Lasse, Sofie og Emilie

DAM assignment -Starting with R (uge 9)

For this assignment, submit here a document that lists your solution and a Github repository URL where you deposited your screenshot :

a) create a Github repository in github.com/Digital-Methods-HASS organization following instructions in slides in Brightspace (you need to acquire a login first, send usename to assistant, and accept an invitation before you can join the organization)

b) present the R code you used to produce answers to the questions below (either write up a text file or post your cleaned up R script in Github) and then

c) take a screenshot of your working environment and put it in Github (via upload file button on top right).

Questions:

1) Use R to figure out how many elements in the vector below are greater than 2.

rooms <- c(1, 5, 2, 3, 1, NA, 3, 1, 3, 2, NA, 1, 8, 3, 1, 4, NA, 1, 3, 1, 2, 1, 7, 1, NA, 4, 3 ,1 ,7 ,2 ,1 ,NA ,1, 1, 3)

Strengen af elementer sættes ind i R's script-vindue og køres via ctrl + enter. Derefter indskrives summen af værelser på mere end 2. For at fjerne de ukendte værdier bruges na.rm = TRUE. Resultatet for antallet af elementer på en værdi større end 2 er 13.

```
>
> rooms <- c(1, 5, 2, 3, 1, NA, 3, 1, 3, 2, NA, 1, 8, 3, 1, 4, NA, 1, 3, 1, 2, 1, 7, 1, NA, 4, 3 ,1 ,7 ,2 ,1 ,NA ,1, 1, 3)
> sum(rooms > 2)
[1] NA
> sum(rooms > 2, na.rm = TRUE)
[1] 13
>
```

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

Når man bruger funktionen **class()** kan man finde ud af hvilken slags data 'rooms' vektoren indeholder. Vi har brugt funktionen til både den vektor der indeholder NA og den hvor NA er blevet fjernet. Vektoren med NA hedder 'rooms' or vektoren uden NA hedder 'rooms_no_nas'.

Lasse, Sofie og Emilie

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

Man kan se, at ved at bruge funktionen ved begge vektorer, får man det resultat: “numeric”. Det fortæller os, at datatypen i 'rooms' og 'rooms_no_nas' er tal.

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

```
> median(rooms_no_nas)
[1] 2
> median(rooms)
[1] NA
```

Ved at bruge **median**() funktionen, så finder man den midterste værdi i datasættet. Medianen for datasættet er 2. Man er nødt til at skabe en ny vektor uden NAerne for at kunne bruge median funktionen. Hvis man bare bruger den ved 'rooms', så får man resultatet NA, hvilket ikke er en reel median.

4) In order to submit a screenshot of RStudio, do the following first: 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

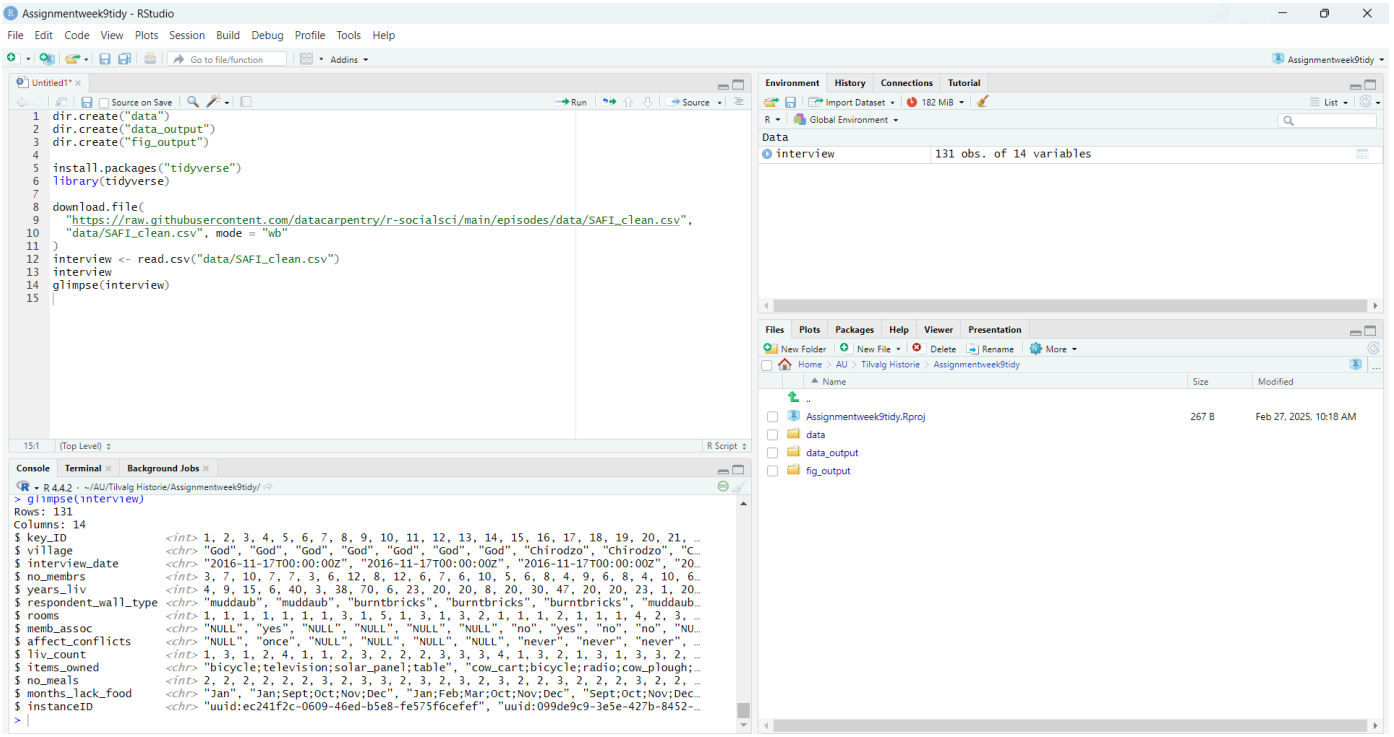
a) the line of code you used to create the object 'interviews', incl. the output in the console,

b) the 'interviews' object in the "Environment" top right pane , and

c) the file structure of **your R project** in the bottom right "Files" pane.

Save the screenshot as a .png image and put it in your **AUID_lastname_firstname** repository inside the class Github organisation (www.github.com/Digital-Methods-HASS) or equivalent. Place here the URL leading to the screenshot in your repository.

Lasse, Sofie og Emilie



Lasses GitHub-URL:

Sofies GitHub-URL:

https://github.com/Digital-Methods-HASS/AU695026_Teilmann_Sofie/blob/main/Tidyverse.png

Emilies GitHub-URL:

https://github.com/EmilieBaernthsen/au767159_Baernthsen_Emilie/blob/main/480904735_1175941560795603_6349374060398969874_n.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 their rule over time. Remember you can reload the dataset infinitely and tweak the loading as you discover issues :)

Lasse, Sofie og Emilie

```
> kongerække <- read_csv("data/kongerække2.csv")
New names:
• `` -> `...8`
Rows: 57 Columns: 11
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr (5): Navn, Kilde_født, Kilde_død, Kilde_StartReg., Kilde_SlutReg.
dbl (5): Født, Død, StartReg., SlutReg., Regeringstid
lgl (1): ...8

ℹ Use `spec()` to retrieve the full column specification for this data.
ℹ Specify the column types or set `show_col_types = FALSE` to quiet this message.
> kongerække["Regeringstid"]
# A tibble: 57 × 1
   Regeringstid
          <dbl>
 1           NA
 2           29
 3           27
 4            4
 5           17
 6            7
 7            5
 8           27
 9            6
10            6
# ℹ 47 more rows
# ℹ Use `print(n = ...)` to see more rows
```

```
> Regeringstid <- kongerække["Regeringstid"]
> mean(Regeringstid)
[1] NA
Advarselsbesked:
I mean.default(Regeringstid) :
  argument er ikke numerisk eller logisk: returnerer NA
> Regeringstid[!is.na(Regeringstid)]
 [1] 29 27  4 17  7  5 27  6  6  9  8 30  3  9 11 11 36 20 39  9  2  7 27 33  7  3  3  8 35 12  9 43  9 33
[35] 32 10 10  3 23 29 60 22 29 31 16 20 42 31  9 15 43  6 35 25 52
> Regeringstid_no_nas <- Regeringstid[!is.na(Regeringstid)]
> mean(Regeringstid_no_nas)
[1] 19.76364
> median(Regeringstid_no_nas)
[1] 16
```

How to submit this homework? Type up the solutions to the tasks (= code that got you the results + the results) , include the URL to the screenshot in a document/pdf and upload to Brightspace.

Første screenshot:

https://github.com/Digital-Methods-HASS/AU695026_Teilmann_Sofie/blob/main/Konger%C3%A6kkesvar1.png

Andet screenshot:
https://github.com/Digital-Methods-HASS/AU695026_Teilmann_Sofie/blob/main/Konger%C3%A6kkesvar2.png