**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)

|  |  |  |
| --- | --- | --- |
| 1) | I firstly define the vector in R |  |
| 2) | I then want to see the content of the vector. This doesn’t show all the content and doesn’t specify the range. |  |
| 3) | I specifically want to see how many elements in the vector are greater than 2. |  |
| 4) | I furthermore want to exclude missing data in order to get a clearer answer.  I can thus conclude that 9 elements are greater than 2 and their value. |  |
| 5) | Alternatively, I could get the same result but a logical answer. |  |

2) Which function tells you the **type** of data the 'rooms' vector above contains?

|  |  |  |
| --- | --- | --- |
| 1) | In order to tells the type of data the 'rooms' vector above contains I use the following function.  I thus see that the vector is numeric, and the NA are recognized as no answer. |  |

3) What is the result of running the median()function on the above 'rooms' vector?

|  |  |  |
| --- | --- | --- |
| 1) | I run the function “median” in R |  |
| 2) | I get the answer NA because median is the middle value in a sorted data set, which in this data set is NA. |  |
| 3) | Alternatively, I could run the median function on the vector room­\_noNA.  In this instance the median 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.

|  |  |  |
| --- | --- | --- |
| 1) | Firstly, I install the 'tidyverse' and 'here' package. |  |
| 2) | I then make folders in my repository called “data” and “output\_data.” |  |
| 2) | I then use the download.file() and read\_csv() function to read the SAFI\_clean.csv dataset into my R project in the data file. |  |
| 3) | I then save the dataset as "interviews." |  |
| 4) | I then save this as an image. |  |
| 5) | I then go to my github repository and upload my image.  This URL leads to the screenshot:  <https://raw.githubusercontent.com/Digital-Methods-HASS/au700785_Elkjaer_Nanna/main/W46%2C%20task%204.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.

|  |  |  |
| --- | --- | --- |
| 1) | I uploaded my dataset of monarchs into R. |  |
| 2) | I then save the dataset as “monarchs” |  |

I can’t finish the task as I don’t know how to calculate the duration of rule over time.