DATA ANALYTICS

Tutorial #1

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 a 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

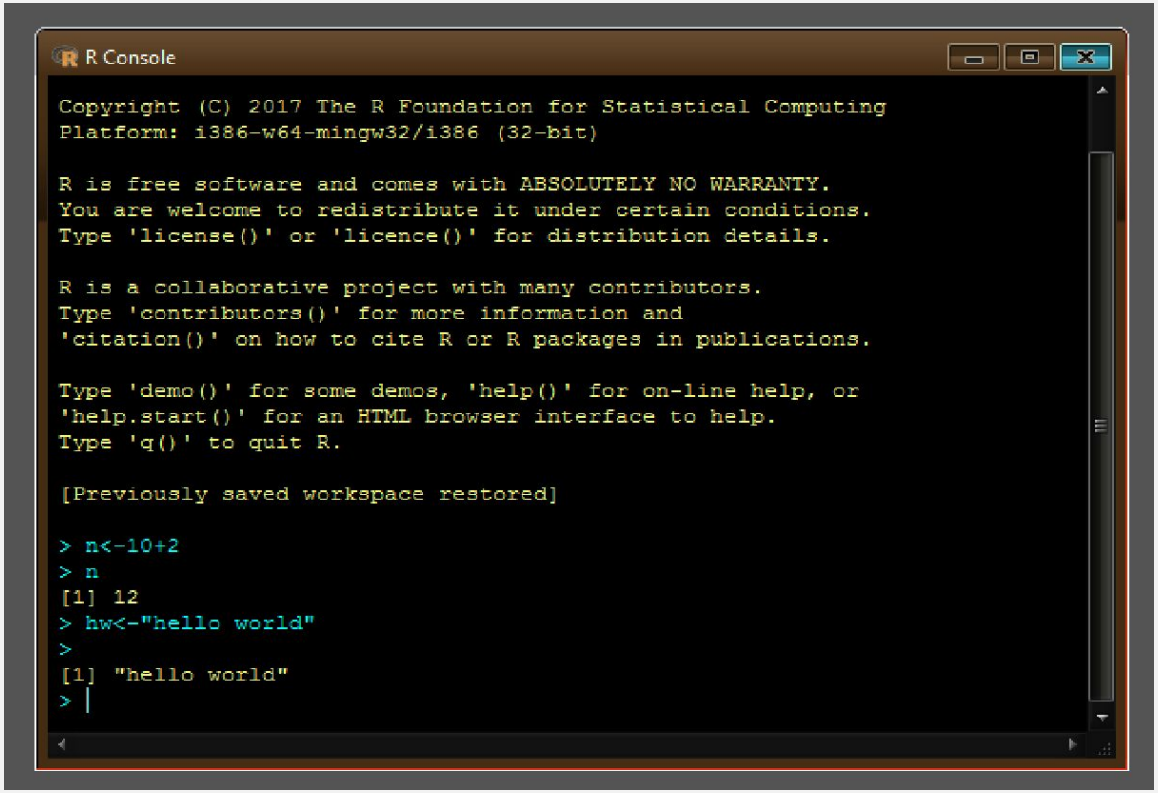
  R.Version()

- Get RStudio version

  RStudio: Toolbar at top > Help > About RStudio

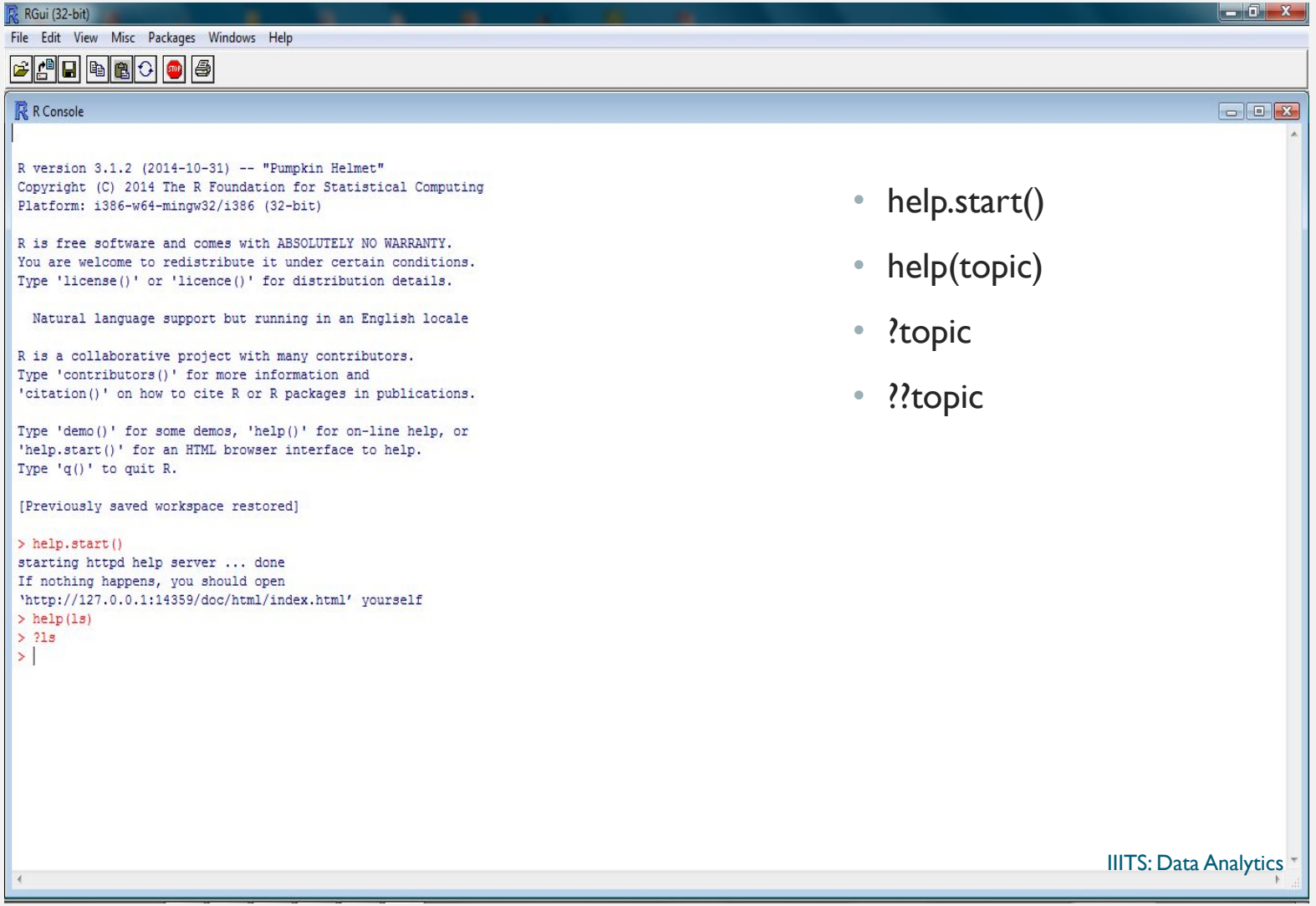
A TEST RUN WITH R IN WINDOWS

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



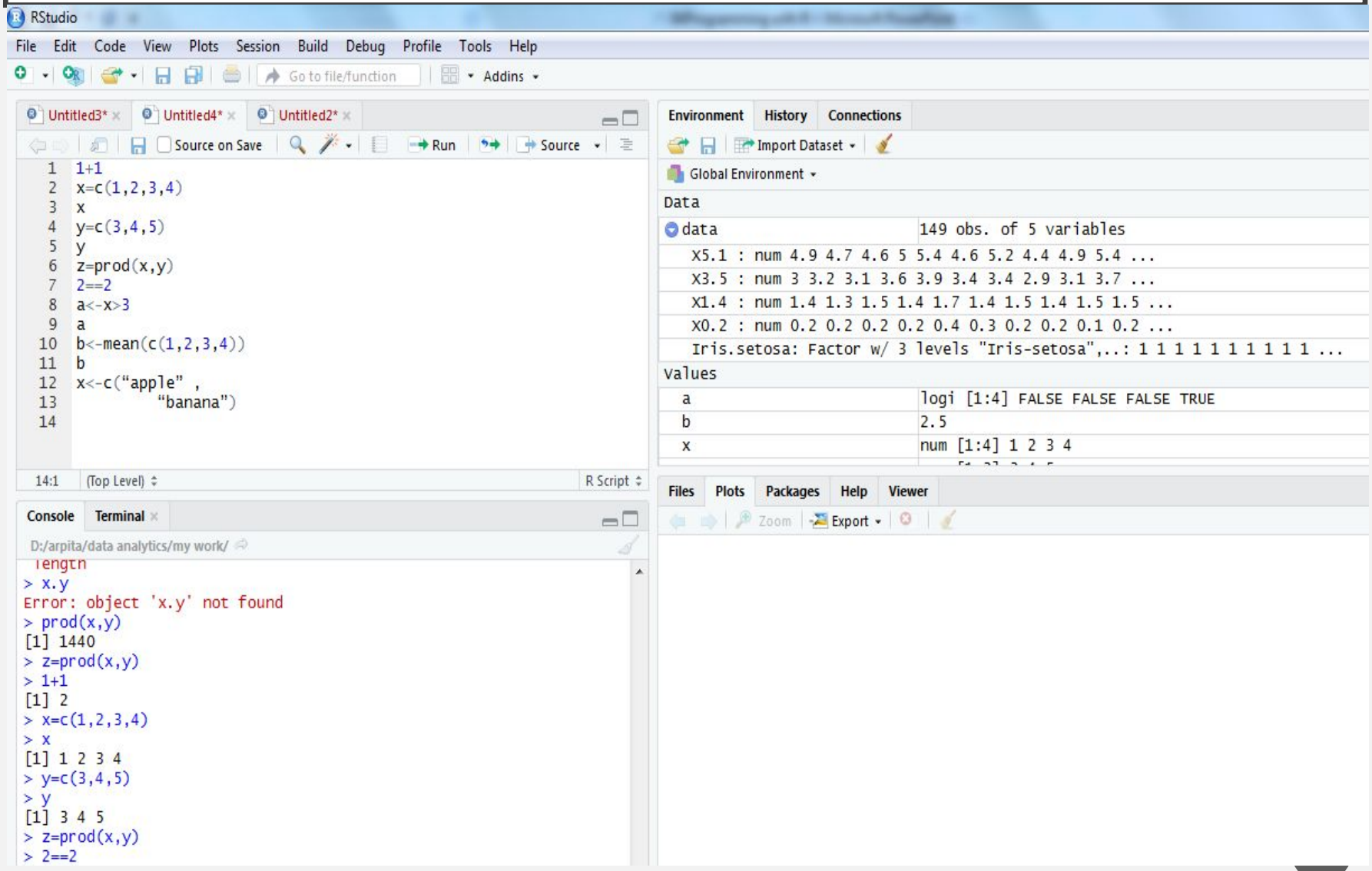
- You can type your own program at the prompt line >.

GETTING HELP FROM R CONSOLE



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

R COMMAND IN INTEGRATED ENVIRONMENT



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?