Initiation to R Software

Pierre Michel

Master AMSE 1st year, 2019

Problem Set II

1) Matrix calculus

Note: In this exercise, use randomly generated matrices and vectors (see Problem Set I).

For example you should quickly get this kind of matrix using the functions `matrix()` and `sample()`:

```
##      [,1] [,2] [,3]
## [1,]    3    5    1
## [2,]    4    2    9
## [3,]    6    7    8
```

a) How to compute the sum of the elements of a vector/matrix ? Mathematically, the sum of the elements of a matrix $A \in \mathbb{R}^{m \times n}$, denoted $\Sigma(A)$, with elements $(a_{i,j}) \in \mathbb{R}$ is defined as follows:

$$\Sigma(A) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{i,j} = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{i,j}$$

b) Compute the sum of the elements of each column of a matrix.
c) Let `M` be a matrix of dimension (4,3). What is the value of M[11] ? Extract an element using another linear index.
d) Generate 2 random vectors `x` and `y`, each with 100 observations draw uniformly from the 5 first integers for `x` and from the 3 first integers for `y`.
e) Compute the contengency table of `x` and `y` (function `table()`), and put it in a matrix `M`.
f) Compute the means by row/column of `M` (function `apply()`).
g) Create a matrix `Q` with the same elements as `M`, but with dimension (2,6).
h) Create a matrix `P` whose elements are `x[i]*y[i]/N`, where `N` is the length of two vectors `x` and `y`. Consider `N = 100`.
i) Extract the submatrix of `P` whose first element of each row is a value greater than 2.

2) Dataframes in R packages

a) Create a `data.frame` with 4 columns (3 numeric, 1 character) and 5 observations (rows). Choose the name of each column. Below is what you should get…

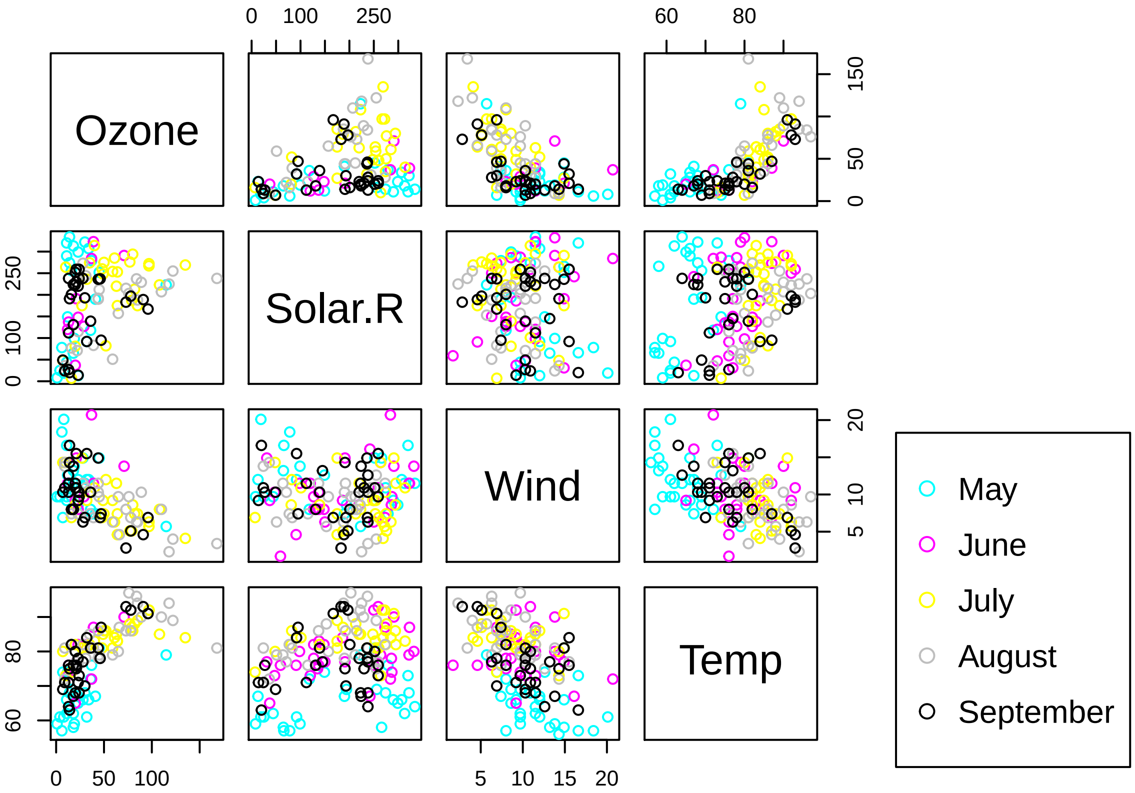
```
##            x        y          z w
## 1 -0.6641500 29.30909 -42.844499 J
## 2  0.3110660 45.27543  -5.390607 Z
## 3 -0.5999580 40.00538  33.112599 W
## 4 -0.3526568 26.37235  -6.900448 G
```

b) Choose the name of each row.
c) Add one numeric column to the `data.frame`, this column should contain the sum of the two first numeric columns values. Choose a name.
d) Add one logical column to the `data.frame`, this column should indicate if the values in the third column are greater than 10. Choose a name.
e) Remove the first and last columns.
f) How many available packages are there in your R session ? Which ones ? Use `library()`.

g) Import in a `data.frame` the data from the file **airquality** in package `datasets`. What does it contain ?

h) Print some descriptive statistics about the columns of this dataset. And plot some graphics. Comment the results. Below is what you should get:

```
##      Ozone           Solar.R           Wind             Temp      
##  Min.   :  1.00   Min.   :  7.0   Min.   : 1.700   Min.   :56.00  
##  1st Qu.: 18.00   1st Qu.:115.8   1st Qu.: 7.400   1st Qu.:72.00  
##  Median : 31.50   Median :205.0   Median : 9.700   Median :79.00  
##  Mean   : 42.13   Mean   :185.9   Mean   : 9.958   Mean   :77.88  
##  3rd Qu.: 63.25   3rd Qu.:258.8   3rd Qu.:11.500   3rd Qu.:85.00  
##  Max.   :168.00   Max.   :334.0   Max.   :20.700   Max.   :97.00  
##  NA's   :37       NA's   :7                                       
##      Month            Day      
##  Min.   :5.000   Min.   : 1.0  
##  1st Qu.:6.000   1st Qu.: 8.0  
##  Median :7.000   Median :16.0  
##  Mean   :6.993   Mean   :15.8  
##  3rd Qu.:8.000   3rd Qu.:23.0  
##  Max.   :9.000   Max.   :31.0  
## 
```


Ozone
Solar.R
Wind
Temp
May
June
July
August
September