Code Book

The script “run_analysis.R” is a code tidying and merging the data from the Samsung Galaxy S smartphone project.

Step 1: prepare, download the data zip and set working directory
- prepare by loading the library dplyr and data.table;
- The dataset was downloaded from https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip;
- set the working directory to “UCI HAR Dataset”.
- assign name for files and name columns, using *read.table()* function.

Step 2: manage the data by instruction

1. Merges the training and the test sets to create one data set.
   - use *rbind()* function to merge data tables: table x is created by merging x_test and x_train; table y is created by merging y_test and y_ train; subject is created by merging subject_test and subject_train; And data1 is created by merging subject, x and y.

2. Extracts only the measurements on the mean and standard deviation for each measurement.
   - data2 is created by extracting the columns of subject, code and columns that contains “mean” and “std” from data1.

3. Uses descriptive activity names to name the activities in the data set
   - the numbers in column “code” are replaced by the descriptive names in *activity* data set

4. Appropriately labels the data set with descriptive variable names.
   - column code in data2 is named activity
   - columns begin with a “t” are replaced by “Time”
   - columns begin with a “f” are replaced by “Frequency”
   - all “Acc” in columns’ name are replaced by “Accelerometer”
   - all “Gyro” in columns’ name are replaced by “Gyroscope”
   - all “Mag” in columns’ name are replaced by “Magnitude”
   - all “mean” in columns’ name are replaced by “Mean”
   - all “std” in columns’ name are replaced by “StandardDeviation”
   - all “-” in columns’ name are replaced by “_”
   - all “BodyBody” in columns’ name are replaced by “Body”
   - all “tBody” in columns’ name are replaced by “TimeBody”

5. From the data set in step 4, creates a second, independent tidy data set with the average of each variable for each activity and each subject.
   - final result “TidyData” is created by summarizing data2, calculating the means of each variable for each activity and each subject.
   - export the file by *write.table()* function