

CSC/SDS 109: Communicating with Data

Fall 2024

Lab 01: Analyzing Data Sources

*This is a **small group assignment**; form teams of ~3.*

Goals:

- Compare and contrast well documented data with poorly documented data
- Think about the downstream effects of data sources and documentation

Instructions

Each group will investigate a dataset. Once you have your group, go to this google sheet to select a dataset:

https://docs.google.com/spreadsheets/d/1XlvcYCxoP2IU84vHoC_IgpLsJM15X78VyovSGlwAbCQ/edit?usp=sharing

Explore your dataset and its documentation to answer the following questions. Do your best to answer each question. If you've given it ~7 minutes and still can't find the answer, note that you couldn't find an answer and move on.

Be prepared to share your findings with the class (you are welcome to just talk, share slides, share a google doc, etc.. Whatever makes it easiest for you to report your findings).

If you get stuck, please ask for help!

Questions:

1. What is the historical context of the data in your dataset?
2. Where did the data come from? Who collected it?
 - a. Do you see any potential biases or data quality issues based on your answer?
3. Why was the data collected?
 - a. Do you see any potential biases or data quality issues based on your answer?
4. What variables are in the data and what information do they hold?
5. What does one observation in the dataset represent?

6. Where is there missing data?
 - a. Do you see any potential biases or data quality issues based on your answer?
7. How is uncertainty in the data handled?
 - a. Do you see any potential biases or data quality issues based on your answer?
8. Who or what larger group does your dataset represent? (Ex. MA residents, Smith College students, Trees on the east coast)
9. What analysis questions could you answer with your dataset? What analysis questions would you be unable to answer?

Submission

Submit some record of your work **as a group** on Gradescope.

(<https://guides.gradescope.com/hc/en-us/articles/21863861823373-Adding-Group-Members-to-a-Submission>)

Rubric

Participation in this lab counts towards your overall participation and engagement grade. Do your best in the time allotted in class and submit on Gradescope for full credit (your grade will simply be did you meaningfully engage, yes or no).