Week 9 - Assignment

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)

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

Trin 1

We have tried to simplify the process by asking ChatGPT for a simpler command that would do the same calculation and achieve the same result by using a simplified formula.

Trin 2 - Kommandoen

The formula used, as suggested by ChatGPT, is "sum(rooms > 2 , na.rm = TRUE)"

Trin 3 - Script

```
##Bruger R til at finde ud af, hvor mange elementer der er større end 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)
rooms

#Tæller elementer større end 2, uden at tage højde for NAs.
sum(rooms > 2, na.rm = TRUE)
```

Trin 4 - Console

```
> sum(rooms>2 , na.rm = TRUE)
[1] 13
```

Trin 5 - Resultatet

The result would then be 13 values, greater than 2

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

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

Trin 1

We have tried to simplify the process by asking ChatGPT for a simpler command that would do the same calculation and achieve the same result by using a simplified formula.

Trin 2 - Kommandoen

By using the function "class(rooms)", R will tell you the definition of *rooms*.

Trin 3 - Script

```
# Which function tells you the type of data the 'rooms' vector above contains?
class(rooms)
```

Trin 4 - Console

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

Trin 5 - Resultatet

R will then classify *rooms* as a numeric value

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

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

Trin 1

We ran the median function on the "rooms" vector in chatGPT and got 2 as the result.

Trin 2 - Kommandoen

By using the function"median(rooms, na.rm = TRUE)", R will tell you the median.

Trin 3 - Script

```
# What is the result of running the median() function on the above 'rooms' vector?
median(rooms, na.rm = TRUE)
```

Trin 4 - Console

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

Trin 5 - Resultatet

The Median is 2

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

Trin 1

We installed the install.packages ("tidyverse")

Trin 2 - Kommandoen

By using the function"interviews <- read_csv("data/SAFI_clean.csv" , na="NULL")"

Trin 3 - Script

```
install.packages("tidyverse")
library(tidyverse)

interviews <- read_csv("data/SAFI_clean.csv" , na="NULL")
```

Trin 4 - Console

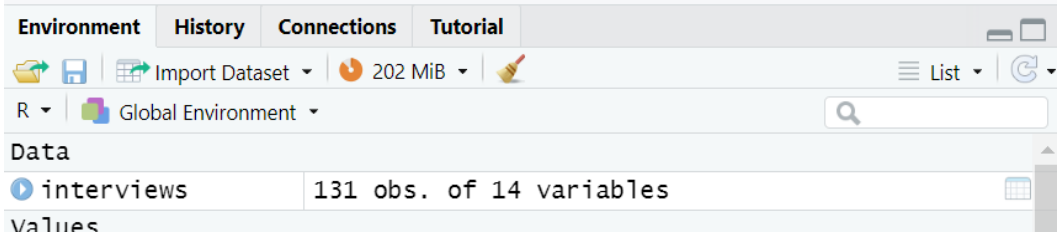
```
> interviews <- read_csv("data/SAFI_clean.csv" , na="NULL")
Rows: 131 Columns: 14
── Column specification ────────────────────────────────────────────────────────
Delimiter: ","
chr  (7): village, respondent_wall_type, memb_assoc, affect_conflicts, ite...
dbl  (6): key_ID, no_membrs, years_liv, rooms, liv_count, no_meals
dttm (1): interview_date

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.
>
```

Trin 5 - Resultatet

See above

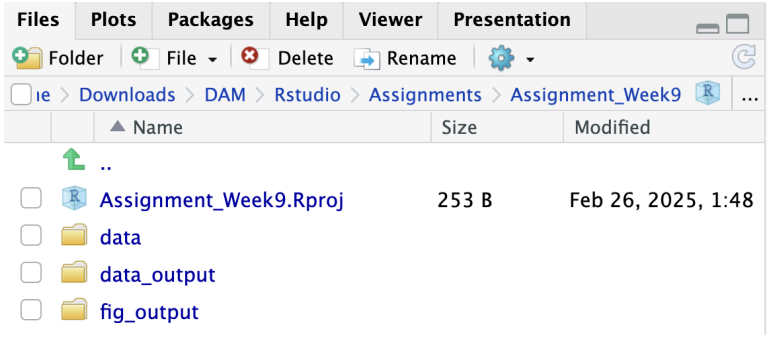
b) the 'interviews' object in the "Environment" top right pane , and

Trin 1



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

Trin 1



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.

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 :)

Challenge

Trin 1

We begin by loading the csv.file into the workspace

Trin 2 - Kommandoen

read_csv2("data/danish_monarchs_clean.csv")

Trin 3 - Script

```
read_csv2("data/danish_monarchs_clean.csv")
konger <- read_csv2("data/danish_monarchs_clean.csv")

Periode <- konger["Periode"]
Periode_uden_NA <- Periode[!is.na(Periode)]

mean(Periode_uden_NA)

median(Periode_uden_NA)
```

Trin 4 - Console

```
Rows: 54 Columns: 14
── Column specification ────────────────────────────────────────────────────────
Delimiter: ";"
chr (5): Sikker_Foedsels_aar, Sikker_doeds_aar, Sikker_start_ regeringsaar, Sikker_slut_regeringsaar, Navn
dbl (9): Foedsels_aar, Foedsels_maaned, Foedsels_dag, Doeds_aar, Doeds_maaned, Doeds_dag, Start_regeringsaar, Slut_regeringsa...

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.
# A tibble: 54 x 14
   Foedsels_aar Foedsels_maaned Foedsels_dag Sikker_Foedsels_aar Doeds_aar Doeds_maaned Doeds_dag Sikker_doeds_aar
          <dbl>           <dbl>        <dbl> <chr>                   <dbl>        <dbl>     <dbl> <chr>
 1          908              NA           NA Usikker                   958           NA        NA Sikker
 2           NA              NA           NA Usikker                   987           NA        NA Sikker
 3          980              NA           NA Usikker                  1014           NA        NA Sikker
 4           NA              NA           NA Usikker                  1018           NA        NA Sikker
 5          995              NA           NA Sikker                   1035           NA        NA Sikker
 6         1020              NA           NA Sikker                   1042           NA        NA Sikker
 7         1024              NA           NA Sikker                   1047           NA        NA Sikker
 8           NA              NA           NA Usikker                  1047            4        28 Sikker
 9           NA              NA           NA Usikker                  1080           NA        NA Sikker
10           NA              NA           NA Usikker                  1086           10         7 Sikker
# i 44 more rows
# i 6 more variables: Start_regeringsaar <dbl>, Slut_regeringsaar <dbl>, Periode <dbl>, `Sikker_start_ regeringsaar` <chr>,
#   Sikker_slut_regeringsaar <chr>, Navn <chr>
# i Use `print(n = ...)` to see more rows

> Periode <- konger["Periode"]
> Periode_uden_NA <- Periode[!is.na(Periode)]
> mean(Periode_uden_NA)
[1] 18.92157
> median(Periode_uden_NA)
[1] 16
>
```

Trin 5 - Resultatet

The Mean is 18.92157

The Median is 16