Jaechul Kim

a. Opening file Boston.csv.

Reading line 1

heading : rm,medv

new length506

Closing file Boston.csv

Number of records: 506

Stats for rm

Sum: 3180.03

Mean: 6.28463

Median: 6.2085

Range: 5.219

Covariance = 4.49345

Correlation = 0.69536

Program terminated.

b. It was way easier to use the built in functions in R because I don't have to make any functions to perform these operations in R because it was built in. I could understand why it is more ML friendly language because it provides features like this in R, which are often used in ML.

c. Mean is when you add up all of the values in data and divide it by the number of observation. It shows the center value of the data.
Median is the middle value of the data set when arranged in order. If the value is even number it provides the average of two middle values.

Range is the difference of largest and smallest values. It shows how spread your data is.

All of these features give basic understanding of the data and can help to do data exploration prior to machine learning.

d. Covariance and Correlation give understanding of the relationship between two objects. Positive Covariance means that two objects move in same direction and negative Covariance means that two objects move in opposite direction. Correlation is measured in -1 to 1 scale where -1 is negative correlation 0 is no correlation and 1 Is largest correlation.

   These features are important to understand the relationship of two objects and it will help us to build better model in Machine Learning.