

Homework 1

Advait Ashtikar

Table of contents

Loading libraries	2
Question 1	2
Question 2	3
Question 3	5

Appendix	8
-----------------	----------

[Link to the Github repository](#)

! Due: Sun, Jan 29, 2023 @ 11:59pm

Please read the instructions carefully before submitting your assignment.

1. This assignment requires you to:
 - Upload your Quarto markdown files to a `git` repository
 - Upload a `PDF` file on Canvas
2. Don't collapse any code cells before submitting.
3. Remember to make sure all your code output is rendered properly before uploading your submission.

Please add your name to the the author information in the frontmatter before submitting your assignment.

Loading libraries

```
library(tidyverse)
```

```
-- Attaching packages ----- tidyverse 1.3.2 --
v ggplot2 3.4.0      v purrr   1.0.1
v tibble  3.1.8      v dplyr  1.0.10
v tidyr   1.2.1      v stringr 1.5.0
v readr   2.1.3      v forcats 0.5.2
-- Conflicts ----- tidyverse_conflicts() --
x dplyr::filter() masks stats::filter()
x dplyr::lag()     masks stats::lag()
```

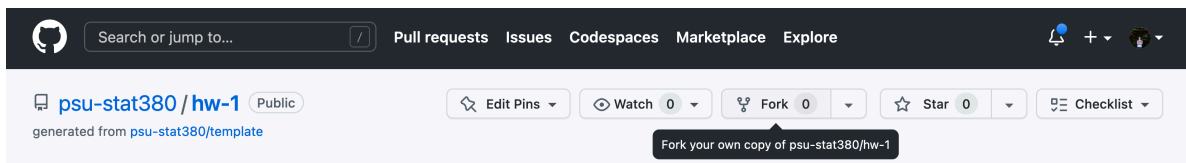
```
library(dplyr)
```

Question 1

💡 20 points

In this question, we will walk through the process of *forking* a **git** repository and submitting a *pull request*.

1. Navigate to the Github repository [here](#) and fork it by clicking on the icon in the top right



Provide a sensible name for your forked repository when prompted.

2. Clone your Github repository on your local machine

```
$ git clone <<insert your repository url here>>
$ cd hw-1
```

Alternatively, you can use Github codespaces to get started from your repository directly.

3. In order to activate the R environment for the homework, make sure you have `renv` installed beforehand. To activate the `renv` environment for this assignment, open an instance of the R console from within the directory and type

```
renv::activate()
```

Follow the instructions in order to make sure that `renv` is configured correctly.

4. Work on the *remaining part* of this assignment as a `.qmd` file.
 - Create a PDF and HTML file for your output by modifying the YAML frontmatter for the Quarto `.qmd` document
5. When you're done working on your assignment, push the changes to your github repository.
6. Navigate to the original Github repository [here](#) and submit a pull request linking to your repository.

Remember to **include your name** in the pull request information!

If you're stuck at any step along the way, you can refer to the [official Github docs here](#)

Question 2

💡 30 points

Consider the following vector

```
my_vec <- c(
  "+0.07",
  "-0.07",
  "+0.25",
  "-0.84",
  "+0.32",
  "-0.24",
  "-0.97",
  "-0.36",
  "+1.76",
  "-0.36"
```

```
)
```

For the following questions, provide your answers in a code cell.

1. What data type does the vector contain?

```
# The vector contains the data type 'Character'
```

1. Create two new vectors called `my_vec_double` and `my_vec_int` which converts `my_vec` to Double & Integer types, respectively,

```
my_vec_double <- as.double(my_vec)
my_vec_int <- as.integer(my_vec)
```

1. Create a new vector `my_vec_bool` which comprises of:

- TRUE if an element in `my_vec_double` is ≤ 0
- FALSE if an element in `my_vec_double` is ≥ 0

How many elements of `my_vec_double` are greater than zero?

```
my_vec_bool <- my_vec_double <= 0

sum(my_vec_double > 0)
```

```
[1] 4
```

```
# There are 4 elements of my_vec_double that are greater than zero
```

1. Sort the values of `my_vec_double` in ascending order.

```
sort(my_vec_double)
```

```
[1] -0.97 -0.84 -0.36 -0.36 -0.24 -0.07  0.07  0.25  0.32  1.76
```

Question 3

💡 50 points

In this question we will get a better understanding of how R handles large data structures in memory.

1. Provide R code to construct the following matrices:

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & \dots & 100 \\ 1 & 4 & 9 & 16 & 25 & \dots & 10000 \end{bmatrix}$$

... {.cell}

```
matrix1 <- matrix(1:9, nrow = 1)
matrix2 <- matrix(c(1:100, (1:100)^2), nrow = 2)
```

...

⚠️ Tip

Recall the discussion in class on how R fills in matrices

In the next part, we will discover how knowledge of the way in which a matrix is stored in memory can inform better code choices. To this end, the following function takes an input n and creates an $n \times n$ matrix with random entries.

```
generate_matrix <- function(n){
  return(
    matrix(
      rnorm(n^2),
      nrow=n
    )
  )
}
```

For example:

```
generate_matrix(4)
```

	[,1]	[,2]	[,3]	[,4]
[1,]	-1.0784276	0.85347990	-0.3651247	0.0750846
[2,]	-1.0963550	0.51869413	2.0545369	-1.1432503
[3,]	2.5093680	0.70496041	2.0003750	-1.5754835
[4,]	0.2247585	0.07561045	-0.4931524	-1.5594556

Let M be a fixed 50×50 matrix

```
M <- generate_matrix(50)
mean(M)
```

```
[1] -0.01403936
```

2. Write a function `row_wise_scan` which scans the entries of M one row after another and outputs the number of elements whose value is ≥ 0 . You can use the following **starter code**

```
row_wise_scan <- function(x){
  n <- nrow(x)
  m <- ncol(x)
  count <- 0
  for(i in 1:n){
    for(j in 1:m){
      if(x[i,j] >= 0){
        count <- count + 1
      }
    }
  }

  return(count)
}
```

3. Similarly, write a function `col_wise_scan` which does exactly the same thing but scans the entries of M one column after another

```
col_wise_scan <- function(x){
  count <- 0
  for(j in 1:ncol(x)){
    for(i in 1:nrow(x)){
      if(x[i,j] >= 0){
        count <- count + 1
      }
    }
  }
}
```

```

    }
  }
  return(count)
}

```

You can check if your code is doing what it's supposed to using the function here¹

4. Between `col_wise_scan` and `row_wise_scan`, which function do you expect to take shorter to run? Why? `:: {cell}`

```

# Between 'col_wise_scan' and 'row_wise_scan', I expect 'col_wise_scan' to take shorter to

```

```

:::

```

5. Write a function `time_scan` which takes in a method `f` and a matrix `M` and outputs the amount of time taken to run `f(M)`

```

time_scan <- function(f, M){
  initial_time <- Sys.time()
  f(M)
  final_time <- Sys.time()

  total_time_taken <- final_time - initial_time
  return(total_time_taken)
}

```

Provide your output to

```

list(
  row_wise_time = time_scan(row_wise_scan, M),
  col_wise_time = time_scan(col_wise_scan, M)
)

```

¹If your code is right, the following code should evaluate to be `TRUE`

```

sapply(1:100, function(i) {
  x <- generate_matrix(100)
  row_wise_scan(x) == col_wise_scan(x)
}) %>% sum == 100

```

```

[1] TRUE

```

```
$row_wise_time  
Time difference of 0.006744862 secs
```

```
$col_wise_time  
Time difference of 0.000164032 secs
```

Which took longer to run? ::: {.cell}

```
# 'row_wise_time' took longer to run
```

:::

6. Repeat this experiment now when:

- M is a 100×100 matrix
- M is a 1000×1000 matrix
- M is a 5000×5000 matrix

What can you conclude? ::: {.cell}

```
# When we repeat the experiment for matrices of different sizes the time taken by 'row_wise'
```

:::

Appendix

Print your R session information using the following command

```
sessionInfo()
```

```
R version 4.2.2 (2022-10-31)  
Platform: x86_64-apple-darwin17.0 (64-bit)  
Running under: macOS Big Sur ... 10.16
```

```
Matrix products: default  
BLAS:   /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRblas.0.dylib  
LAPACK: /Library/Frameworks/R.framework/Versions/4.2/Resources/lib/libRlapack.dylib
```

```
locale:
```



```
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
```

attached base packages:

```
[1] stats      graphics  grDevices datasets  utils      methods    base
```

other attached packages:

```
[1] forcats_0.5.2  stringr_1.5.0  dplyr_1.0.10   purrr_1.0.1
[5] readr_2.1.3    tidyr_1.2.1    tibble_3.1.8   ggplot2_3.4.0
[9] tidyverse_1.3.2
```

loaded via a namespace (and not attached):

```
[1] tidyselect_1.2.0  xfun_0.36      haven_2.5.1
[4] gargle_1.2.1      colorspace_2.0-3 vctrs_0.5.1
[7] generics_0.1.3    htmltools_0.5.4 yaml_2.3.6
[10] utf8_1.2.2        rlang_1.0.6     pillar_1.8.1
[13] withr_2.5.0       glue_1.6.2      DBI_1.1.3
[16] dbplyr_2.2.1      readxl_1.4.1    modelr_0.1.10
[19] lifecycle_1.0.3   munsell_0.5.0   gtable_0.3.1
[22] cellranger_1.1.0  rvest_1.0.3     evaluate_0.20
[25] knitr_1.41        tzdb_0.3.0      fastmap_1.1.0
[28] fansi_1.0.3       broom_1.0.2     renv_0.16.0-53
[31] backports_1.4.1   scales_1.2.1    googlesheets4_1.0.1
[34] jsonlite_1.8.4    fs_1.5.2        hms_1.1.2
[37] digest_0.6.31     stringi_1.7.12  grid_4.2.2
[40] cli_3.6.0         tools_4.2.2     magrittr_2.0.3
[43] crayon_1.5.2      pkgconfig_2.0.3 ellipsis_0.3.2
[46] xml2_1.3.3        reprex_2.0.2    googledrive_2.0.0
[49] lubridate_1.9.0   timechange_0.2.0 assertthat_0.2.1
[52] rmarkdown_2.20    httr_1.4.4      rstudioapi_0.14
[55] R6_2.5.1          compiler_4.2.2
```