

# By OnlineInterviewQuestions.com

**R** is a programming language and free software environment used for multiple purposes such as statistical analysis, data manipulation, predicting and forecasting, etc. With R, well-designed publication plots can be produced. R runs on platforms like UNIX, LINUX, Windows, and MacOS. The code for R is written in C, Fortran and R. R is an interpreted language that can implement a wide variety of statistical and graphical techniques. R makes it easier for the users to follow the algorithm choices as most of the functions are written in R itself.

Due to its interesting benefits, R is used by several companies such as Google, Facebook, Ford, etc. R is used by the Human Rights Data Analysis Group to gauge the impact of war. R is also used by Ford to revamp the designs of its vehicles. R has a promising future because of its open source nature. According to Gartner, the popularity of R will definitely grow further. So, it is the right time to move forward in your career with R. This article covers important **R programming interview questions** that you can take ideas from if you're taking an interview.

## Q1. What is R Programming?

R is a programming language and a software environment meant for statistical analysis and creating graphs. It is used by analysts, statisticians and data scientists for various purposes. R uses data objects for data calculations and it is an alternative to conventional statistical packages such as SAS, SPSS, etc. A lot of companies are incorporating R into their business models to proliferate their revenue. There is a huge career prospect in R such as data scientist, R programmer, Analyst consultant, etc.

## Q2. Enlist some of the functions that R provides?

The functions that R provides are

- Mean- it is calculated by taking the sum of the values and dividing it by a number of values. The function used is **mean()**.
- Median- it is the middle most value in the data series. The function used in R programming is **median()**.

Other functions of R include Regression, GLM, mixed-effects, distribution, GAM, non-linear, etc.

## Q3. What are the different data structures in R?

Data structure is a form of organizing and storing data. It is imperative to have a strong understanding of various data types and data structures in order to make the best use of R languages. R programming supports five types of data structures namely vector, matrix, list, data frame and factor.

- **Vector**– This data structure contains an integer, double, complex, etc. It is a sequence of same data

- elements and `c()` function is used to create a vector in R programming.
- **Matrix-** it is a two-dimensional data structure and is used to bind vectors from the same length. All the elements in the matrix have to be of the same type and it is created using a `matrix()` function. The value of row can be defined using `nrow` and the value of the column can be defined using `ncol`.
  - **List-** list includes data of different types like numbers, strings, vectors, etc. It is somewhat like a vector but it contains mixed elements. A list is created using `()`.
  - **Data frame-** it is a special list where each element is of the same length. A data frame has the features of both, matrices and lists. It is more generic than the matrix as different columns have different data types. It is created using `frame()` function.
  - **Factors-** it is created using `factor()` function and is used to store predefined and categorical data.

#### **Q4. What are the steps involved in building and evaluating a linear regression model in R?**

The following steps are followed to build and evaluate a linear regression model in R

- The first step is to divide the data into train and test sets. This step is crucial because you have to build a model on the train and evaluate its performance on the test set. This can be done with `sample.split()` function from the package.
- The second step is to build the model on the train set. The function used to build the model is `lm()` function.
- Once you've built the model, you can predict the values on the data set with the help of the `predict()` function.
- The last step in the linear regression model is to find out the RMSE. The lower the value of RMSE, better is the prediction.

#### **Q5. What are the different packages in R?**

R packages are the collection of R functions and sample data that are stored under a directory name called library. Initially, R adds a set of packages during installation. However, new packages are added as and when required for specific purposes. The different packages available in R are

- **MICE-** the MICE package deals with missing data. It creates replacement values for the missing data. There are two types of missing data namely MCAR and MNAR. In this package, the `mice()` function looks after the imputing process.
- **Amelia-** this package is used for multiple imputations of missing data. it also produces multiple output datasets for analysis. To use this package, you can either invoke `ameliagui()` command or run `Amelia` function on the data.
- **Mi-** the Mi package provides functions for data manipulation and imputes missing values. This package has several features that allow the users to get into the imputation process and gauge the reasonableness of the resulting model.

#### **Q6. List the functions available in “dplyr” package.**

The functions in dplyr package are-

- **Filter-** Filter() allows you to select a subset of rows in a data frame. the first argument is the tibble and the second argument is the variables within that data frame. it selects the rows where expression is true.
- **Arrange-** Arrange() records the rows on the basis of data frames or a set of column names. Desc() function is used to arrange columns in descending orders.
- **Mutate-** it is used to add new variables to the data. it is also used to add new columns that are functions of existing columns. Dplyr::mutate is used to refer to the newly created column.
- **Select()-** this function is used to zoom in on a useful subset that works on numeric values. With select(), you can use functions like ends\_with(), matches(), starts\_with(), etc.

## Q7. What packages are used for data mining in R?

The packages used for mining in R are-

- Data.table- supports fast reading of large files.
- Arules- used for rule learning.
- Tm- used to perform text mining.
- Forecast- provides functions for time series analysis.

## Q8. Why is clustering required in data analysis?

Clustering refers to the group of objects that belongs to the same class. It is a process to make a group of abstract objects into the class of similar objects. Clustering is required in data analysis due to the following reasons-

- **Scalability-** clustering is required to deal with large databases.
- **Interpretability-** the result of clustering should be comprehensive and usable.
- **Dimensionality-** the clustering algorithm is used to handle high-dimensional space.
- **Deal with noisy data-** Databases contains erroneous data. Algorithms that are sensitive to such data may deliver poor results.

## Q9. Differentiate between K-MEANS clustering and hierarchical clustering.

- **KMEANS clustering** in this method, objects are classified as belonging to K-groups. It is also known as partitioning method. The result of this method is K clusters and in each cluster, there may be a centroid. This method is popular for cluster analysis in data mining. The K-means clustering algorithm is used to find groups which have not been labelled in the data. This method is used to find groups in the data, with the number of groups which are represented by K.
- **Hierarchical clustering** It is a method of cluster analysis which aims to build the hierarchy of clusters. This method has two approaches namely divisive approach and agglomerative approach. In agglomerative approach, each object forms a separate group and keeps on merging the groups that are close to one

another. It is also known as the bottom-up approach. In a divisive approach, we start with all the objects in the same cluster. The cluster is split into smaller clusters. It is also known as a top-down approach.

#### **Q10. What is rattle package in R?**

Rattle gives statistical and visual summaries of data and is a popular GUI for data mining. It transforms data so it can be easily modelled and builds a supervised and unsupervised ML model from the data. It also gives the graphical presentation of the models. Rattle is also used as a teaching facility to learn R languages. The features of Rattle package include clustering, modelling, evaluation, statistical test, etc.

#### **Q11. What is White Noise model in R?**

In R, a white noise model is a basic time series model which is also the basis for more elaborated and defined models. To stimulate the data from a variety of time series model, Arima.sim() function is used. The white noise model has a fixed constant mean, fixed constant variance and no correlation over time.

#### **Q12. What is the Random Walk Model in R?**

In R programming, random walk model is an example of the non-stationary model. A random walk has no fixed mean or variance. It also has a strong dependence over time. There are two types of random walks namely random walk without drift and random walk with drift.

#### **Q13. How can you import data in R?**

There are several ways to import data in R. You can use R commander to import data in R.

- Excel file- If the sample data is in excel format, function read.xls is used from the data package. It returns a data frame. Alternatively, loadWorkbook can also be used to read the entire workbook.
- Minitab file- if the data file is in Minitab format, it can be opened using the function read.mtb. It returns a list of components in the Minitab worksheet.
- SPSS file- the data files in SPSS formats can be opened using the function read.spss. It returns a list of components.
- CSV file- Each cell inside the CSV data file is separated by a special character such as a comma.

#### **Q14. Explain the concept of Principal Component Analysis.**

under the Principal Component Analysis, the data is transformed into a new space. The first principal component takes the maximum amount of variance from the original data. The second principal component captures the amount of variability left. This is true for each component element and they are all uncorrelated. In

R programming, Principal Component Analysis can be done using the function `prcomp()`.

### **Q15. Compare R and Python.**

The following are the few differences between Python and R language

#### **R programming language**

In R programming, model building is similar to python.

It has good model interpretability.

It has a steep learning curve.

It has better data visualisation libraries.

Good commuting support.

#### **Python Programming language**

Model building is similar to R.

It has comparatively low model interpretability.

In python, the learning curve is easier as compared with R.

Data visualisation is not better than R.

Commuting support not better than R.

### **Q16. Differentiate between Library() and Require().**

**Library()**- If the desired package cannot be loaded, this function will display an error message. It loads the package whether it is already loaded or not.

**Require()**- When a particular package is not found, it gives warning messages. Require() is used inside a function. It checks whether it is loaded or not and loads if it is not loaded.

### **Q17. Enlist the advantages of using R languages.**

R language is currently the most sought-after programming languages. It offers several benefits to the users

- R is a comprehensive language used for manipulation and managing of data.
- R has good graphical capabilities.
- It is free and open source software.
- There are no licence restrictions for R.
- R runs on many operating systems and hardware.
- R supports all statistical tests and models.

### **Q18. How many sorting algorithms are available?**

In R, the following sorting algorithms are available

- Bubble sort
- Selection sort
- Merge sort

- Quick sort
- Bucket sort

### **Q19. What are the features of R programming?**

The following are the programming features of R

- There are packages in R. these R packages are useful in collecting functions into a single set.
- R programming includes database input, data export, variable labels, etc.
- R is an interpreted language and support matrices.
- R supports object-oriented programming and procedural programming. Object-oriented programming consists of classes, objects whereas procedural programming includes records, procedures.

### **Q20. What are the applications of R?**

The scope of R as a programming language is high and it has varied applications in various verticals. The important applications are

- R is used as an important tool in finance. It is used by several data analysts and research programmers.
- R deals with a lot of statistics. It is considered the best for data science. R also provides an environment for statistical computing and design.
- R is also used for data importing and cleaning.

Please Visit [OnlineInterviewquestions.com](http://OnlineInterviewquestions.com) to download more pdfs