**Code Book**

**Overview**

This code book explains about the original dataset and the tidy data set populated using the steps as mentioned in run\_analysis.R program. The tidy dataset summarizes a subset of raw data after processing the original data and filtering out the unused data from the original dataset.The raw data was sourced at the UCI Machine Learning Repository found here: [Human Activity Recognition Using Smartphones Data Set](http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones).

This documentation also lists all the variables in the original dataset as listed in features\_info.txt and the variables populated in tidy data set for further analysis after appropriate filtrations.

The original data was populated from the following URL

<https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>

As per the following website, the analysis was performed on Human Activity Recognition Using Smartphones Data Set. <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

**Data Set Information: (This is the extract from the same website)**

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.

As per the same website the Attribute Information provided as follows

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.

**Original Data**

This section is extracted from features\_info.txt of the extracted dataset.

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).

These signals were used to estimate variables of the feature vector for each pattern:

'-XYZ' is used to denote 3-axial signals in the X, Y and Z directions.

tBodyAcc-XYZ

tGravityAcc-XYZ

tBodyAccJerk-XYZ

tBodyGyro-XYZ

tBodyGyroJerk-XYZ

tBodyAccMag

tGravityAccMag

tBodyAccJerkMag

tBodyGyroMag

tBodyGyroJerkMag

fBodyAcc-XYZ

fBodyAccJerk-XYZ

fBodyGyro-XYZ

fBodyAccMag

fBodyAccJerkMag

fBodyGyroMag

fBodyGyroJerkMag

The set of variables that were estimated from these signals are:

mean(): Mean value

std(): Standard deviation

mad(): Median absolute deviation

max(): Largest value in array

min(): Smallest value in array

sma(): Signal magnitude area

energy(): Energy measure. Sum of the squares divided by the number of values.

iqr(): Interquartile range

entropy(): Signal entropy

arCoeff(): Autorregresion coefficients with Burg order equal to 4

correlation(): correlation coefficient between two signals

maxInds(): index of the frequency component with largest magnitude

meanFreq(): Weighted average of the frequency components to obtain a mean frequency

skewness(): skewness of the frequency domain signal

kurtosis(): kurtosis of the frequency domain signal

bandsEnergy(): Energy of a frequency interval within the 64 bins of the FFT of each window.

angle(): Angle between to vectors.

Additional vectors obtained by averaging the signals in a signal window sample. These are used on the angle() variable:

gravityMean

tBodyAccMean

tBodyAccJerkMean

tBodyGyroMean

tBodyGyroJerkMean

The complete list of variables of each feature vector is available in 'features.txt'

**Tidy Data for further Analysis**

The tidy data tasks performed using the training data, test data and corresponding subject data, that are extracted and merged with the activity. There are 30 subjects and 6 (WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING) activities considered for populating the Tidy data. In order to create the Tidy Data, the mean and standard deviation columns are filtered out from the original dataset. The Average of mean and standard deviation column of each variable for each activity and each subject are taken.

The attribute list of the Tidy Data set are listed as follows

|  |
| --- |
| [1] "ActivityNameGroup" "SubjectGroup"  [3] "TimeBodyAccmeanX" "TimeBodyAccmeanY"  [5] "TimeBodyAccmeanZ" "TimeBodyAccSTDX"  [7] "TimeBodyAccSTDY" "TimeBodyAccSTDZ"  [9] "TimeGravityAccmeanX" "TimeGravityAccmeanY"  [11] "TimeGravityAccmeanZ" "TimeGravityAccSTDX"  [13] "TimeGravityAccSTDY" "TimeGravityAccSTDZ"  [15] "TimeBodyAccJerkmeanX" "TimeBodyAccJerkmeanY"  [17] "TimeBodyAccJerkmeanZ" "TimeBodyAccJerkSTDX"  [19] "TimeBodyAccJerkSTDY" "TimeBodyAccJerkSTDZ"  [21] "TimeBodyGyromeanX" "TimeBodyGyromeanY"  [23] "TimeBodyGyromeanZ" "TimeBodyGyroSTDX"  [25] "TimeBodyGyroSTDY" "TimeBodyGyroSTDZ"  [27] "TimeBodyGyroJerkmeanX" "TimeBodyGyroJerkmeanY"  [29] "TimeBodyGyroJerkmeanZ" "TimeBodyGyroJerkSTDX"  [31] "TimeBodyGyroJerkSTDY" "TimeBodyGyroJerkSTDZ"  [33] "TimeBodyAccMagmean" "TimeBodyAccMagSTD"  [35] "TimeGravityAccMagmean" "TimeGravityAccMagSTD"  [37] "TimeBodyAccJerkMagmean" "TimeBodyAccJerkMagSTD"  [39] "TimeBodyGyroMagmean" "TimeBodyGyroMagSTD"  [41] "TimeBodyGyroJerkMagmean" "TimeBodyGyroJerkMagSTD"  [43] "FreqncyBodyAccmeanX" "FreqncyBodyAccmeanY"  [45] "FreqncyBodyAccmeanZ" "FreqncyBodyAccSTDX"  [47] "FreqncyBodyAccSTDY" "FreqncyBodyAccSTDZ"  [49] "FreqncyBodyAccJerkmeanX" "FreqncyBodyAccJerkmeanY"  [51] "FreqncyBodyAccJerkmeanZ" "FreqncyBodyAccJerkSTDX"  [53] "FreqncyBodyAccJerkSTDY" "FreqncyBodyAccJerkSTDZ"  [55] "FreqncyBodyGyromeanX" "FreqncyBodyGyromeanY"  [57] "FreqncyBodyGyromeanZ" "FreqncyBodyGyroSTDX"  [59] "FreqncyBodyGyroSTDY" "FreqncyBodyGyroSTDZ"  [61] "FreqncyBodyAccMagmean" "FreqncyBodyAccMagSTD"  [63] "FreqncyBodyBodyAccJerkMagmean" "FreqncyBodyBodyAccJerkMagSTD"  [65] "FreqncyBodyBodyGyroMagmean" "FreqncyBodyBodyGyroMagSTD"  [67] "FreqncyBodyBodyGyroJerkMagmean" "FreqncyBodyBodyGyroJerkMagSTD" |
|  |
| |  | | --- | | > | |