DATA SCIENCE WITH R

Class 1

Introduction to Analytics and R

DATA SCIENCE WITH R

Class 1 - Introduction to Analytics and R

Topic 1

★ Understanding Analytics & ★

Analytics tools

AGENDA

- **Understanding Analytics & Analytics tools**

  Exploring R

Understanding Analytics

Data Analysis Vs. Reporting

Features of basic and advanced analytics

Pre-requisites to consider while planning an analytics project

What can make a good analytics team ?

Understanding Analytics

Data Analysis Vs. Reporting

Reporting	Analysis
Provides data	Provides answers
Provides what is asked for	Provides what is needed
Is typically standardised	Is typically customised
Does not involve a person	Involves a person leading the process
Is fairly inflexible	Is extremely flexible

Understanding Analytics

Data Analysis Vs. Reporting

Interplay between reporting and analysis is common and necessary.

For Example :

Report : A company generates basic sales summary report showing monthly sales by region.

Analysis : You observe what is the trend in sales over a period of time and you find out that Sales peaks on specific holidays or weekends

Understanding Analytics

Data Analysis Vs. Reporting

Features of basic and advanced analytics

Pre-requisites to consider while planning an analytics project

What can make a good analytics team ?

Understanding Analytics

Features of basic and advanced analytics

Basic Analytics

Not sure what you have? But you think something is of value ?

- Slicing and Dicing of the data
- Monitor large volumes of data in real time
- Anomaly Identification

Advanced Analytics

What happened, what caused it to happen, when it happened, and what the impact was?

- Predictive Modelling
- Text Analytics
- Other Statistical and Data-Mining Algorithms

Understanding Analytics

Data Analysis Vs. Reporting

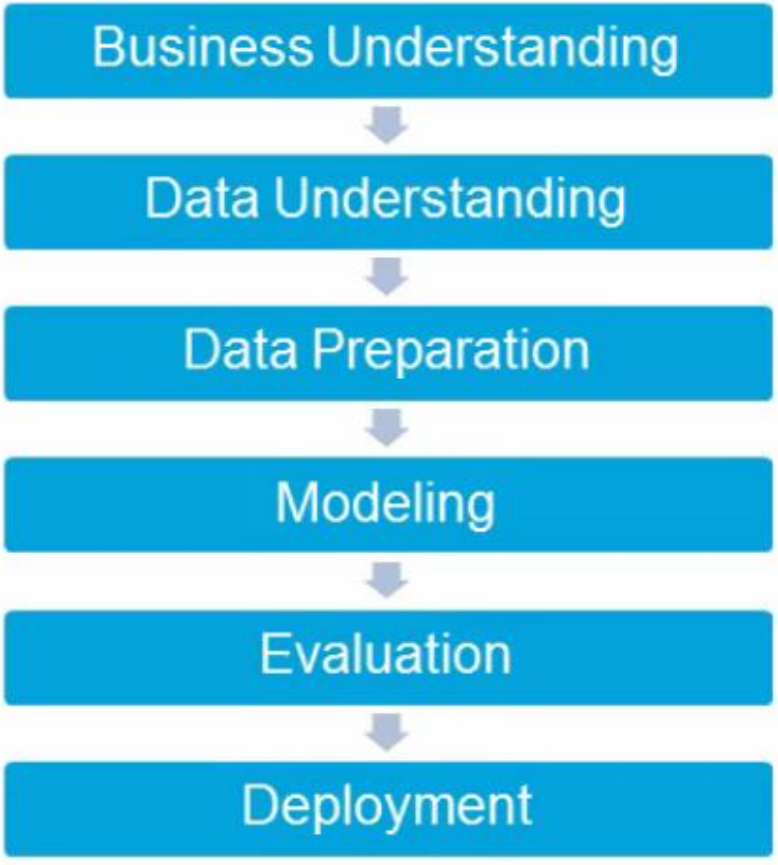
Features of basic and advanced analytics

Pre-requisites to consider while planning an analytics project

What can make a good analytics team ?

Understanding Analytics

Pre-requisites - Analytics Process



Analytical Tools

- Categories of analytical tools
  - Features and importance of R
  - Features and importance of the IBM SPSS tool
  - Features and importance of the SAS tool
- Compare various analytical tools

Analytical Tools

Why do we need specialized analytical tools?

Depends on volume of data and complexity of analysis

For example, maximum number of rows of data allowed in Excel – 1 MM records

- Advanced data mining algorithms are not implemented in Excel
- Specialized data mining, predictive analytics, and visualization tools have been built to effectively analyse large volumes of data

Analytical Tools

Multiple types of Analytical Tools:

- **GUI based tools:** Excel, SPSS, SAS ,RStudio
- **Visualization tools**: Tableau, MicroStrategy
- **Coding based tools**: SAS, R

Analytical Approaches and Tools

Categories of Analytical Tools

Graphic User Interface

- **Advantage** : User interface helps generate code automatically hence making it bug-free and optimized.

- **Limitations** : The ability to generate code quickly also makes it easy to generate bad code faster. Without the ability to understand the code generated, a user will not be able to identify such situations

Analytical Approaches and Tools

Categories of Analytical Tools

Point Solutions

- Analytic point solutions are software packages that address a very specific, narrow set of problems(business) and they often sit on top of analytical tool suites.

- **Limitations** : Point solutions are expensive

Examples : Price optimization applications, fraud applications, and demand forecasting Applications,etc

Analytical Approaches and Tools

Categories of Analytical Tools

Data Visualization Tools

- In the world of analytics, visualization refers to charts, graphs, and tables that display data and help in data analysis

  Example : In a retail store ,find out which parts of the store are most frequently visited by the customers.

  - **Option 1** : Create a bunch of spreadsheets, lay them out on a table, and try to figure out the patterns.
  - **Option 2** :Produce a map of the store floor plan where colour represents the level of activity.

Hence Visualisations are quite powerful

Analytical Approaches and Tools

Popular Analytical Tools

R Project for Statistical Computing

Features:

- Open Source Software
- Object Oriented Design.
- Can be linked with common programming platforms like C++ and Java, which makes it possible to embed R within applications.
- Commercial analytic tools like SAS,etc have even enabled R to be executed within their toolsets.
- Extensibility : Since R codes can be easily written and packaged, its easy to create and distribute them as a package

Analytical Approaches and Tools

Popular Analytical Tools

R Project for Statistical Computing

Disadvantages:

- **Scalability** : Since the base R software runs in memory ,it can only handle datasets of the size of memory available on a machine.

Analytical Approaches and Tools

Popular Analytical Tools

IBM SPSS

Syntax Language + graphic interface of scrolling menus

Important Features :

- Commands are executed line by line to update tables or add results to the Output Editor window
- Statistics can read from and write to ASCII files, databases and tables of other statistical software
- Provides basic data management functions, such as sorting, aggregation, transposition, and table merge
- The file can be in various formats

Analytical Approaches and Tools

Popular Analytical Tools

SAS

A SAS program consists of **DATA steps, procedure steps,** and **macros**

It can be used for the following:

- Statistics
- Data and Text mining
- Data Visualization
- Forecasting
- Optimization
- Model Management and Deployment
- Quality Improvement

Analytical Approaches and Tools

Popular Analytical Tools

Comparisons

	SAS	SPSS	R
User Interface	Has the SAS Enterprise Guide	Graphic interface with scrolling menus	R Graphic interfaces like Rattle, Rstudio, etc
Decision Trees	Expensive, Purchase SAS Miner	Less Expensive, Does not require Data mining suite, offers a wide variety of algorithms	Free, Does not offer many tree algorithms
File Management and Stability	More flexible	Freezes when processing large volumes	Not suitable for large volumes of data
Data Management	No memory Issues	No memory Issues	Memory issues
Documentation	Extensive	Not so extensive	Extensive (CRAN)

DATA SCIENCE WITH R

Class 1 - Introduction to Analytics and R

Topic 2

★ Exploring R ★

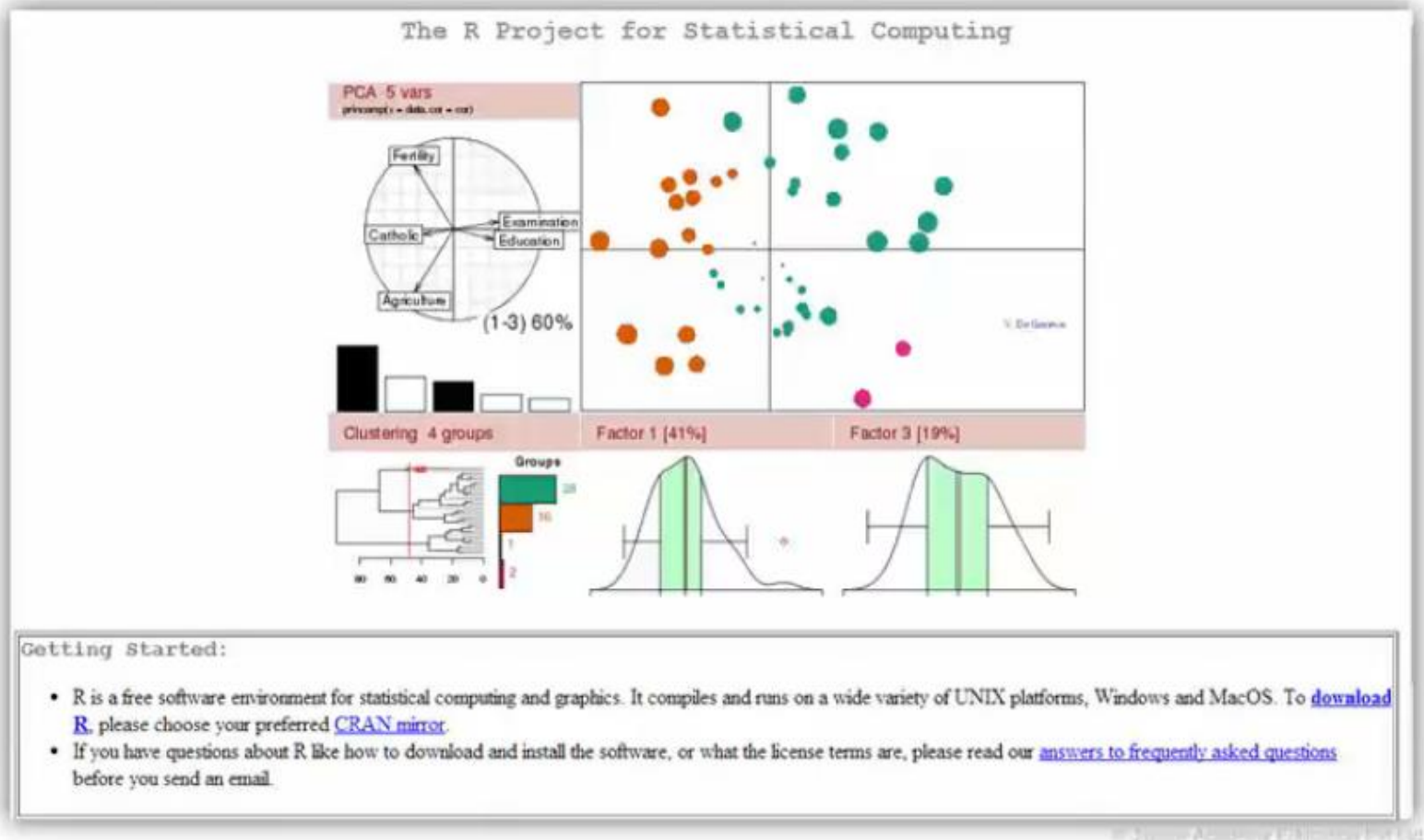
Exploring R

- Installing R
- Working with Scripts
- Execute scripts in R using the various add-on editors available for R

Exploring R

Installing R

You can obtain the installation files for the R program on the official R Website (www.r-project.org)

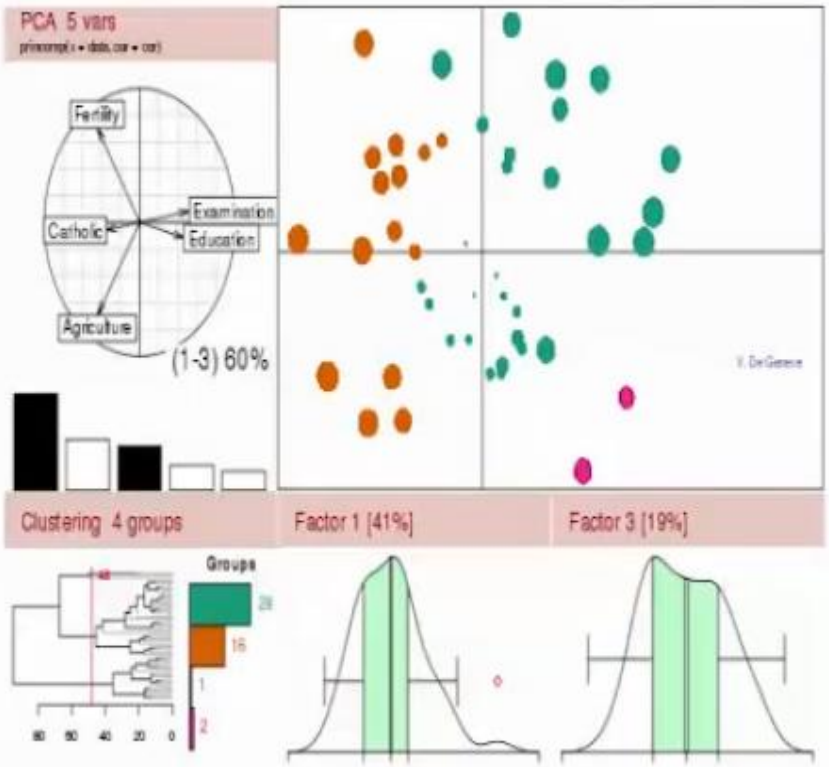


The R Project for Statistic | CRAN - Mirrors | The Comprehensive R Ar... | install rstudio on windows | RStudio - RStudio | RStudio - RStudio

www.r-project.org

Apps | Text Analytics Essen... | DECISION STATS « B... | Distance Matrix by G... | Hierarchical Cluster ... | charuaggarwal.net/t...

The R Project for Statistical Computing



About R
- What is R?
- Contributors
- Screenshots
- What's new?

Download, Packages
- CRAN

R Project
- Foundation
- Members & Donors
- Mailing Lists
- Bug Tracking
- Developer Page
- Conferences
- Search

Documentation
- Manuals
- FAQs
- The R Journal
- Wiki
- Books
- Certification
- Other

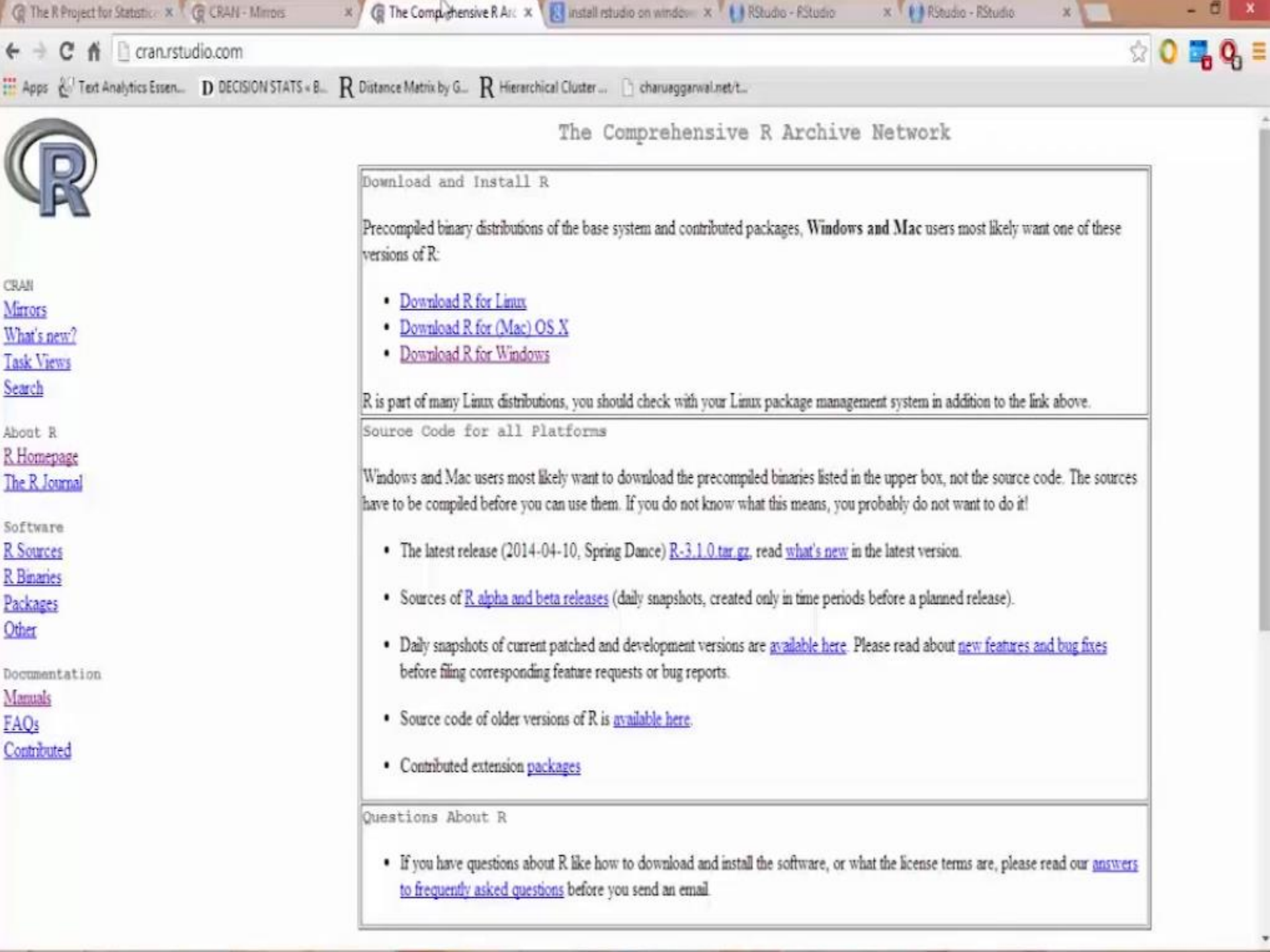
Getting Started:

- 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:

0-Cloud

http://cran.rstudio.com/	Rstudio, automatic redirection to servers worldwide
Argentina	
http://mirror.fcaglp.unlp.edu.ar/CRAN/	Universidad Nacional de La Plata
http://r.mirror.mendoza-conicet.gob.ar/	CONICET Mendoza
Australia	
http://cran.csiro.au/	CSIRO
http://cran.ms.unimelb.edu.au/	University of Melbourne
Austria	
http://cran.at.r-project.org/	Wirtschaftsuniversitaet Wien
Belgium	
http://www.freestatistics.org/cran/	K.U.Leuven Association
Brazil	
http://nbcgib.uesc.br/mirrors/cran/	Center for Comp. Biol. at Universidade Estadual de Santa Cruz
http://cran-r.c3sl.ufpr.br/	Universidade Federal do Parana
http://cran.fiocruz.br/	Oswaldo Cruz Foundation, Rio de Janeiro
http://www.vps.fmvz.usp.br/CRAN/	University of Sao Paulo, Sao Paulo
http://brieger.esalq.usp.br/CRAN/	University of Sao Paulo, Piracicaba
Canada	
http://cran.stat.sfu.ca/	Simon Fraser University, Burnaby
http://mirror.its.dal.ca/cran/	Dalhousie University, Halifax
http://cran.utstat.utoronto.ca/	University of Toronto
http://cran.skazkaforyou.com/	iWeb, Montreal
http://cran.parentingamerica.com/	iWeb, Montreal
Chile	
http://dirichlet.mat.puc.cl/	Pontificia Universidad Catolica de Chile, Santiago
China	
http://ftp.ctex.org/mirrors/CRAN/	CTEX.ORG


The R Project for Statistic
CRAN - Mirrors
The Comprehensive R Arc
install rstudio on windows
RStudio - RStudio
RStudio - RStudio
cran.rstudio.com
Apps
Text Analytics Essen...
DECISION STATS « B...
Distance Matrix by G...
Hierarchical Cluster ...
charuaggarwal.net/t...
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 (2014-04-10, Spring Dance) R-3.1.0.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.
Contributed extension packages
Questions About R
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.
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

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 for Windows

Subdirectories:

base	Binaries for base distribution (managed by Duncan Murdoch). This is what you want to install R for the first time.
contrib	Binaries of contributed packages (managed by Uwe Ligges). There is also information on third party software available for CRAN Windows services and corresponding environment and make variables.
Rtools	Tools to build R and R packages (managed by Duncan Murdoch). This is what you want to build your own packages on Windows, or to build R itself.

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the R FAQ and R for Windows FAQ.

Note: CRAN does some checks on these binaries for viruses, but cannot give guarantees. Use the normal precautions with downloaded executables.

R-3.1.0 for Windows (32/64 bit)

Download R 3.1.0 for Windows (54 megabytes, 32/64 bit)

Installation and other instructions

New features in this version

If you want to double-check that the package you have downloaded exactly matches the package distributed by R, you can compare the md5sum of the .exe to the true fingerprint. You will need a version of md5sum for windows: both graphical and command line versions are available.

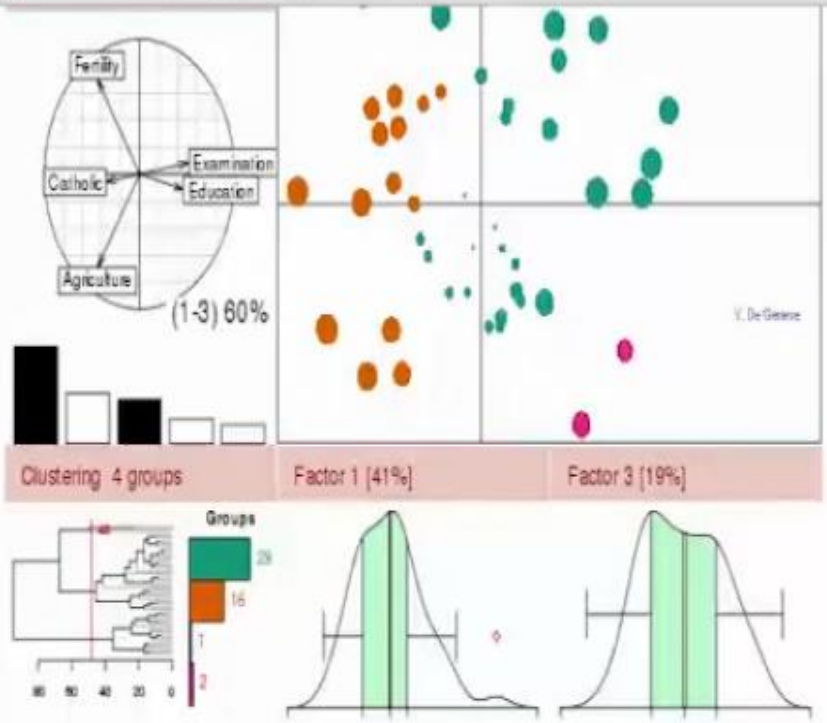
Frequently asked questions

- How do I install R when using Windows Vista?
- How do I update packages in my previous version of R?
- Should I run 32-bit or 64-bit R?

Please see the R FAQ for general information about R and the R Windows FAQ for Windows-specific information.

Other builds

- Patches to this release are incorporated in the r-patched snapshot build
- A build of the development version (which will eventually become the next major release of R) is available in the r-devel snapshot build
- Previous releases



Contributors
Screenshots
What's new?

Download, Packages
CRAN

R Project
Foundation
Members & Donors
Mailing Lists
Bug Tracking
Developer Page
Conferences
Search

Documentation
Manuals
FAQ
The R Journal
Wiki
Books
Certification
Other

Misc
Bioconductor

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:

- **R version 3.1.0** (Spring Dance) has been released on 2014-04-10.
- **R version 3.0.3** (Warm Puppy) has been released on 2014-03-06.
- The R Journal Vol.5/2 is available.
- ...lace at the University of Castilla-La Mancha, Albacete, Spain, July 10-12 2013.

The R Manuals

edited by the R Development Core Team

The following manuals for R were created on Debian Linux and may differ from the manuals for Mac or Windows on platform-specific pages, but most parts will be identical for all platforms. The correct version of the manuals for each platform are part of the respective R installations. The manuals change with R, hence we provide versions for the most recent released R version (R-release), a very current version for the patched release version (R-patched) and finally a version for the forthcoming R version that is still in development (R-devel).

Here they can be downloaded as PDF files, EPUB files, or directly browsed as HTML:

Manual	R-release	R-patched	R-devel
An Introduction to R is based on the former "Notes on R", gives an introduction to the language and how to use R for doing statistical analysis and graphics.	HTML | PDF | EPUB	HTML | PDF | EPUB	HTML | PDF | EPUB
R Data Import/Export describes the import and export facilities available either in R itself or via packages which are available from CRAN.	HTML | PDF | EPUB	HTML | PDF | EPUB	HTML | PDF | EPUB
R Installation and Administration	HTML | PDF | EPUB	HTML | PDF | EPUB	HTML | PDF | EPUB
Writing R Extensions covers how to create your own packages, write R help files, and the foreign language (C, C++, Fortran, ...) interfaces.	HTML | PDF | EPUB	HTML | PDF | EPUB	HTML | PDF | EPUB
A draft of **The R language definition** documents the language *per se*. That is, the objects that it works on, and the details of the expression evaluation process, which are useful to know when programming R functions.	HTML | PDF | EPUB	HTML | PDF | EPUB	HTML | PDF | EPUB
R Internals: a guide to the internal structures of R and coding standards for the	HTML | PDF | EPUB	HTML | PDF | EPUB	HTML | PDF | EPUB

Exploring R

Installing R

The two most useful sections are the Documentation and Download sections.

- These documentations and manuals are especially useful for helping new users to get started with R
- In the Downloads section, links for downloading R for various operating systems are available.

Exploring R

Working with Scripts

Windows R editors are **RGUI** and **RStudio**

RGUI

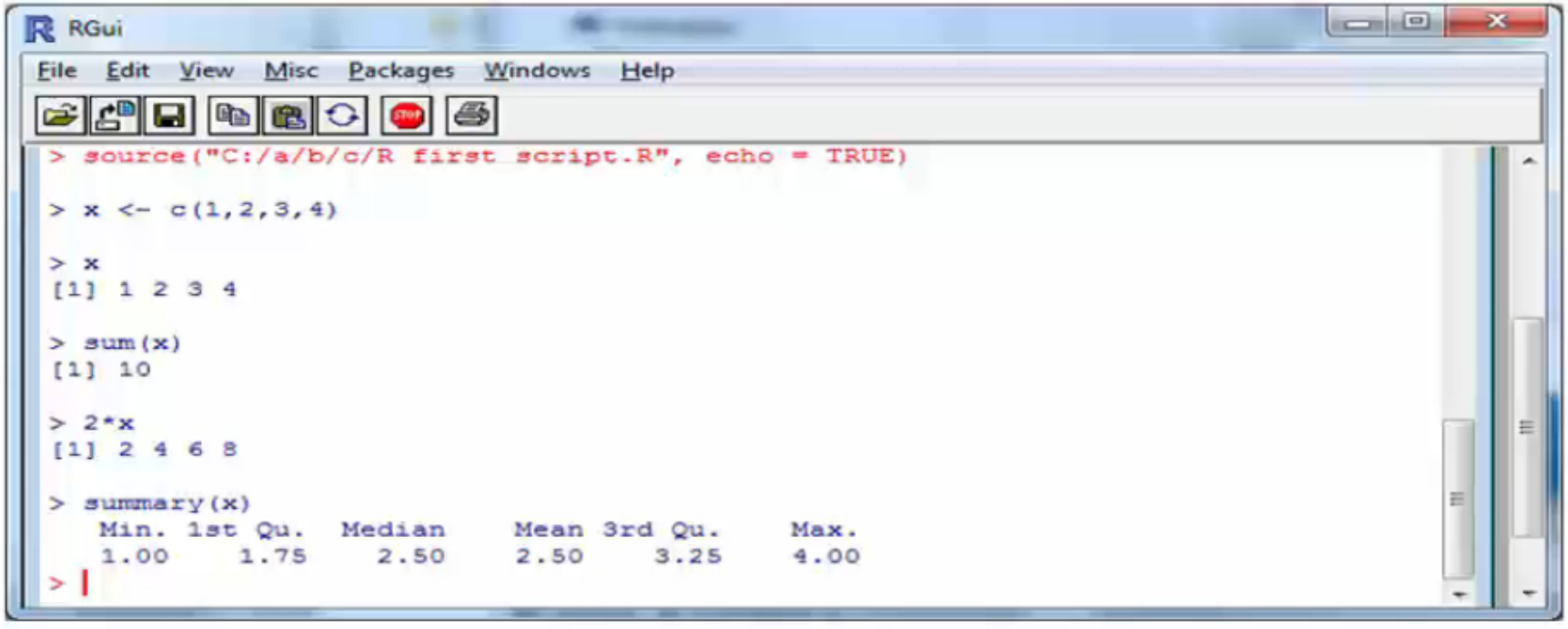
You are provided with a standard Graphical User Interface (GUI) called RGui on downloading and installing R.

Console window :

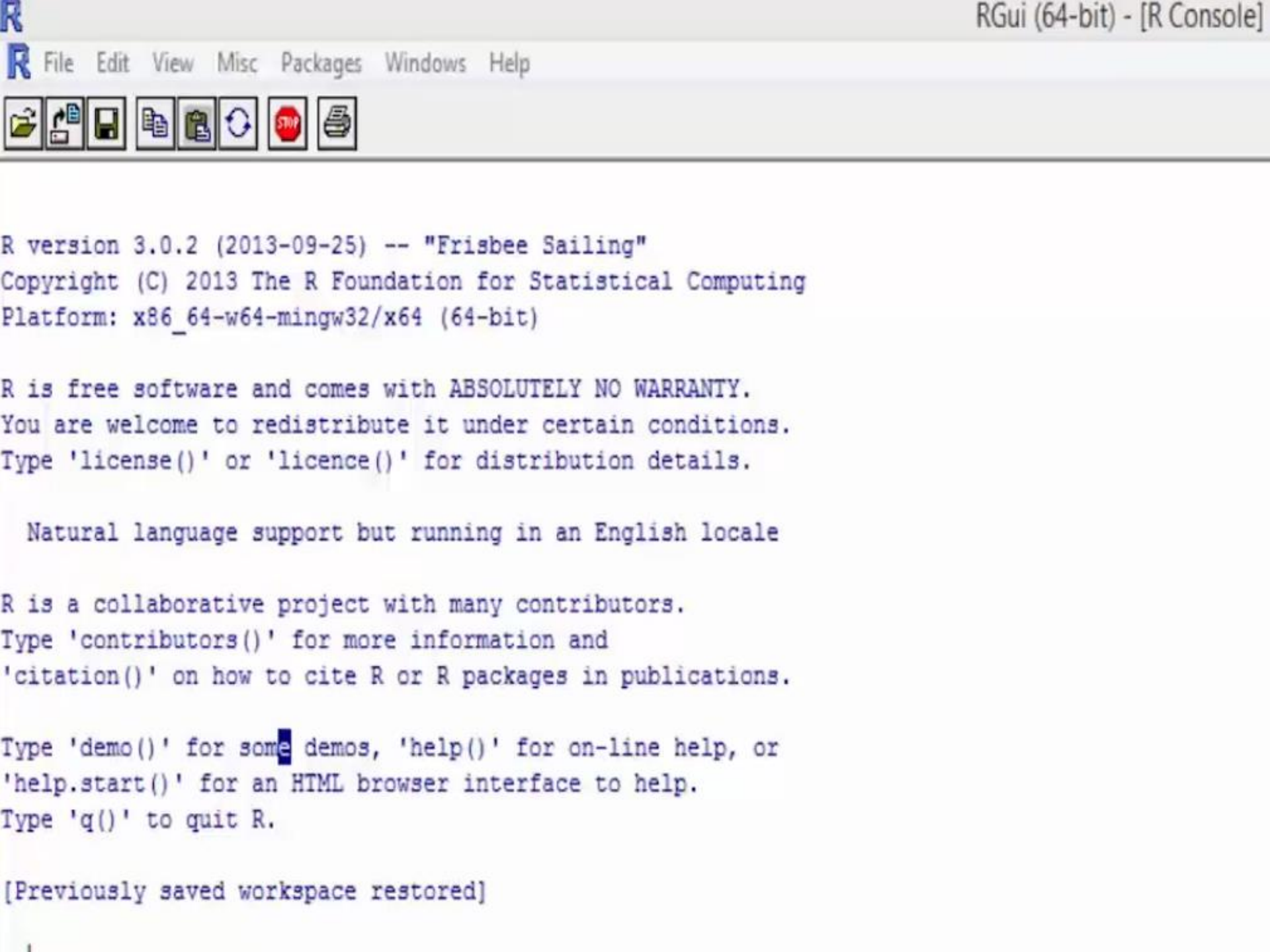
- It appears every time the RGui is opened (once you click the R icon in the start menu)
- This is a place where instructions, scripts, and general R operations are performed
- The '>' symbol is the R prompt, which indicates the place where the user can enter commands

Exploring R

The R Console



- To quit R use : >q()
- Click Yes or No


RGui (64-bit) - [R Console]
File Edit View Misc Packages Windows Help
R version 3.0.2 (2013-09-25) -- "Frisbee Sailing"
Copyright (C) 2013 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.
Natural language support but running in an English locale
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.
[Previously saved workspace restored]

Exploring R

Working with Scripts

RStudio is a code editor and development environment with some nice features that make code development in R easy and fun, such as:

- Code highlighting, making it easier to read
- Automatic bracket matching
- Code completion, so as to reduce the effort of typing the commands in full
- Easy access to R Help
- Easy exploration of variables and values

https://www.google.co.in/?gfe_rd=cr&ei=Ap6aU7bUG6Ow8wf3k4DQBA#q=install+rstudio+on+windows+

Apps | Text Analytics Essen... | DECISION STATS « B... | Distance Matrix by G... | Hierarchical Cluster ... | charuaggarwal.net/t...

Google

install rstudio on windows

+Abhirami | Share

Web | Videos | Images | News | More | Search tools

About 13,70,000 results (0.46 seconds)

Download RStudio - RStudio
www.rstudio.com/ide/download/desktop
Apr 9, 2014 - RStudio 0.98.507 - Windows XP/Vista/7/8. 35.4 MB. 2014-04-25. c4cf4f26d647687b06efdca3da66a419. RStudio 0.98.507 - Mac OS X 10.6+ ...

RStudio - RStudio
www.rstudio.com/products/rstudio/
Apr 4, 2014 - ... and commercial editions and runs on the desktop (Windows, Mac, and Linux) or over the web with RStudio Server and RStudio Server Pro.

RStudio: Home
www.rstudio.com/
Mar 27, 2014 - RStudio IDE is a powerful and productive user interface for R. It's free and open source, and works great on Windows, Mac, and Linux.

Install R Studio Server on Windows - Stack Overflow
stackoverflow.com/questions/.../install-r-studio-server-on-windows
May 28, 2013 - Is it possible to install R Studio Server on a Windows machine? I know ... Thanks! How do you mean "add the iso ... oh yeah, once you have ...

Install R, RStudio, and R Commander in Windows and OS ...

www.andrewheiss.com/.../install-r-rstudio-r-commander-windo...
by Andrew Heiss - in 259 Google+ circles
Apr 17, 2012 - R, RStudio, and R Commander are all powerful open source ...

Results from browsing history

- install rstudio on windows - Google Search
- install rstudio on windows - Google Search
- install rstudio on windows desk - Google Search
- install rstudio on windows desktop - Google Search

View all 4 results

www.rstudio.com/products/rstudio/

Apps | Text Analytics Essen... | DECISION STATS « B... | Distance Matrix by G... | Hierarchical Cluster ... | charuaggarwal.net/t...

RStudio

Products Resources Pricing About Us Blog

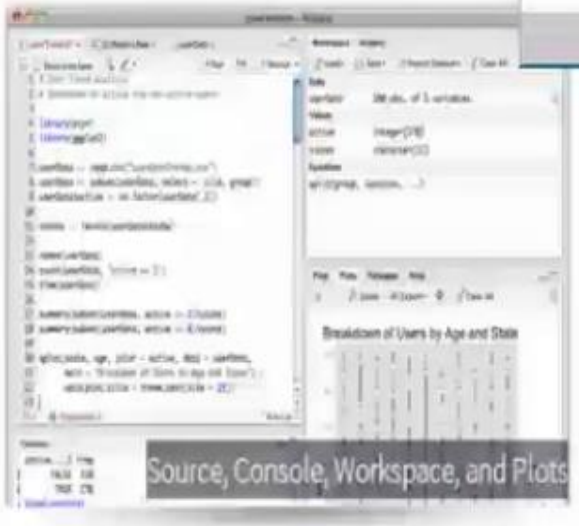
RStudio

Take control of your R code

RStudio is a set of integrated tools designed to help you be more productive with R. It includes a console, syntax-highlighting editor that supports direct code execution, as well as tools for plotting, history, debugging and workspace management. Click here to see more RStudio features.

RStudio is the premier integrated development environment for R. It is available in open source and commercial editions and runs on the desktop (Windows, Mac, and Linux) or over the web with RStudio Server and RStudio Server Pro.

VIEW MORE FEATURES



Run RStudio on your desktop

RStudio Desktop »

Centralize access and computation

RStudio Server »

www.rstudio.com/products/rstudio/#Desk

Apps | Text Analytics Essen... | DECISION STATS « B... | Distance Matrix by G... | Hierarchical Cluster ... | charuaggarwal.net/t...

RStudio

Products Resources Pricing About Us Blog

- Extensive package development tools

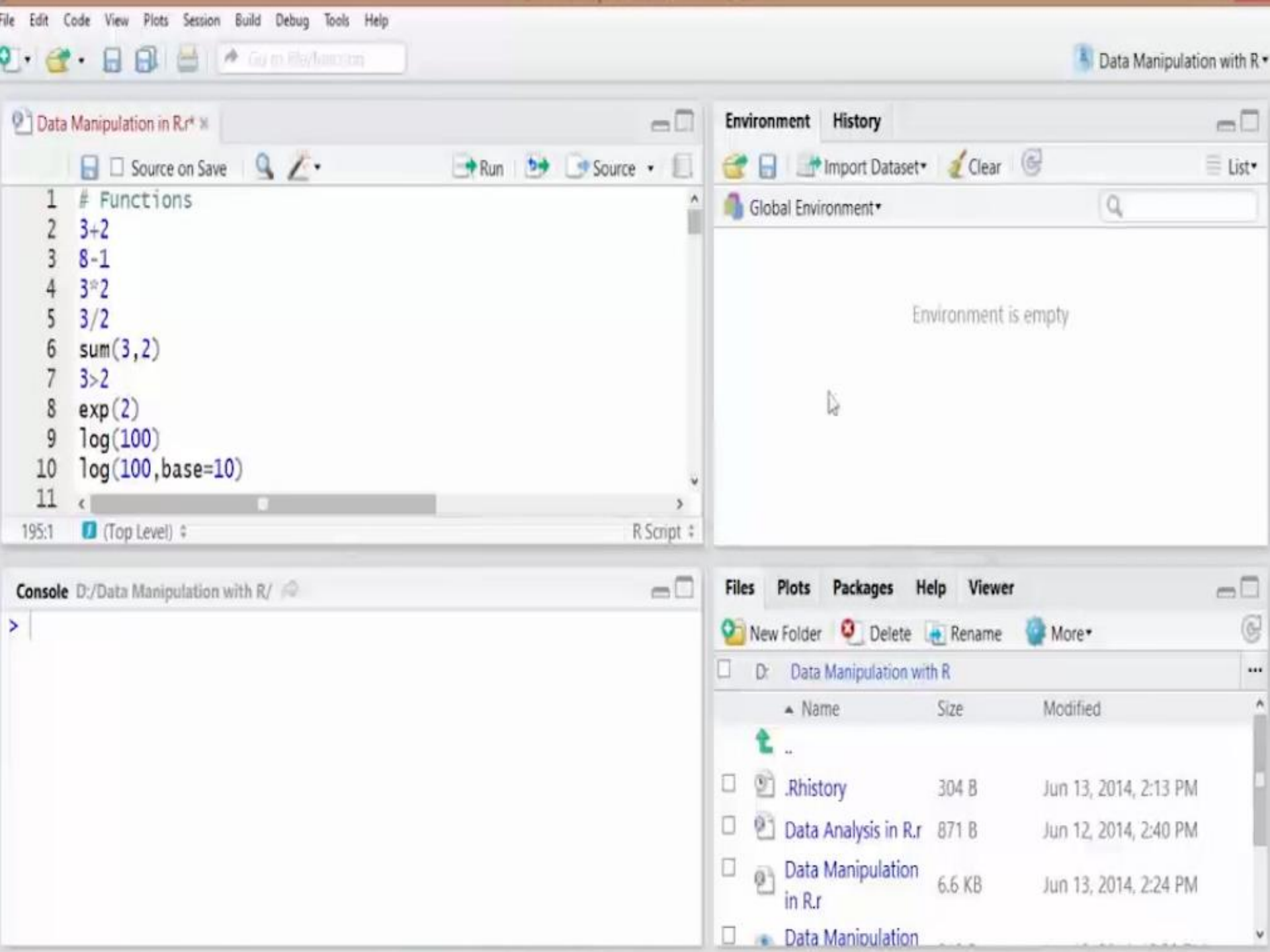
Support	Community forums only	• Priority Email Support • 8 hour response during business hours (ET)
License	AGPL v3	Commercial Desktop License
Pricing	Free	$995/year
	DOWNLOAD RSTUDIO DESKTOP	BUY NOW

RStudio Server

Sundance Web

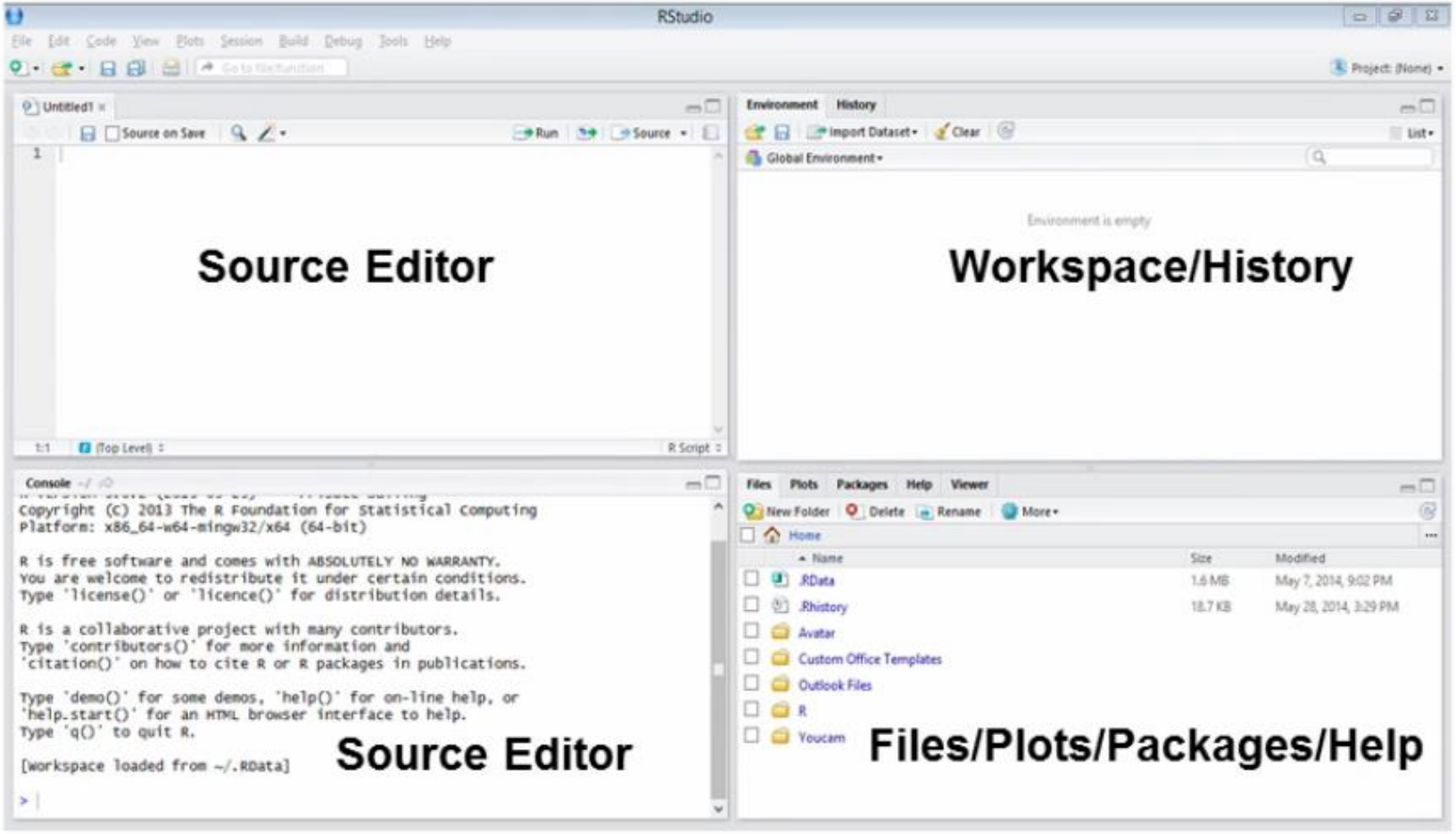
Download

	Open Source Edit	Professional Edition
Overview	• Access via a web browser • Move computation closer to the data • Scale compute and RAM centrally	• All of the features of open source; plus: • Administrative Tools • Enhanced Security and Authentication • Metrics and Monitoring • Advanced Resource Management


File Edit Code View Plots Session Build Debug Tools Help
Data Manipulation with R
Data Manipulation in R.r*
Source on Save
Run
Source
1 # Functions
2 3+2
3 8-1
4 3*2
5 3/2
6 sum(3,2)
7 3>2
8 exp(2)
9 log(100)
10 log(100,base=10)
11
195:1
(Top Level)
R Script
Environment
History
Import Dataset
Clear
List
Global Environment
Environment is empty
Console D:/Data Manipulation with R/
>
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
D: Data Manipulation with R
Name
Size
Modified
..
.Rhistory 304 B Jun 13, 2014, 2:13 PM
Data Analysis in R.r 871 B Jun 12, 2014, 2:40 PM
Data Manipulation in R.r 6.6 KB Jun 13, 2014, 2:24 PM
Data Manipulation

Exploring R

Working with Scripts



Exploring R

Working with Scripts

Source: (top left corner)

- Contains a text editor where multiple lines of code(highlights various elements) can be entered here.
- Users can save script file to disk, and perform other tasks on the script

Console: (bottom left corner)

- The console in RStudio is identical to the console in RGui.
- All the interactive work of R is performed here.

Exploring R

Working with Scripts

Workspace and History: (top right corner)

- Overview of the workspace, where the variables created in the session along with their values can be inspected.
- This is also the area where the user can see a history of the commands issued in R

Exploring R

Working with Scripts

Files, Plots, Package, and Help: (bottom right corner)

- **Files:** This is where the user can browse folders and files on a computer.
- **Plots:** This is where R displays the user's plots.
- **Packages:**
  - This is where the user can view a list of all the installed packages
  - A package is a self-contained set of code that adds functionality to R, similar to the way an add-in adds functionality to Microsoft Excel
- **Help:** This is where you can browse the built-in Help system of R.

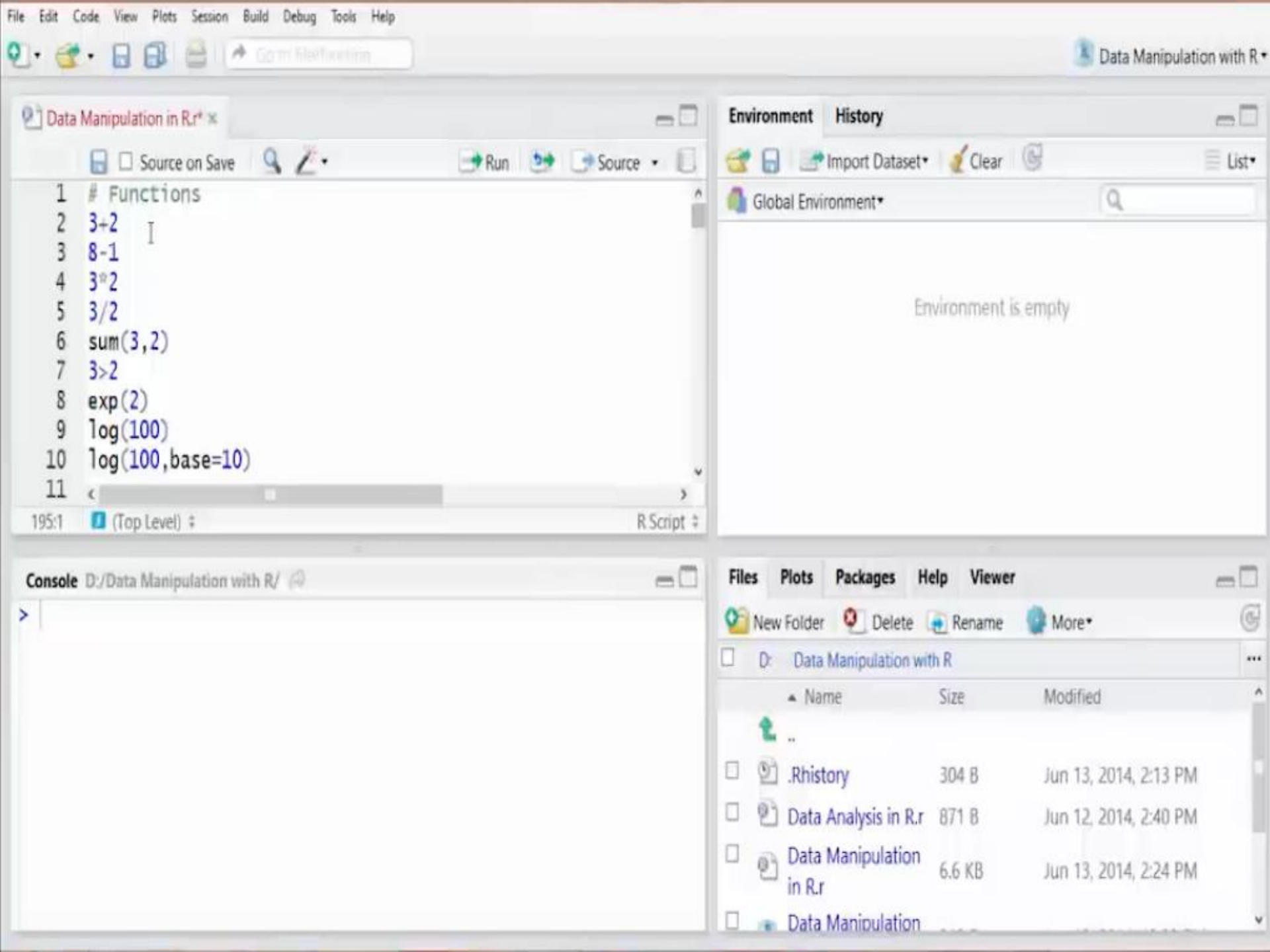
Exploring R

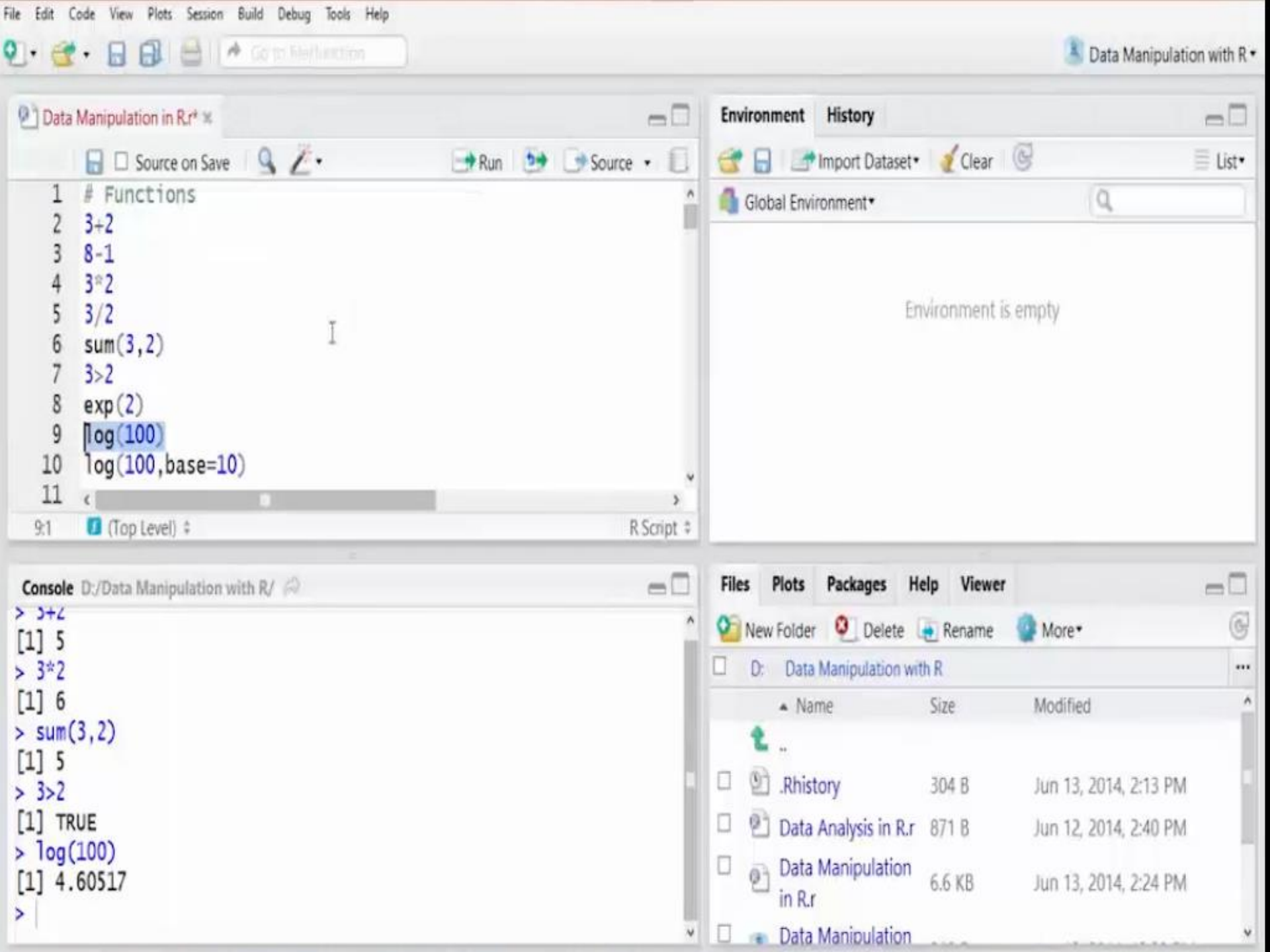
Working with Scripts

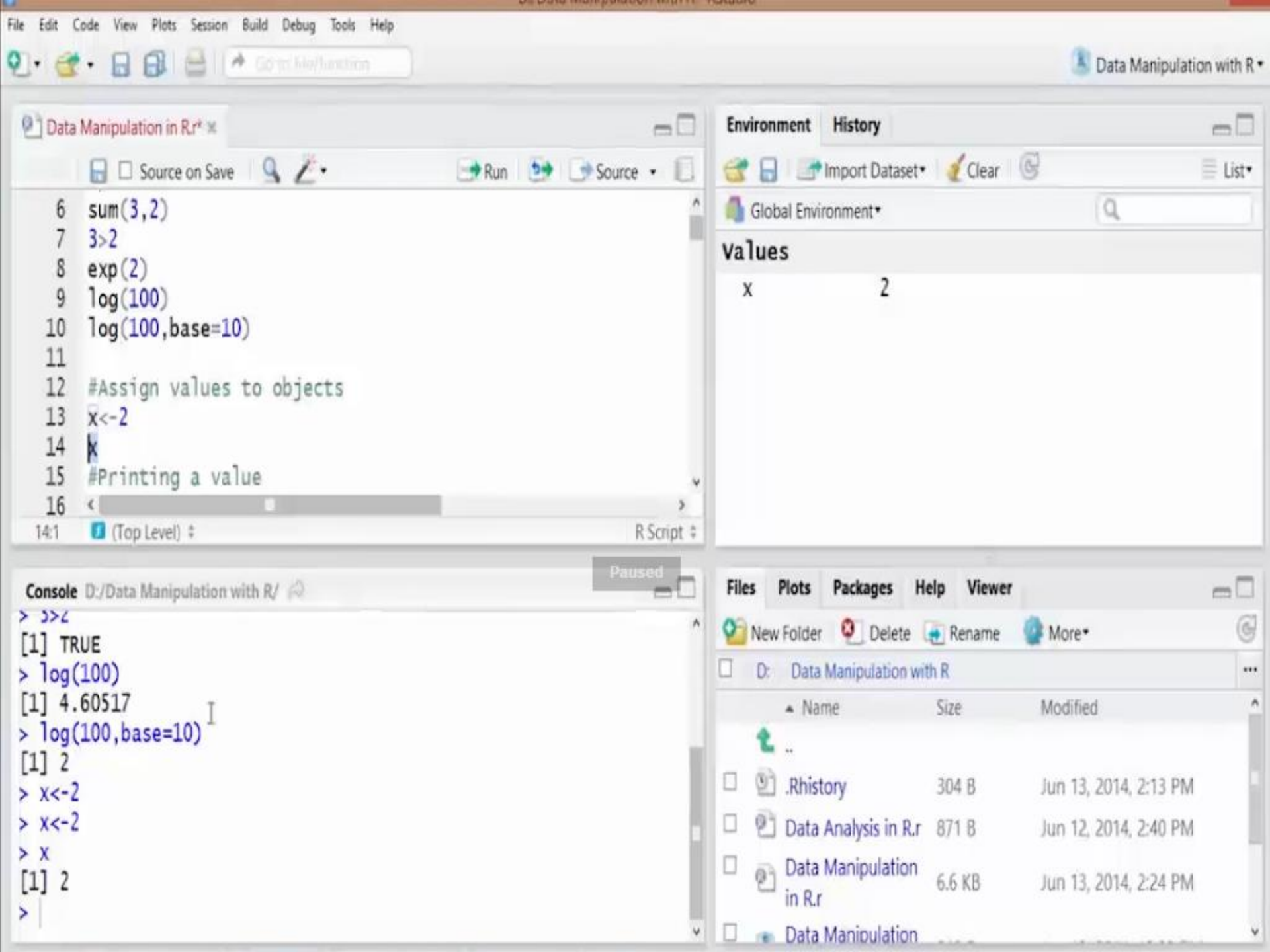
Everything in R is stored as an object.
Lets look at how to create and print objects in R

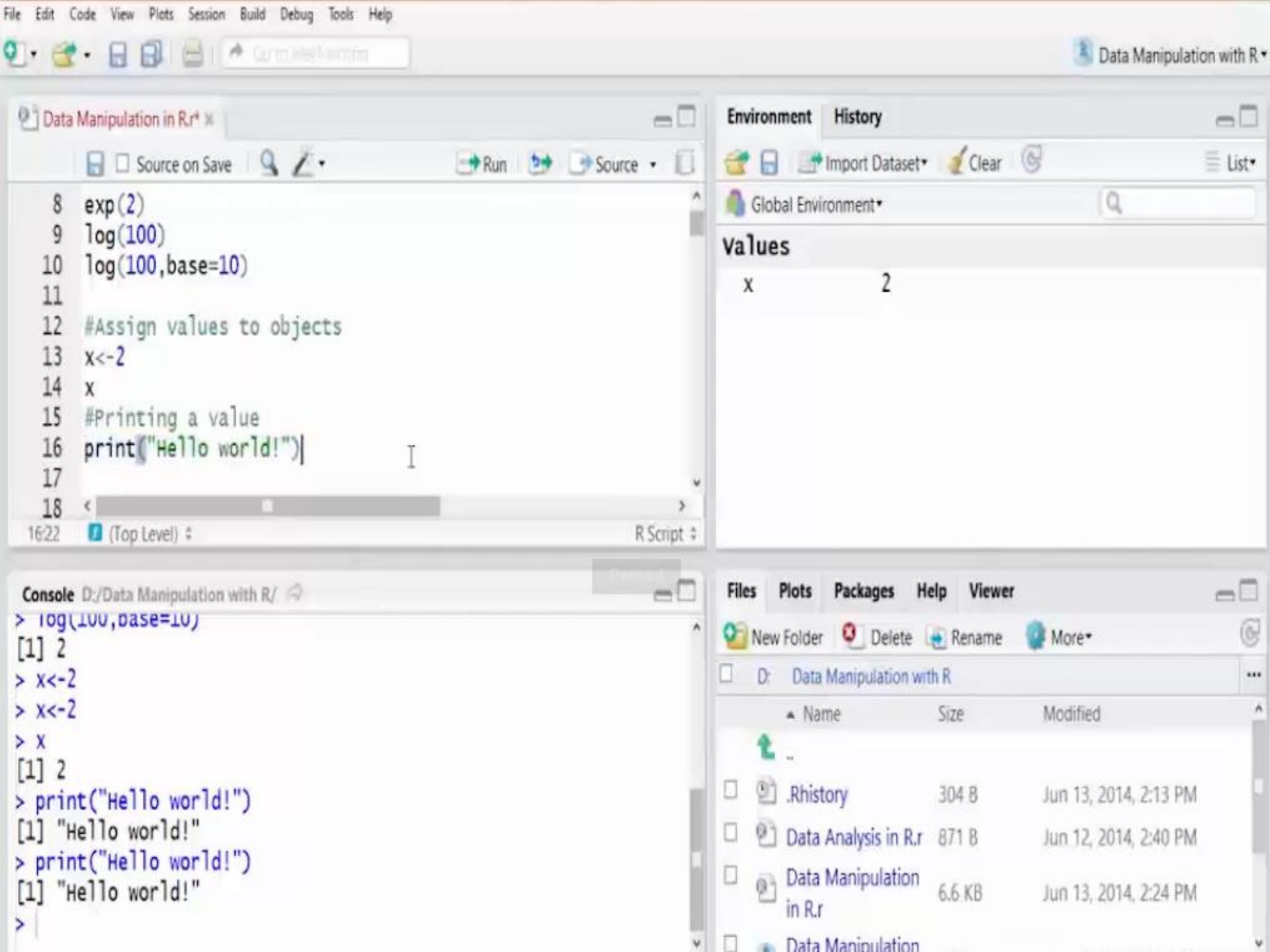
- To print “Hello World” :

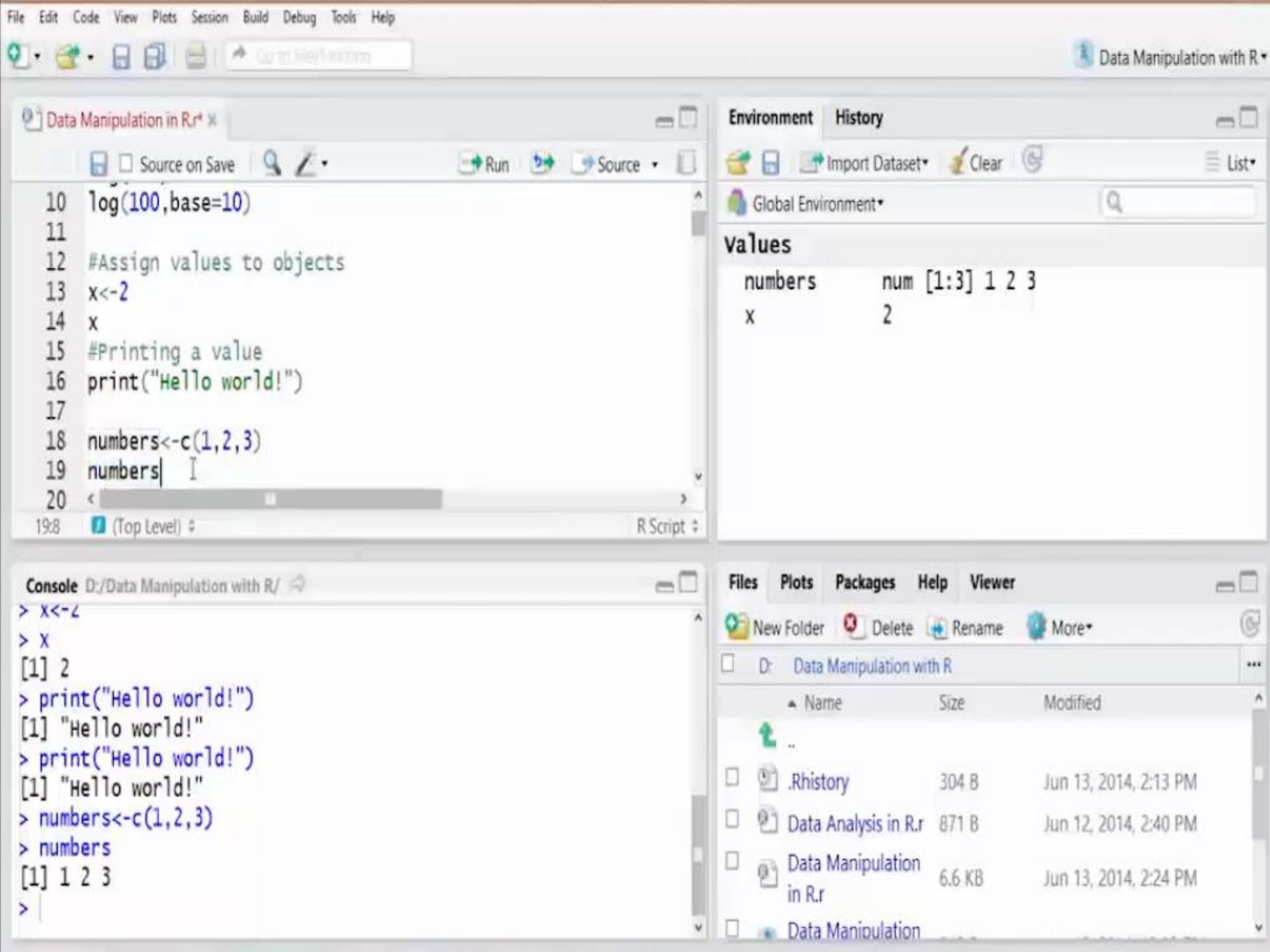
```
>print(“Hello world!”)
[1] “Hello world!”
```

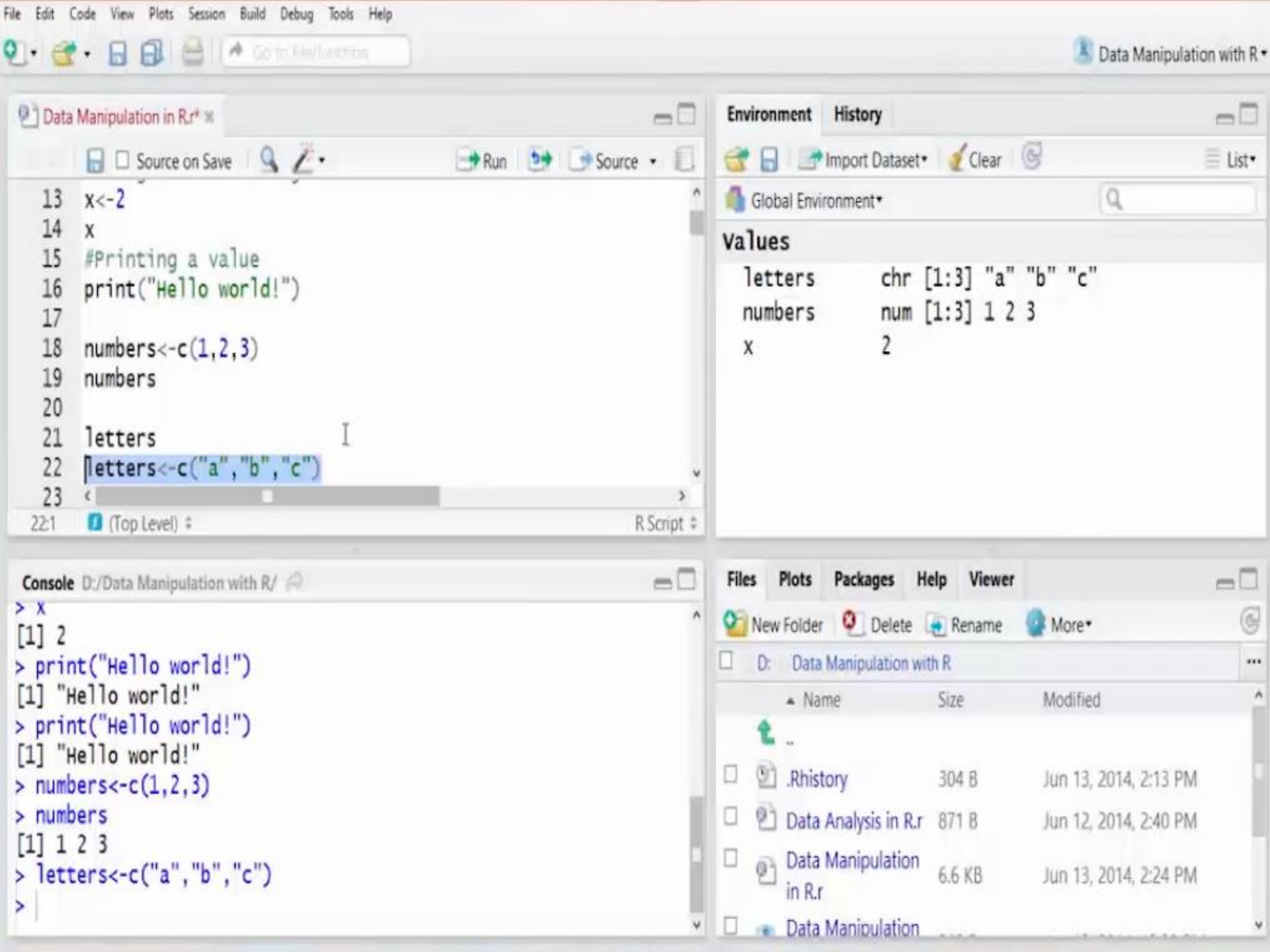

File Edit Code View Plots Session Build Debug Tools Help
Data Manipulation with R
Data Manipulation in R.r*
Source on Save
Run
Source
1 # Functions
2 3+2
3 8-1
4 3*2
5 3/2
6 sum(3,2)
7 3>2
8 exp(2)
9 log(100)
10 log(100,base=10)
11
195:1
(Top Level)
R Script
Environment
History
Import Dataset
Clear
List
Global Environment
Environment is empty
Console D:/Data Manipulation with R/
>
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
D: Data Manipulation with R
Name
Size
Modified
..
.Rhistory 304 B Jun 13, 2014, 2:13 PM
Data Analysis in R.r 871 B Jun 12, 2014, 2:40 PM
Data Manipulation in R.r 6.6 KB Jun 13, 2014, 2:24 PM
Data Manipulation

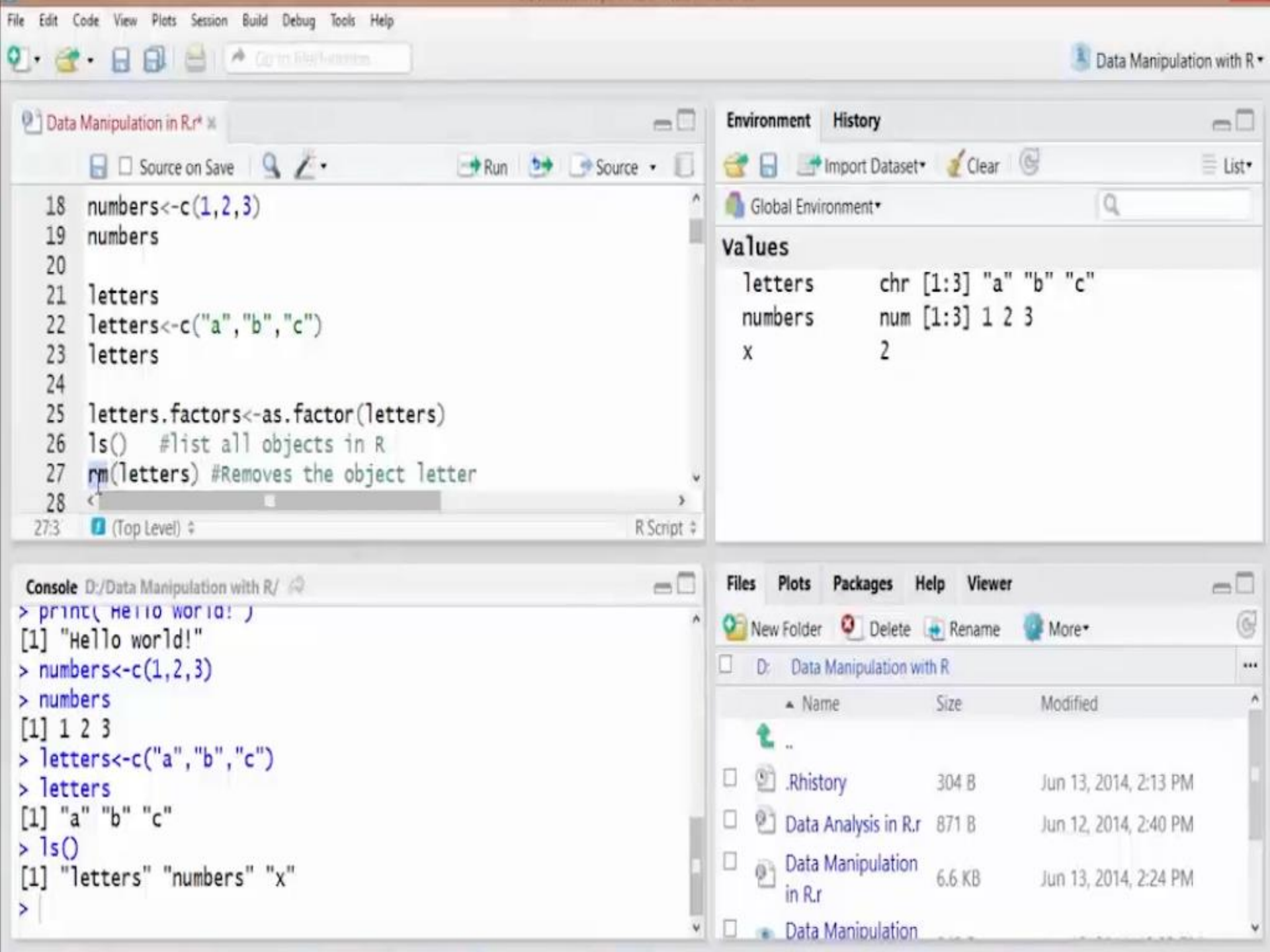

File Edit Code View Plots Session Build Debug Tools Help
Data Manipulation with R
Data Manipulation in R.r*
Source on Save
Run
Source
1 # Functions
2 3+2
3 8-1
4 3*2
5 3/2
6 sum(3,2)
7 3>2
8 exp(2)
9 log(100)
10 log(100,base=10)
11
9:1
(Top Level)
R Script
Environment
History
Import Dataset
Clear
List
Global Environment
Environment is empty
Console D:/Data Manipulation with R/
> 3+2
[1] 5
> 3*2
[1] 6
> sum(3,2)
[1] 5
> 3>2
[1] TRUE
> log(100)
[1] 4.60517
>
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
D: Data Manipulation with R
Name
Size
Modified
.Rhistory 304 B Jun 13, 2014, 2:13 PM
Data Analysis in R.r 871 B Jun 12, 2014, 2:40 PM
Data Manipulation in R.r 6.6 KB Jun 13, 2014, 2:24 PM
Data Manipulation

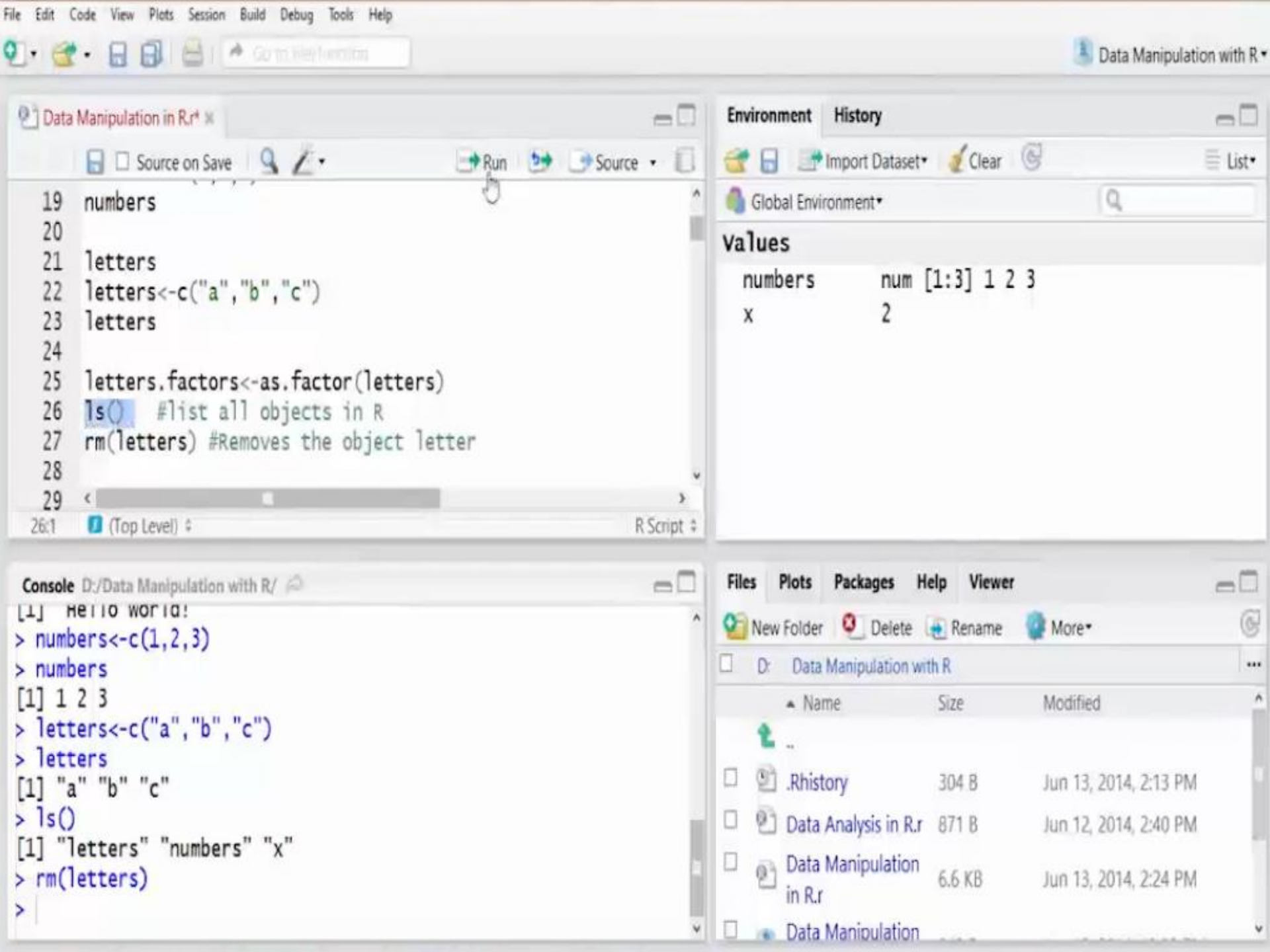

File Edit Code View Plots Session Build Debug Tools Help
Data Manipulation with R
Data Manipulation in R.r*
Source on Save
Run
Source
6 sum(3,2)
7 3>2
8 exp(2)
9 log(100)
10 log(100,base=10)
11
12 #Assign values to objects
13 x<-2
14 x
15 #Printing a value
16
14:1 (Top Level)
R Script
Environment
History
Import Dataset
Clear
List
Global Environment
Values
x 2
Paused
Console D:/Data Manipulation with R/
> 3>2
[1] TRUE
> log(100)
[1] 4.60517
> log(100,base=10)
[1] 2
> x<-2
> x<-2
> x
[1] 2
>
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
D: Data Manipulation with R
Name
Size
Modified
..
.Rhistory 304 B Jun 13, 2014, 2:13 PM
Data Analysis in R.r 871 B Jun 12, 2014, 2:40 PM
Data Manipulation in R.r 6.6 KB Jun 13, 2014, 2:24 PM
Data Manipulation


File Edit Code View Plots Session Build Debug Tools Help
Data Manipulation with R
Data Manipulation in R.r*
Source on Save
Run
Source
8 exp(2)
9 log(100)
10 log(100,base=10)
11
12 #Assign values to objects
13 x<-2
14 x
15 #Printing a value
16 print("Hello world!")
17
18
16:22
(Top Level)
R Script
Environment
History
Import Dataset
Clear
List
Global Environment
Values
x 2
Console D:/Data Manipulation with R/
> log(100,base=10)
[1] 2
> x<-2
> x<-2
> x
[1] 2
> print("Hello world!")
[1] "Hello world!"
> print("Hello world!")
[1] "Hello world!"
>
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
D: Data Manipulation with R
Name
Size
Modified
..
.Rhistory 304 B Jun 13, 2014, 2:13 PM
Data Analysis in R.r 871 B Jun 12, 2014, 2:40 PM
Data Manipulation in R.r 6.6 KB Jun 13, 2014, 2:24 PM
Data Manipulation


File Edit Code View Plots Session Build Debug Tools Help
Data Manipulation with R
Data Manipulation in R.r*
Source on Save
Run
Source
10 log(100,base=10)
11
12 #Assign values to objects
13 x<-2
14 x
15 #Printing a value
16 print("Hello world!")
17
18 numbers<-c(1,2,3)
19 numbers
20
19:8 (Top Level)
R Script
Environment History
Import Dataset
Clear
List
Global Environment
Values
numbers num [1:3] 1 2 3
x 2
Console D:/Data Manipulation with R/
> x<-2
> x
[1] 2
> print("Hello world!")
[1] "Hello world!"
> print("Hello world!")
[1] "Hello world!"
> numbers<-c(1,2,3)
> numbers
[1] 1 2 3
>
Files Plots Packages Help Viewer
New Folder
Delete
Rename
More
D: Data Manipulation with R
Name
Size
Modified
..
.Rhistory 304 B Jun 13, 2014, 2:13 PM
Data Analysis in R.r 871 B Jun 12, 2014, 2:40 PM
Data Manipulation in R.r 6.6 KB Jun 13, 2014, 2:24 PM
Data Manipulation


File Edit Code View Plots Session Build Debug Tools Help
Data Manipulation with R
Data Manipulation in R.r*
Source on Save
Run
Source
13 x<-2
14 x
15 #Printing a value
16 print("Hello world!")
17
18 numbers<-c(1,2,3)
19 numbers
20
21 letters
22 letters<-c("a","b","c")
23
22:1
(Top Level)
R Script
Environment
History
Import Dataset
Clear
List
Global Environment
Values
letters chr [1:3] "a" "b" "c"
numbers num [1:3] 1 2 3
x 2
Console D:/Data Manipulation with R/
> x
[1] 2
> print("Hello world!")
[1] "Hello world!"
> print("Hello world!")
[1] "Hello world!"
> numbers<-c(1,2,3)
> numbers
[1] 1 2 3
> letters<-c("a","b","c")
>
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
D: Data Manipulation with R
Name
Size
Modified
..
.Rhistory 304 B Jun 13, 2014, 2:13 PM
Data Analysis in R.r 871 B Jun 12, 2014, 2:40 PM
Data Manipulation in R.r 6.6 KB Jun 13, 2014, 2:24 PM
Data Manipulation


File Edit Code View Plots Session Build Debug Tools Help
Data Manipulation with R
Data Manipulation in R.r*
Source on Save
Run
Source
18 numbers<-c(1,2,3)
19 numbers
20
21 letters
22 letters<-c("a","b","c")
23 letters
24
25 letters.factors<-as.factor(letters)
26 ls() #list all objects in R
27 rm(letters) #Removes the object letter
28
27:3
(Top Level)
R Script
Environment
History
Import Dataset
Clear
List
Global Environment
Values
letters chr [1:3] "a" "b" "c"
numbers num [1:3] 1 2 3
x 2
Console D:/Data Manipulation with R/
> print("Hello world!")
[1] "Hello world!"
> numbers<-c(1,2,3)
> numbers
[1] 1 2 3
> letters<-c("a","b","c")
> letters
[1] "a" "b" "c"
> ls()
[1] "letters" "numbers" "x"
>
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
D: Data Manipulation with R
Name
Size
Modified
..
.Rhistory 304 B Jun 13, 2014, 2:13 PM
Data Analysis in R.r 871 B Jun 12, 2014, 2:40 PM
Data Manipulation in R.r 6.6 KB Jun 13, 2014, 2:24 PM
Data Manipulation


File Edit Code View Plots Session Build Debug Tools Help
Data Manipulation with R
Data Manipulation in R.r*
Source on Save
Run
Source
19 numbers
20
21 letters
22 letters<-c("a","b","c")
23 letters
24
25 letters.factors<-as.factor(letters)
26 ls() #list all objects in R
27 rm(letters) #Removes the object letter
28
29
26:1
(Top Level)
R Script
Environment
History
Import Dataset
Clear
List
Global Environment
Values
numbers num [1:3] 1 2 3
x 2
Console D:/Data Manipulation with R/
[1] "Hello World!"
> numbers<-c(1,2,3)
> numbers
[1] 1 2 3
> letters<-c("a","b","c")
> letters
[1] "a" "b" "c"
> ls()
[1] "letters" "numbers" "x"
> rm(letters)
>
Files
Plots
Packages
Help
Viewer
New Folder
Delete
Rename
More
D: Data Manipulation with R
Name
Size
Modified
..
.Rhistory 304 B Jun 13, 2014, 2:13 PM
Data Analysis in R.r 871 B Jun 12, 2014, 2:40 PM
Data Manipulation in R.r 6.6 KB Jun 13, 2014, 2:24 PM
Data Manipulation

RECAP

Understanding analytics and analytic tools

Exploring R

DATA SCIENCE WITH R

DATA SCIENCE WITH R

Class 2 – R Programming

DATA SCIENCE WITH R

Class 2 – R Programming

Topic 1

Manipulating Objects and Saving and Loading an Object

AGENDA

- **Manipulating Objects, Saving and Loading an Object**

Data Structures in R

First steps with exploring a dataset

DATA SCIENCE WITH R

RECAP

Manipulating objects, saving and loading an object

DATA SCIENCE WITH R

DATA SCIENCE WITH R

Class 2 – R Programming

DATA SCIENCE WITH R

Class 2 – R Programming

Topic 2

★ Data Structures in R ★

AGENDA

Manipulating Objects, Saving and Loading an Object

Data Structures in R

First steps with exploring a dataset

Manipulating and Data Processing in R

Data Structures in R

Vectors

Manipulating and Data Processing in R

Data Structures in R

Vectors

- Most Simplest structure in R

Manipulating and Data Processing in R

Data Structures in R

Vectors

- Most Simplest structure in R

- If data has only one dimension, like a set of digits, then vectors can be used to represent it**.**

Manipulating and Data Processing in R

Data Structures in R

Vectors

- Most Simplest structure in R
- If data has only one dimension, like a set of digits, then vectors can be used to represent it**.**

Matrices

Manipulating and Data Processing in R

Data Structures in R

Vectors

- Most Simplest structure in R
- If data has only one dimension, like a set of digits, then vectors can be used to represent it**.**

Matrices

- Used when data is a higher dimensional array

Manipulating and Data Processing in R

Data Structures in R

Vectors

- Most Simplest structure in R
- If data has only one dimension, like a set of digits, then vectors can be used to represent it**.**

Matrices

- Used when data is a higher dimensional array
- But contains only data of a single class Eg : only character or numeric

Manipulating and Data Processing in R

Data Structures in R

Data Frames

Manipulating and Data Processing in R

Data Structures in R

Data Frames

- It is like a single table with rows and columns of data

Manipulating and Data Processing in R

Data Structures in R

Data Frames

- It is like a single table with rows and columns of data
- Contains columns or lists of different data

Manipulating and Data Processing in R

Data Structures in R

Data Frames

- It is like a single table with rows and columns of data
- Contains columns or lists of different data

Lists

Manipulating and Data Processing in R

Data Structures in R

Data Frames

- It is like a single table with rows and columns of data
- Contains columns or lists of different data

Lists

- Used when data cannot be represented by data frames

Manipulating and Data Processing in R

Data Structures in R

Data Frames

- It is like a single table with rows and columns of data
- Contains columns or lists of different data

Lists

- Used when data cannot be represented by data frames
- It contain all kinds of other objects, including other lists or data frames

Manipulating and Data Processing in R

Data Structures in R

Data Frames

- It is like a single table with rows and columns of data
- Contains columns or lists of different data

Lists

- Used when data cannot be represented by data frames
- It contain all kinds of other objects, including other lists or data frames
- Very Flexible

DATA SCIENCE WITH R

RECAP

Data structures in R

DATA SCIENCE WITH R

DATA SCIENCE WITH R

Class 2 – R Programming

DATA SCIENCE WITH R

Class 2 – R Programming

Topic 3

★ First steps with exploring a dataset ★

AGENDA

Manipulating Objects, Saving and Loading an Object

Data Structures in R

First steps with exploring a dataset

DATA SCIENCE WITH R

RECAP

First steps with exploring a dataset: Read a dataset, summary of a dataset and also traverse through the dataset

DATA SCIENCE WITH R

DATA SCIENCE WITH R

Class 3 – Importing and Exporting of Data in R

Topic 1

★ Data Import/Export Overview ★

AGENDA

- **Data import/export overview**
- Working with plain text files
- Working with large text files
- Working with excel workbooks
- Working with statistical system outputs
- Working with databases
- Working with web data

Data import/export overview

Need for Data import/export in R

Data import/export overview

Need for Data import/export in R

```
> personId<-c(1,2,3,4,5)
> personWt<-c(60,70,80,65,75)

> person<-data.frame(personId,personWt)
> person
  personId personWt
1        1       60
2        2       70
3        3       80
4        4       65
5        5       75

> plot(person)
```

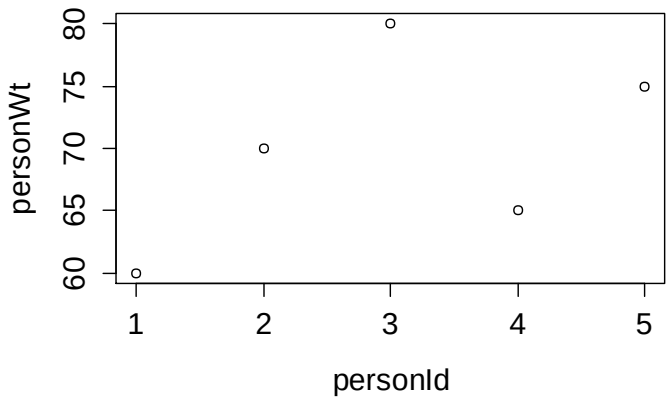
Data import/export overview

Need for Data import/export in R

```
> personId<-c(1,2,3,4,5)
> personWt<-c(60,70,80,65,75)

> person<-data.frame(personId,personWt)
> person
  personId personWt
1        1       60
2        2       70
3        3       80
4        4       65
5        5       75

> plot(person)
```



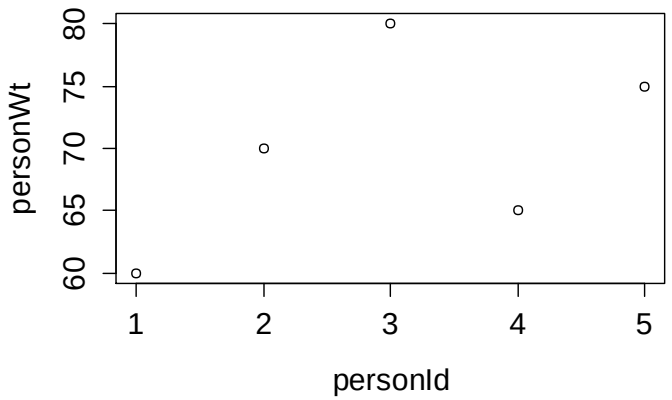
Data import/export overview

Need for Data import/export in R

```
> personId<-c(1,2,3,4,5)
> personWt<-c(60,70,80,65,75)

> person<-data.frame(personId,personWt)
> person
  personId personWt
1        1       60
2        2       70
3        3       80
4        4       65
5        5       75

> plot(person)
```



✓ Approach not practical

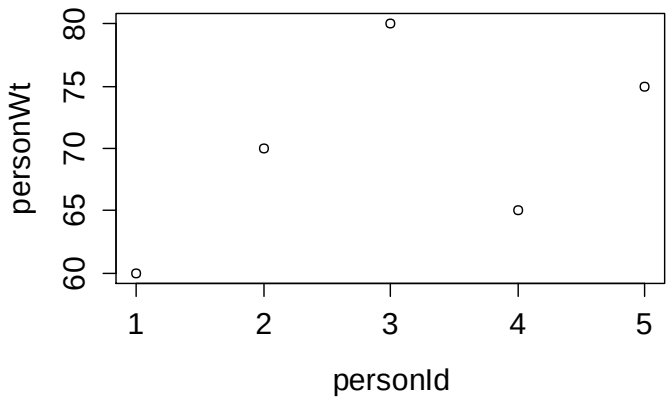
Data import/export overview

Need for Data import/export in R

```
> personId<-c(1,2,3,4,5)
> personWt<-c(60,70,80,65,75)

> person<-data.frame(personId,personWt)
> person
  personId personWt
1        1       60
2        2       70
3        3       80
4        4       65
5        5       75

> plot(person)
```



- ✓ Approach not practical
- ✓ Issues of scalability and typos

Data import/export overview

R Benefits

Data import/export overview

R Benefits

- Data science projects start with a bunch of data

Data import/export overview

R Benefits

- Data science projects start with a bunch of data
- Common way of data sharing is through external files

Data import/export overview

R Benefits

- Data science projects start with a bunch of data
- Common way of data sharing is through external files
- Structured format necessary for application of R functions

personId	personWt
1	60
2	70
3	80
4	65
5	75

Data import/export overview

R Benefits

- Data science projects start with a bunch of data
- Common way of data sharing is through external files
- Structured format necessary for application of R functions
- data.frame format provides spread sheet like look

personId	personWt
1	60
2	70
3	80
4	65
5	75

Data import/export overview

R Benefits

- Data science projects start with a bunch of data
- Common way of data sharing is through external files
- Structured format necessary for application of R functions
- data.frame format provides spread sheet like look
- Presence of R adapters to read data from various formats

personId	personWt
1	60
2	70
3	80
4	65
5	75

Data import/export overview

R Benefits

- Data science projects start with a bunch of data
- Common way of data sharing is through external files
- Structured format necessary for application of R functions
- data.frame format provides spread sheet like look
- Presence of R adapters to read data from various formats
- Facilities available either in R itself or via packages

personId	personWt
1	60
2	70
3	80
4	65
5	75

Data import/export overview

R Benefits

- Data science projects start with a bunch of data
- Common way of data sharing is through external files
- Structured format necessary for application of R functions
- data.frame format provides spread sheet like look
- Presence of R adapters to read data from various formats
- Facilities available either in R itself or via packages

personId	personWt
1	60
2	70
3	80
4	65
5	75

Analyzing data in R after loading is easier with the help of data.frame structure which is ideal for working with structured data

Data import/export overview

Importing Data in R

Data import/export overview

Importing Data in R

- Most of the file formats are supported by R

Data import/export overview

Importing Data in R

- Most of the file formats are supported by R
- Popular ones are delimited files like csv, tab etc.

Data import/export overview

Importing Data in R

- Most of the file formats are supported by R
- Popular ones are delimited files like csv, tab etc.
- These files generally have an extension .txt

Data import/export overview

Importing Data in R

- Most of the file formats are supported by R
- Popular ones are delimited files like csv, tab etc.
- These files generally have an extension .txt
- More recent formats like xml, json etc.

Data import/export overview

Importing Data in R

- Most of the file formats are supported by R
- Popular ones are delimited files like csv, tab etc.
- These files generally have an extension .txt
- More recent formats like xml, json etc.
- Excel files with extensions like .xls, .xlsx etc.

Data import/export overview

Importing Data in R

- Most of the file formats are supported by R
- Popular ones are delimited files like csv, tab etc.
- These files generally have an extension .txt
- More recent formats like xml, json etc.
- Excel files with extensions like .xls, .xlsx etc.
- Other proprietary formats related to SAS, SPSS

Data import/export overview

Importing Data in R

- Most of the file formats are supported by R
- Popular ones are delimited files like csv, tab etc.
- These files generally have an extension .txt
- More recent formats like xml, json etc.
- Excel files with extensions like .xls, .xlsx etc.
- Other proprietary formats related to SAS, SPSS
- Files from Relational Database Management Systems (SQL)

Data import/export overview

Importing Data in R

- Most of the file formats are supported by R
- Popular ones are delimited files like csv, tab etc.
- These files generally have an extension .txt
- More recent formats like xml, json etc.
- Excel files with extensions like .xls, .xlsx etc.
- Other proprietary formats related to SAS, SPSS
- Files from Relational Database Management Systems (SQL)
- Data retrieval from API's like Twitter API, Facebook API etc.

Data import/export overview

Importing Data in R

- Most of the file formats are supported by R
- Popular ones are delimited files like csv, tab etc.
- These files generally have an extension .txt
- More recent formats like xml, json etc.
- Excel files with extensions like .xls, .xlsx etc.
- Other proprietary formats related to SAS, SPSS
- Files from Relational Database Management Systems (SQL)
- Data retrieval from API's like Twitter API, Facebook API etc.

To deal with different file formats, often data can be first exported as a text file like csv and then imported into R

Data import/export overview

Checks before Data Import

Data import/export overview

Checks before Data Import

When reading data from text files, it is necessary to know and understand more about the inherent characteristics

Data import/export overview

Checks before Data Import

When reading data from text files, it is necessary to know and understand more about the inherent characteristics

- Presence of header line

Data import/export overview

Checks before Data Import

When reading data from text files, it is necessary to know and understand more about the inherent characteristics

- Presence of header line
- Kind of value separator

Data import/export overview

Checks before Data Import

When reading data from text files, it is necessary to know and understand more about the inherent characteristics

- Presence of header line
- Kind of value separator
- Representation of missing values

Data import/export overview

Checks before Data Import

When reading data from text files, it is necessary to know and understand more about the inherent characteristics

- Presence of header line
- Kind of value separator
- Representation of missing values
- Notation of comment characters or quotes

Data import/export overview

Checks before Data Import

When reading data from text files, it is necessary to know and understand more about the inherent characteristics

- Presence of header line
- Kind of value separator
- Representation of missing values
- Notation of comment characters or quotes
- Existence of any unfilled or blank lines

Data import/export overview

Checks before Data Import

When reading data from text files, it is necessary to know and understand more about the inherent characteristics

- Presence of header line
- Kind of value separator
- Representation of missing values
- Notation of comment characters or quotes
- Existence of any unfilled or blank lines
- Classes of the variables

Data import/export overview

What will be covered?

Data import/export overview

What will be covered?

Data Science projects involve dealing with data from various sources based on business context

Data import/export overview

What will be covered?

Data Science projects involve dealing with data from various sources based on business context

- ✓ plain text files

Data import/export overview

What will be covered?

Data Science projects involve dealing with data from various sources based on business context

- ✓ plain text files
- ✓ excel spreadsheets

Data import/export overview

What will be covered?

Data Science projects involve dealing with data from various sources based on business context

- ✓ plain text files
- ✓ excel spreadsheets
- ✓ xml & json documents

Data import/export overview

What will be covered?

Data Science projects involve dealing with data from various sources based on business context

- ✓ plain text files
- ✓ excel spreadsheets
- ✓ xml & json documents
- ✓ RDMS and NoSQL databases

Data import/export overview

What will be covered?

Data Science projects involve dealing with data from various sources based on business context

- ✓ plain text files
- ✓ excel spreadsheets
- ✓ xml & json documents
- ✓ RDMS and NoSQL databases
- ✓ Other software file outputs

Data import/export overview

What will be covered?

Data Science projects involve dealing with data from various sources based on business context

- ✓ plain text files
- ✓ excel spreadsheets
- ✓ xml & json documents
- ✓ RDMS and NoSQL databases
- ✓ Other software file outputs
- ✓ API data extraction

Data import/export overview

What will be covered?

Data Science projects involve dealing with data from various sources based on business context

- ✓ plain text files
- ✓ excel spreadsheets
- ✓ xml & json documents
- ✓ RDMS and NoSQL databases
- ✓ Other software file outputs
- ✓ API data extraction

And further, we will also look at exporting data from R into various formats

DATA SCIENCE WITH R

RECAP

Data import/export overview