CodeBook

Samuel Anang

11/7/2020

The run_analysis.R script performs the data preparation and then followed by the steps required as described in the project requirements

1. Download the dataset

- Dataset downloaded and extracted to the folder called UCI HAR Dataset

2. Assign each data to variables

- features <- features.txt : 561 rows, 2 columns The features selected for this database come from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ.
- activities <- activity_labels.txt : 6 rows, 2 columns List of activities performed when the corresponding measurements were taken and its codes (labels)
- subject_test <- test/subject_test.txt : 2947 rows, 1 column contains test data of 9/30 volunteer test subjects being observed
- x_test <- test/X_test.txt : 2947 rows, 561 columns contains recorded features test data
- y_test <- test/y_test.txt : 2947 rows, 1 columns contains test data of activities'code labels
- subject_train <- test/subject_train.txt : 7352 rows, 1 column contains train data of 21/30 volunteer subjects being observed
- x_train <- test/X_train.txt : 7352 rows, 561 columns contains recorded features train data
- y_train <- test/y_train.txt : 7352 rows, 1 columns contains train data of activities'code labels

3. Merge the training and the test sets to create one data set

- merged_x (10299 rows, 561 columns) is created by merging x_train and x_test using rbind()
- merged_y (10299 rows, 1 column) is created by merging y_train and y_test using rbind()
- merged_Subject (10299 rows, 1 column) is created by merging subject_train and subject_test using rbind()
- final_merged (10299 rows, 563 column) is created by merging Subject, Y and X using cbind()

4. Extract only the measurements on the mean and standard deviation for each measurement

- tidy_data (10299 rows, 88 columns) is created by subsetting final_merged, selecting only columns: subject, code and the measurements on the mean and standard deviation (std) for each measurement

5. Use descriptive activity names to name the activities in the data set

- Entire numbers in code column of the tidy_data replaced with corresponding activity taken from second column of the activities variable

6. Appropriately labels the data set with descriptive variable names

- code column in tidy_data renamed into activities
- All Acc in column's name replaced with Accelerometer
- All Gyro in column's name replaced with Gyroscope
- All BodyBody in column's name replaced with Body
- All Mag in column's name replaced with Magnitude
- All start with character f in column's name replaced with Frequency
- All start with character t in column's name replaced with Time

7. From the data set in step 4, create a second, independent tidy data set with the average of each variable for each activity and each subject

- final_independent_tidy_data (180 rows, 88 columns) is created by summarizing tidy_data taking the means of each variable for each activity and each subject, after grouped by subject and activity.
- Export tidy_data into FinalIndependentTidyData.txt file.