**1) Explain what is R?**

R is data analysis software which is used by analysts, quants, statisticians, data scientists and others.

**2) List out some of the function that R provides?**

The function that R provides are

* Mean
* Median
* Distribution
* Covariance
* Regression
* Non-linear
* Mixed Effects
* GLM
* GAM. etc.

**3) Explain how you can start the R commander GUI?**

Typing the command, ("Rcmdr") into the R console starts the R commander GUI.

**4) In R how you can import Data?**

You use R commander to import Data in R, and there are three ways through which you can enter data into it

* You can enter data directly via Data ◊ New Data Set
* Import data from a plain text (ASCII) or other files (SPSS, Minitab, etc.)
* Read a data set either by typing the name of the data set or selecting the data set in the dialog box

**5) Mention what does not ‘R’ language do?**

* Though R programming can easily connects to DBMS is not a database
* R does not consist of any graphical user interface
* Though it connects to Excel/Microsoft Office easily, R language does not provide any spreadsheet view of data

**6) Explain how R commands are written?**

In R, anywhere in the program you have to preface the line of code with a #sign, for example

* # subtraction
* # division
* # note order of operations exists

**7) How can you save your data in R?**

To save data in R, there are many ways, but the easiest way of doing this is

Go to Data > Active Data Set > Export Active Data Set and a dialogue box will appear, when you click ok the dialogue box let you save your data in the usual way.

**8) Mention how you can produce co-relations and covariances?**

You can produce co-relations by the cor () function to produce co-relations and cov () function to produce covariances.

**9) Explain what is t-tests in R?**

In R, the t.test () function produces a variety of t-tests. T-test is the most common test in statistics and used to determine whether the means of two groups are equal to each other.

**10) Explain what is With () and By () function in R is used for?**

* With() function is similar to DATA in SAS, it apply an expression to a dataset.
* BY() function applies a function to each level of factors. It is similar to BY processing in SAS.

**11) What are the data structures in R that is used to perform statistical analyses and create graphs?**

R has data structures like

* Vectors
* Matrices
* Arrays
* Data frames

**12) Explain general format of Matrices in R?**

General format is

Mymatrix< - matrix (vector, nrow=r , ncol=c , byrow=FALSE,

dimnames = list ( char\_vector\_ rowname, char\_vector\_colnames))

**13) In R how missing values are represented ?**

In R missing values are represented by NA (Not Available), why impossible values are represented by the symbol NaN (not a number).

**14) Explain what is transpose?**

For re-shaping data before, analysis R provides various method and transpose are the simplest method of reshaping a dataset. To transpose a matrix or a data frame t () function is used.

**15) Explain how data is aggregated in R?**

By collapsing data in R by using one or more BY variables, it becomes easy. When using the aggregate() function the BY variable should be in the list.

**16) What is the function used for adding datasets in R?**

rbind function can be used to join two data frames (datasets). The two data frames must have the same variables, but they do not have to be in the same order.

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

In R, subset() functions help you to select variables and observations while through sample() function you can choose a random sample of size n from a dataset.

**18) Explain how you can create a table in R without external file?**

Use the code

myTable = data.frame()

edit(myTable)

last to last week you people conducted sql interview questions webinar, I tought it’s the same so keeping that in mind I almost prepared the file.  
Luckily we had discussion, now w

**17)What are the different components of grammar of graphics?**

Broadly speaking these are different components in grammar of graphics:

* Data layer
* Aesthetics layer
* Geometry layer
* Facet layer
* Co-ordinate layer
* Themes layer

18) **What is Rmarkdown? What is the use of it?**

RMarkdown is a reporting tool provided by R. With the help of Rmarkdown, you can create high quality reports of your R code.

The output format of Rmarkdown can be:

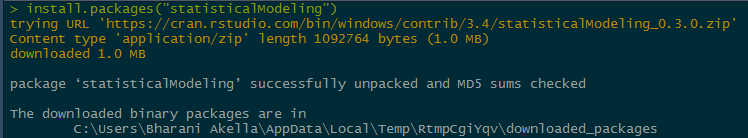
* HTML
* PDF
* WORD

### 19)How do you install a package in R?

he below command is used to install a package in R:

install.packages(“<package\_name>”)

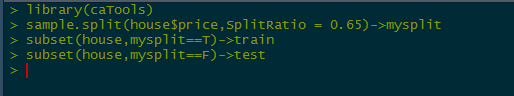
let’s look at an example:



### 20) ****What are the steps to build and evaluate a linear regression model in R?****

These are sequential steps which need to be followed while building a linear regression model:

* Start off by dividing the data into train and test sets, this step is vital because you will be building the model on the train set and evaluating it’s performance on the test set.
  + You can do this using the sample.split() function from the “catools” package. This function gives an option of split-ratio, which you can specify according to your needs.



* Once, you are done splitting the data into training and test sets, You can go ahead and build the model on the train set.
  + The “lm()” function is used to build a model.

21) **Name some packages in R, which can be used for data imputation?**

These are some packages in R which can used for data imputation

* MICE
* Amelia
* missForest
* Hmisc
* Mi
* imputeR

22) **Explain about confusion matrix in R?**

* A confusion matrix can be used to evaluate the accuracy of the model built.  It Calculates a cross-tabulation of observed and predicted classes. This can be done using the “confusionmatrix()” function from the “caTools” package.

### 23) Explain initialize() function in R?

This function is used to initialize the private data members while declaring the object.

### 24) How can we find the mean of one column with respect to another?

In iris dataset, there are five columns, i.e., Sepal.Length, Sepal.Width, Petal.Length, Petal.Width and Species. We will calculate the mean of Sepal-Length across different species of iris flower using the mean() function from the mosaic package.

1. mean(iris$Sepal.Length~iris$Species)

### 25) What is a Random Walk model?

A random walk is the simplest example of a non-stationary process. A random walk has no specified mean or variance, strong dependence over time, and its changes or increments are white noise. Simulating random walk in R:

**arima.sim(model=list(order=c(0,1,0)),n=40)->rw ts.plot(rw)**

### **26)** What is a White Noise model?

It is a basic time series model and a simple example of a stationary process. A white noise model has a fixed constant mean, a fixed constant variance, and no correlation over time. We can simulate a white noise model in the following way:

**arima.sim(model=list(order=c(0,0,0)),n=50)->wn**

### **27)** Give any five features of R.

1. Simple and effective programming language.
2. It is a data analysis software.
3. It gives effective storage facility and data handling.
4. It gives high extensible graphical techniques.
5. It is an interpreted language.

### 28) Differentiate between R and Python in terms of functionality?

For data analysis, R has inbuilt functionality, but in Python, the data analysis functionalities are not inbuilt. They are available by packages like Pandas and Numpy.

### 29) Explain RStudio.

RStudio is an integrated development environment which allows us to interact with R more readily. RStudio is similar to the standard RGui, but it is considered more user-friendly. This IDE has various drop-down menus, windows with multiple tabs, and so many customization processes. The first time when we open RStudio, we will see three Windows. The fourth Window will be hidden by default.

### 30) What are the advantages and disadvantages of R?

**Advantages**

1. Open Source
2. Data Wrangling
3. Array of Packages
4. Platform Independent
5. Machine Learning Operations

**Disadvantages**

1. Weak origin
2. Data Handling
3. Basic Security
4. Complicated Language
5. Lesser Speed

### 31) What is the purpose behind R and Hadoop integration?

1. For executing Hadoop to execute R code.
2. For using R to access the data stored in Hadoop.

### 32) What will be the output of the expression all(NA==NA)?

[1] NA

### 33) What is the difference b/w sample() and subset() in R?

The sample() method is used to choose a random sample of size n from a dataset while the subset method is used to choose variables and observations

### 34) Why do we use the command - install.packages(file.choose(), repos=NULL)?

This command is used to install an R package from the local directory by browsing and selecting the file.

### 35) Give the command to create a histogram and to remove a vector from the R workspace?

hist() and rm() function are used as a command to create a histogram and remove a vector from the R workspace.

### 36) Differentiate b/w "%%" and "%/%".

The "%%" provides a reminder of the division of the first vector with the second, and the "%/%" gives the quotient of the division of the first vector with the second.

### 37) Why do we use apply() function in R?

This 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.

### 38) Differentiate between library() and require() functions.

If the desired package cannot be loaded, then the library() function gives an error message and display while the required () function is used inside the function and throws a warning message whenever a particular package is not found.

### 39) What is the t-test() in R?

The t-test() function is used to determine that the mean of the two groups are equal or not.

### 40) What is the use of with() and by() functions in R?

The with() function applies an expression to a dataset, and the by() function applies a function to each level of factors.

### 41) Differentiate b/w lapply and sapply.

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

### 42) Explain aggregate() function.

The aggregate() function is used to aggregate data in R. There are two methods which are collapsing data by using one or more BY variable and other is an aggregate() function in which By variable should be in the list.

### 43) Explain the doBy package?

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

### 44) Explain fitdistr() function?

This function is used to give the maximum likelihood fitting of univariate distribution and defined under the MASS package.

### 45) What are GGobi and iPlots?

The GGobi is an open-source program for visualization to exploring high dimensional typed data, and the iPlots is a package which provides bar plots, mosaic plots, box plots, parallel plots, histograms, and scatter plots.

### 46) Explain the lattice package.

The lattice package is meant to improve upon the base R graphics by giving better defaults and has the ability to display multivariate relationships easily.

### 47) Explain anova() function.

The anova() function is used for comparing the nested models.

### 48) Explain cv.lm() and stepAIC() function.

The cv.lm() function is defined under the DAAG package used for k-fold validation while the stepAIC() function is defined under the MASS package that performs stepwise model selection under exactAIC.

### 49) Explain leaps() function.

The leaps() function is used to perform the all-subsets regression and defined under the leaps package.

### 50) Explain relaimpo and robust package.

This package is used to measure the relative importance of every predictor in the model, and the robust package gives a library of robust methods, including regression.

### 51) Give full form of MANOVA and what is the use of it.

MANOVA stands for Multivariate Analysis of Variance, and it is used to test more than one dependent variable simultaneously.

### 52) Explain mashapiro.test() and barlett.test().

This function defines in the mvnormtest package and produces the Shapiro-wilk test to multivariate normality. The barlett.test() is used to provide a parametric k-sample test of the equality of variances.

### 53) Explain the use of the forecast package.

The forecast package gives the functions which are used to automatic selection of exponential and ARIMA models.

### 54) Differentiate between qda() and lda() function.

The qda() function prints a quadratic discriminant function while lda() function print the discriminant functions based on the centered variable.

### 55) Explain the auto.arima() and principal() function.

The auto.arima() function handle both the seasonal and non-seasonal ARIMA model and the principal() function used for rotating and extracting the principal components.

### 56) Explain FactoMineR.

The FactoMineR is a package that includes qualitative and quantitative variables. The observations and supplementary variables are also included in these packages.

### 57)What is the full form of SEM and CFA?

CFA stands for Confirmatory Factor Analysis, and SEM stands for Structural Equation Modeling

### 58) Define cluster.stats() and pvclust() function().

The cluster.stats() function define in the fpc package that provides a method for comparing the similarity of two cluster solutions using different validation criteria, and the pvclust() function is defined in the pvclust package that provides p-values for hierarchical clustering.

### 59) Define MATLAB and party packages.

This package includes wrapper functions and variable which are used for replicating Matlab function calls.

### 60) Explain S3 and S4 systems.

In oops, the S3 is used to overload any function. So that we can call the functions with different names, and it depends on the type of input parameter or the number of parameters, and the S4 is the most important characteristic of oops. However, this is a limitation, as it is quite difficult to debug. There is an optional reference class for S4.

### 61) Explain Chi-Square Test

The Chi-Square Test is used to analyze the frequency table (i.e., contingency table), which is formed by two categorical variables. The chi-square test evaluates whether there is a significant relationship between the categories of the two variables.

### 62) Explain Random Forest.

The Random Forest is also known as Decision Tree Forest. It is one of the popular decision tree-based ensemble models. The accuracy of these models is higher than other decision trees. This algorithm is used for both classification and regression applications.

### 63) Explain Time Series Analysis.

Any metric which is measured over regular time intervals creates a time series. Analysis of time series is commercially important due to industrial necessity and relevance, especially with respect to the forecasting (demand, supply, and sale, etc.). A series of data points in which each data point is associated with a timestamp is known as time series.

### 64) Explain Pie chart in R.

R programming language has several libraries for creating charts and graphs. A pie-chart is a representation of values in the form of slices of a circle with different colors.

### 65) Explain Histogram.

A histogram is a type of bar chart which shows the frequency of the number of values which are compared with a set of values ranges. The histogram is used for the distribution, whereas a bar chart is used for comparing different entities. In the histogram, each bar represents the height of the number of values present in the given range.

66) 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.

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

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

1. Which data object in R is used to store and process categorical data?
2. How can you load and use csv file in R?
3. How do you get the name of the current working directory in R?
4. What is R Base package?
5. How R is used in logistic regression?
6. How do you access the element in the 2nd column and 4th row of a matrix named M?
7. What is recycling of elements in a vector?
8. Give an example.What are different ways to call a function in R?
9. What is lazy function evaluation in R?
10. How do you install a package in R?
11. Name a R packages which is used to read XML files.
12. Can we update and delete any of the elements in a list?
13. Give the general expression to create a matrix in R.
14. which function is used to create a boxplot graph in R?
15. In doing time series analysis, what does frequency = 6 means in the ts() function?
16. What is reshaping of data in R?
17. What is the output of runif(4)?
18. How to get a list of all the packages installed in R ?
19. What is expected from running the command - strsplit(x,"e")?
20. Give a R script to extract all the unique words in uppercase from the string - "The quick brown fox jumps over the lazy dog".
21. Vector v is c(1,2,3,4) and list x is list(5:8), what is the output of v\*x[1]?Vector v is c(1,2,3,4) and list x is list(5:8), what is the output of v\*x[[1]]?
22. What does unlist() do?
23. Give the R expression to get 26 or less heads from a 51 tosses of a coin using pbinom.
24. X is the vector c(5,9.2,3,8.51,NA), What is the output of mean(x)?
25. How do you convert the data in a JSON file to a data frame?
26. Give a function in R that replaces all missing values of a vector x with the sum of elements of that vector?
27. What is the use of apply() in R?
28. Is an array a matrix or a matrix an array?
29. How to find the help page on missing values?
30. How do you get the standard deviation for a vector x?
31. How do you set the path for current working directory in R?
32. What is the difference between "%%" and "%/%"?
33. What does col.max(x) do?
34. Give the command to create a histogram.
35. How do you remove a vector from the R workspace?
36. List the data sets available in package "MASS"
37. List the data sets available in all available packages.
38. What is the use of the command - install.packages(file.choose(), repos=NULL)?
39. Give the command to check if the element 15 is present in vector x.
40. Give the syntax for creating scatterplot matrices.
41. What is the difference between subset() function and sample() function in R?
42. How do you check if "m" is a matrix data object in R?
43. What is the output for the below expression all(NA==NA)?
44. How to obtain the transpose of a matrix in R?
45. What is the use of "next" statement in R?
46. **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 ()

1. **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.

1. **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.

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

data.table

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

boxplot () or text ()

1. **What are the different type of sorting algorithms available in R language?**Bucket Sort  
   Selection Sort  
   Quick Sort  
   Bubble Sort  
   Merge Sort
2. **What is the command used to store R objects in a file?**

save (x, file=”x.Rdata”)

1. **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.

1. **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.

1. **How is a Data object represented internally in R language?**

unclass (as.Date (“2016-10-05″))

1. **Which package in R supports the exploratory analysis of genomic data?**

adegenet

1. **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.

1. **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.

1. **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.

1. **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.

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

Scalars, Matrices ad Vectors.

1. **How do you create log linear models in R language**

Using the loglm () function

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

number

1. **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.

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

character

1. **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).

1. **How can you debug and test R programming code?**

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

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

number

1. **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}

1. **What happens if the application object is not able to handle an event?**

The event is dispatched to the delegate for processing.

1. **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)).

1. **How do you write R commands?**

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

1. **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.

1. **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.

1. **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.

1. **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.

1. **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.

1. **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)))

1. **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.

1. **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.

1. 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.

1. How will you check if an element 25 is present in a vector?

There are various ways to do this-

It can be done using the match () function- match () function returns the first appearance of a particular element.

The other is to use %in% which returns a Boolean value either true or false.

Is.element () function also returns a Boolean value either true or false based on whether it is present in a vector or not.

1. **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.

1. **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.

1. **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.

1. **Which function helps you perform sorting in R language?**

Order ()

1. **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))
2. **Which function is used to create a histogram visualisation in R programming language?**  
   Hist()
3. **Write the syntax to set the path for current working directory in R environment.**  
   Setwd(“dir\_path”)
4. **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))

## 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)]

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

It will generate 7 randowm numbers between 0 and 1.

1. **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 = )

1. **What will be the output on executing the following R programming code –**

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

sum(mat)

 8

1. **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="\_")

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

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

1. **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.

1. **What will be the output of the following R programming code ?**

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

**var2**

It will give an error.

1. **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 = ".")

1. **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.

1. **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.

1. **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.

1. **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))

1. **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.

1. **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 -

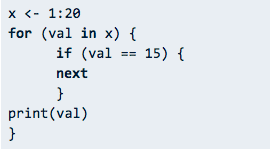
abline(h=yvalues, v=xvalues)

1. **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.

### 174)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:

[](https://res.cloudinary.com/springboard-images/image/upload/w_1080,c_limit,q_auto,f_auto,fl_lossy/wordpress/2018/07/Screen-Shot-2018-07-30-at-11.26.33-AM.png)

### 175)How do you assign a variable in R?

myVar <- 15  
print(myVar

Notice the following is also valid:

myVar = 15  
print(myVar)

### 176) What are the different data types/objects in R?

* Vectors
* Matrices
* Lists
* Arrays
* Factors
* Data frames

### ****How R commands are written?****

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

### ****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.

### ****How you can produce co-relations and covariances?****

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

### ****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.

### ****Which packages are used for exporting of data?****

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

### ****How impossible values are represented in R?****

In R NaN is used to represent impossible values.

### ****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”)

### ****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”)

### ****What is the use of coin package in R?****

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

### ****What is the use of tapply?****

IOS-6.1.3

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

The event will be dispatched to your delegate for processing.

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

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

### ****Explain the purpose of using UIWindow object?****

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

### ****How to create axes in the graph?****

Using axes() function custom axes are created.

### ****Why vcd package is used?****

vcd package provides different methods for visualizing multivariate categorical data.

### ****What is iPlots?****

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

### ****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.

### ****What is fitdistr() function?****

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

### ****What is the use of sink() function?****

It defines the direction of output.

### ****On which type of data binary operators are worked?****

Binary operators are worked on matrices, vectors and scalars.

### ****What is the use of doBY package?****

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

### ****Define loglm() function.****

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

### ****How to create scatterplot matrices?****

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

### ****What is the use of diagnostic plots?****

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