

## Getting-and-cleaning-data-assignment

This is a short description of the data processing implemented in the R script ‘run\_analysis.R’.

1. The data files related to training phase ‘X\_train.txt’, ‘y\_train.txt’ and ‘subject\_train.txt’ have been combined in a single data frame called “train\_data”. Furthermore a new variable called “Phase” has been created and initialized at the value “training”. This new variable takes into account the context change information implied in the following merging of training and test data sets.
2. The data files related to test phase ‘X\_test.txt’, ‘y\_test.txt’ and ‘subject\_test.txt’ have been combined in a single data frame called “test\_data”. Consistently with previous note the new variable called “Phase” has been initialized at the value “test”
3. The two data frame “train\_data” and “test\_data” described above have been merged in a new data frame called “all\_data”
4. The columns names related to the measures included in the original files ‘X\_train.txt’ and ‘X\_test.txt’ have been derived from file ‘features.txt’ assuming the variable order of this file consistent with ‘X\_train.txt’ file columns that now are the 561 rightmost columns of “all\_data” data frame.
5. To obtain a tidy data set only one measure type has been selected choosing all accelerations in the time domain measures. This selection is performed by calculation of the vector “col\_sel”. The selected column names are stored in the character vector “colnames”. Consequently is calculated the vector “col\_ind” as the column index to select the desired columns from “all\_data”.
6. The selected columns by “col\_sel” vector are stored in the new data frame “tidy\_data” and its variable names are assigned taking advantage of the vector “colnames”.
7. Finally the values in the column Activity of “tidy\_data” are replaced with the corresponding literal values derived from file ‘activity\_labels.txt’.
8. Starting from “tidy\_data” the column measures are summarized in a new data frame “sum\_data” with their average values calculated by groups of each Activity and each Subject.
9. The resultant “sum\_data” data frame is saved in the attached file ‘avg\_dataset.txt’