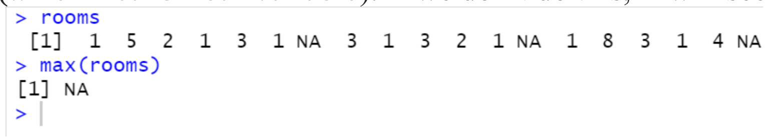
Homework, week 46: "Start with R"

Instructions: *For this assignment, you need to answer a couple questions with code and then take a particular screenshot of your working environment.*

You can submit the solutions including the URL to the screenshot typed up in a doc/pdf to Brightspace OR upload the document with solutions and the screenshot to your repository on Github and submit here (to Brightspace) only your Github URL (make sure your HW files are immediately findable there).

1: Use R to figure out how many elements in the vector below are greater than 2. rooms <- c(1, 5, 2, 1, 3, 1, NA, 3, 1, 3, 2, 1, NA, 1, 8, 3, 1, 4, NA, 1, 3, 1, 2, 1, 7, 1, NA)

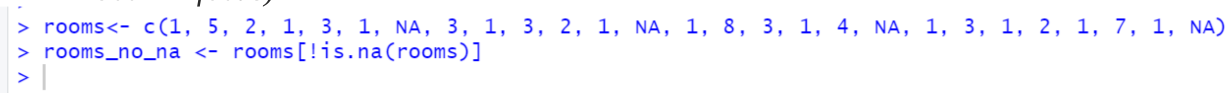
The first thing we want to do, after inserting the vector to RStudio, is to remove the "NA" value (*which means Not Available*). If we don't do this, R will see "NA" as the biggest value: e.g.,



```
> rooms
 [1]  1  5  2  1  3  1 NA  3  1  3  2  1 NA  1  8  3  1  4 NA
> max(rooms)
[1] NA
> 
```

So, in order to see the elements, without the NA, I then write in the script:
rooms_no_na <- rooms[!is.na(rooms)]

- The reason for this, is that by using this, I remove the missing data ("NA") from my dataset:
- [!is.na(rooms)]: this function shows you, if there are any missing data, but because there is an '!' in the beginning, it focusses on the non-missing data, instead of the missing data
- (*If I'd used the same function, but without the '!', it would've been the "NA"s that would have been in focus)*



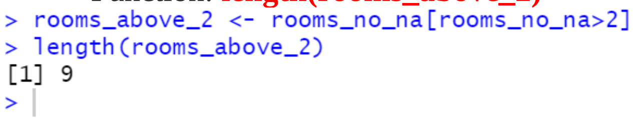
```
> rooms<- c(1, 5, 2, 1, 3, 1, NA, 3, 1, 3, 2, 1, NA, 1, 8, 3, 1, 4, NA, 1, 3, 1, 2, 1, 7, 1, NA)
> rooms_no_na <- rooms[!is.na(rooms)]
> 
```

After I've removed the "NA"s, it's time to figure out, how many rooms there are, that's over 2 rooms. In order to do this, I then use the function: **rooms_above_2 <- rooms_no_na[rooms_no_na>2]**

- By doing this, I make sure that R focus on the number of rooms over 2, but also that it still should not count the "NA" with (the right side of the <-)

Now that the rooms under 2 are being sorted by R, it's now time to see how many rooms that are over 2. For this, I use the "length" function, which is used to show us how many there are of a certain thing – and in this case it's how many rooms there are

- Function: **length(rooms_above_2)**

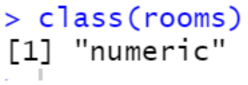


```
> rooms_above_2 <- rooms_no_na[rooms_no_na>2]
> length(rooms_above_2)
[1] 9
> 
```

So, the number of rooms that are over 2, is 9. ← here it's important to notice, that it counts the numbers that appears more than once. So, it counts every single 3 or 5

2: Which function tells you the type of data the 'rooms' vector above contains?
The function you will use to tell, what type of data you are working with, is in this case: **class(rooms)**

- "class" means here "type". As you can see in the picture, "rooms" is a numeric type of data



```
> class(rooms)
[1] "numeric"
```

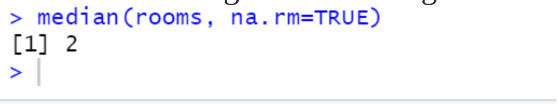
- But there is also another way of figuring this out, without using any function. And that is by looking in the top-right box in RStudio:

```
rooms      num [1:
```

  - Here, you can see the data-name, which is "rooms", and next to that stands: num, which is a short way of saying "numeric"

This way also applies to data that consists of "characters", whereas it would be "chr" instead

3: What is the result of running the median() function on the above 'rooms' vector?
The results that I get after running the **median()** function is:



```
> median(rooms, na.rm=TRUE)
[1] 2
>
```

- Function explanation:
  - Median: the name of the function
  - (rooms,): tells R that it is the rooms dataset we work with
  - Na.rm=TRUE): "Na" "removed" – so R counts the digits, and makes sure that NA's are removed

The result: the median of the vector (without NA) is 2

4: Submit the following image to Github: *Inside your R Project (.Rproj), install the 'tidyverse' package and use the* ***download.file() and read_csv()*** *function to read the SAFI_clean.csv dataset into your R project as 'interviews' digital object (see instructions in https://datacarpentry.org/r-socialsci/setup.html and 'Starting with Data' section). Take a screenshot of your RStudio interface showing*

a) the line of code you used to create the object,
b) the 'interviews' object in the Environment, and
c) the file structure of your R project in the bottom right "Files" pane.
Save the screenshot as an image and put it in your AUID_lastname_firstname repository inside our Github organisation (github.com/Digital-Methods-HASS) or equivalent. Place here the URL leading to the screenshot in your repository.

URL for task 4:
https://raw.githubusercontent.com/Digital-Methods-HASS/AU693867_Kallehauge_Simone/main/Sk%C3%A6rmbillede%20af%20Rstudio%2C%20simone%2019-11-2022.png

5) Challenge: Tidy up your Danish monarchs dataset (you created last week) sufficiently so that you can load it into R as a tibble using the read_csv() and calculate the mean() and median() duration of rule over time.