**Overview:**

This project is done for the Getting and Cleaning Data Coursera course. The purpose of this project is to generate a tidy data set with the average of each variable for each activity and each subject generated from the input training, testing, features and activity datasets.

**Input Data:**

The input data source is available in the following location

<https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>

The input dataset contains the following data files:

* 'features\_info.txt': Shows information about the variables used on the feature vector.
* 'features.txt': List of all features.
* 'activity\_labels.txt': Links the class labels with their activity name.
* 'train/X\_train.txt': Training set.
* 'train/y\_train.txt': Training labels.
* 'test/X\_test.txt': Test set.
* 'test/y\_test.txt': Test labels.

**Run\_Analysis R Script Logic:**

The R script run\_analysis developed as part of this project does the following:

1. Download the File
2. Unzip the file
3. Read Training Data
4. Read Testing Data
5. Read Feature vector
6. Read Activity Labels
7. Assign proper column Names to training and testing data tables from features, activity , and subject
8. Merge All Data into One data table
9. Create Mean and Standard Deviation column name vector to be used to select the mean and standard deviation from the master data table
10. Select subset from the master data table using the Mean and Standard Deviation column vector created above
11. Attach Activity name in the selected subset
12. Create tidy data set grouped by subject id and activity id
13. Write the tidy data set to “ProjectTidyDataset.txt” file

**Variables Used:**

dt\_x\_train: data table to read x\_train.txt

dt\_y\_train: data table to read y\_train.txt

dt\_subject\_train: data table to read subject\_train.txt

dt\_x\_test: data table to read x\_test.txt

dt\_y\_test: data table to read y\_test.txt

dt\_subject\_test: data table to read subject\_test.txt

dt\_features: data table to read features.txt

dt\_activity: data table to read activity\_labels.txt

dt\_merge\_train: merge all training files (x\_train,y\_train,subject\_train)

dt\_merge\_test: merge all testing files (x\_test,y\_test,subject\_test)

dt\_final: merge dt\_merge\_train and dt\_merge\_test data table into one single data table

MeanSDVector: Column vector created to filter the columns ActivityId, SubjectId, mean, and std from merged data table dt\_final

dt\_mean\_SD: subset from master data table dt\_final

dt\_tidy: tidy data set