Elicitation Exercises

STS 101

Name one dataset that a data collector (your government, your university, a private company, a non-profit group, yourself) collects about:

1. You
2. Your family
3. Your community
4. Your state
5. Your country
6. Your globe
7. Your universe

To understand the impact of vehicle emissions on air pollution, the EPA needs counts of the number of cars that travel on each federal highway over the course of the year. Your data team has been hired by the EPA to produce a count of the number of cars that travel on each highway in CA. You are given a limited budget to hire individuals, purchase equipment, and pay for permits to temporarily shut down a highway.

How would you count the cars? What would you physically do?

What are some limits to your approach?

You've been hired by the UN Development Programme to measure "happiness" in 10 countries - Australia, Canada, China, Colombia, India, Israel, Kenya, Russia, United Kingdom, United States.

How would you do it?

How would you define happiness? How would you quantify it?

Let’s say you find a magic lamp, encasing a genie that will grant you 3 wishes.
...for data your government collects.

What 3 datasets would you ask for, and why?

...and no, you cannot ask for more datasets.

You've been hired by the UN Human Settlements Program to conduct a study of affordable housing in 10 countries - Australia, Canada, China, Colombia, India, Israel, Kenya, Russia, United Kingdom, United States.

You've been instructed to compare data and rates of housing affordability and rental stability in each of these countries and to display the differences on a map.

List three data quality issues that you can anticipate in conducting this study.

1. Your local United Way maintains a calling center that, during disasters, accepts phone calls from individuals in need and directs them to relevant social services. Representatives need to be able to quickly navigate a database of services during a disaster. Create a classification system for how they might organize the various types of social services in the database.
2. In the wake of a disaster, all of the phone lines in your community have gone down, and callers are being redirected to a different United Way for help with finding services. Turn to your neighbor and compare the taxonomies. How difficult would it be for a call representative in a different city to navigate the differences in the schemas?

You've been hired by the CalEPA to write a report about the "environmental health" of each county in California.

---

1. What is one concrete dataset that you would use to write the report?
2. How, specifically, would you get that data?
3. What are some of the challenges to collecting and using the data?
4. Who, other than CalEPA, would have stakes with the data?

Example:

Number of asthma-related hospitalizations in 2017 per county	Survey all hospitals across the state about asthma-related hospitalizations	• Rely on hospitals having good information practices • Patient privacy laws • Not all patients go to hospitals in their own counties • Can't capture folks seen by private doctors or not seen at all	• Hospitals • Patients • Insurance Companies • Inhaler Companies • Who else?

You've been given a dataset documenting the income per household in CA since 2000. Each row lists:

- A household,
- Its zip code,
- Its county,
- The year income was reported,
- The household income,
- The number of working adults in the household,
- The average hours individuals worked per week

How might an activist aiming to demonstrate income inequalities filter the data to show inequalities?

Who might try and refute the activist? How might this individual select different variables and filter the data to oppose the claim?

The Department of Education has asked you to develop a metric to determine how well teachers are preparing students for successful futures.

- Describe a evaluation system someone might use to measure this. What data would be collected? What indicators would be considered?
- How might this metric reshape how teachers perform their day-to-day tasks?
- What may be some of the (un)intended consequences of designing this metric?

- You are starting a social networking site where community advocacy groups can advertise various types of civic and political campaigns. You want to give these groups the option to target their campaigns to particular types of users.
- List 5 buckets into which you would categorize users. Why did you select these particular buckets?
- Reflect on how it would affect the users' lives to be classified this way.
- Who might be left out from this classification?

You are part of a local community group that wants to put pressure on your city council member to curtail the rapid development your community has been facing. Construction has been loud. Low and middle income residents are being priced out of their units and small businesses. New bars and restaurants are creating a lot of noise.

Your community group has recently gotten a lot of press. ...and not good press. Your group is being cast as a bunch of angry voices shouting “NIMBY” - Not in My Backyard. You plan to write an op-ed providing evidence of the quality of life impacts the development is having on local residents. How would you gather that evidence?

An earthquake just hit San Francisco, and you are part of the emergency data team responsible for estimating how many individuals are in need of assistance in each SF district. You have 6 hours to come up with the estimates, so that the City can begin directing resources.

How would you come up with the estimates?

What data would consider?

Who might not get included in the count, using the method you've come up with?

I recently heard a rumor that if you rub the eggheads around campus during finals week that it will bring you good luck, and you will pass your finals.

Based on the data below, draw a data visualization that convinces me that this rumor is true. What techniques would you use to make this visualization persuasive? What do these techniques hide?

Rubbed egghead and passed	30% - 9030 students
Rubbed egghead and didn't pass	10% - 3010 students
Didn't rub egghead and passed	40% - 12040 students
Didn't rub egghead and didn't pass	20% - 6020 students

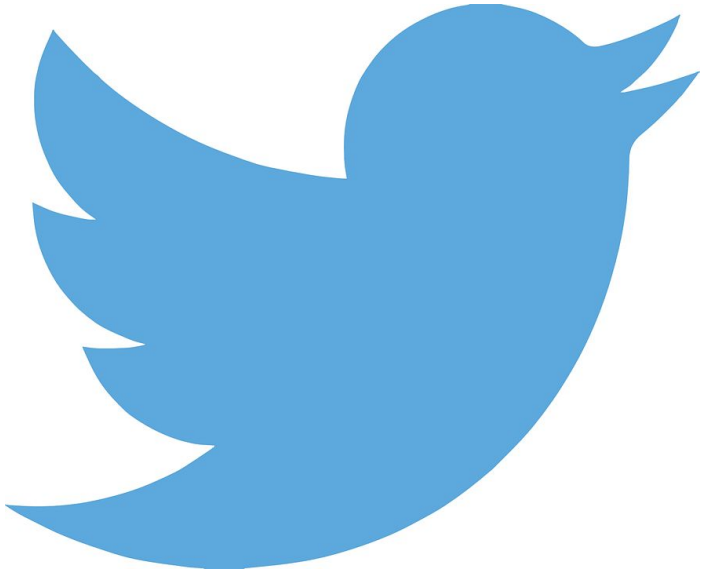
The Environmental Defense Fund has tasked your data team with developing a series of visualizations to represent the human impact on climate change. What data would you collect, and how would you collect the data? How would you visualize the data in a compelling and honest way?

The following lists a week of Air Quality Indexes for the City of Davis. What more would you need to know to communicate air quality risk levels over the past week to the public?

How would you visualize the data in an instructive way?

Monday	33
Tuesday	35
Wednesday	42
Thursday	51
Friday	46
Saturday	35
Sunday	30

Select one data analysis that you have performed in this course, and in 140 characters or less describe what the data shows. What contextual details were you able to to capture in the Tweet? What was glossed over?



Select one dataset that we have talked about in this course or that you've written about in your journal entries. Name another dataset that can be brought together with this dataset in order to expand our understanding of an issue.

1. What might we learn from bringing these datasets together that we might not otherwise have been able to know?
2. What are some of the risks of joining this data together? Can you anticipate unintended consequences of what joining this data might reveal?

The City of San Francisco has asked your data team to come up with a system for predicting where high school dropouts are more likely to happen. Start a long list of factors you might consider.

Why did you select these variables - i.e. why do you believe these factors might be an indicator for dropout rates?

What are the dangers to using these variables to make predictions?