Codebook

Stephan Nguyen

17/4/2017

Content

- The Assignment: Getting and Cleaning Data Course Project
- Codebook

The Assignment:Getting and Cleaning Data Course Project

The purpose of this project is to demonstrate your ability to collect, work with, and clean a data set. The goal is to prepare tidy data that can be used for later analysis. You will be graded by your peers on a series of yes/no questions related to the project. You will be required to submit:
1. a tidy data set as described below;
2. a link to a Github repository with your script for performing the analysis;
3. and a **code book that describes the variables, the data, and any transformations or work that you performed to clean up the data** called CodeBook.md. You should also include a README.md in the repo with your scripts. This repo explains how all of the scripts work and how they are connected.

One of the most exciting areas in all of data science right now is wearable computing - see for example this article. Companies like Fitbit, Nike, and Jawbone Up are racing to develop the most advanced algorithms to attract new users. The data linked to from the course website represent data collected from the accelerometers from the Samsung Galaxy S smartphone.
A full description is available at the site where the data was obtained:
http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones

Here are the data for the project:
https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip

You should 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.

Good luck!

Data

According to the assignment the codebook should have the following information:
"A code book that describes the variables, the data, and any transformations or work that you performed to clean up the data called CodeBook.md".

The variables

Background information

The variables 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.

Signals for the 3-axial directions X, Y and Z	complete description
tBodyAcc-XYZ	time Body acceleration - XYZ
tGravityAcc-XYZ	time Gravity acceleration - XYZ
tBodyAccJerk-XYZ	time Body acceleration Jerk - XYZ
tBodyGyro-XYZ	time Body gyro - XYZ
tBodyGyroJerk-XYZ	time Body gyro Jerk - XYZ
tBodyAccMag	time Body acceleration magnitude
tGravityAccMag	time Gravity acceleration magnitude
tBodyAccJerkMag	time Body acceleration Jerk magnitude
tBodyGyroMag	time Body gyro magnitude
tBodyGyroJerkMag	time Body gyro Jerk magnitude
fBodyAcc-XYZ	frequency Body acceleration - XYZ
fBodyAccJerk-XYZ	frequency Body acceleration Jerk - XYZ
fBodyGyro-XYZ	frequency Body gyro - XYZ
fBodyAccMag	frequency Body acceleration magnitude
fBodyAccJerkMag	frequency Body acceleration Jerk magnitude
fBodyGyroMag	frequency Body gyro magnitude
fBodyGyroJerkMag	frequency Body gyro Jerk magnitude

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

Variable	description
mean()	Mean value
std()	Standard deviation
mad()	Median absolute deviation
max()	Largest value in array
min()	Smallest value in array

Variable	description
sma()	Signal magnitude area
energy()	Energy measure. Sum of the squares divided by the number of values.
iqr()	Interquartile range
entropy()	Signal entropy
arCoeff()	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
tBodyAccMean
tBodyAccJerkMean
tBodyGyroMean
tBodyGyroJerkMean

Data

```
## 'data.frame':    180 obs. of  81 variables:
##  $ subject_ID                    : int  1 1 1 1 1 1 2 2 2 2 ...
##  $ Activity_label                : Factor w/ 6 levels "LAYING","SITTING",..: 1 2 3 4 5 6 1 2 3 4 ..
##  $ tBodyAcc.mean...X             : num  0.222 0.261 0.279 0.277 0.289 ...
##  $ tBodyAcc.mean...Y             : num  -0.04051 -0.00131 -0.01614 -0.01738 -0.00992 ...
##  $ tBodyAcc.mean...Z             : num  -0.113 -0.105 -0.111 -0.111 -0.108 ...
##  $ tBodyAcc.std...X              : num  -0.928 -0.977 -0.996 -0.284 0.03 ...
##  $ tBodyAcc.std...Y              : num  -0.8368 -0.9226 -0.9732 0.1145 -0.0319 ...
##  $ tBodyAcc.std...Z              : num  -0.826 -0.94 -0.98 -0.26 -0.23 ...
##  $ tGravityAcc.mean...X          : num  -0.249 0.832 0.943 0.935 0.932 ...
##  $ tGravityAcc.mean...Y          : num  0.706 0.204 -0.273 -0.282 -0.267 ...
##  $ tGravityAcc.mean...Z          : num  0.4458 0.332 0.0135 -0.0681 -0.0621 ...
##  $ tGravityAcc.std...X           : num  -0.897 -0.968 -0.994 -0.977 -0.951 ...
##  $ tGravityAcc.std...Y           : num  -0.908 -0.936 -0.981 -0.971 -0.937 ...
##  $ tGravityAcc.std...Z           : num  -0.852 -0.949 -0.976 -0.948 -0.896 ...
##  $ tBodyAccJerk.mean...X         : num  0.0811 0.0775 0.0754 0.074 0.0542 ...
##  $ tBodyAccJerk.mean...Y         : num  0.003838 -0.000619 0.007976 0.028272 0.02965 ...
##  $ tBodyAccJerk.mean...Z         : num  0.01083 -0.00337 -0.00369 -0.00417 -0.01097 ...
##  $ tBodyAccJerk.std...X          : num  -0.9585 -0.9864 -0.9946 -0.1136 -0.0123 ...
##  $ tBodyAccJerk.std...Y          : num  -0.924 -0.981 -0.986 0.067 -0.102 ...
##  $ tBodyAccJerk.std...Z          : num  -0.955 -0.988 -0.992 -0.503 -0.346 ...
##  $ tBodyGyro.mean...X            : num  -0.0166 -0.0454 -0.024 -0.0418 -0.0351 ...
##  $ tBodyGyro.mean...Y            : num  -0.0645 -0.0919 -0.0594 -0.0695 -0.0909 ...
##  $ tBodyGyro.mean...Z            : num  0.1487 0.0629 0.0748 0.0849 0.0901 ...
##  $ tBodyGyro.std...X             : num  -0.874 -0.977 -0.987 -0.474 -0.458 ...
##  $ tBodyGyro.std...Y             : num  -0.9511 -0.9665 -0.9877 -0.0546 -0.1263 ...
##  $ tBodyGyro.std...Z             : num  -0.908 -0.941 -0.981 -0.344 -0.125 ...
##  $ tBodyGyroJerk.mean...X        : num  -0.1073 -0.0937 -0.0996 -0.09 -0.074 ...
##  $ tBodyGyroJerk.mean...Y        : num  -0.0415 -0.0402 -0.0441 -0.0398 -0.044 ...
##  $ tBodyGyroJerk.mean...Z        : num  -0.0741 -0.0467 -0.049 -0.0461 -0.027 ...
##  $ tBodyGyroJerk.std...X         : num  -0.919 -0.992 -0.993 -0.207 -0.487 ...
##  $ tBodyGyroJerk.std...Y         : num  -0.968 -0.99 -0.995 -0.304 -0.239 ...
##  $ tBodyGyroJerk.std...Z         : num  -0.958 -0.988 -0.992 -0.404 -0.269 ...
##  $ tBodyAccMag.mean..            : num  -0.8419 -0.9485 -0.9843 -0.137 0.0272 ...
##  $ tBodyAccMag.std..             : num  -0.7951 -0.9271 -0.9819 -0.2197 0.0199 ...
##  $ tGravityAccMag.mean..         : num  -0.8419 -0.9485 -0.9843 -0.137 0.0272 ...
##  $ tGravityAccMag.std..          : num  -0.7951 -0.9271 -0.9819 -0.2197 0.0199 ...
##  $ tBodyAccJerkMag.mean..        : num  -0.9544 -0.9874 -0.9924 -0.1414 -0.0894 ...
##  $ tBodyAccJerkMag.std..         : num  -0.9282 -0.9841 -0.9931 -0.0745 -0.0258 ...
##  $ tBodyGyroMag.mean..           : num  -0.8748 -0.9309 -0.9765 -0.161 -0.0757 ...
##  $ tBodyGyroMag.std..            : num  -0.819 -0.935 -0.979 -0.187 -0.226 ...
##  $ tBodyGyroJerkMag.mean..       : num  -0.963 -0.992 -0.995 -0.299 -0.295 ...
##  $ tBodyGyroJerkMag.std..        : num  -0.936 -0.988 -0.995 -0.325 -0.307 ...
##  $ fBodyAcc.mean...X             : num  -0.9391 -0.9796 -0.9952 -0.2028 0.0382 ...
##  $ fBodyAcc.mean...Y             : num  -0.86707 -0.94408 -0.97707 0.08971 0.00155 ...
##  $ fBodyAcc.mean...Z             : num  -0.883 -0.959 -0.985 -0.332 -0.226 ...
##  $ fBodyAcc.std...X              : num  -0.9244 -0.9764 -0.996 -0.3191 0.0243 ...
##  $ fBodyAcc.std...Y              : num  -0.834 -0.917 -0.972 0.056 -0.113 ...
##  $ fBodyAcc.std...Z              : num  -0.813 -0.934 -0.978 -0.28 -0.298 ...
##  $ fBodyAcc.meanFreq...X         : num  -0.1588 -0.0495 0.0865 -0.2075 -0.3074 ...
##  $ fBodyAcc.meanFreq...Y         : num  0.0975 0.0759 0.1175 0.1131 0.0632 ...
##  $ fBodyAcc.meanFreq...Z         : num  0.0894 0.2388 0.2449 0.0497 0.2943 ...
##  $ fBodyAccJerk.mean...X         : num  -0.9571 -0.9866 -0.9946 -0.1705 -0.0277 ...
##  $ fBodyAccJerk.mean...Y         : num  -0.9225 -0.9816 -0.9854 -0.0352 -0.1287 ...
```

```
##  $ fBodyAccJerk.mean...Z         : num  -0.948 -0.986 -0.991 -0.469 -0.288 ...
##  $ fBodyAccJerk.std...X          : num  -0.9642 -0.9875 -0.9951 -0.1336 -0.0863 ...
##  $ fBodyAccJerk.std...Y          : num  -0.932 -0.983 -0.987 0.107 -0.135 ...
##  $ fBodyAccJerk.std...Z          : num  -0.961 -0.988 -0.992 -0.535 -0.402 ...
##  $ fBodyAccJerk.meanFreq...X     : num  0.132 0.257 0.314 -0.209 -0.253 ...
##  $ fBodyAccJerk.meanFreq...Y     : num  0.0245 0.0475 0.0392 -0.3862 -0.3376 ...
##  $ fBodyAccJerk.meanFreq...Z     : num  0.02439 0.09239 0.13858 -0.18553 0.00937 ...
##  $ fBodyGyro.mean...X            : num  -0.85 -0.976 -0.986 -0.339 -0.352 ...
##  $ fBodyGyro.mean...Y            : num  -0.9522 -0.9758 -0.989 -0.1031 -0.0557 ...
##  $ fBodyGyro.mean...Z            : num  -0.9093 -0.9513 -0.9808 -0.2559 -0.0319 ...
##  $ fBodyGyro.std...X             : num  -0.882 -0.978 -0.987 -0.517 -0.495 ...
##  $ fBodyGyro.std...Y             : num  -0.9512 -0.9623 -0.9871 -0.0335 -0.1814 ...
##  $ fBodyGyro.std...Z             : num  -0.917 -0.944 -0.982 -0.437 -0.238 ...
##  $ fBodyGyro.meanFreq...X        : num  -0.00355 0.18915 -0.12029 0.01478 -0.10045 ...
##  $ fBodyGyro.meanFreq...Y        : num  -0.0915 0.0631 -0.0447 -0.0658 0.0826 ...
##  $ fBodyGyro.meanFreq...Z        : num  0.010458 -0.029784 0.100608 0.000773 -0.075676 ...
##  $ fBodyAccMag.mean..            : num  -0.8618 -0.9478 -0.9854 -0.1286 0.0966 ...
##  $ fBodyAccMag.std..             : num  -0.798 -0.928 -0.982 -0.398 -0.187 ...
##  $ fBodyAccMag.meanFreq..        : num  0.0864 0.2367 0.2846 0.1906 0.1192 ...
##  $ fBodyBodyAccJerkMag.mean..    : num  -0.9333 -0.9853 -0.9925 -0.0571 0.0262 ...
##  $ fBodyBodyAccJerkMag.std..     : num  -0.922 -0.982 -0.993 -0.103 -0.104 ...
##  $ fBodyBodyAccJerkMag.meanFreq.. : num  0.2664 0.3519 0.4222 0.0938 0.0765 ...
##  $ fBodyBodyGyroMag.mean..       : num  -0.862 -0.958 -0.985 -0.199 -0.186 ...
##  $ fBodyBodyGyroMag.std..        : num  -0.824 -0.932 -0.978 -0.321 -0.398 ...
##  $ fBodyBodyGyroMag.meanFreq..   : num  -0.139775 -0.000262 -0.028606 0.268844 0.349614 ...
##  $ fBodyBodyGyroJerkMag.mean..   : num  -0.942 -0.99 -0.995 -0.319 -0.282 ...
##  $ fBodyBodyGyroJerkMag.std..    : num  -0.933 -0.987 -0.995 -0.382 -0.392 ...
##  $ fBodyBodyGyroJerkMag.meanFreq..: num  0.176 0.185 0.334 0.191 0.19 ...
```