# Code Book

# Data

The raw data needed for creation of tidy data is located in the following file;

getdata-projectfiles-UCI HAR Dataset.zip.

The content of the zip file is as follows

**Top level**

**3 files**

**activity\_labels**

This files contains labels for the 6 activities tracked as part of the experiment. Each activity also has a code associated with it that is used in the test and train data files. As part of the tidy data activities, the Labels wil replace the codes in the tidy data set .

**features.txt**

A list of all the variables available in raw data files. The tidy data set makes use of a small subset of data detailed in the code book section

**features\_info.txt**

Information from the authors of the original study

**2 directories**

**test – contents of test directory**

**subject\_test\_txt**

A vector of integers (1-30) that represent one of the 30 subjects that are part of the experiment

**X-test.txt**

Data of observations (features) for each activity and subject.

**Y\_test.txt**

Subset of activities used in test observations

**Inertial Signals** - Not used explicitly in creation of tidy data set

**train– contents of test directory**

**subject\_train\_txt**

A vector of integers (1-30) that represent one of the 30 subjects that are part of the experiment

**X-train.txt**

Data of observations (features) for each activity and subject.

**Y\_train.txt**

Subset of activities used in train observations

Inertial Signals - Not used explicitly in creation of tidy data set

# Info on variables and units

The tidy data set is s subset of all the tracked variables. Specifically only the mean and standard deviation measurements where captured in the tidy data set.

The features selected for this database come from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ. These time domain signals (prefix 't' to denote time) were captured at a constant rate of 50 Hz. Then they were filtered using a median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz to remove noise. Similarly, the acceleration signal was then separated into body and gravity acceleration signals (tBodyAcc-XYZ and tGravityAcc-XYZ) using another low pass Butterworth filter with a corner frequency of 0.3 Hz.

Subsequently, the body linear acceleration and angular velocity were derived in time to obtain Jerk signals (tBodyAccJerk-XYZ and tBodyGyroJerk-XYZ). Also the magnitude of these three-dimensional signals were calculated using the Euclidean norm (tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag).

Finally a Fast Fourier Transform (FFT) was applied to some of these signals producing fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, fBodyGyroJerkMag. (Note the 'f' to indicate frequency domain signals).

**Columns in tidy data set**

|  |  |  |  |
| --- | --- | --- | --- |
| Activities | Subject | 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 |
| fBodyAcc-meanFreq-X | fBodyAcc-meanFreq-Y | fBodyAcc-meanFreq-Z | fBodyAccJerk-mean-X |
| fBodyAccJerk-mean-Y | fBodyAccJerk-mean-Z | fBodyAccJerk-std-X | fBodyAccJerk-std-Y |
| fBodyAccJerk-std-Z | fBodyAccJerk-meanFreq-X | fBodyAccJerk-meanFreq-Y | fBodyAccJerk-meanFreq-Z |
| fBodyGyro-mean-X | fBodyGyro-mean-Y | fBodyGyro-mean-Z | fBodyGyro-std-X |
| fBodyGyro-std-Y | fBodyGyro-std-Z | fBodyGyro-meanFreq-X | fBodyGyro-meanFreq-Y |
| fBodyGyro-meanFreq-Z | fBodyAccMag-mean | fBodyAccMag-std | fBodyAccMag-meanFreq |
| fBodyBodyAccJerkMag-mean | fBodyBodyAccJerkMag-std | fBodyBodyAccJerkMag-meanFreq | fBodyBodyGyroMag-mean |
| fBodyBodyGyroMag-std | fBodyBodyGyroMag-meanFreq | fBodyBodyGyroJerkMag-mean | fBodyBodyGyroJerkMag-std |
| fBodyBodyGyroJerkMag-meanFreq | angletBodyAccMeangravity | angletBodyAccJerkMeangravityMean) | angletBodyGyroMeangravityMean |
| angletBodyGyroJerkMeangravityMean | angleXgravityMean | angleYgravityMean | angleZgravityMean |

**Info on summary choices**

The tidy data set summarization is as follows

For each unique Subject, Activity, and Variable combination we calculate the mean of all the observations

# Experimental Study Design

Please reference <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

# Instruction list

1. The following files need to be in your working directory.
   1. test directory
   2. train directory
   3. activity\_labels.txt
   4. features.txt
   5. run\_analysis.R
2. In an R environment run the run\_analysis.R file
3. The output of this script is a tidy data set in txt format named meanBySubjectAndActivity.txt
4. This file can be read directly into excel or text editor for further analysis
5. In Excel. Use “Text to Columns” to see properly formatted text
   1. A csv version of file is also provided
6. Specific steps of automated script are commented in source file