

DATA SCIENCE LABORATORY LAB MANUAL

1. Download, install and explore the features of R for data analytics.

EXERCISE 1: R LANGUAGE

AIM:

To download, install and explore the features of R/Python for data analytics.

PROCEDURE:

Step 1: Go to www.r-project.org

Step 2: To download R, please choose your preferred CRAN mirror

Step 3: Click CRAN (Comprehensive R Archived Network) on the left side

Step 4: Select India - <https://mirror.niser.ac.in/cran/>

Step 5: Select options to download R for Linux, Mac or for Windows

Step 6: Download the most recent version of R

Step 7: Run it and Install accept the default options

Step 8: After R is installed, download "RStudio"

Step 9: Go to www.rstudio.com

Step 10: Click on Download, then

Step 11: Click to install "RStudio" on the desktop

Step 12: To install "RStudio", drag "RStudio" icon into the applications folder

Step 13: "RStudio" is installed

Step 14: Open "RStudio" in Applications folder

Step 15: Select "RStudio"

Step 16: "RStudio" which is running R is displayed.

FEATURES OF R:

1. Open-source

R is an open-source software environment. It is free of cost and can be adjusted and adapted according to the user's and the project's requirements. Improvements and add packages for additional functionalities can be made. R is freely available. You can learn how to install R, Download and start practicing it.

2. Strong Graphical Capabilities

R can produce static graphics with production quality visualizations and has extended libraries providing interactive graphic capabilities. This makes data visualization and data representation very easy. From concise charts to elaborate and interactive flow diagrams, all are well within R's repertoire. Look at the attractive graphical visualizations in R.

3. Highly Active Community

R has an open-source library which is supported by its growing number of users. The R environment is continuously growing. This growth is due to its large user-base.

4. A Wide Selection of Packages

CRAN or Comprehensive R Archive Network houses more than 10,000 different packages and extensions that help solve all sorts of problems in data science. High-quality interactive graphics, web application development, quantitative analysis or machine learning procedures, there is a package for every scenario available. R contains a sea of packages for all the forms of disciplines like astronomy, biology, etc. While R was originally used for academic purposes, it is now being used in industries as well.

5. Comprehensive Environment

R has a very comprehensive development environment meaning it helps in statistical computing as well as software development. R is an object-oriented programming language. It also has a robust package called Rshiny which can be used to produce full-

fledged web apps. Combined with data analysis and data visualization, R can be used for highly interactive online data-driven storytelling.

6. Can Perform Complex Statistical Calculations

R can be used to perform simple and complex mathematical and statistical calculations on data objects of a wide variety. It can also perform such operations on large data sets.

7. Distributed Computing

In distributed computing, tasks are split between multiple processing nodes to reduce processing time and increase efficiency. R has packages like ddR and multiDplyr that enable it to use distributed computing to process large data sets.

8. Running Code without a Compiler

R is an interpreted language which means that it does not need a compiler to make a program from the code. R directly interprets provided code into lower-level calls and pre-compiled code.

9. Interfacing with Databases

R contains several packages that enable it to interact with databases like Roracle, Open Database Connectivity Protocol, RmySQL, etc.

10. Data Variety

R can handle a variety of structured and unstructured data. It also provides various data modeling and data operation facilities due to its interaction with databases.

11. Machine Learning

R can be used for machine learning as well. The best use of R when it comes to machine learning is in case of exploration or when building one-off models.

12. Data Wrangling

Data wrangling is the process of cleaning complex and inconsistent data sets to enable convenient computation and further analysis. This is a very time taking process. R with its extensive library of tools can be used for database manipulation and wrangling.

13. Cross-platform Support

Cross Platform compatible with R. R is machine-independent. It supports the cross-platform operation. Therefore, it can be used on many different operating systems.

14. Compatible with Other Programming Languages

While most of its functions are written in R itself, C, C++ or FORTRAN can be used for computationally heavy tasks. Java, .NET, Python, C, C++, and FORTRAN can also be used to manipulate objects directly.

15. Data Handling and Storage

R is integrated with all the formats of data storage due to which data handling becomes easy.

16. Vector Arithmetic

Vectors are the most basic data structure in R, and most other data structures are derived from vectors. R uses vectors and vector arithmetic and does not need a lot of looping to process a large set of values. This makes R much more efficient.

17. Compatibility with Other Data Processing Technologies

R can be easily paired with other data processing and distributed computing technologies like Hadoop and Spark. It is possible to remotely use a Spark cluster to process large datasets using R. R and Hadoop can be paired as well to combine Hadoop's large scale data processing and distributing computing capabilities with R's statistical computing power.

18. Generates Report in any Desired Format

R's markdown package is the only report generation package you will ever need when working with R. The markdown package can help produce web pages. It can also generate reports in the form of word documents or PowerPoint presentations. All with your R code and results embedded into them

Some Unique Features of R Programming

Due to a large number of packages available, there are many other handy features as well: Since R can perform operations directly on vectors, it doesn't require too much looping. R can pull data from APIs, servers, SPSS files, and many other formats. R is useful for web scraping. It can perform multiple complex mathematical operations with a single command. Using R Markdown, it can create attractive reports that combine plain text with code and visualizations of the results. Due to a large number of researchers and statisticians using it, new ideas and technologies often appear in the R community first.

Summary

In this exercise, some of the important features of R has been explored. It has been learned why it is the most preferred language for statistical modeling. Its massive repository and sheer versatility make R the most popular statistical language.

Result

Thus, R has been downloaded, installed and important features of R is explored very well for data analytics.

Some of the diagrams that can be drawn on the left sheet:

Features of R Programming

- Various Graphical Libraries
 - Wide Array of Packages
 - Handling all Sorts of Data
 - Fast Calculation
 - Comprehensive Language
 - Open-Source
- 

Data Visualization in R

