

A screenshot of the Microsoft Visual Studio Debug Console window. The window has a title bar with the Visual Studio logo and the text "Microsoft Visual Studio Debug Console". The console output shows the following text:

```
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 = 6.209
Range = 5.219

Stats for medv
Sum = 11401.6
Mean = 22.5328
Median = 21.2
Range = 45

Covariance = 4.49345
Correlation = 3.469

Program terminated.
I:\School Stuff\Seventh Semester\CS 4375.003\Homework\DataExploration\DataExploration\x64\Debug\DataExploration.exe (process 28176)
exited with code 0.
To automatically close the console when debugging stops, enable Tools->Options->Debugging->Automatically close the console when debu
gging stops.
Press any key to close this window . . .
```

In my experience, using the built-in R functions was significantly easier to use than coding my own, especially the covariance and correlation functions. Looking at the statistics was very insightful, the mean told me the average number and what I should be expecting most of the time. The median shows that the data can be skewed, maybe by having several outliers or other tricky data. The range tells us how big of a gap there is between the smallest and largest value, which can help us understand if the data is clustered together or spread out far apart.

The covariance and correlation tell us about the two data sets, the covariance demonstrates how the two variables move with each other, a positive covariance tells us the variables tend to move together, while a negative covariance tells us the variables tend to move in the opposite direction. Correlation works the same way, our correlation was 0.3, which indicates that there is little correlation between the data. This information may be useful to machine learning because it illustrates how the data relates to each other, and can help enforce which data belongs and which doesn't.