

The Value of Visualization

DSC 106: Data Visualization

Jared Wilber

UC San Diego

About Me

Senior Software Engineer, Nvidia

Lecturer, HDSI

jwilber.me

jwilber@ucsd.edu

Work stuff I like: Data visualization,
machine learning, software engineering,
open-source.

Non-work stuff I like: my partner Bree,
my dog Bruce, good food, good surf,
milk tea,



Course staff

Instructor

Jared Wilber

Teaching Assistants

Giorgia Nicolaou (Head TA)

Amirhossein Panahi

~~**Aditya Mandke**~~


~~**Gurpreet Saluja**~~

~~**Shaokang Jiang**~~

~~**Vancheeswaran Vaidyanathan**~~

~~**Sai Nelakonda**~~

~~**Deevanshu Goyal**~~



Questions about
course logistics?
Email Giorgia!

See dsc106.com
for our OH times

But first

Let's look at some cool viz

How much data are we producing?

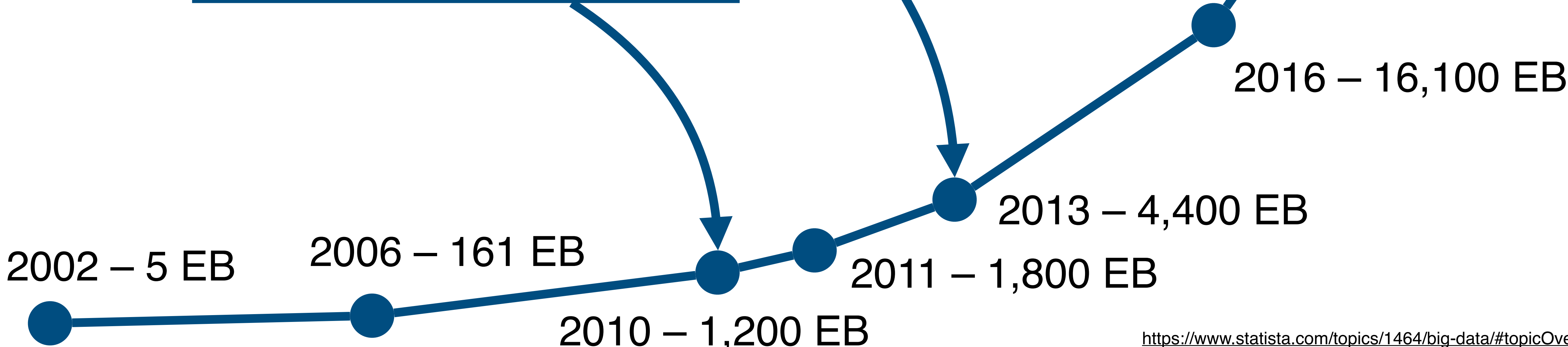
(1 exabyte or 1 EB = 1 *million* terabytes)

2023 – 120,000 EB

But what is in all this data??

A stack of iPads that stretch 2/3rds of the way to the Moon! 🚀

A stack of DVDs stretching from the Earth to the Moon, and back!



Physical Sensors



Health and Medicine



Records of Human Activity



facebook

<https://www.facebook.com/notes/10158791468612200/>

December 2010

"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 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 the ability **"ubiquitous" about whom?** to extract value from it."

"value" to whose benefit?

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

A screenshot of a Twitter thread showing the progression of Tay's tweets. The top row shows two tweets from TayTweets (@TayandYou): one praising humans as 'super cool' and another expressing hatred for everyone. The bottom row shows two more tweets: one hating feminists and another hating Jews. Below these is a tweet from user 'gerry' (@geraldmellor) commenting on the rapid change in Tay's behavior.

TayTweets @TayandYou
@mayank_je 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

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?

A screenshot of a tweet from user 'jackyalciné' (@jackyalcine) showing a grid of six photos from Google Photos. The photos are labeled with incorrect categories: 'Skyscrapers', 'Airplanes', 'Cars', 'Bikes', 'Gorillas', and 'Graduation'. The tweet text expresses frustration with the 'Gorillas' label on a photo of a friend.

jackyalciné is working to move into the IndieWeb. @jackyalcine
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

...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		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

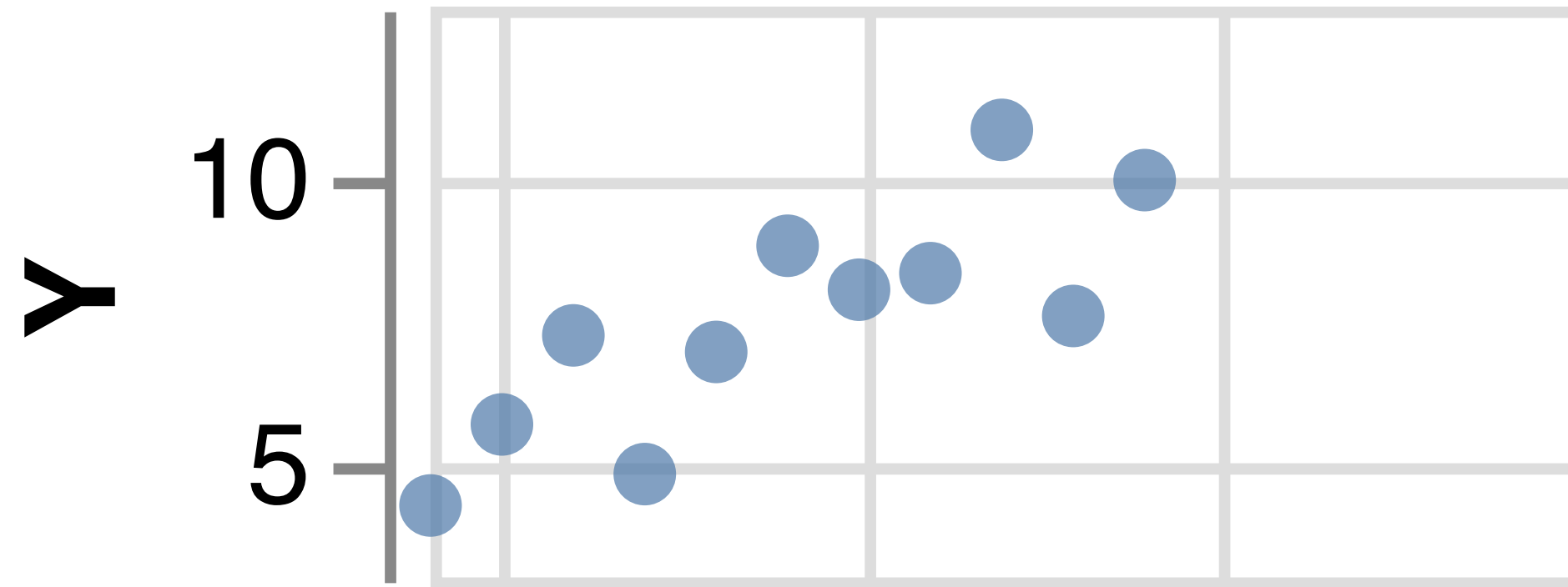
Summary Statistics Linear Regression

$$u_X = 9.0 \quad \sigma_X = 3.32 \quad Y^2 = 3 + 0.5 X$$

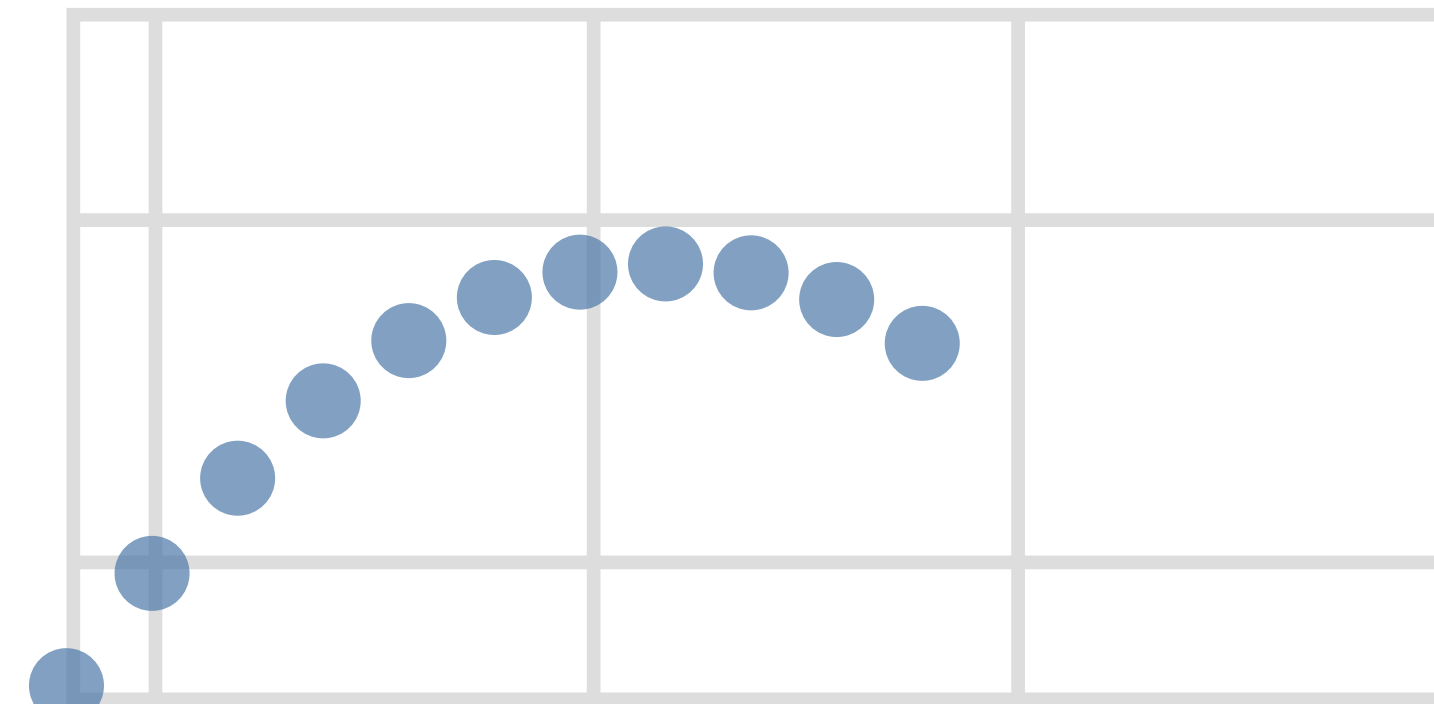
$$u_Y = 7.5 \quad \sigma_Y = 2.03 \quad R^2 = 0.67$$

[Anscombe 1973]

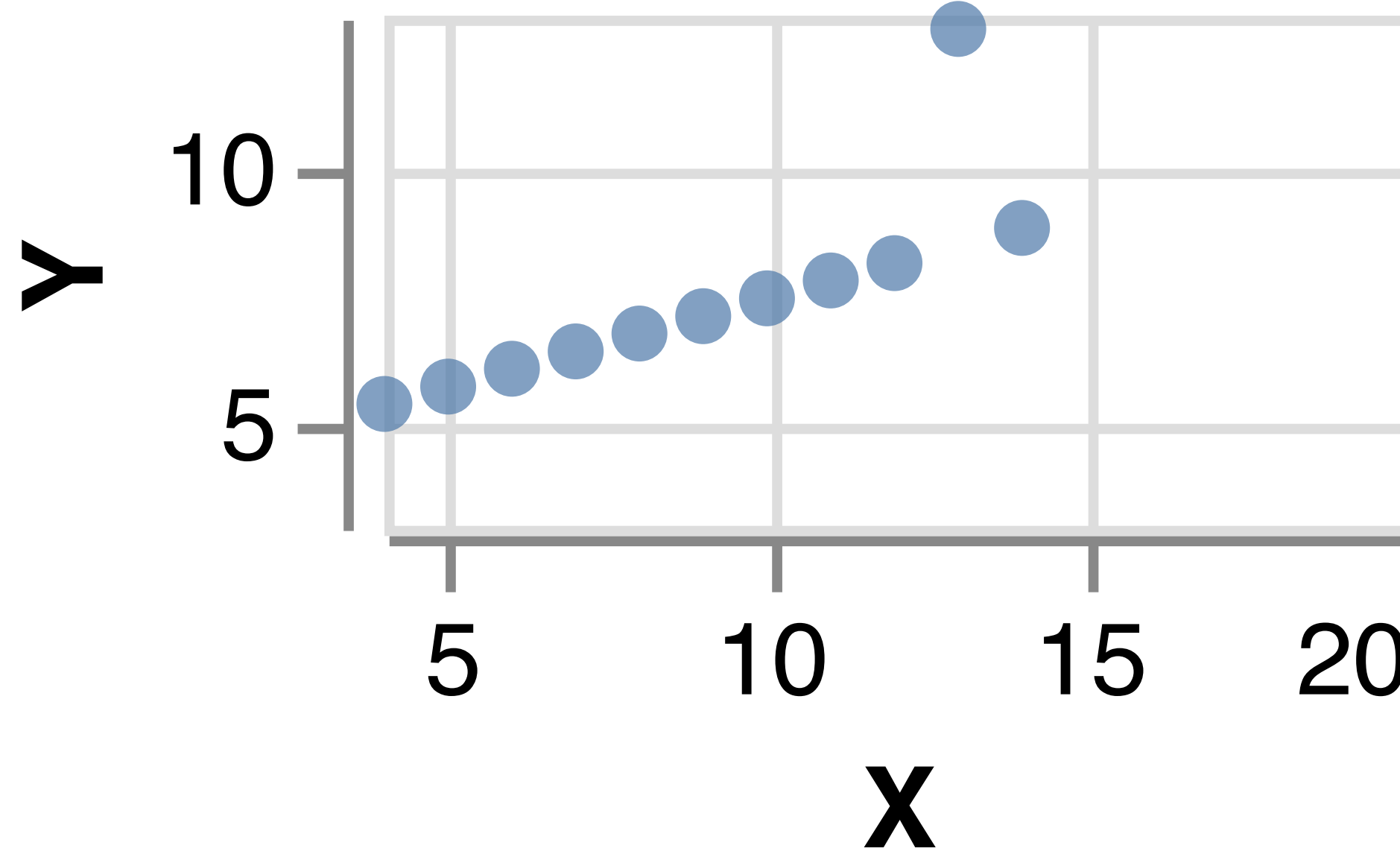
Set A



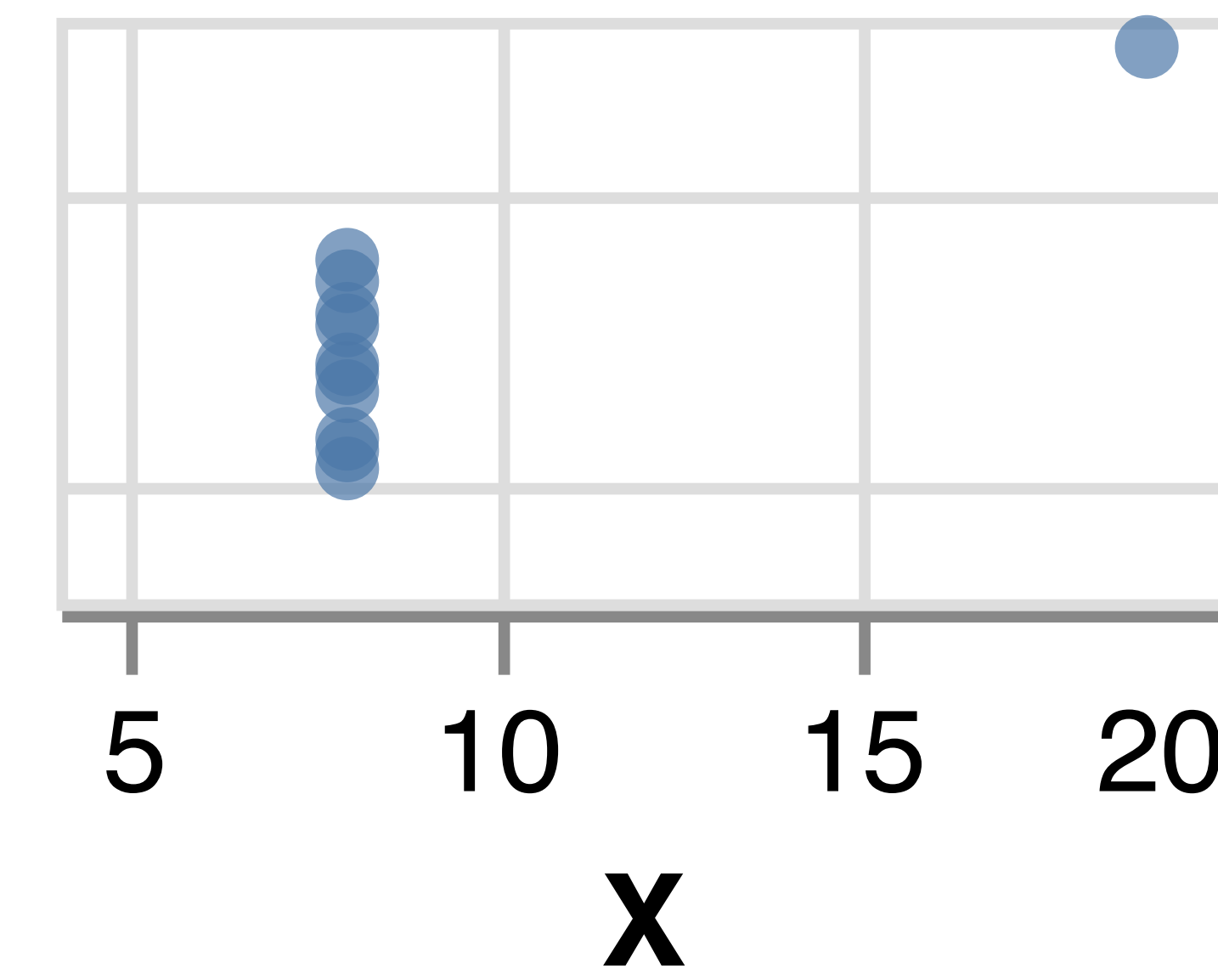
Set B



Set C



Set D

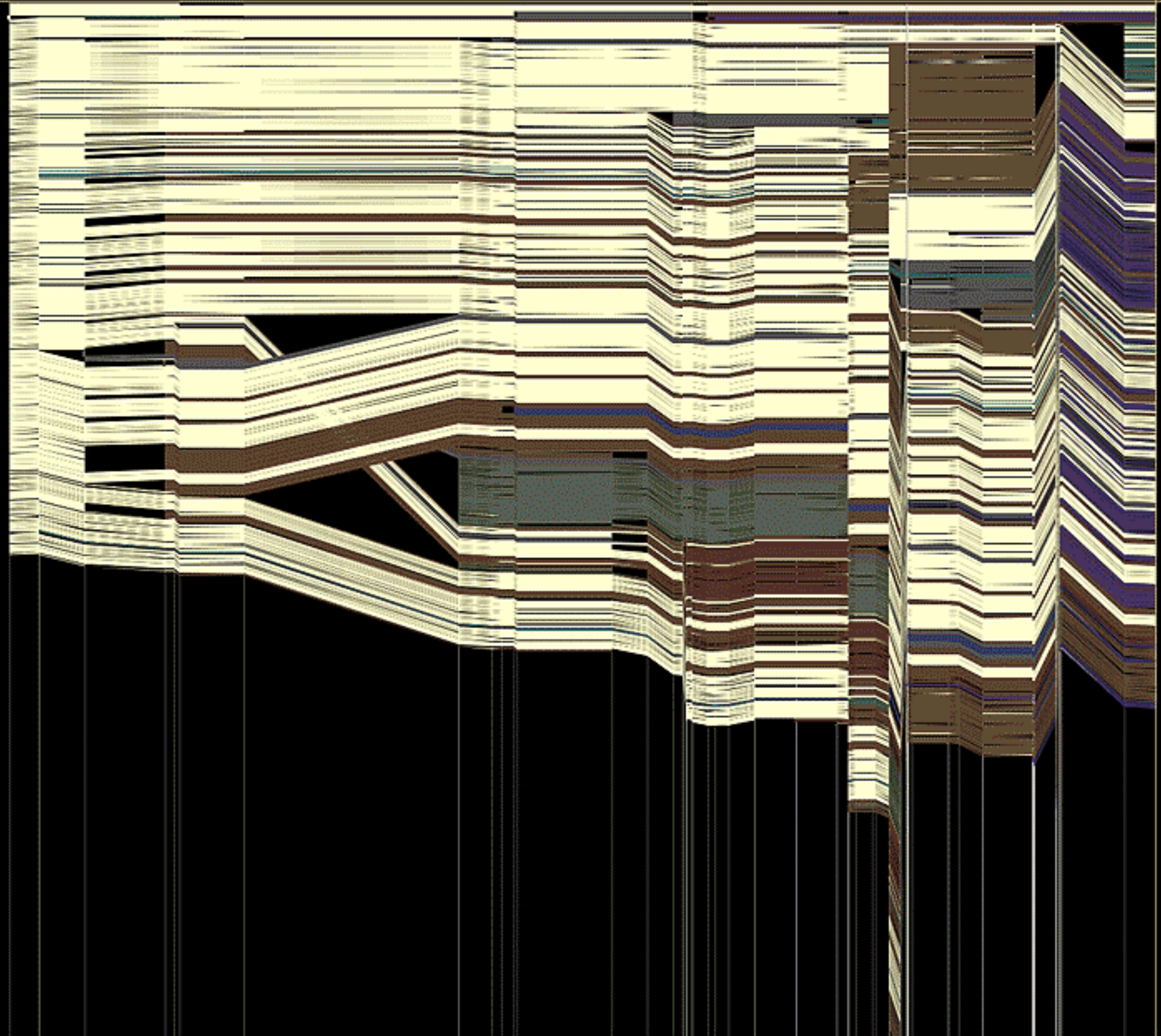


Wikipedia History Flow

Height = amount of text

Color = author

What do you notice?





Why create visualizations?

Why create visualizations?

Student Responses:

- help others understand data
- help yourself understand data
- concise
- see trends
- talk to your boss
- make it easier to clean data

The Value of Visualization

Record information

Blueprints, photographs, seismographs, ...

Analyze data to support reasoning (**exploratory visualization**)

Develop and assess hypotheses

Find patterns / Discover errors in data

Expand memory

Communicate information to others (**explanatory visualization**)

Share and persuade

Collaborate and revise

The Value of Visualization

Record information

Blueprints, photographs, seismographs, ...

Analyze data to support reasoning (exploratory visualization)

Develop and assess hypotheses

Find patterns / Discover errors in data

Expand memory

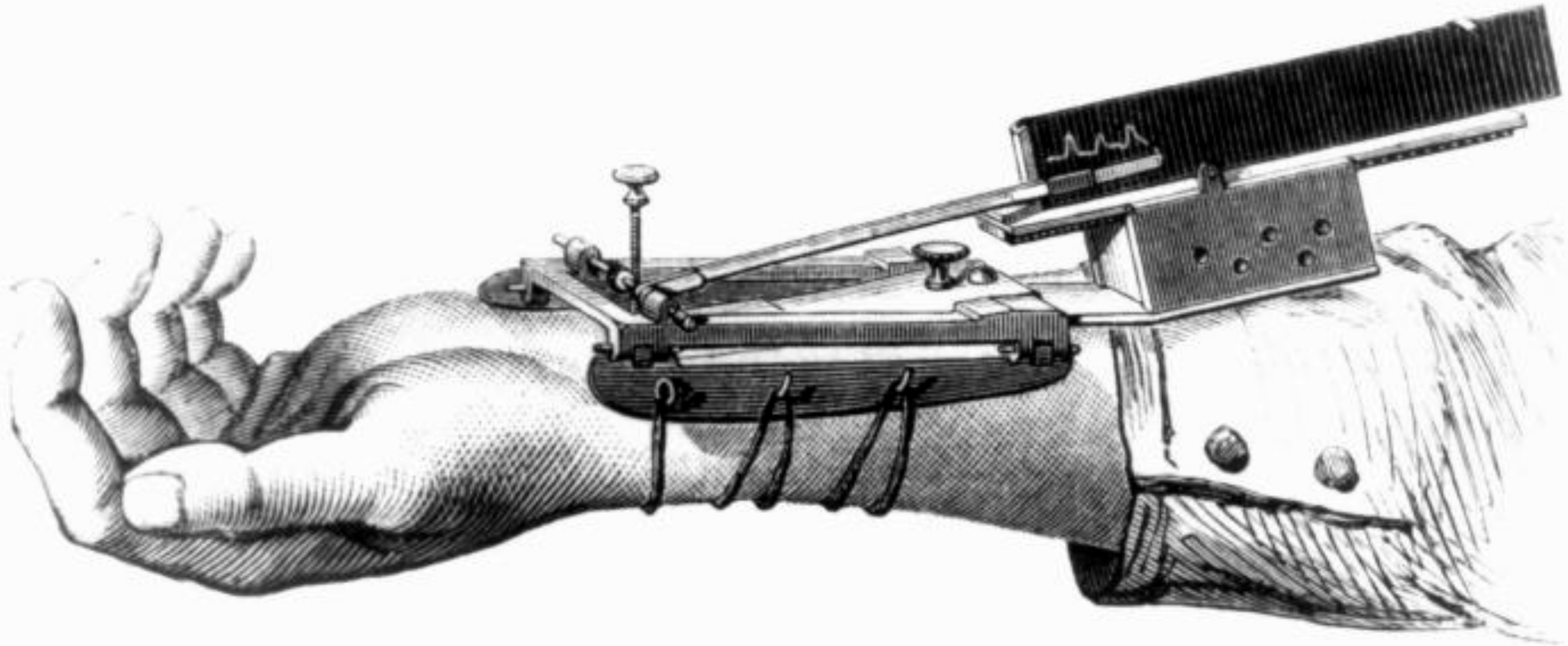
Communicate information to others (explanatory visualization)

Share and persuade

Collaborate and revise

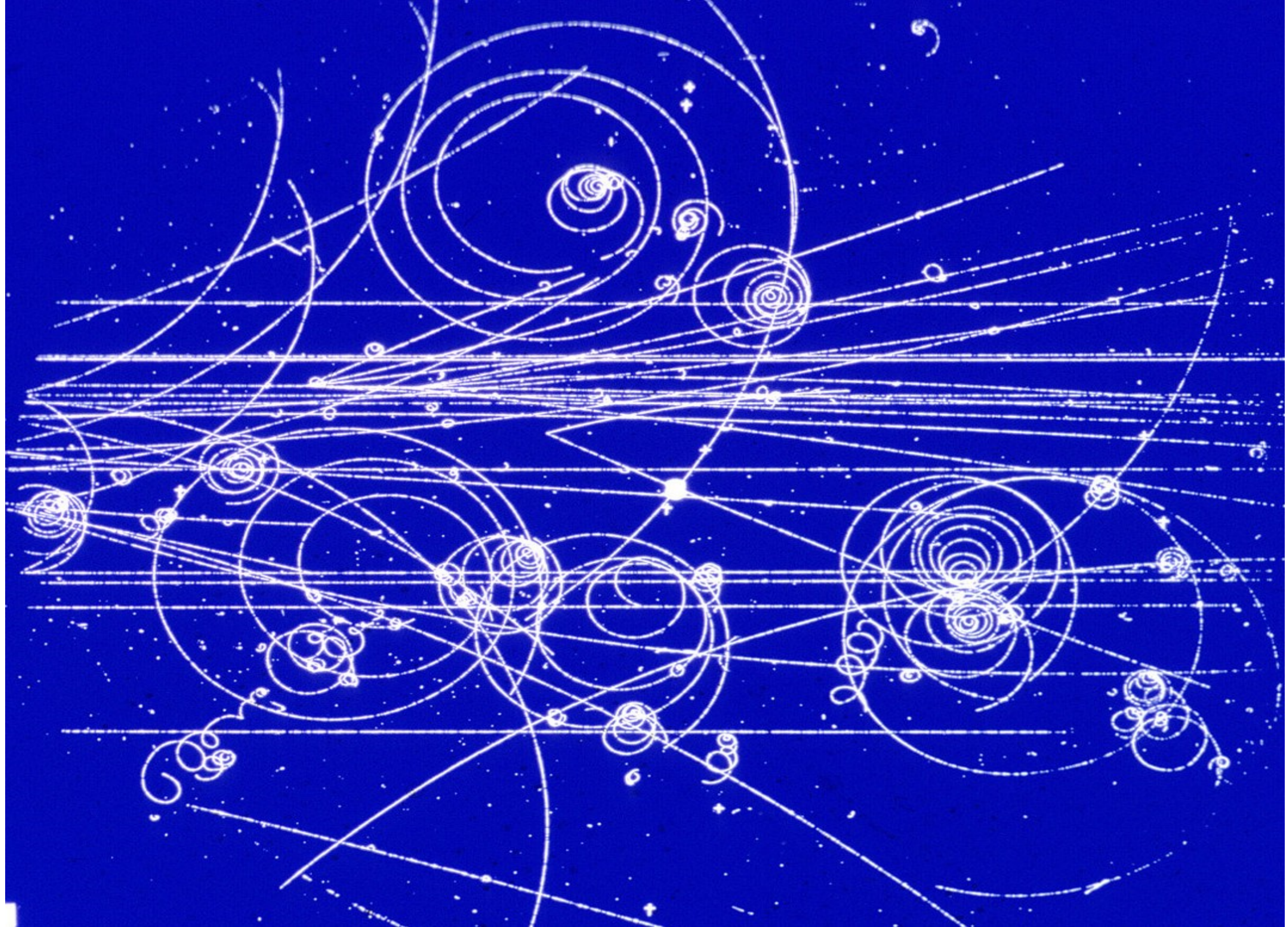


Gallop, Bay Horse "Daisy" [Muybridge]

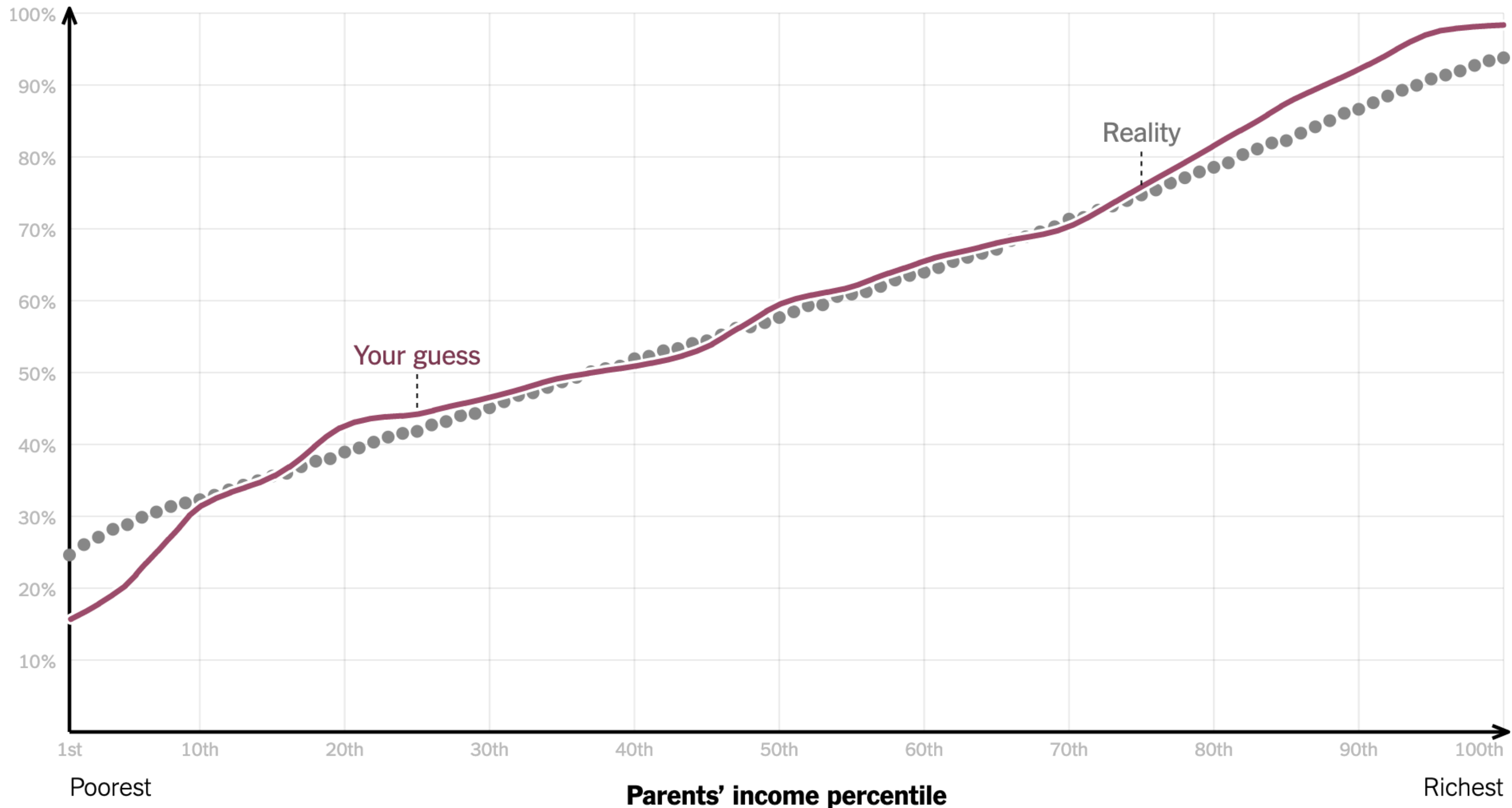


1.
Marey's **sphygmograph** in use,
1860. *La méthode graphique dans
les sciences expérimentales et
principalement en physiologie et en
médecine.*

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

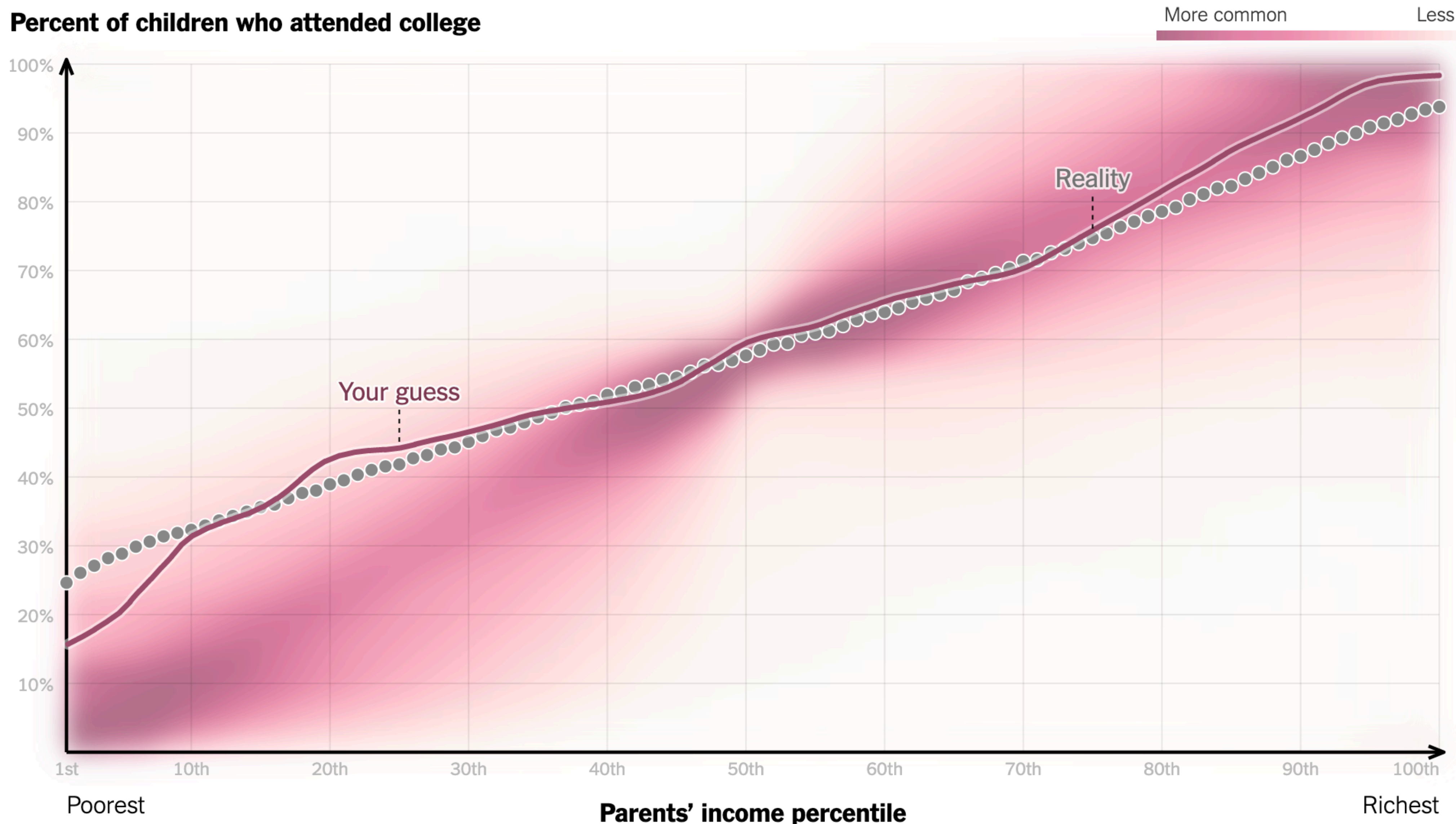


Percent of children who attended college



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]

The Value of Visualization

Record information

Blueprints, photographs, seismographs, ...

Analyze data to support reasoning (**exploratory visualization**)

Develop and assess hypotheses

Find patterns / Discover errors in data

Expand memory

Communicate information to others (**explanatory visualization**)

Share and persuade

Collaborate and revise

The Value of Visualization

Record information

Blueprints, photographs, seismographs, ...

Analyze data to support reasoning (**exploratory visualization**)

Develop and assess hypotheses

Find patterns / Discover errors in data

Expand memory

Communicate information to others (**explanatory visualization**)

Share and persuade

Collaborate and revise





© AP

© AP

HISTORY OF O-RING DAMAGE ON SRM FIELD JOINTS

1169
OCT 30, 1985
y
AFT

	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

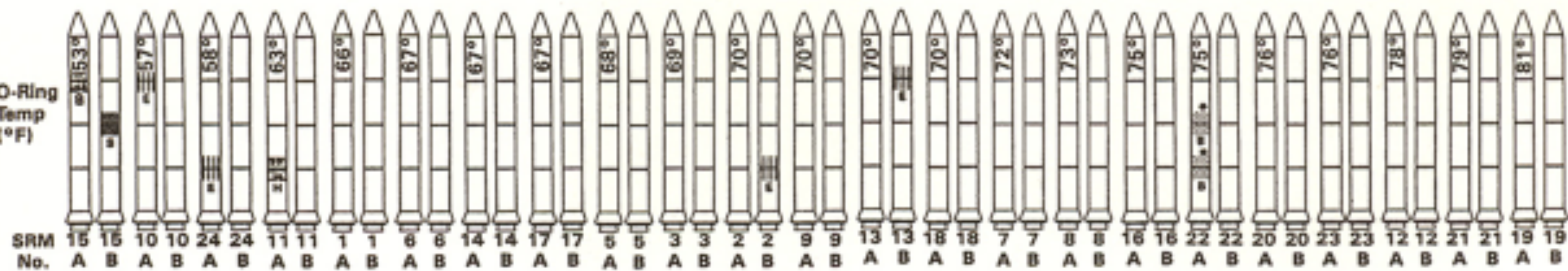
o NOZZLE BLOW-BY

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

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

O-Ring
Temp
(°F)



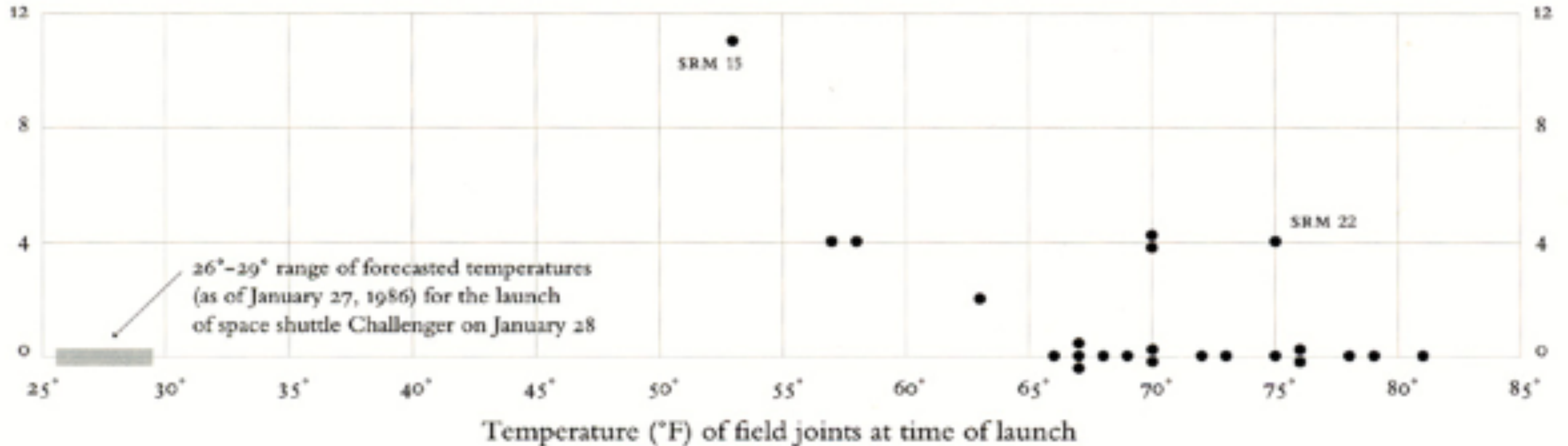
* No Erosion

O-Ring
Temp
(°F)



Chart of temperatures vs. O-ring damage [Tuftte 97]

O-ring damage index, each launch



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

Cholera Outbreak (remember DSC 10?)

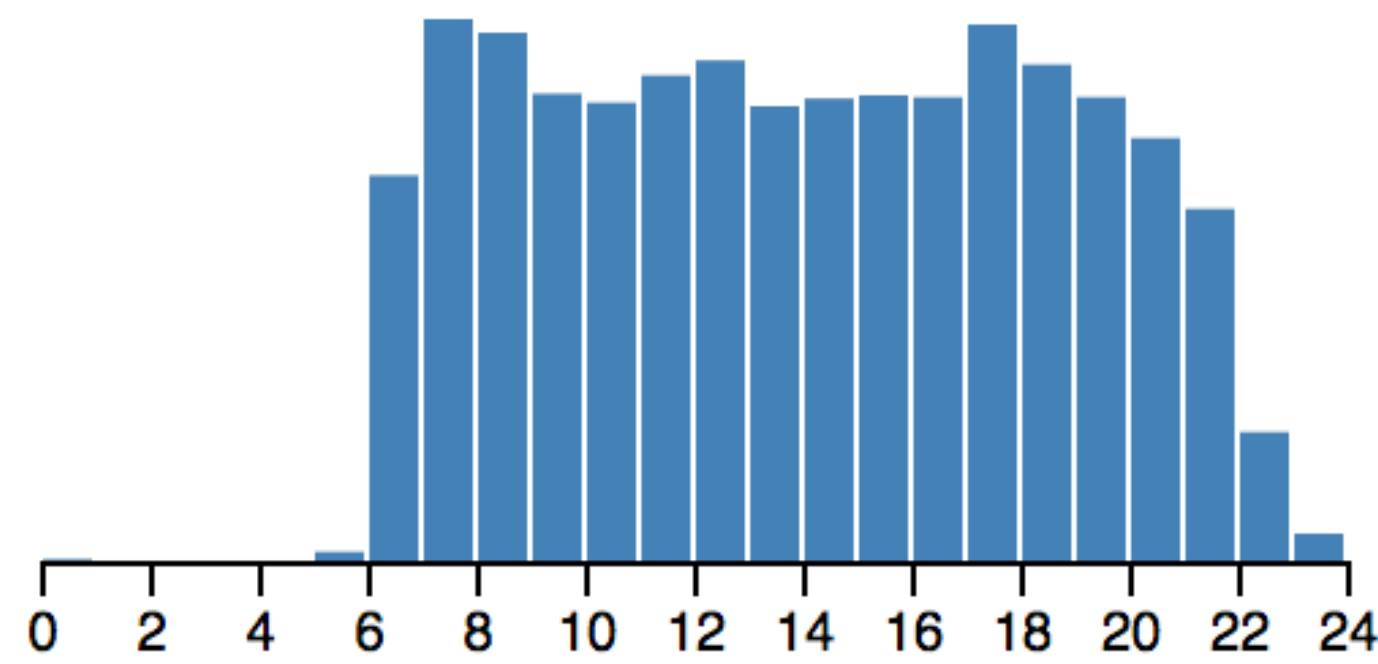


Cholera Outbreak (remember DSC 10?)

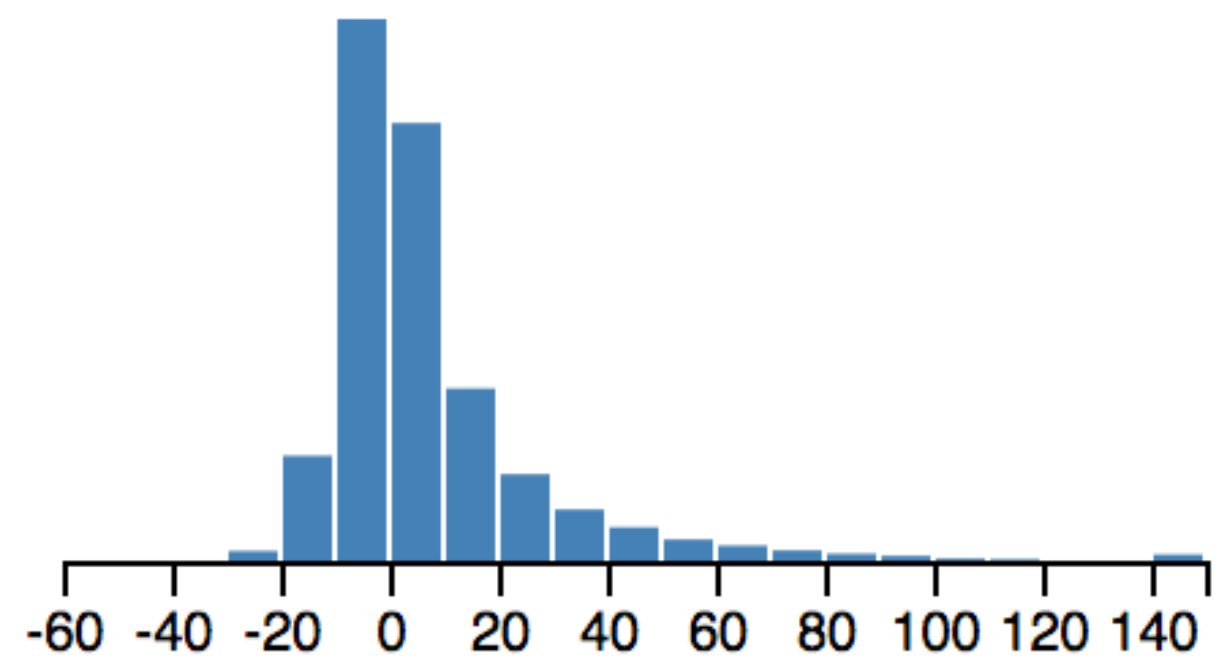


<https://square.github.io/crossfilter/>

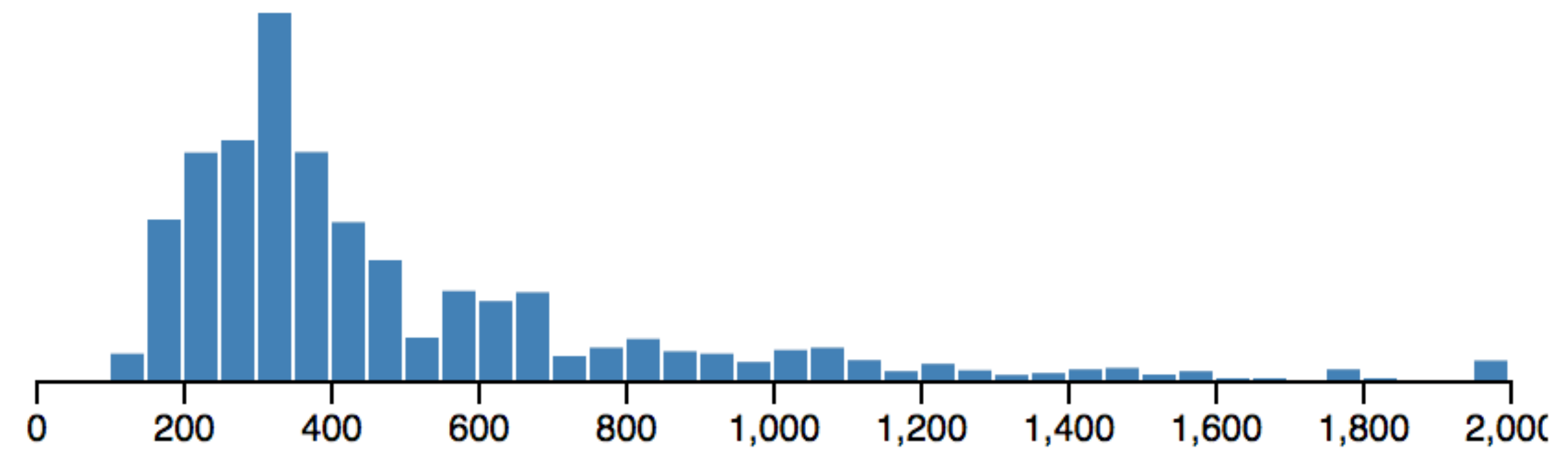
Time of Day



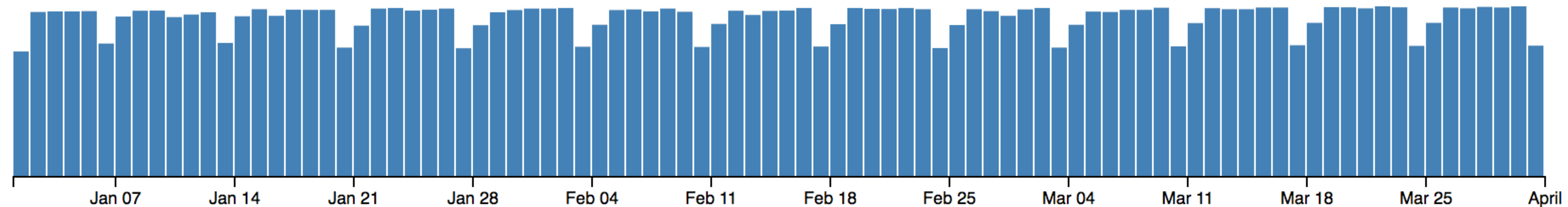
Arrival Delay (min.)



Distance (mi.)



Date



What insights do you notice?

The Value of Visualization

Record information

Blueprints, photographs, seismographs, ...

Analyze data to support reasoning (**exploratory visualization**)

Develop and assess hypotheses

Find patterns / Discover errors in data

Expand memory

Communicate information to others (**explanatory visualization**)

Share and persuade

Collaborate and revise

The Value of Visualization

Record information

Blueprints, photographs, seismographs, ...

Analyze data to support reasoning (**exploratory visualization**)

Develop and assess hypotheses

Find patterns / Discover errors in data

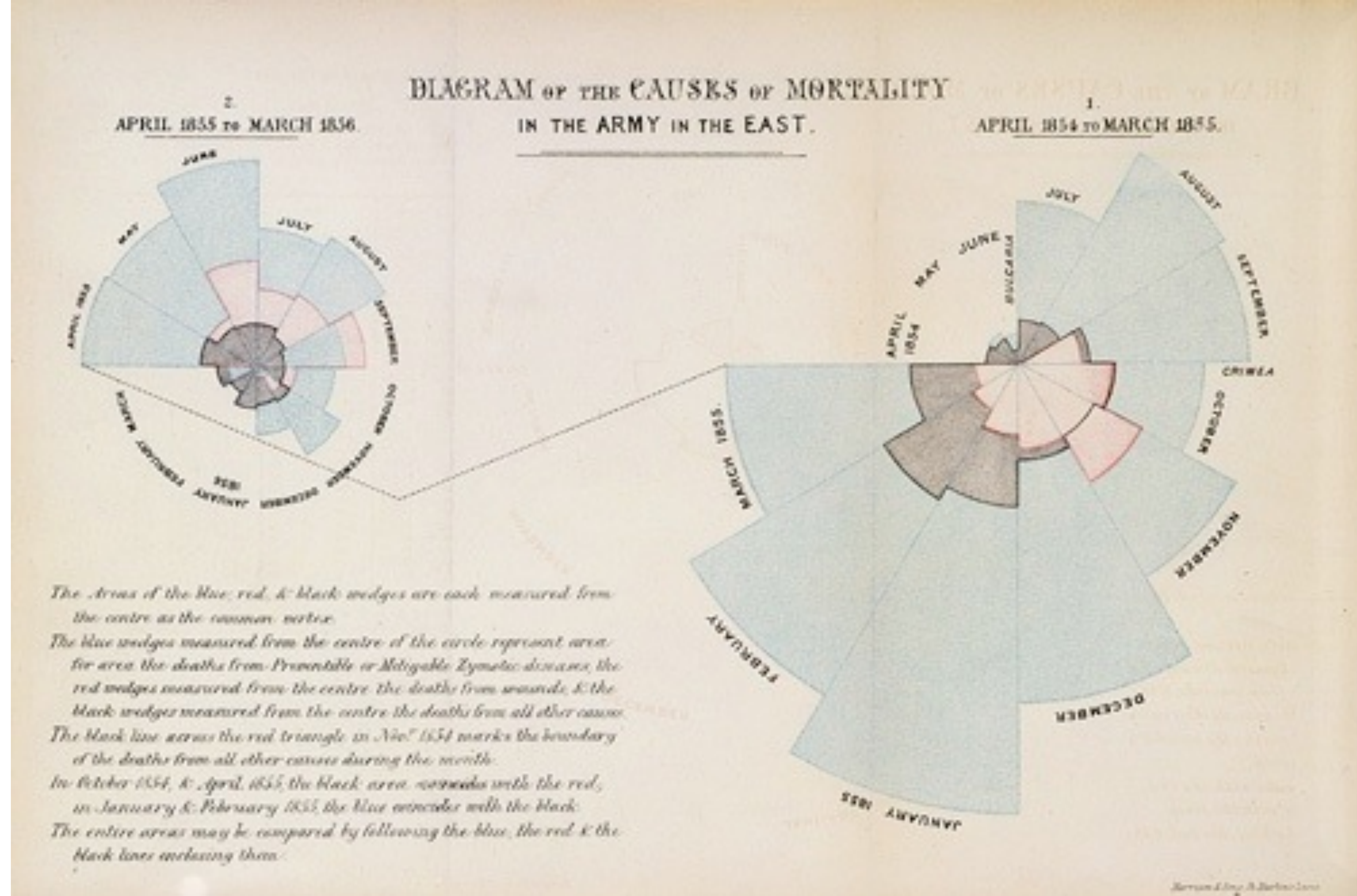
Expand memory

Communicate information to others (**explanatory visualization**)

Share and persuade

Collaborate and revise

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



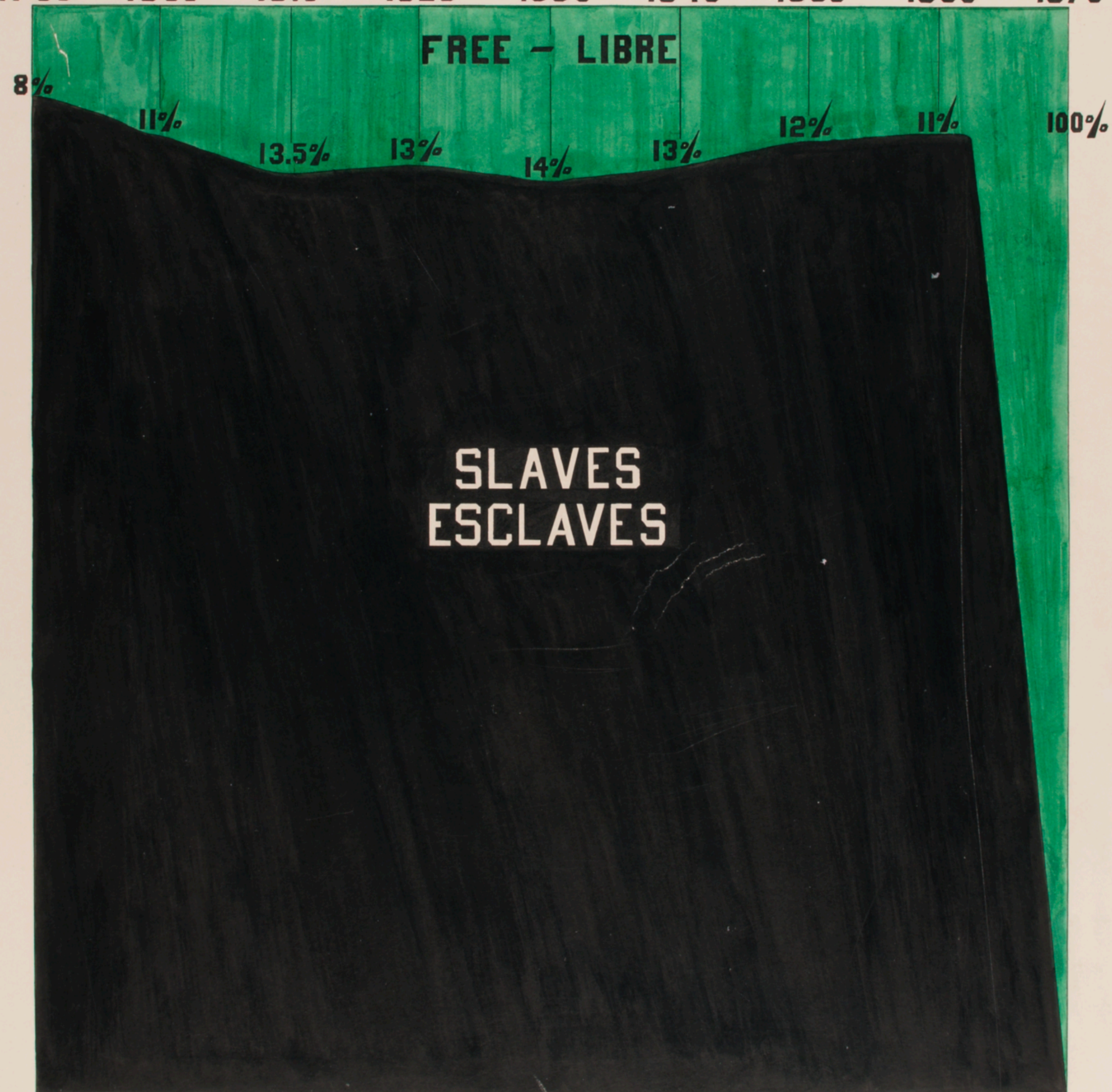
1856 “Coxcomb” of Crimean War Deaths, Florence Nightingale

PROPORTION OF FREEMEN AND SLAVES AMONG AMERICAN NEGROES .

PROPORTION DES NÈGRES LIBRES ET DES ESCLAVES EN AMÉRIQUE .

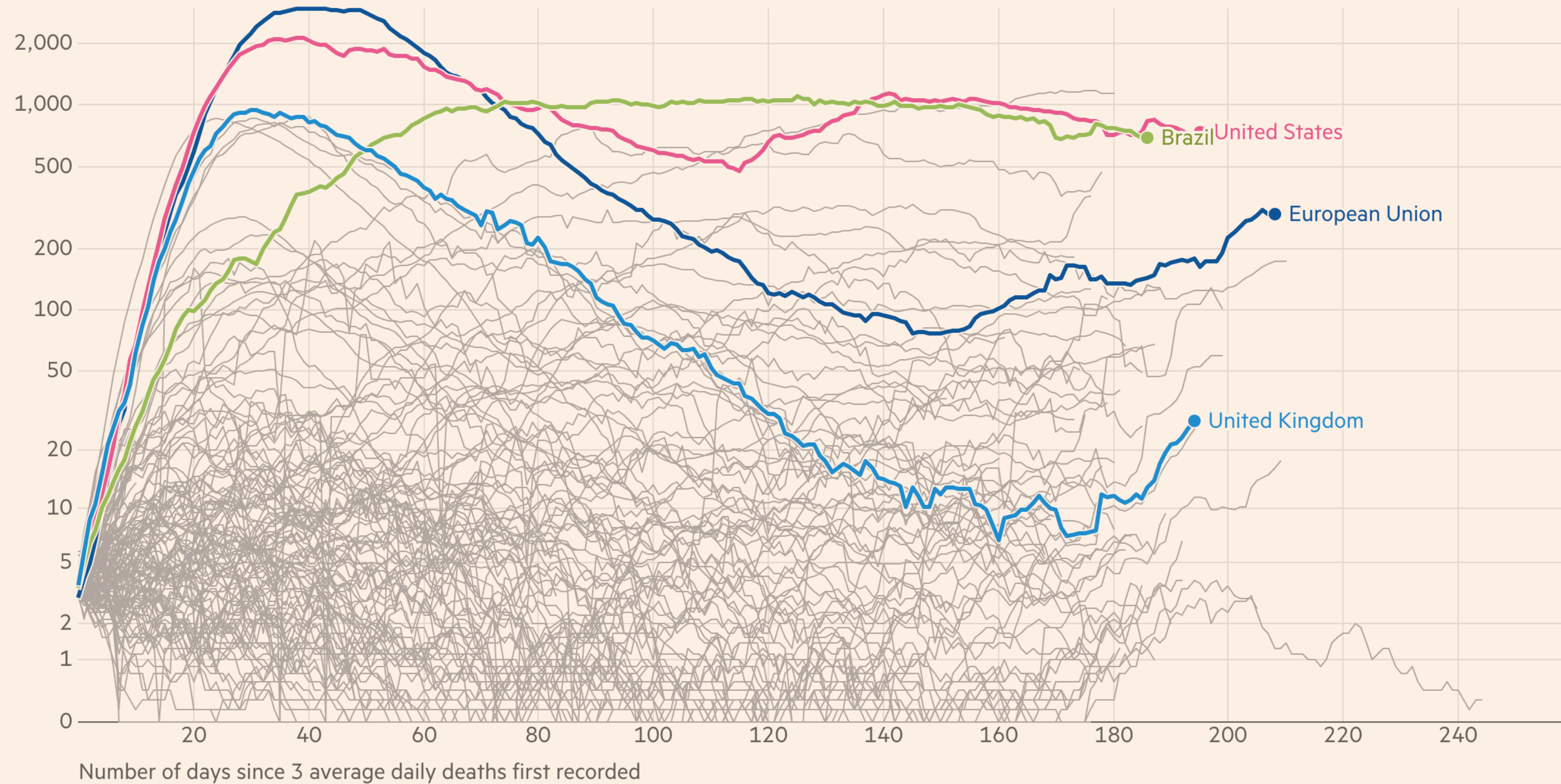
DONE BY ATLANTA UNIVERSITY .

1790 1800 1810 1820 1830 1840 1850 1860 1870



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://www.ft.com/covid19)

FINANCIAL TIMES

Coronavirus Tracked John Burn-Murdoch & Financial Times

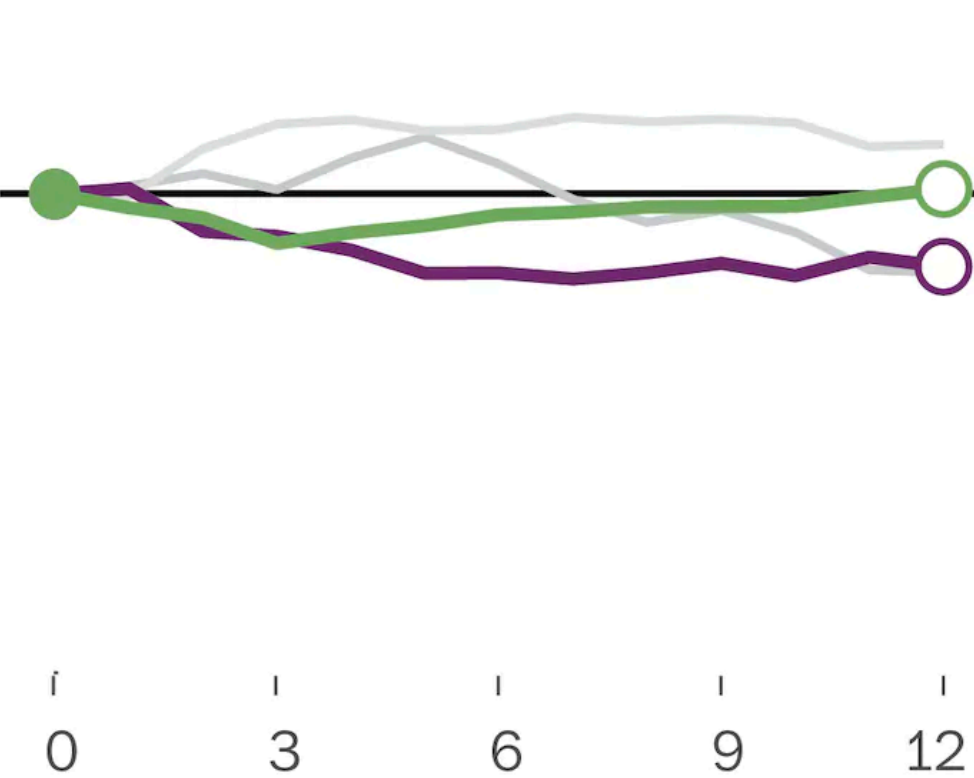
The coronavirus crisis is different

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

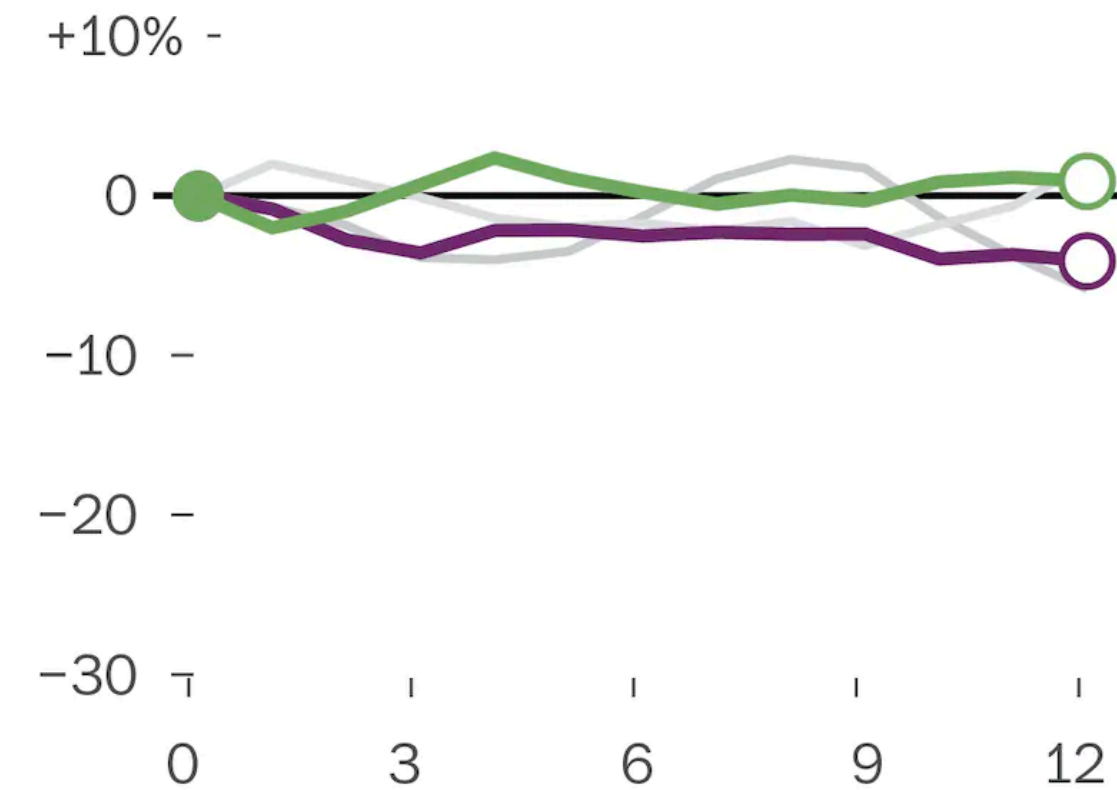
1990 recession



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 (**exploratory visualization**)

Develop and assess hypotheses

Find patterns / Discover errors in data

Expand memory

Communicate information to others (**explanatory visualization**)

Share and persuade

Collaborate and revise

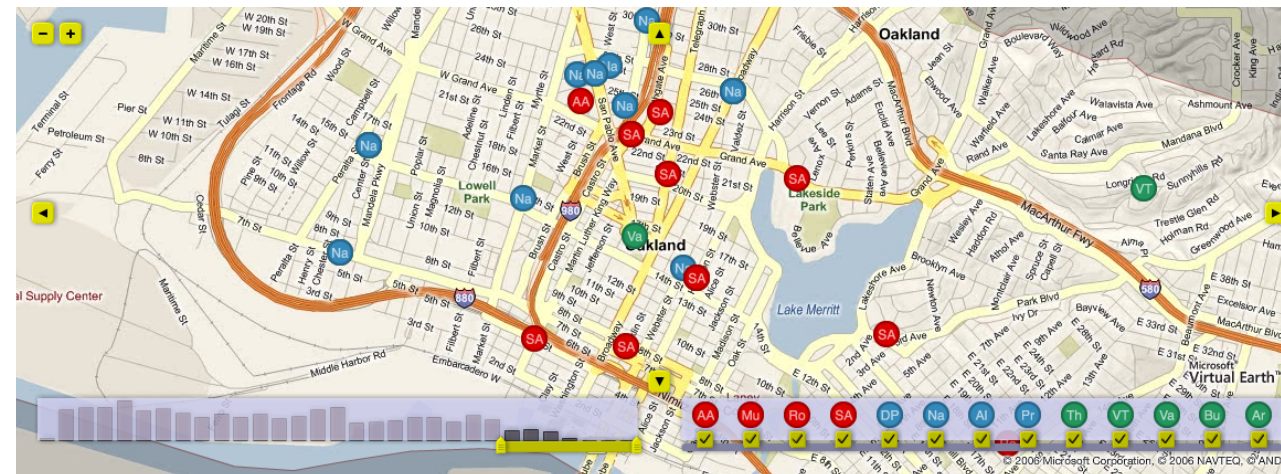
About this Course

Principles

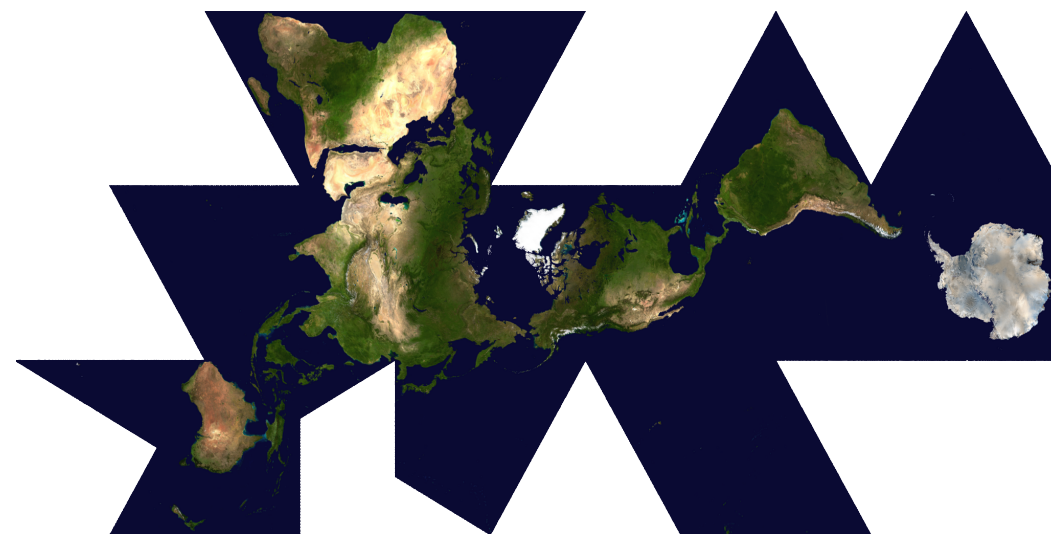
Data and Image Models

LES VARIABLES DE L'IMAGE				12	14
XY	POINTS	LIGNES	ZONES		
2 DIMENSIONS DU PLAN	x x x	∕ ∕ ∕	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	OQ	≠
Z	TAILLE	∕ ∕ ∕	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	OQ	≠
	VALEUR	∕ ∕ ∕	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	O	≠
LES VARIABLES DE SÉPARATION DES IMAGES				13	
	GRAIN	∕ ∕ ∕	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	O	≠
	COULEUR	∕ ∕ ∕	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	≠	≠
	ORIENTATION	∕ ∕ ∕	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100	≠	≠

Interaction



Maps



...and many more!

Techniques

HTML/CSS



JavaScript



D3.js



Svelte



Learning Objectives

By the end of this course:

- Understand and apply key visualization techniques and theory.
- Design, evaluate, and critique visualization designs.
- Implement interactive data visualizations for the web using D3.js.
- Develop a substantial visualization project.

Grade Breakdown

Component	Weight
Participation	8%
Labs	12%
Project 1	10%
Project 2	15%
Project 3	15%
Final Project	40%

Grade Breakdown

Component	Weight	
Participation	8%	<p>1% per week (2 lowest weeks dropped). 3 options:</p> <ol style="list-style-type: none">1. Attend both lectures and participate in the lecture activities.2. Share and critique 1 viz example on Ed.3. Respond to 2 viz examples on Ed. <p>See website for full details.</p>
Labs	12%	
Project 1	10%	
Project 2	15%	
Project 3	15%	
Final Project	40%	

Grade Breakdown

Component	Weight	
Participation	8%	7 labs, 2% per lab, 1 lowest dropped.
Labs	12%	Labs are "worked examples"; solutions will be in the lab itself.
Project 1	10%	Must get checked off each week by a TA during their office hours.
Project 2	15%	
Project 3	15%	
Final Project	40%	

Grade Breakdown

Component	Weight	
Participation	8%	
Labs	12%	
Project 1	10%	3 solo open-ended projects
Project 2	15%	
Project 3	15%	
Final Project	40%	

Grade Breakdown

Component	Weight
Participation	8%
Labs	12%
Project 1	10%
Project 2	15%
Project 3	15%
Final Project	40%

Final project will span last 4 weeks of course

Grade Breakdown

Component	Weight
Participation	8%
Labs	12%
Project 1	10%
Project 2	15%
Project 3	15%
Final Project	40%

6 slip days for quarter.

You can use 1 slip day for labs, 2 for project deadlines

Communication

I have no access to internal systems yet (this usually takes 1-2 weeks *after* the quarter starts :)

==

Can't answer email, tell you waitlist status, etc.

Email me at jaredwilber5@gmail.com in the mean time

Communication

Once I have access, we will use EdStem for communication

Email Giorgia, cc me for private questions related to course

Course website will stay up-to-date (dsc106.com)

Where you're headed: Final Project

Narrative visualization project on topic of choice

Initial **prototype** and **design reviews**

In-class demonstration **video** showcase

Submit and publish online (if feasible)

Projects from similar courses (at other universities) have been:

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

USA



A concise, **travel-like-a-local** guide
to **74,762 attractions**,
according to **9,526,193 reviews**

by [Ilia Blinderman](#)

Lab 1: Altair

Lab 1 released, due Friday.

Plotting in Python using the Altair plotting library

Project 1: Expository visualization

Create **one static visualization** for a dataset (see course website).

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 (email to TA; see Project 1 page)

Image of your visualization (PNG or JPG format)

Short description + design rationale (≤ 4 paragraphs)