



DATA ANALYTICS

Tutorial #I

Programming with R

Dr. Sreeja S R

Assistant Professor

**Indian Institute of Information Technology
IIIT Sri City**

QUOTE OF THE DAY..

- What we think, we become.
- GAUTAMA BUDDHA, Sege



TODAY'S DISCUSSION...

- R is an open source 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 analytics tools



HISTORY OF R

- Modelled after S & S-plus, developed at AT&T labs in late 1980s.
- R project was started by Robert Gentleman and Ross Ihaka Department of Statistics, University of Auckland (1995).
- Currently maintained by R core development team – an international team of volunteer developers (since 1997).

FEATURES OF R PROGRAMMING

- **R** is an open source programming language. R possesses an extensive catalogue of statistical and graphical methods.
- It includes machine learning algorithms, linear regression, time series, statistical inferences.
- Most of the R libraries are written in R, but for heavy computational tasks, C, C++ and Fortran codes are preferred.
- R is not only entrusted by academic, but many large companies also use R programming language, including Uber, Google, Airbnb, Facebook and so on.

FEATURES OF R PROGRAMMING

- Comprehensive Environment
- Can Perform Complex Statistical Calculations
- Distributed Computing
- Running Code Without a Compiler
- Data Variety
- Data Handling and Storage
- Vector Arithmetic
- Generates Report in any Desired Format

R RESOURCES

- <http://www.r-project.org/>
- <http://cran.r-project.org/doc/contrib/Verzani-SimpleR.pdf>

DOWNLOAD R AND RSTUDIO

- Download R :

<http://cran.r-project.org/bin/>

- Download RStudio :

<http://www.rstudio.com/ide/download/desktop>

INSTALLATION

Installing R on windows PC :

- ❑ Use internet browser to point to : <http://mirror.aarnet.edu.au/pub/CRAN>
- ❑ Under the heading Precompiled Binary Distributions, choose the link Windows.
- ❑ Next heading is R for Windows; choose the link base.
- ❑ Click on download option(R 3.4.1 for windows).
- ❑ Save this to the folder C:\R on your PC.
- ❑ When downloading is complete, close or minimize the Internet browser.
- ❑ Double click on R 3.4.1-win32.exe in C:\R to install.

Installing R on Linux:

- ❑ `sudo apt-get install r-base-core`

INSTALLATION

Installing RStudio:

- Go to www.rstudio.com and click on the "Download RStudio" button.
- Click on "Download RStudio Desktop."
- Click on the version recommended for your system, or the latest Windows version, and save the executable file. Run the .exe file and follow the installation instructions.

VERSION

- Get R version

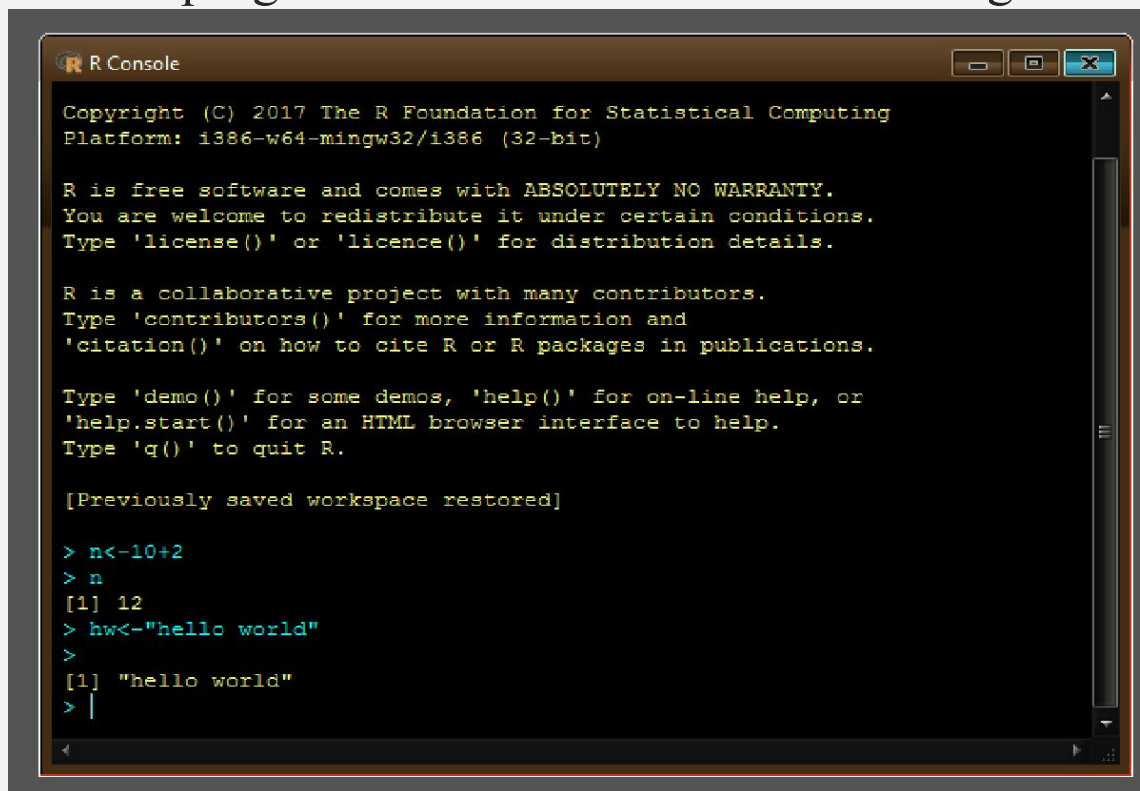
`R.Version()`

- Get RStudio version

RStudio: Toolbar at top > Help > About RStudio

A TEST RUN WITH R IN WINDOWS

- Double click the R icon on the Desktop and the R Console will open.
- Wait while the program loads. You observe something like this.



```
R Console

Copyright (C) 2017 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-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.

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.

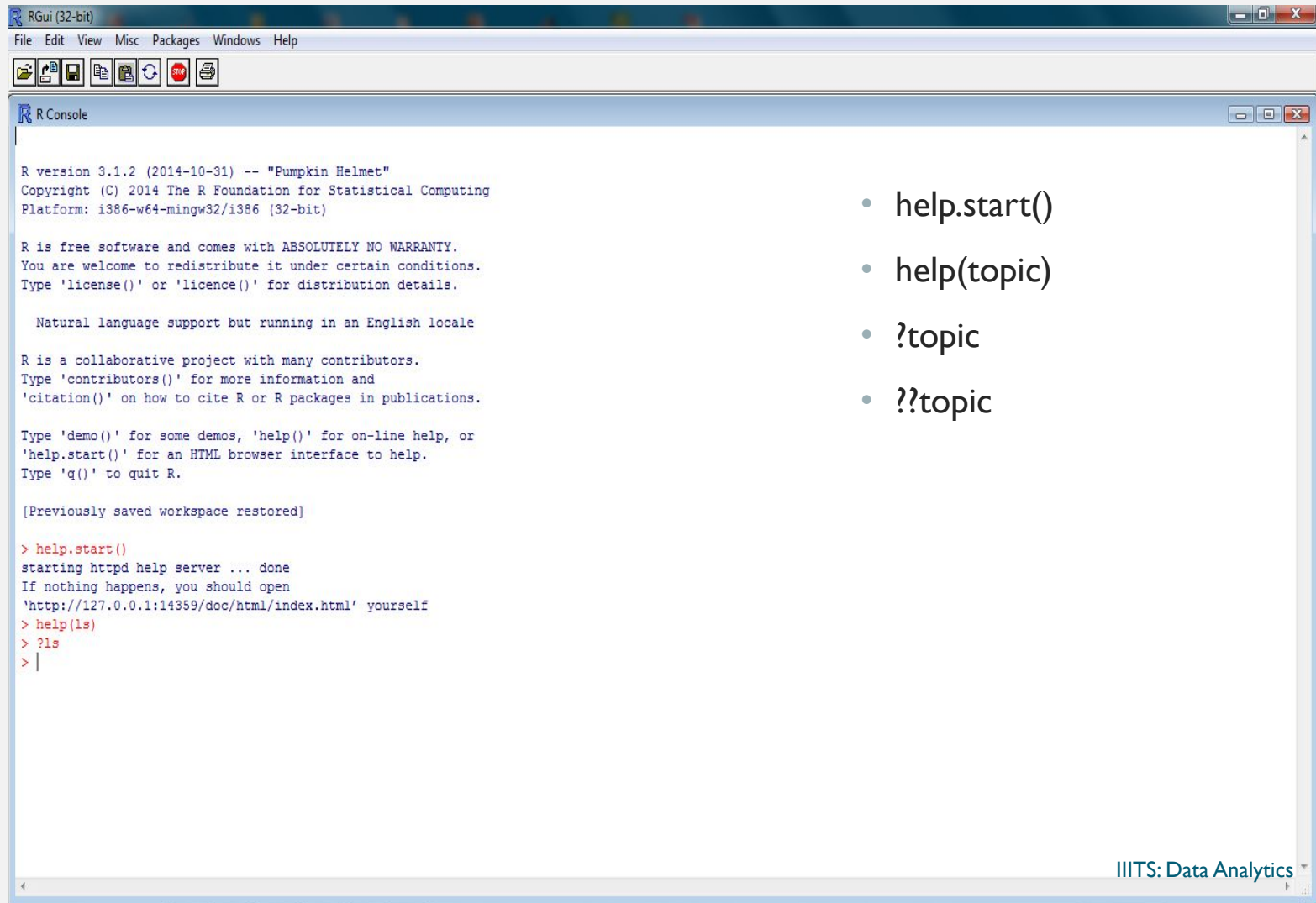
Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

[Previously saved workspace restored]

> n<-10+2
> n
[1] 12
> hw<-"hello world"
>
[1] "hello world"
> |
```

- You can type your own program at the prompt line `>`.

GETTING HELP FROM R CONSOLE



The screenshot shows the RGui (32-bit) window. The menu bar includes File, Edit, View, Misc, Packages, Windows, and Help. The toolbar contains icons for file operations and running code. The R Console window displays the following text:

```
R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet"
Copyright (C) 2014 The R Foundation for Statistical Computing
Platform: i386-w64-mingw32/i386 (32-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]

> help.start()
starting httpd help server ... done
If nothing happens, you should open
'http://127.0.0.1:14359/doc/html/index.html' yourself
> help(ls)
> ?ls
> |
```

- `help.start()`
- `help(topic)`
- `?topic`
- `??topic`

IIITS: Data Analytics

R COMMAND IN INTEGRATED ENVIRONMENT

The screenshot displays the RStudio interface with the following components:

- Script Editor:** Contains an R script with the following code:

```
1 1+1
2 x=c(1,2,3,4)
3 x
4 y=c(3,4,5)
5 y
6 z=prod(x,y)
7 2==2
8 a<-x>3
9 a
10 b<-mean(c(1,2,3,4))
11 b
12 x<-c("apple",
13       "banana")
14
```
- Console:** Shows the execution of the script, with an error message for the first line:

```
D:/arpita/data analytics/my work/
length
> x.y
Error: object 'x.y' not found
> prod(x,y)
[1] 1440
> z=prod(x,y)
> 1+1
[1] 2
> x=c(1,2,3,4)
> x
[1] 1 2 3 4
> y=c(3,4,5)
> y
[1] 3 4 5
> z=prod(x,y)
> 2==2
```
- Environment Pane:** Displays the current environment with the following data and values:

Global Environment	
Data	
data	149 obs. of 5 variables
X5.1	num 4.9 4.7 4.6 5 5.4 4.6 5.2 4.4 4.9 5.4 ...
X3.5	num 3 3.2 3.1 3.6 3.9 3.4 3.4 2.9 3.1 3.7 ...
X1.4	num 1.4 1.3 1.5 1.4 1.7 1.4 1.5 1.4 1.5 1.5 ...
X0.2	num 0.2 0.2 0.2 0.2 0.4 0.3 0.2 0.2 0.1 0.2 ...
Iris.setosa	Factor w/ 3 levels "Iris-setosa",...: 1 1 1 1 1 1 1 1 1 1 ...
values	
a	logi [1:4] FALSE FALSE FALSE TRUE
b	2.5
x	num [1:4] 1 2 3 4

HOW TO USE R FOR SIMPLE MATHS

- `> 3+5`
- `> 12 + 3 / 4 - 5 + 3*8`
- `> (12 + 3 / 4 - 5) + 3*8`
- `> pi * 2^3 - sqrt(4)`
- `> factorial(4)`
- `> log(2, 10)`
- `> log(2, base=10)`
- `> log10(2)`
- `> log(2)`

Note

- R ignores spaces

HOW TO STORE RESULTS OF CALCULATIONS FOR FUTURE USE

- `> x = 3+5`
- `> x`
- `> y = 12 + 3 / 4 - 5 + 3*8`
- `> y`
- `> z = (12 + 3 / 4 - 5) + 3*8`
- `> z`
- `> A <- 6 + 8` `## no space should be between < & -`
- `> a` `## Note: R is case sensitive`
- `>A`

IDENTIFIERS NAMING

- Don't use underscores (_) or hyphens (-) in identifiers.
- The preferred form for variable names is all lower case letters and words separated with dots (variable.name) but variableName is also accepted.
- Examples:

avg.clicks	GOOD
avgClicks	OK
avg_Clicks	BAD
- Function names have initial capital letters and no dots (e.g., **FunctionName**).

Any question?