Code Book

Variables

The following number counts the original merged data frame's variables:

```r
ncol(all_combined)
```

[1] 1718

This dataframe contained all variables provided in the original text files, including:

- Subject label

- Activity code

- Activity label

- 561 features (i.e. measures derived from the base measurements, e.g. standard deviation, Jerk signals, etc.)

- window level data for x, y, z axes for total acceleration (total_acc), body velocity (body_gyro), and body acceleration (body_acc); there are 128 readings/window (**NOTE: each window relates to one activity and one subject, but many windows are recorded for each activity per subject**)

The only variables I created in this original merged dataframe were:

- partition (i.e. train or test)

- window (i.e. counted which window for each subject)

For the second objective, I limited the dataframe to include only the measurements on the mean and the standard deviation for each measurement.

You can see the limited variable names here:

```r
names(combined_final)
```

```
##  [1] "subject"           "activity_code"     "activity"
##  [4] "window"            "partition"         "tBodyAcc-mean()-X"
##  [7] "tBodyAcc-mean()-Y" "tBodyAcc-mean()-Z" "tBodyAcc-std()-X"
## [10] "tBodyAcc-std()-Y"  "tBodyAcc-std()-Z"  "tBodyGyro-mean()-X"
## [13] "tBodyGyro-mean()-Y" "tBodyGyro-mean()-Z" "tBodyGyro-std()-X"
## [16] "tBodyGyro-std()-Y"  "tBodyGyro-std()-Z"  "fBodyAcc-mean()-X"
## [19] "fBodyAcc-mean()-Y"  "fBodyAcc-mean()-Z"  "fBodyAcc-std()-X"
## [22] "fBodyAcc-std()-Y"   "fBodyAcc-std()-Z"   "fBodyGyro-mean()-X"
## [25] "fBodyGyro-mean()-Y" "fBodyGyro-mean()-Z" "fBodyGyro-std()-X"
## [28] "fBodyGyro-std()-Y"  "fBodyGyro-std()-Z"
```

Furthermore, you can see how I descriptively renamed the variables as directed in the fourth objective with the following variable names:

```r
names(combined_rename)
```

```
## [1] "subject"          "activity_code"     "activity"
## [4] "window"           "partition"         "t.bodyacc.mean.X"
## [7] "t.bodyacc.mean.Y" "t.bodyacc.mean.Z" "t.bodyacc.std.X"
## [10] "t.bodyacc.std.Y"  "t.bodyacc.std.Z"  "t.bodygyro.mean.X"
## [13] "t.bodygyro.mean.Y" "t.bodygyro.mean.Z" "t.bodygyro.std.X"
## [16] "t.bodygyro.std.Y"  "t.bodygyro.std.Z"  "f.bodyacc.mean.X"
## [19] "f.bodyacc.mean.Y" "f.bodyacc.mean.Z" "f.bodyacc.std.X"
## [22] "f.bodyacc.std.Y"  "f.bodyacc.std.Z"  "f.bodygyro.mean.X"
## [25] "f.bodygyro.mean.Y" "f.bodygyro.mean.Z" "f.bodygyro.std.X"
## [28] "f.bodygyro.std.Y"  "f.bodygyro.std.Z"
```

My final tidy data set had the following variable names:
NOTE: the prefix 'sub' stands for the subject's averages for each activity and the prefix 'act' stands for the activity's overall averages

```
names(combined_means)
```

```
## [1] "subject"                "activity"               "sub.t.bodyacc.mean.X"
## [4] "sub.t.bodyacc.mean.Y"   "sub.t.bodyacc.mean.Z"   "sub.t.bodyacc.std.X"
## [7] "sub.t.bodyacc.std.Y"    "sub.t.bodyacc.std.Z"    "sub.t.bodygyro.mean.X"
## [10] "sub.t.bodygyro.mean.Y" "sub.t.bodygyro.mean.Z" "sub.t.bodygyro.std.X"
## [13] "sub.t.bodygyro.std.Y"  "sub.t.bodygyro.std.Z"  "sub.f.bodyacc.mean.X"
## [16] "sub.f.bodyacc.mean.Y"  "sub.f.bodyacc.mean.Z"  "sub.f.bodyacc.std.X"
## [19] "sub.f.bodyacc.std.Y"   "sub.f.bodyacc.std.Z"   "sub.f.bodygyro.mean.X"
## [22] "sub.f.bodygyro.mean.Y" "sub.f.bodygyro.mean.Z" "sub.f.bodygyro.std.X"
## [25] "sub.f.bodygyro.std.Y"  "sub.f.bodygyro.std.Z"  "act.t.bodyacc.mean.X"
## [28] "act.t.bodyacc.mean.Y"  "act.t.bodyacc.mean.Z"  "act.t.bodyacc.std.X"
## [31] "act.t.bodyacc.std.Y"   "act.t.bodyacc.std.Z"   "act.t.bodygyro.mean.X"
## [34] "act.t.bodygyro.mean.Y" "act.t.bodygyro.mean.Z" "act.t.bodygyro.std.X"
## [37] "act.t.bodygyro.std.Y"  "act.t.bodygyro.std.Z"  "act.f.bodyacc.mean.X"
## [40] "act.f.bodyacc.mean.Y"  "act.f.bodyacc.mean.Z"  "act.f.bodyacc.std.X"
## [43] "act.f.bodyacc.std.Y"   "act.f.bodyacc.std.Z"   "act.f.bodygyro.mean.X"
## [46] "act.f.bodygyro.mean.Y" "act.f.bodygyro.mean.Z" "act.f.bodygyro.std.X"
## [49] "act.f.bodygyro.std.Y"  "act.f.bodygyro.std.Z"
```

Data

The dataset is from an experiement done on the wearable device, the Samsung Galaxy S smartphone. With the device's accelometer and gyroscope, 30 subjects' total acceleration and velocity were measured as they performed six "activities of daily living (ADL)." These activities were: laying, sitting, standing, walking, walking downstairs, and walking upstairs.

The total acceleration was then transformed to parse out an estimated body acceleration. The experiment designers then processed the data down to captured windows of each activity, all with 128 readings/window.

For information directly from the experiment about the experiment, please reference the following link: http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

The data for this analysis was merged from 32 text files. There were three sections to pull from: the structural files, the training set, and the test set. Below are descriptions of each section's files per the experiment's README.txt file.

The structural files loaded for this analysis include:

- 'activity_labels.txt': Links the class labels with their activity name.

- 'features.txt': List of all features.

"The following files are available for the train and test data. Their descriptions are equivalent.

- 'train/subject_train.txt': Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30.

- 'train/Inertial Signals/total_acc_x_train.txt': The acceleration signal from the smartphone accelerometer X axis in standard gravity units 'g'. Every row shows a 128 element vector. The same description applies for the 'total_acc_x_train.txt' and 'total_acc_z_train.txt' files for the Y and Z axis.

- 'train/Inertial Signals/body_acc_x_train.txt': The body acceleration signal obtained by subtracting the gravity from the total acceleration.

- 'train/Inertial Signals/body_gyro_x_train.txt': The angular velocity vector measured by the gyroscope for each window sample. The units are radians/second."

The tidy data I created in run_analysis.R illustrated each observation as one row, as shown below.

```
## # A tibble: 10 x 50
## # Groups:   subject [2]
##    subject activity sub.t.bodyacc.m... sub.t.bodyacc.m... sub.t.bodyacc.m...
##      <int> <fct>                 <dbl>              <dbl>              <dbl>
## 1        1 LAYING                0.265            -0.0170             -0.100
## 2        1 SITTING               0.268            -0.0125             -0.135
## 3        1 STANDING              0.273            -0.0169             -0.118
## 4        1 WALKING               0.247            -0.0233            -0.0930
## 5        1 WALKING...            0.266            -0.0200             -0.102
## 6        1 WALKING...            0.290            -0.0154             -0.113
## 7        2 LAYING                0.269            -0.0269             -0.121
## 8        2 SITTING               0.276            -0.0178             -0.106
## 9        2 STANDING              0.278            -0.0176             -0.107
## 10       2 WALKING               0.277            -0.0181             -0.107
## # ... with 45 more variables: sub.t.bodyacc.std.X <dbl>,
## #   sub.t.bodyacc.std.Y <dbl>, sub.t.bodyacc.std.Z <dbl>,
## #   sub.t.bodygyro.mean.X <dbl>, sub.t.bodygyro.mean.Y <dbl>,
## #   sub.t.bodygyro.mean.Z <dbl>, sub.t.bodygyro.std.X <dbl>,
## #   sub.t.bodygyro.std.Y <dbl>, sub.t.bodygyro.std.Z <dbl>,
## #   sub.f.bodyacc.mean.X <dbl>, sub.f.bodyacc.mean.Y <dbl>,
## #   sub.f.bodyacc.mean.Z <dbl>, sub.f.bodyacc.std.X <dbl>,
## #   sub.f.bodyacc.std.Y <dbl>, sub.f.bodyacc.std.Z <dbl>,
## #   sub.f.bodygyro.mean.X <dbl>, sub.f.bodygyro.mean.Y <dbl>,
## #   sub.f.bodygyro.mean.Z <dbl>, sub.f.bodygyro.std.X <dbl>,
## #   sub.f.bodygyro.std.Y <dbl>, sub.f.bodygyro.std.Z <dbl>,
## #   act.t.bodyacc.mean.X <dbl>, act.t.bodyacc.mean.Y <dbl>,
## #   act.t.bodyacc.mean.Z <dbl>, act.t.bodyacc.std.X <dbl>,
## #   act.t.bodyacc.std.Y <dbl>, act.t.bodyacc.std.Z <dbl>,
## #   act.t.bodygyro.mean.X <dbl>, act.t.bodygyro.mean.Y <dbl>,
## #   act.t.bodygyro.mean.Z <dbl>, act.t.bodygyro.std.X <dbl>,
## #   act.t.bodygyro.std.Y <dbl>, act.t.bodygyro.std.Z <dbl>,
## #   act.f.bodyacc.mean.X <dbl>, act.f.bodyacc.mean.Y <dbl>,
## #   act.f.bodyacc.mean.Z <dbl>, act.f.bodyacc.std.X <dbl>,
## #   act.f.bodyacc.std.Y <dbl>, act.f.bodyacc.std.Z <dbl>,
## #   act.f.bodygyro.mean.X <dbl>, act.f.bodygyro.mean.Y <dbl>,
## #   act.f.bodygyro.mean.Z <dbl>, act.f.bodygyro.std.X <dbl>,
## #   act.f.bodygyro.std.Y <dbl>, act.f.bodygyro.std.Z <dbl>
```

Transformations

I already documented a lot of my transformations in the variable section of this codebook (see above).

From the get-go, I combined the descriptive actity label with the activity code as required in the third objective of this assignment. I did this to create the tidiest, most descriptive original data set in the initial merging of the training and testing data.