

# Exercise set 2. Introduction to R

## Data Visualization and Modelling

in Master in Modelling for Sciences and Engineering, UAB. September 2022.

2.0 The exercises 2.1–2.7 should be included in a single script:

- (a) Create and save a new script file with your name “yourname-2.R”. As a comment, inside the script, write your name and your NIU.
- (b) Set the working directory to the location of your script.
- (c) Display the list of objects in your workspace.

2.1 Construct the two matrices:

$$A = \begin{pmatrix} 3 & 5 & 7 \\ 9 & 11 & 13 \\ 15 & 17 & 19 \end{pmatrix} \quad B = \begin{pmatrix} 100 & 106 \\ 102 & 108 \\ 104 & 110 \end{pmatrix}$$

- (a) Combine  $A$  and  $B$  into a new matrix  $C$  using `cbind()` (if possible, if not, why?, add o remove a component to  $B$  if necessary).
- (b) Combine  $A$  and  $B$  into a new matrix  $H$  using `rbind()` (if possible, if not, why?, add o remove a component to  $B$  if necessary).
- (c) Combine  $A$  with  $B^t$  (the transpose of  $B$ ) by rows into the new matrix  $G$ .
- (d) Determine the dimensions of  $G$  using `dim()` function.
- (e) Print out the value of the component in the fifth row second column of the transpose of  $G$ .
- (f) Print out the third row of  $C$ . Print out the second column of  $C$ .
- (g) Print out the submatrix of  $G$  given by its rows 1 and 3 and its columns 2 and 4.
- (h) Create a  $4 \times 4 \times 2$  array  $arr$  using the values 1 to 32.
- (i) Print out the value in row 2, column 4 of the first ‘matrix’ (i.e. the element of index 2,4,1).
- (j) Print out the value in the row 1, column 3 of the second ‘matrix’.
- (k) Add these two values together.
- (l) Print out the row 2 of the second ‘matrix’.

2.2 Create a list named “list.1” that contains de following objects.

- A vector of length 4 called “w”
  - A matrix 3 by 3 called “B”
  - the data.frame `anorexia` (it is in the package MASS) (call it “treatment”)
- (a) Make a sublist of the second and third elements (using square brackets)
  - (b) Print out the matrix of the list (using indexing). Calculate its inverse.
  - (c) Use dollar notation (twice!) to print out the variable `Treat` of the data frame contained in the list `list.1`.
  - (d) Add a new data frame to the list `list.1`.
  - (e) Make a new list of two elements and join the two lists.

2.3 Download the `example2.txt` data and save it into your working directory.

- (a) Read this data into R.
- (b) Print out the data for cases 10 to 18.
- (c) Print out the data for column 2, cases 23, 2, and 5 (in that order).
- (d) Find the mean, standard deviation, minimum and maximum for each variable using the smallest number of commands possible.

2.4 Create a data frame called **patients** with the following data.

id	age	gender	diabetes	status
DP1	25	M	Type1	Poor
DP2	34	M	Type2	Improved
DP3	28	F	Type1	Excellent
DP4	52	M	Type2	Poor
DP5	74	M	Type2	Improved
DP6	63	F	Type1	Poor

- (a) The variables **gender** and **diabetes** should be a factor and **status** an ordered factor.
- (b) Calculate the mean of the variable **age** (Hint use **mean()** function).
- (c) Create a subdata frame with data for the males.
- (d) Determine the maximum age of the males.
- (e) Change the age of patient with id DP3 to 29.

2.5 Import data of monkeypox cases from the European Centre for Disease Prevention and Control (<https://www.ecdc.europa.eu/en/publications-data/data-monkeypox-cases-eueea>)

Variable separator is “,” and NA string is empty space. Call it **monkey**

- (a) Convert the variable **CountryCode** into a factor.
- (b) Make a subdataframe of **monkey** called **monkey.h** with the countries with highest incidence (Netherlands, France, Germany and Spain).
- (c) Using indexing by a logical condition, find the total number of cases of each one of the four countries.
- (d) Applying the character function **substr** (see help) to the variable **DateRep** create in the data frame **monkey.h** three new numerical variables: **Year**, **Month** and **Day**
- (e) Find out when (Day, Month and Year) the first case was reported in each of the four countries in **monkey.h**.
- (f) Use the command **split** (see help) on **monkey** to get a list of data frames by countries in Europe to create a list called **bycountry**.
- (g) Use dollar notation (twice) on **bycountry** to find the total number of cases of Belgium.

2.6 Read into R the files in **readdata.zip**. Use proper read data options. When you print out the data in R they should be the same.

2.7 Find monthly precipitation data for Central England in

<https://www.metoffice.gov.uk/hadobs/hadukp/data/download.html>, download it and create a vector with data corresponding to July from year 2012 to 2020.