“Getting and Cleaning Data” Final Project Cookbook

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

Processing
This is what the run_analysis.R does at a high level
1. Load the data from data/UCI_HAR_Dataset
2. Combines test and train data for features, activities, and subjects
3. Merges the combined data from the previous step into a single frame
4. Renames the data frame column names to the appropriate variables names from *features.txt*
5. Selects only subject, activity, mean and standard deviation variables
6. It changes the activity variable from activity id to activity label based on *activity_labels.txt*
7. It renames the column names based to conform to the R naming scheme by using the *fix_col_names()* function 8. Finally, it generates a new data frame that contains an aggregation by subject and activity

Output data
There's a total of 68 variables in the output data set, divided as follows:

subject – Subject id (1-30)
activity – Activity name (Walking, Walking Upstair, Walking Downstairs, Sitting, Standing or Laying)

66 measure variables based on gyroscope and accelerometer measurements of the form `(time|frequency)(accel|gyro)(mean|std)(X|Y|Z)`. Each variable corresponds to the average (mean) of all observation for a given subject and activity for the specific measurement.

`time`	Time domain signal
`frequency`	Frequency domain signal
`accel`	Accelerometer
`gyro`	Gyroscope
`mean`	Mean
`std`	Standard Deviation
`X, Y, X`	Axis

For example the variable name **freq_body_gyro_std_X** stands for the mean of all standard deviations of the frequency domain of the gyroscope on the X axis.

Variable Units

Type	**Unit**
Acceleration (body and gravity)	'g's (9.81m/s^2)
Gyroscope	rad/seg