Description

The following script run\_analysis.R completes five steps to reach its objective.

1. Merge all the similar data using the rbind() function. For this, since they are similar, we are talking about the files that have the same number of columns and have reference to the same units.
2. Next, we note that only the columns with the mean and standard deviation measures are extracted from the entire dataset. Then, after taking these columns, they are renamed the correct names, that are from features.txt.
3. The activity data is issued with the values 1:6, then we use the activity names and IDs from activity\_labels.txt and they are replaced in the dataset.
4. Throughout the dataset, the columns with unclear column names are fixed and corrected.
5. Lastly, we can create a newer dataset with the average measures for each subject and activity type (30 subjects \* 6 activities = 180 rows). The output file is called averages\_data.txt, and uploaded to this repository.

The Variables

* x\_train, y\_train, x\_test, y\_test, subject\_train and subject\_test contain the data from all of the downloaded files.
* x\_data, y\_data and subject\_data merge the previous datasets for further analysis later on.
* These features embody the right names for the x\_data dataset, which are added to the column names stored in mean\_and\_std\_features, a numeric vector used to extract the data that you want.
* Another approach is to take with the activity names through the activities variable.
* all\_data merges x\_data, y\_data and subject\_data in a large dataset.
* Then lastly, averages\_data holds the important averages which will be later stored in a .txt file. ddply() from the plyr package is used to apply colMeans() and ease the development.