

Assignment Four: Data Science Fundamentals Summer 2020

Your fourth assignment aims to build your skills in work with data, thinking about research problems, and developing modeling pipelines. You will also continue to build your skills in R Markdown by crafting another document in that language. For this assignment, you must select at least two variables that you care about from your dataset, hypothesize how they relate, and present a visualization that illustrates the relationship. Your R Markdown document should include:

1. Descriptions of the variables, including the coding of the variables, summary statistics, and, if appropriate, visualizations.
2. Description of the relationship between the variables, including a hypothesis (or hypotheses) about the relationship.
3. Initial demonstration of the relationship, which could include correlation, visualization, or model.

As always, if you are encountering problems, please post questions to Slack or reach out directly to me. Students who are interested in using a different dataset than they have used to this point may do so with the prior consent of the instructor.

DUE: Your R Markdown source code (.rmd file) and output file should be emailed to me at `drrice@umass.edu` by 11:59 p.m. EST on **Friday, August 14th**.