

# run\_analysis.R

*ioannis*

*Mon May 30 00:37:15 2016*

```
##### Libraries #####
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

library(tidyr)

##### Working Directory #####
setwd('~/Documents/Coursera/Getting and Cleaning Data')

##### Folder Initialization and File Download #####
if(!file.exists('./data')){dir.create('./data')}
fileUrl <- 'https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip'
if(!file.exists('./data/Dataset.zip')) { download.file(
  fileUrl, destfile = './data/Dataset.zip', method = 'curl')
}
unzipped <- unzip('./data/Dataset.zip')
# cleanup
rm(fileUrl)

##### Dataframe Initialization #####
train_index <- grep('X_train', unzipped); test_index <- grep('X_test', unzipped)
train <- read.table(unzipped[train_index]); test <- read.table(unzipped[test_index])
train <- tbl_df(train); test <- tbl_df(test)
# cleanup
rm(train_index, test_index)

##### 1. Merging the training and the test sets #####
# Test that all names are in the same order and match: sum(!(names(test) == names(train))) gives 0
merged <- rbind_list(train, test)
# cleanup
rm(train, test)
merged <- tbl_df(merged)
head(merged)

## Source: local data frame [6 x 561]
##
```

```

##          V1          V2          V3          V4          V5          V6
## 1 0.2885845 -0.02029417 -0.1329051 -0.9952786 -0.9831106 -0.9135264
## 2 0.2784188 -0.01641057 -0.1235202 -0.9982453 -0.9753002 -0.9603220
## 3 0.2796531 -0.01946716 -0.1134617 -0.9953796 -0.9671870 -0.9789440
## 4 0.2791739 -0.02620065 -0.1232826 -0.9960915 -0.9834027 -0.9906751
## 5 0.2766288 -0.01656965 -0.1153619 -0.9981386 -0.9808173 -0.9904816
## 6 0.2771988 -0.01009785 -0.1051373 -0.9973350 -0.9904868 -0.9954200
## Variables not shown: V7 (dbl), V8 (dbl), V9 (dbl), V10 (dbl), V11 (dbl),
## V12 (dbl), V13 (dbl), V14 (dbl), V15 (dbl), V16 (dbl), V17 (dbl), V18
## (dbl), V19 (dbl), V20 (dbl), V21 (dbl), V22 (dbl), V23 (dbl), V24 (dbl),
## V25 (dbl), V26 (dbl), V27 (dbl), V28 (dbl), V29 (dbl), V30 (dbl), V31
## (dbl), V32 (dbl), V33 (dbl), V34 (dbl), V35 (dbl), V36 (dbl), V37 (dbl),
## V38 (dbl), V39 (dbl), V40 (dbl), V41 (dbl), V42 (dbl), V43 (dbl), V44
## (dbl), V45 (dbl), V46 (dbl), V47 (dbl), V48 (dbl), V49 (dbl), V50 (dbl),
## V51 (dbl), V52 (dbl), V53 (dbl), V54 (dbl), V55 (dbl), V56 (dbl), V57
## (dbl), V58 (dbl), V59 (dbl), V60 (dbl), V61 (dbl), V62 (dbl), V63 (dbl),
## V64 (dbl), V65 (dbl), V66 (dbl), V67 (dbl), V68 (dbl), V69 (dbl), V70
## (dbl), V71 (dbl), V72 (dbl), V73 (dbl), V74 (dbl), V75 (dbl), V76 (dbl),
## V77 (dbl), V78 (dbl), V79 (dbl), V80 (dbl), V81 (dbl), V82 (dbl), V83
## (dbl), V84 (dbl), V85 (dbl), V86 (dbl), V87 (dbl), V88 (dbl), V89 (dbl),
## V90 (dbl), V91 (dbl), V92 (dbl), V93 (dbl), V94 (dbl), V95 (dbl), V96
## (dbl), V97 (dbl), V98 (dbl), V99 (dbl), V100 (dbl), V101 (dbl), V102
## (dbl), V103 (dbl), V104 (dbl), V105 (dbl), V106 (dbl), V107 (dbl), V108
## (dbl), V109 (dbl), V110 (dbl), V111 (dbl), V112 (dbl), V113 (dbl), V114
## (dbl), V115 (dbl), V116 (dbl), V117 (dbl), V118 (dbl), V119 (dbl), V120
## (dbl), V121 (dbl), V122 (dbl), V123 (dbl), V124 (dbl), V125 (dbl), V126
## (dbl), V127 (dbl), V128 (dbl), V129 (dbl), V130 (dbl), V131 (dbl), V132
## (dbl), V133 (dbl), V134 (dbl), V135 (dbl), V136 (dbl), V137 (dbl), V138
## (dbl), V139 (dbl), V140 (dbl), V141 (dbl), V142 (dbl), V143 (dbl), V144
## (dbl), V145 (dbl), V146 (dbl), V147 (dbl), V148 (dbl), V149 (dbl), V150
## (dbl), V151 (dbl), V152 (dbl), V153 (dbl), V154 (dbl), V155 (dbl), V156
## (dbl), V157 (dbl), V158 (dbl), V159 (dbl), V160 (dbl), V161 (dbl), V162
## (dbl), V163 (dbl), V164 (dbl), V165 (dbl), V166 (dbl), V167 (dbl), V168
## (dbl), V169 (dbl), V170 (dbl), V171 (dbl), V172 (dbl), V173 (dbl), V174
## (dbl), V175 (dbl), V176 (dbl), V177 (dbl), V178 (dbl), V179 (dbl), V180
## (dbl), V181 (dbl), V182 (dbl), V183 (dbl), V184 (dbl), V185 (dbl), V186
## (dbl), V187 (dbl), V188 (dbl), V189 (dbl), V190 (dbl), V191 (dbl), V192
## (dbl), V193 (dbl), V194 (dbl), V195 (dbl), V196 (dbl), V197 (dbl), V198
## (dbl), V199 (dbl), V200 (dbl), V201 (dbl), V202 (dbl), V203 (dbl), V204
## (dbl), V205 (dbl), V206 (dbl), V207 (dbl), V208 (dbl), V209 (dbl), V210
## (dbl), V211 (dbl), V212 (dbl), V213 (dbl), V214 (dbl), V215 (dbl), V216
## (dbl), V217 (dbl), V218 (dbl), V219 (dbl), V220 (dbl), V221 (dbl), V222
## (dbl), V223 (dbl), V224 (dbl), V225 (dbl), V226 (dbl), V227 (dbl), V228
## (dbl), V229 (dbl), V230 (dbl), V231 (dbl), V232 (dbl), V233 (dbl), V234
## (dbl), V235 (dbl), V236 (dbl), V237 (dbl), V238 (dbl), V239 (dbl), V240
## (dbl), V241 (dbl), V242 (dbl), V243 (dbl), V244 (dbl), V245 (dbl), V246
## (dbl), V247 (dbl), V248 (dbl), V249 (dbl), V250 (dbl), V251 (dbl), V252
## (dbl), V253 (dbl), V254 (dbl), V255 (dbl), V256 (dbl), V257 (dbl), V258
## (dbl), V259 (dbl), V260 (dbl), V261 (dbl), V262 (dbl), V263 (dbl), V264
## (dbl), V265 (dbl), V266 (dbl), V267 (dbl), V268 (dbl), V269 (dbl), V270
## (dbl), V271 (dbl), V272 (dbl), V273 (dbl), V274 (dbl), V275 (dbl), V276
## (dbl), V277 (dbl), V278 (dbl), V279 (dbl), V280 (dbl), V281 (dbl), V282
## (dbl), V283 (dbl), V284 (dbl), V285 (dbl), V286 (dbl), V287 (dbl), V288
## (dbl), V289 (dbl), V290 (dbl), V291 (dbl), V292 (dbl), V293 (dbl), V294

```

```
## (dbl), V295 (dbl), V296 (dbl), V297 (dbl), V298 (dbl), V299 (dbl), V300
## (dbl), V301 (dbl), V302 (dbl), V303 (dbl), V304 (dbl), V305 (dbl), V306
## (dbl), V307 (dbl), V308 (dbl), V309 (dbl), V310 (dbl), V311 (dbl), V312
## (dbl), V313 (dbl), V314 (dbl), V315 (dbl), V316 (dbl), V317 (dbl), V318
## (dbl), V319 (dbl), V320 (dbl), V321 (dbl), V322 (dbl), V323 (dbl), V324
## (dbl), V325 (dbl), V326 (dbl), V327 (dbl), V328 (dbl), V329 (dbl), V330
## (dbl), V331 (dbl), V332 (dbl), V333 (dbl), V334 (dbl), V335 (dbl), V336
## (dbl), V337 (dbl), V338 (dbl), V339 (dbl), V340 (dbl), V341 (dbl), V342
## (dbl), V343 (dbl), V344 (dbl), V345 (dbl), V346 (dbl), V347 (dbl), V348
## (dbl), V349 (dbl), V350 (dbl), V351 (dbl), V352 (dbl), V353 (dbl), V354
## (dbl), V355 (dbl), V356 (dbl), V357 (dbl), V358 (dbl), V359 (dbl), V360
## (dbl), V361 (dbl), V362 (dbl), V363 (dbl), V364 (dbl), V365 (dbl), V366
## (dbl), V367 (dbl), V368 (dbl), V369 (dbl), V370 (dbl), V371 (dbl), V372
## (dbl), V373 (dbl), V374 (dbl), V375 (dbl), V376 (dbl), V377 (dbl), V378
## (dbl), V379 (dbl), V380 (dbl), V381 (dbl), V382 (dbl), V383 (dbl), V384
## (dbl), V385 (dbl), V386 (dbl), V387 (dbl), V388 (dbl), V389 (dbl), V390
## (dbl), V391 (dbl), V392 (dbl), V393 (dbl), V394 (dbl), V395 (dbl), V396
## (dbl), V397 (dbl), V398 (dbl), V399 (dbl), V400 (dbl), V401 (dbl), V402
## (dbl), V403 (dbl), V404 (dbl), V405 (dbl), V406 (dbl), V407 (dbl), V408
## (dbl), V409 (dbl), V410 (dbl), V411 (dbl), V412 (dbl), V413 (dbl), V414
## (dbl), V415 (dbl), V416 (dbl), V417 (dbl), V418 (dbl), V419 (dbl), V420
## (dbl), V421 (dbl), V422 (dbl), V423 (dbl), V424 (dbl), V425 (dbl), V426
## (dbl), V427 (dbl), V428 (dbl), V429 (dbl), V430 (dbl), V431 (dbl), V432
## (dbl), V433 (dbl), V434 (dbl), V435 (dbl), V436 (dbl), V437 (dbl), V438
## (dbl), V439 (dbl), V440 (dbl), V441 (dbl), V442 (dbl), V443 (dbl), V444
## (dbl), V445 (dbl), V446 (dbl), V447 (dbl), V448 (dbl), V449 (dbl), V450
## (dbl), V451 (dbl), V452 (dbl), V453 (dbl), V454 (dbl), V455 (dbl), V456
## (dbl), V457 (dbl), V458 (dbl), V459 (dbl), V460 (dbl), V461 (dbl), V462
## (dbl), V463 (dbl), V464 (dbl), V465 (dbl), V466 (dbl), V467 (dbl), V468
## (dbl), V469 (dbl), V470 (dbl), V471 (dbl), V472 (dbl), V473 (dbl), V474
## (dbl), V475 (dbl), V476 (dbl), V477 (dbl), V478 (dbl), V479 (dbl), V480
## (dbl), V481 (dbl), V482 (dbl), V483 (dbl), V484 (dbl), V485 (dbl), V486
## (dbl), V487 (dbl), V488 (dbl), V489 (dbl), V490 (dbl), V491 (dbl), V492
## (dbl), V493 (dbl), V494 (dbl), V495 (dbl), V496 (dbl), V497 (dbl), V498
## (dbl), V499 (dbl), V500 (dbl), V501 (dbl), V502 (dbl), V503 (dbl), V504
## (dbl), V505 (dbl), V506 (dbl), V507 (dbl), V508 (dbl), V509 (dbl), V510
## (dbl), V511 (dbl), V512 (dbl), V513 (dbl), V514 (dbl), V515 (dbl), V516
## (dbl), V517 (dbl), V518 (dbl), V519 (dbl), V520 (dbl), V521 (dbl), V522
## (dbl), V523 (dbl), V524 (dbl), V525 (dbl), V526 (dbl), V527 (dbl), V528
## (dbl), V529 (dbl), V530 (dbl), V531 (dbl), V532 (dbl), V533 (dbl), V534
## (dbl), V535 (dbl), V536 (dbl), V537 (dbl), V538 (dbl), V539 (dbl), V540
## (dbl), V541 (dbl), V542 (dbl), V543 (dbl), V544 (dbl), V545 (dbl), V546
## (dbl), V547 (dbl), V548 (dbl), V549 (dbl), V550 (dbl), V551 (dbl), V552
## (dbl), V553 (dbl), V554 (dbl), V555 (dbl), V556 (dbl), V557 (dbl), V558
## (dbl), V559 (dbl), V560 (dbl), V561 (dbl)
```

```
##### 2. Extract only mean and standard deviation from each observation ##### AND
##### 4. Appropriately label the data set with descriptive variable names. #####
# Roadmap
# 1. Open file features.txt to check how "mean" and "standard deviation" are coded
# 2. Load features into a dataframe
# 3. Keep only feature names that contain "mean" or "standard deviation" (as coded in 1.)
# 4. Keep the corresponding columns of the dataframe
keep_names <- 'mean()|std()'
```

```
features_index <- grep('features.txt', unzipped)
features <- read.table(unzipped[features_index]); features <- tbl_df(features)
features <- filter(features, grepl(keep_names, features$V2))
```

```
names(merged) <- extract_numeric(names(merged))
merged <- select(merged, features$V1)
names(merged) <- features$V2
head(names(merged))
```

```
## [1] "tBodyAcc-mean()-X" "tBodyAcc-mean()-Y" "tBodyAcc-mean()-Z"
## [4] "tBodyAcc-std()-X" "tBodyAcc-std()-Y" "tBodyAcc-std()-Z"
```

##### 3. Use descriptive activity names to name the activities in the data set #####

*# 1. Gather labels*

*# 2. Replace numbers with activity labels*

*# 3. Insert activity labels into the dataset*

```
activity_train_file <- grep('/y_train', unzipped); activity_test_file <- grep('/y_test', unzipped)
activity_train <- read.table(unzipped[activity_train_file]); activity_test <- read.table(unzipped[activity_test_file])
activities <- rbind_list(activity_train, activity_test); activities <- tbl_df(activities)
colnames(activities) <- "Activity"
activity_labels <- read.table(unzipped[grep('activity_labels', unzipped)]); activity_labels <- tbl_df(activity_labels)
colnames(activity_labels)[1] <- "Activity"
merged <- bind_cols(merged, activities)
merged <- tbl_df(merged)
merged <- left_join(merged, activity_labels)
```

## Joining by: "Activity"

```
merged$Activity <- NULL
merged <- rename(merged, Activity = V2)
rm(activity_train_file, activity_test_file)
head(merged$Activity)
```

```
## [1] STANDING STANDING STANDING STANDING STANDING STANDING
## 6 Levels: LAYING SITTING STANDING WALKING ... WALKING_UPSTAIRS
```

##### 5. From the data set in step 4, create a second,

##### independent tidy data set with the average of each variable for each activity and each subject. #

*# First, gather the subjects and add them to the merged dataframe*

```
subject_train_idx <- grep('/subject_train', unzipped); subject_test_idx <- grep('/subject_test', unzipped)
subject_train <- read.table(unzipped[subject_train_idx]); subject_test <- read.table(unzipped[subject_test_idx])
subjects <- rbind_list(subject_train, subject_test); subjects <- tbl_df(subjects)
colnames(subjects) <- 'SubjectID'
merged <- bind_cols(merged, subjects)
# make the new dataframe with summary statistics
grouped_data <- group_by(merged, Activity, SubjectID)
tidy_data <- summarise_each(grouped_data, funs(mean)) # <-- final clean dataset
head(tidy_data)
```

```
## Source: local data frame [6 x 81]
## Groups: Activity
```

```
##
##   Activity SubjectID tBodyAcc-mean()-X tBodyAcc-mean()-Y tBodyAcc-mean()-Z
## 1   LAYING         1      0.2215982      -0.04051395      -0.1132036
## 2   LAYING         2      0.2813734      -0.01815874      -0.1072456
## 3   LAYING         3      0.2755169      -0.01895568      -0.1013005
## 4   LAYING         4      0.2635592      -0.01500318      -0.1106882
## 5   LAYING         5      0.2783343      -0.01830421      -0.1079376
## 6   LAYING         6      0.2486565      -0.01025292      -0.1331196
## Variables not shown: tBodyAcc-std()-X (dbl), tBodyAcc-std()-Y (dbl),
##   tBodyAcc-std()-Z (dbl), tGravityAcc-mean()-X (dbl), tGravityAcc-mean()-Y
##   (dbl), tGravityAcc-mean()-Z (dbl), tGravityAcc-std()-X (dbl),
##   tGravityAcc-std()-Y (dbl), tGravityAcc-std()-Z (dbl),
##   tBodyAccJerk-mean()-X (dbl), tBodyAccJerk-mean()-Y (dbl),
##   tBodyAccJerk-mean()-Z (dbl), tBodyAccJerk-std()-X (dbl),
##   tBodyAccJerk-std()-Y (dbl), tBodyAccJerk-std()-Z (dbl),
##   tBodyGyro-mean()-X (dbl), tBodyGyro-mean()-Y (dbl), tBodyGyro-mean()-Z
##   (dbl), tBodyGyro-std()-X (dbl), tBodyGyro-std()-Y (dbl),
##   tBodyGyro-std()-Z (dbl), tBodyGyroJerk-mean()-X (dbl),
##   tBodyGyroJerk-mean()-Y (dbl), tBodyGyroJerk-mean()-Z (dbl),
##   tBodyGyroJerk-std()-X (dbl), tBodyGyroJerk-std()-Y (dbl),
##   tBodyGyroJerk-std()-Z (dbl), tBodyAccMag-mean() (dbl), tBodyAccMag-std()
##   (dbl), tGravityAccMag-mean() (dbl), tGravityAccMag-std() (dbl),
##   tBodyAccJerkMag-mean() (dbl), tBodyAccJerkMag-std() (dbl),
##   tBodyGyroMag-mean() (dbl), tBodyGyroMag-std() (dbl),
##   tBodyGyroJerkMag-mean() (dbl), tBodyGyroJerkMag-std() (dbl),
##   fBodyAcc-mean()-X (dbl), fBodyAcc-mean()-Y (dbl), fBodyAcc-mean()-Z
##   (dbl), fBodyAcc-std()-X (dbl), fBodyAcc-std()-Y (dbl), fBodyAcc-std()-Z
##   (dbl), fBodyAcc-meanFreq()-X (dbl), fBodyAcc-meanFreq()-Y (dbl),
##   fBodyAcc-meanFreq()-Z (dbl), fBodyAccJerk-mean()-X (dbl),
##   fBodyAccJerk-mean()-Y (dbl), fBodyAccJerk-mean()-Z (dbl),
##   fBodyAccJerk-std()-X (dbl), fBodyAccJerk-std()-Y (dbl),
##   fBodyAccJerk-std()-Z (dbl), fBodyAccJerk-meanFreq()-X (dbl),
##   fBodyAccJerk-meanFreq()-Y (dbl), fBodyAccJerk-meanFreq()-Z (dbl),
##   fBodyGyro-mean()-X (dbl), fBodyGyro-mean()-Y (dbl), fBodyGyro-mean()-Z
##   (dbl), fBodyGyro-std()-X (dbl), fBodyGyro-std()-Y (dbl),
##   fBodyGyro-std()-Z (dbl), fBodyGyro-meanFreq()-X (dbl),
##   fBodyGyro-meanFreq()-Y (dbl), fBodyGyro-meanFreq()-Z (dbl),
##   fBodyAccMag-mean() (dbl), fBodyAccMag-std() (dbl),
##   fBodyAccMag-meanFreq() (dbl), fBodyBodyAccJerkMag-mean() (dbl),
##   fBodyBodyAccJerkMag-std() (dbl), fBodyBodyAccJerkMag-meanFreq() (dbl),
##   fBodyBodyGyroMag-mean() (dbl), fBodyBodyGyroMag-std() (dbl),
##   fBodyBodyGyroMag-meanFreq() (dbl), fBodyBodyGyroJerkMag-mean() (dbl),
##   fBodyBodyGyroJerkMag-std() (dbl), fBodyBodyGyroJerkMag-meanFreq() (dbl)
```

```
write.csv(tidy_data, './tidy_data.csv')
```