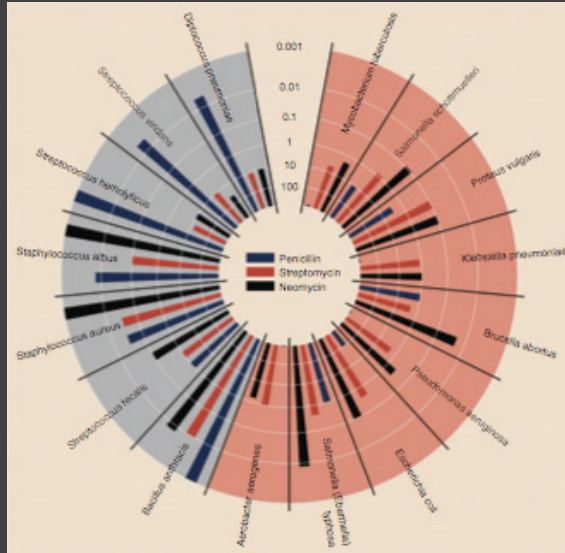


Effectiveness of
**Penicillin,
Neomycin &
Streptomycin vs.
Bacteria Species**

What questions might we ask?

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

Which antibiotic is most effective?



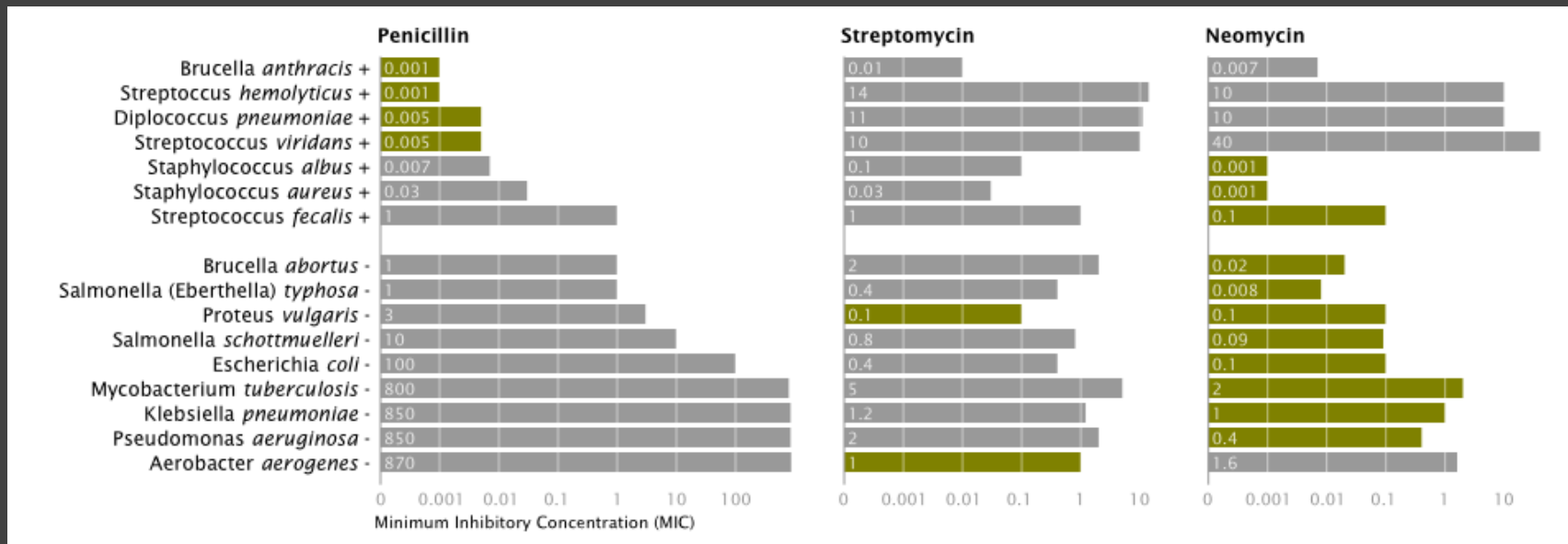
Bacteria	Penicillin	Antibiotic Streptomycin	Neomycin	Gram stain
<i>Aerobacter aerogenes</i>	870	1	1.6	–
<i>Brucella abortus</i>	1	2	0.02	–
<i>Bacillus anthracis</i>	0.001	0.01	0.007	+
<i>Diplococcus pneumoniae</i>	0.005	11	10	+
<i>Escherichia coli</i>	100	0.4	0.1	–
<i>Klebsiella pneumoniae</i>	850	1.2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

Radius: $1 / \log(\text{MIC})$

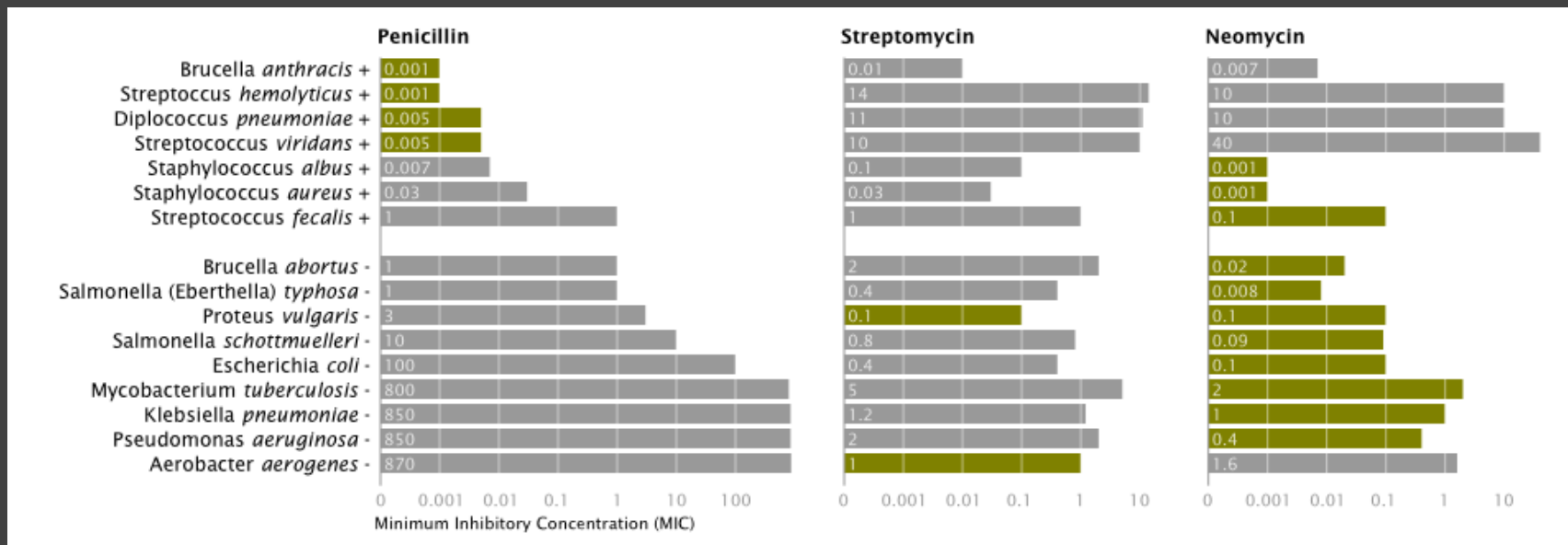
Bar Color: Antibiotic

Background Color: Gram Staining

Which antibiotic is most effective?



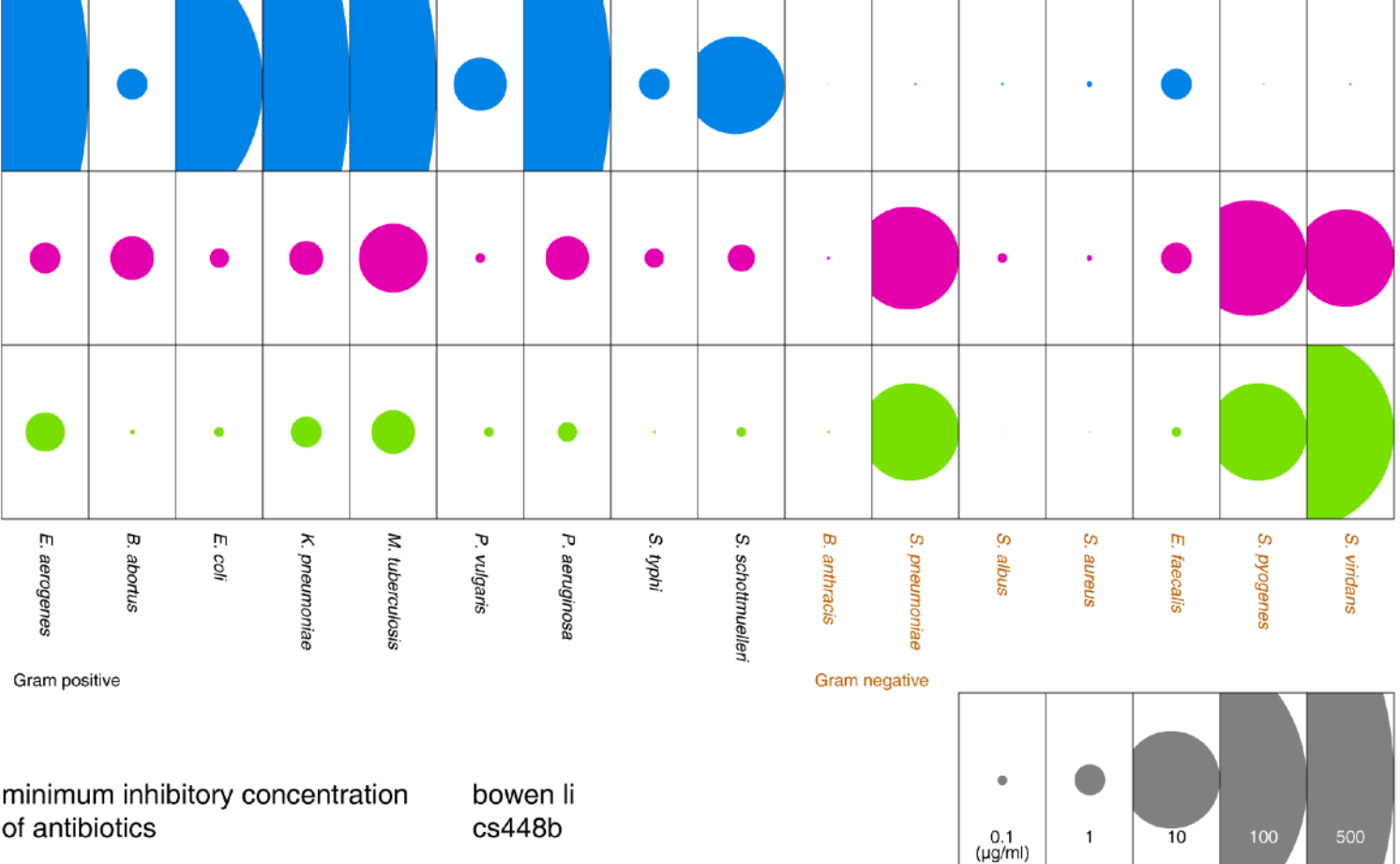
Which antibiotic is most effective?



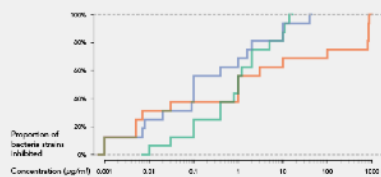
X-axis: Antibiotic | $\log(\text{MIC})$

Y-axis: Gram-Staining | Species

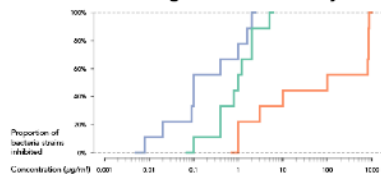
Color: Most-Effective?



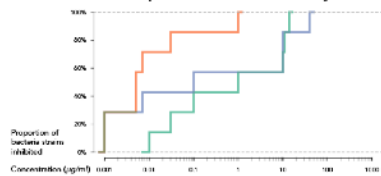
All bacteria



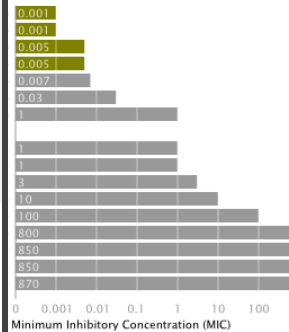
Gram-negative bacteria only



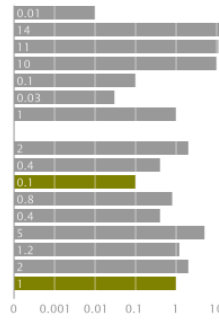
Gram-positive bacteria only



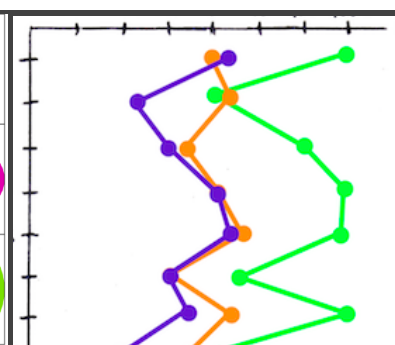
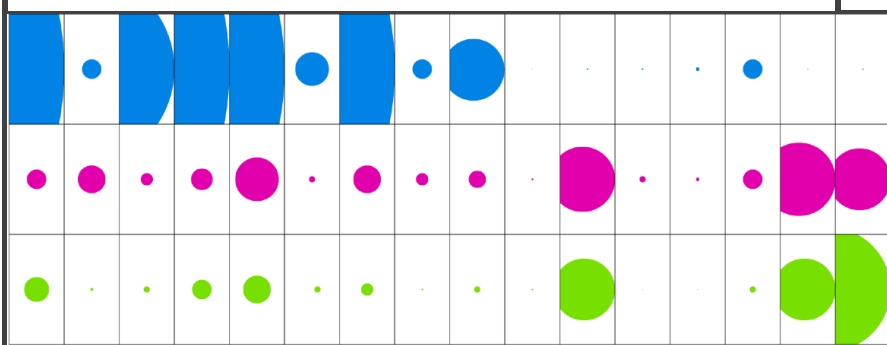
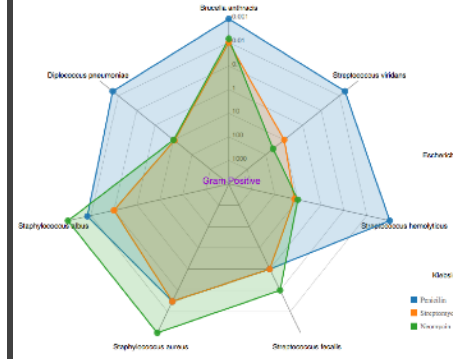
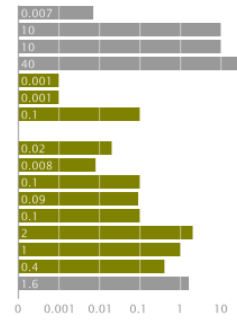
Penicillin



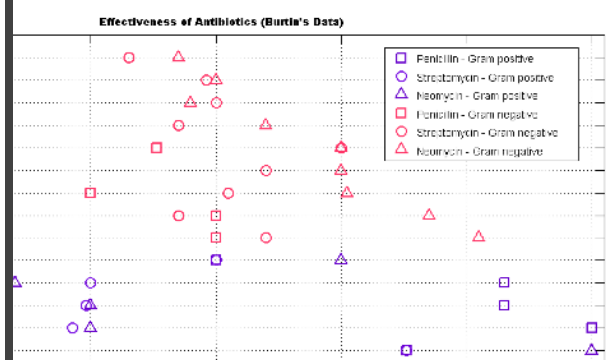
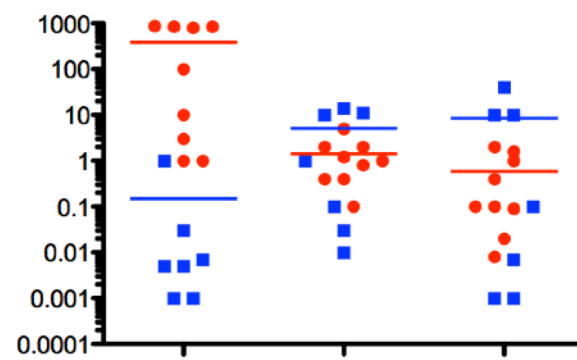
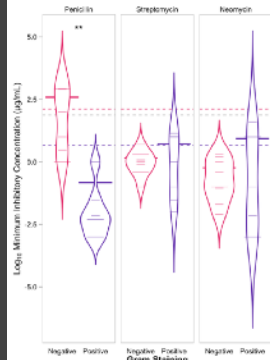
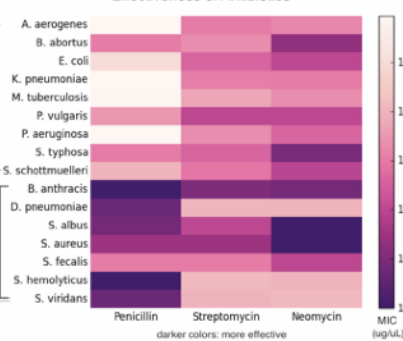
Streptomycin

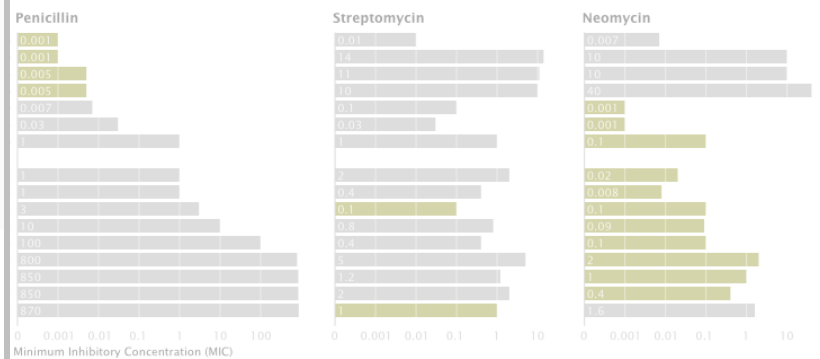
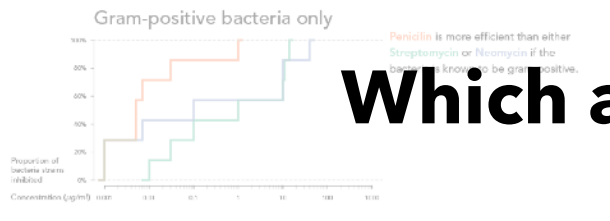
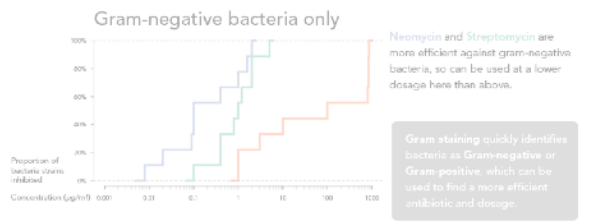
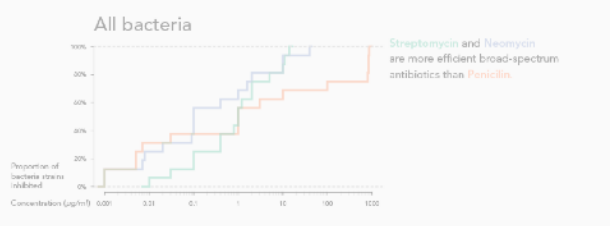


Neomycin

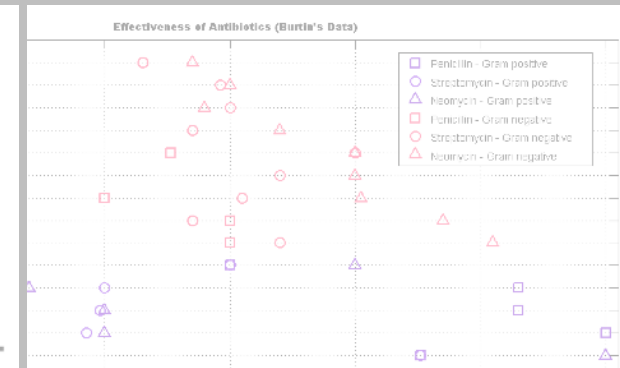
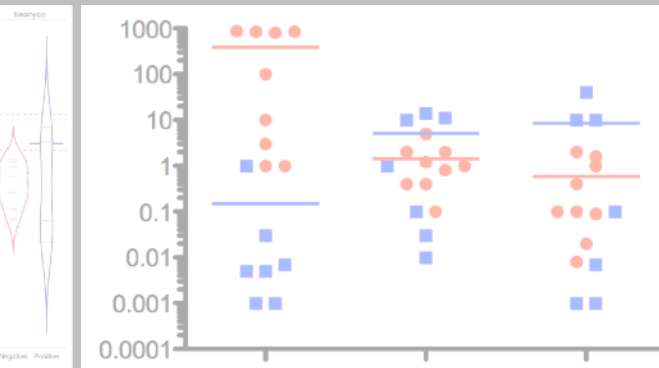
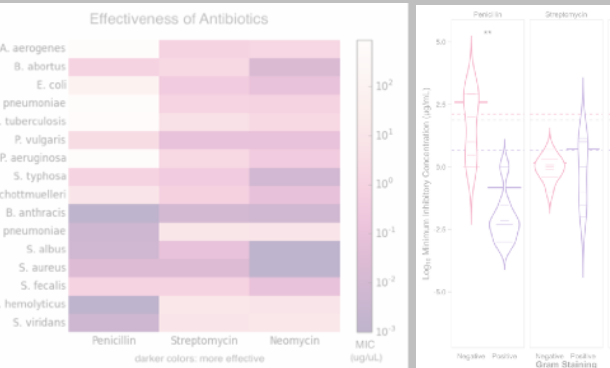


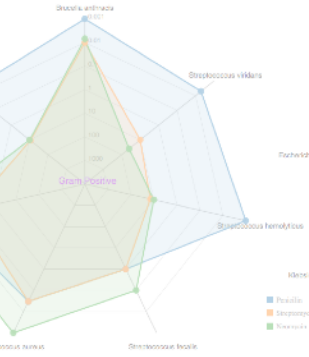
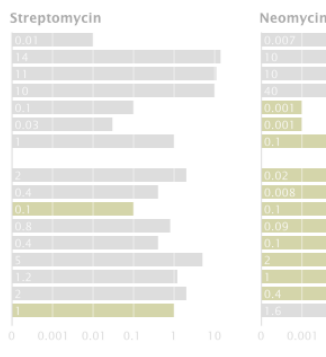
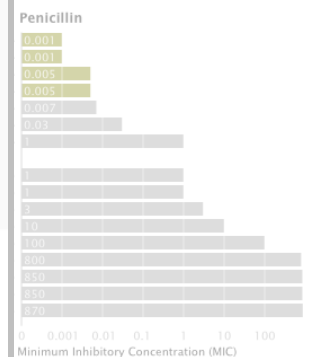
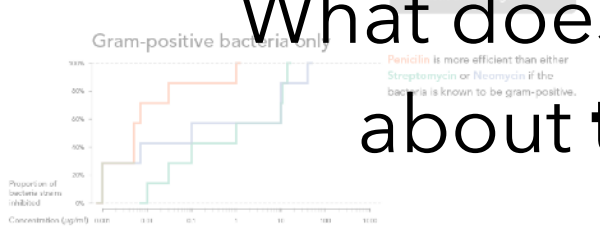
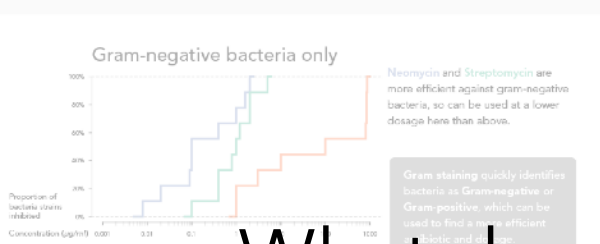
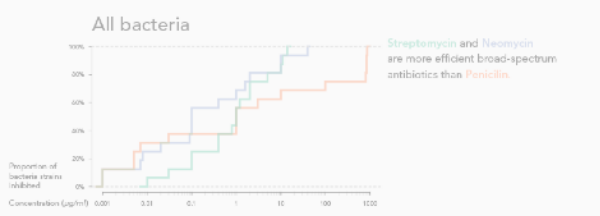
Effectiveness of Antibiotics



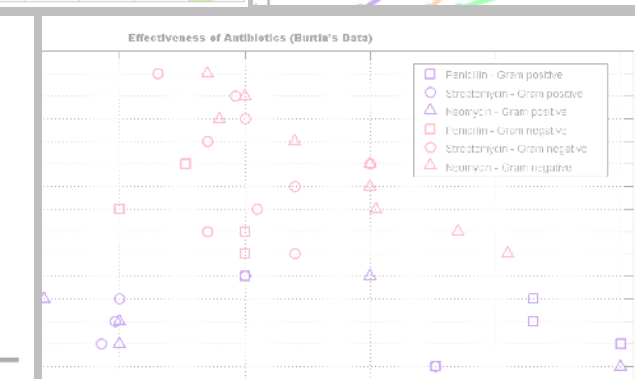
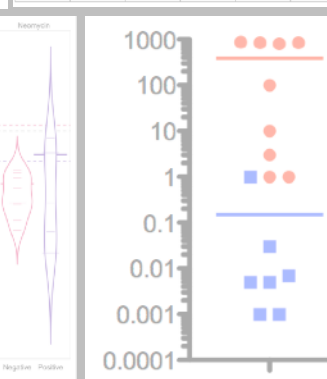
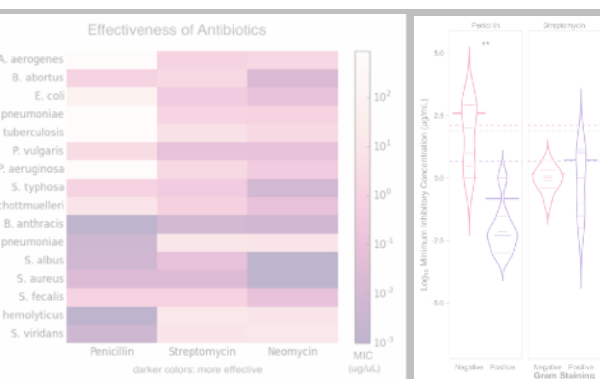


Which antibiotic should one use?

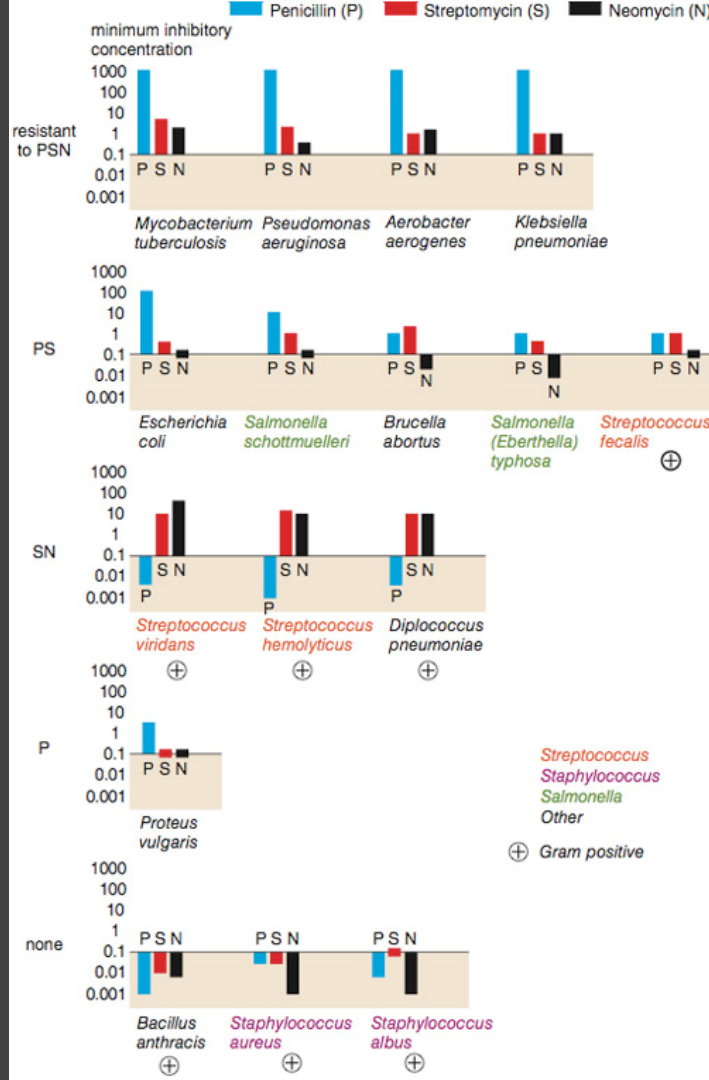




What does antibiotic response reveal about the biology of bacteria?

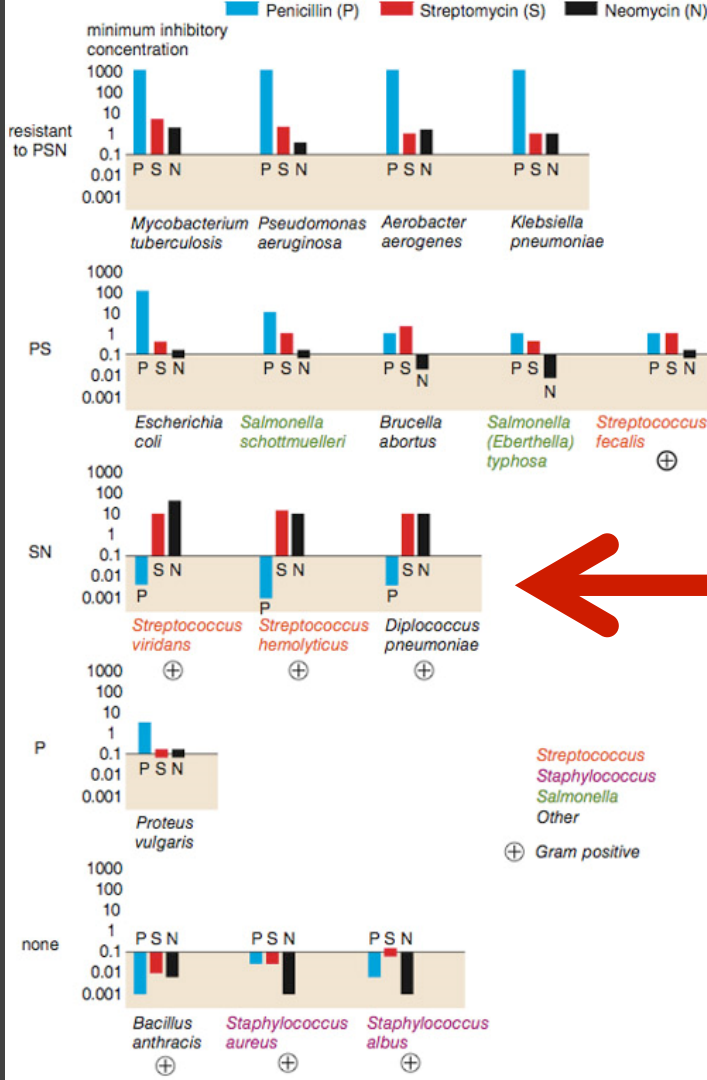


**Do the bacteria
group by antibiotic
resistance?**

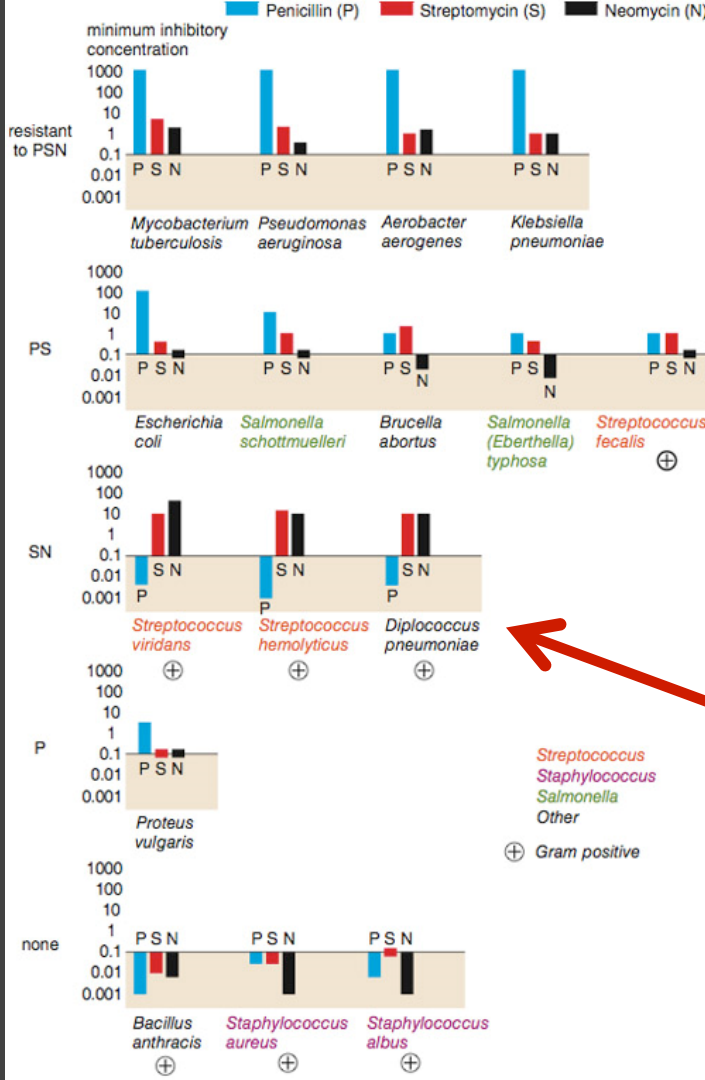


Do the bacteria group by antibiotic resistance?

Wainer & Lysen
American Scientist, 2009



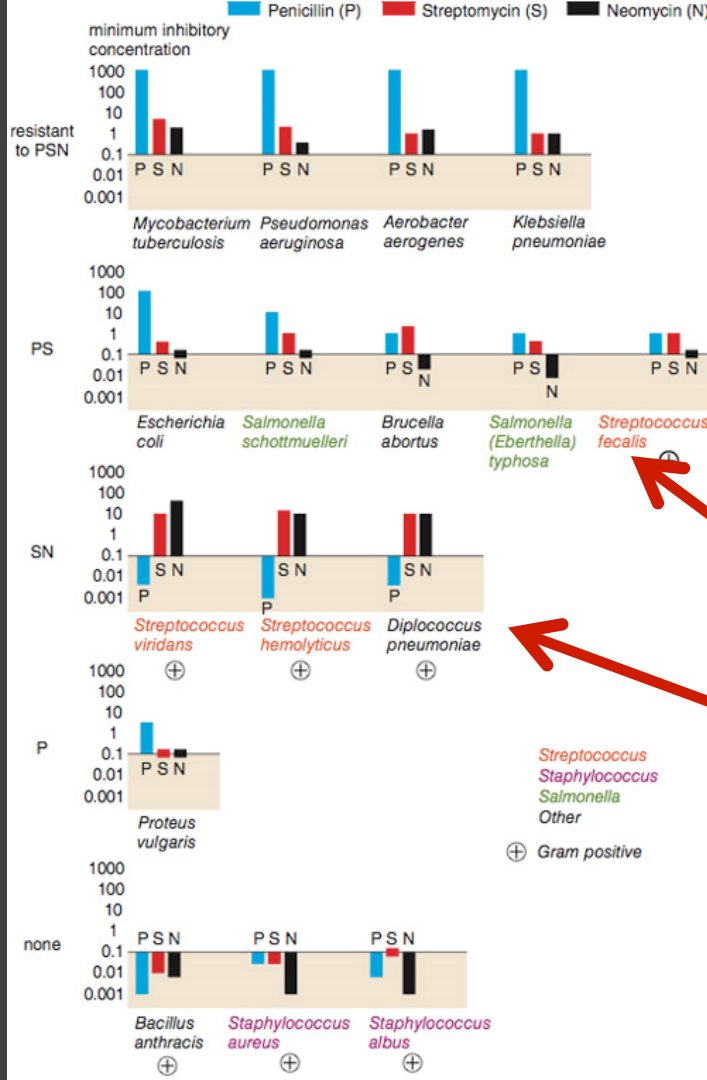
Do the bacteria
group by antibiotic
resistance?



Do the bacteria group by antibiotic resistance?

Really a streptococcus!
(realized ~20 yrs later)

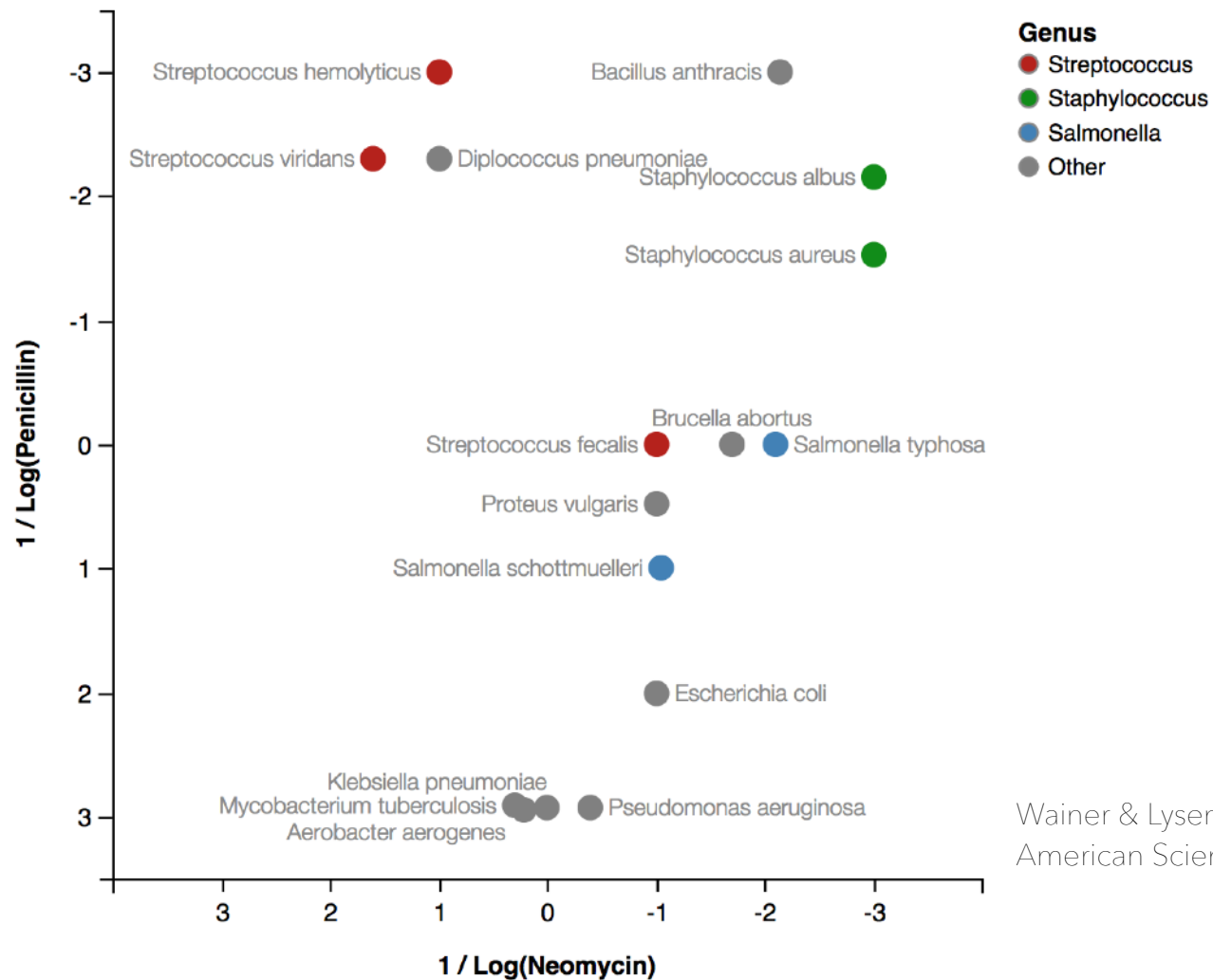
Wainer & Lysen
American Scientist, 2009



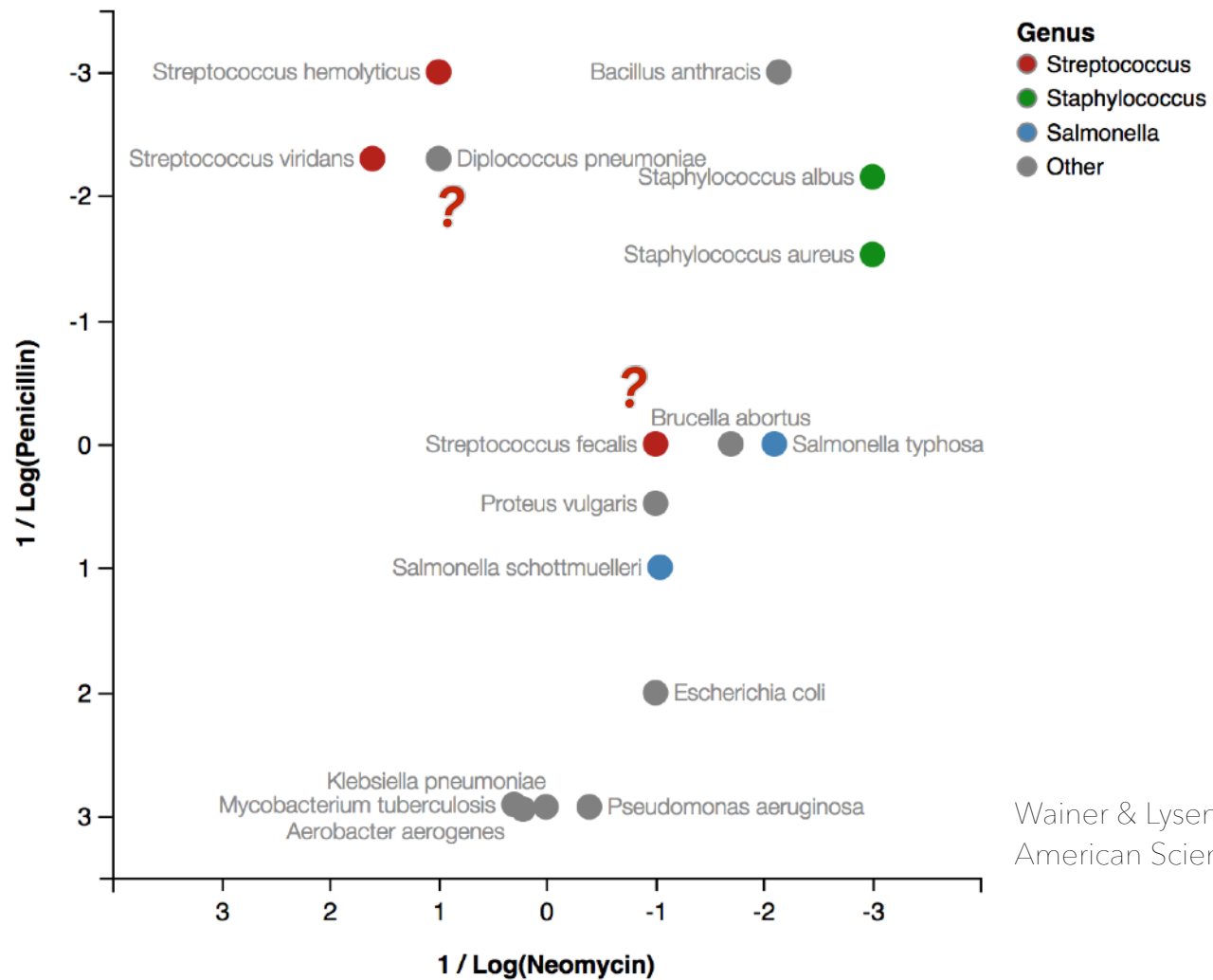
Do the bacteria group by antibiotic resistance?

Not a streptococcus!
(realized ~30 yrs later)

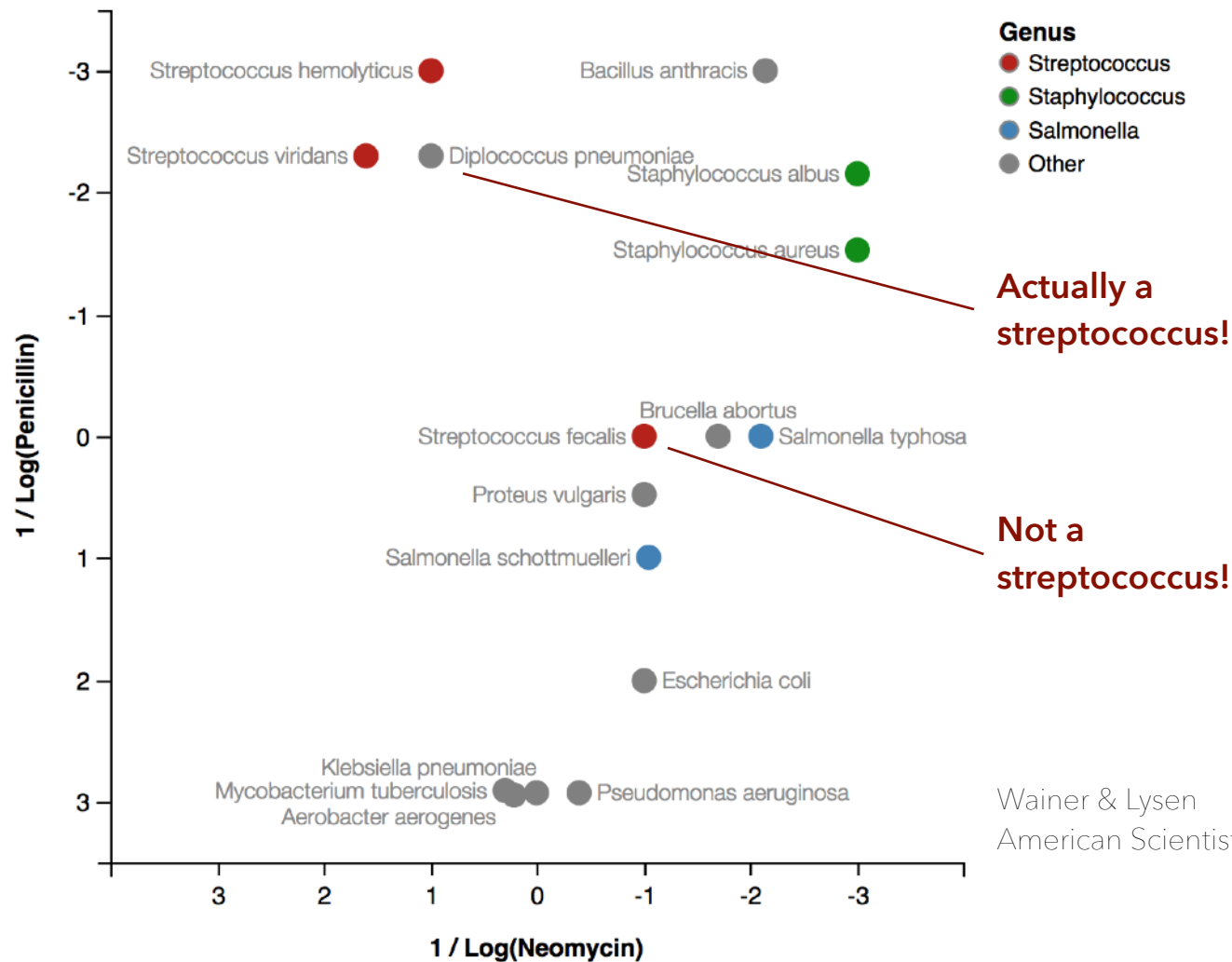
Really a streptococcus!
(realized ~20 yrs later)



Wainer & Lysen
American Scientist, 2009



Wainer & Lysen
American Scientist, 2009



Wainer & Lysen
American Scientist, 2009

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

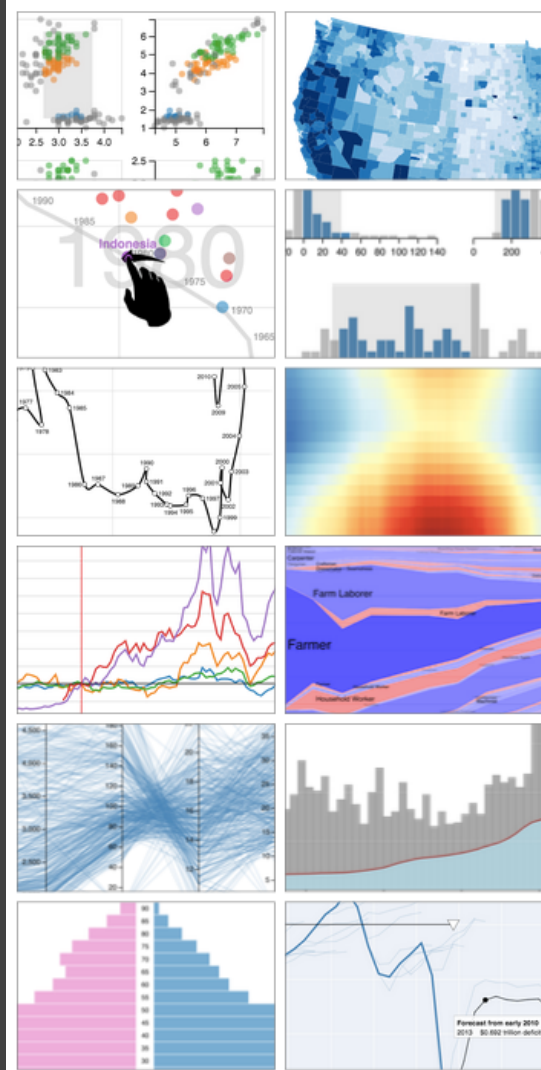
Don't trust your data!

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?



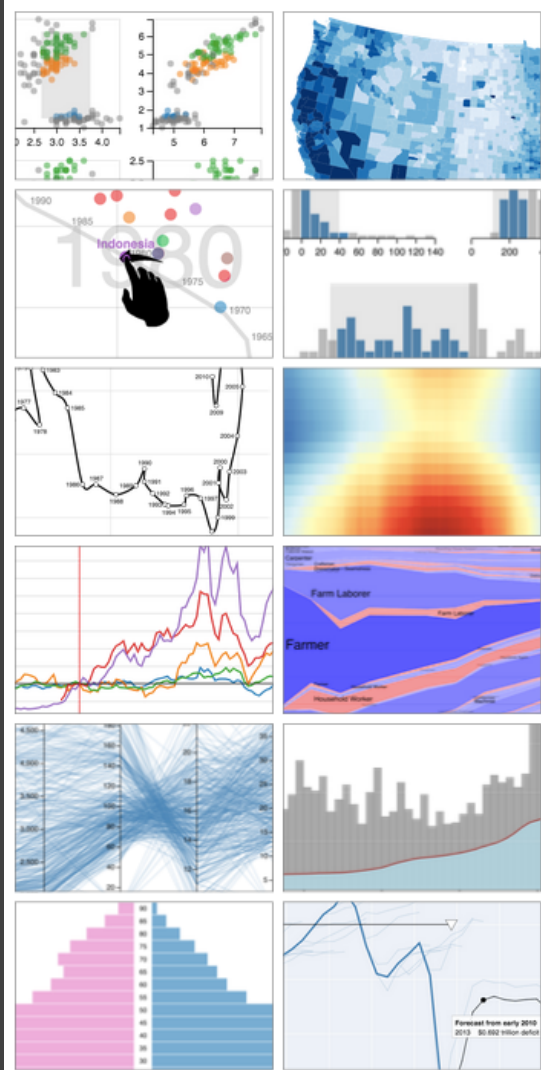
Data Exploration Tasks

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

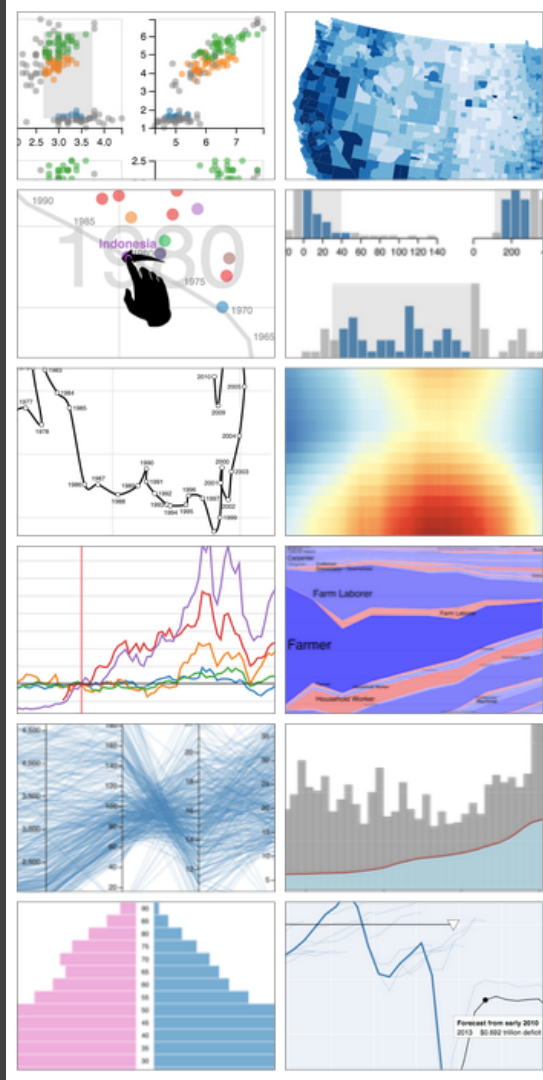
GOAL: Is the data actionable? What can we ask?

Search: identify specific data points or relations of interest to form an evidentiary chain

GOAL: Fact-finding, isolate important points/connections

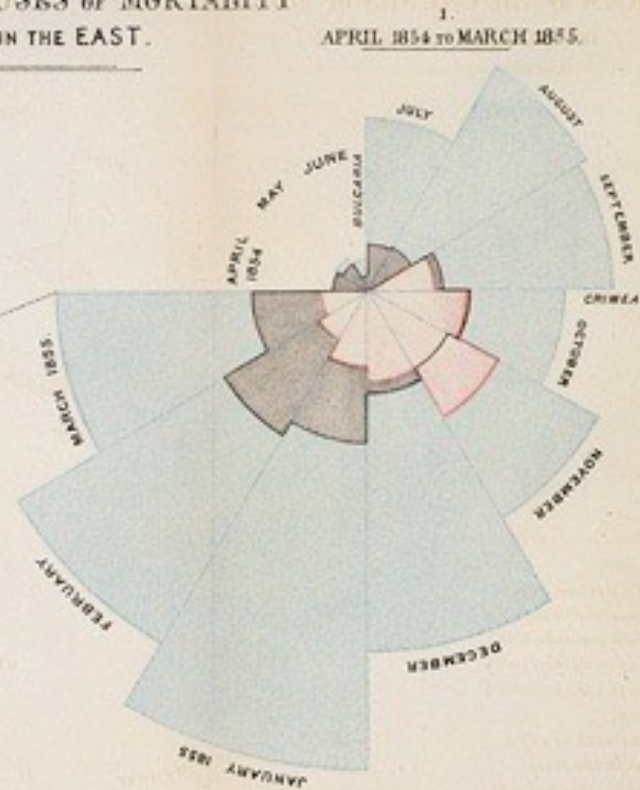
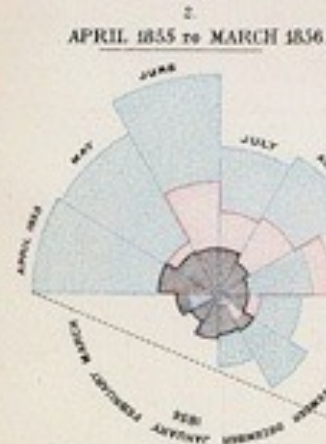
Infer: generalize from observed patterns, ascribe observations to specific factors or causes

GOAL: Inform modeling and decision making



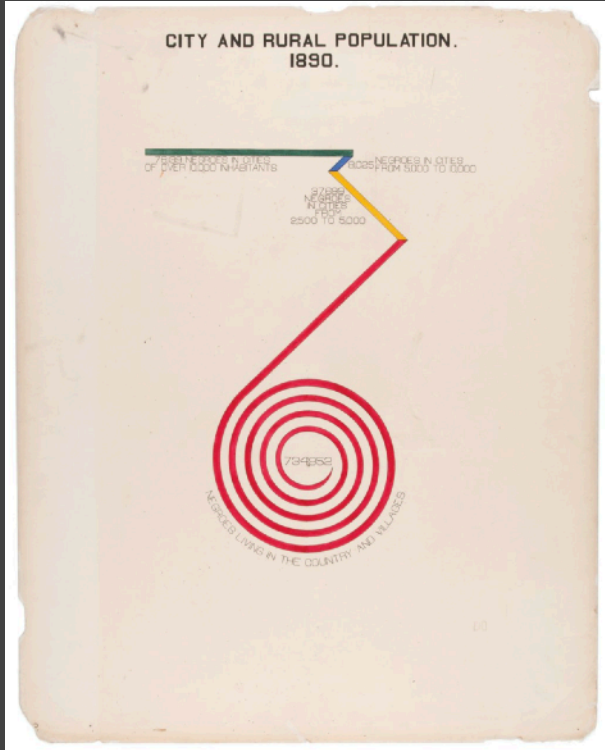
Communication Tasks

DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.

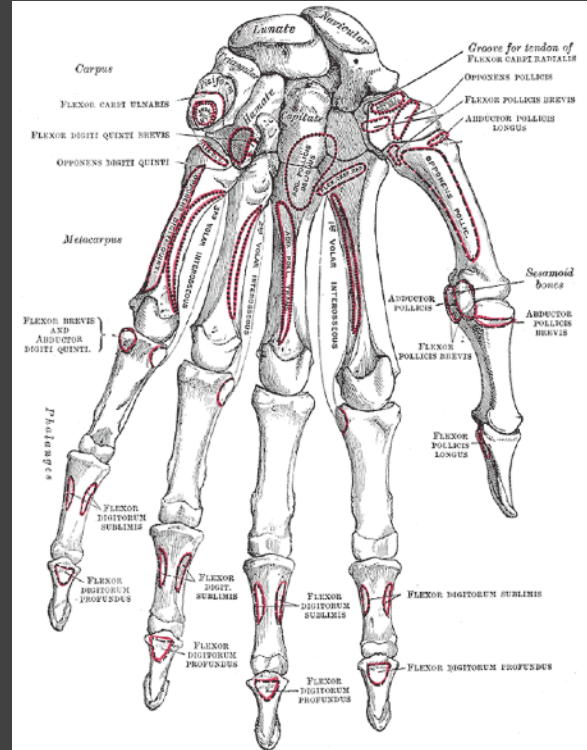


"to affect thro' the Eyes
what we fail to convey to
the public through their
word-proof ears"

Communicate, Inform, Inspire



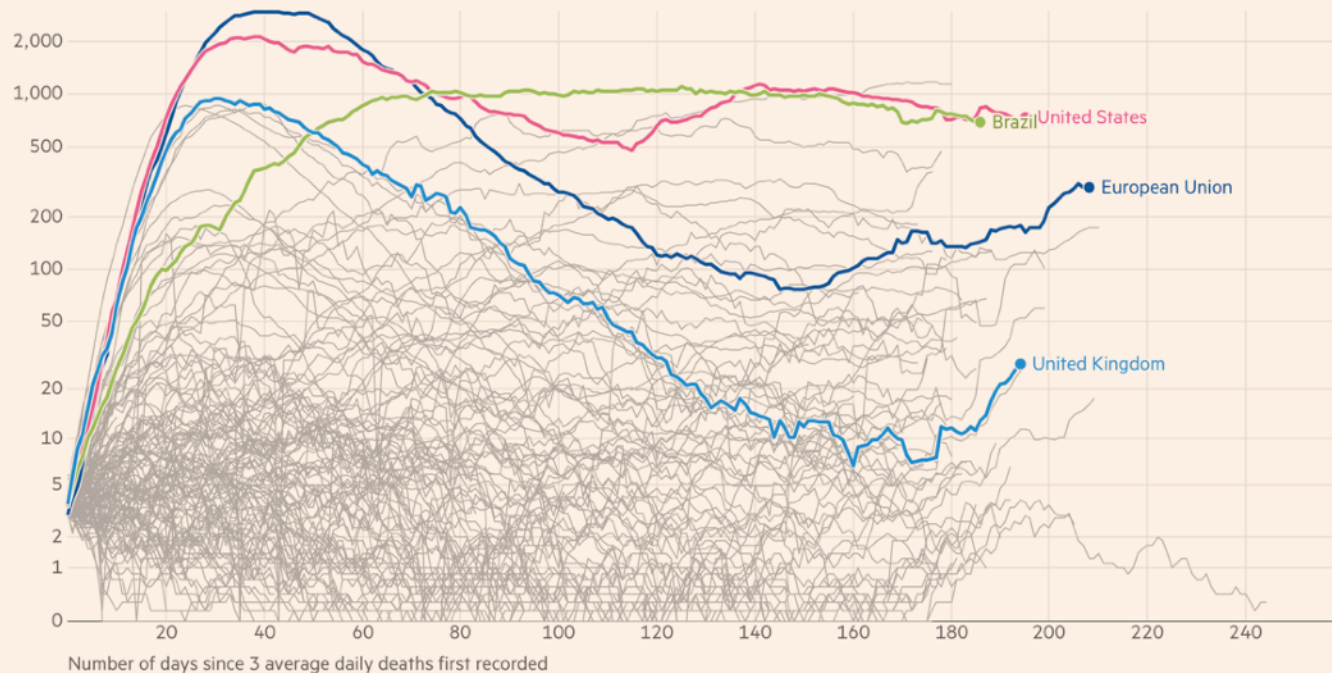
Visualizing Black America, Du Bois et al. 1900



Bones in hand, Gray's Anatomy 1918 ed.

New deaths attributed to Covid-19 in European Union, United States, Brazil and United Kingdom

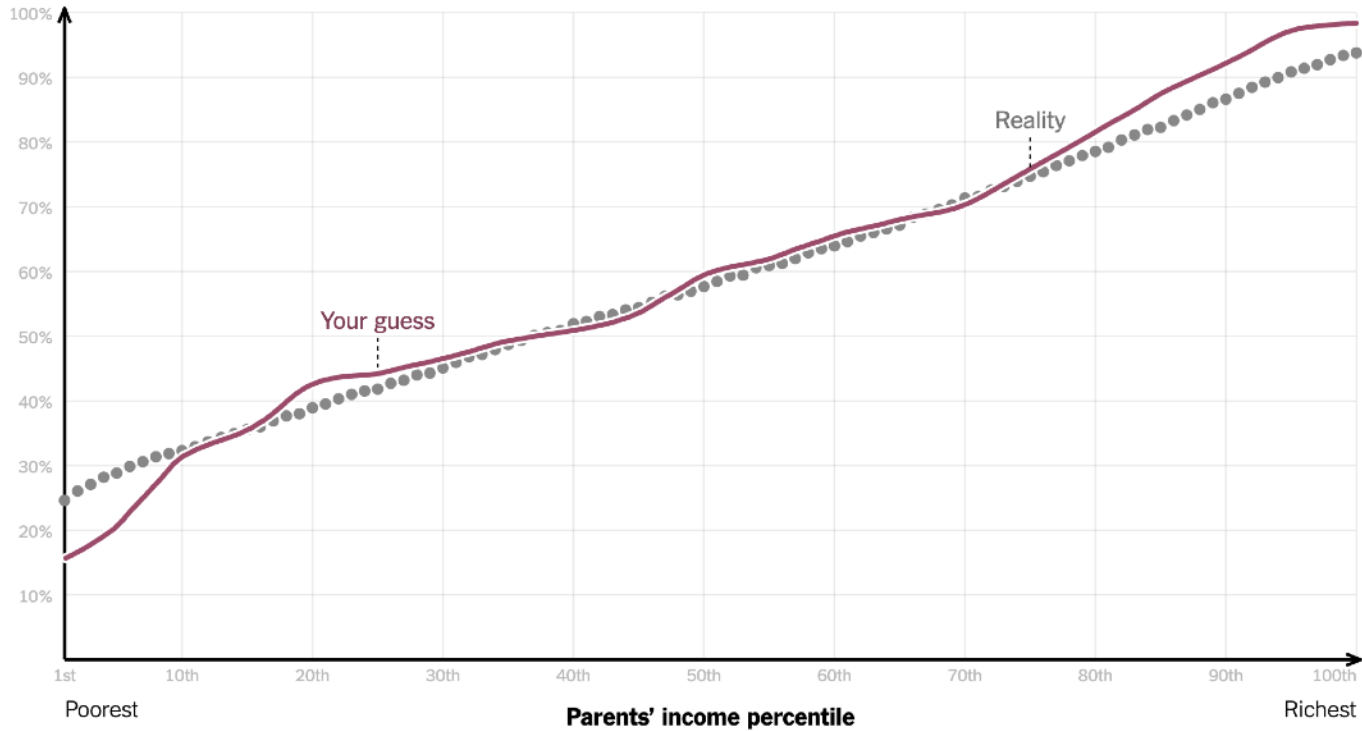
Seven-day rolling average of new deaths, by number of days since 3 average daily deaths first recorded



Source: Financial Times analysis of data from the European Centre for Disease Prevention and Control, the Covid Tracking Project, the UK Dept of Health & Social Care and the Spanish Ministry of Health.
Data updated September 25 2020 12.46pm BST. Interactive version: ft.com/covid19

FINANCIAL TIMES

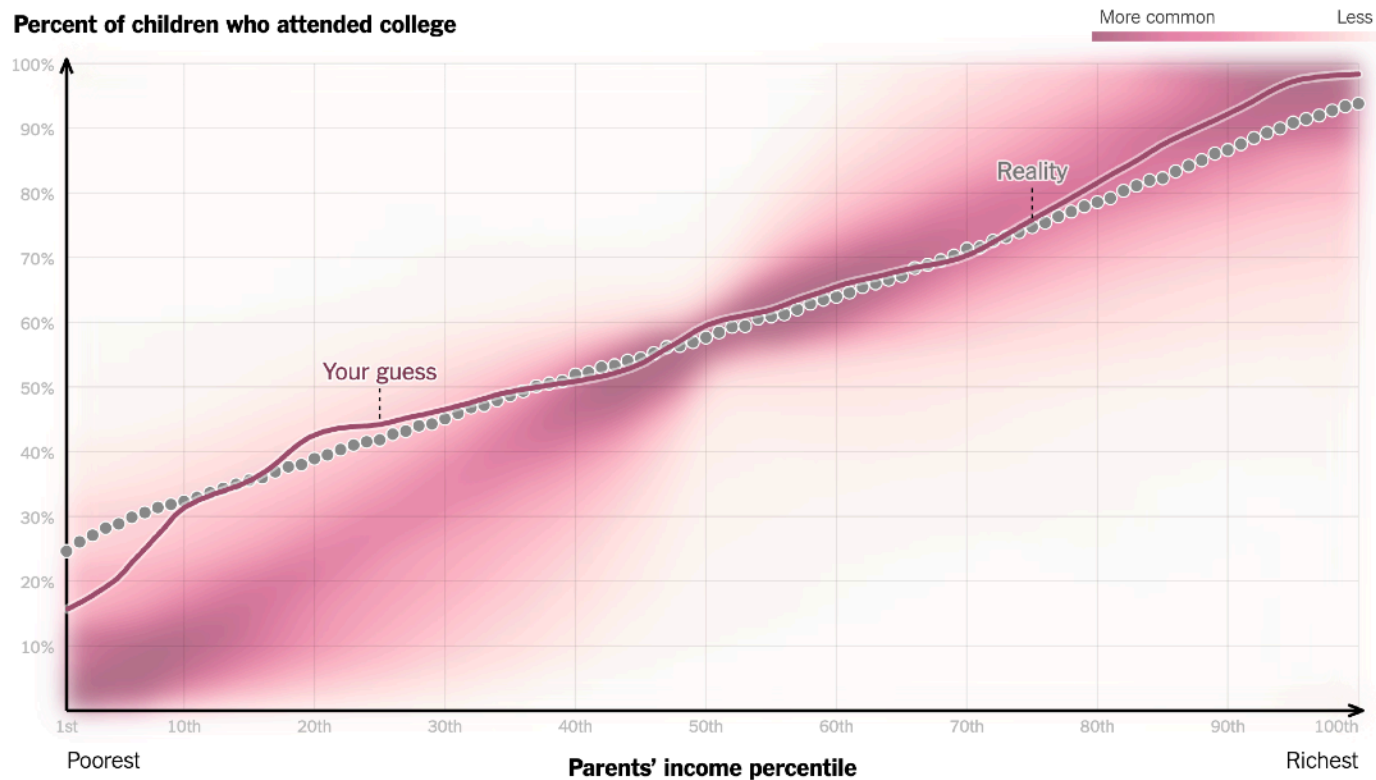
Percent of children who attended college



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

[New York Times, May 28, 2015]

Percent of children who attended college



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

[New York Times, May 28, 2015]

Course Overview

Course Overview

W1: Introduction to Visualization

W2: Visualization Tools, Part 1 & Visual Encoding

W3: Deceptive Visualization Data Transformation

W4: Interaction & Mapping

W5: Visualization Tools, Part 2 & D3.js Tutorial

W6: Animation & Color

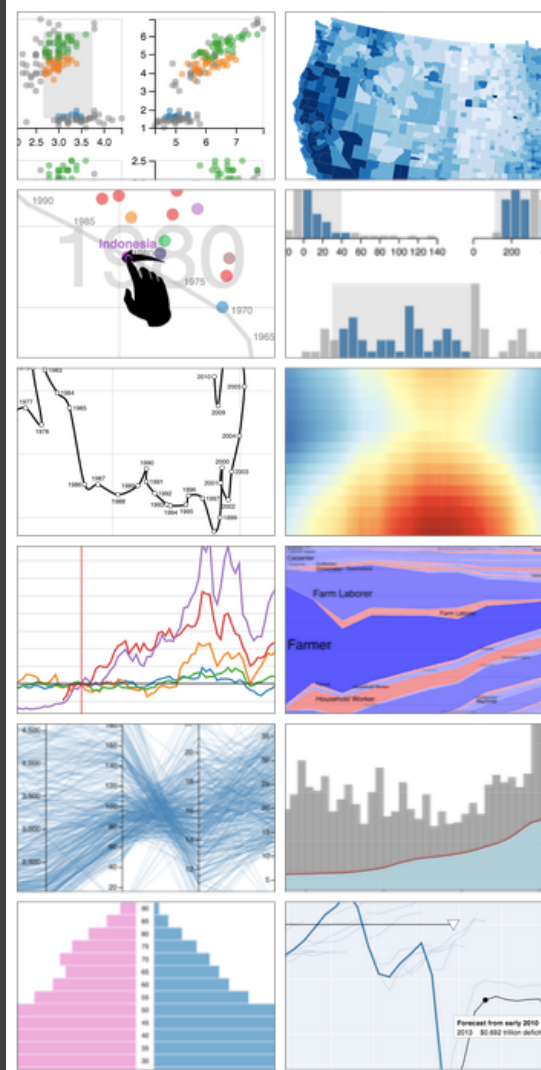
W7: Perception & Final Project Kick-Off

W8: Uncertainty

W9: Networks & Final Project Milestone Review

W10: Scalable Visualization

W11: Evaluation & Final Project Showcase



Visual Encoding

LES VARIABLES DE L'IMAGE										12	14
			POINTS			LIGNES			ZONES		
XY 2 DIMENSIONS DU PLAN Z											
	TAILLE										
	VALEUR										
LES VARIABLES DE SÉPARATION DES IMAGES										13	
GRAIN											
COULEUR											
ORIENTATION											

Data Transformation



Q Latent Dimensions: 32 ▾ Projection: t-SNE ▾ Perplexity: 30 ▾

t-SNE



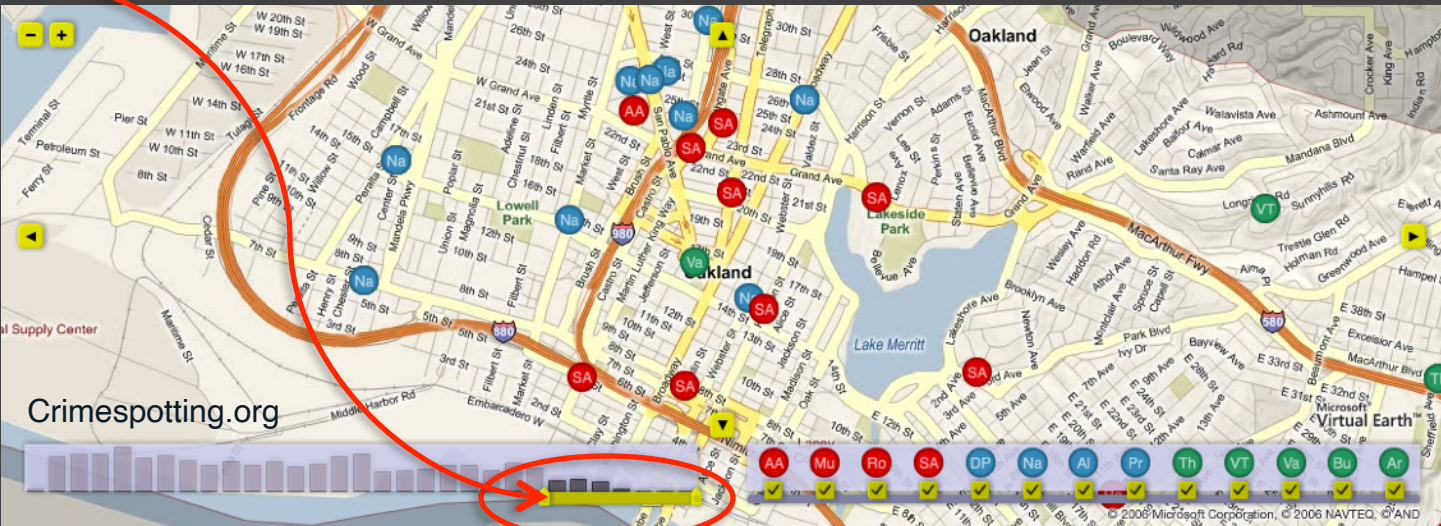
Q Latent Dimensions: 32 ▾ Projection: UMAP ▾ Neighbors: 15 ▾ Distance: 0.1 ▾

UMAP

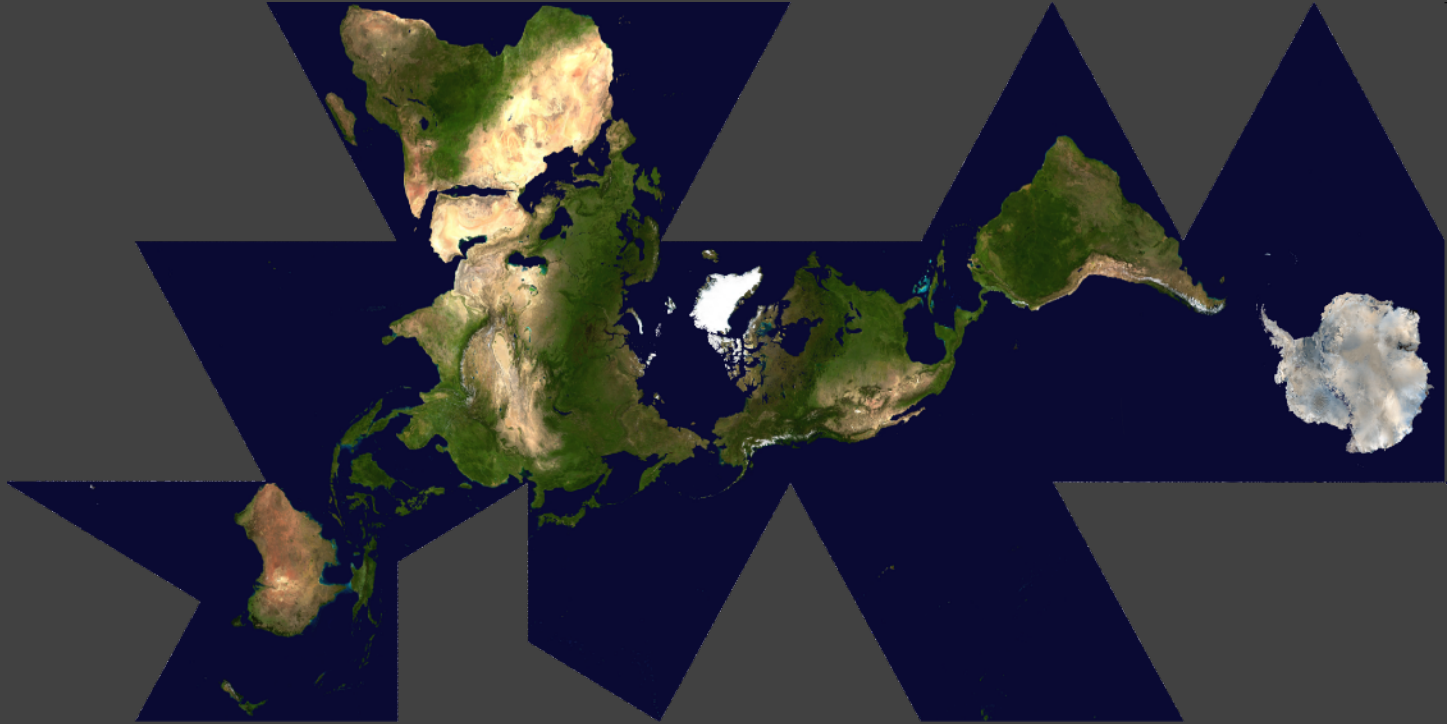


Q Latent Dimensions: 32 ▾ Projection: PCA ▾ X-Axis: PC1 ▾ Y-Axis: PC2 ▾

PCA

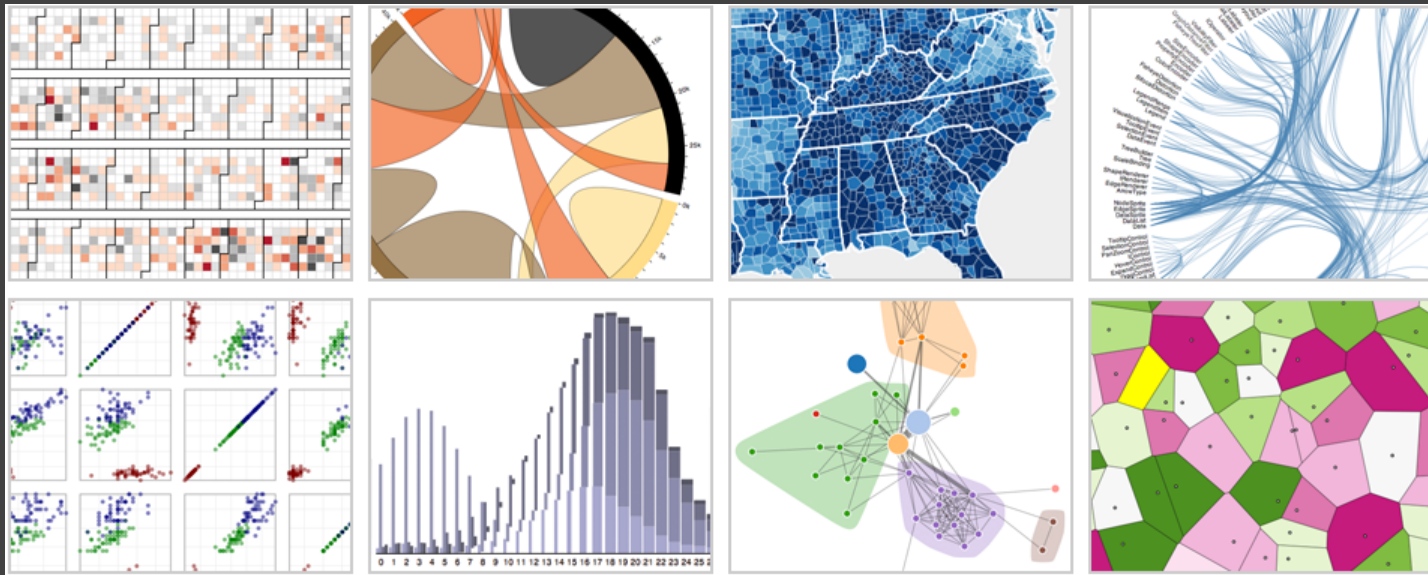


Mapping & Cartography



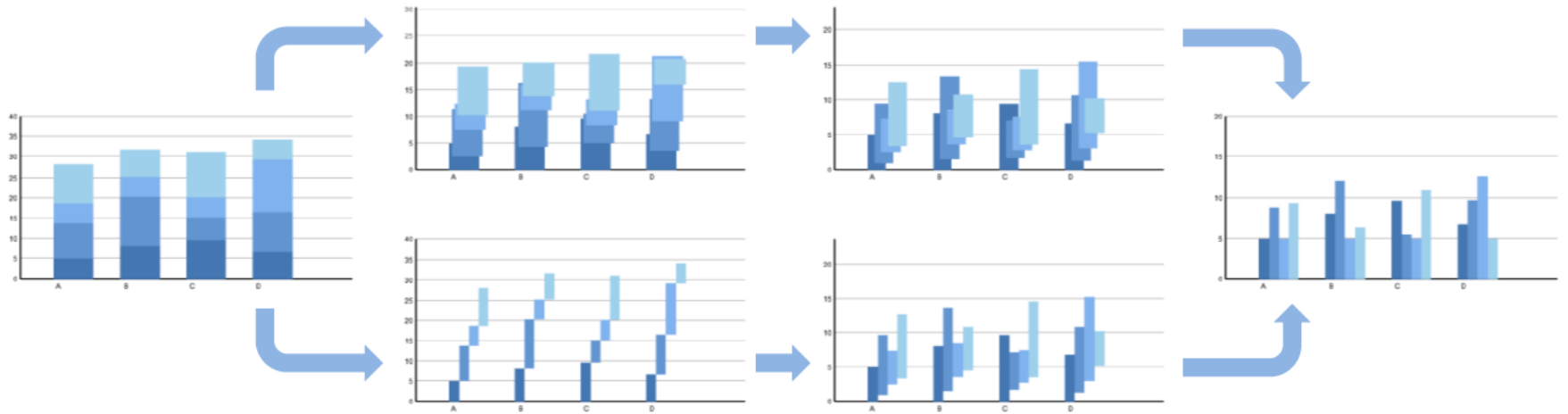
Dymaxion Maps [Fuller 46]

Visualization Tools



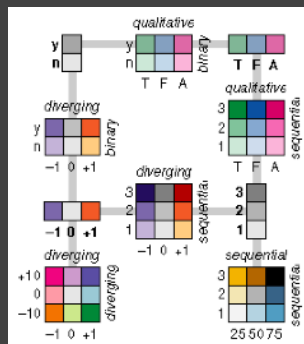
D3: Data-Driven Documents

Animation

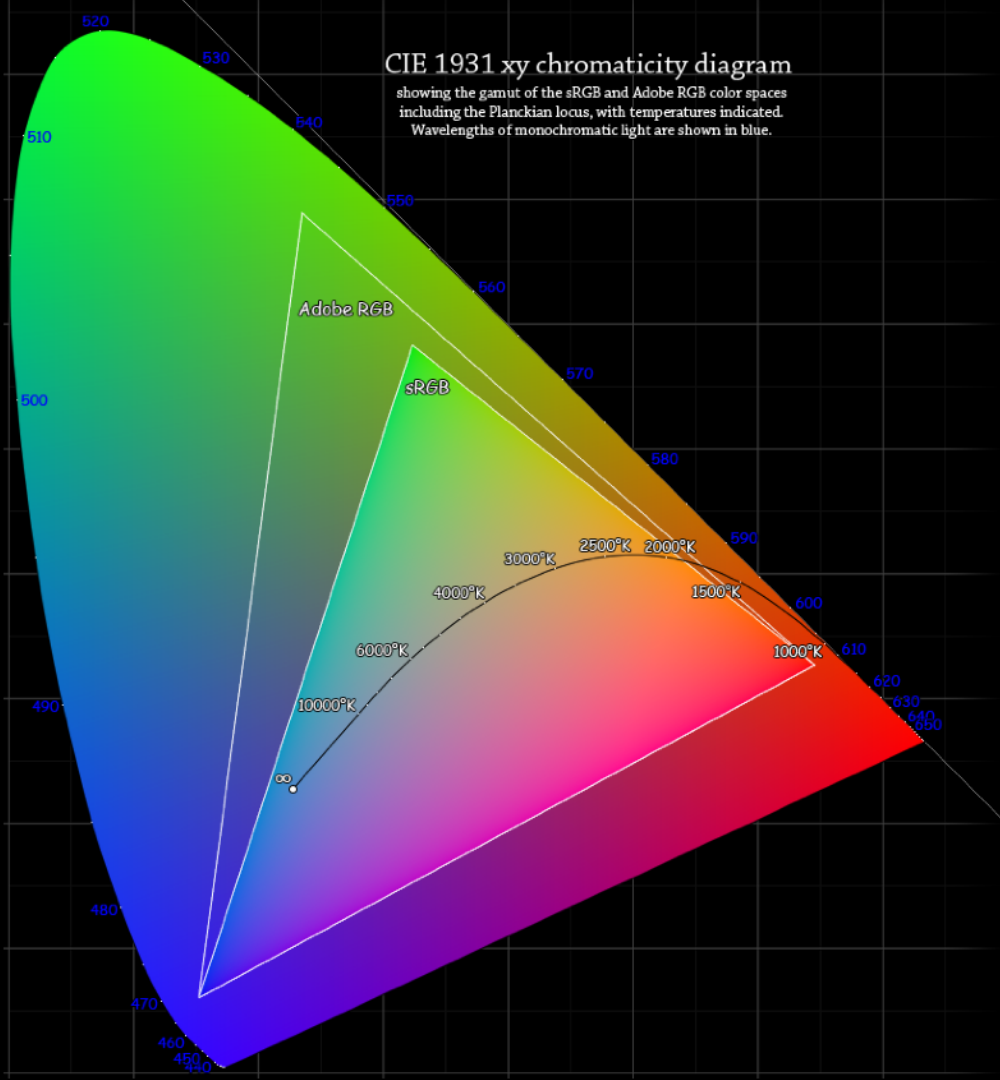


Animated transitions in statistical data graphics [Heer & Robertson 07]

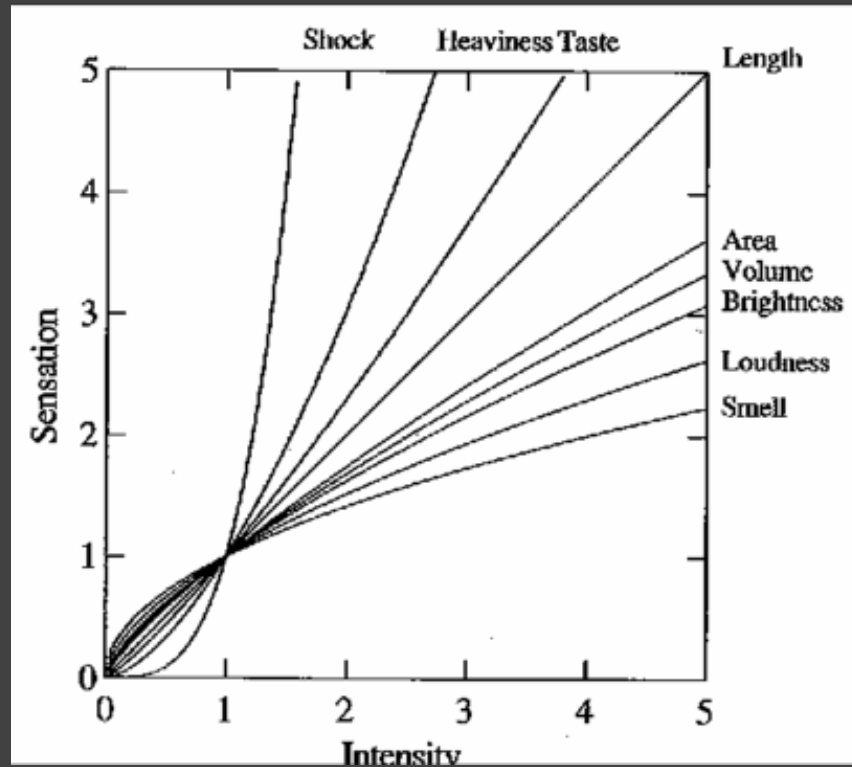
Color



Color Brewer

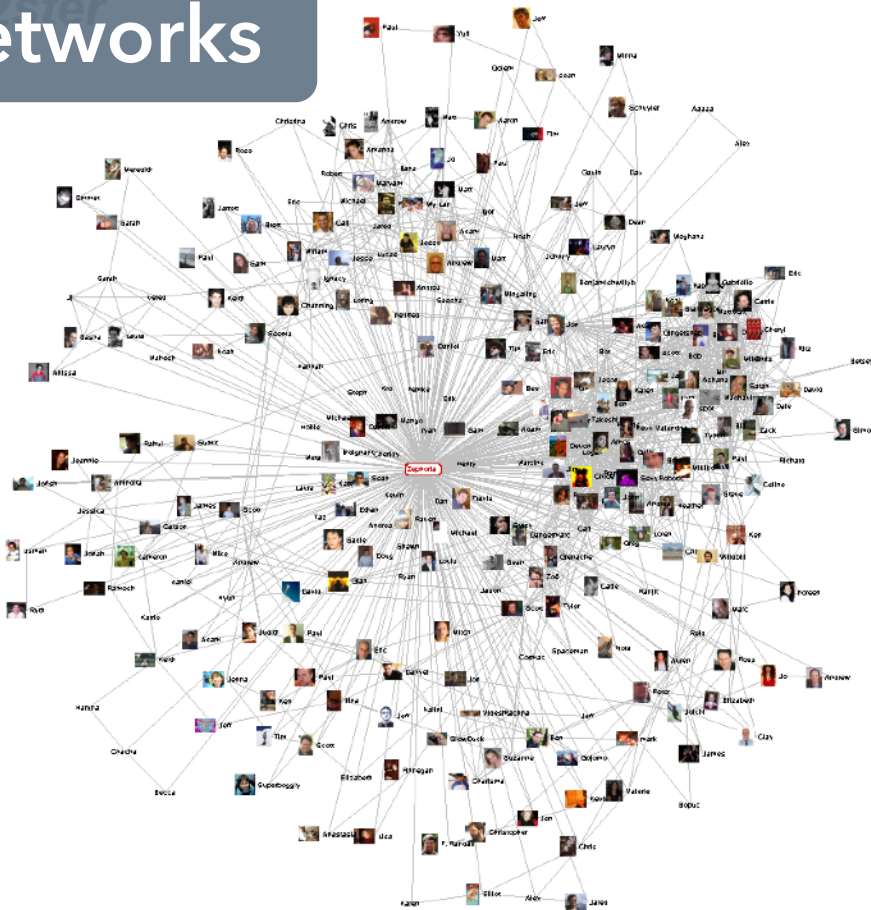


Graphical Perception



The psychophysics of sensory function [Stevens 61]

Networks



Zephoria

User ID 21721

Friends 266

Ago ??

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/goatrance [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/though>

Want to Meet

Someone who makes life's complexities seem simply elegant.

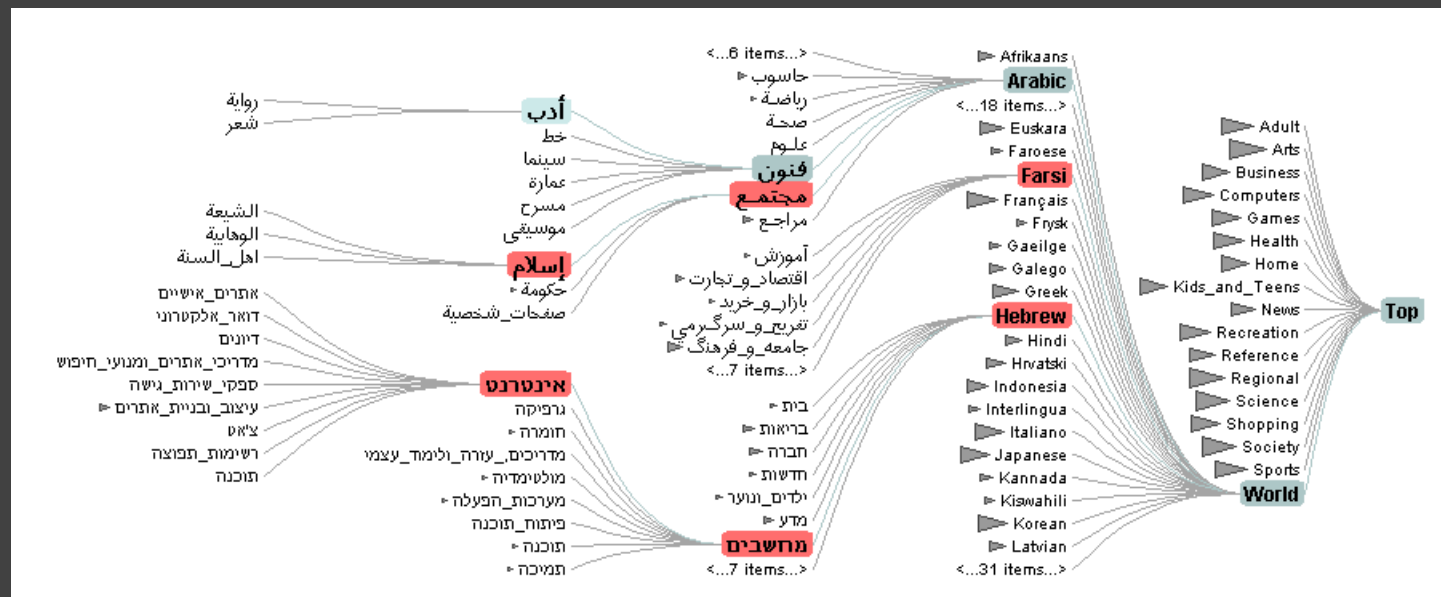
community >>

Enable

search >>

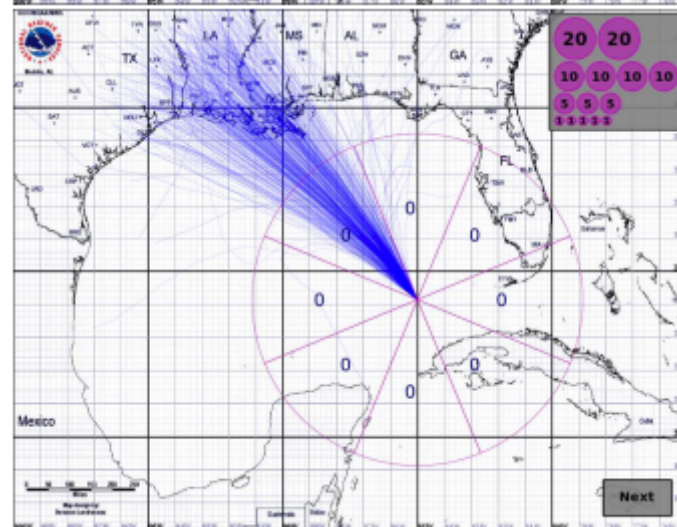
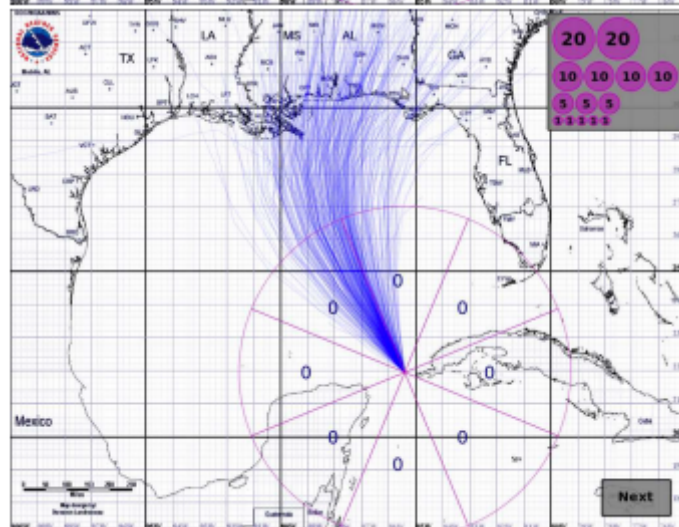
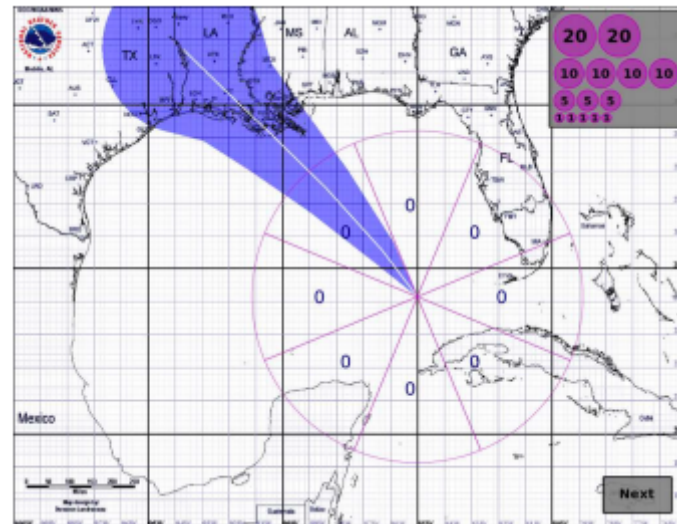
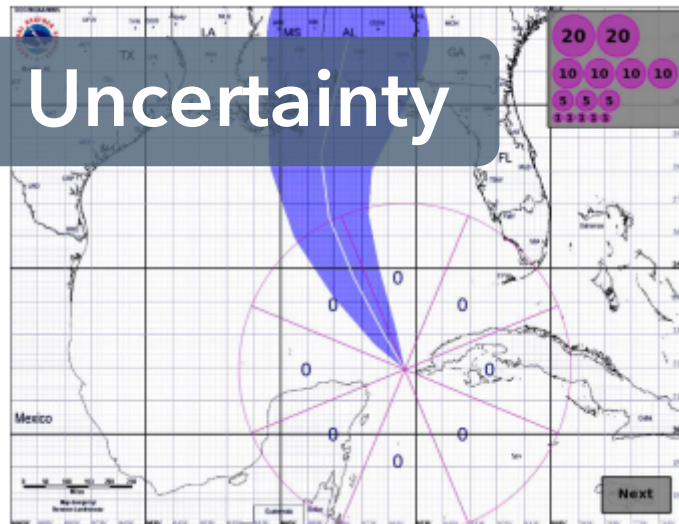
23

Networks

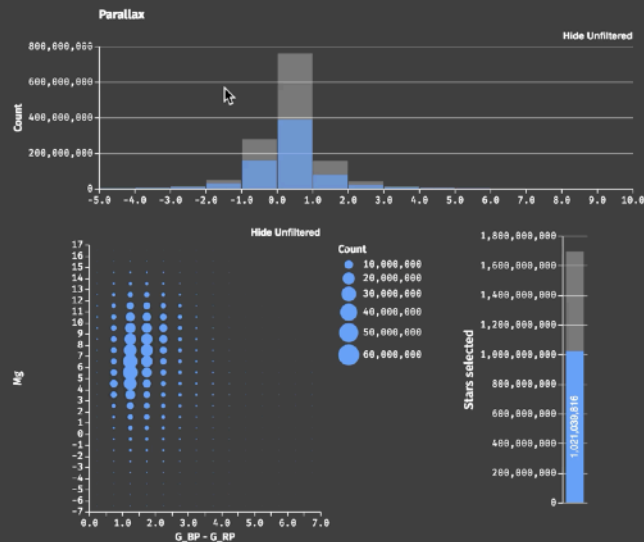
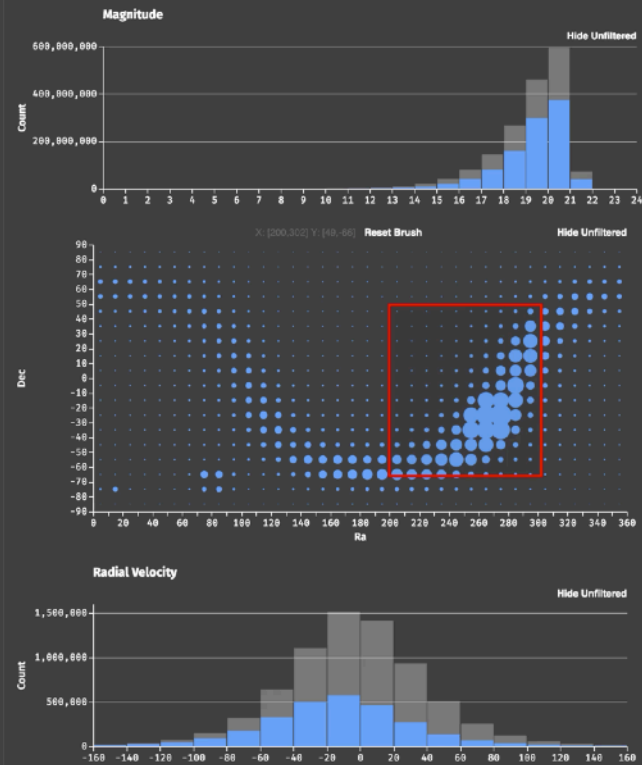


Degree-Of-Interest Trees [Heer & Card 04]

Uncertainty



Scalability



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

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

Instructors

cse442@cs

Lead Instructor

Jeffrey Heer OH: Tue 10-11am, Gates 302

Teaching Assistants

Shaan Chopra OH: Thu 2-3pm, Gates 151

Madeleine OH: Fri 10:15-11am, Gates 374

Grunde-McLaughlin

Sebastin Santy OH: Mon 5:15-6:15pm, Allen 220

Will Wang OH: Online / By Appointment

Sebastin Santy

5th year PhD student

Curious about what intelligence is —

Human and Machine

Shaan Chopra *(she/her)* | 5th year, CSE PhD

Research:

- Human-Computer Interaction
- AI & Personal Health Informatics
- Participatory & Inclusive Design of Health Technologies
- Personal Data Use in Clinical Settings
- Health Equity & Community-Based Health

Personal Interests / Hobbies:

- Outdoor / sports: rowing, hiking, biking, basketball, discus throw, swimming...
- Creative: painting, live sketching, crocheting
- Food-related: drinking mochas, baking & eating desserts





Madeleine Grunde-McLaughlin

Research Background

Cognitive science, Interpretable AI, Data Visualization

Current research

Hypothesis formalization for confirmatory research

Hobbies

dancing, gardening, cooking, biking

5th year PhD student, UW CSE

mgrunde@cs.washington.edu

OH: Friday 10:15-11

Will (Huichen) Wang

wwill@cs.washington.edu

Third-year CS PhD student

Research:

- GenAI for Visualization & Data Science

Fun:

- Ping Pong
- Movies
- Chess and Go



Lectures, Activities & Office Hours

Course sessions will alternate among lecture and in-class exercises (often Thursdays).

All lectures will be in-person and recorded.

Please attend in-person but **NOT** if you feel ill.

Office hours are a mix of in-person and Zoom.

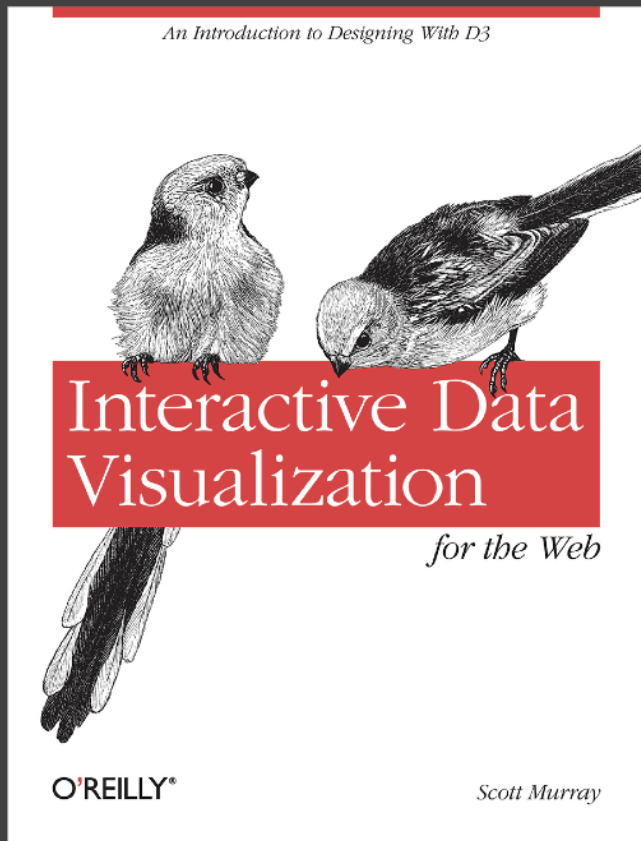
Links for virtual office hours are on Canvas.

Use Ed to post questions and seek help!

Readings

There is no one universal textbook on visualization!
We draw on books, notebooks, and linked articles...
Material in class will loosely follow readings.
Readings should be read by start of class.

Textbook



Interactive Data Visualization for the Web, 2nd Edition

For learning D3!

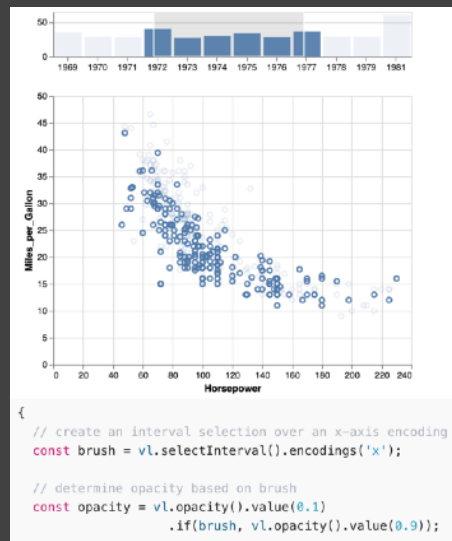
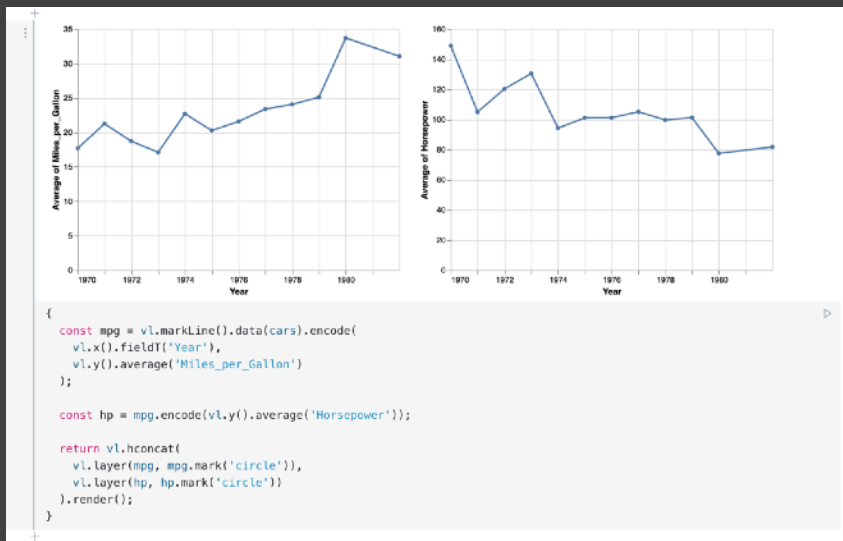
Book available online.

Code / examples on GitHub.

We will be using D3 v7.

<https://d3js.org>

Interactive Notebooks



Hands-on engagement with course concepts and tools using Observable (JavaScript) notebooks.

Assignments

CP Class Participation (10%)

A1 Expository Visualization (10%)

A2 Deceptive Visualization (15%)

Peer Review (5%)

A3 Interactive Prototype (20%)

Peer Review (5%)

FP Final Project (35%)

Proposal

Demonstration Video

Final Prototype

Grading Philosophy

A *great* submission gets a *great grade* (A- to A, 3.6 – 3.8), but an *exceptional* grade (A+, 3.9 – 4.0) requires *exceptional* effort.

Example: Typical A1 grades (out of 10 points).

Everyone starts with a high score (9/10).

We then *deduct* points for errors and also *add* points for going above and beyond the assignment requirements.

The median score for A1 is typically 8.5 out of 10, which maps to an A-.

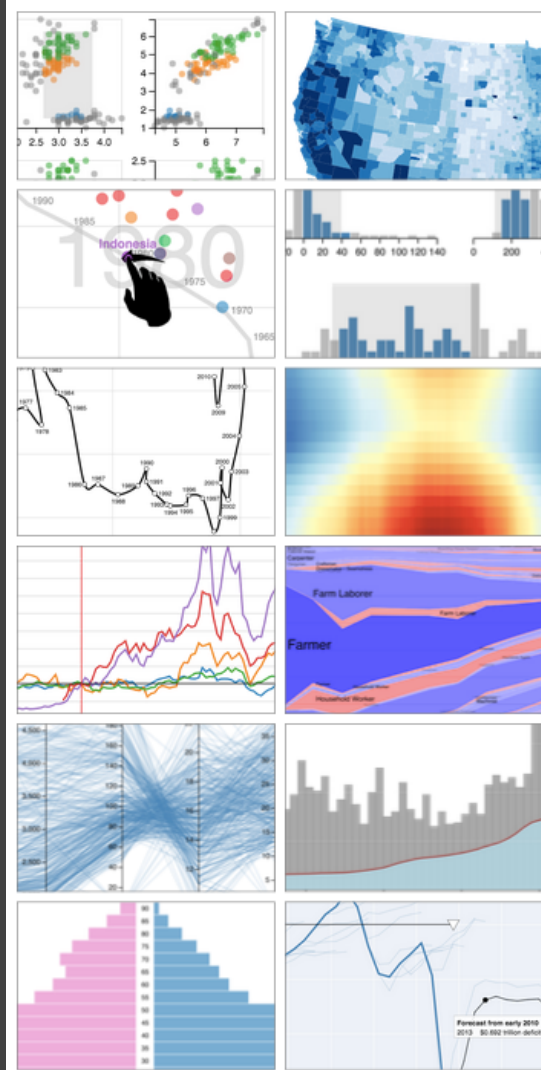
Course Participation

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".



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

task

questions, goals
assumptions

data

physical data type
conceptual data type

domain

metadata
semantics
conventions

processing
algorithms

mapping
visual encoding

image

visual channel
graphical marks



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: =, \neq

O - Ordered

- Operations: =, \neq , $<$, $>$

Q - Interval (location of zero arbitrary)

- Operations: =, \neq , $<$, $>$, -
- Can measure distances or spans

Q - Ratio (zero fixed)

- Operations: =, \neq , $<$, $>$, -, %
- Can measure ratios or proportions

Visual Language is a Sign System



Jacques Bertin

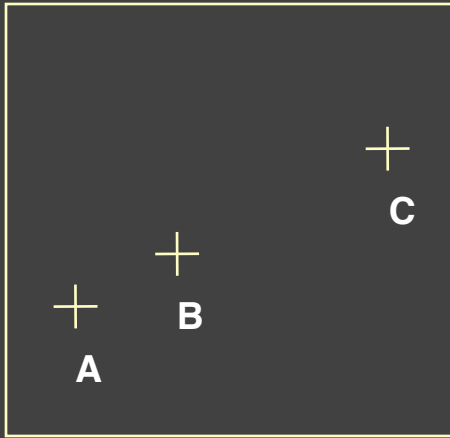
Images perceived as a set of signs

Sender encodes information in signs

Receiver decodes information from signs

Sémiologie Graphique, 1967

Bertin's Semiology of Graphics



1. A, B, C are distinguishable

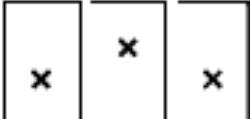
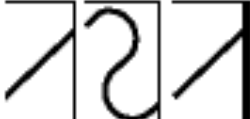



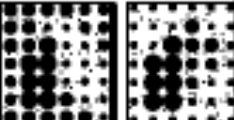



2. B is between A and C.

3. BC is twice as long as AB.



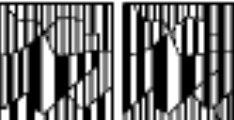




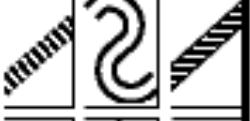




∴ Encode quantitative variables

"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

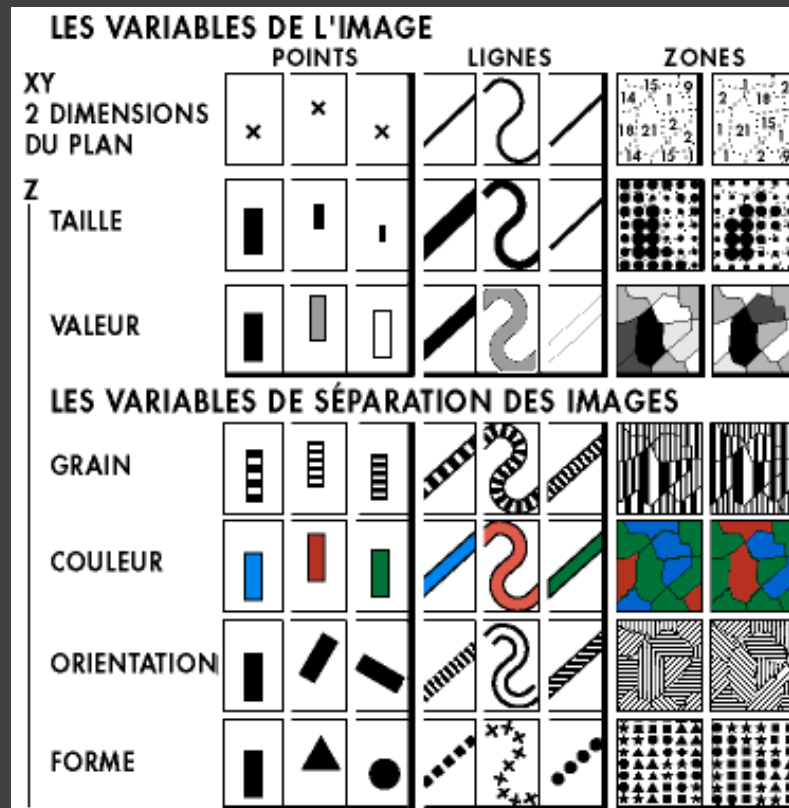
Value

Texture

Color

Orientation

Shape



Visual Encoding Channels

Position

Length

Area

Volume

Value

Texture

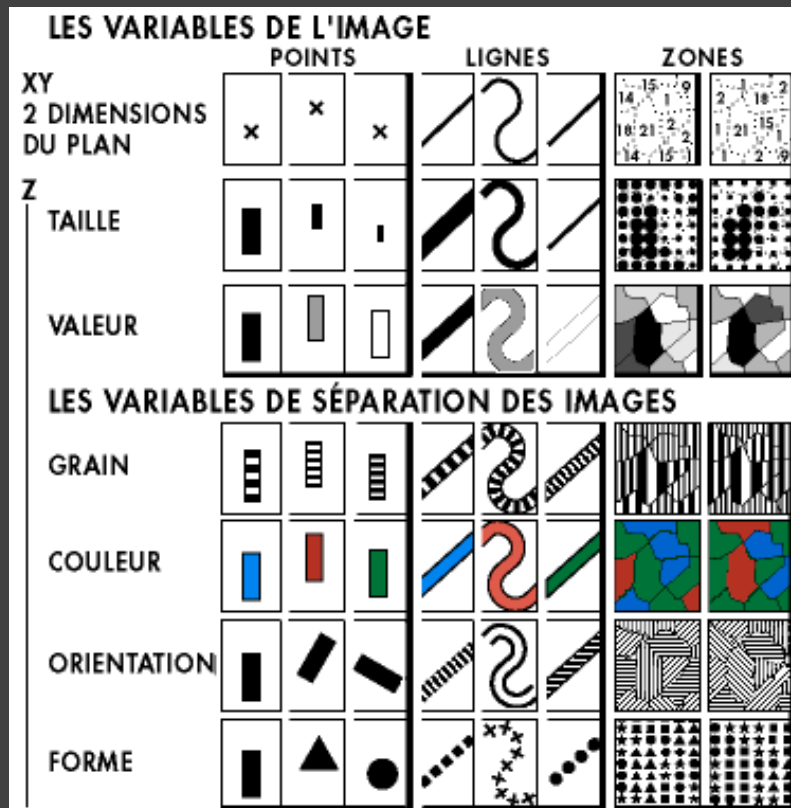
Color

Orientation

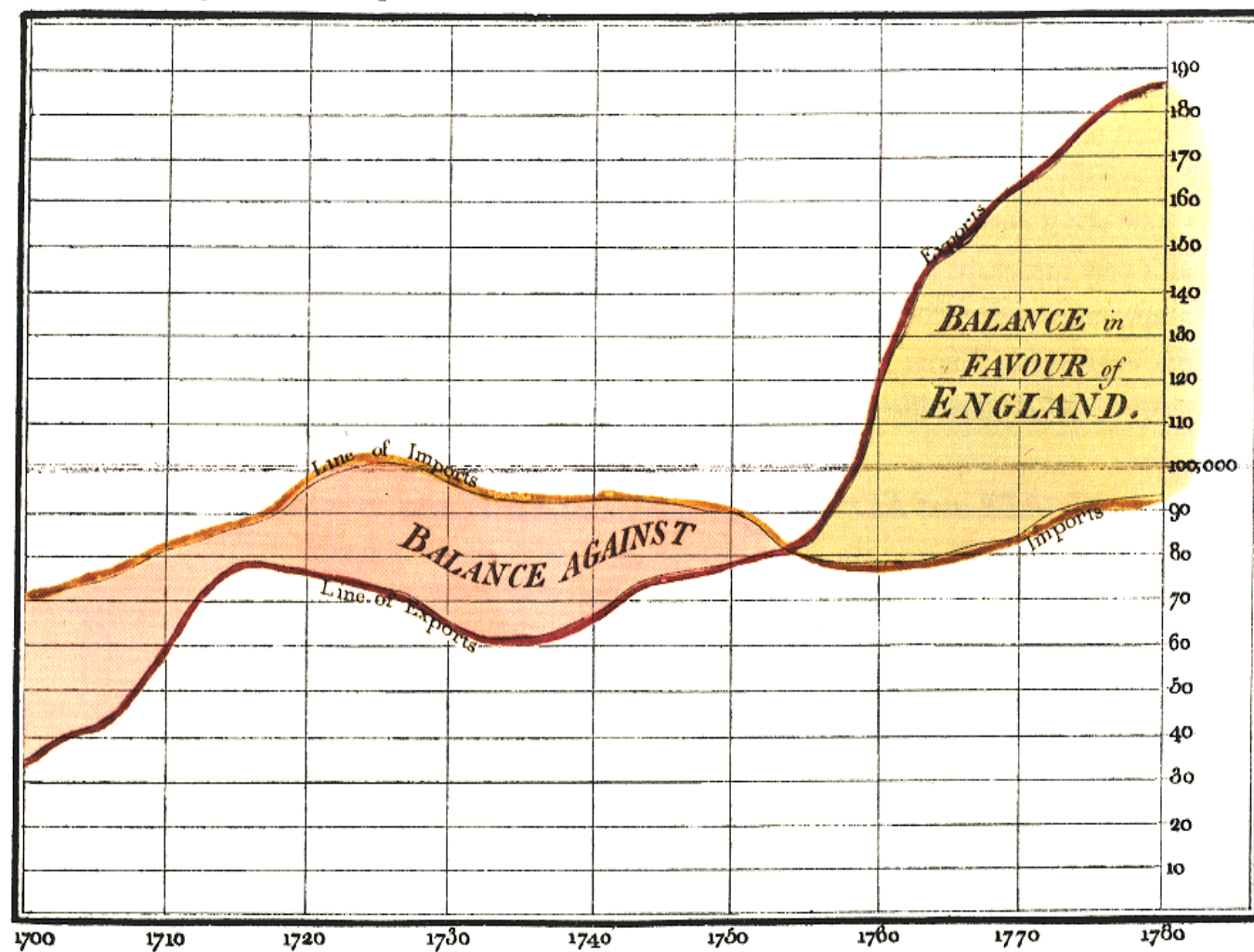
Shape

Transparency

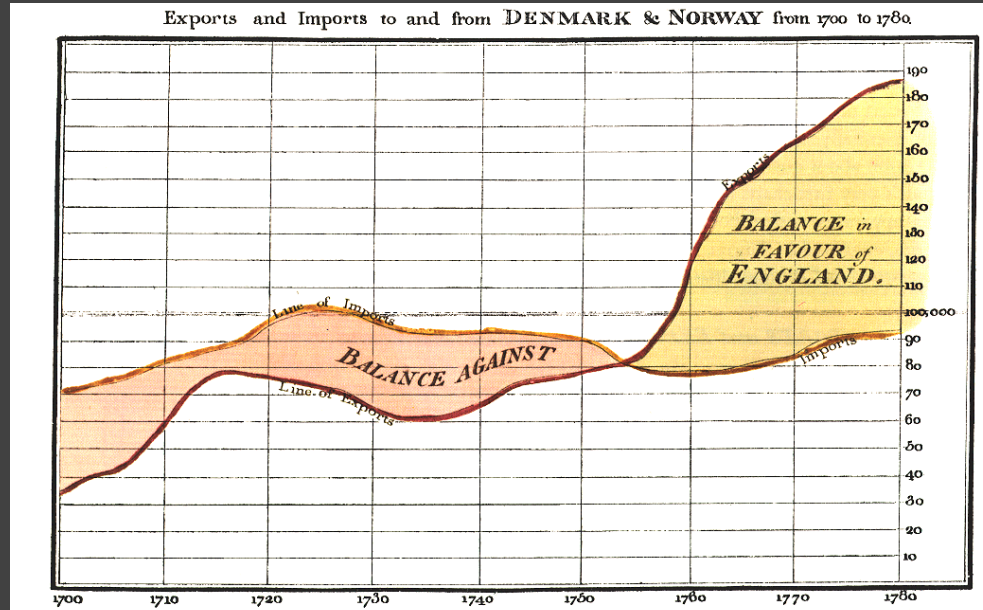
Blur / Focus ...



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

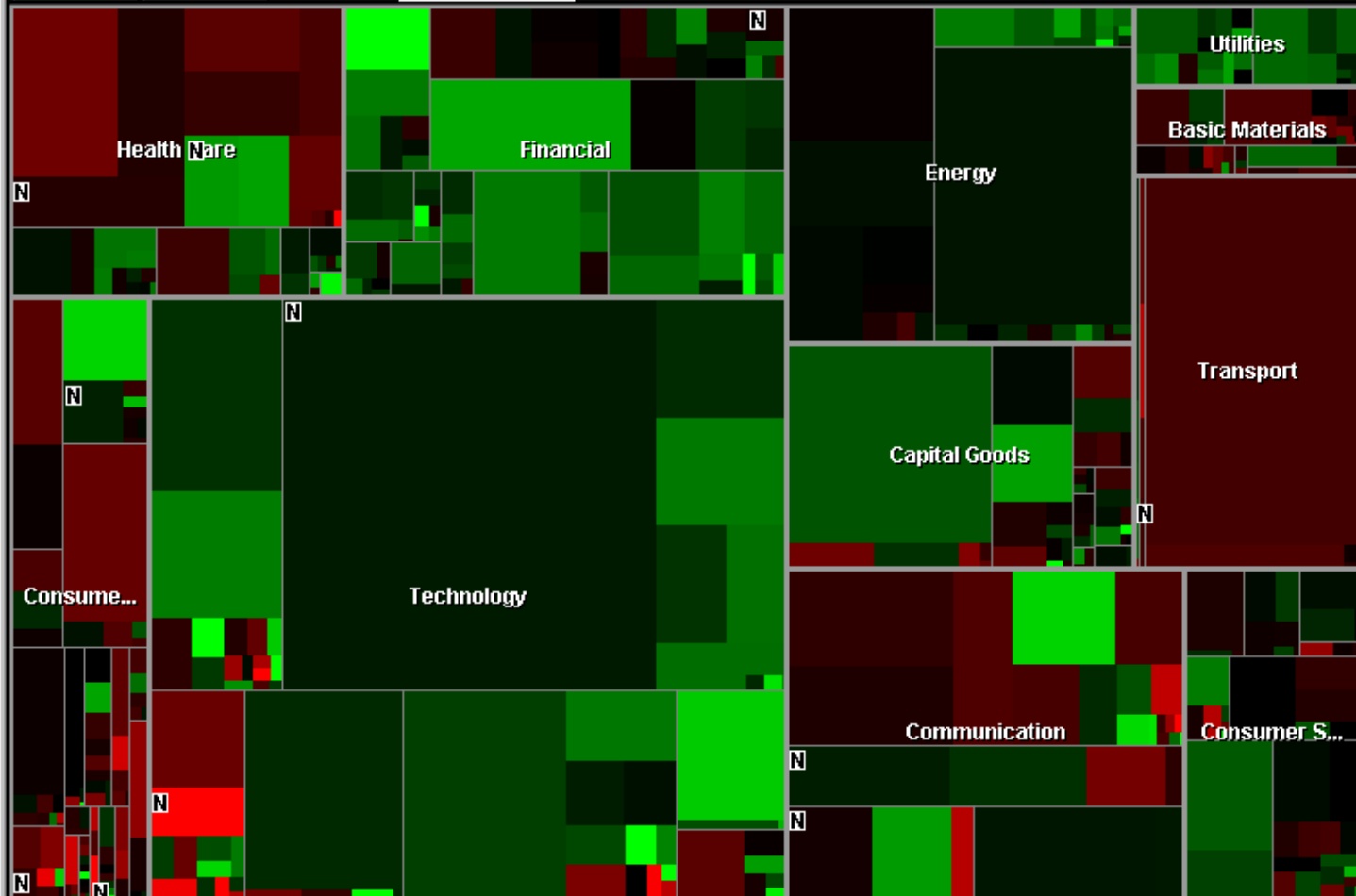
Instructions

Headline Icons

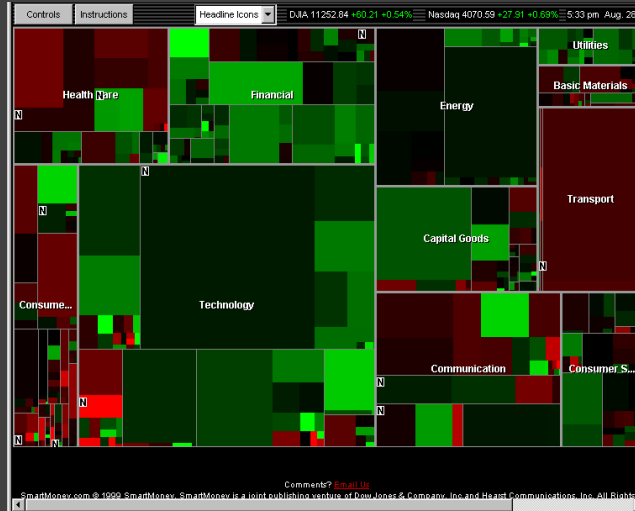
DJIA 11252.84 +60.21 +0.54%

Nasdaq 4070.59 +27.91 +0.69%

5:33 pm Aug. 28



Wattenberg's Map of the Market



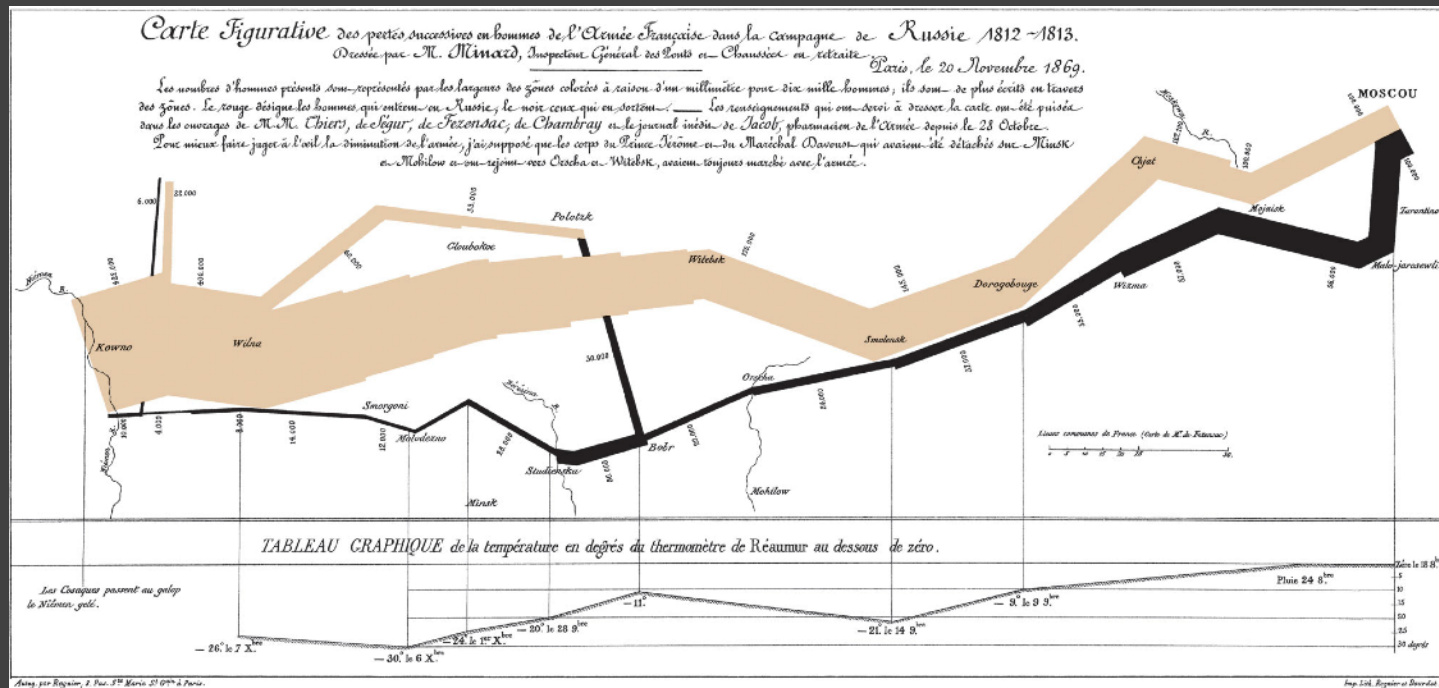
Rectangle Area: market cap (Q)

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

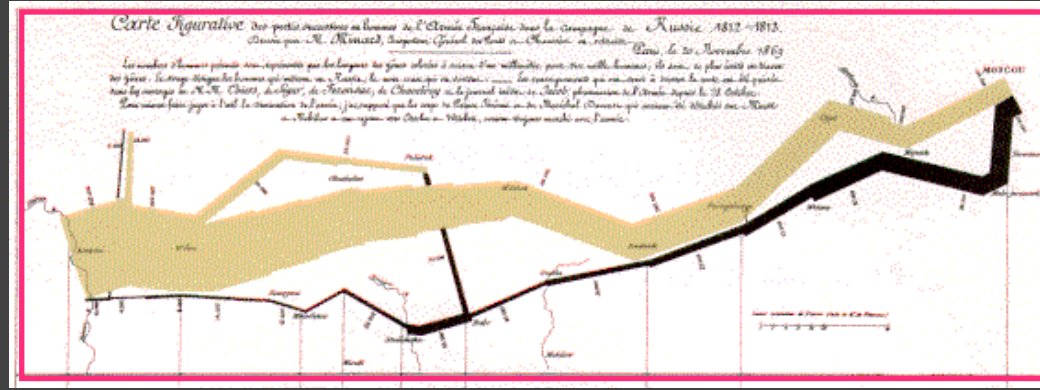
Color Hue: loss vs. gain (N, O)

Color Value: magnitude of loss or gain (Q)

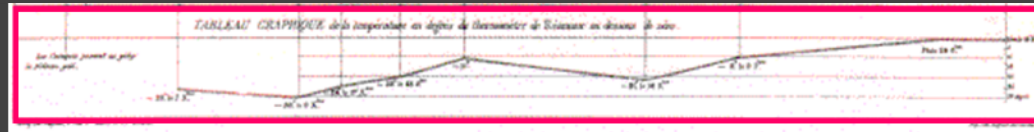
Minard 1869: Napoleon's March



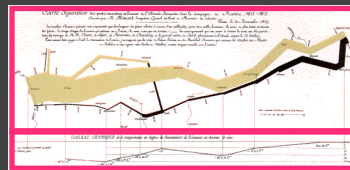
Single-Axis Composition



+



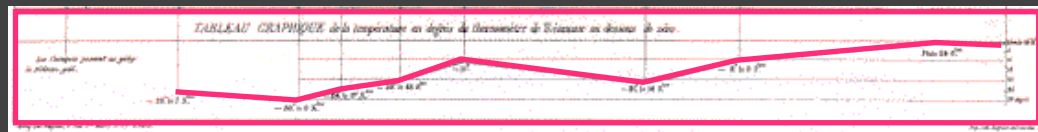
=



Mark Composition

Y-axis: temperature (Q)

X-axis: longitude (Q) / time (O)



Temp over space/time ($\mathbb{Q} \times \mathbb{Q}$)

Mark Composition

Y-axis: latitude (Q)

+ X-axis: longitude (Q)

+ Width: army size (Q)

=

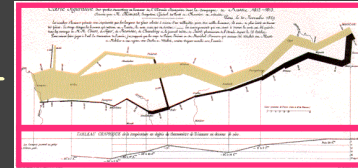


Army position (Q x Q) and army size (Q)

latitude (Q)

longitude (Q)

army size (Q)

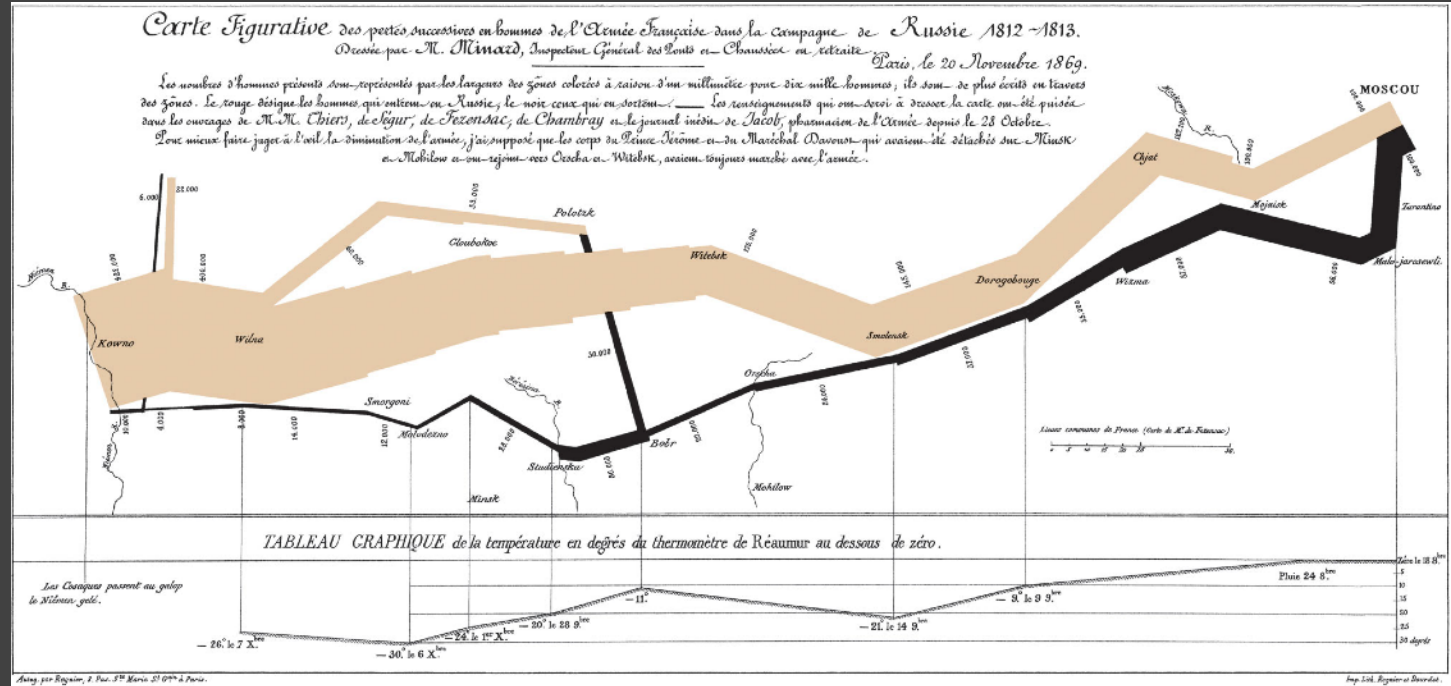


temperature (Q)

longitude (Q) / time (O)



Minard 1869: Napoleon's March



Depicts at least 5 quantitative variables. Any others?

Assignment 1

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.

A1: 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 visualization tool**.

This builds on our next in-class exercise, be sure to read the exercise before diving in!!

Deliverables via Gradescope

Image of your visualization (PNG or JPG format)

Short description + design rationale (≤ 4 paragraphs)

Due by **EOD, Fri October 3**.