ReadMe

Objective: The objective of this ReadMe file is to explain how the Human Activity data collected using Samsung smartphone has been **"tidy"**ed to form a clean dataset conforming to the standards.

What is not explained here is the how the experiment was conducted, the background and what features/parameters have been collected. The **CodeBook** accompanying this script has the detailed definitions of the various parameters studied for the subjects in this experiment.

**The following are the list of raw datasets which were processed:**

1. 'features.txt': List of all features.
2. 'train/X\_train.txt': Training set.
3. 'train/y\_train.txt': Training labels.
4. 'test/X\_test.txt': Test set.
5. 'test/y\_test.txt': Test labels.
6. 'activity\_labels.txt': Links the class labels with their activity name.
7. 'train/subject\_train.txt'

**Enumerated below is the step-by-step process flow of preparing the tidy dataset from the raw files**

1. Read 'subject\_train', 'X\_Train' , 'Y\_Train' files and binded/clipped the datasets one-beside-another
2. Repeated the above for test data
3. Merged both the Test and Train data one-beneath-another to form the consolidated version
4. Read Features and Activity data as well
5. Selected the required features i.e those having mean or standard deeviation within their name
   * Please note that features like meanfreq() have been avoided keeping the spirit of the quesion in mind
6. Collected the required features and their serial #s along.
   * Added a suffix "V" to the serial # so as to make it consistent with the feature names in the Test & Train data
7. The Merged data is in a wide format i.e the variables are spread across columns.
   * Changed this into narrow/long format with one column having the feature name(key) and the other feature value(value)
8. Joined the consolidated data and features data to pull the feature name along
9. Similar exercise was done to pull the activity name along
10. Now, we aggregate the features by taking the mean of each of them
    * This is done for each activity of a subject i.e each subject-activity combination
11. The resulting data is in a narrow format.
    * Although this is a valid tidy data, we are converting it into wide format by spreading the features across columns
12. Thus, we have the desired **TidyData** file