

## Code Book: run\_analysis()

### Variables

#### ACTIVITY

- The type of activity being performed during the measurement
  - WALKING
  - WALKING\_UPSTAIRS
  - WALKING\_DOWNSTAIRS
  - SITTING
  - STANDING
  - LAYING

#### SUBJECT

- Integer identifier to the individual the measurements belong
  - 1 – 30

The following names are the base to multiple variables generating values for standard deviation (std) and mean. Some of which, have separate X, Y, and Z variations.

The following values come from the accelerometer and gyroscope 3-axial raw signals tAcc-XYZ and tGyro-XYZ. These time domain signals (prefix 't' to denote time) were captured at a constant rate of 50 Hz. Then they were filtered using a median filter and a 3rd order low pass Butterworth filter with a corner frequency of 20 Hz to remove noise. Similarly, the acceleration signal was then separated into body and gravity acceleration signals (tBodyAcc-XYZ and tGravityAcc-XYZ) using another low pass Butterworth filter with a corner frequency of 0.3 Hz.

Subsequently, the body linear acceleration and angular velocity were derived in time to obtain Jerk signals (tBodyAccJerk-XYZ and tBodyGyroJerk-XYZ). Also the magnitude of these three-dimensional signals were calculated using the Euclidean norm (tBodyAccMag, tGravityAccMag, tBodyAccJerkMag, tBodyGyroMag, tBodyGyroJerkMag).

Finally a Fast Fourier Transform (FFT) was applied to some of these signals producing fBodyAcc-XYZ, fBodyAccJerk-XYZ, fBodyGyro-XYZ, fBodyAccJerkMag, fBodyGyroMag, fBodyGyroJerkMag. (Note the 'f' to indicate frequency domain signals).

These signals were used to estimate variables of the feature vector for each pattern:  
'-XYZ' is used to denote 3-axial signals in the X, Y and Z directions.

tBodyAcc-XYZ

tGravityAcc-XYZ

tBodyAccJerk-XYZ

tBodyGyro-XYZ

tBodyGyroJerk-XYZ

tBodyAccMag

tGravityAccMag

tBodyAccJerkMag  
tBodyGyroMag  
tBodyGyroJerkMag  
fBodyAcc-XYZ  
fBodyAccJerk-XYZ  
fBodyGyro-XYZ  
fBodyAccMag  
fBodyAccJerkMag  
fBodyGyroMag  
fBodyGyroJerkMag

Standard deviation and mean variables were created from each variation of the above values. Additional variables were created by applying the following mean values on the angle() variable:

gravityMean  
tBodyAccMean  
tBodyAccJerkMean  
tBodyGyroMean  
tBodyGyroJerkMean

## Data

The data aggregated consists of the following:

1. All features variables the test and training data set values represent. (“features” code variable)
2. All test and training values representing the input signal data received. (“testX” and “trainX” in the code)
3. IDs for the activity types that map to the associated test and training values. (“testY” and “trainY” in the code)
4. IDs for the subjects that maps to the associated test and training values. (“testSubjects” and “trainSubjects” in the code)
5. Label mappings for the activity type IDs. (“alabels” within the code)

## Transformations

The data was initially aggregated by the training, test, feature, subject, and activity label data sets into one data frame.

1. The complete set of aggregated data was filtered down to only the ACTIVITY, SUBJECT, and variables relating to mean and standard deviation values.
2. Once a data set containing only mean and standard deviation variables was formed with appropriate variable names it was then melted into a long data set. The ID variables used for this transformation were 'Activity' and 'Subject'. All of the signal data was melted into the 'variable' and 'value' columns. The following line of code represents this transformation:

```
meltDataSet <- melt(completeSet, id=c("Activity","Subject"), measure.vars=names(ds)[3:length(ds)])
```

3. The signal variables have been transformed to display in relation to their 'Activity' and 'Subject'. The data set is then cast to a new set that is organized by showing the average of each variable for each Activity of each Subject

```
castDataSet <- dcast(meltds2, Activity + Subject ~ variable, mean)
```