SOCI 1230

Data Science Across The Disciplines

Winter 2022

Week One Assignment

In this assignment, you will continue exploring the TFS and CSS questions. The assignment will build on the tools we have learned in morning and afternoon sessions to visualize and manipulate data.

You are encouraged to use any of our course materials. You are free to collaborate with other students in our section but each student should submit their own report.

This assignment is due via Canvas by 10:00 AM on Tuesday, January 18, 2022.

Please submit your .Rmd notebook and a knitted PDF of your notebook that includes your output.

Part One

Part One uses the **tfs_question_summary.csv** and **css_question_summary.csv** files we saw in class on Thursday.

Choose three questions from the TFS. For each:

- Make a figure showing the distribution of responses. Use a mix of plot types (dodged and stacked column plots, for example) and remember to label the plots, axes, legends, etc. as appropriate.
- Write a few (3-4) sentences about how you think responses to this variable would be associated with a mobility variable from Opportunity Insights' Table 2. The course readings may be helpful as you consider the possible associations. You do not have to explain every response. For example, take the "firstchoice" response to the "CHOICE" question. Would you expect colleges with a higher proportion of students attending their first choice college would have higher average values for the `k_rank` variable from the mobility dataset? That would be a positive association. Or would colleges with a higher proportion of students attending their first choice college have lower values for the `par_q1` variable from the mobility dataset? That would be a negative association. Remember you can use the **online codebook** for descriptions of the mobility variables. Use a mix of mobility variables (so not `k_rank` for all three of your questions.)

Repeat for three questions from the CSS.

Part Two

Part two uses the **tfs_college_means.csv** and **css_college_means.csv** files.

For the TFS questions you used in Part One:

- Find the mean of one response level (weighted by `n_responses`, the number of survey responses at each college) by type of college (use the `type` variable). The response level can be an existing one (like `CHOICE.firstchoice`) or a new level you think is preferable (like `CHOICE.firstchoice + CHOICE.secondchoice`). All the means can be summarized in the same function.

For the CSS questions you used in Part One:

- Find the mean of one response level (weighted by `n_responses`, the number of survey responses at each college) by tier of college (use the `tier_name` variable). The response level can be an existing one or a new level you think is preferable. All the means can be summarized in the same function.