run\_analysis.R

Gahlot

Sun Nov 22 13:40:09 2015

#download data from http://archive.ics.uci.edu/ml/datasets/Human+Activity+Recognition+Using+Smartphones  
##Load Library httr  
library(httr)   
url <- "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip"  
##save File as HAR Dataset  
file <- "instancia.zip"  
## Download file  
##download.file(url, file)  
  
  
#unzip into UCI HAR Dataset folder check if it exists.  
datafolder <- "UCI HAR Dataset"  
resultsfolder <- "results"  
if(!file.exists(datafolder)){  
 print("unzip file")  
 unzip(file, list = FALSE, overwrite = TRUE)  
}   
  
## create folder if it does not exist  
if(!file.exists(resultsfolder)){  
 print("create results folder")  
 dir.create(resultsfolder)  
}   
  
##Read txt and covnert to data.frame  
gettables <- function (filename,cols = NULL){  
 print(paste("Getting table:", filename))  
 f <- paste(datafolder,filename,sep="/")  
 data <- data.frame()  
 if(is.null(cols)){  
 data <- read.table(f,sep="",stringsAsFactors=F)  
 } else {  
 data <- read.table(f,sep="",stringsAsFactors=F, col.names= cols)  
 }  
 data  
}  
  
#TEst gettables for Features File  
Device\_features <- gettables("features.txt")

## [1] "Getting table: features.txt"

#read data and build database  
  
getdata <- function(type, Device\_features){  
 print(paste("Getting data", type))  
 subject\_data <- gettables(paste(type,"/","subject\_",type,".txt",sep=""),"id")  
 y\_data <- gettables(paste(type,"/","y\_",type,".txt",sep=""),"activity")  
 x\_data <- gettables(paste(type,"/","X\_",type,".txt",sep=""),Device\_features$V2)  
 return (cbind(subject\_data,y\_data,x\_data))  
}  
  
#run and check getdata  
test <- getdata("test", Device\_features)

## [1] "Getting data test"  
## [1] "Getting table: test/subject\_test.txt"  
## [1] "Getting table: test/y\_test.txt"  
## [1] "Getting table: test/X\_test.txt"

train <- getdata("train", Device\_features)

## [1] "Getting data train"  
## [1] "Getting table: train/subject\_train.txt"  
## [1] "Getting table: train/y\_train.txt"  
## [1] "Getting table: train/X\_train.txt"

#save the resulting data in the indicated folder  
saveresults <- function (data,name){  
 print(paste("saving results", name))  
 file <- paste(resultsfolder, "/", name,".csv" ,sep="")  
 write.csv(data,file)  
}  
  
### required activities with Help from internet ###  
  
#1) Merges the training and the test sets to create one data set.  
##install.packages("plyr")  
library(plyr)  
##Merge test and train DB  
data <- rbind(train, test)  
data <- arrange(data, id)  
  
#2) Extracts only the measurements on the mean and standard deviation for each measurement.   
mean\_and\_std <- data[,c(1,2,grep("std", colnames(data)), grep("mean", colnames(data)))]  
saveresults(mean\_and\_std,"mean\_and\_std")

## [1] "saving results mean\_and\_std"

#3) Uses descriptive activity names to name the activities in the data set  
activity\_labels <- gettables("activity\_labels.txt")

## [1] "Getting table: activity\_labels.txt"

#4) Appropriately labels the data set with descriptive variable names.   
data$activity <- factor(data$activity, levels=activity\_labels$V1, labels=activity\_labels$V2)  
  
#5) Creates a second, independent tidy data set with the average of each variable for each activity and each subject.   
tidy\_dataset <- ddply(mean\_and\_std, .(id, activity), .fun=function(x){ colMeans(x[,-c(1:2)]) })  
colnames(tidy\_dataset)[-c(1:2)] <- paste(colnames(tidy\_dataset)[-c(1:2)], "\_mean", sep="")  
saveresults(tidy\_dataset,"tidy\_dataset")

## [1] "saving results tidy\_dataset"

## Finesed with help from Internet  
  
##Create Documentation  
##install.packages("rmarkdown")  
##install.packages("rmarkdown")  
##download and install pandoc from https://github.com/jgm/pandoc/releases/tag/1.15.2  
## Knit created PDF and html