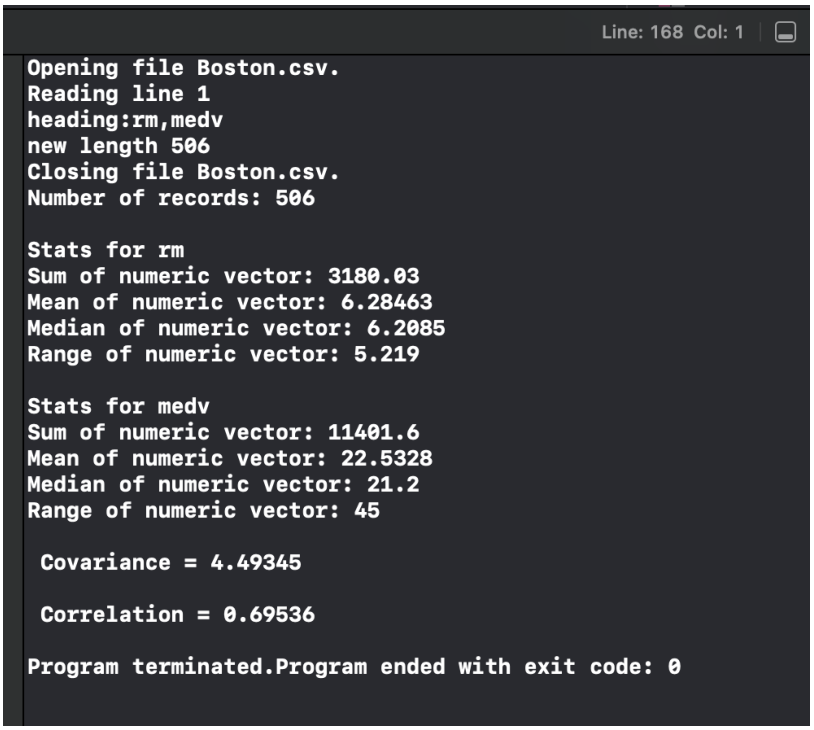
a. copy/paste runs of your code showing the output



b. describing your experience using built-in functions in R versus coding your own functions in C++

It's easier to use R because it has built in functions whereas in C++ we have to code each function which requires more effort and is time consuming, so I would say I prefer R.

c. describe the descriptive statistical measures mean, median, and range, and how these values might be useful in data exploration prior to machine learning

Mean is the sum of all values divided by the total number of values in the dataset. Median is the value that lies in the middle such that half of the values are higher and half of them lower. Range is the difference between the highest and lowest values. These values can be useful because it helps us understand the data better. For example the median tells us where the middle value of the dataset is which can be helpful.

d. describe the covariance and correlation statistics, and what information they give about two attributes. How might this information be useful in machine learning?

Covariance measures how changes in one variable are related to changes in a second variable. It is actually correlation scaled to (-1,1). Correlation measures the extent to which two variables are linearly related. It is a value between -1 and +1. -1 means that it has negative correlation and +1 means it has positive correlation. Covariance and correlation are

useful in machine learning because they can help us understand the relation between two variables.