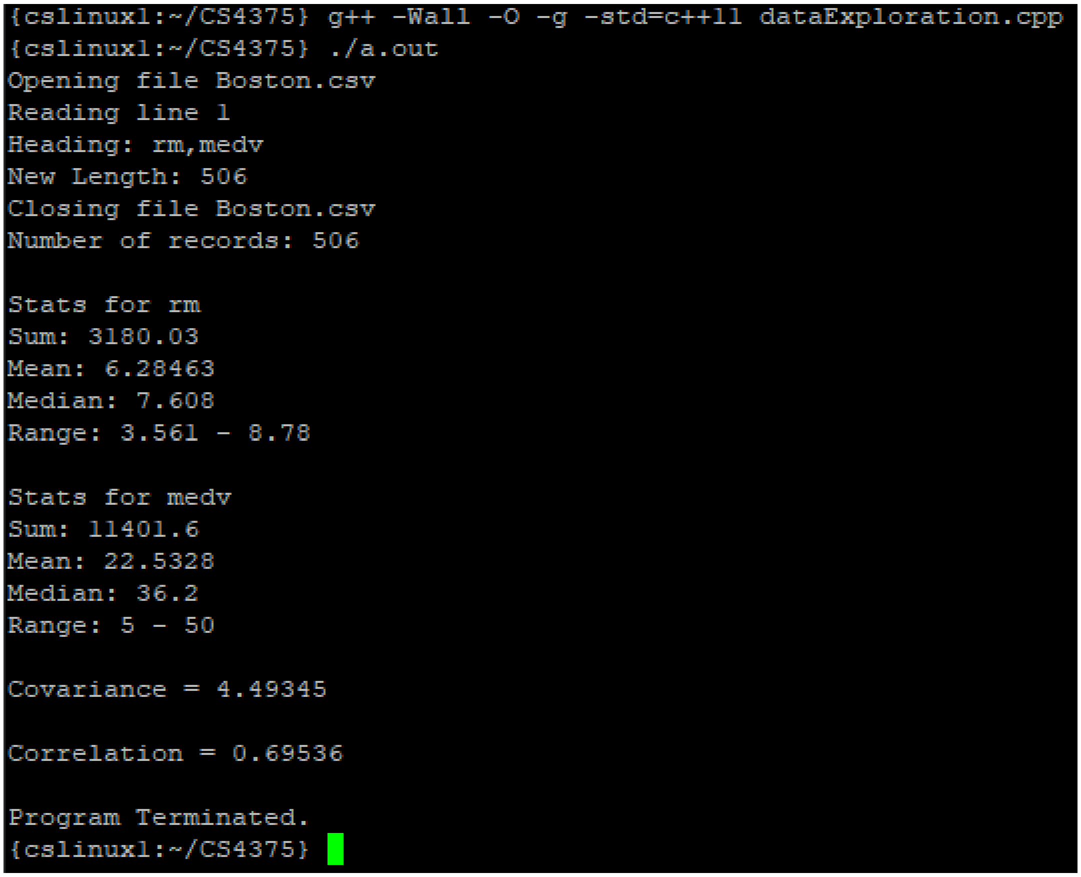
Brandon Runyon



```
{cslinux1:~/CS4375} g++ -Wall -O -g -std=c++11 dataExploration.cpp
{cslinux1:~/CS4375} ./a.out
Opening file Boston.csv
Reading 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: 7.608
Range: 3.561 - 8.78

Stats for medv
Sum: 11401.6
Mean: 22.5328
Median: 36.2
Range: 5 - 50

Covariance = 4.49345

Correlation = 0.69536

Program Terminated.
{cslinux1:~/CS4375}
```

Using built-in functions in R is easier because I don’t have to make them myself.

Mean and median are both good initial descriptions of where you expect the data to be most of the time. Though, they are imprecise and through machine learning, you can obtain better descriptors. Range tells you how wide of an input you can have to the data allowing you to adjust your algorithm appropriately.

Covariance and correlation both describe how likely the Y variable is to change in relation to the X variable. Correlation is more useful to a person because it is normalized. A correlation close to 1 or -1 is more likely to be statistically significant.