

Descriptive Statistics With R Software

Introduction to R Software

Shalabh

Department of Mathematics and Statistics

Indian Institute of Technology Kanpur

R Software

Use of a software is desirable and moreover an essential part of any analysis.

Software – used in all sciences including mathematics and statistics.

Some popular statistical software are SPSS, SAS, Minitab, Stata, Matlab etc.

Another software is R.

Developers of R Software

Currently developed by the R Development Core Team.

Available at www.r-project.org

It is a free software.

It supports many free packages which helps the data scientist and analyst.

What is R?

R is an environment for data manipulation, statistical computing, graphics display and data analysis.

Effective data handling and storage of outputs is possible.

Simple as well as complicated calculations are possible.

Simulations are possible.

What is R?

Graphical display on-screen and hardcopy are possible.

Programming language is effective which includes all possibilities just like any other good programming language.

R has a statistical computing environment.

It has a computer language which is convenient to use for statistical and graphical applications.

Switching to R

R is free (open source) software and therefore is not a black box.

Built in and contributed packages are available, and users are provided tools to make packages.

It is possible to contribute own packages.

The commands can be saved, run and stored in script files.

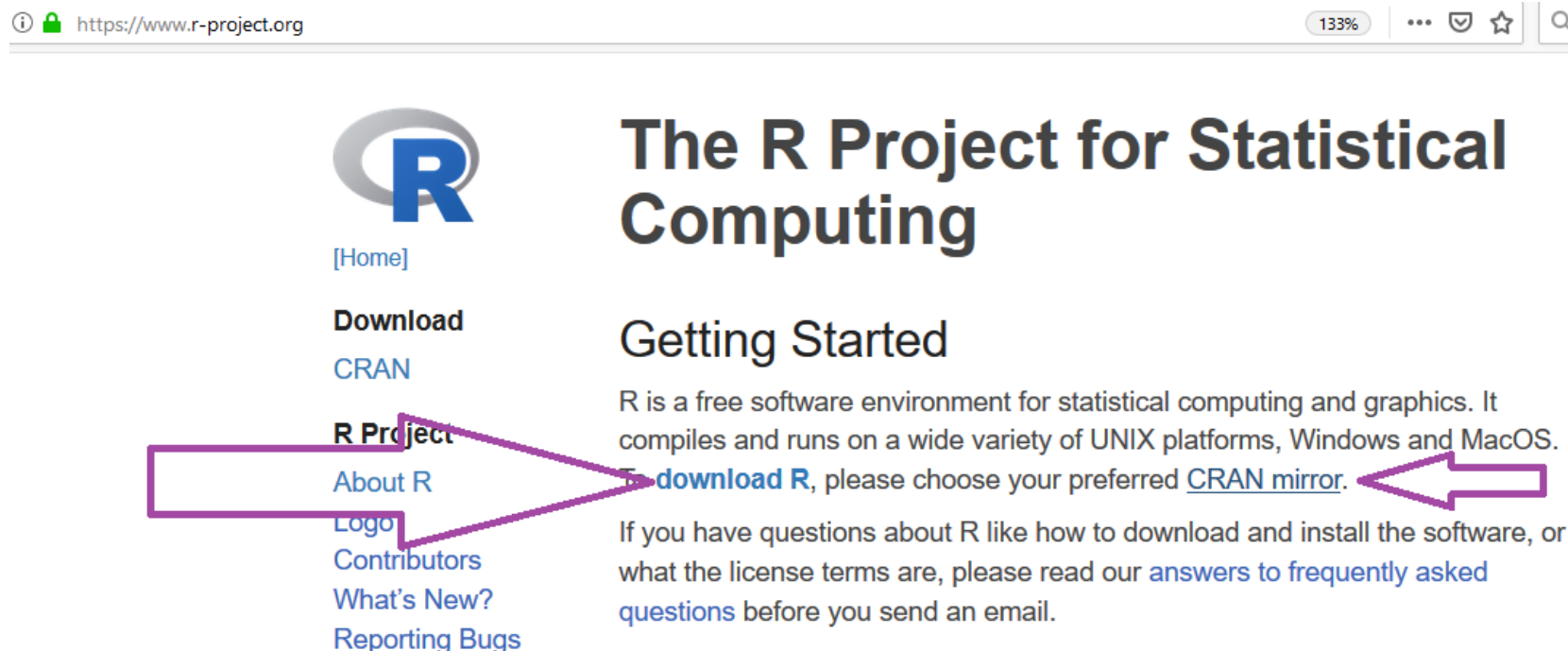
Switching to R

R is available for Windows, Unix, Linux and Macintosh platforms.

Graphics can be directly saved in a Postscript or PDF format.

Installing R

You may install R in a windows or Mac platform by downloading from the Comprehensive R Archive Network (CRAN) website: www.r-project.org or directly from <http://cran.r-project.org/>



The screenshot shows the homepage of The R Project for Statistical Computing. The browser address bar displays 'https://www.r-project.org' with a 133% zoom level. The page features the R logo, a navigation menu on the left, and a main content area with the title 'The R Project for Statistical Computing' and a section 'Getting Started'. A purple arrow points from the 'R Project' link in the navigation menu to the 'download R' link in the main text. Another purple arrow points from the 'CRAN mirror' link in the main text to the right.

[\[Home\]](#)

Download
[CRAN](#)
R Project
[About R](#)
[Logo](#)
[Contributors](#)
[What's New?](#)
[Reporting Bugs](#)

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.

Installing R



icon will appear.

Double click on this icon will start the software.

```
R Console

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]

> | ← COMMAND LINE
    TYPE COMMANDS HERE
```

CONSOLE

Working with R

Use command line to type and execute the commands.

Some free software like R Studio, Tinn R etc. are also available to work with R software. It is an interface between R and us.

Such software are more useful for beginners.

It makes coding and execution of programmes easier.

R Studio is available at <https://www.rstudio.com/>

Tinn R is available at <https://sourceforge.net/projects/tinn-r/>

Installing Packages and Libraries

The base R package contains some necessary libraries only.

Other libraries are required for advanced statistical work which are downloaded and installed.

Run the R program, then use the `install.packages` function to download the libraries.

Examples :

`install.packages("ggplot2")` : installs package `ggplot2`

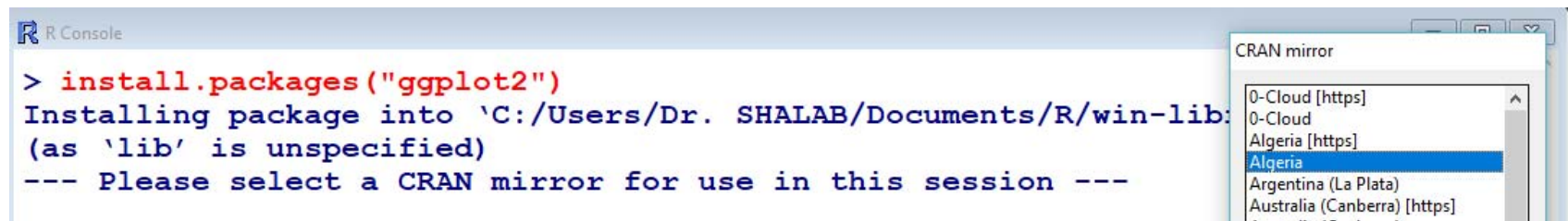
`install.packages("graphics")` : installs package `graphics`

`install.packages("cluster")` : installs package `cluster`

Installing Packages and Libraries

Example

`install.packages("ggplot2")`

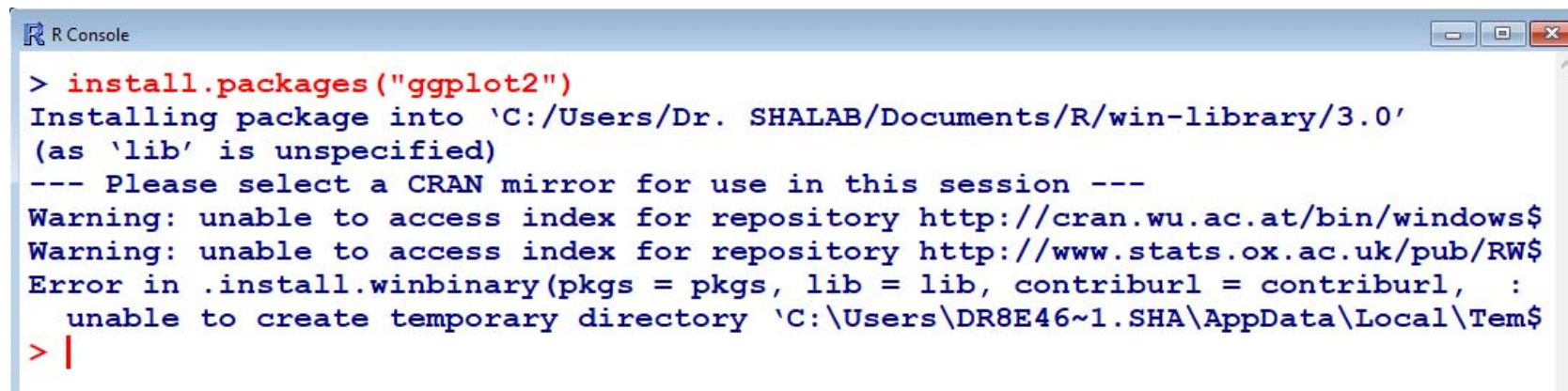


The image shows an R Console window with the following text:

```
> install.packages("ggplot2")
Installing package into 'C:/Users/Dr. SHALAB/Documents/R/win-lib:'
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
```

Overlaid on the console is a "CRAN mirror" dialog box. It contains a list of mirrors with "Algeria" selected. The list includes:

- 0-Cloud [https]
- 0-Cloud
- Algeria [https]
- Algeria
- Argentina (La Plata)
- Australia (Canberra) [https]



The image shows an R Console window with the following text:

```
> install.packages("ggplot2")
Installing package into 'C:/Users/Dr. SHALAB/Documents/R/win-library/3.0'
(as 'lib' is unspecified)
--- Please select a CRAN mirror for use in this session ---
Warning: unable to access index for repository http://cran.wu.ac.at/bin/windows$
Warning: unable to access index for repository http://www.stats.ox.ac.uk/pub/RW$
Error in .install.winbinary(pkgs = pkgs, lib = lib, contriburl = contriburl, :
  unable to create temporary directory 'C:\Users\DR8E46~1.SHA\AppData\Local\Tem$
> |
```

Libraries in R

To use a library, type the `library` function with the name of the library in brackets.

Thus to load the `cluster` library type:

```
library(cluster)
```

Similarly,

```
library(ggplot2) : loads package ggplot2
```

```
library(graphics) : loads package graphics
```

Libraries in R

Examples of libraries that come as a part of base package in R.

MASS : package associated with Venables and Ripley's book entitled *Modern Applied Statistics using S-Plus*.

library(MASS) : loads package **MASS**

mgcv : generalized additive models.

library(mgcv) : loads package **mgcv**

Contents of Libraries

Use `help` function to get the detailed contents of library packages.

Here is how we find out about the contents of the `cluster` library:
`library(help=cluster)` returns the following:

```
Information on package 'cluster'
Description:

Package:           cluster
Version:           1.14.4
Date:              2013-03-26
Priority:           recommended
Author:            Martin Maechler, based on S
                   original by Peter ... ..
... ..
```

followed by a list of all the functions and data sets.

Contents of Libraries

```
R R Console
> library(help=cluster)
> |
```

```
R Documentation for package 'cluster'

                                Information on package 'cluster'

Description:

Package:                cluster
Version:                1.14.4
Date:                  2013-03-26
Priority:               recommended
Author:                Martin Maechler, based on S original by Peter
                        Rousseeuw <rousse@uia.ua.ac.be>,
                        Anja.Struyf@uia.ua.ac.be and
                        Mia.Hubert@uia.ua.ac.be, and initial R port by
                        Kurt.Hornik@R-project.org
Maintainer:            Martin Maechler <maechler@stat.math.ethz.ch>
```


Cleaning up the Windows

We assign names to variables when analyzing any data.

It is good practice to remove the variable names given to any data frame at the end each session in R.

`rm()` command removes variable names

For example,

`rm(x,y,z)` removes the variables `x`, `y` and `z`.

How to quit in R

Type `q()` to quit R.

