Exercise2

Ayelet

3/13/2023

In this exercise, you will submit questions presented on the green slides from our class *in addition* to questions not seen in class.

Remember to submit both the (.rmd) file and the generated pdf file after knitting the document.

The pdf should include both the command and the output of the code chunks.

Don't forget to update the "author" and date information accordingly.

Question 1 - The Iris data set project

1.1 Load the iris built-in dataset (requires the tidyverse library) and describe it - what are the dataset columns? How many observations ?

1.2 Find the maximum sepal length for each species

1.3 Find the row number of the iris with the longest petal, and use it to find its species

1.4 Do the same as in question 1.3, but now for the iris with the longest sepal.

1.5 Create a bar plot (using "barplot()") of the mean sepal length for each species

Question 2 - Some more R

2.1 Print a message asking the user for her/his name and print a personalized (including his/her name) welcoming message

2.2 List all variables in memory (workspace). Save them to a file. Remove all variables from memory. Load the file where the variables have been saved.

2.3 Create 4 vectors:

v1 containing a sequence of numbers from 20 to 50

v2 contain the sum of numbers from 51 to 91

v3 containing 20 random numbers between 1 and 100

v4 containing the 5th to 15th letters of the english alphabet in lower case

2.4 Write a function called "isPrime" that given a number it checks if the number is prime or not. If prime, the function will print "n is prime" (replacing n by the number received), otherwise the function will print all its divisors "n has the following divisors: d1,d2,... " (x is divisor of n if dividing n by x leaves no remainder. E.g. 3 is divisor of 9 bcs 9 %% 3 = 0.).

2.5 Write two functions, bindRows() and bindColumns(), that receive 3 vectors a,b,c, and concatenate them into a matrix. The rows of the matrix returned by bindRows() are the vectors a,b,c. The columns of the matrix returned by bindColumns() are the vectors a,b,c.

For example, given vectors v1= (1,2,3), v2=(4,5,6), v3=(7,8,9) bindRows() will return:

[1,2,3

4,5,6

7,8,9]

bindColumns() will return:

[1, 4, 7

2, 5, 8

3, 6, 9]

2.6.1 Create three vectors: vector "a" containing numeric data, vector "b" containing characters and vector "c" with logical data (TRUE/FALSE). Display the content of the vectors and their type.

2.6.2 Create a data frame from the list (1,2,3). Display its type (with typeOf()) and its class (with class()), and the type and class of its first column.

2.7 Plot as a smooth curve (a line and not dots) the following functions: sin(x), cos(x) and 2sin(x+pi/4) between [0,4pi] (all functions should be displayed on the same plot)

Question 3 - The HR dataset

Note: The questions below refer to the database "HR-Employee-employee" which you should download from the course's website.

3.1 Load the dataset into a dataframe named "df"

3.2 How many employees were surveyed ?

3.3 What are the names of the columns ?

3.4 What's the class of "distanceFromHome" column ?

3.5 List all the JobRoles in alphabetical order

3.6 What is the job role with highest monthly income ?

3.7 How many employees have salaries below the average monthly income ?

3.8.1 What's the education field of employees that have salaries below the average monthly income ?

3.8.2 Is there an education field that every employee with that education field have an average monthly income above average ? If yes, list which education fields

3.8.3 Is there job roles where all employees with this job role have incomes above average? If yes, list which job roles.

3.9 What's the average age of employees that have salaries BELOW the average monthly income ? Is it lower than the average of those who have salaries ABOVE the average monthly income ?

3.10 How many employees have salaries above average ? And below average?

3.11.1 What's the output and meaning of each one of the lines below ?

ind1 <- which(df$MonthlyIncome<meanMonthlyIncome)

ind2 <- df$MonthlyIncome<meanMonthlyIncome

Answer: _____

3.11.2 What's the result of the equation below (ind1 and ind2 as defined in the previous question), and explain the result.

$df EmployeeNumber[ind1] == df EmployeeNumber[ind2]$

3.12 What can we say about the rate of the Job satisfaction ? (select all that is true and justify your answer w code)

 (a) Most employees are satisfied (3) or very satisfied (4)

 (b) Most employees are very satisfied (4)

 (c) The amount of employees that are very satisfied (4) is bigger than the amount of employees that are either not satisfied (1) or medium satisfied (2)

3.13 Prove or contradict the following statement: "In average women have a higher percentage of very high JobSatisfaction (4) than men"

3.14 What is the marital status that is most satisfied with their jobs ? (the % of employees with JobSatisfaction= 4 is the highest across all other levels)

3.15 Plot the monthly income of all employees, adding a line to the median monthly income. What can we say based on the graph ?

3.16 Boxplot the monthly income for each gender separately. What can we say about the difference in salaries between men and women?

3.17 Boxplot the monthly income for each job level separately. What can we conclude regarding the relationship between monthly income and the different job levels?

Good luck!