**IST 5535 Lab 1: College Data**

This exercise relates to the College data set, which is saved in the file College.csv. The dataset contains 777 universities and colleges in the US. Variables include:

- Private: Public/private indicator

- Apps: Number of applications received

- Accept: Number of applicants accepted

- Enroll: Number of new students enrolled

- Top10perc: New students from top 10 % of high school class

- Top25perc: New students from top 25 % of high school class

- F.Undergrad: Number of full-time undergraduates

- P.Undergrad: Number of part-time undergraduates

- Outstate: Out-of-state tuition

- Room.Board: Room and board costs

- Books: Estimated book costs

- Personal: Estimated personal spending

- PhD: Percent of faculty with Ph.D.’s

- Terminal: Percent of faculty with terminal degree

- S.F.Ratio: Student/faculty ratio

- perc.alumni: Percent of alumni who donate

- Expend: Instructional expenditure per student

- Grad.Rate: Graduation rate

**Instruction:** Finish the following steps in R Markdown. Submit your R Markdown and HTML report to assignment on Canvas.

Steps:

1. Use the read.csv() function to read the data into R. Make sure strings are read as factors. Call the loaded data college. Show the structure of the data.
2. The first column is just the name of each university. We can set college names as row names then remove the first column from the dataset.
3. Data Exploration
   1. Show summary statistics of the dataset.
   2. Use pairs() method to create a scatterplot matrix of the first ten variables.
   3. Draw a boxplot of Outstate versus Private.
   4. Create a new qualitative variable, called Elite, from the Top10perc variable. We divide universities into two groups based on whether or not the proportion of students coming from the top 10 % of their high school classes exceeds 50%.Then use the summary() function to see how many elite universities there are.
   5. Draw a 2 by 2 plot which includes four histograms of four variables including Enroll, Outstate, Grad.Rate, and Books.
   6. Use ggplot2 package to visualize the relationship between graduation rate and room and board costs. What is the relationship? Does the relationship change when compare between elite and non-elite universities?