Emmanuel Asante

Dr. Karen Mazidi

CS 4374

1/28/2023

Data processing in R vs Traditional Programming Languages

The built in functions of R make data processing infinitely easier, Aside from the simple graphical interface and low barrier to entry, R abstracts a lot of the internal code required to make all functions work and focusses on providing developers with a set of tools to manipulate their data of choice. Common statistical measures used in data exploration include: mean (the average of the values in a data set), median (the middle value/s of a data set after sorting), and range (the minimum and maximum values in a data set). These values allows us to recognize patterns in our data and points out outliers. These figures help us make predictions as well as determine why the outlier are present. In the context of machine learning, we can use the values the gauge the accuracy of training algorithms as well as train machines to spot discrepancies easily.

Another set of statistics are covariance and correlation. Covariance refers to the relationship between two variables in an observation. The greater the number, the greater the reliance on values has on the other. As an example, height and weight typically coincide with one another. Correlation measures the rate at which two random variables move in sequence. It tracks the changes in variables over time and determines whether their paths are similar or not.