

Communicating with Data – Data Feminism

Dr. Ab Mosca (they/them)

Slides based off *Data Feminism* by Catherine D'Ignazio and Lauren F. Klein

Plan for today

- Recap/Last bits from last class
- Power structures in data science
- Ways of knowing
- Importance of context

Flashback: Late Policy + Revisions

- Keeping up with weekly programming assignments is **crucial** to your success in this class:
 - assignments in this class generally build on one another
 - missing assignments often has 😞 downstream consequences
- That said, we all have outside lives that sometimes take priority over schoolwork – this is **normal and healthy**
- Self-managed extensions on HW + unlimited revisions
 - No budget / limit – take what you need to learn effectively
 - Must be requested **before** the original deadline
 - *Caveat*: late assignments get lower grading priority
- See course website for full instructions on taking an extension and submitting a revision

Flashback: Data

Data is a set of *variables* that capture various aspects of the world

One way to
think about this:

		VARIABLES			
		Tuition	Enrollment	Public vs. Private	...
OBSERVATIONS	Smith College	\$46,288	2,563	private	
	UMass Amherst	\$16,115	28,635	public	
	Hampshire College	\$48,065	1,400	private	
	Mount Holyoke College	\$43,886	2,189	private	
	Amherst College	\$50,562	1,792	private	
	⋮				

Another way to
think about this

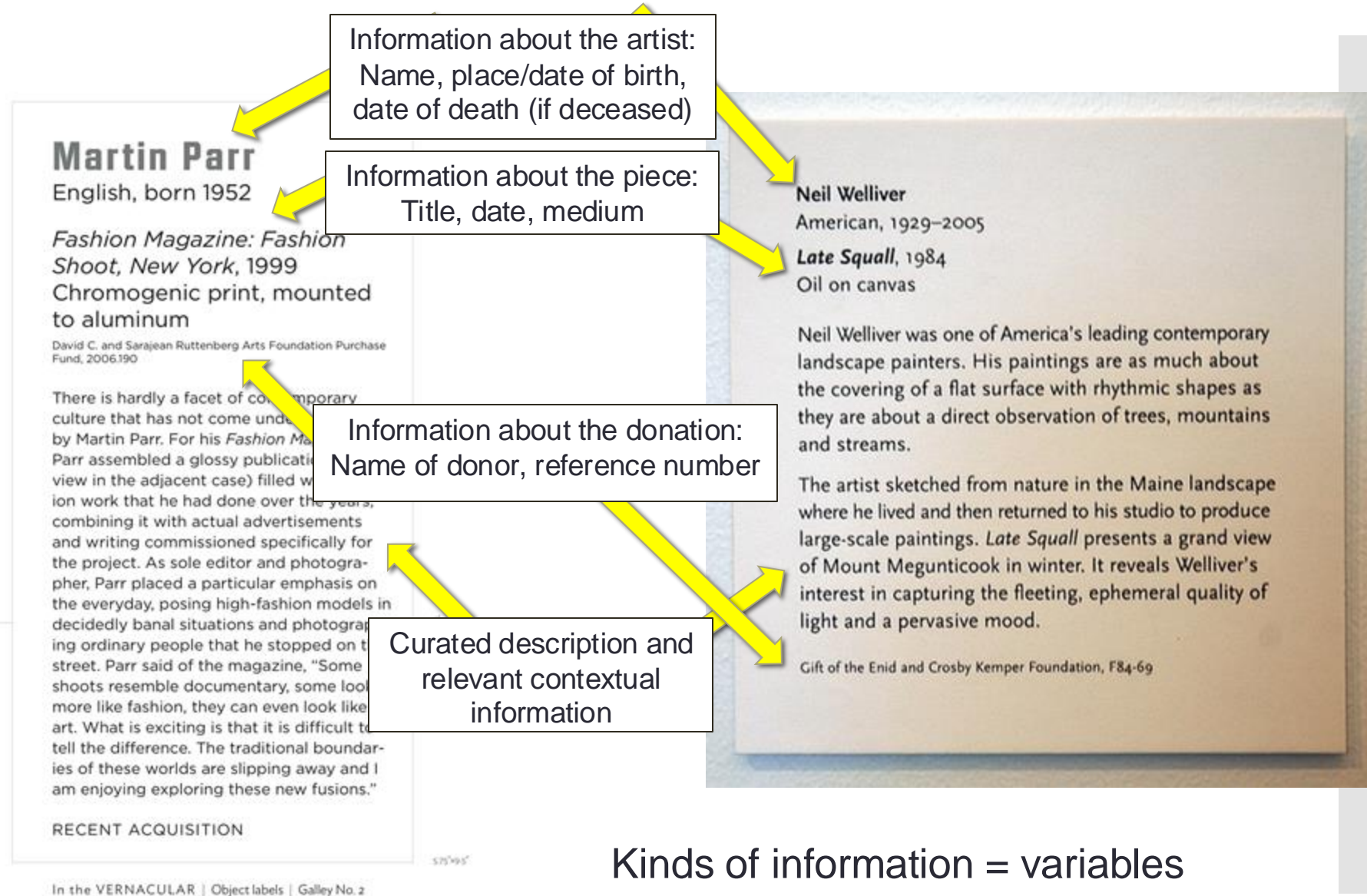
```
class school_obs:
    def __init__(tuition, enrollment,
                 pub_or_priv):
        self.tuition = tuition
        self.enrollment = enrollment
        self.pub_or_priv = pub_or_priv
```

OBSERVATIONS

```
smith = school_obs(46288, 2563, "private")
umass = school_obs(16115, 28635, "public")
```

VARIABLES

Yet another way
to think about
this

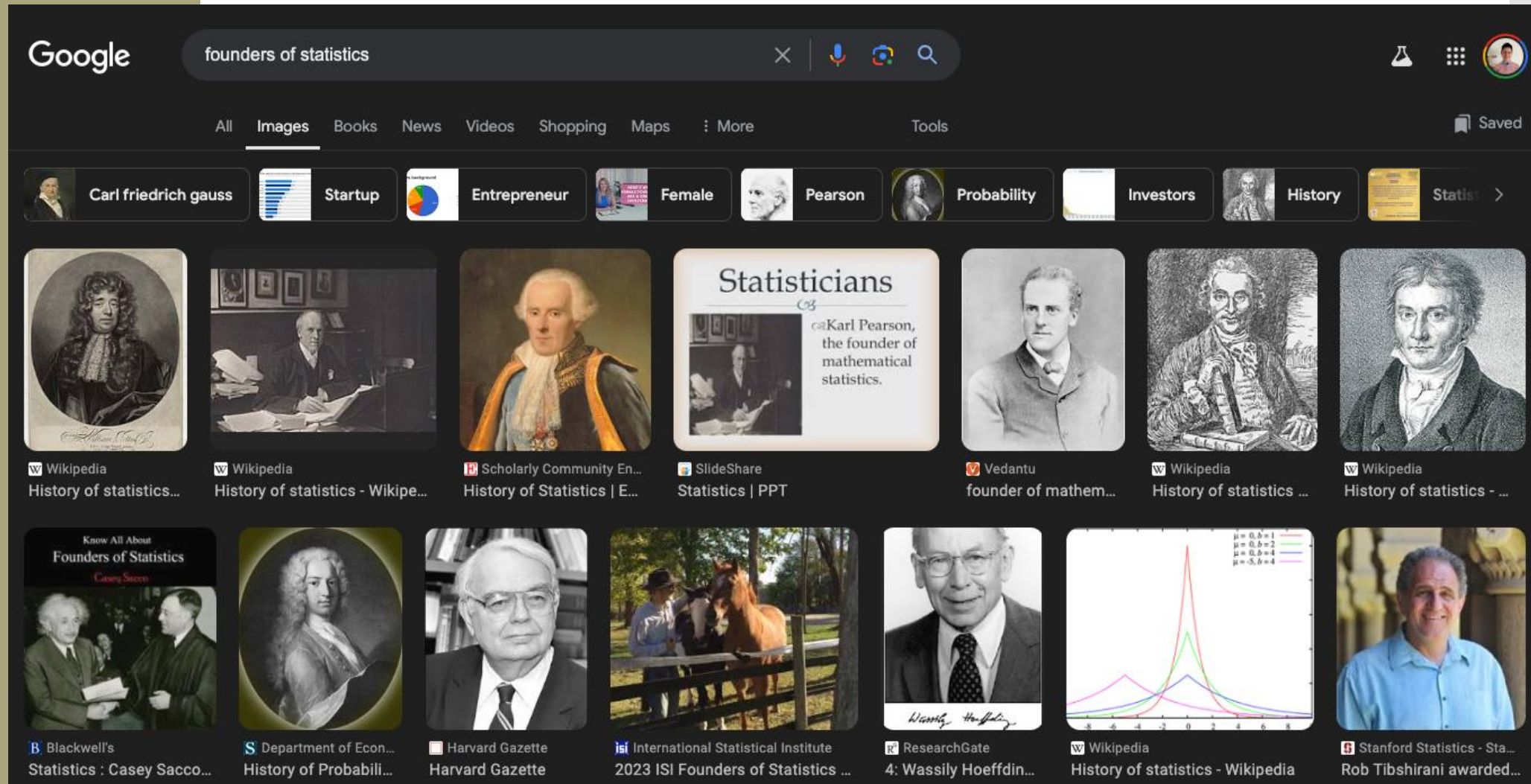


Kinds of information = variables
Actual text on the placard = observations

Discussion

Where does data come from?

Power



Power

Google search results for "founders of computer science".

Search bar: founders of computer science

Navigation: All, Images, Books, Videos, Maps, News, Shopping, More, Tools

Related searches: John mccarthy ai, Charles babbage, Invented, Artificial intelligence, Alan turing, Famous personalities, The father

Results:

- WCAI**: Father of Computer Science AI... (Image of Alan Turing)
- Wikipedia**: History of compute... (Image of Charles Babbage)
- Wikipedia**: History of computer s... (Image of Gottfried Wilhelm Leibniz)
- SUPSI - Dalle Molle Institute for Artificial Intel...**: Leibniz, father of computer science (Collage of historical figures and text)
- SlideShare**: Alan Turing: Founder of Computer ... (Overview of Alan Turing's life and work)
 - Overview
 - Alan Mathison Turing, OBE, FRS (23 June 1912 – 7 June 1954)
 - Mathematician and codebreaker
 - "Founder/father of computer science"
 - Centenary meetings at Cambridge, Bletchley Park, Manchester, Oxford, etc., in 2012
 - Increasingly in the public consciousness
- Britannica**: Charles Babbage | Bio... (Image of Charles Babbage)
- Gresham College**: Alan Turing: The Founder of Computer ... (Image of Alan Turing working)
- Math/Computer Scie...**: Founding Fathers of ... (Image of Alan Turing)
- Initial Commit**: 10 computer scientists who made history (Image of Alan Turing at a computer)
- Advanced Science News**: Pioneers in Science: Alan Turing ... (Image of Alan Turing with a diagram)
- Britannica**: Charles Babbage | Biogr... (Image of Charles Babbage)

Power

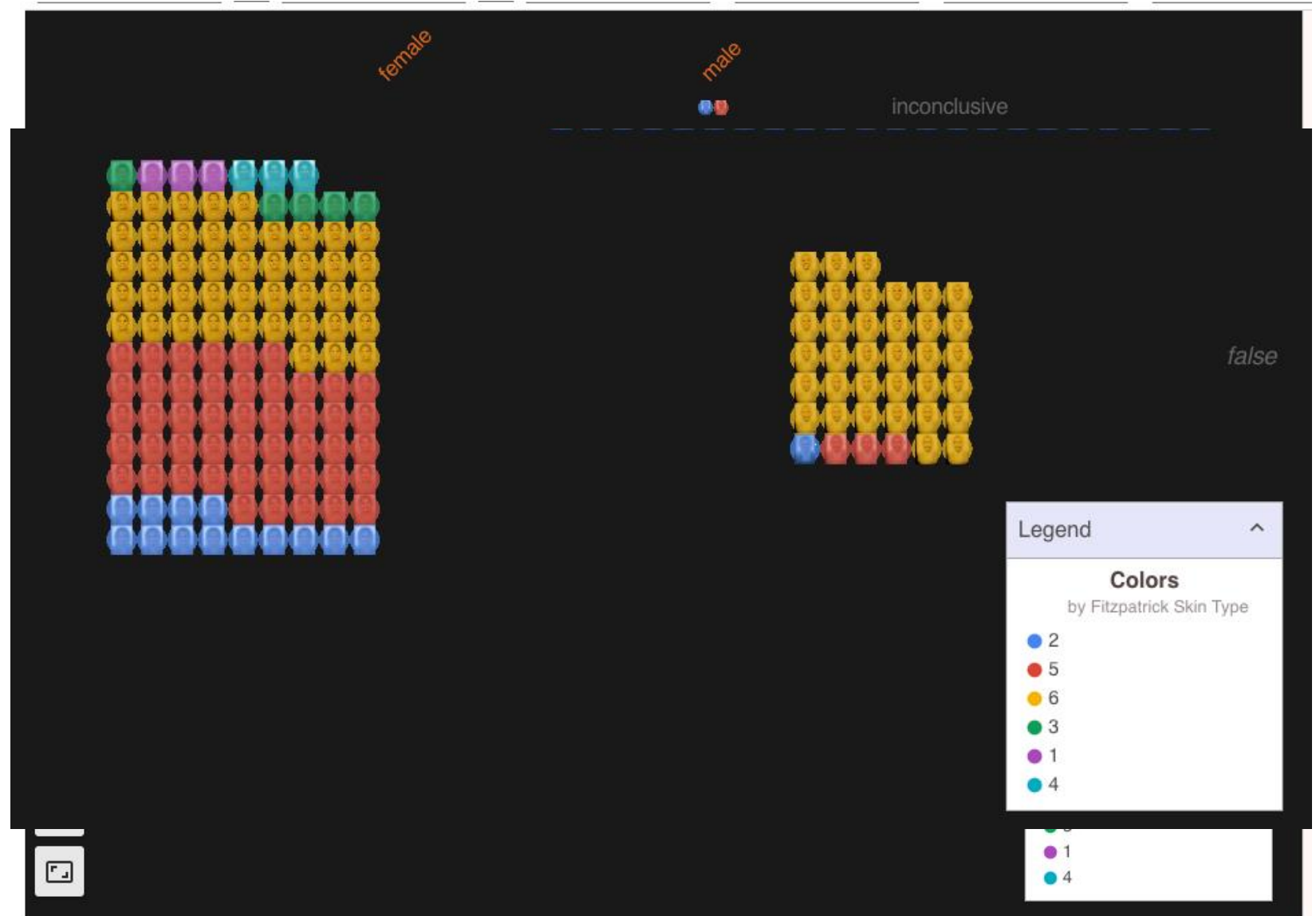
- Hmmmm..... Seems like data science is born from white cis males
 - Fun fact: this is not actually true if you dig deeper! But that's for another class

Example: Facial Recognition



Dr. Joy Buolamwini

Example: Facial Recognition



Example: Facial Recognition

- Training dataset used for most facial recognition systems contains
 - 78% male faces
 - 84% white faces
 - Only 4% were women and dark-skinned

What are downstream real life impacts of this data bias?

Example: Research Funding



NATIONAL CANCER INSTITUTE

- 2022 prostate cancer research funding: 280.5 million
- 2022 uterine cancer research funding: 15.0 million

How does this monetary breakdown
perpetuate systems of power?

<https://www.cancer.gov/about-nci/budget/fact-book/data/research-funding>

<https://data-feminism.mitpress.mit.edu/>

“What gets
counted
counts”

- Data is often used to inform policy and allocate resources
- What is not counted in that data collection can become invisible
 - Ex. Expansive gender

Sign Up
It's free and always will be.

First name Last name

Mobile number or email

New password

Birthday

May 4 1994 Why do I need to provide my birthday?

☐ Female ☐ Male

By clicking Sign Up, you agree to our [Terms](#), [Data Policy](#) and [Cookies Policy](#). You may receive SMS Notifications from us and can opt out any time.

Sign Up

What can we do?

- Subvert norms




What can we do?

- Highlight missing categories

You are ... Enter your details to find yourself in the new Congress	Gender Cis male Cis female Trans + non-binary	Ethnicity White Hispanic Black Asian Native American Middle Eastern Multiracial	Orientation Straight LGB
	Age Under 35 35-49 50-64 Over 65	Religion Christian Jewish Hindu Muslim Mormon Buddhist Non-religious	

**people in Congress are like you**
Race undeclared



[illegible]

<https://data-feminism.mitpress.mit.edu/>

What can we
do?

- Rethink data collection

A2 How do you identify your gender?

☐

Woman
(including trans woman)

☐

Non-binary

☐

In another way

☐

Man
(including trans man)

☐

Prefer not to say

A3 Is this the same gender you were assigned at birth?

☐

Yes

☐

No

☐

Prefer not to say

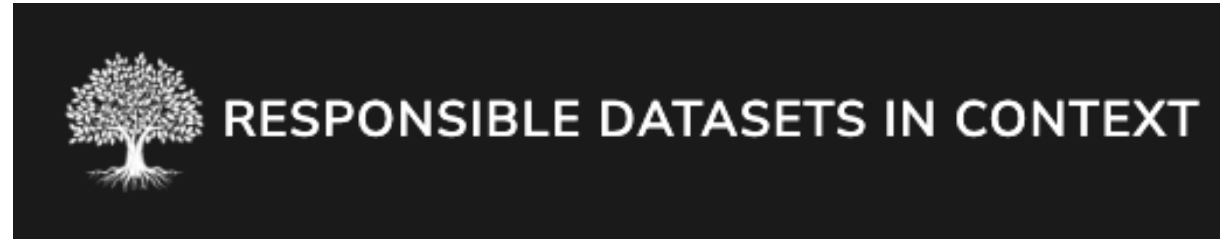
What can we do?

Table 5.1: Features of “data for good” versus data for co-liberation

	“Data for good”	Data for co-liberation
Leadership by members of minoritized groups working in community		✓
Money and resources managed by members of minoritized groups		✓
Data owned and governed by the community		✓
Quantitative data analysis “ground truthed” through a participatory, community-centered data analysis process		✓
Data scientists are not rock stars and wizards, but rather facilitators and guides		✓
Data education and knowledge transfer are part of the project design		✓
Building social infrastructure—community solidarity and shared understanding—is part of the project design		✓

- Many voices are better than a single all-knowing one

Acknowledge context



<https://www.responsible-datasets-in-context.com/datasets.html>

What can we
do?

- What is the historical context of the data?
- Where did the data come from? Who collected it?
- Why was the data collected?
- How was the data collected?
- How is the data used?
- What's in the data?
- What "counts" as a data point?
- What data is missing?
- How is uncertainty handled?

Lab

- Go to the course website to find instructions for today's lab