INTRODUCTION:

This project takes key files from the dataset available at <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>[[1]](#footnote-1) pulls them together into a single data set and produces a final, tidy data set with the average of the mean and standard deviation measurements by subject and activity.

According to the README file that accompanies the data, “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 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. See 'features\_info.txt' for more details.”[[2]](#footnote-2)

The set of observations were divided into two groups as follows: “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.”[[3]](#footnote-3)

FILES USED:

The files used from this dataset[[4]](#footnote-4) (see footnote for link) are:

| File Name | File Description |
| --- | --- |
| x\_train.txt | Set of all observations (see variables in next section) for the training group. |
| x\_test.txt | Set of all observations (see variables in next section) for the test group. |
| y\_train.txt | The activity number (see “activity\_labels.txt”) that corresponds to the variables in the x\_train.txt file. |
| y\_test.txt | The activity number (see “activity\_labels.txt”) that corresponds to the variables in the x\_test.txt file) |
| activity\_labels.txt | References the numbered activity label with its text format: WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING |
| features.txt | Provides the column names for x\_train.txt and x\_test.txt based on the signal type, direction and measurement. (See “Variables” section for more information.) |
| subject\_train.txt | The number of the subject (within the range 1:30) for the observations in the x\_train.txt file. |
| subject\_test.txt | The number of the subject (within the range 1:30) for the observations in the x\_test.txt file. |

VARIABLES:

Each variable is summarized by its average by subject and activity, except for the subject and activity variables, to produce a record set of 180 rows/observations (30 subjects x 6 activities). The original data set had a total of 10,299 observations across the subjects and activities)

| Variable Name | Data Type | Variable Label |
| --- | --- | --- |
| subject | factor | This is a number from 1:30 which represents the person who completed the activities. It was obtained from the subject\_train.txt and subject\_test.txt and then combined to the x\_train.txt and x\_test.txt files, respectively to match the subject to the observations. This factor can be cast to an integer. |
| activity | factor | This variable has one of the six possible activities: WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING (sic). The list of activity names was obtained from the activity\_labels.txt file. |
| AvgTimeBodyAcc-mean-X | numeric | These variables represent the calculated average by subject and activity of the mean of the body acceleration in each of the 3-axial directions (X, Y, Z) |
| AvgTimeBodyAcc-mean-Y |
| AvgTimeBodyAcc-mean-Z |
| AvgTimeBodyAcc-std-X | numeric | These variables represent the calculated average by subject and activity of the standard deviation of the body acceleration in each of the 3-axial directions (X, Y, Z) |
| AvgTimeBodyAcc-std-Y |
| AvgTimeBodyAcc-std-Z |
| AvgTimeGravityAcc-mean-X | numeric | These variables represent the calculated average by subject and activity of the mean of the gravity acceleration in each of the 3-axial directions (X, Y, Z) |
| AvgTimeGravityAcc-mean-Y |
| AvgTimeGravityAcc-mean-Z |
| AvgTimeGravityAcc-std-X | numeric | These variables represent the calculated average by subject and activity of the standard deviation of the gravity acceleration in each of the 3-axial directions (X, Y, Z) |
| AvgTimeGravityAcc-std-Y |
| AvgTimeGravityAcc-std-Z |
| AvgTimeBodyAccJerk-mean-X | numeric | These variables represent the calculated average by subject and activity of the mean of the body linear acceleration derived in time in each of the 3-axial directions(X, Y, Z) |
| AvgTimeBodyAccJerk-mean-Y |
| AvgTimeBodyAccJerk-mean-Z |
| AvgTimeBodyAccJerk-std-X | numeric | These variables represent the calculated average by subject and activity of the standard deviation of the body linear acceleration derived in time in each of the 3-axial directions(X, Y, Z) |
| AvgTimeBodyAccJerk-std-Y |
| AvgTimeBodyAccJerk-std-Z |
| AvgTimeBodyGyro-mean-X | numeric | These variables represent the calculated average by subject and activity of the mean of the body angular velocity derived in time in each of the 3-axial directions(X, Y, Z) |
| AvgTimeBodyGyro-mean-Y |
| AvgTimeBodyGyro-mean-Z |
| AvgTimeBodyGyro-std-X | numeric | These variables represent the calculated average by subject and activity of the standard deviation of the body angular velocity derived in time in each of the 3-axial directions(X, Y, Z) |
| AvgTimeBodyGyro-std-Y |
| AvgTimeBodyGyro-std-Z |
| AvgTimeBodyGyroJerk-mean-X | numeric | These variables represent the calculated average by subject and activity of the mean of the linear acceleration and body angular velocity derived in time in each of the 3-axial directions(X, Y, Z) |
| AvgTimeBodyGyroJerk-mean-Y |
| AvgTimeBodyGyroJerk-mean-Z |
| AvgTimeBodyGyroJerk-std-X | numeric | These variables represent the calculated average by subject and activity of the standard deviation of the linear acceleration and body angular velocity derived in time in each of the 3-axial directions(X, Y, Z) |
| AvgTimeBodyGyroJerk-std-Y |
| AvgTimeBodyGyroJerk-std-Z |
| AvgTimeBodyAccMag-mean | numeric | This variable represents the calculated average by subject and activity of the mean of the magnitude of the 3-dimensional signals of Body Acceleration using the Euclidian norm derived in time |
| AvgTimeBodyAccMag-std | numeric | This variable represents the calculated average by subject and activity of the standard deviation of the magnitude of the 3-dimensional signals of Body Acceleration using the Euclidian norm derived in time |
| AvgTimeGravityAccMag-mean | numeric | This variable represents the calculated average by subject and activity of the mean of the magnitude of the 3-dimensional signals of Gravity Acceleration using the Euclidian norm derived in time |
| AvgTimeGravityAccMag-std | numeric | This variable represents the calculated average by subject and activity of the mean of the magnitude of the 3-dimensional signals of Gravity Acceleration using the Euclidian norm derived in time |
| AvgTimeBodyAccJerkMag-mean | numeric | This variable represents the calculated average by subject and activity of the mean of the magnitude of the 3-dimensional signals of body linear acceleration using the Euclidian norm derived in time |
| AvgTimeBodyAccJerkMag-std | numeric | This variable represents the calculated average by subject and activity of the standard deviation of the magnitude of the 3-dimensional signals of body linear acceleration using the Euclidian norm derived in time |
| AvgTimeBodyGyroMag-mean | numeric | This variable represents the calculated average by subject and activity of the mean of the magnitude of the 3-dimensional signals of body angular velocity using the Euclidian norm derived in time |
| AvgTimeBodyGyroMag-std | numeric | This variable represents the calculated average by subject and activity of the standard deviation of the magnitude of the 3-dimensional signals of body angular velocity using the Euclidian norm derived in time |
| AvgTimeBodyGyroJerkMag-mean | numeric | This variable represents the calculated average by subject and activity of the mean of the magnitude of the 3-dimensional signals of body linear acceleration and angular velocity using the Euclidian norm derived in time |
| AvgTimeBodyGyroJerkMag-std | numeric | This variable represents the calculated average by subject and activity of the standard deviation of the magnitude of the 3-dimensional signals of body linear acceleration and angular velocity using the Euclidian norm derived in time |
| AvgFreqBodyAcc-mean-X | numeric | These variables represent the calculated average by subject and activity of the frequency (using FFT) of the body acceleration in each of the 3-axial directions (X, Y, Z) |
| AvgFreqBodyAcc-mean-Y |
| AvgFreqBodyAcc-mean-Z |
| AvgFreqBodyAcc-std-X | numeric | These variables represent the calculated average by subject and activity of the standard deviation of the frequency (using FFT) of the body acceleration in each of the 3-axial directions (X, Y, Z) |
| AvgFreqBodyAcc-std-Y |
| AvgFreqBodyAcc-std-Z |
| AvgFreqBodyAccJerk-mean-X | numeric | These variables represent the calculated average by subject and activity of the frequency (using FFT) of the body linear acceleration in each of the 3-axial directions (X, Y, Z) |
| AvgFreqBodyAccJerk-mean-Y |
| AvgFreqBodyAccJerk-mean-Z |
| AvgFreqBodyAccJerk-std-X | numeric | These variables represent the calculated average by subject and activity of the standard deviation of the frequency (using FFT) of the body linear acceleration in each of the 3-axial directions (X, Y, Z) |
| AvgFreqBodyAccJerk-std-Y |
| AvgFreqBodyAccJerk-std-Z |
| AvgFreqBodyGyro-mean-X | numeric | These variables represent the calculated average by subject and activity of the frequency (using FFT) of the angular velocity in each of the 3-axial directions (X, Y, Z) |
| AvgFreqBodyGyro-mean-Y |
| AvgFreqBodyGyro-mean-Z |
| AvgFreqBodyGyro-std-X | numeric | These variables represent the calculated average by subject and activity of the standard deviation of the frequency (using FFT) of the angular velocity in each of the 3-axial directions (X, Y, Z) |
| AvgFreqBodyGyro-std-Y |
| AvgFreqBodyGyro-std-Z |
| AvgFreqBodyAccMag-mean | numeric | This variable represents the calculated average by subject and activity of the mean of the magnitude of the frequency of the 3-dimensional signals of body acceleration using the Euclidian norm |
| AvgFreqBodyAccMag-std | numeric | This variable represents the calculated average by subject and activity of the standard deviation of the magnitude of the frequency of the 3-dimensional signals of body acceleration using the Euclidian norm |
| AvgFreqBodyBodyAccJerkMag-mean | numeric | This variable represents the calculated average by subject and activity of the mean of the magnitude of the frequency of the 3-dimensional signals of linear acceleration using the Euclidian norm |
| AvgFreqBodyBodyAccJerkMag-std | numeric | This variable represents the calculated average by subject and activity of the standard deviation of the magnitude of the frequency of the 3-dimensional signals of linear acceleration using the Euclidian norm |
| AvgFreqBodyBodyGyroMag-mean | numeric | This variable represents the calculated average by subject and activity of the mean of the magnitude of the frequency of the 3-dimensional signals of angular velocity using the Euclidian norm |
| AvgFreqBodyBodyGyroMag-std | numeric | This variable represents the calculated average by subject and activity of the standard deviation of the magnitude of the frequency of the 3-dimensional signals of angular velocity using the Euclidian norm |
| AvgFreqBodyBodyGyroJerkMag-mean | numeric | This variable represents the calculated average by subject and activity of the mean of the magnitude of the frequency of the 3-dimensional signals of linear acceleration and angular velocity using the Euclidian norm |
| AvgFreqBodyBodyGyroJerkMag-std | numeric | This variable represents the calculated average by subject and activity of the standard deviation of the magnitude of the frequency of the 3-dimensional signals of linear acceleration and angular velocity using the Euclidian norm |

VARIABLES EXCLUDED:

The original data set had 561 variables. Variables excluded from analysis where those which were not either a mean or a standard deviation. Variables in the original data set with the following descriptive components were thus eliminated from the group:

* 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 (sic) vectors,

1. Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012 [↑](#footnote-ref-1)
2. http://archive.ics.uci.edu/ml/machine-learning-databases/00240/UCI%20HAR%20Dataset.names [↑](#footnote-ref-2)
3. http://archive.ics.uci.edu/ml/machine-learning-databases/00240/UCI%20HAR%20Dataset.names [↑](#footnote-ref-3)
4. https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip [↑](#footnote-ref-4)