

Introduction

Why R and Installation Procedure

Kanahaiya Kumar [Data Mining Central]

What is R?

- R is an environment for data manipulation, statistical computing, graphics display and data analysis
- One can effectively handle the data and storage of outputs
- Simple as well as complicated calculations could be done
- Graphical display on screen and hard copy could be done
- Programming language is effective which includes all possibilities just like any other good programming language
- The R language is very similar in appearance to “S” Language based on a software s-plus

Switching to R?

- Initially R was developed by statistician for statistician and today diverse communities are using for different purposes
- The R language is widely used for research & development, analytics, etc
- The R language is open source and free and many statistical packages are freely available
- R has statistical computing environment and great support of communities
- It has a computer language which is convenient to use for statistical and graphical applications
- The commands can be saved, run and stored in script files

Switching to R?

- R is available for Windows, Unix, Linux, and Macintosh platforms
- Initially R was developed to compete with S-Plus
- Built in and contributed packages are available, and users are provided tools to make packages. It is possible to contribute your own packages.
- This language provides the logical control of branching and looping, and modular programming using functions
- Error messages are provided while executing a programme or a function which are helpful in finding the errors

Switching to R?

- R is interpreted computer language [Not compiler]
- When we use command line interface, each command or expression to be evaluated is typed at the command prompt, and immediately evaluated when the enter key is pressed at the end of a syntactically complete statement
- Tip
 - Press the up-arrow key to recall commands and edit them
 - Use the Esc [Escape] key to cancel a command
- Graphics can be directly saved in a postscript or pdf format

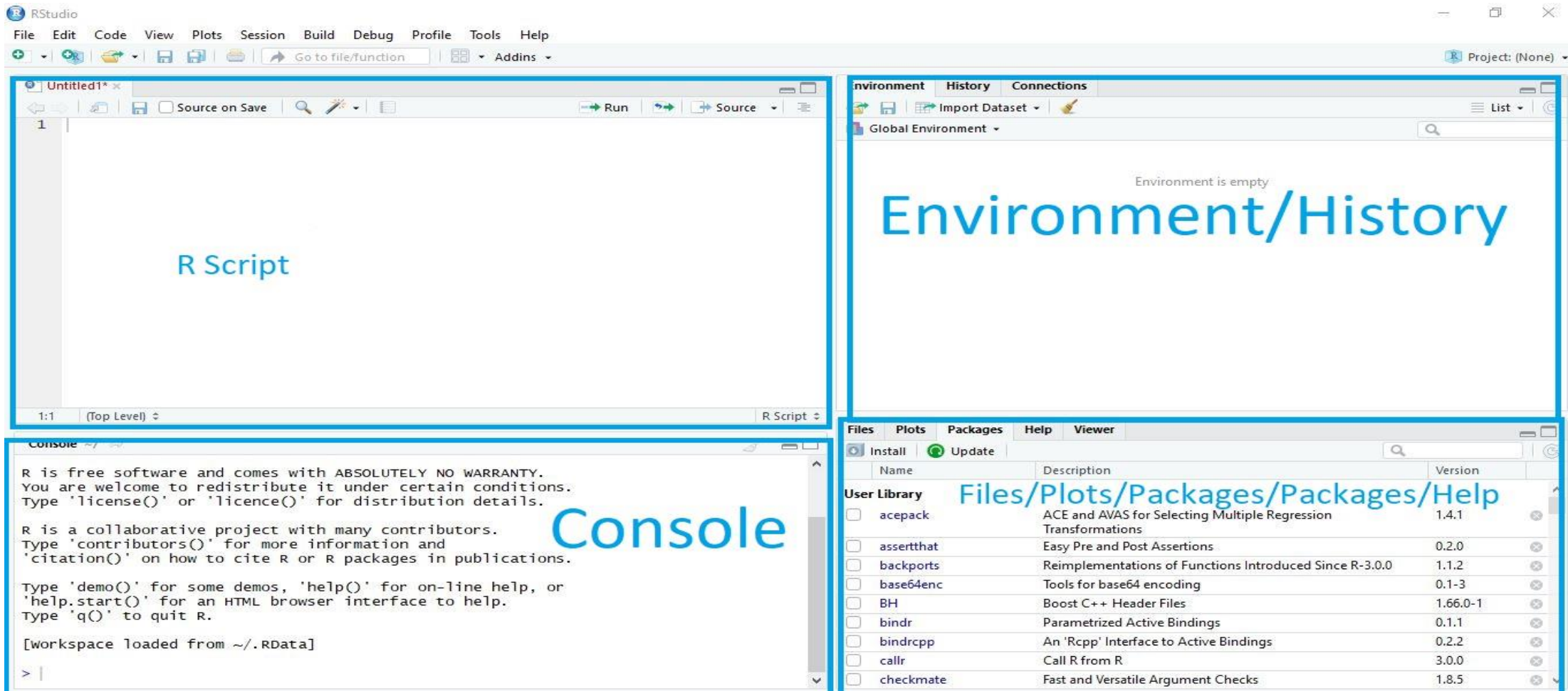
Installing R

- You may install R by downloading it from <https://www.r-project.org>
- Else download it from the Comprehensive R Archive Network [CRAN] website: <http://cran.r-project.org/>
- Installing R Studio
 - R Studio is a software which helps in running the R Software
 - Several such editors are available , e.g. Tinn R [\[https://sourceforge.net/projects/tinn-r\]](https://sourceforge.net/projects/tinn-r)
 - R Studio written in C++ programming language
 - R Studio is a free and open source integrated development environment [IDE] for R
 - Download R Studio software from website <https://www.rstudio.com>

How to

- Set the working directory
- Create an R file and save it
- Execute an R file
- Execute pieces of code

A first look of R Studio

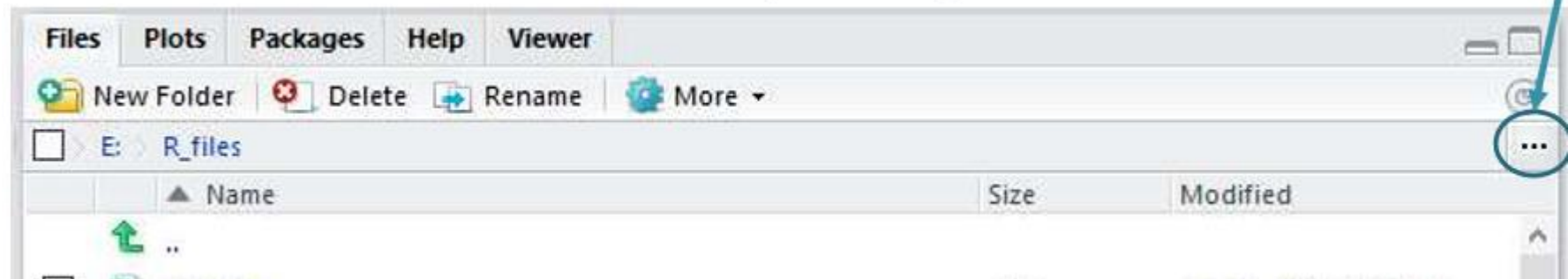


Set the working directory

Enter in console: `setwd("directorypath")`

Or

Step 1: Choose a suitable location by clicking on the indicated button *To choose working directory*



Set the working directory

Step 2: Once directory is chosen, select the more icon and choose “set as working directory”

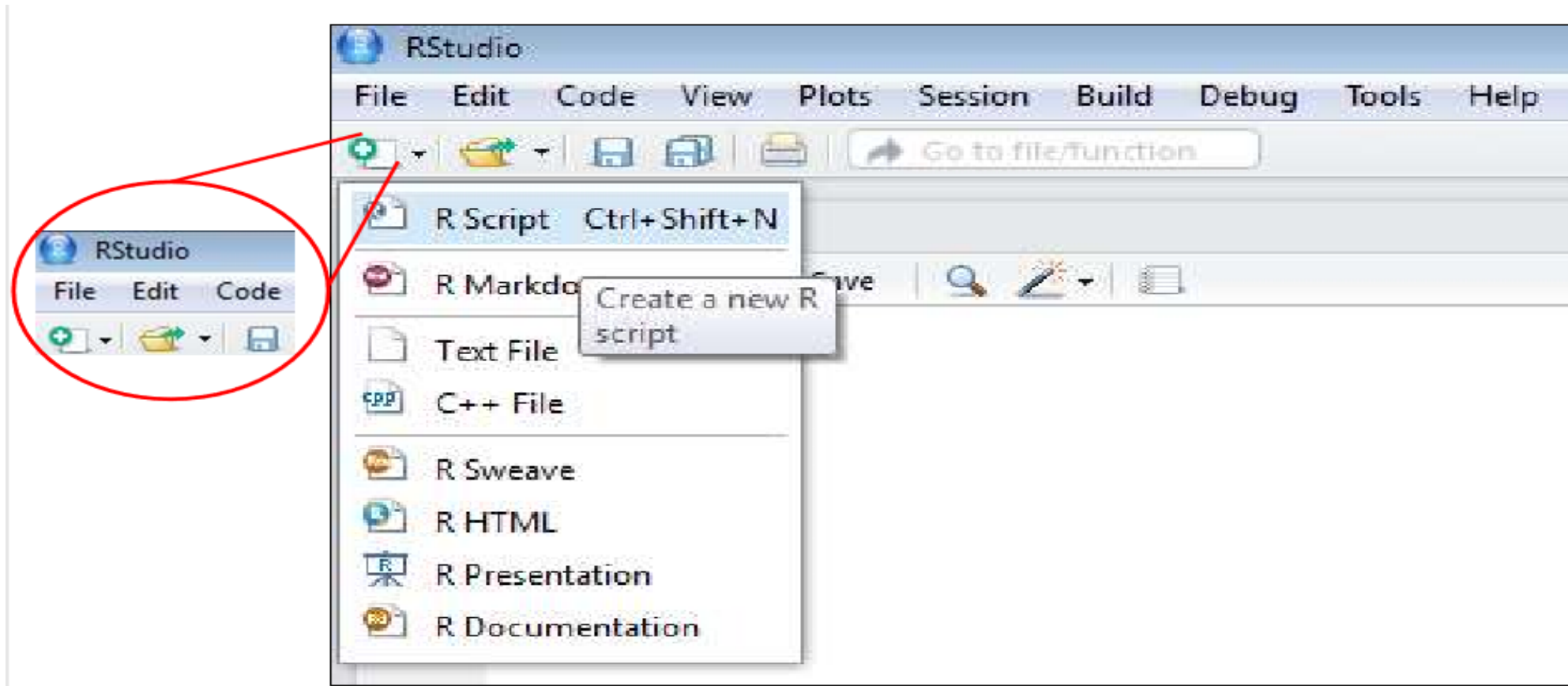
Ctrl + Shift + H → Set the working directory

Or Click on session tab → Set working directory → Choose directory

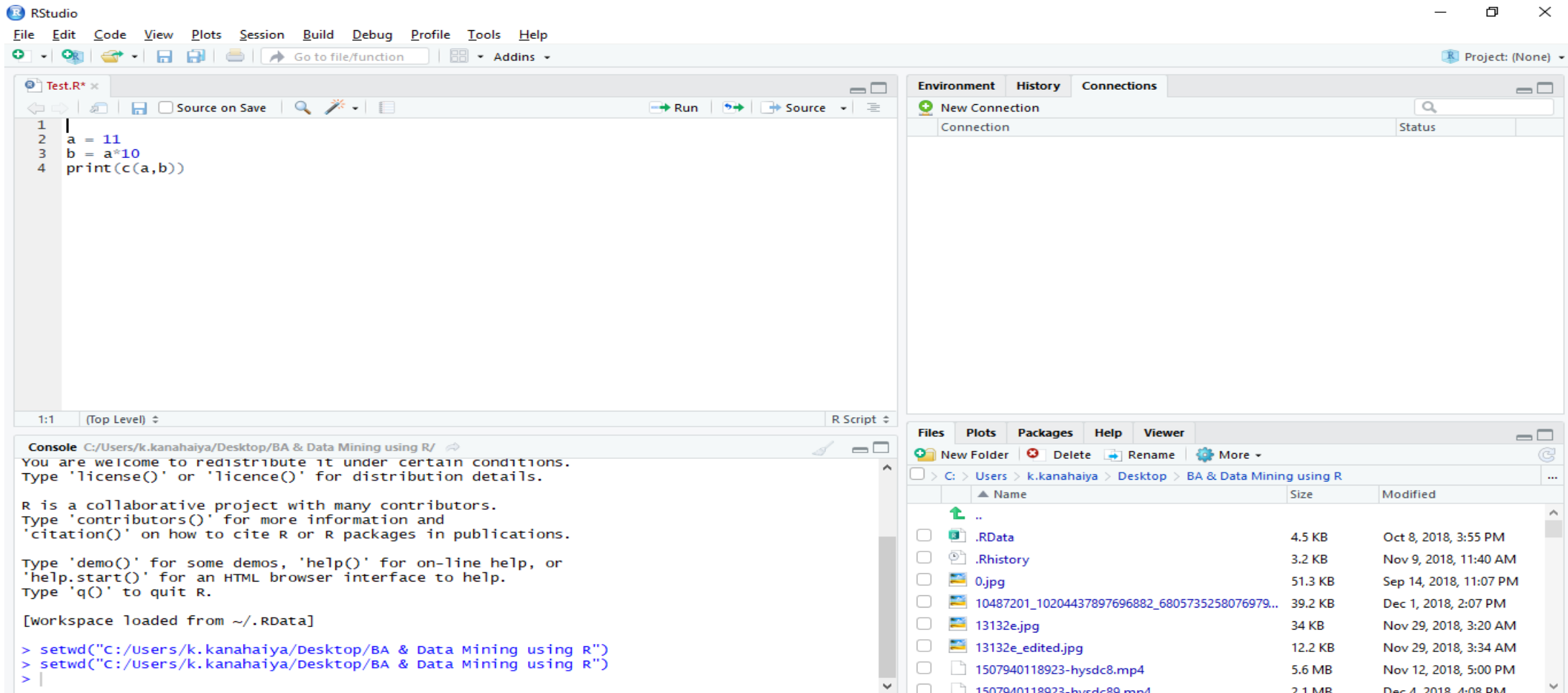
Self Study → Explore all the tabs/buttons to be familiar with R Studio IDE

Create an R file

By Clicking the “+” Button (Below File Tab)



Writing Scripts



The screenshot displays the RStudio environment with the following components:

- Source Editor:** Contains an R script named `Test.R` with the following code:

```
1 |
2 a = 11
3 b = a*10
4 print(c(a,b))
```
- Console:** Shows the R startup message and the execution of the script:

```
C:/Users/k.kanahaiya/Desktop/BA & Data Mining using R/
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

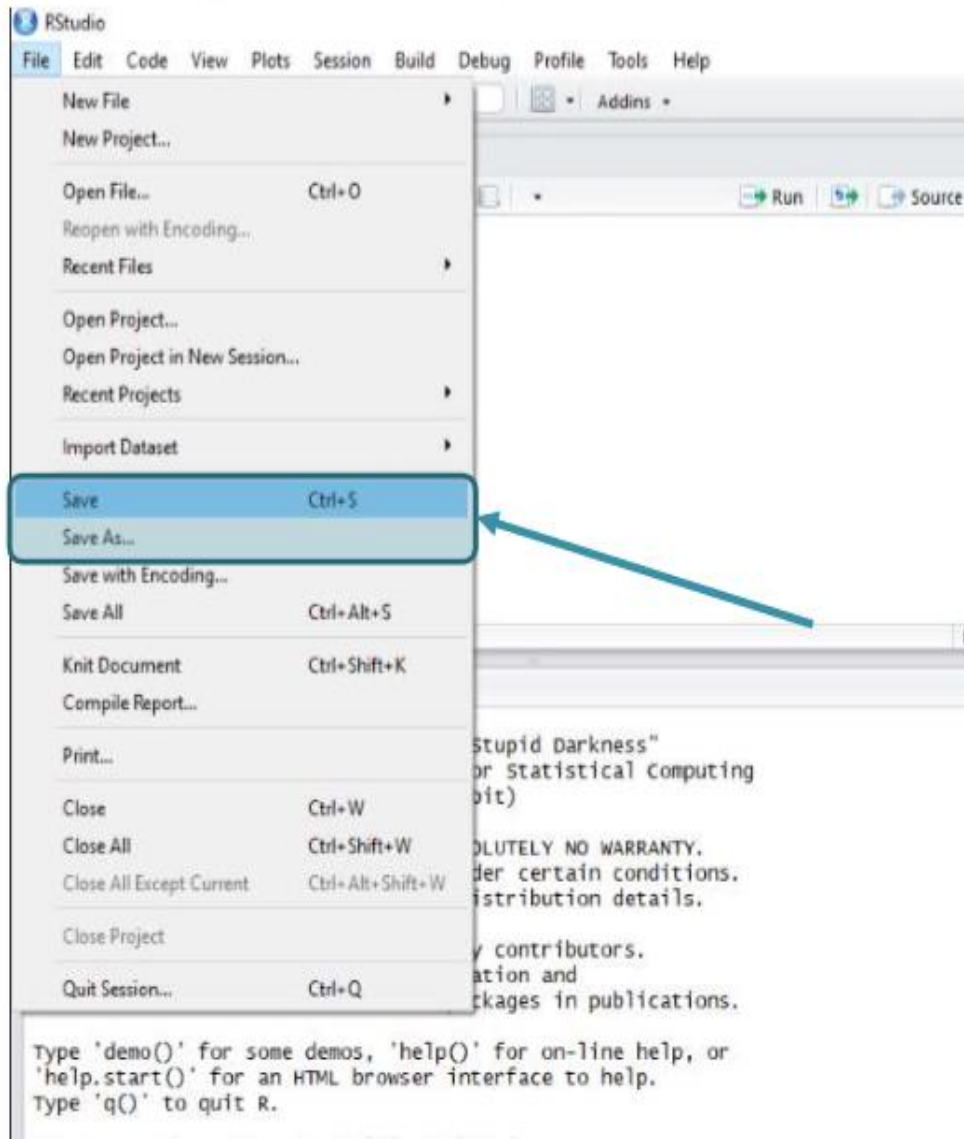
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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.

[workspace loaded from ~/.RData]
> setwd("C:/Users/k.kanahaiya/Desktop/BA & Data Mining using R")
> setwd("C:/Users/k.kanahaiya/Desktop/BA & Data Mining using R")
> |
```
- Environment/History/Connections:** The 'Environment' tab is active, showing a table with columns 'Connection' and 'Status'. It includes a 'New Connection' button and a search bar.
- Files/Plots/Packages/Help/Viewer:** The 'Files' tab is active, showing a file explorer view of the directory `C:/Users/k.kanahaiya/Desktop/BA & Data Mining using R/`. The files listed are:

Name	Size	Modified
..		
.RData	4.5 KB	Oct 8, 2018, 3:55 PM
.Rhistory	3.2 KB	Nov 9, 2018, 11:40 AM
0.jpg	51.3 KB	Sep 14, 2018, 11:07 PM
10487201_10204437897696882_6805735258076979...	39.2 KB	Dec 1, 2018, 2:07 PM
13132e.jpg	34 KB	Nov 29, 2018, 3:20 AM
13132e_edited.jpg	12.2 KB	Nov 29, 2018, 3:34 AM
1507940118923-hysdc8.mp4	5.6 MB	Nov 12, 2018, 5:00 PM
1507940118923-hysdc89.mp4	2.1 MB	Dec 4, 2018, 4:08 PM

Saving R Script File



Executing an R File

1
2 a=11
3 b<-a*10
4 print(c(a,b))

Run
Source Ctrl+Shift+S
Source with Echo Ctrl+Shift+Enter

A) Press Run/ Ctrl + Enter

B) Press Source / Ctrl + Shift + S
Or
C) Press Ctrl + Shift + Enter (Source with echo)

Source will execute all commands in console without displaying them where as source with echo will execute and print automatically

Executing an R File: Example

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains the script `Test.R` with the following code:

```
1 a = 11
2 b = a*10
3 print(c(a,b))
```
- Console:** Shows the execution output:

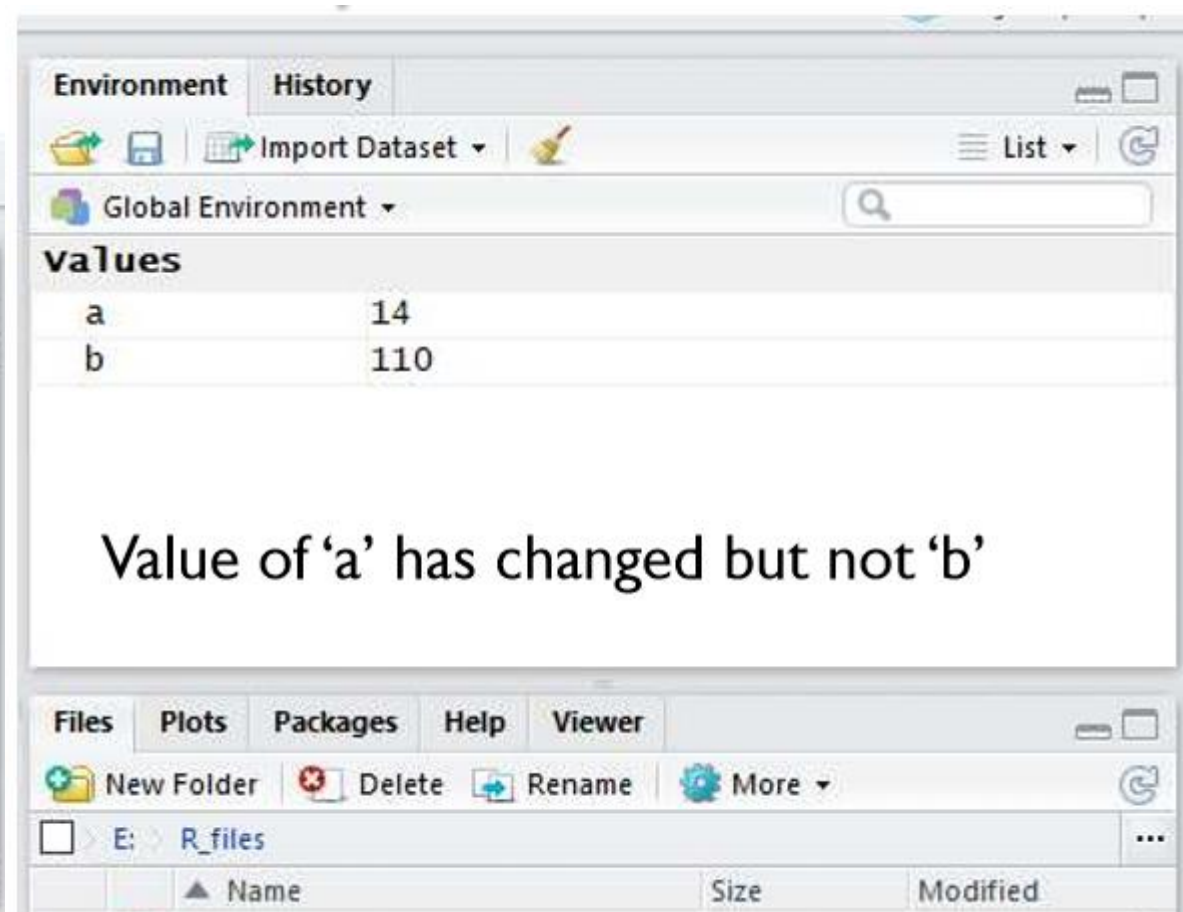
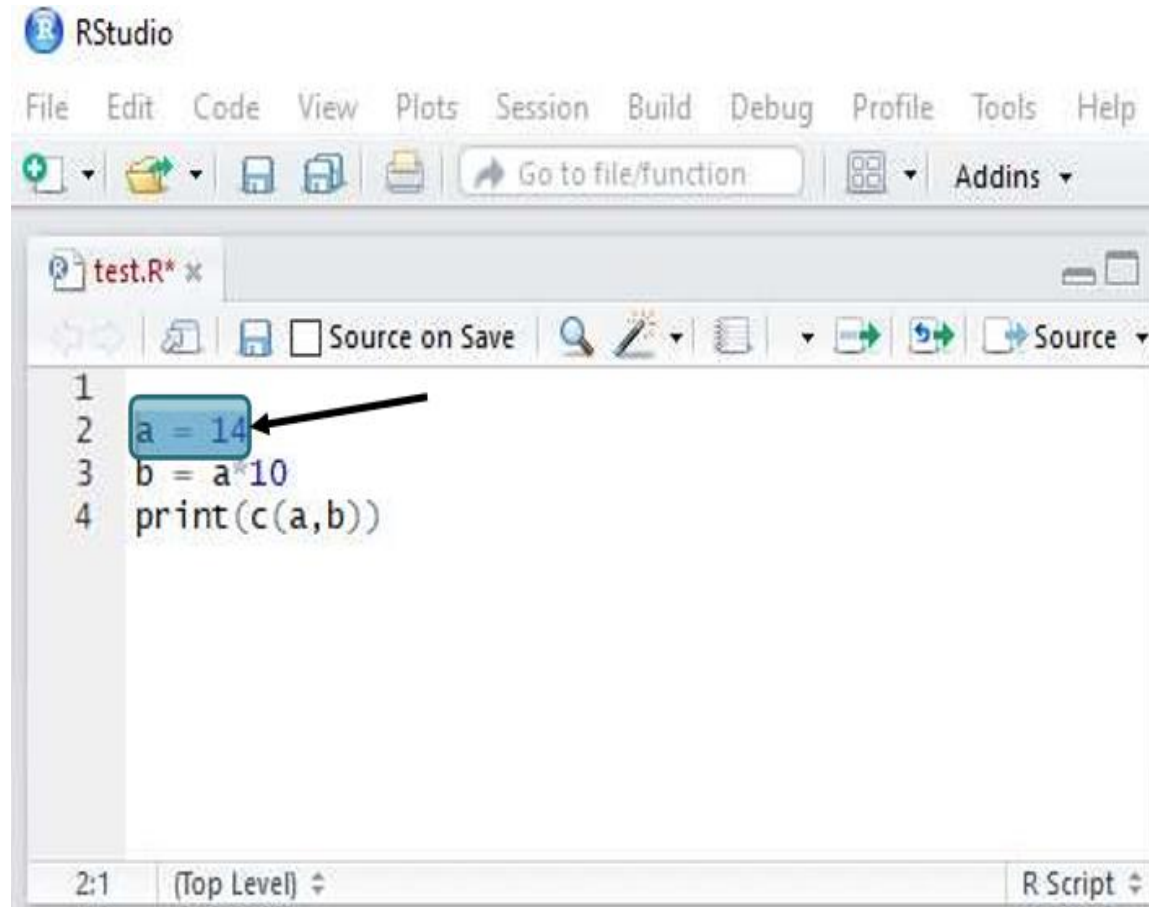
```
C:/Users/k.kanahaiya/Desktop/BA & Data Mining using R/
Type 'q()' to quit R.

[workspace loaded from ~/.RData]
> setwd("C:/Users/k.kanahaiya/Desktop/BA & Data Mining using R")
> setwd("C:/Users/k.kanahaiya/Desktop/BA & Data Mining using R")
> a = 11
> source('C:/Users/k.kanahaiya/Desktop/BA & Data Mining using R/Test.R', echo=TRUE)
> a = 11
> b = a*10
> print(c(a,b))
[1] 11 110
>
```
- Environment:** Shows 'New Connection' and 'Connection' status.
- Files:** Displays a file explorer view of the directory `C:/Users/k.kanahaiya/Desktop/BA & Data Mining using R`. The files listed are:

Name	Size	Modified
..		
.RData	4.5 KB	Oct 8, 2018, 3:55 PM
.Rhistory	3.2 KB	Nov 9, 2018, 11:40 AM
0.jpg	51.3 KB	Sep 14, 2018, 11:07 PM
10487201_10204437897696882_6805735258076979...	39.2 KB	Dec 1, 2018, 2:07 PM
13132e.jpg	34 KB	Nov 29, 2018, 3:20 AM
13132e_edited.jpg	12.2 KB	Nov 29, 2018, 3:34 AM
1507940118923-hysdc8.mp4	5.6 MB	Nov 12, 2018, 5:00 PM
1507940118923-hysdc89.mp4	2.1 MB	Dec 4, 2018, 4:08 PM

A blue arrow points from the text **Result** to the console output `[1] 11 110`.

Execute pieces of code - Run



Executing R Files – Summary

- Run can be used to execute selected lines
- Source / Source with echo is for a whole file
- Advantages – Using Run:
 - Troubleshooting /debugging
- Disadvantages – Using Run:
 - For large section, console will be over populated and messy

How to

- Add Comments
 - For single line comment, insert '#' at the start of the line
 - For multiple line comment, select multiple lines using cursor, then press CTRL+Shift+C
 - Or Select multiple lines using cursor, click on "Code" in menu and select "Comment/Uncomment lines"
- Clear the environment
 - To clear console, CTRL+L
 - To clear environment, Single variable: Enter in console/R Script: `rm(variable)`
 - All variables: Enter in console/R Script: `rm(list=ls())`

How to

- Saving the workspace
 - Workspace data: Workspace information is temporary
 - Workspace data is not retained after the session
 - If you close the R-Session
 - If you restart the computer