README

This repository is a requirement for Course Project of the Coursera course "Getting and Cleaning Data".
This README explains how all of the scripts in the repository work and how they are connected.

Repository Files

There is one R script in this repository **run_analysis.R**. This script creates two tidy data sets:

- **UCI_HAR_DataExtract.txt**

- **UCI_HAR_DataSummary.txt**

The script run_analysis.R requires that the dataset folder **UCI_HAR_Dataset** be in the working directory
There are two R markkdown files which produce the .md and .pdf formats:

- **CodeBook.Rmd**

- **README.Rmd**

run_analysis.R Details

1. Merges the training and the test sets to create one data set. The following files are merged and added to the data frame named x.

    - train/X_train.txt
    - test/X_test.txt
    - train/y_train.txt
    - test/y_test.txt
    - train/subject_train.txt
    - test/subject_test.txt

2. Appropriately labels the data set with descriptive variable names. The data in the y and test files is named 'Activity' and the subject data is named Subject. The x column labels are read in from the file **features.txt** and are cleaned up to make them easier to work with.

3. Uses descriptive activity names to name the activities in the data set. These are hard-coded from the information contained in the file activity_labels.txt.

    1. WALKING
    2. WALKING_UPSTAIRS
    3. WALKING_DOWNSTAIRS
    4. SITTING
    5. STANDING
    6. LAYING

4. The script extracts only the measurements on the mean and standard deviation for each measurement to create a data frame named **dataExtract**. The output file is **UCI_HAR_DataExtract.txt**.

The values:

- mean(): Mean value
- std(): Standard deviation

The variables:

- 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
- Activity
- Subject

6. Creates a second, independent tidy data set with the average of each variable for each activity and each subject. This data frame is named dataSummary and output file is **UCI_HAR_DataSummary.txt**.