CodeBook Final Proyect

Edgar Manuel Martinez Santana

24/8/2021

Getting and cleaning data Course Proyect

The data source was the set provided by the instructor, which can be downloaded at the following link: https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip.

Steps to get the requested result

1. The libraries "data.table" and "dplyr" were loaded

```
library(data.table)
library(dplyr)
```

2. the data set was loaded in the variable "urlFile", and it was unzipped in the root folder

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

3. the activity and features files were read, the variables, "activityLab", "features" and "featuresfin" were created

```
activityLab <- read.table("./UCI HAR Dataset/activity_labels.txt", header = FALSE)
names(activityLab) <- c("ActivityID","ActivityName")

features <- read.table("./UCI HAR Dataset/features.txt", header = FALSE)
names(features) <- c("featuresID","featuresName")

featuresfilter <- grep("(mean|std)\\(\\)",features[,"featuresName"])
featuresfin <- features[featuresfilter, "featuresName"]
featuresfin <- gsub("[()]","",featuresfin)
```

4. The training data was read and the variables, "tsubjet", "tX", "ty", "dataTrain" were created.

```
tsubjet <- read.table("./UCI HAR Dataset/train/subject_train.txt", header = FALSE)
names(tsubjet) <- c("subjetID")

tX <- read.table("./UCI HAR Dataset/train/X_train.txt", header = FALSE)
cleaname <- gsub("[()]","",features$featuresName)
names(tX) <- c(cleaname)
tX <- tX[,featuresfin]

ty <- read.table("./UCI HAR Dataset/train/y_train.txt", header = FALSE)
names(ty) <- c("activity")

dataTrain <- cbind(tsubjet,ty,tX)
```

5.The test data was read and the variables, "tstsubjet", "tstX", "tsty", "dataTest" were created.

```
tstsubjet <- read.table("./UCI HAR Dataset/test/subject_test.txt", header = FALSE)
names(tstsubjet) <- c("subjetID")

tstX <- read.table("./UCI HAR Dataset/test/X_test.txt", header = FALSE)
cleaname <- gsub("[()]","",features$featuresName)
names(tstX) <- c(cleaname)
tstX <- tstX[,featuresfin]

tsty <- read.table("./UCI HAR Dataset/test/y_test.txt", header = FALSE)
names(tsty) <- c("activity")

dataTest <- cbind(tstsubjet,tsty,tstX)
```

6. the dataset "data Train" and "data Test" were combined to create the complete file "dataFinal"

```
dataFinal <- rbind(dataTrain, dataTest)
```

7. Finally, the grouping variables are converted into a factor, the file "tidyData" is created with the means of all the columns grouped by "activity" and "subjectID" and saved with that name.

```
dataFinal$activity <- factor(dataFinal$activity
                             , levels = activityLab$ActivityID
                             , labels = activityLab$ActivityName)


data1 <- data.table(dataFinal)
tidyDat  %>% group_by(activity, subjetID) %>% summarize_all(mean)


data.table::fwrite(x = tidyDat, file = "tidyData.csv", quote = FALSE)
rm(list = ls())
```