Code Book

James Watt

11/1/2020

The variables in **run_analysis.R** are extracted from wearable computer data generated by the UCI Machine Learning Repository and represent movement data across three dimensions for various sensor signals from an accelerometer and gyroscope. Thirty participants wore a Samsung device and completed six activities. During these activities, repeated measurements of a total of 561 individual variables were measured or calculated by the investigators.

For the purposes of this analysis, specific variables were extracted (described below). These extracted variables were merged with subject and activity data. These data were further reduced to the mean value of the repeated measures of each variable by subject and activity into a tidy data set for further analysis of motion parameters.

Subject Variable   Subjects are numbered 1-30. Subjects 2,4,9,10,12,13,18,20 and 24 were listed in the 'test' data set; remaining subjects were listed in the 'train' data set. These were merged into the 'full' data set (see README). No further re-coding was done for the subject variable.

Activity Variables   Activities were numbered 1-6 and corresponding descriptive activities were listed in the 'activity_labels.txt' file. Following the full data merge (see README.md), descriptive names were added as 'activity'. The list of 'activities' codes and their respective 'activity' descriptives are:

```
##   activities          activity
## 1          1           WALKING
## 2          2   WALKING_UPSTAIRS
## 3          3 WALKING_DOWNSTAIRS
## 4          4           SITTING
## 5          5          STANDING
## 6          6            LAYING
```

Extracted Variables   The variables extracted from the measured or calculated features were those that reported a mean value or a standard deviation value (std). To extract these variables, variable names were extracted, using the **grepl()** function, from the full, merged data set (see README.md) where names included "mean()" or "std()".

```
subset <- full[,grepl("mean()", names(full))|grepl("std()", names(full))]
```

The extracted variables (n = 79) are as follows:

```
##  [1] "tBodyAcc-mean()-X"        "tBodyAcc-mean()-Y"
##  [3] "tBodyAcc-mean()-Z"        "tBodyAcc-std()-X"
##  [5] "tBodyAcc-std()-Y"         "tBodyAcc-std()-Z"
##  [7] "tGravityAcc-mean()-X"     "tGravityAcc-mean()-Y"
##  [9] "tGravityAcc-mean()-Z"     "tGravityAcc-std()-X"
## [11] "tGravityAcc-std()-Y"      "tGravityAcc-std()-Z"
## [13] "tBodyAccJerk-mean()-X"    "tBodyAccJerk-mean()-Y"
## [15] "tBodyAccJerk-mean()-Z"    "tBodyAccJerk-std()-X"
```

```
## [17] "tBodyAccJerk-std()-Y"          "tBodyAccJerk-std()-Z"
## [19] "tBodyGyro-mean()-X"            "tBodyGyro-mean()-Y"
## [21] "tBodyGyro-mean()-Z"            "tBodyGyro-std()-X"
## [23] "tBodyGyro-std()-Y"             "tBodyGyro-std()-Z"
## [25] "tBodyGyroJerk-mean()-X"        "tBodyGyroJerk-mean()-Y"
## [27] "tBodyGyroJerk-mean()-Z"        "tBodyGyroJerk-std()-X"
## [29] "tBodyGyroJerk-std()-Y"         "tBodyGyroJerk-std()-Z"
## [31] "tBodyAccMag-mean()"            "tBodyAccMag-std()"
## [33] "tGravityAccMag-mean()"         "tGravityAccMag-std()"
## [35] "tBodyAccJerkMag-mean()"        "tBodyAccJerkMag-std()"
## [37] "tBodyGyroMag-mean()"           "tBodyGyroMag-std()"
## [39] "tBodyGyroJerkMag-mean()"       "tBodyGyroJerkMag-std()"
## [41] "fBodyAcc-mean()-X"             "fBodyAcc-mean()-Y"
## [43] "fBodyAcc-mean()-Z"             "fBodyAcc-std()-X"
## [45] "fBodyAcc-std()-Y"              "fBodyAcc-std()-Z"
## [47] "fBodyAcc-meanFreq()-X"         "fBodyAcc-meanFreq()-Y"
## [49] "fBodyAcc-meanFreq()-Z"         "fBodyAccJerk-mean()-X"
## [51] "fBodyAccJerk-mean()-Y"         "fBodyAccJerk-mean()-Z"
## [53] "fBodyAccJerk-std()-X"          "fBodyAccJerk-std()-Y"
## [55] "fBodyAccJerk-std()-Z"          "fBodyAccJerk-meanFreq()-X"
## [57] "fBodyAccJerk-meanFreq()-Y"     "fBodyAccJerk-meanFreq()-Z"
## [59] "fBodyGyro-mean()-X"            "fBodyGyro-mean()-Y"
## [61] "fBodyGyro-mean()-Z"            "fBodyGyro-std()-X"
## [63] "fBodyGyro-std()-Y"             "fBodyGyro-std()-Z"
## [65] "fBodyGyro-meanFreq()-X"        "fBodyGyro-meanFreq()-Y"
## [67] "fBodyGyro-meanFreq()-Z"        "fBodyAccMag-mean()"
## [69] "fBodyAccMag-std()"             "fBodyAccMag-meanFreq()"
## [71] "fBodyBodyAccJerkMag-mean()"    "fBodyBodyAccJerkMag-std()"
## [73] "fBodyBodyAccJerkMag-meanFreq()" "fBodyBodyGyroMag-mean()"
## [75] "fBodyBodyGyroMag-std()"        "fBodyBodyGyroMag-meanFreq()"
## [77] "fBodyBodyGyroJerkMag-mean()"   "fBodyBodyGyroJerkMag-std()"
## [79] "fBodyBodyGyroJerkMag-meanFreq()"
```

To clarify the variable names, the variables (features) were re-named using descriptive titles based on information from the 'features_info.txt' file provided by the data source. This was accomplished by the **gsub()** function and piping using the **magrittr** package. Special characters (e.g. '()') were eliminated; abbreviations were extended, 'time domain' and 'frequency domain' were added where appropriate; and "mean of" was appended to the front of each variable to indicate that the reported variables (transformed below) are mean values of the repeated measures. All variable names were changed to lowercase. The code follows ('featuresbot' is a vector of variable names):

```
featuresbot %<>%
        gsub("Acc", "accelerometer ", .) %>%
        gsub("Gyro", "gyroscope ", .) %>%
        gsub("Jerk", "jerksignal ", .) %>%
        gsub("Mag", "magnitude ", .) %>%
        gsub("^tBody", "time domain body ", .) %>%
        gsub("^fBody", "frequency domain body ", .) %>%
        gsub("Body", "", .) %>%
        gsub("^tGravity", "time domain gravity ", .) %>%
        gsub("-mean", "mean", .) %>%
        gsub("-std", "std", .) %>%
        gsub("\\()", "", .) %>%
        gsub("\\-", " ", .) %>%
        gsub("meanFreq", "mean frequency", .) %>%
```

```
    tolower() %>%
    paste("mean of", ., sep = " ")
```

Updated variable names (n = 79) are as follows, where "x", "y", and "z" are measurements in the x, y, and z direction, respectively:

```
##  [1] "mean of time domain body accelerometer mean x"
##  [2] "mean of time domain body accelerometer mean y"
##  [3] "mean of time domain body accelerometer mean z"
##  [4] "mean of time domain body accelerometer std x"
##  [5] "mean of time domain body accelerometer std y"
##  [6] "mean of time domain body accelerometer std z"
##  [7] "mean of time domain gravity accelerometer mean x"
##  [8] "mean of time domain gravity accelerometer mean y"
##  [9] "mean of time domain gravity accelerometer mean z"
## [10] "mean of time domain gravity accelerometer std x"
## [11] "mean of time domain gravity accelerometer std y"
## [12] "mean of time domain gravity accelerometer std z"
## [13] "mean of time domain body accelerometer jerksignal mean x"
## [14] "mean of time domain body accelerometer jerksignal mean y"
## [15] "mean of time domain body accelerometer jerksignal mean z"
## [16] "mean of time domain body accelerometer jerksignal std x"
## [17] "mean of time domain body accelerometer jerksignal std y"
## [18] "mean of time domain body accelerometer jerksignal std z"
## [19] "mean of time domain body gyroscope mean x"
## [20] "mean of time domain body gyroscope mean y"
## [21] "mean of time domain body gyroscope mean z"
## [22] "mean of time domain body gyroscope std x"
## [23] "mean of time domain body gyroscope std y"
## [24] "mean of time domain body gyroscope std z"
## [25] "mean of time domain body gyroscope jerksignal mean x"
## [26] "mean of time domain body gyroscope jerksignal mean y"
## [27] "mean of time domain body gyroscope jerksignal mean z"
## [28] "mean of time domain body gyroscope jerksignal std x"
## [29] "mean of time domain body gyroscope jerksignal std y"
## [30] "mean of time domain body gyroscope jerksignal std z"
## [31] "mean of time domain body accelerometer magnitude mean"
## [32] "mean of time domain body accelerometer magnitude std"
## [33] "mean of time domain gravity accelerometer magnitude mean"
## [34] "mean of time domain gravity accelerometer magnitude std"
## [35] "mean of time domain body accelerometer jerksignal magnitude mean"
## [36] "mean of time domain body accelerometer jerksignal magnitude std"
## [37] "mean of time domain body gyroscope magnitude mean"
## [38] "mean of time domain body gyroscope magnitude std"
## [39] "mean of time domain body gyroscope jerksignal magnitude mean"
## [40] "mean of time domain body gyroscope jerksignal magnitude std"
## [41] "mean of frequency domain body accelerometer mean x"
## [42] "mean of frequency domain body accelerometer mean y"
## [43] "mean of frequency domain body accelerometer mean z"
## [44] "mean of frequency domain body accelerometer std x"
## [45] "mean of frequency domain body accelerometer std y"
## [46] "mean of frequency domain body accelerometer std z"
## [47] "mean of frequency domain body accelerometer mean frequency x"
## [48] "mean of frequency domain body accelerometer mean frequency y"
```

```
## [49] "mean of frequency domain body accelerometer mean frequency z"
## [50] "mean of frequency domain body accelerometer jerksignal mean x"
## [51] "mean of frequency domain body accelerometer jerksignal mean y"
## [52] "mean of frequency domain body accelerometer jerksignal mean z"
## [53] "mean of frequency domain body accelerometer jerksignal std x"
## [54] "mean of frequency domain body accelerometer jerksignal std y"
## [55] "mean of frequency domain body accelerometer jerksignal std z"
## [56] "mean of frequency domain body accelerometer jerksignal mean frequency x"
## [57] "mean of frequency domain body accelerometer jerksignal mean frequency y"
## [58] "mean of frequency domain body accelerometer jerksignal mean frequency z"
## [59] "mean of frequency domain body gyroscope mean x"
## [60] "mean of frequency domain body gyroscope mean y"
## [61] "mean of frequency domain body gyroscope mean z"
## [62] "mean of frequency domain body gyroscope std x"
## [63] "mean of frequency domain body gyroscope std y"
## [64] "mean of frequency domain body gyroscope std z"
## [65] "mean of frequency domain body gyroscope mean frequency x"
## [66] "mean of frequency domain body gyroscope mean frequency y"
## [67] "mean of frequency domain body gyroscope mean frequency z"
## [68] "mean of frequency domain body accelerometer magnitude mean"
## [69] "mean of frequency domain body accelerometer magnitude std"
## [70] "mean of frequency domain body accelerometer magnitude mean frequency"
## [71] "mean of frequency domain body accelerometer jerksignal magnitude mean"
## [72] "mean of frequency domain body accelerometer jerksignal magnitude std"
## [73] "mean of frequency domain body accelerometer jerksignal magnitude mean frequency"
## [74] "mean of frequency domain body gyroscope magnitude mean"
## [75] "mean of frequency domain body gyroscope magnitude std"
## [76] "mean of frequency domain body gyroscope magnitude mean frequency"
## [77] "mean of frequency domain body gyroscope jerksignal magnitude mean"
## [78] "mean of frequency domain body gyroscope jerksignal magnitude std"
## [79] "mean of frequency domain body gyroscope jerksignal magnitude mean frequency"
```

Calculation of Mean Values   The data were first rearranged to form a *long* data set using **melt()**

```
subset6m <- melt(subset3, id.vars = c("subject", "activities", "activity"))
```

The data were grouped by "subject" x "activity" x "variable", where "variable" was a vector of variable names (as above). Using the **summarize()** function, the mean value of each group was calculated into a data frame named 'mean_subset'.

```
mean_subset <- summarize(subset6m, mean(value))
```

This generated a *long* data frame consisting of subject, activity, variable, and the mean value calculated of that variable's repeated measurement by subject and activity (4 variables). The *long* dataset **mean_subset** consists of 14220 observations.

```
dim(mean_subset)
```

```
## [1] 14220     4
```

```
head(mean_subset)
```

```
## # A tibble: 6 x 4
## # Groups:   subject, activity [1]
##   subject activity variable                                      `mean(value)`
##     <int> <chr>    <fct>                                                 <dbl>
## 1       1 LAYING   mean of time domain body accelerometer mean x         0.222
```

```
## 2          1 LAYING    mean of time domain body accelerometer mean y       -0.0405
## 3          1 LAYING    mean of time domain body accelerometer mean z       -0.113
## 4          1 LAYING    mean of time domain body accelerometer std x        -0.928
## 5          1 LAYING    mean of time domain body accelerometer std y        -0.837
## 6          1 LAYING    mean of time domain body accelerometer std z        -0.826
```

```
tail(mean_subset)
```

```
## # A tibble: 6 x 4
## # Groups:   subject, activity [1]
##   subject activity       variable                                    `mean(value)`
##     <int> <chr>          <fct>                                               <dbl>
## ## 1      30 WALKING_UPST~ mean of frequency domain body gyroscope m~       -0.449
## ## 2      30 WALKING_UPST~ mean of frequency domain body gyroscope m~       -0.151
## ## 3      30 WALKING_UPST~ mean of frequency domain body gyroscope m~       -0.457
## ## 4      30 WALKING_UPST~ mean of frequency domain body gyroscope j~       -0.774
## ## 5      30 WALKING_UPST~ mean of frequency domain body gyroscope j~       -0.791
## ## 6      30 WALKING_UPST~ mean of frequency domain body gyroscope j~       -0.0714
```

To help clarify the data for spreadsheets, a *wide* dataset was generated using **dcast()**.

```
mean_subsetw <- dcast(mean_subset, subject + activity ~ variable,
                      value.var = "mean(value)")
```

This transformed the dataset such that each row contains the subject, activity, and every mean value calculation across the variables. The *wide* dataset **mean_subsetw** has 81 variables (subject, activity, and 79 features) and 180 observations.

```
str(mean_subsetw)
```

```
## 'data.frame':    180 obs. of  81 variables:
## $ subject                                                   : int  1 1 1 1 1 1
## $ activity                                                  : chr  "LAYING" "S~
## $ mean of time domain body accelerometer mean x             : num  0.222 0.261
## $ mean of time domain body accelerometer mean y             : num  -0.04051 -0~
## $ mean of time domain body accelerometer mean z             : num  -0.113 -0.1~
## $ mean of time domain body accelerometer std x              : num  -0.928 -0.9~
## $ mean of time domain body accelerometer std y              : num  -0.8368 -0.~
## $ mean of time domain body accelerometer std z              : num  -0.826 -0.9~
## $ mean of time domain gravity accelerometer mean x          : num  -0.249 0.83~
## $ mean of time domain gravity accelerometer mean y          : num  0.706 0.204
## $ mean of time domain gravity accelerometer mean z          : num  0.4458 0.33~
## $ mean of time domain gravity accelerometer std x           : num  -0.897 -0.9~
## $ mean of time domain gravity accelerometer std y           : num  -0.908 -0.9~
## $ mean of time domain gravity accelerometer std z           : num  -0.852 -0.9~
## $ mean of time domain body accelerometer jerksignal mean x  : num  0.0811 0.07~
## $ mean of time domain body accelerometer jerksignal mean y  : num  0.003838 -0~
## $ mean of time domain body accelerometer jerksignal mean z  : num  0.01083 -0.~
## $ mean of time domain body accelerometer jerksignal std x   : num  -0.9585 -0.~
## $ mean of time domain body accelerometer jerksignal std y   : num  -0.924 -0.98~
## $ mean of time domain body accelerometer jerksignal std z   : num  -0.955 -0.98~
## $ mean of time domain body gyroscope mean x                 : num  -0.0166 -0.~
## $ mean of time domain body gyroscope mean y                 : num  -0.0645 -0.~
## $ mean of time domain body gyroscope mean z                 : num  0.1487 0.06~
## $ mean of time domain body gyroscope std x                  : num  -0.874 -0.9~
## $ mean of time domain body gyroscope std y                  : num  -0.9511 -0.~
## $ mean of time domain body gyroscope std z                  : num  -0.908 -0.94~
```

```
##  $ mean of time domain body gyroscope jerksignal mean x                              : num  -0.1073 -0.(
##  $ mean of time domain body gyroscope jerksignal mean y                              : num  -0.0415 -0.(
##  $ mean of time domain body gyroscope jerksignal mean z                              : num  -0.0741 -0.(
##  $ mean of time domain body gyroscope jerksignal std x                               : num  -0.919 -0.99
##  $ mean of time domain body gyroscope jerksignal std y                               : num  -0.968 -0.99
##  $ mean of time domain body gyroscope jerksignal std z                               : num  -0.958 -0.98
##  $ mean of time domain body accelerometer magnitude mean                             : num  -0.8419 -0.9
##  $ mean of time domain body accelerometer magnitude std                              : num  -0.7951 -0.9
##  $ mean of time domain gravity accelerometer magnitude mean                          : num  -0.8419 -0.9
##  $ mean of time domain gravity accelerometer magnitude std                           : num  -0.7951 -0.9
##  $ mean of time domain body accelerometer jerksignal magnitude mean                  : num  -0.9544 -0.9
##  $ mean of time domain body accelerometer jerksignal magnitude std                   : num  -0.9282 -0.9
##  $ mean of time domain body gyroscope magnitude mean                                 : num  -0.8748 -0.9
##  $ mean of time domain body gyroscope magnitude std                                  : num  -0.819 -0.93
##  $ mean of time domain body gyroscope jerksignal magnitude mean                      : num  -0.963 -0.99
##  $ mean of time domain body gyroscope jerksignal magnitude std                       : num  -0.936 -0.98
##  $ mean of frequency domain body accelerometer mean x                                : num  -0.9391 -0.9
##  $ mean of frequency domain body accelerometer mean y                                : num  -0.86707 -0
##  $ mean of frequency domain body accelerometer mean z                                : num  -0.883 -0.95
##  $ mean of frequency domain body accelerometer std x                                 : num  -0.9244 -0.9
##  $ mean of frequency domain body accelerometer std y                                 : num  -0.834 -0.9
##  $ mean of frequency domain body accelerometer std z                                 : num  -0.813 -0.93
##  $ mean of frequency domain body accelerometer mean frequency x                      : num  -0.1588 -0.(
##  $ mean of frequency domain body accelerometer mean frequency y                      : num  0.0975 0.075
##  $ mean of frequency domain body accelerometer mean frequency z                      : num  0.0894 0.238
##  $ mean of frequency domain body accelerometer jerksignal mean x                     : num  -0.9571 -0.9
##  $ mean of frequency domain body accelerometer jerksignal mean y                     : num  -0.9225 -0.9
##  $ mean of frequency domain body accelerometer jerksignal mean z                     : num  -0.948 -0.98
##  $ mean of frequency domain body accelerometer jerksignal std x                      : num  -0.9642 -0.9
##  $ mean of frequency domain body accelerometer jerksignal std y                      : num  -0.932 -0.98
##  $ mean of frequency domain body accelerometer jerksignal std z                      : num  -0.961 -0.98
##  $ mean of frequency domain body accelerometer jerksignal mean frequency x           : num  0.132 0.257
##  $ mean of frequency domain body accelerometer jerksignal mean frequency y           : num  0.0245 0.04
##  $ mean of frequency domain body accelerometer jerksignal mean frequency z           : num  0.02439 0.09
##  $ mean of frequency domain body gyroscope mean x                                    : num  -0.85 -0.97(
##  $ mean of frequency domain body gyroscope mean y                                    : num  -0.9522 -0.9
##  $ mean of frequency domain body gyroscope mean z                                    : num  -0.9093 -0.9
##  $ mean of frequency domain body gyroscope std x                                     : num  -0.882 -0.97
##  $ mean of frequency domain body gyroscope std y                                     : num  -0.9512 -0.9
##  $ mean of frequency domain body gyroscope std z                                     : num  -0.917 -0.94
##  $ mean of frequency domain body gyroscope mean frequency x                          : num  -0.00355 0.
##  $ mean of frequency domain body gyroscope mean frequency y                          : num  -0.0915 0.0(
##  $ mean of frequency domain body gyroscope mean frequency z                          : num  0.010458 -0
##  $ mean of frequency domain body accelerometer magnitude mean                        : num  -0.8618 -0.9
##  $ mean of frequency domain body accelerometer magnitude std                         : num  -0.798 -0.91
##  $ mean of frequency domain body accelerometer magnitude mean frequency              : num  0.0864 0.23(
##  $ mean of frequency domain body accelerometer jerksignal magnitude mean             : num  -0.9333 -0.9
##  $ mean of frequency domain body accelerometer jerksignal magnitude std              : num  -0.922 -0.98
##  $ mean of frequency domain body accelerometer jerksignal magnitude mean frequency: num  0.2664 0.35
##  $ mean of frequency domain body gyroscope magnitude mean                            : num  -0.862 -0.95
##  $ mean of frequency domain body gyroscope magnitude std                             : num  -0.824 -0.93
##  $ mean of frequency domain body gyroscope magnitude mean frequency                  : num  -0.139775 -(
##  $ mean of frequency domain body gyroscope jerksignal magnitude mean                 : num  -0.942 -0.99
##  $ mean of frequency domain body gyroscope jerksignal magnitude std                  : num  -0.933 -0.98
```

```
##  $ mean of frequency domain body gyroscope jerksignal magnitude mean frequency    : num  0.176 0.185
```