Getting and Cleaning Data Course Project CodeBook

This file describes the variables, the data, and any transformations or work that was performed to clean up the data.

- The data for the project:
  - https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip
- Unpacking the file, it will create a directory called `UCI HAR Dataset`

The run_analysis.R script performs the following steps to clean the data:

1. Read Train related information (`X_train.txt, y_train.txt and subject_train.txt`) from the `./UCI HAR Dataset/train` directory and store them in the variables: `trainData`, `trainLabel` and `trainSubject` respectively.
2. Repeat previous step to the Test data, read `X_test.txt, y_test.txt and subject_test.txt` files from the `./UCI HAR Dataset/test` directory and store them in variables: `testData`, `testLabel` and `testsubject` respectively.
3. Concatenate the information:
   - Bind by row the the variables `testData` and `trainData` creating a new data frame called `joinData`.
   - Bind by row the the variables `testLabel` and `trainLabel` creating a new data frame called `joinLabel`.
   - Bind by row the variables `testSubject` and `trainSubject` creating a new data frame called `joinSubject`.
4. Read the `features.txt` file from the `./UCI HAR Dataset` directory and store the data in a variable called `features`. We only extract the measurements on the mean and standard deviation. This results in a 66 indices list. We get a subset of `joinData` with the 66 corresponding columns.
5. Clean the column names of the subset by removing the "()" and "-" symbols in the names. Also changes the first letter of "mean" and "std" with their's respectivy capital letters.
6. Read the activity_labels.txt file from the "./UCI HAR Dataset"" directory and store the data in a variable called `activity`.
7. Normalize the activity names in the second column of `activity` by changing all names to lower cases, removing the underscores and capitalize the letter immediately after the underscore.
8. Transform the values of `joinLabel` according to the `activity` data frame.
9. Combine the `joinSubject`, `joinLabel` and `joinData` by column to get a new cleaned 10299x68 data frame, `cleanedData`. Assign the correct names to the first two columns, `subject` and `activity`. The "subject" column contains integers that range from 1 to 30 inclusive; the "activity" column contains 6 kinds of activity names; the last 66 columns contain measurements that range from -1 to 1 exclusive.
10. Write the `cleanedData` variable to `DS1_merged_data.txt` file in the directory `./UCI HAR Dataset`. ###### First answer
11. Finally, generate a tidy second dataset with the average of each measurement for each activity and each subject. There is a 180 combinations based in the 30 subjects and 6 activities. For each combination it is calculated the mean of each measurement with the corresponding combination. The new dataset it is assigned to the `result` variable.

12. Write the `result` out to `DS2_data_with_means.txt` file in current working directory. ###### second answer

Note: The script has two variables: `inDir` and `outDir`. The first is used to point to the original datasets directory `(./UCI HAR Dataset)`, the second is used to point to the project's github directory. This was done only to facilitate the task.