BSE658A-Assignment-03

Abin Thomas (Roll No. 21218261)

2022-08-22

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.2 --
## v ggplot2 3.3.6     v purrr   0.3.4
## v tibble  3.1.8     v dplyr   1.0.9
## v tidyr   1.2.0     v stringr 1.4.0
## v readr   2.1.2     v forcats 0.5.1
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

Questions

1. If you open the R console, the first line in the console shows the version of R you have installed on your computer. Please specify which R version you have by opening R console. (1 point)

Answer: *4.2.1*

2. There are 'base' packages in R which you don't have to install separately after installing the latest version of R - they are bundled with R (like the furniture or dining table you might get when you rent a new house.) One of these base packages/libraries in R is `datasets`. Type `help(datasets)` to read more about it. View the complete list of datasets in `datasets` and load the dataset `AirPassengers` in your RStudio. Write the code below which you will use to solve question 3 and 4. (1 point)

Answer:

```
AirPassengers
```

```
##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 1949 112 118 132 129 121 135 148 148 136 119 104 118
## 1950 115 126 141 135 125 149 170 170 158 133 114 140
## 1951 145 150 178 163 172 178 199 199 184 162 146 166
## 1952 171 180 193 181 183 218 230 242 209 191 172 194
## 1953 196 196 236 235 229 243 264 272 237 211 180 201
## 1954 204 188 235 227 234 264 302 293 259 229 203 229
## 1955 242 233 267 269 270 315 364 347 312 274 237 278
## 1956 284 277 317 313 318 374 413 405 355 306 271 306
## 1957 315 301 356 348 355 422 465 467 404 347 305 336
## 1958 340 318 362 348 363 435 491 505 404 359 310 337
## 1959 360 342 406 396 420 472 548 559 463 407 362 405
## 1960 417 391 419 461 472 535 622 606 508 461 390 432
```

```
help("AirPassengers")
```

```
## starting httpd help server ... done
```

3. Mention the source of the dataset `AirPassengers`. (1 point)

Answer: *Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1976) Time Series Analysis, Forecasting and Control. Third Edition. Holden-Day. Series G*

4. How many rows and columns does the dataset `AirPassengers` have? (1 point)

Answer: Number of rows: *12*, number of columns: *12*. #The above answer was based only on manually counting, is there any function that can directly count rows and columns from a time series dataframe?

```
apmatrix <- matrix(AirPassengers, ncol = 12, byrow =TRUE,
                   dimnames = list( as.character(1949:1960),month.abb))

apmatrix
```

```
##      Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
## 1949 112 118 132 129 121 135 148 148 136 119 104 118
## 1950 115 126 141 135 125 149 170 170 158 133 114 140
## 1951 145 150 178 163 172 178 199 199 184 162 146 166
## 1952 171 180 193 181 183 218 230 242 209 191 172 194
## 1953 196 196 236 235 229 243 264 272 237 211 180 201
## 1954 204 188 235 227 234 264 302 293 259 229 203 229
## 1955 242 233 267 269 270 315 364 347 312 274 237 278
## 1956 284 277 317 313 318 374 413 405 355 306 271 306
## 1957 315 301 356 348 355 422 465 467 404 347 305 336
## 1958 340 318 362 348 363 435 491 505 404 359 310 337
## 1959 360 342 406 396 420 472 548 559 463 407 362 405
## 1960 417 391 419 461 472 535 622 606 508 461 390 432
```

```
aptibble<- as_tibble(apmatrix)
nrow(apmatrix)
```

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

```
ncol(apmatrix)
```

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

5. There are multiple columns in `AirPassengers`. Can you create a new column `Passengers` by adding all the values in different columns in a given row? The number of rows in the original dataset and the new dataset will be the same. (3 points)

Answer:

```
apt1 <- aptibble %>% mutate(Passengers = rowSums(.),.before=Jan)
apt1
```

```
## # A tibble: 12 x 13
##    Passengers   Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov
##         <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1       1520   112   118   132   129   121   135   148   148   136   119   104
##  2       1676   115   126   141   135   125   149   170   170   158   133   114
##  3       2042   145   150   178   163   172   178   199   199   184   162   146
##  4       2364   171   180   193   181   183   218   230   242   209   191   172
##  5       2700   196   196   236   235   229   243   264   272   237   211   180
##  6       2867   204   188   235   227   234   264   302   293   259   229   203
##  7       3408   242   233   267   269   270   315   364   347   312   274   237
##  8       3939   284   277   317   313   318   374   413   405   355   306   271
##  9       4421   315   301   356   348   355   422   465   467   404   347   305
## 10       4572   340   318   362   348   363   435   491   505   404   359   310
## 11       5140   360   342   406   396   420   472   548   559   463   407   362
## 12       5714   417   391   419   461   472   535   622   606   508   461   390
## # ... with 1 more variable: Dec <dbl>
## # i Use `colnames()` to see all variable names
```

6. How will you check the number of `NA` values in the dataset `AirPassengers`? (1 point)

Answer:

```
sum(is.na(apt1))
```

```
## [1] 0
```

7. You might have used `Tidyverse` package to solve the above questions. Can you create a new column `PercentPassengers` which shows the percentage of passengers who flew in a given month of a given year? The value of hundred percentage denotes the total number of passengers who flew in a given year (which you have already calculated in the column `Passengers`). Make sure the column `PercentPassengers` is the first column in the dataset.(3 points)

Answer:

```
apt2 <- apt1 %>% mutate(across()*100/apt1$Passengers )
Pss=apt1$Passengers
#stored the list of passengers across years in a list 'Pss'
#because haven't discovered how to use 'across' with exceptions
apt2 <- apt2 %>% mutate(Passengers = Pss)
#plugged the number of passengers for each year back into the final tibble
apt2
```

```
## # A tibble: 12 x 13
##    Passengers   Jan   Feb   Mar   Apr   May   Jun   Jul   Aug   Sep   Oct   Nov
##         <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
##  1       1520  7.37  7.76  8.68  8.49  7.96  8.88  9.74  9.74  8.95  7.83  6.84
##  2       1676  6.86  7.52  8.41  8.05  7.46  8.89 10.1  10.1   9.43  7.94  6.80
##  3       2042  7.10  7.35  8.72  7.98  8.42  8.72  9.75  9.75  9.01  7.93  7.15
```

```
##  4       2364  7.23  7.61  8.16  7.66  7.74  9.22  9.73 10.2   8.84  8.08  7.28
##  5       2700  7.26  7.26  8.74  8.70  8.48  9     9.78 10.1   8.78  7.81  6.67
##  6       2867  7.12  6.56  8.20  7.92  8.16  9.21 10.5  10.2   9.03  7.99  7.08
##  7       3408  7.10  6.84  7.83  7.89  7.92  9.24 10.7  10.2   9.15  8.04  6.95
##  8       3939  7.21  7.03  8.05  7.95  8.07  9.49 10.5  10.3   9.01  7.77  6.88
##  9       4421  7.13  6.81  8.05  7.87  8.03  9.55 10.5  10.6   9.14  7.85  6.90
## 10       4572  7.44  6.96  7.92  7.61  7.94  9.51 10.7  11.0   8.84  7.85  6.78
## 11       5140  7.00  6.65  7.90  7.70  8.17  9.18 10.7  10.9   9.01  7.92  7.04
## 12       5714  7.30  6.84  7.33  8.07  8.26  9.36 10.9  10.6   8.89  8.07  6.83
## # ... with 1 more variable: Dec <dbl>
## # i Use ‘colnames()‘ to see all variable names
```