HW3-W35-start-with-R

Mina Almasi

2022-08-31

Instructions: For this assignment, you need to answer a couple questions with code and then take a screenshot of your working environment.

Submit the solutions including the URL to the screenshot in a doc/pdf to Brightspace.

(1) Use R to figure out how many elements in the vector below are **greater than 2** and then tell me what their **sum** (of the larger than 2 elements) is.

```
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)
```

Firstly, we omit NA values:

```
rooms <- na.omit(rooms)
```

Then, we subset only the integers greater than 2:

```
# subsetting
greater_than_2 <- rooms[rooms>2]

# printing the vector
greater_than_2
```

```
##  [1] 4 5 3 3 3 8 3 4 3 7 9 3
```

We take the length of the vector *greater_than_2* which we just created:

```
length(greater_than_2)
```

```
## [1] 12
```

Finally, we take the sum of said vector:

```
sum(greater_than_2)
```

```
## [1] 55
```

We conclude that there are 12 elements in the vector *rooms* that are greater than 2 and the sum of them vector is 55.

(2) What **type** of data is in the 'rooms' vector?

The type of data in the *rooms* vector is integers (i.e., whole numbers that is either positive, negative or zero - in this case only positive).

(3) 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 (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.

The URL is here: https://github.com/Digital-Methods-HASS/au675000_Almasi_Mina/blob/main/homework/HW3-TASK-MINA3.png (https://github.com/Digital-Methods-HASS/au675000_Almasi_Mina/blob/main/homework/HW3-TASK-MINA3.png)

(4) Challenge: Attached to this assignment you will find Danish monarchs dataset that might need to be cleaned up a bit. Load the provided 'kings.csv" dataset into R as a tibble. Calculate the mean() and median() duration of rule over time and find the three mondarchs ruling the longest. How many days did they rule (accounting for transition year?)

Firstly, we loading the package *tidyverse* and the kings data

```
# package
library(tidyverse)

# loading kings data
kings_data <- read_csv2("data/kings.csv")
```

Then we do some preprocessing of the data. In this case, we filter the NA's away("Unknowns in this case) and make the column"Yearasruler" numeric.

```
# filtering the unknown columns away
kings_data <- kings_data %>% filter(Yearasruler != "Unknown")

# making the yearasruler into numeric
kings_data$Yearasruler <- as.numeric(kings_data$Yearasruler)
```

We calculate the median and mean duration rule over time:

```
# calculating the median and mean duration rule over time
mean_rule_time <- mean(kings_data$Yearasruler)
median_rule_time <- median(kings_data$Yearasruler)

# printing the mean and median rule time
tibble(mean_rule_time, median_rule_time)
```

```
## # A tibble: 1 × 2
##   mean_rule_time median_rule_time
##            <dbl>            <dbl>
## 1           18.7               14
```

The mean rule time is thus 18.7 and the median is 14 years.

We find the three monarchs who have ruled for the longest duration. We do this by using the arrange function from the dplyr package to sort the data according to the amount of years they ruled (from highest to lowest which is denoted by *desc* for descending).

```
# sorting the kings_data according to Yearasruler
kings_data_sorted <- kings_data %>% arrange(desc(Yearasruler))

# we save the three monarchs who have ruled for the longest duration (top three)
top_three <- head(kings_data_sorted, 3)

# printing the three highest
top_three
```

```
## # A tibble: 3 × 4
##   Kings              Start_date End_date Yearasruler
##   <chr>                   <dbl>    <dbl>       <dbl>
## 1 Christian 4.             1588     1648          60
## 2 Erik 7. af Pommern       1396     1439          43
## 3 Christian 7.             1766     1808          42
```

The three monarchs who ruled for the longest amount of duration are thus Christian the 4th, Erik 7 af Pommeren and Christian the 7th.

Finally, we calculate how many days they ruled. To account for transition/leap years, we could go the historical route and see when exactly there were transition years, but for the sake of this exercise, we will simply calculated based on the knowledge that leap years occur every 4 years.

```
top_three$leap_years <- (top_three$Yearasruler/4)
```

For each transition/leap year, we know that there will be an extra day (366 instead of 365). So instead of multiplying all years by 365, we would have to multiply the leap years by 366 days and the normal years by 365.

A short cut is to multiply all of the years by 365 and simply add the leap years to this calculation as done below.

```
top_three$days_as_ruler <- top_three$Yearasruler * 365 + top_three$leap_years
```

We print final table which includes the calculation of the days as ruler as the final column:

```
top_three
```

```
## # A tibble: 3 × 6
##   Kings              Start_date End_date Yearasruler leap_years days_as_ruler
##   <chr>                   <dbl>    <dbl>       <dbl>      <dbl>         <dbl>
## 1 Christian 4.             1588     1648          60         15         21915
## 2 Erik 7. af Pommern       1396     1439          43       10.8        15706.
## 3 Christian 7.             1766     1808          42       10.5        15340.
```