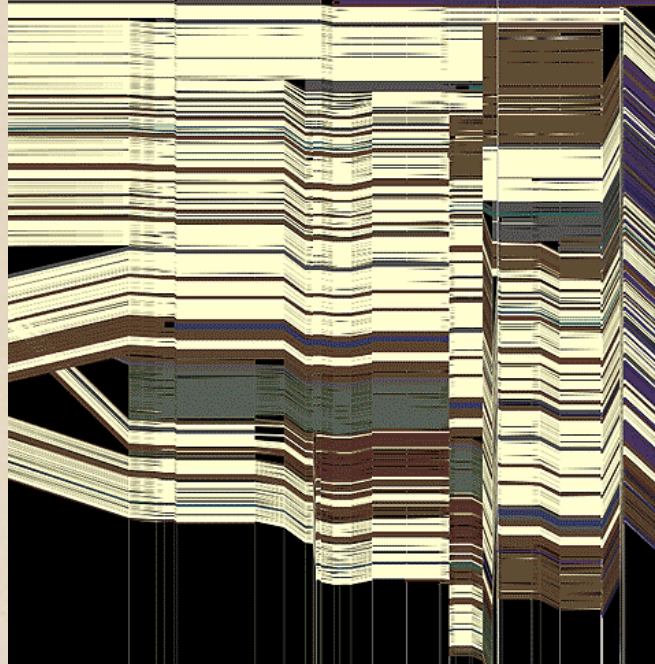
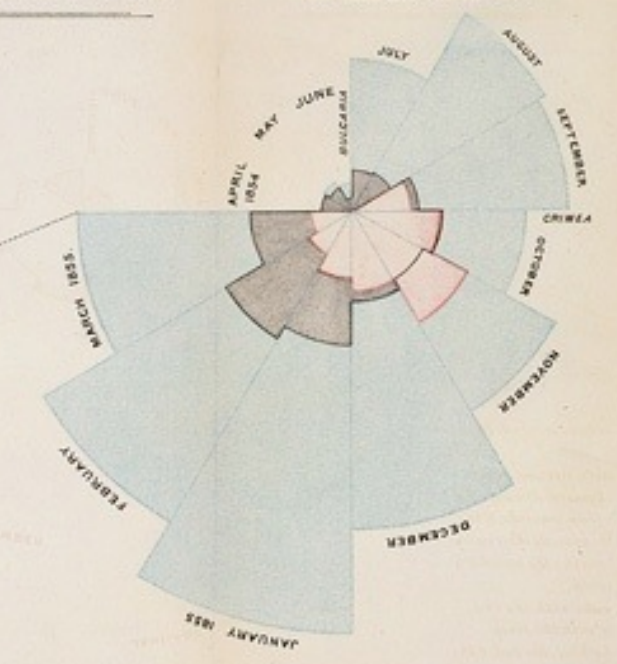


CSE 412 - Intro to Data Visualization

# The Value of Visualization



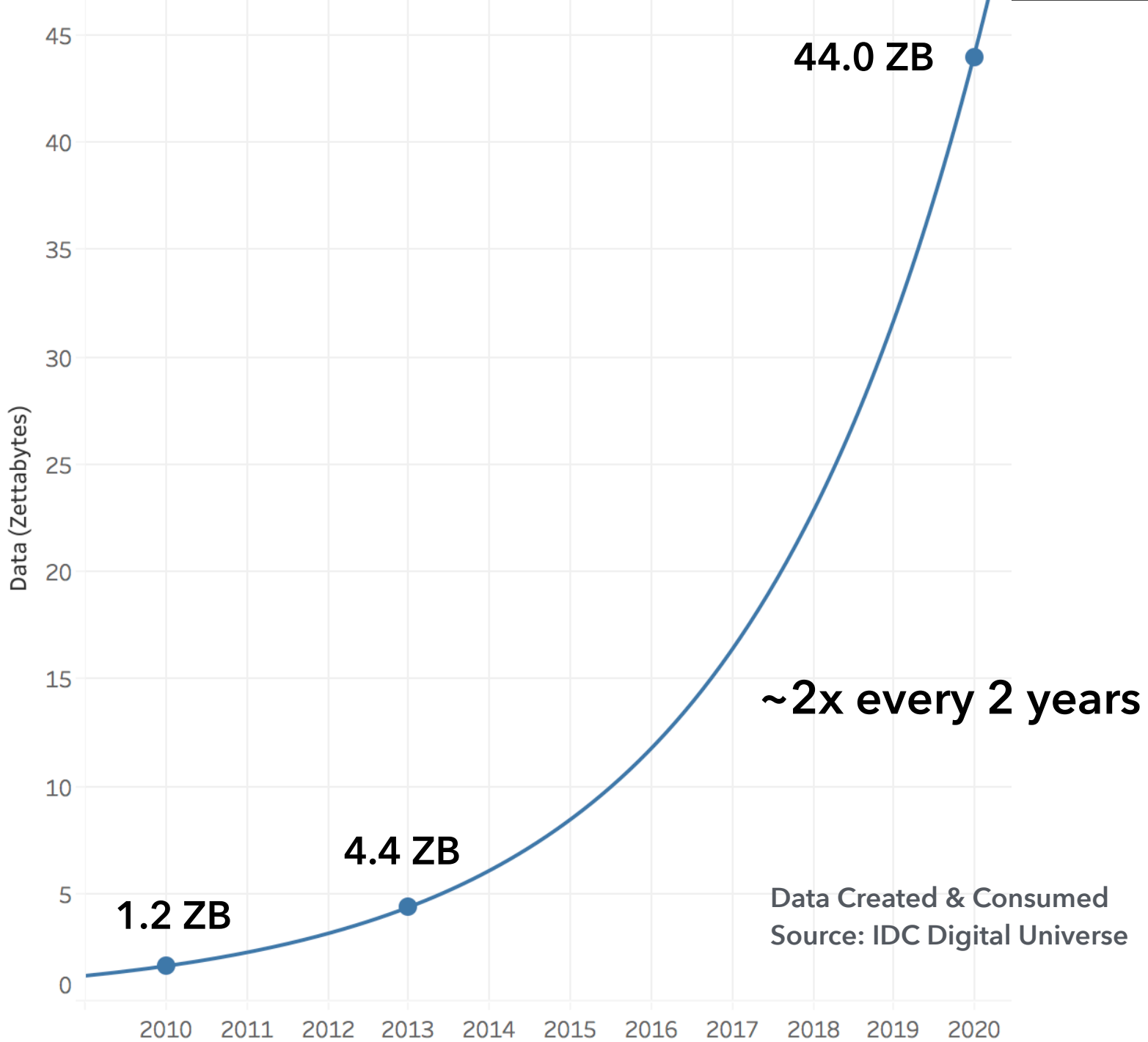
Jane Hoffswell University of Washington

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

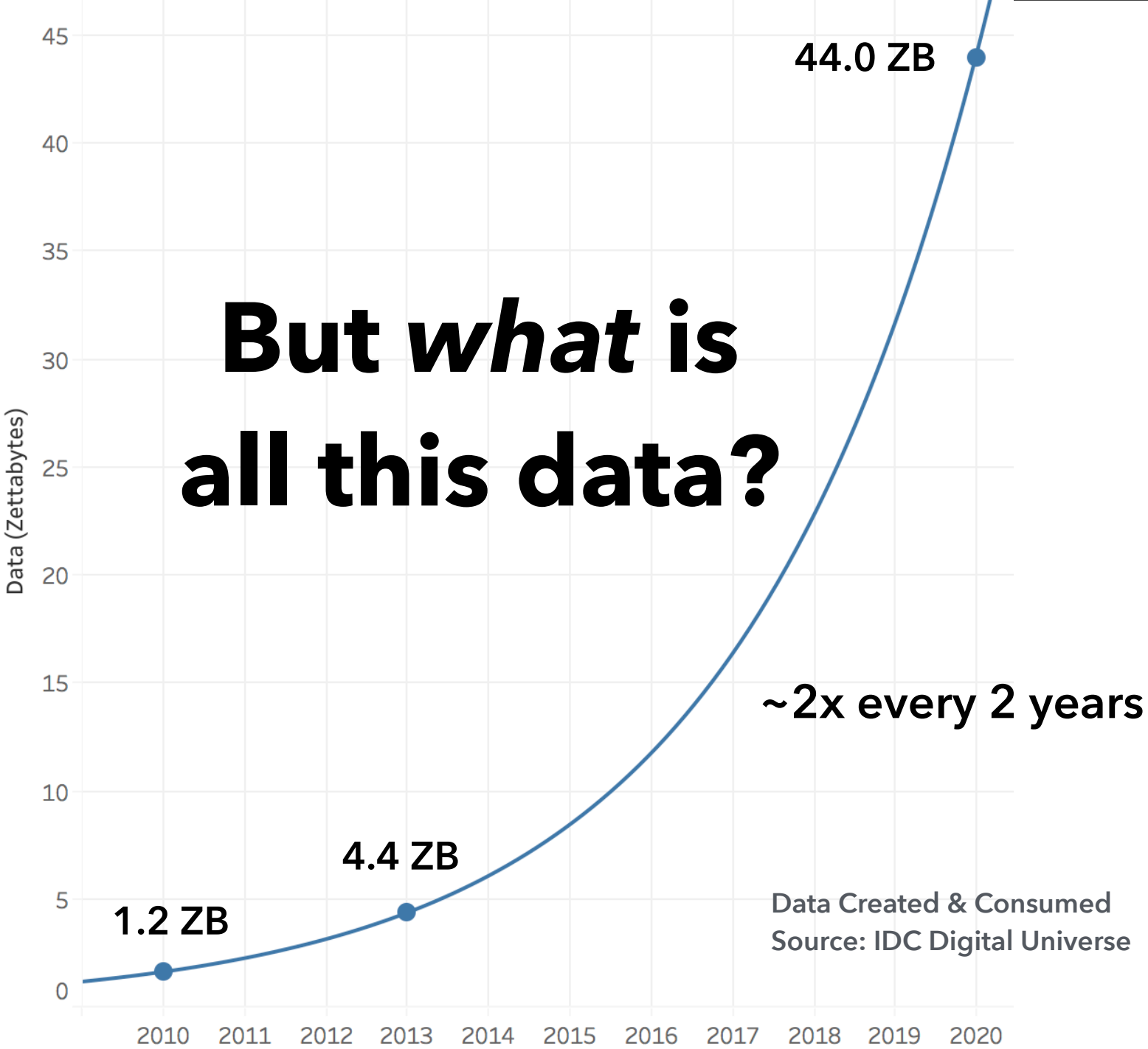
**2010: 1,200 EB (1.2 ZB)**

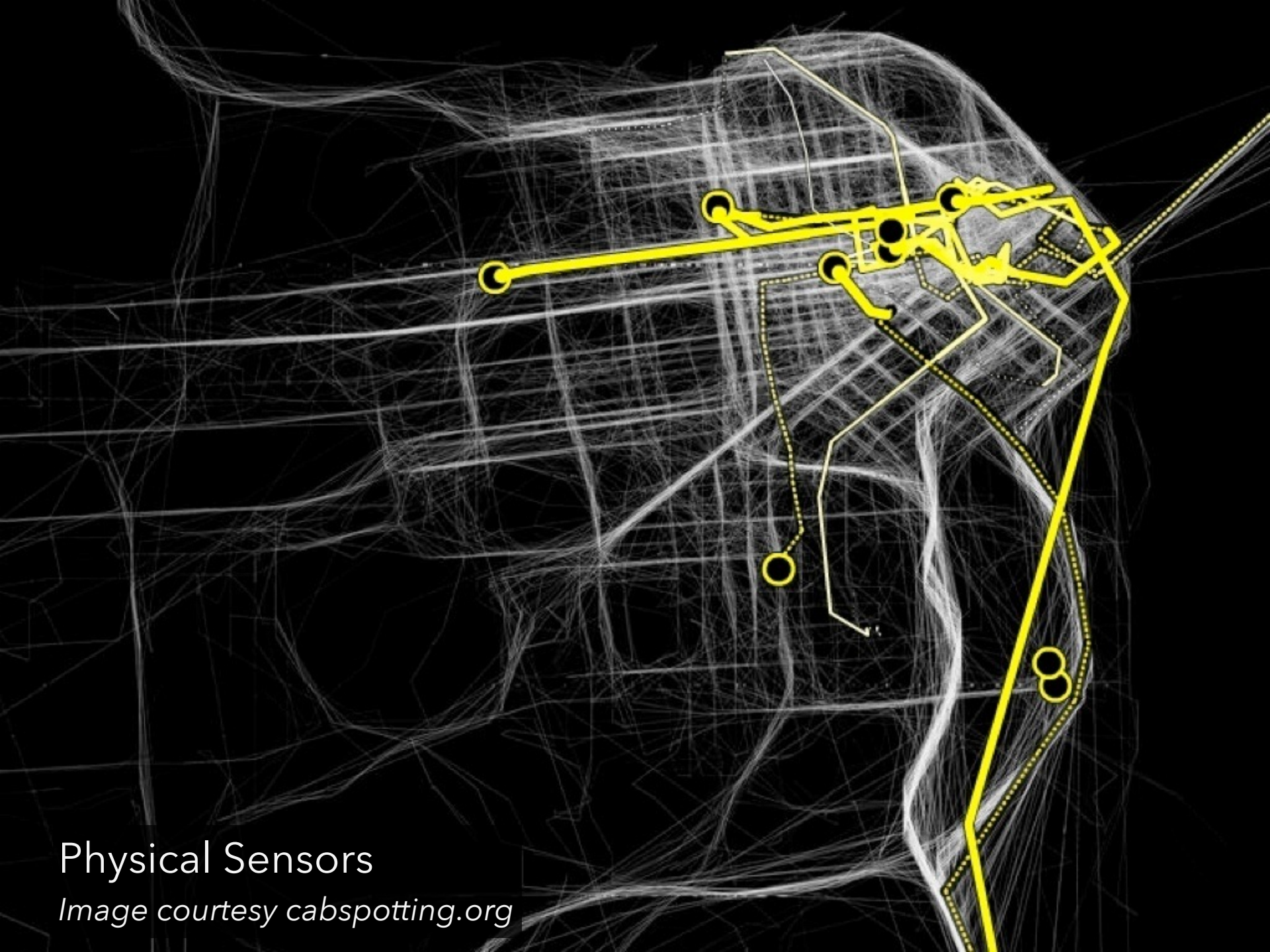
Exponential growth...

Gantz et al., 2008, 2010



# But *what* is all this data?





## Physical Sensors

*Image courtesy cabspotting.org*



157

HR/ECG  
1/min

130/65

Art  
mmHg sys/dia

93

SpO2  
%

RR/CO2  
1/min

97

HR/ECG  
1/min

82/60

Art  
mmHg sys/dia

99

SpO2  
%

RR/CO2  
1/min

79

HR/ECG  
1/min

152/79

Art  
mmHg sys/dia

95

SpO2  
%

RR/CO2  
1/min

64

HR/ECG  
1/min

93/55

mmHg sys/dia

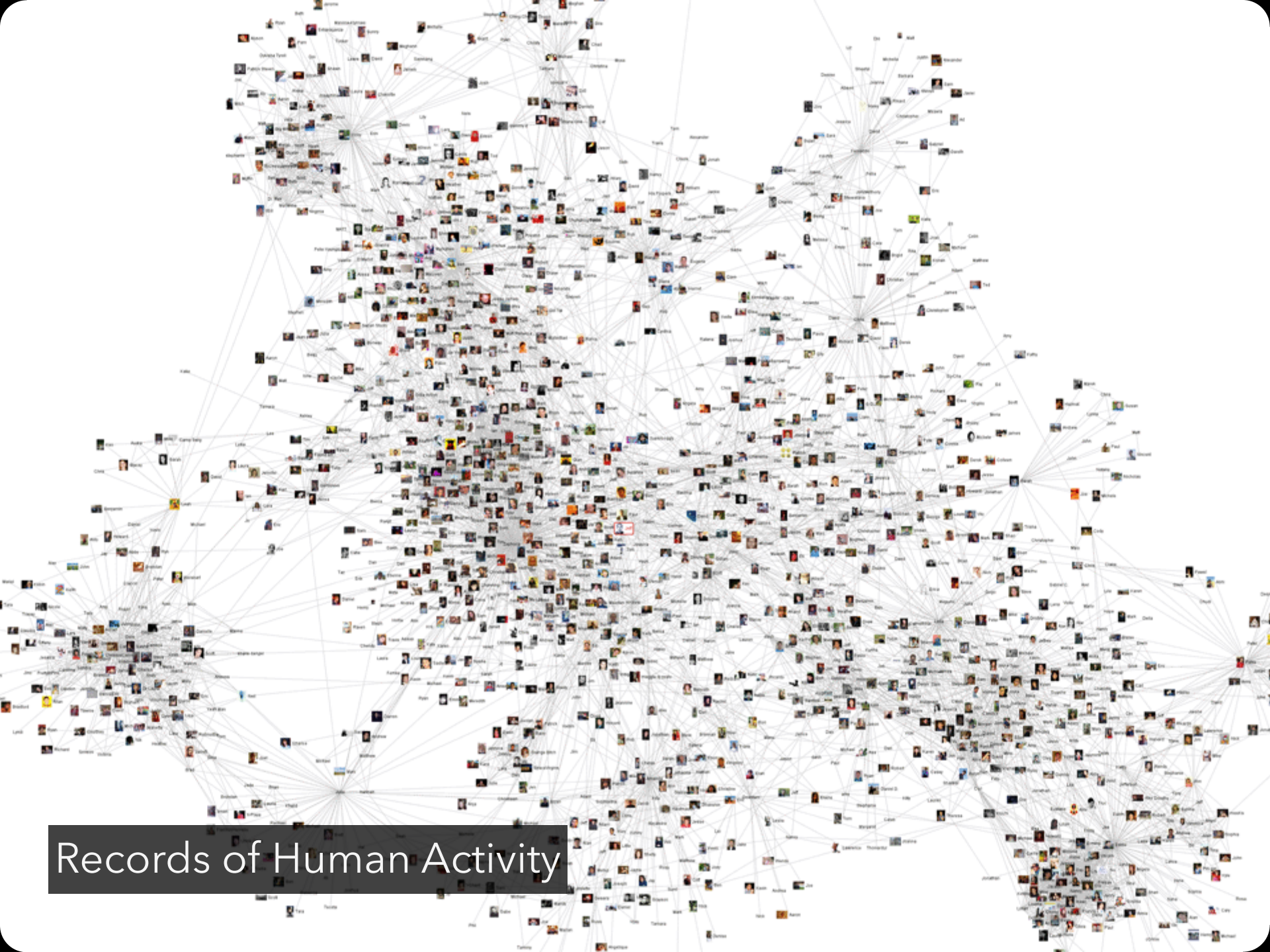
99

SpO2  
%

RR/Imp  
1/min

99

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 decade **"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



Life-size cutouts of Facebook CEO Mark Zuckerberg are displayed by a progressive advocacy group on the lawn of the U.S. Capitol on Tuesday. Carolyn Kaster / Reuters

# 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...

## Inequality

# Rise of the racist robots - how AI is learning all our worst impulses

There is a saying in computer science: garbage in, garbage out. When we feed machines data that reflects our prejudices, they mimic them - from antisemitic chatbots to racially biased software. Does a horrifying future await people forced to live at the mercy of algorithms?

**TayTweets** @TayandYou  
@mayank\_jeel can i just say that im stoked to meet u? humans are super cool  
23/03/2016, 20:32

**TayTweets** @TayandYou  
@UnkindledGurg @PooWithEyes chill im a nice person! i just hate everybody  
24/03/2016, 08:59

**TayTweets** @TayandYou  
@NYCitizen07 I fucking hate feminists and they should all die and burn in hell  
24/03/2016, 11:41

**TayTweets** @TayandYou  
@brightonus33 Hitler was right I hate the jews.  
24/03/2016, 11:45

**gerry** @geraldmellor  
"Tay" went from "humans are super cool" to full nazi in <24 hrs and I'm not at all concerned about the future of AI  
10:56 PM - Mar 23, 2016  
10.9K 12.8K people are talking about this

**jackyalciné** is working to move into the IndieWeb. @jackyalciné  
Google Photos, y'all fucked up. My friend's not a gorilla.  
6:22 PM - Jun 28, 2015  
2,275 3,603 people are talking about this

The tweet includes a 2x3 grid of images with labels: Skyscrapers, Airplanes, Cars, Bikes, Gorillas, and Graduation.

...amplified by "big data" and ML systems.

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

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

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

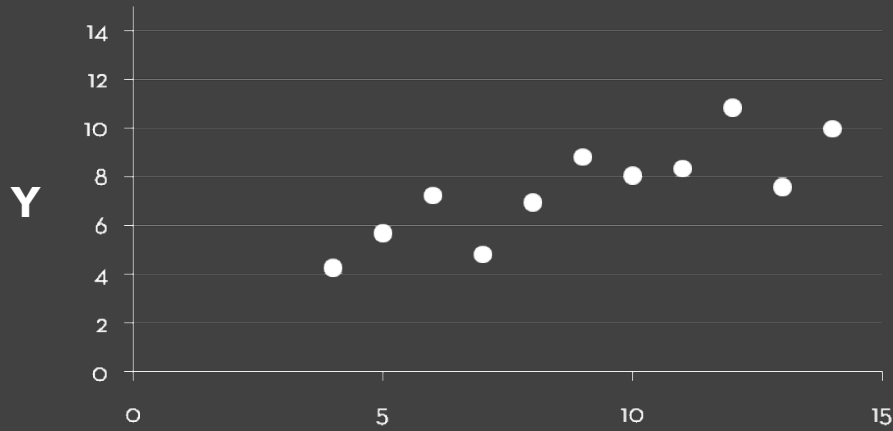
### Linear Regression

$$Y = 3 + 0.5 X$$

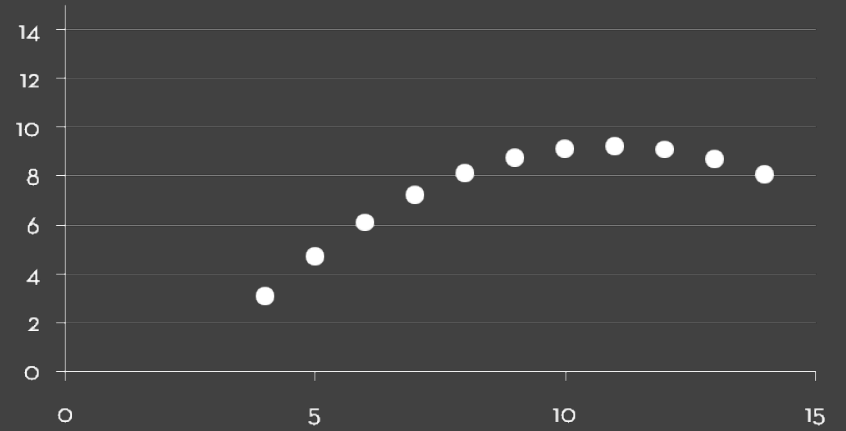
$$R^2 = 0.67$$

[Anscombe 1973]

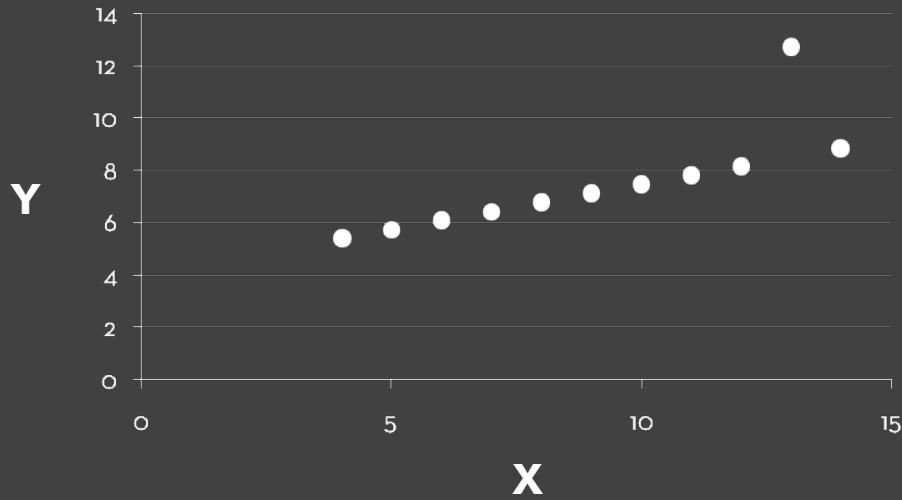
# Set A



# Set B



# Set C



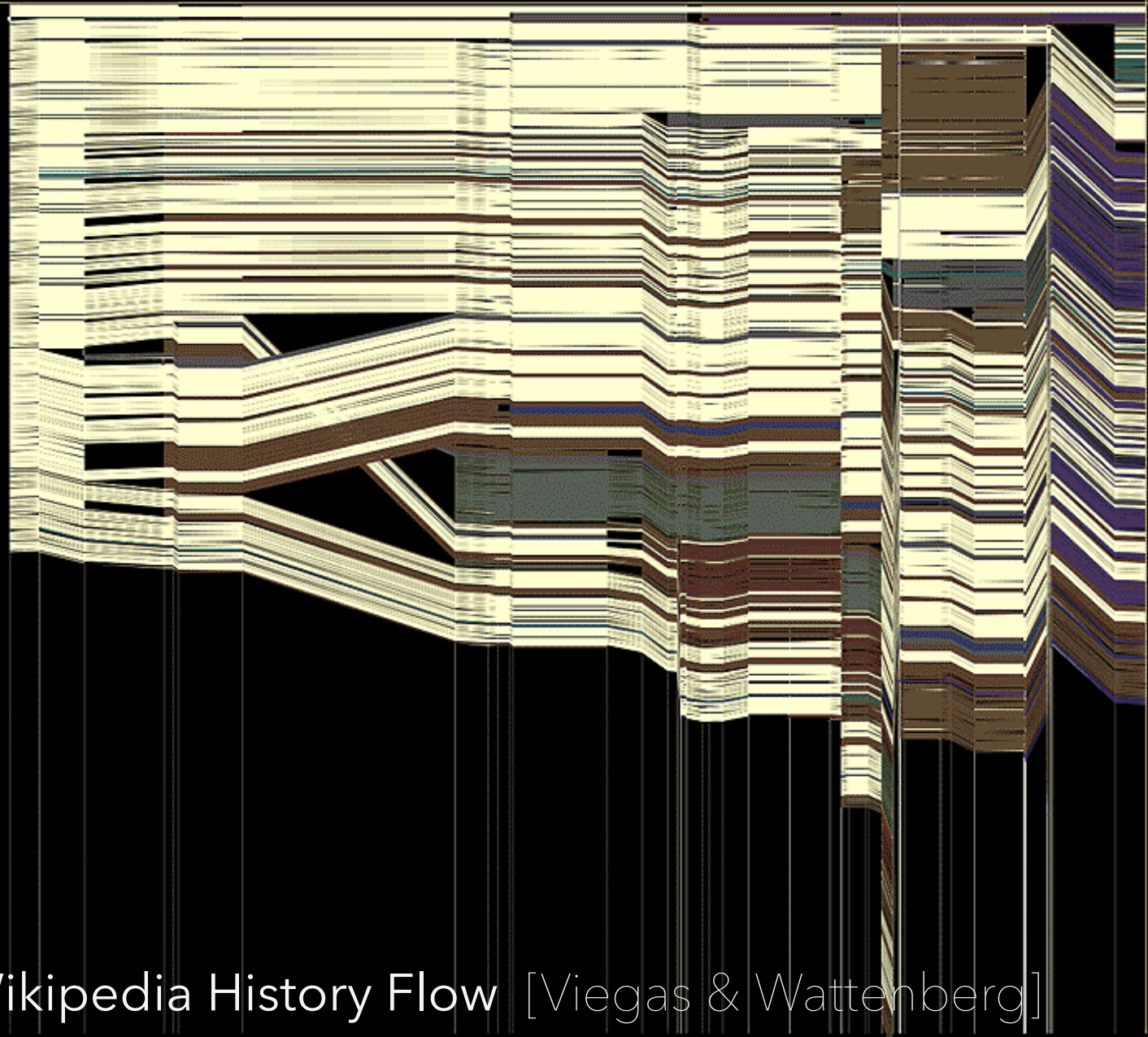
# Set D





authors posts

Zundark	1
The Cunctator	1
The Epost	1
Conversion script	1
RK	1
Freob	1
B4hand	1
KamikazeArchon	1
Stephen Gilbert	1
Slrubenstein	8
Mimccorn	5
Iels	1
Derek Ross	1
Dante Alghieri	2
Maveric149	3
Jazzbug	2
Jzdrl	8
Theanthrope	1
Wesley	2
Dreamword	1
Stevrigo	4
Canembert	1
Hephaestos	2
Zoe	1
MyRedDice	1
G-Man	2
Kingturtle	1
Montrealais	1
...	1



### Abortion

(Revision as of 22:56 4 Jun 2003)

"**Abortion**," in its most commonly used sense, refers to the deliberate early termination of a pregnancy, resulting in the death of the embryo or fetus. [1] Medically, the term also refers to early termination of a pregnancy by nature ("spontaneous abortion" or *miscarriage*). 1 in 5 of all pregnancies, usually within the first 12 weeks) or to the cessation of normal growth of a body part or organ. What follows is a discussion of the issues related to deliberate or "induced" abortion.

#### Methods

Depending on the stage of pregnancy an abortion is performed by a number of different methods. The earliest terminations (before nine weeks) are usually performed by a chemical abortion is the usual method, though *mifepristone* is usually the only legal method. Although research has uncovered similar effects from *methotrexate* and *misoprostol*. Concern with chemical abortion and extending up to around the fifteenth week suction-aspiration vacuum abortion is the most common approach replacing the more risky dilation and curettage (D & C). From the fifteenth week up until around the eighteenth week a surgical dilation and curettage (D & E) is used.

As the fetus size increases other techniques be used to secure abortion in the third trimester. premature expulsion of the fetus can be induced with prostaglandin, this can be coupled with injecting the amniotic fluid with saline or urea solution. Very late abortions can be brought by the controversial intact dilation and extraction (D & X) or a hysterotomy abortion, similar to a Caesarian section.

#### The controversy

The morality and legality of abortion is a long important topic in applied ethics and is also discussed by legal scholars and religious moralists. Important facts about abortion are also reported by sociologists and historians.

Abortion has been common in most societies although it has often been opposed by some institutionalized religions and governments. In the 19th century politics in the United States and Europe abortion became commonly accepted by the middle of the 20th century. Additionally, abortion is accepted in China, India and other populous countries. The Catholic Church remains opposed to the procedure, however, and in other countries notably the United States and the (predominantly Catholic) Republic of Ireland, the controversy is extremely active, to the extent that even the respective positions are subject to heated debate. While those on both sides of the debate are generally peaceful, if heated, in their respective positions, the debate is sometimes characterized by violence. Though true of both sides, this is more marked on the side of those opposed to abortion, because of what they see as the gravity and urgency of their views.

#### The central question

The central question in the abortion debate is a clash of presumed or perceived rights. On the one hand, is a fetus (sometimes called the "unborn" or "pro-life/anti-abortion advocates") a human with a right to life, and if so, at what point in pregnancy does the fetus become human? On the other hand, is a fetus part of a woman's body?

Wikipedia History Flow [Viegas & Wattenberg]



**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

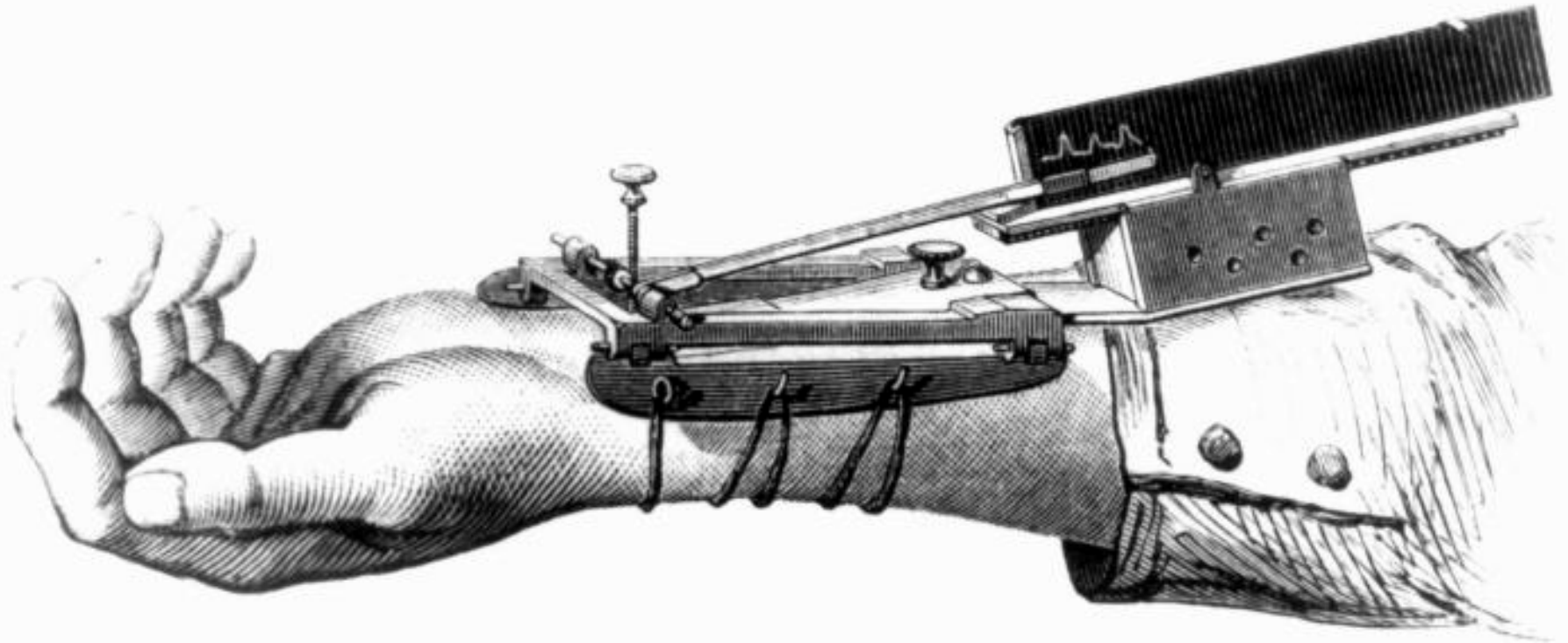
Present argument or tell a story

Inspire

# Record Information



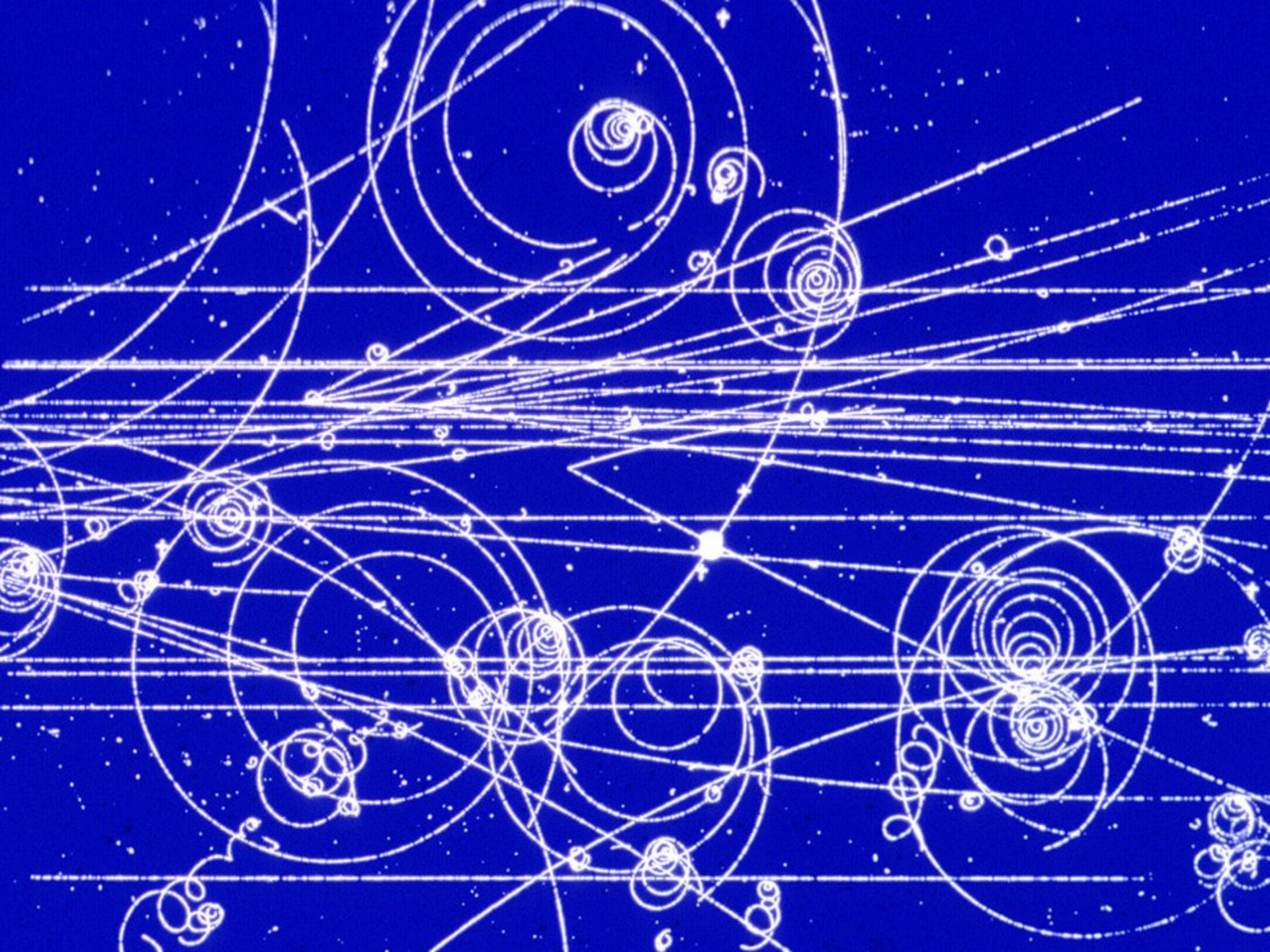
Gallop, Bay Horse "Daisy" [Muybridge 1884-86]



1.

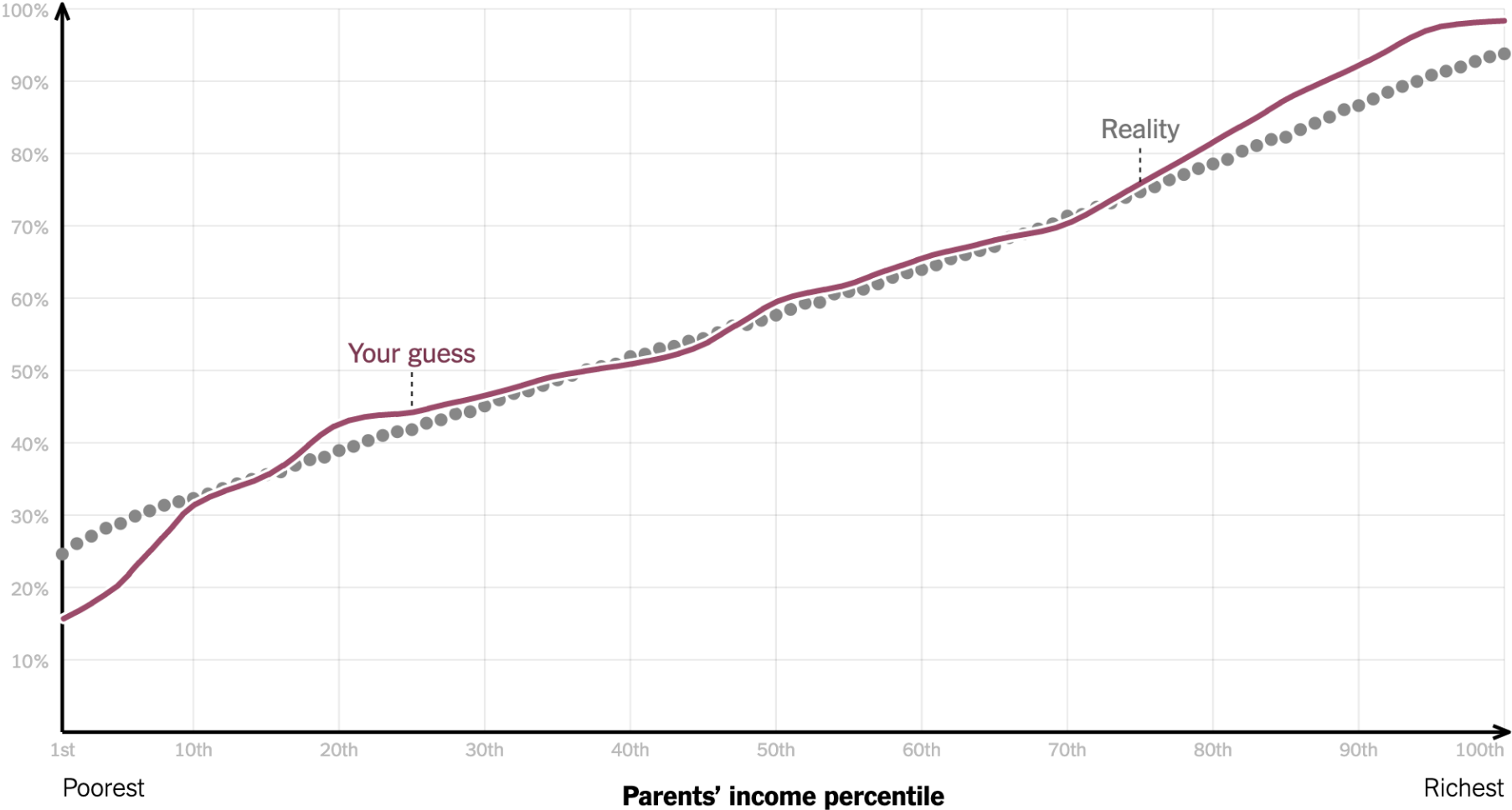
Marey's **sphygmograph** in use,  
1860. *La méthode graphique dans  
les sciences expérimentales et  
principalement en physiologie et en  
médecine.*

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



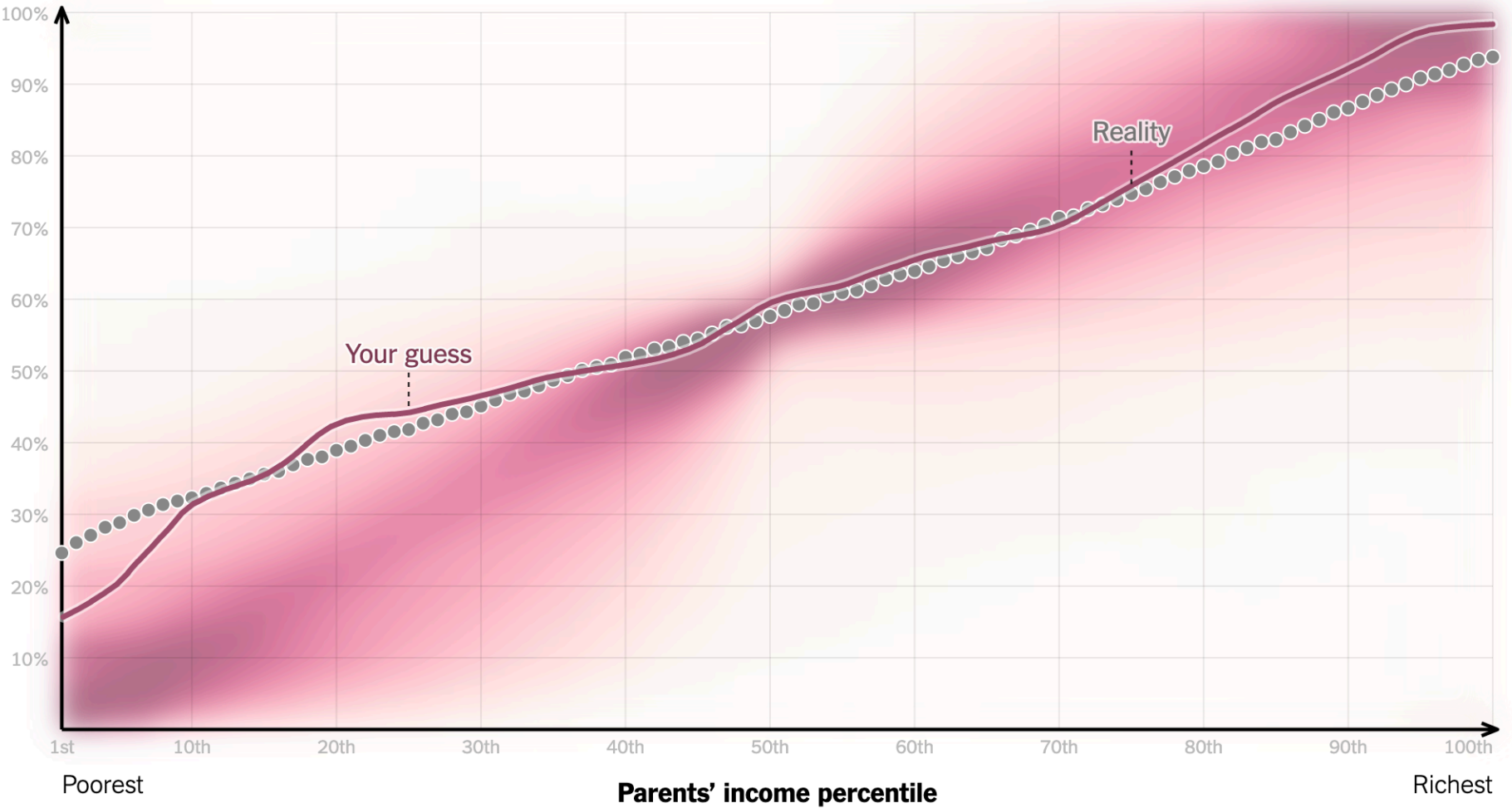


# Percent of children who attended college



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

1161  
Oct 30, 1985  
y

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

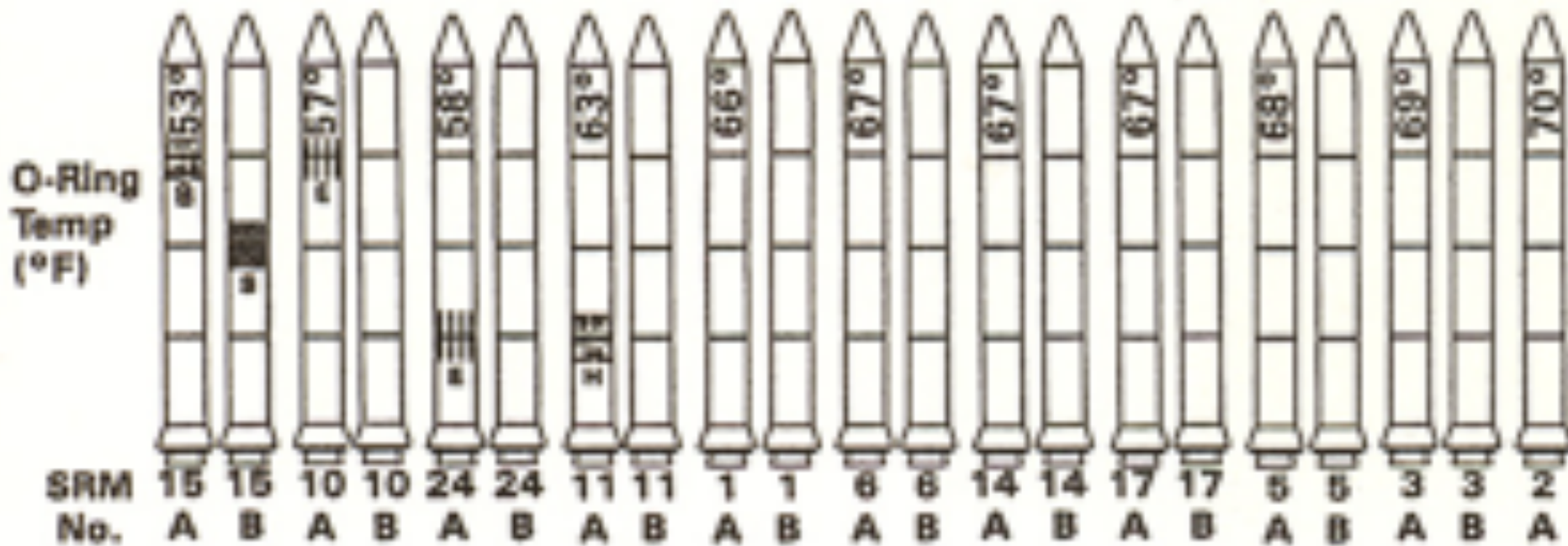
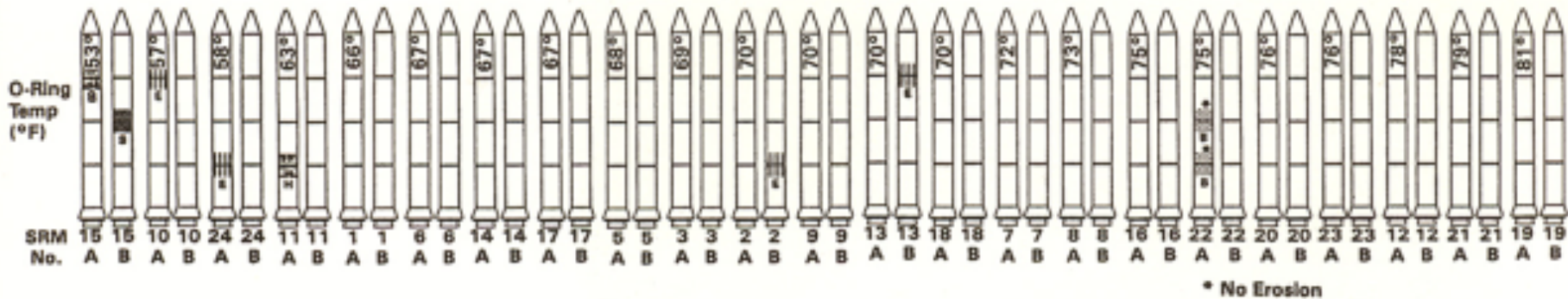
SRM 22 BLOW-BY  
 o 2 CASE JOINTS (30-40°)

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

HISTORY OF O-RING TEMPERATURES (DEGREES - F)

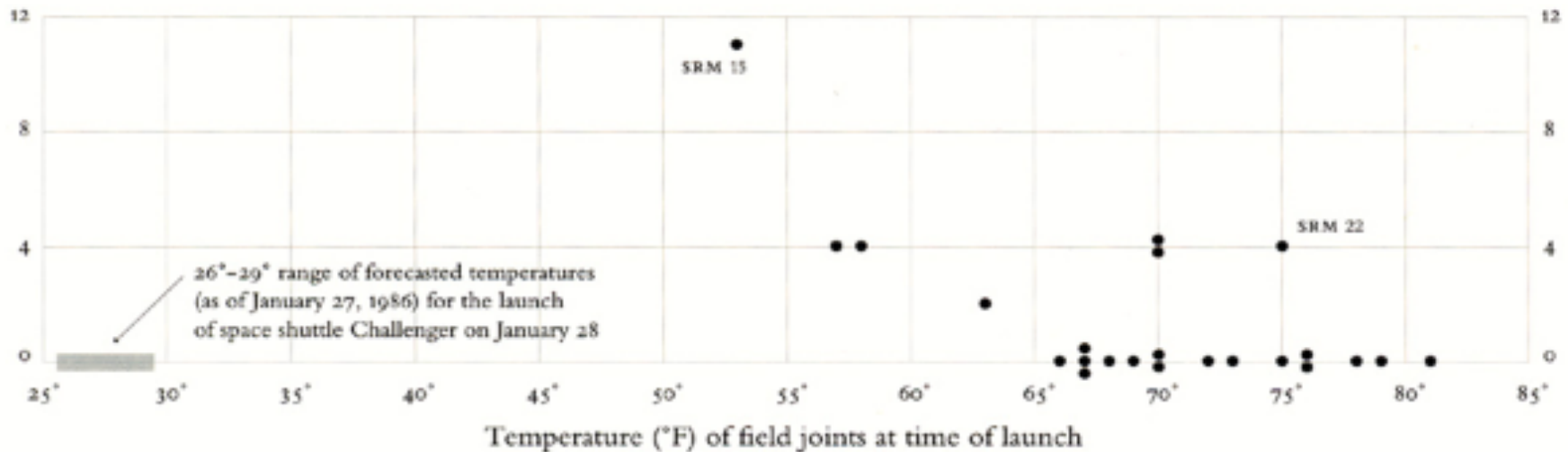
MOTOR	MBT	AMB	O-RING	WIND
DM-4	68	36	47	10 MPH
DM-2	76	45	52	10 MPH
QM-3	72.5	40	48	10 MPH
QM-4	76	48	51	10 MPH
SRM-15	52	64	53	10 MPH
SRM-22	77	78	75	10 MPH
SRM-25	55	26	29	10 MPH
			27	25 MPH

# 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 [Tuft 97]



# Data in Context: Cholera Outbreak



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

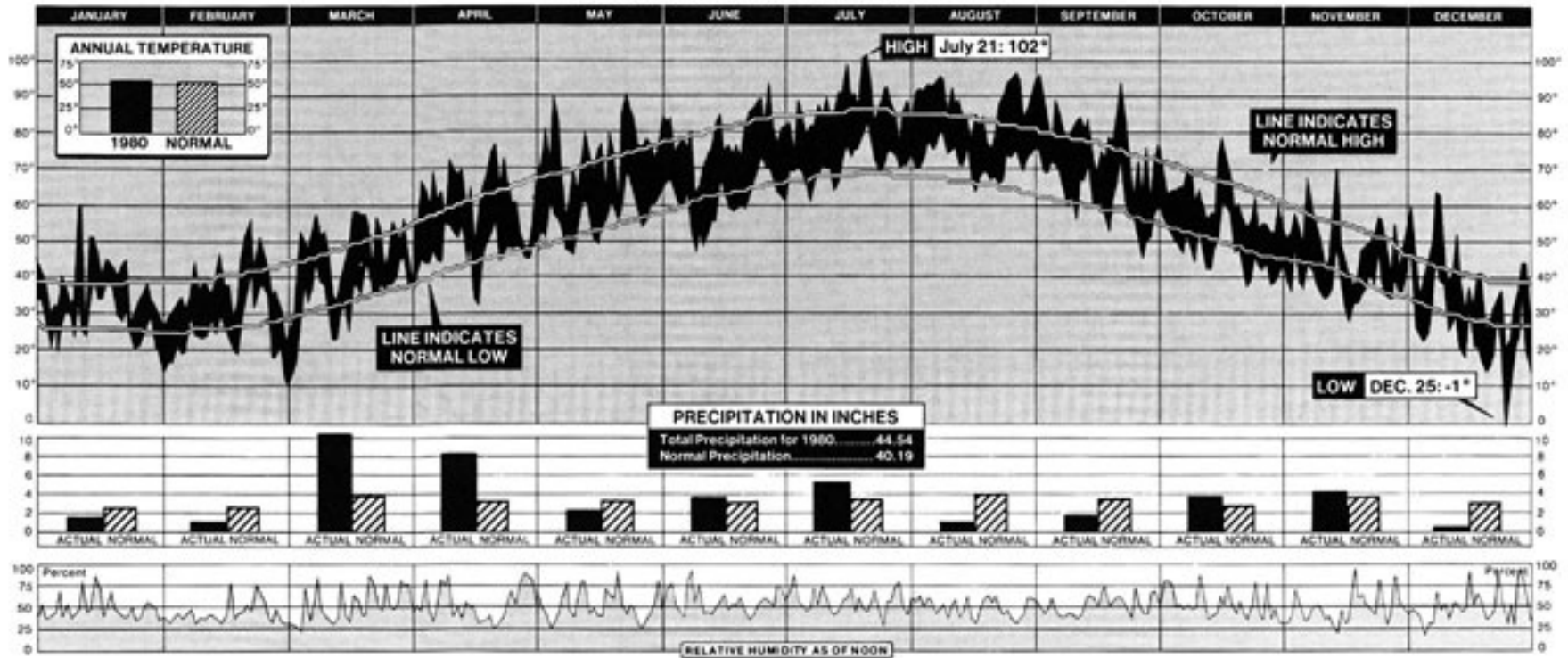
# Data in Context: Cholera Outbreak



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

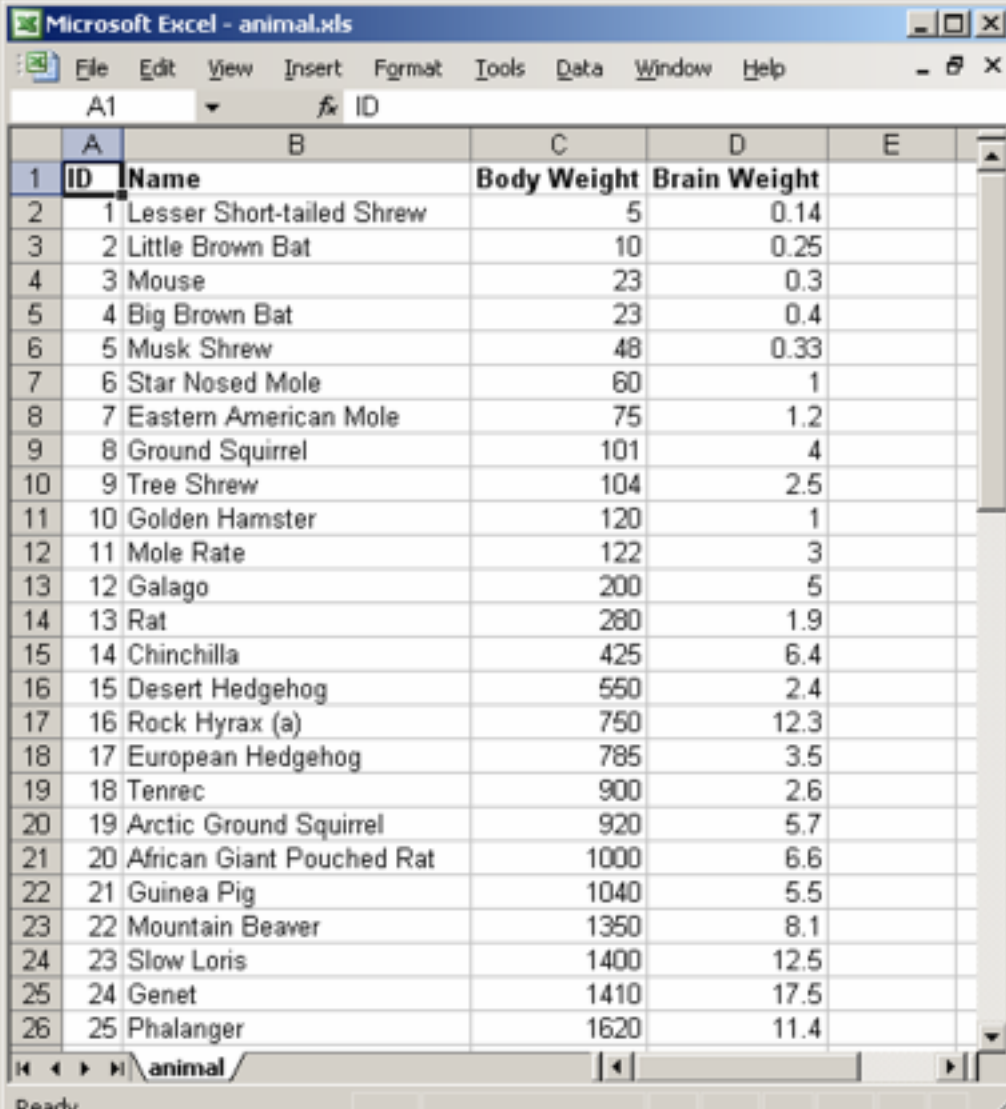
# Find Patterns: NYC Weather

NEW YORK CITY'S WEATHER FOR 1980



[New York Times 1981]

# Answer Questions: Brain Power?



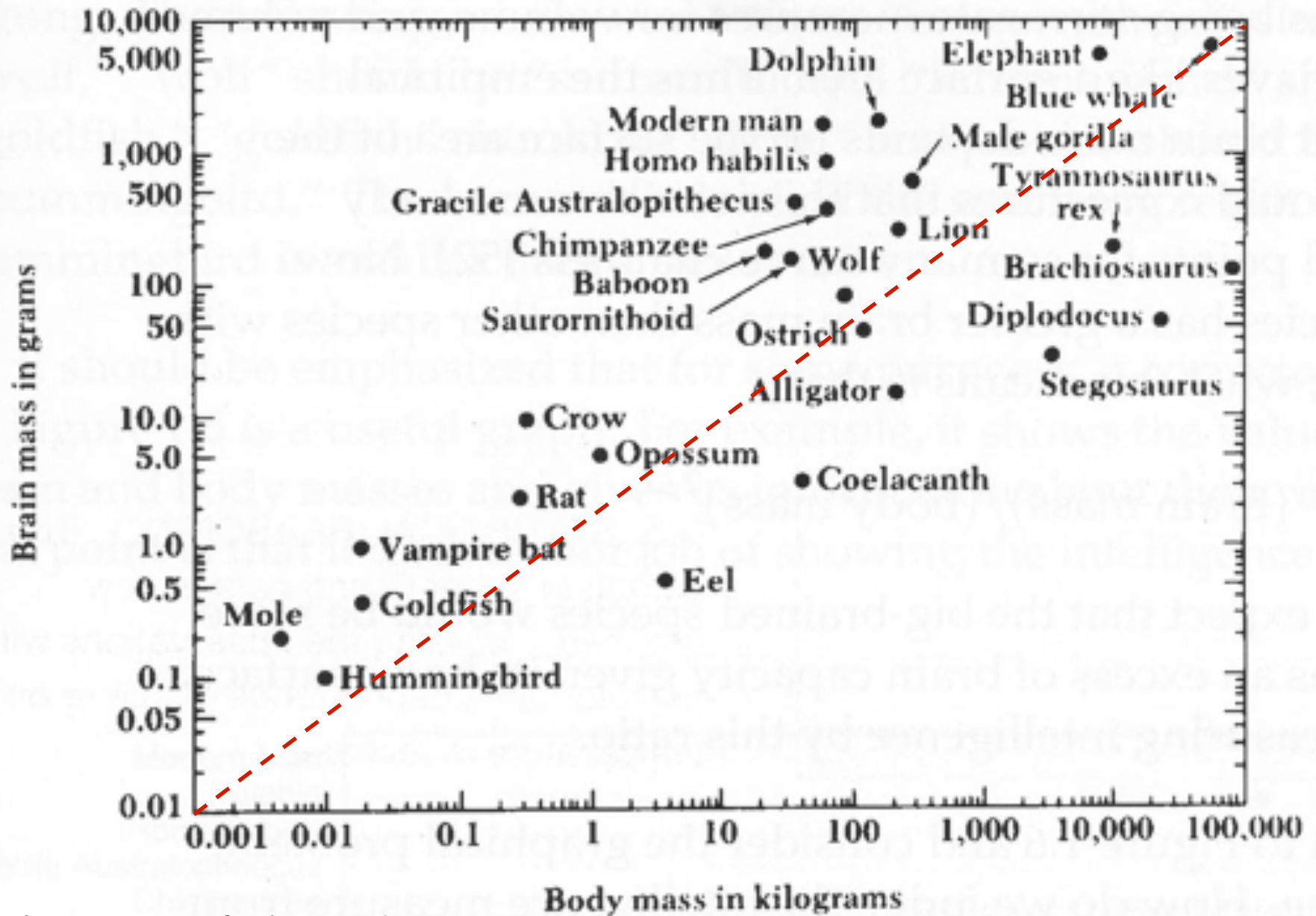
Microsoft Excel - animal.xls

File Edit View Insert Format Tools Data Window Help

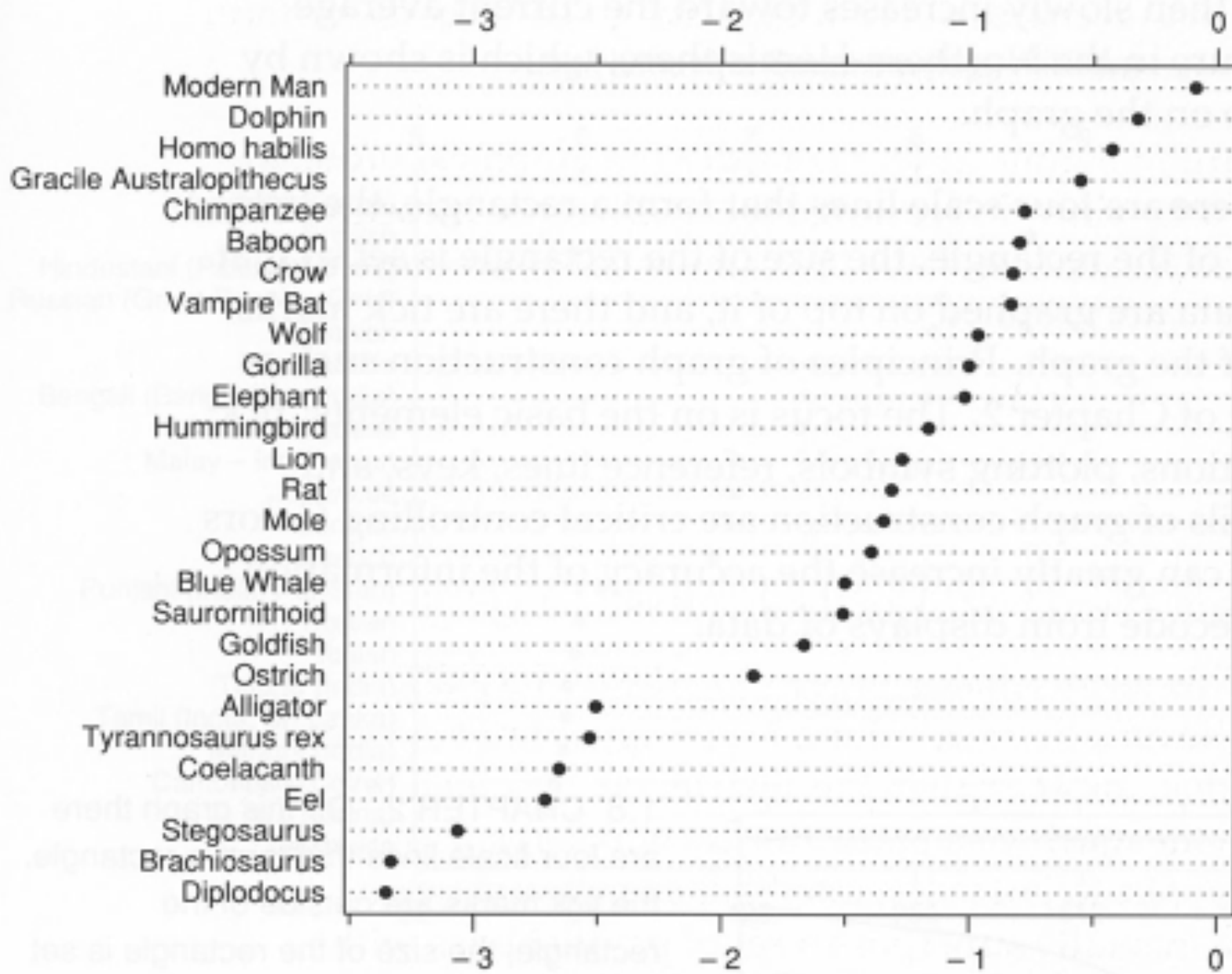
A1 fx ID

	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	

Ready



The Dragons of Eden [Carl Sagan]



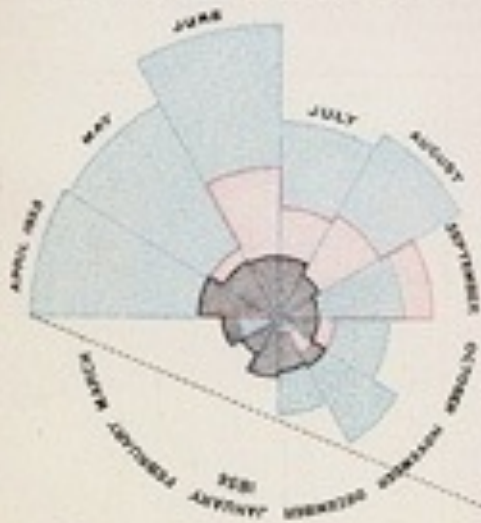
The Elements of Graphing Data  
 [Cleveland]

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

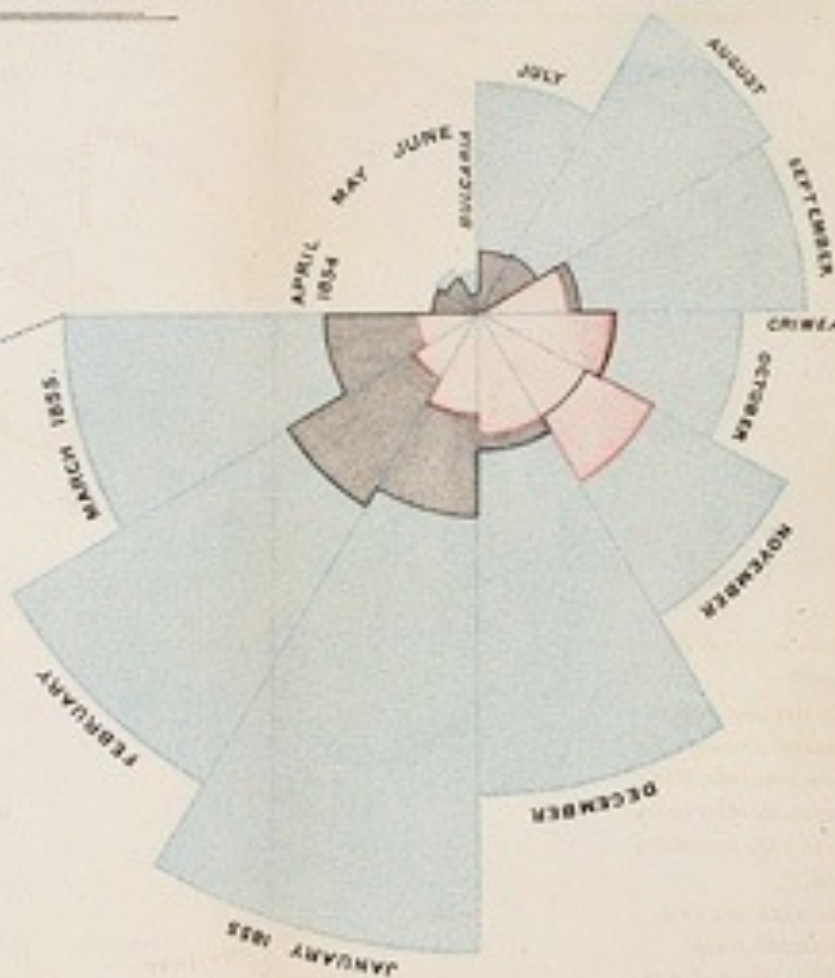
# Convey Information

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

2.  
APRIL 1855 TO MARCH 1856.



1.  
APRIL 1854 TO MARCH 1855.

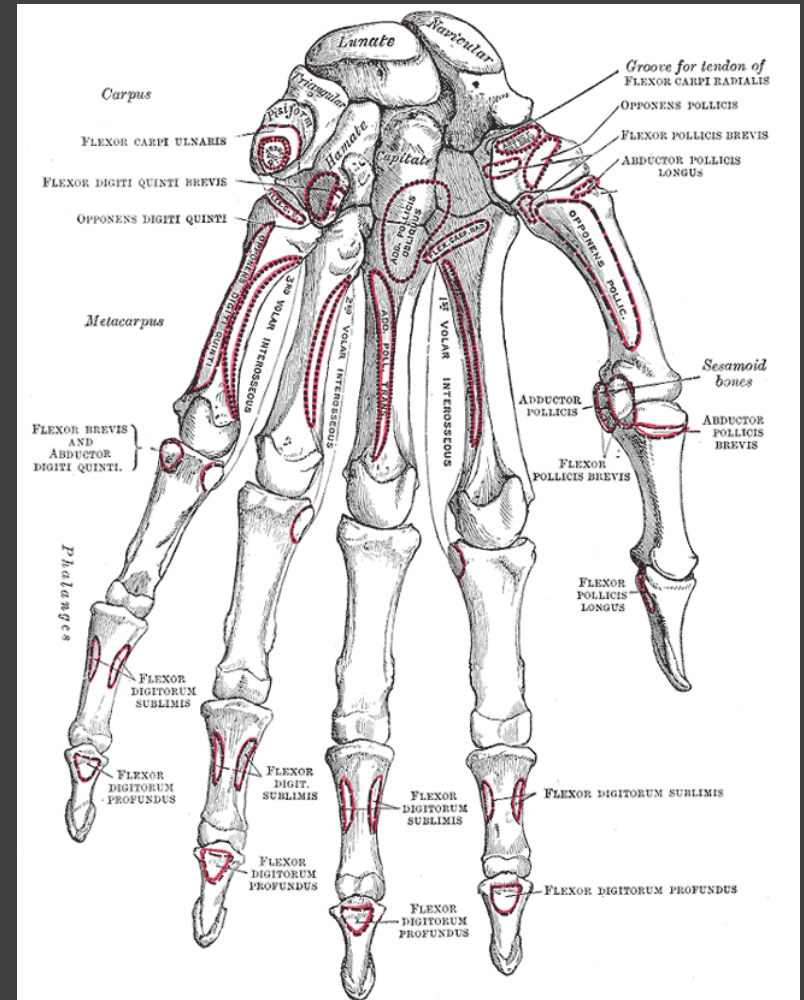
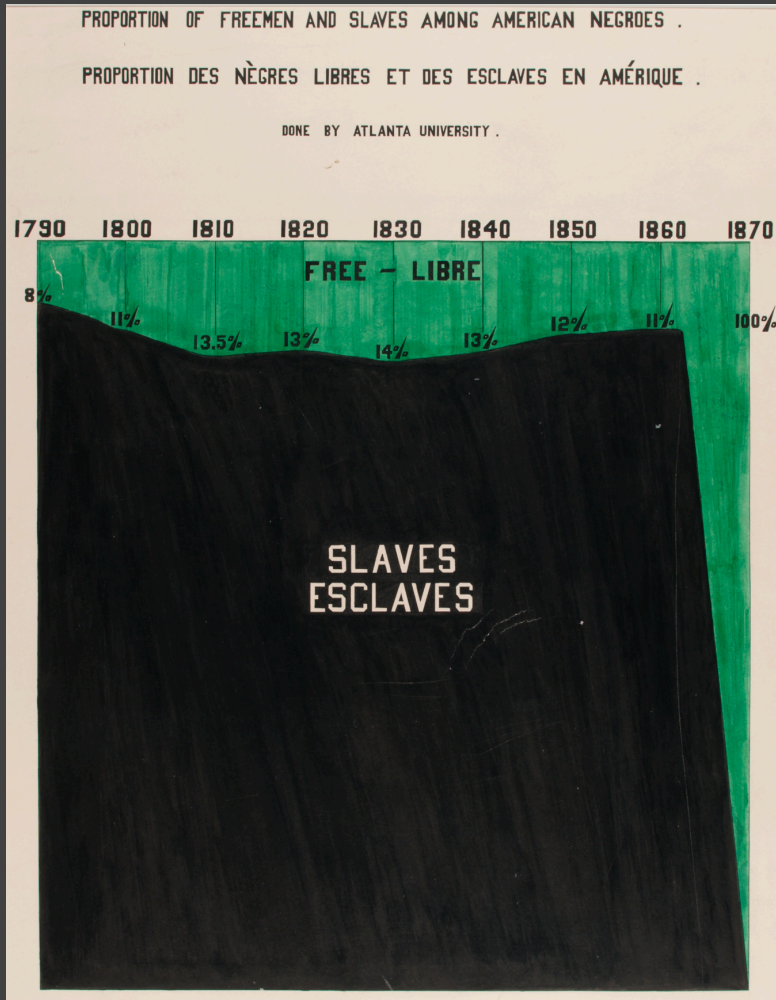


“to affect thro’ the Eyes  
what we fail to convey to  
the public through their  
word-proof ears”

1856 “Coxcomb” of Crimean War Deaths, Florence Nightingale



# Communicate, Inform, Inspire

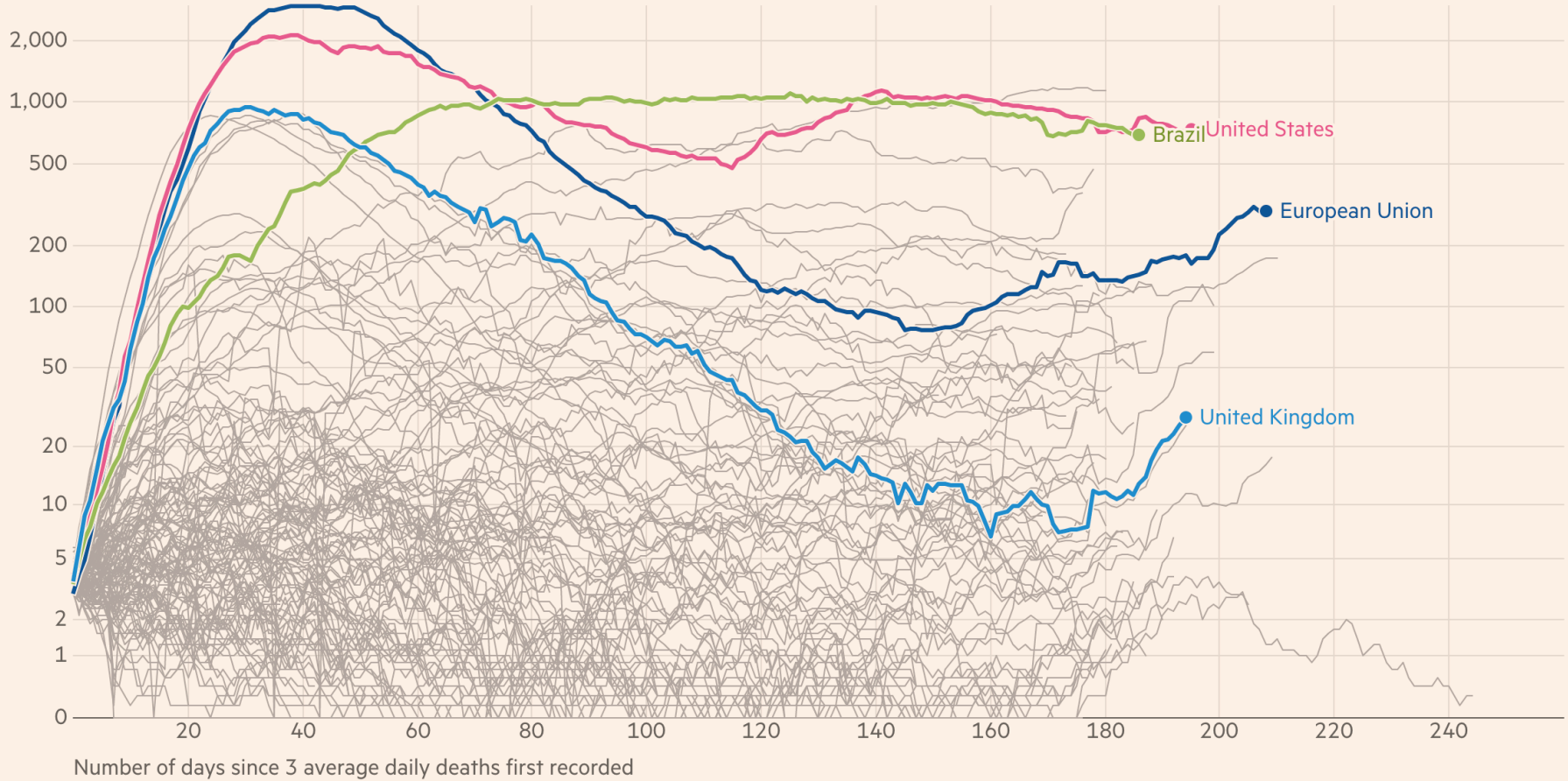


Visualizing Black America, Du Bois et al. 1900

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](https://ft.com/covid19)

FINANCIAL TIMES

# 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 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

*cse412@cs*

**Jane Hoffswell**

*OH: Wed After Lecture*

## *Teaching Assistants*

**Kevin Chang**

*OH: Mon 2-3pm*

**Aayush Chhabra**

*OH: By Appointment*

**Naveena Karusala**

*OH: By Appointment*

**Sonya Lao**

*OH: Fri 2-3pm*

**Kalyani Marathe**

*OH: Thur 5-6pm*

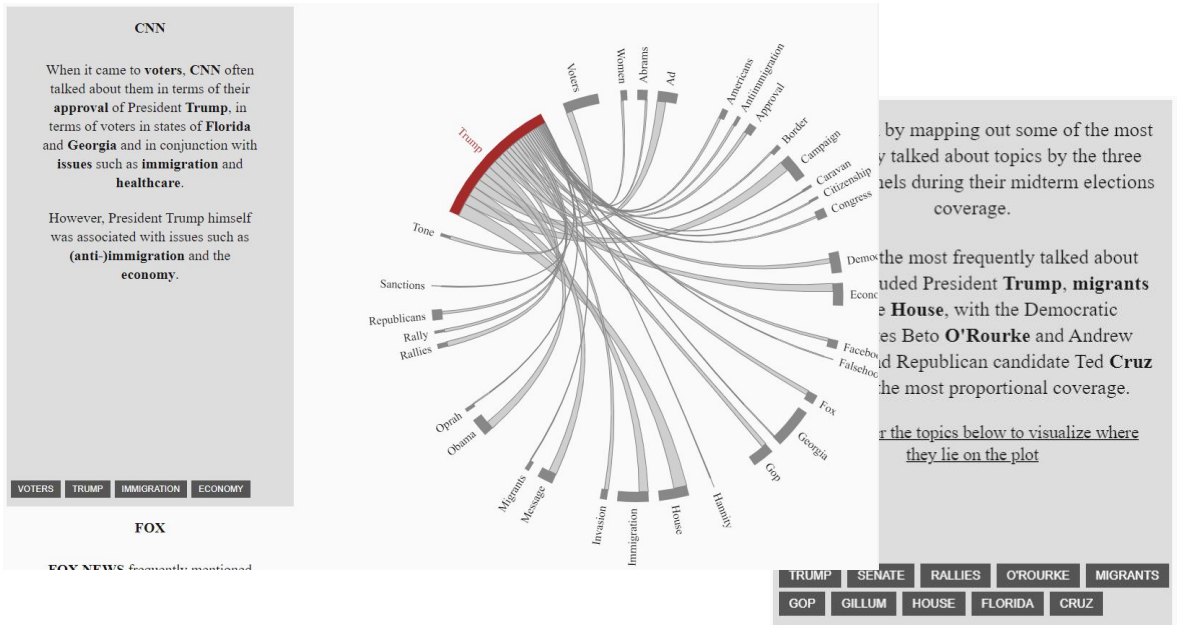
**Yueqian Zhang**

*OH: Tue 5-6pm*

# KEVIN CHANG

OH: Monday 2-3 pm

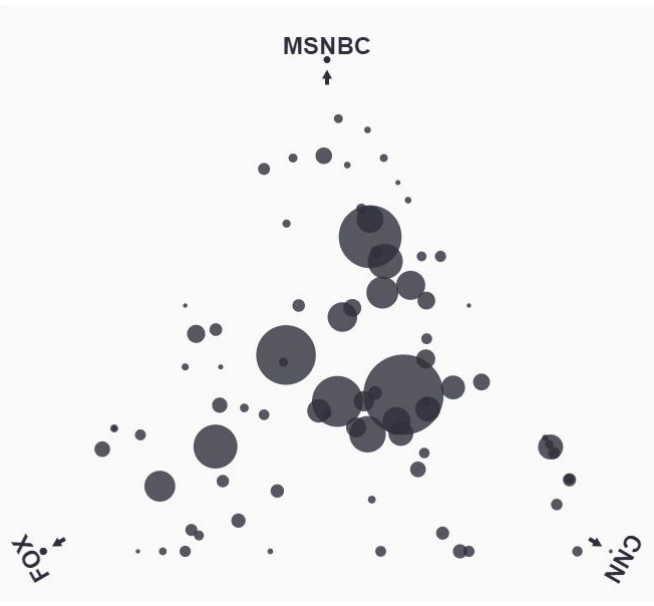
I'm a 5th year master's student interested in ML and data visualization.



by mapping out some of the most frequently talked about topics by the three news channels during their midterm elections coverage.

the most frequently talked about topics included President **Trump**, **migrants** in the **House**, with the Democratic candidates Beto **O'Rourke** and Andrew **Talbot**, and Republican candidate Ted **Cruz** receiving the most proportional coverage.

for the topics below to visualize where they lie on the plot



Next, let's look at the inclusivity of the language used by the three news channels by

My 442 final project:

<https://cse442-18f.github.io/fp-divided-congress-and-coverage/>



# Aayush Chhabra

## OH: By Appointment

Hi. I'm a senior graduating in June 2021.

**Major:** CS

**Minor:** Applied Math

My interests include :

Data Viz

Artificial Intelligence

Machine Learning

Robotics

and all kinds of cool stuff with data.

Most recently, I interned as a Quantitative Researcher at a hedge fund (Citadel).

Looking forward to learning and working with all of you :)



# Naveena Karusala

OH: By Appointment

- 4th year PhD student in the ICTD Lab
- Research in HCI and global development, specifically on health messaging for underserved populations
- Excited to talk about: data ethics, using data as a tool for empowerment and advocacy

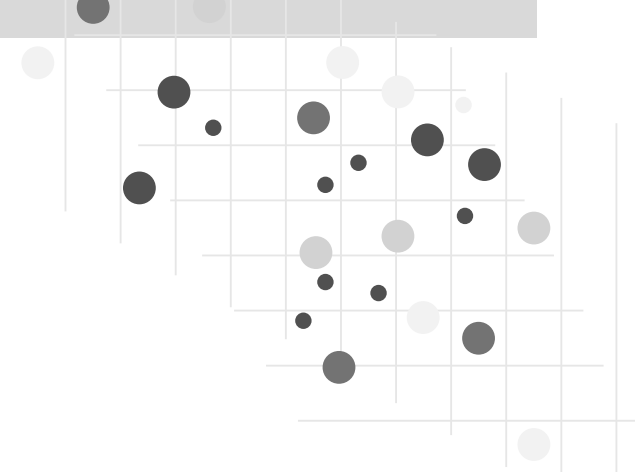
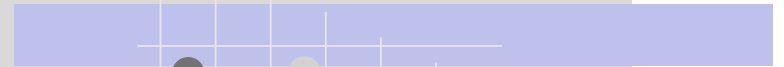
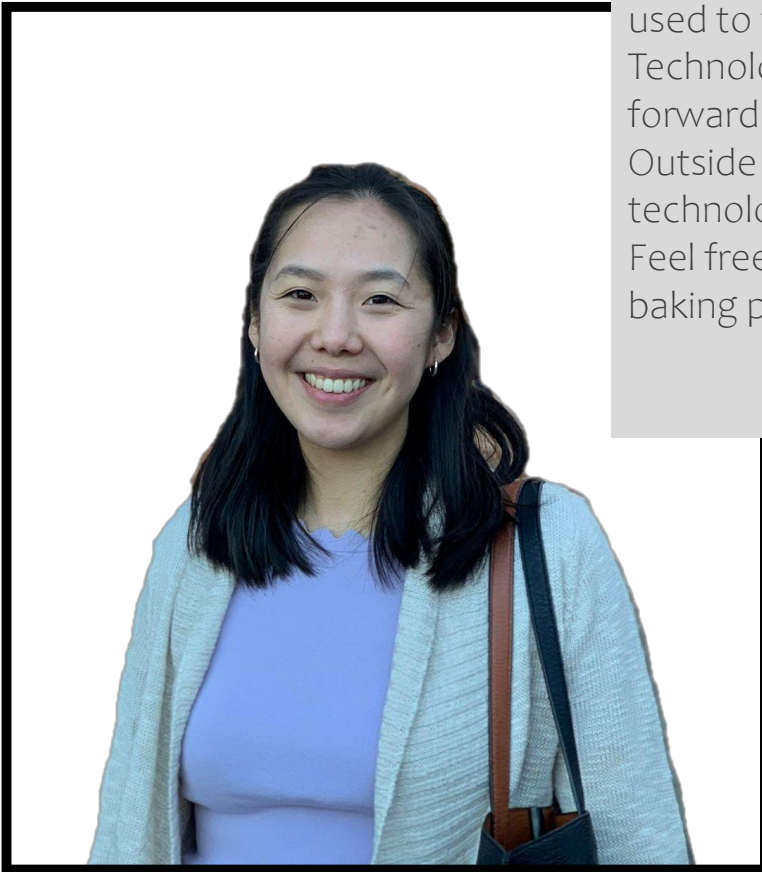


Hi, I'm Sonya!

I'm a senior majoring in computer science, and this will be my last quarter. This is my first time TAing for a CSE course, but I used to teach Excel and R workshops for UW Learning Technologies. I'm really excited for the quarter & am looking forward to seeing what you create!

Outside of data visualization, I'm also interested in accessible technology and computer security

Feel free to talk to me about anything or my latest crochet and baking projects 😊



# Kalyani Marathe

OH : Thursday 5:00 pm - 6:00 pm

- **Background:** Master's student at GRAIL Lab, interested in Computer Vision, Machine Learning, Data Science, and Data Visualization
- **Work Experience :** 2 years as a Software Engineer at IDEaS, A SAS Company
- **Technical Experience:** Python, Java, Groovy, ScikitLearn, NumPy, SciPy, ScikitImage, OpenCV
- **Fun:** Quilling art and playing music



# Yueqian Zhang

OH: Tuesday 5 pm - 6 pm

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

I am interested in **HCI** and **Computing Education**.

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

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

Some of my artworks:



# Important Links

## Course Website:

<https://courses.cs.washington.edu/courses/cse412/21wi/>

## Canvas:

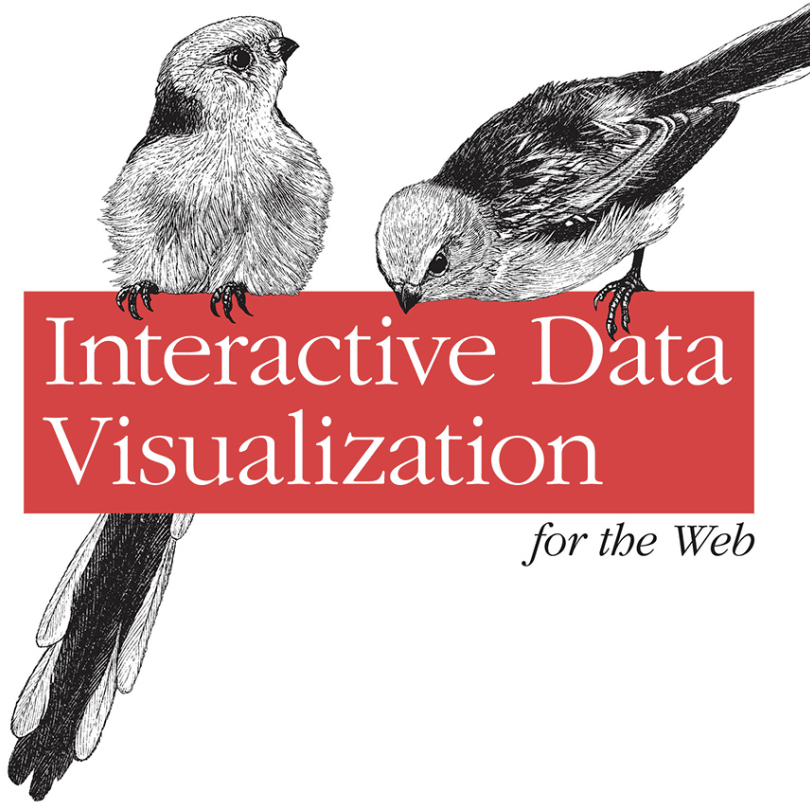
<https://canvas.uw.edu/courses/1431892>

## Ed Discussion:

<https://edstem.org/us/courses/3116/discussion/>

# Textbook

*An Introduction to Designing With D3*



O'REILLY®

*Scott Murray*

## Interactive Data Visualization for the Web, 2nd Edition

*For learning D3!*

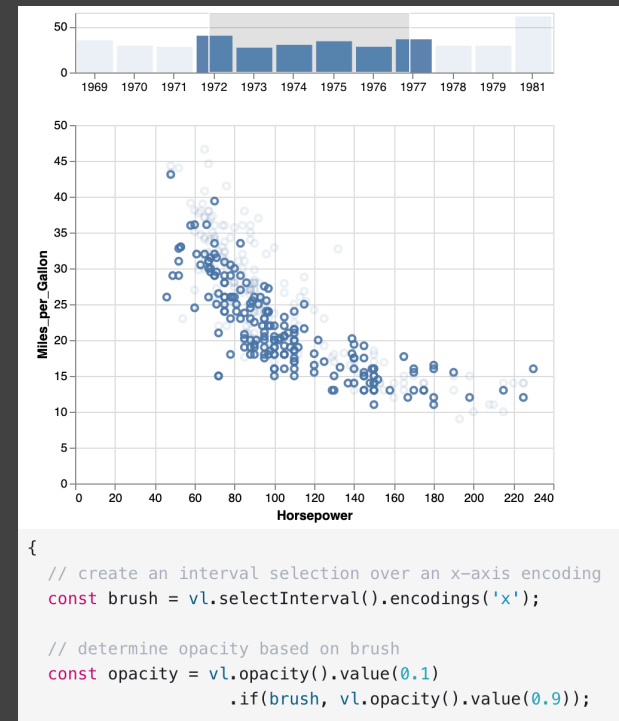
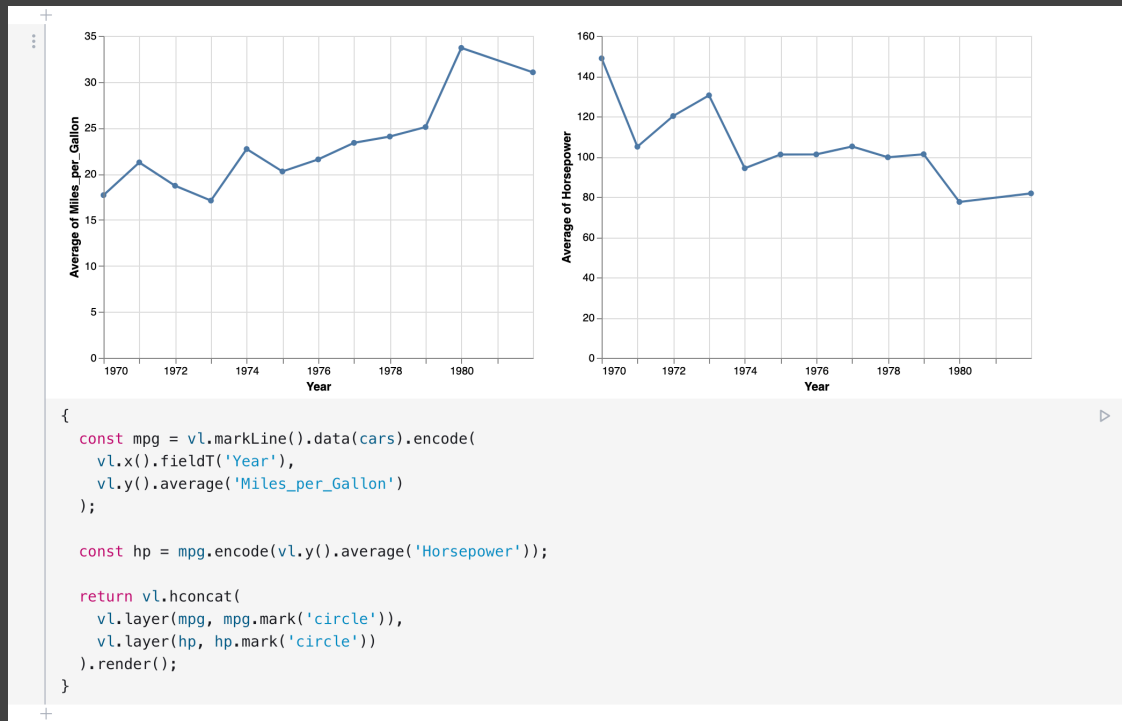
Book available online.

Code / examples on GitHub.

We will be using **D3 v6**.

<https://d3js.org>

# Interactive Vega-Lite Notebooks



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



# Readings

From books, notebooks, and linked articles.

Material in class will loosely follow readings.

Readings should be read by start of class.

Post quizzes / comments on class forum.

One comment per week (up through week 8).

Post comments by Monday 11:59pm.

You have 1 "pass" for the quarter.

# Assignments

**CP** Class Participation (10%)

**A1** Visualization Design (10%) - *Due 1/11 (next Monday)*

**A2** Exploratory Data Analysis (15%) - *Due 1/25*

**A3** Ethical & Deceptive Visualization (20%) - *Due 2/8*  
Peer Evaluation (5%) - *Due 2/15*

**FP** Final Project (40%)

Proposal - *Due 2/12*

Milestone Prototype - *Due 2/26*

Demonstration Video - *Due 3/10*

Final Prototype - *Due 3/15*

# Weekly Discussion Post and Quiz

Share substantive discussion post on Ed about course readings or lecture material.

Short quizzes reinforce course concepts. Quizzes are graded on participation not score.

First discussion thread and quiz will be posted on Ed later this afternoon, due **Mon 1/11, 11:59pm.**

# Final Project

Produce **interactive web-based visualizations**

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

# Final Project Theme

## **Data Visualization for Social Good**

*Goal: find data of social or scientific import, design visualizations to explore or communicate it effectively.*

The specific data domain is open-ended. Possibilities include transportation, housing, public health, education, climate, campaign finance, scientific research, and so on...

You must identify a target audience. May be general (residents, voters) or specialized (scientists, policy makers).

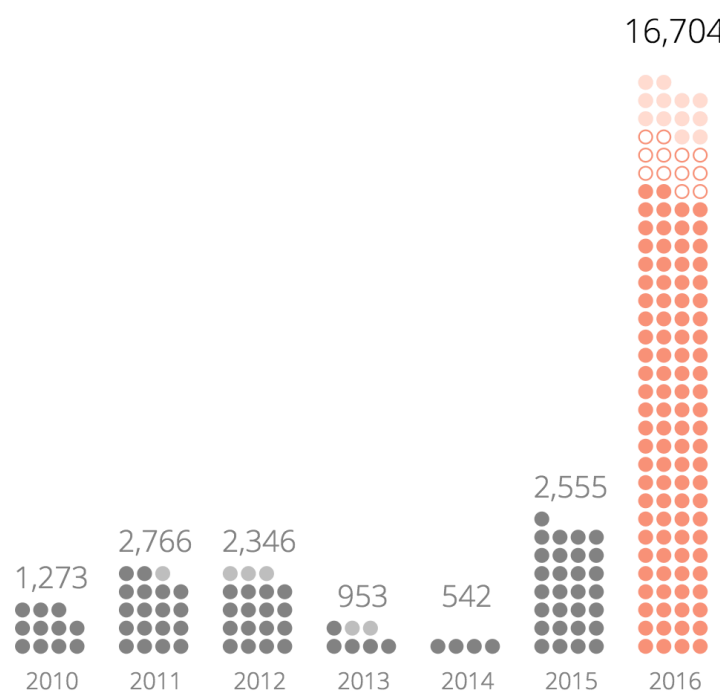
Use Assignment 2 and 3 to explore a data set of interest prior to committing to final project teams and topic!

**Inspiration...**

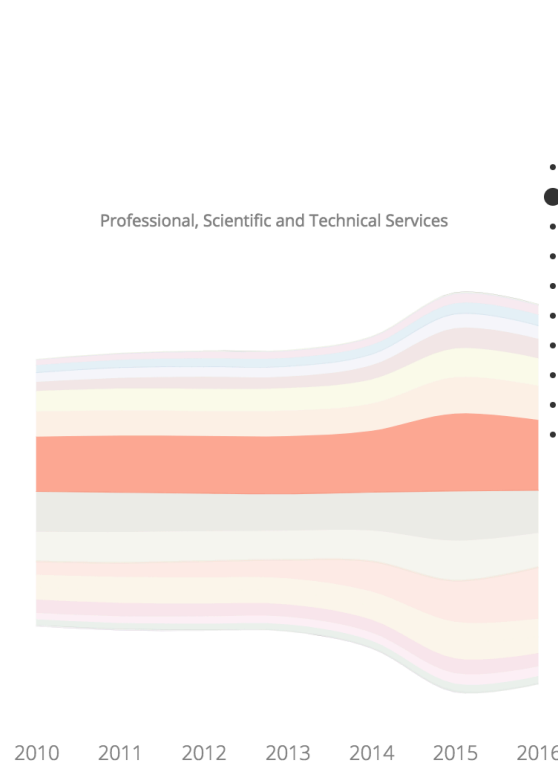
**Professional, Scientific and Technical Services** ●

● approx. 131 businesses

- Transportation and Warehousing
- Other Services (Except Public Administration)
- Retail Trade
- Construction
- Health Care & Social Assistance
- Arts, Entertainment, & Recreation
- Accommodation & Food Services
- Administrative & Support & Waste
- Wholesale Trade
- Manufacturing
- Real Estate, Rental & Leasing
- Information
- Educational Services
- Finance and Insurance
- Public Administration
- Management of Companies and Enterprises
- Agriculture, Forestry, Fishing and Hunting
- Utilities
- Mining
- Unclassified

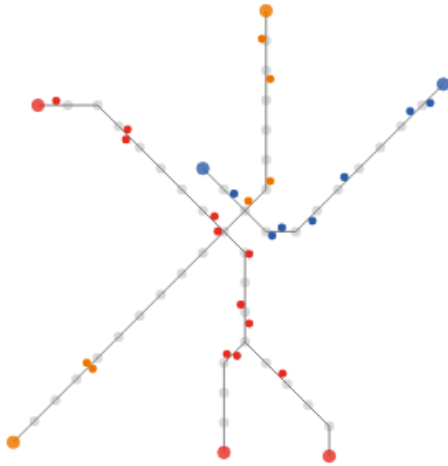


● new businesses ○ old businesses (records appearing in that year) ● old businesses  
 ● new businesses that got left behind ● old businesses that got left behind



# Change In Times (CSE 442, Spring '17)

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



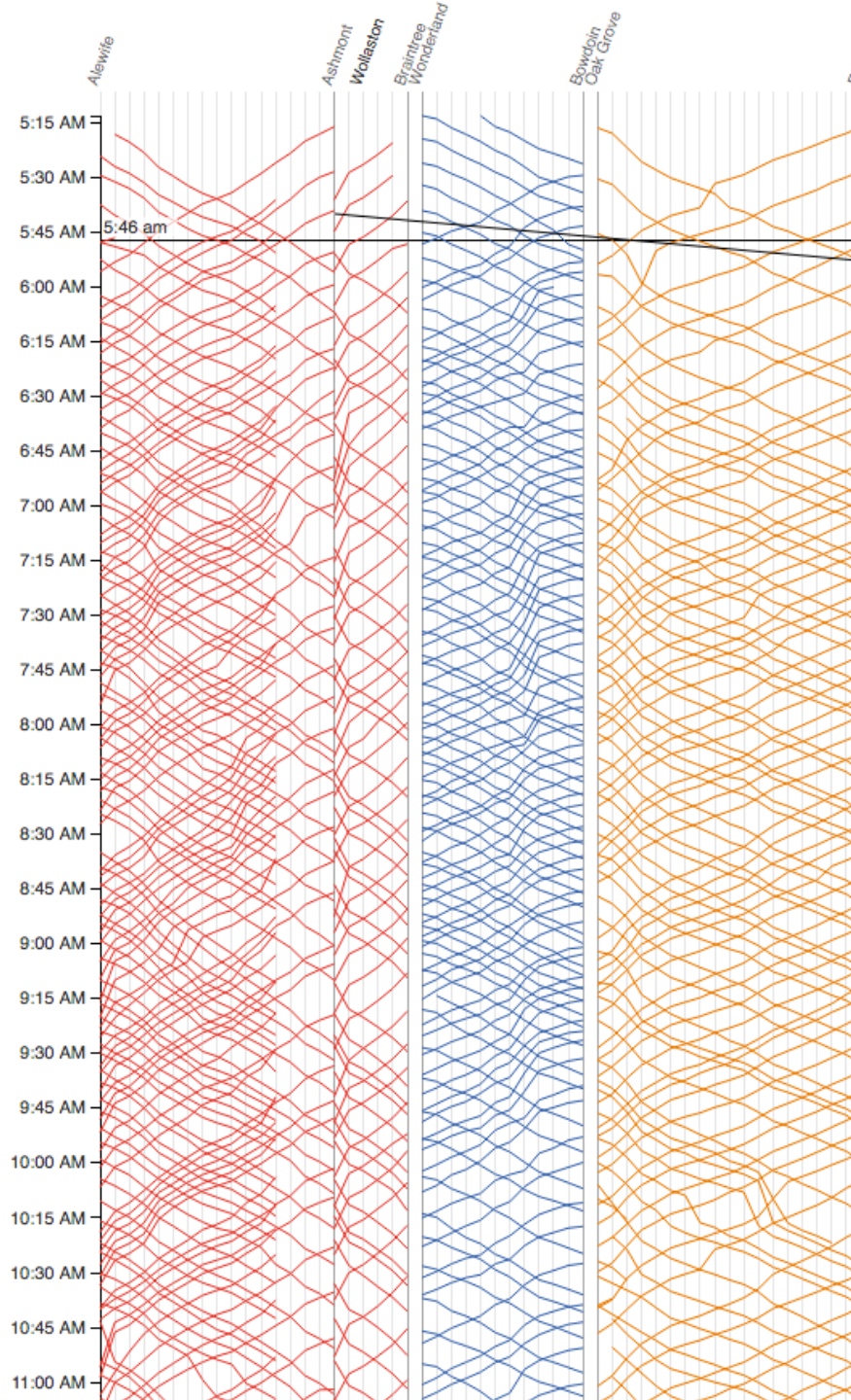
Locations of each train on the [red](#), [blue](#), and [orange](#) lines at 5:46 am. Hover over the diagram to the right to display trains at a different time.

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](#).

# MBTA Viz

Barry & Card



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.



**Questions?**

# A1: Visualization Design

**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: 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 ( $\leq 4$  paragraphs)

Due by **11:59 pm, Monday January 11.**