Homework 1

Data: **Cars.csv** for Exercise 1-2 and **airquality** (R built-in) for Exercise 3.
Use the significance level of .05

Cars.csv will be used for Exercise 1 and 2. The variables in the data are included below in the table. The variables in the data set are the following attributes of cars in the year 2004:

- **Make** – the auto manufacturer
- **Model** – name of the vehicle
- **Type** – SUV, sedan, sports, truck, or wagon
- **Origin** – continent of the manufacturer; Europe, Asia, or USA
- **Invoice** – price (dollars) that the manufacturer sends to the dealer upon delivery of the car
- **Horsepower** – amount of the car's power
- **MPG_City** – miles per gallon (fuel efficiency) during city driving
- **MPG_Highway** – miles per gallon during highway driving
- **Wheelbase** – distance (inches) between the centers of the front and rear wheels
- **Length** – distance (inches) from the nose to the tail of the car

Exercise 1: Descriptive Statistics (20 points)

a) Create a combined mpg variable called **MPG_Combo** which combines 60% of the **MPG_City** and 40% of the **MPG_Highway**. Obtain a box plot for **MPG_Combo** and comment on what the plot tells us about fuel efficiencies.

b) Obtain box plots for **MPG_Combo** by **Type** and comment on any differences you notice between the different vehicle types combined fuel efficiency.

c) Obtain basic descriptive statistics for **Horsepower** for all vehicles. Comment on any general features and statistics of the data. Use visual and quantitative methods to comment on whether an assumption of Normality would be reasonable for **Horsepower** variable.

d) Use visual and quantitative methods to comment on whether an assumption of normality would be reasonable for **Horsepower** variable by **Type**, especially for Sedan, *Sports, and SUV* (i.e., check normality of **Horsepower** from Type of i) Sports, ii) SUV, and iii) Truck.

Exercise 2: Hypothesis Testing (15 points)

Perform a hypothesis test of whether **SUV** has different MPG_City than **Truck**, and state your conclusions

a) Which test should we perform, and why? Justify your answer based on findings on Exercise 1 (d).

b) Specify null and alternative hypotheses.

c) State the conclusion based on the test result.

The **airquality** data will be used for Exercise 3. The information about variables in the dataset can be found in R.

Exercise 3: Hypothesis Testing (15 points)

Perform a hypothesis test -whether Wind in **July** has a different speed (mph) than Wind in **August.**

a) Which test should we perform, and why? See QQ-plot and perform the Shapiro-Wilk test for normality check.

b) Specify null and alternative hypotheses

c) State the conclusion based on the test result.