Assignment #1: Exploratory Data Analysis (4 pts) 4-5 members

Group Submission

Due: Wednesday Sep 19 2:00 pm.

In this assignment, we will predict the Graduation rate (Grad.Rate) using the other variables in the College data set. The primary goal is: "Which variables (predictors) better predict the Graduation rate?" You can find a detailed description for this data set on Page 54 in our textbook.

In this assignment, please use R to answer the following six questions.

1. How many records in this data set? How many variables?

2. Which of the predictors are quantitative, and which are qualitative?

3. What is the minimum, maximum, mean and median of each quantitative predictor?

4. Use the hist() function to produce histograms for Grad.Rate, using different numbers of bins if necessary. Describe your findings.

5. Make boxplots of Grad.Rate for Private. Comment on your findings.
   Hint: use plot() with Private on x-axis and Grad.Rate on y-axis.

6. Use scatterplots to examine the predictors. Could you find any of the other variables are the potential predictors of Grad.Rate from your plots? Justify your answer.

Deliverables:

1. Group submission. Each group submits one set of report and code. Please include a cover page listing all team members' names.
2. Two files: R code file and the report are submitted to Blackboard.
3. The report should contain the answers for each question. Please do not include R code or R raw outputs except figures in the report. Only perform exploratory data analysis for this data without model development. Please concisely summarize the major findings and pay attention to the presentation of your tables and figures.

The assignment will be graded on neatness of the report, comprehensiveness of the analysis, and clarity of results presentation.