Codebook

\*\* In order to work, the test and training data MUST be subdirectories of the working directory.

-------------------------------------------------------------------------------------

Transformation of raw data into tidy dataset (tidy\_HAR\_data.R)

Transform Script (test data set)

* Test subject identification data is read from raw data file.
* Test activity identification data is read from raw data file.
* Test result data (accelerometer mean and standard deviation) is read from data file.
* Data frame created for consolidation and manipulation of data.
* Column 1 created with a list of all subject identification numbers. Identification numbers range from 1-30.
* Data from column 1 is copied to column 2.
* Column 1 is filled with the text “test\_set” to identify that the data is from the test set of data.
* Column 3 is populated with associated activities by converting the activity identification code (which ranges from 1-7) and translating it into an associated textual description.
* Columns 4-9 are populated with accelerometer mean and standard deviation data in x-y-z order.

Transform Script (train data set)

* Subject identification data is read from raw data file.
* Training activity identification data is read from raw data file.
* Training result data (accelerometer mean and standard deviation) is read from data file.
* Data frame created for consolidation and manipulation of data.
* Column 1 created with a list of all subject identification numbers. Identification numbers range from 1-30.
* Data from column 1 is copied to column 2.
* Column 1 is filled with the text “train\_set” to identify that the data is from the test set of data.
* Column 3 is populated with associated activities by converting the activity identification code (which ranges from 1-7) and translating it into an associated textual description.
* Columns 4-9 are populated with accelerometer mean and standard deviation data in x-y-z order.

Transformation script for combined data

* The test data and the training data frames are combined.
* The combined data frame is sorted based on the subject identification number.
* The subject identification number is converted from an integer to a factor.

The transformed data set is composed of the following columns:

1 – Activity set (Test or Train)

2 – Subject identification number (a factor variable between 1 – 30)

3 – Activity description (Walking, Walking\_upstairs, Walking\_downstairs, Sitting, Standing, or Laying)

4-6 – Mean Accelerometer data (g’s) for X, Y, and Z axes.

7-9 – Standard Deviation of acceleration readings (g’s squared) for X, Y, and Z axes.

The results are written to a csv file titled “tidy\_HAR\_dataset.csv” and returned to the calling function.

-------------------------------------------------------------------------------------

Calculate the means for each subject and activity from the tidy dataset (tidy\_HAR\_means.R)

The csv file created by tidy\_HAR\_data.R is read into a data frame.

The Aggregate function is used on the data frame to calculate the means using subject and activity columns to subset the data.

The results are written to a csv file titled “tidy\_HAR\_means.csv” and returned to the calling function.