

R Programming

R.MENAHA,
Asst.Professor
IT Department

Syllabus

INTRODUCTION TO R: 10

Basic R features- Script, Commands, Objects, Data Structures, Data Types, Vector, Matrix & Data Frames, Strings, Functions - Data visualization using R

STATISTICAL ANALYSIS: 10 Correlation- Types & Measures -Regression - Linear Regression, Logistic Regression, Principal Component Analysis - Dimensionality Reduction using PCA, Representation of PC

DATA MINING TECHNIQUES: 10

Association Rule mining, Classification using R -Decision Tree, Support Vector Machine, Cluster Analysis Using R- Partitioning Clustering -Hierarchical Clustering

References

1. Advanced R by Hadley Wickham
2. The Art of R Programming by Norman Matloff
3. ggplot2 by Hadley Wickham (pdf available via SpringerLink)
4. E-book: R Programming for Data Science
5. Dirk Hunniger :
http://DE.WIKIBOOKS.ORG/WIKI/BENUTZER:DIRK_HUENNIGER/WB2PDF

Web Resources:

1. R tool available at: <http://www.r-project.org/>
2. <http://www.r-tutor.com/>

Course Outcome

- Use and understand the R data types
- Write their own functions in R and break a problem into a set of functions
- Reshape data and use visual exploratory graphics
- Preparation and process the prepared data

Contents of Unit -1

- Basic R features
- Script
- Commands,
- Objects,
- Data Structures,
- Data Types,
- Vector, Matrix & Data Frames,
- Strings, Functions
- Data visualization using R

Session 1- Basic R Features

Outline

- Introduction to R Programming
- Why learn R
- Installation steps
- R Programming IDE
- R Packages
- What R does
- What R does not
- Features
- Advantages & Disadvantages
- Applications

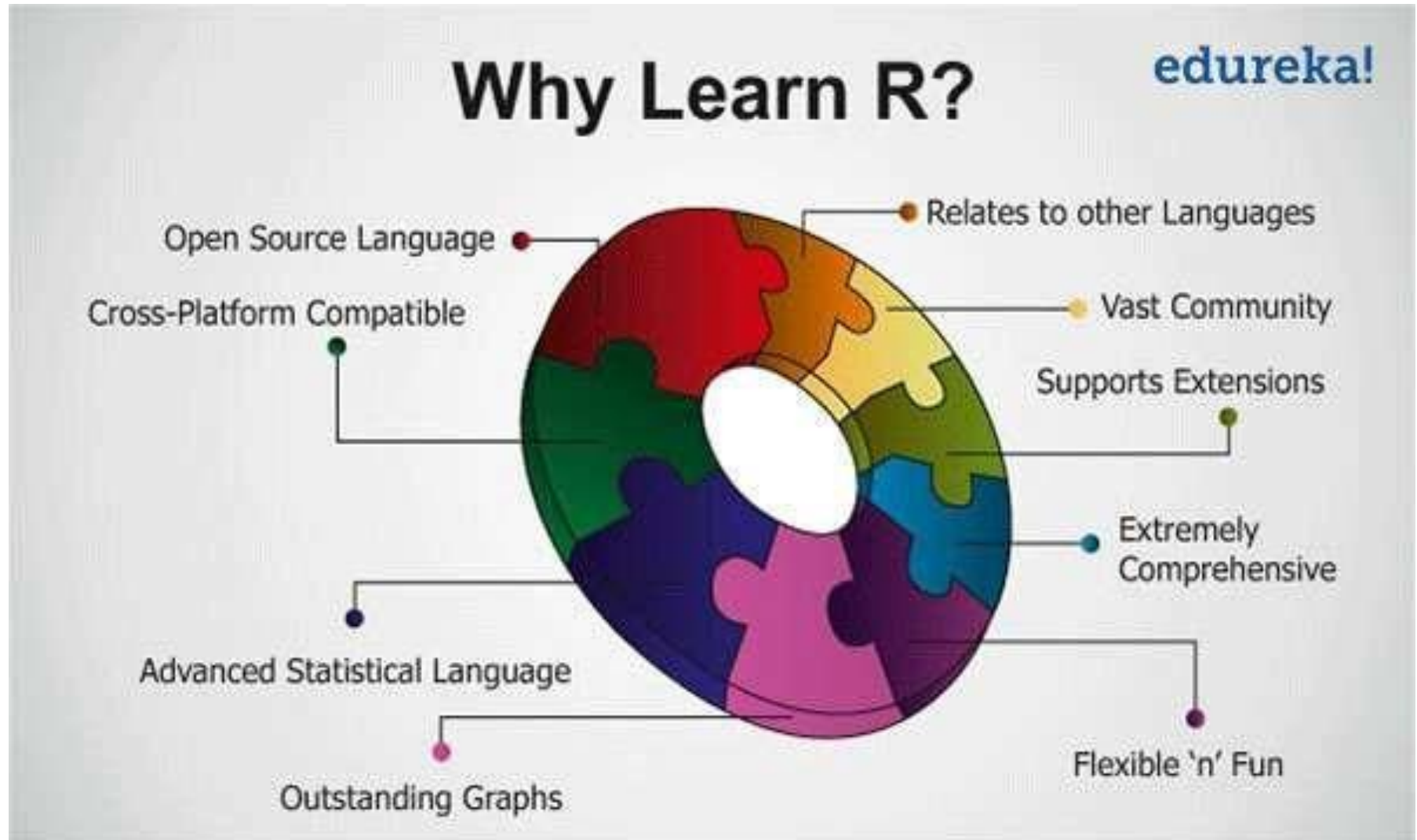
Introduction to R programming

- R is a programming language and software environment for statistical analysis, graphics representation and reporting.
- R was created by Ross Ihaka and Robert Gentleman at the University of Auckland, New Zealand, and is currently developed by the R Development Core Team.

Introduction to R programming

- The core of R is an interpreted computer language which allows branching and looping as well as modular programming using functions.
- R allows integration with the procedures written in the C, C++, .Net, Python or FORTRAN languages for efficiency.

Why Learn R?



INSTALLATION STEPS OF “R”

Getting Started with R

- [To install R GUI go to http://www.r-project.org/](http://www.r-project.org/)
- To install R Studio go to <http://www.rstudio.com/>



[Home]

Download

CRAN

R Project

About R

Logo

Contributors

What's New?

Reporting Bugs

Development Site

Conferences

Search

R Foundation

Foundation

Board

Members

Donors

Donate

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred CRAN mirror.

If you have questions about R like how to download and install the software, or what the license terms are, please read our answers to frequently asked questions before you send an email.

News

- **The R Journal Volume 8/1** is available.
- The **useR! 2017** conference will take place in Brussels, July 4 - 7, 2017, and details will be appear here in due course.
- **R version 3.3.1 (Bug in Your Hair)** has been released on Tuesday 2016-06-21.
- **R version 3.2.5 (Very, Very Secure Dishes)** has been released on 2016-04-14. This is a rebadging of the quick-fix release 3.2.4-revised.
- **Notice XQuartz users (Mac OS X)** A security issue has been detected with the Sparkle update mechanism used by XQuartz. Avoid updating over insecure



Select CRAN Mirrors (Click on India CRAN)

Greece

<http://cran.cc.uoc.gr/mirrors/CRAN/>

University of Crete

Hungary

<http://cran.rapporter.net/>

Rapporter.net, Budapest

Iceland

<https://cran.hafro.is/>

Marine Research Institute

<http://cran.hafro.is/>

Marine Research Institute

India

<https://ftp.iitm.ac.in/cran/>

Indian Institute of Technology Madras

<http://ftp.iitm.ac.in/cran/>

Indian Institute of Technology Madras

Iran

<http://cran.um.ac.ir/>

Ferdowsi University of Mashhad

Ireland

<https://ftp.heanet.ie/mirrors/cran.r-project.org/>

HEAnet,Dublin

<http://ftp.heanet.ie/mirrors/cran.r-project.org/>

HEAnet,Dublin

Italy

<http://cran.mirror.garr.it/mirrors/CRAN/>

Garr Mirror, Milano

<https://cran.stat.unipd.it/>

University of Padua

<http://cran.stat.unipd.it/>

University of Padua

<http://dssm.unipa.it/CRAN/>

Universita degli Studi di Palermo

Japan

<https://cran.ism.ac.jp/>

The Institute of Statistical Mathematics, Tokyo

<http://cran.ism.ac.jp/>

The Institute of Statistical Mathematics, Tokyo

Korea

<http://cran.nexr.com/>

NexR Corporation, Seoul

<http://healthstat.snu.ac.kr/CRAN/>

Graduate School of Public Health, Seoul National University, Seoul



Select OS Version which you want to download and install



The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)



R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release (Tuesday 2016-06-21, Bug in Your Hair) [R-3.3.1.tar.gz](#), read [what's new](#) in the latest version.
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).

CRAN

[Mirrors](#)

[What's new?](#)

[Task Views](#)

[Search](#)

About R

[R Homepage](#)

[The R Journal](#)

Software

[R Sources](#)

[R Binaries](#)

[Packages](#)

[Other](#)

Documentation

[Manuals](#)

[FAQs](#)

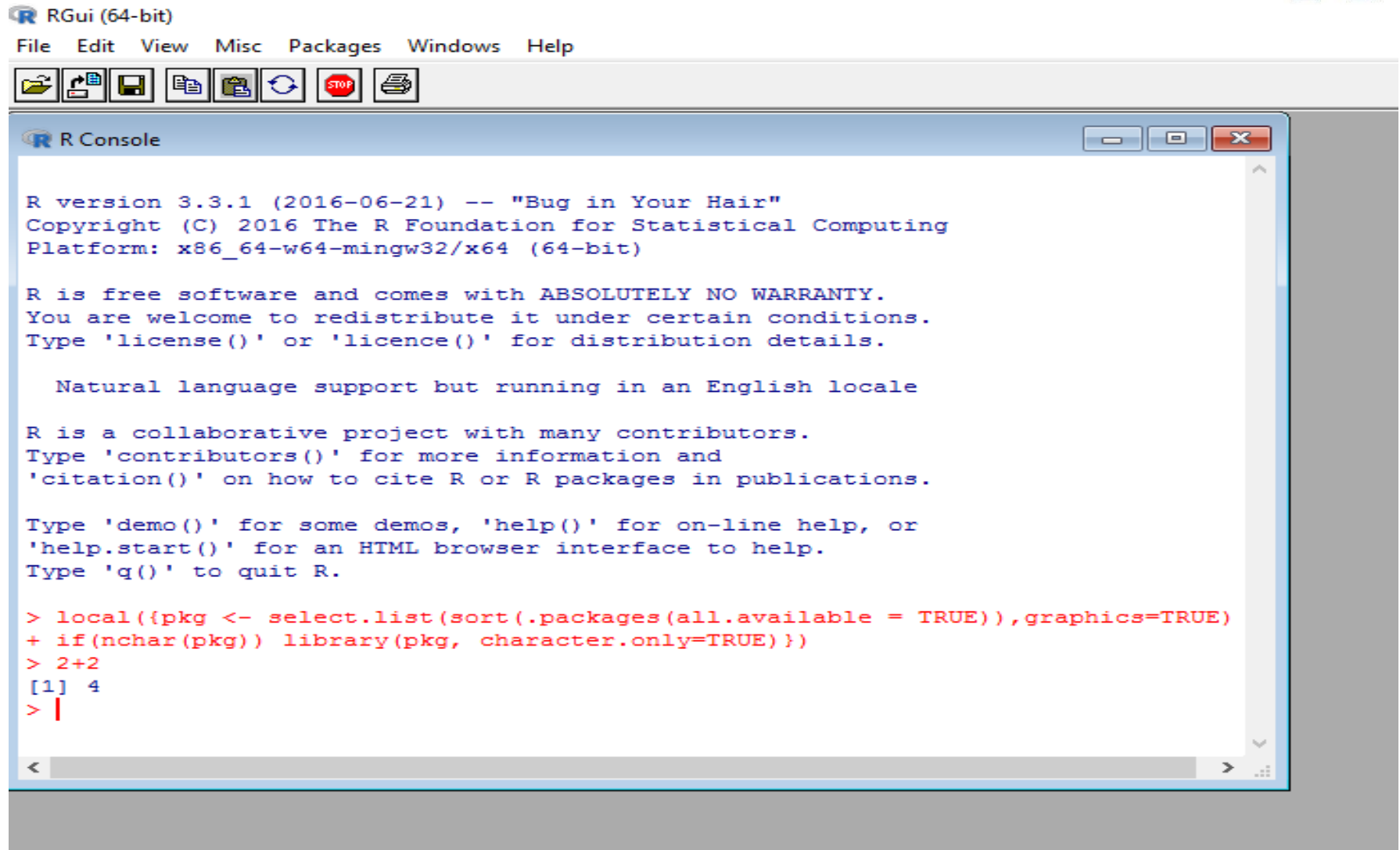
[Contributed](#)

R PROGRAMMING IDE

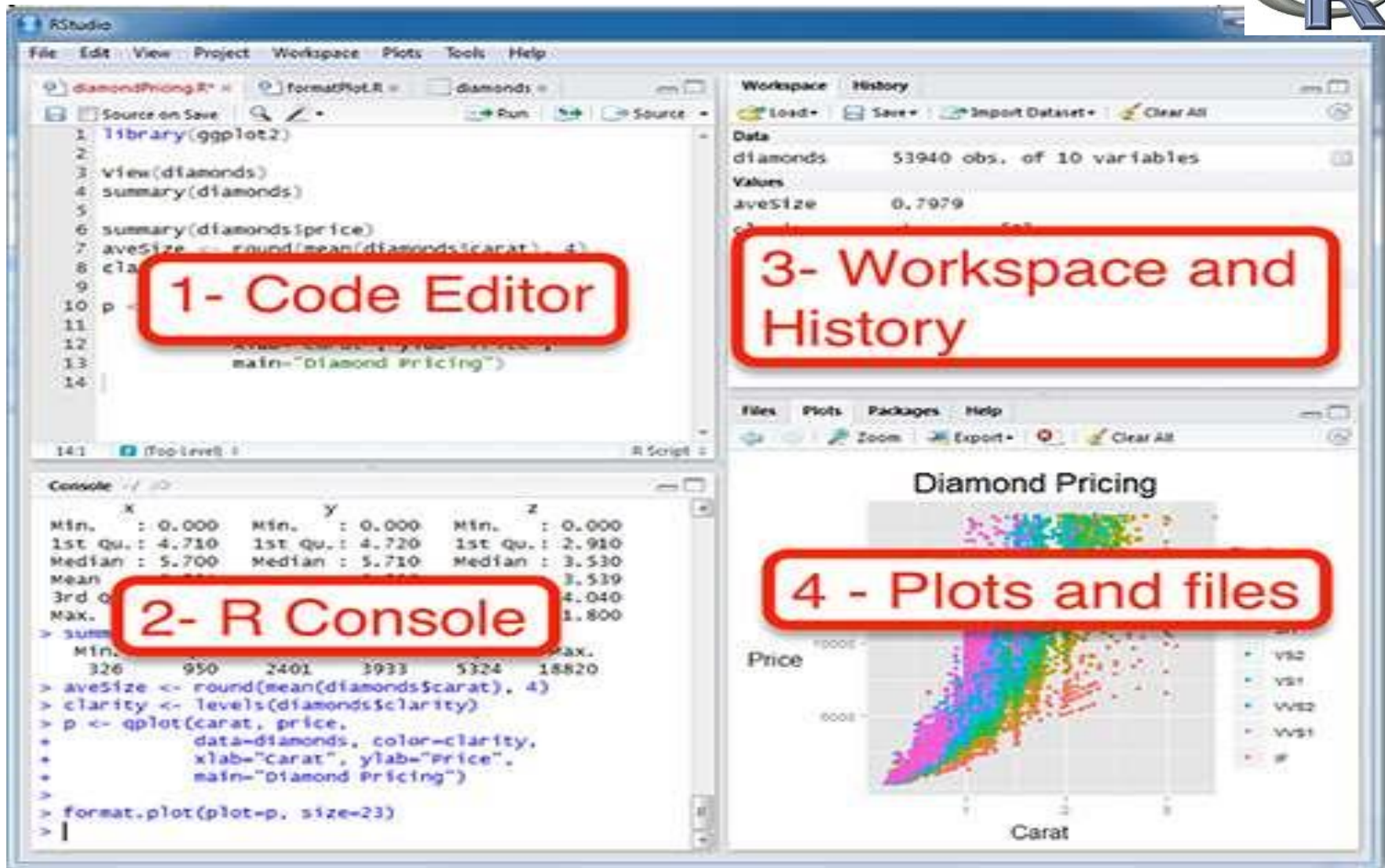
□ There are two Integrated Development Environments for R Programming:-

- i) R GUI (Graphical User Interface)
- ii) R Studio

R GUI



R Studio



The screenshot displays the RStudio environment with four key components highlighted by red boxes and numbers:

- 1- Code Editor:** The top-left pane shows R code for loading ggplot2, viewing the 'diamonds' dataset, and creating a scatter plot of Price vs. Carat, colored by clarity.
- 2- R Console:** The bottom-left pane shows the output of the code, including summary statistics for the 'diamonds' dataset and the execution of the plot command.
- 3- Workspace and History:** The top-right pane shows the 'Workspace' tab with the 'diamonds' dataset loaded, and the 'History' tab showing the executed commands.
- 4- Plots and files:** The bottom-right pane shows the 'Plots' tab with a scatter plot titled 'Diamond Pricing' showing Price vs. Carat, colored by clarity.

What R does

- data handling and storage: numeric, textual
- matrix algebra
- hash tables
- high-level data analytic and statistical functions
- classes (“Object Oriented”)
- graphics
- programming language: loops, branching, subroutines

What R does not

- R is not a database, but connects to DBMSs
- has no graphical user interfaces, but connects to Java
- language interpreter can be very slow, but allows to call own C/C++ code
- no spreadsheet view of data, but connects to Excel/MsOffice
- no professional /commercial support

Features of R

- R is a well-developed, simple and effective programming language which includes conditionals, loops, user defined recursive functions and input and output facilities.
- R has an effective data handling and storage facility
- R provides a suite of operators for calculations on arrays, lists, vectors and matrices.
- R provides a large, coherent and integrated collection of tools for data analysis.
- R provides graphical facilities for data analysis and display either directly at the computer or printing at the papers.

Advantages of R



- R is **free** and **open source** software.
- R has no **license restrictions**.
- 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 plays well with many other tools, importing data, for example, from CSV or directly from Microsoft Excel, Microsoft Access, Oracle, MySQL, and SQLite.
- It can also produce graphics output in PDF, JPG, PNG, and SVG formats, and table output for LATEX and HTML.

How R Programming Is Applied To Real World



R Programming has turned into the most prevalent language for data science and a fundamental tool for Finance and analytics-driven organizations, for example, Google, Facebook, and LinkedIn.

