

Peer-graded Assignment: Getting and Cleaning Data Course Project

Itamara

4/26/2021

The `run_analysis.R` script performs the data preparation and then performance the 5 steps required for the assignment.

In order to make more easy and understandable to code is divide in 8 steps as described below:

1. Call necessary libraries and download the data

```
library(dplyr)
```

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

Download the dataset using the URL info and save in the current directory.

Unzip the dataset

2. Read datafiles

In this case the unzip directory has 4 files and 2 more directories:

- '*README.txt*'

- '*features_info.txt*': Shows information about the variables used on the feature vector.

- '*features.txt*': List of all features.

- '*activity_labels.txt*': Links the class labels with their activity name.

The directories *train* and *test* has more files inside. Their descriptions are equivalent. For more info about the files please read *'*README.txt*'.

Creating temporary variables inside on R Environment.

3. Merging datasets from train and test

The train and the test directories has the same file name with difference observation and equivalent descriptions. Those files are merged

- *test/X_test.txt* & *train/X_train.txt*

- *test/y_test.txt* & *train/Y_train.txt*

- *test/subject_test.txt* & *train/subject_train.txt*

Remove temporary variables to save memory.

4. Performance the calculation of Standard Deviation and Mean for the measurement

Before performance the calculation it is necessary filter the data, selecting the column with the variable the it is desired to keep.

5. Describe Activities that were select in the step 4

Replace activity values with named factor levels

6. Labeling the data

Cleaning, remove typos and proper named the column data to avoid misunderstanding, which means remove the abbreviation and replace by the entire name of the attributes, if is necessary.

7. Creating new data

Group by subject and activity and summarize using mean.

8. Save the final data

Save as external file call: tidy_data.txt.

When is run in the Console the output should be something like:

```
[1] "STEP 1: GET THE DATA"
trying URL 'https://d396qusza40orc.cloudfront.net/getdata%2Fprojectfiles%2FUCI%20HAR%20Dataset.zip'
Content type 'application/zip' length 62556944 bytes (59.7 MB)
downloaded 59.7 MB

[1] "STEP 2: READ DATAFILES FROM DATA"
[1] "STEP 3: MERGING DATA"
[1] "STEP 4: CALCULATION ..."
[1] "STEP 5: DESCRIPTIVE ACTIVITY"
[1] "STEP 6: LABEL THE DATA"
[1] "STEP 7: CREATE NEW DATA "
[1] "STEP 8: SAVE OUTPUT --> tidy_data.txt "
```

If any step is missing the code failed in this specif stage.