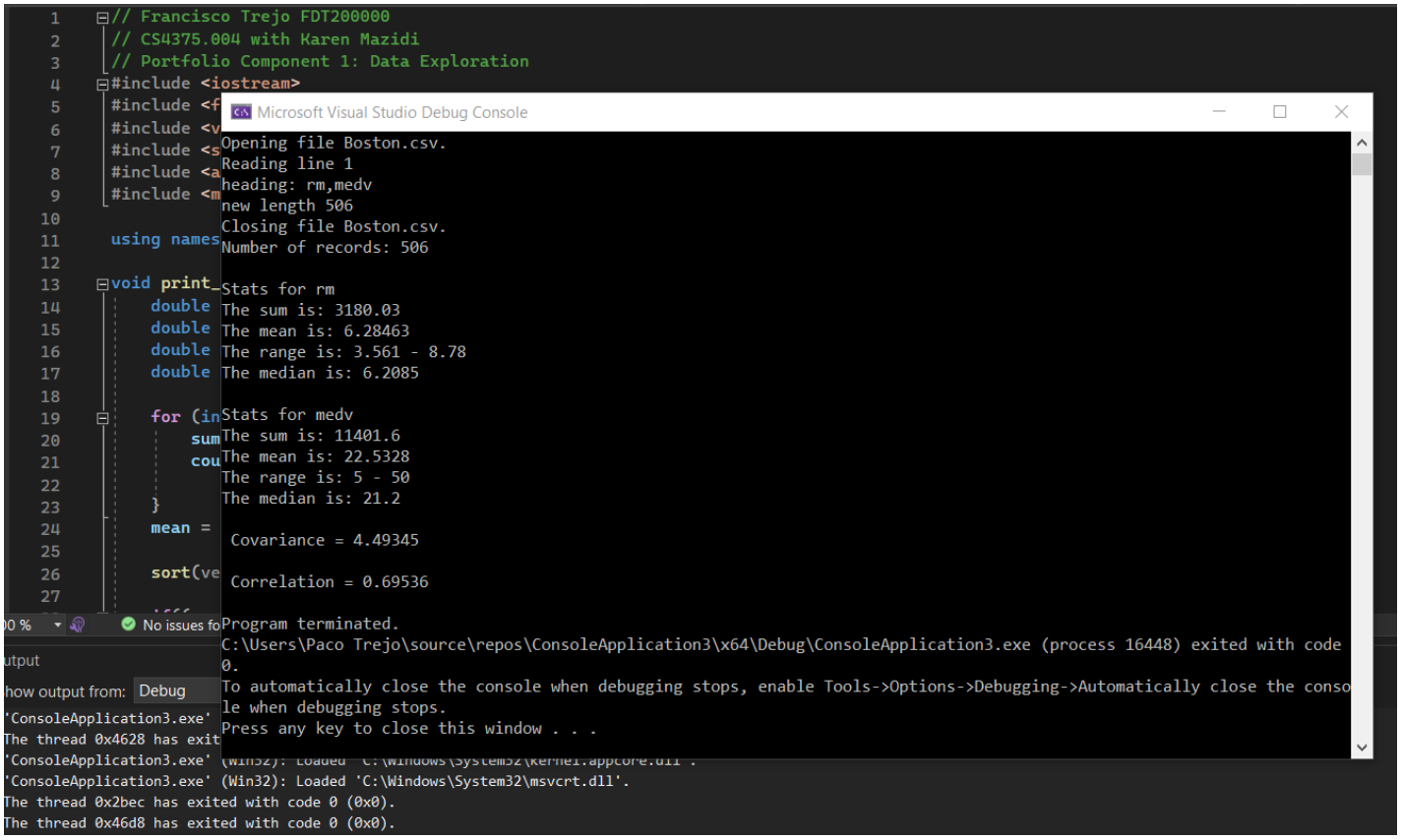
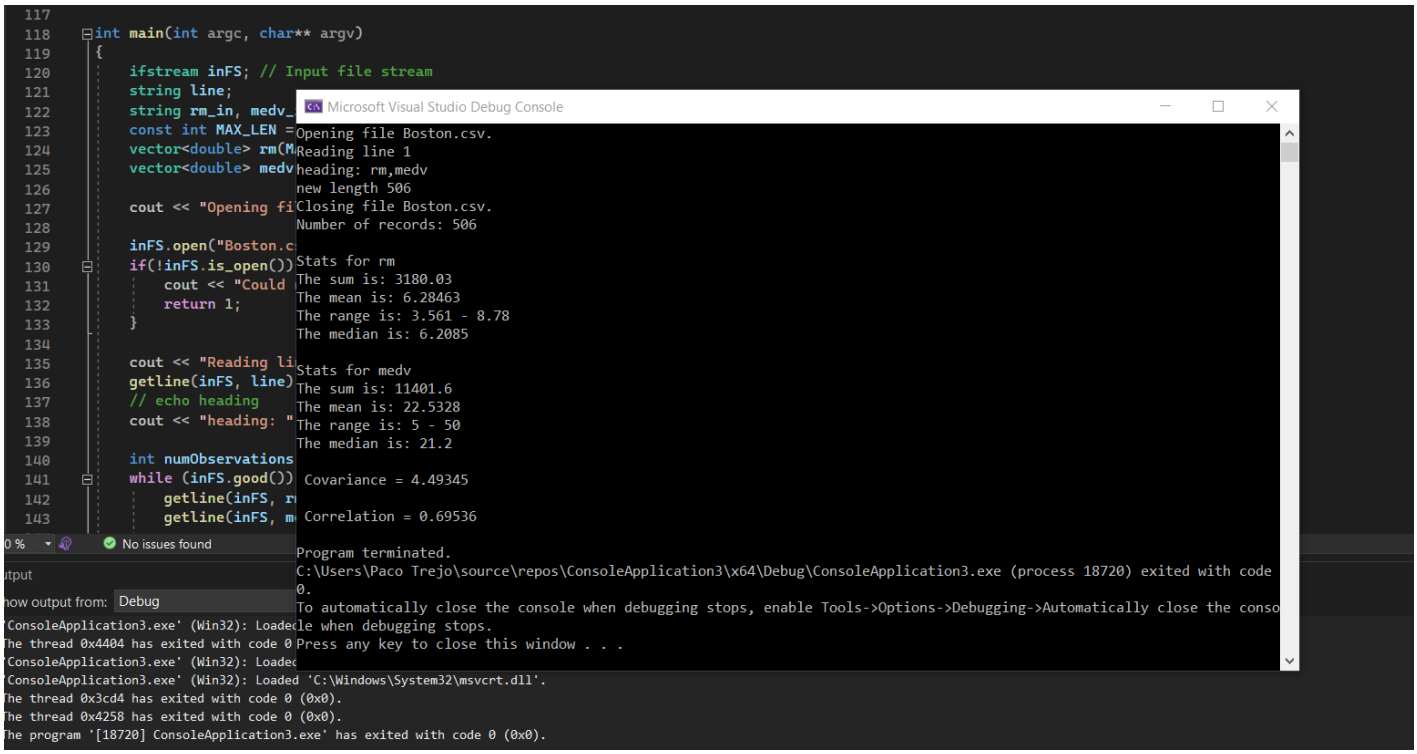
Francisco Trejo
2/4/2023
CS 4375.004
Karen Mazidi

Portfolio Component 1: Data Exploration





I feel it was definitely harder to code it on C++ versus coding it in R. In C++ you have to write several functions to accomplish the same thing a few lines of R would do. R really doesn't seem like too much coding and those built in functions are easier to use and understand.

The mean gives you the average value of the elements of the vector. The range gives you the highest and lowest values of the vector and the median value gives you the value that is right in the middle of the sorted vector. These are useful in data exploration because it pretty much gives an overview of the data. The range gives you a finite boundary on where the data limits itself. The mean and median gives you a value that will most likely come close to or come up in that set of data.

The covariance and correlation is how much two attributes are related to each other and gives a numeric representation of their relationship. Covariance gives the direction and extent of the relationship while correlation gives the strength of that relationship. These are important in

machine learning because since both find relationships between two attributes then it can help find trends and reliability of those trends to predict future data.

https://ftrejo2013.github.io/Class_Portfolio/