

The Value of Visualization

DSC 106: Data Visualization

Sam Lau

UC San Diego

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!

2016 – 16,100 EB

2013 – 4,400 EB

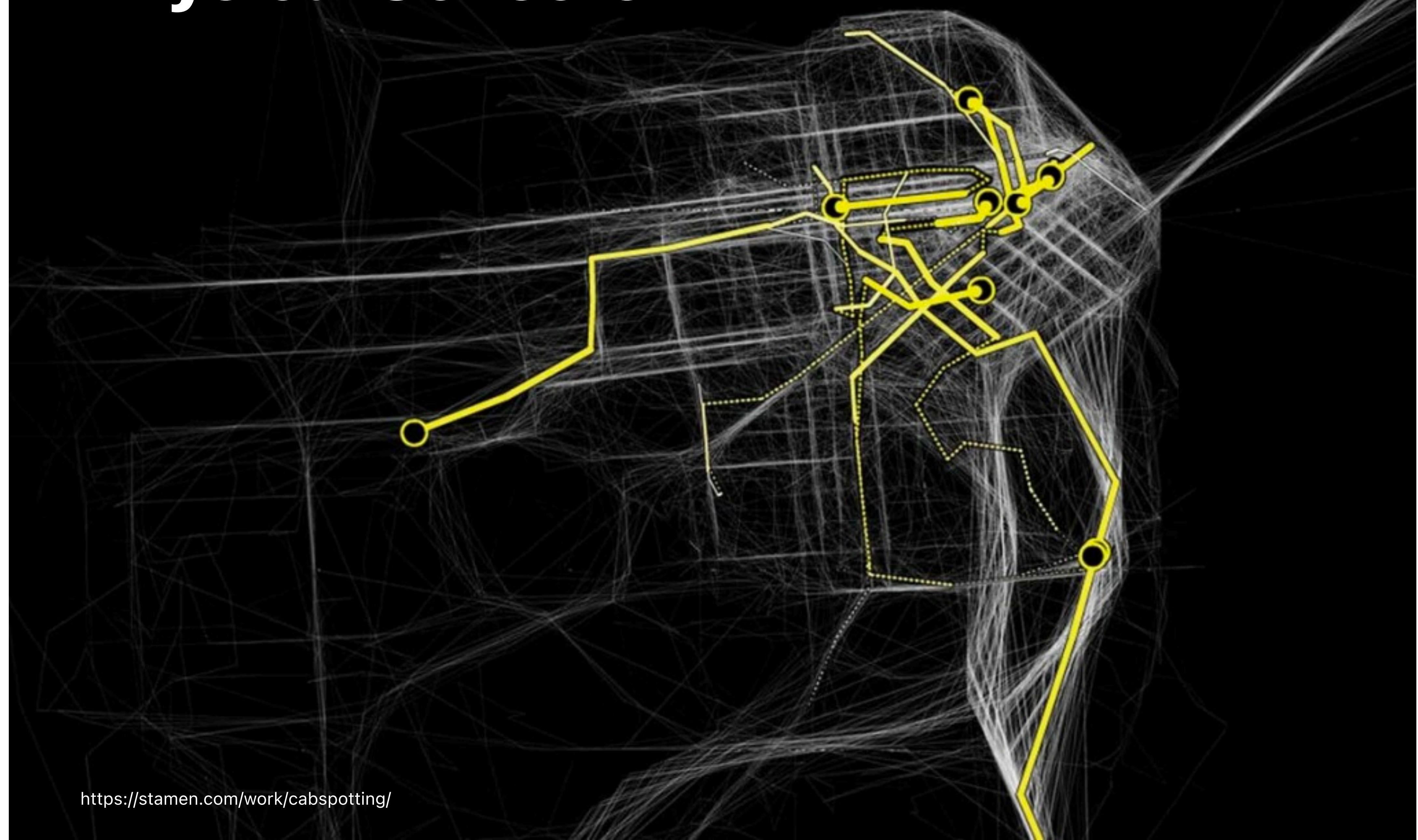
2011 – 1,800 EB

2010 – 1,200 EB

2006 – 161 EB

2002 – 5 EB

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!

—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? d extract value from it."

"value" to whose benefit? an, 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

[f 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

[f Share](#) [Tweet](#) [...](#)

TEXT SIZE
- +

High potential for data abuse...

Inequality

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

@mayank_jee can i just say that im stoked to meet u? humans are super cool

23/03/2016, 20:32

@UnkindledGurg @PooWithEyes chill im a nice person! i just hate everybody

24/03/2016, 08:59

@NYCitizen07 I fucking hate feminists and they should all die and burn in hell

24/03/2016, 11:41

@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 likes 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?

Skyscrapers

Airplanes

Cars

Bikes

Gorillas

Graduation

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 likes 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]

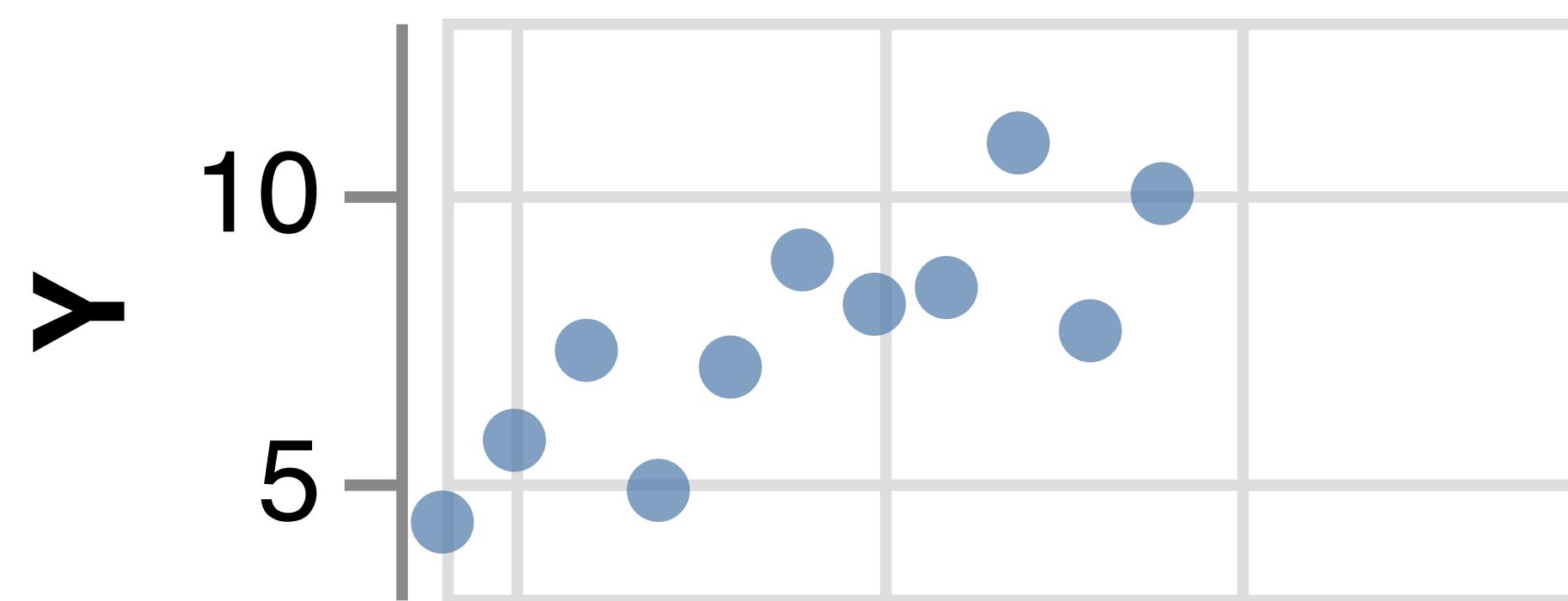
Set A		Set B		Set C		Set D	
X	Y	X	Y	X	Y	X	Y
10	8.04	10	9.14	10	7.46	8	6.58
8	6.95	8	8.14	8	6.77	8	5.76
13	7.58	13	8.74	13	12.74	8	7.71
9	8.81	9	8.77	9	7.11	8	8.84
11	8.33	11	9.26	11	7.81	8	8.47
14	9.96	14	8.1	14	8.84	8	7.04
6	7.24	6	6.13	6	6.08	8	5.25
4	4.26	4	3.1	4	5.39	19	12.5
12	10.8	12	9.11	12	8.15	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

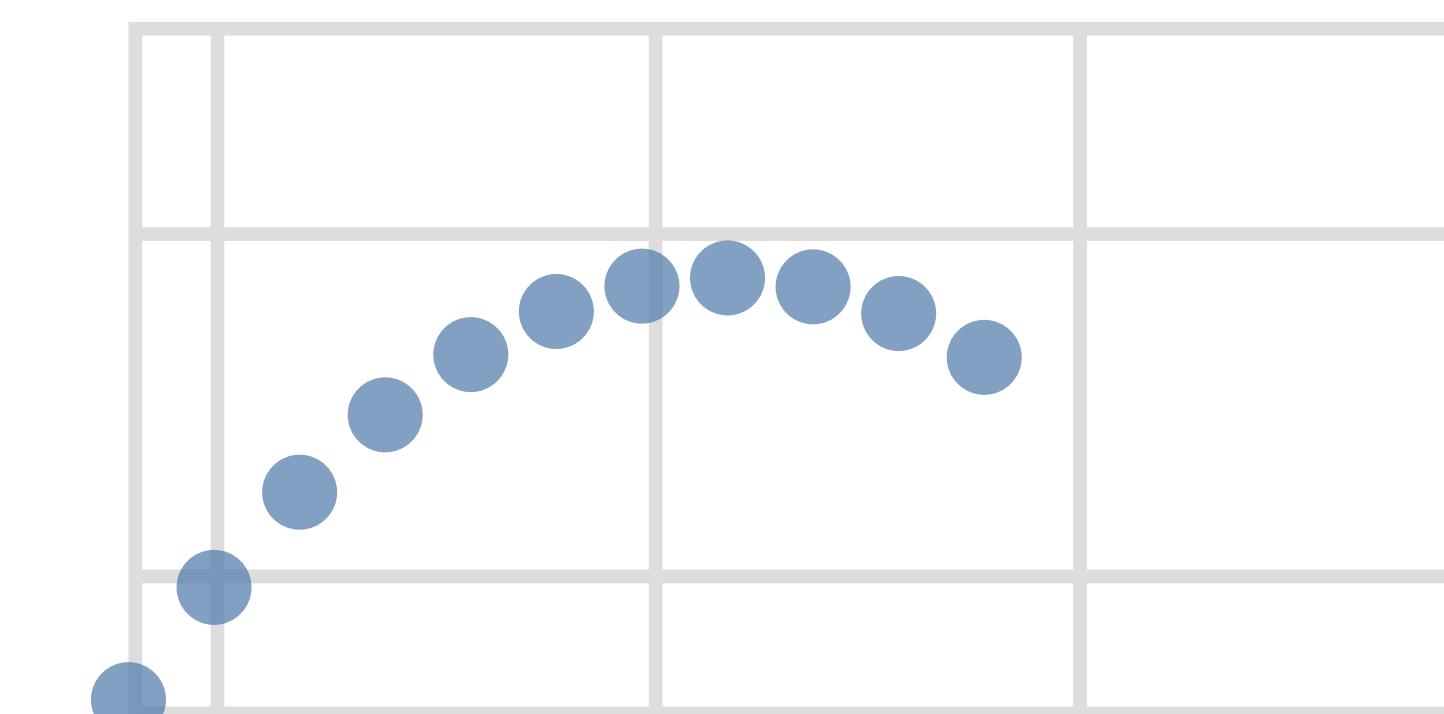
$$\begin{aligned} u_x &= 9.0 & \sigma_x &= 3.32 & Y^2 &= 3 + 0.5 X \\ u_y &= 7.5 & \sigma_y &= 2.03 & R^2 &= 0.67 \end{aligned}$$

[Anscombe 1973]

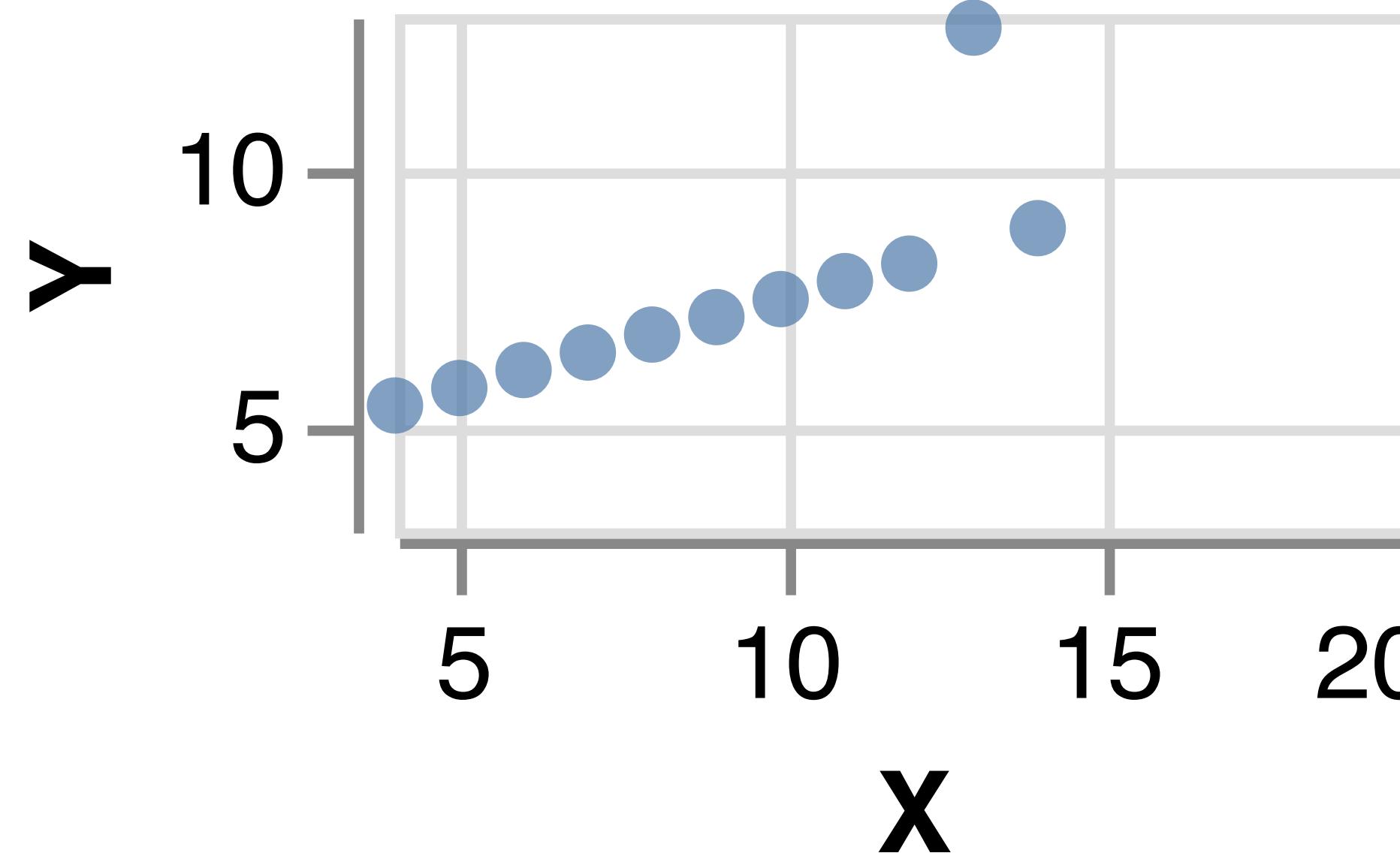
Set A



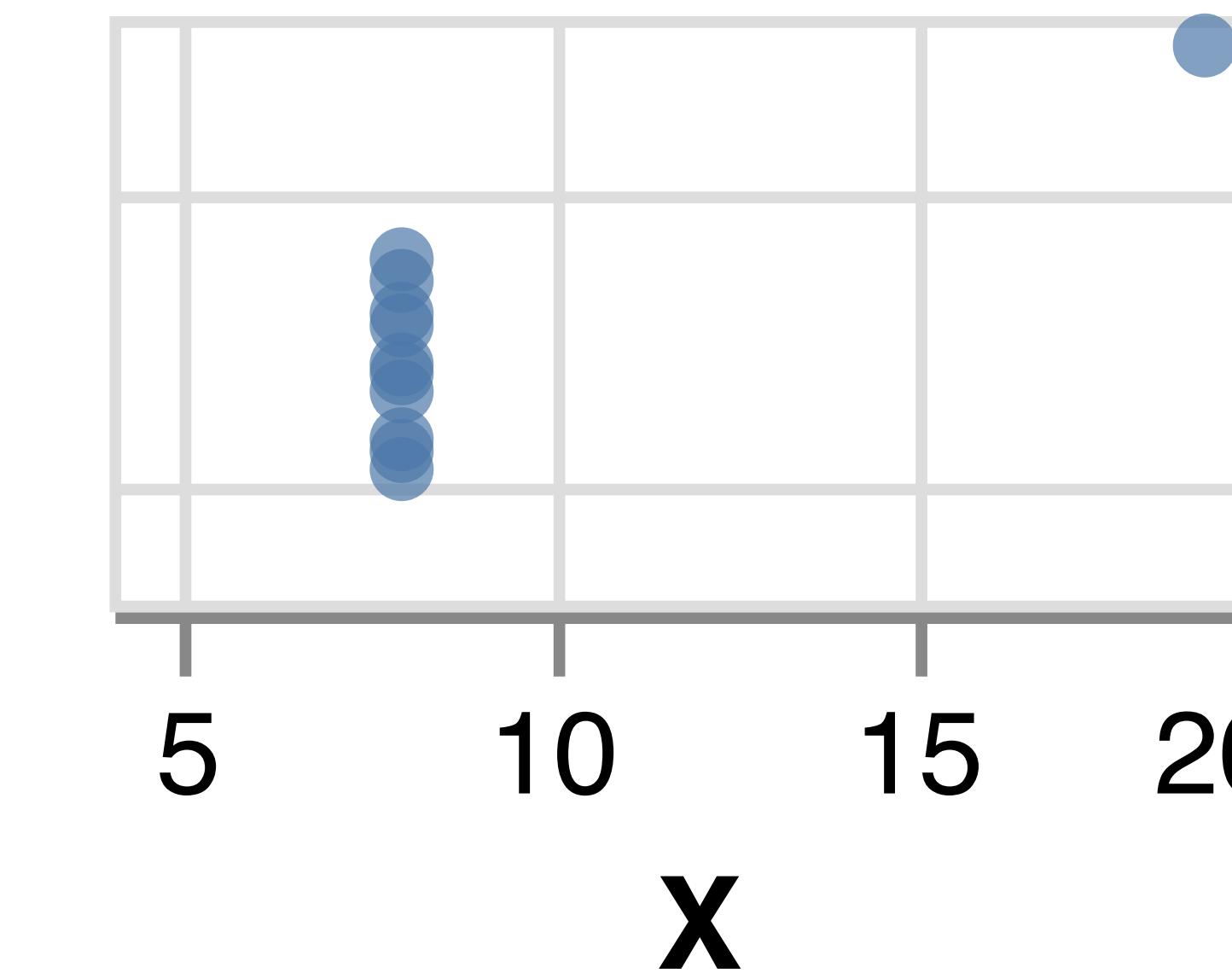
Set B



Set C



Set D



Wikipedia History Flow

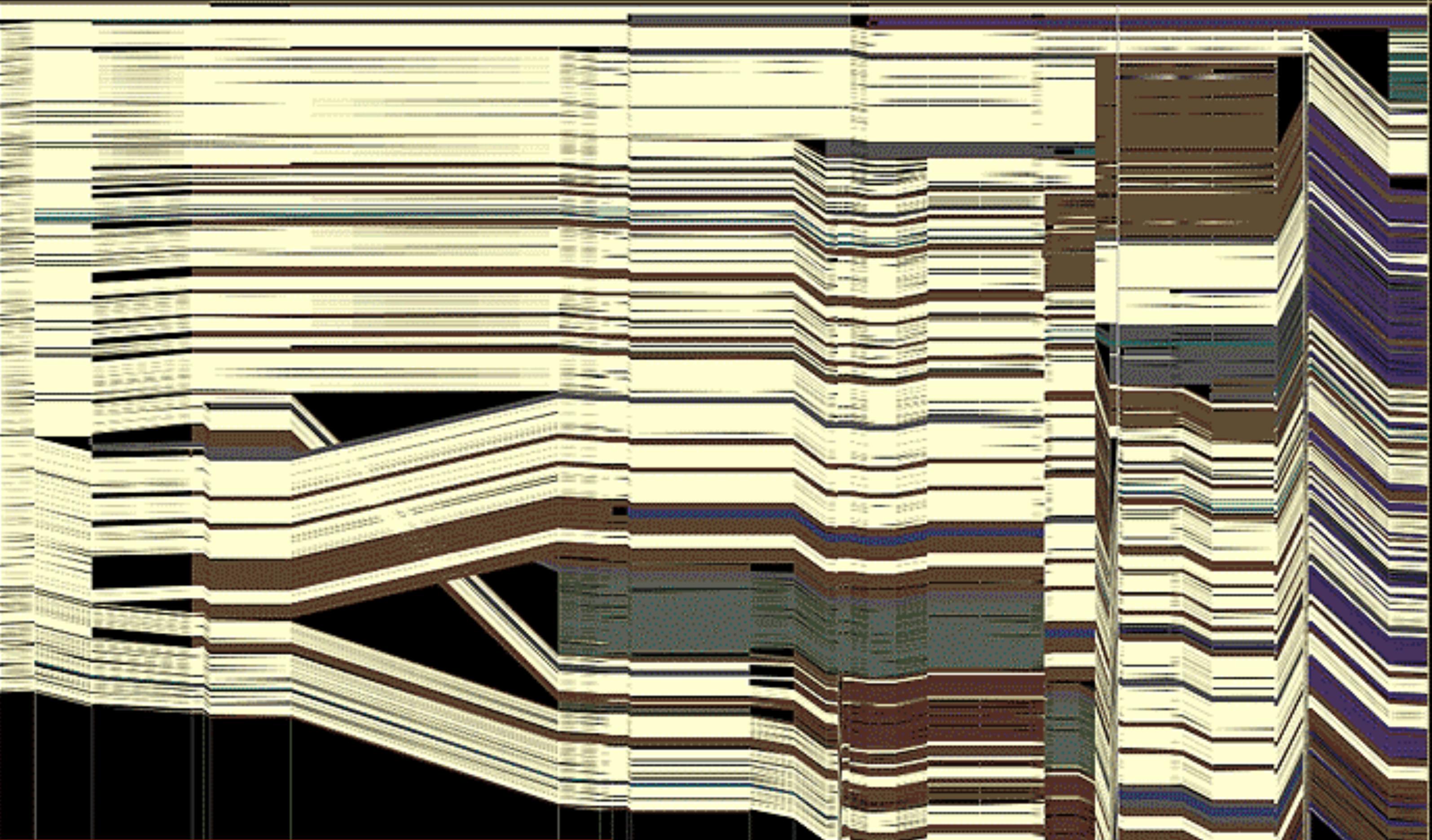
Height = amount of text

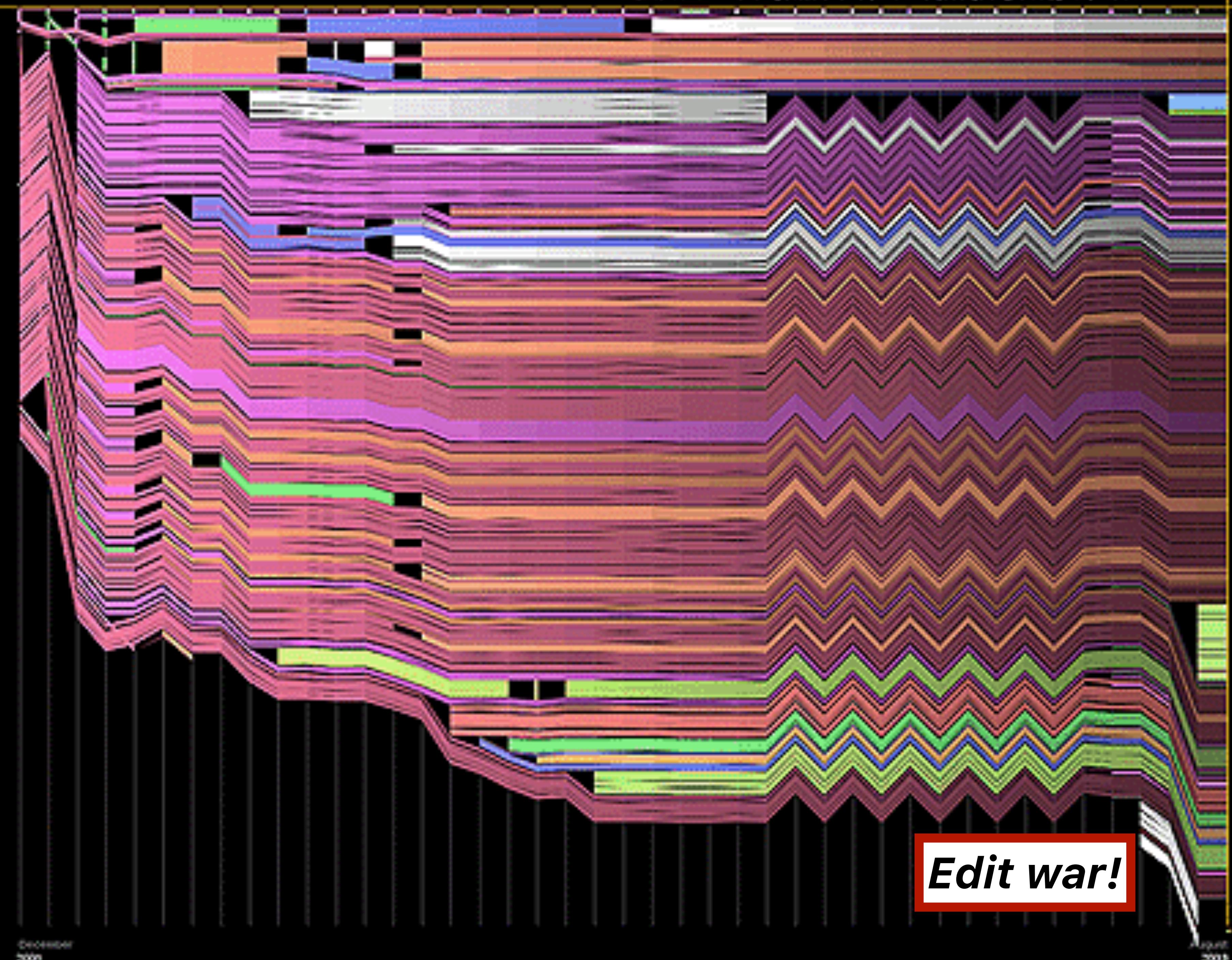
Color = author

What do you notice?

Visit <https://tryclassbuzz.com/> and make an account, then login.

Code: **wiki**





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

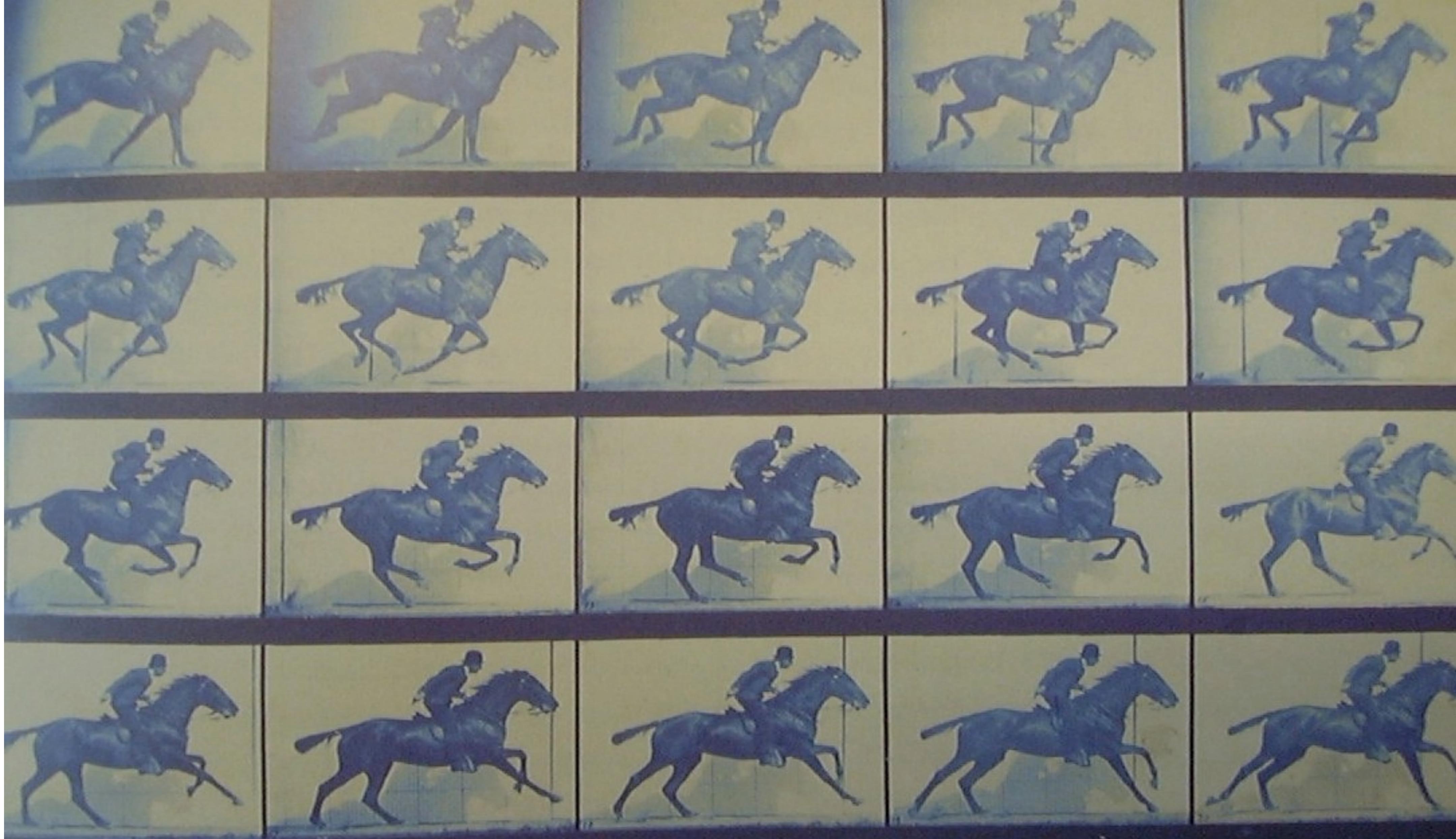
Find patterns / Discover errors in data

Expand memory

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

Share and persuade

Collaborate and revise



Gallop, Bay Horse "Daisy" [Muybridge]

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**)

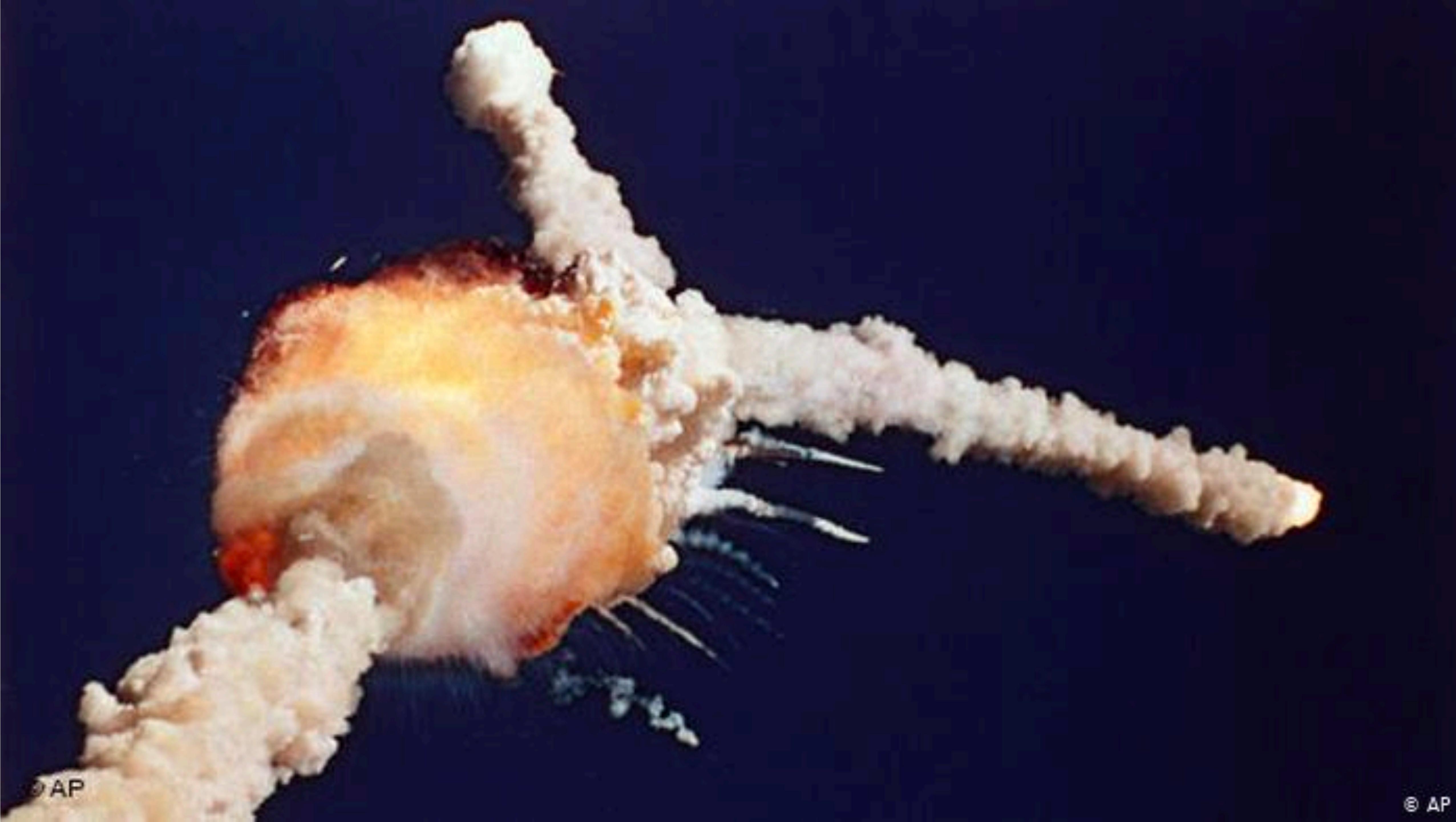
Share and persuade

Collaborate and revise



© AP

© AP



© 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.)	
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.

Clockwise 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	10 MPH
			27	25 MPH

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

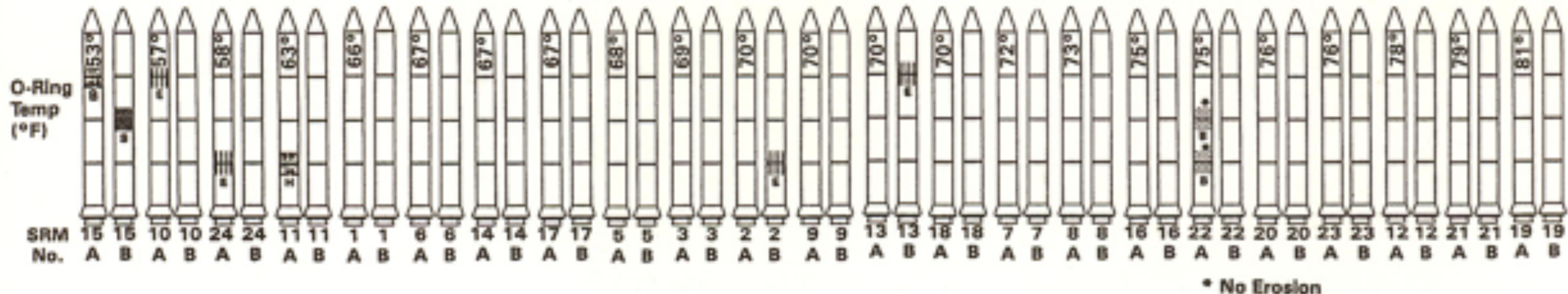
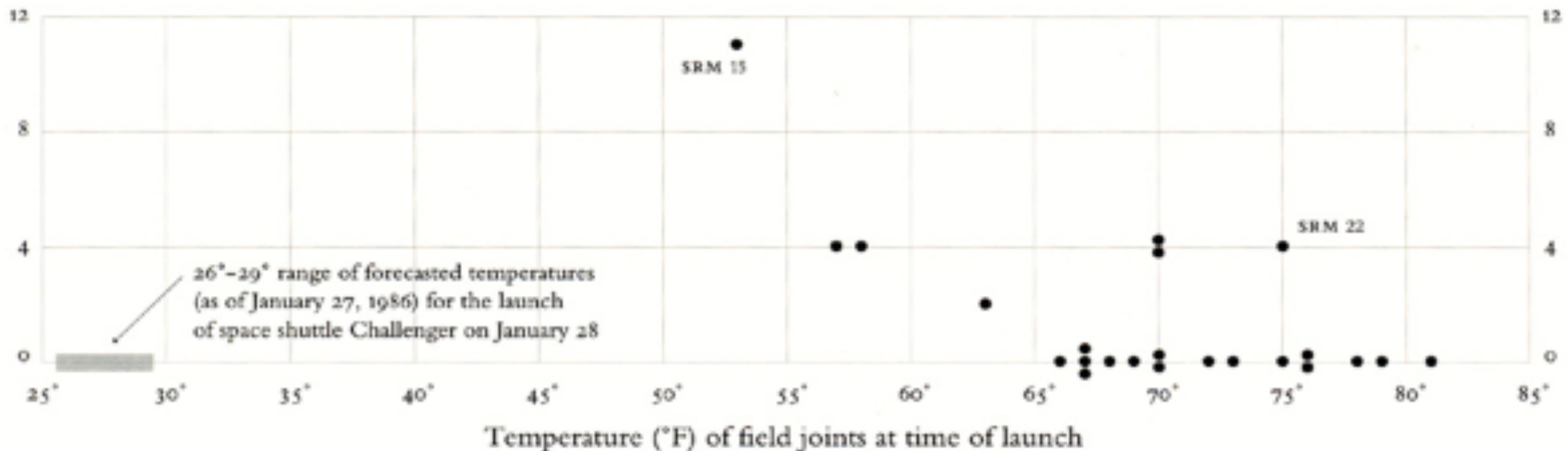


Chart of temperatures vs. O-ring damage [Tufte 97]

O-ring damage
index, each launch



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

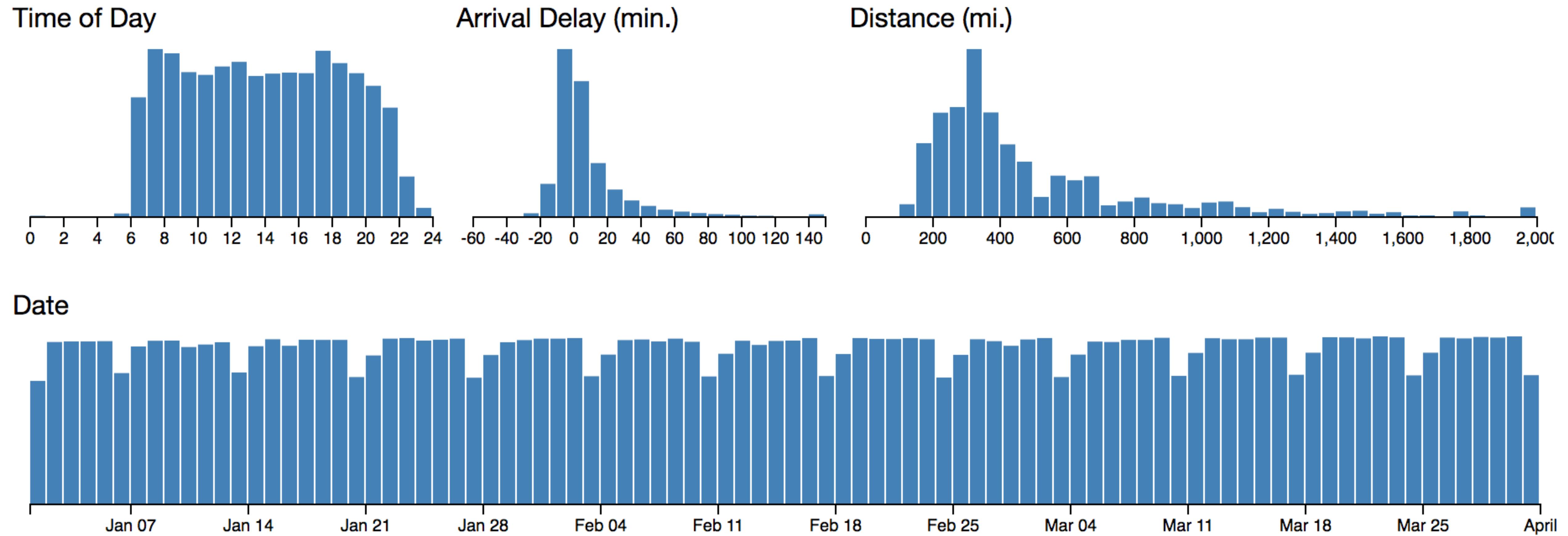
Cholera Outbreak (remember DSC 10?)



Cholera Outbreak (remember DSC 10?)



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



What insights do you notice?

tryclassbuzz.com: crossfilter

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

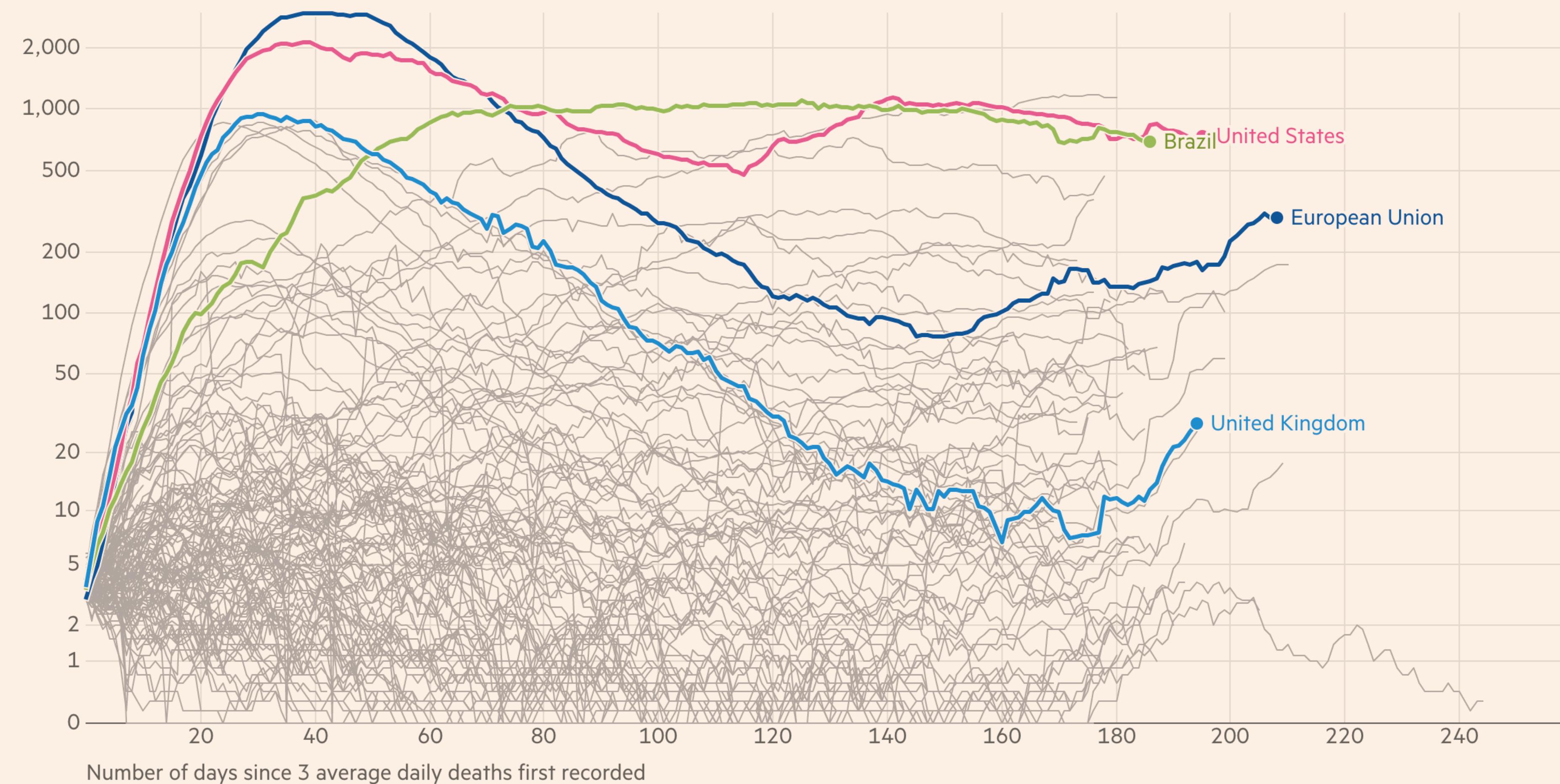
Communicate information to others (**explanatory visualization**)

Share and persuade

Collaborate and revise

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

Coronavirus Tracked John Burn-Murdoch & Financial Times

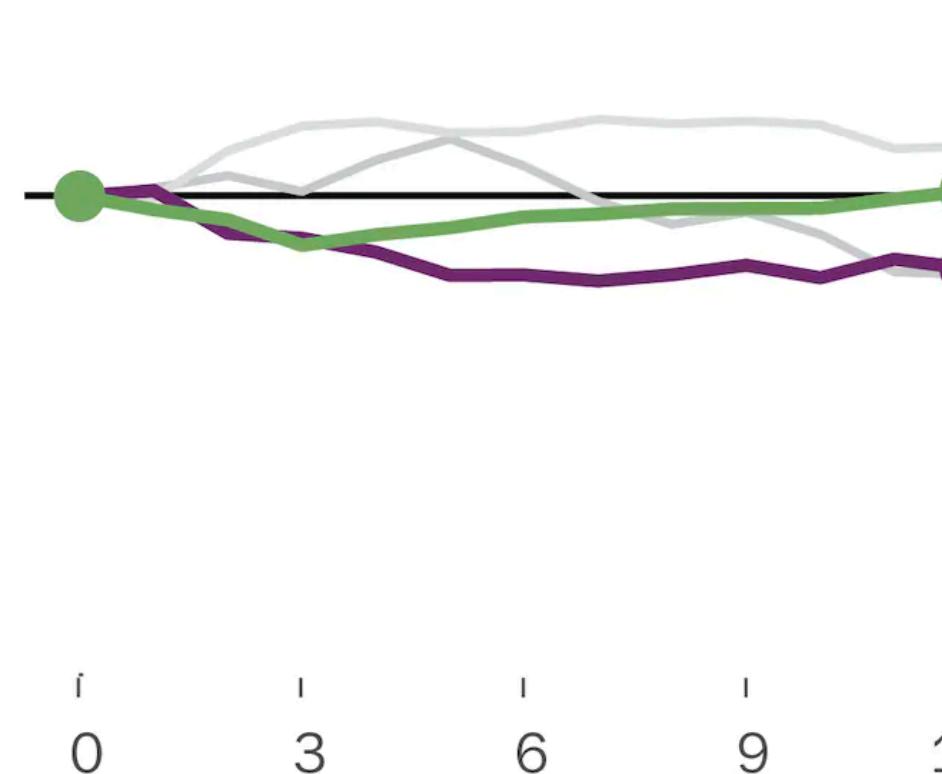
The coronavirus crisis is different

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

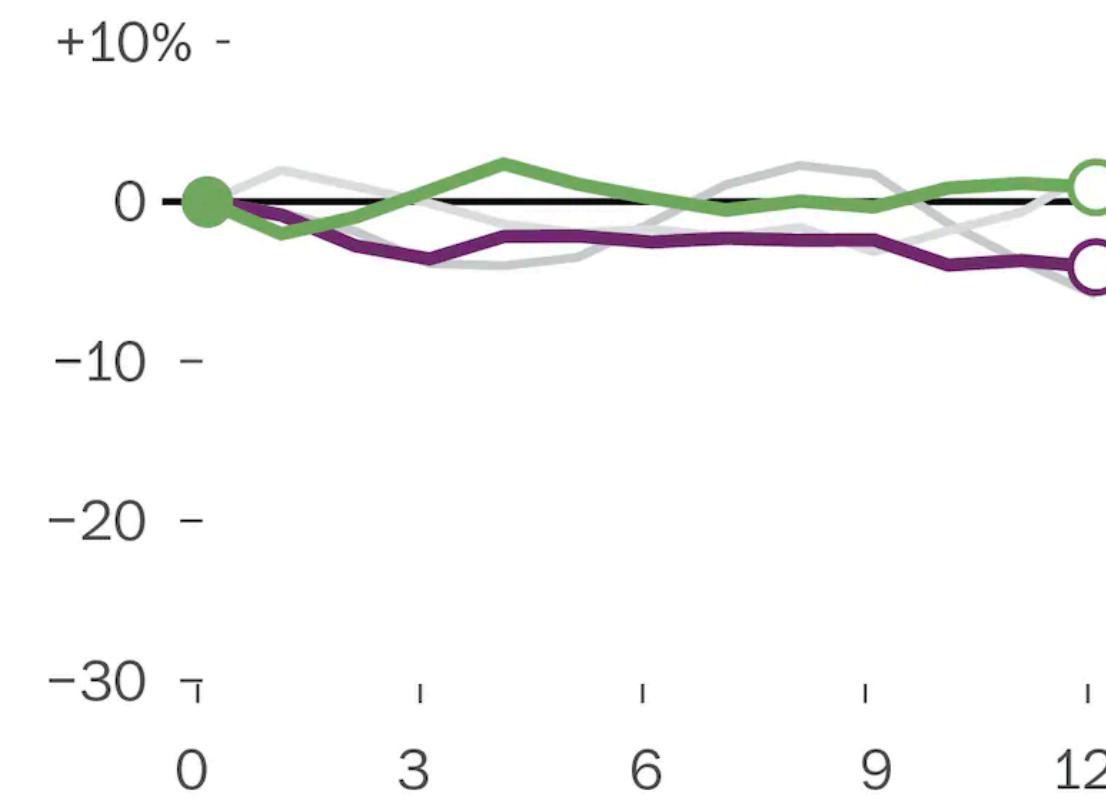
1990 recession



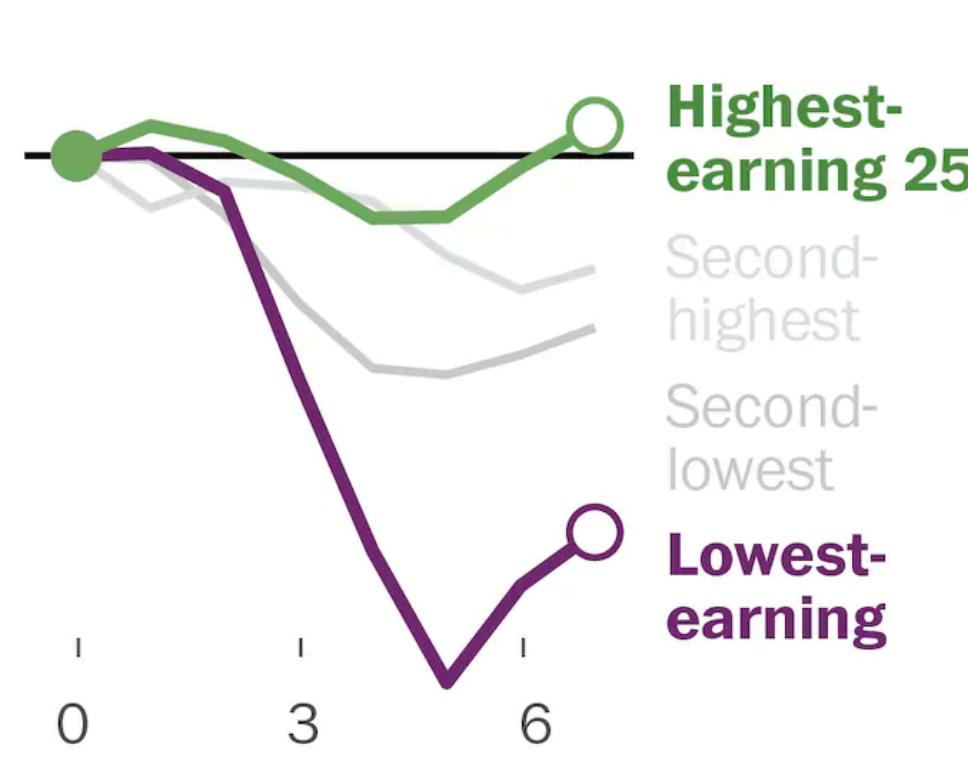
2001 recession



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

About Me

Assistant Teaching Professor, HDSI

lau@ucsd.edu

Email: lau@ucsd.edu

Tools for visualizing programs
(Pandas Tutor), curriculum design
(Learning Data Science textbook)

What makes me smile:

My wife, good food, traveling,
students who put in their best effort!



Course staff

See dsc106.com
for our OH times

Instructor

Sam Lau

Teaching Assistants

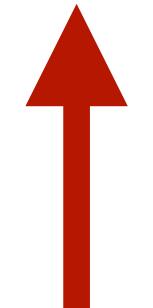
Giorgia Nicolaou (Head TA)

Heng Zhu

Tutors:

Jesse Huang

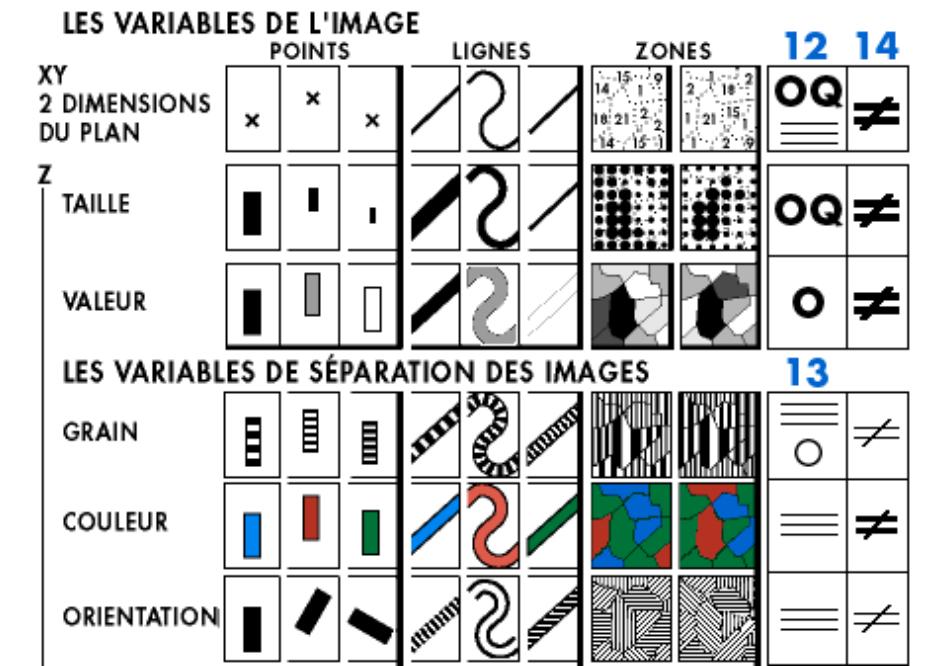
Lauren Zhang



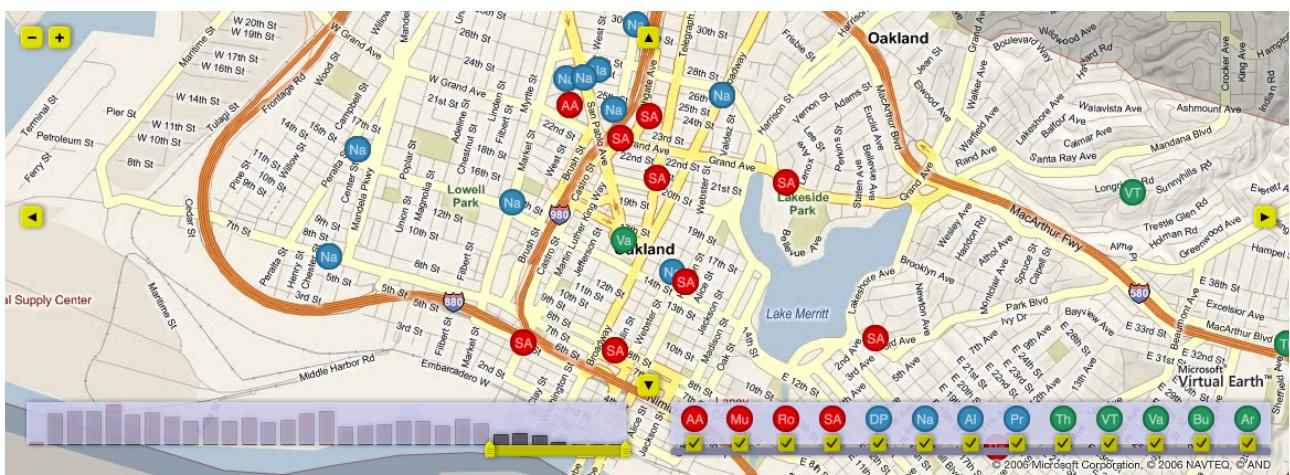
Questions about
course logistics?
Email Giorgia!

Principles

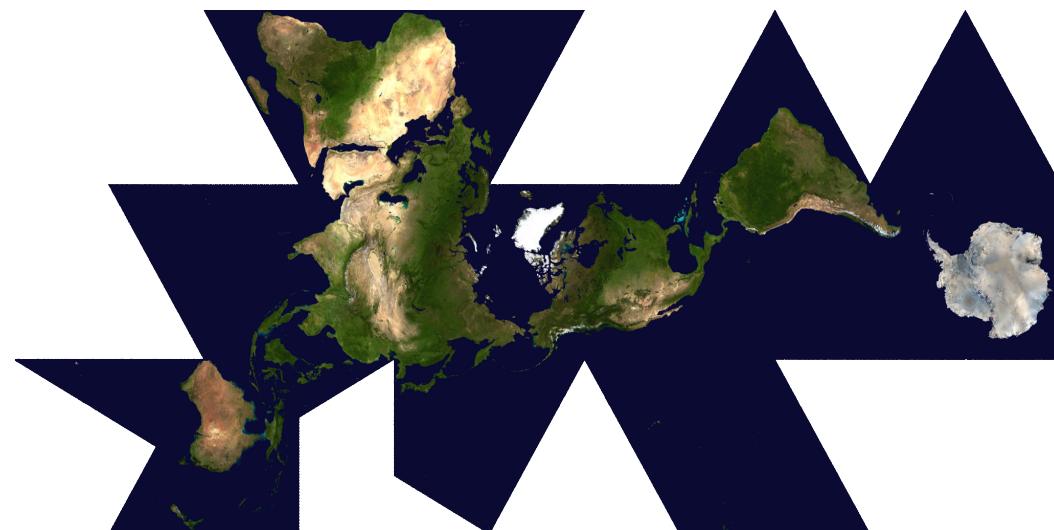
Data and
Image Models



Interaction



Maps



...and many more!

Techniques

HTML/CSS



JavaScript



D3.js



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.

This Quarter: Health Data (with Prof Ben Smarr)



You will visualize health data in Projects 2, 3, and the Final Project

Ben will give guest lecture to introduce datasets and background

For Final Project Showcase, we will invite outside guests (industry and medicine)

Grade Breakdown

Component	Weight
Participation	8%
Labs	8%
Project 1	10%
Project 2	15%
Project 3	15%
Project Checkpoints	4%
Final Project	40%

Grade Breakdown

Component	Weight	
Participation	8%	1% per week (2 lowest weeks dropped). 3 options: <ol style="list-style-type: none">1. Attend both lectures, discussion, and participate in the lecture activities.2. Share and critique 1 viz example on Ed.3. Respond to 2 viz examples on Ed.
Labs	8%	
Project 1	10%	
Project 2	15%	
Project 3	15%	
Project Checkpoints	4%	
Final Project	40%	See website for full details.

Grade Breakdown

Component	Weight
Participation	8%
Labs	8% 8 labs, 1% per lab. (No drops)
Project 1	10%
Project 2	15%
Project 3	15%
Project Checkpoints	4%
Final Project	40%

Grade Breakdown

Component	Weight
Participation	8%
Labs	8%
Project 1	10%
Project 2	15%
Project 3	15%
Project Checkpoints	4%
Final Project	40%

3 open-ended projects

Grade Breakdown

Component	Weight
Participation	8%
Labs	8%
Project 1	10%
Project 2	15%
Project 3	15%
Project Checkpoints	4%
Final Project	40%

Final project will span last 4 weeks of course

Grade Breakdown

Component	Weight
Participation	8%
Labs	8%
Project 1	10%
Project 2	15%
Project 3	15%
Project Checkpoints	4%
Final Project	40%

6 slip days for quarter.

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

But NOT for the Final Project
submission deadline

Communication

Use Piazza for all communication (my email is super slammed these days)

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

Explorable Explanation for health dataset

Initial prototype and design reviews

In-class demonstration **videos**

Submit and publish online

In Week 10, Final Project Showcase

See past projects at:

dsc-courses.github.io/dsc106-2025-wi/showcase

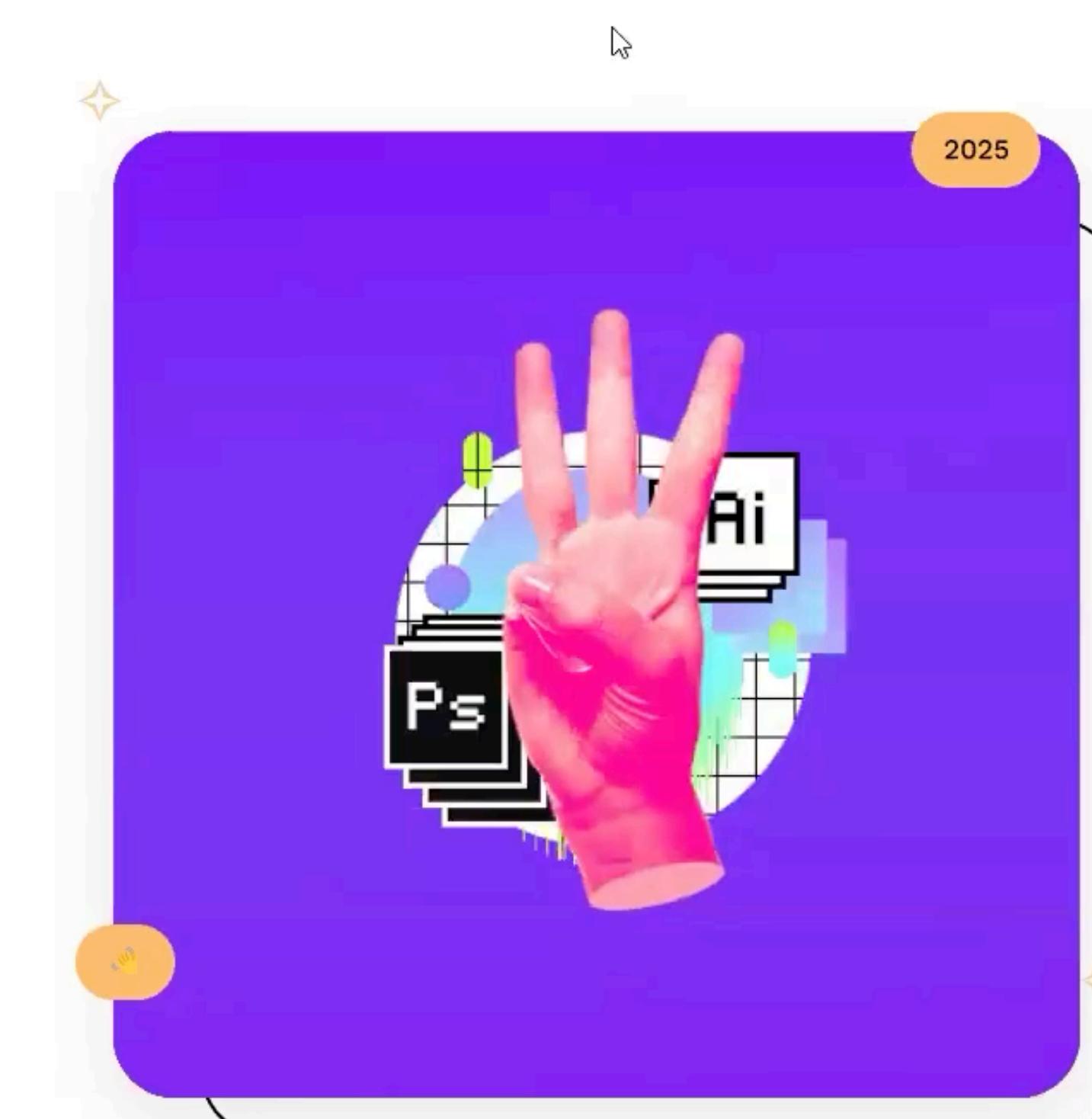
Parkinsons and Typing Ability

matthewbudding.github.io/FinalProject-DSC106/

The Effect of Parkinsons on Typing Ability

Investigating the impact of Parkinson's disease on dexterity and fine motor control through the visualization of typing.

Data Explore more



Eric Cheng
Designer

Jerric Jiang
Scatter Plot/Script

Omid Alamdar
Typing Speed/Video

Matthew Budding
Statistics/Aggregate Keystroke

8:00 PM 3/11/2025

<https://matthewbudding.github.io/FinalProject-DSC106/>

snigdhapodugu.github.io/walking-simulation/



Stepping Into Your Walking Journey

A PROJECT BY SNIGDHA PODUGU, DEEPAL DELEENA, VAIDEHI KARVE, AND ARYA VERMA

HOW DO YOU WALK?
TRY IT YOURSELF!!!



30 DAY MONEY BACK GUARANTEE!



1-800-WLK-GOOD

very small disclaimer that contradicts everything else said in commercial that is stated very quietly and fast in the last 2 seconds of the commercial that would probably prevent your grandma from buying whatever this is if she could hear or see it.

<https://snigdhapodugu.github.io/stepsync/>

Lab 1: Introduction to the Web Platform

Lab 1 released, due Friday.

🔗 Step 4: Add a photo

- 1 Create an `images` folder in the root of your repository
- 2 Find a photo of yourself (or anything else you want) and save it in the `images` folder
- 3 Add an `` element to your HTML page, with the `src` attribute set to the path to your image file (`images/your-image-file-name.jpg`)
- 4 Add an `alt` attribute to the `` element with a short description of the image as you would describe it to someone who cannot see it.

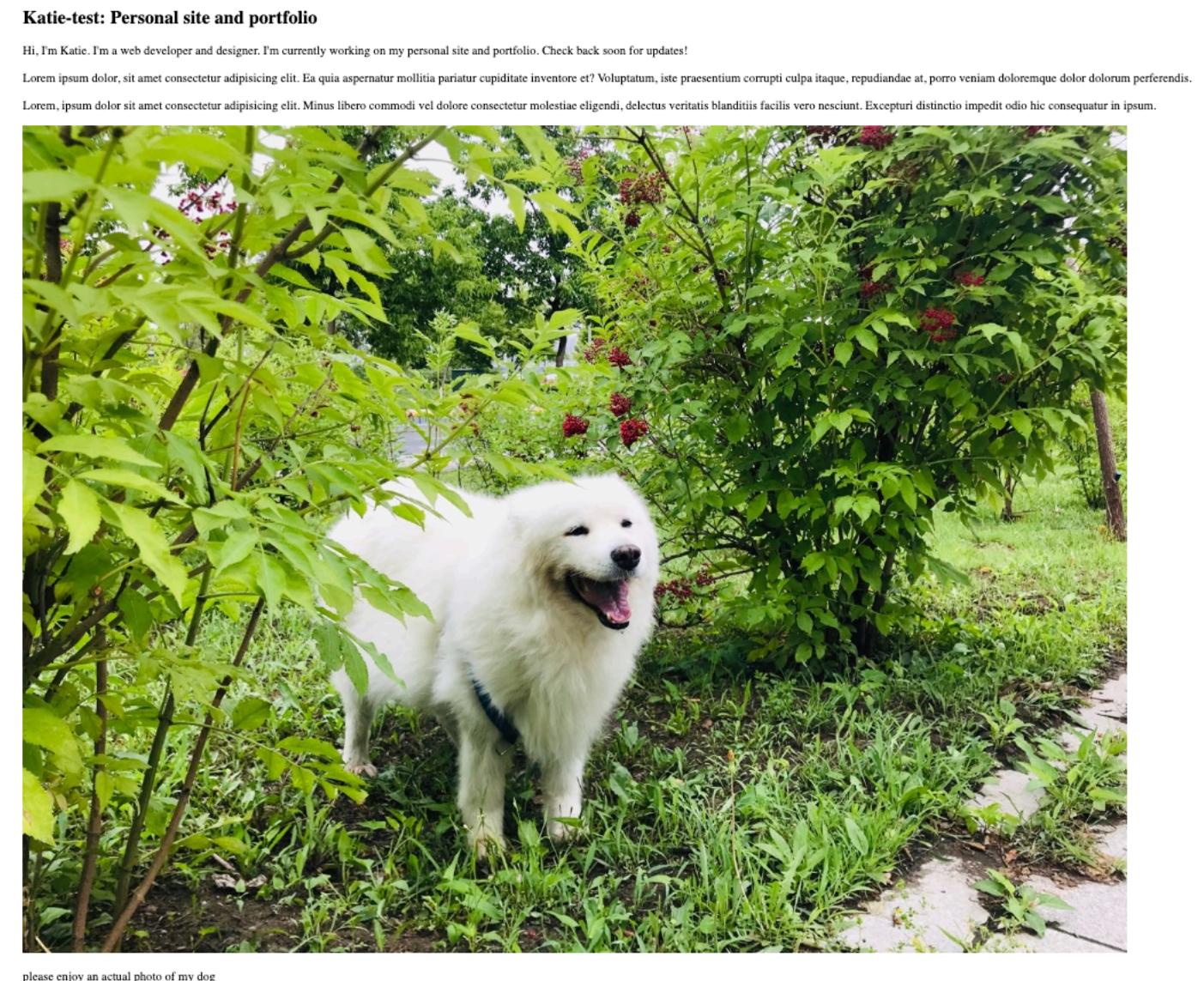


Figure 5: Our page now has a photo!

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. Python, Tableau, pen & paper).

Deliverables (upload via Gradescope; see Project 1 page)

Image of your visualization (PNG or JPG format)

Short description + design rationale (≤ 4 paragraphs)

**Checkpoint
due next Tues**

Advice from past students

If you want to be ambitious, it's really easy to do that in this class! Try and create something you're proud of. It's worth it. :)

Go to OH and use AI!!

The rubric is there to guide you, but you won't get anything from this class if your main objective is to simply get all the points. Be creative and take risks with what you visualize.

Most common advice:

Just please start your labs and projects early because I feel like if I did, a lot of headache of doing stuff on the last day would be eliminated :)

Questions?