Getting and Cleaning Data course Project

Downloading and unzip the Data

```
if(!file.exists("./data")){dir.create("./data")}
fileUrl <- "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip"
download.file(fileUrl,destfile="./data/Dataset.zip")
unzip(zipfile="./data/Dataset.zip",exdir="./data")
```

```
library(plyr)
```

```
## Warning: package 'plyr' was built under R version 3.2.5
```

Getting the List of Files

```
filepath<- file.path("./data" , "UCI HAR Dataset")
list<-list.files(filepath, recursive=TRUE)
list
```

```
##  [1] "activity_labels.txt"
##  [2] "features.txt"
##  [3] "features_info.txt"
##  [4] "README.txt"
##  [5] "test/Inertial Signals/body_acc_x_test.txt"
##  [6] "test/Inertial Signals/body_acc_y_test.txt"
##  [7] "test/Inertial Signals/body_acc_z_test.txt"
##  [8] "test/Inertial Signals/body_gyro_x_test.txt"
##  [9] "test/Inertial Signals/body_gyro_y_test.txt"
## [10] "test/Inertial Signals/body_gyro_z_test.txt"
## [11] "test/Inertial Signals/total_acc_x_test.txt"
## [12] "test/Inertial Signals/total_acc_y_test.txt"
## [13] "test/Inertial Signals/total_acc_z_test.txt"
## [14] "test/subject_test.txt"
## [15] "test/X_test.txt"
## [16] "test/y_test.txt"
## [17] "train/Inertial Signals/body_acc_x_train.txt"
## [18] "train/Inertial Signals/body_acc_y_train.txt"
## [19] "train/Inertial Signals/body_acc_z_train.txt"
## [20] "train/Inertial Signals/body_gyro_x_train.txt"
## [21] "train/Inertial Signals/body_gyro_y_train.txt"
## [22] "train/Inertial Signals/body_gyro_z_train.txt"
## [23] "train/Inertial Signals/total_acc_x_train.txt"
## [24] "train/Inertial Signals/total_acc_y_train.txt"
## [25] "train/Inertial Signals/total_acc_z_train.txt"
## [26] "train/subject_train.txt"
## [27] "train/X_train.txt"
## [28] "train/y_train.txt"
```

Merging the training and the test sets to create one data set.

Creating y Data set

```
y_test  <- read.table(file.path(filepath, "test" , "Y_test.txt"),header = FALSE)
y_train <- read.table(file.path(filepath, "train", "Y_train.txt"),header=FALSE)
y_dataset<-rbind(y_train, y_test)
str(y_dataset)
```

```
## 'data.frame':    10299 obs. of  1 variable:
##  $ V1: int  5 5 5 5 5 5 5 5 5 5 ...
```

Creating subject data set

```
sub_train <- read.table(file.path(filepath, "train", "subject_train.txt"),header=FALSE)
sub_test  <- read.table(file.path(filepath, "test" , "subject_test.txt" ),header=FALSE)
sub_dataset<-rbind(sub_train,sub_test)
```

Creating x data set

```
x_test  <- read.table(file.path(filepath, "test" , "X_test.txt" ),header = FALSE)
x_train <- read.table(file.path(filepath, "train", "X_train.txt"),header = FALSE)
x_dataset<-rbind(x_train, x_test)
```

Extracting only the measurements on the mean and standard deviation for each measurement

```
features<-read.table(file.path(filepath, "features.txt"),head=FALSE)
mean_std <- grep("-(mean|std)\\(\\)", features[, 2])
mn_std<-features$V2[grep("mean\\(\\)|std\\(\\)", features$V2)]
head(mean_std)
```

```
## [1] 1 2 3 4 5 6
```

```
head(mn_std)
```

```
## [1] tBodyAcc-mean()-X tBodyAcc-mean()-Y tBodyAcc-mean()-Z tBodyAcc-std()-X
## [5] tBodyAcc-std()-Y  tBodyAcc-std()-Z
## 477 Levels: angle(tBodyAccJerkMean),gravityMean) ...
```

Subseting and correcting the desired columns

```
x_data <- x_dataset[, mean_std]
names(x_data) <- features[mean_std, 2]
names(x_data)
```

```
##  [1] "tBodyAcc-mean()-X"           "tBodyAcc-mean()-Y"
##  [3] "tBodyAcc-mean()-Z"           "tBodyAcc-std()-X"
##  [5] "tBodyAcc-std()-Y"            "tBodyAcc-std()-Z"
##  [7] "tGravityAcc-mean()-X"        "tGravityAcc-mean()-Y"
##  [9] "tGravityAcc-mean()-Z"        "tGravityAcc-std()-X"
## [11] "tGravityAcc-std()-Y"         "tGravityAcc-std()-Z"
## [13] "tBodyAccJerk-mean()-X"       "tBodyAccJerk-mean()-Y"
## [15] "tBodyAccJerk-mean()-Z"       "tBodyAccJerk-std()-X"
## [17] "tBodyAccJerk-std()-Y"        "tBodyAccJerk-std()-Z"
## [19] "tBodyGyro-mean()-X"          "tBodyGyro-mean()-Y"
## [21] "tBodyGyro-mean()-Z"          "tBodyGyro-std()-X"
## [23] "tBodyGyro-std()-Y"           "tBodyGyro-std()-Z"
## [25] "tBodyGyroJerk-mean()-X"      "tBodyGyroJerk-mean()-Y"
## [27] "tBodyGyroJerk-mean()-Z"      "tBodyGyroJerk-std()-X"
## [29] "tBodyGyroJerk-std()-Y"       "tBodyGyroJerk-std()-Z"
## [31] "tBodyAccMag-mean()"          "tBodyAccMag-std()"
## [33] "tGravityAccMag-mean()"       "tGravityAccMag-std()"
## [35] "tBodyAccJerkMag-mean()"      "tBodyAccJerkMag-std()"
## [37] "tBodyGyroMag-mean()"         "tBodyGyroMag-std()"
## [39] "tBodyGyroJerkMag-mean()"     "tBodyGyroJerkMag-std()"
## [41] "fBodyAcc-mean()-X"           "fBodyAcc-mean()-Y"
## [43] "fBodyAcc-mean()-Z"           "fBodyAcc-std()-X"
## [45] "fBodyAcc-std()-Y"            "fBodyAcc-std()-Z"
## [47] "fBodyAccJerk-mean()-X"       "fBodyAccJerk-mean()-Y"
## [49] "fBodyAccJerk-mean()-Z"       "fBodyAccJerk-std()-X"
## [51] "fBodyAccJerk-std()-Y"        "fBodyAccJerk-std()-Z"
## [53] "fBodyGyro-mean()-X"          "fBodyGyro-mean()-Y"
## [55] "fBodyGyro-mean()-Z"          "fBodyGyro-std()-X"
## [57] "fBodyGyro-std()-Y"           "fBodyGyro-std()-Z"
## [59] "fBodyAccMag-mean()"          "fBodyAccMag-std()"
## [61] "fBodyBodyAccJerkMag-mean()"  "fBodyBodyAccJerkMag-std()"
## [63] "fBodyBodyGyroMag-mean()"     "fBodyBodyGyroMag-std()"
## [65] "fBodyBodyGyroJerkMag-mean()" "fBodyBodyGyroJerkMag-std()"
```

Use descriptive activity names to name the activities in the data set

```
activitylabels <- read.table(file.path(filepath, "activity_labels.txt"),header = FALSE)
y_dataset[, 1] <- activitylabels[y_dataset[, 1], 2]
names(y_dataset) <- "activity"
```

Appropriately label the data set with descriptive variable names

```
names(sub_dataset) <- "subject"
```

Creating a single data set

```
dataset <- cbind(x_data, y_dataset, sub_dataset)
```

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

```
newdata <- ddply(dataset, .(subject, activity), function(x) colMeans(x[, 1:65]))
head(newdata)
```

```
##   subject           activity tBodyAcc-mean()-X tBodyAcc-mean()-Y
## 1       1             LAYING         0.2215982      -0.040513953
## 2       1            SITTING         0.2612376      -0.001308288
## 3       1           STANDING         0.2789176      -0.016137590
## 4       1            WALKING         0.2773308      -0.017383819
## 5       1 WALKING_DOWNSTAIRS         0.2891883      -0.009918505
## 6       1   WALKING_UPSTAIRS         0.2554617      -0.023953149
##   tBodyAcc-mean()-Z tBodyAcc-std()-X tBodyAcc-std()-Y tBodyAcc-std()-Z
## 1        -0.1132036      -0.92805647     -0.836827406      -0.82606140
## 2        -0.1045442      -0.97722901     -0.922618642      -0.93958629
## 3        -0.1106018      -0.99575990     -0.973190056      -0.97977588
## 4        -0.1111481      -0.28374026      0.114461337      -0.26002790
## 5        -0.1075662       0.03003534     -0.031935943      -0.23043421
## 6        -0.0973020      -0.35470803     -0.002320265      -0.01947924
##   tGravityAcc-mean()-X tGravityAcc-mean()-Y tGravityAcc-mean()-Z
## 1           -0.2488818            0.7055498           0.44581772
## 2            0.8315099            0.2044116           0.33204370
## 3            0.9429520           -0.2729838           0.01349058
## 4            0.9352232           -0.2821650          -0.06810286
## 5            0.9318744           -0.2666103          -0.06211996
## 6            0.8933511           -0.3621534          -0.07540294
##   tGravityAcc-std()-X tGravityAcc-std()-Y tGravityAcc-std()-Z
## 1          -0.8968300          -0.9077200          -0.8523663
## 2          -0.9684571          -0.9355171          -0.9490409
## 3          -0.9937630          -0.9812260          -0.9763241
## 4          -0.9766096          -0.9713060          -0.9477172
## 5          -0.9505598          -0.9370187          -0.8959397
## 6          -0.9563670          -0.9528492          -0.9123794
##   tBodyAccJerk-mean()-X tBodyAccJerk-mean()-Y tBodyAccJerk-mean()-Z
## 1            0.08108653          0.0038382040           0.010834236
## 2            0.07748252         -0.0006191028          -0.003367792
## 3            0.07537665          0.0079757309          -0.003685250
## 4            0.07404163          0.0282721096          -0.004168406
## 5            0.05415532          0.0296504490          -0.010971973
## 6            0.10137273          0.0194863076          -0.045562545
##   tBodyAccJerk-std()-X tBodyAccJerk-std()-Y tBodyAccJerk-std()-Z
## 1          -0.95848211           -0.9241493           -0.9548551
## 2          -0.98643071           -0.9813720           -0.9879108
## 3          -0.99460454           -0.9856487           -0.9922512
## 4          -0.11361560            0.0670025           -0.5026998
```

```
## 5          -0.01228386           -0.1016014           -0.3457350
## 6          -0.44684389           -0.3782744           -0.7065935
##   tBodyGyro-mean()-X tBodyGyro-mean()-Y tBodyGyro-mean()-Z
## 1        -0.01655309        -0.06448612         0.14868944
## 2        -0.04535006        -0.09192415         0.06293138
## 3        -0.02398773        -0.05939722         0.07480075
## 4        -0.04183096        -0.06953005         0.08494482
## 5        -0.03507819        -0.09093713         0.09008501
## 6         0.05054938        -0.16617002         0.05835955
##   tBodyGyro-std()-X tBodyGyro-std()-Y tBodyGyro-std()-Z
## 1       -0.8735439      -0.951090440        -0.9082847
## 2       -0.9772113      -0.966473895        -0.9414259
## 3       -0.9871919      -0.987734440        -0.9806456
## 4       -0.4735355      -0.054607769        -0.3442666
## 5       -0.4580305      -0.126349195        -0.1247025
## 6       -0.5448711       0.004105184        -0.5071687
##   tBodyGyroJerk-mean()-X tBodyGyroJerk-mean()-Y tBodyGyroJerk-mean()-Z
## 1            -0.10727095            -0.04151729            -0.07405012
## 2            -0.09367938            -0.04021181            -0.04670263
## 3            -0.09960921            -0.04406279            -0.04895055
## 4            -0.08999754            -0.03984287            -0.04613093
## 5            -0.07395920            -0.04399028            -0.02704611
## 6            -0.12223277            -0.04214859            -0.04071255
##   tBodyGyroJerk-std()-X tBodyGyroJerk-std()-Y tBodyGyroJerk-std()-Z
## 1            -0.9186085            -0.9679072            -0.9577902
## 2            -0.9917316            -0.9895181            -0.9879358
## 3            -0.9929451            -0.9951379            -0.9921085
## 4            -0.2074219            -0.3044685            -0.4042555
## 5            -0.4870273            -0.2388248            -0.2687615
## 6            -0.6147865            -0.6016967            -0.6063320
##   tBodyAccMag-mean() tBodyAccMag-std() tGravityAccMag-mean()
## 1        -0.84192915       -0.79514486           -0.84192915
## 2        -0.94853679       -0.92707842           -0.94853679
## 3        -0.98427821       -0.98194293           -0.98427821
## 4        -0.13697118       -0.21968865           -0.13697118
## 5         0.02718829        0.01988435            0.02718829
## 6        -0.12992763       -0.32497093           -0.12992763
##   tGravityAccMag-std() tBodyAccJerkMag-mean() tBodyAccJerkMag-std()
## 1          -0.79514486            -0.95439626           -0.92824563
## 2          -0.92707842            -0.98736420           -0.98412002
## 3          -0.98194293            -0.99236779           -0.99309621
## 4          -0.21968865            -0.14142881           -0.07447175
## 5           0.01988435            -0.08944748           -0.02578772
## 6          -0.32497093            -0.46650345           -0.47899162
##   tBodyGyroMag-mean() tBodyGyroMag-std() tBodyGyroJerkMag-mean()
## 1         -0.87475955         -0.8190102              -0.9634610
## 2         -0.93089249         -0.9345318              -0.9919763
## 3         -0.97649379         -0.9786900              -0.9949668
## 4         -0.16097955         -0.1869784              -0.2987037
## 5         -0.07574125         -0.2257244              -0.2954638
## 6         -0.12673559         -0.1486193              -0.5948829
##   tBodyGyroJerkMag-std() fBodyAcc-mean()-X fBodyAcc-mean()-Y
## 1             -0.9358410       -0.93909905      -0.867065205
## 2             -0.9883087       -0.97964124      -0.944084550
```

```
## 3             -0.9947332       -0.99524993      -0.977070848
## 4             -0.3253249       -0.20279431       0.089712726
## 5             -0.3065106        0.03822918       0.001549908
## 6             -0.6485530       -0.40432178      -0.190976721
##   fBodyAcc-mean()-Z fBodyAcc-std()-X fBodyAcc-std()-Y fBodyAcc-std()-Z
## 1        -0.8826669      -0.92443743      -0.83362556      -0.81289156
## 2        -0.9591849      -0.97641231      -0.91727501      -0.93446956
## 3        -0.9852971      -0.99602835      -0.97229310      -0.97793726
## 4        -0.3315601      -0.31913472       0.05604001      -0.27968675
## 5        -0.2255745       0.02433084      -0.11296374      -0.29792789
## 6        -0.4333497      -0.33742819       0.02176951       0.08595655
##   fBodyAccJerk-mean()-X fBodyAccJerk-mean()-Y fBodyAccJerk-mean()-Z
## 1           -0.95707388           -0.92246261            -0.9480609
## 2           -0.98659702           -0.98157947            -0.9860531
## 3           -0.99463080           -0.98541870            -0.9907522
## 4           -0.17054696           -0.03522552            -0.4689992
## 5           -0.02766387           -0.12866716            -0.2883347
## 6           -0.47987525           -0.41344459            -0.6854744
##   fBodyAccJerk-std()-X fBodyAccJerk-std()-Y fBodyAccJerk-std()-Z
## 1           -0.9641607           -0.9322179           -0.9605870
## 2           -0.9874930           -0.9825139           -0.9883392
## 3           -0.9950738           -0.9870182           -0.9923498
## 4           -0.1335866            0.1067399           -0.5347134
## 5           -0.0863279           -0.1345800           -0.4017215
## 6           -0.4619070           -0.3817771           -0.7260402
##   fBodyGyro-mean()-X fBodyGyro-mean()-Y fBodyGyro-mean()-Z
## 1         -0.8502492        -0.95219149        -0.90930272
## 2         -0.9761615        -0.97583859        -0.95131554
## 3         -0.9863868        -0.98898446        -0.98077312
## 4         -0.3390322        -0.10305942        -0.25594094
## 5         -0.3524496        -0.05570225        -0.03186943
## 6         -0.4926117        -0.31947461        -0.45359721
##   fBodyGyro-std()-X fBodyGyro-std()-Y fBodyGyro-std()-Z fBodyAccMag-mean()
## 1        -0.8822965       -0.95123205        -0.9165825        -0.86176765
## 2        -0.9779042       -0.96234504        -0.9439178        -0.94778292
## 3        -0.9874971       -0.98710773        -0.9823453        -0.98535636
## 4        -0.5166919       -0.03350816        -0.4365622        -0.12862345
## 5        -0.4954225       -0.18141473        -0.2384436         0.09658453
## 6        -0.5658925        0.15153891        -0.5717078        -0.35239594
##   fBodyAccMag-std() fBodyBodyAccJerkMag-mean() fBodyBodyAccJerkMag-std()
## 1        -0.7983009                -0.93330036                -0.9218040
## 2        -0.9284448                -0.98526213                -0.9816062
## 3        -0.9823138                -0.99254248                -0.9925360
## 4        -0.3980326                -0.05711940                -0.1034924
## 5        -0.1865303                 0.02621849                -0.1040523
## 6        -0.4162601                -0.44265216                -0.5330599
##   fBodyBodyGyroMag-mean() fBodyBodyGyroMag-std()
## 1              -0.8621902             -0.8243194
## 2              -0.9584356             -0.9321984
## 3              -0.9846176             -0.9784661
## 4              -0.1992526             -0.3210180
## 5              -0.1857203             -0.3983504
## 6              -0.3259615             -0.1829855
##   fBodyBodyGyroJerkMag-mean()
```

```
## 1                  -0.9423669
## 2                  -0.9897975
## 3                  -0.9948154
## 4                  -0.3193086
## 5                  -0.2819634
## 6                  -0.6346651
```

```r
write.table(newdata, "mynewdata.txt", row.name=FALSE)
```