Code Book

for

Getting and Cleaning Data Course Project

Peer-graded Assignment

Information for this assignment was downloaded from the following URL: https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip and then unzipped into a folder on my computer. Unfortunately, my computer will not allow me to download the files as per the course lessons (gave me a “permissions denied” warning), however, once I unzipped the files on my computer, I could use the files.

The zip file contained a combination of 6 folders and text documents. Once all folders were delved into, a total of 28 text files were discovered. Therefore, the first step (after unzipping) was to load text files into R.

The following variables were assigned information from these text files:

* test\_features: this information is from "X\_test.txt" . According to the data’s README file, this information is the “test set”
* test\_activity: this information is from "y\_test.txt" . According to the data’s README file, this information is the “test labels”
* test\_subject : this information is from "subject\_test.txt" . According to the data’s README file, the information in each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30.
* train\_features: this information is from "X\_train.txt" . According to the data’s README file, this information is the “training set”
* train\_activity: this information is from "y\_train.txt" . According to the data’s README file, this information is the “training labels”
* train\_subject: this information is from "subject\_train.txt" . According to the data’s README file, the information in each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30.
* name\_features: this information is from "features.txt" . According to the data’s README file, this includes a “List of all features”
* activity\_labels this information is from "activity\_labels.txt" . According to the data’s README file, this “links the class labels with their activity name”

The first requirement for this assignment is to merge the training and data sets to create one data set. This was accomplished by using R’s “rbind” feature and the test and train data sets were merged according to “subject”, “activity”, and “features”. Columns were built according to the data set’s “features”, to include the Activity Name and Subject. cbind was used to merge “subject”, “activity”, and “features” into the single data set, designated as “merged\_data” for this project.

The next requirement (#2) was to extract only the measurements on the mean and standard deviation for each measurement. “grep” was used to pattern match the “mean” and “std” (standard deviation) into “mean\_std\_columns”. The variable “need\_columns” was built to aid the final variable of “extract\_data”, which contains only the measurements on the mean and standard deviation for each measurement.

Requirement #3 directed to use descriptive activity names to name the activities in the data set. A “for” loop was built to accomplish this.

Requirement #4 directed assigning labels the data set with descriptive variable names. Shortened names were substituted with more descriptive names using “gsub” which replaced of the all matches respectively. The variable names were checked before and after the gsub iterations to confirm that the substitutions were completed.

The final requirement, #5, instructed to, from the data set in step 4, create a second, independent tidy data set (“tidy\_data”) with the average of each variable for each activity and each subject. “aggregate” was utilized to split the data into subsets (SubjectName and ActivityName), compute summary statistics (the mean) for each, and return the result in a convenient form. This tidy data set was captured in the “Tidy\_data\_set” text file and is submitted with this project.