# Code Book

This is the code book for Data Science Cleaning data course.

## Study Design

How you collected the data?

The data that was used for this assignment can be found: http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

Here are the data for the project:

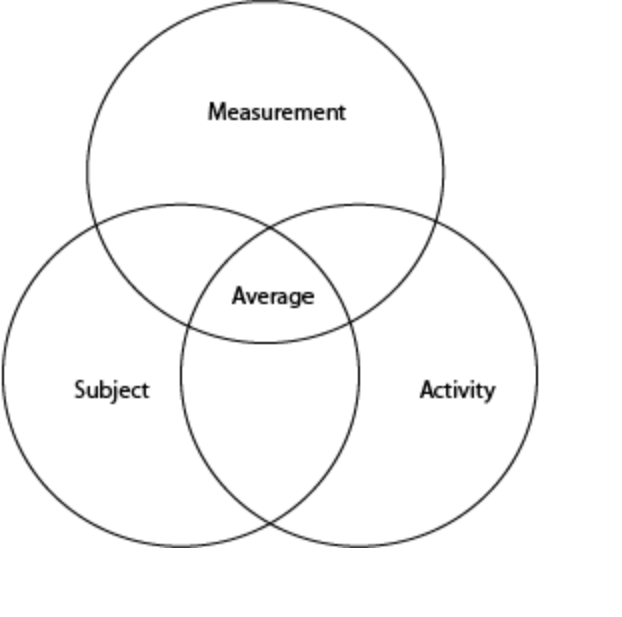
https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip

I spent a considerable amount of time examining the data after reading the ReadMe file. It was clear that the data was very disconnected and would require some work to get things in a format that was meaningful. Thankfully there were some commonalities on the files that allowed the pulling together of data.

I also made use of the following paper that was mentioned in the community forums. <https://thoughtfulbloke.wordpress.com/2015/09/09/getting-and-cleaning-the-assignment/>

I found this to be extremely useful and something we should all do…do as much research on something before getting to far into it.

This was especially helpful:



## Each Variable including units

Units as mentioned in the readme from the original dataset is

Features are normalized and bounded within [-1,1].

- Each feature vector is a row on the text file.

- The units used for the accelerations (total and body) are 'g's (gravity of earth -> 9.80665 m/seg2).

- The gyroscope units are rad/seg.

A table of each variable is below:

|  |  |
| --- | --- |
| Column Name | Additional Information |
| subjectnumber | Subject number. Integer between 1-30 |
| Activityname | Activity performed by the subject. String and one of the following values:  WALKING  WALKING\_UPSTAIRS  WALKING\_DOWNSTAIRS  SITTING  STANDING  LAYING |
| tBodyAcc-mean()-X  tBodyAcc-mean()-Y  tBodyAcc-mean()-Z  tBodyAcc-std()-X  tBodyAcc-std()-Y  tBodyAcc-std()-Z  tGravityAcc-mean()-X  tGravityAcc-mean()-Y  tGravityAcc-mean()-Z  tGravityAcc-std()-X  tGravityAcc-std()-Y  tGravityAcc-std()-Z  tBodyAccJerk-mean()-X  tBodyAccJerk-mean()-Y  tBodyAccJerk-mean()-Z  tBodyAccJerk-std()-X  tBodyAccJerk-std()-Y  tBodyAccJerk-std()-Z  tBodyGyro-mean()-X  tBodyGyro-mean()-Y  tBodyGyro-mean()-Z  tBodyGyro-std()-X  tBodyGyro-std()-Y  tBodyGyro-std()-Z  tBodyGyroJerk-mean()-X  tBodyGyroJerk-mean()-Y  tBodyGyroJerk-mean()-Z  tBodyGyroJerk-std()-X  tBodyGyroJerk-std()-Y  tBodyGyroJerk-std()-Z  tBodyAccMag-mean()  tBodyAccMag-std()  tGravityAccMag-mean()  tGravityAccMag-std()  tBodyAccJerkMag-mean()  tBodyAccJerkMag-std()  tBodyGyroMag-mean()  tBodyGyroMag-std()  tBodyGyroJerkMag-mean()  tBodyGyroJerkMag-std()  fBodyAcc-mean()-X  fBodyAcc-mean()-Y  fBodyAcc-mean()-Z  fBodyAcc-std()-X  fBodyAcc-std()-Y  fBodyAcc-std()-Z  fBodyAcc-meanFreq()-X  fBodyAcc-meanFreq()-Y  fBodyAcc-meanFreq()-Z  fBodyAccJerk-mean()-X  fBodyAccJerk-mean()-Y  fBodyAccJerk-mean()-Z  fBodyAccJerk-std()-X  fBodyAccJerk-std()-Y  fBodyAccJerk-std()-Z  fBodyAccJerk-meanFreq()-X  fBodyAccJerk-meanFreq()-Y  fBodyAccJerk-meanFreq()-Z  fBodyGyro-mean()-X  fBodyGyro-mean()-Y  fBodyGyro-mean()-Z  fBodyGyro-std()-X  fBodyGyro-std()-Y  fBodyGyro-std()-Z  fBodyGyro-meanFreq()-X  fBodyGyro-meanFreq()-Y  fBodyGyro-meanFreq()-Z  fBodyAccMag-mean()  fBodyAccMag-std()  fBodyAccMag-meanFreq()  fBodyBodyAccJerkMag-mean()  fBodyBodyAccJerkMag-std()  fBodyBodyAccJerkMag-meanFreq()  fBodyBodyGyroMag-mean()  fBodyBodyGyroMag-std()  fBodyBodyGyroMag-meanFreq()  fBodyBodyGyroJerkMag-mean()  fBodyBodyGyroJerkMag-std()  fBodyBodyGyroJerkMag-meanFreq() | The STD and MEAN values for each subject per activity take as the average. 79 in total. For units see original documentation: The units used for the accelerations (total and body) are 'g's (gravity of earth -> 9.80665 m/seg2).  - The gyroscope units are rad/seg. |
|  |

## Summary Choices Made

The run\_anlysis.r file contains steps on producing the results. I made the choice to limit the features at the beginning of the process vs. the second step in the assignment. The results were written to a file tidydata.txt that is a combination of Subject, activity and feature data that is unique, and is the average per variable.

It is interesting to note that the format of the txt file is a bit unreadable in raw format. To import into Excel a few extra steps are required:

* Replace ‘” “ ’ with ‘,’
* Replace ‘”’ with ‘’
* Replace ‘ ‘ with ‘,’

A snapshot of the produced data is below:

