

CodeBook

Vanessa

13/08/2020

The run_analysis.R script performs the data preparation and then followed by the 5 steps required as described in the course project's definition.

Download the dataset

Dataset downloaded and extracted under the folder called UCI HAR Dataset

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

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() function Y (10299 rows, 1 column) is created by merging y_train and y_test using rbind() function Subject (10299 rows, 1 column) is created by merging subject_train and subject_test using rbind() function Merged_Data (10299 rows, 563 column) is created by merging Subject, Y and X using cbind() function

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 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 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

FinalData (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 FinalData into FinalData.txt file.