Andrew Gerungan
CS4375.003
awg190000

```
PS C:\Users\gerun> cd "c:\Users\gerun\Desktop\stuff\School\UTD FALL 2022\CS4375\" ; if ($?) { g++ pfc1.cpp -o pfc1 } ; if ($?) { .\pf
c1 }
Opening file Boston.csv
Reaing line 1
heading: rm,medv
new length 506
Closing file Boston.csv.
Number of records: 506

Stats for rm
sum      3180.03
mean     6.28463
median   6.209
min      3.561    max      8.78

Stats for medv
sum      11401.6
mean     22.5328
median   21.2
min      5         max      50

 Covariance = 4.49345

 Correlation = 0.696737

Program terminated.
PS C:\Users\gerun\Desktop\stuff\School\UTD FALL 2022\CS4375> 
```

a.

b. Using built-in functions in R is far easier and more convenient that coding my own function in C++, largely because I don't really have to process and compute all the data but rather allow the computer to do it for me. I will say that coding the function in C++ did give me a refresher for C++ as well as more solidify my understanding of the statistical functions.

c. Mean and median are both useful values, both being used in finding a general estimate of the data in the dataset. The range is useful in understanding the extremities of the data. Both can be utilized in data exploration prior to machine learning since they generalize the data and help us get a grasp on the scope of the dataset.

d. Covariance is how much a variable is related to the other, which helps determine how much a given variable changes alongside the other. Meanwhile, correlation scales that number to a range of [-1, 1], where the closer to either 1 signifies that the 2 attributes are closely related and the further from either signifies less relation; the positive or negative just shows negative or positive correlation. This information can be useful in machine learning as it helps algorithms determine which set of attributes to utilize in calculations, as attributes that are not related are probably not useful to compare.