CodeBook

Sergio

29/6/2020

Getting and Cleaning Data Project

Description

This project is part of the Getting and Cleaning Data course from Johns Hopkins University on Coursera.org.

The purpose of this project is to demonstrate your ability to collect, work with, and clean a data set. The goal is to prepare tidy data that can be used for later analysis.

1. A tidy data set as described below,
2. A link to a Github repository with your script for performing the analysis,
3. A code book that describes the variables, the data, and any transformations or work that you performed to clean up the data called CodeBook.md. You should also include a README.md in the repo with your scripts. This repo explains how all of the scripts work and how they are connected.

Source Data

One of the most exciting areas in all of data science right now is wearable computing - see for example this article . Companies like Fitbit, Nike, and Jawbone Up are racing to develop the most advanced algorithms to attract new users. The data linked to from the course website represent data collected from the accelerometers from the Samsung Galaxy S smartphone. A full description is available at the site where the data was obtained:

Description can be found here UCI Machine Learning Repository

Here are the data for the project: Data Set

Loading required packages

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

Downloaded the dataset

```
filename <- "Getting_Cleaning_Dataset.zip"
# Checking if archieve already exists.
if (!file.exists(filename)){
  fileURL <- "https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip"
  download.file(fileURL, filename, method="curl")
```

```
}

# Checking if folder exists
if (!file.exists("UCI HAR Dataset")) {
  unzip(filename)
}
```

Read .txt files into data frames

```
#features and activities labels
features <- data.table::fread("UCI HAR Dataset/features.txt", col.names = c("n","functions"))
activities <- data.table::fread("UCI HAR Dataset/activity_labels.txt", col.names = c("code", "activity"))
```

An alternative is: features <- data.table::fread("UCI HAR Dataset/features.txt", col.names = c("n","functions"))

features and activities labels

```
features <- data.table::fread("UCI HAR Dataset/features.txt", col.names = c("n","functions"))
activities <- data.table::fread("UCI HAR Dataset/activity_labels.txt", col.names = c("code",
"activity"))
```

#Subjects group of 30 volunteers subject_test <- data.table::fread("UCI HAR Dataset/test/subject_test.txt", col.names = "subject") subject_train <- data.table::fread("UCI HAR Dataset/train/subject_train.txt", col.names = "subject")

#Activity (1 to 6) (WALKING, WALKING_UPSTAIRS, WALKING_DOWNSTAIRS, SITTING, STANDING, LAYING) y_test <- data.table::fread("UCI HAR Dataset/test/y_test.txt", col.names = "code") y_train <- data.table::fread("UCI HAR Dataset/train/y_train.txt", col.names = "code")

#Data where 70% of the volunteers was selected for generating the training data and 30% the test data x_train <- data.table::fread("UCI HAR Dataset/train/X_train.txt", col.names = $features functions)x_test < -data.table :: fread("UCIHARDataset/test/X_test.txt", col.names = features$functions)

Including Plots

You can also embed plots, for example:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.