Introduction to R-part1

Sameh MAGDELDIN, M.V.Sc, 2 Ph.D

November, 2024

my graph

Before we start,, why **R** you here? [Opening statement]

==

What is R?

Program language, widely used among researchers used for statistical computing,big data processing, and elegant graphics

#Creator; Ross Ihaka

A language called R. And as those millions use it to predict shifts in the stock market, track the behaviour of marine life, or search for a cure for cancer, the effects of this statistical programming language – developed by Associate Professor Ross Ihaka and his colleagues at the University of Auckland – reach far beyond millions.

That said, what Ross is focused on now is the future. As datasets grow larger, so too does the need for R to become more powerful.

So whether it's being used by an internet giant in California, or a little-known institute near Cape Town, this adaptive, open-source platform known by just a single letter, will allow the new challenges we face to be understood more deeply, and in turn bring about amazing new breakthroughs.

For more about Ross and R, or how others are achieving the amazing at the University of Auckland, visit **achievetheamazing.ac.nz/research**

my graph

Why R?

1. It is completely free

2. It is open-source

3. useful for large data “unlike speadsheets”

4. Top quality graphics

5. big community of users among researchers and easy support

6. Superiority by far over current statistical and graphical softwares available such as SPSS , SAS, Matlab and others

http://www.burns-stat.com/pages/Tutor/R_relative_statpack.pdf (http://www.burns-stat.com/pages/Tutor/R_relative_statpack.pdf)

From my opinion, 3.5 reason to switch to R

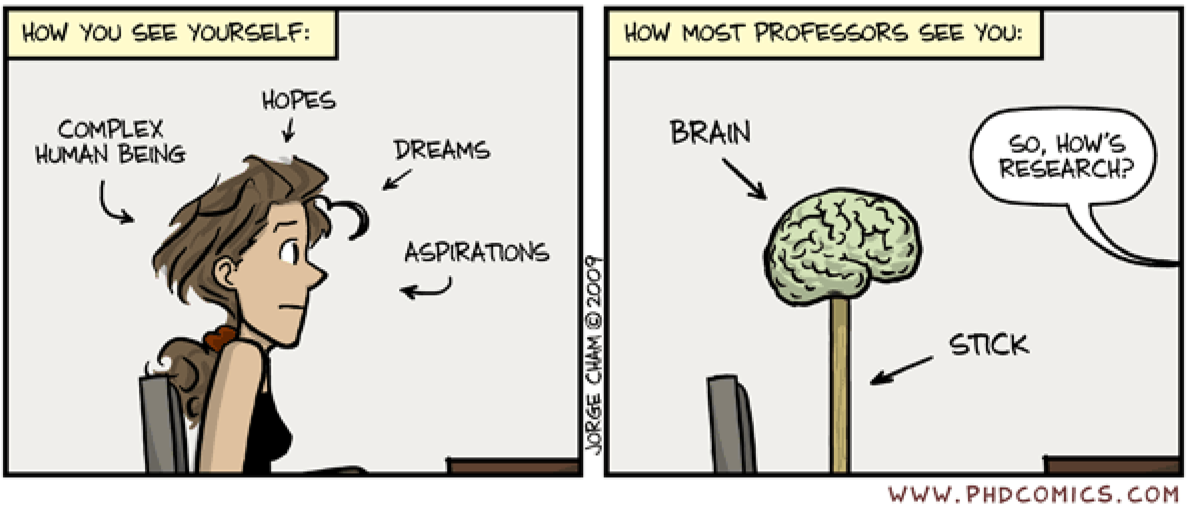
1. Safety

2. Speed

3. No reason spread sheet have poor graphics and statistics, but simply, good graphics and statistics haven’t happened

0.5. I like it (its classy) like this cat:)

Additionally, R will solve this problem with your professor



my graph

Researchers were asked in a marketing survey, tell me what you want? but guess what, majority of them (~73%) don’t know their

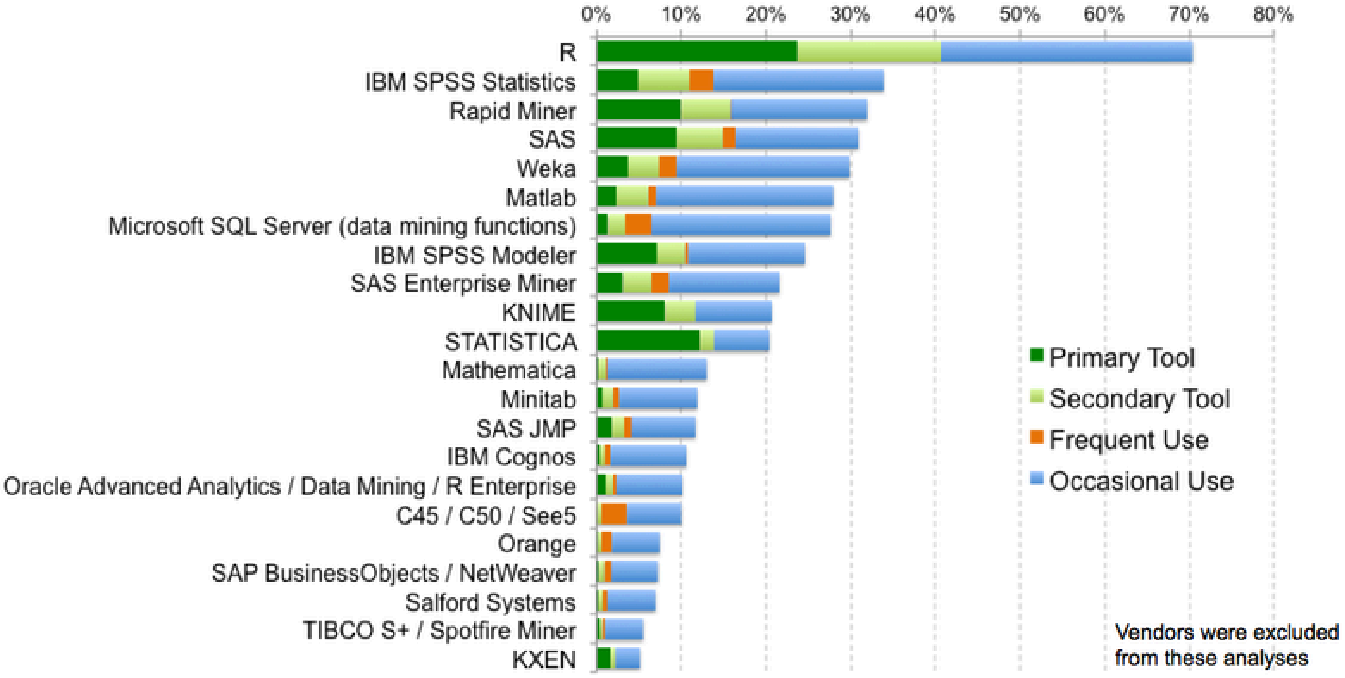
needs!!

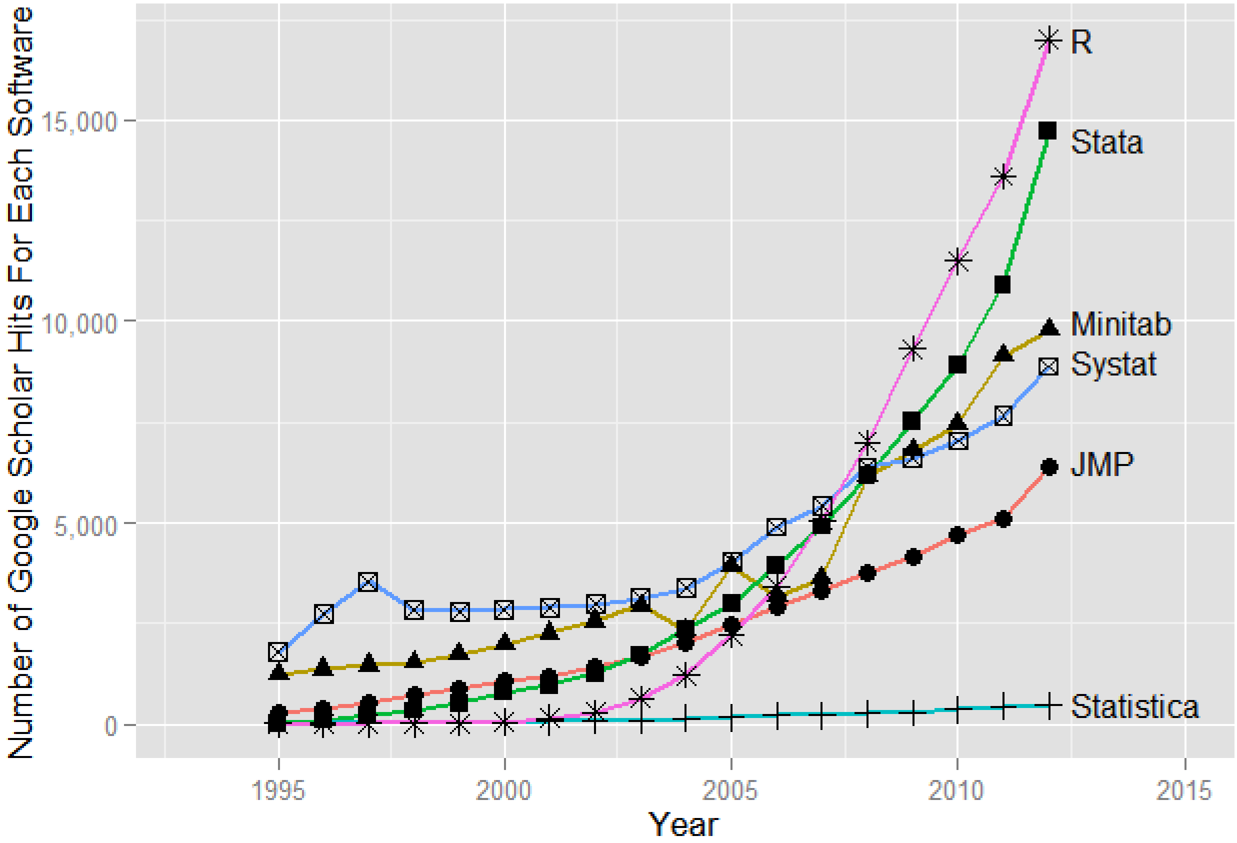
“People don’t know what they want until you show it to them”

Steve Jobs

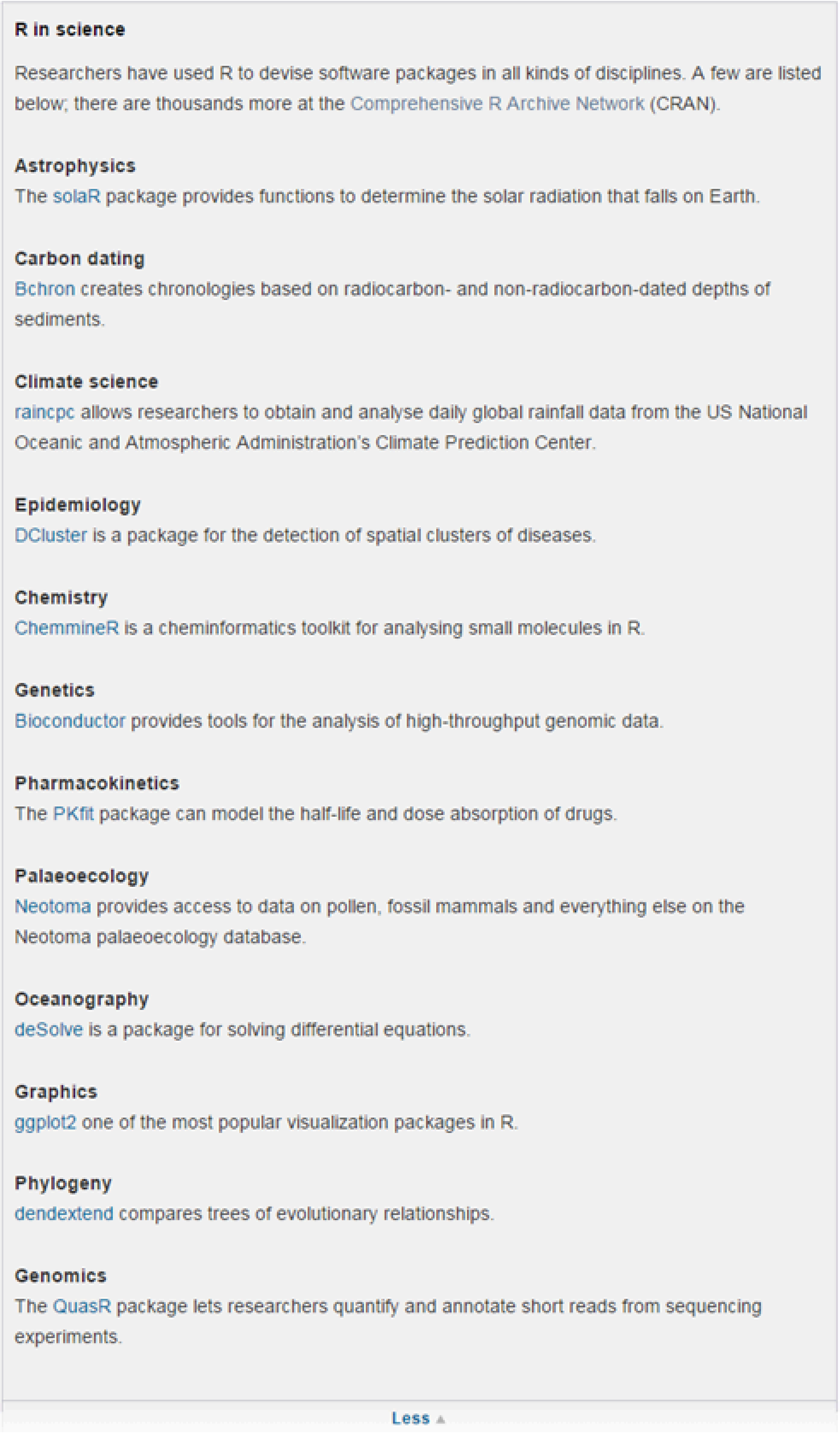
my graph

R users statistics





** You can find solution packages for your professional needs



R in science

Researchers have used R to devise software packages in all kinds of disciplines. A few are listed below; there are thousands more at the Comprehensive R Archive Network (CRAN).

Astrophysics
The solaR package provides functions to determine the solar radiation that falls on Earth.

Carbon dating
Bchron creates chronologies based on radiocarbon- and non-radiocarbon-dated depths of sediments.

Climate science
raincpc allows researchers to obtain and analyse daily global rainfall data from the US National Oceanic and Atmospheric Administration's Climate Prediction Center.

Epidemiology
DCluster is a package for the detection of spatial clusters of diseases.

Chemistry
ChemmineR is a cheminformatics toolkit for analysing small molecules in R.

Genetics
Bioconductor provides tools for the analysis of high-throughput genomic data.

Pharmacokinetics
The PKfit package can model the half-life and dose absorption of drugs.

Palaeoecology
Neotoma provides access to data on pollen, fossil mammals and everything else on the Neotoma palaeoecology database.

Oceanography
deSolve is a package for solving differential equations.

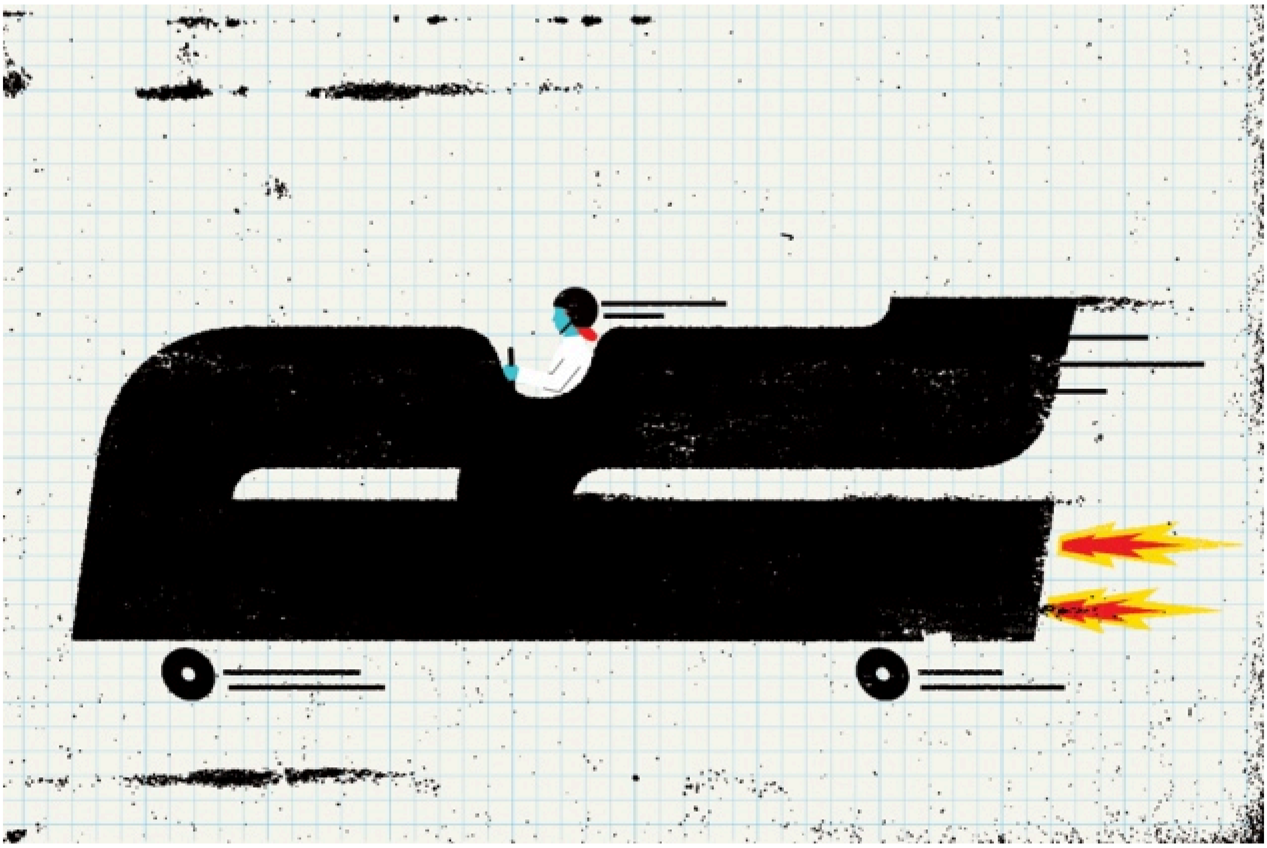
Graphics
ggplot2 one of the most popular visualization packages in R.

Phylogeny
dendextend compares trees of evolutionary relationships.

Genomics
The QuasR package lets researchers quantify and annotate short reads from sequencing experiments.

my graph

So working with R is an adventure



my graph

==

#Getting started (Installation)

==

We need to download R from “Comprehensive R Archive Network- CRAN”

http://cran.r-project.org/ (http://cran.r-project.org/)

Install it !

Hint1. Druing R installation, use default setting, such as C as destination folder.

Hint2. a list of components to install will appear, choose, core files, message translation and select one of 32 bit or 64 bit depend on your operating system.

Hint3. For startup option, select “No (Accept dafaults)”

Hint4. In additional tasks select “create desktop icon, save version number, associate R with .RData”

Now you have R on your PC machine - navigate for the R logo

You can work with R as it is, this is called the naked version, however, another software called R studio makes life more easy in term of viewing and dealing with R. Note that u need to install R first then R studio, R studio can not work alone but uses R as a core.

you can download R studio from their webpage

https://www.rstudio.com/products/rstudio/download/ (https://www.rstudio.com/products/rstudio/download/)

Install it !

navigate for R studio “Blue circle with R”

==

Working with R studio

==

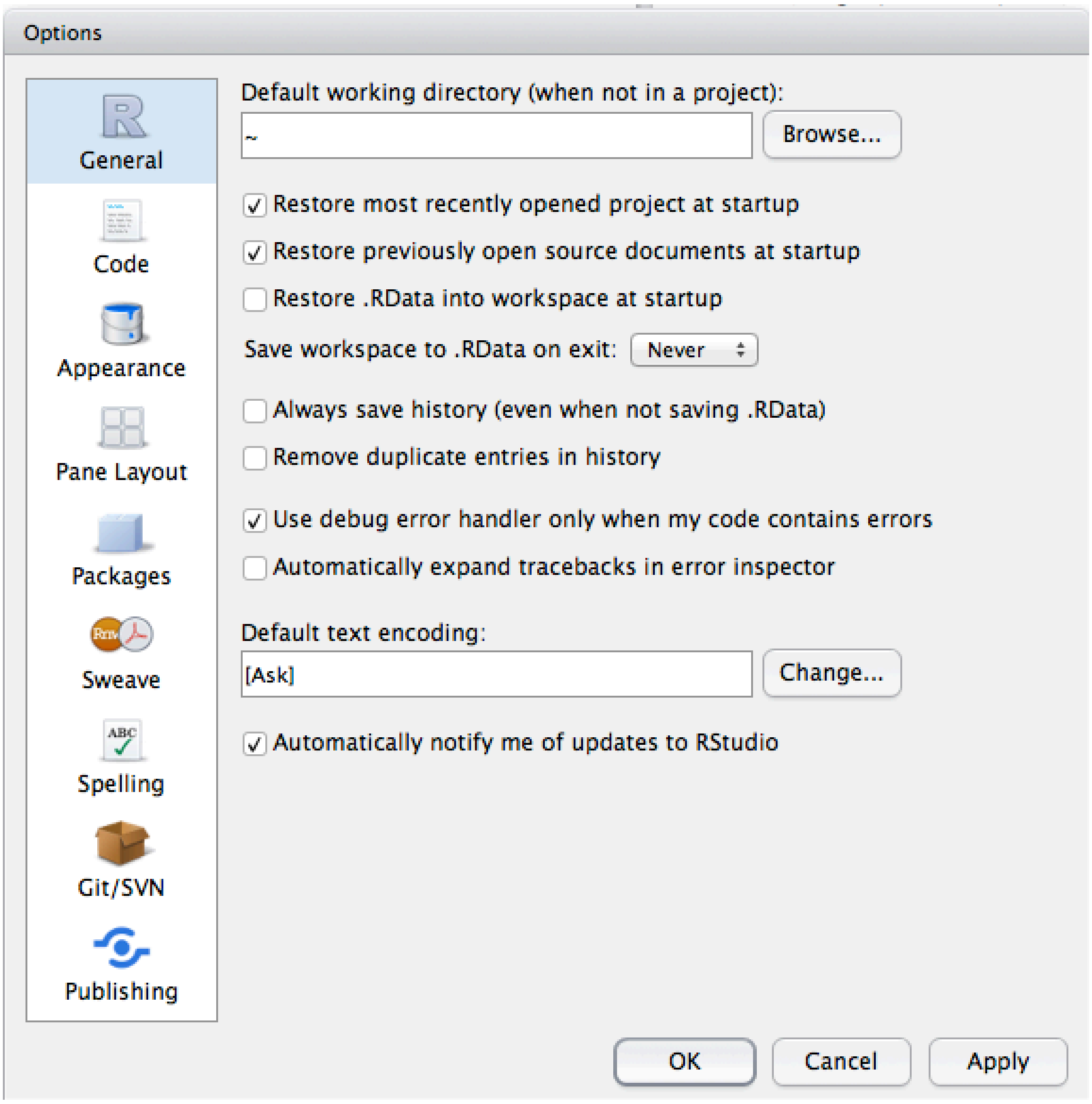
##Basic setting

###Before we dig deep in R , i would like to pass through some setting in R studio

###1. open R studio -> Tools-> global setting-General

If you would like R studio to store the session and load it again, u need to check “restore .RData on exit”. some ppl like to start on a clean base, in that case “unckeck it and also set”save

workspace on exit to NEVER””



Default working directory (when not in a project): ~

☑ Restore most recently opened project at startup

☑ Restore previously open source documents at startup

□ Restore .RData into workspace at startup

Save workspace to .RData on exit: Never

□ Always save history (even when not saving .RData)

□ Remove duplicate entries in history

☑ Use debug error handler only when my code contains errors

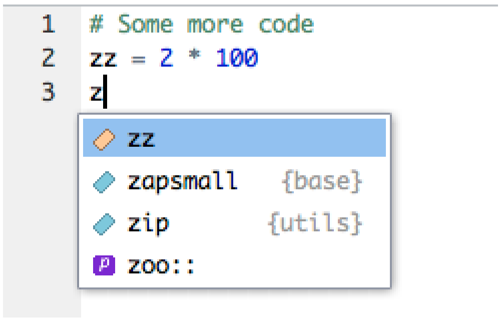
□ Automatically expand tracebacks in error inspector

Default text encoding: [Ask]

☑ Automatically notify me of updates to RStudio

my graph

R studio can do auto complete whenever you start typing then press “Tab” key a pop up menu will appear for you.



```
# Some more code
zz = 2 * 100
z
```

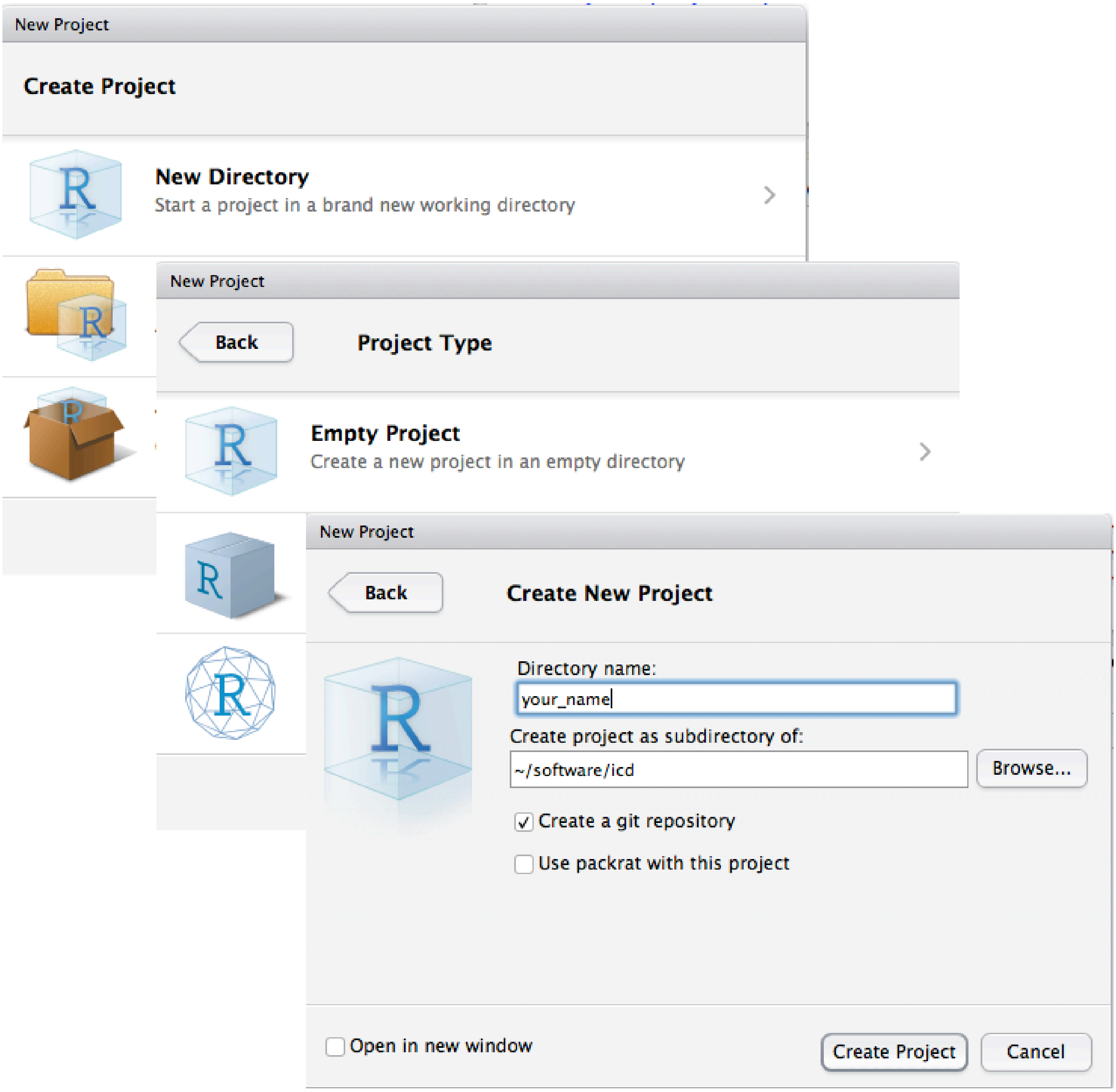
my graph

###R studio is very powerfull, using projects management, you can control several project, working with several project at once.

Lets create a new R Studio project

Select File > New project

###Select New Directory ###Select Empty project ###Select a name and sub directory of your project.



New Project

Create Project

New Directory

Start a project in a brand new working directory

New Project

Project Type

Empty Project

Create a new project in an empty directory

New Project

Create New Project

Directory name: your_name

Create project as subdirectory of: ~/software/icd

☑ Create a git repository

☐ Use packrat with this project

☐ Open in new window

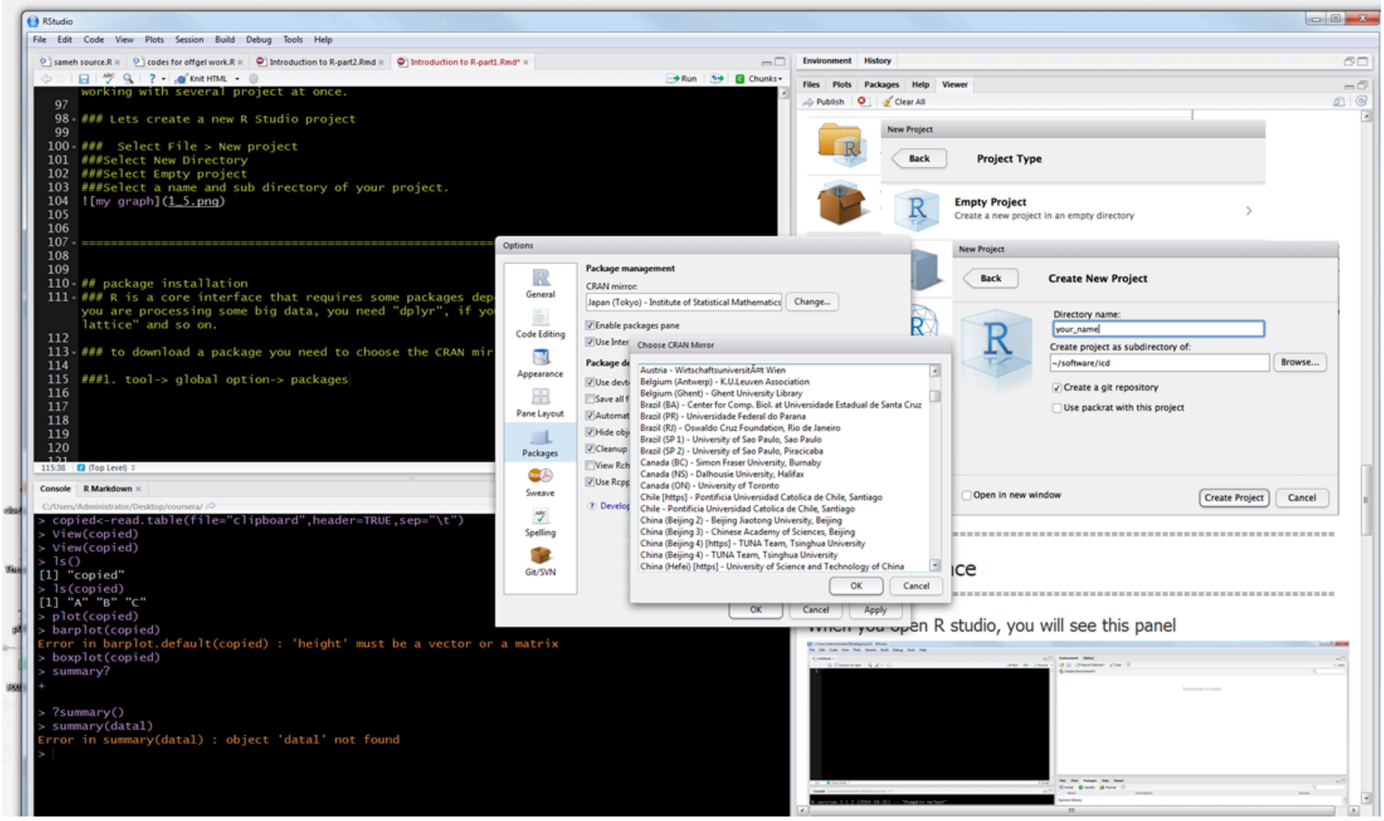
==

package installation

R is a core interface that requires some packages depending on what you do. For example if you are processing some big data, you need “dplyr”, if you want graphics you need “ggplot2 or lattice” and so on.

to download a package you need to choose the CRAN mirror

###1. tool-> global option-> packages



```
working with several project at once.

### Lets create a new R Studio project

###  Select File > New project
###Select New Directory
###Select Empty project
###Select a name and sub directory of your project.
![my graph](1_5.png)


===================================================================


## package installation
### R is a core interface that requires some packages dep
you are processing some big data, you need "dplyr", if yo
lattice" and so on.

### to download a package you need to choose the CRAN mir

###1. tool-> global option-> packages
```

C:/Users/Administrator/Desktop/coursera/

```
> copied<-read.table(file="clipboard",header=TRUE,sep="\t")
> View(copied)
> View(copied)
> ls()
[1] "copied"
> ls(copied)
[1] "A" "B" "C"
> plot(copied)
> barplot(copied)
Error in barplot.default(copied) : 'height' must be a vector or a matrix
> boxplot(copied)
> summary?
+

> ?summary()
> summary(datal)
Error in summary(datal) : object 'datal' not found
>
```

Package management

CRAN mirror: Japan (Tokyo) - Institute of Statistical Mathematics

☑ Enable packages pane

☑ Use Inter

Package de

☑ Use devt

□ Save all f

☑ Automat

☑ Hide obj

☑ Cleanup

□ View Rch

☑ Use Rcpp

Develop

Austria - WirtschaftsuniversitÃ¤t Wien
Belgium (Antwerp) - K.U.Leuven Association
Belgium (Ghent) - Ghent University Library
Brazil (BA) - Center for Comp. Biol. at Universidade Estadual de Santa Cruz
Brazil (PR) - Universidade Federal do Parana
Brazil (RJ) - Oswaldo Cruz Foundation, Rio de Janeiro
Brazil (SP 1) - University of Sao Paulo, Sao Paulo
Brazil (SP 2) - University of Sao Paulo, Piracicaba
Canada (BC) - Simon Fraser University, Burnaby
Canada (NS) - Dalhousie University, Halifax
Canada (ON) - University of Toronto
Chile [https] - Pontificia Universidad Catolica de Chile, Santiago
Chile - Pontificia Universidad Catolica de Chile, Santiago
China (Beijing 2) - Beijing Jiaotong University, Beijing
China (Beijing 3) - Chinese Academy of Sciences, Beijing
China (Beijing 4) [https] - TUNA Team, Tsinghua University
China (Beijing 4) - TUNA Team, Tsinghua University
China (Hefei) [https] - University of Science and Technology of China

Empty Project
Create a new project in an empty directory

Create New Project

Directory name: your_name

Create project as subdirectory of: ~/software/icd

☑ Create a git repository

□ Use packrat with this project

□ Open in new window

ce

When you open R studio, you will see this panel

my graph

alternatively, you can search the package name (ex: ggplot2 download) download the zip or tar.gz file to your pc, then load it

packages -> Install-> Install from package archive-> browse your file-> install

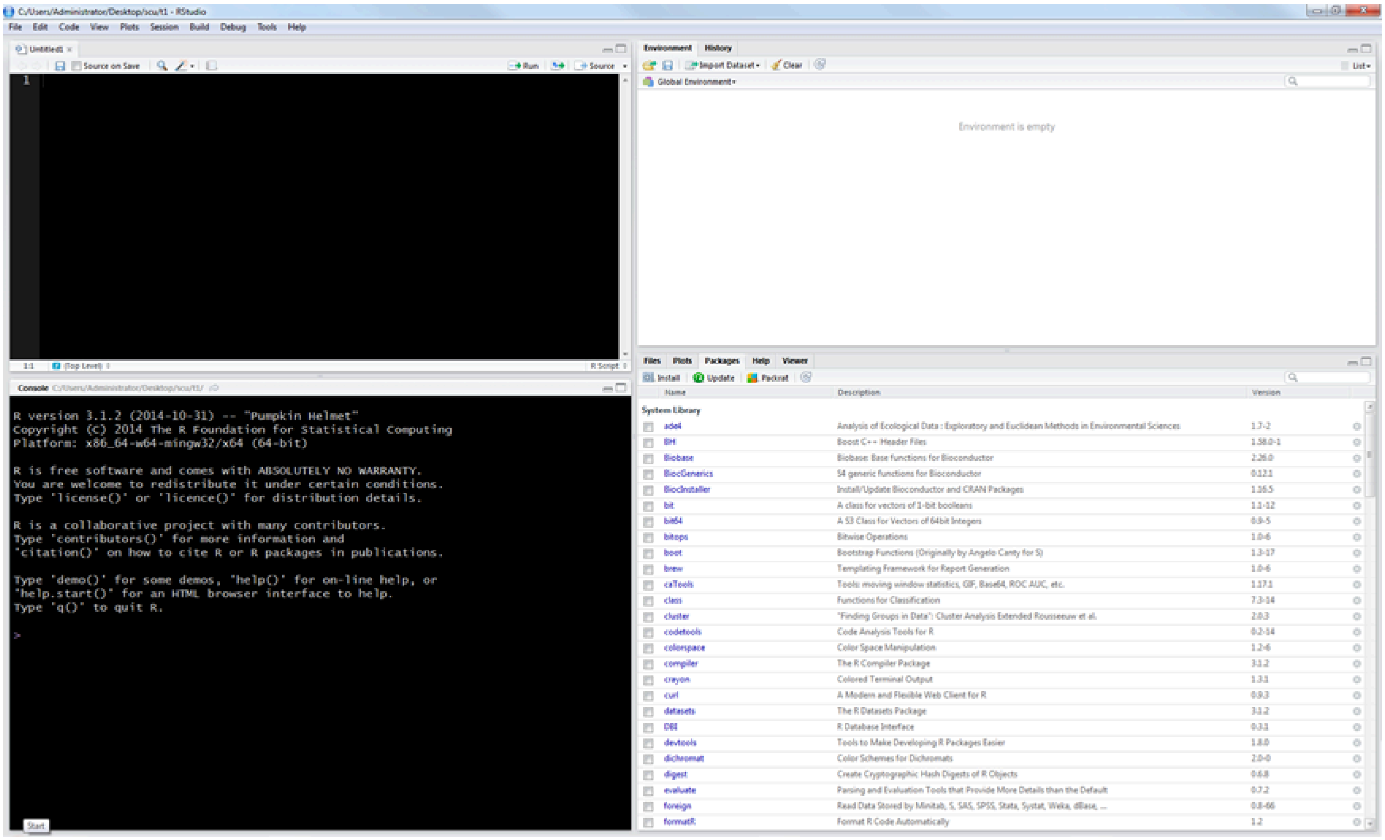
###[in some cases, some packages require dependencies ‘other prior packages to be installed, please read the error messages if its the case’]

===

R studio interface

===

###When you open R studio, you will see this panel


C:/Users/Administrator/Desktop/scu/t1 - RStudio
File Edit Code View Plots Session Build Debug Tools Help
Untitled1
Source on Save
Run
Source
1:1
(Top Level)
R Script
Console C:/Users/Administrator/Desktop/scu/t1/
Environment History
Import Dataset
Clear
List
Global Environment
Environment is empty
Files Plots Packages Help Viewer
Install Update Packrat
Start

```
R version 3.1.2 (2014-10-31) -- "Pumpkin Helmet"
Copyright (C) 2014 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.

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.

>
```

	Name	Description	Version
	System Library		
□	ade4	Analysis of Ecological Data : Exploratory and Euclidean Methods in Environmental Sciences	1.7-2
□	BH	Boost C++ Header Files	1.58.0-1
□	Biobase	Biobase: Base functions for Bioconductor	2.26.0
□	BiocGenerics	S4 generic functions for Bioconductor	0.12.1
□	BiocInstaller	Install/Update Bioconductor and CRAN Packages	1.16.5
□	bit	A class for vectors of 1-bit booleans	1.1-12
□	bit64	A S3 Class for Vectors of 64bit Integers	0.9-5
□	bitops	Bitwise Operations	1.0-6
□	boot	Bootstrap Functions (Originally by Angelo Canty for S)	1.3-17
□	brew	Templating Framework for Report Generation	1.0-6
□	caTools	Tools: moving window statistics, GIF, Base64, ROC AUC, etc.	1.17.1
□	class	Functions for Classification	7.3-14
□	cluster	"Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al.	2.0.3
□	codetools	Code Analysis Tools for R	0.2-14
□	colorspace	Color Space Manipulation	1.2-6
□	compiler	The R Compiler Package	3.1.2
□	crayon	Colored Terminal Output	1.3.1
□	curl	A Modern and Flexible Web Client for R	0.9.3
□	datasets	The R Datasets Package	3.1.2
□	DBI	R Database Interface	0.3.1
□	devtools	Tools to Make Developing R Packages Easier	1.8.0
□	dichromat	Color Schemes for Dichromats	2.0-0
□	digest	Create Cryptographic Hash Digests of R Objects	0.6.8
□	evaluate	Parsing and Evaluation Tools that Provide More Details than the Default	0.7.2
□	foreign	Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ...	0.8-66
□	formatR	Format R Code Automatically	1.2

===

##Basic method to load your data

===

The standard way to load your sample is to load it as comma separated files (.csv).

###There is some points you need to take care about ###1. from spreadsheet-> save your file as (.csv), use only 1 sheet in the excel workbook ###2. copy and paste your csv file in the same folder where you are working ###3. always try to use simple file names without any complex tags

here is your first code:)

```
data <- read.csv("test1.csv",header=T)
```

###lets try it !

==

The golden code

This R code accept the data in the clipboard of your PC. so you copy your data and submit the code, the code gets what in the clipboard.

Here is the code

```
copied <- read.table(file="clipboard",header=TRUE,sep="\t")
```

The R built-in function

The function file.choose() allows you to browse files in your computer

```
Copied <- read.csv(file.choose())
```

###after loading your file to the R, you can check if your data is probarly loaded by several ways

1. type ls() in the console

2. check global environment

##Now, you can start !

===

###In general, and as routine work, i would like you to always save your codes in a script file. the extension is usally .R to do that

###1. File -> new file-> **new R script** # i recommend to create a script store file for your own use. ###2. type your code, select it (or stop the cursor on the line) and press run ###3. the code will be moved to the console and evaluated by R lets try this on a file you loaded to R

1. open spreadsheet and create a table from 3 column named C, A, B

2. add 10 random values in each column

3. save the sheet as .csv (save as-> CSV comma delimited)

4. copy and paste the sheet in the folder you are working in in R (extension written below console) [most users forget this!]

5. Load your file into R (copy and paste the code in the R script then open the spreed sheet, marke the table, copy it, then push run in R studio "top right of your script window")

6. Ok, lets check if its loaded approbiatly or not

```
# ignore this part
A <- c(2,4,5,6,4,5,6,42,2)
B <- c(34,5,6,4,3,5,98,5,3)
C <- c(43,23,32,23,52,23,54,23,54)
copied <- data.frame(A,B,C)
copied
```

```
##    A  B  C
## 1  2 34 43
## 2  4  5 23
## 3  5  6 32
## 4  6  4 23
## 5  4  3 52
## 6  5  5 23
## 7  6 98 54
## 8 42  5 23
## 9  2  3 54
```

```
str(copied)
```

```
## 'data.frame':    9 obs. of  3 variables:
##  $ A: num  2 4 5 6 4 5 6 42 2
##  $ B: num  34 5 6 4 3 5 98 5 3
##  $ C: num  43 23 32 23 52 23 54 23 54
```

```
class(copied)
```

```
## [1] "data.frame"
```

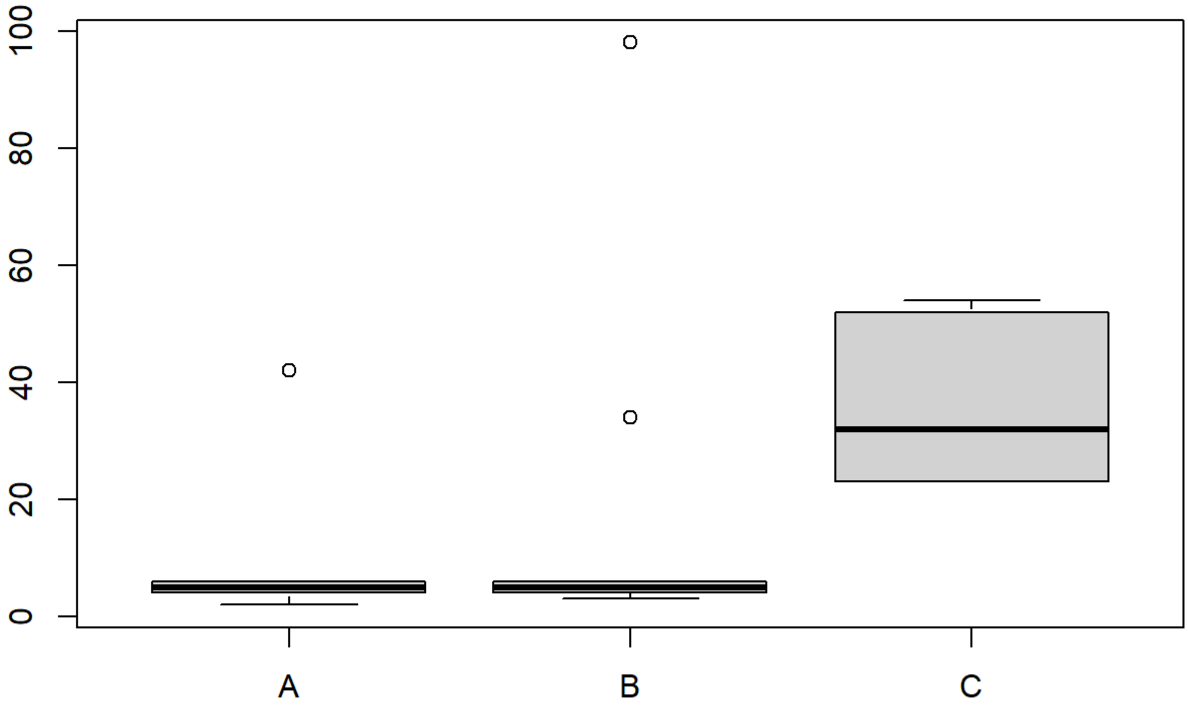
Lets see summary statistics for your data by passing this code

```
summary(copied)
```

```
##        A                B               C        
##  Min.   : 2.000   Min.   : 3.00   Min.   :23.00  
##  1st Qu.: 4.000   1st Qu.: 4.00   1st Qu.:23.00  
##  Median : 5.000   Median : 5.00   Median :32.00  
##  Mean   : 8.444   Mean   :18.11   Mean   :36.33  
##  3rd Qu.: 6.000   3rd Qu.: 6.00   3rd Qu.:52.00  
##  Max.   :42.000   Max.   :98.00   Max.   :54.00
```

after loading your data lets draw a boxplot by passing this simple code

```
boxplot(copied)
```



The help system

R includes a help system. for example if you are using a function like summary and you want to know its options

type

?summary

help(summary)

or ??summary

Congratulations!