**Code Book – Final Project**

**Getting and Cleaning Data Course**

**Johns Hopkins University by Coursera**

**Author: Rejane Rodrigues de Carvalho Pereira (30/10/2020, Brasília-DF, Brazil)**

**1) General information about the project:**

This codebook intends to document the final project that was done in October 2020 and applied to students who were enrolled in the Johns Hopkins University Getting and Cleaning Data Course by Coursera. The final project consists of creating a script, in addition to this Code Book and the README\_Project\_Rejane.txt file.

This script used like input data set that results of the research carried out through the experiments with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities described in the third section wearing a smartphone (Samsung Galaxy S II) on the waist.

The script called “run\_analysis.R” intends to do five steps to follow:

STEP 1: Merges the train and the test sets to create data set;

STEP 2: Extracts the measurements concerning mean and standard deviation;

STEP 3: Uses descriptive activity names for activities;

STEP 4: Appropriately labels with descriptive variable names in data sets;

STEP 5: Creates an independent tidy data set with the average of each variable for each activity and each subject.

This script is attached to that document.

**2) Experiment Method**

Using its embedded accelerometer and gyroscope, the data were 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 were 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.

Check other information for further details about this project in site: <http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones>

Here are the data for the project:

[https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip](https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip" \t "_blank)

**3) Information on the variables of the main data set created:**

**new\_dataset\_average**

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable names** | **Variable description** | **Data Type** | **Source Files and additional information** |
| **activity** | **Name of the activity performed in the experiment.**  Can be 6 activities:  - LAYING  - STANDING  - SITTING  - WALKING  - WALKING\_DOWNSTAIRS  - WALKING\_UPSTAIRS | String | The activity list for each subject and parameter came from the merged files y**\_train.txt and y\_test.txt.**  Furthermore, the activities names came from the file activity\_labels.txt. |
| **subject** | **Sequential number of the experiment volunteer. Its range is from 1 to 6.** | Integer | The subject list for each activity and parameter came from the merged files subject\_train.txt and subject\_test.txt |
| **measure** | **Each measured parameter within the experiment.**  **The script filtered only the measures concerning to mean and standard deviation.**  **Clean name of the measure such as:**   * **removed characteres “[(),]|-”;** * **standardized substrings like “Mean” or “Std”.** | **String** | **Variables names came from the file features.txt with applied filters "mean|Mean|std|Std" in its second column.** |
| **average** | **Average of the variables concerning to mean and standard deviation for each parameter by activity and by subject.** | **Double with two decimal places** | **The measures of each parameter came from the merged files X\_train.txt and X\_test.txt with the same previous filter applied.**  Furthermore, **they were grouped by activity and by subject and after was applied the mean.** |