DATA SCIENCE USING R - INTRODUCTION



J.JANSI PREMA M.Sc.,B.ed., M.Phil.,
DEPARTMENT OF COMPUTER SCIENCE,
JEYARAJ ANNAPACKIAM COLLEGE FOR WOMEN (AUTONOMOUS),
PERIAKUALM.





R Programming





TABLE OF CONTENTS



- History
- Introduction
- R Basics
- Features
- Comparison with other programming languages
- Merits
- Demerits
- Conclusion







R is a programming language it was an implementation over S language. **R** was first designed by Ross Ihaka and Robert Gentleman at the University of Auckland in 1993.

It was stable released on October 31st 2014 the 4 months ago ,by R Development Core Team Under GNU General Public License.



Introduction



R is a programming language and software environment for statistical computing and graphics. The R language is widely used among statisticians and data miners for developing statistical software and data analysis.

It compiles and runs on a wide variety of UNIX platforms, Windows and Mac OS. R can be downloaded and installed from CRAN website, CRAN stands for Comprehensive R Archive Network.



R Basics



Why R?

- The most extensive modeling resources in scientific research.
- The fine publishing quality graphs.
- Easy to develop your own model.
- R is freely available under GNU General Public License.



R Package

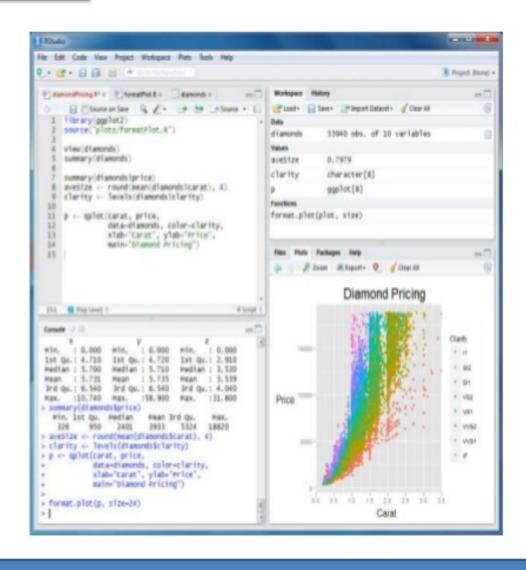


- A Package is a collection of R functions with comprehensive documents.
- A Package includes: R functions, Data Example, Help Files, Namespace and Description.
- The default installation is kept as minimum.
- The function of R could be extent by loading R packages.



R GUI







<u>Features</u>



Open Source:

The source code of R program and the extensions could be examined line by line.

Integrating with other Programming Language:

R is an interpreting language, can be rather slow, but could integrate with high efficient languages such as C, C++ or Fortran.

OS Independence:

UNIX, Linux, Windows, MacOS, FreeBSD...

Command line Driven:

You have to write Commands...



Comparison with other languages

R Programming	Python	Java
•It was stably released in 2014.	•It was stably released in 1996.	•It was stably released in 1995.
•It has more functions and packages.	 It has less functions and packages. 	 It has large number of inbuilt functions and packages.
•It is an interpreter base language	•It is an interpreter base language	 It is interpreter and compiled based language.
•It is statistical design and graphics programming language.	•It is general purpose language.	•It is general purpose programming language designed for web applications .
 It is difficult to learn and understand. 	 It is easy to understand. 	•It is easy to learn and understand.
•R is mostly use for data analysis.	•Generic programming tasks such as design of software's or	•Java is mostly used in design of windows applications and web





Merits



- R is the most comprehensive statistical analysis package available. It incorporates all of the standard statistical tests, models, and analyses, as well as providing a comprehensive language for managing and manipulating data.
- R is a programming language and environment developed for statistical analysis by practising statisticians and researchers.
- The graphical capabilities of R are outstanding, providing a fully programmable graphics language that surpasses most other statistical and graphical packages. ^





- R is free and open source software, allowing anyone to use and, importantly, to modify it. R is licensed under the GNU General Public License, with copyright held by The R Foundation for Statistical Computing.
- R has over 4800 packages available from multiple repositories specializing in topics like econometrics, data mining, spatial analysis, and bio-informatics.
- R is cross-platform. R runs on many operating systems and different hardware. It is popularly used on GNU/Linux, Macintosh, and Microsoft Windows, running on both 32 and 64 bit processors. ^



Demerits



R is slow:

Is an interpreting language and is not very fast. Could be 1/40 of C.

Limitation of Memory:

All the objects are in memory.

R is hard to learn:

One has to memorize the commands/functions, and understand the logics of programming. The fluency in R requires great time and energy.

THANK YOU

