README

YZ

September 20, 2019

## Getting and Cleaning Data Course Project

set wd

setwd("C:/Users/Public/Dropbox (Personal)/ETELSE/coursera/data science specialization/ds")

downloading and unzipping the dataset

if(!file.exists(".")){dir.create("./data")}  
fileUrl <- "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip"  
download.file(fileUrl,destfile="./Dataset.zip")

Unzip dataSet to /data directory

unzip(zipfile="./Dataset.zip",exdir="./data")

Reading trainings tables

x\_train <- read.table("./data/UCI HAR Dataset/train/X\_train.txt")  
y\_train <- read.table("./data/UCI HAR Dataset/train/y\_train.txt")  
subject\_train <- read.table("./data/UCI HAR Dataset/train/subject\_train.txt")

Reading testing tables

x\_test <- read.table("./data/UCI HAR Dataset/test/X\_test.txt")  
y\_test <- read.table("./data/UCI HAR Dataset/test/y\_test.txt")  
subject\_test <- read.table("./data/UCI HAR Dataset/test/subject\_test.txt")

Reading feature vector

features <- read.table('./data/UCI HAR Dataset/features.txt')

Reading activity labels

activityLabels = read.table('./data/UCI HAR Dataset/activity\_labels.txt')

assigning column names

colnames(x\_train) <- features[,2]   
colnames(y\_train) <-"activityId"  
colnames(subject\_train) <- "subjectId"  
  
colnames(x\_test) <- features[,2]   
colnames(y\_test) <- "activityId"  
colnames(subject\_test) <- "subjectId"  
  
colnames(activityLabels) <- c('activityId','activityType')

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

mrg\_train <- cbind(y\_train, subject\_train, x\_train)  
mrg\_test <- cbind(y\_test, subject\_test, x\_test)  
setAllInOne <- rbind(mrg\_train, mrg\_test)

Reading column names

colNames <- colnames(setAllInOne)

Create vector for defining ID, mean and standard deviation

mean\_and\_std <- (grepl("activityId" , colNames) |   
 grepl("subjectId" , colNames) |   
 grepl("mean.." , colNames) |   
 grepl("std.." , colNames)   
)

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

setForMeanAndStd <- setAllInOne[ , mean\_and\_std == TRUE]

### 3. Using descriptive activity names to name the activities in the data set

setWithActivityNames <- merge(setForMeanAndStd, activityLabels,  
 by='activityId',  
 all.x=TRUE)

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

This is done as part of the data prep for q.3

### 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.

secTidySet <- aggregate(. ~subjectId + activityId, setWithActivityNames, mean)  
secTidySet <- secTidySet[order(secTidySet$subjectId, secTidySet$activityId),]  
write.table(secTidySet, "secTidySet.txt", row.name=FALSE)