**CodeBook for preparing Tidy Data from the given UCI HAR Dataset**

**Introduction**

This CodeBook describes the data contained in the output of the Run\_analysis.R script contained in this repository. The tidy flat text file can be read using data.table to create a data table for further analysis.

tidy\_data<-data\_table(“tidy.txt”)

Use of this script creates a tidy, condensed version of the UCIs dataset for Human Activity Recognition (HAR) using smartphones which can be used for further research and analysis.

The original UCI HAR Dataset is a public domain dataset built from the recordings of subjects performing activities of daily living (ADL) while carrying a waist-mounted smartphone with embedded inertial sensor.

(see<http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>).

As per the above referenced document, the original data comes from experiments that were 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, and laying) wearing a Samsung Galaxy S II smartphone on the waist. Using its embedded accelerometer and gyroscope, 3-axial linear acceleration and 3-axial angular velocity were captured at a constant rate of 50Hz. The experiments were video-recorded to label the data manually.

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 was assumed to have only low frequency components, so a filter with a 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.

The script given in the file generates a combined subset of the original data by extracting the mean and standard deviation features for each of the 33 processed signals, for a total of 66 features (out of the 561 available features from the original feature vector). This combined subset contains 10299 observations of 68 variables, with activity and subject appended to the 66 features.

The combined subset is further reduced by calculating the mean of the observations by activity and subject pair to generate 180 observations (6 activities \* 30 subjects) of the same 68 variables. This dataset is tidied to generate a narrow and lean dataset containing 11880 observations with 4 variables each and is saved as a text file in the current working directory with the name  hgt.Tidy.txt

**Variable name cleanup**

During the *tidying* process the variable names are cleaned up using the following transformations.

filtered\_feature\_names <- gsub("\\(\\)", "", filtered\_feature\_names)

filtered\_feature\_names <- gsub("Acc", "-acceleration", filtered\_feature\_names)

filtered\_feature\_names <- gsub("Mag", "-Magnitude", filtered\_feature\_names)

filtered\_feature\_names <- gsub("^t(.\*)$", "\\1-time", filtered\_feature\_names)

filtered\_feature\_names <- gsub("^f(.\*)$", "\\1-frequency", filtered\_feature\_names)

filtered\_feature\_names <- gsub("(Jerk|Gyro)", "-\\1", filtered\_feature\_names)

filtered\_feature\_names <- gsub("BodyBody", "Body", filtered\_feature\_names)

And finally all variable names are converted into lowercase.

**Description of the data variables**

The Tidy dataset consists of 11880 observations summarized by activity (6 categories) and subject (30 volunteers) pairs. For each observation (row) in the Tidy dataset, the following 4 columns are provided:

* subject
* activity
* measurement
* mean

**subject**

A numeric identifier (1-30) of the subject who carried out the experiment.

**activity**

The activity name with the following possible values.

* 'laying',
* 'sitting',
* 'standing',
* 'walking',
* 'walking\_downstairs'
* 'walking\_upstairs'

**measurement**

The name of the measurement for which the mean is calculated. This variable can contain one of the following 66 variables. Please refer the codebook with the original dataset for the explanation of these different variables.

* body-acceleration-jerk-magnitude-mean-frequency
* body-acceleration-jerk-magnitude-mean-time
* body-acceleration-jerk-magnitude-std-frequency
* body-acceleration-jerk-magnitude-std-time
* body-acceleration-jerk-mean-x-frequency
* body-acceleration-jerk-mean-x-time
* body-acceleration-jerk-mean-y-frequency
* body-acceleration-jerk-mean-y-time
* body-acceleration-jerk-mean-z-frequency
* body-acceleration-jerk-mean-z-time
* body-acceleration-jerk-std-x-frequency
* body-acceleration-jerk-std-x-time
* body-acceleration-jerk-std-y-frequency
* body-acceleration-jerk-std-y-time
* body-acceleration-jerk-std-z-frequency
* body-acceleration-jerk-std-z-time
* body-acceleration-magnitude-mean-frequency
* body-acceleration-magnitude-mean-time
* body-acceleration-magnitude-std-frequency
* body-acceleration-magnitude-std-time
* body-acceleration-mean-x-frequency
* body-acceleration-mean-x-time
* body-acceleration-mean-y-frequency
* body-acceleration-mean-y-time
* body-acceleration-mean-z-frequency
* body-acceleration-mean-z-time
* body-acceleration-std-x-frequency
* body-acceleration-std-x-time
* body-acceleration-std-y-frequency
* body-acceleration-std-y-time
* body-acceleration-std-z-frequency
* body-acceleration-std-z-time
* body-gyro-jerk-magnitude-mean-frequency
* body-gyro-jerk-magnitude-mean-time
* body-gyro-jerk-magnitude-std-frequency
* body-gyro-jerk-magnitude-std-time
* body-gyro-jerk-mean-x-time
* body-gyro-jerk-mean-y-time
* body-gyro-jerk-mean-z-time
* body-gyro-jerk-std-x-time
* body-gyro-jerk-std-y-time
* body-gyro-jerk-std-z-time
* body-gyro-magnitude-mean-frequency
* body-gyro-magnitude-mean-time
* body-gyro-magnitude-std-frequency
* body-gyro-magnitude-std-time
* body-gyro-mean-x-frequency
* body-gyro-mean-x-time
* body-gyro-mean-y-frequency
* body-gyro-mean-y-time
* body-gyro-mean-z-frequency
* body-gyro-mean-z-time
* body-gyro-std-x-frequency
* body-gyro-std-x-time
* body-gyro-std-y-frequency
* body-gyro-std-y-time
* body-gyro-std-z-frequency
* body-gyro-std-z-time
* gravity-acceleration-magnitude-mean-time
* gravity-acceleration-magnitude-std-time
* gravity-acceleration-mean-x-time
* gravity-acceleration-mean-y-time
* gravity-acceleration-mean-z-time
* gravity-acceleration-std-x-time
* gravity-acceleration-std-y-time
* gravity-acceleration-std-z-time

**mean**

The mean of the measurements.

**Codebook (table format)**

| **variable** | **description** | **range** |
| --- | --- | --- |
| subject | ID of the subject. | string in {walking, walking\_upstairs, walking\_downstairs, sitting, standing, laying} |
| activity | Activity type performed. | integer in [1, 30] |
| tBodyAcc-mean()-X | Mean of means of accelerometer body data in time domain in the X direction. | float in [-1,1] |
| tBodyAcc-mean()-Y | Mean of means of accelerometer body data in time domain in the Y direction. | float in [-1,1] |
| tBodyAcc-mean()-Z | Mean of means of accelerometer body data in time domain in the Z direction. | float in [-1,1] |
| tBodyAcc-std()-X | Mean of standard deviations of accelerometer body data in time domain in the X direction. | float in [-1,1] |
| tBodyAcc-std()-Y | Mean of standard deviations of accelerometer body data in time domain in the Y direction. | float in [-1,1] |
| tBodyAcc-std()-Z | Mean of standard deviations of accelerometer body data in time domain in the Z direction. | float in [-1,1] |
| tGravityAcc-mean()-X | Mean of means of accelerometer gravity data in time domain in the X direction. | float in [-1,1] |
| tGravityAcc-mean()-Y | Mean of means of accelerometer gravity data in time domain in the Y direction. | float in [-1,1] |
| tGravityAcc-mean()-Z | Mean of means of accelerometer gravity data in time domain in the Z direction. | float in [-1,1] |
| tGravityAcc-std()-X | Mean of standard deviations of accelerometer gravity data in time domain in the X direction. | float in [-1,1] |
| tGravityAcc-std()-Y | Mean of standard deviations of accelerometer gravity data in time domain in the Y direction. | float in [-1,1] |
| tGravityAcc-std()-Z | Mean of standard deviations of accelerometer gravity data in time domain in the Z direction. | float in [-1,1] |
| tBodyAccJerk-mean()-X | Mean of means of jerk signal of accelerometer body data in time domain in the X direction. | float in [-1,1] |
| tBodyAccJerk-mean()-Y | Mean of means of jerk signal of accelerometer body data in time domain in the Y direction. | float in [-1,1] |
| tBodyAccJerk-mean()-Z | Mean of means of jerk signal of accelerometer body data in time domain in the Z direction. | float in [-1,1] |
| tBodyAccJerk-std()-X | Mean of standard deviations of jerk signal of accelerometer body data in time domain in the X direction. | float in [-1,1] |
| tBodyAccJerk-std()-Y | Mean of standard deviations of jerk signal of accelerometer body data in time domain in the Y direction. | float in [-1,1] |
| tBodyAccJerk-std()-Z | Mean of standard deviations of jerk signal of accelerometer body data in time domain in the Z direction. | float in [-1,1] |
| tBodyGyro-mean()-X | Mean of means of giroscope body data in time domain in the X direction. | float in [-1,1] |
| tBodyGyro-mean()-Y | Mean of means of giroscope body data in time domain in the Y direction. | float in [-1,1] |
| tBodyGyro-mean()-Z | Mean of means of giroscope body data in time domain in the Z direction. | float in [-1,1] |
| tBodyGyro-std()-X | Mean of standard deviations of giroscope body data in time domain in the X direction. | float in [-1,1] |
| tBodyGyro-std()-Y | Mean of standard deviations of giroscope body data in time domain in the Y direction. | float in [-1,1] |
| tBodyGyro-std()-Z | Mean of standard deviations of giroscope body data in time domain in the Z direction. | float in [-1,1] |
| tBodyGyroJerk-mean()-X | Mean of means of jerk signal of giroscope body data in time domain in the X direction. | float in [-1,1] |
| tBodyGyroJerk-mean()-Y | Mean of means of jerk signal of giroscope body data in time domain in the Y direction. | float in [-1,1] |
| tBodyGyroJerk-mean()-Z | Mean of means of jerk signal of giroscope body data in time domain in the Z direction. | float in [-1,1] |
| tBodyGyroJerk-std()-X | Mean of standard deviations of jerk signal of giroscope body data in time domain in the X direction. | float in [-1,1] |
| tBodyGyroJerk-std()-Y | Mean of standard deviations of jerk signal of giroscope body data in time domain in the Y direction. | float in [-1,1] |
| tBodyGyroJerk-std()-Z | Mean of standard deviations of jerk signal of giroscope body data in time domain in the Z direction. | float in [-1,1] |
| tBodyAccMag-mean() | Mean of means of the magnitude of accelerometer body data in time domain. | float in [-1,1] |
| tBodyAccMag-std() | Mean of standard deviations of the magnitude of accelerometer body data in time domain. | float in [-1,1] |
| tGravityAccMag-mean() | Mean of means of the magnitude of accelerometer gravity data in time domain. | float in [-1,1] |
| tGravityAccMag-std() | Mean of standard deviations of the magnitude of accelerometer gravity data in time domain. | float in [-1,1] |
| tBodyAccJerkMag-mean() | Mean of means of the magnitude of jerk signal of accelerometer body data in time domain. | float in [-1,1] |
| tBodyAccJerkMag-std() | Mean of standard deviations of the magnitude of jerk signal of accelerometer body data in time domain. | float in [-1,1] |
| tBodyGyroMag-mean() | Mean of means of the magnitude of giroscope body data in time domain. | float in [-1,1] |
| tBodyGyroMag-std() | Mean of standard deviations of the magnitude of giroscope body data in time domain. | float in [-1,1] |
| tBodyGyroJerkMag-mean() | Mean of means of the magnitude of jerk signal of giroscope body data in time domain. | float in [-1,1] |
| tBodyGyroJerkMag-std() | Mean of standard deviations of the magnitude of jerk signal of giroscope body data in time domain. | float in [-1,1] |
| fBodyAcc-mean()-X | Mean of means of accelerometer body data in frequency domain in the X direction. | float in [-1,1] |
| fBodyAcc-mean()-Y | Mean of means of accelerometer body data in frequency domain in the Y direction. | float in [-1,1] |
| fBodyAcc-mean()-Z | Mean of means of accelerometer body data in frequency domain in the Z direction. | float in [-1,1] |
| fBodyAcc-std()-X | Mean of standard deviations of accelerometer body data in frequency domain in the X direction. | float in [-1,1] |
| fBodyAcc-std()-Y | Mean of standard deviations of accelerometer body data in frequency domain in the Y direction. | float in [-1,1] |
| fBodyAcc-std()-Z | Mean of standard deviations of accelerometer body data in frequency domain in the Z direction. | float in [-1,1] |
| fBodyAccJerk-mean()-X | Mean of means of jerk signal of accelerometer body data in frequency domain in the X direction. | float in [-1,1] |
| fBodyAccJerk-mean()-Y | Mean of means of jerk signal of accelerometer body data in frequency domain in the Y direction. | float in [-1,1] |
| fBodyAccJerk-mean()-Z | Mean of means of jerk signal of accelerometer body data in frequency domain in the Z direction. | float in [-1,1] |
| fBodyAccJerk-std()-X | Mean of standard deviations of jerk signal of accelerometer body data in frequency domain in the X direction. | float in [-1,1] |
| fBodyAccJerk-std()-Y | Mean of standard deviations of jerk signal of accelerometer body data in frequency domain in the Y direction. | float in [-1,1] |
| fBodyAccJerk-std()-Z | Mean of standard deviations of jerk signal of accelerometer body data in frequency domain in the Z direction. | float in [-1,1] |
| fBodyGyro-mean()-X | Mean of means of giroscope body data in frequency domain in the X direction. | float in [-1,1] |
| fBodyGyro-mean()-Y | Mean of means of giroscope body data in frequency domain in the Y direction. | float in [-1,1] |
| fBodyGyro-mean()-Z | Mean of means of giroscope body data in frequency domain in the Z direction. | float in [-1,1] |
| fBodyGyro-std()-X | Mean of standard deviations of giroscope body data in frequency domain in the X direction. | float in [-1,1] |
| fBodyGyro-std()-Y | Mean of standard deviations of giroscope body data in frequency domain in the Y direction. | float in [-1,1] |
| fBodyGyro-std()-Z | Mean of standard deviations of giroscope body data in frequency domain in the Z direction. | float in [-1,1] |
| fBodyAccMag-mean() | Mean of means of the magnitude of accelerometer body data in frequency domain. | float in [-1,1] |
| fBodyAccMag-std() | Mean of standard deviations of the magnitude of accelerometer body data in frequency domain. | float in [-1,1] |
| fBodyBodyAccJerkMag-mean() | Mean of means of the magnitude of jerk signal of accelerometer body data in frequency domain. | float in [-1,1] |
| fBodyBodyAccJerkMag-std() | Mean of standard deviations of the magnitude of jerk signal of accelerometer body data in frequency domain. | float in [-1,1] |
| fBodyBodyGyroMag-mean() | Mean of means of the magnitude of giroscope body data in frequency domain. | float in [-1,1] |
| fBodyBodyGyroMag-std() | Mean of standard deviations of the magnitude of giroscope body data in frequency domain. | float in [-1,1] |
| fBodyBodyGyroJerkMag-mean() | Mean of means of the magnitude of jerk signal of giroscope body data in frequency domain. | float in [-1,1] |
| fBodyBodyGyroJerkMag-std() | Mean of standard deviations of the magnitude of jerk signal of giroscope body data in frequency domain. | float in [-1,1] |