

DATA SCIENCE FOR ENGINEERS

R Programming

1. Is it mandatory to go through the documents in prerequisite section or it will be covered in lectures?

Answer:

It will be covered in lectures. But it is better if you can go through it before the day's lecture.

2. Why all the numeric value are stored as double?

`typeof(2)` - double

`typeof(-1.09)` -double

`typeof(1.09)` -double

Answer:

That is how R does internally. Please refer to the link below

<https://stat.ethz.ch/R-manual/R-devel/library/base/html/numeric.html>

3. What is the meaning of `names(df)`?

`df3 = df[,!names(df) %in% c("vec3")]`

Answer:

`names(df)` refers to the column names of the dataframe (df).

4. What is a working directory? Which folder needs to be the working directory?

Answer:

Working directory in R is the user home directory.

To read files from a specific location or write files to a specific location you have to set working directory in R by locating that particular file.

5. What are levels in a dataframe?

Answer:

Levels represent the categories in a vector.

While accessing the numerical variable (vec1) the levels will not come into play.

But when accessing the categorical variable (vec2), the levels available in that variable will be displayed along with the accessed element.

6. When trying to convert string into an integer using following command:

```
as.integer('happy')
```

```
[1] NA
```

Warning message:

NAs introduced by coercion

The result is NA with the above-mentioned warning message, so what is the meaning of the above warning message?

Answer:

Not all coercions are possible and if attempted will return “NA” as output. Since it is of character data type, the coercion is not possible.

7. \$ is used in accessing components by name. Does the symbol \$ have a particular role there? (Like, c is used for concatenation).

Answer:

\$ is used to access the variables from a data frame/list.

8. In the deleting of row or column in a dataframe df[-4,] is just removing the data from the view actually it is not getting deleted.

Answer:

The command df[-4,] deletes the 4th row of the dataframe, if there exist 4th row.

9. Whenever we close an R session, the working directory changes to the default one. Is there a way to set it permanently for some days?

Answer:

Kindly find below the link which will be useful to set the working directory.

<https://support.rstudio.com/hc/en-us/articles/200711843-Working-Directories-and-Workspaces>

10. While using print(list\$listelement) to access the elements of a list, the console shows NULL instead of showing the list element.

Answer:

NULL represents the null object in R: it is a reserved word.

NULL is often returned by expressions and functions whose values are undefined. Please check if you have defined the elements of the list properly.

11. In vectors, it is said that elements of a vector must be of the same data type but the following code doesn't give any error > x=c(2.3,4.5,"vinu").

Answer:

The elements of a vector must be of same data type. Since you have given vinu within quotes, by default R considers all other elements in that variable are of character data type. That is why it does not throw any errors.

12. Is it correct to say that the "source(ctrl+shift+S)" as compilation of program and not displaying and that "Source with echo (ctrl+Shift+Enter)" as execution of program?

Answer:

Source will execute all commands in console without displaying them whereas source with echo will execute and print automatically.

13. Is "sum" an inbuilt function? What are some other functions in R?

Answer:

Yes, "sum" is an inbuilt function. Any R- Built-in functions can be used as arguments such as min, max, sqrt, mean, median.

14. How to join data frames pd, pd_new if not having common column names?
How to join the 2 columns Name and Name1?

Answer:

full_join() returns all rows and all columns from both x and y. Where there are not matching values, returns NA for the one missing. You can try this out.

15. How to print the ascii value of a character in R?.

Answer:

Kindly find below the example to print the ASCII value of a character in R.

```
x<-"base"
```

```
asc <-function(x) { strtoi(charToRaw(x),16L) }
```

```
asc(x)
```

16. When running RStudio on a mac, a pop up showing unable to locate R binary by scanning standard locations appears.

Answer:

Try executing the following commands in a terminal:

```
$ echo $PATH
```

```
$ which R
```

```
$ ls -la /usr/bin/R
```

```
$ ls -la /usr/local/bin/R
```

```
$ ls -la /opt/local/bin/R
```

Finally, if you launch RStudio from the command line

```
$ open -a RStudio
```

17. Whenever executing the following code print(object) it shows object not found.

Answer:

The object is nothing but any objects in R such as matrices and data frames. The error might come when you have not run that particular object before printing it.

18. We can edit a particular string element using slicing operator `(df[[2]][2] = "R")`, are there other ways?

Answer:

There are several ways to edit the data frame, some of them have explained in the videos. `(df[[2]][2] = "R")` explains editing an element alone. `$` operator can also be used to edit directly. Every editing commands does the same.

19. While editing, the values are not saved after closing the particular dialogue box.

Answer:

The edited values need to be saved to a variable, otherwise it will not get saved.

20. Can one use `%/%` to do regular division of matrices?

Answer:

You can either use `%/%` or just `/`. Both will do the element wise division. You can be even more specific with decimal points when using `/`. Hence, we suggest using `/` to get specific values.

21. What are the scenarios where one can use Recasting?

Answer:

Ultimately every row corresponds to the observation and every column to variable. The observations should be unique. There should not be any repetitions. By using Recasting method we can have a meaningful data frame.

22. While having option to save entire R program code and why one should save individual data?

Answer:

Let us suppose that you have done some imputation on your data and when coming back after saving that you do not have to run all the codes again, it is enough to load the data into R, and can start from where you have left.

23. Why are `seq(from=1,to=10,by=3)` and `seq(from=1,to=10,length=4)` having the same output?

Answer:

In `seq(from=1,to=10,by=3)`, you are incrementing the elements by 3 and in `seq(from=1,to=10,length=4)`, you are specifying the length. Hence R automatically does by calculating the given length the incremental value will be 3. So the outputs are same.

24. `X = c(2.3,4.5,6.7,8.9)` print (X) output is `[1] 2.3,4.5,6.7,8.9` what is `[1]` here and how does it come?

Answer:

`[1]` represents the layer. Since we have only one vector /row it shows `[1]`.

25. When running the above code it shows error `plot(x,y,type=l)` Error in `plot.xy(xy, type, ...)` : object 'l' not found.

Answer:

In `plot(x,y,type=l)` , l should be with in the quotes ('l')

26. While trying to read a particular file, getting these errors

```
Error in read.table(path = "E:\file") : unused argument (path = "E:\file") > newDF =  
read.table(path = "C:\Users\HP\Anaconda3\Library\share\rstudio\R")
```

```
Error: '\U' used without hex digits in character string starting ""C:\U" >
```

I have saved the file in C:\Users\HP\Anaconda3\Library\share\rstudio\R location..

Answer:

Use forward slash, it will solve the issue

27. While trying to concatenate lists, what will happen if one doesn't have the same number of elements? ie., in my employee list I have 5 elements and the newly added ages lists, that is being concatenated has 4 elements.

Answer:

It just concatenates irrespective of the size of the lists.

28. Example: While assigning values for a as follows:-

1) `a <- 10` 2) `a. <- 20`, 3) `a_ <- 30`, 4) `.a <- 40`

in the above example, all variables are accepted in R, however only `a`, `a.` & `a_` is reflected in the environment, But `.a` is not reflected in the environment window but when u type this variables in console u get the value as below :

```
.a <- 40
```

```
> .a
```

```
[1] 40 .
```

why does this happen?

Answer:

Dot notation is used for hidden objects in R

You can use `ls(all.names = TRUE)` to get all the object names

29. Are the commands used case sensitive?

Answer:

Yes commands are case sensitive.

30. How to import an excel file ?

Answer:

```
install.packages("readxl")
```

```
library(readxl)
```

```
mileage.xlsx<-read_xlsx("Mileage.xlsx")
```

31. How to create a matrix with different values for eg. negative values?

Answer:

```
a<-matrix(c(1,-1,2),nrow=1,ncol=3)
```

32. In the command `apply(A,1 or 2,sum)`, 1 indicates row and 2 indicates column. Does that remain constant for all situations?

Answer:

Yes it remains constant.

33. In `lapply`, whether the determinant function is predefined in R or has to be created?

Answer:

To calculate the determinant, `det` is the predefined function in R

34. The code below describes the problem:

Creating a data frame:

#Vectors defined below

`SerialNum = c(1, 2, 3)`

`Name = c("Kunal", "Indranil", "Harsh")`

`EmployedStatus = c("Accenture", "Capgemini", "No")`

#Data Frame creation syntax

`Info = data.frame(SerialNum, Name, EmployedStatus)`

`print(Info)`

#Assigning new value to a data frame using direct assignment

`Info[[2]][1] = "Kunal Gaurav"`

`print(Info)`

When running the code, it's replacing the value "Kunal" by "N/A":

➤ `Info[[2]][1] = "Kunal Gaurav"`

Warning message: In ``[<-.factor`(*tmp*, 1, value = c(NA, 2L, 1L))` : invalid factor level, NA generated.

Answer:

while creating a dataframe, set `stringsAsFactors = F`.

`SerialNum = c(1, 2, 3)`

`Name = c("Kunal", "Indranil", "Harsh")`

`EmployedStatus = c("Accenture", "Capgemini", "No")`

#Data Frame creation syntax

`Info = data.frame(SerialNum, Name, EmployedStatus,stringsAsFactors = F)`

`print(Info)` #Assigning new value to a data frame using direct assignment

```
Info[[2]][2] = "Kunal Gaurav"  
Info
```

35. How to know when a loop breaks?

Answer:

Please specify the condition, once the condition fails it will break out of the loop.
Break function can be used to break out of the loop.

36. `ID = c(1, 2, 3)` `emp.name = c("Prachi", "Mamta", "Rakshita")` Here `ID` and `emp.name` are variable names, what is `c`?

Answer:

`c` is used to combine values into a vector. Please find the below link to know more about this:

<https://www.rdocumentation.org/packages/base/versions/3.5.2/topics/c>

37. `rm(list=ls(b,end_point))`

shows below error,

please guide

Error in `as.environment(pos)` : invalid 'pos' argument.

Answer:

To clear a specific variable : `rm(variablename)`

To clear all the variables : `rm(list=ls())`

38. How to increase the number of rows and column in edit table?

Answer:

It is like a spreadsheet, you can add the values to the rows and columns.

39. While trying to execute the left join got an error message:

Warning message: Column `Name` joining factors with different levels, coercing to character vector.

Observed that the `dataframe1 (pd)` and `dataframe2(pd_new)` has different factor levels. How to get rid of this error?

Answer:

In R, You cannot add/ delete existing factors / categories in a data frame. If you want to do so, you will have to convert factors to character data type before doing that. Similarly, while joining data frames it automatically converts factor data type to character data type since `Name` has different levels. Please find below the snippet for your reference.

40. What is the difference between 'print(a)' or 'print(c(a))' where a is any vector or list.

Answer:

print()- it prints the argument

print(c(a)) - it converts the data frame to a vector or list

41. When executing the following code:

```
id=c(29,30,31,32)
➤ names=c("bruh","blah","bleh","bloh")
➤ score=c(95,85,84,87)
➤ gender=c("f","m","f","m")
➤ df=data.frame(id,names,score,gender)
➤ print(df)
id names score gender
1 29 bruh 95 f
2 30 blah 85 m
3 31 bleh 84 f
4 32 bloh 87 m
➤ df1=df[4,3]
➤ print(df1)
[1] 87
```

Expecting the output of df[4,3] as the value of 4th row and 3rd column, which is, "f".

But the output turned out to be the value of 3rd row and 4th column, i.e. 87.

How does this work?

Answer:

The elements along the horizontal axis are considered as rows and the elements along the vertical axis are considered as columns.

42. Is it necessary to set the working directory every time to write a program or code?.

Answer:

While reading the data, it is necessary to set the working directory.

43. Inline function showing error.

```
func = function(x) 2x^2+3x+17
```

Error: unexpected symbol in "func = function(x) 2x"

How to rectify the error?

Answer:

Please include multiplication operator (*) in between 2 and x.

44. Is it necessary to use " " for numbers? If " " is used will it automatically takes it as a character?

Answer:

For numerical variable double quotes is not required. If double quotes are added in the numerical variable, R automatically converts to character.

45. Why does `as.numeric("a")` return NA? Why doesn't it return its ASCII value?.

Answer:

R does not return ASCII value.

Coercions from character to numeric variable is not possible so it returns NA

46. What is `anti_join`? where it is used?

Answer:

`anti_join` returns the row of the first dataframe where it cannot find a match in the second dataframe.

It can be used if you want to join 2 data frames based on certain condition.

47. What is tabular data?

Answer:

Tabular data is data arranged in rows and columns.

48. What are factor variables?

Answer:

Factor variable is a categorical variable.

For Example: To predict the price of the car based on Fuel type, Kilometer, Mileage. In this case Fuel type is a categorical variable with two levels Petrol, Diesel.

In predictive modelling module, we will be explaining more with examples.

49. In the command : `df3=df[,!name(df)%in%c("vec3")]`
what is the meaning of : `name(df)%in%c`.

Answer:

`%in%` - it basically returns a vector of the positions of (first) matches of its first argument in its second.

50. Why is dataframe important for storing data? Why can't we store using hashtable?

Answer:

You can also store the data using hashtable.

51. Unable to create dataframe from a file.

Getting below error.

Even tried setting working directory still c:/Rworkspace.

```
newdf=read.table(file="c:/Rworkspace/dataframe.csv")
```

Error in file(file, "rt") : cannot open the connection In addition: Warning message:

In file(file, "rt") : cannot open file 'c:/Rworkspace/dataframe.csv': No such file or directory.

Answer:

Kindly set the particular file/folder as working directory.

52. While installing the packages it says, package 'plyr' is not available (for R version 3.5.2) 'lib = "dplyr"' is not writable.

Answer:

Kindly install dplyr.

```
install.packages("dplyr")
```

53. When we enter `as.character(2)` the result is `2+0i` why?.

Answer:

`as.character(2)` returns "2".

`as.complex(2)` returns `2+0i`.

54. What is the difference between list and vector.

Answer:

vector can hold numeric, categorical or logical values. Elements in the vector will be of same data type.

list - it allows to combine all the data types and create as a list.

55. How do we rename the column names of a dataframe?

Answer:

Kindly use the `colnames` command and specify the column number.

56. The argument taken for margin for adding rows and columns are '1' and '2'. Is there any explanation for this ?

Answer:

To calculate row sum please specify `margin=1` and for column sum, `margin= 2`.

Apart from calculating sum we can also find mean and other operations.

57. When trying manipulation using `stringsAsFactors` it was not removing the invalid factor message.

Answer:

Kindly set the `stringAsFactors = F`.

58. Is it correct to say that inner join is similar to left join?

Answer:

Inner join : returns only the rows in which the left table have matching keys in the right table.

left join: returns all rows from the left table and any rows with matching keys from the right table.

59. What does "in" keyword do in for loop?.

Answer:

These are the basic control-flow constructs of the R language. They function in much the same way as control statements in any Algol-like language. They are all reserved words.

To know more about please use the `help` command in R

60. The `mutate` command is adds a column in a Dataframe. What is the difference between `cbind` and `mutate` command.

Answer:

`cbind` is used to combine the columns in the dataframe.

`mutate()` adds new variables and preserves existing ones.

61. Can we add any row without entry of a specific column using `rbind` command ? If yes then what is the syntax?.

Answer:

You can add a row without entry for a specific column using `""` . It will be considered as NA.

Please find attached snapshot for your reference.

```

> vector1=c(1,2,3)
> vector2=c("R","JAvA","Python")
> Vector3=c("DataScience" ,"WEbService" ,"NetworkAutomation")
> data=data.frame(vector1,vector2,Vector3,stringsAsFactors = F)
> print(data)
  vector1 vector2      Vector3
1        1      R    DataScience
2        2    JAvA    WEbService
3        3 Python NetworkAutomation
> data=rbind(data,data.frame(vector1=4,vector2="",Vector3="Datasci"))
> data
  vector1 vector2      Vector3
1        1      R    DataScience
2        2    JAvA    WEbService
3        3 Python NetworkAutomation
4        4              Datasci

```

62. Why do we use c before () in any vector or data frame?

Answer:

c() - combines the values to a vector or a list.

63. install.packages("reshape2")

trying URL 'https://cloud.r-project.org/bin/windows/contrib/3.4/reshape2_1.4.3.zip'

Content type 'application/zip' length 611622 bytes (597 KB)

downloaded 597 KB

package 'reshape2' successfully unpacked and MD5 sums checked

➤ library("reshape2")

Error: unexpected input in "library(""

Answer:

To load the library please use the following command:

library(reshape2)