Data Cleaning

group

September 17, 2019

Project on Scraping and Cleaning Data

This data is acquired from 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.

Information on the data may be found:

https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones#

Preparing workspace

This cleans up the workspace, removing exiting items:

```
rm(list = ls())
```

Getting and loading the required packages

This loads the required packages for the data cleaning process including a condition to install the packages if they are not already installed:

```
if ("dplyr" %in% row.names(installed.packages()) == FALSE) {
    install.packages("dplyr")
}
if ("tidyr" %in% row.names(installed.packages()) == FALSE) {
    install.packages("tidyr")
}

library("dplyr")
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
##     intersect, setdiff, setequal, union
```

```
library("tidyr")
```

Extracting data from the web

Here the data is sourced from the website and stored. A check is also done to know if the data already exists.

```
file_url <- "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip"

if (!dir.exists("data2")) {
    dir.create("data2")
}

data_path = "./data2/UCI HAR Dataset"

if (!file.exists(data_path)) {
    download.file(file_url, "./data2/UCI_data.zip")
    unzip("./data2/UCI_data.zip", exdir = "./data2")

}
```

Merging the test and train datasets to create one dataset

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 different files are read in and training and test sets are merged for subsequent cleaning:

```
list.files("./data2/UCI HAR Dataset/")
```

```
## [1] "activity_labels.txt" "features.txt"        "features_info.txt"
## [4] "README.txt"          "test"                "train"
```

```
# reading in train set values
x_train <- read.table(file = "./data2/UCI HAR Dataset/train/X_train.txt")
y_train <- read.table(file = "./data2/UCI HAR Dataset/train/y_train.txt",
    col.names = c("activity"))
sub_train <- read.table(file = "./data2/UCI HAR Dataset/train/subject_train.txt",
    col.names = c("subject"))

# combining the columns to make training data
dt_train <- cbind(sub_train, y_train, x_train)

# reading in the test set values
```

```r
x_test <- read.table(file = "./data2/UCI HAR Dataset/test/X_test.txt")
y_test <- read.table(file = "./data2/UCI HAR Dataset/test/y_test.txt",
    col.names = c("activity"))
sub_test <- read.table(file = "./data2/UCI HAR Dataset/test/subject_test.txt",
    col.names = c("subject"))

# combinig the columns to make the training data
dt_test <- cbind(sub_test, y_test, x_test)

# merging test and train data sets
dt_merged <- rbind(dt_train, dt_test)

# displaying a sample data entry
print(head(dt_merged, 1))
```

```
##   subject activity        V1          V2         V3         V4         V5
## 1       1        5 0.2885845 -0.02029417 -0.1329051 -0.9952786 -0.9831106
##           V6         V7         V8        V9        V10        V11
## 1 -0.9135264 -0.9951121 -0.9831846 -0.923527 -0.9347238 -0.5673781
##          V12       V13       V14       V15        V16        V17       V18
## 1 -0.7444125 0.8529474 0.6858446 0.8142628 -0.9655228 -0.9999446 -0.999863
##          V19        V20        V21       V22        V23        V24
## 1 -0.9946122 -0.9942308 -0.9876139 -0.94322 -0.4077471 -0.6793375
##          V25       V26        V27       V28         V29       V30
## 1 -0.6021219 0.9292935 -0.8530111 0.3599098 -0.05852638 0.2568915
##          V31       V32         V33       V34        V35      V36
## 1 -0.2248476 0.2641057 -0.09524563 0.2788514 -0.4650846 0.491936
##          V37       V38       V39       V40       V41        V42       V43
## 1 -0.1908836 0.3763139 0.4351292 0.6607903 0.9633961 -0.1408397 0.1153749
##          V44        V45       V46        V47        V48        V49
## 1 -0.9852497 -0.9817084 -0.877625 -0.9850014 -0.9844162 -0.8946774
##         V50        V51       V52       V53        V54        V55       V56
## 1 0.8920545 -0.1612655 0.1246598 0.9774363 -0.1232134 0.05648273 -0.375426
##         V57        V58        V59        V60        V61        V62 V63 V64
## 1 0.8994686 -0.9709052 -0.9755104 -0.9843254 -0.9888491 -0.9177426  -1  -1
##         V65       V66       V67        V68       V69        V70       V71
## 1 0.1138061 -0.590425 0.5911463 -0.5917735 0.5924693 -0.7454488 0.7208617
##          V72    V73        V74       V75        V76       V77       V78
## 1 -0.7123724 0.7113 -0.9951116 0.9956749 -0.9956676 0.9916527 0.5702216
##         V79       V80        V81         V82         V83        V84
## 1 0.4390273 0.9869131 0.07799634 0.005000803 -0.06783081 -0.9935191
##        V85       V86        V87        V88        V89        V90
## 1 -0.98836 -0.993575 -0.9944876 -0.9862066 -0.9928183 -0.9851801
##          V91        V92       V93       V94       V95       V96        V97
## 1 -0.9919942 -0.9931189 0.9898347 0.9919569 0.9905192 -0.993522 -0.9999349
##          V98        V99      V100       V101       V102       V103
## 1 -0.9998204 -0.9998785 -0.994364 -0.9860249 -0.9892336 -0.8199492
##         V104       V105 V106      V107      V108      V109      V110
## 1 -0.7930464 -0.8888529    1 -0.220747 0.6368308 0.3876436 0.2414015
##          V111      V112      V113      V114       V115      V116
## 1 -0.05225285 0.2641772 0.3734395 0.3417775 -0.5697912 0.2653988
##         V117       V118       V119       V120         V121        V122
## 1 -0.4778749 -0.3853005 0.03364394 -0.1265108 -0.006100849 -0.03136479
##        V123       V124       V125       V126       V127       V128
```

```
## 1 0.1077254 -0.9853103 -0.9766234 -0.9922053 -0.9845863 -0.9763526
##         V129       V130      V131       V132     V133      V134      V135
## 1 -0.9923616 -0.8670437 -0.933786 -0.7475662 0.847308 0.9148953 0.8308405
##         V136       V137       V138       V139       V140      V141
## 1 -0.9671843 -0.9995783 -0.9993543 -0.9997634 -0.9834381 -0.978614
##         V142       V143      V144       V145       V146       V147
## 1 -0.9929656 0.08263168 0.2022676 -0.1687567 0.09632324 -0.2749851
##        V148       V149 V150       V151      V152      V153      V154
## 1 0.4986442 -0.2203169    1 -0.9729714 0.3166545 0.3757264 0.7233992
##        V155      V156      V157      V158      V159      V160       V161
## 1 -0.771112 0.6902132 -0.331831 0.7095838 0.1348734 0.3010995 -0.0991674
##          V162       V163       V164       V165       V166       V167
## 1 -0.05551737 -0.0619858 -0.9921107 -0.9925193 -0.9920553 -0.9921648
##         V168      V169       V170       V171       V172      V173
## 1 -0.9949416 -0.992619 -0.9901558 -0.9867428 -0.9920416 0.9944288
##        V174      V175       V176       V177       V178       V179
## 1 0.9917558 0.9893519 -0.9944534 -0.9999375 -0.9999535 -0.9999229
##         V180       V181      V182      V183      V184       V185      V186
## 1 -0.9922997 -0.9969389 -0.992243 -0.589851 -0.688459 -0.5721069 0.2923763
##        V187      V188        V189      V190       V191      V192      V193
## 1 -0.361998 0.4055427 -0.03900695 0.9892838 -0.4145605 0.3916025 0.2822509
##        V194     V195      V196      V197      V198        V199      V200
## 1 0.9272698 -0.57237 0.6916192 0.4682898 -0.131077 -0.08715969 0.3362475
##         V201       V202       V203       V204       V205       V206
## 1 -0.9594339 -0.9505515 -0.9579929 -0.9463052 -0.9925557 -0.9594339
##         V207       V208       V209       V210        V211       V212
## 1 -0.9984928 -0.9576374 -0.2325816 -0.1731787 -0.02289666 0.09483157
##        V213       V214       V215       V216       V217       V218
## 1 0.1918171 -0.9594339 -0.9505515 -0.9579929 -0.9463052 -0.9925557
##         V219       V220       V221       V222       V223        V224
## 1 -0.9594339 -0.9984928 -0.9576374 -0.2325816 -0.1731787 -0.02289666
##         V225      V226       V227       V228       V229      V230
## 1 0.09483157 0.1918171 -0.9933059 -0.9943364 -0.9945004 -0.992784
##         V231       V232       V233       V234       V235      V236
## 1 -0.9912085 -0.9933059 -0.9998919 -0.9929337 -0.8634148 0.2830852
##         V237       V238        V239       V240       V241       V242
## 1 -0.2373087 -0.1054322 -0.03821231 -0.9689591 -0.9643352 -0.9572448
##         V243       V244       V245       V246       V247       V248
## 1 -0.9750599 -0.9915537 -0.9689591 -0.9992865 -0.9497658 0.07257904
##        V249       V250      V251     V252       V253       V254      V255
## 1 0.5725114 -0.7386022 0.2125778 0.433405 -0.9942478 -0.9913676 -0.993143
##         V256      V257       V258      V259       V260       V261
## 1 -0.9889356 -0.993486 -0.9942478 -0.999949 -0.9945472 -0.6197676
##        V262       V263       V264       V265       V266       V267
## 1 0.2928405 -0.1768892 -0.1457792 -0.1240723 -0.9947832 -0.9829841
##         V268       V269      V270      V271       V272       V273
## 1 -0.9392687 -0.9954217 -0.983133 -0.906165 -0.9968886 -0.9845193
##        V274       V275       V276       V277       V278       V279
## 1 -0.932082 -0.9937563 -0.9831629 -0.8850542 -0.9939619 -0.9934461
##         V280       V281       V282       V283       V284      V285
## 1 -0.9234277 -0.9747327 -0.9999684 -0.9996891 -0.9948915 -0.995926
##         V286       V287       V288       V289       V290 V291 V292 V293
## 1 -0.9897089 -0.9879912 -0.9463569 -0.9047478 -0.5913025   -1   -1   -1
##        V294      V295        V296      V297       V298       V299
```

```
## 1 0.2524829 0.1318358 -0.05205025 0.1420506 -0.1506825 -0.2205469
##         V300      V301         V302       V303       V304       V305
## 1 -0.5587385 0.2467687 -0.007415521 -0.9999628 -0.9999865 -0.9999791
##         V306       V307       V308       V309       V310       V311
## 1 -0.9999624 -0.9999322 -0.9997251 -0.9996704 -0.9999858 -0.9999687
##         V312       V313       V314       V315       V316       V317
## 1 -0.9999769 -0.9998697 -0.9997761 -0.9999712 -0.9999193 -0.9996568
##         V318      V319      V320       V321       V322       V323
## 1 -0.9998605 -0.999867 -0.999863 -0.9997378 -0.9997322 -0.9994926
##         V324       V325       V326       V327      V328       V329
## 1 -0.9998136 -0.9996818 -0.9998394 -0.9997382 -0.999612 -0.9996872
##         V330       V331       V332      V333       V334       V335
## 1 -0.9998386 -0.9935923 -0.9994758 -0.999662 -0.9996423 -0.9992934
##         V336       V337       V338       V339       V340       V341
## 1 -0.9978922 -0.9959325 -0.9951464 -0.9947399 -0.9996883 -0.9989246
##         V342       V343       V344       V345       V346       V347
## 1 -0.9956713 -0.9948773 -0.9994544 -0.9923325 -0.9871699 -0.9896961
##         V348       V349       V350       V351       V352       V353
## 1 -0.9958207 -0.9909363 -0.9970517 -0.9938055 -0.9905187 -0.9969928
##         V354       V355       V356       V357       V358       V359
## 1 -0.9967369 -0.9919752 -0.9932417 -0.9983491 -0.9911084 -0.9598854
##        V360       V361       V362       V363       V364       V365
## 1 -0.990515 -0.9999347 -0.9998205 -0.9998845 -0.9930263 -0.9913734
##         V366 V367 V368 V369 V370  V371 V372      V373     V374      V375
## 1 -0.9962396   -1   -1   -1    1 -0.24   -1 0.8703845 0.210697 0.2637079
##         V376       V377       V378       V379       V380       V381
## 1 -0.7036858 -0.9037425 -0.5825736 -0.9363101 -0.5073447 -0.8055359
##         V382       V383       V384       V385       V386       V387
## 1 -0.9999865 -0.9999796 -0.9999748 -0.9999551 -0.9999186 -0.9996401
##         V388       V389       V390       V391      V392       V393
## 1 -0.9994833 -0.9999609 -0.9999823 -0.9999707 -0.999811 -0.9994847
##         V394       V395       V396       V397       V398       V399
## 1 -0.9999808 -0.9998519 -0.9999326 -0.9998999 -0.9998244 -0.9998598
##         V400       V401       V402       V403       V404       V405
## 1 -0.9997275 -0.9997288 -0.9995671 -0.9997652 -0.9999002 -0.9998149
##         V406       V407       V408       V409       V410       V411
## 1 -0.9997098 -0.9995961 -0.9998522 -0.9998221 -0.9993999 -0.9997656
##         V412       V413       V414       V415       V416       V417
## 1 -0.9999585 -0.9999495 -0.9998385 -0.9998135 -0.9987805 -0.9985778
##         V418       V419       V420       V421       V422       V423
## 1 -0.9996197 -0.9999836 -0.9998281 -0.9986807 -0.9998442 -0.9999279
##         V424       V425       V426       V427       V428       V429
## 1 -0.9865744 -0.9817615 -0.9895148 -0.9850326 -0.9738861 -0.9940349
##         V430       V431      V432       V433       V434       V435
## 1 -0.9865308 -0.9836164 -0.992352 -0.9804984 -0.9722709 -0.9949443
##         V436       V437       V438       V439       V440       V441
## 1 -0.9975686 -0.9840851 -0.9943354 -0.9852762 -0.9998637 -0.9996661
##         V442       V443       V444       V445       V446       V447
## 1 -0.9999346 -0.9903439 -0.9948357 -0.9944116 -0.7124023 -0.6448424
##        V448 V449 V450 V451       V452       V453     V454      V455
## 1 -0.838993   -1   -1   -1 -0.2575489 0.09794711 0.547151 0.3773112
##        V456     V457        V458       V459       V460      V461
## 1 0.1340915 0.273372 -0.09126183 -0.4843465 -0.7828507 -0.999865
##         V462       V463       V464       V465       V466      V467
```

```
## 1 -0.9999318 -0.9999729 -0.9999702 -0.9999301 -0.9999586 -0.999929
##         V468       V469       V470       V471       V472       V473
## 1 -0.9999847 -0.9998633 -0.9999681 -0.9999361 -0.9999536 -0.9998644
##        V474       V475       V476       V477       V478       V479
## 1 -0.999961 -0.9994537 -0.9999781 -0.9999915 -0.9999901 -0.9999686
##         V480       V481       V482       V483       V484       V485
## 1 -0.9998066 -0.998346 -0.9989612 -0.9996187 -0.9999893 -0.9999354
##         V486       V487       V488       V489       V490       V491
## 1 -0.9983875 -0.9996426 -0.9999727 -0.9999554 -0.9999763 -0.9999058
##         V492       V493       V494       V495       V496       V497
## 1 -0.9999855 -0.9999372 -0.9997512 -0.9990723 -0.9999275 -0.9999516
##         V498       V499       V500       V501       V502       V503
## 1 -0.9999058 -0.9998927 -0.9994443 -0.999941 -0.9999586 -0.9521547
##        V504       V505       V506       V507       V508       V509
## 1 -0.956134 -0.9488701 -0.9743206 -0.9257218 -0.9521547 -0.9982852
##         V510       V511       V512        V513       V514       V515
## 1 -0.9732732 -0.6463764 -0.7931035 -0.08843612 -0.436471 -0.7968405
##         V516       V517       V518       V519       V520       V521
## 1 -0.9937257 -0.993755 -0.9919757 -0.9933647 -0.9881754 -0.9937257
##         V522       V523 V524       V525       V526       V527      V528
## 1 -0.9999184 -0.9913637   -1 -0.9365079 0.3469885 -0.5160801 -0.80276
##         V529       V530       V531       V532       V533       V534
## 1 -0.9801349 -0.9613094 -0.9736534 -0.9522638 -0.9894981 -0.9801349
##         V535       V536       V537 V538       V539       V540       V541
## 1 -0.9992403 -0.9926555 -0.7012914   -1 -0.1289889 0.5861564 0.3746046
##         V542       V543       V544       V545       V546       V547
## 1 -0.9919904 -0.9906975 -0.9899408 -0.9924478 -0.9910477 -0.9919904
##         V548       V549       V550 V551        V552       V553       V554
## 1 -0.9999368 -0.9904579 -0.8713058   -1 -0.07432303 -0.2986764 -0.7103041
##         V555       V556       V557        V558       V559       V560
## 1 -0.1127543 0.03040037 -0.4647614 -0.01844588 -0.8412468 0.1799406
##          V561
## 1 -0.05862692
```

Extracting the feature names

The feature names are extracted from the relevant file here and additional feature names (`"subject"`, `"activity"`) are included:

```
features <- read.table(file = "./data2/UCI HAR Dataset/features.txt",
    stringsAsFactors = FALSE, col.names = c("id", "name"))
feature_extended <- c("subject", "activity", features$name)

# displaying some of the features
print(feature_extended[1:20])
```

```
##  [1] "subject"             "activity"            "tBodyAcc-mean()-X"
##  [4] "tBodyAcc-mean()-Y"   "tBodyAcc-mean()-Z"   "tBodyAcc-std()-X"
##  [7] "tBodyAcc-std()-Y"    "tBodyAcc-std()-Z"    "tBodyAcc-mad()-X"
## [10] "tBodyAcc-mad()-Y"    "tBodyAcc-mad()-Z"    "tBodyAcc-max()-X"
## [13] "tBodyAcc-max()-Y"    "tBodyAcc-max()-Z"    "tBodyAcc-min()-X"
## [16] "tBodyAcc-min()-Y"    "tBodyAcc-min()-Z"    "tBodyAcc-sma()"
## [19] "tBodyAcc-energy()-X" "tBodyAcc-energy()-Y"
```

Extracting only measurements on mean and standard deviation

A decision is made here to extract only features (columns) which includes `mean,std,subject,activity` in the naming while dropping those which include `freq`. This is an arbitrary choice to simplify the data:

```
feature_select <- feature_extended[grepl("mean|std|subject|activity",
    feature_extended) & !grepl("Freq|freq", feature_extended)]
feature_select_index <- grepl("mean|std|subject|activity", feature_extended) &
    !grepl("Freq|freq", feature_extended)

# filtering the needed variables in merged dataset by the
# position of the required features
dt_merged_filtered <- dt_merged[, feature_select_index]

head(dt_merged_filtered, 1)
```

```
##   subject activity        V1          V2         V3         V4         V5
## 1       1        5 0.2885845 -0.02029417 -0.1329051 -0.9952786 -0.9831106
##           V6       V41        V42       V43        V44        V45
## 1 -0.9135264 0.9633961 -0.1408397 0.1153749 -0.9852497 -0.9817084
##         V46        V81         V82         V83        V84      V85
## 1 -0.877625 0.07799634 0.005000803 -0.06783081 -0.9935191 -0.98836
##         V86         V121        V122      V123       V124       V125
## 1 -0.993575 -0.006100849 -0.03136479 0.1077254 -0.9853103 -0.9766234
##         V126       V161        V162       V163       V164       V165
## 1 -0.9922053 -0.0991674 -0.05551737 -0.0619858 -0.9921107 -0.9925193
##         V166       V201       V202       V214       V215       V227
## 1 -0.9920553 -0.9594339 -0.9505515 -0.9594339 -0.9505515 -0.9933059
##         V228       V240       V241       V253       V254       V266
## 1 -0.9943364 -0.9689591 -0.9643352 -0.9942478 -0.9913676 -0.9947832
##         V267       V268       V269      V270      V271       V345
## 1 -0.9829841 -0.9392687 -0.9954217 -0.983133 -0.906165 -0.9923325
##         V346       V347       V348       V349       V350       V424
## 1 -0.9871699 -0.9896961 -0.9958207 -0.9909363 -0.9970517 -0.9865744
##         V425       V426       V427       V428       V429       V503
## 1 -0.9817615 -0.9895148 -0.9850326 -0.9738861 -0.9940349 -0.9521547
##        V504       V516      V517       V529       V530       V542
## 1 -0.956134 -0.9937257 -0.993755 -0.9801349 -0.9613094 -0.9919904
##         V543
## 1 -0.9906975
```

Cleaning up variable names

The activity labels are extracted from the relevant file and merged with the data:

```
# reading in activity names
activity_labels <- read.table(file = "./data2/UCI HAR Dataset/activity_labels.txt",
    stringsAsFactors = FALSE, col.names = c("id", "name"))

# 3. fixing corresponding descriptive activity name to the
# values in the merged dataset as a variabe called activity
# name
dt_merged_filtered <- mutate(dt_merged_filtered, activity_name = as.character(factor(dt_merged_filtered$
    labels = activity_labels$name)))
```

Cleaning feature names and using descriptive names

The variable names are cleaned up to provide a more meaningful representation:

```
feature_select_clean <- gsub("\\(\\)", "", feature_select)
feature_select_clean <- gsub("^t(.*)$", "\\1-timedomain", feature_select_clean)
feature_select_clean <- gsub("^f(.*)$", "\\1-freqdomain", feature_select_clean)
feature_select_clean <- gsub("(Jerk|Gyro)", "-\\1", feature_select_clean)
feature_select_clean <- gsub("Mag", "-Magnitude", feature_select_clean)
feature_select_clean <- gsub("Acc", "-Acceleration", feature_select_clean)
feature_select_clean <- gsub("BodyBody", "Body", feature_select_clean)
feature_select_clean <- tolower(feature_select_clean)

# fixing variable names for the filtered dataset
names(dt_merged_filtered) <- c(feature_select_clean, "activity_name")

# displaying some of the features
print(feature_select_clean[1:20])
```

```
##  [1] "subject"
##  [2] "activity"
##  [3] "body-acceleration-mean-x-timedomain"
##  [4] "body-acceleration-mean-y-timedomain"
##  [5] "body-acceleration-mean-z-timedomain"
##  [6] "body-acceleration-std-x-timedomain"
##  [7] "body-acceleration-std-y-timedomain"
##  [8] "body-acceleration-std-z-timedomain"
##  [9] "gravity-acceleration-mean-x-timedomain"
## [10] "gravity-acceleration-mean-y-timedomain"
## [11] "gravity-acceleration-mean-z-timedomain"
## [12] "gravity-acceleration-std-x-timedomain"
## [13] "gravity-acceleration-std-y-timedomain"
## [14] "gravity-acceleration-std-z-timedomain"
## [15] "body-acceleration-jerk-mean-x-timedomain"
## [16] "body-acceleration-jerk-mean-y-timedomain"
## [17] "body-acceleration-jerk-mean-z-timedomain"
## [18] "body-acceleration-jerk-std-x-timedomain"
## [19] "body-acceleration-jerk-std-y-timedomain"
## [20] "body-acceleration-jerk-std-z-timedomain"
```

```
# displaying a sample data entry
print(head(dt_merged_filtered, 1))
```

```
##   subject activity body-acceleration-mean-x-timedomain
## 1       1        5                           0.2885845
##   body-acceleration-mean-y-timedomain body-acceleration-mean-z-timedomain
## 1                         -0.02029417                          -0.1329051
##   body-acceleration-std-x-timedomain body-acceleration-std-y-timedomain
## 1                         -0.9952786                         -0.9831106
##   body-acceleration-std-z-timedomain
## 1                         -0.9135264
##   gravity-acceleration-mean-x-timedomain
## 1                              0.9633961
##   gravity-acceleration-mean-y-timedomain
## 1                             -0.1408397
##   gravity-acceleration-mean-z-timedomain
```

```
## 1                              0.1153749
##   gravity-acceleration-std-x-timedomain
## 1                            -0.9852497
##   gravity-acceleration-std-y-timedomain
## 1                            -0.9817084
##   gravity-acceleration-std-z-timedomain
## 1                             -0.877625
##   body-acceleration-jerk-mean-x-timedomain
## 1                              0.07799634
##   body-acceleration-jerk-mean-y-timedomain
## 1                             0.005000803
##   body-acceleration-jerk-mean-z-timedomain
## 1                            -0.06783081
##   body-acceleration-jerk-std-x-timedomain
## 1                             -0.9935191
##   body-acceleration-jerk-std-y-timedomain
## 1                               -0.98836
##   body-acceleration-jerk-std-z-timedomain body-gyro-mean-x-timedomain
## 1                              -0.993575                -0.006100849
##   body-gyro-mean-y-timedomain body-gyro-mean-z-timedomain
## 1                -0.03136479                   0.1077254
##   body-gyro-std-x-timedomain body-gyro-std-y-timedomain
## 1                -0.9853103                 -0.9766234
##   body-gyro-std-z-timedomain body-gyro-jerk-mean-x-timedomain
## 1                -0.9922053                       -0.0991674
##   body-gyro-jerk-mean-y-timedomain body-gyro-jerk-mean-z-timedomain
## 1                      -0.05551737                       -0.0619858
##   body-gyro-jerk-std-x-timedomain body-gyro-jerk-std-y-timedomain
## 1                      -0.9921107                      -0.9925193
##   body-gyro-jerk-std-z-timedomain
## 1                      -0.9920553
##   body-acceleration-magnitude-mean-timedomain
## 1                                -0.9594339
##   body-acceleration-magnitude-std-timedomain
## 1                               -0.9505515
##   gravity-acceleration-magnitude-mean-timedomain
## 1                                 -0.9594339
##   gravity-acceleration-magnitude-std-timedomain
## 1                                -0.9505515
##   body-acceleration-jerk-magnitude-mean-timedomain
## 1                                  -0.9933059
##   body-acceleration-jerk-magnitude-std-timedomain
## 1                                 -0.9943364
##   body-gyro-magnitude-mean-timedomain body-gyro-magnitude-std-timedomain
## 1                          -0.9689591                          -0.9643352
##   body-gyro-jerk-magnitude-mean-timedomain
## 1                              -0.9942478
##   body-gyro-jerk-magnitude-std-timedomain
## 1                             -0.9913676
##   body-acceleration-mean-x-freqdomain body-acceleration-mean-y-freqdomain
## 1                          -0.9947832                          -0.9829841
##   body-acceleration-mean-z-freqdomain body-acceleration-std-x-freqdomain
## 1                          -0.9392687                          -0.9954217
##   body-acceleration-std-y-freqdomain body-acceleration-std-z-freqdomain
```

```
## 1                          -0.983133                           -0.906165
##   body-acceleration-jerk-mean-x-freqdomain
## 1                              -0.9923325
##   body-acceleration-jerk-mean-y-freqdomain
## 1                              -0.9871699
##   body-acceleration-jerk-mean-z-freqdomain
## 1                              -0.9896961
##   body-acceleration-jerk-std-x-freqdomain
## 1                             -0.9958207
##   body-acceleration-jerk-std-y-freqdomain
## 1                             -0.9909363
##   body-acceleration-jerk-std-z-freqdomain body-gyro-mean-x-freqdomain
## 1                             -0.9970517                   -0.9865744
##   body-gyro-mean-y-freqdomain body-gyro-mean-z-freqdomain
## 1                  -0.9817615                  -0.9895148
##   body-gyro-std-x-freqdomain body-gyro-std-y-freqdomain
## 1                 -0.9850326                 -0.9738861
##   body-gyro-std-z-freqdomain body-acceleration-magnitude-mean-freqdomain
## 1                 -0.9940349                                   -0.9521547
##   body-acceleration-magnitude-std-freqdomain
## 1                                  -0.956134
##   body-acceleration-jerk-magnitude-mean-freqdomain
## 1                                       -0.9937257
##   body-acceleration-jerk-magnitude-std-freqdomain
## 1                                       -0.993755
##   body-gyro-magnitude-mean-freqdomain body-gyro-magnitude-std-freqdomain
## 1                          -0.9801349                           -0.9613094
##   body-gyro-jerk-magnitude-mean-freqdomain
## 1                              -0.9919904
##   body-gyro-jerk-magnitude-std-freqdomain activity_name
## 1                             -0.9906975      STANDING
```

Creating a second, independent tidy data set with the average of each variable for each activity and each subject.

Here the observations for each subject and activity combination is grouped together and the mean is computed such that only one value is given for person 1 doing activity walking for example (as opposed to having several entries for person 1 walking):

```
# dt_tidy <- dt_merged_filtered %>% group_by(subject,
# activity) %>% summarise_each(funs(mean)) #%>% gather(key =
# measure, value = mean, -subject, -activity)

dt_tidy <- dt_merged_filtered %>% group_by(subject, activity,
    activity_name) %>% summarise_each(funs = list(mean))

# displaying a sample data entry
print(head(dt_tidy, 1))
```

```
## # A tibble: 1 x 69
## # Groups:   subject, activity [1]
##   subject activity activity_name `body-accelerat~ `body-accelerat~
##     <int>    <int> <chr>                    <dbl>            <dbl>
## 1       1        1 WALKING                  0.277          -0.0174
## # ... with 64 more variables: `body-acceleration-mean-z-timedomain` <dbl>,
```

```
## #   `body-acceleration-std-x-timedomain` <dbl>,
## #   `body-acceleration-std-y-timedomain` <dbl>,
## #   `body-acceleration-std-z-timedomain` <dbl>,
## #   `gravity-acceleration-mean-x-timedomain` <dbl>,
## #   `gravity-acceleration-mean-y-timedomain` <dbl>,
## #   `gravity-acceleration-mean-z-timedomain` <dbl>,
## #   `gravity-acceleration-std-x-timedomain` <dbl>,
## #   `gravity-acceleration-std-y-timedomain` <dbl>,
## #   `gravity-acceleration-std-z-timedomain` <dbl>,
## #   `body-acceleration-jerk-mean-x-timedomain` <dbl>,
## #   `body-acceleration-jerk-mean-y-timedomain` <dbl>,
## #   `body-acceleration-jerk-mean-z-timedomain` <dbl>,
## #   `body-acceleration-jerk-std-x-timedomain` <dbl>,
## #   `body-acceleration-jerk-std-y-timedomain` <dbl>,
## #   `body-acceleration-jerk-std-z-timedomain` <dbl>,
## #   `body-gyro-mean-x-timedomain` <dbl>,
## #   `body-gyro-mean-y-timedomain` <dbl>,
## #   `body-gyro-mean-z-timedomain` <dbl>,
## #   `body-gyro-std-x-timedomain` <dbl>,
## #   `body-gyro-std-y-timedomain` <dbl>,
## #   `body-gyro-std-z-timedomain` <dbl>,
## #   `body-gyro-jerk-mean-x-timedomain` <dbl>,
## #   `body-gyro-jerk-mean-y-timedomain` <dbl>,
## #   `body-gyro-jerk-mean-z-timedomain` <dbl>,
## #   `body-gyro-jerk-std-x-timedomain` <dbl>,
## #   `body-gyro-jerk-std-y-timedomain` <dbl>,
## #   `body-gyro-jerk-std-z-timedomain` <dbl>,
## #   `body-acceleration-magnitude-mean-timedomain` <dbl>,
## #   `body-acceleration-magnitude-std-timedomain` <dbl>,
## #   `gravity-acceleration-magnitude-mean-timedomain` <dbl>,
## #   `gravity-acceleration-magnitude-std-timedomain` <dbl>,
## #   `body-acceleration-jerk-magnitude-mean-timedomain` <dbl>,
## #   `body-acceleration-jerk-magnitude-std-timedomain` <dbl>,
## #   `body-gyro-magnitude-mean-timedomain` <dbl>,
## #   `body-gyro-magnitude-std-timedomain` <dbl>,
## #   `body-gyro-jerk-magnitude-mean-timedomain` <dbl>,
## #   `body-gyro-jerk-magnitude-std-timedomain` <dbl>,
## #   `body-acceleration-mean-x-freqdomain` <dbl>,
## #   `body-acceleration-mean-y-freqdomain` <dbl>,
## #   `body-acceleration-mean-z-freqdomain` <dbl>,
## #   `body-acceleration-std-x-freqdomain` <dbl>,
## #   `body-acceleration-std-y-freqdomain` <dbl>,
## #   `body-acceleration-std-z-freqdomain` <dbl>,
## #   `body-acceleration-jerk-mean-x-freqdomain` <dbl>,
## #   `body-acceleration-jerk-mean-y-freqdomain` <dbl>,
## #   `body-acceleration-jerk-mean-z-freqdomain` <dbl>,
## #   `body-acceleration-jerk-std-x-freqdomain` <dbl>,
## #   `body-acceleration-jerk-std-y-freqdomain` <dbl>,
## #   `body-acceleration-jerk-std-z-freqdomain` <dbl>,
## #   `body-gyro-mean-x-freqdomain` <dbl>,
## #   `body-gyro-mean-y-freqdomain` <dbl>,
## #   `body-gyro-mean-z-freqdomain` <dbl>,
## #   `body-gyro-std-x-freqdomain` <dbl>,
## #   `body-gyro-std-y-freqdomain` <dbl>,
```

```
## #   `body-gyro-std-z-freqdomain` <dbl>,
## #   `body-acceleration-magnitude-mean-freqdomain` <dbl>,
## #   `body-acceleration-magnitude-std-freqdomain` <dbl>,
## #   `body-acceleration-jerk-magnitude-mean-freqdomain` <dbl>,
## #   `body-acceleration-jerk-magnitude-std-freqdomain` <dbl>,
## #   `body-gyro-magnitude-mean-freqdomain` <dbl>,
## #   `body-gyro-magnitude-std-freqdomain` <dbl>,
## #   `body-gyro-jerk-magnitude-mean-freqdomain` <dbl>,
## #   `body-gyro-jerk-magnitude-std-freqdomain` <dbl>
```

Saving the data and clearing up the workspace

The now tidied data is saved in the relevant workspave as a text file:

```
# saving independent tidied data to a file
write.table(dt_tidy, file = "./data2/tidied_data.txt")


# Clearing the workspace
rm(list = ls())
```