COURSERA: DEVELOPING DATA PRODUCS

DATA REPRODUCIBILITY AND QUALITY

To run the application you must have a matrix saved in .CSV. The columns must be separated by a tab. The application must be run with runApp() in R where server.R and ui.R are located. The application will display the top of the matrix by using head() and will then plot two histograms, of the first numeric column and last column for the data quality purpose. To check the reproducibility of the data, I also included a calculation of the mean and standard deviation for each of the column in our matrix. That will help to detect samples/columns that might be outliers. In the next step of the application, a correlation between samples is performed following by a hierarchical clustering algorithm. Columns that are in close proximity and have higher correlation can be further analyzed.