DATA 23700 Autumn 2023

Exercise 5: Project check-in

Due November 3, 2023

For the project (due November 20), students will produce an original data analysis and a short written report. The first steps involved in the project are to **find a dataset to analyze** and to **motivate a problem or question** the student will address using this dataset. These first steps are the emphasis of the project check-in (a.k.a. Exercise 5).

Students should submit a one-page PDF document both: (1) motivating a problem or question they wish to address in their project; and (2) describing the dataset they plan to use in terms of what variables are included, provenance, and why the data are suitable to address the motivated question or problem. Students should include at least one visualization of their chosen dataset to demonstrate its value for the purpose they have identified.

It is important to choose a dataset that can support analysis of sufficient depth to demonstrate skills aquired in the course and of sufficient interest to support a narrative about the analysis akin to a technical report or a piece of data journalism. Choosing a dataset that has no interesting structure or which cannot support the intended analysis would set students on shakey footing for the project, so we are checking in with you to help avoid this (and to help students avoid procrastinating).

Students must work alone. Knowingly choosing the same dataset as a peer in order to reduce effort is an academic honesty violation. Recall that the Project (unlike Exercises in general) is intended to be individual work.

To get credit for Exercise 5, **students must submit a PDF document to Gradescope** by the stated deadline.