

Codebook

wintics

10/23/2020

Getting and Cleaning Data Course Project

Description

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.

The task is to create 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.

Study design and data processing

Collection of the raw data

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.

Used raw data files for project:

- 'train/subject_train.txt': Each row identifies the subject who performed the activity for each window sample (training set). Subjects (volunteers) codes range is from 1 to 30. Activities are represented by codes from 1 to 6. Dimensions: 7352 X 1
- 'train/X_train.txt': Training set (measurements). Dimensions: 7352 X 561
- 'train/y_train.txt': Training labels. Dimensions: 7352 X 1
- 'test/subject_test.txt': Each row identifies the subject who performed the activity for each window sample (test set). Subjects (volunteers) codes range is from 1 to 30. Activities are represented by codes from 1 to 6. Dimensions: 2947 X 1
- 'test/X_test.txt': Test set (measurements). Dimensions: 2947 x 561
- 'test/y_test.txt': Test labels. Dimensions: 2947 x 1
- 'activity_labels.txt': Activity codes and labels
- 'features.txt': Measurements labels. Raw data set has 561 features.

Notes on the original (raw) data

Data for the assignment project (includes more descriptions):

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

Creating the tidy datafile

Guide to create the output file

1. download all data and extract from zip-file,
2. read measurements data and merge training and test data,
3. for extracting the measurements on the mean and standard deviation for each measurement select these columns which have "mean()" or "std()" in names,
4. substitute activity codes to activity labels from activity_labels file, rename the column accordingly,
5. calculate mean of each variable for each activity and each subject,
6. write result in to text file "tidy_data.txt"

Output data

Description of columns

- **activity** - activity label refers to activity of volunteer
- **subjectID** - volunteers' numeric code from 1 to 30
- **tBodyAcc-mean()-X**
....
- **fBodyBodyGyroJerkMag-std()** - calculated means of different measurements from raw data for this project all measurements in this dataset have still their original names from raw data

Activity labels:

- WALKING - subject was walking during the test
- WALKING_UPSTAIRS - subject was walking up a staircase during the test
- WALKING_DOWNSTAIRS - subject was walking down a staircase during the test
- SITTING - subject was sitting during the test
- STANDING - subject was standing during the test
- LAYING - subject was laying down during the test

More information

```
## DIMENSIONS
```

```
## [1] 180 68
```

```
## ALL COLUMNS
```

```
##      1
## 1  "activity"
## 2  "subjectID"
## 3  "tBodyAcc-mean()-X"
## 4  "tBodyAcc-mean()-Y"
## 5  "tBodyAcc-mean()-Z"
## 6  "tGravityAcc-mean()-X"
## 7  "tGravityAcc-mean()-Y"
## 8  "tGravityAcc-mean()-Z"
## 9  "tBodyAccJerk-mean()-X"
## 10 "tBodyAccJerk-mean()-Y"
## 11 "tBodyAccJerk-mean()-Z"
```

```
## 12 "tBodyGyro-mean()-X"
## 13 "tBodyGyro-mean()-Y"
## 14 "tBodyGyro-mean()-Z"
## 15 "tBodyGyroJerk-mean()-X"
## 16 "tBodyGyroJerk-mean()-Y"
## 17 "tBodyGyroJerk-mean()-Z"
## 18 "tBodyAccMag-mean()"
## 19 "tGravityAccMag-mean()"
## 20 "tBodyAccJerkMag-mean()"
## 21 "tBodyGyroMag-mean()"
## 22 "tBodyGyroJerkMag-mean()"
## 23 "fBodyAcc-mean()-X"
## 24 "fBodyAcc-mean()-Y"
## 25 "fBodyAcc-mean()-Z"
## 26 "fBodyAccJerk-mean()-X"
## 27 "fBodyAccJerk-mean()-Y"
## 28 "fBodyAccJerk-mean()-Z"
## 29 "fBodyGyro-mean()-X"
## 30 "fBodyGyro-mean()-Y"
## 31 "fBodyGyro-mean()-Z"
## 32 "fBodyAccMag-mean()"
## 33 "fBodyBodyAccJerkMag-mean()"
## 34 "fBodyBodyGyroMag-mean()"
## 35 "fBodyBodyGyroJerkMag-mean()"
## 36 "tBodyAcc-std()-X"
## 37 "tBodyAcc-std()-Y"
## 38 "tBodyAcc-std()-Z"
## 39 "tGravityAcc-std()-X"
## 40 "tGravityAcc-std()-Y"
## 41 "tGravityAcc-std()-Z"
## 42 "tBodyAccJerk-std()-X"
## 43 "tBodyAccJerk-std()-Y"
## 44 "tBodyAccJerk-std()-Z"
## 45 "tBodyGyro-std()-X"
## 46 "tBodyGyro-std()-Y"
## 47 "tBodyGyro-std()-Z"
## 48 "tBodyGyroJerk-std()-X"
## 49 "tBodyGyroJerk-std()-Y"
## 50 "tBodyGyroJerk-std()-Z"
## 51 "tBodyAccMag-std()"
## 52 "tGravityAccMag-std()"
## 53 "tBodyAccJerkMag-std()"
## 54 "tBodyGyroMag-std()"
## 55 "tBodyGyroJerkMag-std()"
## 56 "fBodyAcc-std()-X"
## 57 "fBodyAcc-std()-Y"
## 58 "fBodyAcc-std()-Z"
## 59 "fBodyAccJerk-std()-X"
## 60 "fBodyAccJerk-std()-Y"
## 61 "fBodyAccJerk-std()-Z"
## 62 "fBodyGyro-std()-X"
## 63 "fBodyGyro-std()-Y"
## 64 "fBodyGyro-std()-Z"
## 65 "fBodyAccMag-std()"
## 66 "fBodyBodyAccJerkMag-std()"
## 67 "fBodyBodyGyroMag-std()"
## 68 "fBodyBodyGyroJerkMag-std()"
```

SUMMARY

```
##      activity          subjectID  tBodyAcc.mean...X tBodyAcc.mean...Y
## Length:180          Min.    : 1.0  Min.    :0.2216  Min.    :-0.040514
## Class :character    1st Qu.: 8.0  1st Qu.:0.2712  1st Qu.: -0.020022
## Mode  :character    Median :15.5  Median :0.2770  Median : -0.017262
##                                     Mean  :15.5  Mean   :0.2743  Mean   : -0.017876
##                                     3rd Qu.:23.0  3rd Qu.:0.2800  3rd Qu.: -0.014936
##                                     Max.   :30.0  Max.   :0.3015  Max.   : -0.001308
## tBodyAcc.mean...Z  tGravityAcc.mean...X tGravityAcc.mean...Y
## Min.    :-0.15251  Min.    :-0.6800  Min.    :-0.47989
## 1st Qu.: -0.11207  1st Qu.: 0.8376  1st Qu.: -0.23319
## Median : -0.10819  Median : 0.9208  Median : -0.12782
## Mean    : -0.10916  Mean    : 0.6975  Mean    : -0.01621
## 3rd Qu.: -0.10443  3rd Qu.: 0.9425  3rd Qu.: 0.08773
## Max.    : -0.07538  Max.    : 0.9745  Max.    : 0.95659
## tGravityAcc.mean...Z tBodyAccJerk.mean...X tBodyAccJerk.mean...Y
## Min.    :-0.49509  Min.    :0.04269  Min.    :-0.0386872
## 1st Qu.: -0.11726  1st Qu.:0.07396  1st Qu.: 0.0004664
## Median : 0.02384  Median :0.07640  Median : 0.0094698
## Mean    : 0.07413  Mean    :0.07947  Mean    : 0.0075652
## 3rd Qu.: 0.14946  3rd Qu.:0.08330  3rd Qu.: 0.0134008
## Max.    : 0.95787  Max.    :0.13019  Max.    : 0.0568186
## tBodyAccJerk.mean...Z tBodyGyro.mean...X tBodyGyro.mean...Y tBodyGyro.mean...Z
## Min.    :-0.067458  Min.    :-0.20578  Min.    :-0.20421  Min.    :-0.07245
## 1st Qu.: -0.010601  1st Qu.: -0.04712  1st Qu.: -0.08955  1st Qu.: 0.07475
## Median : -0.003861  Median : -0.02871  Median : -0.07318  Median : 0.08512
## Mean    : -0.004953  Mean    : -0.03244  Mean    : -0.07426  Mean    : 0.08744
## 3rd Qu.: 0.001958  3rd Qu.: -0.01676  3rd Qu.: -0.06113  3rd Qu.: 0.10177
## Max.    : 0.038053  Max.    : 0.19270  Max.    : 0.02747  Max.    : 0.17910
## tBodyGyroJerk.mean...X tBodyGyroJerk.mean...Y tBodyGyroJerk.mean...Z
## Min.    :-0.15721  Min.    :-0.07681  Min.    :-0.092500
## 1st Qu.: -0.10322  1st Qu.: -0.04552  1st Qu.: -0.061725
## Median : -0.09868  Median : -0.04112  Median : -0.053430
## Mean    : -0.09606  Mean    : -0.04269  Mean    : -0.054802
## 3rd Qu.: -0.09110  3rd Qu.: -0.03842  3rd Qu.: -0.048985
## Max.    : -0.02209  Max.    : -0.01320  Max.    : -0.006941
## tBodyAccMag.mean.. tGravityAccMag.mean.. tBodyAccJerkMag.mean..
## Min.    :-0.9865  Min.    :-0.9865  Min.    :-0.9928
## 1st Qu.: -0.9573  1st Qu.: -0.9573  1st Qu.: -0.9807
## Median : -0.4829  Median : -0.4829  Median : -0.8168
## Mean    : -0.4973  Mean    : -0.4973  Mean    : -0.6079
## 3rd Qu.: -0.0919  3rd Qu.: -0.0919  3rd Qu.: -0.2456
## Max.    : 0.6446  Max.    : 0.6446  Max.    : 0.4345
## tBodyGyroMag.mean.. tBodyGyroJerkMag.mean.. fBodyAcc.mean...X
## Min.    :-0.9807  Min.    :-0.99732  Min.    :-0.9952
## 1st Qu.: -0.9461  1st Qu.: -0.98515  1st Qu.: -0.9787
## Median : -0.6551  Median : -0.86479  Median : -0.7691
## Mean    : -0.5652  Mean    : -0.73637  Mean    : -0.5758
## 3rd Qu.: -0.2159  3rd Qu.: -0.51186  3rd Qu.: -0.2174
## Max.    : 0.4180  Max.    : 0.08758  Max.    : 0.5370
## fBodyAcc.mean...Y fBodyAcc.mean...Z fBodyAccJerk.mean...X
## Min.    :-0.98903  Min.    :-0.9895  Min.    :-0.9946
## 1st Qu.: -0.95361  1st Qu.: -0.9619  1st Qu.: -0.9828
## Median : -0.59498  Median : -0.7236  Median : -0.8126
## Mean    : -0.48873  Mean    : -0.6297  Mean    : -0.6139
## 3rd Qu.: -0.06341  3rd Qu.: -0.3183  3rd Qu.: -0.2820
```

```

## Max. : 0.52419 Max. : 0.2807 Max. : 0.4743
## fBodyAccJerk.mean...Y fBodyAccJerk.mean...Z fBodyGyro.mean...X
## Min. :-0.9894 Min. :-0.9920 Min. :-0.9931
## 1st Qu.:-0.9725 1st Qu.:-0.9796 1st Qu.:-0.9697
## Median :-0.7817 Median :-0.8707 Median :-0.7300
## Mean :-0.5882 Mean :-0.7144 Mean :-0.6367
## 3rd Qu.:-0.1963 3rd Qu.:-0.4697 3rd Qu.:-0.3387
## Max. : 0.2767 Max. : 0.1578 Max. : 0.4750
## fBodyGyro.mean...Y fBodyGyro.mean...Z fBodyAccMag.mean..
## Min. :-0.9940 Min. :-0.9860 Min. :-0.9868
## 1st Qu.:-0.9700 1st Qu.:-0.9624 1st Qu.:-0.9560
## Median :-0.8141 Median :-0.7909 Median :-0.6703
## Mean :-0.6767 Mean :-0.6044 Mean :-0.5365
## 3rd Qu.:-0.4458 3rd Qu.:-0.2635 3rd Qu.:-0.1622
## Max. : 0.3288 Max. : 0.4924 Max. : 0.5866
## fBodyBodyAccJerkMag.mean.. fBodyBodyGyroMag.mean.. fBodyBodyGyroJerkMag.mean..
## Min. :-0.9940 Min. :-0.9865 Min. :-0.9976
## 1st Qu.:-0.9770 1st Qu.:-0.9616 1st Qu.:-0.9813
## Median :-0.7940 Median :-0.7657 Median :-0.8779
## Mean :-0.5756 Mean :-0.6671 Mean :-0.7564
## 3rd Qu.:-0.1872 3rd Qu.:-0.4087 3rd Qu.:-0.5831
## Max. : 0.5384 Max. : 0.2040 Max. : 0.1466
## tBodyAcc.std...X tBodyAcc.std...Y tBodyAcc.std...Z tGravityAcc.std...X
## Min. :-0.9961 Min. :-0.99024 Min. :-0.9877 Min. :-0.9968
## 1st Qu.:-0.9799 1st Qu.:-0.94205 1st Qu.:-0.9498 1st Qu.:-0.9825
## Median :-0.7526 Median :-0.50897 Median :-0.6518 Median :-0.9695
## Mean :-0.5577 Mean :-0.46046 Mean :-0.5756 Mean :-0.9638
## 3rd Qu.:-0.1984 3rd Qu.:-0.03077 3rd Qu.:-0.2306 3rd Qu.:-0.9509
## Max. : 0.6269 Max. : 0.61694 Max. : 0.6090 Max. :-0.8296
## tGravityAcc.std...Y tGravityAcc.std...Z tBodyAccJerk.std...X
## Min. :-0.9942 Min. :-0.9910 Min. :-0.9946
## 1st Qu.:-0.9711 1st Qu.:-0.9605 1st Qu.:-0.9832
## Median :-0.9590 Median :-0.9450 Median :-0.8104
## Mean :-0.9524 Mean :-0.9364 Mean :-0.5949
## 3rd Qu.:-0.9370 3rd Qu.:-0.9180 3rd Qu.:-0.2233
## Max. :-0.6436 Max. :-0.6102 Max. : 0.5443
## tBodyAccJerk.std...Y tBodyAccJerk.std...Z tBodyGyro.std...X tBodyGyro.std...Y
## Min. :-0.9895 Min. :-0.99329 Min. :-0.9943 Min. :-0.9942
## 1st Qu.:-0.9724 1st Qu.:-0.98266 1st Qu.:-0.9735 1st Qu.:-0.9629
## Median :-0.7756 Median :-0.88366 Median :-0.7890 Median :-0.8017
## Mean :-0.5654 Mean :-0.73596 Mean :-0.6916 Mean :-0.6533
## 3rd Qu.:-0.1483 3rd Qu.:-0.51212 3rd Qu.:-0.4414 3rd Qu.:-0.4196
## Max. : 0.3553 Max. : 0.03102 Max. : 0.2677 Max. : 0.4765
## tBodyGyro.std...Z tBodyGyroJerk.std...X tBodyGyroJerk.std...Y
## Min. :-0.9855 Min. :-0.9965 Min. :-0.9971
## 1st Qu.:-0.9609 1st Qu.:-0.9800 1st Qu.:-0.9832
## Median :-0.8010 Median :-0.8396 Median :-0.8942
## Mean :-0.6164 Mean :-0.7036 Mean :-0.7636
## 3rd Qu.:-0.3106 3rd Qu.:-0.4629 3rd Qu.:-0.5861
## Max. : 0.5649 Max. : 0.1791 Max. : 0.2959
## tBodyGyroJerk.std...Z tBodyAccMag.std.. tGravityAccMag.std..
## Min. :-0.9954 Min. :-0.9865 Min. :-0.9865
## 1st Qu.:-0.9848 1st Qu.:-0.9430 1st Qu.:-0.9430
## Median :-0.8610 Median :-0.6074 Median :-0.6074
## Mean :-0.7096 Mean :-0.5439 Mean :-0.5439
## 3rd Qu.:-0.4741 3rd Qu.:-0.2090 3rd Qu.:-0.2090
## Max. : 0.1932 Max. : 0.4284 Max. : 0.4284
## tBodyAccJerkMag.std.. tBodyGyroMag.std.. tBodyGyroJerkMag.std..

```

```

## Min.      :-0.9946      Min.      :-0.9814      Min.      :-0.9977
## 1st Qu.   :-0.9765      1st Qu.   :-0.9476      1st Qu.   :-0.9805
## Median    :-0.8014      Median    :-0.7420      Median    :-0.8809
## Mean      :-0.5842      Mean      :-0.6304      Mean      :-0.7550
## 3rd Qu.   :-0.2173      3rd Qu.   :-0.3602      3rd Qu.   :-0.5767
## Max.      : 0.4506      Max.      : 0.3000      Max.      : 0.2502
## fBodyAcc.std...X fBodyAcc.std...Y fBodyAcc.std...Z fBodyAccJerk.std...X
## Min.      :-0.9966      Min.      :-0.99068      Min.      :-0.9872      Min.      :-0.9951
## 1st Qu.   :-0.9820      1st Qu.   :-0.94042      1st Qu.   :-0.9459      1st Qu.   :-0.9847
## Median    :-0.7470      Median    :-0.51338      Median    :-0.6441      Median    :-0.8254
## Mean      :-0.5522      Mean      :-0.48148      Mean      :-0.5824      Mean      :-0.6121
## 3rd Qu.   :-0.1966      3rd Qu.   :-0.07913      3rd Qu.   :-0.2655      3rd Qu.   :-0.2475
## Max.      : 0.6585      Max.      : 0.56019      Max.      : 0.6871      Max.      : 0.4768
## fBodyAccJerk.std...Y fBodyAccJerk.std...Z fBodyGyro.std...X fBodyGyro.std...Y
## Min.      :-0.9905      Min.      :-0.993108      Min.      :-0.9947      Min.      :-0.9944
## 1st Qu.   :-0.9737      1st Qu.   :-0.983747      1st Qu.   :-0.9750      1st Qu.   :-0.9602
## Median    :-0.7852      Median    :-0.895121      Median    :-0.8086      Median    :-0.7964
## Mean      :-0.5707      Mean      :-0.756489      Mean      :-0.7110      Mean      :-0.6454
## 3rd Qu.   :-0.1685      3rd Qu.   :-0.543787      3rd Qu.   :-0.4813      3rd Qu.   :-0.4154
## Max.      : 0.3498      Max.      :-0.006236      Max.      : 0.1966      Max.      : 0.6462
## fBodyGyro.std...Z fBodyAccMag.std.. fBodyBodyAccJerkMag.std..
## Min.      :-0.9867      Min.      :-0.9876      Min.      :-0.9944
## 1st Qu.   :-0.9643      1st Qu.   :-0.9452      1st Qu.   :-0.9752
## Median    :-0.8224      Median    :-0.6513      Median    :-0.8126
## Mean      :-0.6577      Mean      :-0.6210      Mean      :-0.5992
## 3rd Qu.   :-0.3916      3rd Qu.   :-0.3654      3rd Qu.   :-0.2668
## Max.      : 0.5225      Max.      : 0.1787      Max.      : 0.3163
## fBodyBodyGyroMag.std.. fBodyBodyGyroJerkMag.std..
## Min.      :-0.9815      Min.      :-0.9976
## 1st Qu.   :-0.9488      1st Qu.   :-0.9802
## Median    :-0.7727      Median    :-0.8941
## Mean      :-0.6723      Mean      :-0.7715
## 3rd Qu.   :-0.4277      3rd Qu.   :-0.6081
## Max.      : 0.2367      Max.      : 0.2878

```