

# CodeBook

Gianni Enas

2023-01-22

## Details of observations

This database was first conceived to analyse human activities using smartphones. 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 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.

## Credits

The credits for this work goes to: Jorge L. Reyes-Ortiz, Davide Anguita, Alessandro Ghio, Luca Oneto. Smartlab - Non Linear Complex Systems Laboratory DITEN - Università degli Studi di Genova. Via Opera Pia 11A, I-16145, Genoa, Italy. [activityrecognition@smartlab.ws](mailto:activityrecognition@smartlab.ws) further infos at: [smartlab](http://smartlab.ws)

## My implementation

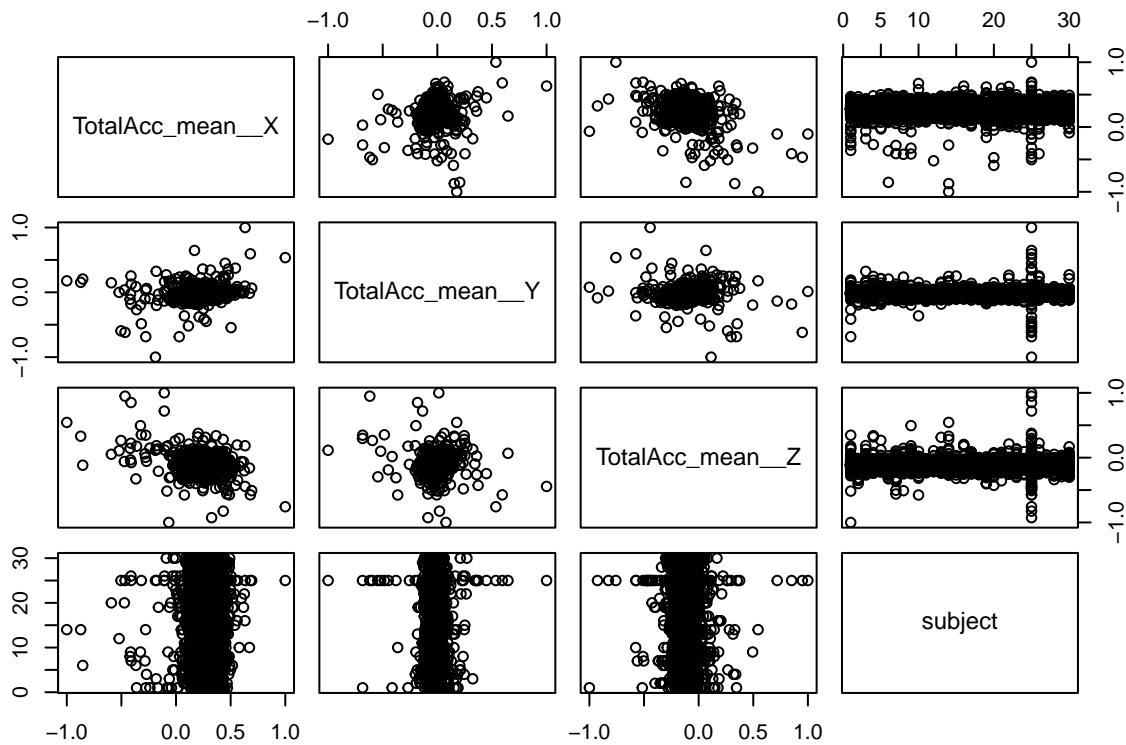
First thing I had to join the train data with the test data, at the same time I've also included an extra column with integer values indicating each one of the 30 subjects that took part at the experiment. The column was labeled "subject". Later on I've implemented a function to assign a text label to the numeric variables for each activity in order to render the database more readable, and added the column to the database labeling it "str\_activity" later I've tried to clean the dataset selecting only the features for the means and standard deviations, including the last columns "subject, activity and str\_activity", I've also re-formatted the features names into a more readable way, I didn't want to change the names entirely but I got rid of some redundant symbols and characters. Here below and extract of the database with the first and last columns:

## Dataset extract

```
##   TotalAcc_mean__X TotalAcc_mean__Y TotalAcc_mean__Z subject str_activity
## 1      -1.0000000    0.1775221600     0.54393929      14    laying
## 2      -0.8723954    0.1546078000     0.33075342      14    laying
## 3      -0.8538482    0.2053651600    -0.11634455       6    laying
## 4      -0.5920043    0.1469832700     0.05256077      20    laying
## 5      -0.5210621    -0.0001832748    0.10661589      12    laying
## 6      -0.5038227    -0.5942073800     0.26480435      25    laying
```

## Including Plots

Sample plot for the mean of the 3 axis Total Acceleration and each subject in all the rows:



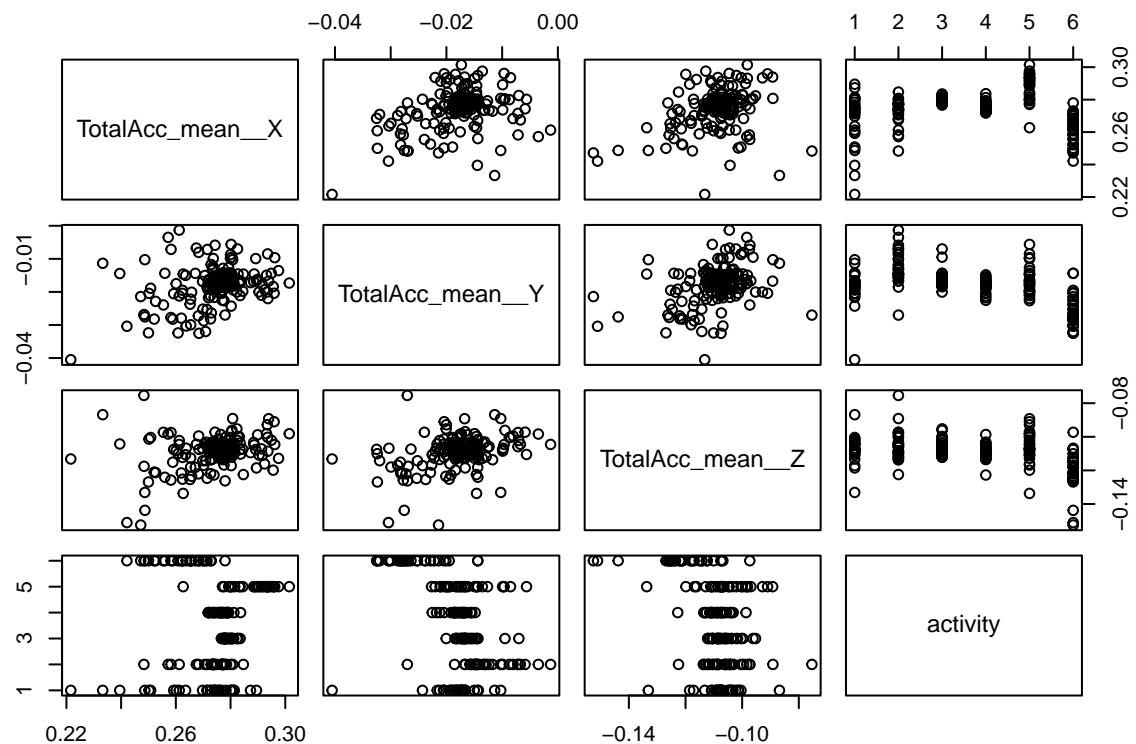
## Building a new dataset with the mean and std for each subject and each activity:

Continuing the assignment, I sorted out the previous dataset for Subject and activity and I've implemented another function to filter all the values for each subject and each activity. The function "sort\_activity" goes through each value and build a brand new database, with the column activity as descriptive character values. The result is a new database and below is an extract of the firsts subject for all the activities:

### Extract of the new dataset

	TotalAcc_mean_X	TotalAcc_mean_Y	TotalAcc_mean_Z	subject	activity
## 1	0.2773308	-0.017383819	-0.1111481	1	walking
## 2	0.2554617	-0.023953149	-0.0973020	1	walking_upstairs
## 3	0.2891883	-0.009918505	-0.1075662	1	walking_downstairs
## 4	0.2612376	-0.001308288	-0.1045442	1	sitting
## 5	0.2789176	-0.016137590	-0.1106018	1	standing
## 6	0.2215982	-0.040513953	-0.1132036	1	laying

including plot



## conclusions I hope that my work has well interpreted the instruction of the Monitor.