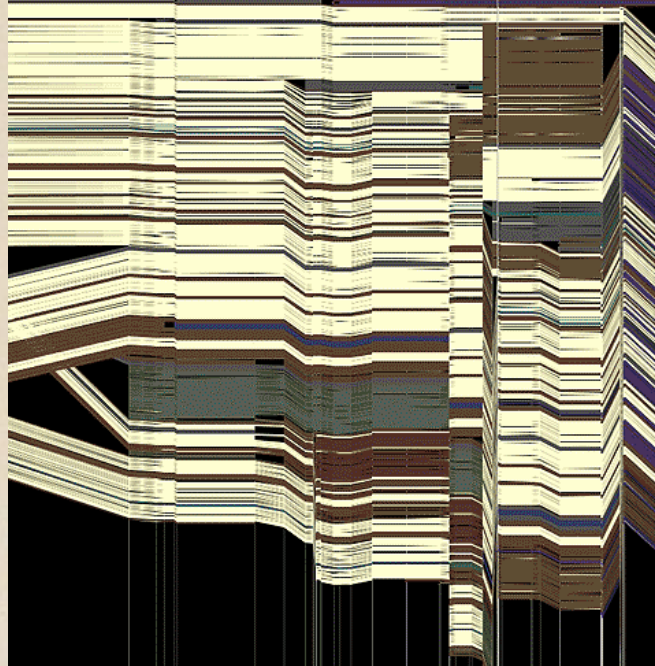
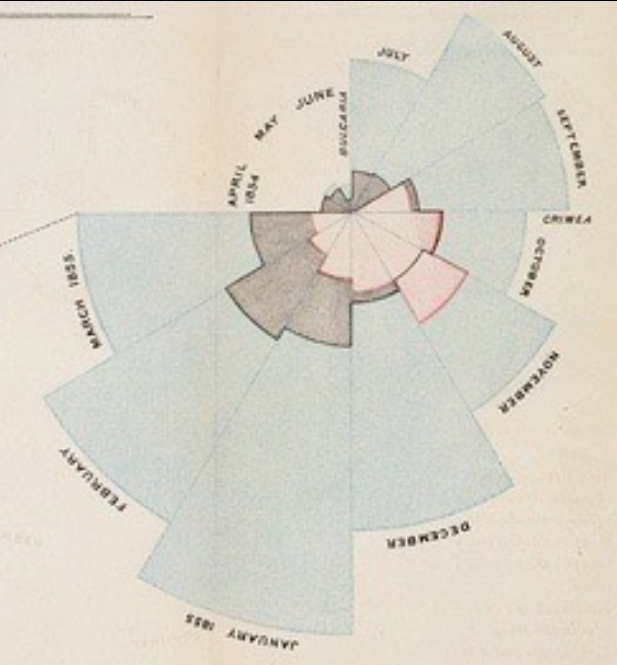


CSE 442 - Data Visualization

The Value of Visualization

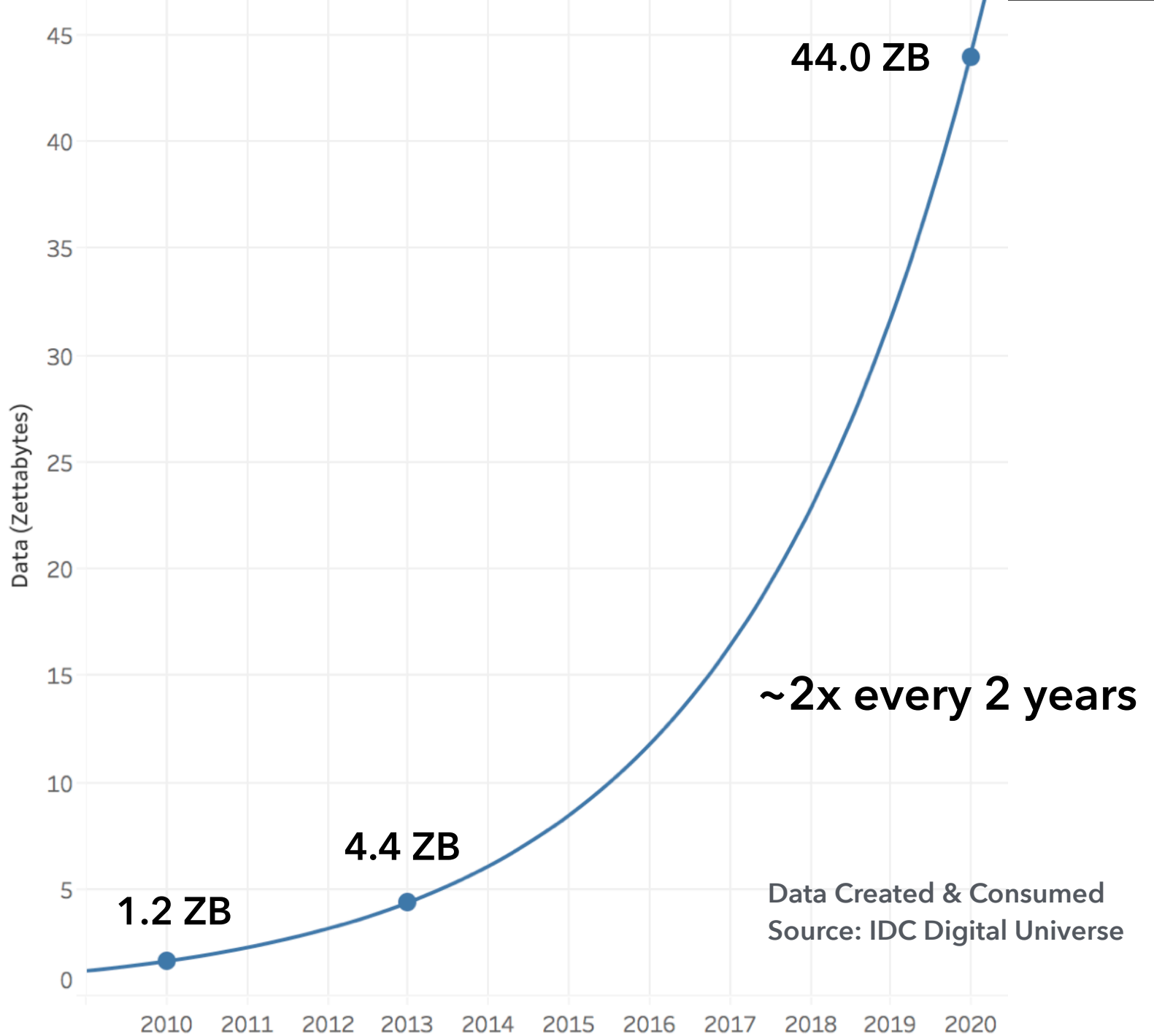


Jeffrey Heer University of Washington

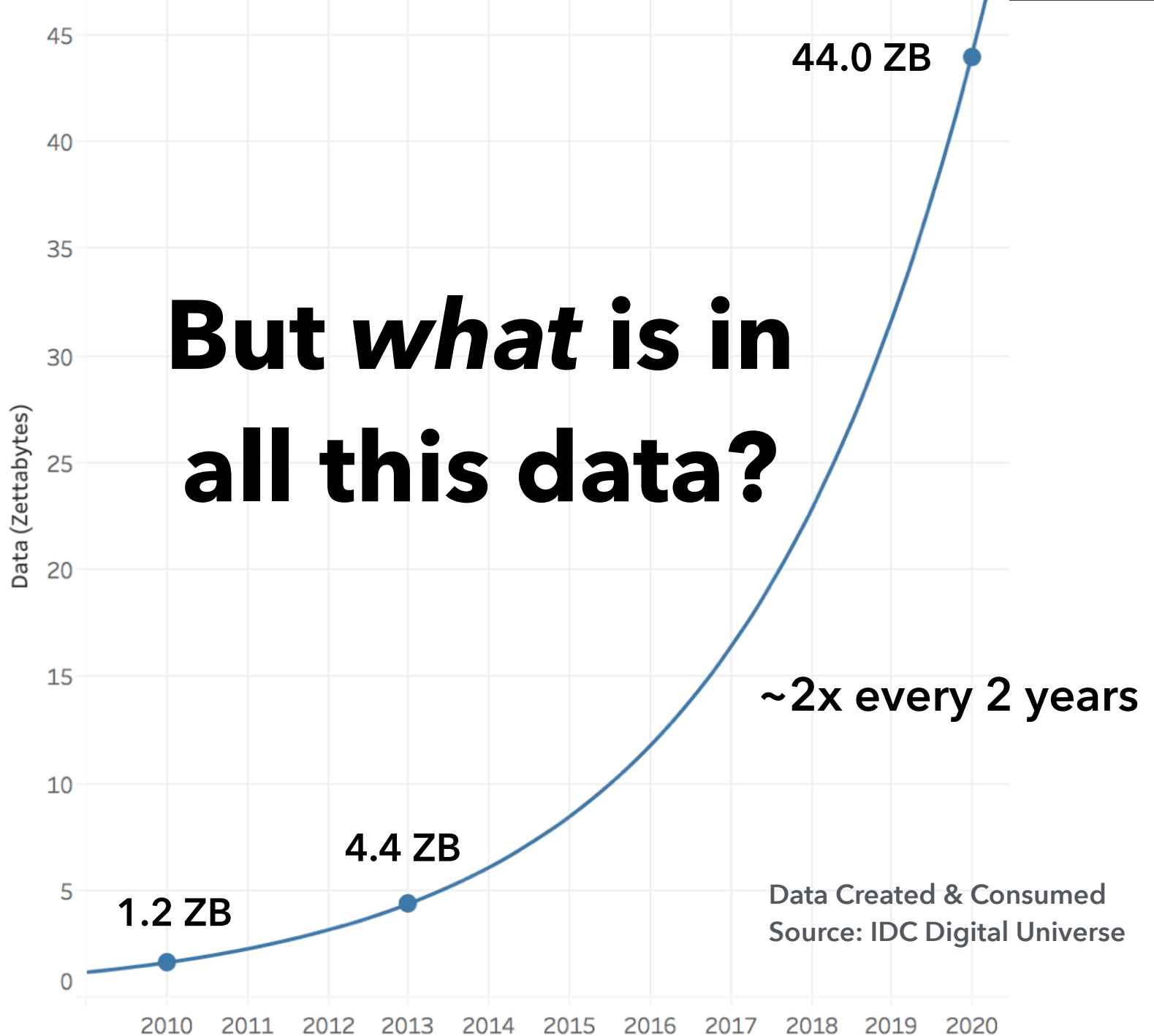
**How much data (bytes)
did we produce in 2010?**

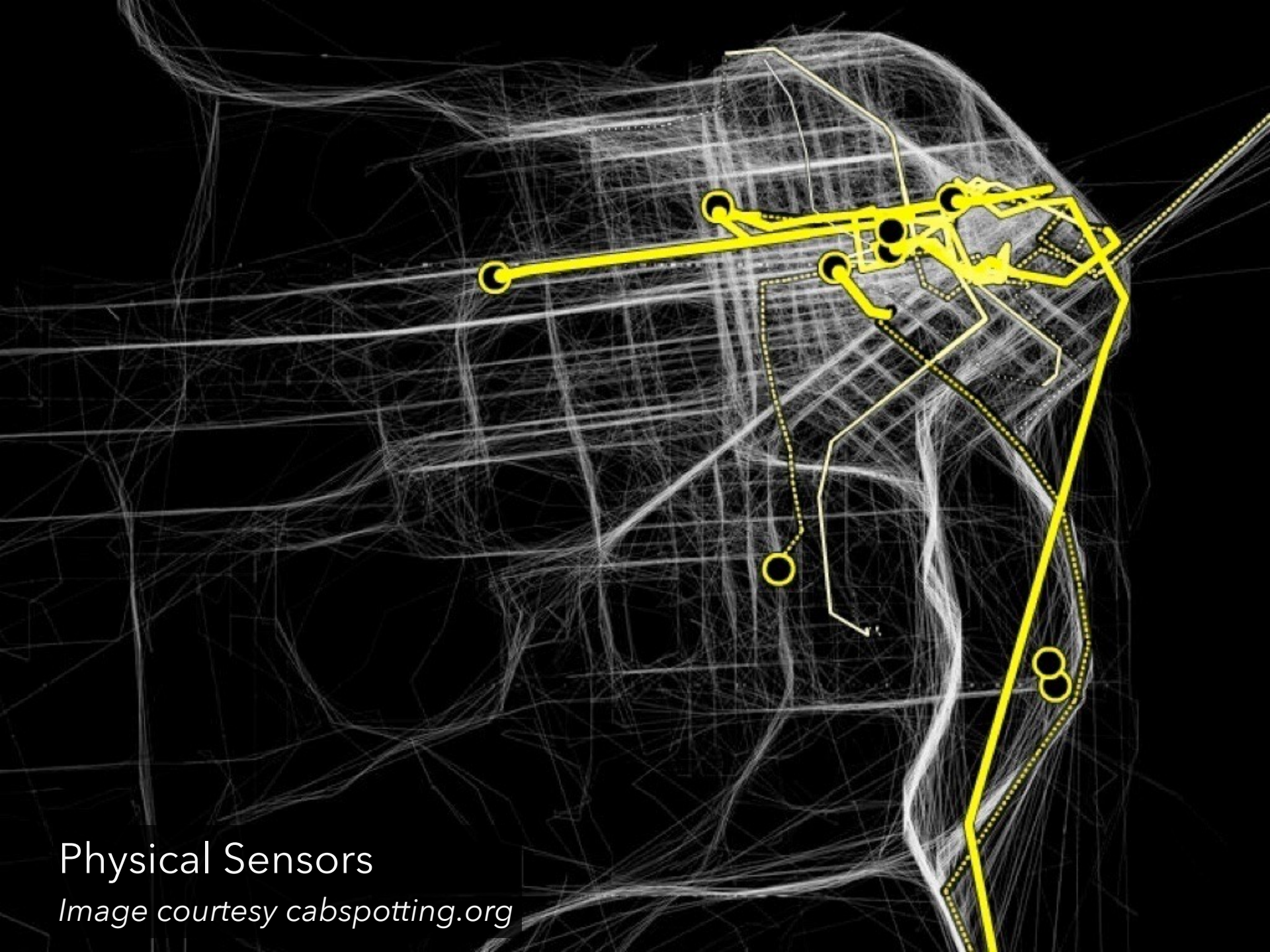
2010: 1,200 exabytes
and exponential growth...

Gantz et al., 2008, 2010



**But *what* is in
all this data?**



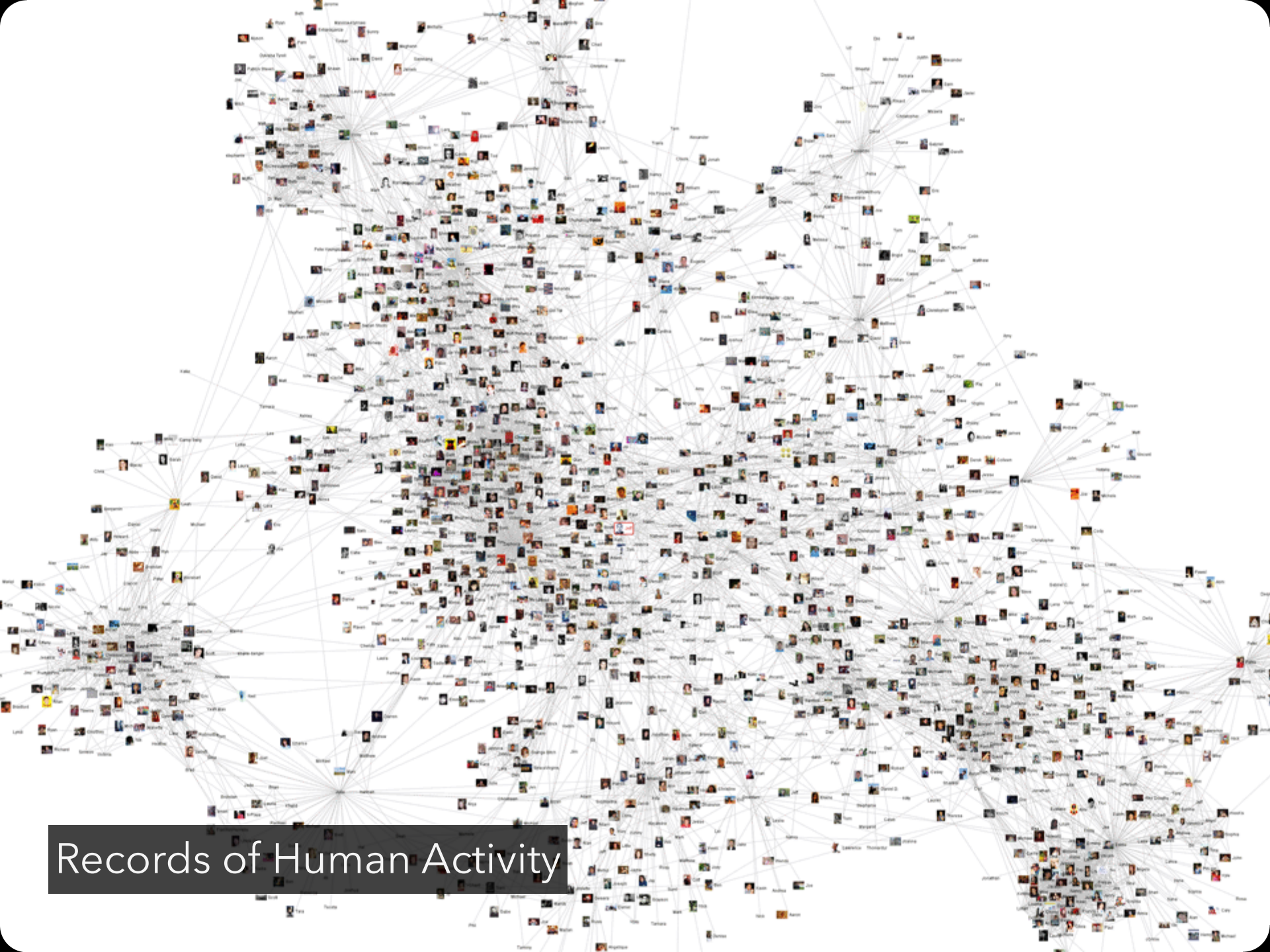


Physical Sensors

Image courtesy cabspotting.org



Health & Medicine



Records of Human Activity

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

But wait!

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, **"free" to whom?** because now we really do have **essentially free and ubiquitous data**. So the complimentary scarce factor is **"ubiquitous" about whom?** and extract value from it. **...to whose benefit?**

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



My Facebook Was Breached by Cambridge Analytica. Was Yours?

How to find out if you are one of the 87 million victims

ROBINSON MEYER | APR 10, 2018 | TECHNOLOGY

Share Tweet ...

TEXT SIZE
- +



Psychology's Replication Crisis Can't Be Wished Away

It has a real and heartbreaking cost.

ED YONG | MAR 4, 2016 | SCIENCE

Share Tweet ...

TEXT SIZE
- +

High potential for data abuse...

How might we use **visualization**
to **empower understanding** of
data and analysis processes?

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

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

Set B

X	Y
10	9.14
8	8.14
13	8.74
9	8.77
11	9.26
14	8.1
6	6.13
4	3.1
12	9.11
7	7.26
5	4.74

Set C

X	Y
10	7.46
8	6.77
13	12.74
9	7.11
11	7.81
14	8.84
6	6.08
4	5.39
12	8.15
7	6.42
5	5.73

Set D

X	Y
8	6.58
8	5.76
8	7.71
8	8.84
8	8.47
8	7.04
8	5.25
19	12.5
8	5.56
8	7.91
8	6.89

Summary Statistics

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

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

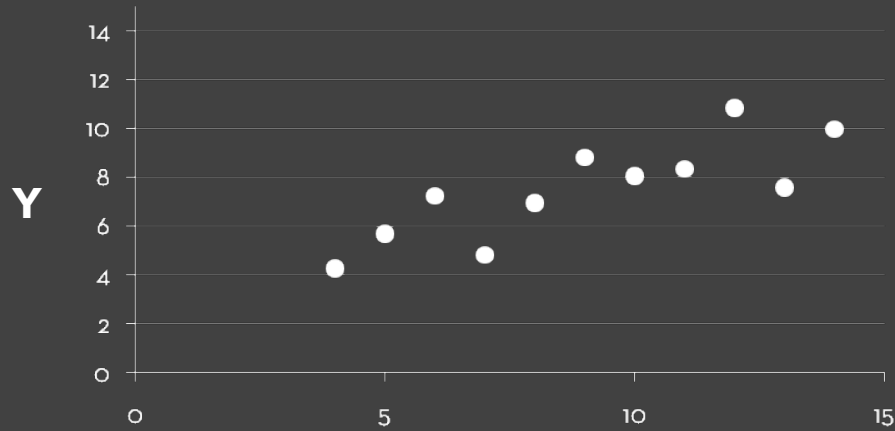
Linear Regression

$$Y = 3 + 0.5 X$$

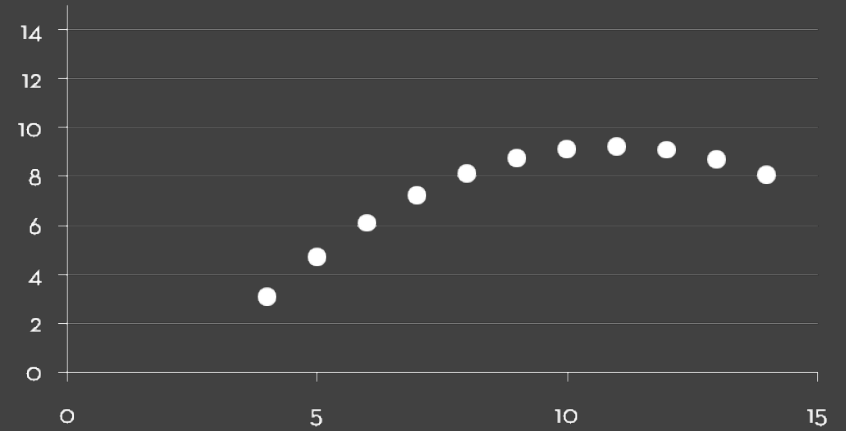
$$R^2 = 0.67$$

[Anscombe 1973]

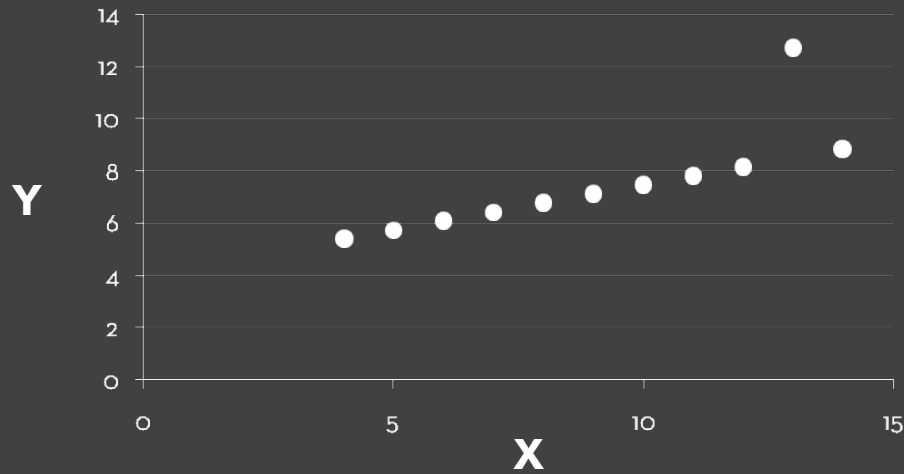
Set A



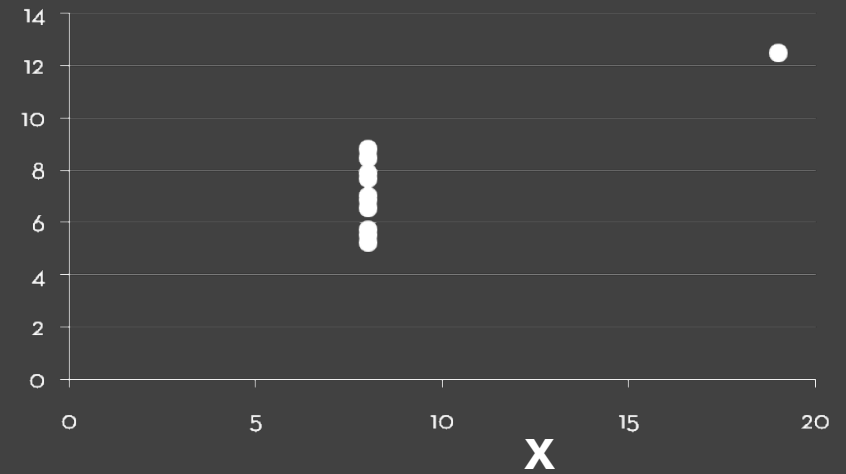
Set B

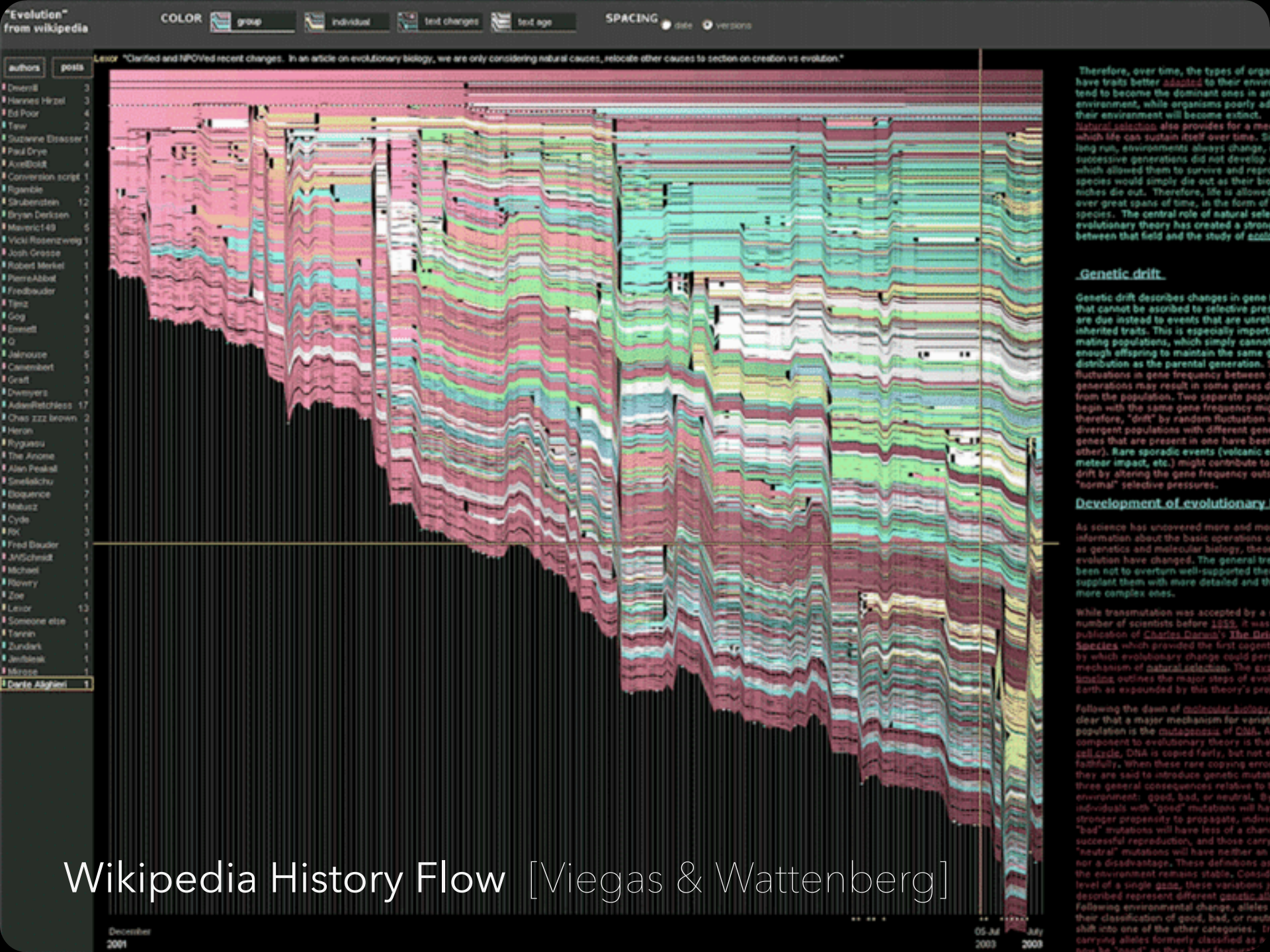


Set C



Set D





Wikipedia History Flow [Viegas & Wattenberg]

A complex visualization of Wikipedia's edit history. The image shows a dense, multi-layered structure of horizontal lines in various colors (purple, pink, orange, green, blue, and black) that form a jagged, wavy pattern. The lines are stacked vertically, creating a sense of depth and complexity. The overall shape is roughly rectangular but with many irregular edges, suggesting a large volume of data. The colors are vibrant and contrast sharply against the dark background.

Edit War...

Wikipedia History Flow [Viegas & Wattenberg]

Why Create Visualizations?

Why Create Visualizations?

Answer questions (or discover them)

Make decisions

See data in context

Expand memory

Support graphical calculation

Find patterns

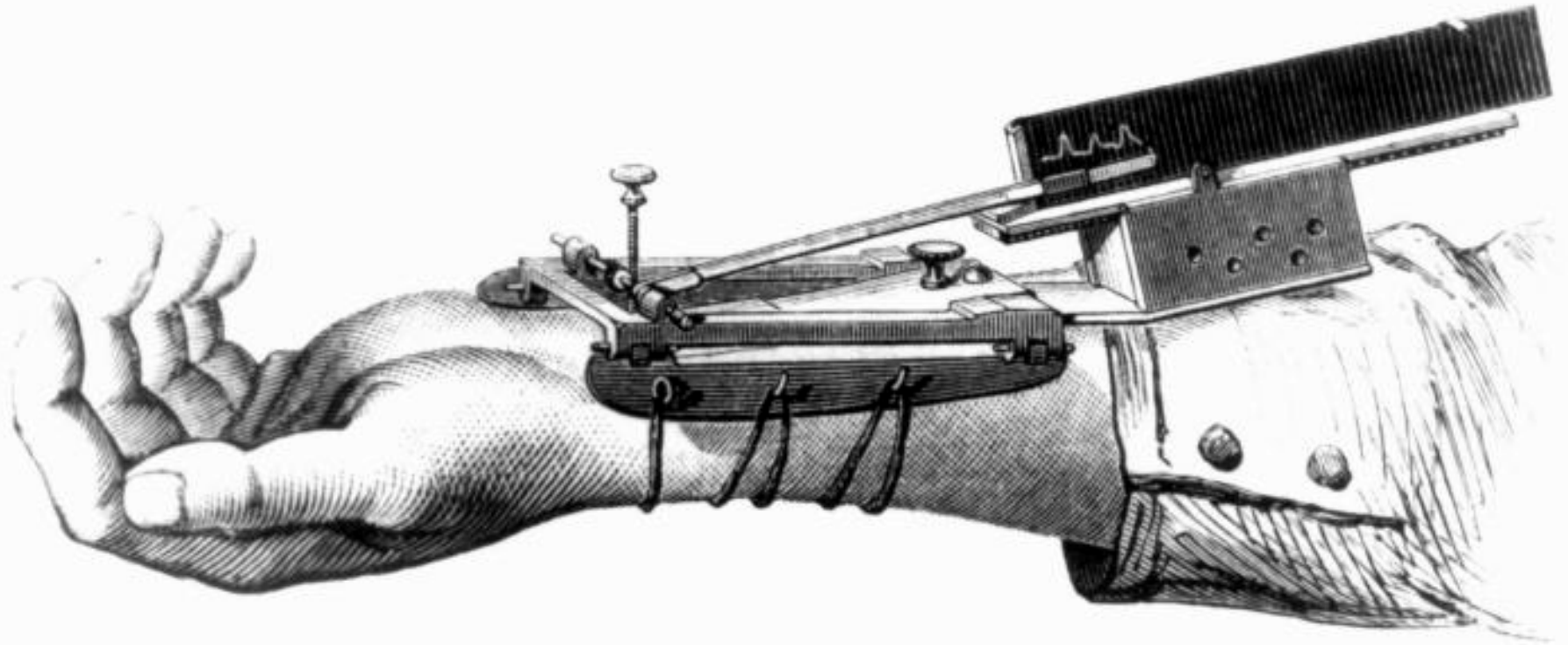
Present argument or tell a story

Inspire

Record Information



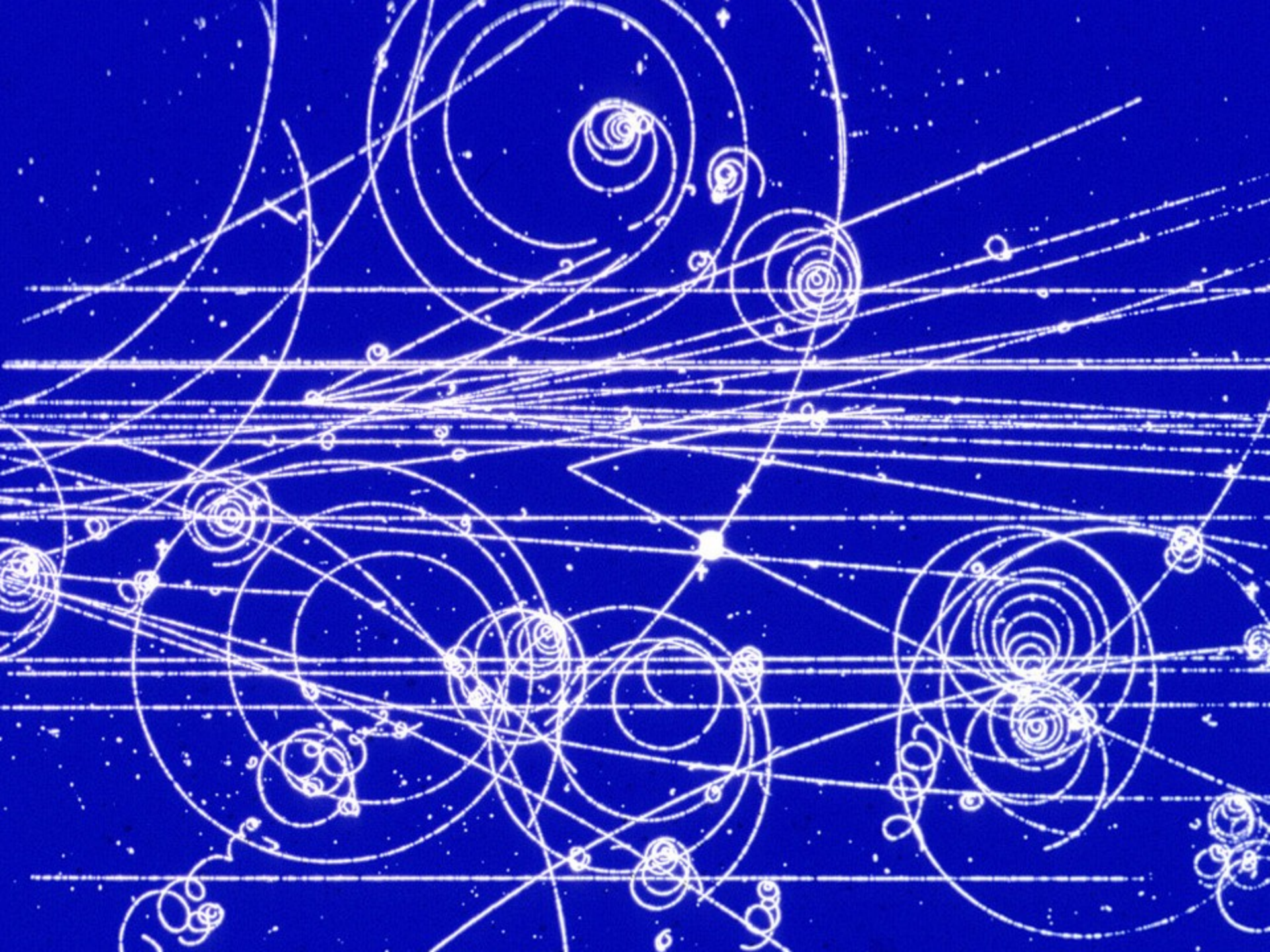
Gallop, Bay Horse "Daisy" [Muybridge]



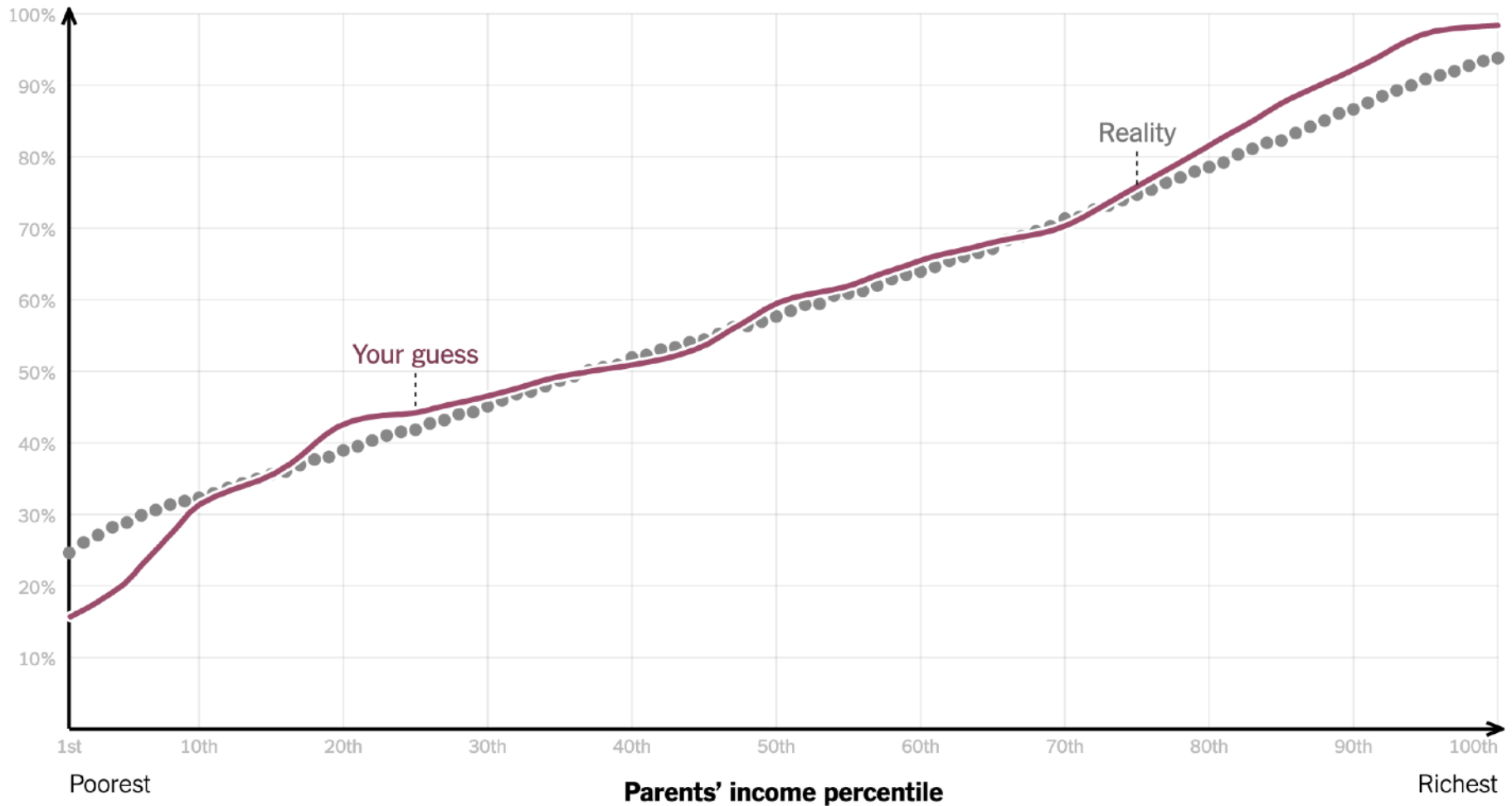
1.

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]



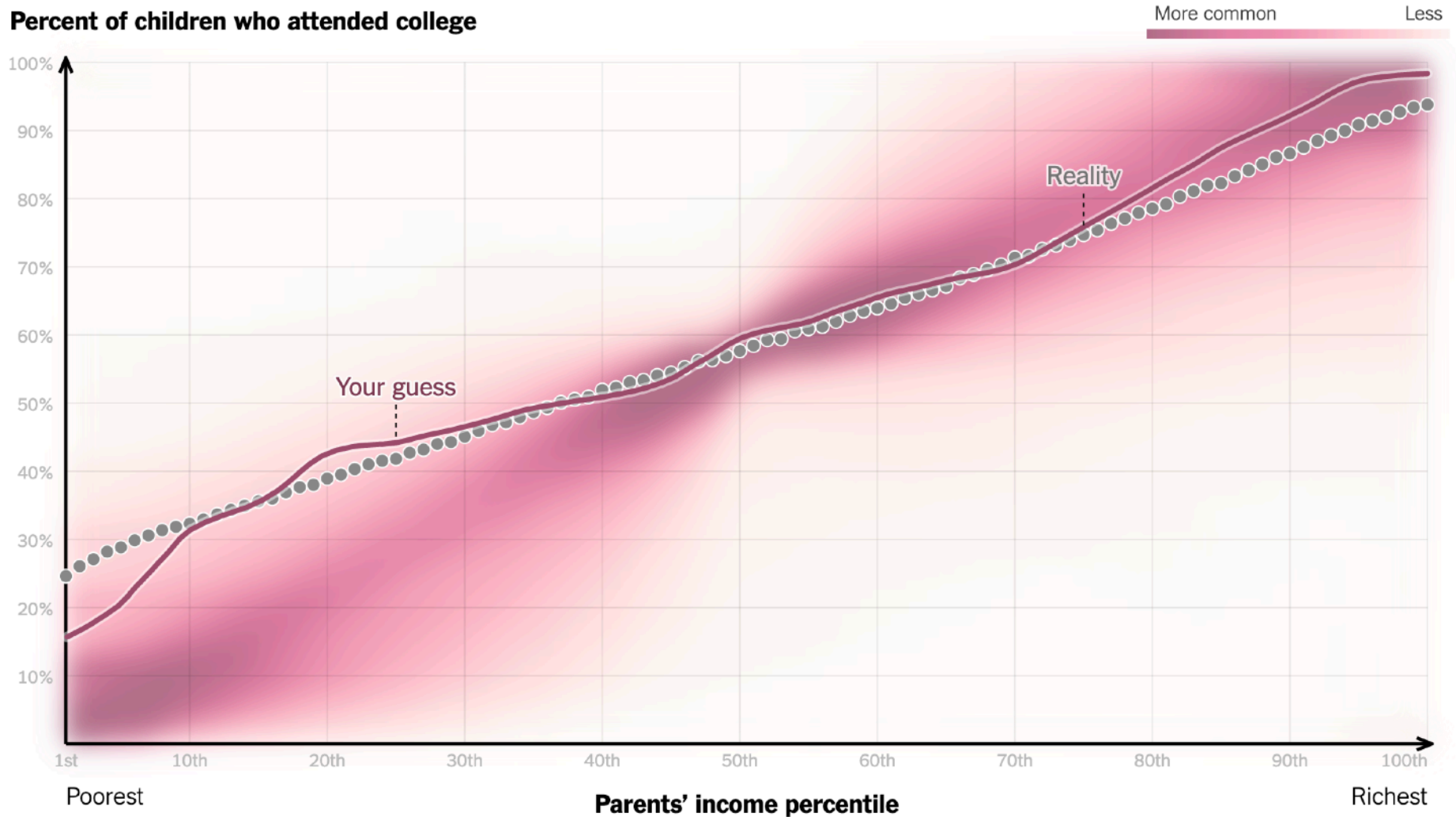
Percent of children who attended college



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

[New York Times, May 28, 2015]

Percent of children who attended college



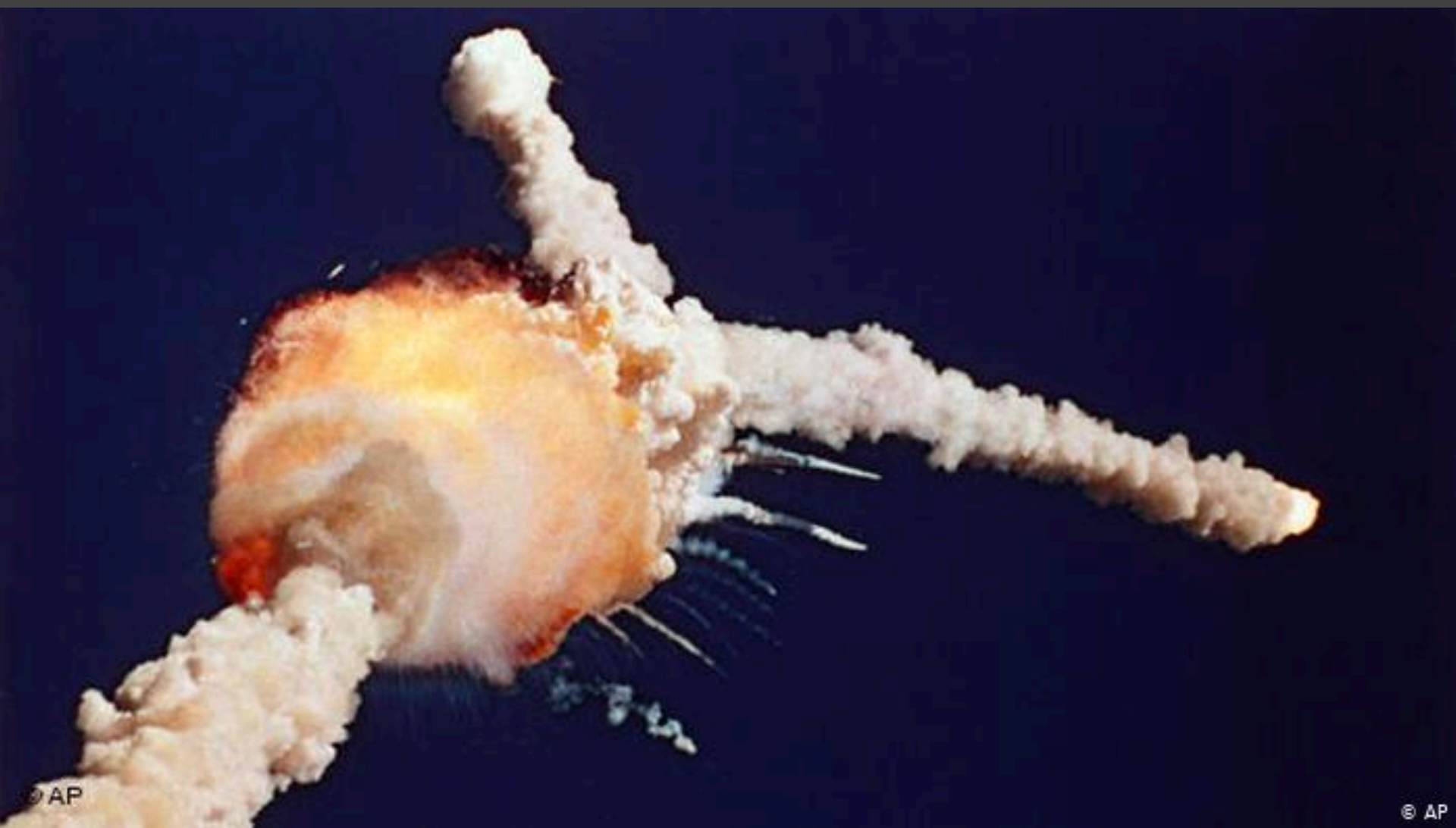
You Draw It: How Family Income Predicts Children's College Chances
[New York Times, May 28, 2015]

Support Reasoning



© AP

© AP



© AP

© AP

HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

	SRM No.	Cross Sectional View			Top View		Clocking Location (deg)
		Erosion Depth (in.)	Perimeter Affected (deg)	Nominal Dia. (in.)	Length Of Max Erosion (in.)	Total Heat Affected Length (in.)	
Oct 30, 1985	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°
8-	51C LH Forward Field**	15A 0.010	154.0	0.280	4.25	5.25	163
5-	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
July	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

- o 2 CASE JOINTS (80°), (110°) ARC
- o MUCH WORSE VISUALLY THAN SRM-22

SRM 22 BLOW-BY

- o 2 CASE JOINTS (30-40°)

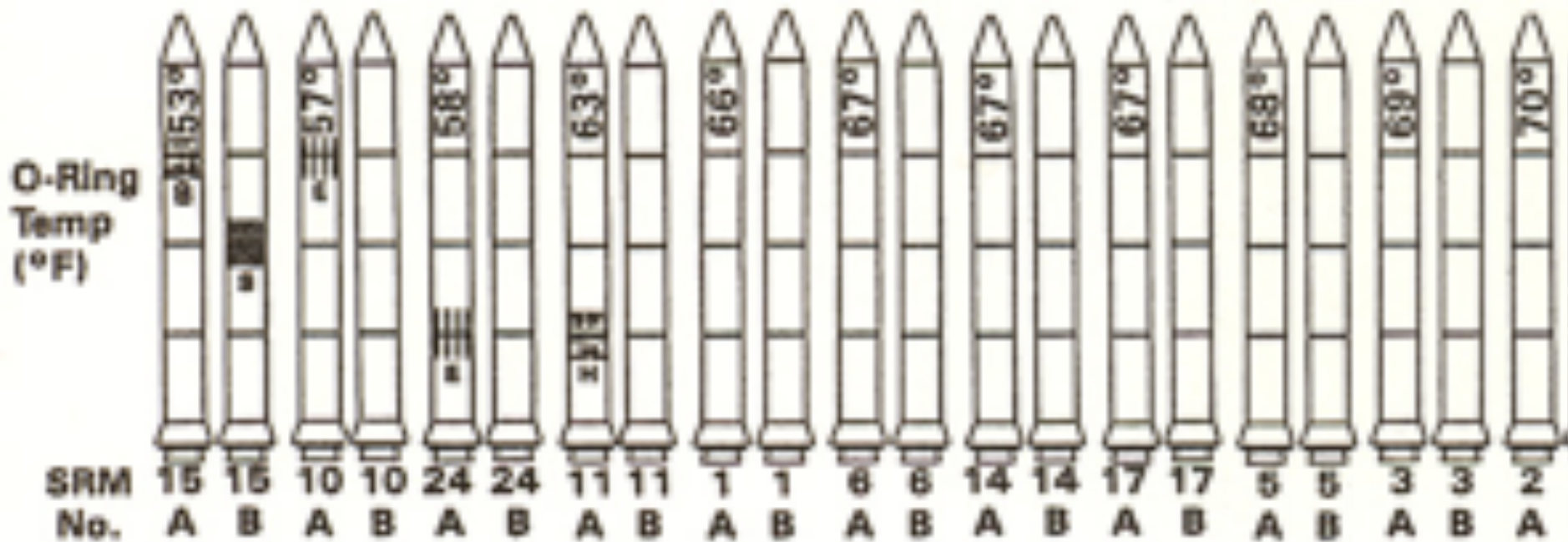
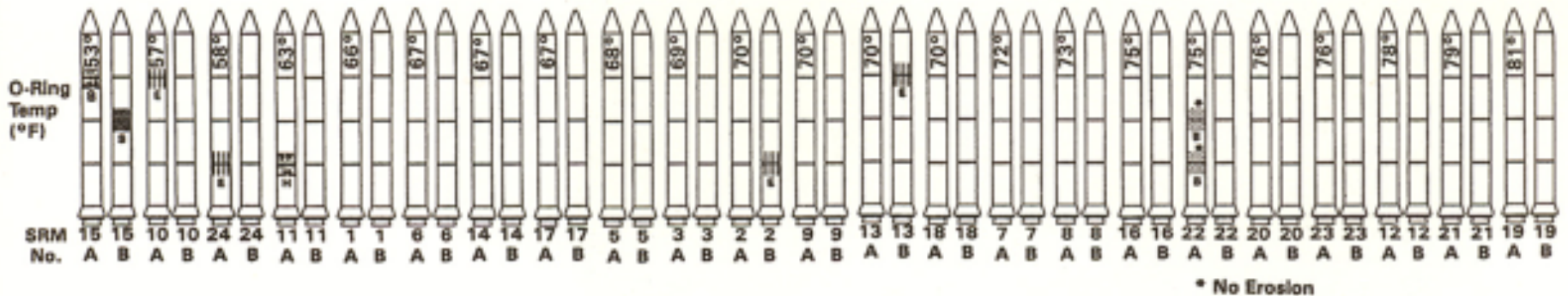
SRM-13A, 15, 16A, 18, 23A 24A

- o NOZZLE BLOW-BY

HISTORY OF O-RING TEMPERATURES (DEGREES - F)

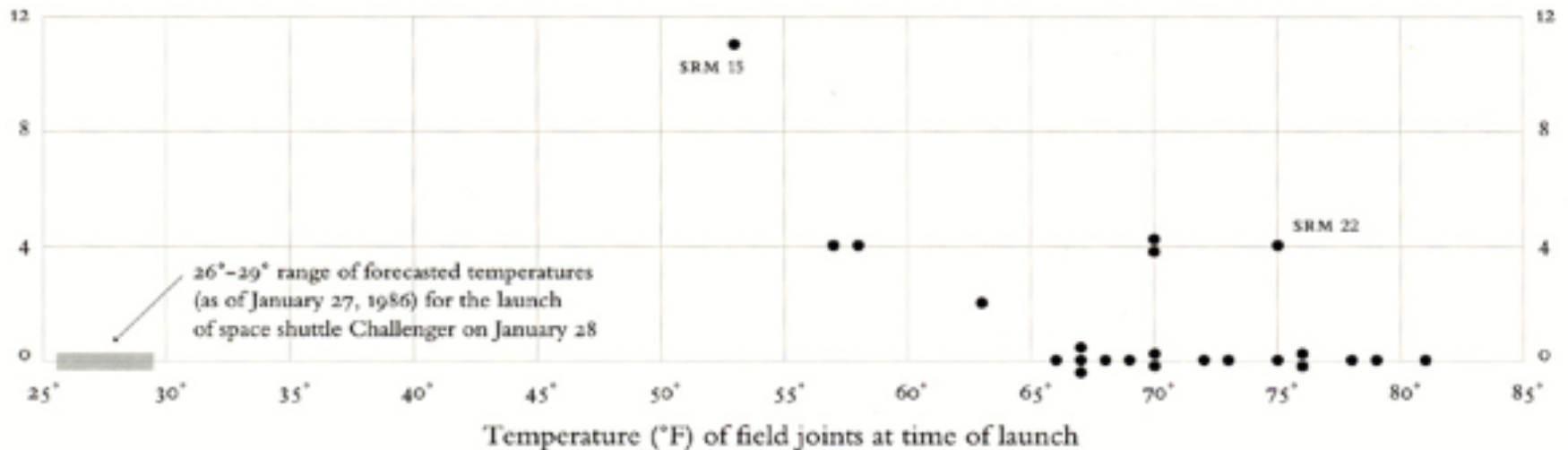
MOTOR	MBT	AMB	O-RING	WIND
DM-4	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29	10 MPH
			27	25 MPH

Make Decisions: Challenger



Make Decisions: Challenger

O-ring damage
index, each launch



**But wait! What is an appropriate "damage index"?
Which temperatures, O-ring or outside air?**

Chart of temperatures vs. O-ring damage [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]

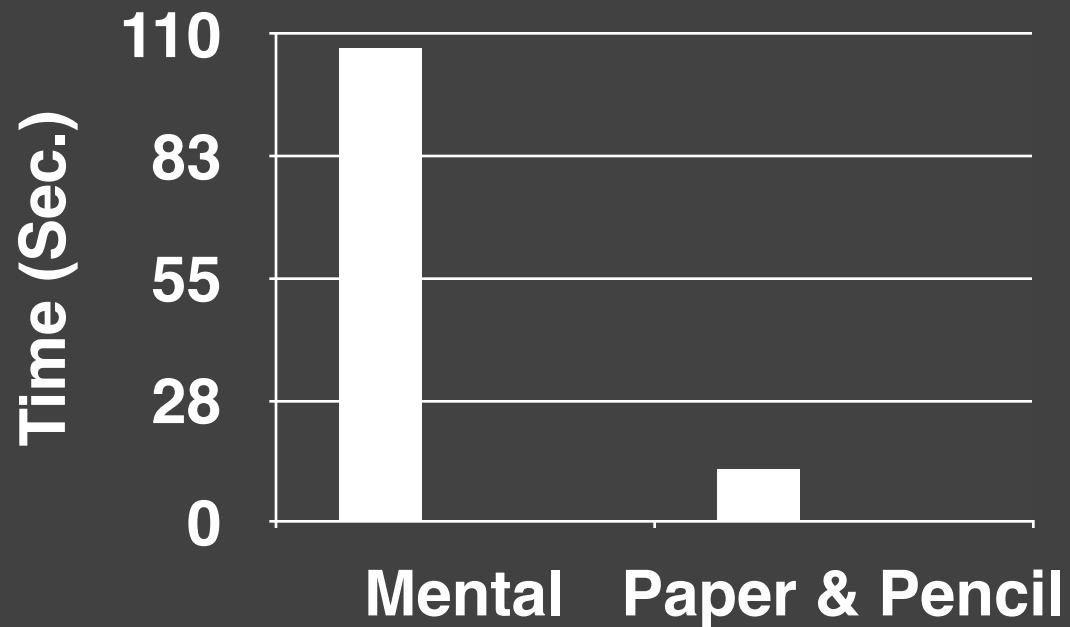
Expand Memory: Multiplication

Expand Memory: Multiplication

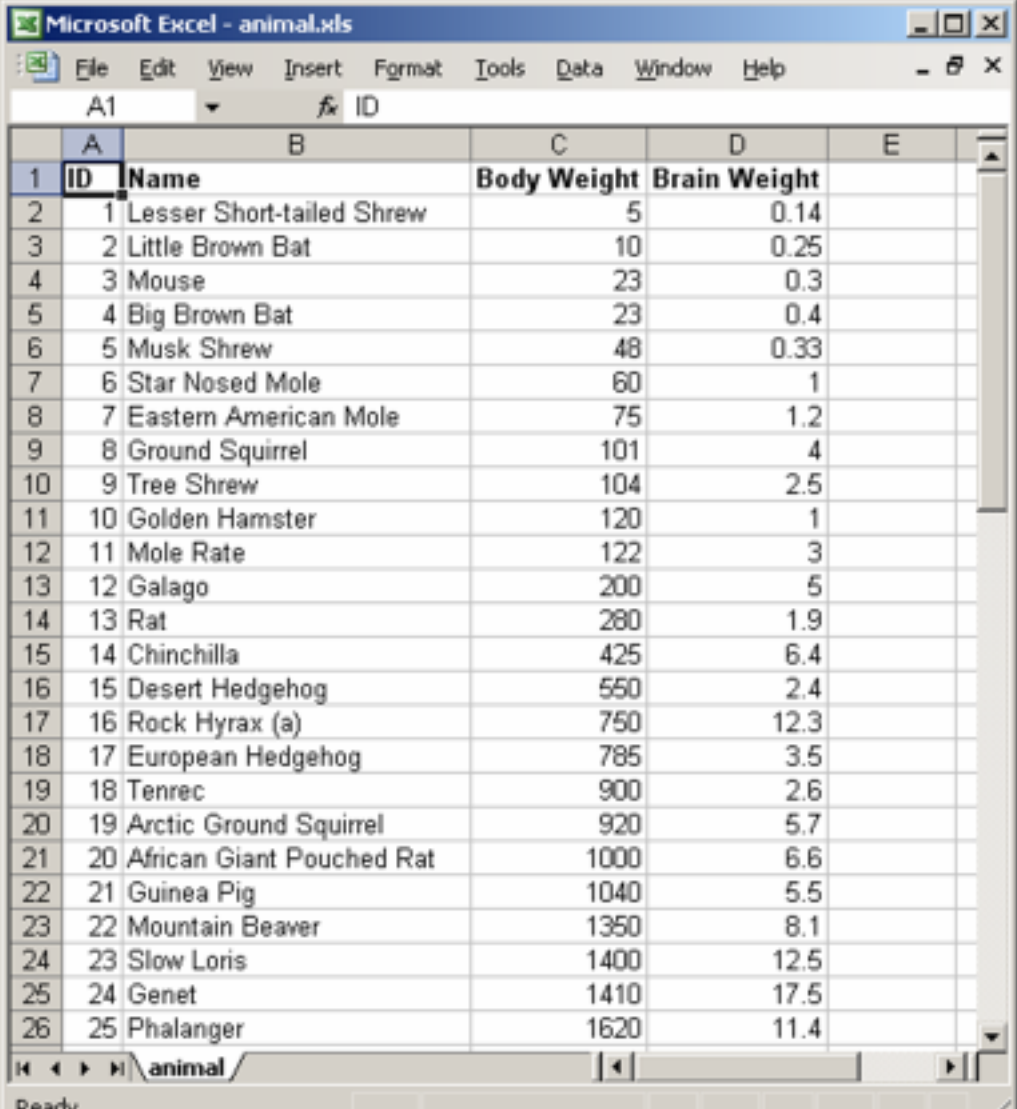
$$\begin{array}{r} 34 \\ \times 72 \\ \hline \end{array}$$

Expand Memory: Multiplication

$$\begin{array}{r} 34 \\ \times 72 \\ \hline 68 \\ 2380 \\ \hline 2448 \end{array}$$

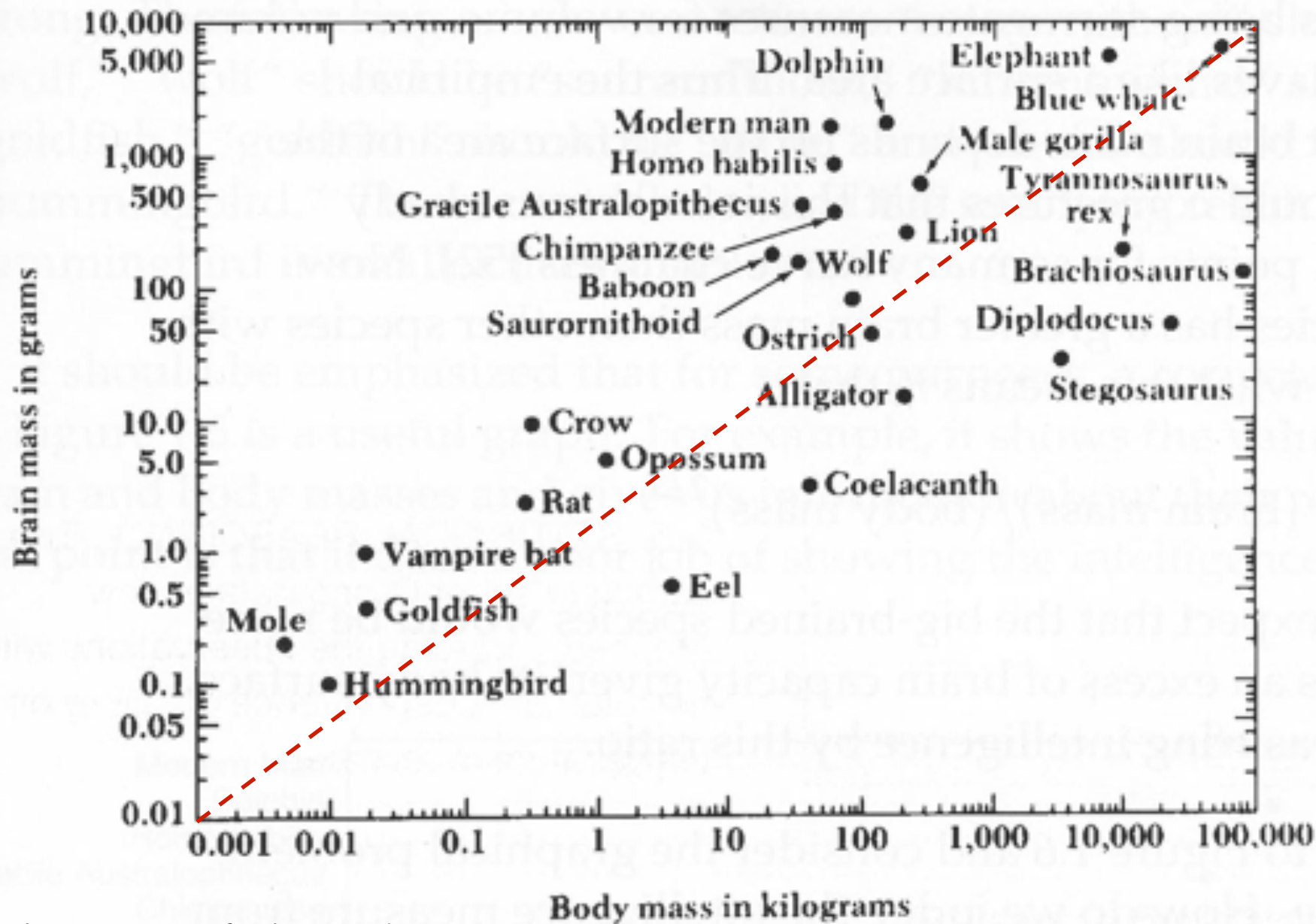


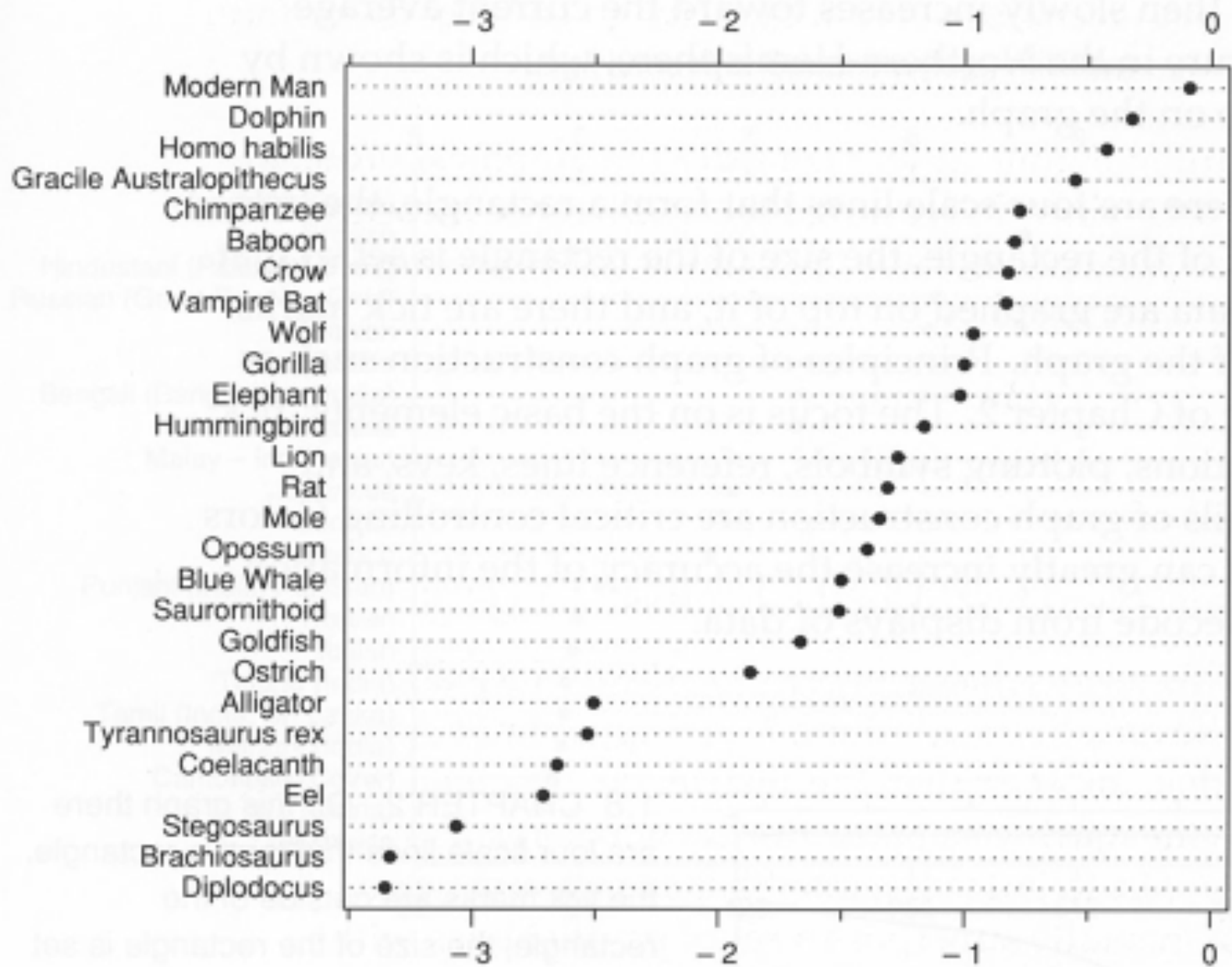
Answer Questions: Brain Power?



A screenshot of a Microsoft Excel spreadsheet titled "animal.xls". The spreadsheet contains a table with 4 columns: ID, Name, Body Weight, and Brain Weight. The data is organized into rows, with the first row (row 1) serving as the header. The table lists 25 different animals, each with a unique ID, a name, a body weight, and a brain weight. The interface shows the standard Excel menu bar (File, Edit, View, Insert, Format, Tools, Data, Window, Help) and the status bar at the bottom indicates "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





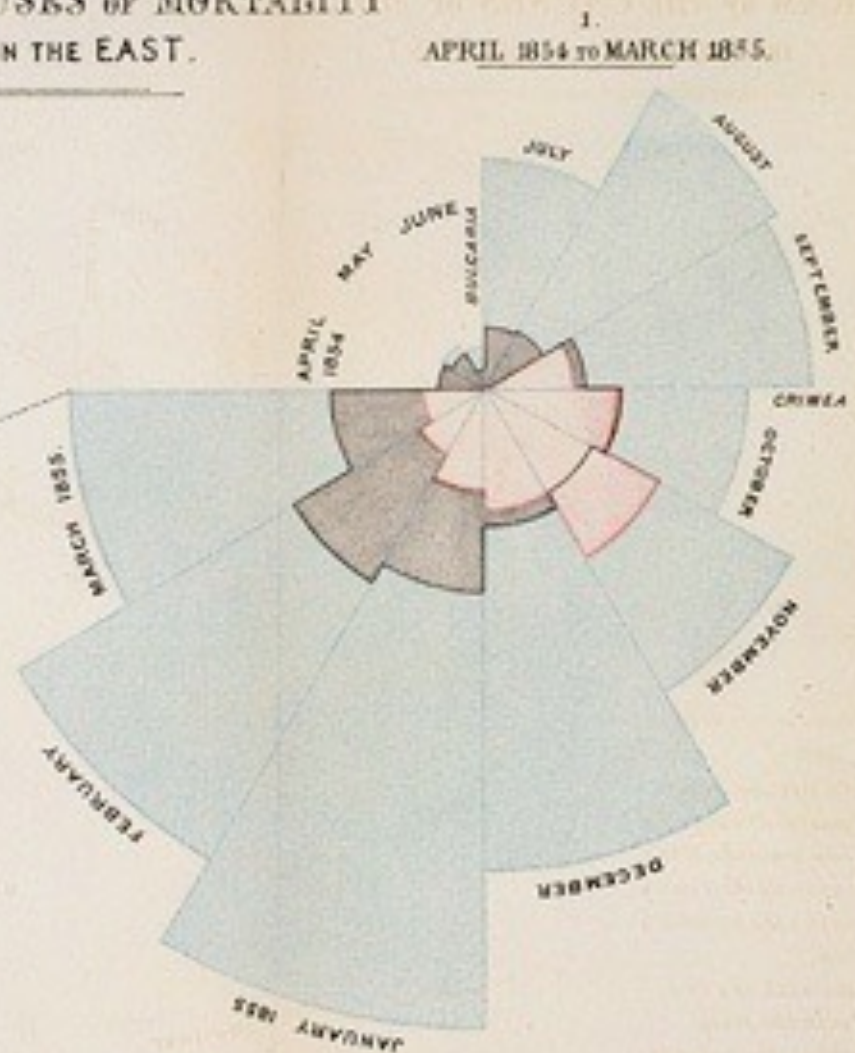
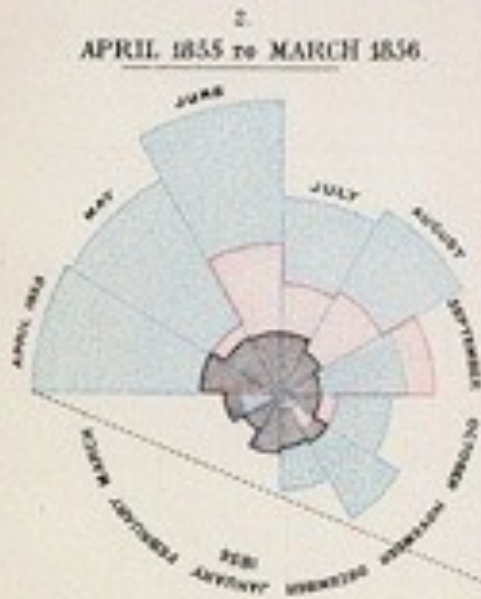
The Elements of Graphing Data

[Cleveland]

$\log_{10} \text{ Brain Weight} - \frac{2}{3} \log_{10} \text{ Body Weight}$

Convey Information

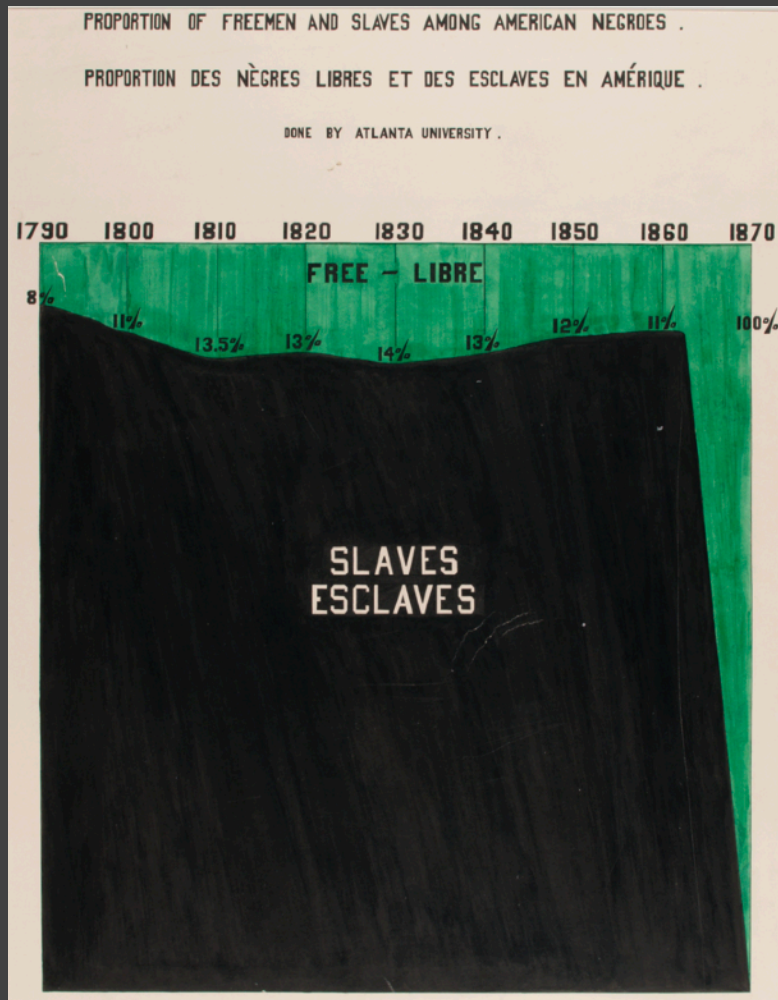
DIAGRAM OF THE CAUSES OF MORTALITY IN THE ARMY IN THE EAST.



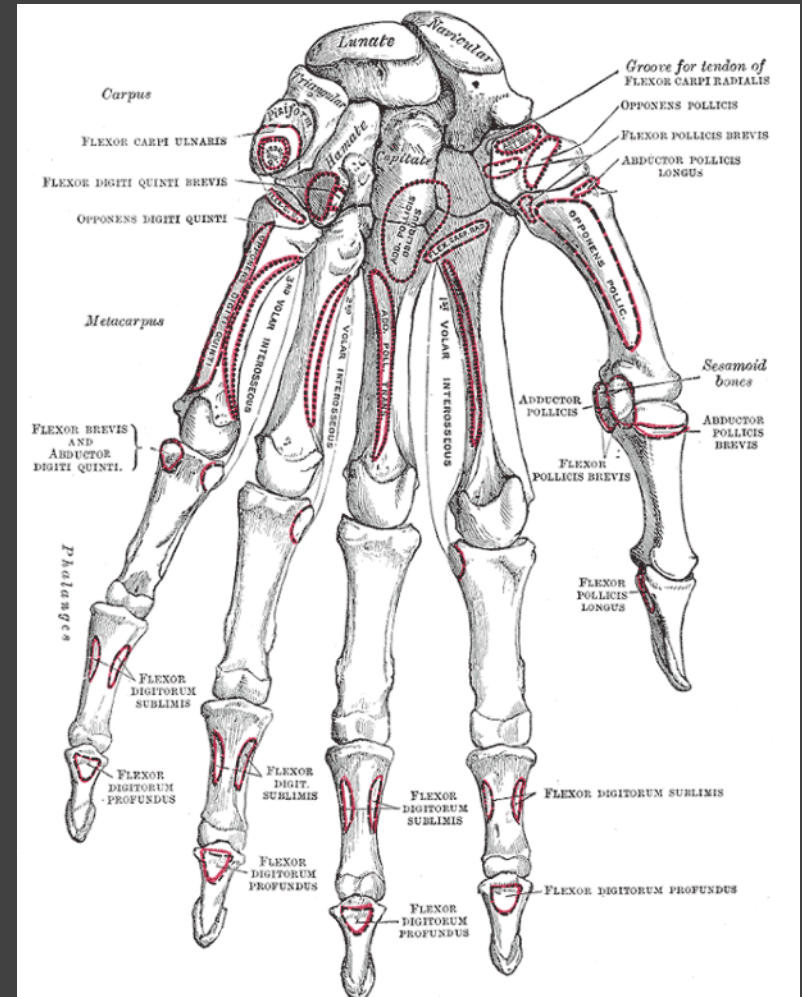
"to affect thro' the Eyes
what we fail to convey to
the public through their
word-proof ears"

1856 "Coxcomb" of Crimean War Deaths, Florence Nightingale

Communicate, Inform, Inspire



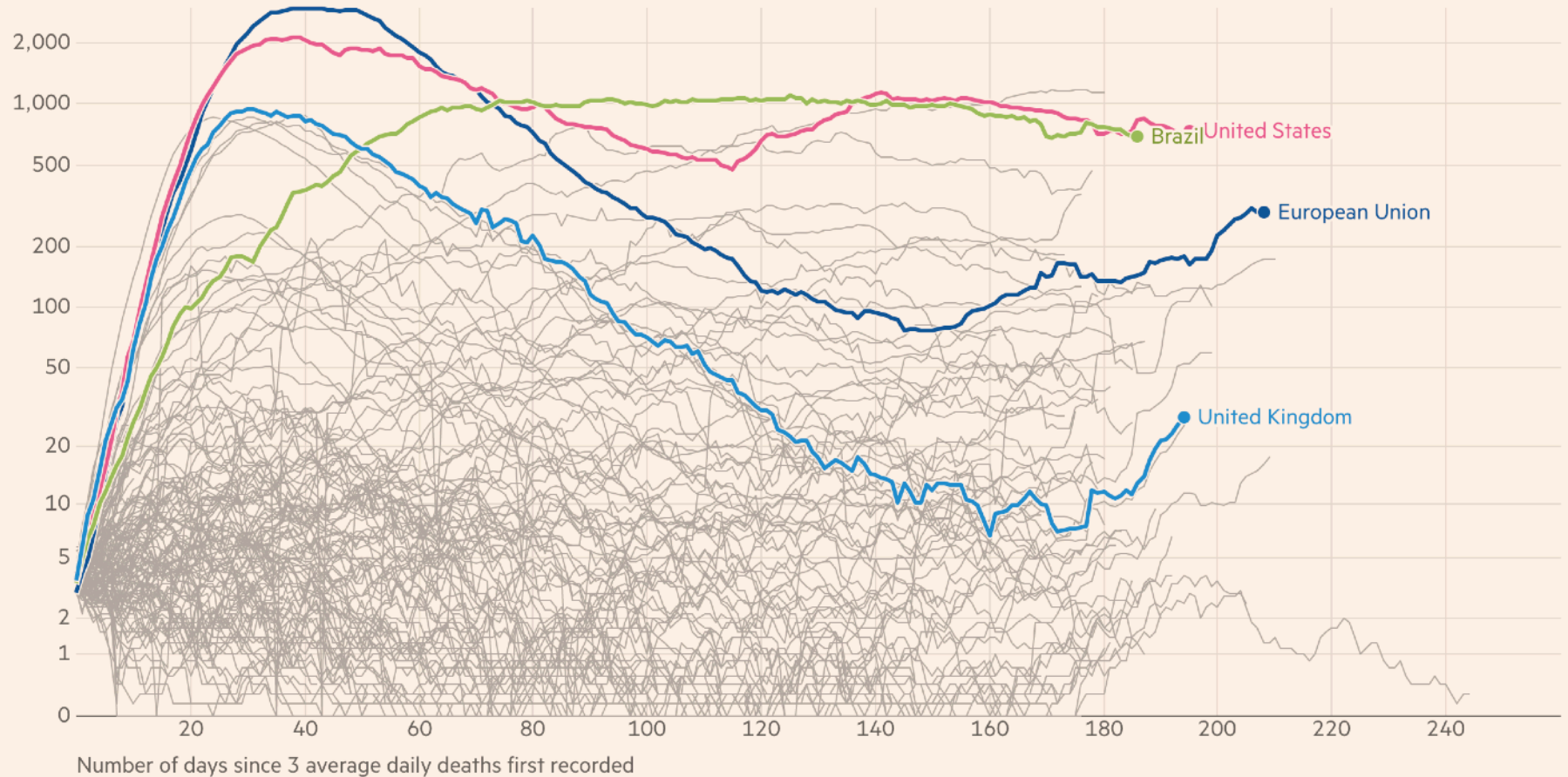
Visualizing Black America, Du Bois et al. 1900



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



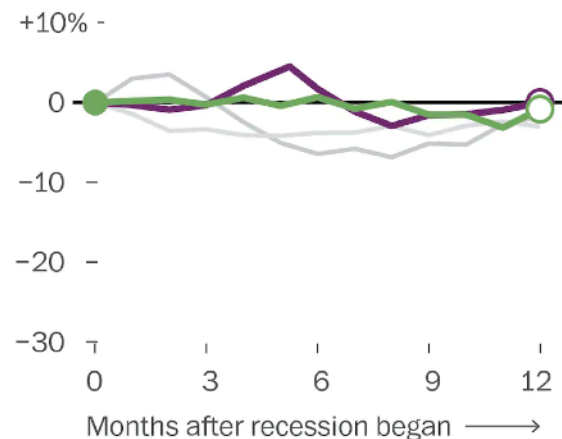
Source: Financial Times analysis of data from the European Centre for Disease Prevention and Control, the Covid Tracking Project, the UK Dept of Health & Social Care and the Spanish Ministry of Health.
Data updated September 25 2020 12.46pm BST. Interactive version: ft.com/covid19

FINANCIAL TIMES

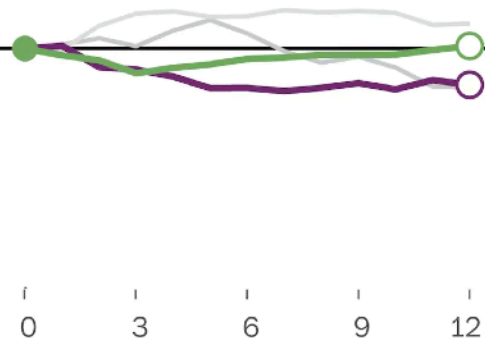
The coronavirus crisis is different

Job growth (or loss) since each recession began, based on weekly earnings

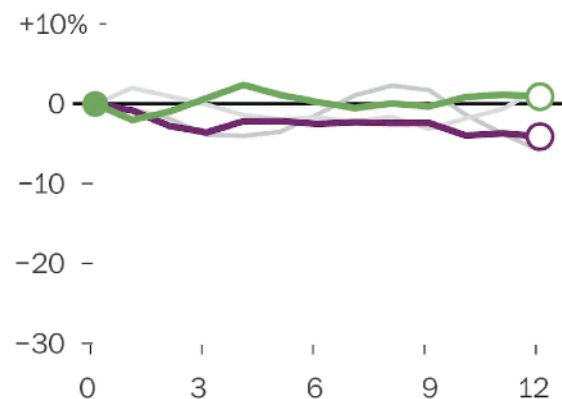
1990 recession



2001 recession



2008 recession



Coronavirus crisis



Notes: Based on a three-month average to show the trend in volatile data.

Source: Labor Department via IPUMS, with methodology assistance from Ernie Tedeschi of Evercore ISI

THE WASHINGTON POST

The Covid Economy

Washington Post

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

Convey information

Communicate, inform, inspire

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

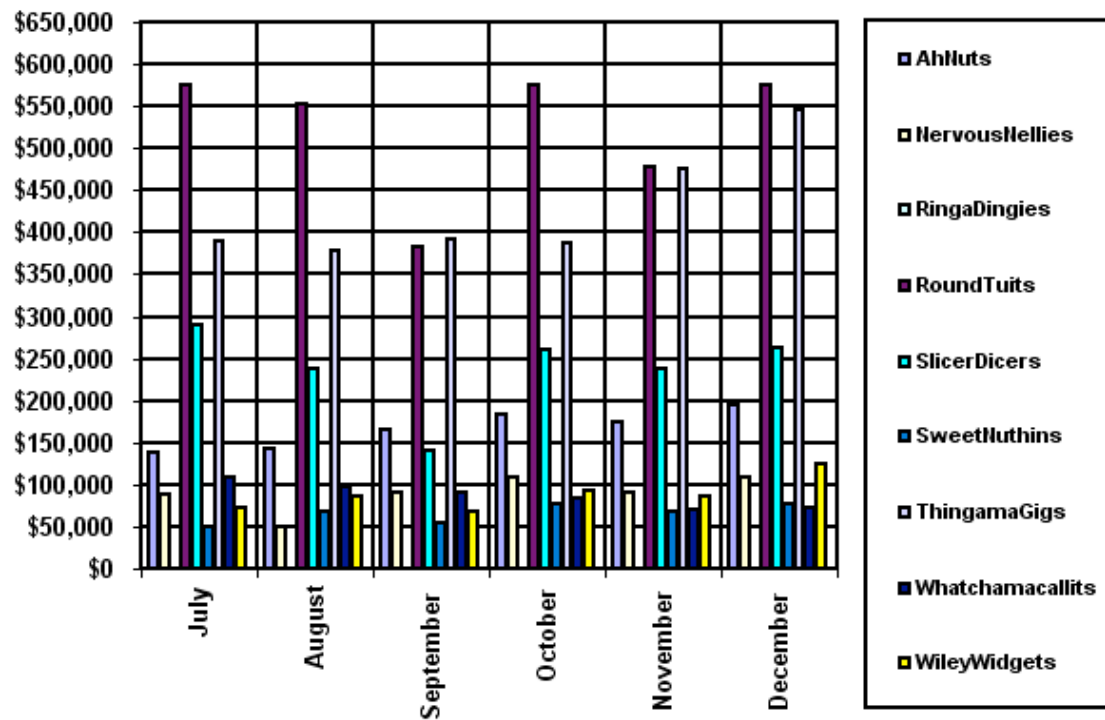
Course Topics

Data and Image Models

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

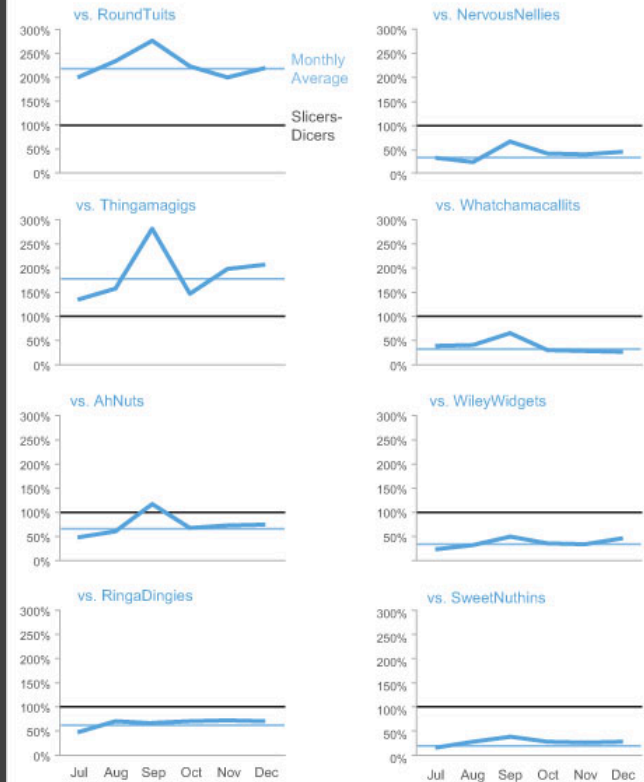
Visualization Design

SlicerDicers' Sales Compared to Other Products



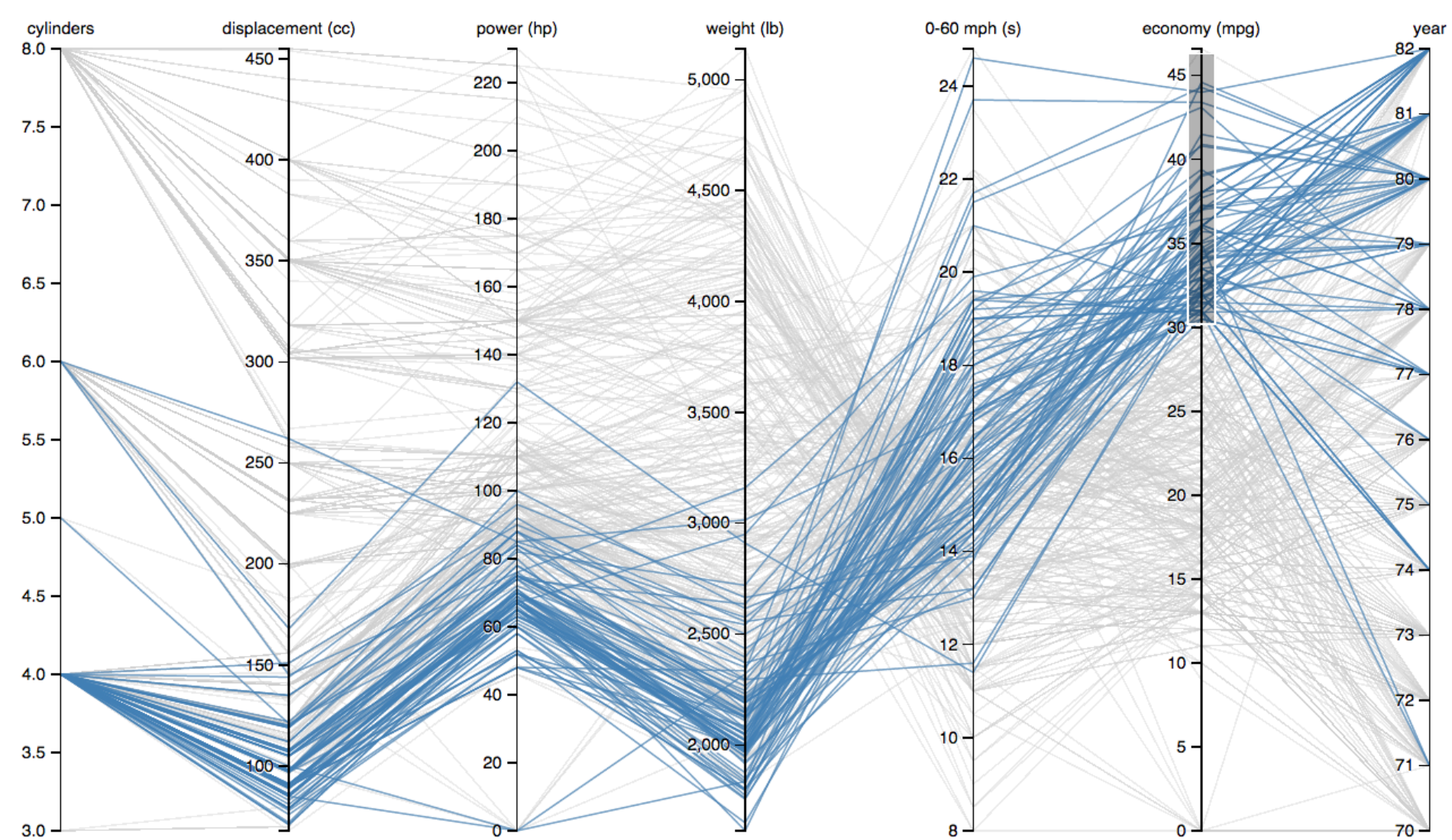
Problematic design

Sales of SlicersDicers Compared to Sales of Other Products
July - December, 2011



Redesign

Exploratory Data Analysis



Interaction

Friday, December 12, 2008
154 reports

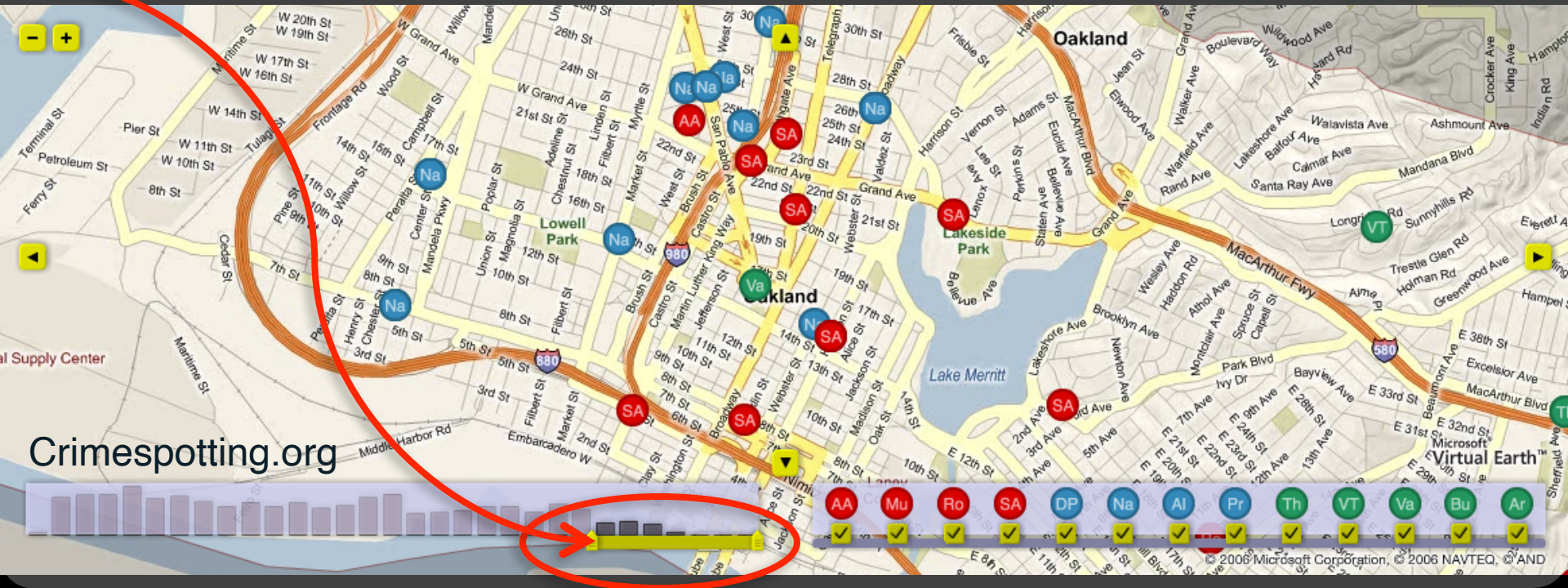
Na SA Ro Th Bu VT Va DP Pr Th VT Va Bu Ar

© 2006 Microsoft Corporation. © 2006 NAVTEQ. © AND

154 reports



© 2006 Microsoft Corporation, © 2006 NAVTEQ, © AND

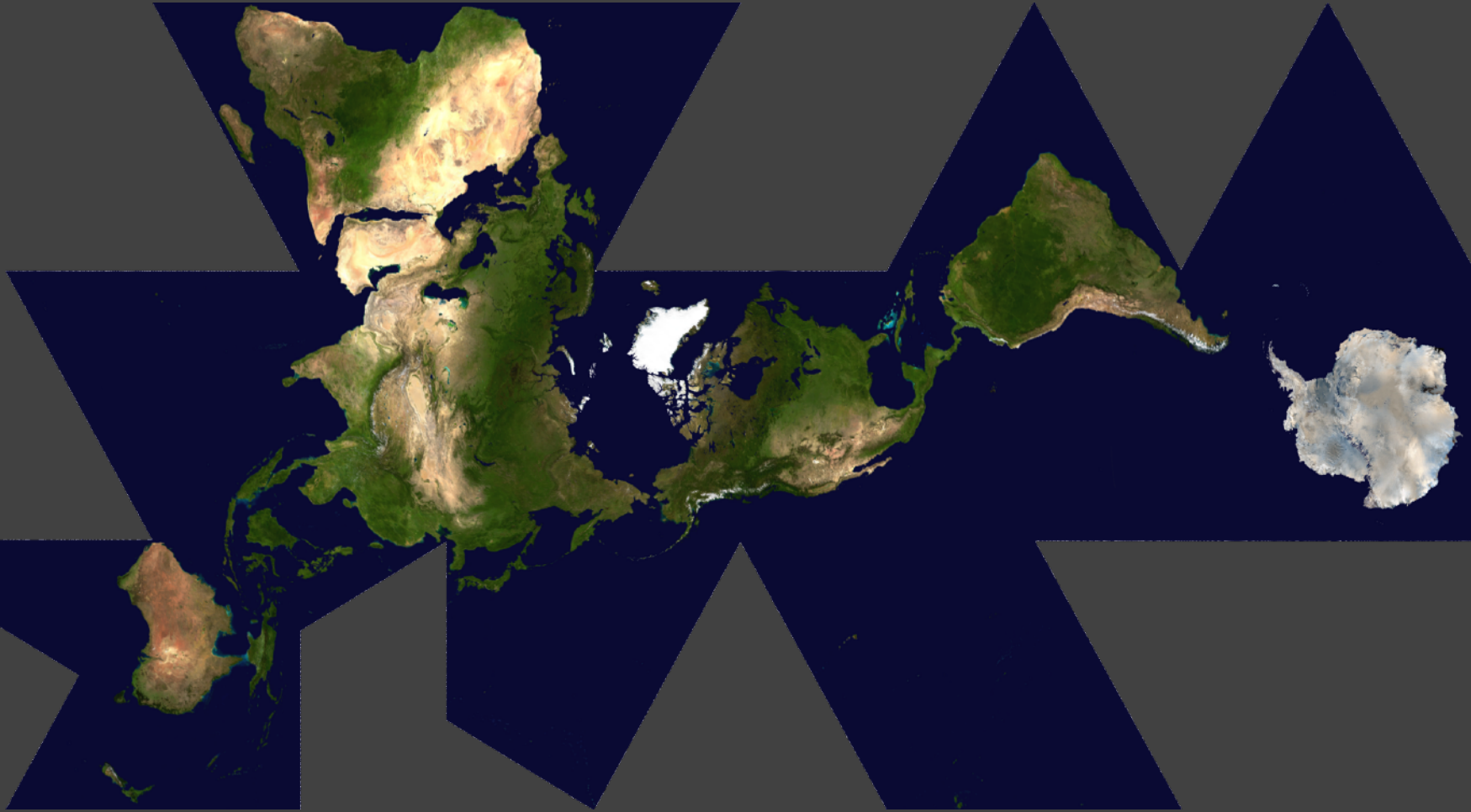


Crimespotting.org



© 2006 Microsoft Corporation, © 2006 NAVTEQ, © AND

Maps

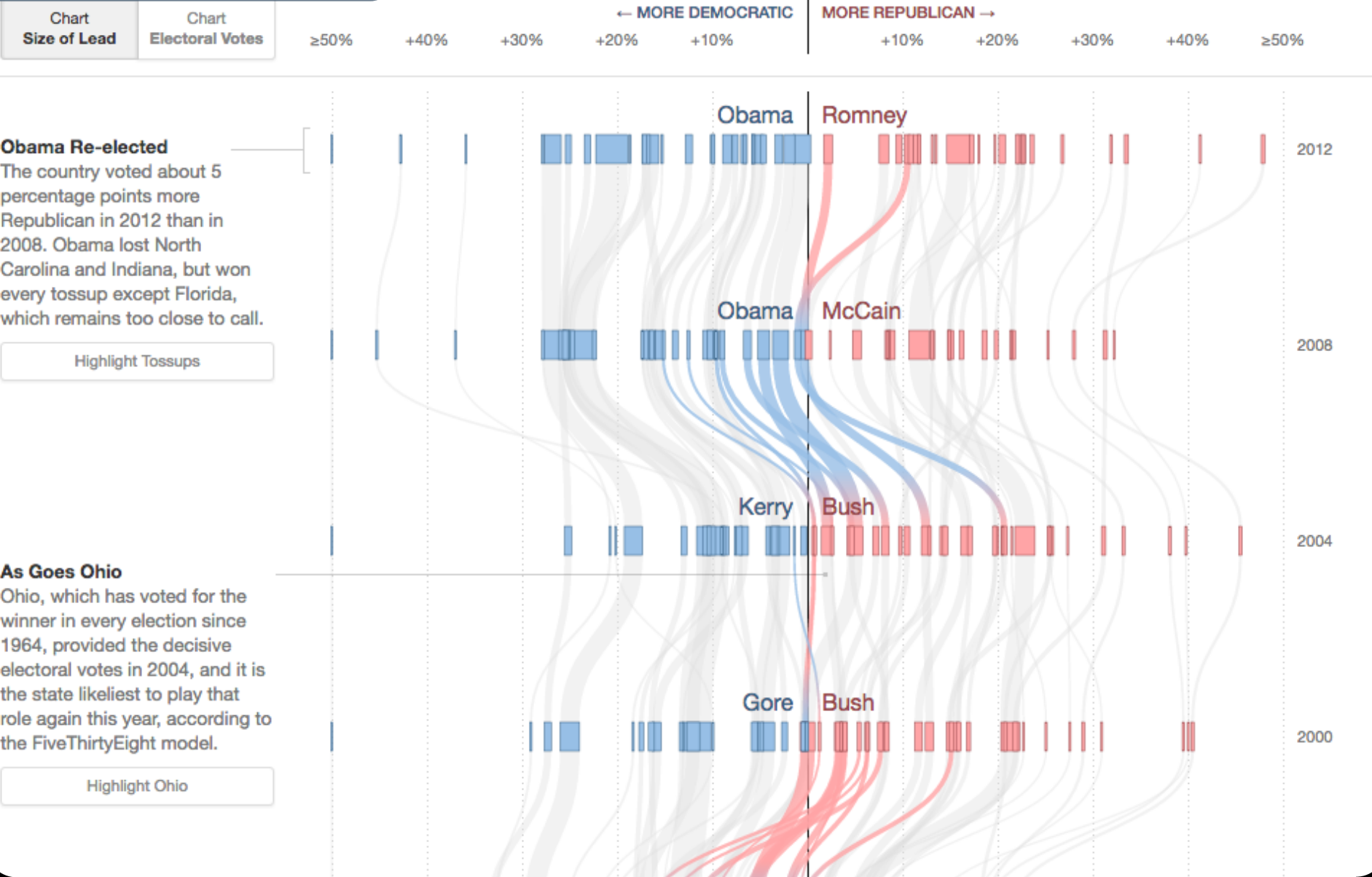


Dymaxion Maps [Fuller 46]

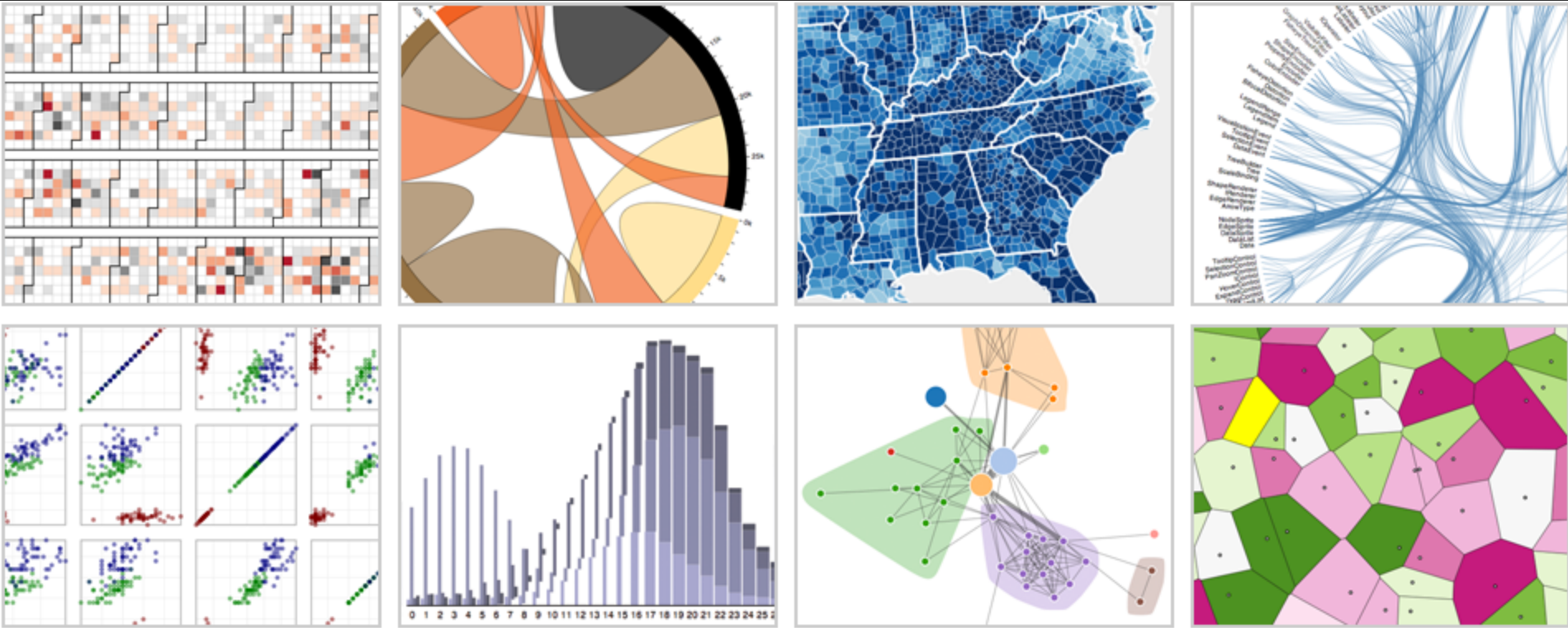
Narrative

Recent elections have placed a heavy emphasis on “swing states” — Ohio, Florida and the other competitive states. Yet in the past, many more states shifted between the Democratic and Republican parties. A look at how the states stuck or shifted between elections and how 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

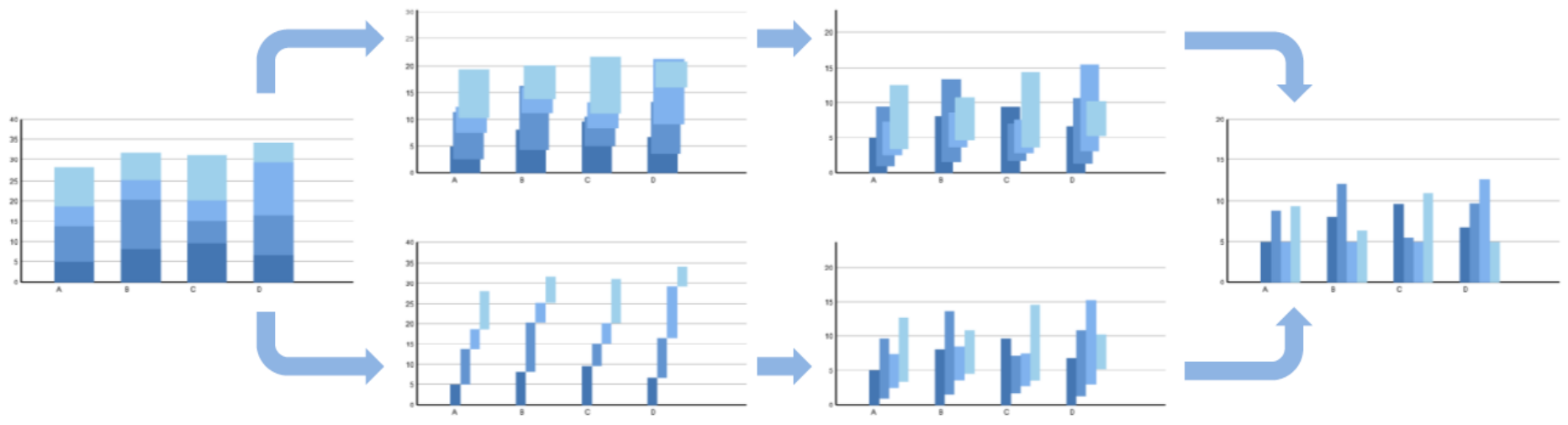


Visualization Software



D3: Data-Driven Documents
Vega-Lite / Altair

Animation

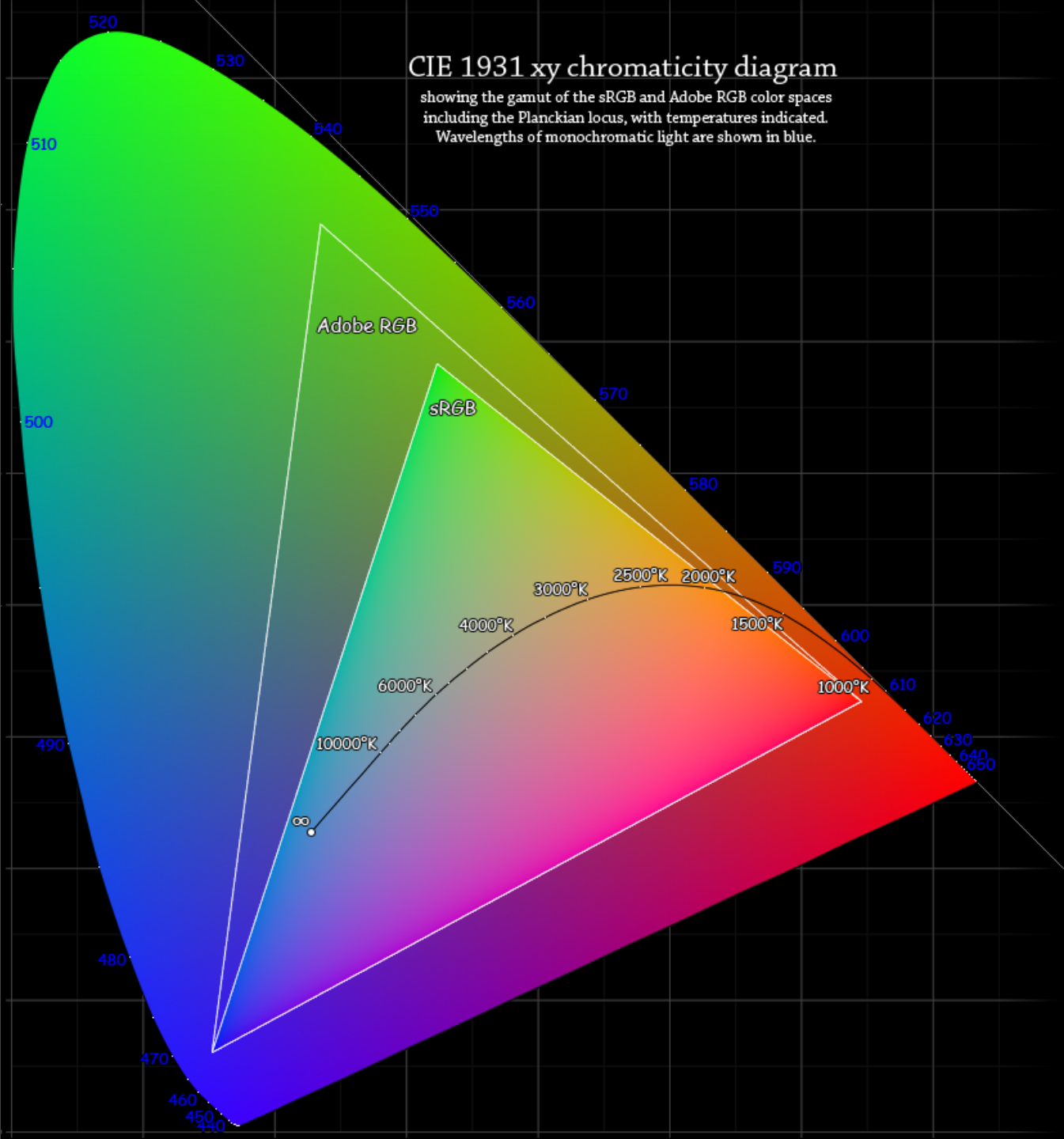


Animated transitions in statistical data graphics [Heer & Robertson 07]

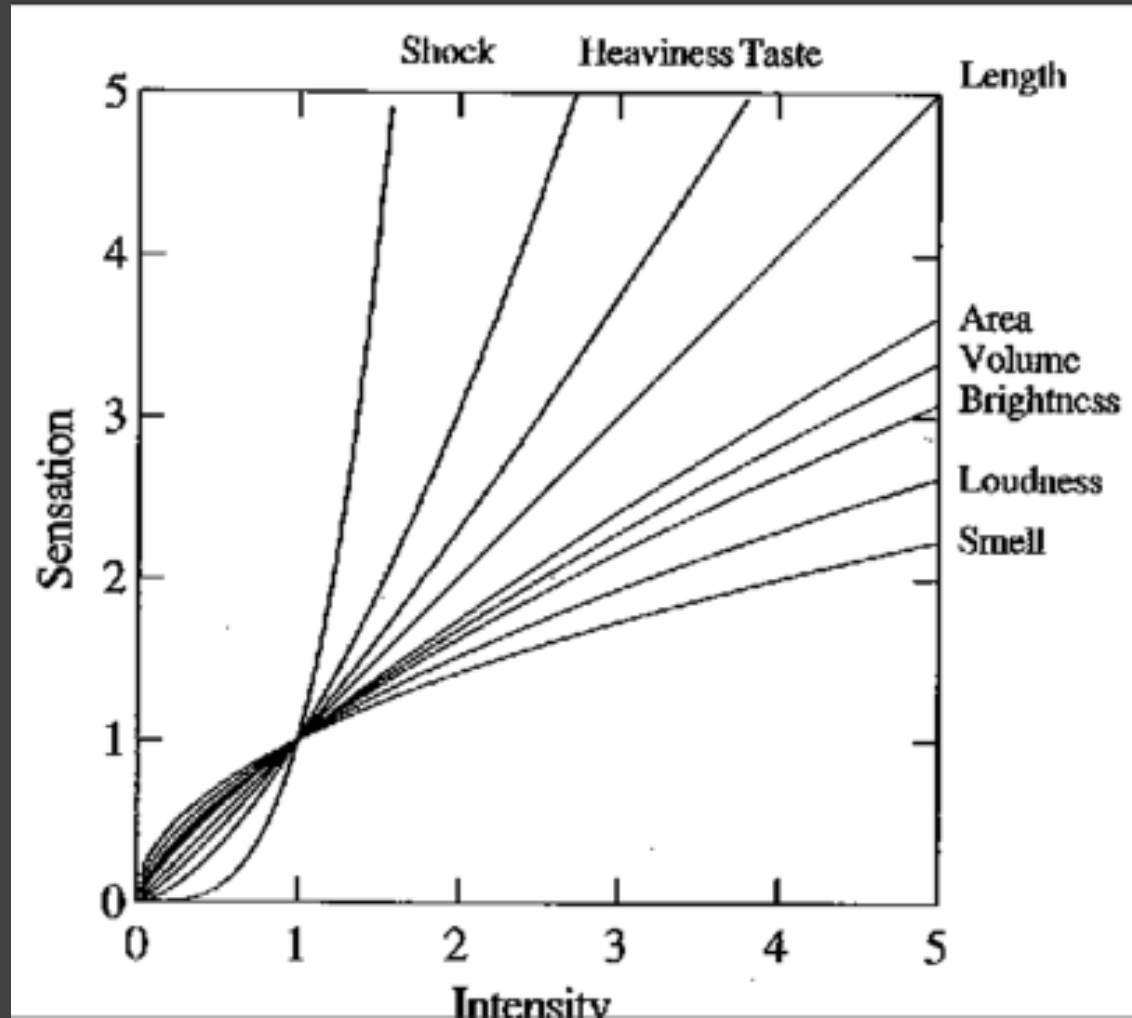
[illegible]

CIE 1931 xy chromaticity diagram
 showing the gamut of the sRGB and Adobe RGB color spaces
 including the Planckian locus, with temperatures indicated.
 Wavelengths of monochromatic light are shown in blue.

The diagram illustrates the CIE 1931 xy chromaticity space. The outer boundary represents the visible spectrum of monochromatic light, with wavelengths labeled in blue (430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650 nm). The Planckian locus, representing black-body radiation, is shown as a curve with temperatures labeled: 10000°K, 6000°K, 4000°K, 3000°K, 2500°K, 2000°K, 1500°K, and 1000°K. The Adobe RGB color space is outlined as a large triangle, and the sRGB color space is outlined as a smaller triangle nested within it. The diagram is set against a dark gray grid.

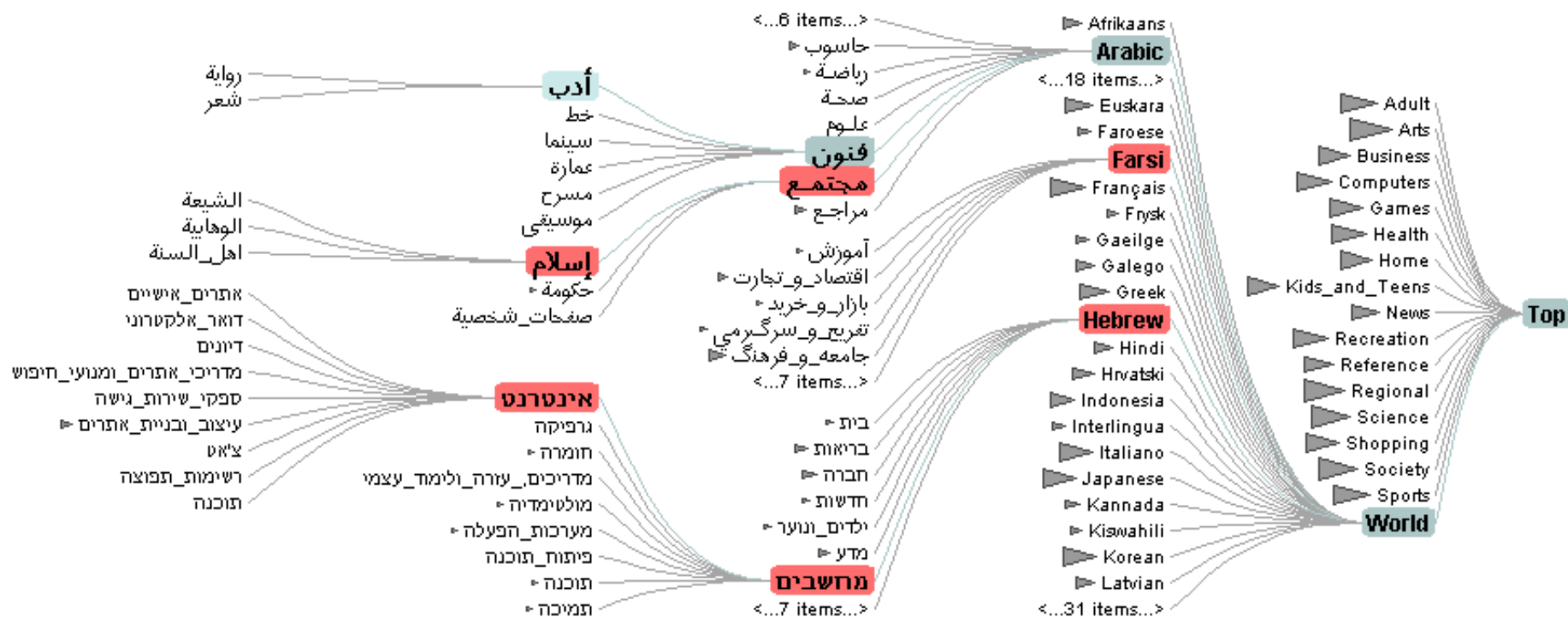


Graphical Perception



The psychophysics of sensory function [Stevens 61]

Hierarchies



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

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/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

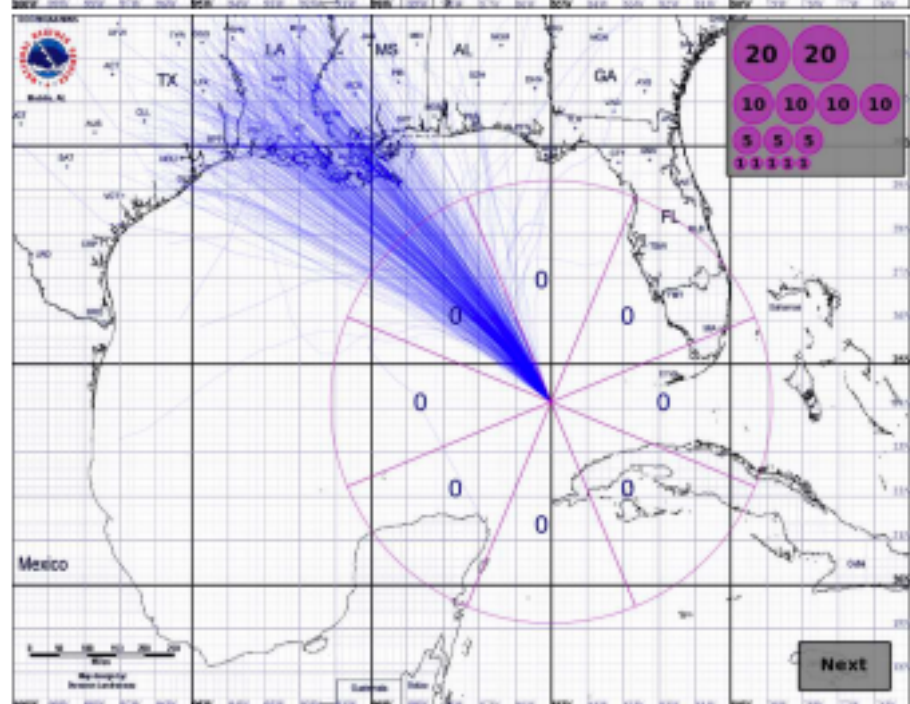
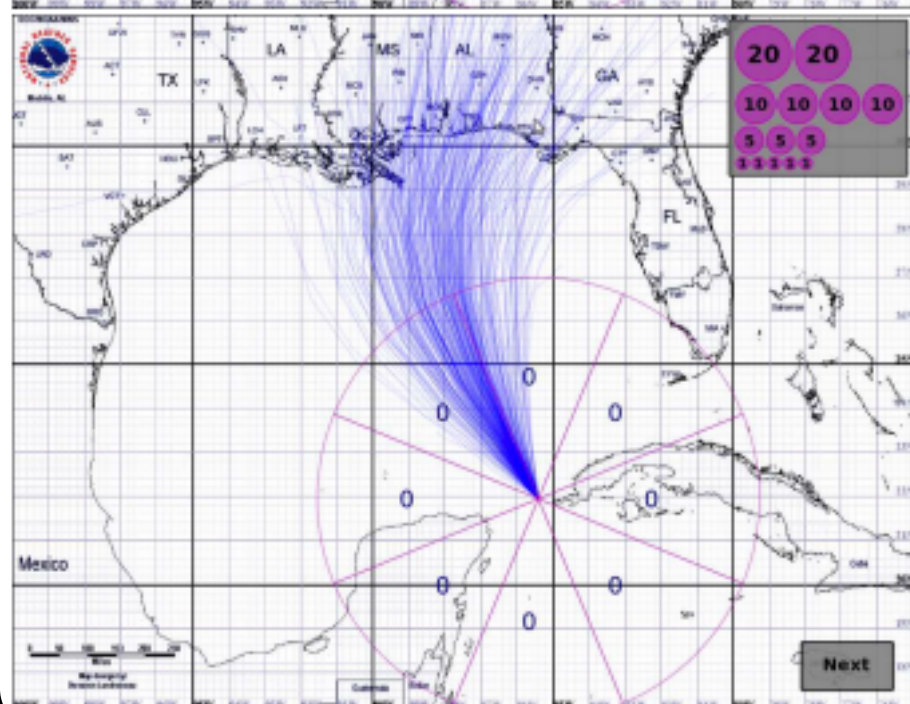
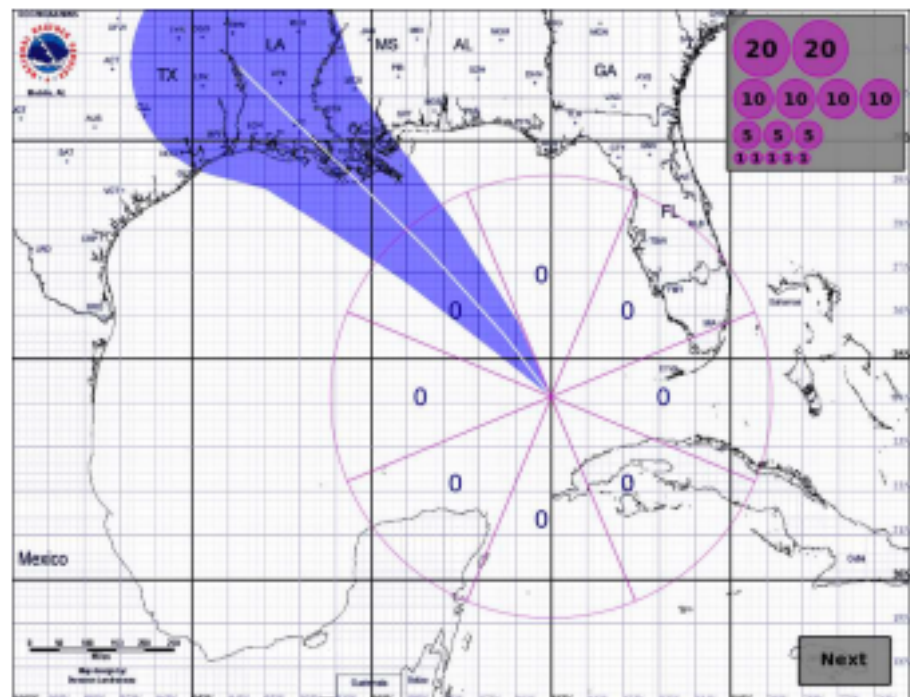
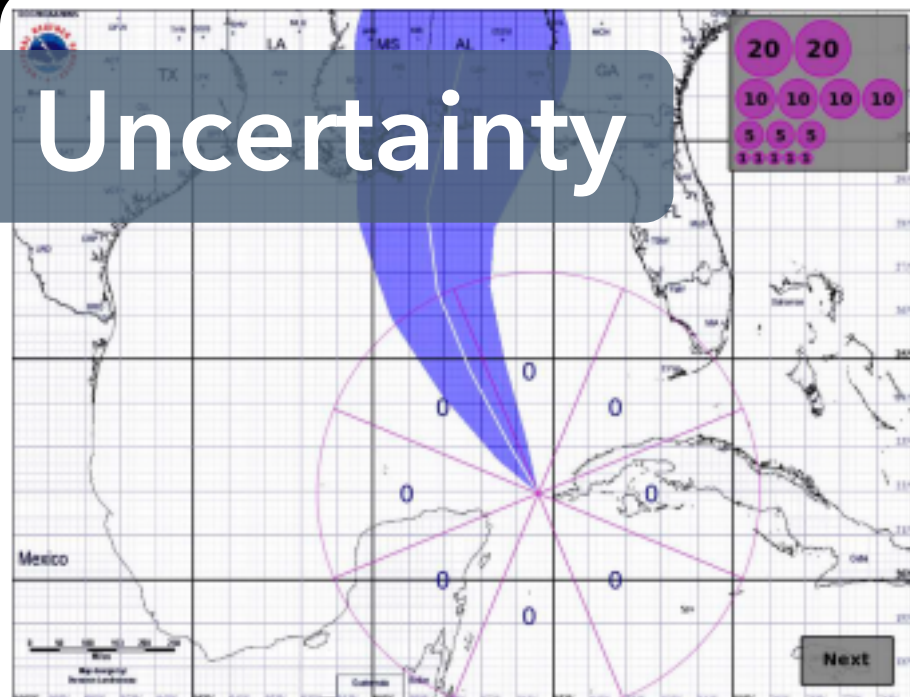
About [Some know me as danah...]

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.

Uncertainty

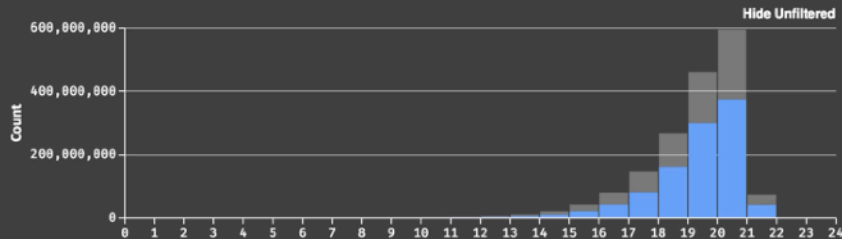


Scalability

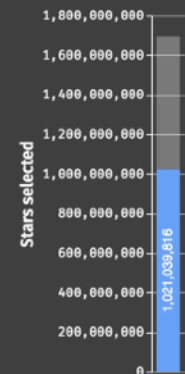
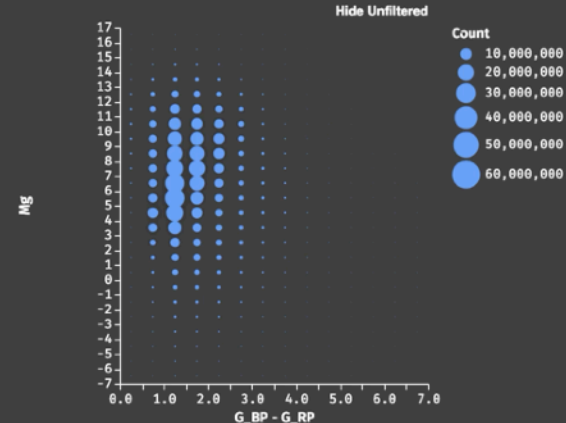
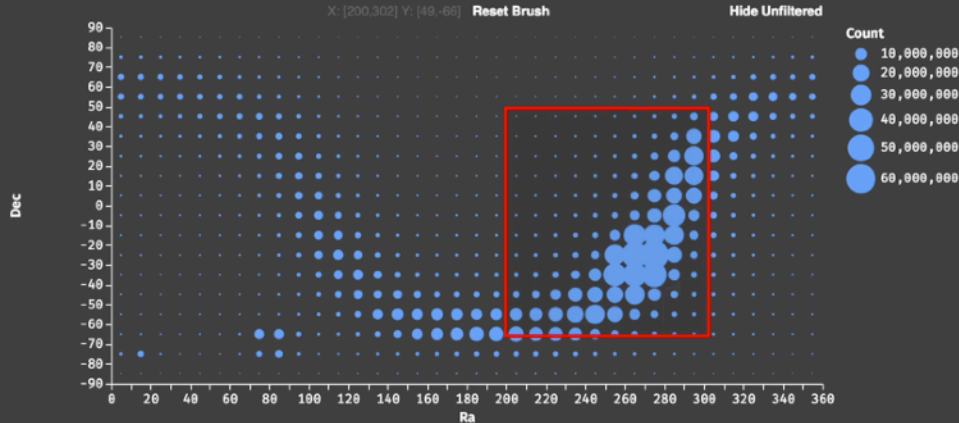
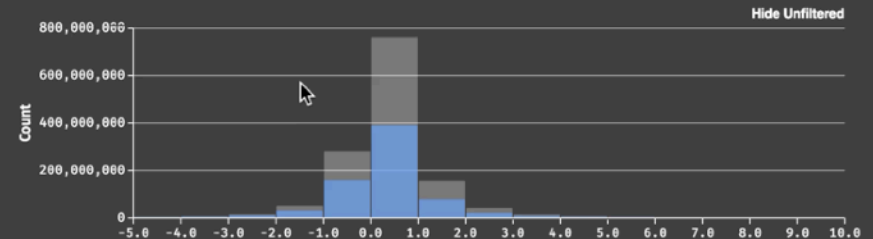
localhost:1234



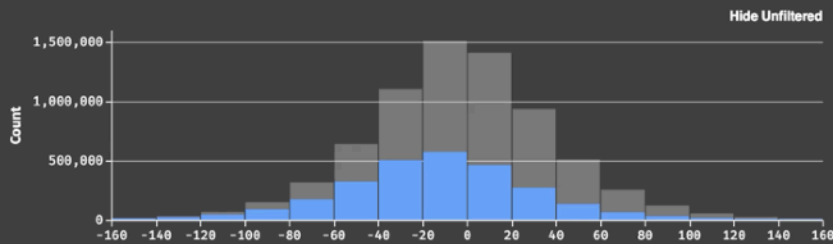
Magnitude



Parallax



Radial Velocity



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

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

Instructors

cse442@cs

Instructor

Jeffrey Heer

OH: *Tue 10-11am*

Teaching Assistants

Abhishek Babu

OH: *Mon 1:30-2:30pm*

Vishal Devireddy

OH: *Online / Ed*

Naveena Karusala

OH: *By appointment*

Xia Su

OH: *Online / Ed*

Firn Teanklin

OH: *Wed 1:30-2:30pm*

Yueqian Zhang

OH: *Fri 2:30-3:30pm*

Abhishek Babu

babua @ cs

Office hours: Mondays 1:30 - 2:30 PM

I'm an MS student in the BS/MS program and I'm interested in machine learning and data engineering.

Relevant technical skills: JavaScript, TypeScript, HTML/CSS, D3.js, general web dev (and other data wrangling libraries in Python)

I enjoy Spikeball as well! (check out Husky Roundnet Club)



Vishal Devireddy

Email vishald@cs

Office hours by appointment / Ed

I'm a BS/MS student interested in interactive web development, HCI, and cool graphs.

Ask me about

- JavaScript + CSS
- Web design
- Web accessibility
- Observable
- D3
- Idyll





Nussara (Firn) Tieanklin

nussara@cs.washington.edu

Office hour: Wed 1:30 - 2:30 PM

Research

- Seattle Community Networks
providing internet access to resource-constrained community
- Evaluate the impact of Air Pollution on Grab motorcycle drivers/riders

Technical Experience

User research, Design process, Data Management, Web-programming

(Covid) Hobbies

- Gardening
- Badminton
- Playing computer games
- Travelling

Naveena

- ▶ 5th year PhD student in the ICTD Lab
- ▶ Research in HCI and global development, specifically on health messaging to underserved populations
- ▶ Excited to talk about: design of visualizations





Xia Su

xiasu@cs.washington.edu

- Hi! I'm a first year PhD student researching into HCI with special interest in design and creativity.
- Have background in design. Feel free to discuss your designs with me!
- No office hour, but I manage the online discussions with Vishal. So chat with me online!
- Will lead a Tableau Tutorial with Naveena on Oct. 8th.

Yueqian Zhang

OH: Friday 2:30 pm - 3:30 pm

I am a senior **majoring in CS** and **minoring in Education**.

I am interested in **Accessibility** and **Graphics**.

Outside of CS I enjoy **design** and **drawing**.

Come talk to me! I'd love to meet everyone :)

Some of my artworks:



Readings

From books, notebooks, and linked articles.

Material in class will loosely follow readings.

Readings should be read by start of class.

Complete non-scored quizzes on Canvas.

Post comments on class forum.

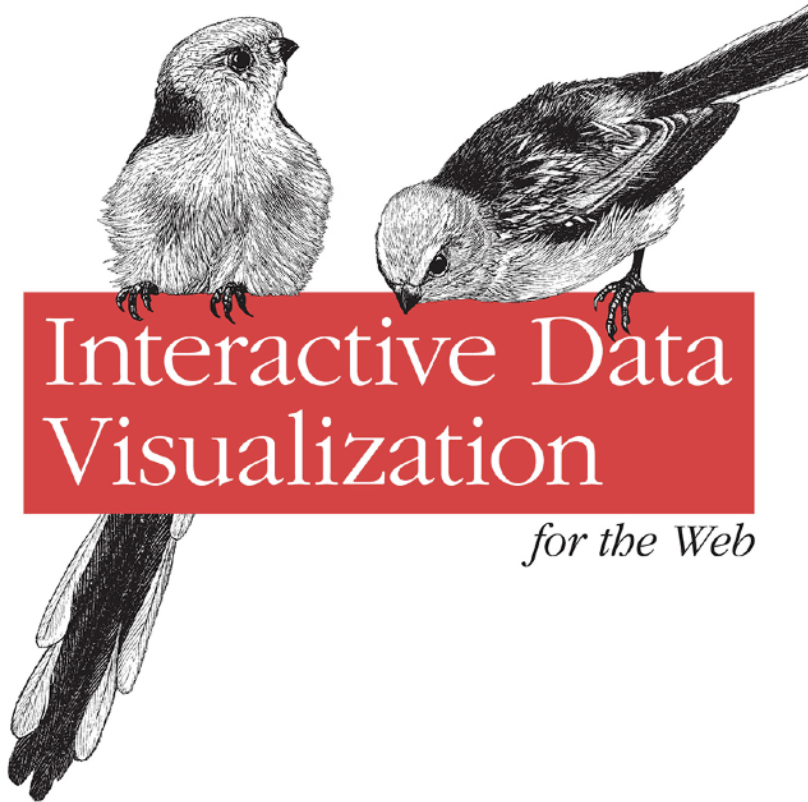
One comment or peer reviews each week.

Post comments by Friday 11:59pm.

You have 1 discussion “pass” for the quarter.

Textbook

An Introduction to Designing With D3



O'REILLY®

Scott Murray

Interactive Data Visualization for the Web, 2nd Edition

For learning D3!

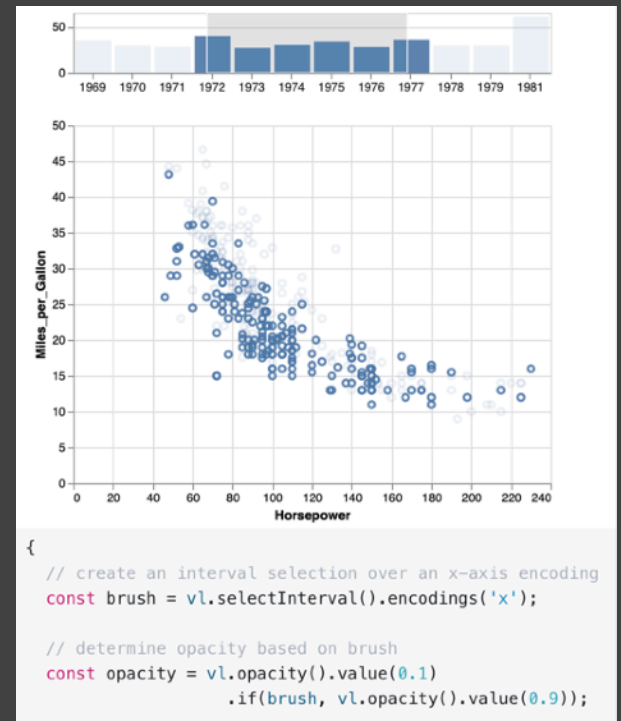
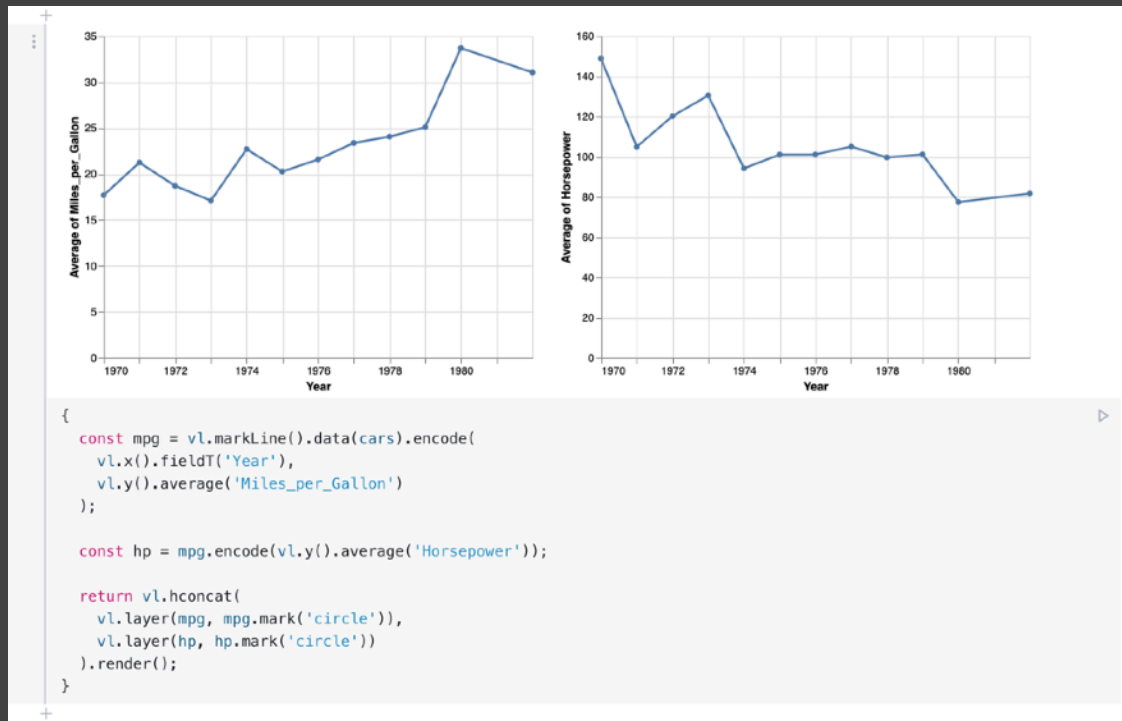
Book available online.

Code / examples on GitHub.

We will be using **D3 v7**.

<https://d3js.org>

Interactive Vega-Lite Notebooks



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

Assignments

CP Class Participation (10%)

A1 Expository Visualization (10%) - *Due 10/11*

A2 Deceptive Visualization (15%) - *Due 10/20*

Peer Review (5%) - *Due 10/25*

A3 Interactive Prototype (20%) - *Due 11/8*

Peer Review (5%) - *Due 11/15*

FP Final Project (35%)

Proposal - *Due 11/19*

Demonstration Video - *Due 12/8*

Final Prototype - *Due 12/14*

Final Project

Produce an **explorable visual explanation**

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

Final deliverables and **video presentation**

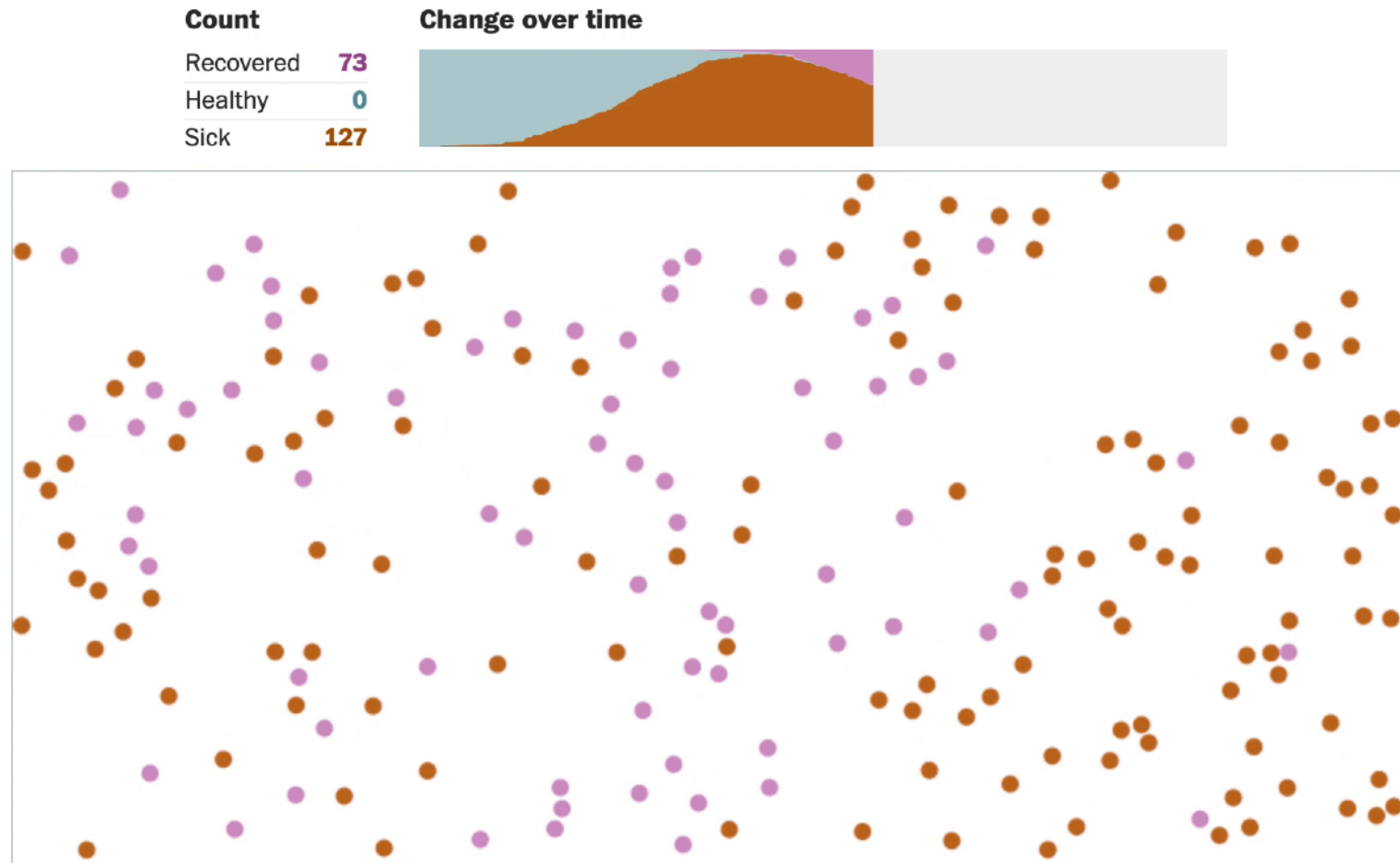
Submit and **publish online** (GitHub)

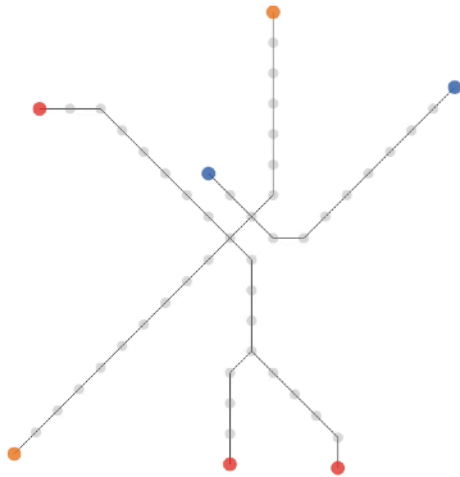
Projects from **previous classes** have been:

- Published as research papers
- Shared widely (some in the New York Times!)
- Released as successful open source projects

Why outbreaks like coronavirus spread exponentially, and how to “flatten the curve”

Harry Stevens, Washington Post 2020

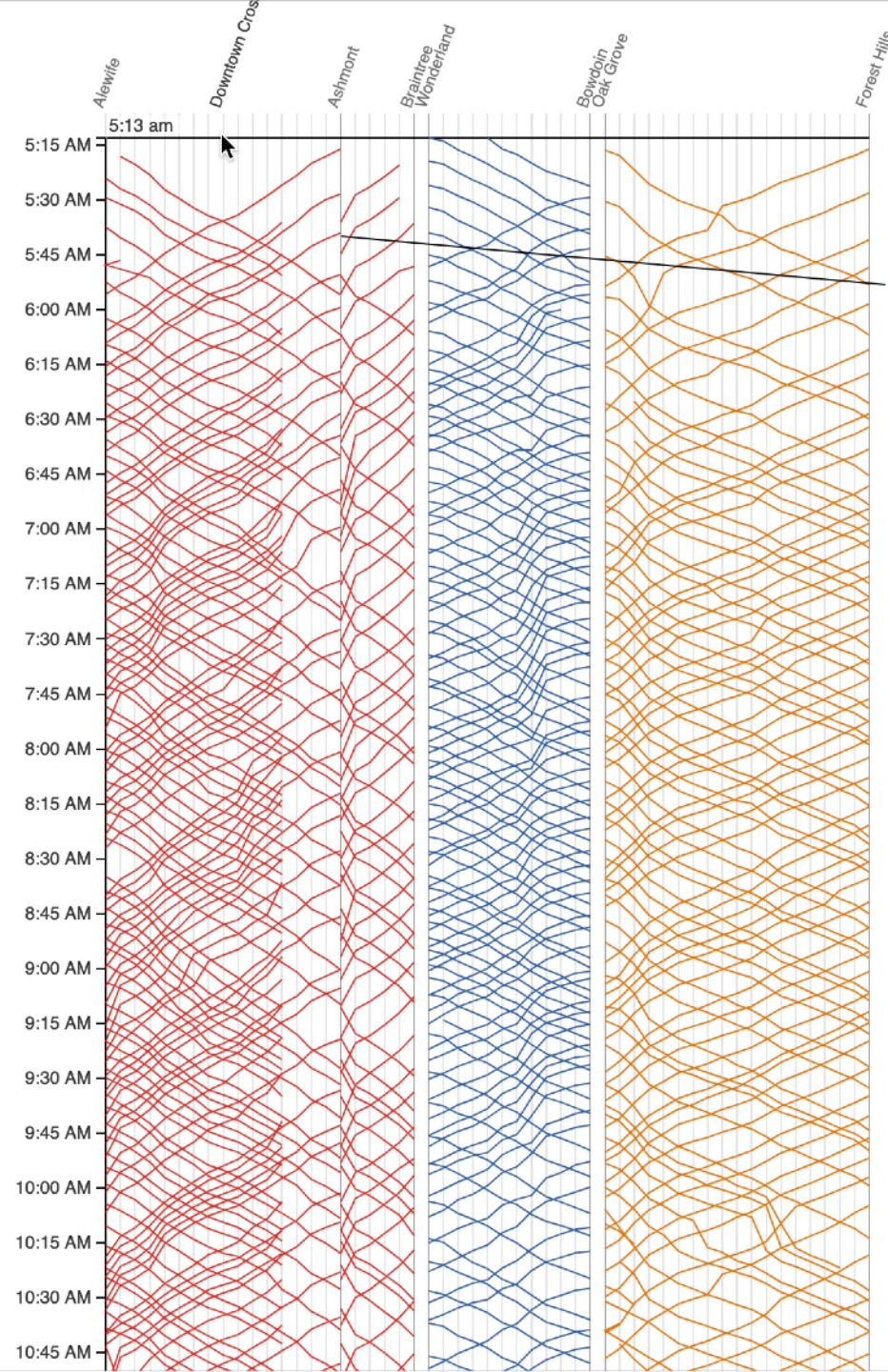




Locations of each train on the [red](#), [blue](#), and [orange](#) lines at 5:13 am. Hover over the diagram to the right to display trains at a different time.

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.

KEYBOARD WALKING

Passwords with a “keyboard walking” pattern start at an arbitrary key, then move in a direction (usually right or down) while continuing to hit keys. Sometimes this is combined with holding down the **SHIFT** key, so that some characters are uppercase or symbols to improve complexity.

While the generated password may seem to be random and unhackable, password crackers [check for these keyboard patterns](#) and guess them early on.

Many passwords in the leaked passwords dataset have a spatial pattern. Other than the numeric passwords like **123456**, common keyboard walking offenders include **qwerty** and **1qaz@wsx**.



Semantic Passwords

Vishal Devireddy (CSE 512, Spring '21)

Questions?

Observable + Data Tutorial

Friday Oct. 1, 4:30-6pm on Zoom

Introduction to Observable notebooks, JavaScript basics, and data management and transformation, led by Firn & Yueqian.

Zoom link is available on Canvas.

The tutorial will be recorded.

A1: Expository Visualization

Design a static visualization for a data set.

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.

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: Expository Visualization

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, Monday October 11.**