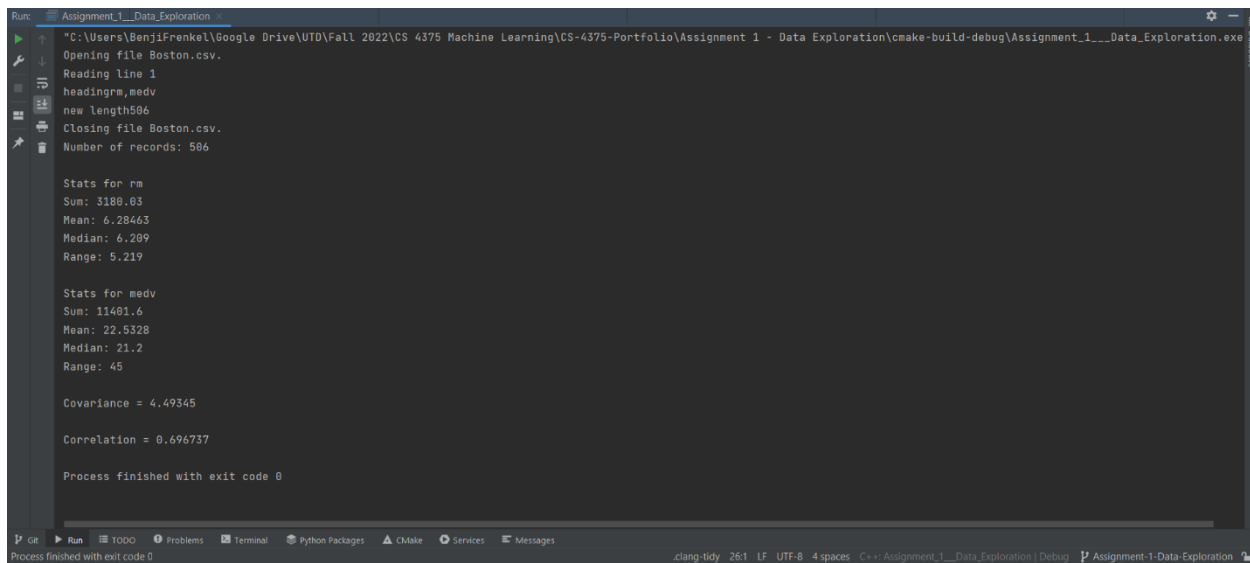


a)



```
Run: Assignment_1_Data_Exploration
"C:\Users\BenjiFrenkel\Google Drive\UT0\Fall 2022\CS 4375 Machine Learning\CS-4375-Portfolio\Assignment 1 - Data Exploration\cmake-build-debug\Assignment_1_Data_Exploration.exe"
Opening file Boston.csv.
Reading line 1
headingrm,medv
new length586
Closing file Boston.csv.
Number of records: 586

Stats for rm
Sum: 3188.03
Mean: 6.28463
Median: 6.209
Range: 5.219

Stats for medv
Sum: 11401.6
Mean: 22.5328
Median: 21.2
Range: 45

Covariance = 4.49345

Correlation = 0.696737

Process finished with exit code 0
```

b)

Using built in functions in R is way faster and more convenient compared to having to code your own functions in C++

c)

The mean is the average of the data, or the sum of the data divided by the size of the data set. The median is the middle value in the data set. The range is the largest value minus the smallest value in the data set, or otherwise known as the spread. These values allow us to interpret sets of data.

d)

Covariance is the measure of how two random variables vary together. Correlation is the measure of the extent of how much two variables are linearly related. Both statistics are useful for figuring out if variables have a relationship. These statistics are useful in machine learning because machine learning is about finding relationships between data and making predictions.