

Code Book

This Codebook provides variable names and steps taken for cleaning and merging data.

1. Downloaded Dataset

- a. Dataset downloaded and uploaded under the folder named UCI HAR Dataset

2. Assigned Variables for Data

- a. `features <- features.txt` : 561 rows, 2 columns
Parameters selected for this dataset were recorded on a accelerometer and gyroscope with a 3-axial raw signal `tAcc-XYZ` and `tGyro-XYZ`
- b. `activities <- activity_labels.txt` : 6 rows, 2 columns
The list of activities performed by 9/30 of the volunteer subjects, that created the corresponding data
- c. `subject_test <- test/subject_test.txt` : 2947 rows, 1 column
The volunteer test data observed
- d. `x_test <- test/X_test.txt` : 2947 rows, 561 columns
Contains features of the test data
- e. `y_test <- test/y_test.txt` : 2947 rows, 1 column
The activities labels for the test data
- f. `subject_train <- test/subject_train.txt` : 7352 rows, 1 column
Train data of 21/30 volunteer observed volunteers
- g. `x_train <- test/X_train.txt` : 7352 rows, 561
Contains features of the train data
- h. `y_train <- test/y_train.txt` : 7352 rows, 1 column
- i. The activities labels for the train data

3. Merging the two data sets, Train and Test, making one data set

- a. `X` (10299 rows, 561 column) merging `x_train` and `x_test`, using the `rbind()` function
- b. `Y` (10299 rows, 1 column) merging `y_train` and `y_test`, using the `rbind()` function
- c. `Subject` (10299 rows, 1 column) merging `subject_train` and `subject_test` using `rbind()`
- d. `Merge_Data` (10299 rows, 563 columns) merging `Y` and `X`, using `cbind()`

4. Extracting the mean and standard deviation of the measurements

- a. TidyData (10299 rows, 88 columns), subsetting Merge_data, select columns: subject, code and measurements on the mean/standard deviation (std) for each measurement

5. More descriptive activity names for the activities in the data set

- a. code column i TidyData replaced with corresponding activity marked in the second column in the activities variables

6. Descriptive variable names for the data set

- a. activities replaced code (in TidyData)
- b. Accelerometer replaced Acc
- c. Gyroscope replaced Gyro
- d. Body replaced BodyBody
- e. Magnitude replaced Mag
- f. Frequency replaced character f
- g. Time replaced character t

7. Second independent data set, containing average of each variable for each activity and each subject

- a. FinalData (180 rows, 88 columns) a summarized version of TidyData, mean of each variable for each activity and each subject, after grouped by subject and activity.
- b. Export FinalData into FinalData.txt file