# Codebook for assignment of Course "Getting and Cleaning Data", John Hopkins University Data Science Program

## Data Set Information

The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years[1]. 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.

## Analysis Process

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

For each record 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 cleansing is done by the following process:

* Load the labels and features, strip off invalid and unused characters
* Read measurements which are std or means
* Read training data and combine with their subjects
* Read test data and combine with their subjects
* Combine training and test datasets
* Calculate the average of each acvitivity and each subject
* OUput the dataset into "cleanData.txt"

## The variables in output file

The output data file contains the following columns:

* SubjectNum,
* ActivityName,
* tBodyAcc-mean-X,
* tBodyAcc-mean-Y,
* tBodyAcc-mean-Z,
* tBodyAcc-std-X,
* tBodyAcc-std-Y,
* tBodyAcc-std-Z,
* tGravityAcc-mean-X,
* tGravityAcc-mean-Y,
* tGravityAcc-mean-Z,
* tGravityAcc-std-X,
* tGravityAcc-std-Y,
* tGravityAcc-std-Z,
* tBodyAccJerk-mean-X,
* tBodyAccJerk-mean-Y,
* tBodyAccJerk-mean-Z,
* tBodyAccJerk-std-X,
* tBodyAccJerk-std-Y,
* tBodyAccJerk-std-Z,
* tBodyGyro-mean-X,
* tBodyGyro-mean-Y,
* tBodyGyro-mean-Z,
* tBodyGyro-std-X,
* tBodyGyro-std-Y,
* tBodyGyro-std-Z,
* tBodyGyroJerk-mean-X,
* tBodyGyroJerk-mean-Y,
* tBodyGyroJerk-mean-Z,
* tBodyGyroJerk-std-X,
* tBodyGyroJerk-std-Y,
* tBodyGyroJerk-std-Z,
* tBodyAccMag-mean,
* tBodyAccMag-std,
* tGravityAccMag-mean,
* tGravityAccMag-std,
* tBodyAccJerkMag-mean,
* tBodyAccJerkMag-std,
* tBodyGyroMag-mean,
* tBodyGyroMag-std,
* tBodyGyroJerkMag-mean,
* tBodyGyroJerkMag-std,
* fBodyAcc-mean-X,
* fBodyAcc-mean-Y,
* fBodyAcc-mean-Z,
* fBodyAcc-std-X,
* fBodyAcc-std-Y,
* fBodyAcc-std-Z,
* fBodyAccJerk-mean-X,
* fBodyAccJerk-mean-Y,
* fBodyAccJerk-mean-Z,
* fBodyAccJerk-std-X,
* fBodyAccJerk-std-Y,
* fBodyAccJerk-std-Z,
* fBodyGyro-mean-X,
* fBodyGyro-mean-Y,
* fBodyGyro-mean-Z,
* fBodyGyro-std-X,
* fBodyGyro-std-Y,
* fBodyGyro-std-Z,
* fBodyAccMag-mean,
* fBodyAccMag-std,
* fBodyBodyAccJerkMag-mean,
* fBodyBodyAccJerkMag-std,
* fBodyBodyGyroMag-mean,
* fBodyBodyGyroMag-std,
* fBodyBodyGyroJerkMag-mean,
* fBodyBodyGyroJerkMag-std

## Reference

[1]<https://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>