**R Lesson 3 - Descriptive Statistics**

**References:**  
Black - Chapter 3 Descriptive Statistics (pp. 52-95)  
Verzani - Chapter 2 Univariate Data (pp.77-79)  
Lander - Chapter 11 Group Manipulation (pp.117-139) and Chapter 15 Basic Statistics (pp. 187-189)  
Stowell - Chapter 5 Summary Statistics for Continuous Variables (pp. 59-62)

**Exercises:** Solve the following problems using the data sets as indicated.

[mileage.csvView in a new window](https://canvas.northwestern.edu/courses/38799/files/2178901/download?wrap=1) is derived from a 1991 U.S EPA study of passenger car mileage. This file includes information on sixty cars: HP (engine horsepower), MPG (average miles per gallon) WT (vehicle weight in 100 lb units) and CLASS (vehicle weight class C1,…,C6).

1. For each weight class determine the mean and standard deviation of MPG. What can you conclude from these calculations?
2. For each weight class determine the mean and standard deviation of HP. What can you conclude from these calculations?

[shoppers.csvView in a new window](https://canvas.northwestern.edu/courses/38799/files/2178900/download?wrap=1) contains the dollar amounts spent in a store by individual shoppers during one day. Find the mean, median, range, standard deviation, variance, Q1, Q3 and P10. Plot the histogram and describe the distribution.

[pontus.csvView in a new window](https://canvas.northwestern.edu/courses/38799/files/2178902/download?wrap=1) lists the ages of USA Presidents at the time of their inauguration. Also listed are the heights of the Presidents and their opponents.

1. Find the mean, median, range, standard deviation, Q1, Q3 and P10 of the Presidents’ ages..
2. Find the mean, median, range, standard deviation, Q1, Q3 and P10 of the heights of the Presidents and also their opponents. .
3. Calculate the difference between each President’s height and that of his opponent. Plot a histogram of these differences. Construct a boxplot.

What do you conclude from your calculations? Why is the difference of average heights calculated in (2) different from the average of the pairwise differences calculated in (3)?

[geyser.csvView in a new window](https://canvas.northwestern.edu/courses/38799/files/2178903/download?wrap=1) contains the intervals (in minutes) between eruptions of Old Faithful Geyser in Yellowstone National Park. The data were taken on two consecutive weeks: WEEK1 and WEEK2. Compare the two sets of data using summary(), hist() and boxplot(). What do you conclude?

**Running into Trouble?** Check out these solutions to help guide you along.

* [Lesson\_03\_Solutions .pdfPreview the documentView in a new window](https://canvas.northwestern.edu/courses/38799/files/2178836/download?wrap=1)
* [Lesson\_03\_Code\_Solution.rView in a new window](https://canvas.northwestern.edu/courses/38799/files/2178759/download?wrap=1)