Homework 3 week 35: Start with R

Task 1: use R to figure out how many elements in the vector is greater than 2 and find sum of those elements

At first, I create a vector called "rooms" containing the elements from the assignment description:

```
rooms <- c(1, 2, 4, 5, 1, 3, 1, NA, 3, 1, 3, 2, 1, NA, 1, 8,
3, 1, 4, NA, 1, 3, 1, 2, 1, 7, 1, 9, 3, NA)
```

Now I want to remove the NA elements from the "rooms" vector. I create a new vector called "rooms_nona" containing the elements from the "rooms" vector but without the NAs:

```
rooms_nona <- rooms[!is.na(rooms)]
```

Then I create a new vector called "rooms_above_2" which only contains the values from the "rooms_nona" vector which is greater than 2:

```
list(rooms_nona[rooms_nona>2])
```

```
rooms_above_2 <- rooms_nona[rooms_nona>2]
```

I use the sum() function on the "rooms_above_2" vector to find the sum of the elements greater than two:

```
sum(rooms_above_2)
```

The sum is 55.

Task 2: What type of data is the "rooms" vector?

The data in the "rooms" vector is I use the function is.numeric() to check, if the data is numeric data:

```
Is.numeric(rooms)
```

The printed answer is TRUE, which means that the data is numeric.

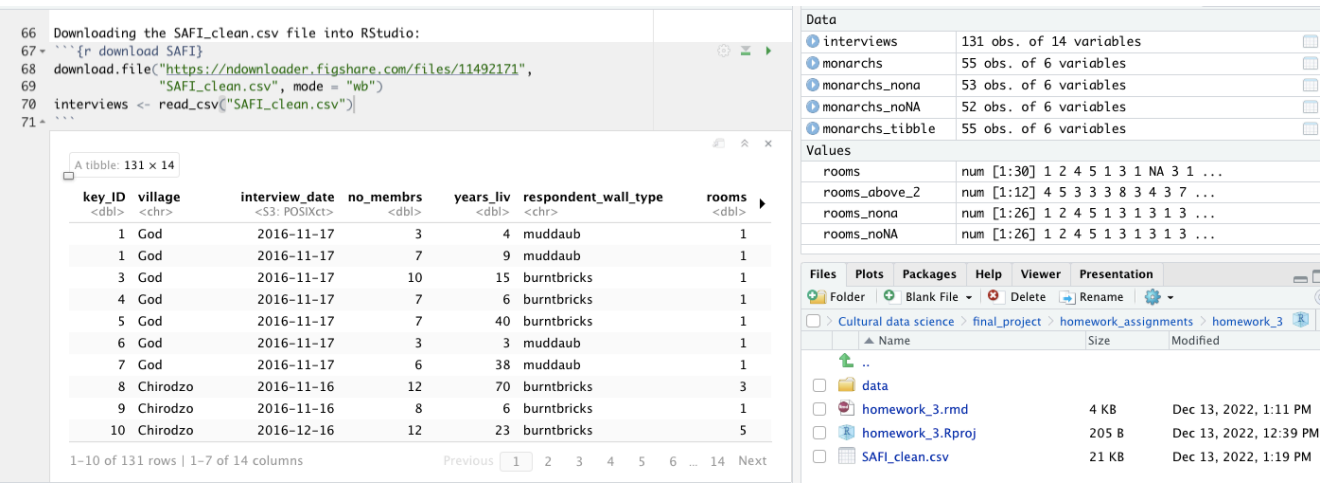
```
> #TASK 2: what type of data is the "rooms" vrector?
> is.numeric(rooms)
[1] TRUE
>
```

Task 3: turn the SAFI data into a digital object called "interviews" and take a screenshot of:
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.**

To begin this task, I install the tidyverse package in RStudio with the library() function. Then I download the file SAFI_clean.csv with the download.file() and read_csv() functions:

```
download.file("https://ndownloader.figshare.com/files/11492171",
              "SAFI_clean.csv", mode = "wb")
interviews <- read_csv("SAFI_clean.csv")
```

I have taken a screenshot of my RStudio interface, which shows the line of code I used to create the object, the 'interviews' object in the Environment, and the file structure of my R project in the bottom right "Files" pane. The screenshot can be accessed on Github here: https://github.com/Digital-Methods-HASS/au672638_Jorgensen_Emma-Marie/blob/72d46fc0cce4729ec2a1a72bacdd780008285e6d/screenshot_homework_3.png
I can also be found in the in the "homework_3" (final_project /homework_assignments/ homework_3) or seen down below:



Task 4

I started out by installing the tidyverse packages to be able to create tibble from my data set:

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

```
library(tidyverse)
```

For the homework assignment 2 I created a tidy spreadsheet with data about the Danish monarchs called "monarchs.csv", which I will use in this task.

```
monarchs <- read_csv("data/monarchs.csv")
```

I check the data type of the "monarchs" dataset with the class() function:

```
class(monarchs)
```

The data is a tibble!

The missing data is called NULL in my data set, but R doesn't interpret NULL as missing data. Before I can calculate anything from my data set, I therefore must transform the NULL values into NA:

```
monarchs[monarchs == "NULL"] <- NA
```

When I print the tibble again, the missing data is now named NA:

danish_monarchs <chr>	birth_year <chr>	death_year <chr>	reign_start_year <chr>	reign_end_year <chr>
Gorm den Gamle	NA	958	NA	958
Harald (1.) Blåtand	NA	987	958	987
Svend (1.) Tveskæg	NA	1014	987	1014
Harald (2.) Svensen	NA	1018	1014	1018
Knud (2.) den Store	995	1035	1018	1035
Hardeknud (Knud 3.)	1020	1042	1035	1042
Magnus (1.) den Gode	1024	1047	1042	1047
Svend (2.) Estridsen	NA	1076	1074	1080
Harald (3.) Hén	NA	1080	1074	1080
Knud (4.) den Hellige	NA	1086	1080	1086

I check the data type of the column "years_ruled", which I will use to calculate the mean and median of ruling time:

```
class(monarchs$years_ruled)
```

The data type isn't numeric but character, and that is a problem when I will calculate the mean and median. I must change the data type, but before changing the data type, I remove the NAs from the data set:

```
monarchs_nona <- monarchs %>% filter(years_ruled != "NA")
```

Now I can change the data type into numeric with the as.numeric() function:

```
monarchs_nona$years_ruled <-
as.numeric(monarchs_nona$years_ruled)
class(monarchs_nona$years_ruled)
```

I calculate the mean and median duration of rule over time with the mean() and the median() functions:

```
mean(monarchs_nona$years_ruled)
median(monarchs_nona$years_ruled)
```

The mean duration of rule over time is 19.72 years, and the median durations of rule over time is 17 years.

To find the three monarchs, who have been ruling the longest, I sorted the "years_ruled" column by size with the sort() function to be able to see the three greatest values:

```
sort(monarchs_nona$years_ruled)
```

The greatest values are 60, 50 and 43 years.

```{r}
sort(monarchs_nona$years_ruled)
```

```
 [1]  2  2  3  4  4  5  6  6  6  6  7  7  8  8  9  9  9  9 10 10 10 11 11 11 15 16 17 20 20 22 25
[32] 25 25 25 27 27 27 29 29 29 30 31 31 32 33 33 35 35 39 42 43 50 60
```

Then I found out to which rows the three values belonged using the which() function:

```
which(monarchs == 60, arr.ind=TRUE)
which(monarchs == 50, arr.ind=TRUE)
which(monarchs == 43, arr.ind=TRUE)
```

```{r find rows}
which(monarchs == 60, arr.ind=TRUE)
which(monarchs == 50, arr.ind=TRUE)
which(monarchs == 43, arr.ind=TRUE)
```

```
     row col
[1,]  40   6
     row col
[1,]  54   6
     row col
[1,]  50   6
```

The biggest values belong to row 40, 54 and 50. I print those rows to find the names of the three monarchs I'm looking for:

```
print(monarchs[40,])
print(monarchs[54,])
print(monarchs[50,])
```

danish_monarchs	birth_year	death_year	reign_start_year	reign_end_year	years_ruled
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
Christian 4.	1577	1648	1588	1648	60

1 row

danish_monarchs	birth_year	death_year	reign_start_year	reign_end_year	years_ruled
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
Margrethe 2.	1940	NA	1972	NA	50

1 row

danish_monarchs	birth_year	death_year	reign_start_year	reign_end_year	years_ruled
<chr>	<chr>	<chr>	<chr>	<chr>	<chr>
Christian 9.	1818	1906	1863	1906	43

1 row

The three monarchs with the longest duration of rule over time is Christian 4th, Margrethe 2nd and Christian 9th.

To find the number of days the three monarchs have ruled, I multiply the number of years by 365, because a year consists of 365 days in average. It must be noted that I haven't taken leap years into account:

```
60*365 #Christian 4th
50*365 # Margrethe 2nd
43*365 # Christian 9th
```

```{r days of ruling}
#Christian 4th (row 39)
60*365
# Margrethe 2nd (row 53)
50*365
# Christian 9th (row 49)
43*365
```

```
[1] 21900
[1] 18250
[1] 15695
```

Christian 9th has ruled for 21900 days, Margrethe 2nd for 18250 days and Christian 9th for 15695 days.