

Problem Set #1

Marwin Carmo

2023-10-05

Overview

Welcome to Data Cleaning and Management using R! This problem set is intended to give you some practice becoming familiar with using R. In this problem set, I'm asking you to: create an R project; render to pdf; load and investigate an R data frame that is stored on the web; and apply some basic functions to atomic vectors.

- Note: Change the values of the YAML header above to your name and the date.

Question 1: Creating an R project

Create an R project

- Create a folder where you want to save files associated with problem set 1. Let's call that folder "problemset1", but you can name it whatever you want.
 - For instance, it could be psc290-fq23 » problem-sets » problemset1.
- In RStudio, click on "File" » "New Project" » "Existing Directory" » "Browse".
- Browse to find and select your problem set 1 folder.
- Click on "Create Project".
 - An R project file has the extension ".Rproj".
 - The name of the file should be "problemset1.Rproj", or whatever you named the folder.

Save this problemset1.Rmd file anywhere in the folder named problemset1.

- Use this naming convention "lastname-firstname-ps#" for your .qmd files (e.g. beck_emorye_ps1.qmd).
 - If you want, you can change the name of this file to include your first and last name.
- Run the `getwd()` function and the `list.files()` function in the code chunk below.

- What is the output? Why?

```
getwd()
list.files()
```

ANSWER:

```
[1] "C:/Users/marwin/Documents/Rprojects/phd/23-Fall/psc290-fq23/problem-sets/problemset1"
[1] "carmo_marwin_ps1.qmd" "problemset1.Rproj"
```

The first line of the output shows the current working directory of the R process. It was given by the function `getwd()`. The second output line says which files are in the current working directory, and was produced by the function `list.files()`. At the moment there are only two, this `.qmd` file and a shortcut to the problem set 1 project.

Question 2: Knit to pdf

- At the top of this `.qmd` file, type in your first and last name in the appropriate place in the YAML header (e.g. “Hadley Wickham”).
- In the date field of the YAML header, insert the date within quotations (any date format is fine).
- Now click the “Render” button near the top of your RStudio window (icon with blue yarn ball).
 - Alternatively you can use the shortcut: **Cmd/Ctrl + Shift + k**.
 - *Note:* One goal of this assignment is to make sure you are able to render to a PDF without running into errors.

Question 3: Load .Rdata directly with url and then investigate the data frame

1. This question asks you to load a dataframe by specifying the `url()` function within the `load()` function.
 - Url link for data frame: <https://github.com/emoriebeck/psc290-data-FQ23/raw/main/05-assignments/01-ps1/pwe-ps1-small.RData>
 - Hint: to load `.Rdata` use the `load()` and `url()` functions because you are using a link. follow this approach: `load(url("url_link"))`.
 - * Note: the `url_link` is put within quotes

Load the dataframe within this code chunk below.

```
#?load
```

2. Print the data frame `df_pwe` by typing its name.
3. Use the `typeof()` function to investigate the type of data frame `df_pwe`.
4. Apply the `length()` function to the data frame `df_pwe`. What does this output mean in your own words?

ANSWER:

5. Use the `str()` function to investigate the structure of the data frame `df_pwe`.
6. Use the `names` function to list the names of the elements (variables) within `df_pwe`.
7. Wrap your answer above — `names(data_frame_name)` — within the `typeof()` function. Do the same for the `length()` function, and the `str()` function as well. Interpret what the output means in your own words.

ANSWER:

Question 4: Applying basic functions to atomic vectors

1. Create an atomic vector object named `age` with the following values: 3, 6, 41, 43.
2. Apply the `typeof()`, `length()`, and `str()` functions to the object `age`.
3. Apply the `sum()` function to `age`.
4. Apply the `sum()` function to `age` but this time include the argument `na.rm = FALSE`.
5. In general, what is a function “argument name” and what is an “argument value”? What does the argument `na.rm` do?

ANSWER:

6. Create a new object `age2` with the following values: 3, 6, 41, 43, NA. Now calculate the sum of `age2` using the argument `na.rm = FALSE` and then calculate the sum using the argument `na.rm = TRUE`. Explain why the outputs of these two `sum()` functions differ.

ANSWER:

7. Create a vector `tf` using the following code: `tf <- c(TRUE,FALSE,TRUE,FALSE,TRUE)`. Next apply the `typeof()`, `length()`, and `str()` functions to the object `tf`. Based on this output, briefly describe the object `tf` in your own words (one sentence is fine).

ANSWER:

8. Apply the `sum()` function to the object, using the option to remove `NA` values prior to calculation. What numeric value do mathematical calculations in R assign to `TRUE` values and what do they assign to `FALSE` values?

ANSWER:

9. This is the syntax of the `mean()` function that includes both argument names and the default values for arguments: `mean(x, trim = 0, na.rm = FALSE)`.

When using a function, R requires you to type the values you assign to each argument, but typing in the argument names is usually optional. Even though it takes a bit more time, I usually like typing in both argument names and argument values, because it forces me to be more conscious about what value I am assigning to which argument, especially when a function is new to me.

Use the `mean()` function to calculate the mean of object `tf` (removing `NA` values prior to calculation). In your function call, include both the argument name and the argument value for each argument (argument value for the `trim` argument can be 0). Then run the same function, but without typing any argument names.

Render to pdf and submit problem set

Render to pdf by clicking the “Render” button near the top of your RStudio window (icon with blue yarn ball) or drop down and select “Knit to PDF”.

- Go to the Canvas and under the “Assignments”, submit to the Problem Set 1 Assignment.
- Submit both `.qmd` and `.pdf` files.
- Use this naming convention “lastname_firstname_ps#” for your `.qmd` and `pdf` files (e.g. `beck-emorie-ps1.qmd` & `beck-emorie-ps1.pdf`).