Yazan Abughazaleh
CS 4375.004

Data Exploration

a) Program Output:

```
Opening Boston.csv File
Header: rm,medv
vector size: 506
Closing Boston.csv
Number of observations: 506

rmSum: 3180.03

medvSum: 11401.6

rmMean: 6.28463

medvMean: 22.5328

rmMedian: 6.2085

medvMedian: 21.2

rmRange: 5.219

medvRange: 45

Covariance 4.49345

Correlation 0.69536
Program ended with exit code: 0
```

b) Implementing the functions in C++ was a relatively quick and fun exercise to perform. I felt as though I better understood how the metrics work and how they are calculated. I do however understand the reason why languages like R are preferred for this type of work. Having to implement these functions for each project in C++ is a tedious task and it also introduces opportunities for errors and miscalculations in a program if a function is incorrectly implemented. R provides these functions built in to allow users to use these metrics in building larger projects, providing a solid and verified foundation. It makes large projects easier to work on by cutting down the time required to implement the project and this is noticeable when I personally work with R, since I can instantly find the metrics I need.

c) The statistical mean is the average of all values in a data set. The median represents the value in the middle of a sorted data set. Lastly, the range is indicative of the spread of the data. The mean and the median help find the center point of the data, while the range helps show how far apart values in a dataset are from each other.

d) The covariance formula is used to determine how the changes in one variable affect another variable. Correlation is used to determine whether a relationship exists between the two variables. Covariance is used to determine whether a correlation between variables exists. These two metrics are very helpful in machine learning because they indicate whether variables are good predictors for a given target. The goal of machine learning is to ensure that models use the best predictors to find the targets.