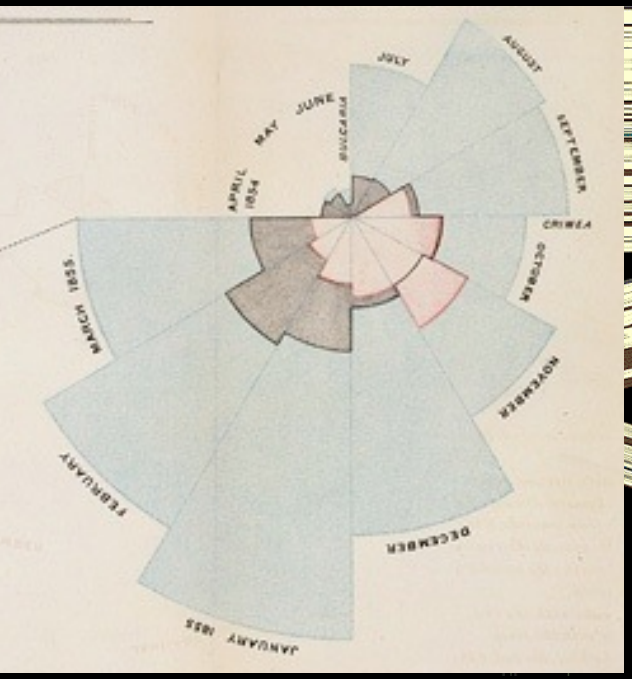
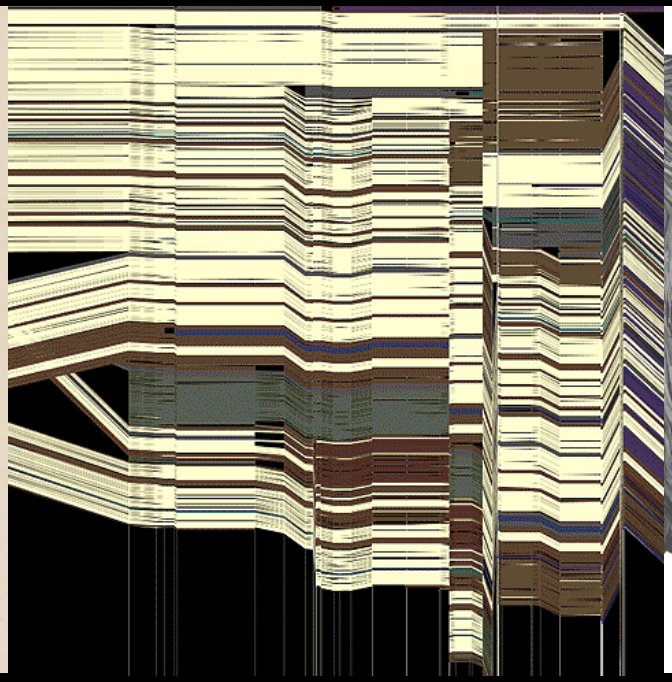
CSE 442 - Data Visualization

The Value of Visualization





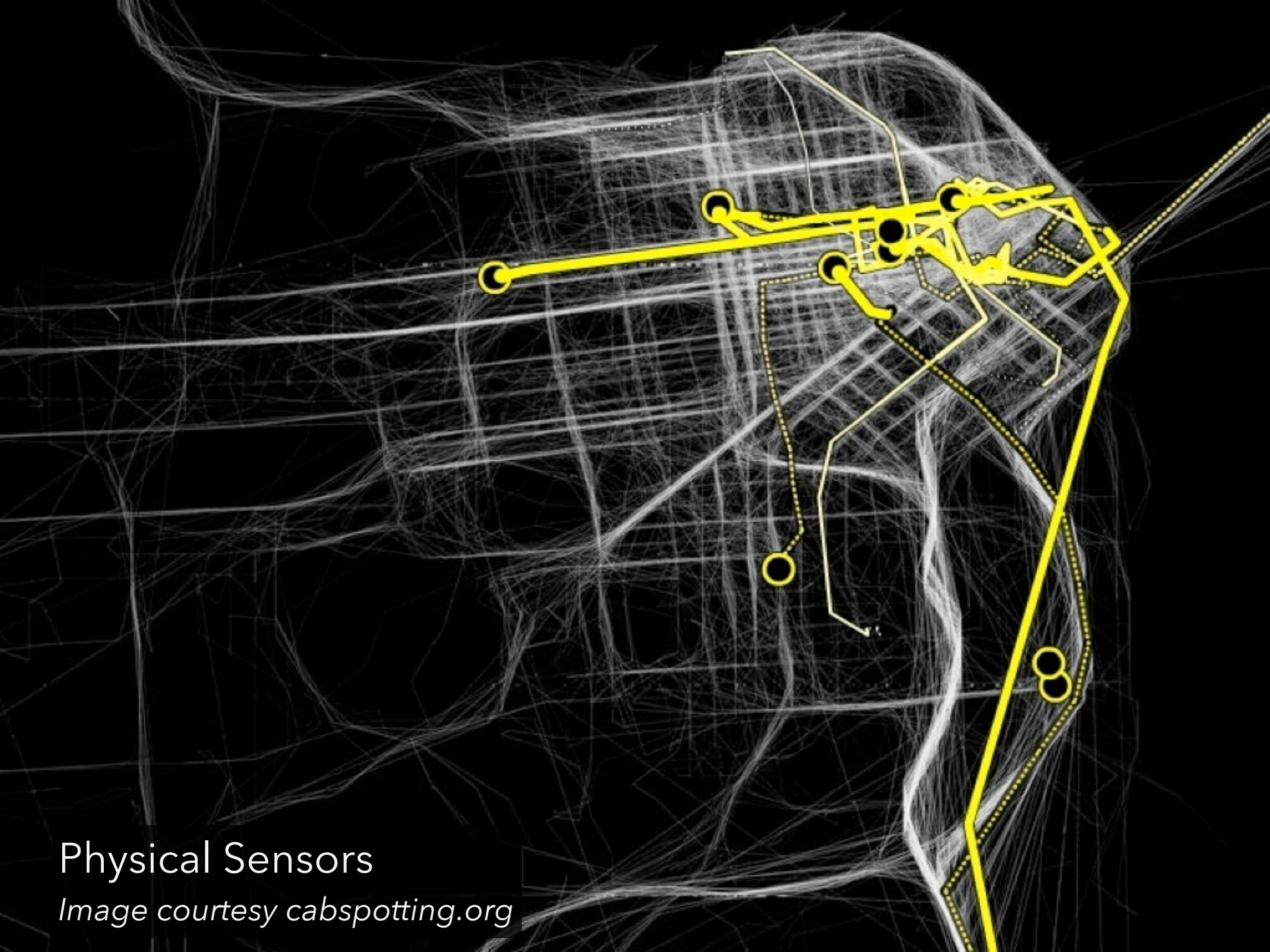
Matthew Conlen University of Washington

How much data (bytes) did we produce in 2018?

2018: 33 zettabytes

Up from 1.2 zettabyte (2010)

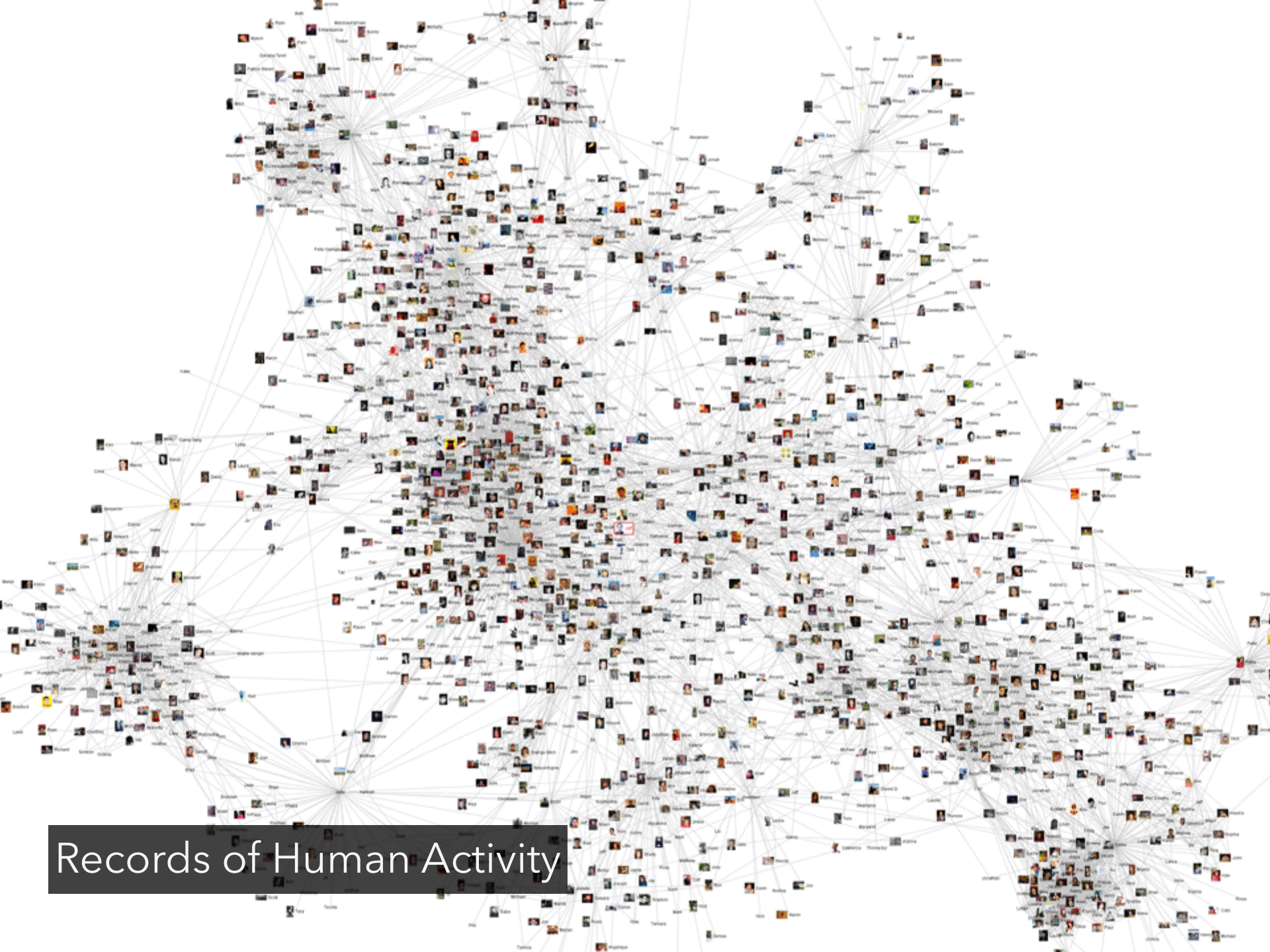
Statista 2018, Gantz et al 2010

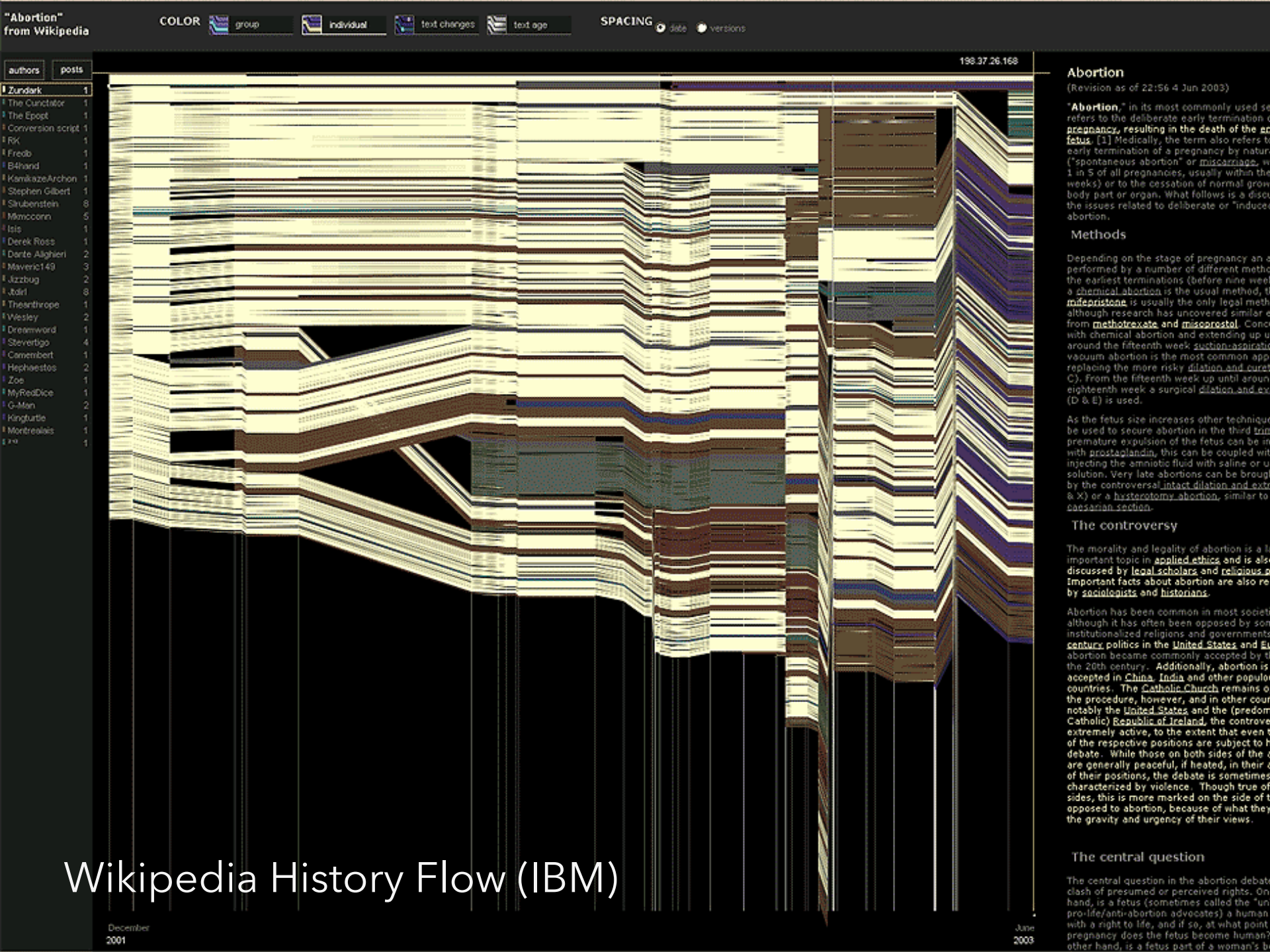


Physical Sensors

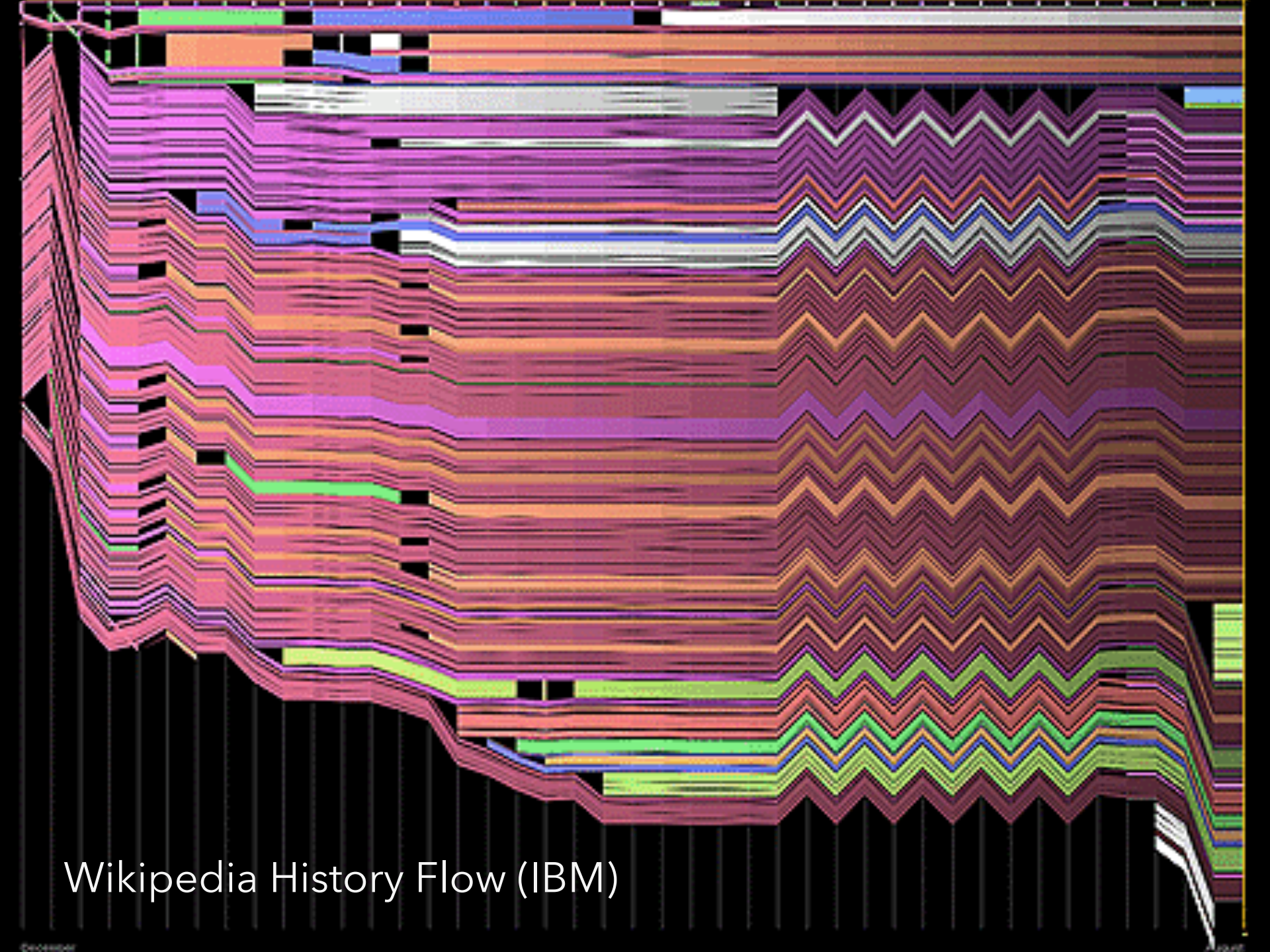
Image courtesy cabspotting.org

Health & Medicine


Records of Human Activity



Wikipedia History Flow (IBM)

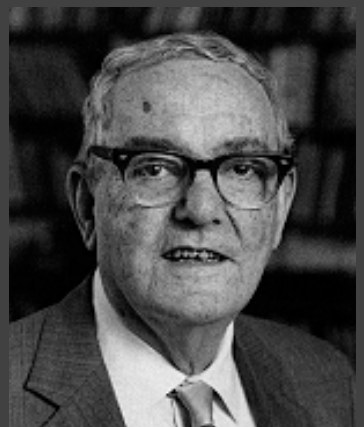

Wikipedia History Flow (IBM)

The ability to take data—to be able to **understand** it, to **process** it, to **extract value** from it, to **visualize** it, to **communicate** it—that's going to be a hugely important skill in the next decades, … because now we really do have **essentially free and ubiquitous data**. So the complimentary scarce factor is the ability to understand that data and extract value from it.

Hal Varian, Google's Chief Economist
The McKinsey Quarterly, Jan 2009

A Poverty of Attention

"What information consumes is rather obvious: it consumes the attention of its recipients. Hence a wealth of information creates a poverty of attention, and a need to allocate that attention efficiently among the overabundance of information sources that might consume it."



Herb Simon
as quoted by Hal Varian
Scientific American
September 1995

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]

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

$u_X = 9.0$ $\sigma_X = 3.317$

$u_Y = 7.5$ $\sigma_Y = 2.03$

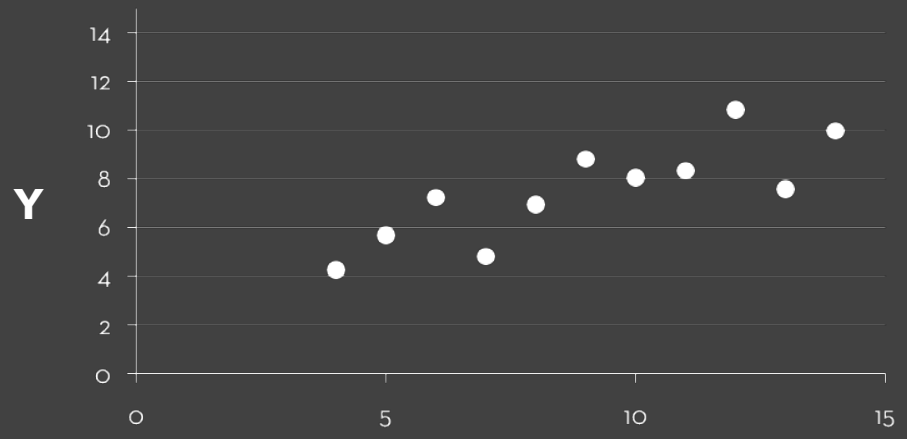
Linear Regression

$Y = 3 + 0.5\ X$

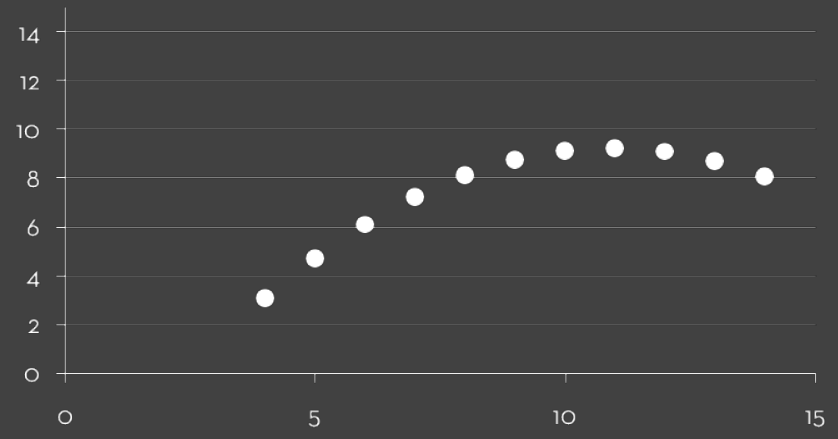
$R^2 = 0.67$

[Anscombe 1973]

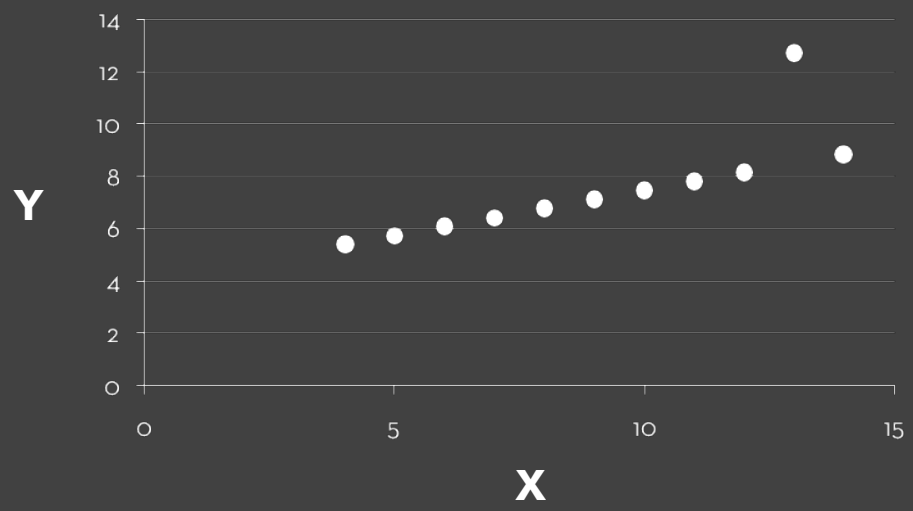
Set A



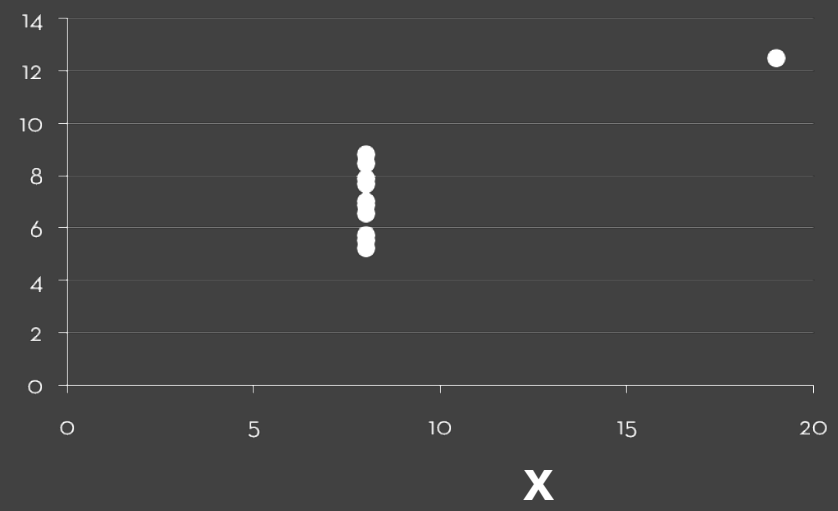
Set B



Set C



Set D

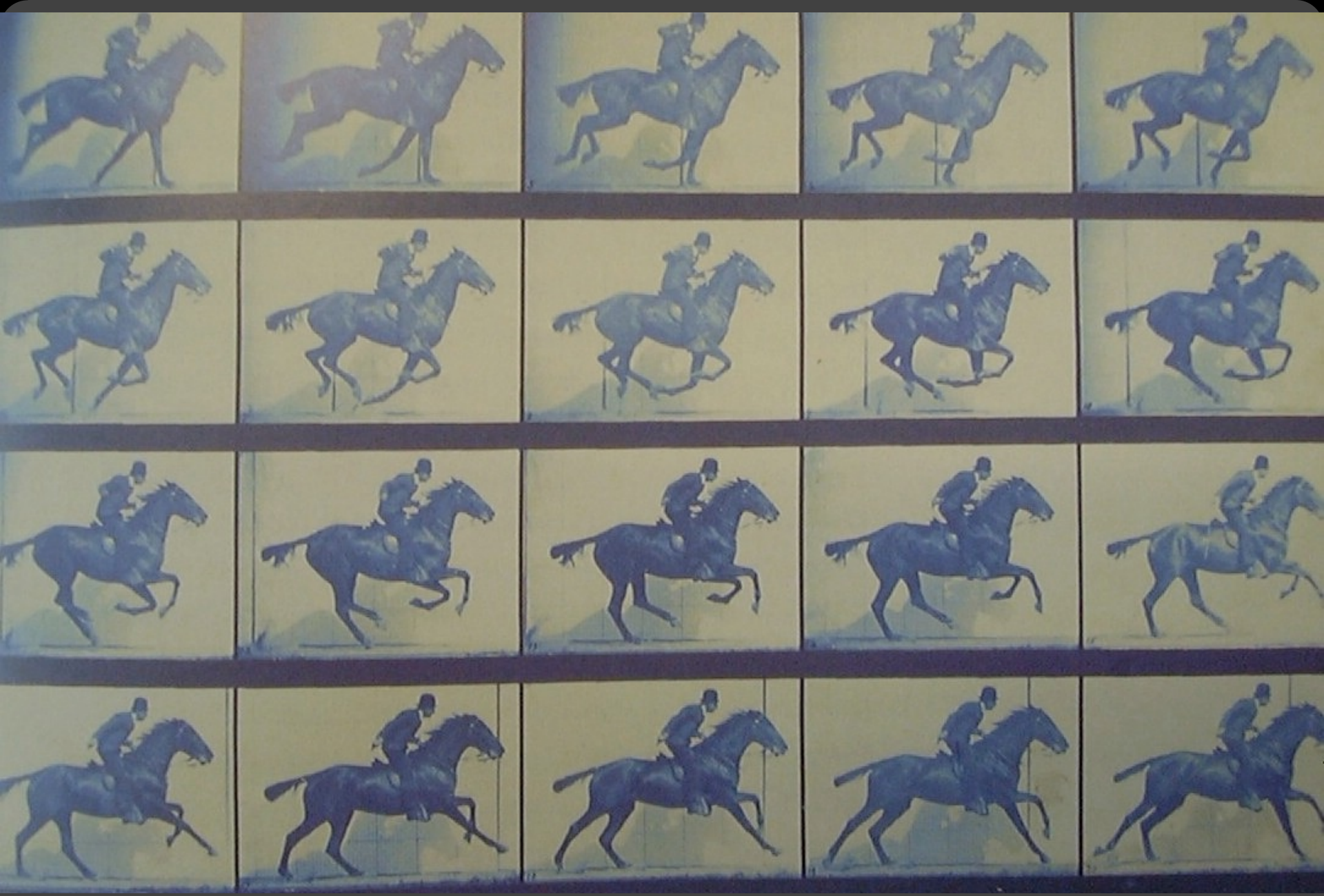


Why Create Visualizations?

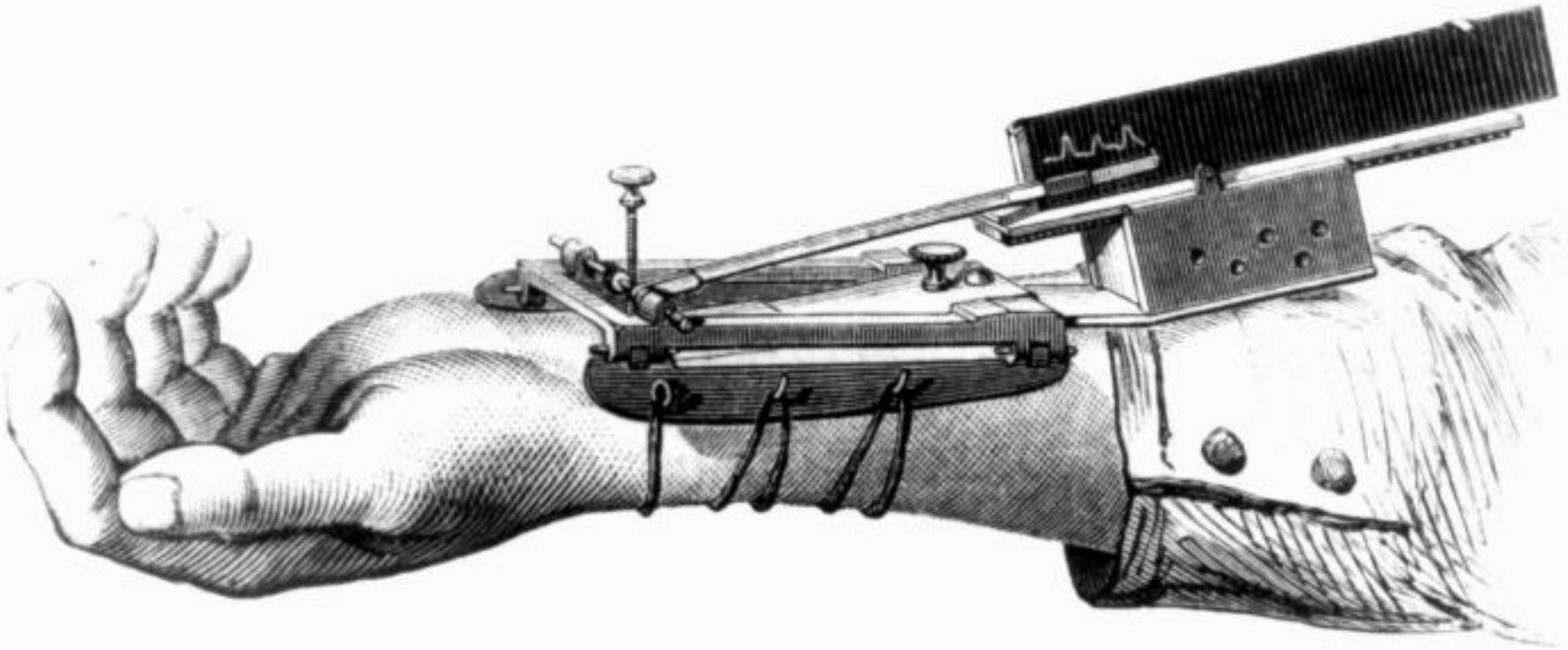
Why Create Visualizations?

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

Record Information

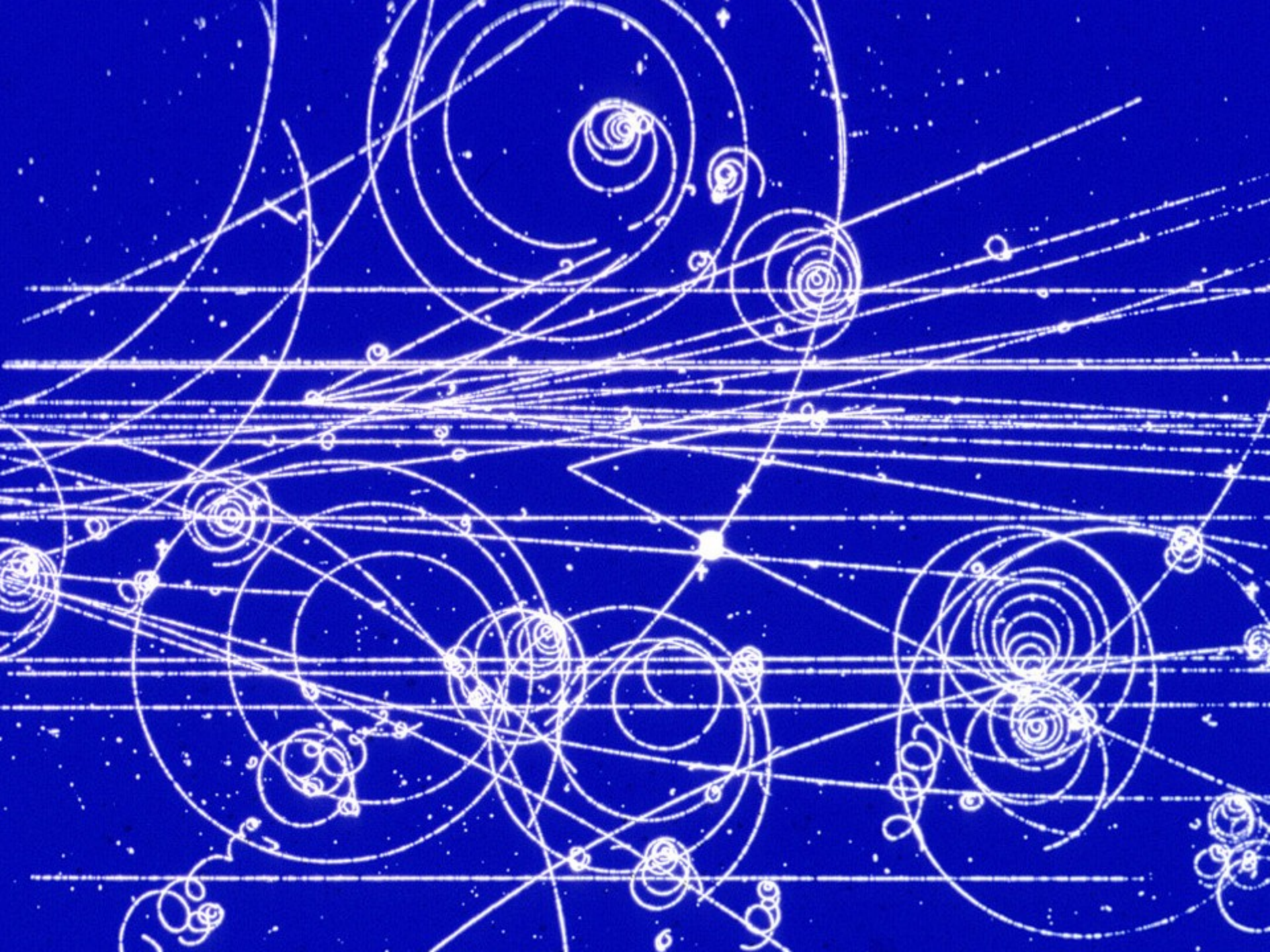


Gallop, Bay Horse “Daisy” [Muybridge 1884-86]



1.
Marey's sphygmograph in use, 1860. *La méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine.*

E.J. Marey's sphygmograph [from Braun 83]




~150 KM

Edge of Frame

DN

50-45 DARK

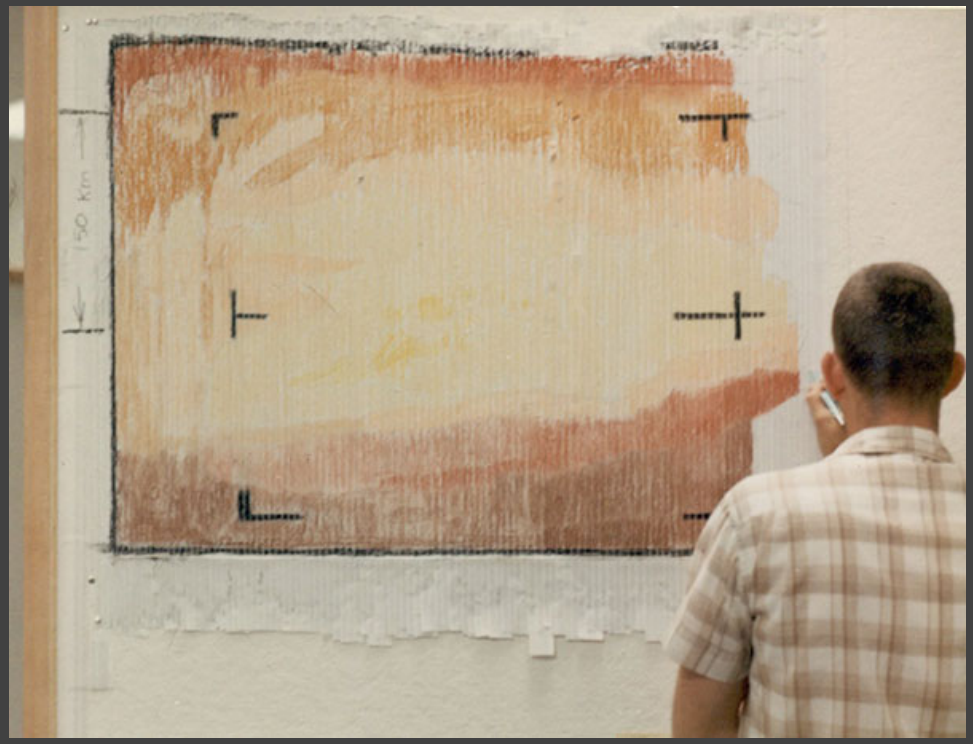
45-40

40-35

35-30

30-25

25-20 LIGHT



Support Reasoning

HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

	SRM No.	Cross Sectional View: Erosion Depth (in.)	Cross Sectional View: Perimeter Affected (deg)	Cross Sectional View: Nominal Dia. (in.)	Top View: Length Of Max Erosion (in.)	Top View: Total Heat Affected Length (in.)	Clocking Location (deg)
61A LH Center Field**	22A	None	None	0.280	None	None	36° --66°
61A LH ~~CENTER~~ FIELD**	22A	NONE	NONE	0.280	NONE	NONE	338°-18°
51C LH Forward Field**	15A	0.010	154.0	0.280	4.25	5.25	163
51C RH Center Field (prim)***	15B	0.038	130.0	0.280	12.50	58.75	354
51C RH Center Field (sec)***	15B	None	45.0	0.280	None	29.50	354
41D RH Forward Field	13B	0.028	110.0	0.280	3.00	None	275
41C LH Aft Field*	11A	None	None	0.280	None	None	--
41B LH Forward Field	10A	0.040	217.0	0.280	3.00	14.50	351
STS-2 RH Aft Field	2B	0.053	116.0	0.280	--	--	90

*Hot gas path detected in putty. Indication of heat on O-ring, but no damage.
**Soot behind primary O-ring.
***Soot behind primary O-ring, heat affected secondary O-ring.

Clocking location of leak check port - 0 deg.

OTHER SRM-15 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY AND NO SOOT NEAR OR BEYOND THE PRIMARY O-RING.

SRM-22 FORWARD FIELD JOINT HAD PUTTY PATH TO PRIMARY O-RING, BUT NO O-RING EROSION AND NO SOOT BLOWBY. OTHER SRM-22 FIELD JOINTS HAD NO BLOWHOLES IN PUTTY.

BLOW BY HISTORY

SRM-15 WORST BLOW-BY
- 2 CASE JOINTS (80°), (110°) ARC
- MUCH WORSE VISUALLY THAN SRM-22

SRM 22 BLOW-BY
- 2 CASE JOINTS (30-40°)

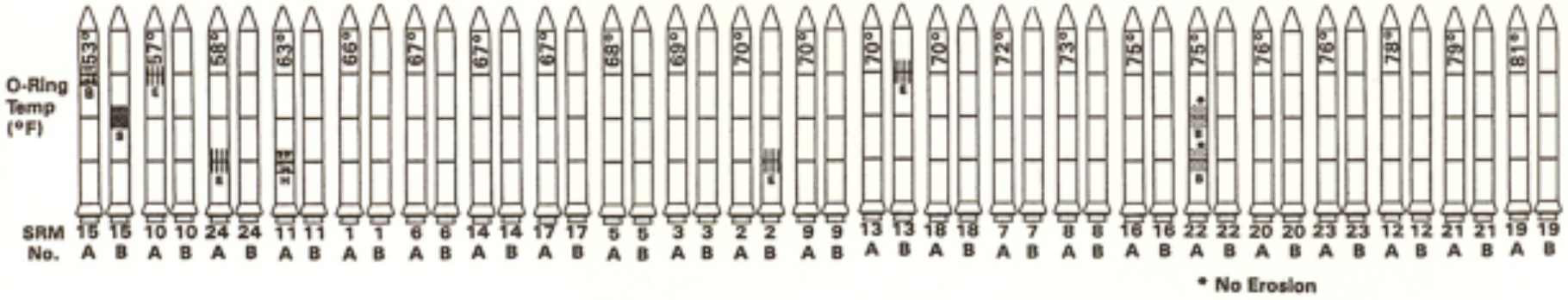
SRM-13A, 15, 16A, 18, 23A 24A
- NOZZLE BLOW-BY

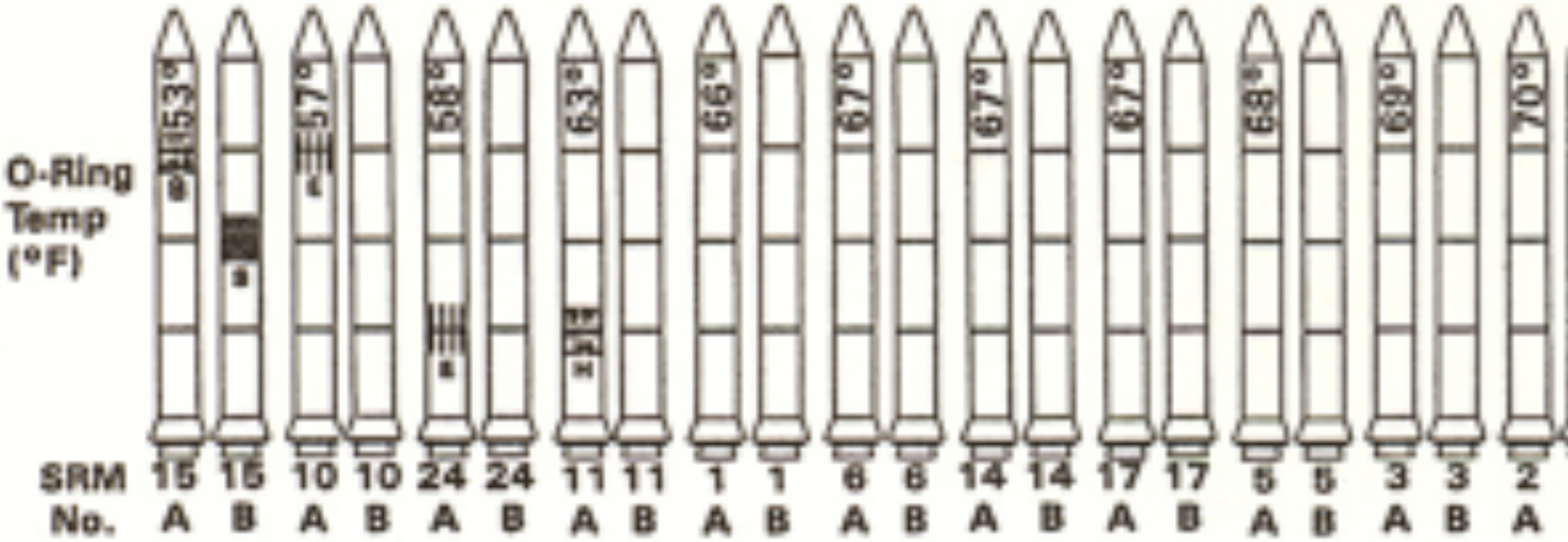
HISTORY OF O-RING TEMPERATURES (DEGREES - F)

MOTOR	MBT	AMB	O-RING	WIND
DM-1	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29 27	10 MPH 25 MPH

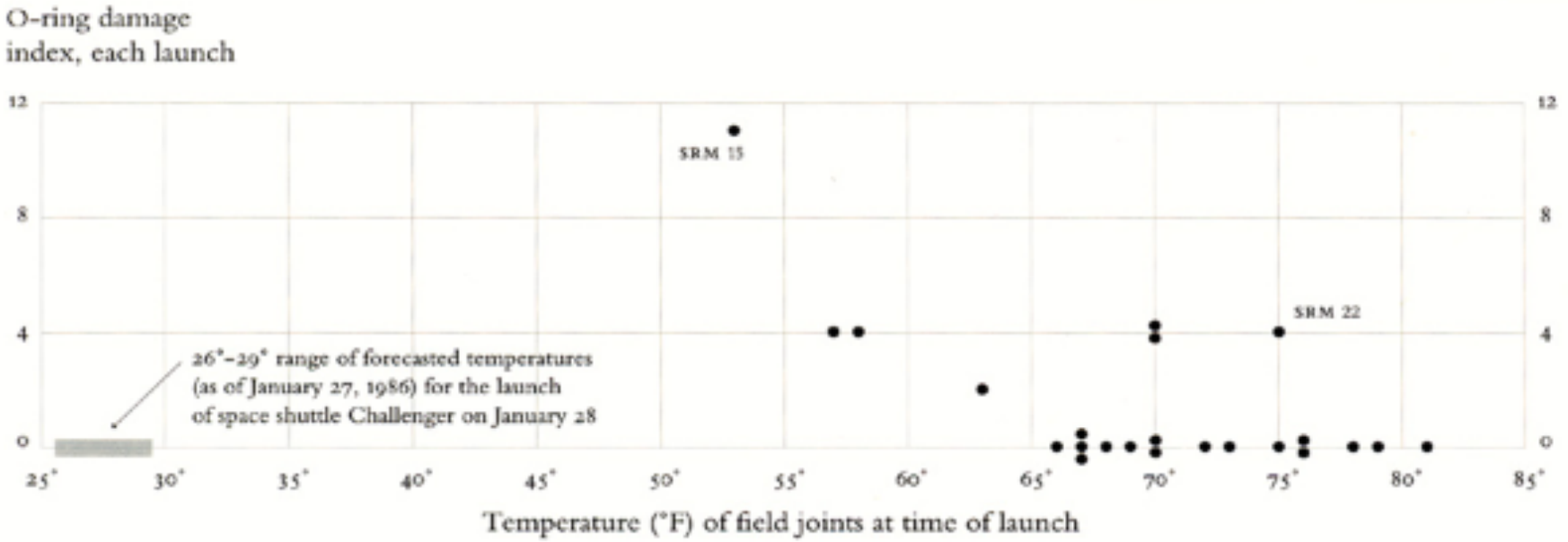
2 of 13 pages of material faxed to NASA by Morton Thiokol [from Tufte 1997]

Make a Decision: Challenger





Make a Decision: Challenger



Visualizations drawn by Tufte show how low temperatures damage O-rings [Tufte 97]

Data in Context: Cholera Outbreak

In 1854 John Snow plotted the position of each cholera case on a map. [from Tufte 83]

Data in Context: Cholera Outbreak

Used map to hypothesize that pump on Broad St. was the cause. [from Tufte 83]

Expand Memory: Multiplication

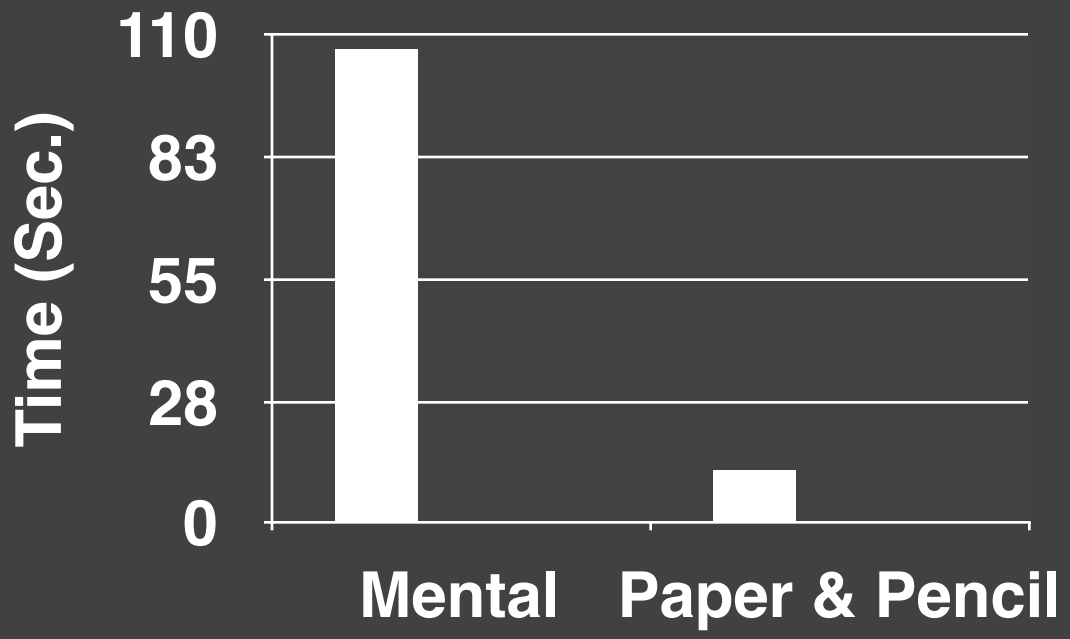
Class Exercise!

Expand Memory: Multiplication

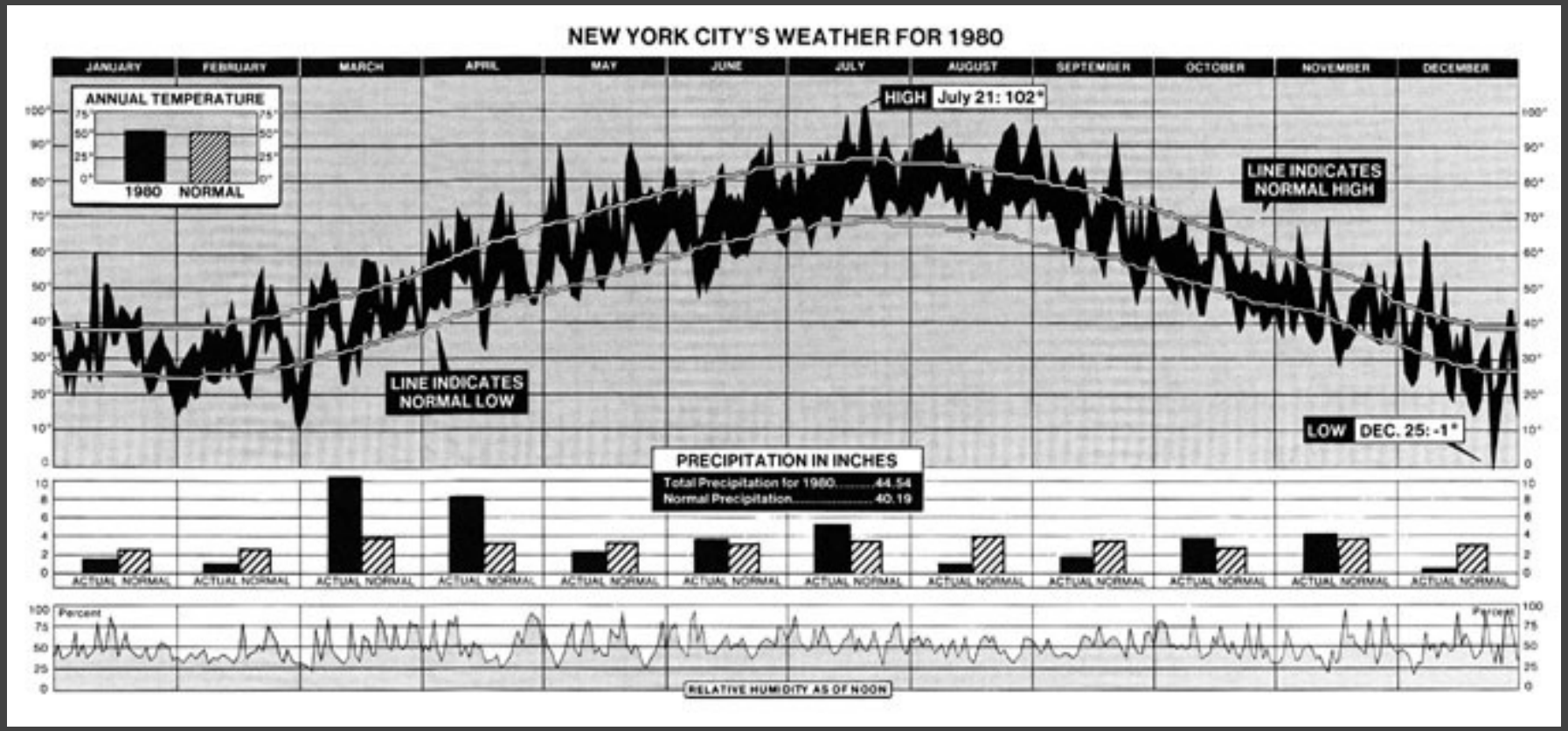
```
   34
x  72
```

Expand Memory: Multiplication

```
   34
x  72
-----
   68
 2380
-----
 2448
```

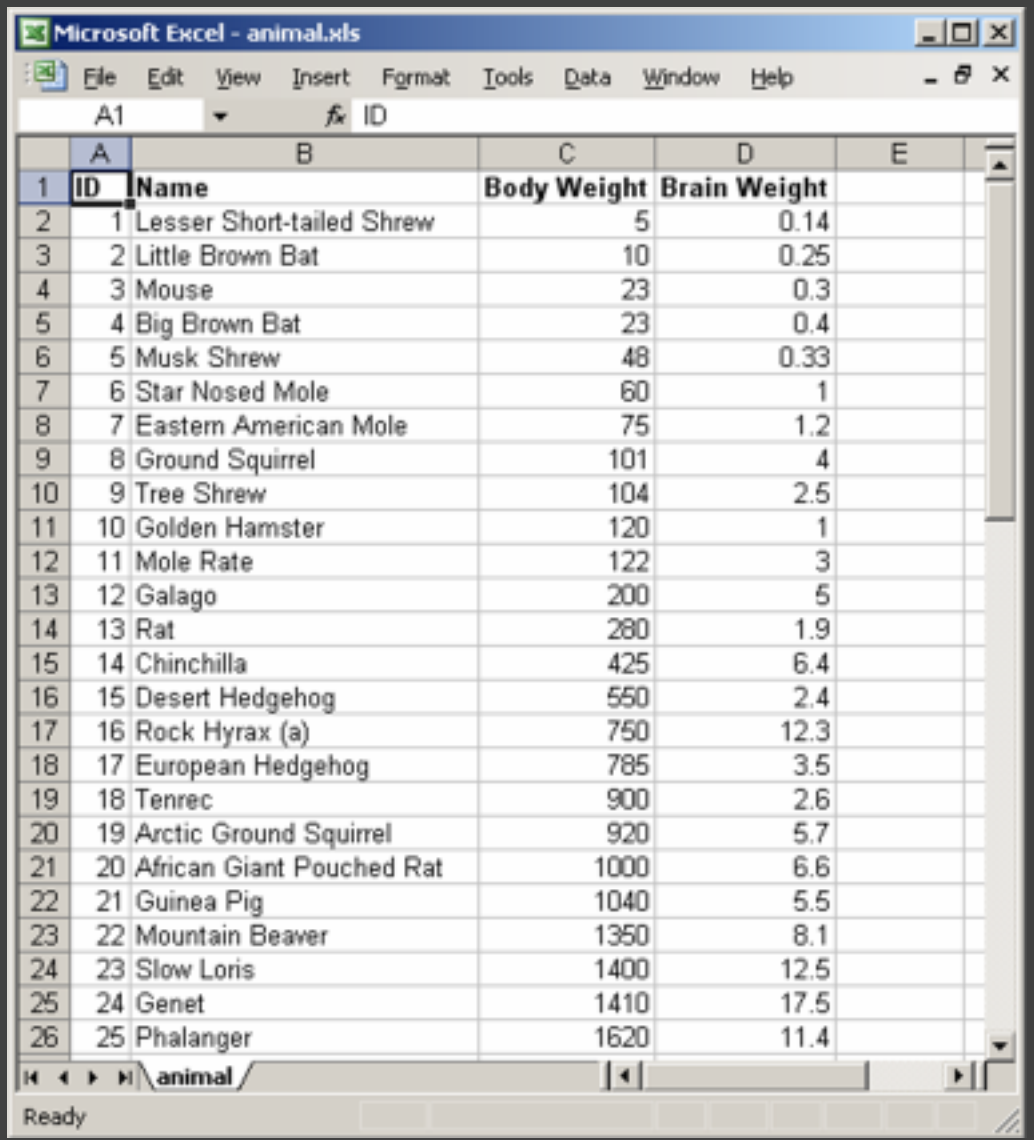


Find Patterns: NYC Weather



[New York Times 1981]

The Most Powerful Brain?


Microsoft Excel - animal.xls

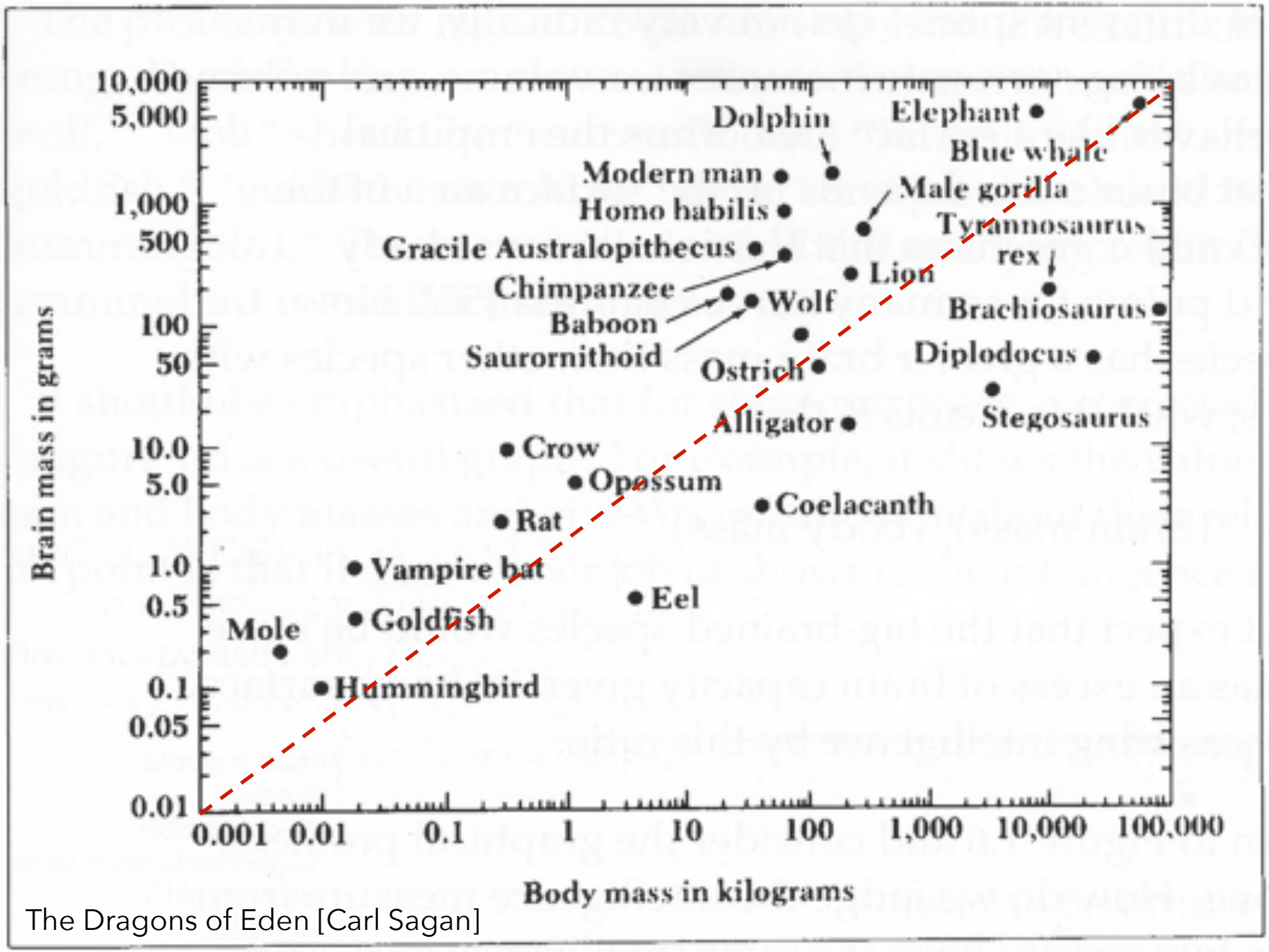
File Edit View Insert Format Tools Data Window Help

A1 ID

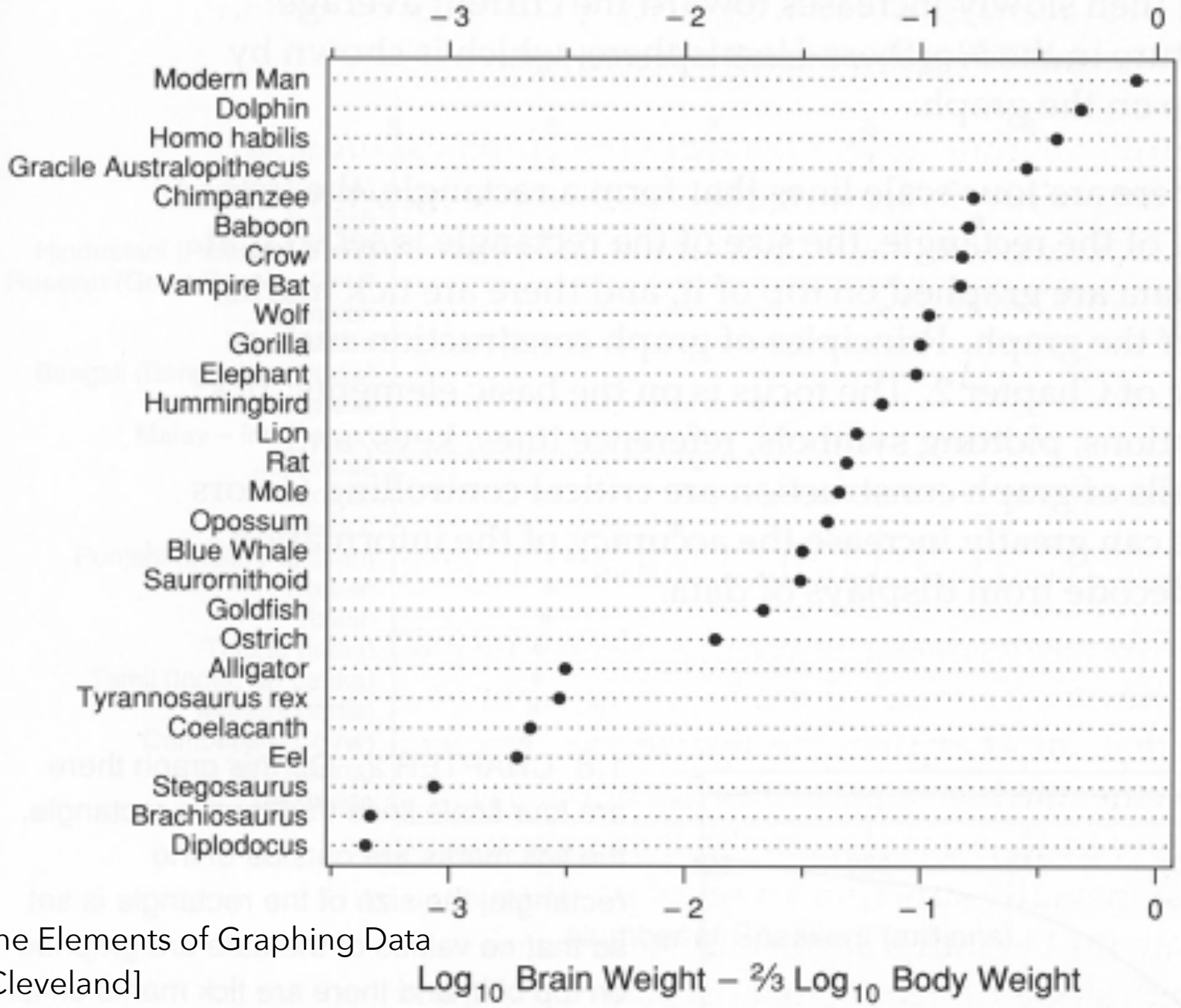
	A	B	C	D	E
1	ID	Name	Body Weight	Brain Weight	
2	1	Lesser Short-tailed Shrew	5	0.14	
3	2	Little Brown Bat	10	0.25	
4	3	Mouse	23	0.3	
5	4	Big Brown Bat	23	0.4	
6	5	Musk Shrew	48	0.33	
7	6	Star Nosed Mole	60	1	
8	7	Eastern American Mole	75	1.2	
9	8	Ground Squirrel	101	4	
10	9	Tree Shrew	104	2.5	
11	10	Golden Hamster	120	1	
12	11	Mole Rate	122	3	
13	12	Galago	200	5	
14	13	Rat	280	1.9	
15	14	Chinchilla	425	6.4	
16	15	Desert Hedgehog	550	2.4	
17	16	Rock Hyrax (a)	750	12.3	
18	17	European Hedgehog	785	3.5	
19	18	Tenrec	900	2.6	
20	19	Arctic Ground Squirrel	920	5.7	
21	20	African Giant Pouched Rat	1000	6.6	
22	21	Guinea Pig	1040	5.5	
23	22	Mountain Beaver	1350	8.1	
24	23	Slow Loris	1400	12.5	
25	24	Genet	1410	17.5	
26	25	Phalanger	1620	11.4	

animal

Ready



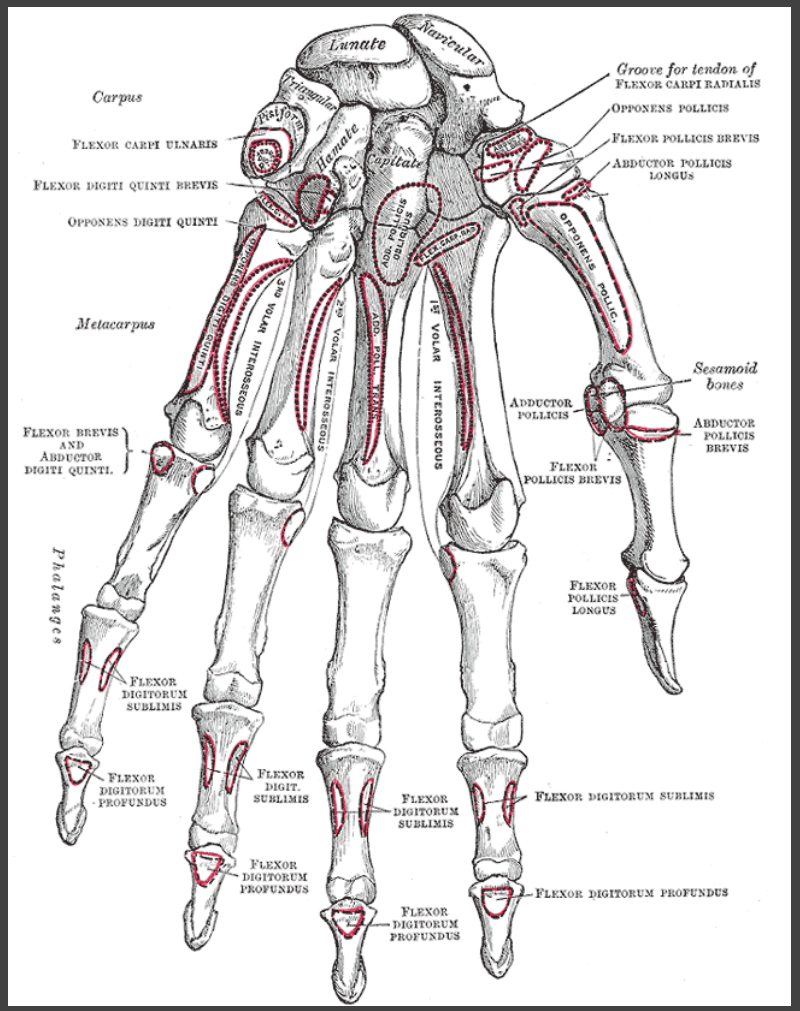
The Dragons of Eden [Carl Sagan]



The Elements of Graphing Data
[Cleveland]

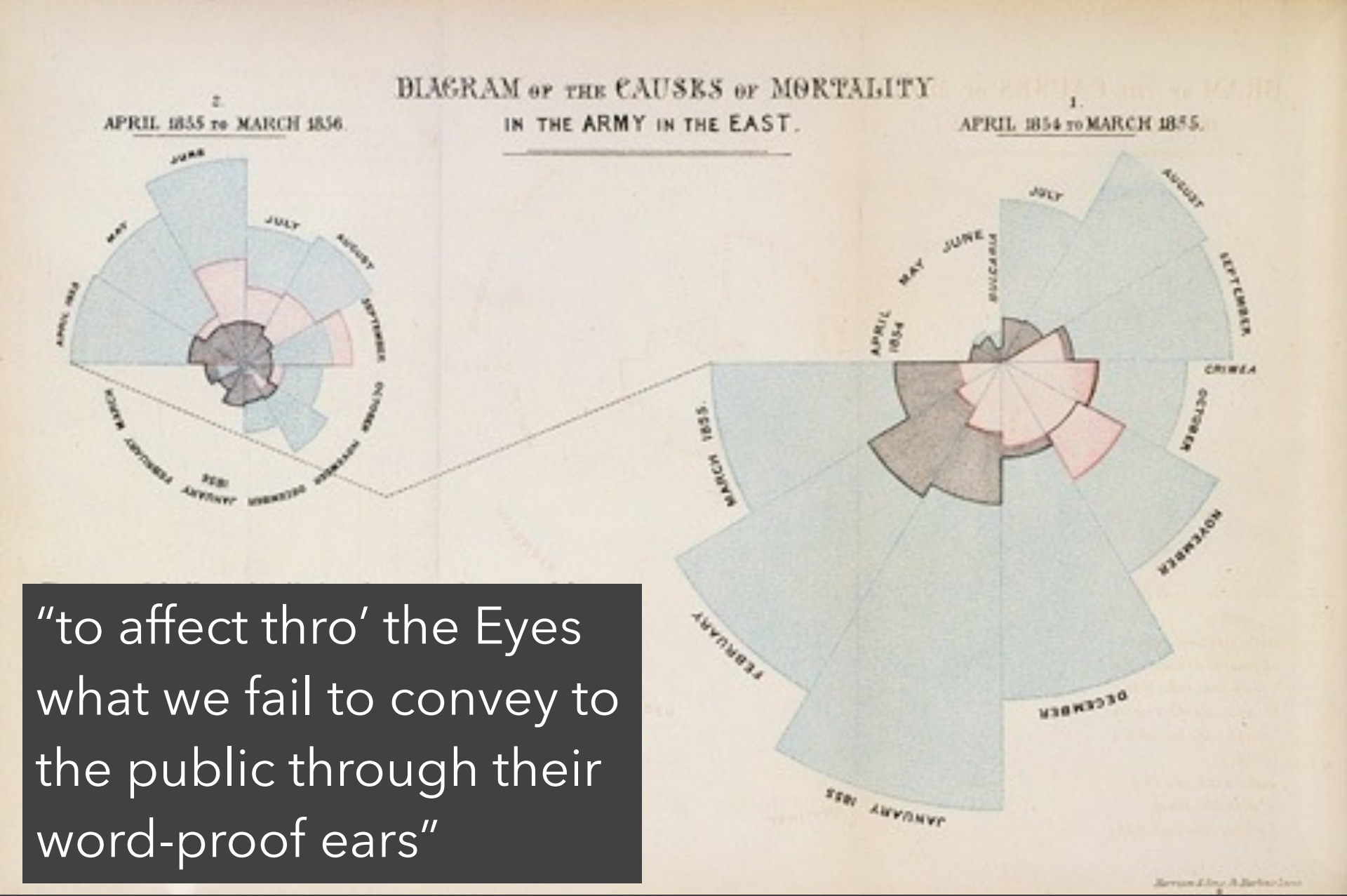
Convey Information to Others

Inspire



Bones in hand [from 1918 edition]

Double helix model [Watson and Crick 53]



"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

The Value of Visualization

Record information
- Blueprints, photographs, seismographs, …

Analyze data to support reasoning
- Develop and assess hypotheses
- Find patterns / Discover errors in data
- Expand memory

Communicate information to others
- Share and persuade
- Collaborate and revise

Goals of Visualization Research

1. **Understand** how visualizations convey information

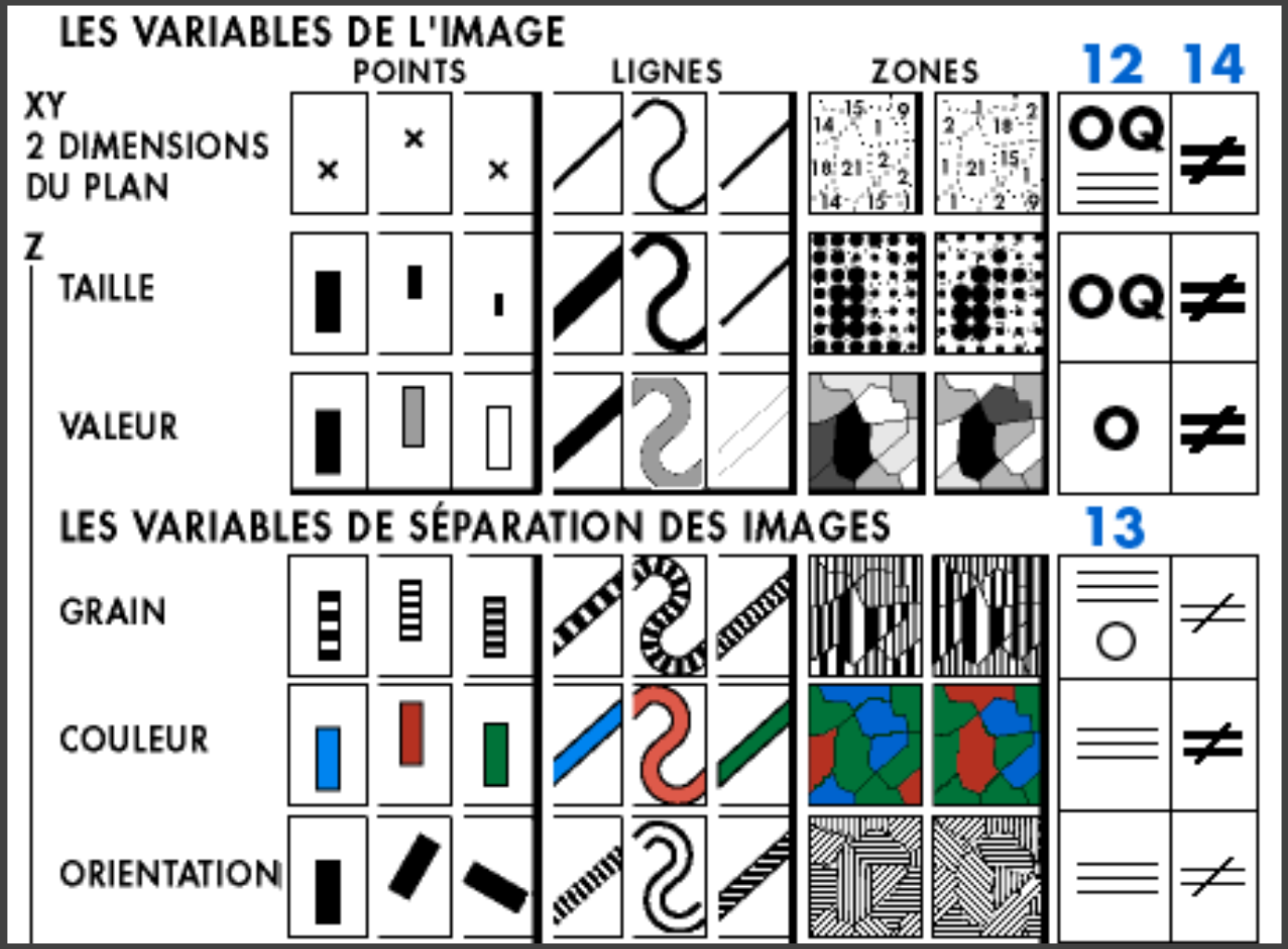
   What do people perceive/comprehend?
   How do visualizations inform mental models?

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

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

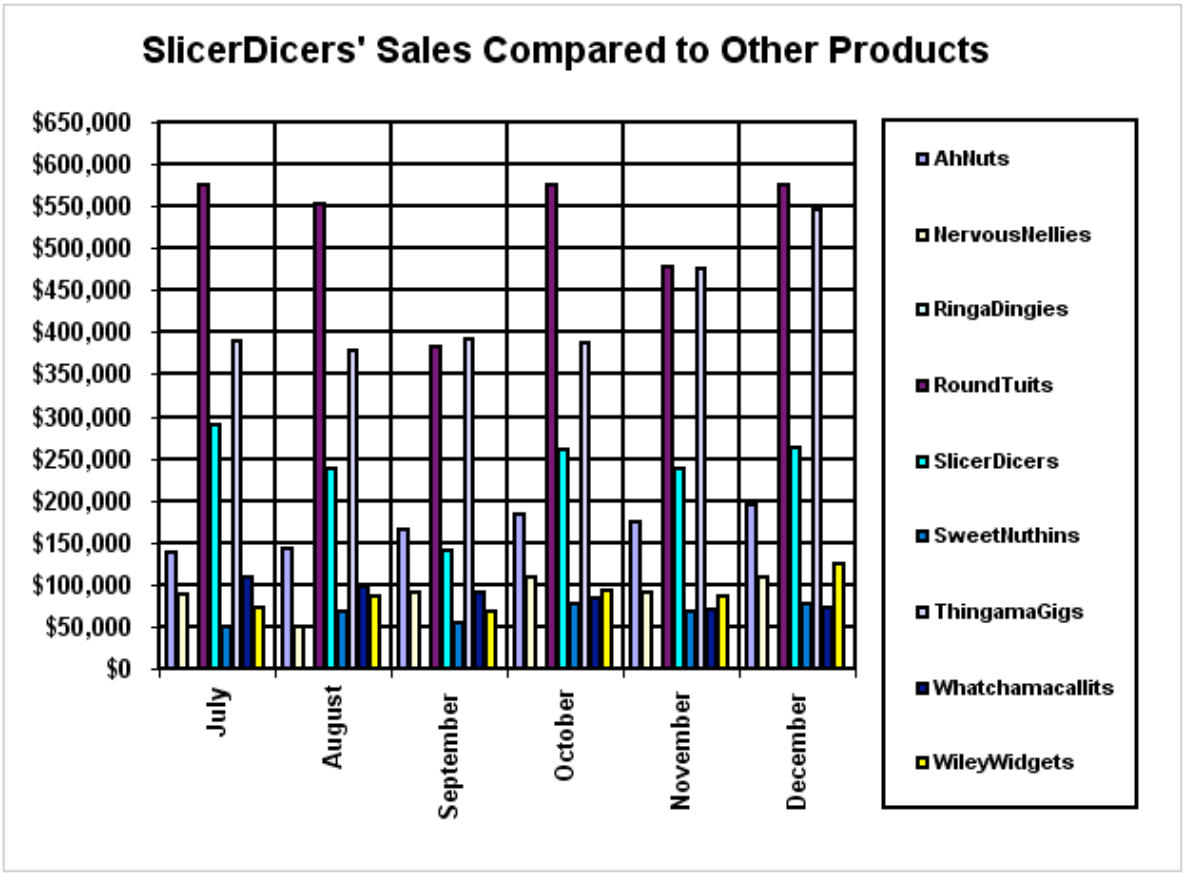
Course Topics

Data and Image Models

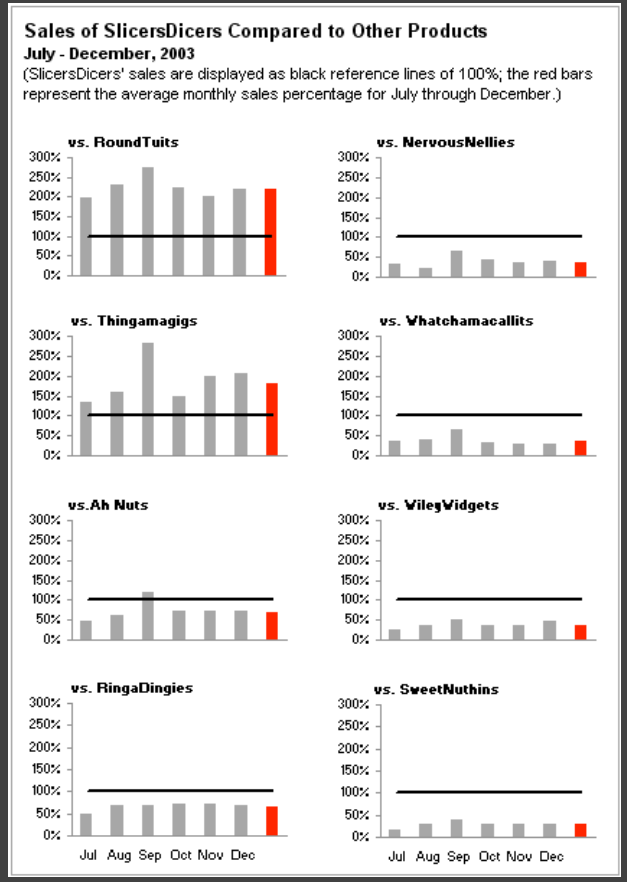


Sémiologie Graphique [Bertin 67]

Visualization Design

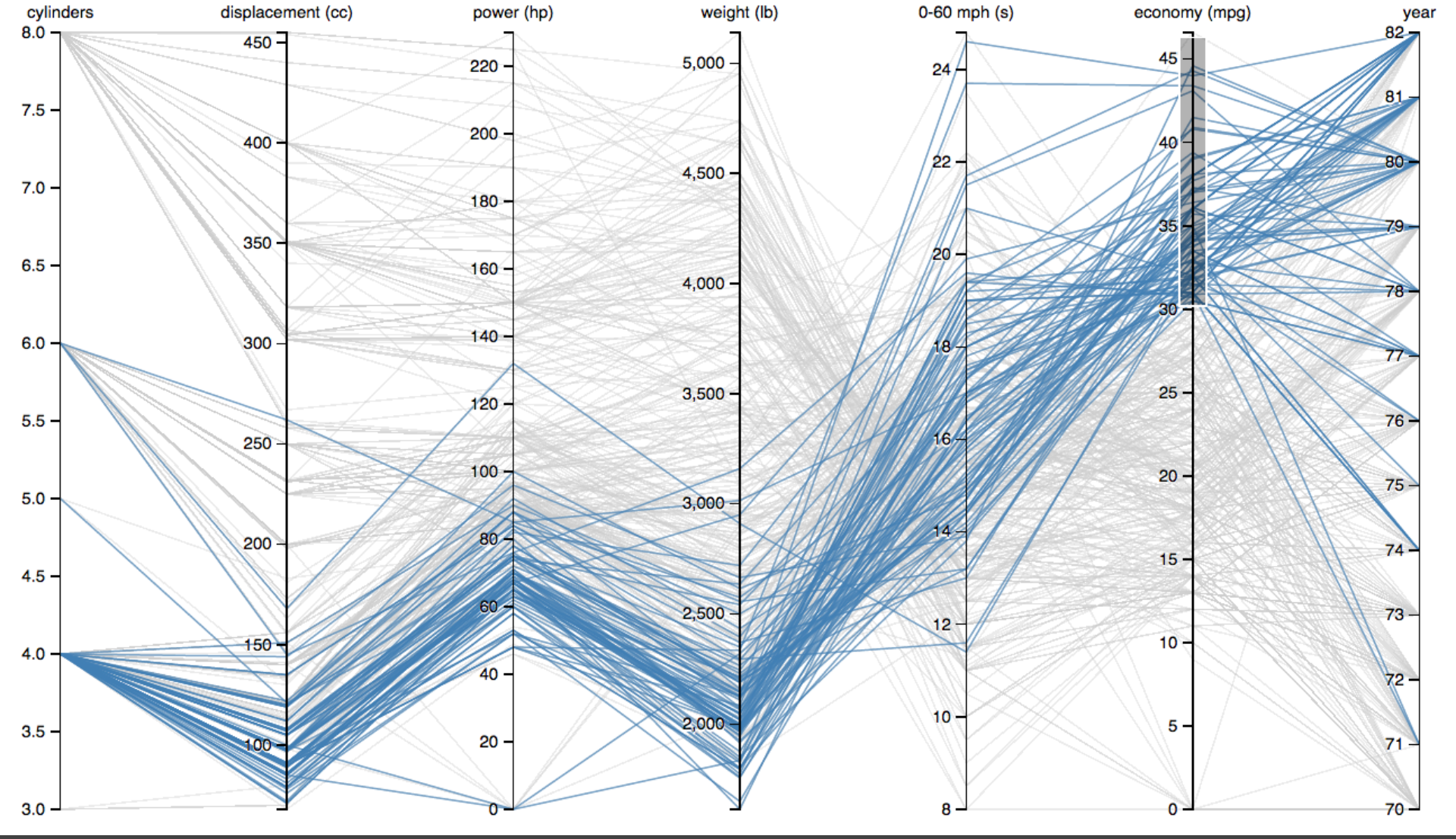


Problematic design



Redesign

Exploratory Data Analysis



Narrative

Recent elections have placed a heavy emphasis on "swing states" — Ohio, Florida and the other competitive states. Y[illegible] the past, many more states shifted between the Democratic and Republican parties. A look at how the states st[illegible] they have shifted over past elections.

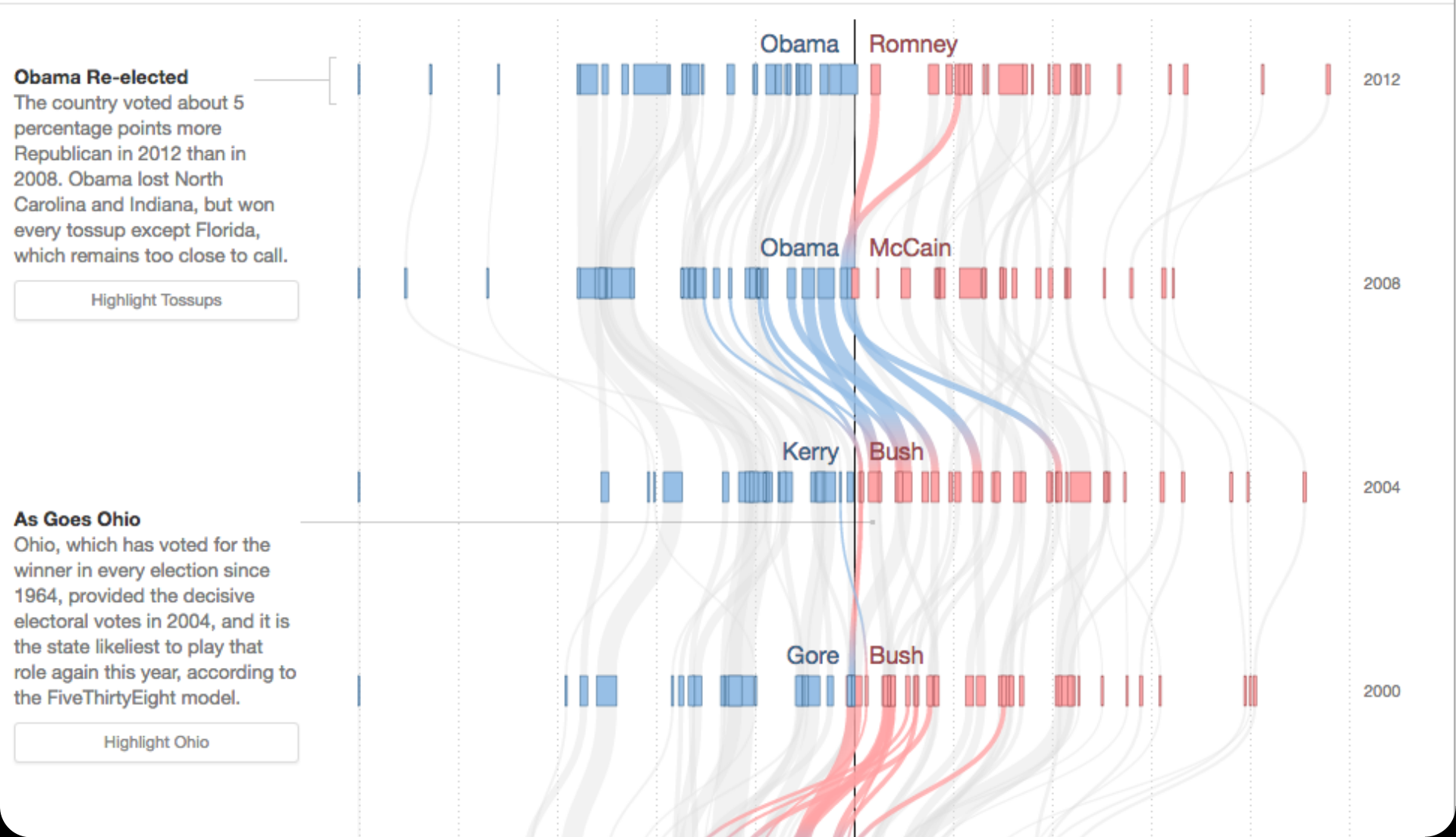
Each box represents a state sized by number of electoral votes.

Each curve shows how much it shifted left or right between elections

Chart **Size of Lead** | Chart **Electoral Votes**

← MORE DEMOCRATIC | MORE REPUBLICAN →

≥50% +40% +30% +20% +10% | +10% +20% +30% +40% ≥50%



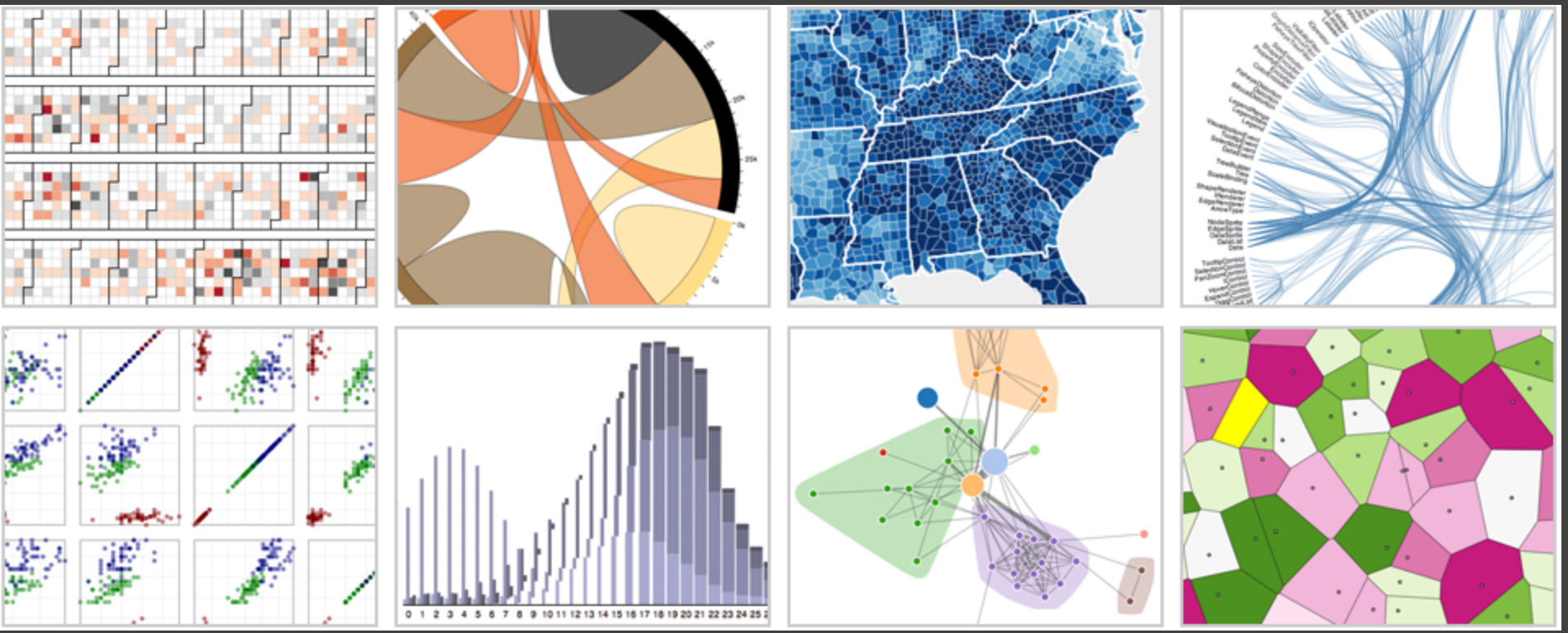
Obama Re-elected
The country voted about 5 percentage points more Republican in 2012 than in 2008. Obama lost North Carolina and Indiana, but won every tossup except Florida, which remains too close to call.

Highlight Tossups

As Goes Ohio
Ohio, which has voted for the winner in every election since 1964, provided the decisive electoral votes in 2004, and it is the state likeliest to play that role again this year, according to the FiveThirtyEight model.

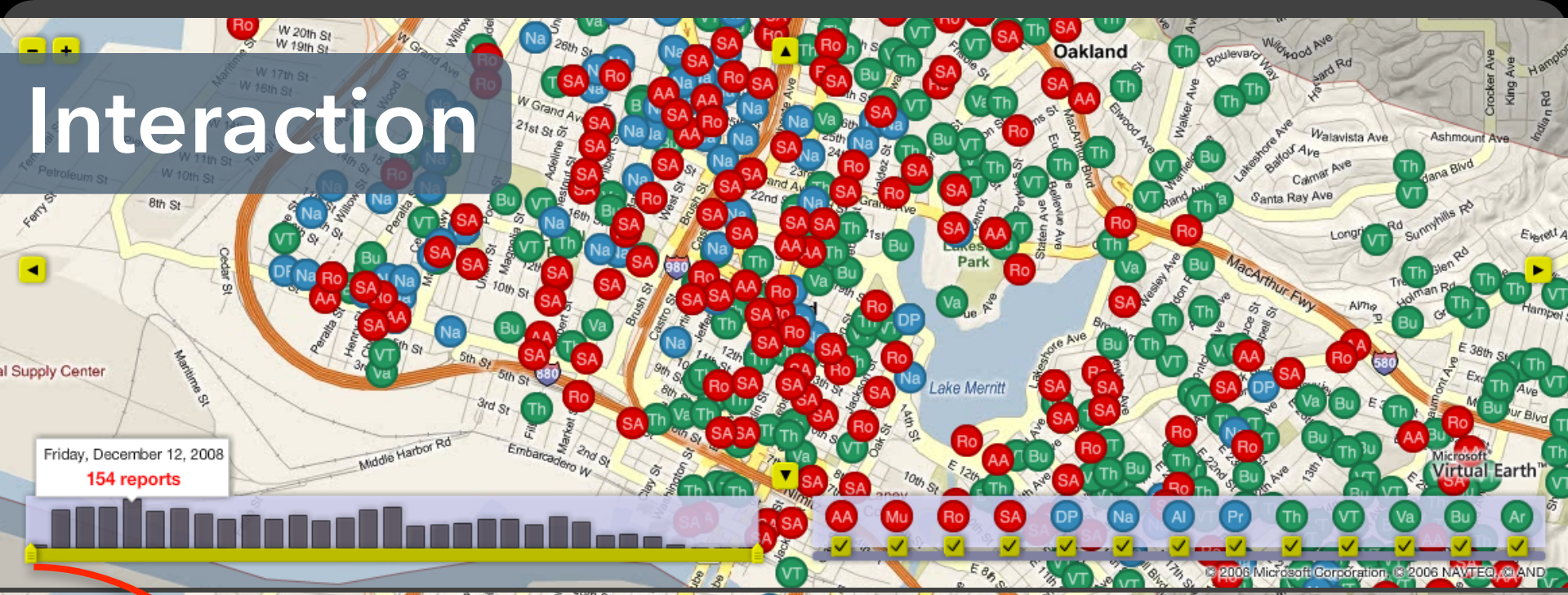
Highlight Ohio

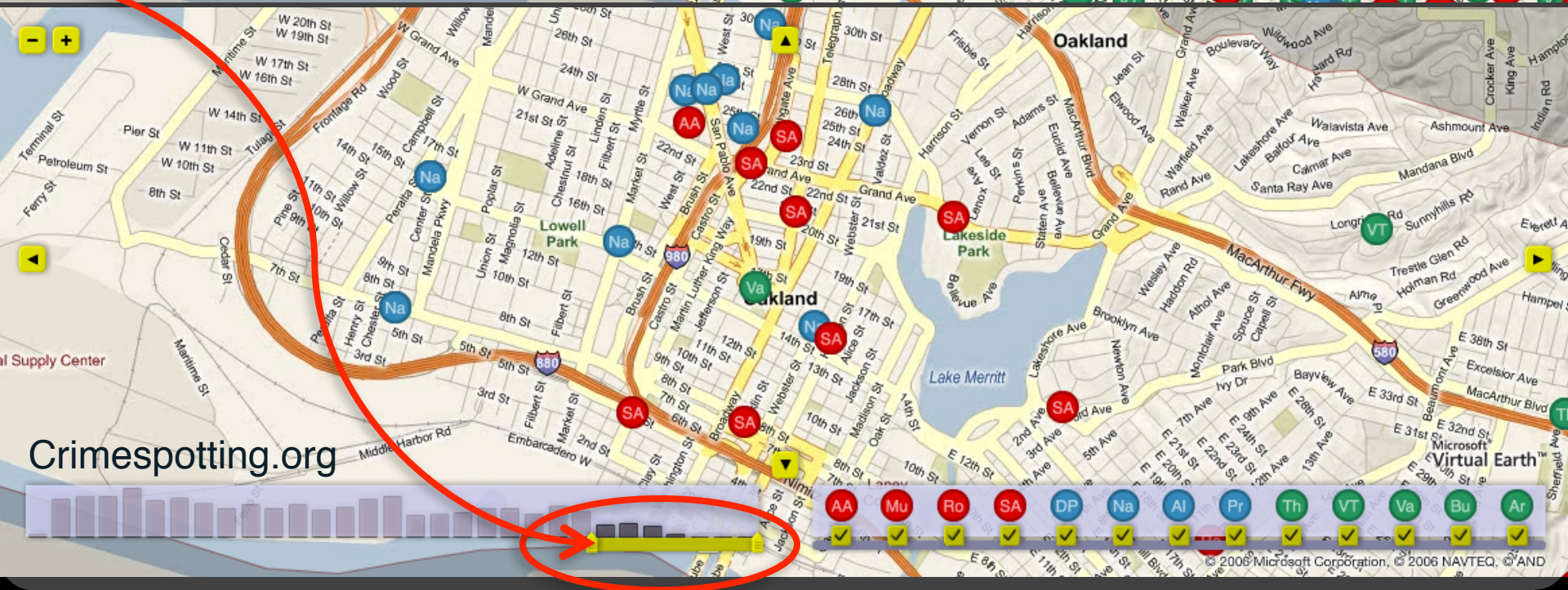
Visualization Software



D3: Data-Driven Documents

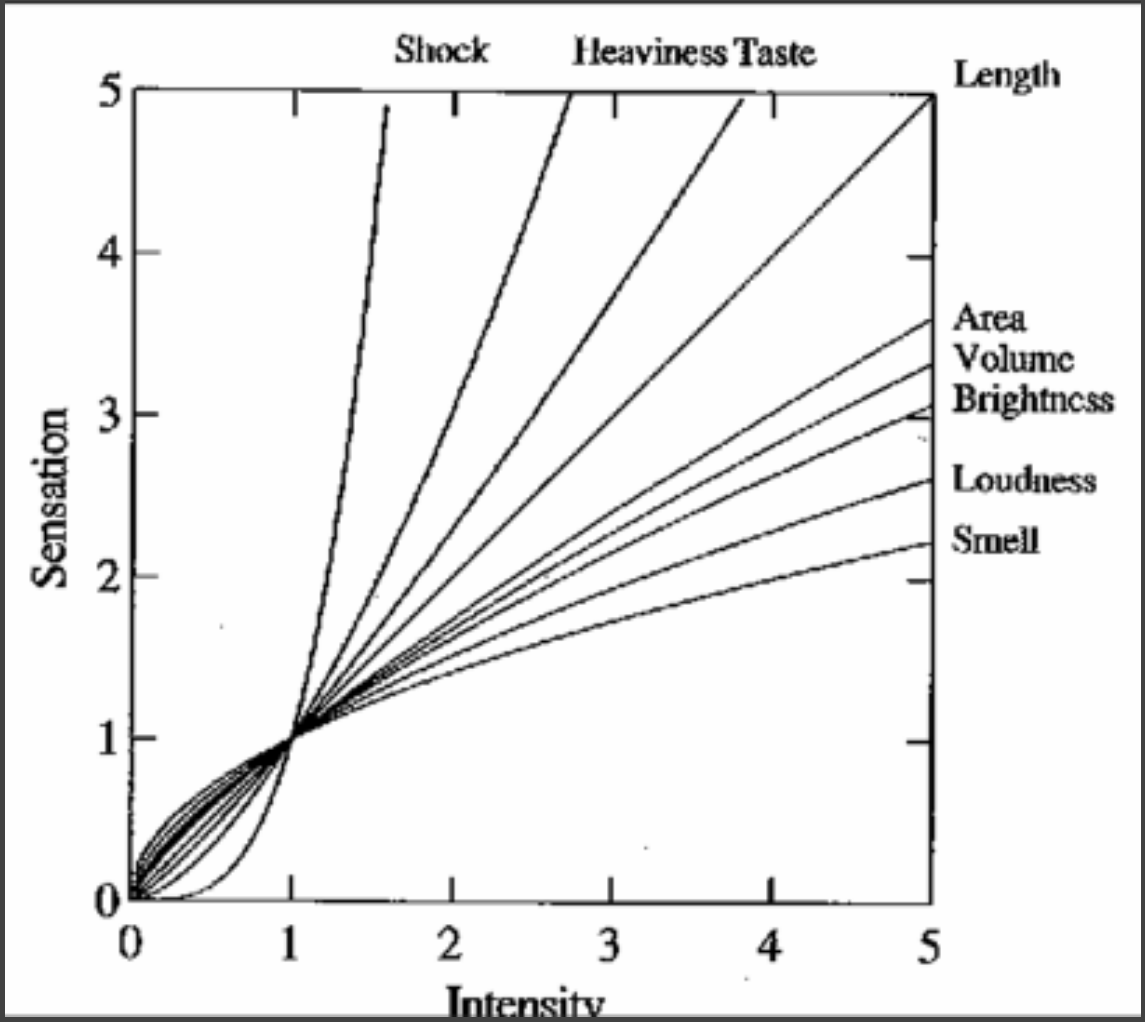
Interaction





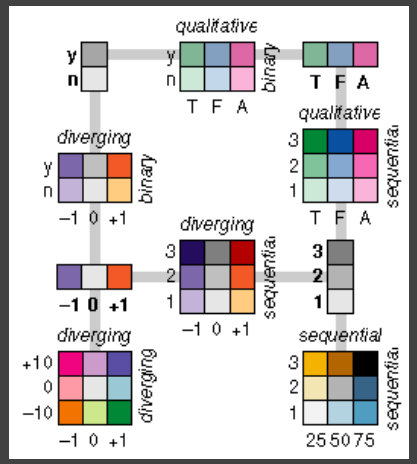
Crimespotting.org

Graphical Perception

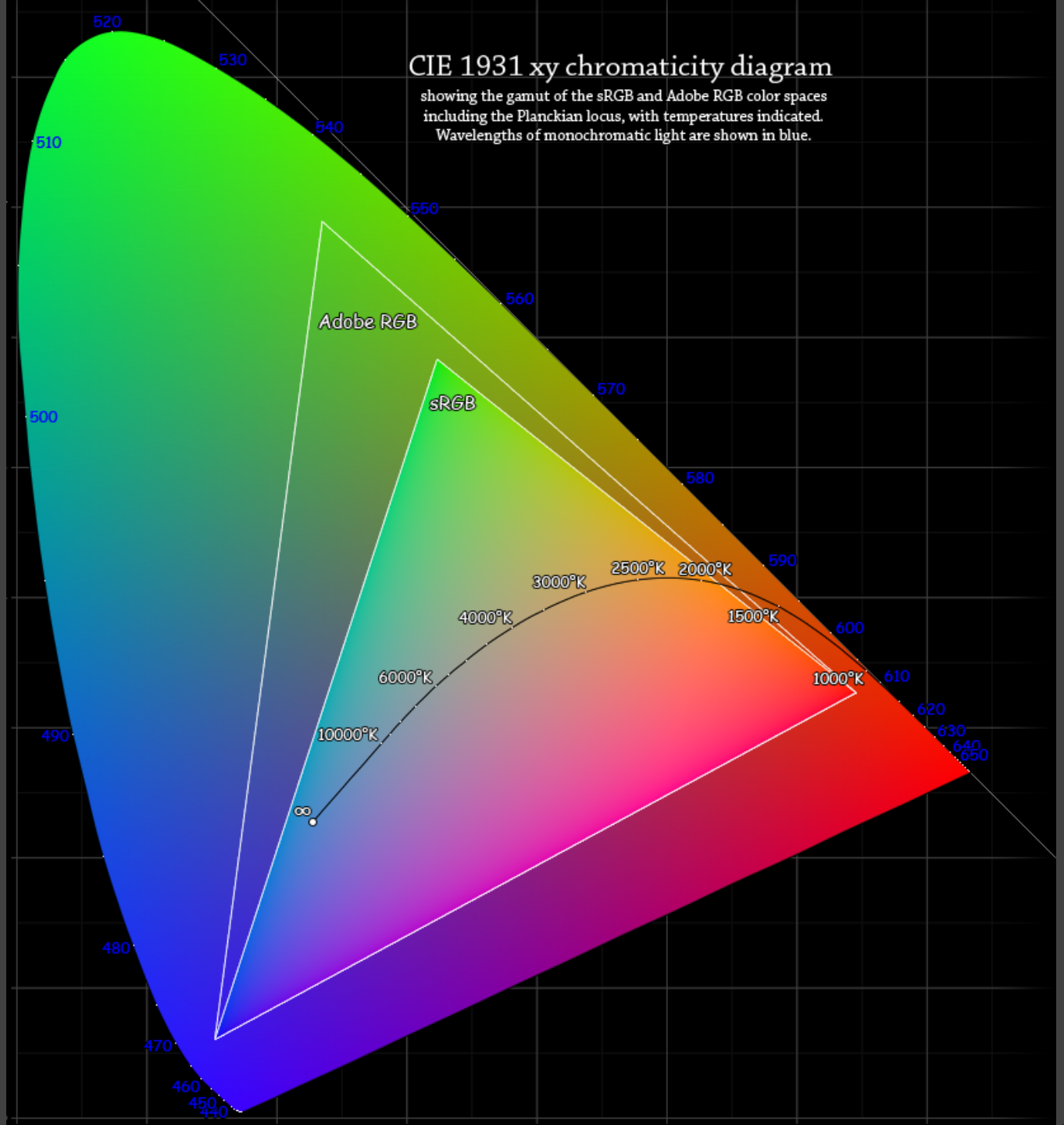


The psychophysics of sensory function [Stevens 61]

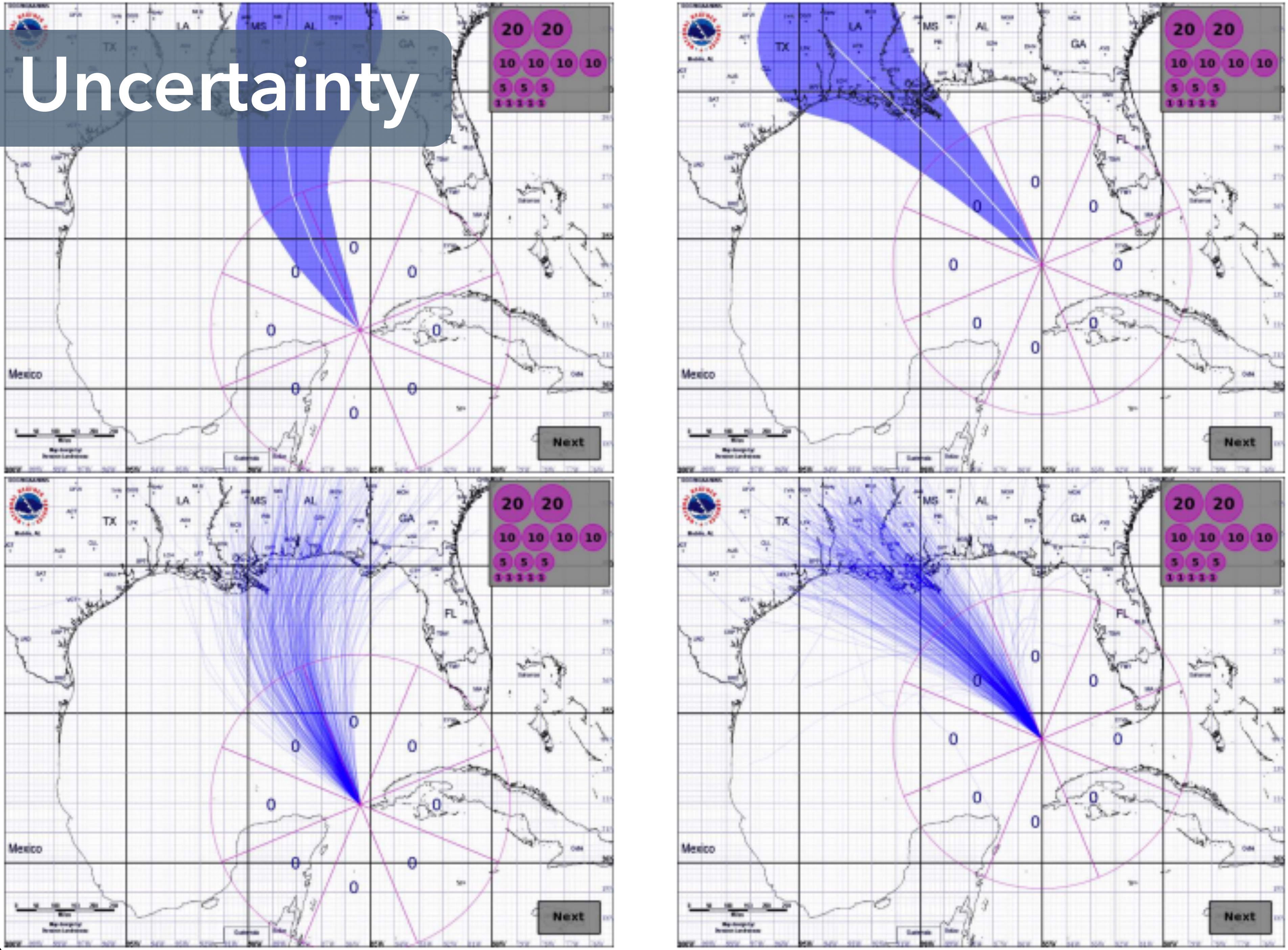
Color



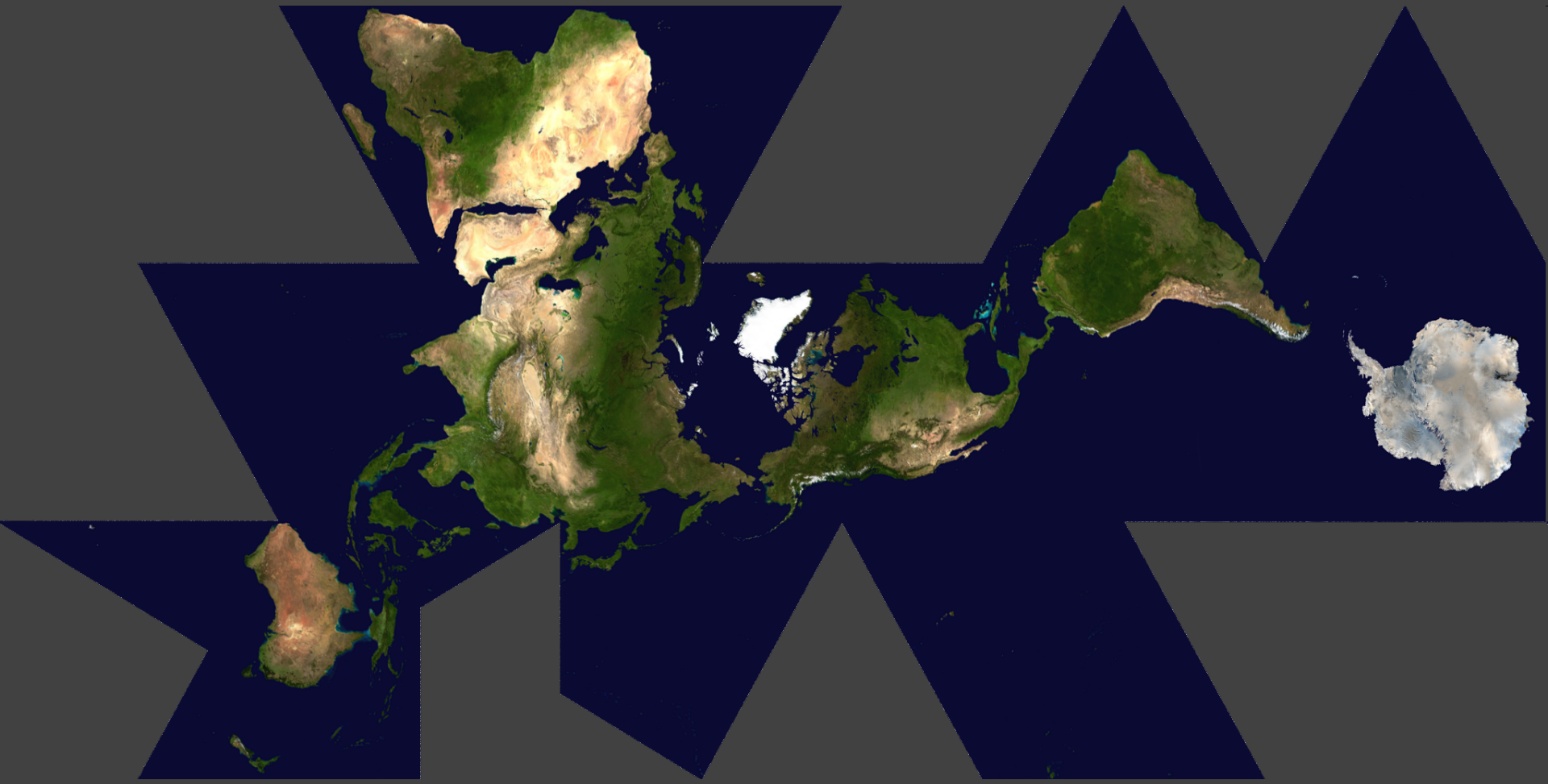
Color Brewer



Uncertainty

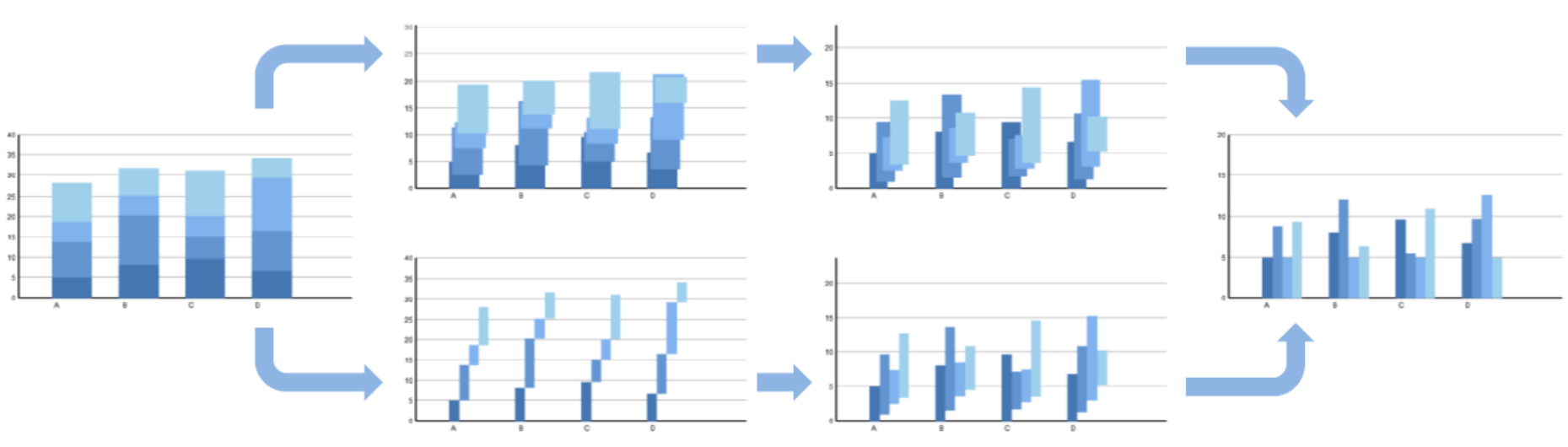


Maps



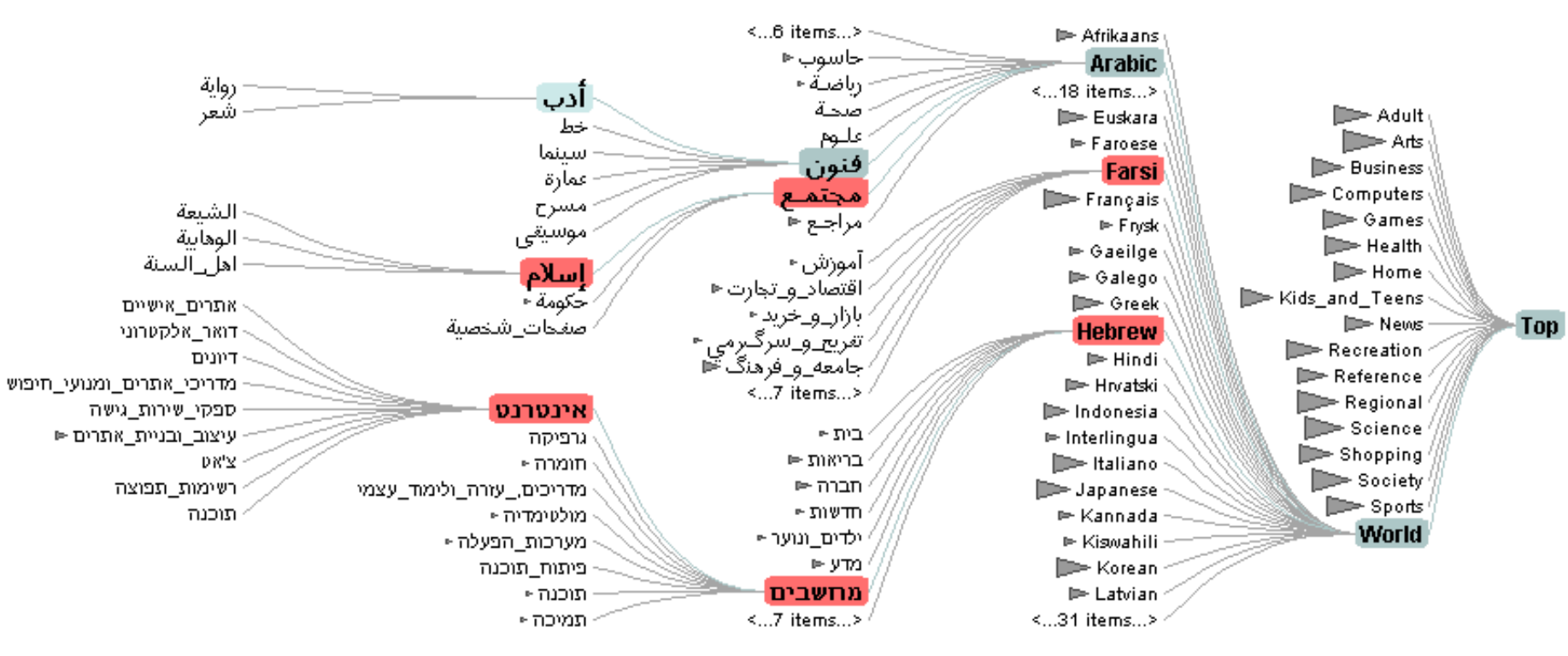
Dymaxion Maps [Fuller 46]

Animation



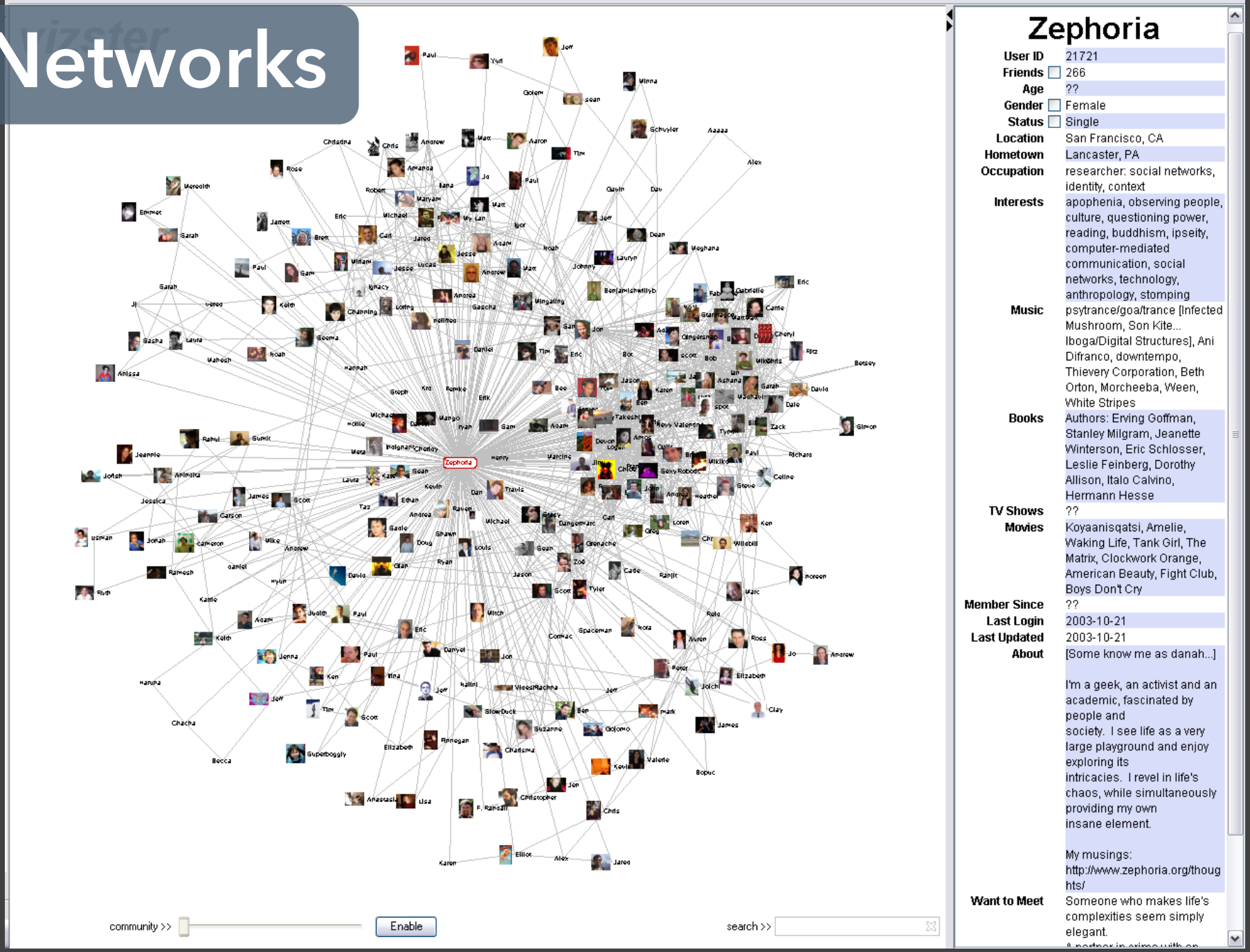
Animated transitions in statistical data graphics [Heer & Robertson 07]

Hierarchies



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

Networks



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.

Course Mechanics

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

Matthew Conlen

mconlen@cs.washington.edu

Research on interactive documents and data-driven storytelling.

I maintain an interactive markup language called Idyll, publish a data-driven digital magazine called the Parametric Press, and work with NASA's climate communications team.

THE NEW YORKER

D 48.1% R 44.2%

Change voters' political loyalties to build a winning coalition.

White evangelical Protestant, no college degree
15% of voters, expected to shrink
Members of this group are strongly anti-immigrant, not accepting of gay identity, and do not believe that racial discrimination is a significant issue. They are very skeptical of government and prefer military action to diplomacy. They are concerned about corporate greed.

White women, college education or more
13% of voters, expected to shrink
Members of this group support immigrants, environmental protections, and business regulation. They prefer diplomacy to military action and are very accepting of gay identity.

White men, college education or more
12% of voters, expected to shrink
Members of this group support immigrants. They are very accepting of gay identity and are more pro-business than any other group.

58% D
White, no religious affiliation, no college degree
9% of voters, minimal change expected
Members of this group are the most supportive of gay identity. They prefer diplomacy to military action and support business regulations and environmental protections.

White non-evangelical Protestant, no college degree
9% of voters, expected to shrink
Members of this group are skeptical of government and the least sympathetic to the poor. They are strongly anti-immigrant and do not believe that racial discrimination is a significant issue; they are accepting of gay identity. They are concerned about corporate greed.

White Catholics, no college degree
8% of voters, expected to shrink
Members of this group are skeptical of government and are concerned about corporate greed. They do not believe that racial discrimination is a significant issue; they are accepting of gay identity.

FiveThirtyEight

Who Should Get Parole?

Even the best risk assessments yield probabilities, not certainties. That means they label as "high risk" some people who won't commit another crime and label as "low risk" some people who will. This simulation lets you sort offenders into risk categories based on the results of an assessment. Think we should rarely lock up anyone who wouldn't reoffend? Set the "low risk" threshold high and the "high risk" threshold even higher. Have little tolerance for recidivism? Try the opposite. In the real world, policymakers have to strike a balance. Read more »

1 The prisoners in this simulation are up for parole. Some will reoffend if released and some won't. They each take an assessment, which estimates the chance they will reoffend.

2 Prisoners are placed in one of three categories based on these estimates. **Move the slider** to change the cutoffs for each category. "Low risk" prisoners will be awarded parole. "High risk" prisoners will be denied.

33 67%

3 Some people you let out reoffended. Some people you left in prison wouldn't have. **Are you OK with the results?**

Stop

PAROLE-ELIGIBLE PRISONERS

LOW RISK

MEDIUM RISK

AWARDED PAROLE

42% AWARDED PAROLE AND THEN REOFFENDED

DENIED PAROLE

CNN politics 45 Congress SCOTUS Facts First 2020 Candidates

NATIONAL RESULTS STATE RESULTS EXIT POLLS BALLOT MEASURES Key

THE FORECAST

with Harry Enten

Predictions of the 2018 election results HOUSE SENATE

HARRY'S LATEST UPDATE • (EARLIER TODAY)

Republican Chris Collins predicted to win New York-27

My best estimate gives Chris Collins a four-point win. The worst case for Chris Collins is to lose by 8 points. That means Chris Collins is favored, but the margin of error is wide enough that we shouldn't be surprised if Nate McMurray comes from behind.

MOST LIKELY OUTCOME **Collins** is forecasted to beat **McMurray** by 4pts

This seat is forecasted to swing towards the **Democrats** and to remain under Republican control

Guide to prediction band

Less likely election result More likely

OUTSIDE CHANCE	MOST LIKELY	OUTSIDE CHANCE
McMurray	Collins	Collins
8 POINT WIN	4 POINT WIN	17 POINT WIN

+40 +30 +20 +10 +10 +20 +30 +40

POSSIBLE WITHIN MARGIN OF ERROR

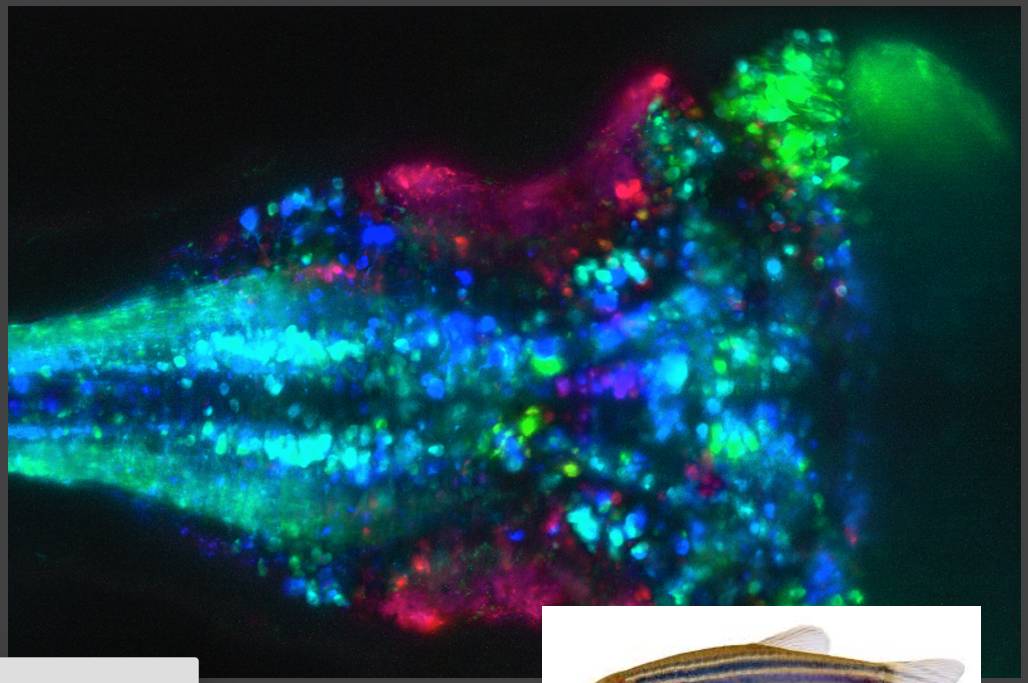
M | OneZero

Listen to this story

Leak of Microsoft Salaries Shows Fight for Higher Compensation

The numbers range from $40,000 to $320,000 and reveal key details about how pay works at big tech companies

Dave Gershgorn Follow
Sep 13, 2019 · 6 min read ★



budo

192.168.1.5:9966/#/results?solStart=4776&solEnd=null&yaw=0&pitch=0&roll=0

SOLS 4776
YPR 0, 0, 0

GTP EMP 128

CLEAR ESTIMATE SEND

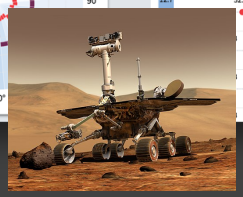
ODY 47761
182 15:34 2017
4776 17:26 SOL
EAST DIRECTION
35.69° MAX EL.
128 RATE
HIDE

MRO 57761
182 12:27 2017
4776 14:25 SOL
WEST DIRECTION
44.5° MAX EL.
128 RATE
HIDE

ODY 47761
182 15:34 2017
4776 17:26 SOL
EAST DIRECTION
35.69° MAX EL.
128 RATE
HIDE

MRO 57761
182 12:27 2017
4776 14:25 SOL
WEST DIRECTION
20.05° MAX EL.
128 RATE
HIDE

47761 10°
47761 10°
47761 20°
47761 20°
47761 30°
47761 30°
57761 10°
57761 20°



Instructors

cse442@cs

Matthew Conlen
PhD Student, CSE
https://mathisonian.com

Assistants

Eunice Jun

Chanwut (Mick) Kittivorawong

Andrew Wang

Zhu (Ruby) Li

JL (Jialiang) Liu

Instructors

cse442@cs

Matthew Conlen
PhD Student, CSE
https://mathisonian.com

Assistants

Eunice Jun

Chanwut (Mick) Kittivorawong

Andrew Wang

Zhu (Ruby) Li

JL (Jialiang) Liu

Office hours
posted tomorrow.

Eunice Jun

emjun@cs.washington.edu

OH: Friday 1-2pm GATES 152 or by appointment

Research

Languages and tools for statistical data analysis designed for non-experts* 🍵 *tea-lang.org*

Ask me about...

Much experience with Python, R, backend web programming

Some experience with Javascript, D3, general web programming

Visualization + functional programming

*Interested in research? Contact me!

Chanwut (Mick) Kittivorawong

chanwutk@cs.washington.edu

Research Contributions:

Vega: Labeling automation in visualizations

Vega-Lite: Enabling syntax for creating composite marks

Technical Experience:

TypeScript, Vega-Lite/Vega, and web development

Andrew Wang

aywang@cs.washington.edu
OH: TBD

Work experience in data vis @ *Microsoft Gaming*

Other: data infrastructure @ *Citadel*, machine learning infrastructure @ *Stripe*

Research experience in computer vision

Experienced in JavaScript, D3, and general web programming

Zhu (Ruby) Li

liz67@cs.washington.edu

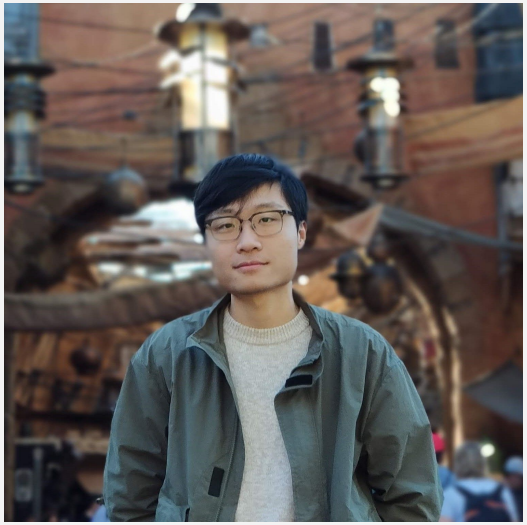
Research on interactive mapping tool for accessibility

Experience with JavaScript, D3, prototyping, and general web programming

Interest in perceptual and cognitive psychology

JL (Jialiang) Liu

jl262@cs.washington.edu

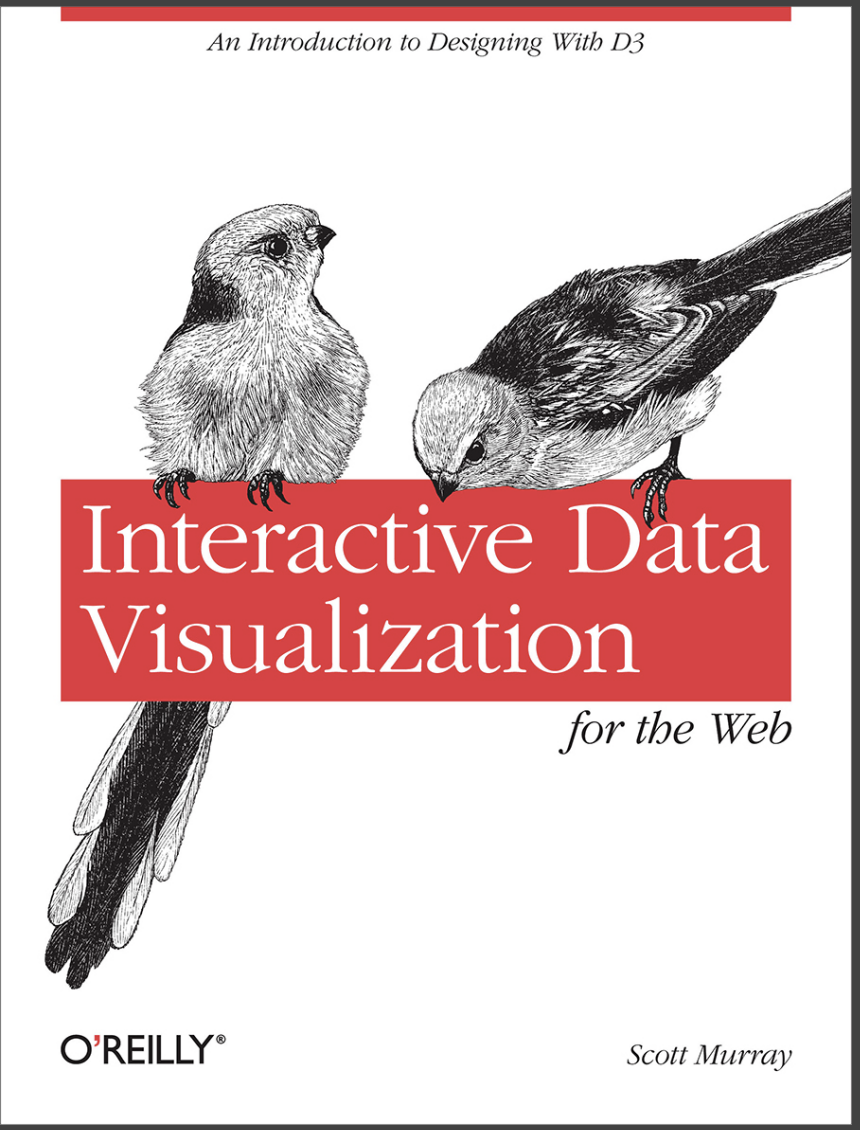


Love making interactive and meaningful websites.

Taken INFO and HCDE courses related to design.

Experience with JavaScript, D3, and general web programming.

Textbook



Interactive Data Visualization for the Web, 2nd Edition

For learning D3!
Book available online.
Code / examples on GitHub.

We will be using **D3 v5**.
https://d3js.org

Readings

Some from D3 book, others from papers & web.

Material in class will loosely follow readings.

Readings should be read by start of class.

Post discussion comments on class Canvas forum.

- One comment per week (ending week 8).
- Comments posted by ***Friday 11:59pm***.
- You have 1 “pass” for the quarter.

Assignments

Class Participation (10%)

A1: Visualization Design (10%)

A2: Exploratory Data Analysis (15%)

A3: Interactive Prototype (25%)

- Peer Evaluation

FP: Final Project (40%)

- Initial Prototypes
- Project Deliverables

Final Project

Produce **interactive web-based visualizations**

Initial **prototype** and **design review**

Final deliverables and **video presentation**

Submit and publish on GitHub

Projects from **previous classes** have been:

- Published as research papers
- Featured in the New York Times
- Released as successful open source projects

Final Project Theme

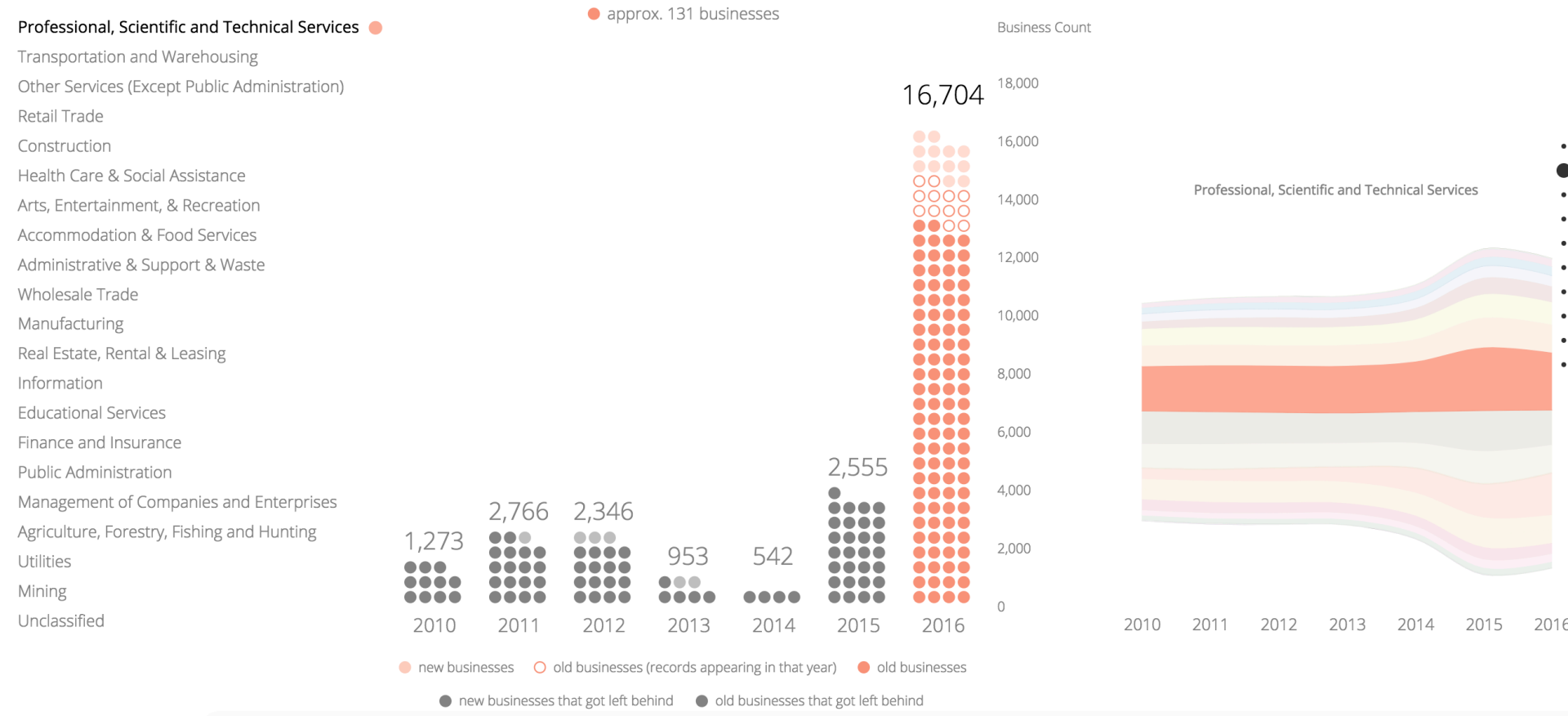
Interactive dashboard for a public audience.

Goal: find data of public interest, design visualizations to explore and communicate it effectively.

- Politics
- Sports
- Climate
- …

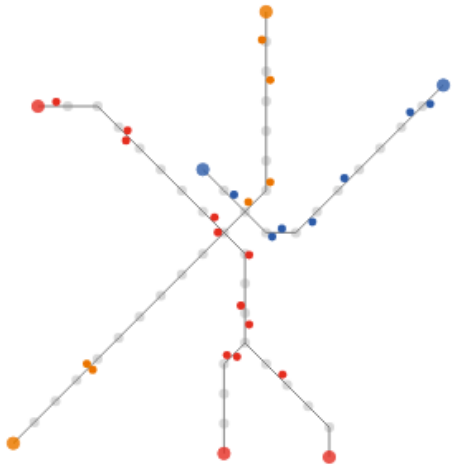
You must identify a topic, dataset, and target audience. We'll provide some examples and potential datasets.

Inspiration…



Change In Times (CSE 442, Spring 2017)

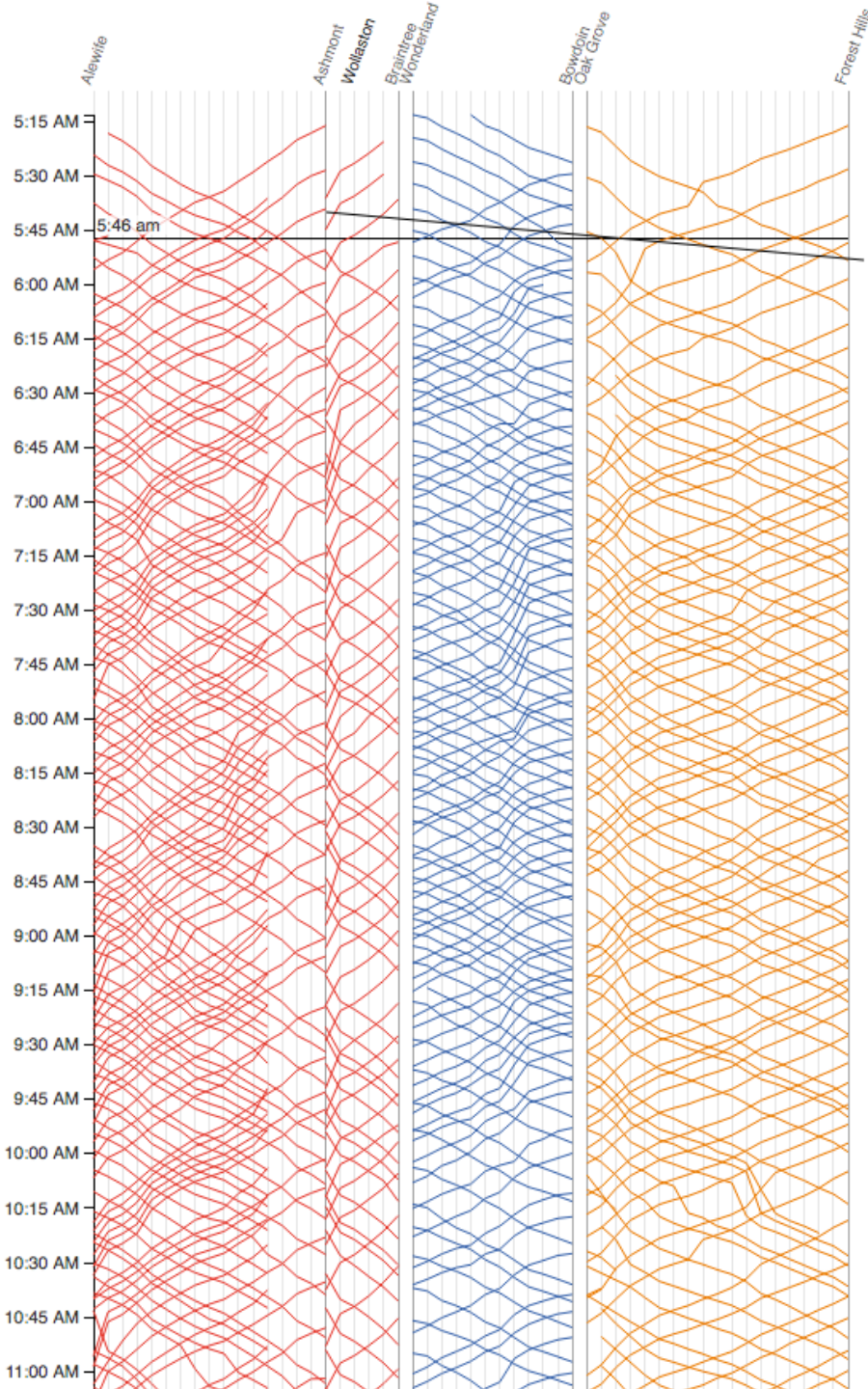
Gunnar Olson, Halden Lin, Lilian Liang, and Shobhit Hathi



Locations of each train on the red, blue, and orange lines at 5:46 am. Hover over the diagram to the right to display trains at a different time.

Trains are on the right side of the track relative to the direction they are moving.

See the morning rush-hour, midday lull, afternoon rush-hour, and the evening lull.



Service starts at 5AM on Monday morning. Each line represents the path of one train. Time continues downward, so steeper lines indicate slower trains.

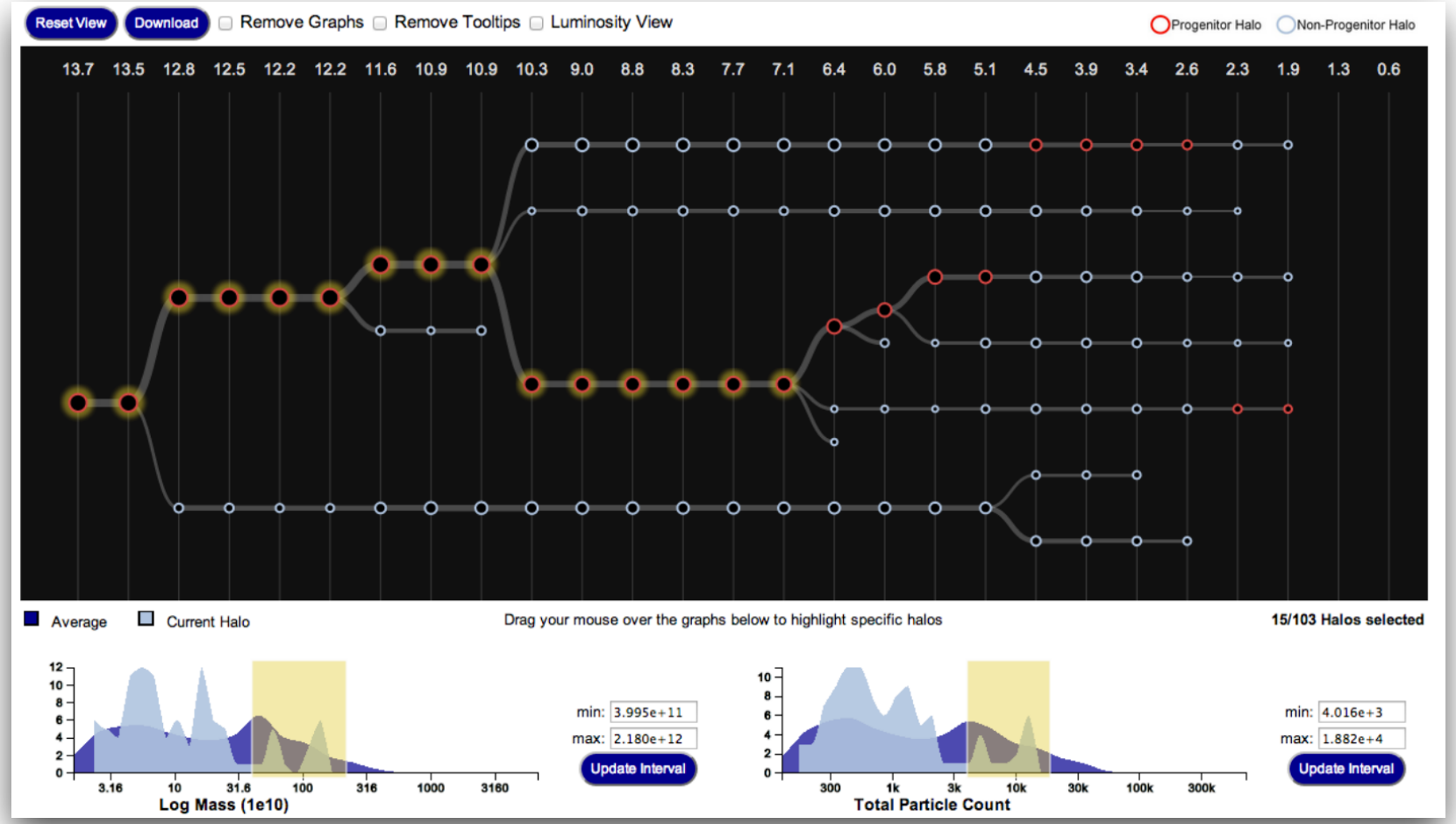
Since the red line splits, we show the Ashmont branch first then the Braintree branch. Trains on the Braintree branch "jump over" the Ashmont branch.

Train frequency increases around 6:30AM as morning rush hour begins.

MBTA Viz

Barry & Card

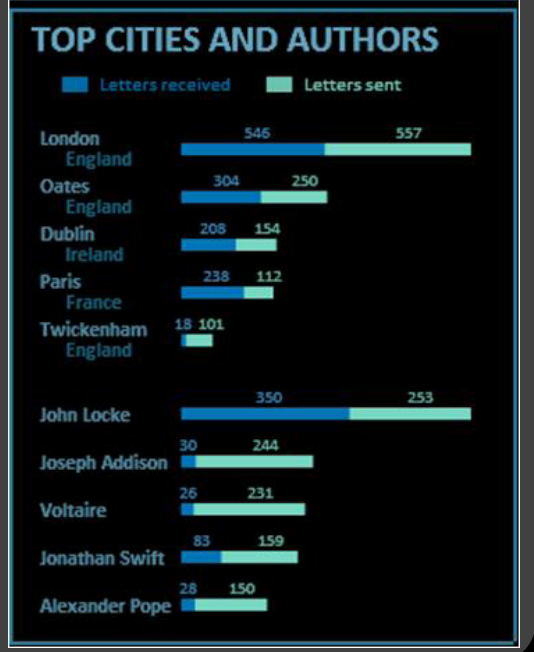
Visualizing Galaxy Merger Trees



S. Loebman, **J. Ortiz**, **L. Orr**, M. Balazinska, T. Quinn et al. [SIGMOD '14]

Visualizing the Republic of Letters

Daniel Chang, Yuankai Ge, Shiwei Song



Questions?

A1: Visualization Design

Design a static visualization for a data set.

Every 10 years, the census documents the demographic make-up of the U.S., influencing congressional districting and social services. This dataset contains a summary of census data for two years a century apart: 1900 and 2000.

You must choose the message you want to convey. What question(s) do you want to answer? What insight do you want to communicate?

A1: Visualization Design

Pick a **guiding question**, use it to title your vis.
Design a **static visualization** for that question.
You are free to **use any tools** (inc. pen & paper).

Deliverables (upload via Canvas; see A1 page)
Image of your visualization (PNG or JPG format)
Short description + design rationale (≤ 4 paragraphs)

Due by **11:59 pm, Tues Jan 14**.