Readme

This document describes what my script run_analysis.R does.

1. Merges the training and the test sets to create one data set.

- Load training and test sets: X_train.txt, X_test.txt.
- Do quick inspection of the two sets to see whether the variable orders are likely the same and that they do contain 561 features
- Check to see whether the the train file does have roughly 7/3 as many obs as the test
- Merge the 2 data sets by appending the training set to the test set and check merged set

2. Extracts only the measurements on the mean and standard deviation for each measurement.

- Read in the features from features.txt, inspect data, and then use the feature names as the column names using a for-loop. Features were in factor, change to character
- Extract all columns with mean() or std() to form new dataframe using a for-loop.

3. Uses descriptive activity names to name the activities in the data set

- Read in integer activity labels from y_train.txt and y_test.txt and then combine them.
- Read in activity definition from activity_labels.txt.
- Add the Activity column to the overall data set.
- Fix the column name from mergedActivity$Activity

4. Appropriately labels the data set with descriptive variable names.

- 3 transforms will be used to make the names more intuitive: mean()=>Average, std()=>StandarDev(), gyro=>AngularVel

5. From the data set in step 4, creates a second, independent tidy data set with the average of each variable for each activity and each subject.

- First, subject identity labels from subject_train.txt and subject_test.txt need to be added to the big data set
- New data set creation, relying on carMelt example in lecture on reshaping data
- Final data set is output as FinalDataSet.txt