**Getting and Cleaning Data Project Overview**

For creating a tidy data set of wearable computing data originally from <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

**Files in this Repo**

* README.md – This is the file
* CodeBook.md -- codebook describing variables, the data and transformations
* run\_analysis.R -- actual R code

**R Script Goals: run\_analysis.R**

Objective is to create one R script called run\_analysis.R that does the following:

1. Merge the training and the test sets to create one data set.

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

3. Use descriptive activity names to name the activities in the data set

4. Appropriately label the data set with descriptive activity names.

5. Create a second, independent tidy data set with the average of each variable for each activity and each subject.

It should run in a folder of the Samsung data (the zip had this folder: UCI HAR Dataset). Please note that this script assumes it has in its working directory the following files and folders:

* **activity\_labels.txt**
* **features.txt**
* **test/**
* **train/**

The final output file called **Tidy2.txt** output is created and stored in the designated working directory.

*Note:* the R script is built to run without including any libraries for the purpose of this course.

**Run\_analysis.R Step-wise Procedure**

The R script executes the following objectives in a methodical and sequential manner.

* **Step 1:**
  + Read all the test and training files: y\_test.txt, subject\_test.txt and X\_test.txt.
  + Combine the files to a data frame in the form of subjects, labels, the rest of the data.
* **Step 2:**
  + Read the features from features.txt and filter it to only leave features that are either means ("mean()") or standard deviations ("std()"). The reason for leaving out meanFreq() is that the goal for this step is to only include means and standard deviations of measurements, of which meanFreq() is neither.
  + A new data frame is then created that includes subjects, labels, and the described features.
* **Step 3:**
  + Read the activity labels from activity\_labels.txt and replace the numbers with the text.
* **Step 4:**
  + Make a column list (including "subjects" and "label" at the start)
  + Tidy the list by removing all non-alphanumeric characters and converting the result to lowercase
  + Apply the now-good-column names to the data frame
* **Step 5:**
  + Create a new data frame by finding the mean for each combination of subject and label. It's done by aggregate() function
* **Step 6:**
  + Write the new tidy set into a text file called tidy2.txt, formatted similarly to the original files.