

# Homework 5 (due 3/2)

DS-210 @ Boston University

Spring 2022

## Before you start...

**Collaboration policy:** You may verbally collaborate on required homework problems. However, you must write your solutions independently without showing them to other students. If you choose to collaborate on a problem, you are allowed to discuss it with at most 2 other students currently enrolled in the class.

The header of each assignment you submit must include the field “Collaborators:” with the names of the students with whom you have had discussions concerning your solutions. If you didn’t collaborate with anyone, write “Collaborators: none.” A failure to list collaborators may result in a credit deduction.

You may use external resources such as software documentation, textbooks, lecture notes, and videos to supplement your general understanding of the course topics. You may use references such as books and online resources for well known facts. However, you must always cite the source.

You may **not** look up answers to a homework assignment in the published literature or on the web. You may **not** share written work with anyone else.

**Submitting:** Solutions should be submitted via Gradescope. The entry code is 3Y85PZ.

**Grading:** Whenever we ask for a solution, you may receive partial credit if your solution is not sufficiently efficient or close to optimal. For instance, if we ask you to solve a specific problem that has a polynomial-time algorithm that is easy to implement, but the solution you provide is exponentially slower, you are likely to receive partial credit.

**Late submission policy:** No extensions, except for extraordinary circumstances. We accept submissions submitted up to one day late, but we may deduct 10% of points.

## Questions

Please submit a solution to this homework as a single PDF file. (Another document format is fine as well as long Gradescope accepts it and can display it.)

1. **(20 points)** Find two publicly available data sets that you like. One of these data sets could become the data set that you analyze in your final project, and the motivation for this task is to explore options for such a data set. However, there is no commitment at this point and you are welcome to use a different data set later.

For each data set:

- Clearly specify its name and provide a URL to where it can be obtained.
- Explain why you selected it, using *at least three* and *at most ten* sentences.

To help you with this task, we provide a list of links to public data sets in Piazza (pinned post “Finding a data set for the final project”).

## 2. (Optional, no credit)

- Install a Rust kernel for Jupyter. We will use it for our lecture notes, but you will be asked to submit your code in Rust as `.rs` files, so this will not be directly useful for your homework. The instructions are available at [https://github.com/google/evcxr/tree/main/evcxr\\_jupyter](https://github.com/google/evcxr/tree/main/evcxr_jupyter)

**Warning:** This may not be easy!

- Pronounce “evcxr.”