Data Code Book for tidy\_data.csv and tidy\_data.txt

The tidy\_data.csv and tidy\_data.txt file is generated from data linked in the ***Getting and Cleaning Data Course Project*** in Coursera. The data was provided through links at :

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

The directions for the assignment stated:

You should create one R script called run\_analysis.R that does the following.

Merges the training and the test sets to create one data set.

Extracts only the measurements on the mean and standard deviation for each measurement.

Uses descriptive activity names to name the activities in the data set

Appropriately labels the data set with descriptive variable names.

From the data set in step 4, creates a second, independent tidy data set with the average of each variable for each activity and each subject.

The data downloaded, and analyzed using R. The provided readme file and code book explained the data and its format.

A script called run\_analysis.r was written to read the data and process it. A HTML and MD file are provided in the repository for the script. The script was broken down into 5 major functions. The functions core functionality correlate to:

* Merge the training and the test sets to create one data set.

merge\_ds

* Extract only the measurements on the mean and standard deviation for each measurement.

extract\_mean\_sd

* Use descriptive activity names to name the activities in the data set

descriptive\_names

* Appropriately label the data set with descriptive variable names.

appropriate\_labels

* From the data set above, creates a second, independent tidy data set with the average of each variable for each activity and each subject.

independent\_tidy\_data

The data was processed in each function consecutively in the following manner:

- read the file with the features

- the features table does not need tidying

- read the file with the activity labels

- add tidy column names

- now read files from the /train/ directory

- read the file with the training subjects

- add tidy column names in human readable text

- add tidy column names from features

- add tidy column names in human readable text

- combine the training data

- now read files from the /test/ directory

- read the file with the test subjects

- add tidy column names in human readable text

- add tidy column names from features

- add tidy column names in human readable text

- combine the test data

- combine the tidy test and training data

- extract the means standard, deviations with the subject and activity ID

- read the file with the activity labels

- add tidy column names

- get rid of parentheses

- A syntactically valid name consists of letters,

- numbers and the dot or underline characters and starts with a letter or

- the dot not followed by a number.

- Fix abbreviated col\_headers

- change Mag for Magnitude

- change ^t for Magnitude

- change Freq for Time

- change Acc for Frequency

- change BodyBody for Acceleration

- change mean for Body

- change ^f for Mean

- change std for Frequency

All means, including “frequency mean” were left as part of the tidy\_data.csv and tidy\_data.txt output in accordance with the assignment directions. “Extracts only the measurements on the mean and standard deviation for each measurement.” Is very specific.