National University of Singapore
School of Computing

SWS3018 Predictive Analytics
Lab 2

Learning Objectives

- Manipulate data in R
- Generate charts/graphics on R

1. In this exercise, we will still be using the `College` (`College.csv`) data set from Tutorial/Lab 1

   a) Use the `summary()` function to produce a numerical summary of the variables in the data set.

   b) Using the `pairs()` function to produce a scatterplot matrix of the following columns: **Apps**, **Accept**, **Enroll**, **Grad.Rate**.

   c) Generate the same plot as (b) but differentiating the the Private and Non-private University using different colors.

   d) Use the `plot()` function to produce side-by-side boxplots of **Apps** vs **Private**. Label your axes and the plot and provide a title for your plot.

   e) Produce a scatterplot based on **Top10perc** (percentage of new students from top 10% of the high school). Using different colors for the different percentages (<=30, <=50, <=70, >70)

   f) Use the `sort()` function to sort the values in ascending order and generate the same plot as (e) on this sorted data

   g) Use the `hist()` function to generate a histogram of **Top10perc**. Add in the same color coding scheme as (e)

   h) The command `par(mfrow=c(1,2))` allows you to divide the plotting window into multiple regions for plotting multiple plot. Try typing it on the console window together with the plotting commands from (f) and the (g). How can you increase the regions to allow 4 diagrams instead?

   i) Try exploring the data with different charts.