

Problem Set #1

INSERT YOUR NAME HERE

Invalid Date

Grade /20

Overview

Welcome to Introduction to Programming and Data Management using R! This problem set is intended to give you some practice becoming familiar with using R. In this problem set, we are asking you to: create an R project; knit a .QMD file to pdf [IF WE WANT THEM TO SUBMIT R SCRIPT, DO THEY NEED TO KNIT?]; load and investigate an R data frame that is stored on the web [GIVE THEM FILE?]; apply some basic functions to atomic vectors; and create a GitHub issue.

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

Question 1: Creating an R project & R script

/2 pts

Create an R project

- Create a folder where you want to save files associated with problem set 1. Make sure to name the folder “problemset1”.
 - For instance, the file directory/structure could be EDUC260a » 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”.

Create an R script

- In RStudio, click on “File” » “New File” » “R Script” - Make sure to save your R script in the folder you just created and name it “problemset1.R”

- In your R script, run the `getwd()` function and the `list.files()` function.
- What is the output? Why? Make sure to comment out your text using a `#`

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

Answer: The output are files: “problemset1.R”, “problemset1.qmd”, and “problemset1.Rproj” because these are the files currently saved in the folder where this .R file is saved.

Question 2: Knit to pdf

/1 pts

- Open the “problemset.qmd” file in your folder. At the top of this .Rmd file, type in your first and last name in the appropriate place in the YAML header (e.g. “Joe Bruin”).
- 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 (blue arrow icon) or drop down and select “Render Document”.
 - Alternatively you can use the shortcut: **Cmd/Ctrl + Shift + k**.
 - *Note:* One goal of this assignment is to make sure you are able to knit to a PDF without running into errors.

Question 3: Load .Rdata directly with url and then investigate the data frame [GIVE THEM R DATE OR CSV FILE?]

/7 pts

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/anyone-can-cook/rclass1/raw/master/data/recruiting/recruit>
 - 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
load(url("https://github.com/anyone-can-cook/rclass1/raw/master/data/recruiting/recruit_ps
```

2. Print the data frame `df_recruiting` by typing its name.

```
df_recruiting
```

	state_code	school_type	ncessch	name
1	AK	public	020000100208	Bethel Regional High School
2	AK	public	020000100211	Ayagina'ar Elitnaurvik
3	AK	public	020000100212	Kwigillingok School
4	AK	public	020000100213	Nelson Island Area School
5	AK	public	020000300216	Alakanuk School
6	AK	public	020000300217	Emmonak School

	address	city
1	1006 Ron Edwards Memorial Dr	Bethel
2	106 Village Road	Kongiganak
3	108 Village Road	Kwigillingok
4	118 Village Road	Toksook Bay
5	9 School Road	Alakanuk
6	General Delivery	Emmonak

3. Use the `typeof()` function to investigate the type of data frame `df_recruiting`.

```
typeof(df_recruiting)
```

```
[1] "list"
```

4. Apply the `length()` function to the data frame `df_recruiting`. What does this output mean in your own words?

```
length(df_recruiting)
```

```
[1] 6
```

Answer: the data frame object `df_recruiting` has a length of 6, meaning that there are 6 elements in the object.

5. Use the `str()` function to investigate the structure of the data frame `df_recruiting`.

```
str(df_recruiting)
```

Classes 'tbl_df', 'tbl' and 'data.frame': 6 obs. of 6 variables:

```
$ state_code : chr "AK" "AK" "AK" "AK" ...
```

```
$ school_type: Factor w/ 2 levels "private","public": 2 2 2 2 2 2
```

```
$ ncessch : chr "020000100208" "020000100211" "020000100212" "020000100213" ...
```

```
$ name : chr "Bethel Regional High School" "Ayagina'ar Elitnaurvik" "Kwigillingok Sch
```

```
$ address : chr "1006 Ron Edwards Memorial Dr" "106 Village Road" "108 Village Road" "1
```

```
$ city : chr "Bethel" "Kongiganak" "Kwigillingok" "Toksook Bay" ...
```

6. Use the `names` function to list the names of the elements (variables) within `df_recruiting`.

```
names(df_recruiting)
```

```
[1] "state_code" "school_type" "ncessch" "name" "address"
```

```
[6] "city"
```

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.

```
typeof(names(df_recruiting))
```

```
[1] "character"
```

```
length(names(df_recruiting))
```

```
[1] 6
```

```
str(names(df_recruiting))
```

```
chr [1:6] "state_code" "school_type" "ncessch" "name" "address" "city"
```

Answer: the names of the elements of the data frame object `df_recruiting` are a character atomic vector of 6 elements

Question 4: Applying basic functions to atomic vectors

/9 pts

1. Create an atomic vector object named `age` with the following values: 3, 6, 41, 43.

```
age <- c(3,6,41,43)
```

2. Apply the `typeof()`, `length()`, and `str()` functions to the object `age`.

```
typeof(age)
```

```
[1] "double"
```

```
length(age)
```

```
[1] 4
```

```
str(age)
```

```
num [1:4] 3 6 41 43
```

3. Apply the `sum()` function to `age`.

```
sum(age)
```

```
[1] 93
```

4. Apply the `sum()` function to `age` but this time include the argument `na.rm = FALSE`.

```
sum(age, na.rm = FALSE)
```

```
[1] 93
```

5. In general, what is a function “argument name” and what is an “argument value”? What does the argument `na.rm` do?

Answer: When calling a function, the argument name tells R which argument you are referring to (e.g., the `trim` argument) and the argument value is the value you are assigning to that argument. the argument `na.rm` controls whether NA values are removed or not prior to performing calculations.

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.

```
age2 <- c(3,6,41,43,NA)
sum(age2, na.rm = FALSE)
```

```
[1] NA
```

```
sum(age2, na.rm = TRUE)
```

```
[1] 93
```

Answer: The function `sum(age2, na.rm = FALSE)` returns NA because NA values are not removed prior to calculation and any calculation that includes an NA value returns NA. The function `sum(age2, na.rm = TRUE)` does not return NA because all NA values are removed prior to calculation.

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).

```
tf <- c(TRUE,FALSE,TRUE,FALSE,TRUE,NA)

typeof(tf)
```

```
[1] "logical"
```

```
length(tf)
```

```
[1] 6
```

```
str(tf)
```

```
logi [1:6] TRUE FALSE TRUE FALSE TRUE NA
```

Answer: the object `tf` is an atomic vector of logical type with 5 elements.

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?

```
sum(tf, na.rm = TRUE)
```

```
[1] 3
```

Answer: Mathematical calculations in R assign the value of 0 to FALSE and the value of 1 to TRUE

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.

```
mean(x=tf, trim = 0, na.rm = TRUE)
```

```
[1] 0.6
```

```
mean(tf, 0, TRUE)
```

```
[1] 0.6
```

Question 5: Create a GitHub issue

/1 pts

- Go to the [class repository](#) and create a new issue.
- Refer to [rclass1 student issues readme](#) for instructions on how to post questions or reflections.
- You are also required to respond to at least one issue posted by another student.
- Paste the url to your issue here:
- Paste the url to the issue you responded to here:

Knit to pdf and submit problem set

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

- Go to the [class website](#) and under the “Readings & Assignments” » “Week 1” tab, click on the “Problem set 1 submission link”.
- Submit both .Rmd and .pdf files.
- Use this naming convention “lastname_firstname_ps#” for your .Rmd and pdf files (e.g. jaquette_ozan_ps1.Rmd & jaquette_ozan_ps1.pdf).