CodeBook.md

This codebook details the transformations and work performed on cleaning up the data for analysis.

Source Dataset

The source dataset is originally from the following location:

`http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones`

This is made available for Coursera students via an AWS CloudFront content delivery network, at the following location:

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

This file was downloaded via the following R command into the dataset.zip file:

```
download.file("https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Da
```

Once unzipped, the data is contained in a directory with the following structure:

- UCI HAR Dataset
  - test
    - Inertial Signals
  - train
    - Inertial Signals

The data is described in the file "UCI HAR Dataset/README.txt" - which refers to other fiels such as features.txt, activity_labels.txt, etc. For this particular assignment we are only interested in the mean and standard deviation for each measurement.

Transformations

Below are the steps used to transform and analyze the data

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

- In order to complete this step it is first necessary to extract the data, then read the different components of the test and train data sets and combine them, and finally merge the training and test data sets. Unfortunately, the each of the test and train statsets are broken up into several different files, so we have some work to do before they are ready for merging.
- First we unzip the dataset and extract all files with the following command:

  ```
  unzip("dataset.zip")
  ```

- Next we read information on data.
  - The features.txt file contains the feature name for each element in the vector X_text.txt. It will be used as the descriptive column names for the tidy data set. We load the table and rename the colums with the following commands:

    ```
    feature_names <- read.table("UCI HAR Dataset/features.txt", col.names=c("elemen
    ```

  - The activity_lablels.txt file is a short table that contains the activite codes and names. Like features, it will be used to label the activities with descriptive names. We load the table and rename the columns with the following command:

    ```
    activities <- read.table("UCI HAR Dataset/activity_labels.txt", col.names=c("co
    ```

- Read test data
  - subject_test.txt contains the subject number for each corresponding row in X_text.txt

    ```
    test_subject <- read.table("UCI HAR Dataset/test/subject_test.txt")
    ```

  - y_test.txt contains the activity code for each corresponding row in X_text.txt

    ```
    test_activity <- read.table("UCI HAR Dataset/test/y_test.txt")
    ```

  - X_test.txt is the 561 element vector, one per row, for each test performed

    ```
    test_feature <- read.table("UCI HAR Dataset/test/X_test.txt")
    ```

- Read the train data
  - subject_train.txt contains the subject number for each corresponding row in X_train.txt

    ```
    train_subject <- read.table("UCI HAR Dataset/train/subject_train.txt")
    ```

  - y_train.txt contains the activity code for each corresponding row in X_train.txt

    ```
    train_activity <- read.table("UCI HAR Dataset/train/y_train.txt")
    ```

  - X_train.txt is the 561 element vector, one per row, for each train performed

```
train_feature <- read.table("UCI HAR Dataset/train/X_train.txt")
```

- Actually merge the test and train data sets

```
subject <- rbind(test_subject, train_subject)
activity <- rbind(test_activity, train_activity)
feature <- rbind(test_feature, train_feature)
```

- Delete component parts, no longer needed

```
rm(train_subject, train_activity, train_feature, test_subject, test_activity, test_
```

2. Extract only the measurements on the mean and standard deviation for each measurement / feature.

   Since the featuers.txt from the original data set says that the mean and standard deviation contain mean() and std() in the feature names, we can use that fact in determining which of the columns we need to extract. grepl provides a logical vector of whether a particular element matches or not. Since the feature names directly correspond to the vector position in the feature table, we can use grepl to select which columns to keep

   ```
   feature <- feature[ , grepl("(mean|std)\\(", feature_names$name)]
   ```

   We also don't need to keep the feature_names that we no longer need, so we can drop them

   ```
   feature_names <- feature_names[grepl("(mean|std)\\(", feature_names$name), c("name")]
   ```

   Replace - with _ in feature names, and remove () in feature names

   ```
   feature_names <- gsub("[()]", "", gsub("[-]", "_", feature_names))
   ```

3. Use descriptive names to name the activities in the data set

   For this step we need to replace the activity code with the activity name. We can then delete the activities codes, as we no longer need them

   ```
   activity <- sapply(activity$V1, function(f) activities[activities$code == f, c("activit
   rm(activities)
   ```

4. Appropriately label the data set with descriptive variable names

   We can combine the feature, activity, and subject now, and label the varaiables appropriately

   ```
   data <- cbind(feature, activity, subject)
   colnames(data) <- c(feature_names, "activity", "subject")
   rm(feature,activity,subject,feature_names)
   write.csv(data, file="tidy_data.csv")
   ```

We have now tidied up our data set!

- Each variable measured is in one column
- Each different observation is in a different row
- One table for each "kind" of variable
- We don't need multiple tables.
  The activity and the subject are the "primary key" to the table as with a database. If there were additional information about the subject, such as name, address, phone number, etc, then that would be in a separate table. If there were additional data about each activity, such as the documentation for performing the activity or who is the designated author of the activity, then that would be in a separate table.

Analysis

The analysis is actually the fifth step of the assignement. Up until now there have been no changing of data or summarizations of the data; we have just been formatting the data into a tidy data set.

5. Create a second, independent tidy data set with the average of each variable for each activity and subject

   ```
   mean_data <- data %>% group_by(activity, subject) %>% summarise_each(funs(mean))
   ```

   Fixup column names!

   ```
   colnames(mean_data) <- c("activity", "subject", paste0("Mean_", colnames(mean_data)[3:n
   ```

   Save new tidy data set

   ```
   write.csv(mean_data, file="mean_tidy_data.csv")
   ```

Data Dictionary

Below is a data dictionary for the data in both the original tidy data set, and the new tidy data set produced in step 5.

Original Tidy Data Set

data table contains statistics of various meter readings normalized and bounded within [-1,1].

Variable	Contents
tBodyAcc-mean()-X	mean of time domain body accelerometer X data
tBodyAcc-mean()-Y	mean of time domain body accelerometer Y data
tBodyAcc-mean()-Z	mean of time domain body accelerometer Z data
tBodyAcc-std()-X	standard deviation of time domain body accelerometer X data
tBodyAcc-std()-Y	standard deviation of time domain body accelerometer Y data
tBodyAcc-std()-Z	standard deviation of time domain body accelerometer Z data
tGravityAcc-mean()-X	mean of time domain gravity accelerometer X data
tGravityAcc-mean()-Y	mean of time domain gravity accelerometer Y data
tGravityAcc-mean()-Z	mean of time domain gravity accelerometer Z data
tGravityAcc-std()-X	standard deviation of time domain gravity accelerometer X data
tGravityAcc-std()-Y	standard deviation of time domain gravity accelerometer Y data
tGravityAcc-std()-Z	standard deviation of time domain gravity accelerometer Z data
tBodyAccJerk-mean()-X	mean of time domain body accelerometer jerk X data
tBodyAccJerk-mean()-Y	mean of time domain body accelerometer jerk Y data
tBodyAccJerk-mean()-Z	mean of time domain body accelerometer jerk Z data
tBodyAccJerk-std()-X	standard deviation of time domain body accelerometer jerk X data
tBodyAccJerk-std()-Y	standard deviation of time domain body accelerometer jerk Y data
tBodyAccJerk-std()-Z	standard deviation of time domain body accelerometer jerk Z data
tBodyGyro-mean()-X	mean of time domain body gyroscope X data
tBodyGyro-mean()-Y	mean of time domain body gyroscope Y data
tBodyGyro-mean()-Z	mean of time domain body gyroscope Z data
tBodyGyro-std()-X	standard deviation of time domain body gyroscope X data
tBodyGyro-std()-Y	standard deviation of time domain body gyroscope Y data
tBodyGyro-std()-Z	standard deviation of time domain body gyroscope Z data
tBodyGyroJerk-mean()-X	mean of time domain gravity gyroscope X data
tBodyGyroJerk-mean()-Y	mean of time domain gravity gyroscope Y data
tBodyGyroJerk-mean()-Z	mean of time domain gravity gyroscope Z data
tBodyGyroJerk-std()-X	standard deviation of time domain gravity gyroscope X data
tBodyGyroJerk-std()-Y	standard deviation of time domain gravity gyroscope Y data
tBodyGyroJerk-std()-Z	standard deviation of time domain gravity gyroscope Z data
tBodyAccMag-mean()	mean of magnitude of time domain body accelerometer via Ecludean norm
tBodyAccMag-std()	standard deviation of magnitude of time domain body accelerometer via E
tGravityAccMag-mean()	mean of magnitude of time domain gravity accelerometer via Ecludean no
tGravityAccMag-std()	standard deviation of magnitude of time domain gravity accelerometer via
tBodyAccJerkMag-mean()	mean of magnitude of time domain body accelerometer jerk via Ecludean
tBodyAccJerkMag-std()	standard deviation of magnitude of time domain body accelerometer jerk
tBodyGyroMag-mean()	mean of magnitude of time domain body gyroscope via Ecludean norm
tBodyGyroMag-std()	standard deviation of magnitude of time domain body gyroscope via Eclu
tBodyGyroJerkMag-mean()	mean of magnitude of time domain body gyroscope jerk via Ecludean nor
tBodyGyroJerkMag-std()	standard deviation of magnitude of time domain body gyroscope jerk via
fBodyAcc-mean()-X	mean of Fast Fourier Transform of body accelerometer X data
fBodyAcc-mean()-Y	mean of Fast Fourier Transform of body accelerometer Y data
fBodyAcc-mean()-Z	mean of Fast Fourier Transform of body accelerometer Z data
fBodyAcc-std()-X	standard deviation of Fast Fourier Transform of body accelerometer X dat

Variable	Contents
fBodyAcc-std()-Y	standard deviation of Fast Fourier Transform of body accelerometer Y dat
fBodyAcc-std()-Z	standard deviation of Fast Fourier Transform of body accelerometer Z dat
fBodyAccJerk-mean()-X	mean of Fast Fourier Transform of body accelerometer jerk X data
fBodyAccJerk-mean()-Y	mean of Fast Fourier Transform of body accelerometer jerk Y data
fBodyAccJerk-mean()-Z	mean of Fast Fourier Transform of body accelerometer jerk Z data
fBodyAccJerk-std()-X	standard deviation of Fast Fourier Transform of body accelerometer jerk X
fBodyAccJerk-std()-Y	standard deviation of Fast Fourier Transform of body accelerometer jerk Y
fBodyAccJerk-std()-Z	standard deviation of Fast Fourier Transform of body accelerometer jerk Z
fBodyGyro-mean()-X	mean of Fast Fourier Transform of body gyroscope X data
fBodyGyro-mean()-Y	mean of Fast Fourier Transform of body gyroscope Y data
fBodyGyro-mean()-Z	mean of Fast Fourier Transform of body gyroscope Z data
fBodyGyro-std()-X	standard deviation of Fast Fourier Transform of body gyroscope X data
fBodyGyro-std()-Y	standard deviation of Fast Fourier Transform of body gyroscope Y data
fBodyGyro-std()-Z	standard deviation of Fast Fourier Transform of body gyroscope Z data
fBodyAccMag-mean()	mean of Fast Fourier Transform of magnitude of body accelerometer via E
fBodyAccMag-std()	standard deviation of Fast Fourier Transform of magnitude of body accele
fBodyBodyAccJerkMag-mean()	mean of Fast Fourier Transform of magnitude of body accelerometer jerk
fBodyBodyAccJerkMag-std()	standard deviation of Fast Fourier Transform of magnitude of body accele
fBodyBodyGyroMag-mean()	mean of Fast Fourier Transform of magnitude of body gyroscope via Eclu
fBodyBodyGyroMag-std()	standard deviation of Fast Fourier Transform of magnitude of body gyrosc
fBodyBodyGyroJerkMag-mean()	mean of Fast Fourier Transform of magnitude of body gyroscope jerk via
fBodyBodyGyroJerkMag-std()	standard deviation of Fast Fourier Transform of magnitude of body gyrosc
activity	The activity performed for the observed statistis
subject	The subject performing the activity

New Tidy Data Set

mean_data table contains the mean of all observations by the same subject and activity. Statistics of various meter readings normalized and bounded within [-1,1].

Variable	Contents
activity	The activity performed for the observed statistis
subject	The subject performing the activity
Mean_tBodyAcc-mean()-X	mean of mean of time domain body accelerometer X data
Mean_tBodyAcc-mean()-Y	mean of mean of time domain body accelerometer Y data
Mean_tBodyAcc-mean()-Z	mean of mean of time domain body accelerometer Z data
Mean_tBodyAcc-std()-X	mean of standard deviation of time domain body accelerometer X
Mean_tBodyAcc-std()-Y	mean of standard deviation of time domain body accelerometer Y
Mean_tBodyAcc-std()-Z	mean of standard deviation of time domain body accelerometer Z
Mean_tGravityAcc-mean()-X	mean of mean of time domain gravity accelerometer X data
Mean_tGravityAcc-mean()-Y	mean of mean of time domain gravity accelerometer Y data

Variable	Contents
Mean__tGravityAcc-mean()-Z	mean of mean of time domain gravity accelerometer Z data
Mean__tGravityAcc-std()-X	mean of standard deviation of time domain gravity accelerometer X
Mean__tGravityAcc-std()-Y	mean of standard deviation of time domain gravity accelerometer Y
Mean__tGravityAcc-std()-Z	mean of standard deviation of time domain gravity accelerometer Z
Mean__tBodyAccJerk-mean()-X	mean of mean of time domain body accelerometer jerk X data
Mean__tBodyAccJerk-mean()-Y	mean of mean of time domain body accelerometer jerk Y data
Mean__tBodyAccJerk-mean()-Z	mean of mean of time domain body accelerometer jerk Z data
Mean__tBodyAccJerk-std()-X	mean of standard deviation of time domain body accelerometer jer
Mean__tBodyAccJerk-std()-Y	mean of standard deviation of time domain body accelerometer jer
Mean__tBodyAccJerk-std()-Z	mean of standard deviation of time domain body accelerometer jer
Mean__tBodyGyro-mean()-X	mean of mean of time domain body gyroscope X data
Mean__tBodyGyro-mean()-Y	mean of mean of time domain body gyroscope Y data
Mean__tBodyGyro-mean()-Z	mean of mean of time domain body gyroscope Z data
Mean__tBodyGyro-std()-X	mean of standard deviation of time domain body gyroscope X data
Mean__tBodyGyro-std()-Y	mean of standard deviation of time domain body gyroscope Y data
Mean__tBodyGyro-std()-Z	mean of standard deviation of time domain body gyroscope Z data
Mean__tBodyGyroJerk-mean()-X	mean of mean of time domain gravity gyroscope X data
Mean__tBodyGyroJerk-mean()-Y	mean of mean of time domain gravity gyroscope Y data
Mean__tBodyGyroJerk-mean()-Z	mean of mean of time domain gravity gyroscope Z data
Mean__tBodyGyroJerk-std()-X	mean of standard deviation of time domain gravity gyroscope X da
Mean__tBodyGyroJerk-std()-Y	mean of standard deviation of time domain gravity gyroscope Y da
Mean__tBodyGyroJerk-std()-Z	mean of standard deviation of time domain gravity gyroscope Z da
Mean__tBodyAccMag-mean()	mean of mean of magnitude of time domain body accelerometer via
Mean__tBodyAccMag-std()	mean of standard deviation of magnitude of time domain body acc
Mean__tGravityAccMag-mean()	mean of mean of magnitude of time domain gravity accelerometer
Mean__tGravityAccMag-std()	mean of standard deviation of magnitude of time domain gravity a
Mean__tBodyAccJerkMag-mean()	mean of mean of magnitude of time domain body accelerometer jer
Mean__tBodyAccJerkMag-std()	mean of standard deviation of magnitude of time domain body acc
Mean__tBodyGyroMag-mean()	mean of mean of magnitude of time domain body gyroscope via Eu
Mean__tBodyGyroMag-std()	mean of standard deviation of magnitude of time domain body gyr
Mean__tBodyGyroJerkMag-mean()	mean of mean of magnitude of time domain body gyroscope jerk v
Mean__tBodyGyroJerkMag-std()	mean of standard deviation of magnitude of time domain body gyr
Mean__fBodyAcc-mean()-X	mean of mean of Fast Fourier Transform of body accelerometer X d
Mean__fBodyAcc-mean()-Y	mean of mean of Fast Fourier Transform of body accelerometer Y d
Mean__fBodyAcc-mean()-Z	mean of mean of Fast Fourier Transform of body accelerometer Z d
Mean__fBodyAcc-std()-X	mean of standard deviation of Fast Fourier Transform of body acce
Mean__fBodyAcc-std()-Y	mean of standard deviation of Fast Fourier Transform of body acce
Mean__fBodyAcc-std()-Z	mean of standard deviation of Fast Fourier Transform of body acce
Mean__fBodyAccJerk-mean()-X	mean of mean of Fast Fourier Transform of body accelerometer jer
Mean__fBodyAccJerk-mean()-Y	mean of mean of Fast Fourier Transform of body accelerometer jer
Mean__fBodyAccJerk-mean()-Z	mean of mean of Fast Fourier Transform of body accelerometer jer
Mean__fBodyAccJerk-std()-X	mean of standard deviation of Fast Fourier Transform of body acce
Mean__fBodyAccJerk-std()-Y	mean of standard deviation of Fast Fourier Transform of body acce
Mean__fBodyAccJerk-std()-Z	mean of standard deviation of Fast Fourier Transform of body acce

Variable	Contents
Mean_fBodyGyro-mean()-X	mean of mean of Fast Fourier Transform of body gyroscope X data
Mean_fBodyGyro-mean()-Y	mean of mean of Fast Fourier Transform of body gyroscope Y data
Mean_fBodyGyro-mean()-Z	mean of mean of Fast Fourier Transform of body gyroscope Z data
Mean_fBodyGyro-std()-X	mean of standard deviation of Fast Fourier Transform of body gyro
Mean_fBodyGyro-std()-Y	mean of standard deviation of Fast Fourier Transform of body gyro
Mean_fBodyGyro-std()-Z	mean of standard deviation of Fast Fourier Transform of body gyro
Mean_fBodyAccMag-mean()	mean of mean of Fast Fourier Transform of magnitude of body acc
Mean_fBodyAccMag-std()	mean of standard deviation of Fast Fourier Transform of magnitud
Mean_fBodyBodyAccJerkMag-mean()	mean of mean of Fast Fourier Transform of magnitude of body acc
Mean_fBodyBodyAccJerkMag-std()	mean of standard deviation of Fast Fourier Transform of magnitud
Mean_fBodyBodyGyroMag-mean()	mean of mean of Fast Fourier Transform of magnitude of body gyr
Mean_fBodyBodyGyroMag-std()	mean of standard deviation of Fast Fourier Transform of magnitud
Mean_fBodyBodyGyroJerkMag-mean()	mean of mean of Fast Fourier Transform of magnitude of body gyr
Mean_fBodyBodyGyroJerkMag-std()	mean of standard deviation of Fast Fourier Transform of magnitud