MATH113/CAIS105: Intro to Data Science

Fall 2023

Homework 01

Homework is DUE before class on the day indicated on the course schedule.

Learning Objectives:

- Observe and critically reason about data science "in the wild"
- Identify parts of a dataset

Part 1

Find an example of data science "in the wild". Some places you might look are:

- News outlets (ex. FiveThirtyEight, NY Times, Washington Post, Boston Globe)
- City or state websites (ex. boston.gov, mass.gov)
- Websites for areas of interest to you (ex. nasa.gov, noaa.gov, nrel.gov)

Once you have your example, read it critically. Can you:

a) Find and access the original, raw data?
b) Find who collected the data?
c) Identify any potential issues with the data such as missing groups, potential biases?
d) Understand what analyses were done on the data?
e) Identify the takeaway message of the work? What do you think the author(s) wanted you to get out of reading it?

Start a document for this assignment and write up your answers to these questions in it. Be sure to include the link to your example of data science "in the wild". (Note: it's okay if your answer to d is "no", or "not entirely" – unclear data science communication happens often and it's one of the things we will practice doing well in this class!)

Part 2

Choose an area of interest to you and find a related dataset. Add a link to the data and at least the first 10 rows of the dataset to the document you started for Part 1. (Please add headings for part 1 and 2 to your document).

Label the following on your dataset:

- Variables
- Observations
- Values

For each variable, record its type and if applicable, it's ordering direction.

Submission

Submit your work through PLATO.