***Cleaning and Getting Data* code book**

**Analysis**

The script, run\_analysis.R performs a number of functions to create tidy data to perform some further analysis.

* First step was to download and read the data from the data source provided.
* Second step was to determine how to get the data into more useable form
* Determined that several of the data inputs provided were names to be assigned to the primary datasets. So these column names were assigned.
* Column names needed to be assigned to the activity\_labels file including activityid and activityname as well as the subject\_test and subject\_train file which I labeled volunteer.
* Next merged the columns of the X,Y,Subject files for both train and test data.
* Then merged test and train data into a master data file.
* Next, only variables containing the mean or standard deviation measurements were extracted and saved to a new file. I interpreted the instruction to be applied to the initial set of variables listed in features\_info.txt
* Then activityname values were merged using activityid.
* Variable names in the dataset were cleansed to take out troublesome punctuation and duplication errors including: “(),-,BodyBody”, converting std and mean to “Std and Mean”,converting leading “t” to “time”, and “f” to “freq”.
* From this file now tidy file, a second independent file was created which contained the average of each variable for each activity and each subject.
* Finally, the new dataset contain the mean is written to an output file, tidydata.csv.
* Refer to run\_analysis.R for more specifics.

**Variable names in output file**

The columns included in the output file are listed below:

|  |
| --- |
| volunteer: subject id |
| activityname: activity labels such as walking,  these variable names are the cleansed names that are detailed in features.txt and outlined in features.info.txt. |
| |  | | --- | | timeBodyAccMeanX | | timeBodyAccMeanY | | timeBodyAccMeanZ | | timeBodyAccStdX | | timeBodyAccStdY | | timeBodyAccStdZ | | timeGravityAccMeanX | | timeGravityAccMeanY | | timeGravityAccMeanZ | | timeGravityAccStdX | | timeGravityAccStdY | | timeGravityAccStdZ | | timeBodyAccJerkMeanX | | timeBodyAccJerkMeanY | | timeBodyAccJerkMeanZ | | timeBodyAccJerkStdX | | timeBodyAccJerkStdY | | timeBodyAccJerkStdZ | | timeBodyGyroMeanX | | timeBodyGyroMeanY | | timeBodyGyroMeanZ | | timeBodyGyroStdX | | timeBodyGyroStdY | | timeBodyGyroStdZ | | timeBodyGyroJerkMeanX | | timeBodyGyroJerkMeanY | | timeBodyGyroJerkMeanZ | | timeBodyGyroJerkStdX | | timeBodyGyroJerkStdY | | timeBodyGyroJerkStdZ | | timeBodyAccMagMean | | timeBodyAccMagStd | | timeGravityAccMagMean | | timeGravityAccMagStd | | timeBodyAccJerkMagMean | | timeBodyAccJerkMagStd | | timeBodyGyroMagMean | | timeBodyGyroMagStd | | timeBodyGyroJerkMagMean | | timeBodyGyroJerkMagStd | | freqBodyAccMeanX | | freqBodyAccMeanY | | freqBodyAccMeanZ | | freqBodyAccStdX | | freqBodyAccStdY | | freqBodyAccStdZ | | freqBodyAccJerkMeanX | | freqBodyAccJerkMeanY | | freqBodyAccJerkMeanZ | | freqBodyAccJerkStdX | | freqBodyAccJerkStdY | | freqBodyAccJerkStdZ | | freqBodyGyroMeanX | | freqBodyGyroMeanY | | freqBodyGyroMeanZ | | freqBodyGyroStdX | | freqBodyGyroStdY | | freqBodyGyroStdZ | | freqBodyAccMagMean | | freqBodyAccMagStd | | freqBodyAccJerkMagMean | | freqBodyAccJerkMagStd | | freqBodyGyroMagMean | | freqBodyGyroMagStd | | freqBodyGyroJerkMagMean | | freqBodyGyroJerkMagStd | |
|  |
|  |
|  |
|  |
|  |
|  |
|  |

**Additional Information**

Refer to the README.txt included in the UCI HAR Dataset download for more detailed information on the experiment as well as the read.md document on git hub for more details on the project.