

Getting and Cleaning Data course Project

1. 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"
```

2. 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 ...
```

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)
```

3.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 ...
```

Subsetting 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()"

```

4. 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"

```

5. 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)
```

6.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]))  
write.table(newdata, "mynewdata.txt", row.name=FALSE)
```