Coursera Assignment for Getting & Cleaning Data based on Human Activity Recognition Using Smartphones Dataset

This README file explains details around what files are included and what are their features.

Original Data Source

Data for analysis is downloaded from the below URL  
<https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>

Files Included in Repository

This repo includes following files - run\_analysis.R - CodeBook.md –HumanActivitylog\_SummaryData.txt

Details of *run\_analysis.R* script

This is the script used to perform analysis on raw data to create a tidy datafile called HumanActivitylog\_SummaryData.txt

Functions of run\_analysis.R Script

* Downloads the dataset from the URL mentioned above and unzips it to create UCI HAR Dataset folder
* Imports test and train datasets and creates data frames from then and then Merges the training and the test sets to create one data frame.
* Extracts a subset of data with only the measurements on the mean mean() and standard deviation std()for each measurement. Also excludes meanFreq()-X measurements or angle measurements where the term mean exists resulting in 66 measurement variables.
* Updates the variable names in dataframe variable to improve readability
* Appropriately labels the data set with descriptive activity names in place of activity Ids
* Reshapes dataset to create a data frame with average of each measurement variable for each activity and each subject
* Writes new tidy data frame to a text file to create the required tidy data set file of 180 observations and68 columns(2 columns for ActivityDesc and SubjectId and 66 columns for measurement variables)

Running the script

To run the script, you just have to download the script and source the script from your working directory in R. source(run\_analysis.R)

Details of *CodeBook.md*

The code book file describes the variables, the data, and any transformations and work performed to clean up the data.

Details of tidy Dataset file HumanActivitylog\_SummaryData.txt

This is the tidy data file created after running run\_analysis.R script on the original data downloaded from this [URL](https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip)

This tidy dataset contains following

* 180 observations and 69 columns(2 columns for  SubjectId, ActivityId and ActivityDesc and 66 columns for measurement variables)
* Each measurement variable column is average value for each combination of SubjectId andActivityDesc