**Code book**

Loaded packages: dplyr

Variables

The original dataset was loaded into the following variables:

* trainData/testData – data.frame with numeric values of the measurements taken from the X\_train.txt/X\_test.txt files respectively
* trainSubjects/testSubjects – data.frame with one numeric column, consisting of a list of subject numbers that correspond to the rows of trainData/testData, taken from subject\_train.txt/subject\_test.txt
* trainLabels/testLabels – data.frame with one numeric column, consisting of a list of activity label numbers that correspond to the rows of trainData/testData, taken from y\_train.txt/y\_test.txt
* activityLabels – data.frame with two columns, a numeric column corresponding to unique activity label numbers and a text column with the label names
* features - data.frame with two columns, a numeric column corresponding to column numbers of the data and a text column with the names of the columns (features)

Manipulation of data

Based on these variables, a merged dataset was created:

mergedData – data.frame with one numeric column of the subject, one char column with the activity label of the measurement, one char column indicating whether this is from the training or test set and 66 columns with the mean (33) and std (33) features, using the column names from the features variable. Note, prior to using the features as column names they were cleaned (parentheses and dashes removed).

Tidy data set

A tidy data set was then created with the summary of the average measurement for each combination of subjectXactivity (combining data from training and test), using group\_by to first group the data by activity and subject.

dataSummary – data.frame with one numeric column of the subject, one char column with the activity label , and 66 columns corresponding to the 66 features, same column names as mergedData. One row per each combination of activity and subject (6 activities X 30 subjects, total of 180 rows)