Getting and Cleaning Data Project

Harshad B.

07/09/2020

Introduction

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 Packages and get the Data

```
packages <- c("data.table", "reshape2")
sapply(packages, require, character.only=TRUE, quietly=TRUE)
```

```
##
## Attaching package: 'reshape2'

## The following objects are masked from 'package:data.table':
##
##     dcast, melt

## data.table   reshape2
##       TRUE       TRUE
```

```
path <- getwd()
url <- "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip"
download.file(url, file.path(path, "dataFiles.zip"))
unzip(zipfile = "dataFiles.zip")
```

Load activity labels + features

```
activityLabels <- fread(file.path(path, "UCI HAR Dataset/activity_labels.txt")
                        , col.names = c("classLabels", "activityName"))
features <- fread(file.path(path, "UCI HAR Dataset/features.txt")
                  , col.names = c("index", "featureNames"))
featuresWanted <- grep("(mean|std)\\(\\)", features[, featureNames])
measurements <- features[featuresWanted, featureNames]
measurements <- gsub('[()]', '', measurements)
```

Load train datasets

```
train <- fread(file.path(path, "UCI HAR Dataset/train/X_train.txt"))[, featuresWanted, with = FALSE]
data.table::setnames(train, colnames(train), measurements)
trainActivities <- fread(file.path(path, "UCI HAR Dataset/train/Y_train.txt")
                         , col.names = c("Activity"))
trainSubjects <- fread(file.path(path, "UCI HAR Dataset/train/subject_train.txt")
                         , col.names = c("SubjectNum"))
train <- cbind(trainSubjects, trainActivities, train)
```

Load test datasets

```
test <- fread(file.path(path, "UCI HAR Dataset/test/X_test.txt"))[, featuresWanted, with = FALSE]
data.table::setnames(test, colnames(test), measurements)
testActivities <- fread(file.path(path, "UCI HAR Dataset/test/Y_test.txt")
                          , col.names = c("Activity"))
testSubjects <- fread(file.path(path, "UCI HAR Dataset/test/subject_test.txt")
                        , col.names = c("SubjectNum"))
test <- cbind(testSubjects, testActivities, test)
```

merge datasets

```
combined <- rbind(train, test)
```

Convert classLabels to activityName basically. More explicit.

```
combined[["Activity"]] <- factor(combined[, Activity]
                                 , levels = activityLabels[["classLabels"]]
                                 , labels = activityLabels[["activityName"]])

combined[["SubjectNum"]] <- as.factor(combined[, SubjectNum])
combined <- reshape2::melt(data = combined, id = c("SubjectNum", "Activity"))
combined <- reshape2::dcast(data = combined, SubjectNum + Activity ~ variable, fun.aggregate = mean)

data.table::fwrite(x = combined, file = "tidyData.txt", quote = FALSE)
```