CodeBook

Jens

6 12 2020

The R script *run_analysis.R* performs the data preparation, and exports the required tidy data set

1. Read the data and assign to variables

- *features <- features.txt*
  - List of measured features
- *activities <- activity_labels.txt*
  - List of activities and its codes
- *subject_test <- test/subject_test.txt*
  - containts the 9 subjects of the test data
- *x_test <- test/x_test.txt*
  - containts recorded features test data
- *y_test <- test/y_test.txt*
  - contains activity codes for test data
- *subject_train <- train/subject_train.txt*
  - containts the 21 subjects of the train data
- *x_train <- train/x_train.txt*
  - containts recorded features train data
- *y_train <- train/y_train.txt*
  - contains activity codes for train data

2. Merges the training and the test sets to create one data set

- *x* is created by merging *x_train* and *x_test* using **rbind()** function
- *y* is created by merging *y_train* and *y_test* using **rbind()** function
- *Subject* is created by merging *subject_train* and *subject_test* using **rbind()** function
- *Merged_Data* is created by merging *Subject,y* and *x* using **cbind()** function

3. Extracts only the measurements on the mean and standard deviation for each measurement

- *TidyData* is created by subsetting *Merged_Data*, selecting only columns: *subject*, *code* and the columns which name containts *mean* or *std*

4. Uses descriptive activity names to name the activities in the data set

- Replace the numbers in *code* in the *TidyData* with the corresponding names

5. Appropriately labels the data set with descriptive variable names

- Change the *code* column-name into *activities*
- Change several abbreviations in the names

6. 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* is created by taking the means of each variable after grouping by subject and activity
- Export the data into *FinalData.txt*