Code Book for Peer-graded Assignment: Getting and Cleaning Data Course Project

The run\_analysis.R script includes data preparation stage followed by 5 stages as required in the course project’s description.

* Download all the datasets from the folder: UCI HAR Dataset
* Assign each dataset to a 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 labels

subject\_test = test/subject\_test.txt : 2947 rows, 1 column

contains test data for volunteer test subjects

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 for volunteer subjects

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

* Merges the training and the test sets to create one data set

x (10299 rows, 561 columns) is created by merging x\_train and x\_test using rbind()

y (10299 rows, 1 column) is created by merging y\_train and y\_test using rbind()

Subject (10299 rows, 1 column) is created by merging subject\_train and subject\_test using rbind()

Merged\_Data (10299 rows, 563 column) is created by merging Subject, y and x using cbind()

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

TidyData (10299 rows, 88 columns) is created by subsetting Merged\_Data, selecting only columns: subject, code and the measurements on the mean and standard deviation (std) for each measurement

* Uses descriptive activity names to name the activities in the data set

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

* Appropriately labels the data set with descriptive variable names

code column in TidyData renamed into activity

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 start with character f in column’s name replaced by Frequency

All start with character t in column’s name replaced by Time

* 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\_Data (180 rows, 88 columns) is created by summarizing TidyData taking the means of each variable for each activity and each subject, after grouped by subject and activity.

Export Final\_Data into Final\_Data.txt file.