DATA 22700 Spring 2023

Exercise 6: Project check-in

Due May 5, 2023

For the project (due May 19), students will produce an original data analysis and a short written report. The first task involved in the project is to **find a dataset to analyze**. This is the emphasis of the project check-in (a.k.a. Exercise 6). Students should come to class on Thursday, April 27 with a few ideas of what dataset(s) they might use for the project. Pratham (one of our TAs) will be in our usual classroom at the usual time and will check with students individually that they have identified a suitable dataset.

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 6, **students must check in with Pratham, who will mark the Exercise as complete on Gradescope.** If you miss class on April 27, please make time to visit Pratham's office hours to get your choice of dataset approved.