DS2001 - CS Practicum

Spring 2023

Project #1

Data scientists are great communicators. We find insights from big piles of data and use them to convince other people about what's important. Most of our DS2000 homeworks and DS2001 practicum assignments focus on writing great code, but data scientists also need to be able to create informative visualizations and present them to an audience. That is what we're practicing with this project.

This project is worth 10% of your DS2001 grade (remember that DS2000 and DS2001 have separate grades!). It's a chance to practice and get feedback on these additional, creative aspects of data science: creating visualizations and communicating about them.

Timeline

All of the work for this project should be done during practicum. No need to take anything home.

- Feb 8 and 9 -- Choose your dataset, select your question to ask/answer with guidance from instructors/TAs, and generate your plot.
- Feb 15 and 16 -- Present your plot to Ab or a TA during practicum. (When not presenting, you'll have a regular practicum assignment to work on.)

Requirement #1: Pick a Dataset

Choose a dataset you've seen in class, in practicum, or on a homework. You can also choose an entirely new data that you've downloaded (e.g., from kaggle, data.boston.gov, data.gov, BARI, etc.).

Either way, we expect you to choose a reasonable dataset -- at least 10 points of data to work with. But remember we have only one practicum to work on this! If you haven't found your absolute favorite dataset after 20 minutes, use your second-favorite and move on to step 2.

Requirement #2: Ask a question about the data

Given your data, what are you curious to know? Think carefully about what is included in your dataset and questions you can reasonable ask and answer given the data and what you've learned so far in class.

Requirement #3: Generate a Plot

Use matplotlib to generate a plot (e.g., a scatterplot, bar chart, line chart, etc.) that answers your question. Your plot has to be new in some way -- something we didn't do in practicum, and

something that wasn't specifically assigned in a homework. You can use data we've seen before, but should ask and answer a different question about it.

We expect your plot to be labeled, easy to follow, and use colors that are not matplotlib's defaults.

Requirement #3: Present Your Plot

In practicum, you'll present your plot to Ab or a TA. We will not look at your code, and you will be graded entirely on your plot and presentation.

Your presentation should be about 2-3 minutes long, and it should assume your audience knows nothing about the data. You don't need to create any slides; just show the plot and describe it:

- What data the plot is based on.
- What question you've asked about the data.
- How your visualization answers that question.

Submission + Grading

You'll submit a screenshot of your plot on Gradescope. **Submit your plot on gradescope before your presentation time arrives.** You'll be graded on a scale of 1-5, where:

- 5 = plot is interesting and informative; presentation is clear.
- 3 = plot needs improvement, and/or presentation is hard to follow
- 1 = plot is hard to understand, and/or presentation is not useful or informative.