**What is R Programming?**

R is a programming language meant for statistical analysis and creating graphs for this purpose.Instead of data types, it has data objects which are used for calculations. It is used in the fields of data mining, Regression analysis, Probability estimation etc., using many packages available in it.

**What are the different data objects in R?**

There are 6 data objects in R. They are vectors, lists, arrays, matrices, data frames and tables.

**What makes a valid variable name in R?**

A valid variable name consists of letters, numbers and the dot or underline characters. The variable name starts with a letter or the dot not followed by a number.

**What is the main difference between an Array and a matrix?**

A matrix is always two dimensional as it has only rows and columns. But an array can be of any number of dimensions and each dimension is a matrix. For example a 3x3x2 array represents 2 matrices each of dimension 3x3.

**Which data object in R is used to store and process categorical data?**

The Factor data objects in R are used to store and process categorical data in R.

**How can you load and use csv file in R?**

A csv file can be loaded using the read.csv function. R creates a data frame on reading the csv files using this function.

**How do you get the name of the current working directory in R?**

The command getwd() gives the current working directory in the R environment.

**What is R Base package?**

This is the package which is loaded by default when R environment is set. It provides the basic functionalities like input/output, arithmetic calculations etc. in the R environment.

**How do you access the element in the 2nd column and 4th row of a matrix named M?**

The expression M[4,2] gives the element at 4th row and 2nd column.

**What is recycling of elements in a vector? Give an example.**

When two vectors of different length are involved in a operation then the elements of the shorter vector are reused to complete the operation. This is called element recycling. Example - v1 <- c(4,1,0,6) and V2 <- c(2,4) then v1\*v2 gives (8,4,0,24). The elements 2 and 4 are repeated.

**What are different ways to call a function in R?**

We can call a function in R in 3 ways. First method is to call by using position of the arguments. Second method id to call by using the name of the arguments and the third method is to call by default arguments.

**What is lazy function evaluation in R?**

The lazy evaluation of a function means, the argument is evaluated only if it is used inside the body of the function. If there is no reference to the argument in the body of the function then it is simply ignored.

**How do you install a package in R?**

To install a package in R we use the below command.

install.packages("package Name")

**Can we update and delete any of the elements in a list?**

We can update any of the element but we can delete only the element at the end of the list.

**Give the general expression to create a matrix in R.**

The general expression to create a matrix in R is - matrix(data, nrow, ncol, byrow, dimnames)

**What is reshaping of data in R?**

In R the data objects can be converted from one form to another. For example we can create a data frame by merging many lists. This involves a series of R commands to bring the data into the new format. This is called data reshaping.

**How to get a list of all the packages installed in R ?**

Use the command

installed.packages()

**What is expected from running the command - strsplit(x,"e")?**

It splits the strings in vector x into substrings at the position of letter e.

**Give a R script to extract all the unique words in uppercase from the string - "The quick brown fox jumps over the lazy dog".**

x <- "The quick brown fox jumps over the lazy dog"

split.string <- strsplit(x, " ")

extract.words <- split.string[[1]]

result <- unique(tolower(extract.words))

print(result)

**Vector v is c(1,2,3,4) and list x is list(5:8), what is the output of v\*x[1]?**

Error in v \* x[1] : non-numeric argument to binary operator

**Vector v is c(1,2,3,4) and list x is list(5:8), what is the output of v\*x[[1]]?**

[1] 5 12 21 32s

**What does unlist() do?**

It converts a list to a vector.

**Give the R expression to get 26 or less heads from a 51 tosses of a coin using pbinom.**

x <- pbinom(26,51,0.5)

print(x)

X is the vector c(5,9.2,3,8.51,NA), What is the output of mean(x)?

NA

**Give a function in R that replaces all missing values of a vector x with the sum of elements of that vector?**

function(x) { x[is.na(x)] <- sum(x, na.rm = TRUE); x }

**What is the use of apply() in R?**

It is used to apply the same function to each of the elements in an Array. For example finding the mean of the rows in every row.

**Is an array a matrix or a matrix an array?**

Every matrix can be called an array but not the reverse. Matrix is always two dimensional but array can be of any dimension.

**How to find the help page on missing values?**

?NA

**How do you set the path for current working directory in R?**

setwd("Path")

**What is the difference between "%%" and "%/%"?**

"%%" gives remainder of the division of first vector with second while "%/%" gives the quotient of the division of first vector with second.

**What does col.max(x) do?**

Find the column has the maximum value for each row.

**How do you remove a vector from the R workspace?**

rm(x)

**List the data sets available in package "MASS"**

data(package = "MASS")

**List the data sets available in all available packages.**

data(package = .packages(all.available = TRUE))

**What is the use of the command - install.packages(file.choose(), repos=NULL)?**

It is used to install a r package from local directory by browsing and selecting the file.

**Give the command to check if the element 15 is present in vector x.**

15 %in% x

**What is the difference between subset() function and sample() function in R?**

The subset() functions is used to select variables and observations. The sample() function is used to choose a random sample of size n from a dataset.

**How do you check if "m" is a matrix data object in R?**

is.matrix(m) should retrun TRUE.

**What is the output for the below expression all(NA==NA)?**

[1] NA

**How to obtain the transpose of a matrix in R?**

The function t() is used for transposing a matrix. Example - t(m) , where m is a matrix.

**What is the use of "next" statement in R?**

The "next" statement in R programming language is useful when we want to skip the current iteration of a loop without terminating it.

**When should you apply “next” statement in R? When is it appropriate to use the “next” statement in R?**

A data scientist will use next to skip an iteration in a loop. As an example:

1. > val1 <- 1:30
2. > **for**(val in val1){
3. + **if**(val == 25){
4. + next
5. + }
6. + **print**(val)
7. + }

This piece of code will iterate through the numbers from 1 to 30. It will skip 25 as we have used the next statement to skip the iteration from which we move on to the next value. We will obtain the output from 1-24 and 26-30.

**What is the value of equation1(3) for the following R code?**

1. > num <- 4
2. > equation1 <- **function** (val)
3. + {
4. + num <- 3
5. + num^3 + **g** (val)
6. + }
7. > equation2 <- **function** (val)
8. + {
9. + val\*num
10. + }
11. }

For the above code snippet, we obtain the output as 39.

**A csv file consists of missing values that are represented by hashtags (“#”) and ampersands (“&”). How can you read this type of a csv file in R?**

We can read this type of data with the following line of code –

1. **csv2**(‘data-file.csv’,header=FALSE,sep=’,’,na.strings=**c**(‘#’,’&’))

**You have two tables “employee\_salary” and “employee\_experience”. The first table consists of two columns “Name” and “Salary”. The second table consists of “Name” and “Experience” columns. How will you merge these two tables to create a single table that does not have any redundant column in it?**

In order to join these two tables, we will have to perform their cross-product. R provides the function “merge” that can be used for the same.

**merge**(employee\_salary,employee\_experience,by=NULL)

**Suppose that I want to know the values in c(1, 2, 6, 3, 19) that are not present in c(2, 6, 14, 3, 15). How can you carry this out using built-in function as well as without it?**

There are two methods to execute this problem –

* By using setdiff() function – setdiff(c(1, 2, 6, 3, 19), c(2, 6, 14, 3, 15)) and,
* Through the %in% as – c(1, 4, 5, 9, 10)[!c(1, 4, 5, 9, 10) %in% c(1, 5, 10, 11, 13)

**You have a dataset from which you want to extract a subset. For example, for a data consisting of employee details, you want to create a subset of employees who are above 30 years and who make salary less than 10,000. How will you execute this in R? The salary dataframe be denoted by the variable ‘age’ and the data is contained within the variable ‘employee\_data’.**

To create a subset of employees older than 30, we will make use of the filter() function as follows –

1. **filter**(employee\_data, age >30 1, salary < 10000)