Project Requirement

Create one R script called run_analysis.R that does the following:

1. Merges the training and the test sets to create one data set.
2. Extracts only the measurements on the mean and standard deviation for each measurement.
3. Uses descriptive activity names to name the activities in the data set
4. Appropriately labels the data set with descriptive variable names.
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.

General Approach

Keeping the fundamental goal of making tidy data in mind, my approach was to ensure that:

1. Each variable forms a column and each observation forms a row.
2. Variable names are:

i) Lower case.
ii) Descriptive.
iii) Do not contain any special characters such as "_","()", etc.
iv) No duplicated variables or data.

I made more in-depth comments throughout my code below, but here is an outline of the steps I took to clean and compile data set.

1. Cleaned the data related to the feature variables. I renamed the variables given in the features_info.txt file according to tidy data principles. The updated variable names are in the codebook. I checked see if there were any duplicated variables as well.
2. Cleaned the activities data file according to tidy data principles.
3. Built the test and training data sets removing duplicated data.
4. Merged the test and training data sets into a complete data set.
5. Melted the data set to a narrow format.

Data Source

https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip

Background information.

==========

Human Activity Recognition Using Smartphones Dataset
Version 1.0

==========

Jorge L. Reyes-Ortiz, Davide Anguita, Alessandro Ghio, Luca Oneto.
Smartlab - Non Linear Complex Systems Laboratory
DITEN - Università degli Studi di Genova.
Via Opera Pia 11A, I-16145, Genoa, Italy.
activityrecognition@smartlab.ws

www.smartlab.ws
==========

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. See 'features_info.txt' for more details.

For each record it is provided:
==========

- Triaxial acceleration from the accelerometer (total acceleration) and the estimated body acceleration.
- Triaxial Angular velocity from the gyroscope.
- A 561-feature vector with time and frequency domain variables.
- Its activity label.
- An identifier of the subject who carried out the experiment.

The dataset includes the following files:
==========

- 'README.txt'

- 'features_info.txt': Shows information about the variables used on the feature vector.

- 'features.txt': List of all features.

- 'activity_labels.txt': Links the class labels with their activity name.

- 'train/X_train.txt': Training set.

- 'train/y_train.txt': Training labels.

- 'test/X_test.txt': Test set.

- 'test/y_test.txt': Test labels.

The following files are available for the train and test data. Their descriptions are equivalent.

- 'train/subject_train.txt': Each row identifies the subject who performed the activity for each window sample. Its range is from 1 to 30.

- 'train/Inertial Signals/total_acc_x_train.txt': The acceleration signal from the smartphone accelerometer X axis in standard gravity units 'g'. Every row shows a 128 element vector. The same description applies for the 'total_acc_x_train.txt' and 'total_acc_z_train.txt' files for the Y and Z axis.

- 'train/Inertial Signals/body_acc_x_train.txt': The body acceleration signal obtained by subtracting the gravity from the total acceleration.

- 'train/Inertial Signals/body_gyro_x_train.txt': The angular velocity vector measured by the gyroscope for each window sample. The units are radians/second.

Notes:
======
- Features are normalized and bounded within [-1,1].
- Each feature vector is a row on the text file.

For more information about this dataset contact: activityrecognition@smartlab.ws

License:
========
Use of this dataset in publications must be acknowledged by referencing the following publication [1]

[1] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. Human Activity Recognition on Smartphones using a Multiclass Hardware-Friendly Support Vector Machine. International Workshop of Ambient Assisted Living (IWAAL 2012). Vitoria-Gasteiz, Spain. Dec 2012

Jorge L. Reyes-Ortiz, Alessandro Ghio, Luca Oneto, Davide Anguita. November 2012.

Code Reference

Feature Selection
==========

The features selected for this database 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.

timebodyacceleration- xyz
timegravityacceleration- xyz
timebodyacceleration - xyz
timebodygyro- xyz
timebodygyrojerk-xyz
timebodyacceleration mag
timegravityaccelerationmag
timebodyaccelerationjerkmag
timebodyacceleration gyromag
timebodygyrojerkmag
frequecncybodyacceleration- xyz
frequecncybodyacceleration jerk- xyz
frequecncybodygyro-xyz
frequecncybodyaccelerationmag
frequecncybodyaccelerationmag
frequecncybodygyromag
frequecncybodygyroJerkmag

The set of variables that were estimated from these signals are:

mean: Mean value
std: Standard deviation
mad: Median absolute deviation
max: Largest value in array
min: Smallest value in array
sma: Signal magnitude area
energy: Energy measure. Sum of the squares divided by the number of values.
iqr: Interquartile range
entropy: Signal entropy
autoregressioncoefficient: Autorregresion coefficients with Burg order equal to 4
correlation: correlation coefficient between two signals
maxinds: index of the frequency component with largest magnitude
meanfreq: Weighted average of the frequency components to obtain a mean frequency
skewness: skewness of the frequency domain signal
kurtosis: kurtosis of the frequency domain signal
bandsenergy: Energy of a frequency interval within the 64 bins of the FFT of each window.
angle: Angle between to vectors.

Additional vectors obtained by averaging the signals in a signal window sample. These are used on the angle() variable:

gravitymean
timebodyacceleration
timebodyacceleration jerkmean
timebodygyromean
timebodygyrojerkmean