Homework week 46: Start with R

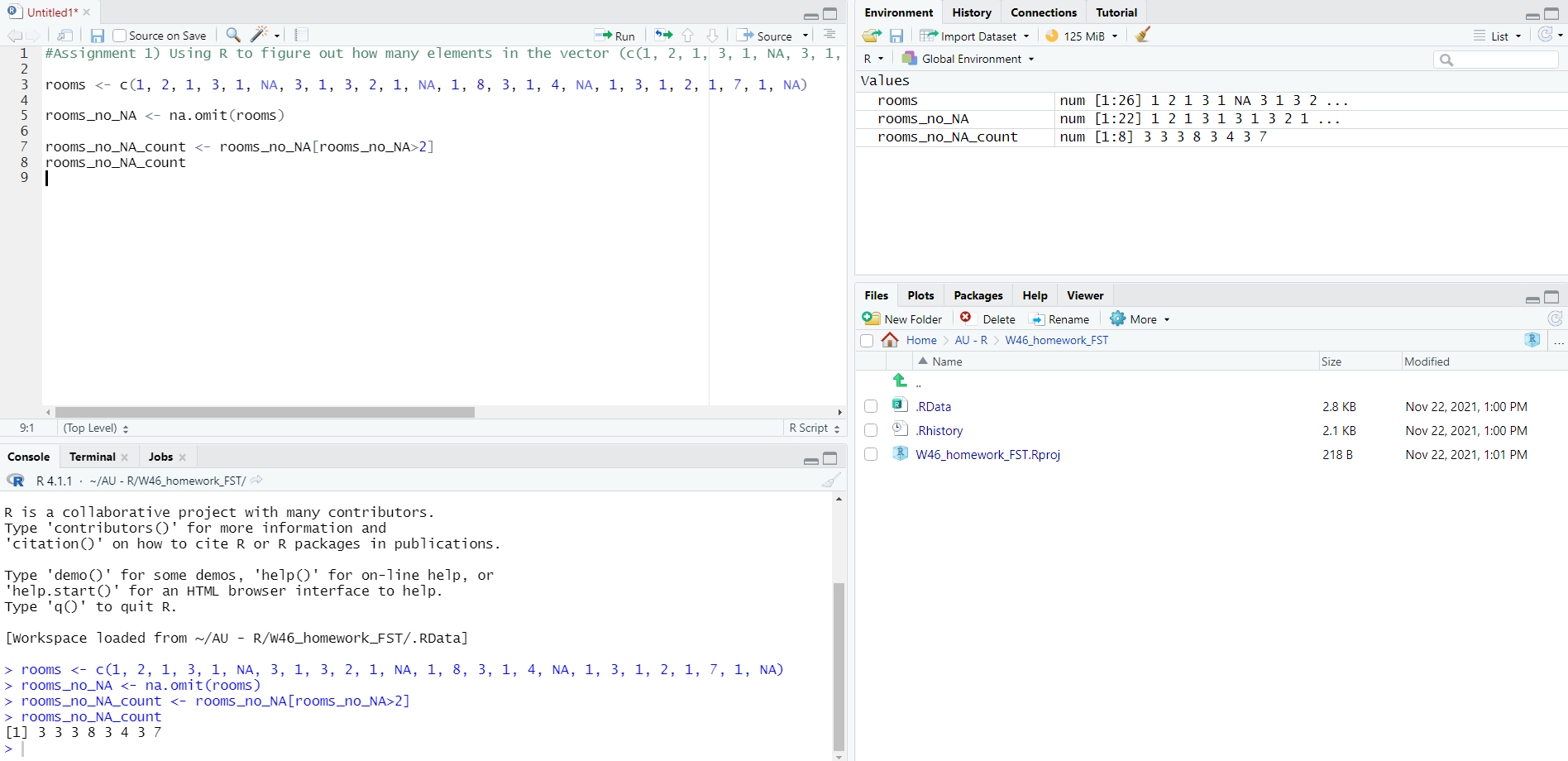
# DESCRIPTION

**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).

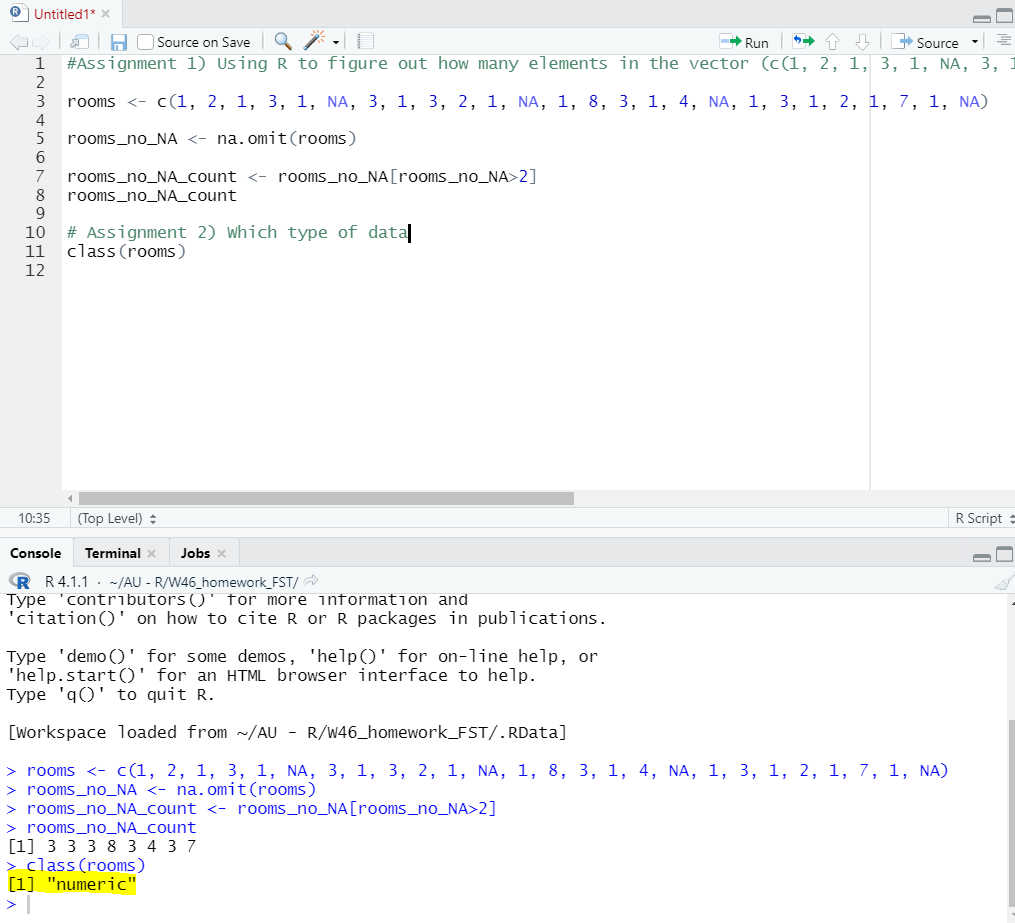
# Use R to figure out how many elements in the vector below are greater than 2. rooms <- c(1, 2, 1, 3, 1, NA, 3, 1, 3, 2, 1, NA, 1, 8, 3, 1, 4, NA, 1, 3, 1, 2, 1, 7, 1, NA)

First, I load the data sequence into R and name it “rooms”. Since there are a few fields that are missing, I rewrite the sequence to avoid confusion. I call the new sequence “rooms\_no\_NA”. I had a difficult time figuring out a code that would give me the count of rooms that are greater than two rooms. One could count them manually, but that would be less efficient. I tried using the max() function, however even though it showed 8, which is the count of rooms greater than two, it could also be the highest number in the sequence. I instead made it into a new value, where it shows the interval of values to be [1:8]. **The are 8 households that have more than 2 rooms.**



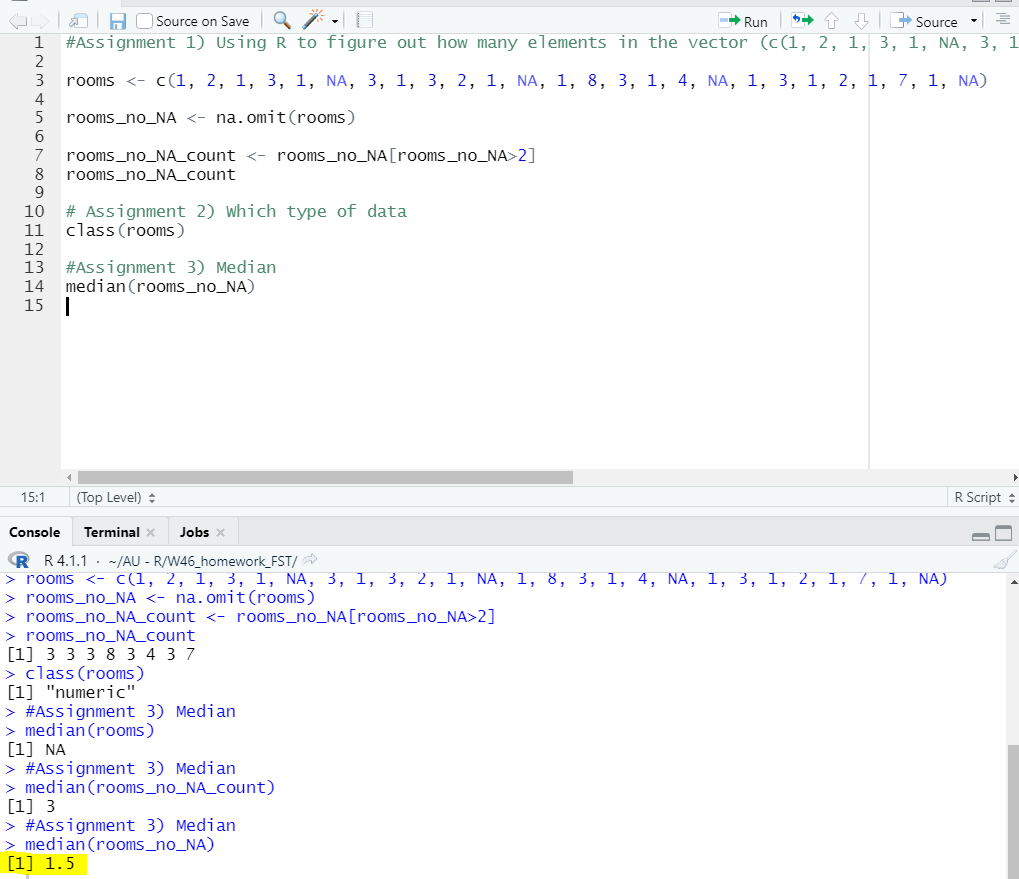
# What **type** of data is in the 'rooms' vector?

To find out what type of data is in the “rooms” vector I write the function class(rooms) and it then tells me that the data is numeric. – Marked with yellow



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

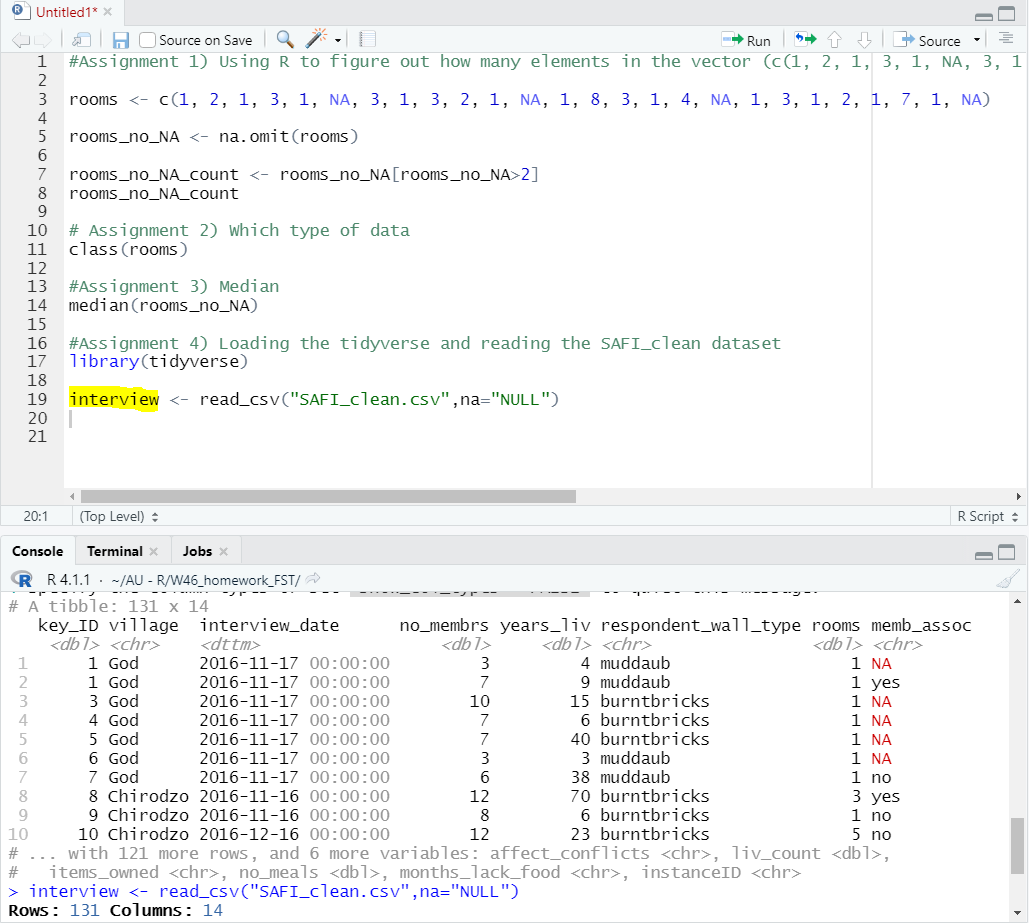
When finding the median one needs to use the rewritten sequence to avoid the “NA”. The median found is 1,5, which means that the middle value in the sorted dataset is 1,5.



# 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

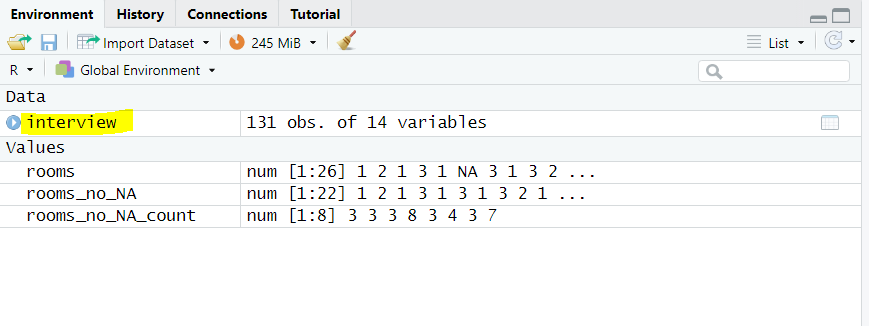
### the line of code you used to create the object,

The line of code is interview <- read\_csv("SAFI\_clean.csv",na="NULL") as showed in the screenshot below. The <- marks the creation as an object.



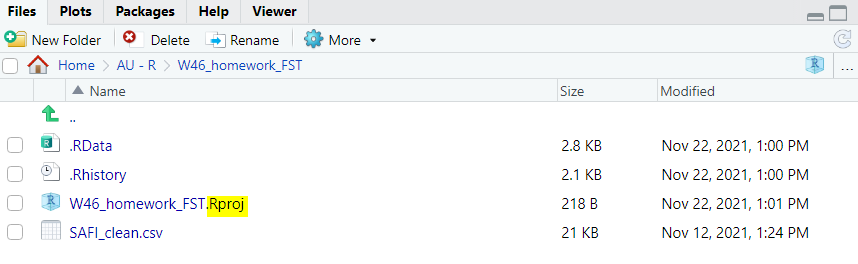
### the 'interviews' object in the Environment, and

Since creating the object, the “interviews” will be showed in the Environment as seen on the screenshot.



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

I saved the file as an R-project as seen in “Files”.



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. URL for screenshot repository: <https://github.com/Digital-Methods-HASS/AU677820-Thoroe-Freja.git> (copy it into the search field, URL does not work if you click on it).

# 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() or some other:) function and calculate the mean() and median() duration of rule over time

I did not understand how I could tidy my data better, so I just tried to upload the spreadsheet anyways. I had some trouble uploading the file, because it was placed in a different folder than the folder I was working on. I therefore moved the document into the correct folder to upload it. I read the csv file and marked that the values are separated by “;”.

