**Code Book**

This code book will describe the data used in this project, as well as the processing required to create the resulting tidy data set.

**Overview**

30 volunteers performed 6 different activities while wearing a smartphone. The smartphone captured various data about their movements.

**Explanation of each file**

* features.txt: Names of the 561 features.
* activity\_labels.txt: Names and IDs for each of the 6 activities.
* X\_train.txt: 7352 observations of the 561 features, for 21 of the 30 volunteers.
* subject\_train.txt: A vector of 7352 integers, denoting the ID of the volunteer related to each of the observations in X\_train.txt.
* y\_train.txt: A vector of 7352 integers, denoting the ID of the activity related to each of the observations in X\_train.txt.
* X\_test.txt: 2947 observations of the 561 features, for 9 of the 30 volunteers.
* subject\_test.txt: A vector of 2947 integers, denoting the ID of the volunteer related to each of the observations in X\_test.txt.
* y\_test.txt: A vector of 2947 integers, denoting the ID of the activity related to each of the observations in X\_test.txt.

More information about the files is available in README.txt

**Data files that were not used**

This analysis was performed using only the files above, and did not use the raw signal data. Therefore, the data files in the "Inertial Signals" folders were ignored.

**Processing steps**

1. All of the relevant data files were read into data frames, appropriate column headers were added, and the training and test sets were combined into a single data set.
2. All feature columns were removed that did not contain the exact string "mean()" or "std()". This left 66 feature columns, plus the subjectID and activity columns.
3. The activity column was converted from a integer to a factor, using labels describing the activities.
4. A tidy data set was created containing the mean of each feature for each subject and each activity. Thus, subject #1 has 6 rows in the tidy data set (one row for each activity), and each row contains the mean value for each of the 66 features for that subject/activity combination. Since there are 30 subjects, there are a total of 180 rows.
5. The tidy data set was output to a txt file.