

```
=====
Data Cleaning Project using
Human Activity Recognition Using Smartphones Dataset
=====
Yap Pow Look
Data Science Specialization Course
Getting and Cleaning Data
21st March 2015
```

```
=====

This project describe the process of cleaning the abovementioned data and
outline the necessary steps and procedures anyone who will like to use
this process to clean the data
```

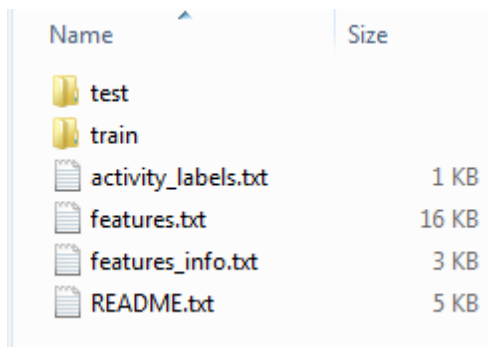
## 1. Pre-processing

Download the following dataset to your local storage and extract the zip folder to its default description.

Here are the data for the project:

<https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip>

When the zip folder is unzipped, you will see the following files and sub-folders origination from your newly created folder :



Name	Size
test	
train	
activity_labels.txt	1 KB
features.txt	16 KB
features_info.txt	3 KB
README.txt	5 KB

In your data clean process, please see the directory of these files and folder as your working directory.

The directory includes the following files:

- ```
=====
```
- 'README.txt' : explanation about the data collection project
  - 'features\_info.txt': Shows information about the variables used on the feature vector.
  - 'features.txt': List of all features.
  - 'activity\_labels.txt': Links the class labels with their activity name.
  - Test folder - This is the results of measurement of the test data
  - Train folder : This is the results of measurement of the train data

Click on the test sub-folder and you will see these files and folder

| Name             | Size      |
|------------------|-----------|
| Inertial Signals |           |
| subject_test.txt | 8 KB      |
| X_test.txt       | 25,839 KB |
| y_test.txt       | 6 KB      |

Description of files in the test sub-folder :

- 'X\_test.txt': Test set.
- 'y\_test.txt': Test labels.
- 'subject\_test.txt': Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30.
- 'Inertial Signals folder': see explanation below

Click on the train sub-folder and you will see these files and folder

| Name              | Size      |
|-------------------|-----------|
| Inertial Signals  |           |
| subject_train.txt | 20 KB     |
| X_train.txt       | 64,460 KB |
| y_train.txt       | 15 KB     |

- 'X\_train.txt': Training set.
- 'y\_train.txt': Training labels.
- 'subject\_test.txt': Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30.
- 'Inertial Signals folder': see explanation below.

The following files are available in the test and train Inertial folders. Their descriptions are equivalent.

- '../Inertial Signals/total\_acc\_x\_train.txt': The acceleration signal from the smartphone accelerometer X axis in standard gravity units 'g'. Every row shows a 128 element vector. The same description applies for the 'total\_acc\_x\_train.txt' and 'total\_acc\_z\_train.txt' files for the Y and Z axis.

- '../Inertial Signals/body\_acc\_x\_train.txt': The body acceleration signal obtained by subtracting the gravity from the total acceleration.

- '../Inertial Signals/body\_gyro\_x\_train.txt': The angular velocity vector measured by the gyroscope for each window sample. The units are radians/second.

Notes:

=====

- Features are normalized and bounded within [-1,1].
- Each feature vector is a row on the text file.

For more information about this dataset contact:  
activityrecognition@smartlab.ws

## 2. Git-hub

The following are the files found in the Git-hub folder from which you have download this README file. The files available in Git-hub are the following :

- a. README - TidyData.rtf
- b. run\_analysis.R : Download this file to your local working directory and execute the as source in your R Console. The version of the R you are running should be R Version 3.1.1 and later.
- c. Codebook - Data Dictionary of the variables found in the data.

## 3. Results files

If you are to execute the run\_analysis.R script, you will also have the following file generated in your working directory. They are :

- a. TidyData.txt

License:

=====

Use of this dataset in publications must be acknowledged by referencing the following publication [1]

[1] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012

This dataset is distributed AS-IS and no responsibility implied or explicit can be addressed to the authors or their institutions for its use or misuse. Any commercial use is prohibited.

Jorge L. Reyes-Ortiz, Alessandro Ghio, Luca Oneto, Davide Anguita.  
November 2012.