Code book

=================

This document provides information about the generated sets by the script `run\_analysis.R`, provided in this repository.

The dataset is the result of cleaning and selecting data from data from an experiment in which measurements were collected from the accelerometer and gyroscope from the Samsung Galaxy S smartphone used by 30 subjects carrying out a variety of 6 activities.

More information about the measurements in the raw dataset is best obtained from the website where the dataset was taken from:

http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

For your convenience we have included information about the original data in the section [Original data set](#original-data-set).

1. The first variable `subject` denotes the subject number that performed an activity. In total there are 30 subjects.

2. The second variable `activity` denotes the activity performed by the subject. There are six activities, listed here:

\* WALKING

\* WALKING\_UPSTAIRS

\* WALKING\_DOWNSTAIRS

\* SITTING

\* STANDING

\* LAYING

All these labels are collected in **activityLabels** var.

3) **allData** variable :- Train (X\_trainy\_train,subject\_train) and Test (X\_test,y\_test,subject\_test) were merged into this variable using rbind and cbind functions.

4) A total of 79 features were selected from the original data in **mean\_std\_features** variable: only the estimated mean and standard deviations, using a `grep` on `"mean|std"`. These features form the 79 other variables in the dataset, making a total of 81 columns.

The feature names from the original data have been rewritten, using the following rules:

* The prefix `t` was rewritten into `time`, to make it clear the feature corresponds to the time domain.
* The prefix `f` was rewritten into `freq`, to make it clear the feature corresponds to the frequency domain
* The prefix 'Acc' was rewritten into 'Accelerometer' to make it descriptive.
* The prefix 'Gyro' was rewritten into 'Gyroscope' to make it descriptive.
* The prefix 'Mag' was rewritten into 'Magnitude' to make it descriptive.
* dashes and parentheses have been removed.
* BodyBody has been replaced by Body

For example:

\* `tBodyAcc-mean()-X` becomes `timeBodyAccMeanX`

\* `tBodyAcc-std()-Y` becomes `timeBodyAccStdY`

\* `fBodyAcc-mean()-Z` becomes `freqBodyAccMeanZ`

\* `fBodyBodyGyroMag-mean()` becomes `freqBodyGyroMagMean`

\* `fBodyBodyGyroJerkMag-meanFreq()` becomes `freqBodyGyroJerkMagMeanFreq`

A full description of the features is best obtained from the file `features\_info.txt`, included with the original dataset (I found it unnecessary to duplicate all the information here).

5) The data collected in allData variable were grouped by subject and activity and then aggregated using the mean function. This yields 180 rows (30 subjects times 6 activities) and 81 columns (1 for subject, 1 for activity plus 79 aggregated features). This is collected in **tidyData** variable.

6) The tidy data is collected in a txt file named **tidydataset.txt.**

## Original data set

Taken from http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones:

"The experiments have been carried out with a group of 30 volunteers within an age bracket of 19-48 years. Each person performed six activities (WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING) wearing a smartphone (Samsung Galaxy S II) on the waist. Using its embedded accelerometer and gyroscope, we captured 3-axial linear acceleration and 3-axial angular velocity at a constant rate of 50Hz. The experiments have been video-recorded to label the data manually. The obtained dataset has been randomly partitioned into two sets, where 70% of the volunteers was selected for generating the training data and 30% the test data.

The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window). The sensor acceleration signal, which has gravitational and body motion components, was separated using a Butterworth low-pass filter into body acceleration and gravity. The gravitational force is assumed to have only low frequency components, therefore a filter with 0.3 Hz cutoff frequency was used. From each window, a vector of features was obtained by calculating variables from the time and frequency domain."

For more information read visit http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones, the website where the original data came from.