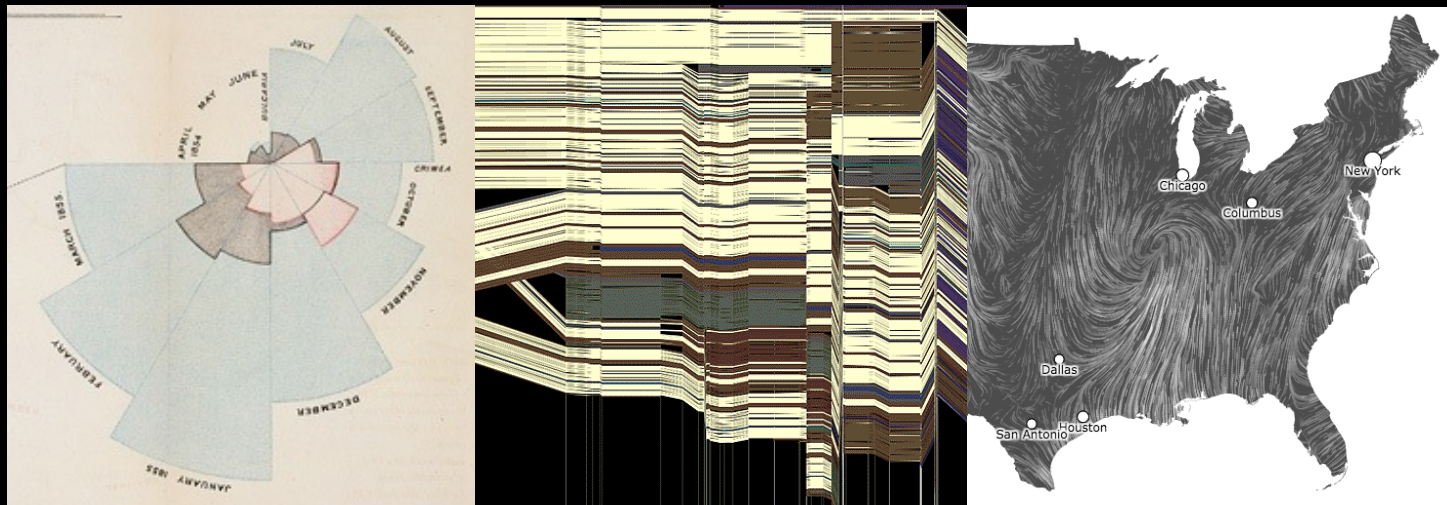


CSE 512 - Data Visualization

Introduction



Leilani Battle University of Washington

What is Information Visualization?

Or “Data Visualization” or simply “Visualization”

Discuss with your neighbors

Form a definition

Then we will share out!

Three Formal Definitions

“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]

Abstract Data

Symbolic

Data

Information Visualizations (Graphics)

Geometric

Artificial Memory

Visual
Representation

Human Reasoning

Human
Perception

Human Cognition

**Information Visualizations Are Graphics
That Leverage Human Perception to
Illustrate Abstract Data and Thereby
Enhance Human Reasoning**

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

Why Does Visualization Matter?

Discuss with your neighbors

Then we will share out!

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

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

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

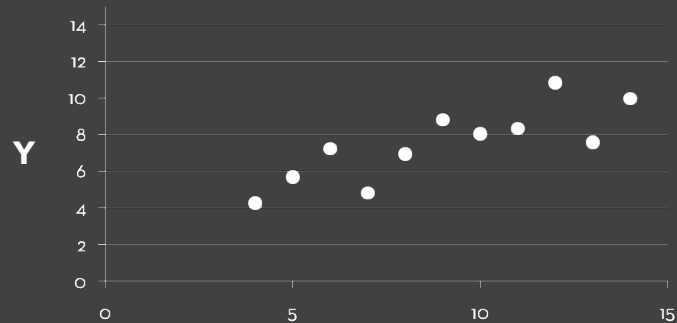
Linear Regression

$$Y = 3 + 0.5 X$$

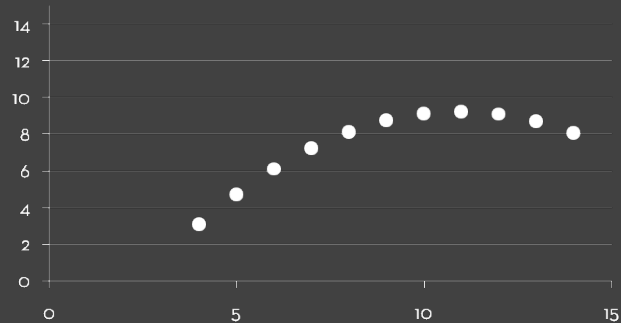
$$R^2 = 0.67$$

[Anscombe 1973]

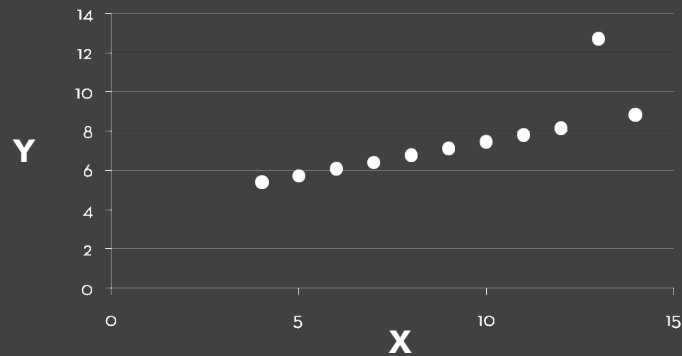
Set A



Set B

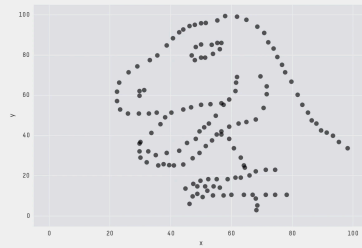


Set C

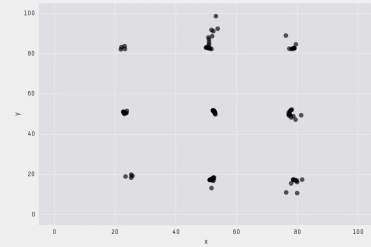
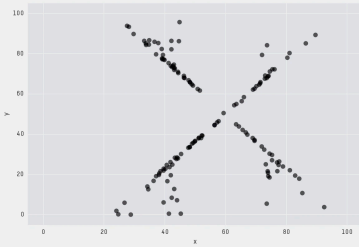
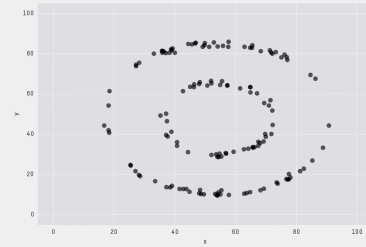
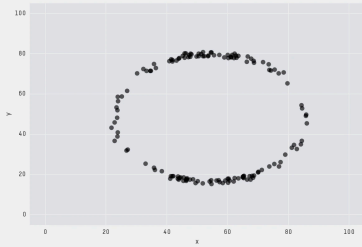
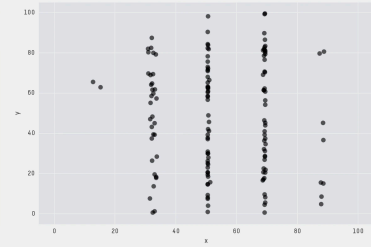
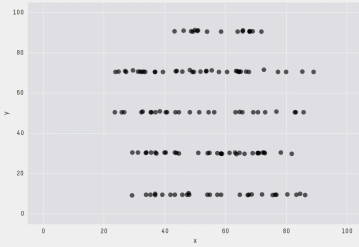
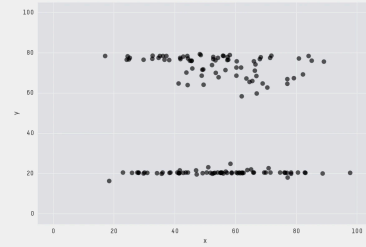
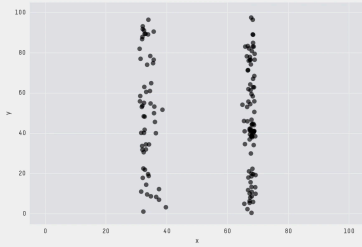
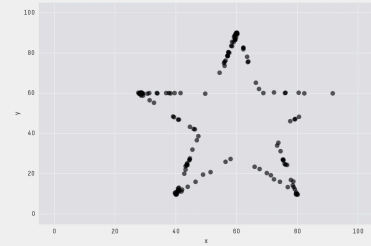
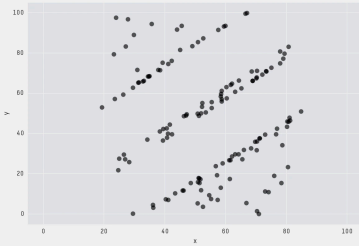
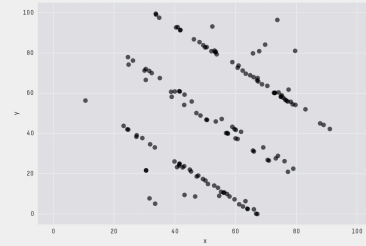
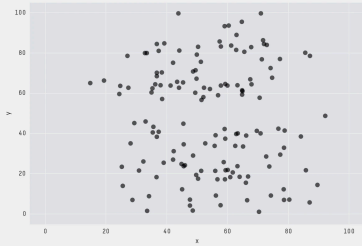


Set D





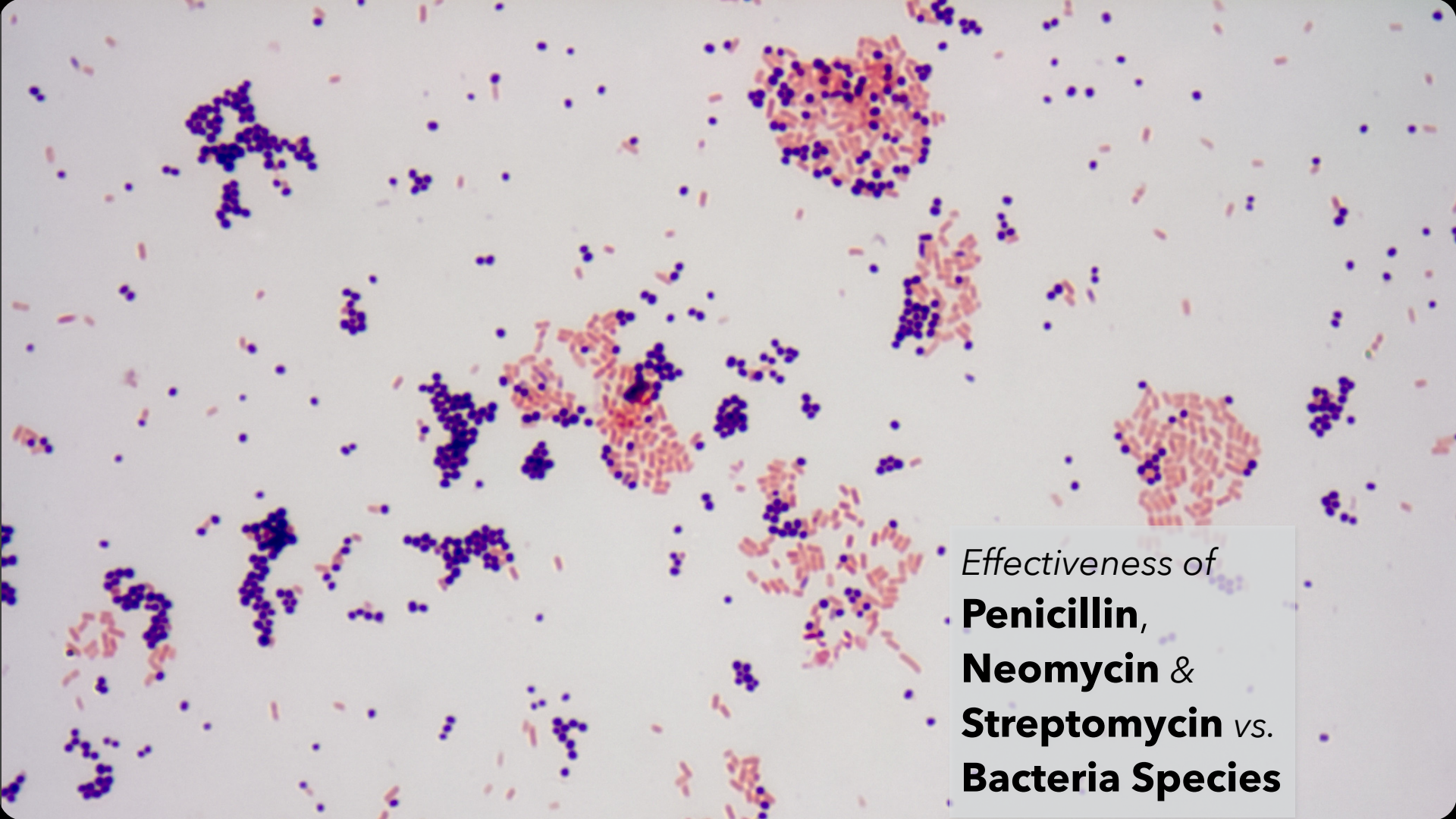
X Mean: 54.26
Y Mean: 47.83
X SD : 16.76
Y SD : 26.93
Corr. : -0.06





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.**

Example:
Antibiotic Effectiveness



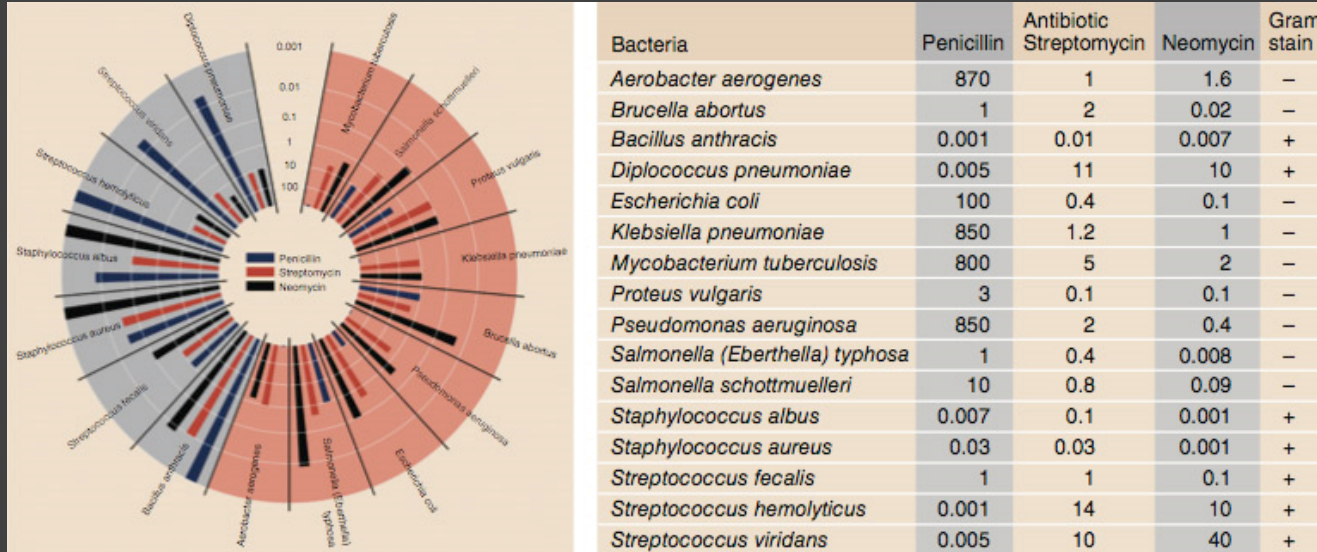
Effectiveness of
Penicillin,
Neomycin &
Streptomycin vs.
Bacteria Species

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?

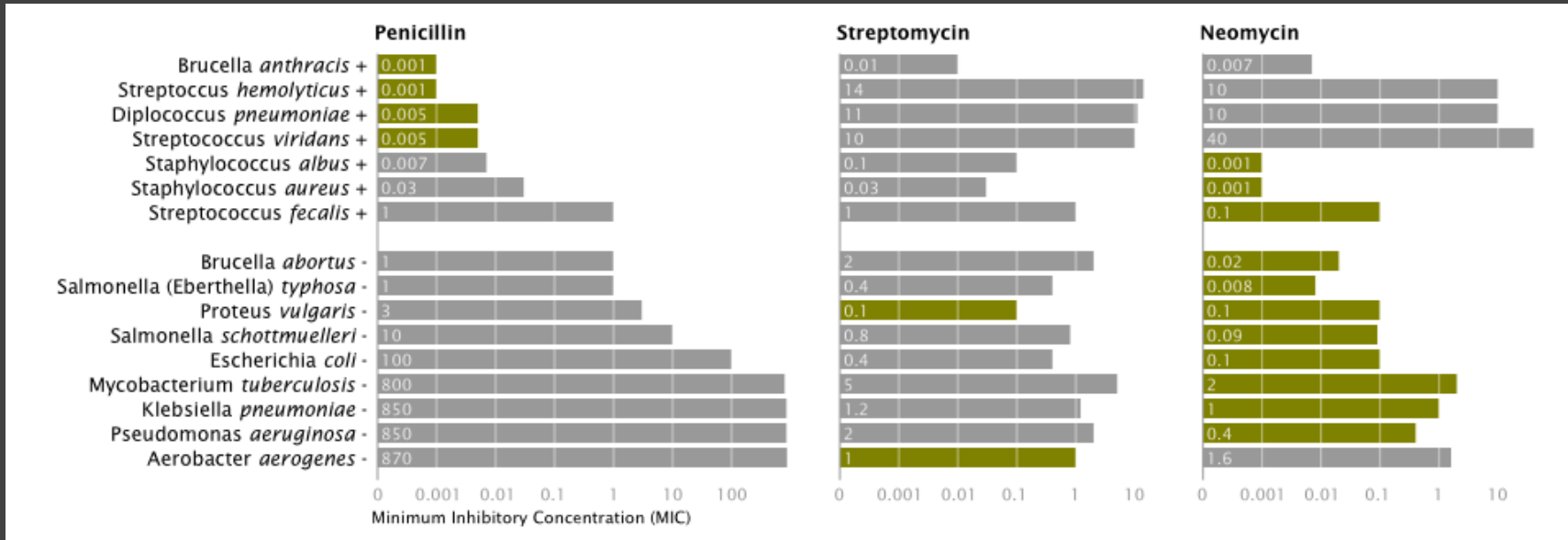


Radius: $1 / \log(\text{MIC})$

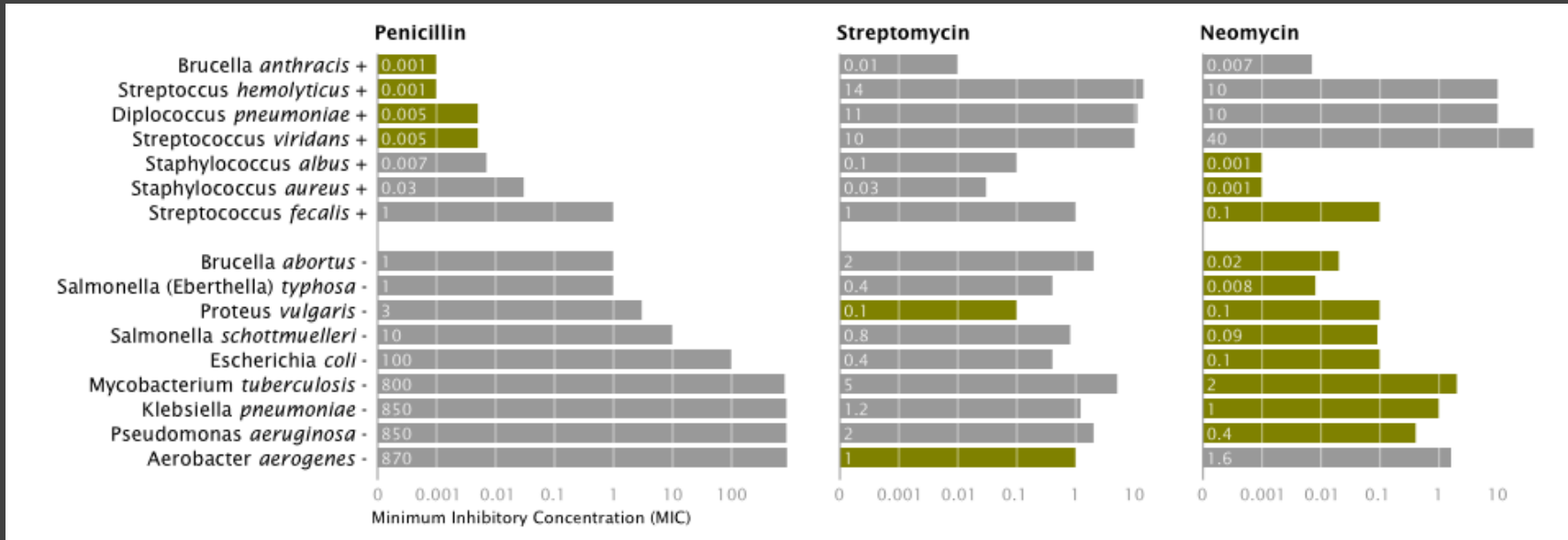
Bar Color: Antibiotic

Background Color: Gram Staining

Which antibiotic is most effective?



Which antibiotic is most effective?



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

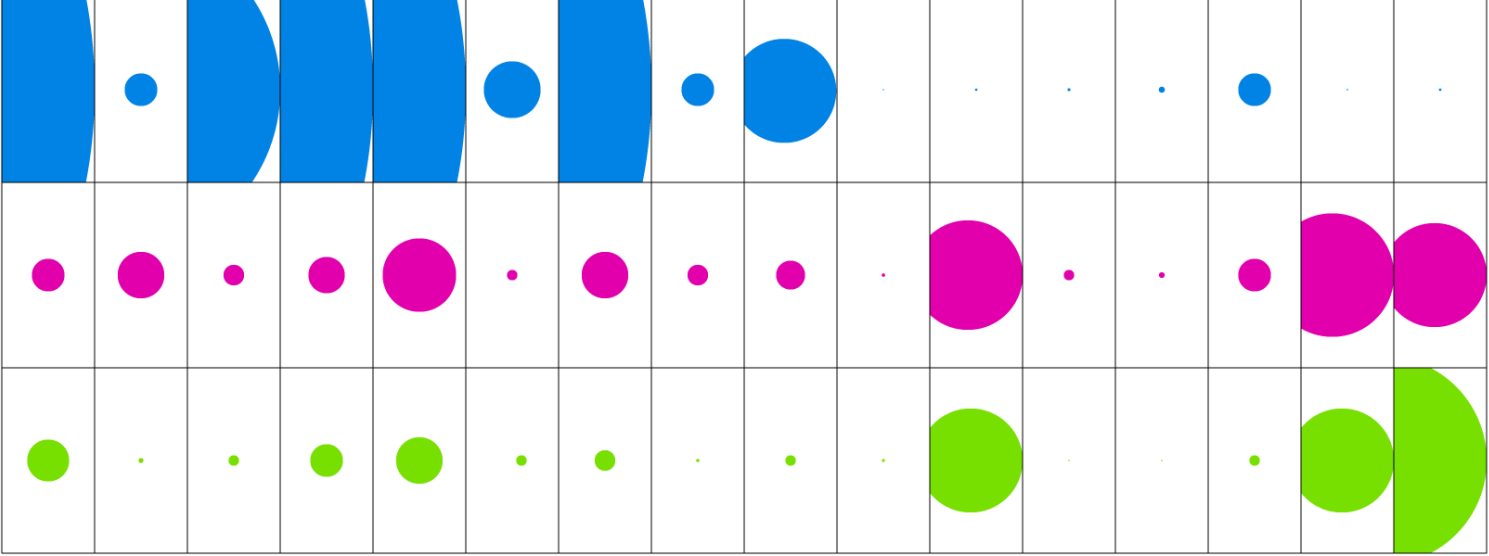
Y-axis: Gram-Staining | Species

Color: Most-Effective?

penicillin

streptomycin

neomycin



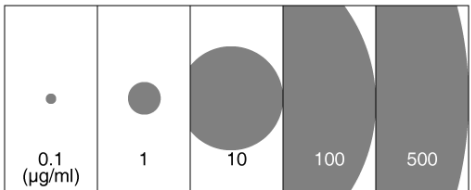
E. aerogenes
B. abortus
E. coli
K. pneumoniae
M. tuberculosis
P. vulgaris
P. aeruginosa
S. typhi
S. schottmulleri
B. anthracis
S. pneumoniae
S. albus
S. aureus
E. faecalis
S. pyogenes
S. viridans

Gram positive

Gram negative

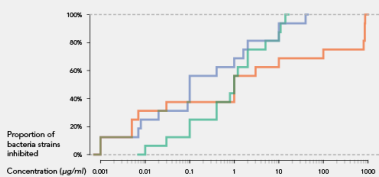
minimum inhibitory concentration of antibiotics

bowen li cs448b



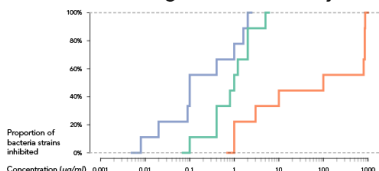
Bowen Li

All bacteria



Streptomycin and Neomycin are more efficient broad-spectrum antibiotics than Penicillin.

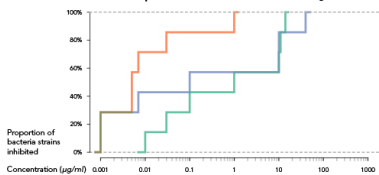
Gram-negative bacteria only



Neomycin and Streptomycin are more efficient against gram-negative bacteria, so can be used at a lower dosage here than above.

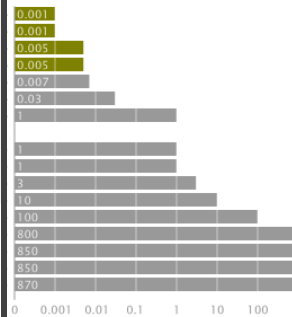
Gram staining quickly identifies bacteria as Gram-negative or Gram-positive, which can be used to find a more efficient antibiotic and dosage.

Gram-positive bacteria only



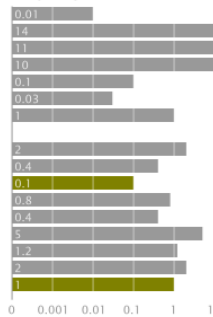
Penicillin is more efficient than either Streptomycin or Neomycin if the bacteria is known to be gram-positive.

Penicillin

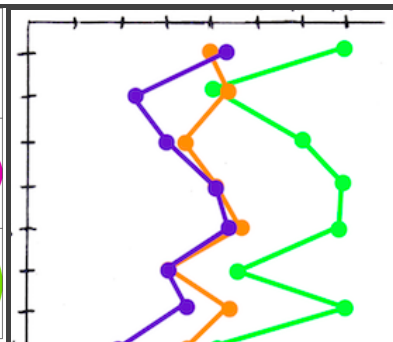
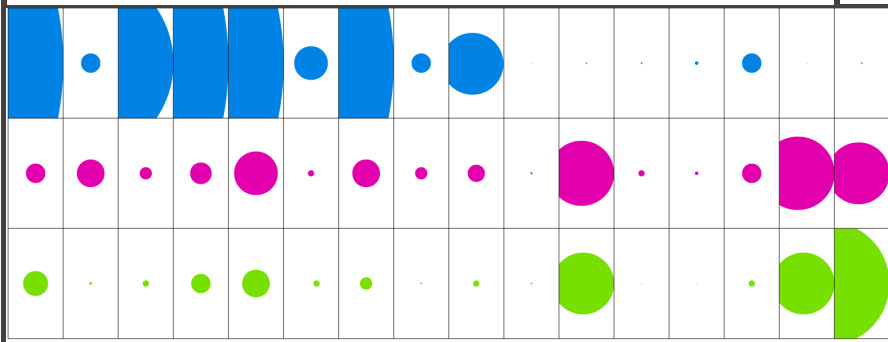
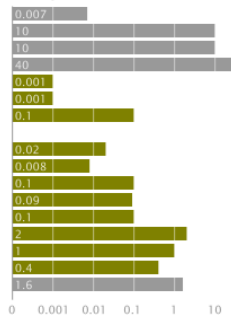


Minimum Inhibitory Concentration (MIC)

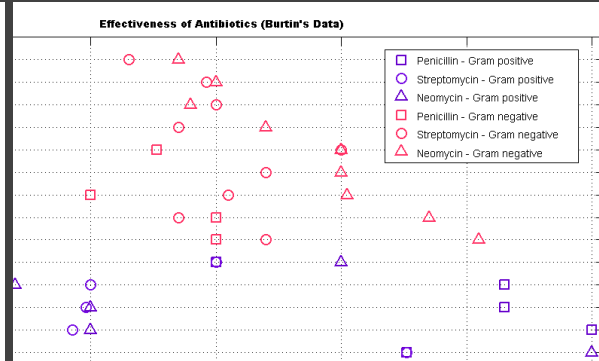
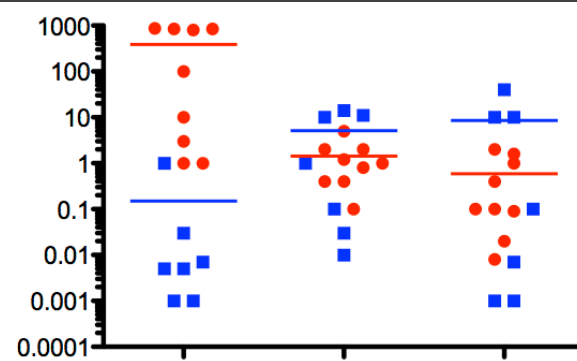
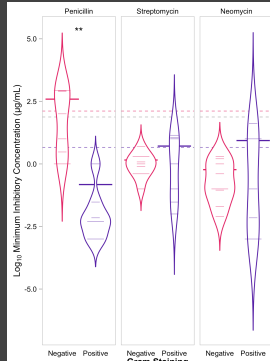
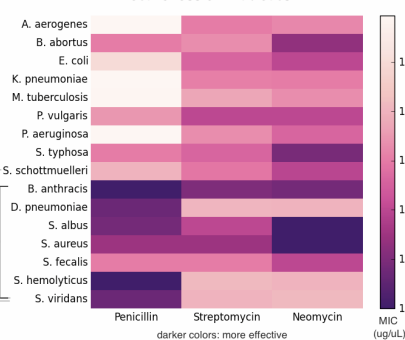
Streptomycin



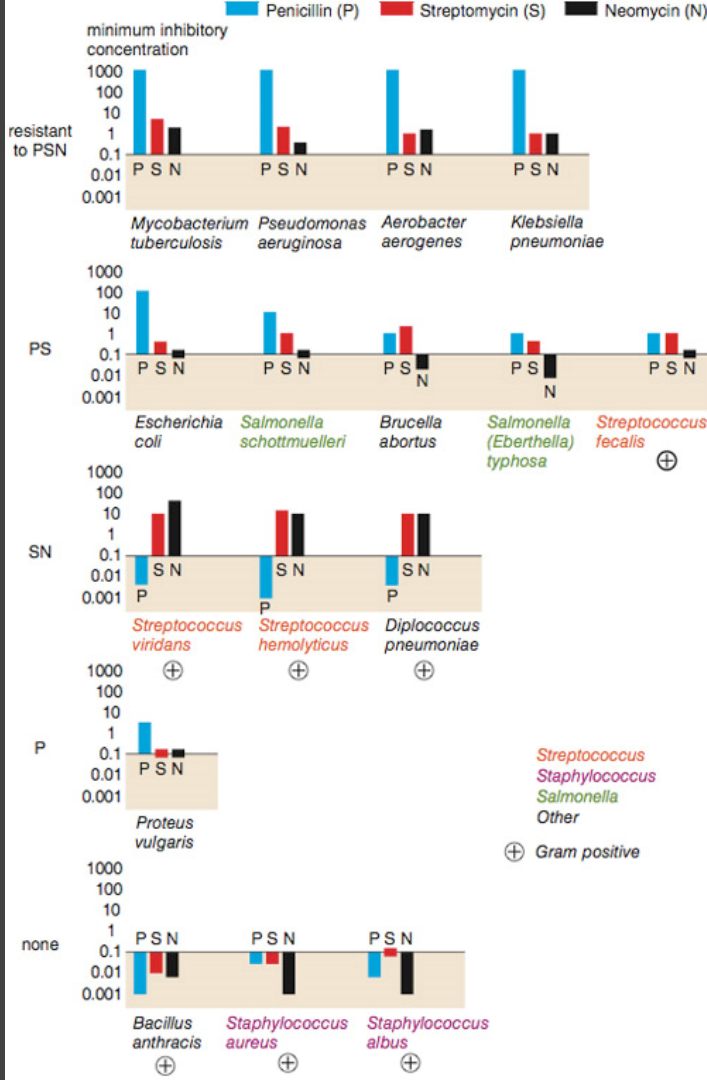
Neomycin



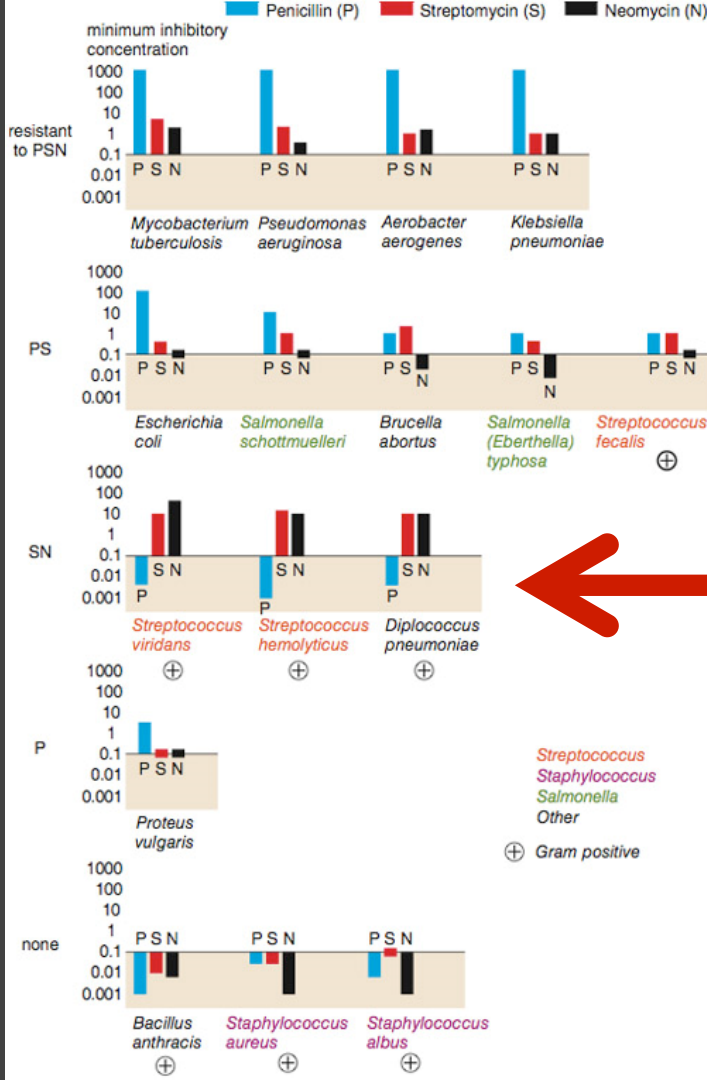
Effectiveness of Antibiotics



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



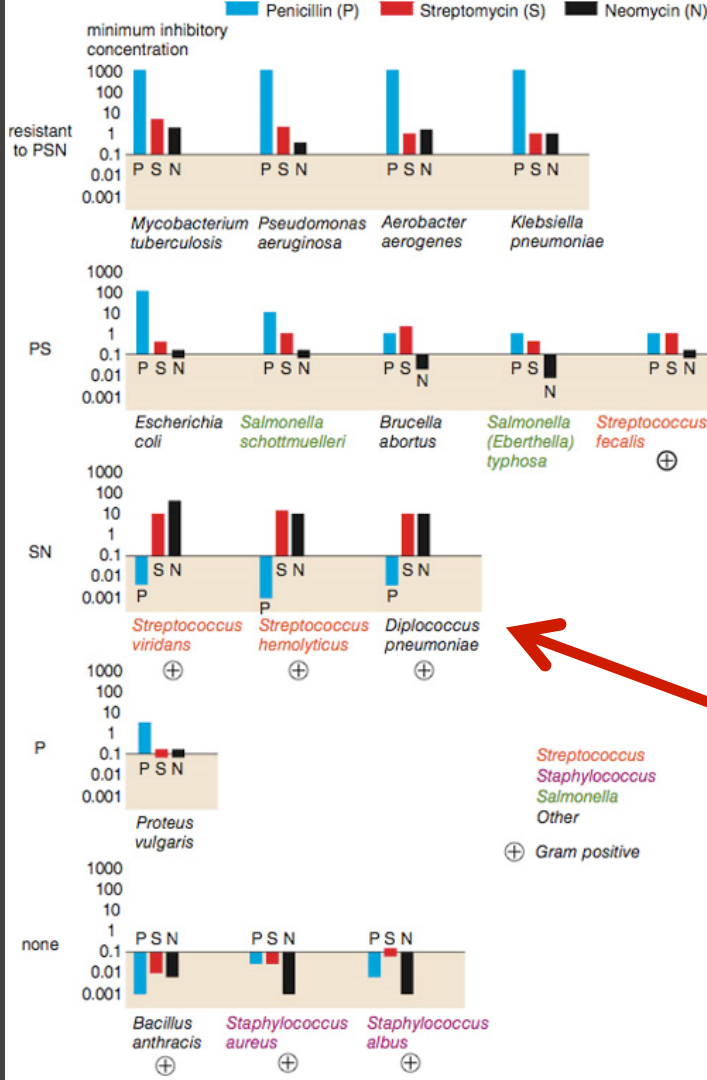
Do the bacteria group by antibiotic resistance?



Do the bacteria group by antibiotic resistance?

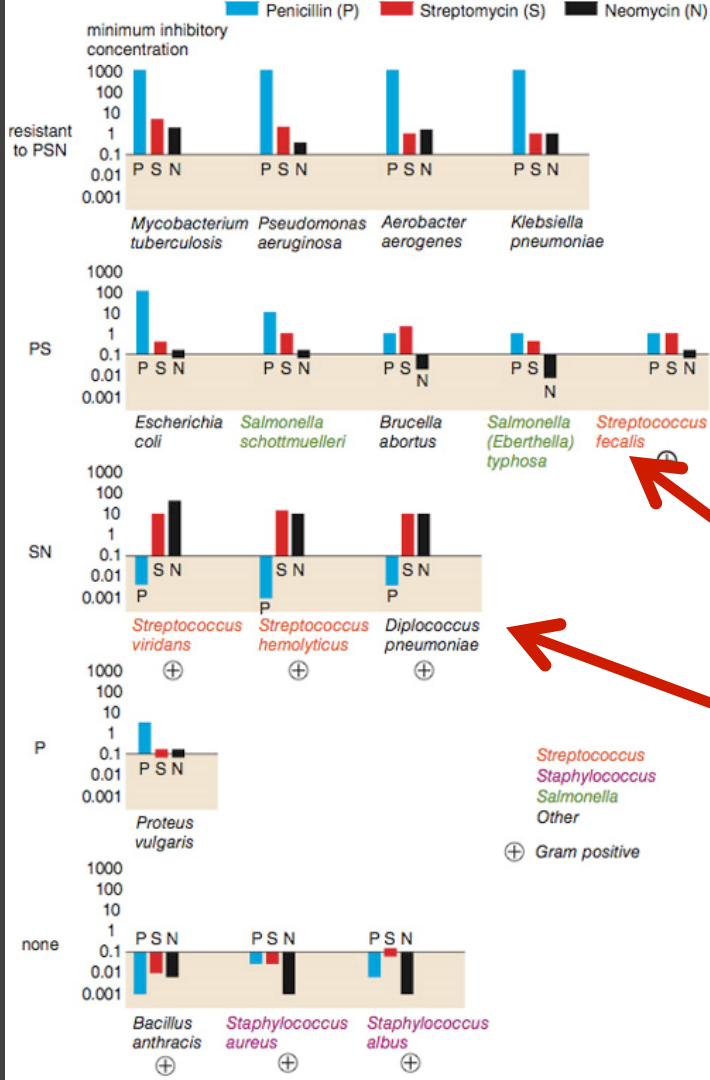


Wainer & Lysen
American Scientist, 2009



Do the bacteria group by antibiotic resistance?

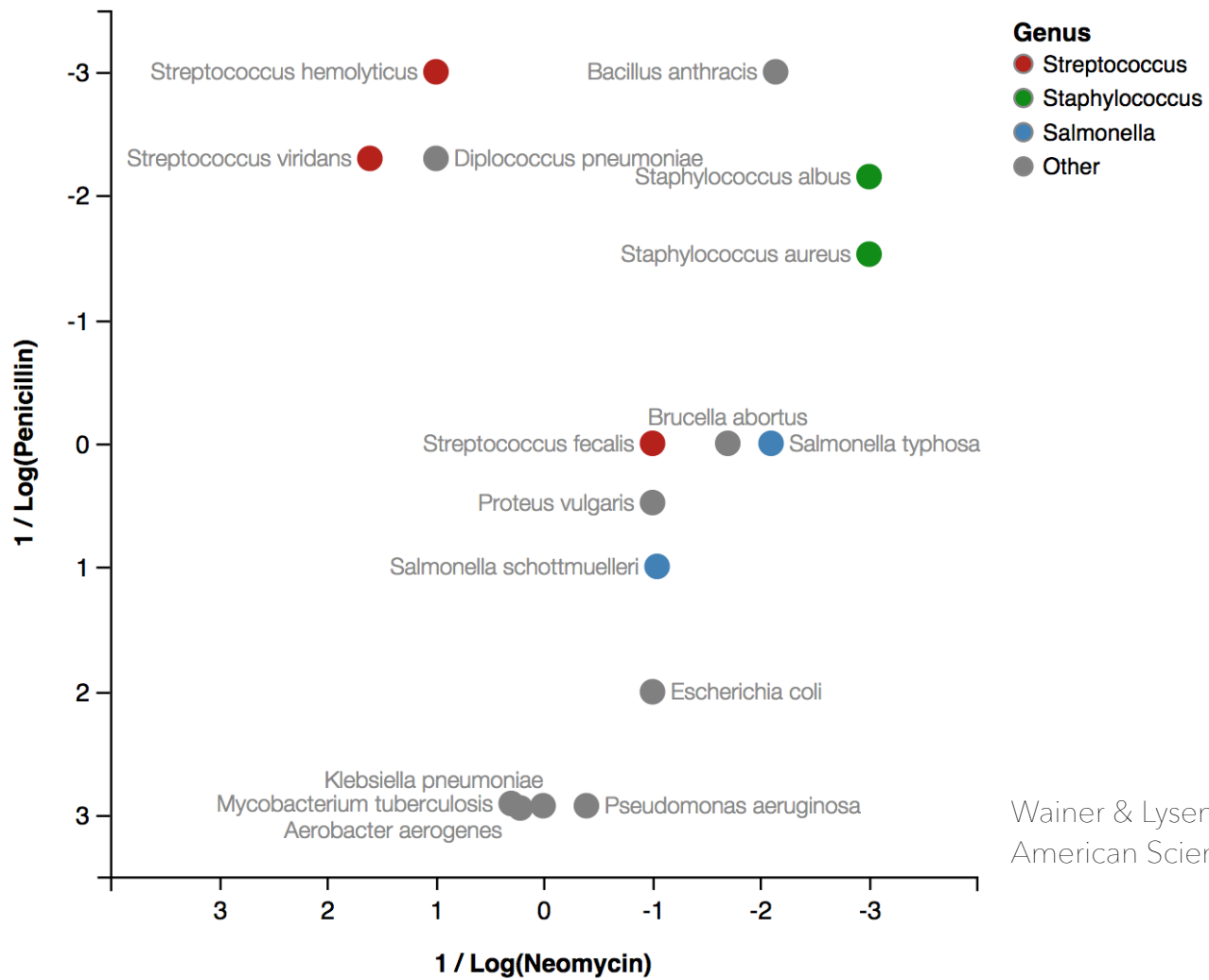
Really a streptococcus!
(realized ~20 yrs later)



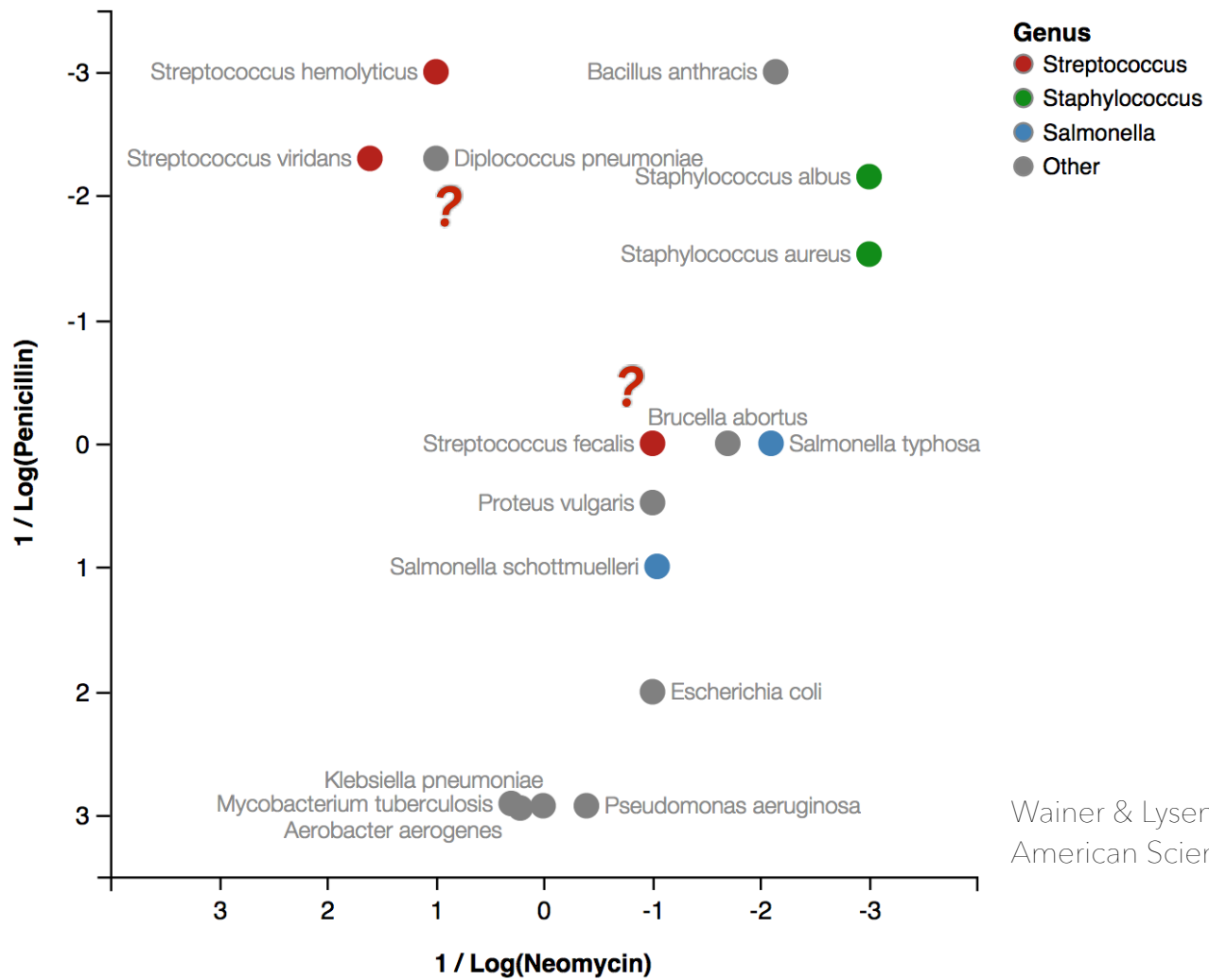
Do the bacteria group by antibiotic resistance?

Not a streptococcus!
(realized ~30 yrs later)

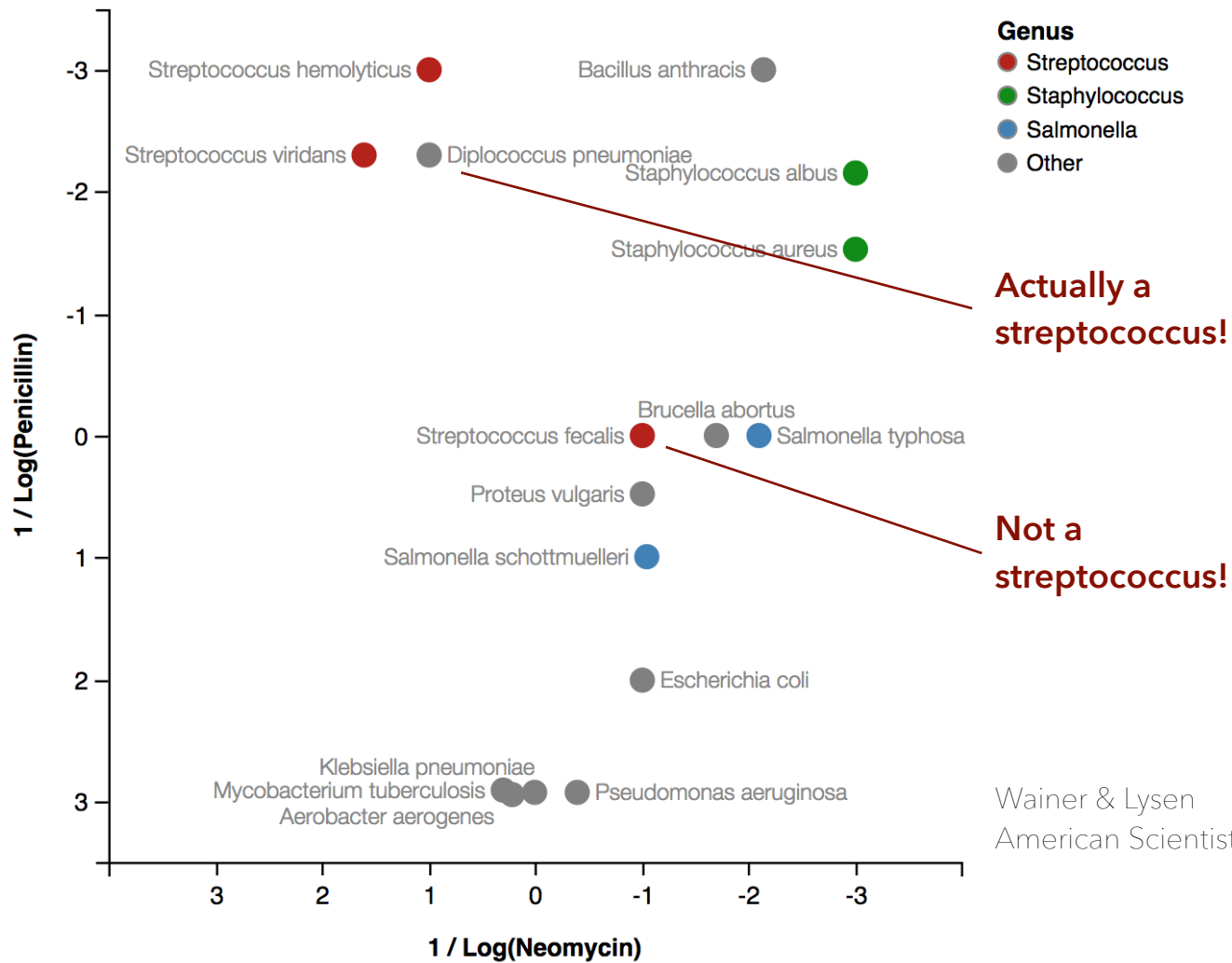
Really a streptococcus!
(realized ~20 yrs later)



Wainer & Lysen
American Scientist, 2009



Wainer & Lysen
American Scientist, 2009



Wainer & Lysen
American Scientist, 2009

Lesson: Iterative Exploration

Exploratory Process

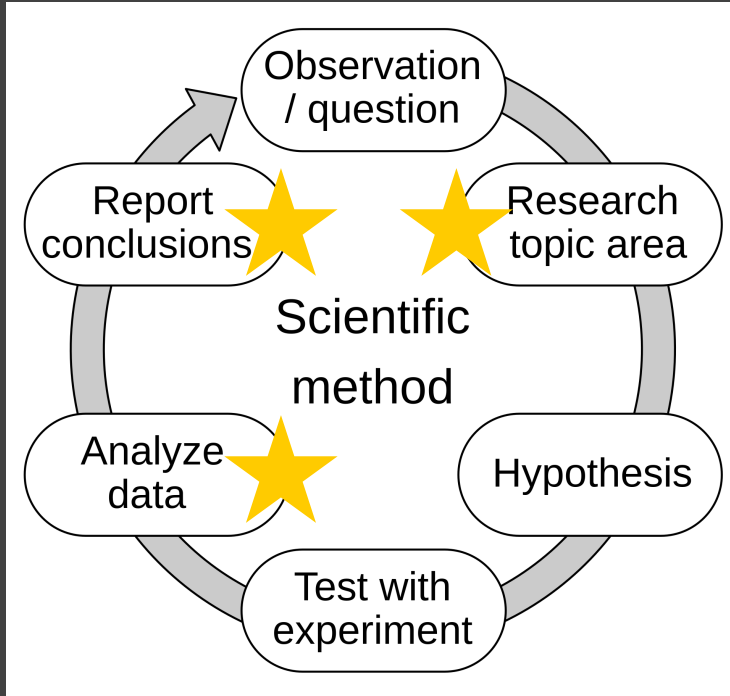
- 1 Construct graphics to address questions
- 2 Inspect “answer” and assess new questions
- 3 Repeat...

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

Formulate clear analysis questions & goals

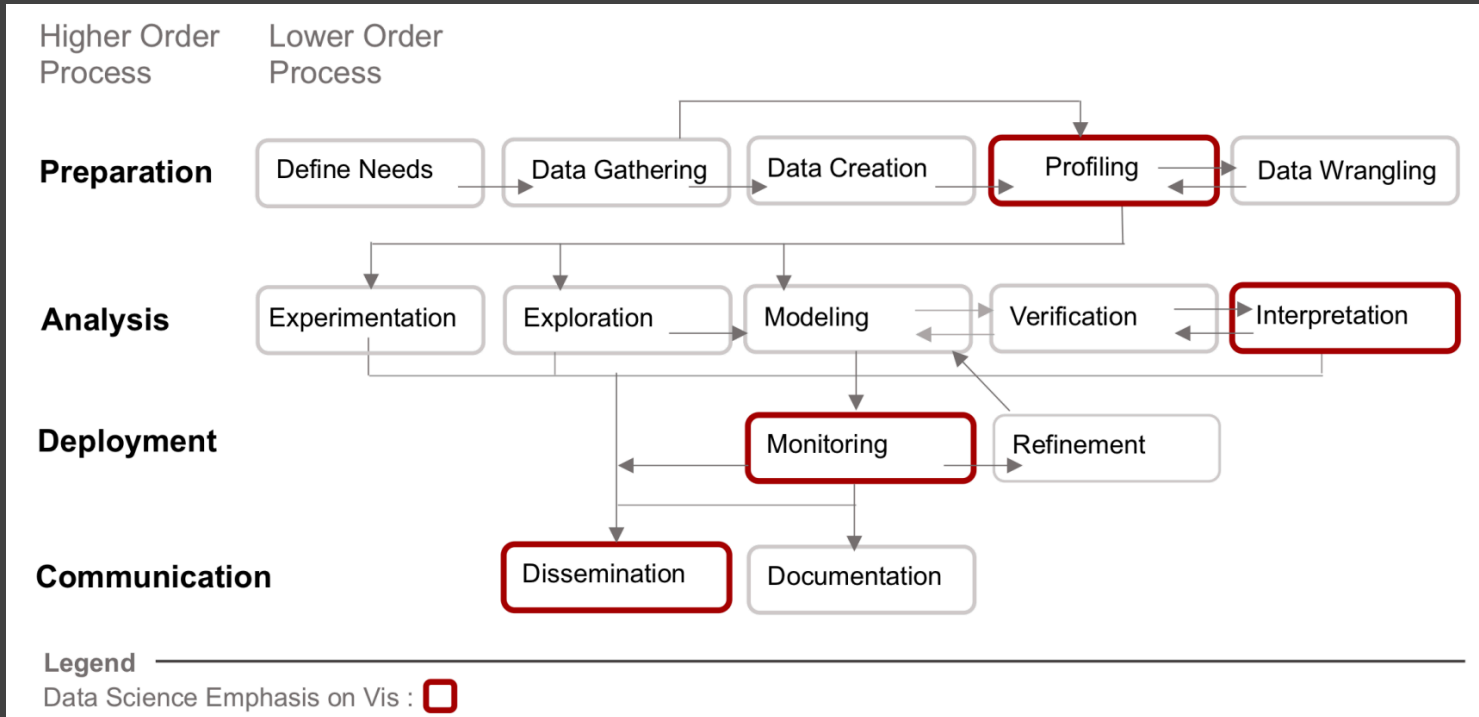
Don't trust your data!

Visualizations are Key to Advancing Science



- 1) Form a question/observe a phenomenon.
- 2) Gather information/resources.
- 3) Form a hypothesis.
- 4) Test the hypothesis with an experiment.
- 5) Analyze the results.
- 6) Interpret the data, report conclusions.

Visualizations are Key to Advancing Science

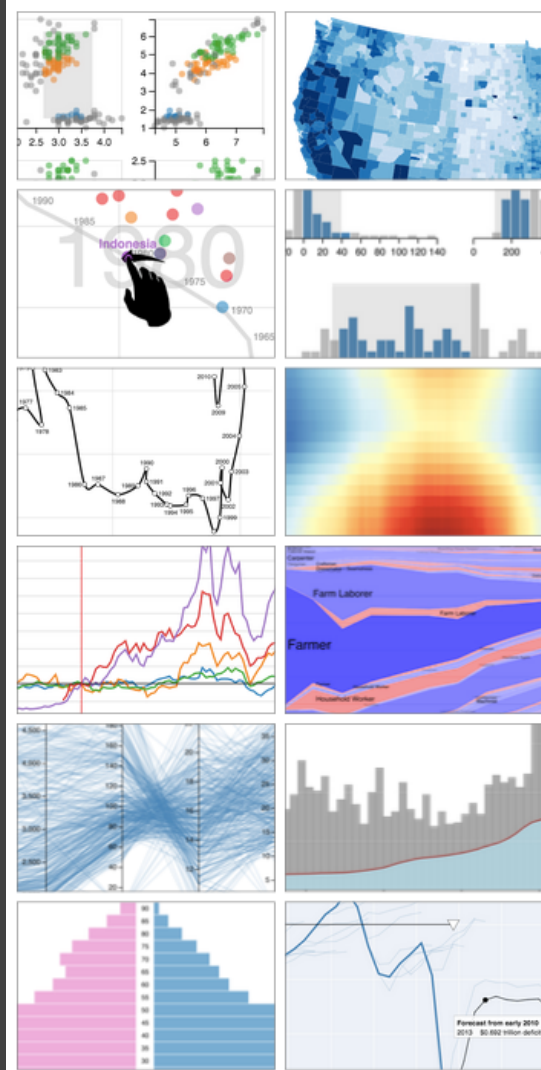


Exploration Tasks

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?



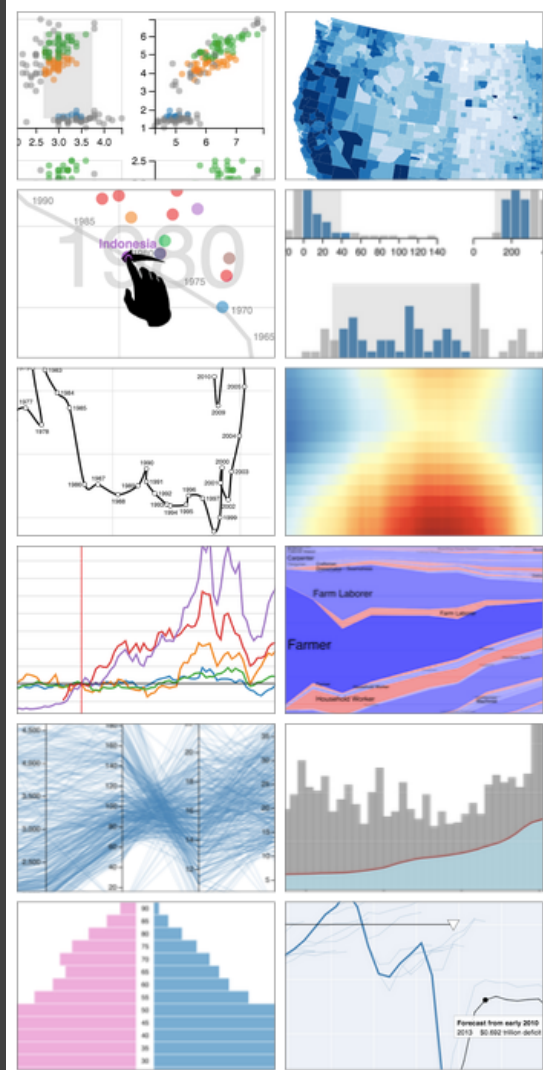
Data Exploration Tasks

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

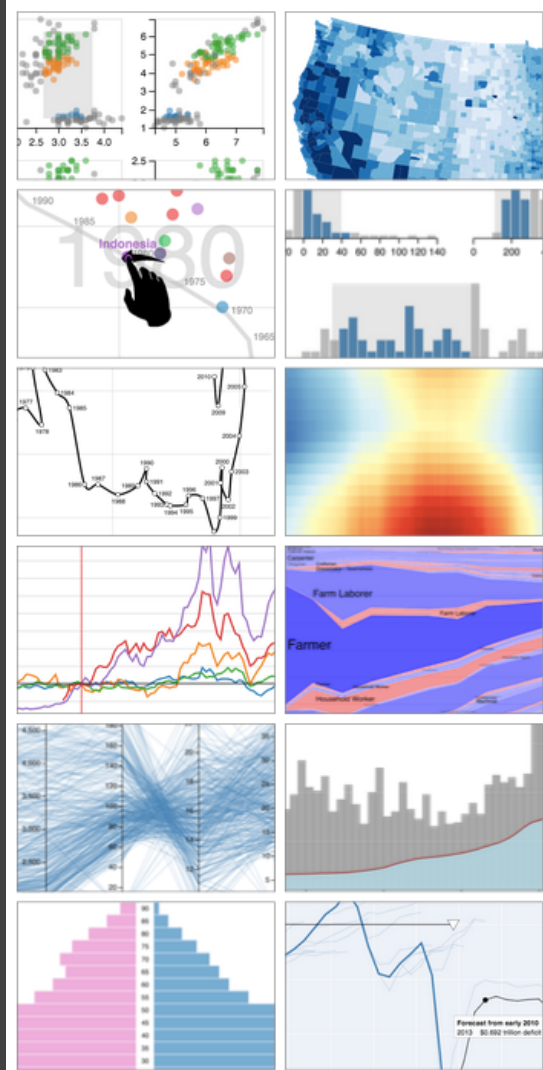
GOAL: Is the data actionable? What can we ask?

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

GOAL: Fact-finding, isolate important points/connections

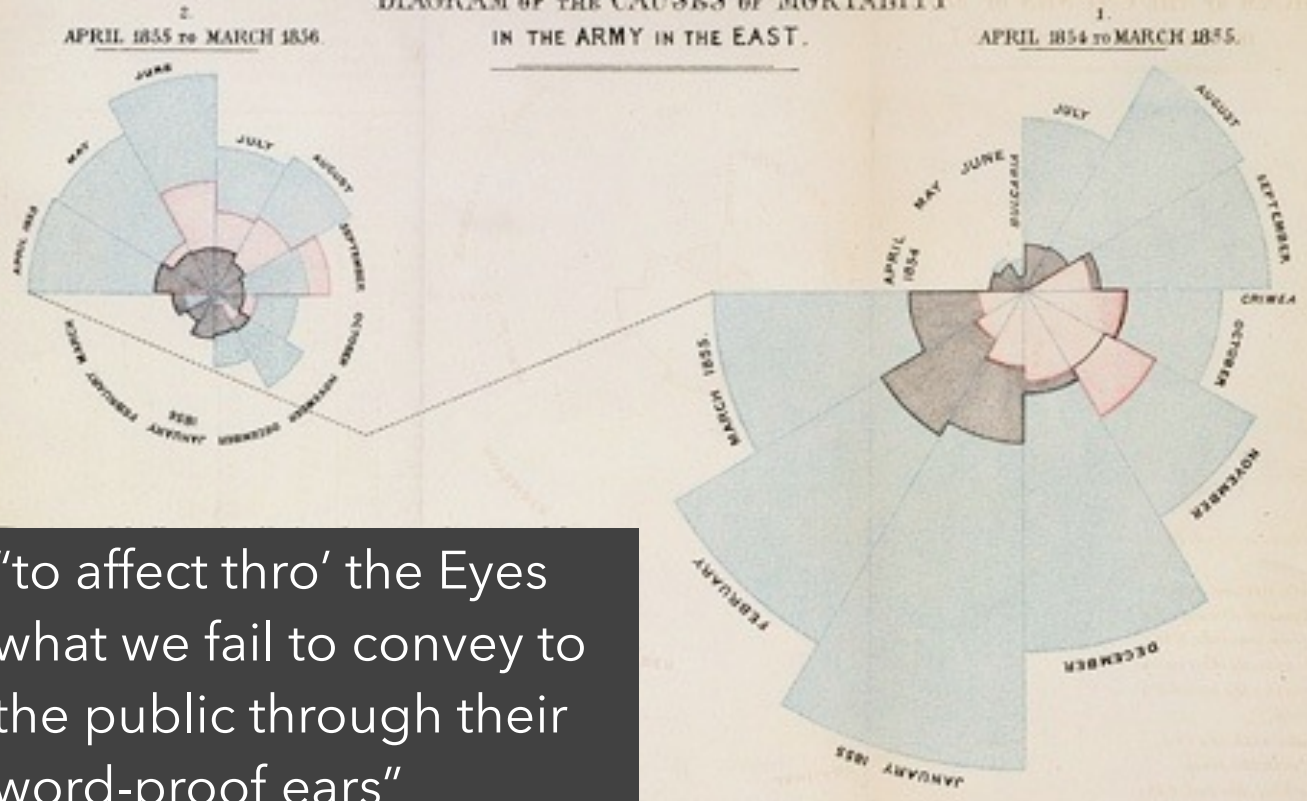
Infer: generalize from observed patterns, ascribe observations to specific factors or causes

GOAL: Inform modeling and decision making



Communication Tasks

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

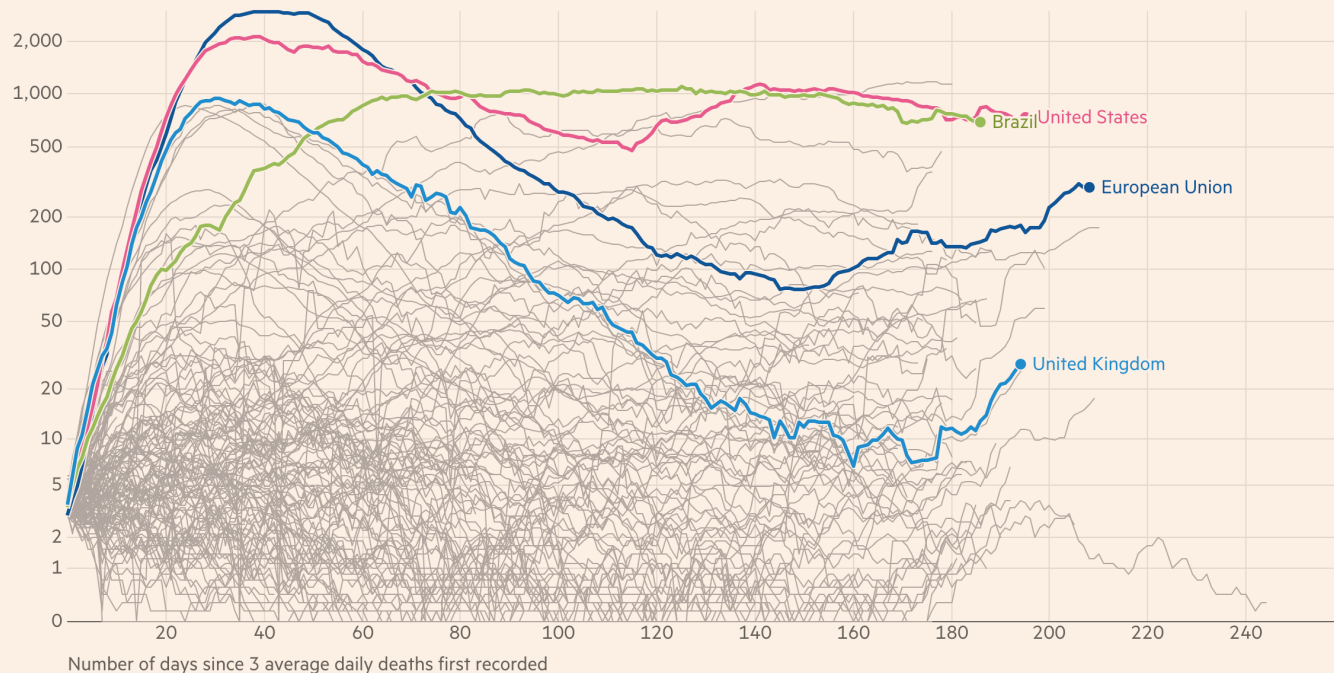


“to affect thro’ the Eyes
what we fail to convey to
the public through their
word-proof ears”

1856 “Coxcomb” of Crimean War Deaths, Florence Nightingale

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

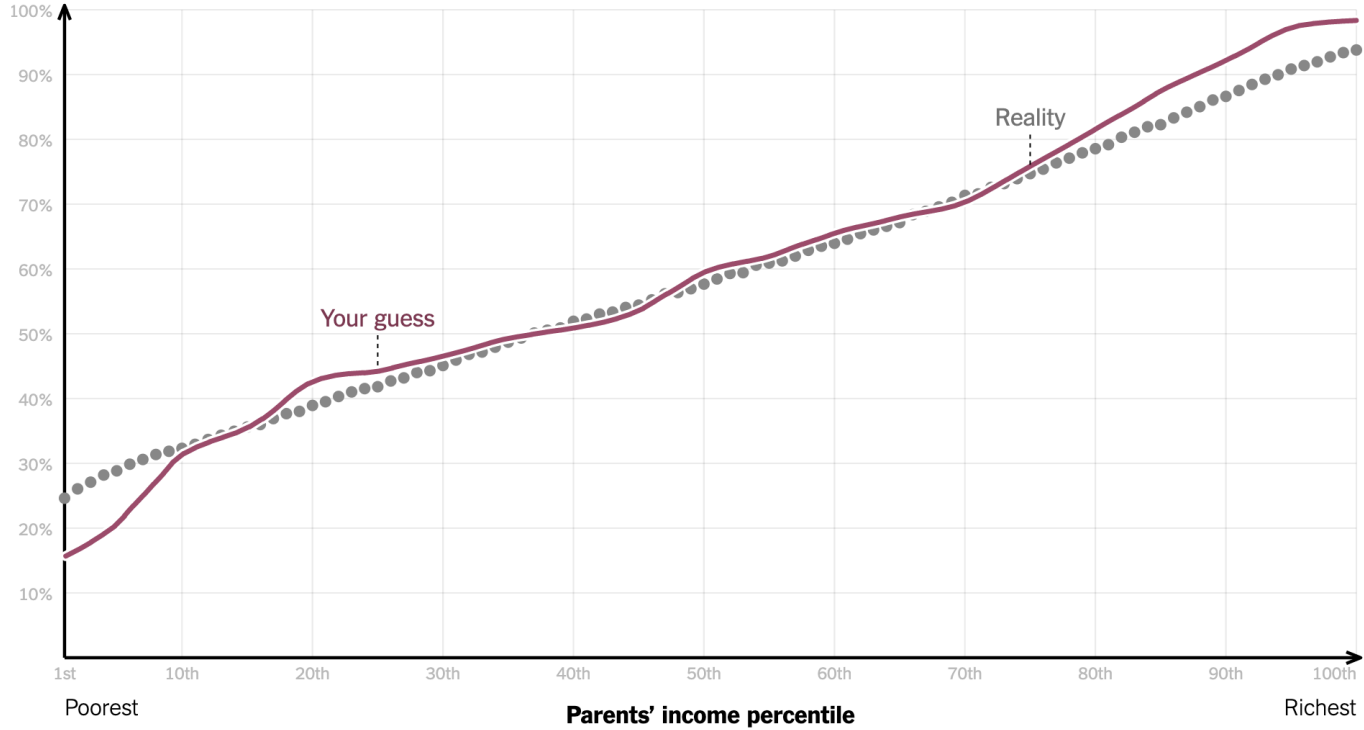
Seven-day rolling average of new deaths, by number of days since 3 average daily deaths first recorded



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

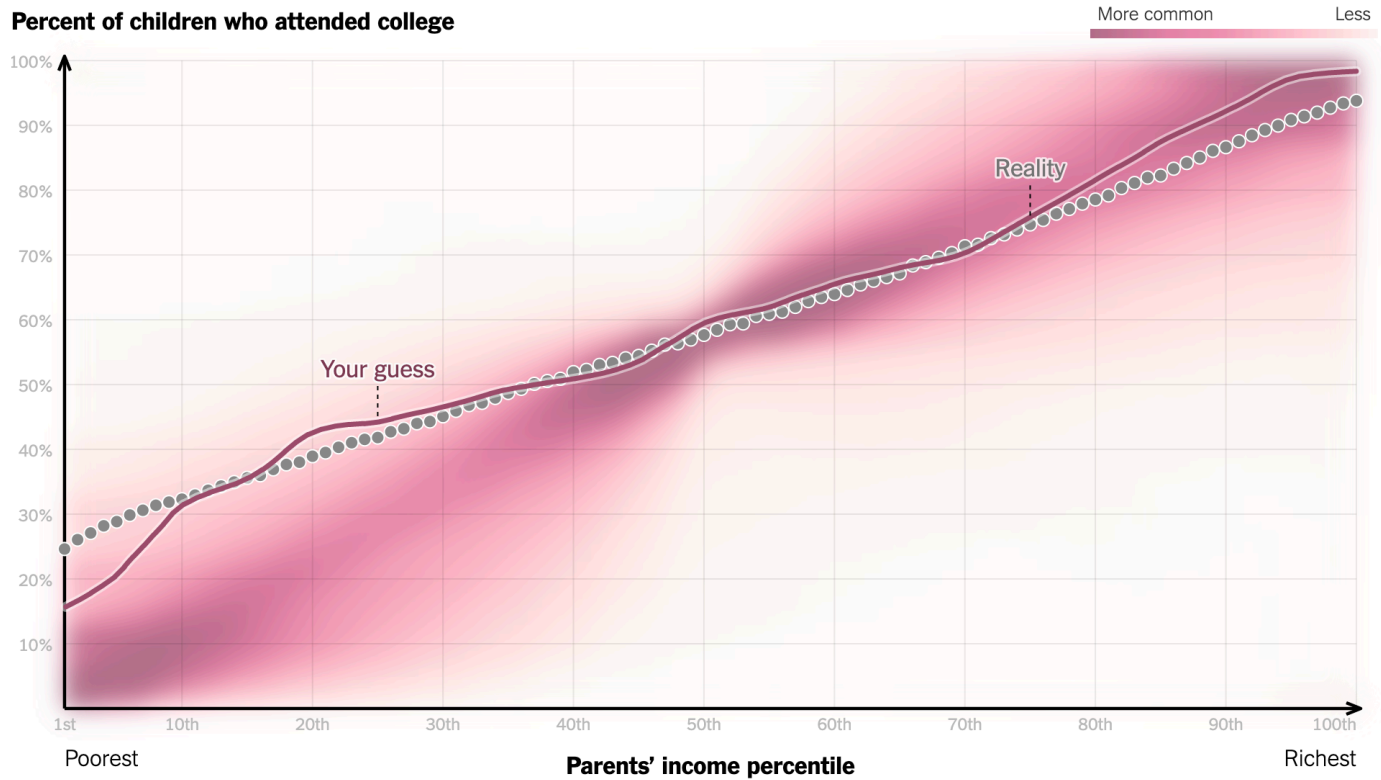
Percent of children who attended college



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 and Vega-Lite

W2: Visual Encoding & Deceptive Visualization

W3: Data Transformation & Dim. Reduction

W4: Interaction & Mapping

W5: Visualization Tools (D3.js)

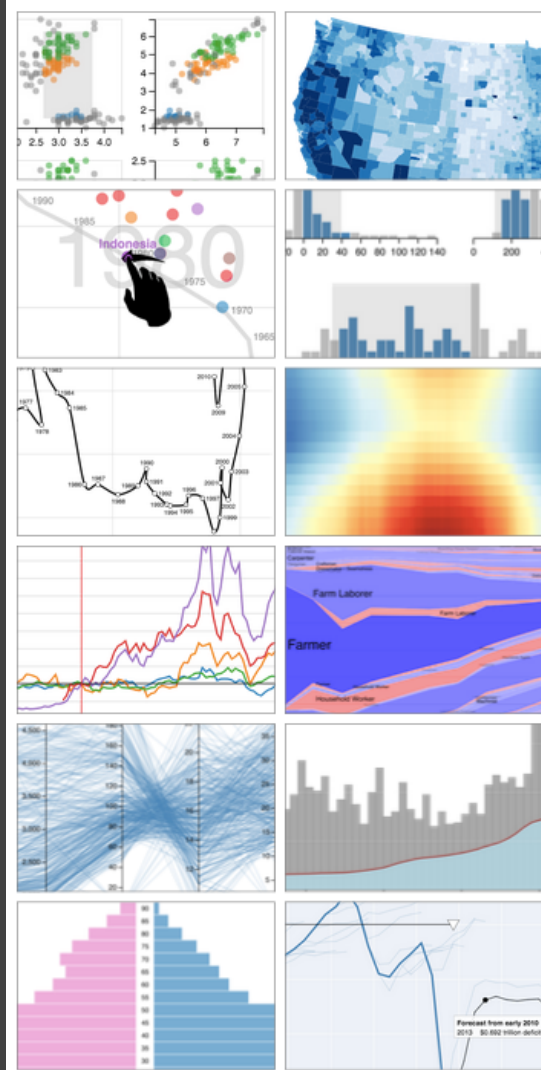
W6: Animation & Color

W7: Perception & Final Project Kick-Off

W8: Networks & Uncertainty

W9: Scalability & Final Project Review

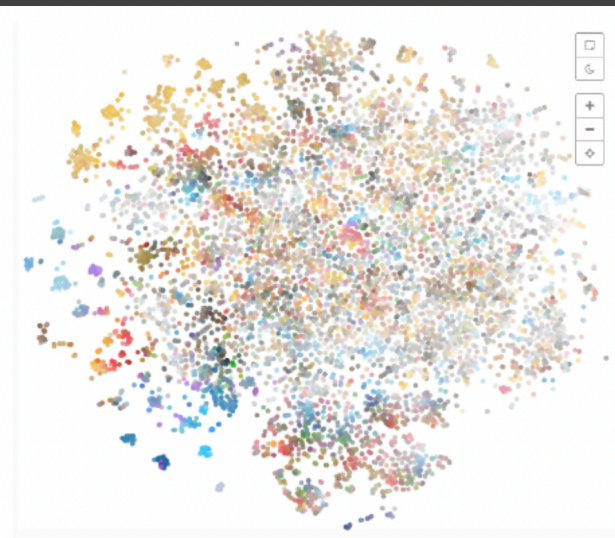
W10: Evaluation & Final Project Showcase



Visual Encoding

LES VARIABLES DE L'IMAGE										12	14
	POINTS			LIGNES			ZONES				
XY 2 DIMENSIONS DU PLAN	x	x	x	/	~	/	14 15 9 1 1 18 21 2 14 15 1	2 1 18 2 1 21 15 1 2 9	OQ	≠	
Z TAILLE	█	█	█	/	~	/	█	█	OQ	≠	
VALEUR	█	█	█	/	~	/	█	█	O	≠	
LES VARIABLES DE SÉPARATION DES IMAGES										13	
GRAIN	█	█	█	/	~	/	█	█	≡	○	≠
COULEUR	█	█	█	/	~	/	█	█	≡	≠	
ORIENTATION	█	█	█	/	~	/	█	█	≡	≠	

Data Transformation



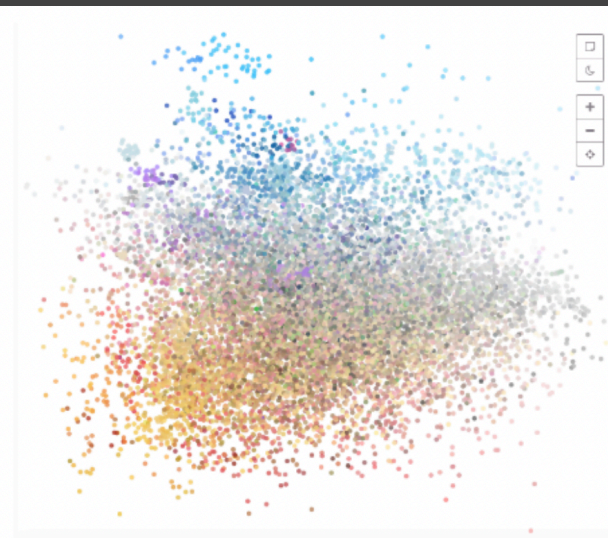
Q Latent Dimensions: 32 ▾ Projection: t-SNE ▾ Perplexity: 30 ▾

t-SNE



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

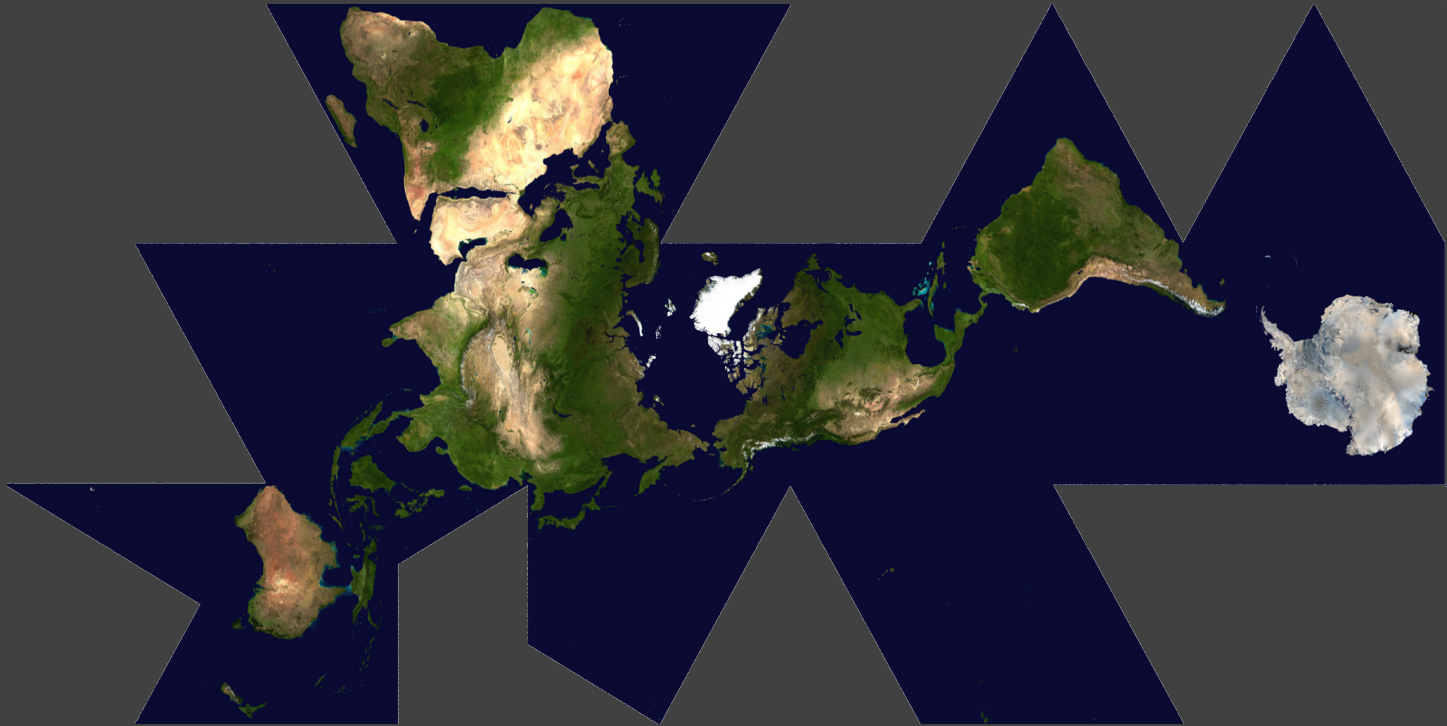
UMAP



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

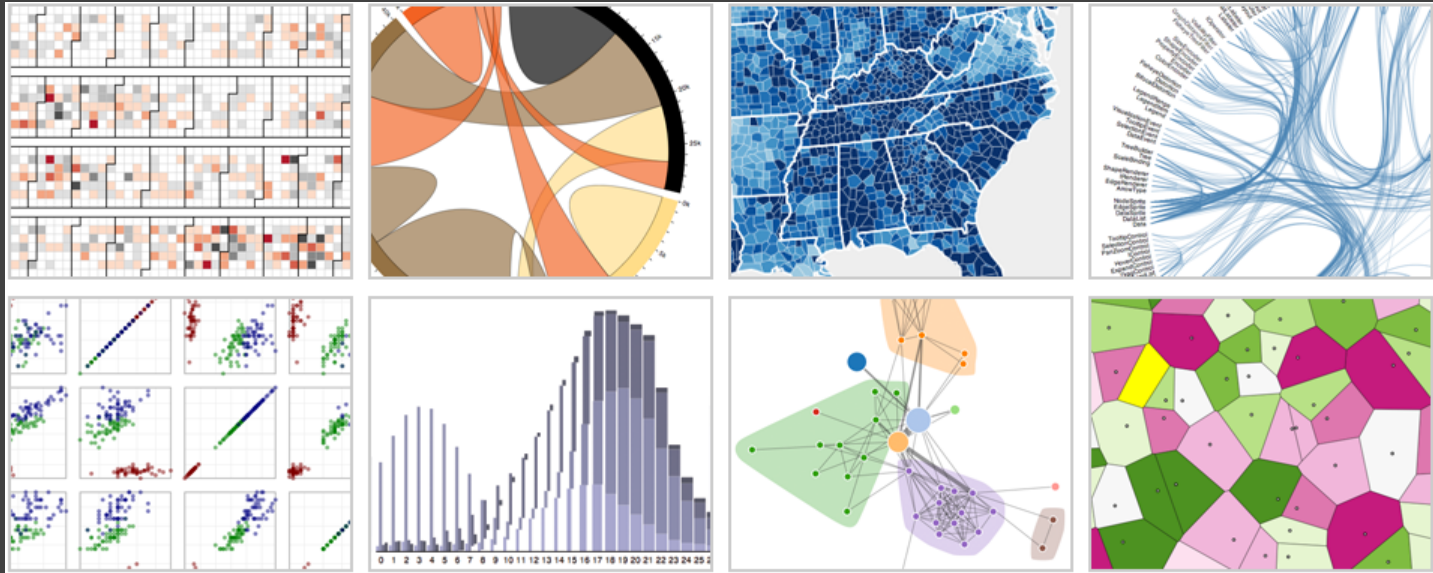
PCA

Mapping & Cartography



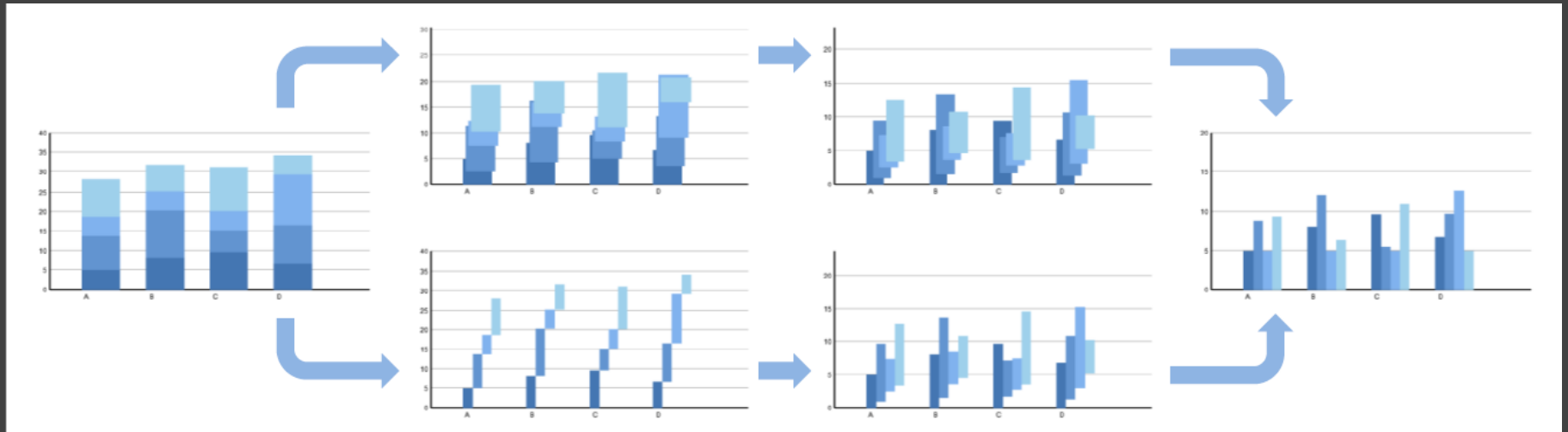
Dymaxion Maps [Fuller 46]

Visualization Tools



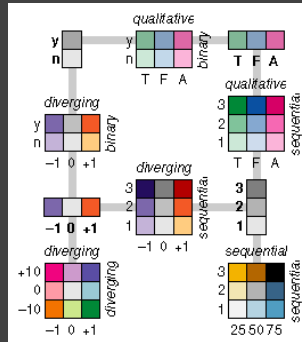
D3: Data-Driven Documents

Animation

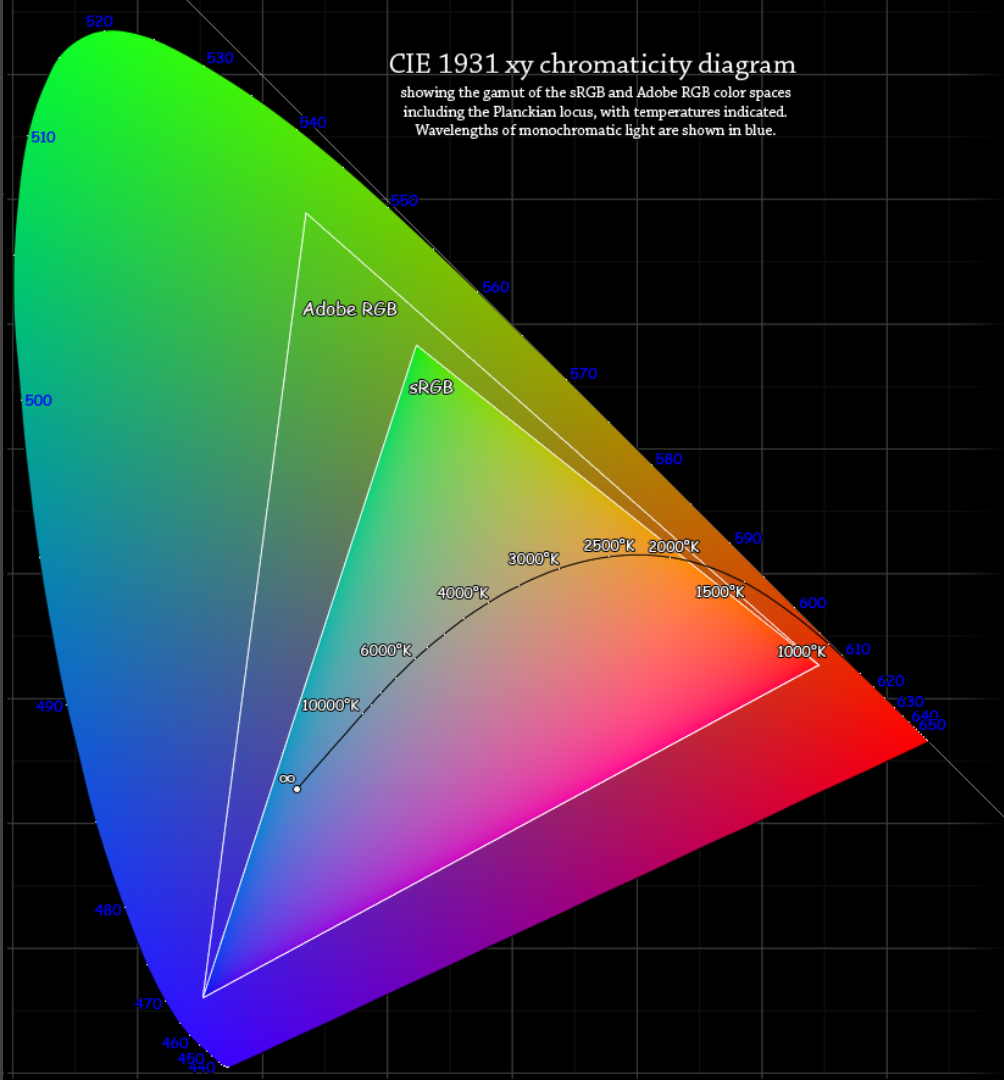


Animated transitions in statistical data graphics [Heer & Robertson 07]

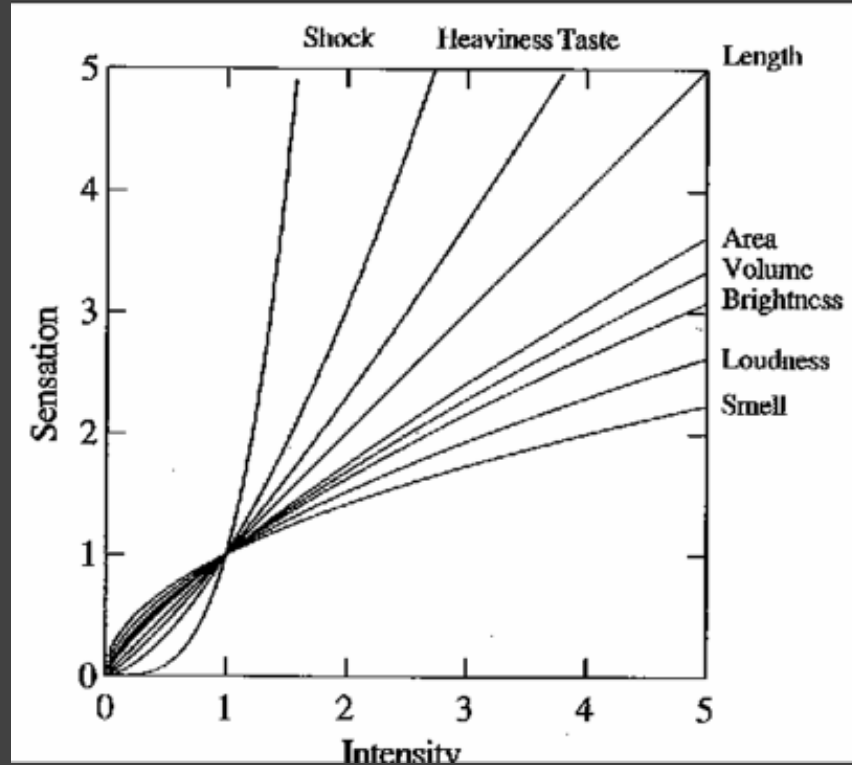
Color



Color Brewer

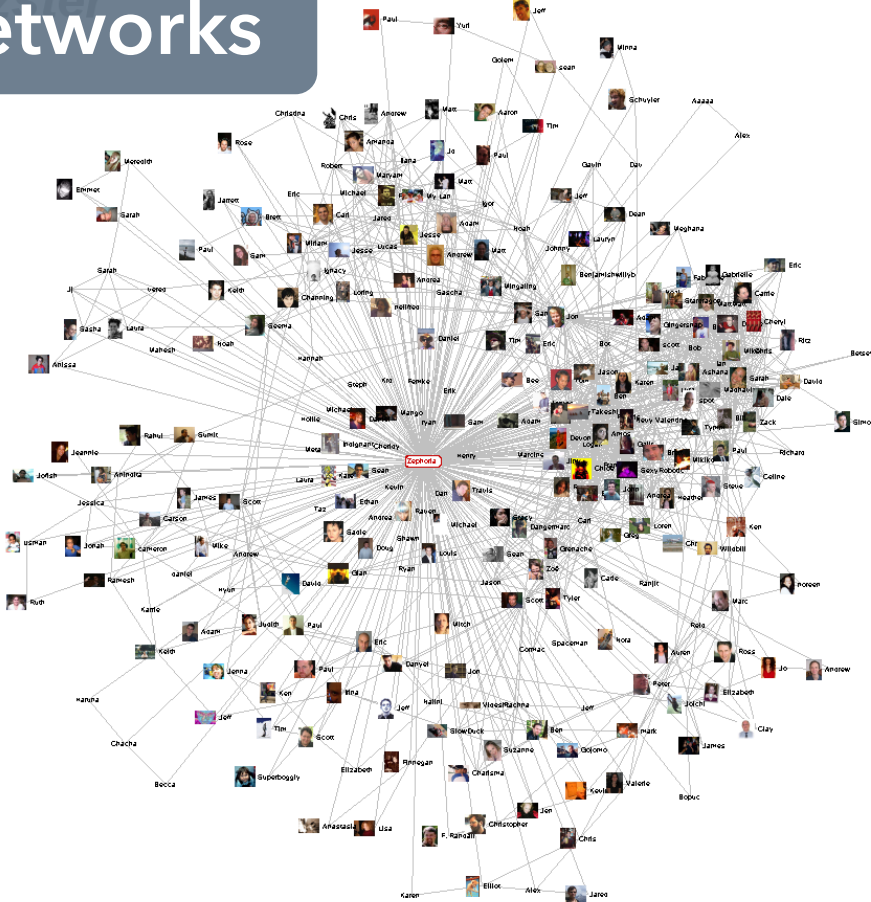


Graphical Perception



The psychophysics of sensory function [Stevens 61]

Networks



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/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 Dont Cry

Member Since ??

Last Login 2003-10-21

Last Updated 2003-10-21

About

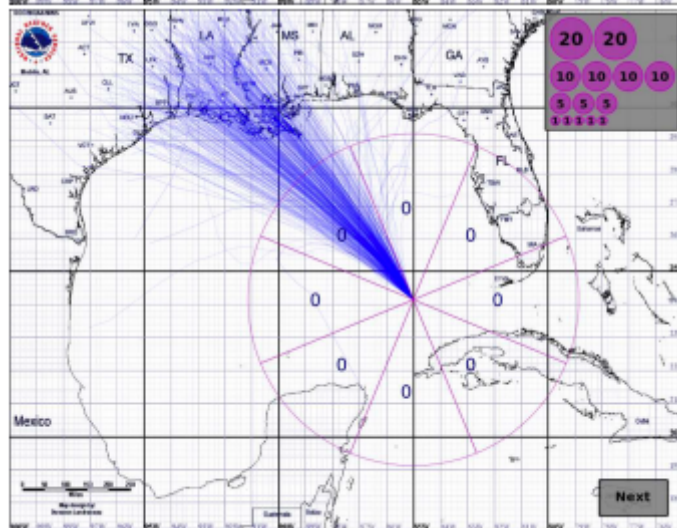
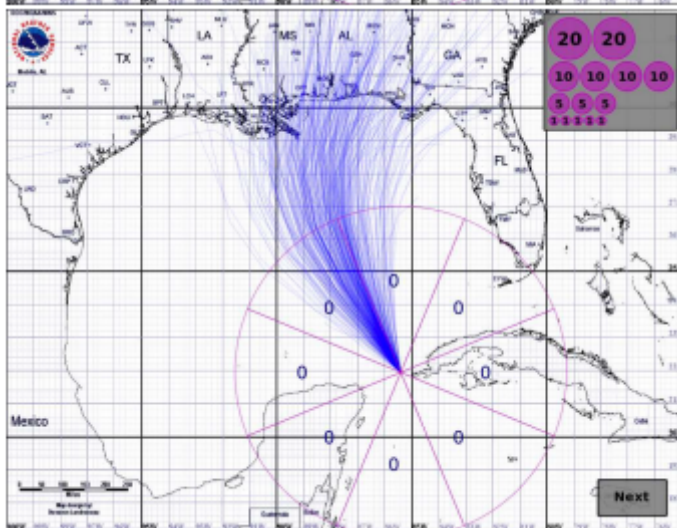
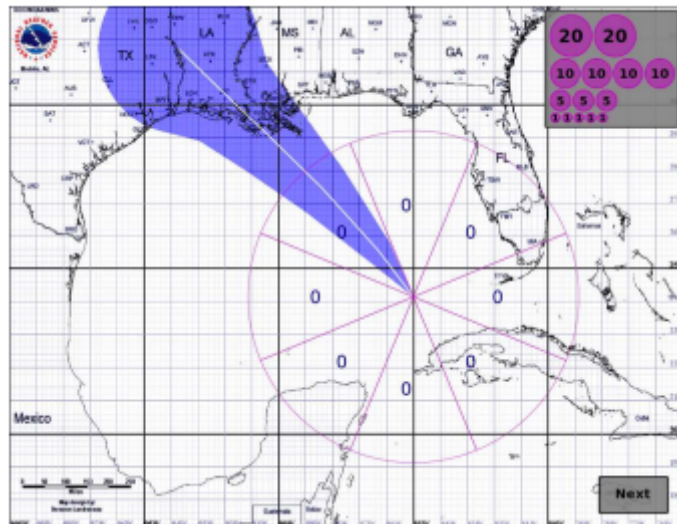
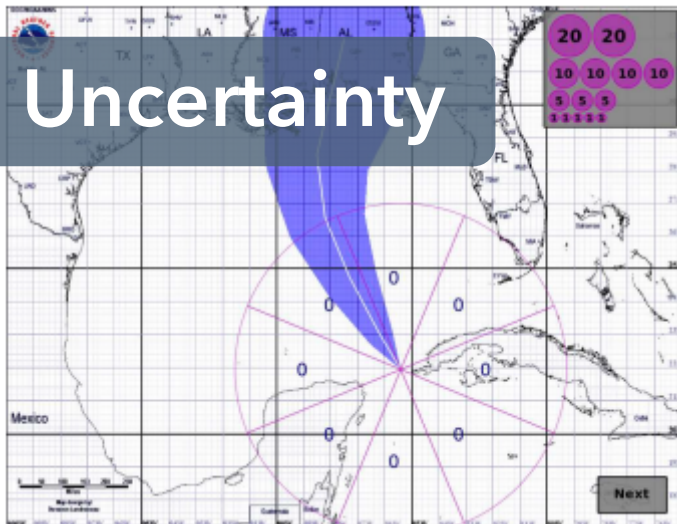
I'm a geek, an activist and an academic, fascinated by people and society. I see life as a very large playground and enjoy exploring its intricacies. I revel in life's chaos, while simultaneously providing my own insane element.

My musings:
<http://www.zephoria.org/thoughts/>

Want to Meet

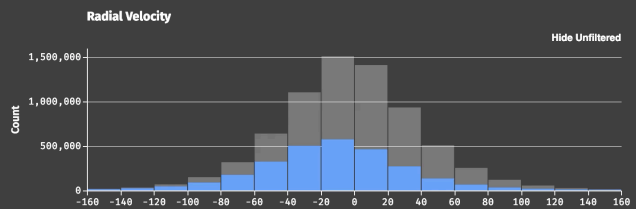
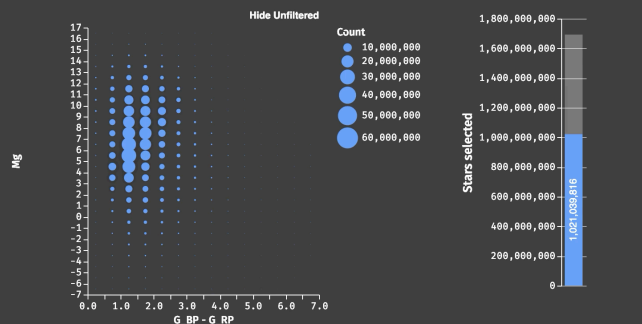
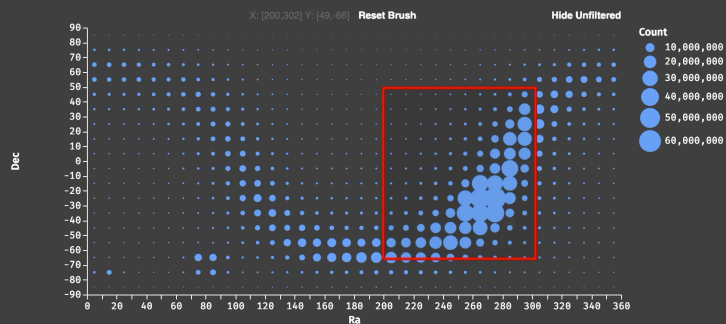
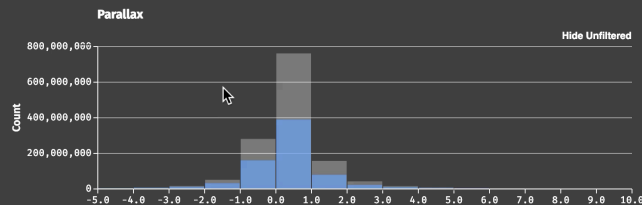
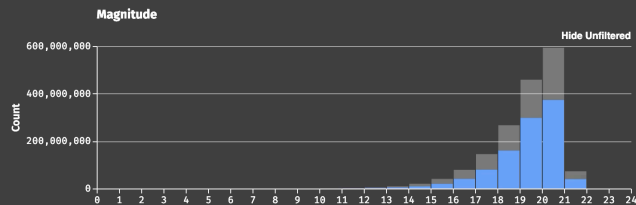
Someone who makes life's complexities seem simply elegant.

Uncertainty



Scalability

localhost:1234



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

cse512@cs.washington.edu

Lead Instructor

Leilani Battle

Teaching Assistants

Katherine Juarez

Ameya Patil

Luke Snyder

All office hours are by appointment this quarter.

We will focus instead on Q&A via Ed!



Leilani Battle (she/her)

Associate Professor, UW CSE

Co-Director, CSE Interactive Data Lab

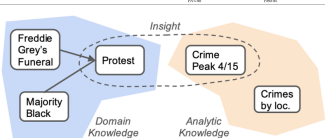
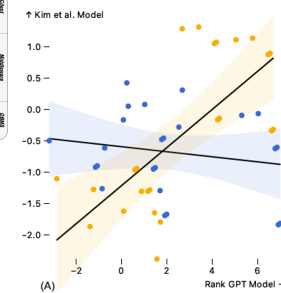
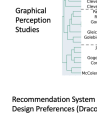
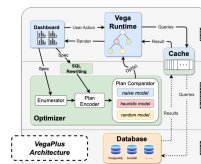
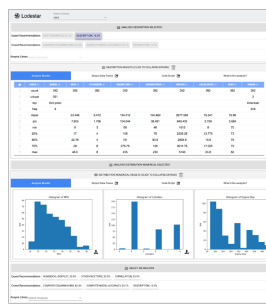
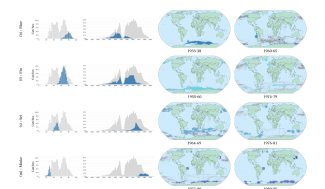
<https://homes.cs.washington.edu/~leibatt/>

Visualization / HCI / Data management / Data Science

I model how people interact with data analysis systems.

I use these models to build *behavior-driven* optimizations, UI features, and performance benchmarks for interactive data analysis

Hobbies: disc golf, reading, cooking, travel, video/board games.



Katherine Juarez

3rd-ish year PhD student

Research Interests: HCI, Technology use in Latin America

Hobbies:

- Kayaking
- CrossFit
- Hiking
- Learning new languages



Ameya Patil

6th Year PhD advised by Dr. Battle

Research Interests: Interactive Network Data Analytics, Visualization
Perception, Applications in Environmental Sciences

Hobbies: Photography, Music, Sports and exercise, Exploring the great outdoors

Website: <https://ameyabp.github.io/>



Luke Snyder

5th year CSE PhD

Research

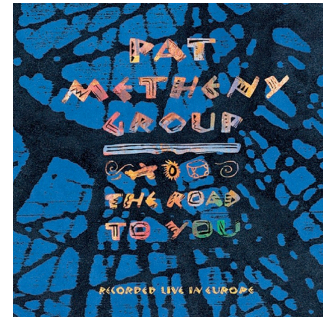
- Data visualization & HCI
- Tools for interactive visualization
- Human-AI knowledge discovery & synthesis

Technical Expertise

- JavaScript / web programming, D3, Vega-Lite

For Fun

- Music, running, travel
- Progressive Jazz, Weather Report, Pat Metheny Group



Lectures, Activities & Office Hours

Course sessions will alternate among lecture and in-class exercises. Thursdays will typically be activity days.

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!

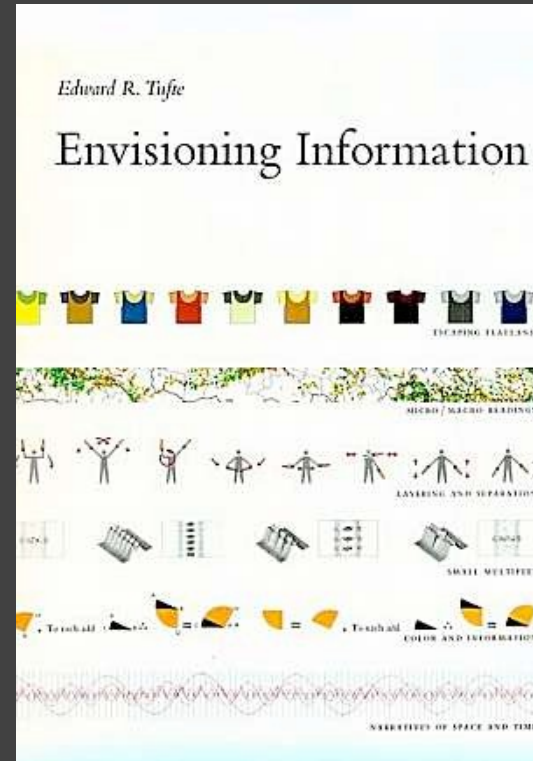
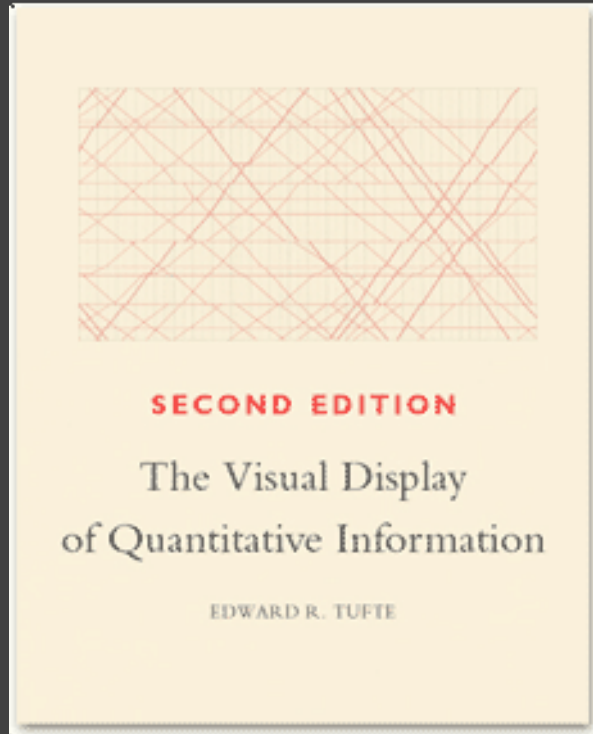
Readings

There is no one universal textbook on visualization!
So we will draw on books, notebooks, and linked articles...

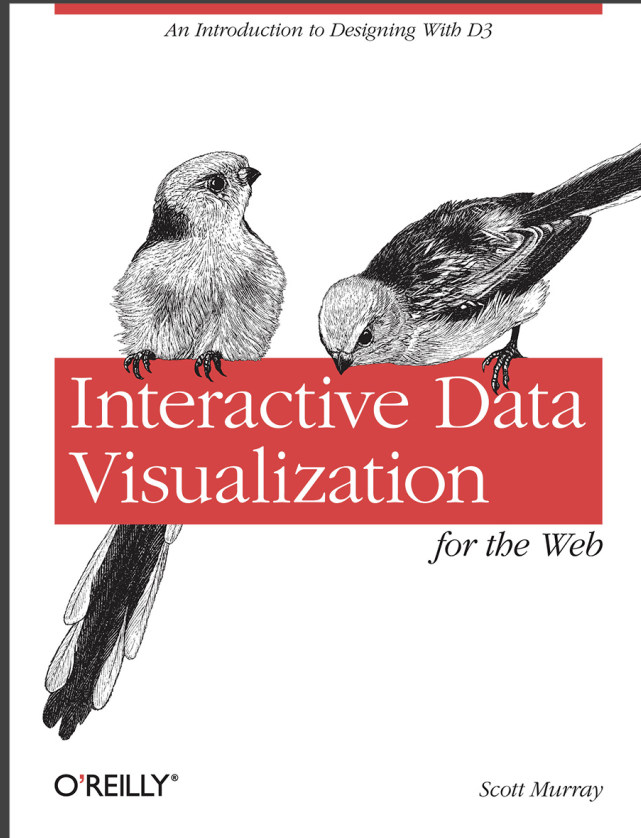
Material in class will loosely follow readings.

Readings should be read by start of class.

"Textbooks"



Optional Book



Interactive Data Visualization for the Web, 2nd Edition

For learning D3!

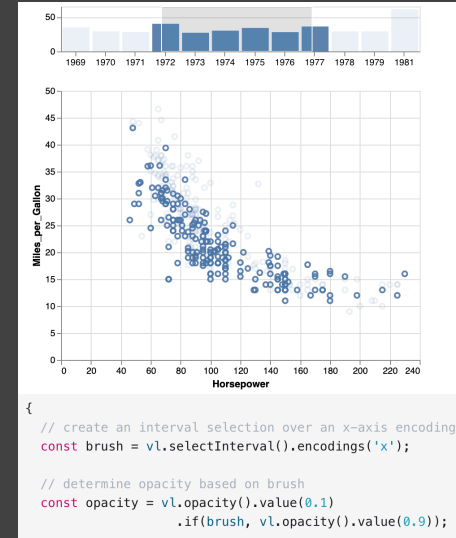
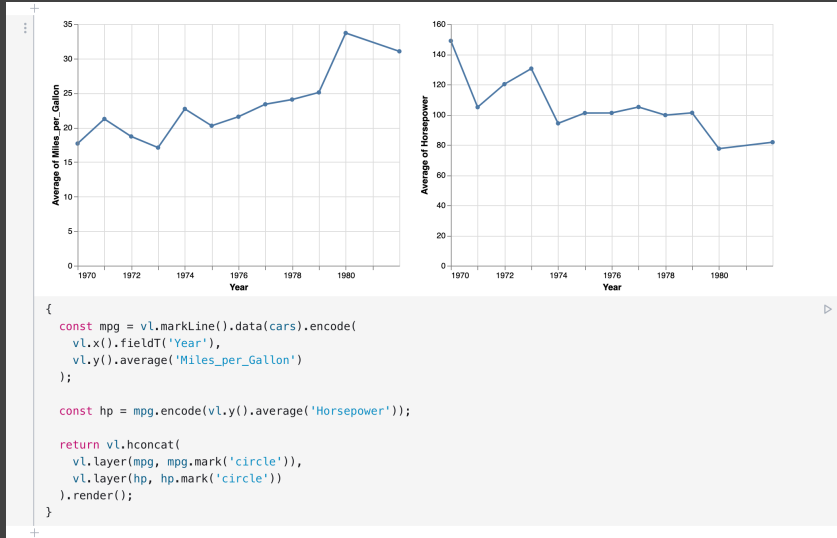
Book available online.

Code / examples on GitHub.

We will be using D3 v7.

<https://d3js.org>

Interactive Notebooks



Hands-on engagement with course concepts and visualization tools (Vega-Lite / Altair), in both JavaScript (Observable) and Python (Jupyter).

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-.

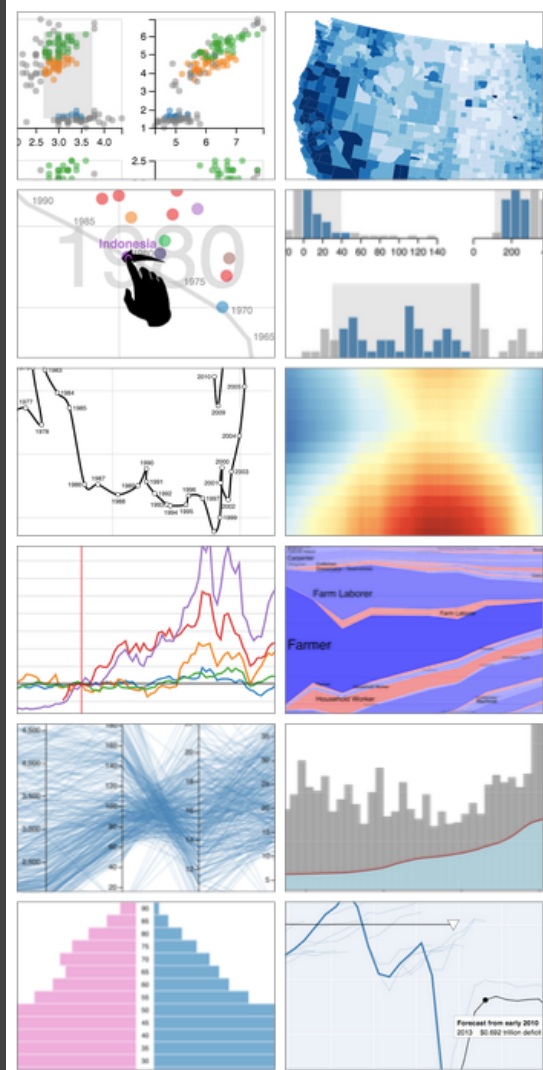
Course Participation

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?

Questions?