

Assignment 5

EDH7916 | Summer C 2020

Benjamin Skinner

Using the `hsls_small.dta` data set (the Stata version) and the online codebook, make plots that help answer each of the following questions. You **do not** need to save the final plot. Throughout, you **should** account for missing values by dropping them.

For each question, show your data work and write 1-2 sentence(s) in a comment that describes the relationship your plot shows (*i.e.*, how it answers the question).

Questions

1. What is the distribution of household size among students in the sample?
2. How does student socioeconomic status differ between students who ever attended college and those who did not?
3. How do parental educational expectations differ across region?
4. How does the relationship between socioeconomic status and math test score differ across region (use a smoothing line to help show any relationship)?
5. Among students who ever attended college, how does socioeconomic status differ between those who delayed postsecondary enrollment and those who did not delay, when delay is defined as:
 - more than 6 months between high school graduation and postsecondary enrollment?
 - more than 12 months?
6. **BONUS** Using the `plotly` library, recreate one of the plots from the lesson, but add an interactive component. For example, in one of the scatter plots, make it so that hovering over a point produces a pop-up that shows its value and the student to which it belongs.

Submission details

- Upload your script (`<lastname>_assignment_5.R`) to our CANVAS course site no later than 11:59 p.m. EDT on the due date.