**Homework 1: Due Thursday, February 11 at 11:59pm**

There are three parts to this homework assignment, each with multiple questions. Please insert answers and plots under their corresponding questions, then save the document as a pdf and upload it to CANVAS. *Providing your R code is not required, but may be helpful when assigning partial credit.*

We will be using the **College** dataset in the **ISLR** package for this assignment.

1. First, we will explore the dataset.
2. [5 points] How many colleges are in the dataset?
3. [5 points] How many features are there for each college?
4. [5 points] Which feature(s) is(are) categorical?
5. [5 points] Which feature(s) is(are) numerical?
6. [5 points] How many colleges are private?
7. [5 points] What is the mean graduation rate across colleges?
8. [5 points] What is the maximum number of undergraduate students at a college? *Hint:* Make sure that you include both full-time and part-time students in your calculation.
9. Second, we will examine the relationship between the percentage of students from the top of their high school class and the out-of-state tuition at a college.
10. [10 points] Create a scatterplot with the percentage of students from the top of their high school class on the axis and the out-of-state tuition on the axis. Overlay the points with a smoothed line and confidence bands. Remember to avoid overplotting.
11. [5 points] Is the correlation between the percentage of students from the top of their high school class and the out-of-state tuition positive or negative?
12. [5 points] Explain what this correlation means about the relationship between the percentage of students from the top 10% of their high school class and the out-of-state tuition at a college.

1. [5 points] If you were to perform a hypothesis test to evaluate this relationship, what would be the null hypothesis?
2. [5 points] Why is the confidence interval widest when the percentage of students from the top of their high school class is largest?
3. Last, we will compare relationships between the percentage of students from the top of their high school class and out-of-state tuition at public and private colleges.
4. [10 points] Using different colored points for public and private colleges, create a scatterplot with the percentage of students from the top of their high school class on the axis and the out-of-state tuition on the axis. Overlay each set of points with a smoothed line of the same color with confidence bands. Remember to avoid overplotting.
5. [10 points] Using faceting, create side-by-side scatterplots for public and private schools, with the number of students from the top of their high school class on each axis and the out-of-state tuition on each axis. Overlay each scatterplot with a smoothed line with confidence bands. Remember to avoid overplotting.
6. [5 points] Is the out-of-state tuition generally higher at public or private colleges?
7. [5 points] Is the correlation between the percentage of students in the top of their high school class and the out-of-state tuition stronger for public or private colleges?
8. [5 points] Why are the confidence bands generally wider for public than for private colleges?