**Getting and Cleaning Data Course Project**

**Study Design:**

The Data Set generated is called Human Activity Recognition Using Smartphones

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.   
  
The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain.

**Attribute Information:**

For each record in the dataset it is provided:   
- Triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration.   
- Triaxial Angular velocity from the gyroscope.   
- A 561-feature vector with time and frequency domain variables.   
- Its activity label.   
- An identifier of the subject who carried out the experiment

The data was downloaded from: <https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>

**Code book**

1. Load activity labels and features from
   1. activity\_labels.txt
   2. features.txt

Changed activity labels and features to characters

1. Extract only the measurements on the mean and standard deviation for each measurement

Used grep to search files for the occurrence of a string of characters that matched mean and std in the features table.

Used the gsub function for to clean up the data labels, specifically replaced –mean with Mean and –std with Std

1. Load the datasets by reading in the tables of data located in
   1. subject\_train.txt
   2. x\_train.txt
   3. y\_train.txt
   4. subject\_test.txt
   5. x\_test.txt
   6. y\_test.txt
2. Merge the training and the test sets to create one data set and use descriptive activity names to name the activities in the data set
   1. First used cbind to combine the columns from X\_train.txt using trainSubject and trainActivity.
   2. Second used cbind to combine the columns from X\_test.txt using testSubject and testActivities Assign column names and merge to create one data set.
   3. The used rbind to combine the rows from the train and test data sets to make one data set
   4. Then renamed the column names
3. Change activities & subjects into factors
4. Using melt change the id to subject and activity and then use dcast to make the data frame output
5. Create a second independent tidy data set
   1. It produced a tidy data set with the average of each variable for each activity and subject
   2. The file was saved as a txt file