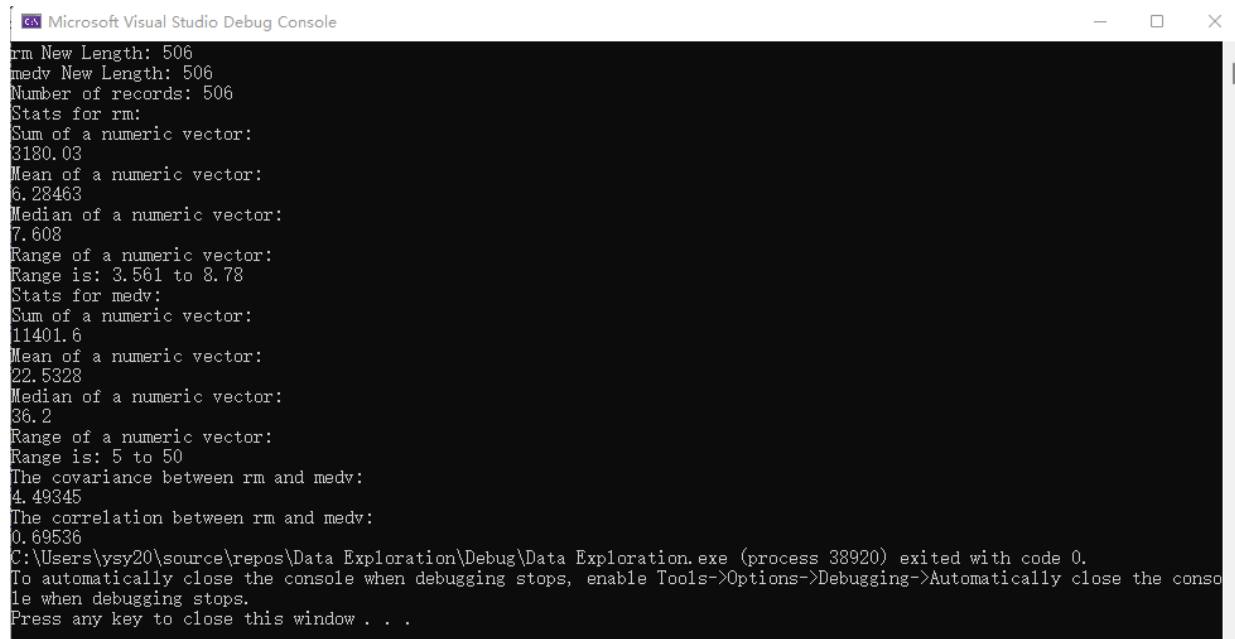


C++ Data Exploration

1. Copy/paste runs of your code showing the output.



```
Microsoft Visual Studio Debug Console
rm New Length: 506
medv New Length: 506
Number of records: 506
Stats for rm:
Sum of a numeric vector:
3180.03
Mean of a numeric vector:
6.28463
Median of a numeric vector:
7.608
Range of a numeric vector:
Range is: 3.561 to 8.78
Stats for medv:
Sum of a numeric vector:
11401.6
Mean of a numeric vector:
22.5328
Median of a numeric vector:
36.2
Range of a numeric vector:
Range is: 5 to 50
The covariance between rm and medv:
4.49345
The correlation between rm and medv:
0.69536
C:\Users\ysy20\source\repos\Data Exploration\Debug\Data Exploration.exe (process 38920) exited with code 0.
To automatically close the console when debugging stops, enable Tools->Options->Debugging->Automatically close the console when debugging stops.
Press any key to close this window . . .
```

2. Describing your experience using built-in functions in R versus coding your own functions in C++.

In R versus, it is easy to get answers for sum, average, median, correlation, and covariance from function like median(), sum(), and so on. However, in C++, I cannot get answers directly from function. I need to calculate each part of the formula and combine them to get the final answer.

3. Describes the descriptive statistical measures mean, median, and range, and how these values might be useful in data exploration prior to machine learning.
Mean in descriptive statistical is adding all figures within the data set and divide the number of figures. Median in descriptive statistical is the figure in the middle of the data set. Range in descriptive statistical is the lowest to the highest value in the data set. They are considered as different parameters in the machine learning.
4. Describe the covariance and correlation statistics, and what information they give about two attributes. How might this information be useful in machine learning?
Covariance is to measure of how much random variables vary together. It can tell us what's the relationship between variables. The large covariance means strong relationship, otherwise means weak relationship. Correlation is to indicates how strongly two variables are related. Range is -1 to 1. The figure of correlation is closer to the 1, it means that two variables have strong relevance. These information can be used in Image Recognition.