1. **Compare R and Python programming languages for Predictive Modelling.**

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **Python is Better** | **R Language is Better** |  |
| Model Building | Both are Similar | Both are Similar |  |
| Model Interpretability | Not better than R. | R is better |  |
| Production | Python is Better | Not better than Python |  |
| Community Support | Not better than R. | R  has good community support over Python. |  |
| Data Science Libraries | Both are similar. | Both are similar |  |
| Data Visualizations | Not better than R | R has good data visualizations libraries and tools. |  |
| Learning Curve | Learning Python is easier than learning R. | R has a steep learning curve. |  |
| **Python vs R for Predictive Modelling** | | |  |



[**CLICK HERE**](https://docs.google.com/forms/d/1QButFq8cP2wjT-jwNulA1qsFYTBrZM-yAOUHJHOHKMk/viewform) to get the data scientist salary report delivered to your inbox!

**2) Explain about data import in R language**

R Commander is used to import data in R language. To start the R commander GUI, the user must type in the command Rcmdr into the console. There are 3 different ways in which data can be imported in R language-

•           Users can select the data set in the dialog box or enter the name of the data set (if they know).

•           Data can also be entered directly using the editor of R Commander via Data->New Data Set. However, this works well when the data set is not too large.

•           Data can also be imported from a URL or from a plain text file (ASCII), from any other statistical package or from the clipboard.

**3) Two vectors X and Y are defined as follows – X <- c(3, 2, 4) and Y <- c(1, 2). What will be output of vector Z that is defined as Z <- X\*Y.**

In R language when the vectors have different lengths, the multiplication begins with the smaller vector and continues till all the elements in the larger vector have been multiplied.

The output of the above code will be –

Z <- (3, 4, 4)

**4) How missing values and impossible values are represented in R language?**

NaN (Not a Number) is used to represent impossible values whereas NA (Not Available) is used to represent missing values. The best way to answer this question would be to mention that deleting missing values is not a good idea because the probable cause for missing value could be some problem with data collection or programming or the query. It is good to find the root cause of the missing values and then take necessary steps handle them.

**5) R language has several packages for solving a particular problem. How do you make a decision on which one is the best to use?**

CRAN package ecosystem has more than 6000 packages. The best way for beginners to answer this question is to mention that they would look for a package that follows good software development principles. The next thing would be to look for user reviews and find out if other data scientists or analysts have been able to solve a similar problem.

[](https://goo.gl/Estvuq)

**6) Which function in R language is used to find out whether the means of 2 groups are equal to each other or not?**

t.tests ()

**7) What is the best way to communicate the results of data analysis using R language?**

The best possible way to do this is combine the data, code and analysis results in a single document using knitr for reproducible research. This helps others to verify the findings, add to them and engage in discussions. Reproducible research makes it easy to redo the experiments by inserting new data and applying it to a different problem.

**8) How many data structures does R language have?**

R language has Homogeneous and Heterogeneous data structures. Homogeneous data structures have same type of objects – Vector, Matrix ad Array. Heterogeneous data structures have different type of objects – Data frames and lists.

**9) What is the value of f (2) for the following R code?**

b <- 4

f <- function (a)

{

b <- 3

b^3 + g (a)

}

g <- function (a)

{

a\*b

}

The answer to the above code snippet is 35. The value of “a” passed to the function is 2 and the value for “b” defined in the function f (a) is 3. So the output would be 3^3 + g (2). The function g is defined in the global environment and it takes the value of b as 4(due to lexical scoping in R) not 3 returning a value 2\*4= 8 to the function f. The result will be 3^3+8= 35.

**10) What is the process to create a table in R language without using external files?**

MyTable= data.frame ()

 edit (MyTable)

The above code will open an Excel Spreadsheet for entering data into MyTable.

Learn [Data Science in R Programming](https://www.dezyre.com/data-science-in-R-programming-online-training/37) to land a top gig as an Enterprise Data Scientist!

**11) Explain about the significance of transpose in R language**

Transpose t () is the easiest method for reshaping the data before analysis.

**12) What are with () and BY () functions used for?**

With () function is used to apply an expression for a given dataset and BY () function is used for applying a function each level of factors.

**13) dplyr package is used to speed up data frame management code. Which package can be integrated with dplyr for large fast tables?**

data.table

**14) In base graphics system, which function is used to add elements to a plot?**

boxplot () or text ()

**15) What are the different type of sorting algorithms available in R language?**

Bucket Sort

Selection Sort

Quick Sort

Bubble Sort

Merge Sort

**15) What is the command used to store R objects in a file?**

save (x, file=”x.Rdata”)

**16) What is the best way to use Hadoop and R together for analysis?**

HDFS can be used for storing the data for long-term. MapReduce jobs submitted from either Oozie, Pig or Hive can be used to encode, improve and sample the data sets from HDFS into R. This helps to leverage complex analysis tasks on the subset of data prepared in R.

**17) What will be the output of log (-5.8) when executed on R console?**

Executing the above on R console will display a warning sign that NaN (Not a Number) will be produced because it is not possible to take the log of negative number.

**18) How is a Data object represented internally in R language?**

unclass (as.Date (“2016-10-05″))

**19) What will be the output of the below code -**

printmessage <- function (a) {

                  if (is.na (a))

                             print ("a is a missing value!")

                     else if (a < 0)

                             print ("a is less than zero")

                     else

                            print ("a is greater than or equal to zero")

                    invisible (a)

}

printmessage (NA)

The output for the above R programming code will be “a is a missing value.” The function is.na () is used to check if the input passed is a missing value.

**20) Which package in R supports the exploratory analysis of genomic data?**

adegenet

**21) What is the difference between data frame and a matrix in R?**

Data frame can contain heterogeneous inputs while a matrix cannot. In matrix only similar data types can be stored whereas in a data frame there can be different data types like characters, integers or other data frames.

**22) How can you add datasets in R?**

rbind () function can be used add datasets in R language provided the columns in the datasets should be same.

**23) How do you split a continuous variable into different groups/ranks in R?**

**24) What are factor variable in R language?**

Factor variables are categorical variables that hold either string or numeric values. Factor variables are used in various types of graphics and particularly for statistical modelling where the correct number of degrees of freedom is assigned to them.

**25) What is the memory limit in R?**

8TB is the memory limit for 64-bit system memory and 3GB is the limit for 32-bit system memory.

**26) What are the data types in R on which binary operators can be applied?**

Scalars, Matrices ad Vectors.

**27) How do you create log linear models in R language?**

Using the loglm () function

**28) What will be the class of the resulting vector if you concatenate a number and NA?**

number

**29) What is meant by K-nearest neighbour?**

K-Nearest Neighbour is one of the simplest machine learning classification algorithms that is a subset of supervised learning based on lazy learning. In this algorithm the function is approximated locally and any computations are deferred until classification.

**30) What will be the class of the resulting vector if you concatenate a number and a character?**

character

**31) Write code to build an R function powered by C?**

**32) If you want to know all the values in c (1, 3, 5, 7, 10) that are not in c (1, 5, 10, 12, 14). Which in-built function in R can be used to do this? Also, how this can be achieved without using the in-built function.**

Using in-built function - setdiff(c (1, 3, 5, 7, 10), c (1, 5, 10, 11, 13))

Without using in-built function - c (1, 3, 5, 7, 10) [! c (1, 3, 5, 7, 10) %in% c (1, 5, 10, 11, 13).

**33) How can you debug and test R programming code?**

 R code can be tested using Hadley’s testthat package.

**34) What will be the class of the resulting vector if you concatenate a number and a logical?**

number

**35) Write a function in R language to replace the missing value in a vector with the mean of that vector.**

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

**36) What happens if the application object is not able to handle an event?**

The event is dispatched to the delegate for processing.

**37) Differentiate between lapply and sapply.**

If the programmers want the output to be a data frame or a vector, then sapply function is used whereas if a programmer wants the output to be a list then lapply is used. There one more function known as vapply which is preferred over sapply as vapply allows the programmer to specific the output type. The disadvantage of using vapply is that it is difficult to be implemented and more verbose.

**38) Differentiate between seq (6) and seq\_along (6)**

Seq\_along(6) will produce a vector with length 6 whereas seq(6) will produce a sequential vector from 1 to 6  c( (1,2,3,4,5,6)).

**39) How will you read a .csv file in R language?**

read.csv () function is used to read a .csv file in R language. Below is a simple example –

filcontent <-read.csv (sample.csv)

print (filecontent)

**40) How do you write R commands?**

The line of code in R language should begin with a hash symbol (#).

**41) How can you verify if a given object “X” is a matric data object?**

If the function call is.matrix(X ) returns TRUE then X can be termed as a matrix data object.

**42) What do you understand by element recycling in R?**

If two vectors with different lengths perform an operation –the elements of the shorter vector will be re-used to complete the operation. This is referred to as element recycling.

Example – Vector A <-c(1,2,0,4) and Vector B<-(3,6) then the result of A\*B will be ( 3,12,0,24). Here 3 and 6 of vector B are repeated when computing the result.

**43) How can you verify if a given object “X” is a matrix data object?**

If the function call is.matrix(X) returns true then X can be considered as a matrix data object otheriwse not.

**44) How will you measure the probability of a binary response variable in R language?**

Logistic regression can be used for this and the function glm () in R language provides this functionality.

**45) What is the use of sample and subset functions in R programming language?**

Sample () function can be used to select a random sample of size ‘n’ from a huge dataset.

Subset () function is used to select variables and observations from a given dataset.

**46) There is a function fn(a, b, c, d, e) a + b \* c - d / e. Write the code to call fn on the vector c(1,2,3,4,5) such that the output is same as fn(1,2,3,4,5).**

do.call (fn, as.list(c (1, 2, 3, 4, 5)))

[](https://goo.gl/ZV5sYP)

**47) How can you resample statistical tests in R language?**

Coin package in R provides various options for re-randomization and permutations based on statistical tests. When test assumptions cannot be met then this package serves as the best alternative to classical methods as it does not assume random sampling from well-defined populations.

**48) What is the purpose of using Next statement in R language?**

If a developer wants to skip the current iteration of a loop in the code without terminating it then they can use the next statement. Whenever the R parser comes across the next statement in the code, it skips evaluation of the loop further and jumps to the next iteration of the loop.

**49) How will you create scatterplot matrices in R language?**

A matrix of scatterplots can be produced using pairs. Pairs function takes various parameters like formula, data, subset, labels, etc.

The two key parameters required to build a scatterplot matrix are –

* formula- A formula basically like ~a+b+c . Each term gives a separate variable in the pairs plots where the terms should be numerical vectors. It basically represents the series of variables used in pairs.
* data- It basically represents the dataset from which the variables have to be taken for building a scatterplot.

**50) How will you check if an element 25 is present in a vector?**

There are various ways to do this-

1. It can be done using the match () function- match () function returns the first appearance of a particular element.
2. The other is to use %in% which returns a Boolean value either true or false.
3. Is.element () function also returns a Boolean value either true or false based on whether it is present in a vector or not.

**51) What is the difference between library() and require() functions in R language?**

There is no real difference between the two if the packages are not being loaded inside the function. require () function is usually used inside function and throws a warning whenever a particular package is not found. On the flip side, library () function gives an error message if the desired package cannot be loaded.

**52) What are the rules to define a variable name in R programming language?**

A variable name in R programming language can contain numeric and alphabets along with special characters like dot (.) and underline (-). Variable names in R language can begin with an alphabet or the dot symbol. However, if the variable name begins with a dot symbol it should not be a followed by a numeric digit.

**53) What do you understand by a workspace in R programming language?**

The current R working environment of a user that has user defined objects like lists, vectors, etc. is referred to as Workspace in R language.

**54) Which function helps you perform sorting in R language?**

Order ()

**55) How will you list all the data sets available in all R packages?**  
Using the below line of code-  
data(package = .packages(all.available = TRUE))  
**56) Which function is used to create a histogram visualisation in R programming language?**  
Hist()  
**57)  Write the syntax to set the path for current working directory in R environment.**  
Setwd(“dir\_path”)

**58) How will you drop variables using indices in a data frame?**

Let’s take a dataframe df<-data.frame(v1=c(1:5),v2=c(2:6),v3=c(3:7),v4=c(4:8))

df

##   v1 v2 v3 v4

## 1  1  2  3  4

## 2  2  3  4  5

## 3  3  4  5  6

## 4  4  5  6  7

## 5  5  6  7  8

Suppose we want to drop variables v2 & v3 , the variables v2 and v3 can be dropped using negative indicies as follows-

df1<-df[-c(2,3)]

df1

##   v1 v4

## 1  1  4

## 2  2  5

## 3  3  6

## 4  4  7

## 5  5  8

**59) What will be the output of runif (7)?**

It will generate 7 randowm numbers between 0 and 1.

**60) What is the difference between rnorm and runif functions ?**

rnorm function generates "n" normal random numbers based on the mean and standard deviation arguments passed to the function.

**Syntax of rnorm function -**

rnorm(n, mean = , sd = )

runif function generates "n" unform random numbers in the interval of minimum and maximum values passed to the function.

**Syntax of runif function -**

 runif(n, min = , max = )

**61) What will be the output on executing the following R programming code –**

mat<-matrix(rep(c(TRUE,FALSE),8),nrow=4)

sum(mat)

 8

**62) How will you combine multiple different string like “Data”, “Science”, “in” ,“R”, “Programming” as a single string “Data\_Science\_in\_R\_Programmming” ?**

paste(“Data”, “Science”, “in” ,“R”, “Programming”,sep="\_")

**63) Write a function to extract the first name from the string “Mr. Tom White”.**

substr (“Mr. Tom White”,start=5, stop=7)

**64) Can you tell if the equation given below is linear or not ?**

**Emp\_sal= 2000+2.5(emp\_age)2**

Yes it is a linear equation as the coefficients are linear.

**65) What will be the output of the following R programming code ?**

**var2<- c("I","Love,"DeZyre")**

**var2**

 It will give an error.

**66) What will be the output of the following R programming code?**

**x<-5**

**if(x%%2==0)**

**print("X is an even number")**

**else**

**print("X is an odd number")**

Executing the above code will result in an error as shown below -

## Error: :4:1: unexpected 'else'

## 3:   print("X is an even number")

## 4: else

##    ^

 R programming language does not know if the else related to the first ‘if’ or not as the first if() is a complete command on its own.

**67) I have a string "**[**contact@dezyre.com**](mailto:contact@dezyre.com)**". Which string function can be used to split the string into two different strings “contact@dezyre” and “com” ?**

This can be accomplished using  the strsplit function which splits a string based on the identifier given in the function call. The output of strsplit() function is a list.

strsplit("[contact@dezyre.com](mailto:contact@dezyre.com)",split = ".")

Output of the strsplit function is -

## [[1]]

## [1] " contact@dezyre" "com"

**68) What is R Base package?**

R Base package is the package that is loaded by default whenever R programming environent is loaded .R base package provides basic fucntionalites in R environment like arithmetic calcualtions, input/output.

**69)  How will you merge two dataframes in R programming language?**

Merge () function is used to combine two dataframes and it identifies common rows or columns between the 2 dataframes. Merge () function basically finds the intersection between two different sets of data.

Merge () function in R language takes a long list of arguments as follows –

Syntax for using Merge function in R language -

 merge (x, y, by.x, by.y, all.x  or all.y or all )

* X represents the first dataframe.
* Y represents the second dataframe.
* by.X- Variable name in dataframe X that is common in Y.
* by.Y- Variable name in dataframe Y that is common in X.
* all.x - It is a logical value that specifies the type of merge. all.X should be set to true, if we want all the observations from dataframe X . This results in Left Join.
* all.y - It is a logical value that specifies the type of merge. all.y should be set to true , if we want all the observations from dataframe Y . This results in Right Join.
* all – The default value for this is set to FALSE which means that only matching rows are returned resulting in Inner join. This should be set to true if you want all the observations from dataframe X and Y resulting in Outer join.

**70) Write the R programming code for an array of words so that the output is displayed in decreasing frequency order.**

R Programming Code to display output in decreasing frequency order -

tt <- sort(table(c("a", "b", "a", "a", "b", "c", "a1", "a1", "a1")), dec=T)

depth <- 3

tt[1:depth]

**Output -**

1) a a1 b

2) 3 3 2

**71) How to check the frequency distribution of a categorical variable?**

The frequency distribution of a categorical variable can be checked using the table function in R language. Table () function calculates the count of each categories of a categorical variable.

gender=factor(c(“M”,”F”,”M”,”F”,”F”,”F”))

table(sex)

**Output of the above R Code –**

Gender

F  M

4  2

Programmers can also calculate the % of values for each categorical group by storing the output in a dataframe and applying the column percent function as shown below -

t = data.frame(table(gender))  
t$percent= round(t$Freq / sum(t$Freq)\*100,2)

|  |  |  |
| --- | --- | --- |
| **Gender** | **Frequency** | **Percent** |
| F | 4 | 66.67 |
| M | 2 | 33.33 |

**72) What is the procedure to check the cumulative frequency distribution of any categorical variable?**

The cumulative frequency distribution of a categorical variable can be checked using the cumsum () function in R language.

**Example –**

gender = factor(c("f","m","m","f","m","f"))  
y = table(gender)  
cumsum(y)

**Output of the above R code-**

Cumsum(y)

f m

3 3

**73) What will be the result of multiplying two vectors in R having different lengths?**

The multiplication of the two vectors will be performed and the output will be displayed with a warning message like – “Longer object length is not a multiple of shorter object length.” Suppose there is a vector a<-c (1, 2, 3) and vector b <- (2, 3) then the multiplication of the vectors a\*b will give the resultant as 2 6 6 with the warning message. The multiplication is performed in a sequential manner but since the length is not same, the first element of the smaller vector b will be multiplied with the last element of the larger vector a.

**74) R programming language has several packages for data science which are meant to solve a specific problem, how do you decide which one to use?**

CRAN package repository in R has more than 6000 packages, so a data scientist needs to follow a well-defined process and criteria to select the right one for a specific task. When looking for a package in the CRAN repository a data scientist should list out all the requirements and issues so that an ideal R package can address all those needs and issues.

The best way to answer this question is to look for an R package that follows good software development principles and practices. For example, you might want to look at the quality documentation and unit tests. The next step is to check out how a particular R package is used and read the reviews posted by other users of the R package. It is important to know if other data scientists or data analysts have been able to solve a similar problem as that of yours. When you in doubt choosing a particular R package, I would always ask for feedback from R community members or other colleagues to ensure that I am making the right choice.

**75) How can you merge two data frames in R language?**

Data frames in R language can be merged manually using cbind () functions or by using the merge () function on common rows or columns.

**76)  Explain the usage of which() function in R language.**

which() function determines the postion of elemnts in a logical vector that are TRUE. In the below example, we are finding the row number wherein the maximum value of variable v1 is recorded.

mydata=data.frame(v1 = c(2,4,12,3,6))  
which(mydata$v1==max(mydata$v1))  
It returns 3 as 12 is the maximum value and it is at 3rd row in the variable x=v1.

**77) How will you convert a factor variable to numeric in R language ?**

A factor variable can be converted to numeric using the as.numeric() function in R language. However, the variable first needs to be converted to character before being converted to numberic because the as.numeric() function in R does not return original values but returns the vector of the levels of the factor variable.

X <- factor(c(4, 5, 6, 6, 4))  
X1 = as.numeric(as.character(X))

**78) Explain the significance of R programming language for Data Science ?**

i) Most of the calculations can be done with the help of vector so it is easy for data scientists to add functions to a single vector without having to put them in a loop.

ii) A turning complete language that can be used for any kind of data science task whether it is in the field of genetics, statistics or biology.

iii) Being an interpreted language , it does not require any compiler-making development of code easier.

**79)  What is power analysis ?**

Power analysis is the process used to determine the effect of a given sample size and is generally used for experimental design.Pwr package in R is used for power analysis.

**80) Explain the usage of abline() function.**

abline function in R used to add reference line to a graph. Below is the syntax of using abline function -

abline(h=yvalues, v=xvalues)

**81)  What is the usage of lattice package in R ?**

Lattice package helps enhance base R graphics by providing better defaults and helps easily display multi-variate relationships.

**1. Compare R & Python**

|  |  |
| --- | --- |
| **R  programming Language** | **Python programming language** |
| Model Building is similar to  Python | Model Building is similar to R. |
| Model Interpretability is  good | Model Interpretability is  not good |
| Production is not better than Python. | Production is good |
| R has good community support over Python. | Community Support  is not better than R |
| Data Science Libraries are same as Python. | Data Science Libraries are same as R. |
| R has good data visualizations libraries and tools | Data visualization is not better than R |
| R has a steep learning curve. | Learning Curve in Python is easier than learning R. |

**2. Explain the data import in R language.**

R provides to import data in R language. To begin with the R commander GUI, user should type the commands in the command Rcmdr into the console. Data can be imported in R language in 3 ways such as:

* Select the data set in the dialog box or enter the name of the data set as required.
* Data is entered directly using the editor of R Commander via Data->New Data Set. This works good only when the data set is not too large.
* Data can also be imported from a URL or from plain text file (ASCII), or from any statistical package or from the clipboard.

**3. Explain how to communicate the outputs of data analysis using R language.**

Combine the data, code and analysis results in a single document using knitr for Reproducible research done. Helps to verify the findings, add to them and engage in conversations. Reproducible research makes it easy to redo the experiments by inserting new data values and applying it to different various problems.

**GET PDF**

**4. Difference between library () and require () functions in R language.**

|  |  |
| --- | --- |
| **library()** | **require()** |
| Library () function gives an error message display, if the desired package cannot be loaded. | Require () function is used inside function and throws a warning messages whenever a particular package is not Found |
| It loads the packages whether it is already loaded or not, | It just checks that it is loaded, or loads it if it isn’t (use in functions that rely on a certain package). The documentation explicitly states that neither function will reload an already loaded package. |

Consider a related program for the above differentiation.

if(!require(package, character.only=T, quietly=T)) {

install.packages (package)

library(package, character.only=T)

}

For multiple packages you can use

for(package in c(”, ”)) {

if(!require(package, character.only=T, quietly=T)) {

install.packages (package)

library(package, character.only=T)

}

}

**5. What is R?**

R is a programming language which is used for developing statistical software and data analysis. It is being increasingly deployed for machine learning applications as well.

**6. How R commands are written?**

By using # at the starting of the line of code like #division commands are written.

**7. What is t-tests() in R?**

It is used to determine that the means of two groups are equal or not by using t.test() function.

**8. What are the disadvantages of R Programming?**

The disadvantages are:-

* Lack of standard GUI
* Not good for big data.
* Does not provide spreadsheet view of data.

**9. What is the use of With () and By () function in R?**

with() function applies an expression to a dataset.

#with(data,expression)

By() function applies a function t each level of a factors.

#by(data,factorlist,function)

**10. In R programming, how missing values are represented?**

In R missing values are represented by NA which should be in capital letters.

**11. What is the use of subset() and sample() function in R?**

Subset() is used to select the variables and observations and sample() function is used to  generate  a random sample of the size n from a dataset.

**12. Explain what is transpose.**

Transpose is used for reshaping of the data which is used for analysis. Transpose is performed by t() function.

**13. What are the advantages of R?**

* The advantages are:-
* It is used for managing and manipulating of data.
* No license restrictions
* Free and open source software.
* Graphical capabilities of R are good.
* Runs on many Operating system and different hardware and also run on 32 & 64 bit processors etc.

Now that you are aware of the benefits of R programming, [Check R Programming training](https://intellipaat.com/r-programming-certification-training/#course-content).

**14. What is the function used for adding datasets in R?**

For adding two datasets rbind() function is used but the column of two datasets must be same.

Syntax: rbind(x1,x2……) where x1,x2: vector, matrix, data frames.

**15. How you can produce co-relations and covariances?**

Cor-relations is produced by cor() and covariances is produced by cov() function.

**16. What is difference between matrix and dataframes?**

Dataframe can contain different type of data but matrix can contain only similar type of data.

**17. What is difference between lapply and sapply?**

lapply is used to show the output in the form of list whereas sapply is used to show the output in the form of vector or data frame

**18. What is the difference between seq(4) and seq\_along(4)?**

Seq(4) means vector from 1 to 4 (c(1,2,3,4)) whereas seq\_along(4) means a vector of the  length(4) or 1(c(1)).

**19. Explain how you can start the R commander GUI.**

rcmdr command is used to start the R commander GUI.

**20. What is the memory limit of R?**

In 32 bit system memory limit is 3Gb but most versions limited to 2Gb and in 64 bit system memory limit is 8Tb.

**21. How many data structures R has?**

There are 5 data structure in R i.e. vector, matrix, array which are of homogenous type and other two are list and data frame which are heterogeneous.

Learn more about [data structure](https://intellipaat.com/tutorial/r-programming/data-structures-r-programming/) in R programming tutorial.

**22. Explain how data is aggregated in R.**

There are two methods that is collapsing data by using one or more BY variable and other is aggregate() function in which BY variable should be in list.

**23. How many sorting algorithms are available?**

There are 5 types of sorting algorithms are used which are:-

* Bubble Sort
* Selection Sort
* Merge Sort
* Quick Sort
* Bucket Sort

**24. How to create new variable in R programming?**

For creating new variable assignment operator ‘<-’ is used  
For e.g. mydata$sum <- mydata$x1 + mydata$x2

**25. What are R packages?**

Packages are the collections of data, R functions and compiled code in a well-defined format and these packages are stored in library. One of the strengths of R is the user-written function in R language.

**26. What is the workspace in R?**

Workspace is the current R working environment which includes any user defined objects like vector, lists etc.

**27. What is the function which is used for merging of data frames horizontally in R?**

Merge()function is used to merge two data frames  
Eg. Sum<-merge(data frame1,data frame 2,by=’ID’)

**28. what is the function which is used for merging of data frames vertically in R?**

rbind() function is used to merge two data frames vertically.  
Eg.

Sum<- rbind(data frame1,data frame 2)

**29. What is the power analysis?**

It is used for experimental design .It is used to determine the effect of given sample size.

**30. Which package is used for power analysis in R?**

Pwr package is used for power analysis in R.

**31. Which method is used for exporting the data in R?**

There are many ways to export the data into another formats like SPSS, SAS , Stata , Excel Spreadsheet.

**32. Which packages are used for exporting of data?**

For excel xlsReadWrite package is used and for sas,spss ,stata foreign package is implemented.

**33. How impossible values are represented in R?**

In R NaN is used to represent impossible values.

**34. Which command is used for storing R object into a file?**

Save command is used for storing R objects into a file.  
Syntax: >save(z,file=”z.Rdata”)

**35. Which command is used for restoring R object from a file?**

load command is used for storing R objects from a file.  
Syntax: >load(”z.Rdata”)

**36. What is the use of coin package in R?**

Coin package is used to achieve the re randomization or permutation based statistical tests.

**37. Which function is used for sorting in R?**

order() function is used to perform the sorting.

**38. What is the use of tapply?**

IOS-6.1.3

**39. What happens when the application object does not handle an event?**

The event will be dispatched to your delegate for processing.

**40. Explain app specific objects which store the app contents.**

The app specific objects are Data model objects that store app’s contents.

**41. Explain the purpose of using UIWindow object?**

UIWindow object coordinates the one or more views presenting on the screen.

**42. Tell me the super class of all view controller objects.**

UIView Controller class.

**43. How to create axes in the graph?**

Using axes() function custom axes are created.

**44. What is the use of abline() function?**

abline() function is add the reference line to a graph.  
Syntax:-

abline(h=yvalues, v=xvalues)

**45. Why vcd package is used?**

vcd package provides different methods for visualizing multivariate categorical data.

**46. What is GGobi?**

GGobi is an open source program for visualization for exploring high dimensional typed data.

**47. What is iPlots?**

It is a package which provide bar plots, mosaic plots, box plots, parallel plots, scatter plots and histograms.

**48. What is the use of lattice package?**

lattice package is to improve on base R graphics by giving better defaults and it have the ability to easily display multivariate relationships.

**49. What is fitdistr() function?**

It is used to provide the maximum likelihood fitting of univariate distributions. It is defined under the MASS package.

**50. Which data structures are used to perform statistical analysis and create graphs.**

Data structures are vectors, arrays, data frames and matrices.

**51. What is the use of sink() function?**

It defines the direction of output.

**52. Why library() function is used?**

This function is used to show the packages which are installed.

**53. Why search() function is used?**

By this function we see that which packages are currently loaded.

**54. On which type of data binary operators are worked?**

Binary operators are worked on matrices, vectors and scalars.

**55. What is the use of doBY package?**

It is used to define the desired table using function and model formula.

**56. Which function is used to create frequency table?**

Frequency table is created by table() function.

**57. Define loglm() function.**

Loglm() function is used to create log-linear models.

**58. What is the use of corrgram() function?**

corrgram() function is used to plot correlograms.

**59. How to create scatterplot matrices?**

Pair() or splom() function is used for create scatterplot matrices.

**60. What is npmc?**

It is a package which gives nonparametric multiple comparisons.

**61. What is the use of diagnostic plots?**

It is used to check the normality, heteroscedasticity and influential observations.

**62. Define anova() function.**

anova() is used to compare the nested models.

**63. What is cv.lm() function?**

It is defined under the DAAG package which is used for k-fold validation.

**64. Define stepAIC() function.**

It is define under the MASS package which performs stepwise model selection under exact AIC.

**65. Define leaps().**

It is used to perform the all-subsets regression and it is defined under the leaps package.

**66. Define relaimpo package.**

It is used to measure the relative importance of each of the predictor in the model.

**67. Why car package is used?**

It  provide a variety of regression including scatter plots, variable plots and it also enhanced diagnostic.

**68. Define robust package.**

It provides a library of robust methods including regression.

**69. What is robustbase?**

It is a package which provides basic robust statistics including model selection methods.

**70. Define plotmeans().**

It is define under gplots package which includes confidence intervals and it produces mean plot for single factors.

**71. What is the full form of MANOVA?**

MANOVA stands for multivariate analysis of variance.

**72. What is the use of MANOVA?**

By using MANOVA we can test more than one dependent variable simultaneously.

**73. Define mshapiro.test( ).**

It is a function which defines in mvnormtest package. It produces the Shapiro-wilk test for multivariate normality.

**74. Define barlett.test().**

Barlett.test() is used to provide a parametric k-sample test of the equality of variances.

**75. What is fligner.test()?**

It is a function which provides a non-parametric k sample test of the equality of variances.

**76. Define hovplot().**

It is define in HH package which provides a graphic test of homogeneity of variance based on brown forsyth.

**77. Which variables are represented by lower case letters?**

Numerical variables are represented by lower case letters.

**78. Which variables are represented by upper case letters?**

Categorical factors are represented by upper case letters.

**79. What is logistic regression?**

Logistic regression is used to predict the binary outcome from the given set of continuous predictor variables.

**80. Define Poison regression.**

It is used to predict the outcome variable which represents counts from the given set of continuous predictor variable.

**81. Define Survival analysis.**

It includes number of techniques which is used for modeling the time to an event.

**82. What is the use survfit() function?**

It estimates a survival distribution one or more groups.

**83. Define survdiff().**

It determines the differences in survival distribution between two or more groups.

**84. What is coxph()?**

It is a function which is used to model the hazard function on the set of predictor variable.

**85. In which package survival analysis is defined?**

Survival analysis is defined under the survival package.

**86. What is the use of MASS package?**

MASS functions include those functions which performs linear and quadratic discriminant function analysis.

**87. Define qda().**

qda() prints a quadratic discriminant function.

**88. Define lda().**

lda() is used to print the discriminant functions which is based on centered variable.

**89. What is the use of forecast package?**

It provides the functions which are used for automatic selection of ARIMA and exponential models.

**90. Define auto.arima().**

It is used to handle the seasonal as well as non-seasonal ARIMA models.

**91. What is principal() function?**

It is define in psych package which is used to rotate and extract the principal components.

**92. What is FactoMineR?**

It is a package which includes quantitative and qualitative variables. It also includes supplementary variables and observations.

**93. What is the full form of CFA?**

CFA stands for Confirmatory Factor Analysis.

**94. What is the use of boot.sem() function?**

It is used to bootstrap the structural equation model.

**95. What is the full form of SEM?**

SEM stands for Structural Equation Modeling.

**96. Which function performs classical multidimensional scaling?**

cmdscale() function is used to perform classical multidimensional scaling.

**97. Define isoMDS().**

This function is defined under the MASS package which performs nonmetric multidimensional scaling.

**98. Which function perform individual difference scaling?**

It is done by indscal() function.

**99. What is pvclust() function ?**

It comes under the pvclust package which provides p-values for hierarchical clustering.

**100. Define cluster.stats() ?**

It is define in fpc package which provide a method for comparing the similarity of two clusters solution using different validation criteria.

**101. What we use party package?**

It is used to provide a non-parametric regression for ordinal, nominal, censored and multivariate responses.

**102. Which package provide the bootstrapping?**

boot package is used which provide bootstrapping.

**103. Define MATLAB package.**

Matlab package includes those wrapper functions and variable which are used to replicate matlab function calls.

**104. What is the of use Matrix package?**

Matrix package includes those function which support sparse and dense matrices like Lapack, BLAS etc.