Getting and Cleaning Project
Version 1.0

Javier Augusto Díaz
javieraugsutodiaz@gmail.com

Objective: To get a tidy data from a raw data set of Human Activity Recognition for smartphones. This data set was obtained from
http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones and converted for these project to the data set in
https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip
This project provides an scrip that obtain a tidy data set for analysis and an application of mean value variables evaluation by subject and by activity is also provided following the steps was specify on the project definition.

The experiments: the data was measured through signal sensors (accelerometer and gyroscope) on each out of 30 volunteers **Subjects** while they were doing 6 **Activities**: WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, and LAYING. The measurements capture 3 axial (XYZ) linear acceleration and 3 axial angular velocity on a 128 windows. These raw information was filtered and transformed to vectors of features of variables from the time frequency domain. See README.txt in mentioned data set.

Each record provides:
- 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.

Raw data set: As it is mention in the README.txt, the dataset includes the following files
- 'README.txt'

- 'features_info.txt': Shows information about the variables used on the feature vector.

- 'features.txt': List of all features.

- 'activity_labels.txt': Links the class labels with their activity name.

- 'train/X_train.txt': Training set.

- 'train/y_train.txt': Training labels.

- 'test/X_test.txt': Test set.

- 'test/y_test.txt': Test labels.

The following files are available for the train and test data. Their descriptions are equivalent.

- 'train/subject_train.txt': Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30.

- 'train/Inertial Signals/total_acc_x_train.txt': The acceleration signal from the smartphone accelerometer X axis in standard gravity units 'g'. Every row shows a 128 element vector. The same description applies for the 'total_acc_x_train.txt' and 'total_acc_z_train.txt' files for the Y and Z axis.

- 'train/Inertial Signals/body_acc_x_train.txt': The body acceleration signal obtained by subtracting the gravity from the total acceleration.

- 'train/Inertial Signals/body_gyro_x_train.txt': The angular velocity vector measured by the gyroscope for each window sample. The units are radians/second.

From these information the following data set structure was drown

./Project

/UCI HARD Dataset

/test

/Inertial Signals

body_acc_x_test.txt	grav acc 128
body_acc_y_test.txt	grav acc 128
body_acc_z_test.txt	grav acc 128
body_gyro_x_test.txt	Velocidad rad/sec 128
body_gyro_y_test.txt	Velocidad rad/sec128
body_gyro_z_test.txt	Velocidad rad/sec128
total_acc_x_test.txt	128 col
total_acc_y_test.txt	
total_acc_z_test.txt	
subject_test.txt	**30**
X_test.txt	**test DS**
Y_test.txt	**test labels**

/train

/Inertial Signals

body_acc_x_train.txt	grav acc 128
body_acc_y_train.txt	grav acc 128
body_acc_z_train.txt	grav acc 128
body_gyro_x_train.tx.	Velocidad rad/sec 128
body_gyro_y_train.txt	Velocidad rad/sec 128
body_gyro_z_train.txt	Velocidad rad/sec 128
total_acc_x_train.txt	total acc 128 col (windows)
total_acc_y_train.txt	total acc 128
total_acc_z_train.txt	total acc 128
subject_train.txt	**30**
X_train.txt	**training DS Col=features**
Y_test.txt	**training labels**
activity_labels.txt	**6 activities**
features.txt	**561=(3x3+3x2+5+3x3+4)17**
futures_info.txt	Document
README.txt	Document

5 variables of mean

gravityMean
tbodyAccMean
tBodyAccJerkMean
tBodyGyroMean
tBodyGyroJerkMean

The scrip follows 4 step :

1 Merge the training and the test sets: **X_test.txt** and **X_train.txt,** and **Y_test.txt** and **Y_test.txt.**

At the end, when required are merged **subject_test.txt** and **subject_train.txt**

2 Extract the mean and standard deviation of each measurements are of the 3-axial vector for velocity and accelerations and Jerk components.

3.Names of activities are added observations of data set as column from merge in 1 for Yfiles

4Names to variables are added to columns using **features.txt**

---

5 The data set obtained is grouped by activity and the mean value of variables selected is obtained. See file “ActMean.csv”
In addition, data set is grouped by subject and the mean value of variables selected is obtained. See file “SubjMean.csv”

Following files are included

README.txt
CodeBook.txt
Run_analysis.R Contains the scrip
/Data/ActivityMeans .csv Contains the mean values of variables .mean() and .std() for each **Activity**
/Data/SubjectMean.csv Contains the mean values of variables .mean() and .std() for each **Subject**

Reference

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