

Final project: data set

EDH7916 | Spring 2020

Benjamin Skinner

***NOTE** This assignment needs to be completed by the start of the next class. That means everything pushed to your remote GitHub repo before class starts. If you are unsure whether you have successfully pushed your changes, check the online version of your repo at [GitHub.com](https://github.com). If you can see your changes there, I can see them too.*

Remember, I encourage you to save your work, commit smaller changes, and push to your remote GitHub repo often rather than wait until the last minute.

Your final project will be to produce a 3-5 page reproducible report on an education topic of your choosing. It should:

- contain a combination of writing, tables, and figures
- have minimal citations
- be fully reproducible with minimal effort

The sophistication of your analyses should be commensurate with your statistical training so far. That said, I'm generally expecting descriptive analyses targeted at sophisticated, but non-technical audiences. By far the most important aspects of the report are that your underlying code follows the good practices we learn over the semester and that **the final report is reproducible (by me) with minimal effort**.

To facilitate the last item it is a requirement that **everyone must use publicly available data**. For purposes of this project, a publicly available data set is one that anyone can, today, download from a website. This restriction, therefore, excludes data you've received from your advisor, that you've collected yourself, or that one could theoretically get by emailing an administrator or author of a paper.

To help you get started, your assignment this week is to

1. brainstorm a topic of interest
2. find a publicly-available data set that will allow you examine that topic in a short descriptive analysis
3. begin reading through the technical report and codebook to see which variables you might use

I would recommend using of the NCES panel data sets or IPEDS, but I'm open to others.

Create a new folder called `final_project` in your student repo. Submit a Markdown file called `final_project_dataset_<last_name>.md` with the following information:

- Name and link to data set you plan to use
- 1-2 paragraph(s) on your topic idea, with at least a sentence or two about how the data set you've chosen will help you examine your topic
- A Markdown table with 5-10 key variables and their descriptions that you plan to use to examine your topic