

# Short Introduction to R

Chao-Lung Yang, Ph.D.  
Department of Industrial Management  
National Taiwan University of Science and Technology

# Notes

- The following materials are borrowed and modified based on slides and examples from
- 1. Dr. Hung Chen  
<http://www.math.ntu.edu.tw/~hchen/Prediction/notes/R-programming.ppt>
- 2. Dr. Henry Horng-Shing Lu  
[http://www.stat.nctu.edu.tw/~misg/hslu/course/statistics/An\\_Introduction\\_of\\_R.pdf](http://www.stat.nctu.edu.tw/~misg/hslu/course/statistics/An_Introduction_of_R.pdf)
- 3. Oscar Torres-Reyna  
[http://dss.princeton.edu/training/\*\*RStudio101.pdf\*\*](http://dss.princeton.edu/training/RStudio101.pdf)

# R and S-Plus

- S: an interactive environment for data analysis developed at Bell Laboratories since 1976
  - 1988 - S2: RA Becker, JM Chambers, A Wilks
  - 1992 - S3: JM Chambers, TJ Hastie
  - 1998 - S4: JM Chambers
- Exclusively licensed by *AT&T/Lucent* to *Insightful Corporation*, Seattle WA. Product name: “S-plus”.
- Implementation languages C, Fortran.
- See:  
<http://cm.bell-labs.com/cm/ms/departments/sia/S/history.html>
- R: initially written by Ross Ihaka and Robert Gentleman at Dep. of Statistics of U of Auckland, New Zealand during 1990s.
- Since 1997: international “R-core” team of ca. 15 people with access to common CVS archive.

# R

- R is “GNU S” — A language and environment for data manipulation, calculation and graphical display.
  - R is similar to the award-winning S system, which was developed at Bell Laboratories by John Chambers et al.
  - a suite of operators for calculations on arrays, in particular matrices,
  - a large, coherent, integrated collection of intermediate tools for interactive data analysis,
  - graphical facilities for data analysis and display either directly at the computer or on hardcopy
  - a well developed programming language which includes conditionals, loops, user defined recursive functions and input and output facilities.

# What R Does and Does Not

- data handling and storage: numeric, textual
- matrix algebra
- hash tables and regular expressions
- high-level data analytic and statistical functions
- classes (“OO”)
- graphics
- programming language: loops, branching, subroutines
- is not a database, but connects to DBMSs
- has no graphical user interfaces, but connects to Java, TclTk
- language interpreter can be very slow, but allows to call own C/C++ code
- no spreadsheet view of data, but connects to Excel/MsOffice
- no professional / commercial support

# R and Statistics

- Packaging: a crucial infrastructure to efficiently produce, load and keep consistent software libraries from (many) different sources / authors
- Statistics: most packages deal with statistics and data analysis
- State of the art: many statistical researchers provide their methods as R packages

# Data Analysis and Presentation

- The R distribution contains functionality for large number of statistical procedures.
  - linear and generalized linear models
  - nonlinear regression models
  - time series analysis
  - classical parametric and nonparametric tests
  - clustering
  - smoothing
- R also has a large set of functions which provide a flexible graphical environment for creating various kinds of data presentations.

# Installation of R



# Installation of R

<http://www.r-project.org/>

The R Project for Statistical Computing - Microsoft Internet Explorer


檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

網址(D) <http://www.r-project.org/>

Google G R

開始 2 已擱截 拼字檢查 翻譯

## The R Project for Statistical Computing



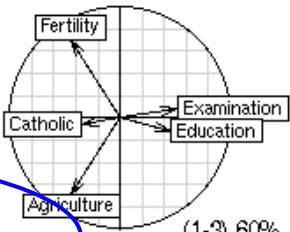
About R  
[What is R?](#)  
[Contributors](#)  
[Screenshots](#)  
[What's new?](#)

Download **CRAN**

Step 1: Click here

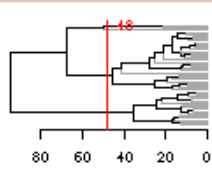
R Project  
[Foundation](#)  
[Members & Donors](#)  
[Mailing Lists](#)  
[Bug Tracking](#)  
[Developer Page](#)  
[Conferences](#)  
[Search](#)

PCA 5 vars  
princomp(x = data, cor = cor)

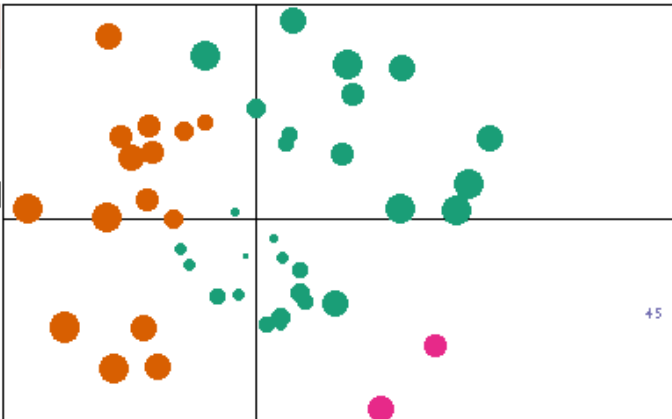


(1-3) 60%


Clustering 4 groups



Factor 1 [41%]



Factor 3 [19%]



C.-L. Yang, NTUST IM

網際網路

# Selection a Mirror Site

The screenshot shows the Microsoft Internet Explorer browser window titled "The R Project for Statistical Computing - Microsoft Internet Explorer". The address bar shows "http://www.r-project.org/". The page content includes the R logo, a left sidebar with links like "About R", "What is R?", "Contributors", "Screenshots", "What's new?", "Download CRAN", "R Project Foundation", "Members & Donors", "Mailing Lists", "Bug Tracking", "Developer Page", "Conferences", and "Search". The main content area lists mirror sites by country:

Country	URL	Host
Spain	<a href="http://cran.es.r-project.org/">http://cran.es.r-project.org/</a>	Spanish National Research Network, Madrid
Sweden	<a href="http://ftp.sunet.se/pub/lang/CRAN/">http://ftp.sunet.se/pub/lang/CRAN/</a>	Swedish University Computer Network, Uppsala
Switzerland	<a href="http://cran.ch.r-project.org/">http://cran.ch.r-project.org/</a>	ETH Zuerich
	<a href="http://www.imsv.unibe.ch/cran/">http://www.imsv.unibe.ch/cran/</a>	Universitaet Bern
	<a href="http://cran.prokmu.com/">http://cran.prokmu.com/</a>	Prokmu Hosting, Bern
Turkey	<a href="http://godel.cs.bilgi.edu.tr/mirror/cran/">http://godel.cs.bilgi.edu.tr/mirror/cran/</a>	Istanbul Bilgi University
Taiwan	<a href="http://cran.cs.pu.edu.tw/">http://cran.cs.pu.edu.tw/</a>	Providence University, Taichung
	<a href="http://cran.csie.ntu.edu.tw/">http://cran.csie.ntu.edu.tw/</a>	National Taiwan University, Taipei
UK	<a href="http://cran.uk.r-project.org/">http://cran.uk.r-project.org/</a>	University of Bristol
	<a href="http://www.sourcekeg.co.uk/cran/">http://www.sourcekeg.co.uk/cran/</a>	Sourcekeg, London
USA	<a href="http://cran.cnr.Berkeley.edu">http://cran.cnr.Berkeley.edu</a>	University of California, Berkeley, CA
	<a href="http://cran.stat.ucla.edu/">http://cran.stat.ucla.edu/</a>	University of California, Los Angeles, CA
	<a href="http://cran.ssds.ucdavis.edu/">http://cran.ssds.ucdavis.edu/</a>	University of California, Davis, CA
	<a href="http://rh-mirror.linux.iastate.edu/CRAN/">http://rh-mirror.linux.iastate.edu/CRAN/</a>	Iowa State University, Ames, IA
	<a href="http://www.stathv.com/cran/">http://www.stathv.com/cran/</a>	Stathv, Inc., Chicago, IL

A blue oval highlights the Taiwan section, and a text box labeled "Step 2: Selection a mirror site" points to it.

The Comprehensive R Archive Network - Microsoft Internet Explorer

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

← 上一頁 → 搜尋 我的最愛

網址(D) http://cran.csie.ntu.edu.tw/ 移至 連結 >>

Google G-R 開始 RS 書籤 2 已擱截 ABC 拼字檢查 設定

# The Comprehensive R Archive Network

## Frequently used pages

CRAN

- [Mirrors](#)
- [What's new?](#)
- [Task Views](#)
- [Search](#)

About R

- [R Homepage](#)

Software

- [R Sources](#)
- [R Binaries](#)
- [Packages](#)
- [Other](#)

Documentation

- [Manuals](#)
- [FAQs](#)

### Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Linux](#)
- [MacOS X](#)
- [Windows \(95 and later\)](#)

### Step 3: Selection OS for R

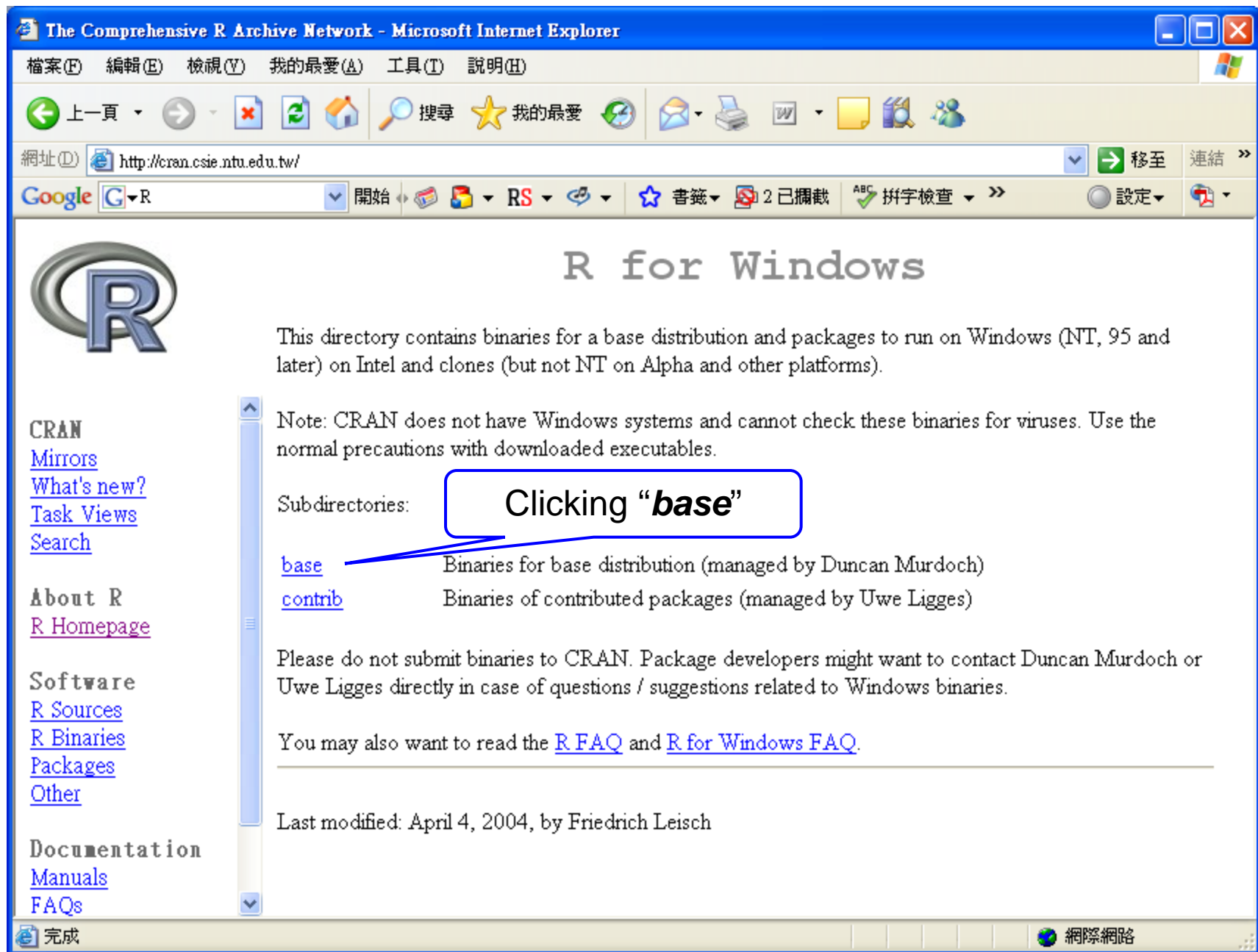
**e.g., Windows**

### Source Code for all Platforms

Windows and Mac users most likely want the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- The latest release** (2006-12-18): [R-2.4.1.tar.gz](#) (read [what's new](#) in the latest version).

網際網路



The Comprehensive R Archive Network - Microsoft Internet Explorer

檔案(F) 編輯(E) 檢視(V) 我的最愛(A) 工具(T) 說明(H)

← 上一頁 → 搜尋 我的最愛

網址(D) http://cran.csie.ntu.edu.tw/ Google R

移至 連結 >> 設定 >>

**檔案下載 - 安全性警告**

是否要執行或儲存這個檔案?

名稱: R-2.4.1-win32.exe  
類型: 應用程式, 28.0 MB  
來自: cran.csie.ntu.edu.tw

執行(R) 儲存(S) 取消

雖然來自網際網路的檔案可能是有用的, 但是這個檔案類型有可能會傷害您的電腦。如果您不信任其來源, 請不要執行或儲存這個軟體。有什麼樣的風險?

CRAN  
[Mirrors](#)  
[What's new?](#)  
[Task Views](#)  
[Search](#)

About R  
[R Homepage](#)

Software  
[R Sources](#)  
[R Binaries](#)  
[Packages](#)  
[Other](#)

Documentation  
[Manuals](#)  
[FAQs](#)

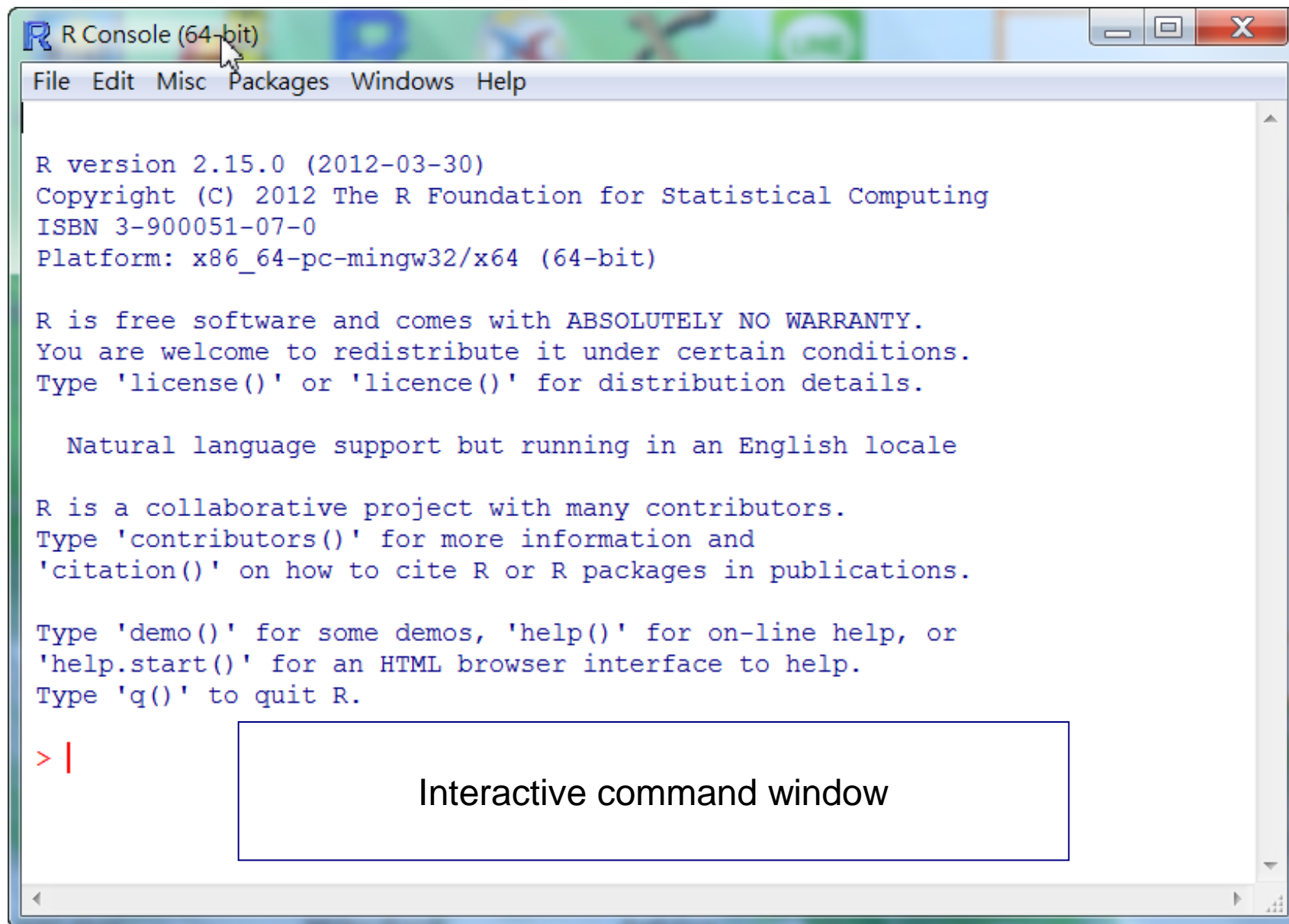
This directory contains the source code for R, as well as the binaries for Windows, Mac OS X, and Linux. A build of the development version (which will eventually become R 2.6.0 in the fall) is available in the [R-devel snapshot build](#).

In this directory:

- [README.R-2.4.1](#) Installation and other instructions.
- [CHANGES](#) New features of this Windows version.
- [NEWS](#) New features of all versions.
- [R-2.4.1-win32.exe](#) Setup program (about 28 megabytes). Please download this from a [mirror near you](#). This corresponds to the file named **SetupR.exe** or **rwXXXX.exe** in pre-2.2.0 releases.

Clicking

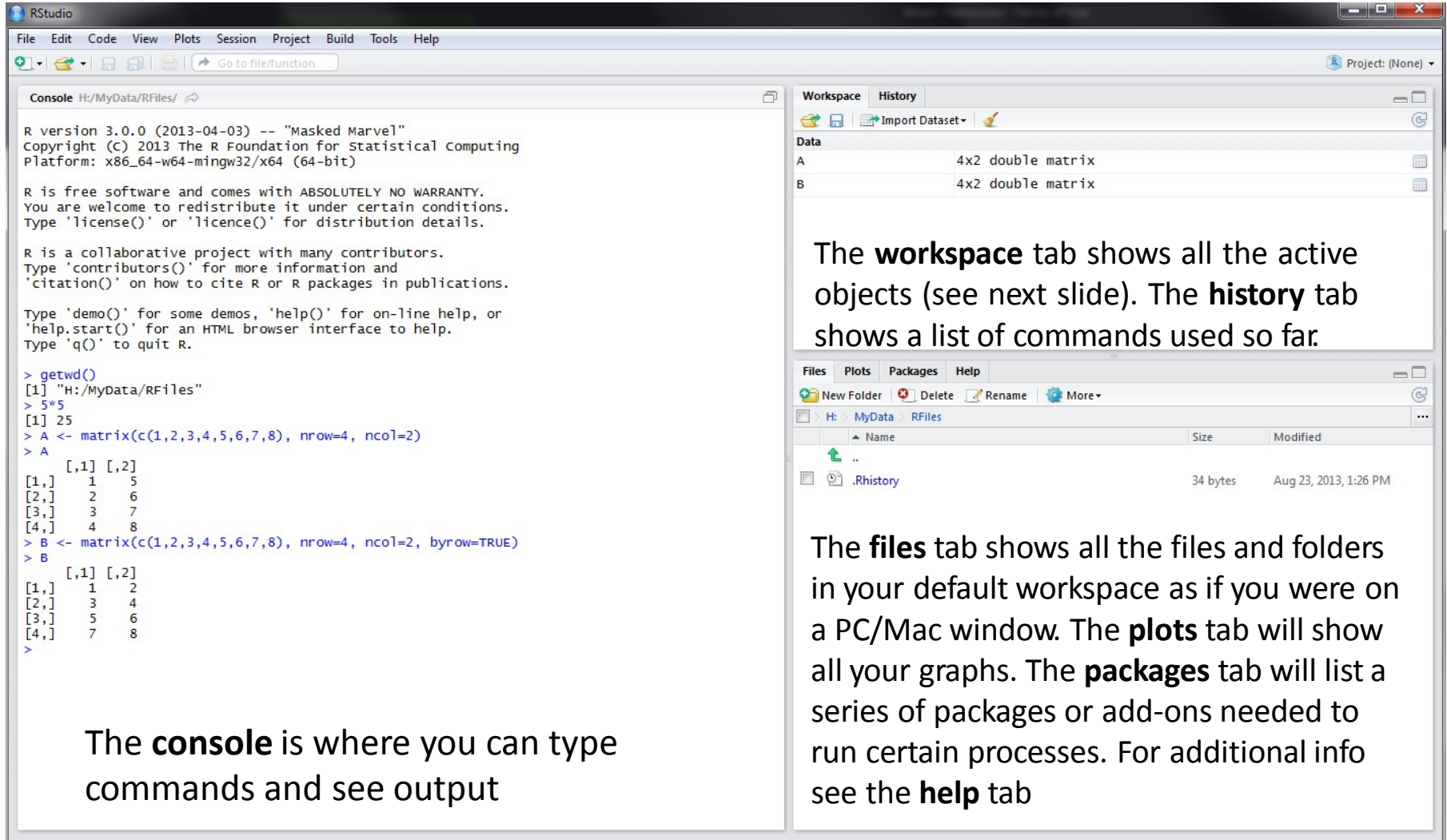
網際網路



# RStudio

- RStudio allows the user to run R in a more user-friendly environment. It is open-source (i.e. free) and available at <http://www.rstudio.com/>

# RStudio screen



The screenshot displays the RStudio application window. The top menu bar includes File, Edit, Code, View, Plots, Session, Project, Build, Tools, and Help. Below the menu is a toolbar with icons for file operations and a search bar labeled 'Go to file/function'. The main interface is divided into four panes: Console, Workspace, History, and Files. The Console pane on the left shows the R version (3.0.0), copyright information, and the results of several R commands, including creating two 4x2 matrices, A and B. The Workspace pane on the right shows the active objects A and B, both 4x2 double matrices. The History pane shows a list of commands used. The Files pane at the bottom shows the file explorer with a tree view of the workspace contents, including a folder named .Rhistory.

**Console** H:/MyData/RFiles/ ↗

```
R version 3.0.0 (2013-04-03) -- "Masked Marvel"
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.

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.

> getwd()
[1] "H:/MyData/RFiles"
> 5*5
[1] 25
> A <- matrix(c(1,2,3,4,5,6,7,8), nrow=4, ncol=2)
> A
      [,1] [,2]
[1,]    1    5
[2,]    2    6
[3,]    3    7
[4,]    4    8
> B <- matrix(c(1,2,3,4,5,6,7,8), nrow=4, ncol=2, byrow=TRUE)
> B
      [,1] [,2]
[1,]    1    2
[2,]    3    4
[3,]    5    6
[4,]    7    8
>
```

The **workspace** tab shows all the active objects (see next slide). The **history** tab shows a list of commands used so far.

**Files** | **Plots** | **Packages** | **Help**

New Folder | Delete | Rename | More ▾

H: > MyData > RFiles

Name	Size	Modified
..		
.Rhistory	34 bytes	Aug 23, 2013, 1:26 PM

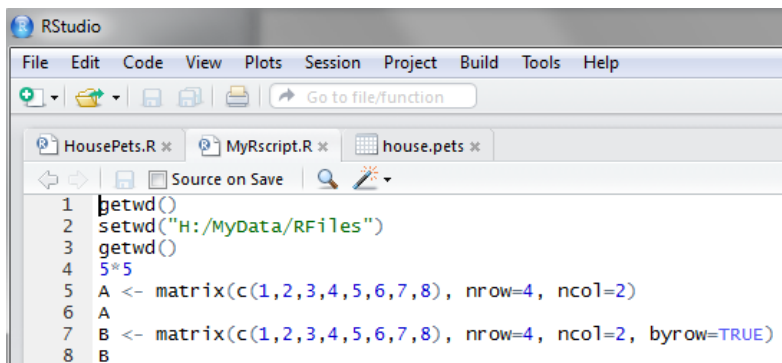
The **files** tab shows all the files and folders in your default workspace as if you were on a PC/Mac window. The **plots** tab will show all your graphs. The **packages** tab will list a series of packages or add-ons needed to run certain processes. For additional info see the **help** tab

The **console** is where you can type commands and see output

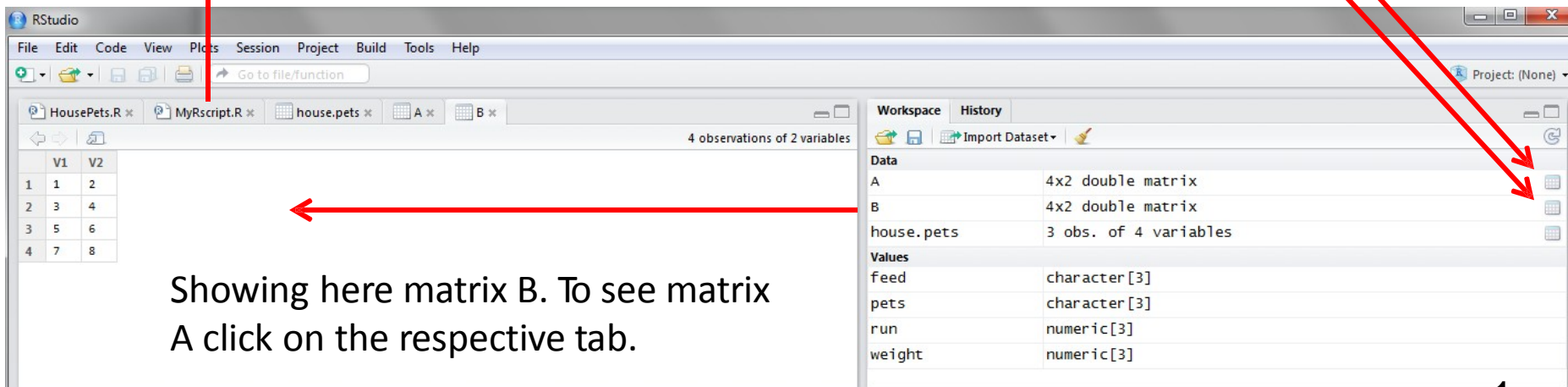


# Workspace tab (1)

The workspace tab stores any object, value, function or anything you create during your R session. In the example below, if you click on the dotted squares you can see the data on a screen to the left.



```
1 betwd()
2 setwd("H:/MyData/RFiles")
3 getwd()
4 5*5
5 A <- matrix(c(1,2,3,4,5,6,7,8), nrow=4, ncol=2)
6 A
7 B <- matrix(c(1,2,3,4,5,6,7,8), nrow=4, ncol=2, byrow=TRUE)
8 B
```



4 observations of 2 variables

	V1	V2
1	1	2
2	3	4
3	5	6
4	7	8

Showing here matrix B. To see matrix A click on the respective tab.

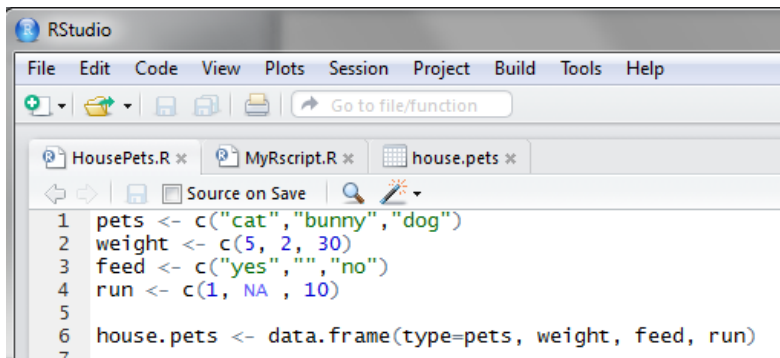
Workspace History

Data	
A	4x2 double matrix
B	4x2 double matrix
house.pets	3 obs. of 4 variables

Values	
feed	character[3]
pets	character[3]
run	numeric[3]
weight	numeric[3]

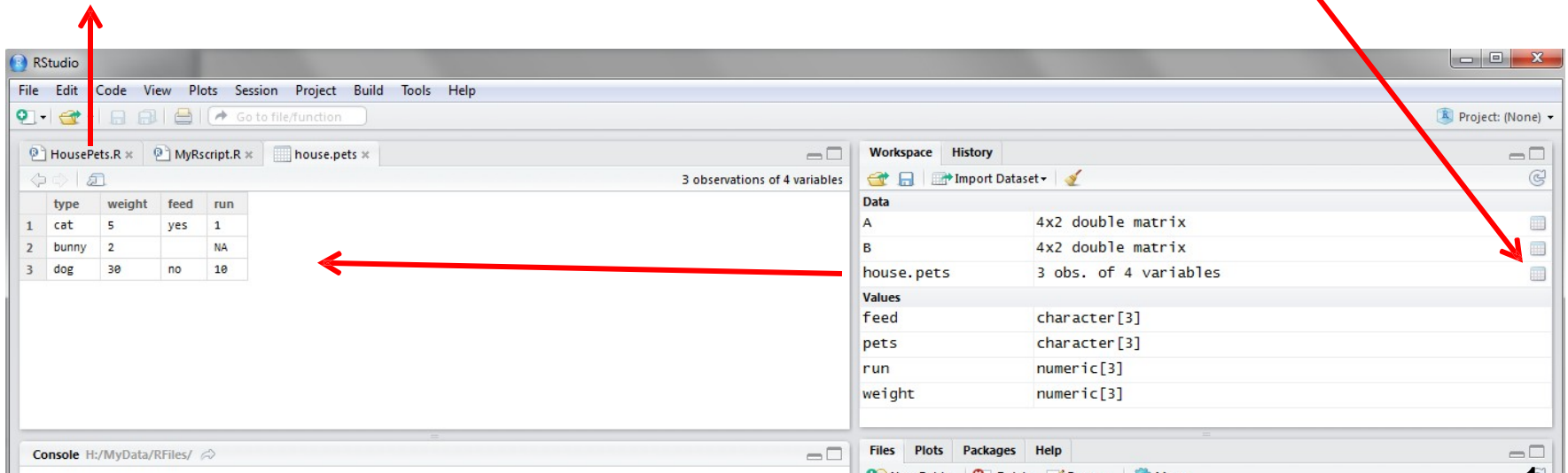
# Workspace tab (2)

Here is another example on how the workspace looks like when more objects are added. Notice that the data frame `house.pets` is formed from different individual values or vectors.



```
1 pets <- c("cat", "bunny", "dog")
2 weight <- c(5, 2, 30)
3 feed <- c("yes", "", "no")
4 run <- c(1, NA, 10)
5
6 house.pets <- data.frame(type=pets, weight, feed, run)
7
```

Click on the dotted square to look at the dataset in a spreadsheet form.



The screenshot shows the RStudio interface with the `house.pets` data frame loaded. The `Source` tab is active, showing the code from the previous block. The `Workspace` tab is also visible, showing the objects in the workspace. The `house.pets` data frame is highlighted, and its structure is shown in the `Values` pane. A red arrow points from the text 'Click on the dotted square to look at the dataset in a spreadsheet form.' to the dotted square icon in the `Workspace` pane. Another red arrow points from the `house.pets` data frame in the `Workspace` pane to the spreadsheet view of the data frame.

	type	weight	feed	run
1	cat	5	yes	1
2	bunny	2	NA	NA
3	dog	30	no	10

3 observations of 4 variables

Workspace History

Data

- A: 4x2 double matrix
- B: 4x2 double matrix
- house.pets: 3 obs. of 4 variables

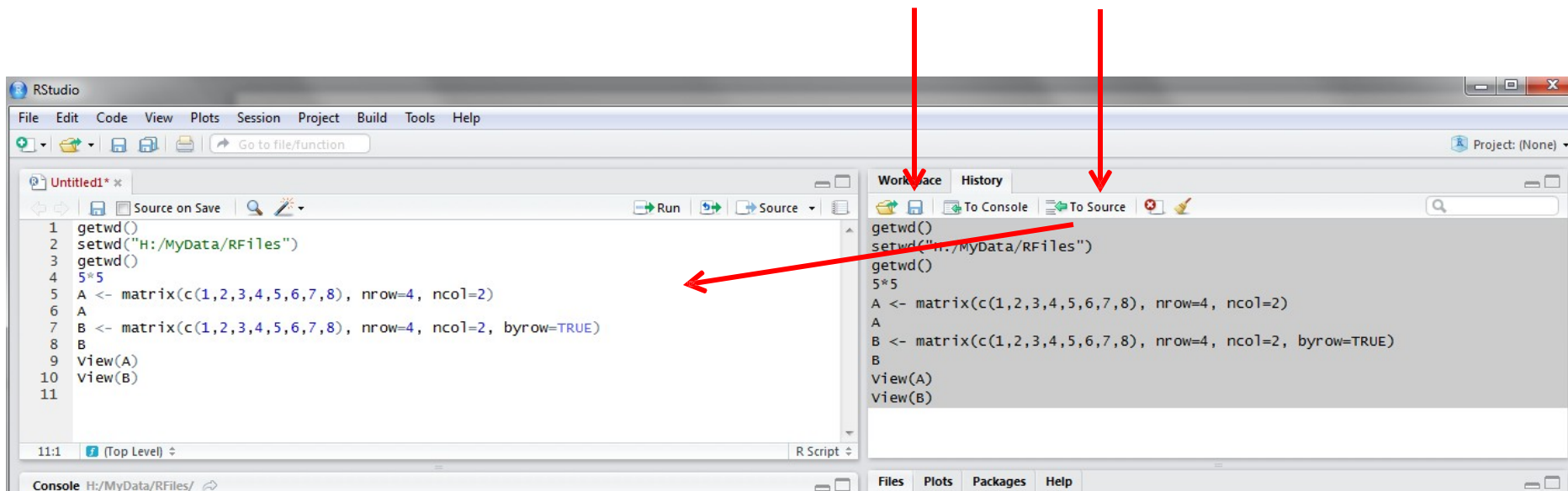
Values

- feed: character[3]
- pets: character[3]
- run: numeric[3]
- weight: numeric[3]

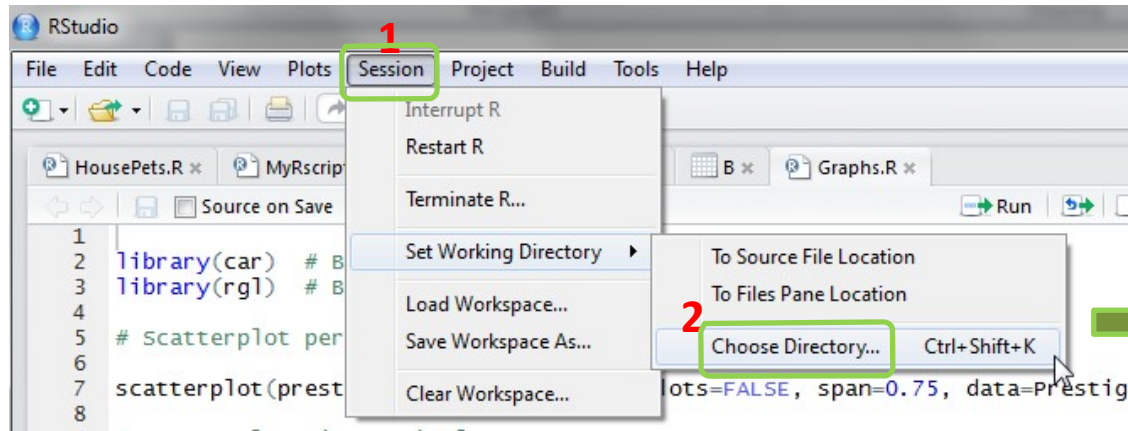
# History tab

The history tab keeps a record of all previous commands. It helps when testing and running processes. Here you can either **save** the whole list or you can **select** the commands you want and send them to an R script to keep track of your work.

In this example, we select all and click on the “To Source” icon, a window on the left will open with the list of commands. Make sure to save the ‘untitled1’ file as an \*.R script.



# Changing the working directory



If you have different projects you can change the working directory for that session, see above. Or you can type:

# Shows the working directory (wd)

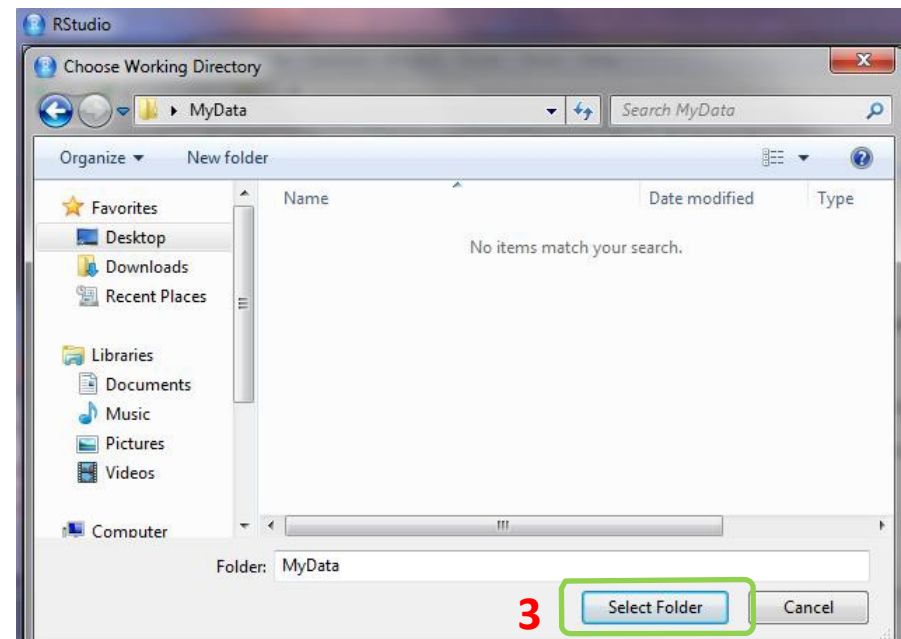
```
getwd()
```

# Changes the wd

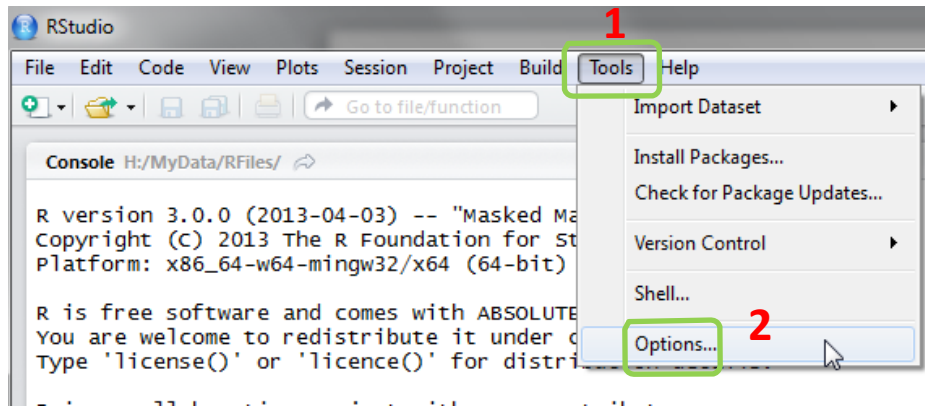
```
setwd("C:/myfolder/data")
```

More info see the following document:

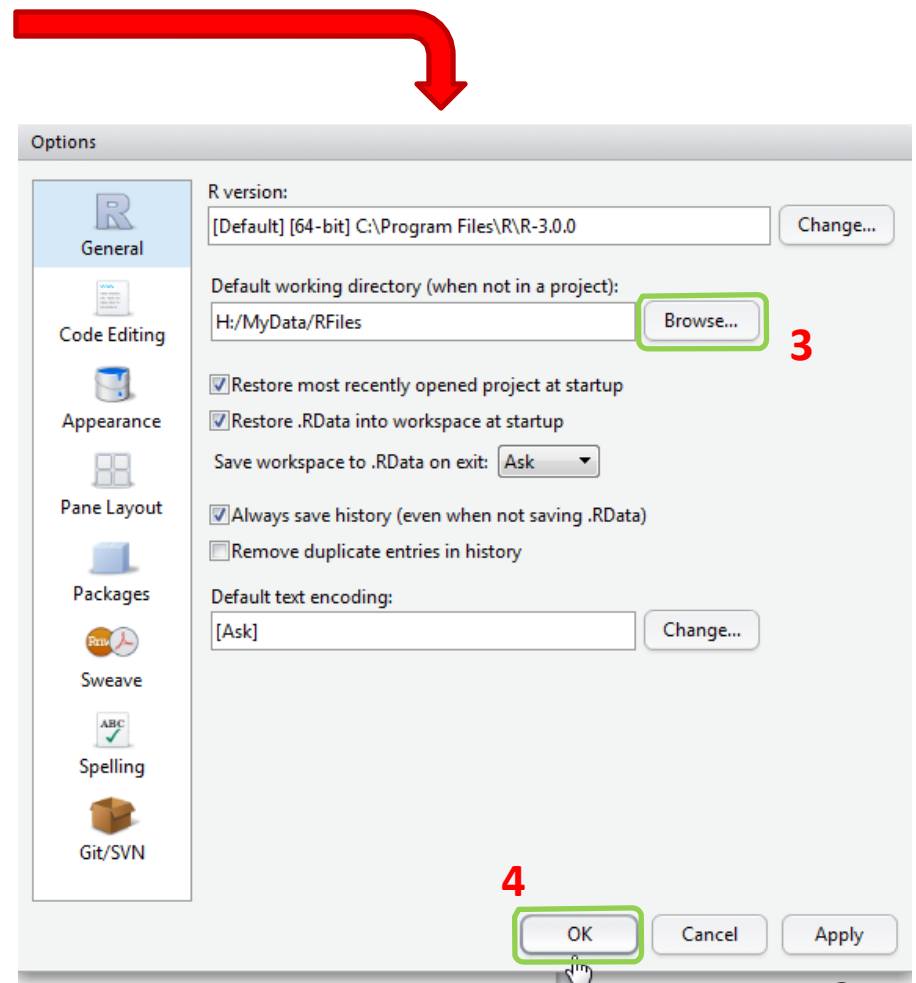
<http://dss.princeton.edu/training/RStata.pdf>



# Setting a default working directory



Every time you open RStudio, it goes to a default directory. You can change the default to a folder where you have your datafiles so you do not have to do it every time. In the menu go to Tools->Options

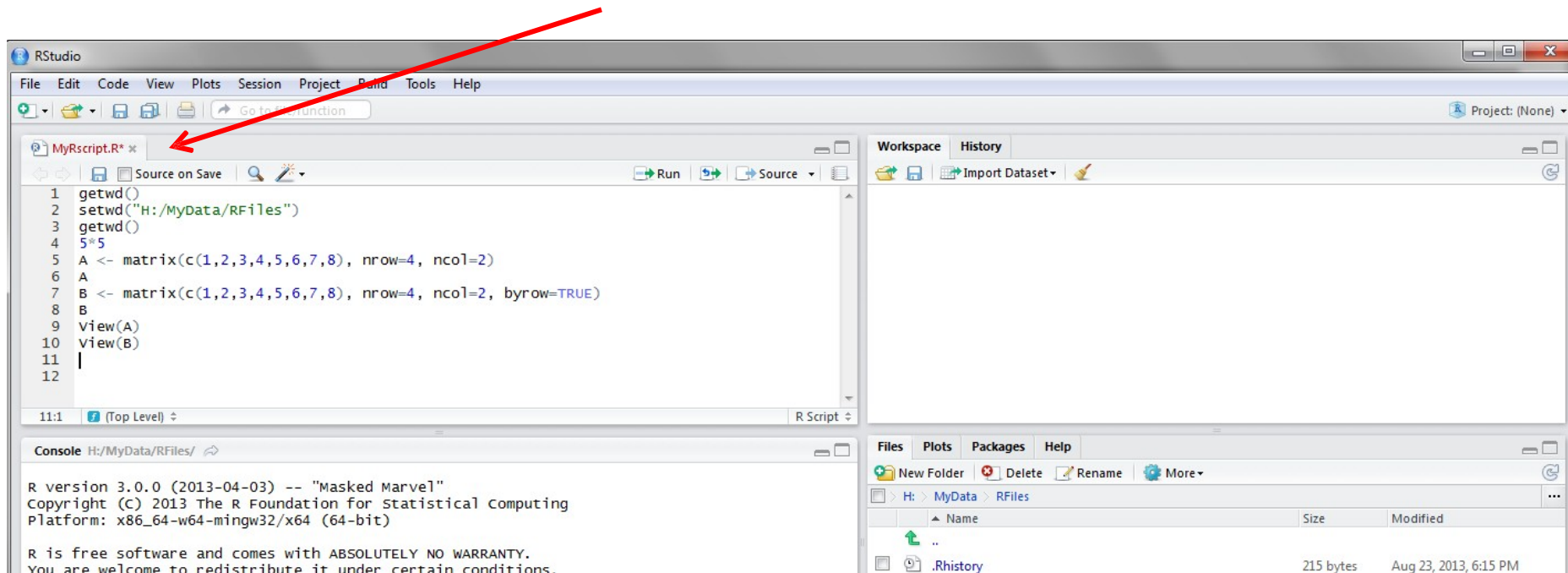


# R script (1)

The usual Rstudio screen has four windows:

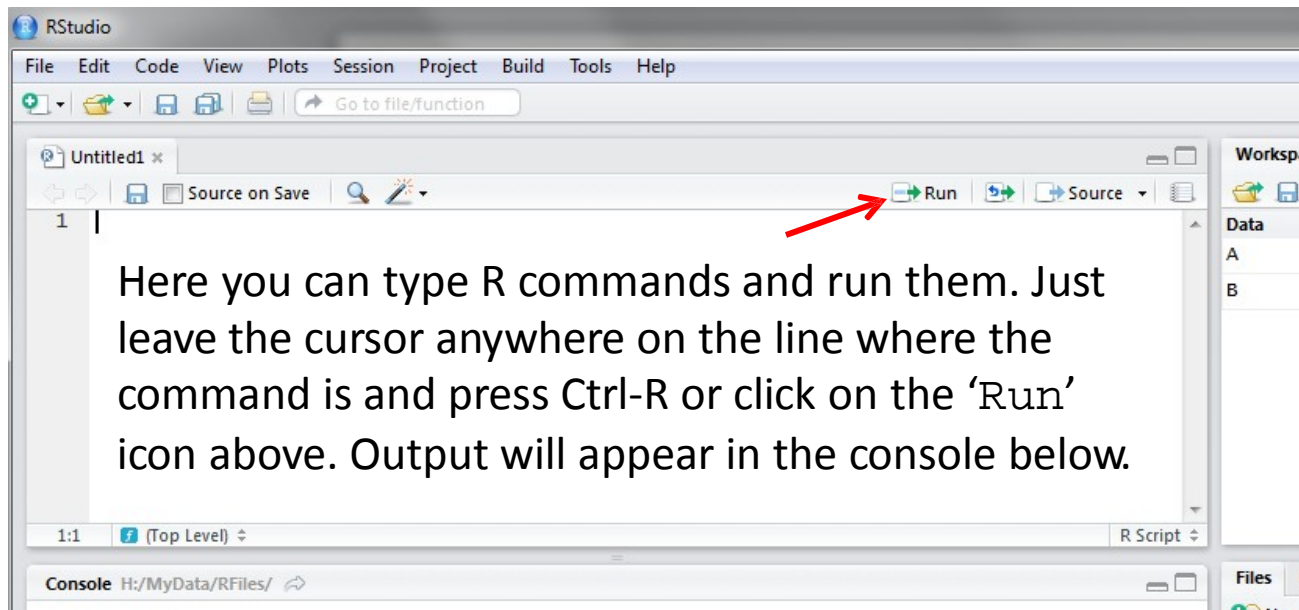
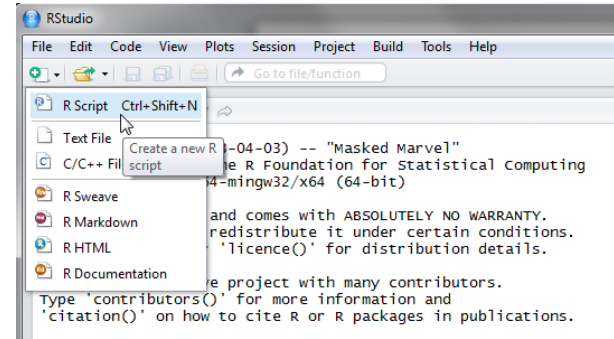
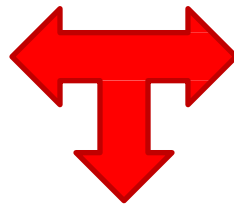
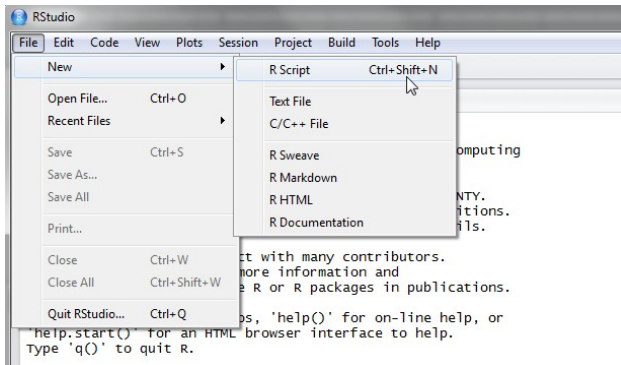
1. Console.
2. Workspace and history.
3. Files, plots, packages and help.
4. The R script(s) and data view.

The R script is where you keep a record of your work. For Stata users this would be like the do-file, for SPSS users is like the syntax and for SAS users the SAS program.



# R script (2)

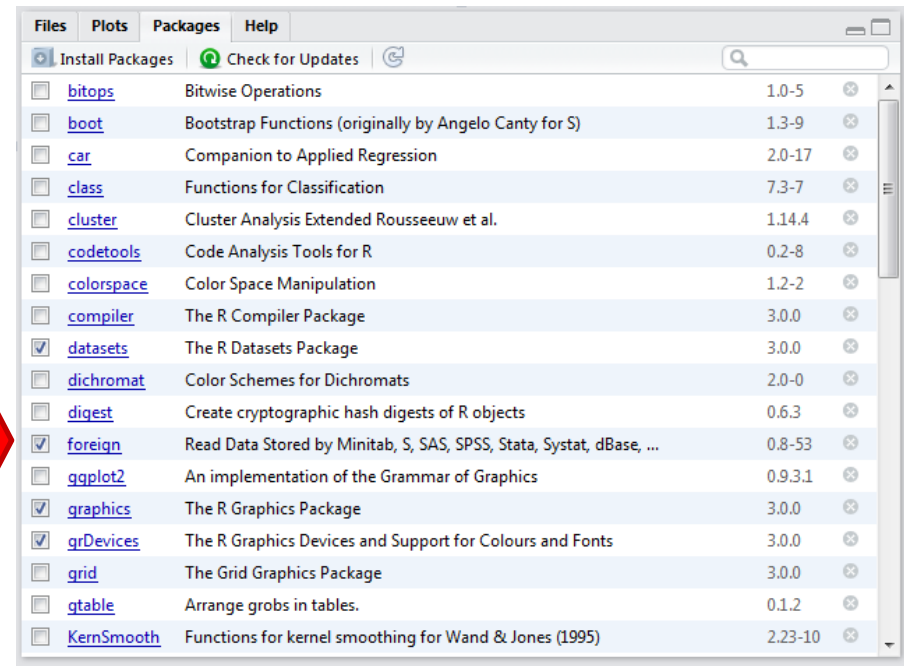
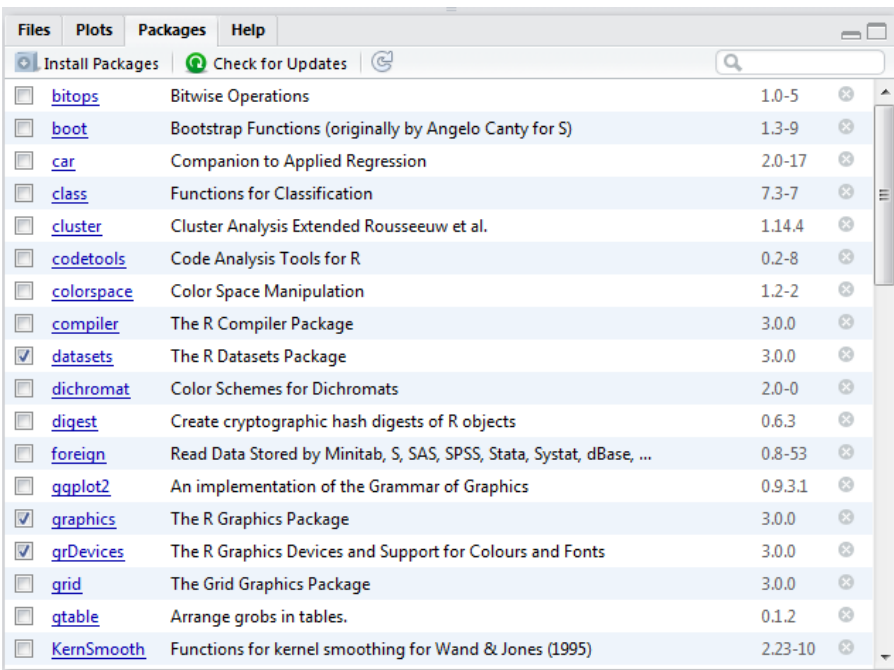
To create a new R script you can either go to **File -> New -> R Script**, or click on the icon with the “+” sign and select “R Script”, or simply press **Ctrl+Shift+N**. Make sure to save the script.





# Packages tab

The package tab shows the list of add-ons included in the installation of RStudio. If checked, the package is loaded into R, if not, any command related to that package won't work, you will need select it. You can also install other add-ons by clicking on the 'Install Packages' icon. Another way to activate a package is by typing, for example, `library(foreign)`. This will automatically check the `--foreign` package (it helps bring data from proprietary formats like Stata, SAS or SPSS).

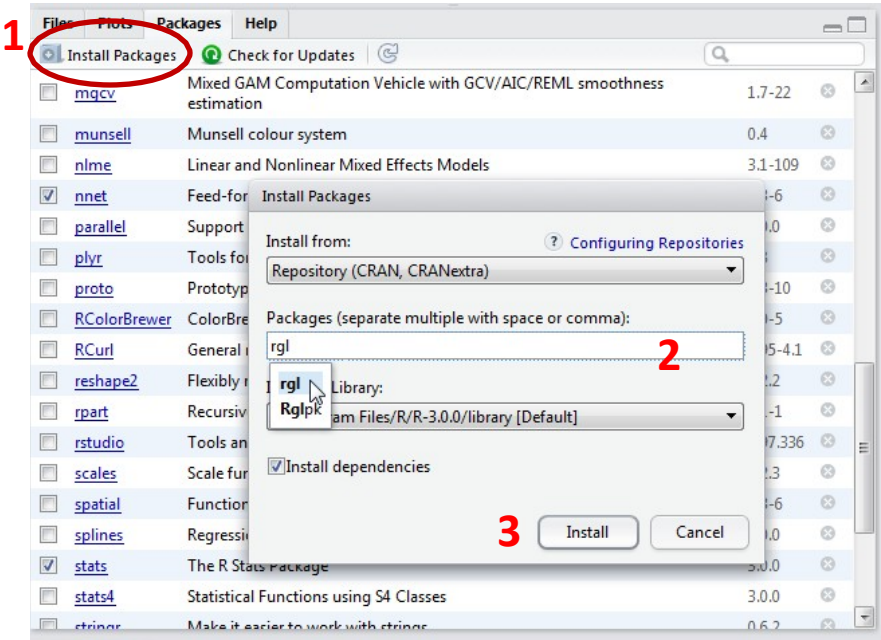




# Installing a package

<input type="checkbox"/>	<a href="#">RCurl</a>	General network (HTTP/FTP/...) client interface for R	1.95-4.1	✕
<input type="checkbox"/>	<a href="#">reshape2</a>	Flexibly reshape data: a reboot of the reshape package.	1.2.2	✕
<input type="checkbox"/>	<a href="#">rpart</a>	Recursive Partitioning	4.1-1	✕

Before



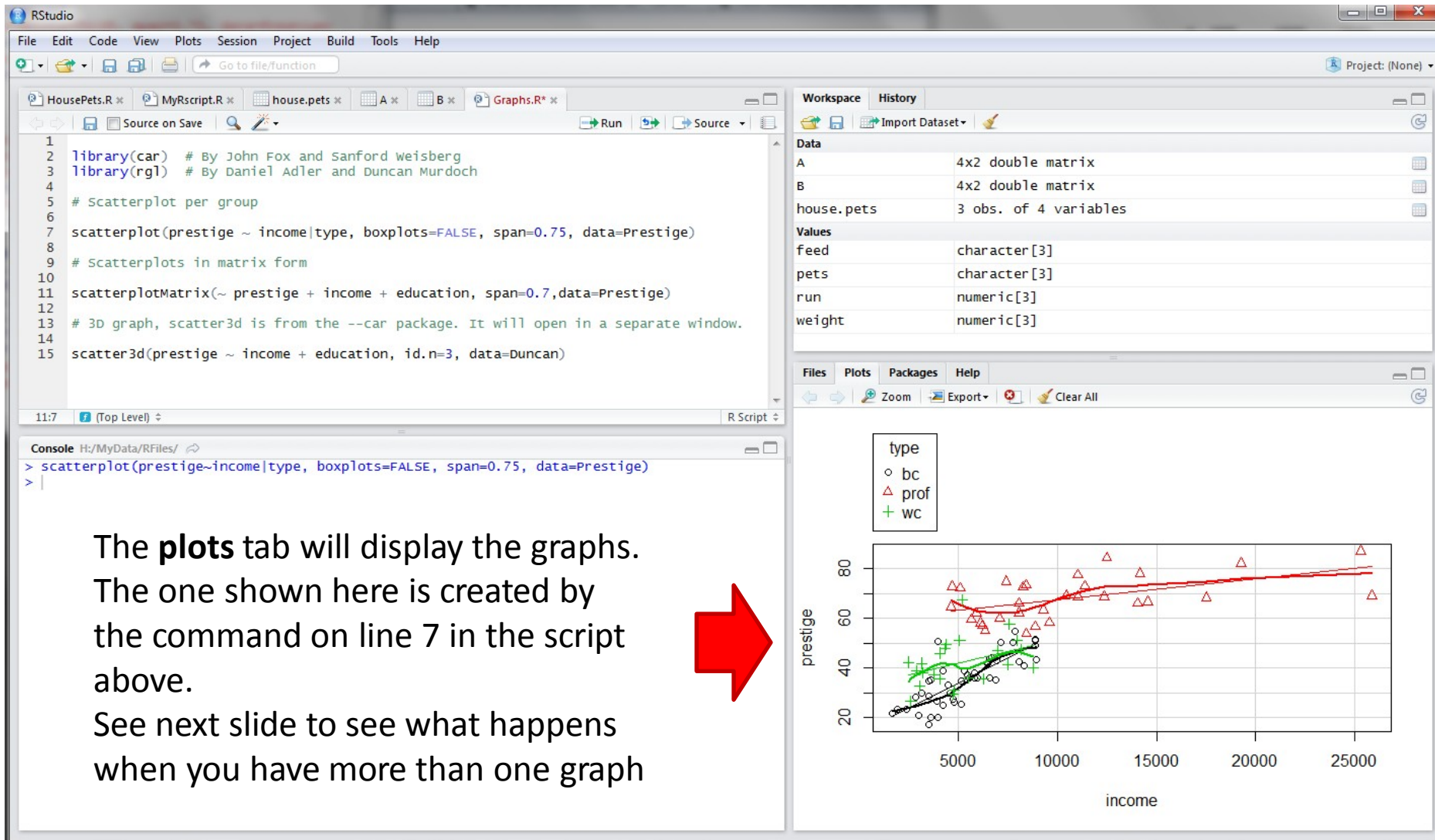
We are going to install the package – `rgl` (useful to plot 3D images). It does not come with the original R install.

Click on “Install Packages”, write the name in the pop-up window and click on “Install”.

After

<input type="checkbox"/>	<a href="#">RCurl</a>	General network (HTTP/FTP/...) client interface for R	1.95-4.1	✕
<input type="checkbox"/>	<a href="#">reshape2</a>	Flexibly reshape data: a reboot of the reshape package.	1.2.2	✕
<input type="checkbox"/>	<a href="#">rgl</a>	3D visualization device system (OpenGL)	0.93.952	✕
<input type="checkbox"/>	<a href="#">rpart</a>	Recursive Partitioning	4.1-1	✕

# Plots tab (1)



# Plots tab (2)

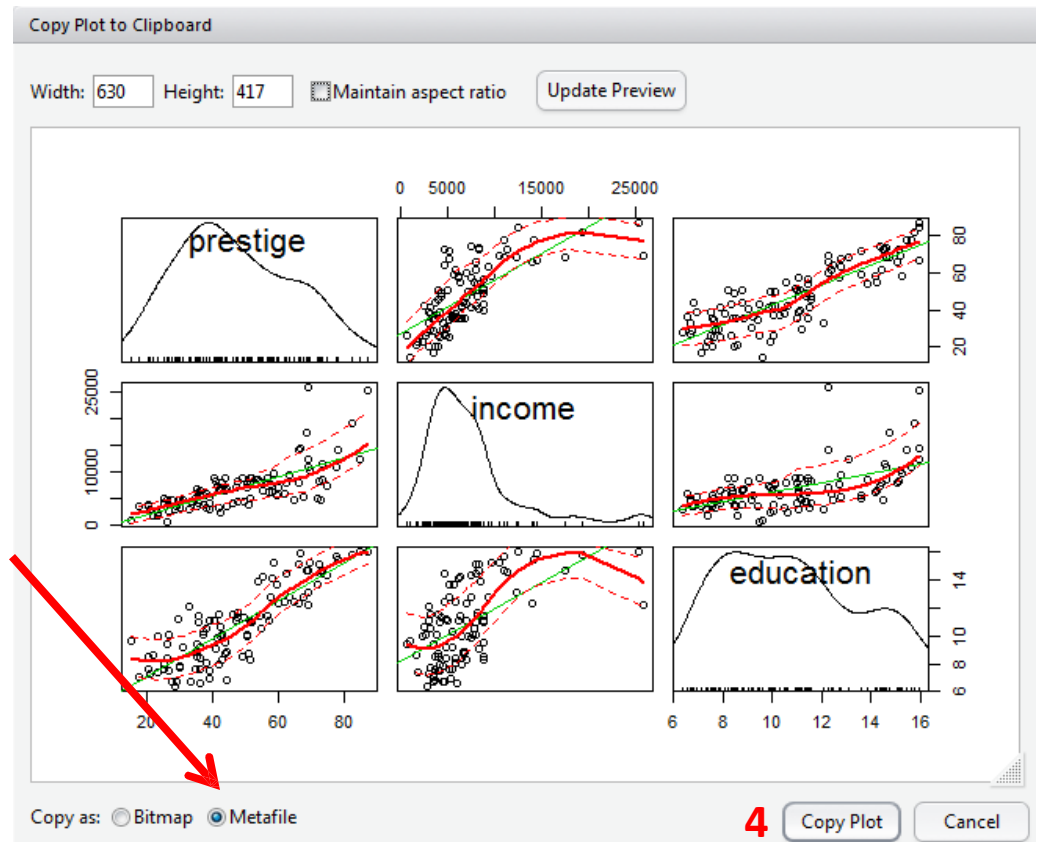
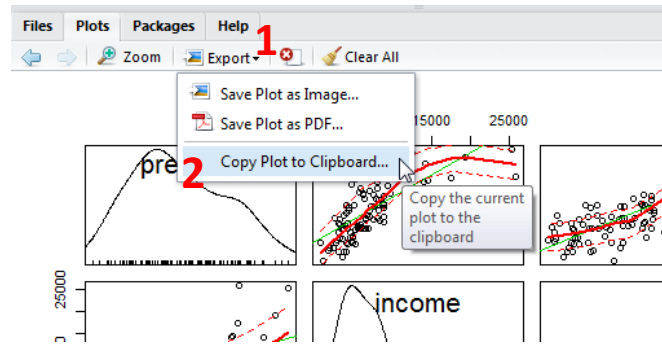
The screenshot shows the RStudio interface with the following components:

- Source Editor:** Contains R code for plotting. Line 11 is highlighted, showing `scatterplotMatrix(~ prestige + income + education, span=0.7, data=Prestige)`.
- Workspace:** Lists objects: A (4x2 double matrix), B (4x2 double matrix), house.pets (3 obs. of 4 variables), feed (character[3]), pets (character[3]), run (numeric[3]), and weight (numeric[3]).
- Plots Tab:** Displays a 3x3 grid of diagnostic plots for variables 'prestige', 'income', and 'education'. The plots include histograms on the diagonal and scatter plots with regression lines on the off-diagonals.
- Console:** Shows the execution of the plotting commands from the source editor.
- Toolbar:** Includes icons for Files, Plots, Packages, Help, Zoom, Export, and Clear All. A red arrow points to the left arrow icon.

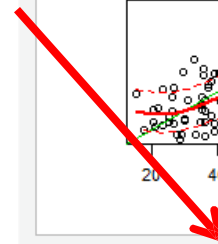
Here there is a second graph (see line 11 above). If you want to see the first one, click on the left-arrow icon.

# Plots tab (3) – Graphs export

To extract the graph, click on “Export” where you can save the file as an image (PNG, JPG, etc.) or as PDF, these options are useful when you only want to share the graph or use it in a LaTeX document. Probably, the easiest way to export a graph is by copying it to the clipboard and then paste it directly into your Word document.

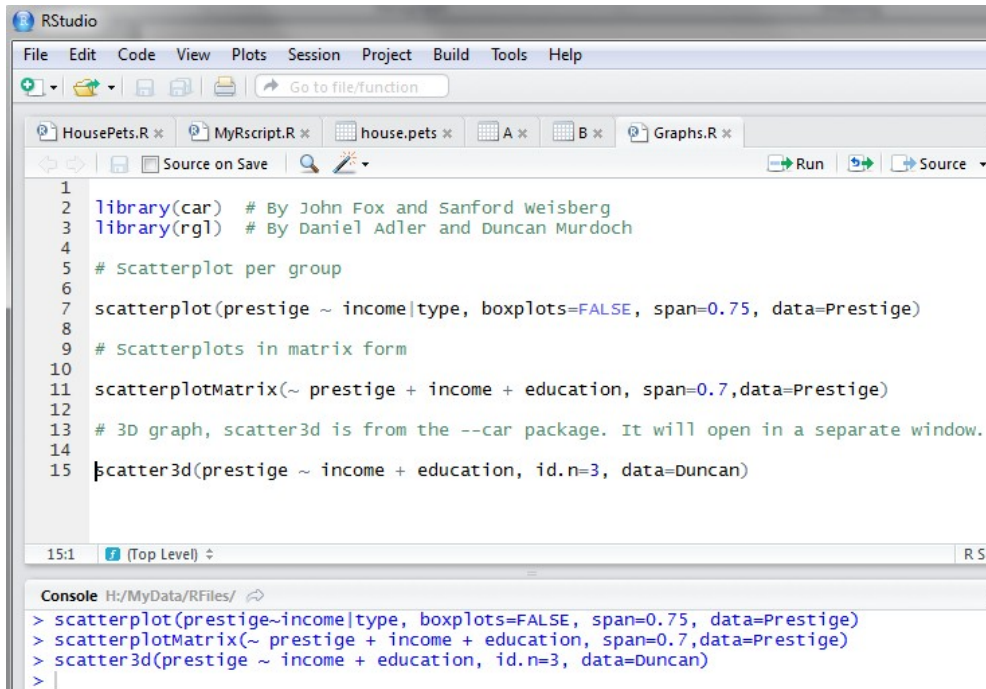


3 Make sure to select 'Metafile'



5 Paste it into your Word document

# 3D graphs

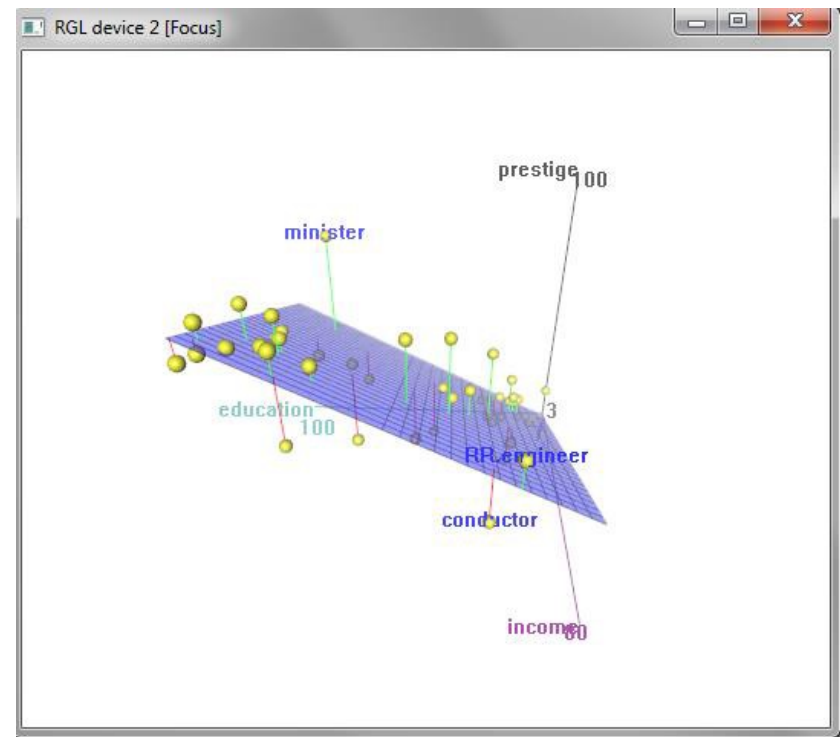


```
1 library(car) # By John Fox and Sanford Weisberg
2 library(rgl) # By Daniel Adler and Duncan Murdoch
3
4 # Scatterplot per group
5
6 scatterplot(prestige ~ income|type, boxplots=FALSE, span=0.75, data=Prestige)
7
8 # Scatterplots in matrix form
9
10 scatterplotMatrix(~ prestige + income + education, span=0.7, data=Prestige)
11
12 # 3D graph, scatter3d is from the --car package. It will open in a separate window.
13
14
15 scatter3d(prestige ~ income + education, id.n=3, data=Duncan)
```

Console H:/MyData/RFiles/

```
> scatterplot(prestige~income|type, boxplots=FALSE, span=0.75, data=Prestige)
> scatterplotMatrix(~ prestige + income + education, span=0.7, data=Prestige)
> scatter3d(prestige ~ income + education, id.n=3, data=Duncan)
```

3D graphs will display on a separate screen (see line 15 above). You won't be able to save it, but after moving it around, once you find the angle you want, you can screenshot it and paste it to you Word document.



# Basic of R Programming

# Environment Commands

> `search()` # loaded packages

```
[1] ".GlobalEnv"      "package:stats"    "package:graphics"  
[4] "package:grDevices" "package:utils"    "package:datasets"  
[7] "package:methods" "Autoloads"        "package:base"
```

> `ls()` # Used objects

```
[1] "bb"      "col"      "colorlut"  "EdgeList"  
[5] "Edges"   "EXP"      "f"
```

> `rm(bb)` # remove object "bb"

> `?lm` # ? Function name = look function

> `args(lm)` # look arguments in "lm"

> `help(lm)` # See detail of function (lm)



# Operators

- Mathematic operators: + - \* / ^
  - Mod: %%
  - sqrt, exp, log, log10, sin, cos, tan, .....
- Other operators:
  - \$ component selection HIGH
  - [ , [[ subscripts, elements
  - : sequence operator
  - %\*% matrix algebra
  - <, >, <=, >= inequality
  - ==, != comparison
  - ! not
  - &, |, &&, || and, or
  - ~ formulas
  - <- assignment (or = 1.9.1 later)



# Arithmetic Operators

Operator	Description
+	addition
-	subtraction
*	multiplication
/	division
<b>^ or **</b>	exponentiation
<b>x %% y</b>	modulus (x mod y) 5%%2 is 1
<b>x %/% y</b>	integer division 5%/2 is 2

# Logical Operators

Operator	Description
<	less than
<=	less than or equal to
>	greater than
>=	greater than or equal to
==	exactly equal to
!=	not equal to
!x	Not x
x   y	x OR y
x & y	x AND y
isTRUE(x)	test if x is TRUE

# Demo Algebra, Operators and Functions

```
> 1+2
[1] 3
> 1 > 2
[1] FALSE
> 1 > 2 | 2 > 1
[1] TRUE
> 1:3
[1] 1 2 3
> A = 1:3
> A
[1] 1 2 3
> A*6
[1] 6 12 18
> A/10
[1] 0.1 0.2 0.3
> A %% 2
[1] 1 0 1
```

```
> B=4:6
> A*B
[1] 4 10 18
> A%*%B
      [,1]
[1,]    32
> A %*% t(B)
      [,1] [,2] [,3]
[1,]    4    5    6
[2,]    8   10   12
[3,]   12   15   18
> A / B
[1] 0.25 0.40 0.50
> sqrt(A)
[1] 1.000000 1.414214 1.732051
> log(A)
[1] 0.0000000 0.6931472 1.0986123
```

```
> round(sqrt(A),2)
[1] 1.00 1.41 1.73
> ceiling(sqrt(A))
[1] 1 2 2
> floor(sqrt(A))
[1] 1 1 1
> eigen( A%*% t(B))
$values
[1] 3.200000e+01 5.835176e-16 2.480655e-16

$vectors
      [,1]      [,2]      [,3]
[1,] 0.2672612 0.3273463 -0.8890009
[2,] 0.5345225 -0.8217055 0.2540003
[3,] 0.8017837 0.4665237 0.3810004
> eigen( A%*% t(B))$values
[1] 3.200000e+01 5.835176e-16 2.480655e-16
```

# Variables

```
> a = 49  
> sqrt(a)  
[1] 7
```

numeric

```
> a = "The dog ate my homework"  
> sub("dog","cat",a)  
[1] "The cat ate my homework"
```

character  
string

```
> a = (1+1==3)  
> a  
[1] FALSE
```

logical

# Vectors, Matrices, and Arrays

- **vector**: an ordered collection of data of the same type
- `> a = c(1,2,3)`
- `> a*2`
- `[1] 2 4 6`
- In R, a single number is the special case of a vector with 1 element.
- Other vector types: character strings, logical

# Vectors, Matrices, and Arrays

- **matrix**: a rectangular table of data of the same type

```
> A = matrix(1:12,3,4)
> A
      [,1] [,2] [,3] [,4]
[1,]    1    4    7   10
[2,]    2    5    8   11
[3,]    3    6    9   12
>
```

- **array**: 3-,4-,...dimensional matrix

```
> A[1,]
[1]  1  4  7 10
>
```

# Lists

- **vector**: an ordered collection of data of the same type.

```
> a = c(7,5,1)
> a[2]
[1] 5
```

- **list**: an ordered collection of data of arbitrary types.

```
> doe = list(name="john",age=28,married=F)
> doe$name
[1] "john"
> doe$age
[1] 28
```

- Typically, vector elements are accessed by their index (an integer), list elements by their name (a character string). But both types support both access methods.

# Data Frame

- **data frame:** is supposed to represent the typical data table that researchers come up with – like a spreadsheet.
- It is a rectangular table with rows and columns; data within each column has the same type (e.g. number, text, logical), but different columns may have different types.



# Branching

```
if (logical expression) {  
    statements  
} else {  
    alternative statements  
}
```

**else** branch is optional

# Loops

- When the same or similar tasks need to be performed multiple times; for all elements of a list; for all columns of an array; etc.
  - Monte Carlo Simulation
  - Cross-Validation (delete one and etc)

```
> for(i in 1:10) {  
+   print(i*i)  
+ }  
[1] 1  
[1] 4  
[1] 9  
[1] 16  
[1] 25  
[1] 36  
[1] 49  
[1] 64  
[1] 81  
[1] 100
```

```
> i=1  
> while(i<=10) {  
+   print(i*i)  
+   i=i+sqrt(i)  
+ }  
[1] 1  
[1] 4  
[1] 11.65685  
[1] 27.68836  
[1] 57.0912
```

# lapply, sapply, apply

- When the same or similar tasks need to be performed multiple times for all elements of a list or for all columns of an array.
  - May be easier and faster than “for” loops
- `lapply(li, function )`
  - To each element of the list `li`, the function *function* is applied.
  - The result is a list whose elements are the individual *function* results.

```
> li = list("klaus", "martin", "georg")
> lapply(li, toupper)
[[1]]
[1] "KLAUS"

[[2]]
[1] "MARTIN"

[[3]]
[1] "GEORG"
```

# lapply, sapply, apply

- `sapply( li, fct )`
- Like `apply`, but tries to simplify the result, by converting it into a vector or array of appropriate size

```
> li = list("klaus", "martin", "georg")
> sapply(li, toupper)
[1] "KLAUS" "MARTIN" "GEORG"
```

```
> fct = function(x) { return(c(x, x*x, x*x*x)) }
> sapply(1:5, fct)
      [,1] [,2] [,3] [,4] [,5]
[1,]    1    2    3    4    5
[2,]    1    4    9   16   25
[3,]    1    8   27   64  125
```

# apply

`apply( arr, margin, fct )`

Apply the function `fct` along some dimensions of the array `arr`, according to `margin`, and return a vector or array of the appropriate size.

```
> A
      [,1] [,2] [,3] [,4]
[1,]     1     4     7    10
[2,]     2     5     8    11
[3,]     3     6     9    12
> apply(A, 1, sum)
[1] 22 26 30
> apply(A, 2, sum)
[1]  6 15 24 33
```

# Functions and Operators

Functions do things with data

“Input”: function arguments (0,1,2,...)

“Output”: function result (exactly one)

Example:

```
add = function(a,b)
{ result = a+b
  return(result) }
```

Operators:

Short-cut writing for frequently used functions of one or two arguments.

Examples: + - \* / ! & | %%

# Functions and Operators

- Functions do things with data
  - “Input”: function arguments (0,1,2,...)
  - “Output”: function result (exactly one)

## Exceptions to the rule:

- Functions may also use data that sits around in other places, not just in their argument list: “scoping rules”\*
- Functions may also do other things than returning a result. E.g., plot something on the screen: “side effects”

# Numeric Functions

Function	Description
<b>abs(<math>x</math>)</b>	absolute value
<b>sqrt(<math>x</math>)</b>	square root
<b>ceiling(<math>x</math>)</b>	ceiling(3.475) is 4
<b>floor(<math>x</math>)</b>	floor(3.475) is 3
<b>trunc(<math>x</math>)</b>	trunc(5.99) is 5
<b>round(<math>x</math>, digits=<math>n</math>)</b>	round(3.475, digits=2) is 3.48
<b>signif(<math>x</math>, digits=<math>n</math>)</b>	signif(3.475, digits=2) is 3.5
<b>cos(<math>x</math>), sin(<math>x</math>), tan(<math>x</math>)</b>	also acos( $x$ ), cosh( $x$ ), acosh( $x$ ), etc.
<b>log(<math>x</math>)</b>	natural logarithm
<b>log10(<math>x</math>)</b>	common logarithm
<b>exp(<math>x</math>)</b>	$e^x$



# Character Functions

Function	Description
<b>substr</b> ( <i>x</i> , <b>start</b> = <i>n1</i> , <b>stop</b> = <i>n2</i> )	Extract or replace substrings in a character vector. <code>x &lt;- "abcdef"</code> <code>substr(x, 2, 4)</code> is "bcd" <code>substr(x, 2, 4) &lt;- "22222"</code> is "a222ef"
<b>grep</b> ( <i>pattern</i> , <i>x</i> , <b>ignore.case</b> =FALSE, <b>fixed</b> =FALSE)	Search for <i>pattern</i> in <i>x</i> . If <b>fixed</b> =FALSE then <i>pattern</i> is a <a href="#">regular expression</a> . If <b>fixed</b> =TRUE then <i>pattern</i> is a text string. Returns matching indices. <code>grep("A", c("b","A","c"), fixed=TRUE)</code> returns 2
<b>sub</b> ( <i>pattern</i> , <i>replacement</i> , <i>x</i> , <b>ignore.case</b> =FALSE, <b>fixed</b> =FALSE)	Find <i>pattern</i> in <i>x</i> and replace with <i>replacement</i> text. If <b>fixed</b> =FALSE then <i>pattern</i> is a regular expression. If <b>fixed</b> = T then <i>pattern</i> is a text string. <code>sub("\\s", ".", "Hello There")</code> returns "Hello.There"
<b>strsplit</b> ( <i>x</i> , <i>split</i> )	Split the elements of character vector <i>x</i> at <i>split</i> . <code>strsplit("abc", "")</code> returns 3 element vector "a","b","c"
<b>paste</b> (..., <b>sep</b> ="")	Concatenate strings after using <i>sep</i> string to separate them. <code>paste("x",1:3,sep="")</code> returns <code>c("x1","x2" "x3")</code> <code>paste("x",1:3,sep="M")</code> returns <code>c("xM1","xM2" "xM3")</code> <code>paste("Today is", date())</code>
<b>toupper</b> ( <i>x</i> )	Uppercase
<b>tolower</b> ( <i>x</i> )	Lowercase

# Date Values

**Dates are represented as the number of days since 1970-01-01, with negative values for earlier dates.**

```
> # use as.Date( ) to convert strings to dates
> mydates <- as.Date(c("2007-06-22", "2004-02-13"))
> # number of days between 6/22/07 and 2/13/04 days <- mydates[1] - mydates[2]
> mydates
[1] "2007-06-22" "2004-02-13"
```

**Sys.Date( ) returns today's date.**

**Date() returns the current date and time.**

# Date Values

The following symbols can be used with the `format( )` function to print dates.

Symbol	Meaning	Example
<b>%d</b>	day as a number (0-31)	01-31
<b>%a</b>	abbreviated weekday	Mon
<b>%A</b>	unabbreviated weekday	Monday
<b>%m</b>	month (00-12)	00-12
<b>%b</b>	abbreviated month	Jan
<b>%B</b>	unabbreviated month	January
<b>%y</b>	2-digit year	07
<b>%Y</b>	4-digit year	2007

# Date Values

```
# print today's date  
today <- Sys.Date()  
format(today, format="%B %d %Y")  
"June 20 2007"
```