project-w3-final.R

Myriam

Thu Mar 19 18:59:19 2015

```
## load package used for project
library(plyr)
library(dplyr)
library(tidyr)
library(stringr)


url= "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip"
dfile = "./W3projectdata/data.zip"
destdir = "./W3projectdata/"
datadir = "./UCI HAR Dataset/"


## download the zipped file
if(!file.exists(destdir)){dir.create(destdir)}
download.file(url, destfile = dfile, mode= "wb", method="curl")


## identify which files to unzip and unzip
filelist = c("UCI HAR Dataset/train/subject_train.txt",
             "UCI HAR Dataset/train/X_train.txt",
             "UCI HAR Dataset/train/y_train.txt",
             "UCI HAR Dataset/train/subject_train.txt",
             "UCI HAR Dataset/test/subject_test.txt",
             "UCI HAR Dataset/test/X_test.txt",
             "UCI HAR Dataset/test/y_test.txt",
             "UCI HAR Dataset/activity_labels.txt",
             "UCI HAR Dataset/features.txt",
             "UCI HAR Dataset/features_info.txt",
             "UCI HAR Dataset/README.txt")
unzip(dfile, files = filelist)


## read the unzipped files
trainX <- read.table("UCI HAR Dataset/train/X_train.txt")
trainY <- read.table("UCI HAR Dataset/train/y_train.txt")
testX <- read.table("UCI HAR Dataset/test/X_test.txt")
```

```
testY <- read.table("UCI HAR Dataset/test/y_test.txt")
testsubject <- read.table("UCI HAR Dataset/test/subject_test.txt")
trainsubject <- read.table("UCI HAR Dataset/train/subject_train.txt")


## create a train file, a test file and a combined file called data
train <- cbind(trainX,trainY,trainsubject)
test <- cbind(testX,testY,testsubject)
data <- rbind(train, test)

## clean up un-needed data frames
rm(trainX,trainY, testX , testY, train, test,trainsubject, testsubject  )

# read the features file
features <- read.table("UCI HAR Dataset/features.txt")

## use the feature file to name data column,
## adding 2 add names for activity and subject columns
names(data) <- c(as.character(features[,2]), "activity" ,"subject")

## cleaning column names
valid_column_names <- make.names(names=names(data), unique=TRUE, allow_ = TRUE)
valid_column_names = str_replace_all(valid_column_names, pattern = "[[:punct:]]", " ")
valid_column_names = str_trim(valid_column_names)
names(data)<- valid_column_names

## clean up un-needed data frames
rm(features)

## create a dataframe that only contains columns with mean but not meanFreq within column name
D1 <- data %>%
    select(contains("mean")) %>%
    select(-contains("meanFreq"))
## create a dataframe that only contains columns with std within column name
D2 <- data %>%
    select(contains("std"))
## create a dataframe that only contains columns activity & subject
D3 <- select(data, activity: subject)
```

```
## create tidy data set
cleandata <- cbind(D3,D1,D2)

## clean up un-needed data frames
rm(D1,D2,D3)

## read the activity labels table
activitylabel <- read.table("UCI HAR Dataset/activity_labels.txt")
## set column names for activity labels table
names(activitylabel) <- c("activity", "activity_label")

## join activitylabel table and clean data by activity column (activity number)
cleandata <- merge( activitylabel, cleandata, by.x="activity", by.y="activity" , sort = FALSE, all.x)

## remove the acticity number column
cleandata = select(cleandata, -activity)


## Group by activity_label and student
bt <- group_by(cleandata, activity_label, subject )


## Create summary table (tidy data set step5)
bt = ddply(bt,.(activity_label, subject),  colwise(mean))

## write table to working directory
write.table(bt, file = "tidysummary.txt", row.names = FALSE)
```