run\_analysis.R

cburgess44

Fri Aug 21 09:55:10 2015

## cleanup workspace  
rm(list = ls())  
  
## load plyr  
library(plyr)  
  
##Set WD as data--files have been unzipped--UCI HAR Dataset folder  
setwd("~/data/")  
##Set file path   
pathdata <-file.path("./data" , "UCI HAR Dataset")  
  
## Read the data; Y data is the actvity data; x is feature data   
## Test data  
ytestdata <- read.table(file.path(pathdata, "test" , "Y\_test.txt" ),header = FALSE)  
xtestdata <- read.table(file.path(pathdata, "test" , "X\_test.txt" ),header = FALSE)  
subtestdata <- read.table(file.path(pathdata, "test" , "subject\_test.txt"),header = FALSE)  
## Train data  
ytraindata <- read.table(file.path(pathdata, "train", "Y\_train.txt"),header = FALSE)  
xtraindata <- read.table(file.path(pathdata, "train", "X\_train.txt"),header = FALSE)  
subtraindata <- read.table(file.path(pathdata, "train", "subject\_train.txt"),header = FALSE)  
## the column varibale for the Feature Data  
featuresHead <- read.table(file.path(pathdata, "features.txt"), header=FALSE)   
## the Activities labels  
Activities <- read.table(file.path(pathdata, "activity\_labels.txt"), header=FALSE)  
  
## vertically group and name the data into three; activity data, subject data, feature data  
  
activitydata <- rbind(ytraindata,ytestdata)  
names(activitydata) <- c("Activity")  
  
subjectdata <- rbind(subtraindata,subtestdata)  
names(subjectdata) <- c("Subject")  
## 561 variables in the feature data (x) set will need to name the varibales based on the Features.txt data  
featuredata <- rbind(xtraindata,xtestdata)  
names(featuredata) <- featuresHead$V2  
  
## Combine all the data into one data set (#1)  
  
group1 <- cbind(subjectdata,activitydata)  
All\_data <- cbind(featuredata, group1)  
  
## limit the data to data with measurelements on the Mean and STD (#2)  
  
DataMeanStd <- All\_data[ ,grepl("mean|std|Subject|Activity", names(All\_data))]  
  
## descriptive activity names to name the activities in the data set (#3)--I don't think this is right  
  
  
ActivityNames <- Activities$V2  
  
## labels the data set with descriptive variable names (#4)  
  
names(DataMeanStd) <- gsub('\\(|\\)',"",names(DataMeanStd), perl = TRUE)  
names(DataMeanStd) <- make.names(names(DataMeanStd))  
  
names(DataMeanStd) <- gsub('Acc',"Accelerometer",names(DataMeanStd))  
names(DataMeanStd) <- gsub('GyroJerk',"Angle Acceleration",names(DataMeanStd))  
names(DataMeanStd) <- gsub('Gyro',"Gyroscope",names(DataMeanStd))  
names(DataMeanStd) <- gsub('Mag',"Magnitude",names(DataMeanStd))  
names(DataMeanStd) <- gsub('^t',"Time",names(DataMeanStd))  
names(DataMeanStd) <- gsub('^f',"FrequencyDomain.",names(DataMeanStd))  
names(DataMeanStd) <- gsub('\\.mean',".Mean",names(DataMeanStd))  
names(DataMeanStd) <- gsub('\\.std',".StandardDeviation",names(DataMeanStd))  
names(DataMeanStd) <- gsub('Freq\\.',"Frequency.",names(DataMeanStd))  
names(DataMeanStd) <- gsub('Freq$',"Frequency",names(DataMeanStd))  
names(DataMeanStd) <- gsub('BodyBody',"Body",names(DataMeanStd))  
  
## create a second, independent tidy data set with the Average of each variable   
  
Tidy <- ddply(DataMeanStd, c("Subject","Activity"), numcolwise(mean))  
write.table(Tidy, file = "Tidy.txt", row.names = FALSE)