|  |
| --- |
| # Script for Course #3 Project |
|  | # |
|  | # You should create one R script called run\_analysis.R that does the following. |
|  | # |
|  | # 1. Merges the training and the test sets to create one data set. |
|  | # 2. Extracts only the measurements on the mean and standard deviation for each measurement. |
|  | # 3. Uses descriptive activity names to name the activities in the data set |
|  | # 4. Appropriately labels the data set with descriptive variable names. |
|  | # 5. From the data set in step 4, creates a second, independent tidy data set with the average of each |
|  | # variable for each activity and each subject. |
|  |  |
|  | # load the packages |
|  | library(dplyr) |
|  | setwd("~/DataScience") |
|  |  |
|  | # Create data folder |
|  | if(!file.exists("data")) { |
|  | dir.create("data") |
|  | } |
|  |  |
|  |  |
|  | # get the data, load it into data.frames |
|  | fileurl <- "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip" |
|  | download.file(fileurl,destfile = "./data/rundata.zip") |
|  | unzip("./data/rundata.zip") # unzips to the home dir |
|  | # C:\Users\lorit\Documents\DataScience\UCI HAR Dataset\test |
|  | # C:\Users\lorit\Documents\DataScience\UCI HAR Dataset\train |
|  | fileurl <- "http://archive.ics.uci.edu/ml/machine-learning-databases/00240/UCI%20HAR%20Dataset.names" |
|  | download.file(fileurl,destfile = "./data/rundata.names") |
|  |  |
|  | col\_names <- read.table("./UCI HAR Dataset/features.txt") # the variable names |
|  |  |
|  | X\_train <- read.table("./UCI HAR Dataset/train/X\_train.txt", col.names = col\_names[,2]) # the values from the activity |
|  | y\_train <- read.table("./UCI HAR Dataset/train/y\_train.txt", col.names = as.factor(c("activity"))) # the activity |
|  | subject\_train <- read.table("./UCI HAR Dataset/train/subject\_train.txt", col.names = as.factor(c("subject"))) # the subject |
|  |  |
|  | #Activites: WALKING, WALKING\_UPSTAIRS, WALKING\_DOWNSTAIRS, SITTING, STANDING, LAYING |
|  | X\_test <- read.table("./UCI HAR Dataset/test/X\_test.txt", col.names = col\_names[,2]) # the values from the activity |
|  | y\_test <- read.table("./UCI HAR Dataset/test/y\_test.txt", col.names = c("activity")) |
|  | subject\_test <- read.table("./UCI HAR Dataset/test/subject\_test.txt", col.names = c("subject")) # the subject |
|  |  |
|  | # --------------------------------------------- |
|  | # 1. Merges the training and the test sets to create one data set. |
|  | test <- cbind(X\_test,y\_test,subject\_test) |
|  | train <- cbind(X\_train,y\_train,subject\_train) |
|  | all <- rbind(test,train) |
|  |  |
|  | # --------------------------------------------- |
|  | # 2. Extracts only the measurements on the mean and standard deviation for each measurement. |
|  | mean\_std <- select(all, contains("mean"),contains("std"),activity,subject) |
|  |  |
|  | # --------------------------------------------- |
|  | # 3. Uses descriptive activity names to name the activities in the data set |
|  | xlate\_act <- function(x) |
|  | { |
|  | activities <- read.table("./UCI HAR Dataset/activity\_labels.txt") |
|  | activities[x,2] |
|  | } |
|  | mean\_std <- mutate(mean\_std, activity = xlate\_act(mean\_std$activity)) |
|  |  |
|  | # --------------------------------------------- |
|  | # 4. Appropriately labels the data set with descriptive variable names. |
|  |  |
|  | names(mean\_std) <- gsub("\\.","",tolower(names(mean\_std))) |
|  |  |
|  |  |
|  | # --------------------------------------------- |
|  | # 5. From the data set in step 4, creates a second, independent tidy data set with the average of each |
|  | # variable for each activity and each subject. |
|  | mean\_std\_grouped <- group\_by(mean\_std,activity,subject) |
|  | tidy\_data <- summarise\_all(mean\_std\_grouped,mean) |
|  | View(tidy\_data) |
|  | write.table(tidy\_data,"./data/tidy\_data.txt",quote = FALSE, row.names = FALSE) |
|  |  |
|  |  |
|  | # To view the tidy\_data set provided you can run the follow commands |
|  | # data <- read.table("./data/tidy\_data.txt", header = TRUE) |
|  | # View(data) |