Getting and Cleaning Data Course Project

CodeBook by Chen Lianghe

19/01/2020

The run_analysis.R script performs the data preparation and the 5 steps required as defined in the Peer-graded Assignment: Getting and Cleaning Data Course Project.

1. Merges the training and the test sets to create one data set.
   - Dataset was downloaded and extracted under the folder named UCI HAR Dataset.
   - Data was assigned 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 with its code labels.
     - subject_test <- test/subject_test.txt : 2947 rows, 1 column. Contains test data of 9/30 volunteer test subjects that were observed.
     - x_test <- test/X_test.txt : 2947 rows, 561 columns. Contains recorded features from test data.
     - y_test <- test/y_test.txt : 2947 rows, 1 column. 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 that were observed.
     - x_train <- test/X_train.txt : 7352 rows, 561 columns. Contains recorded features from train data.
     - y_train <- test/y_train.txt : 7352 rows, 1 column. Contains train data of activities' code labels.
   - Merges the training and the test sets to create one data set.
     - X (10299 rows, 561 columns) created by merging x_train and x_test using rbind() function.
     - Y (10299 rows, 1 column) created by merging y_train and y_test using rbind() function.
     - Subject (10299 rows, 1 column) created by merging subject_train and subject_test using rbind() function.
     - Merged_Data (10299 rows, 563 columns) created by merging Subject, Y and X using cbind() function.
2. Extracts only the measurements on the mean and standard deviation for each measurement.
   - TidyData (10299 rows, 88 columns) created by subsetting Merged_Data, selecting only columns: subject, code and measurements of the mean and standard deviation (std).
3. Uses descriptive activity names to name the activities in the data set.
   - Entire numbers in code column of TidyData replaced with corresponding activity taken from second column of activities variable.
4. Appropriately labels the data set with descriptive variable names.
   - code column in TidyData renamed activities.
   - All Acc in column's name replaced by Accelerometer.

- All Gyro in column's name replaced by Gyroscope.
- All BodyBody in column's name replaced by Body.
- All Mag in column's name replaced by Magnitude.
- All labels starting with character f in column's name replaced by Frequency.
- All labels starting with character t in column's name replaced by Time.

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.

    - FinalData (180 rows, 88 columns) created by summarizing TidyData. This is done by taking the means of each variable for each activity and subject, after grouping by activity and subject.
    - Export FinalData into FinalData.txt file.