Problem Set #1 - The College Dataset

Econ 224

Due Date: Sunday, September 2nd 11:59pm

Instructions

Complete both of the following parts. For part 2, submit your code and solutions as an RMarkdown document. In particular, upload both the `.Rmd` file you used to generate your report and the resulting `.html` output to *canvas* by the due date listed above. Late work will only be accepted in exceptional circumstances, so you are better off submitting an incomplete problem set rather than nothing at all. You may discuss this problem set with your classmates, provided that you adhere to the *empty hands* policy: after any such discussion, all parties must leave the room *empty-handed* i.e. without code files or written notes or any kind. In other words, the final code and write-up that you produce must be entirely your own work. If you discuss the problem set with any other students be sure to list their names at the top of your problem set. Copying of problem set code or solutions in violation of the empty hands policy constitutes cheating and will be reported to the Office of Student Conduct.

Part 1

Complete Chapters 1-3 of the course "Reporting with R Markdown" on *Datacamp*.

Part 2

1. Install the R package `ISLR` using the command `install.packages('ISLR')`. Load the package using `library(ISLR)` and then read the help file for the `College` dataset using the command `?College`.
2. Pose 3 empirical questions that could be answered by constructing plots and calculating summary statistics for the `College` dataset.
3. For each of your 3 questions, write `dplyr` and `ggplot2` code to produce the summary statistics and plots you will need to determine the answer, and write a brief summary of your findings.

Note:

You will get 100% for Part 1 as long as you complete the required chapters before the deadline. For Part 2 you will be graded both on the quality of your questions and the quality of your answers to them. A question such as "what is the average graduation rate across colleges?" is not very interesting and does not give you an opportunity to show what you've learned in class. In contrast, a question like "how does the average graduation rate vary with selectivity and whether a school is public?" is much more interesting. Be sure to write clean code following the guidelines discussed in class.