run_analysis.R

Justin Cassidy

Fri Oct 23 15:45:42 2015

```r
# Scripts to create a tidy data set - COURSERA PROJECT



# part 1 Merges the training and the test sets
# to create one data set.

setwd("C:/Users/Justin Cassidy/Desktop/COURSERA_TidyData_class/UCI HAR Dataset")
features <- read.table("features.txt")

setwd("C:/Users/Justin Cassidy/Desktop/COURSERA_TidyData_class/UCI HAR Dataset/test")
x_test <- read.table("X_test.txt")
y_test <- read.table("y_test.txt")
subject_test <- read.table("subject_test.txt")

setwd("C:/Users/Justin Cassidy/Desktop/COURSERA_TidyData_class/UCI HAR Dataset/train")
x_train <- read.table("X_train.txt")
y_train <- read.table("y_train.txt")
subject_train <- read.table("subject_train.txt")

#add labels to 'x' datasets and merge them by adding rows
colnames(x_train) <- features$V2
colnames(x_test) <- features$V2
total <- rbind(x_train,x_test)

#merge labels column with each other and x datasets
ybound <- rbind(y_train,y_test)
total$labels <- ybound

#merge subjects column with each other and x datasets
subjectbound <- rbind(subject_train,subject_test)
total$subject <- subjectbound



# part 2 Extracts only the measurements on the mean
# and standard deviation for each measurement.

# use grepl to search column names for mean and std
meancols <- total[ , grepl( "mean" , names( total ) ) ]
stdcols <- total[ , grepl( "std" , names( total ) ) ]
newtotal <- cbind(meancols,stdcols)



# part 3  Uses descriptive activity names to name
# the activities in the data set
```

```r
# first converts y_test activity labels to data frame, then factors to descriptive names
y_testDF <- as.data.frame(y_test)
y_testDF$V2<- factor(y_testDF$V1,levels = c(1,2,3,4,5,6),labels = c("WALKING", "WALKING_UPSTAIRS", "WALK

# first converts y_train activity labels to data frame, then factors to descriptive names
y_trainDF <- as.data.frame(y_train)
y_trainDF$V2<- factor(y_trainDF$V1,levels = c(1,2,3,4,5,6),labels = c("WALKING", "WALKING_UPSTAIRS", "WA

# adds these new sets to each other then to the previously merged set (all mean and std data).
ybound2 <- rbind(y_trainDF,y_testDF)
newtotal$activity <- ybound2$V2


# part 4  Appropriately labels the data set
# with descriptive variable names.

# NOTE that this part was done already when the data was input; see part 1
# See below Part 5 for some commands that cleaned up the names a bit


# part 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.
library(reshape2)
newtotalMELT <- melt(newtotal,id="activity")
donecast <- dcast(newtotalMELT,activity~variable,mean)


# (Part 4 again) Clean up column names to make a bit more 'tidy'
names(donecast) <- gsub("Mean()-", "Mean", names(donecast))
names(donecast) <- gsub("tBody", "Body", names(donecast))
names(donecast) <- gsub("tGravity", "Gravity", names(donecast))
names(donecast) <- gsub("fBody", "Body", names(donecast))
names(donecast) <- gsub("Acc", "Acceleration", names(donecast))
names(donecast) <- gsub("Freq", "Frequency", names(donecast))
names(donecast) <- gsub("std", "StandardDeviation", names(donecast))
names(donecast) <- gsub("Mag", "Magnitude", names(donecast))


#  Output for submission
write.table(donecast, "TidyData.txt", row.name=FALSE)
```