

Data Analytics for Materials Science

A.D. (Tony) Rollett

& unknown author "TB"

Dept. Materials Sci. Eng.
Carnegie Mellon University

Introduction to 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)

← 上一頁 → 搜尋 我的最愛

網址(D) <http://www.r-project.org/> 移至 連結 >>

Google [G](#) R 開始 [RS](#) 2 已擱載 拼字檢查 a i 翻譯 >> 設定

The R Project for Statistical Computing

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

Download
[CRAN](#)

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

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

Clustering 4 groups

Factor 1 [41%]

Factor 3 [19%]

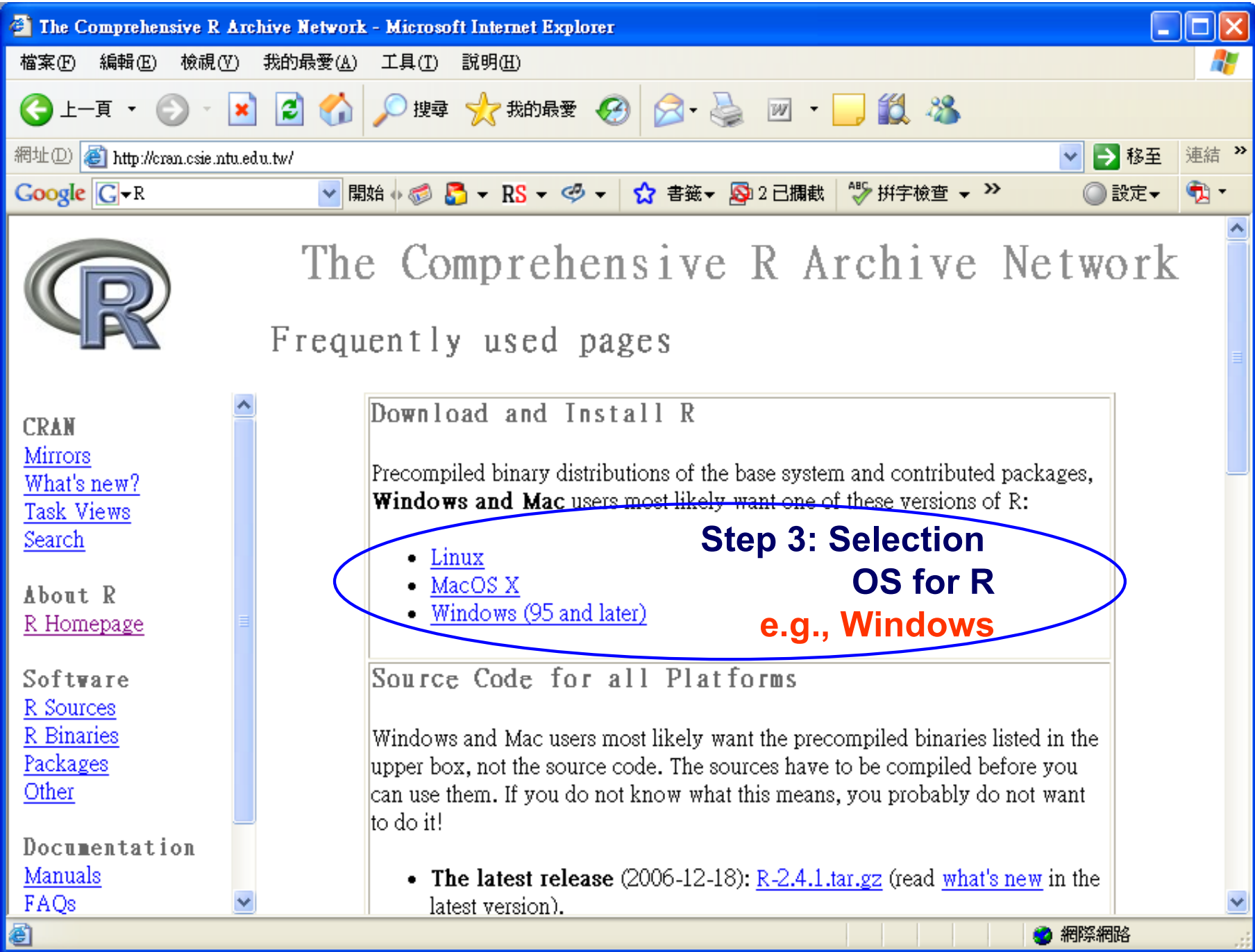
45

網際網路

Selection a Mirror Site

The screenshot shows the Microsoft Internet Explorer browser window displaying the R Project website. The address bar shows <http://www.r-project.org/>. The page content includes a navigation menu on the left and a list of mirror sites. A blue oval highlights the Taiwan mirror sites, and the text "Step 2: Selection a mirror site" is overlaid on the oval.

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



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).



R for Windows

This directory contains binaries for a base distribution and packages to run on Windows (NT, 95 and later) on Intel and clones (but not NT on Alpha and other platforms).

Note: CRAN does not have Windows systems and cannot check these binaries for viruses. Use the normal precautions with downloaded executables.

Subdirectories:

Clicking "**base**"

- [base](#) Binaries for base distribution (managed by Duncan Murdoch)
- [contrib](#) Binaries of contributed packages (managed by Uwe Ligges)

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Last modified: April 4, 2004, by Friedrich Leisch

CRAN

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

About R

- [R Homepage](#)

Software

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

Documentation

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


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



This directory is for Windows XP on...

A build of the development version (which will eventually become R 2.6.0 in the fall) is available in the [R-level snapshot build](#)

available [here](#). Please

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？

執行(R) 儲存(S) 取消

Clicking

[R-2.4.1-win32.exe](#)

Installation and other instructions.

New features of this Windows version.

New features of all versions.

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 versions.

網際網路

Installation of R

The image shows a grid of 13 file icons arranged in three rows. The icons are as follows:

- Row 1: NAV10CHT.exe (yellow folder with arrow), Osiris419.pc.zip (yellow folder), PhotoKing.exe (computer with CD), **R-2.4.1-win32.exe** (computer with CD, highlighted with a red box),
- Row 2: RealOnePlayerUpdate... (yellow folder with window), SkypeSetup.exe (blue cube with arrow), vobsub_2.23.exe (blue globe with arrow), weka-3-4-10jre.exe (blue globe with arrow),
- Row 3: winrar310tc.exe (yellow folder with arrow), xmedcon-0.9.9.4-win3... (yellow folder), x-win543.exe (blue square with white shape).

A blue callout box with a white background and a blue border points to the 'R-2.4.1-win32.exe' icon. The text inside the callout box is:

**Double clicking
R-icon for
installation *R***

Update if You Already Have R

R Console

R version 3.6.2 (2019-12-12) -- "Dark and Stormy Night"
Copyright (C) 2019 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin15.6.0 (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.

Natural language support but running in an English locale

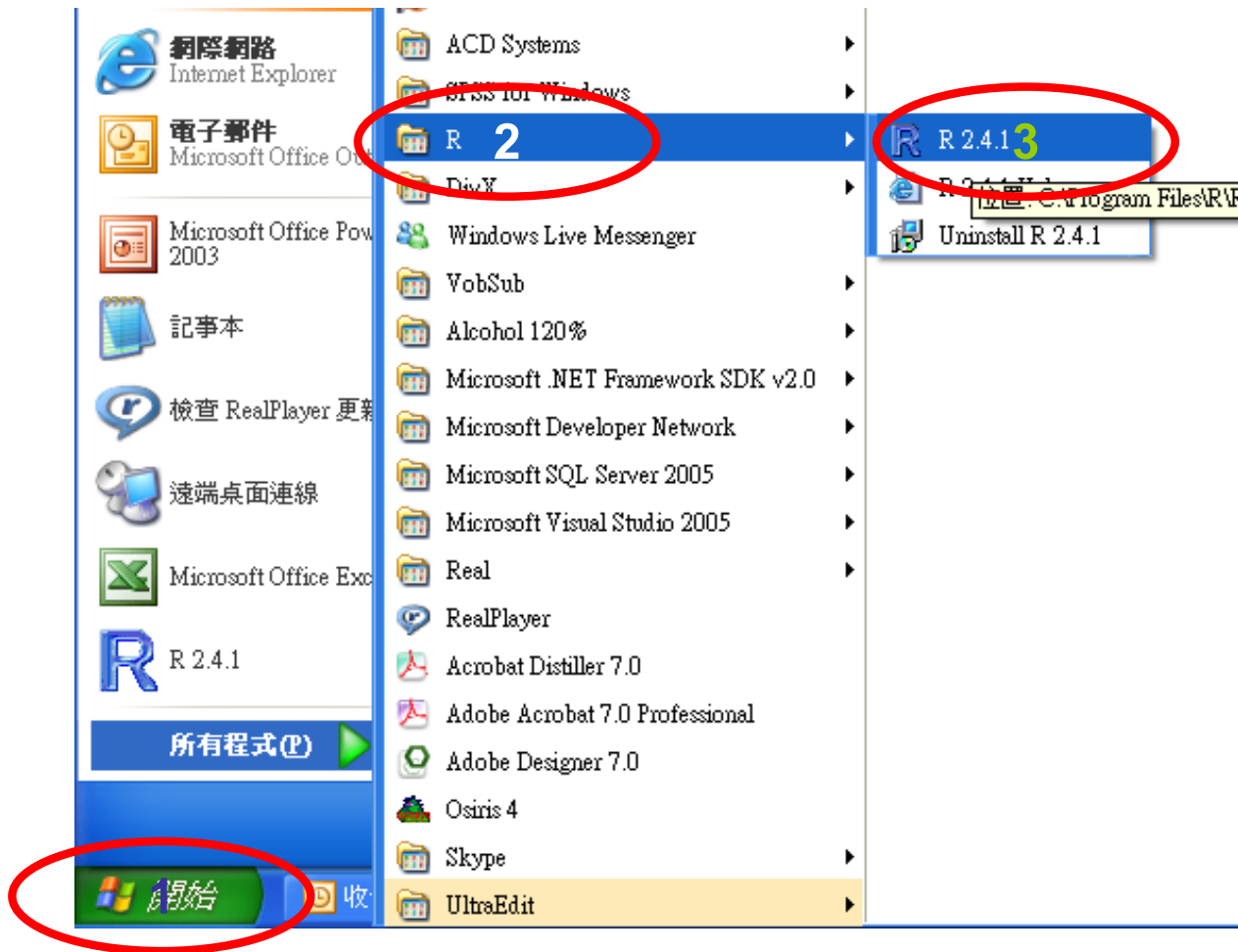
R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

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.

[R.app GUI 1.70 (7735) x86_64-apple-darwin15.6.0]
[Workspace restored from /Users/rollett/.RData]
[History restored from /Users/rollett/.Rapp.history]

This version of R is 3.6.2
There is a newer version of R on CRAN which is 4.0.3
> Do you want to visit CRAN now? yes

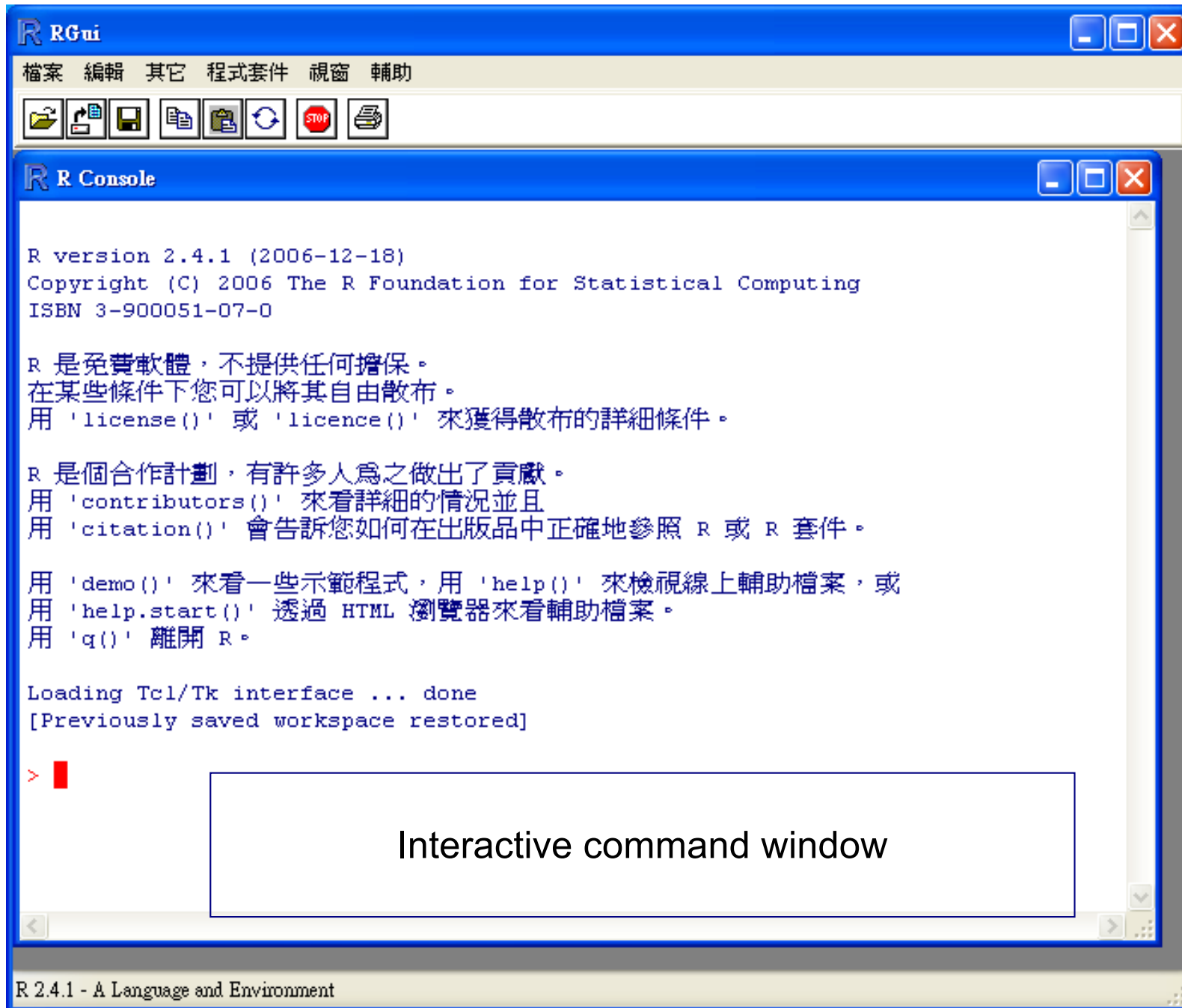
Using R



Or, on a Mac,
double-click "R"
in the Applications
window.

You can also put the
R icon in your Dock
or in the side-bar.

Same for R Studio.



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)

From a spreadsheet of data ...

Recap-18Ech-traction-fw-ao-ADR.xlsx

Search in Sheet

Home Layout Tables Charts SmartArt Formulas Data Review

Edit Font Alignment Number Format Cells Themes

Calibri (Body) 11

General

F28

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
3	Recapitulation: 18 Ti specimens																	
4																		
5		Ech.	serie	Rolling	TT	ReX	D_my	1/VD	Tr//	YS	L_P1	σ_{e_max}	ϵ^*	$\sigma_{max} - \sigma_e$	L_P2	ϵ_r		
6							(in μm)			MPa	%	MPa	%	MPa	%	%		
7		R1	1	75% //RD	740°C-2h	100%	11.7	0.292	RD	315	-	466	19.5	151	22.5	35.0		
8		R2	1	75% //RD	500°C-1h	91%	1.8	0.745	RD	420	1.9	510	19.6	90	21.4	32.4		
9		R3	1	75% //RD	470°C-2h	62%	1.2	0.913	RD	490	-	600	14.6	110	20.4	27.6		
10		R4	1	75% //RD	500°C-40min	80.00%	1.7	0.767	RD	439	1.0	538	17.3	99		33.7		
11		R5	1	75% //RD	550°C-40min	98.00%	2.8	0.598	RD	375	1.8	485	21	110		40.2		
12																		
13		RR1	2	75% //RD	740°C-2h	98.30%	9.7	0.321	RD	332	~0	468	14	136	14.6	24.3		
14		RR2	2	75% //RD	500°C-1h	69.20%	1.4	0.845	RD	404		480	7.2		9.5	13.5	<< too thin	
15		RR3	2	75% //RD	470°C-2h	46.00%	1.0	1	RD	544	~0	645	10.3	101	17.5	22.1		
16		RR4	2	75% //RD	500°C-40min				RD	526		674	12.3	148		22.6	<< essai méca. loupé	
17																		
18		RT1	2	75% //RD	740°C-2h	98.30%	9.7	0.321	TD	379	1.5	456	9	77	14.7	28.5		
19		RT2	2	75% //RD	500°C-1h	69.20%	1.4	0.845	TD	511		523.5	0.3	12.5	10.5	21.2		
20		RT3	2	75% //RD	470°C-2h	46.00%	1.0	1	TD	621		672	1.1	51	4.9	14.3		
21		RT4	2	75% //RD	500°C-40min				TD	522	-	579	2.2	57		14.4		
22																		
23		T1	1	75% //TD	740°C-2h	100%	10.9	0.303	RD'	300	-	425	19.0	125	24.0	35.0		
24		T2	1	75% //TD	500°C-1h	98%	1.8	0.745	RD'	400	3.7	450	19.5	50	29.4	43.0		
25		T3	1	75% //TD	470°C-2h	80%	1.2	0.913	RD'	485	2.1	512	17.0	27	26.2	36.5		
26		T4	1	75% //TD	500°C-40min	94.00%	1.7	0.767	RD'	390	3.0	463	19.3	73		44.5		
27		T5	1	75% //TD	550°C-40min	98.00%	2.6	0.62	RD'	375	3.9	436	21	61		44.8		
28																		
29		Tr// : direction of extension (tensile test with 'small' samples)																
30		L_P1: length (in %) of the plateau after the YS																
31		σ_{e_max} : max Eng. stress																
32		ϵ^* : elongation (in %) for the max stress (Considere's criterium)																

Courtesy of Prof. Francis Wagner, Univ. Lorraine, Metz, France

Feuil1 Feuil2 Feuil3 +

Such a spreadsheet requires "clean up"

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	Anneal_time	Anneal_Tem	Rolling_Dire	Test_Directio	Yield_Stress	L_P1	Eng_Stress	Strain_at_pe	stress_max	Strain_to_fa	per_cent_rec	Grain_Diam	1/sqrtD	L_P2	
2	120	740	1	1	315	0	466	19.5	151	35	1	11.7	0.29235267	22.5	
3	60	500	1	1	420	1.9	510	19.6	90	32.4	0.91	1.8	0.74535599	21.4	
4	120	470	1	1	490	0	600	14.6	110	27.6	0.62	1.2	0.91287093	20.4	
5	40	500	1	1	439	1	538	17.3	99	33.7	0.8	1.7	0.76696499	0	
6	40	550	1	1	375	1.8	485	21	110	40.2	0.98	2.8	0.59761431	0	
7	120	740	1	1	332	0	468	14	136	24.3	0.98	9.7	0.32108065	14.6	
8	60	500	1	1	404	0	480	7.2	90	13.5	0.69	1.4	0.84515426	9.5	
9	120	470	1	1	544	0	645	10.3	101	22.1	0.46	1	1	17.5	
10	40	500	1	1	526	0	674	12.3	148	22.6	0.8	1.7	0.76696499	0	
11	120	740	1	2	379	1.5	456	9	77	28.5	0.98	9.7	0.32108065	14.7	
12	60	500	1	2	511	0	523.5	0.3	12.5	21.2	0.69	1.4	0.84515426	10.5	
13	120	470	1	2	621	0	672	1.1	51	14.3	0.46	1	1	4.9	
14	40	500	1	2	522	0	579	2.2	57	14.4	0.8	1.7	0.76696499	0	
15	120	740	2	1	300	0	425	19	125	35	1	10.9	0.30289127	24	
16	60	500	2	1	400	3.7	450	19.5	50	43	0.98	1.8	0.74535599	29.4	
17	120	470	2	1	485	2.1	512	17	27	36.5	0.8	1.2	0.91287093	26.2	
18	40	500	2	1	390	3	463	19.3	73	44.5	0.94	1.7	0.76696499	0	
19	40	550	2	1	375	3.9	436	21	61	44.8	0.98	2.6	0.62017367	0	
20															
21															
22															

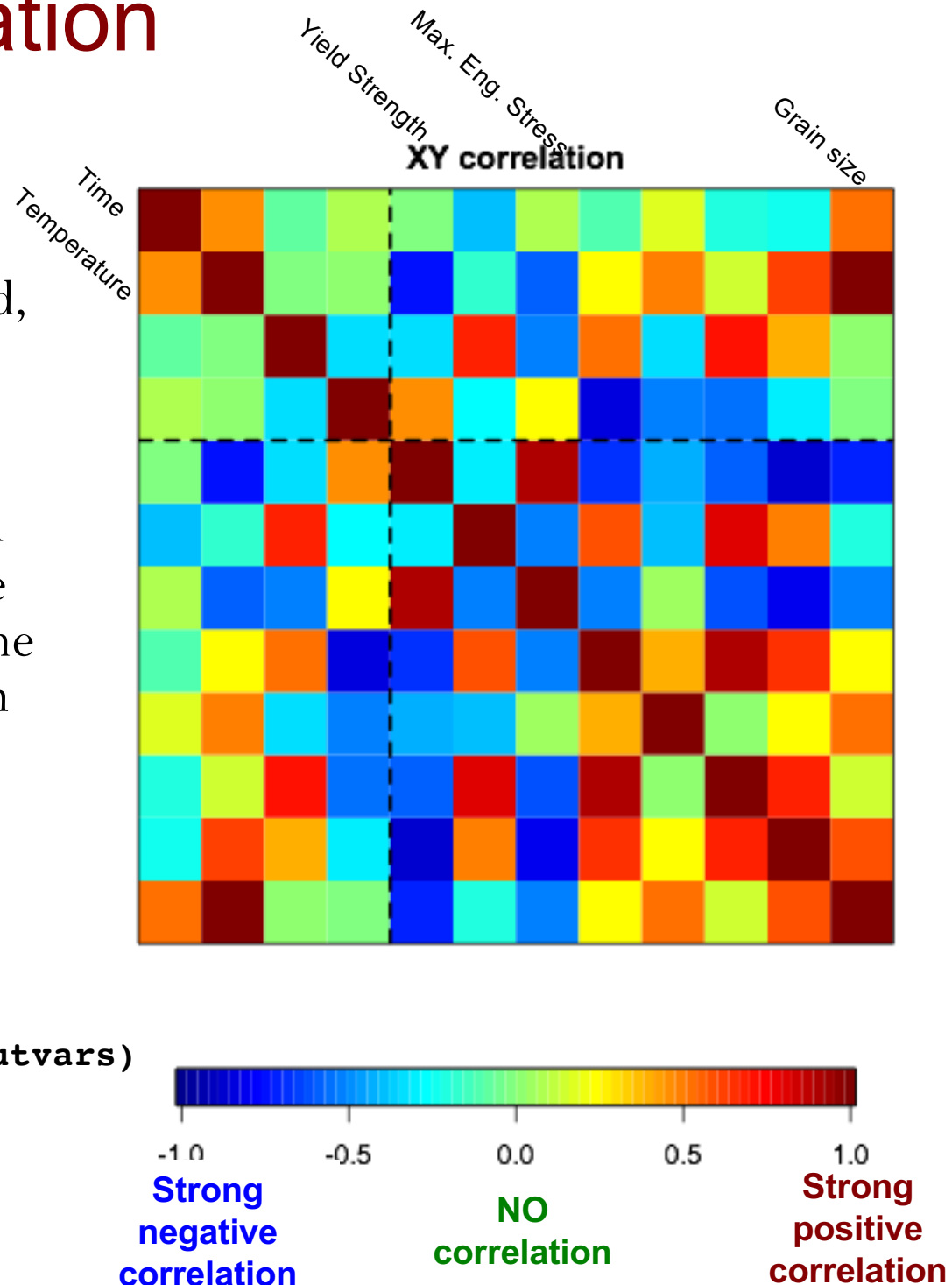
Column names must be simplified. Check that each column is numeric, not text. No comments or extraneous text allowed.

... to a correlation analysis

For a very straightforward, visual way to visualize correlations and cross-correlations, one can use *matcor*. The top right and bottom left quadrants are mirror images and show the cross-correlation between input & output.

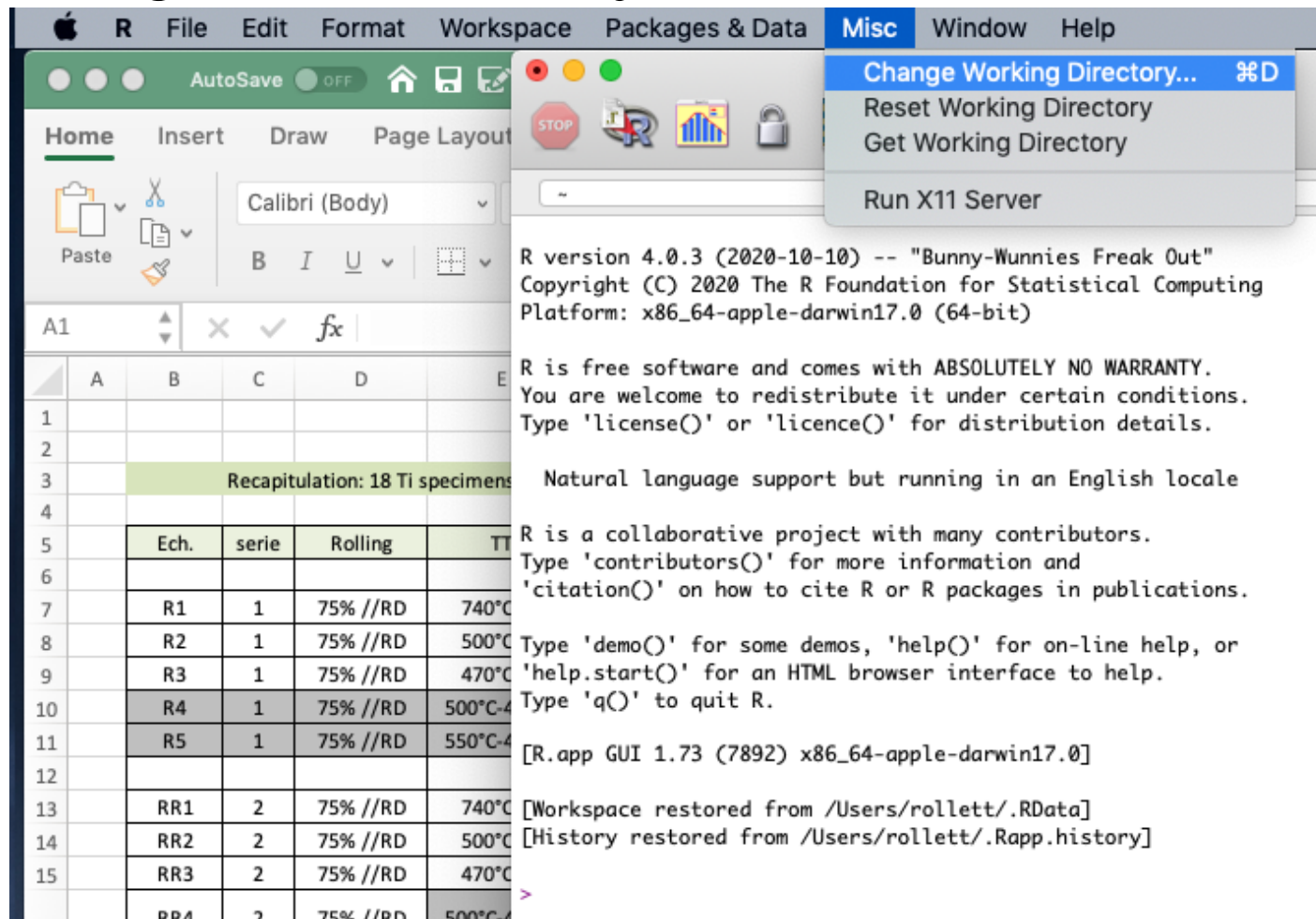
- `Data = read.table('file')`
- `outvars=data[,1:2]`
- `invars = data[,3:12]`
- > `simpleCorr=matcor(invars,outvars)`
- > `img.matcor(simpleCorr)`

Thanks to Francis Wagner, Univ. Metz, France



Change Working Directory

Navigate to where you have data* stored ...



The screenshot shows the RStudio application window. The 'Misc' menu is open, and the 'Change Working Directory...' option is highlighted. The background shows a spreadsheet with data and the R console output.

	A	B	C	D	E
1					
2					
3		Recapitulation: 18 Ti specimens			
4					
5		Ech.	serie	Rolling	TT
6					
7		R1	1	75% //RD	740°C
8		R2	1	75% //RD	500°C
9		R3	1	75% //RD	470°C
10		R4	1	75% //RD	500°C-4
11		R5	1	75% //RD	550°C-4
12					
13		RR1	2	75% //RD	740°C
14		RR2	2	75% //RD	500°C
15		RR3	2	75% //RD	470°C
		RR4	2	75% //RD	500°C

```
R version 4.0.3 (2020-10-10) -- "Bunny-Wunnies Freak Out"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-apple-darwin17.0 (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.

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.

[R.app GUI 1.73 (7892) x86_64-apple-darwin17.0]

[Workspace restored from /Users/rollett/.RData]
[History restored from /Users/rollett/.Rapp.history]
>
```

* cp_Ti-Tensile-tests-2015-simplified-ADR.xlsx

Even so ...

```
> Data = read.xlsx('cp_Ti-Tensile-tests-2015-simplified-ADR.xlsx')
```

```
Error: unexpected input in "Data = read.xlsx("
```

```
> Data = read.xlsx('cp_Ti-Tensile-tests-2015-simplified-ADR.xlsx')
```

```
Error: unexpected input in "Data = read.xlsx("
```

```
> Data = read.xlsx(cp_Ti-Tensile-tests-2015-simplified-ADR.xlsx)
```

```
Error in read.xlsx(cp_Ti - Tensile - tests - 2015 - simplified - ADR.xlsx) :
```

```
object 'cp_Ti' not found
```

```
> Data = read.xlsx("cp_Ti-Tensile-tests-2015-simplified-ADR.xlsx")
```

In-class exercise: how did we circumvent these errors?!

Hint: look very carefully at all the characters used to specify the file

Get a "summary" of the Data

```
> summary(Data)
```

```
Anneal_time  Anneal_Temperature_degrC  Rolling_Direction  Test_Direction  Yield_Stress
Min.   : 40  Min.   :470.0  Min.   :1.000  Min.   :1.000  Min.   :300.0
1st Qu.: 40  1st Qu.:500.0  1st Qu.:1.000  1st Qu.:1.000  1st Qu.:376.0
Median : 60  Median :500.0  Median :1.000  Median :1.000  Median :412.0
Mean   : 80  Mean   :552.2  Mean   :1.278  Mean   :1.222  Mean   :434.9
3rd Qu.:120  3rd Qu.:550.0  3rd Qu.:1.750  3rd Qu.:1.000  3rd Qu.:505.8
Max.   :120  Max.   :740.0  Max.   :2.000  Max.   :2.000  Max.   :621.0

  L_P1  Eng_Stress_max  Strain_at_peak_stress  stress_max-s_yield  Strain_to_failure
Min.   :0.000  Min.   :425.0  Min.   : 0.300  Min.   : 12.50  Min.   :13.50
1st Qu.:0.000  1st Qu.:463.8  1st Qu.: 9.325  1st Qu.: 58.00  1st Qu.:22.23
Median :0.000  Median :497.5  Median :15.800  Median : 90.00  Median :30.45
Mean   :1.050  Mean   :521.2  Mean   :13.567  Mean   : 87.14  Mean   :29.64
3rd Qu.:1.875  3rd Qu.:568.8  3rd Qu.:19.450  3rd Qu.:110.00  3rd Qu.:36.12
Max.   :3.900  Max.   :674.0  Max.   :21.000  Max.   :151.00  Max.   :44.80

per_cent_recrystallized  Grain_Diameter  1/sqrtD  L_P2
Min.   :0.4600  Min.   : 1.000  Min.   :0.2924  Min.   : 0.00
1st Qu.:0.7175  1st Qu.: 1.400  1st Qu.:0.6033  1st Qu.: 0.00
Median :0.8550  Median : 1.700  Median :0.7670  Median :12.55
Mean   :0.8261  Mean   : 3.611  Mean   :0.6961  Mean   :11.98
3rd Qu.:0.9800  3rd Qu.: 2.750  3rd Qu.:0.8452  3rd Qu.:21.15
Max.   :1.0000  Max.   :11.700  Max.   :1.0000  Max.   :29.40
```

Access a single column with "\$"

Note that almost any statistical quantity that you need can be used as function, e.g., mean, median etc.

Suffixing with \$ and the column name will access that part of the data

```
> mean(Data$Anneal_time)
[1] 80
```

Operators

- Mathematical 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)

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
```

Distributions in R

- Notation:
 - Probability Density Function: **p**
 - Distribution Function: **p**
 - Quantile function: **q**
 - Random generation for distribution: **r**
- Example:
 - Normal distribution:
 - `dnorm(x, mean=0, sd=1, log = FALSE)`
 - `pnorm(q, mean=0, sd=1, lower.tail = TRUE, log.p = FALSE)`
 - `qnorm(p, mean=0, sd=1, lower.tail = TRUE, log.p = FALSE)`
 - `rnorm(n, mean=0, sd=1)`

- Weibull Distribution

- `dweibull(x, shape, scale = 1, log = FALSE)`
- `pweibull(q, shape, scale = 1, lower.tail = TRUE, log.p = FALSE)`
- `qweibull(p, shape, scale = 1, lower.tail = TRUE, log.p = FALSE)`
- `rweibull(n, shape, scale = 1)`

- Log Normal Distribution

- `dlnorm(x, meanlog = 0, sdlog = 1, log = FALSE)`
- `plnorm(q, meanlog = 0, sdlog = 1, lower.tail = TRUE, log.p = FALSE)`
- `qlnorm(p, meanlog = 0, sdlog = 1, lower.tail = TRUE, log.p = FALSE)`
- `rlnorm(n, meanlog = 0, sdlog = 1)`

Statistical Functions

Excel

NORMSDIST

NORMSINV

LOGNORMDIST

LOGINV

GAMMADIST

GAMMAINV

GAMMALN

WEIBULL

BINOMDIST

POISSON

R

`pnorm(7.2,mean=5,sd=2)`

`qnorm(0.9,mean=5,sd=2)`

`plnorm(7.2,meanlog=5,sdlog=2)`

`qlnorm(0.9,meanlog=5,sdlog=2)`

`pgamma(31, shape=3, scale =5)`

`qgamma(0.95, shape=3, scale =5)`

`lgamma(4)`

`pweibull(6, shape=3, scale =5)`

`pbinom(2,size=20,p=0.3)`

`ppois(2, lambda =3)`

Write Data to a TXT File

- Usage:
- ***write***(x, file, ...)

- `x<-matrix(c(1.0, 2.0, 3.0, 4.0, 5.0, 6.0), 2, 3) ; x`
- `[,1] [,2] [,3]`
- `[1,] 1 3 5`
- `[2,] 2 4 6`
- `write(t(x), file="d:/out2.txt", ncolums=3)`
- `write(x, file="d:/out3.txt", ncolums=3)`

d:/out2.txt

1 3 5
2 4 6

d:/out3.txt

1 2 3
4 5 6

Write Data to a CSV File

- Usage:
 - ***write.table***(x, file = "foo.csv", sep=",", ",...)
 - Example:
 - `x<-matrix(c(1.0, 2.0, 3.0, 4.0, 5.0, 6.0), 2, 3) ; x`
 - `[,1] [,2] [,3]`
 - `[1,] 1 3 5`
 - `[2,] 2 4 6`
 - `write.table(t(x), file="d:/out4.txt", sep=",", col.names=FALSE, row.names=FALSE)`
 - `write.table(x, file="d:/out5.txt", sep=",", col.names=FALSE, row.names=FALSE)`
- d:/out4.txt
1,2
3,4
5,6
- d:/out5.txt
1,3,5
2,4,6

Read TXT and CSV File

- Usage:
- ***read.table*** (file,...)
- X=read.table(file="d:/out2.txt"); X
- V1 V2 V3
- 1 1 3 5
- 2 2 4 6
- X=read.table(file="d:/out5.txt",sep=",",header=FALSE); X

d:/out2.txt

1 3 5
2 4 6

d:/out5.csv

1,3,5
2,4,6

Demo Reading CSV File

```
> Data = read.table(file="d:/01.csv",header=TRUE, sep=",")
```

```
> Data
```

```
      Y      X1      X2
1 2.651680 13.808986 26.75896
2 1.875039 17.734523 37.89857
3 1.523964 19.891025 26.03624
4 2.984314 15.574261 30.21754
5 10.423087  9.293612 28.91459
6  0.840065  8.830160 30.38578
7  8.126936  9.615875 32.69579
```

```
> mean(Data$Y) # $ is used to identify which variable
```

```
[1] 4.060726
```

```
> mean(Data$X1)
```

```
[1] 13.53549
```

```
> mean(Data$X2)
```

```
[1] 30.41535
```

```
> sd(Data$Y)
```

```
[1] 3.691044
```

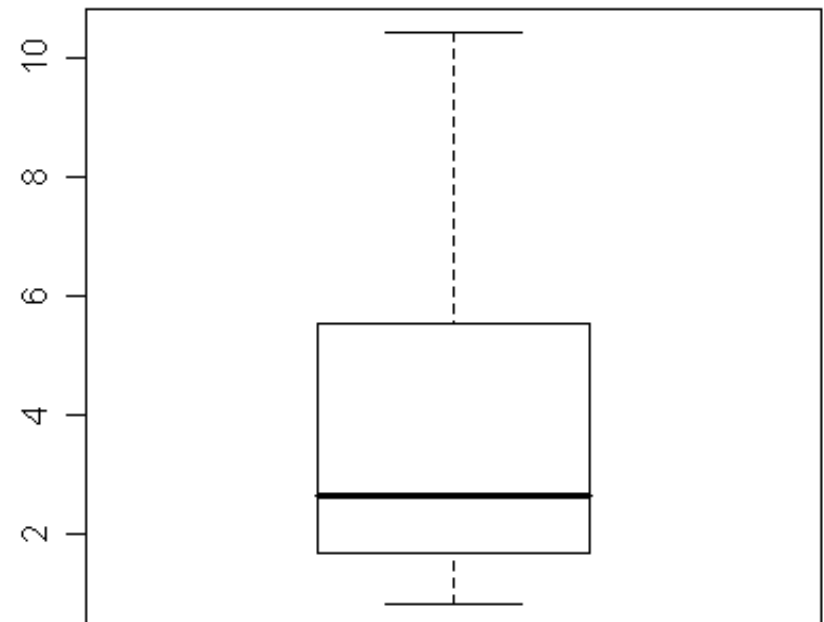
```
> summary(Data$Y)
```

```
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.8401  1.7000  2.6520  4.0610  5.5560 10.4200
```

```
> boxplot(Data$Y)
```

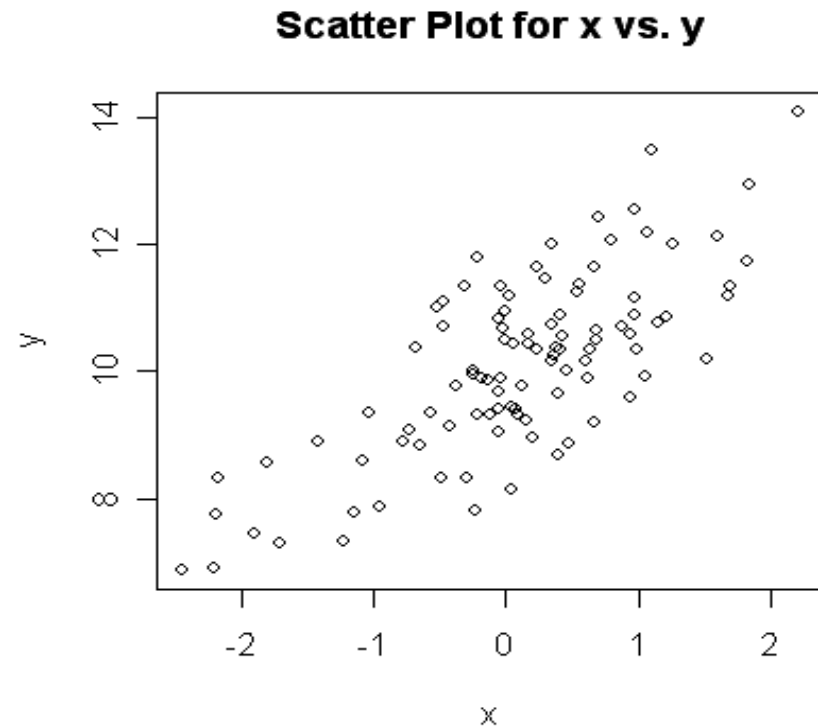
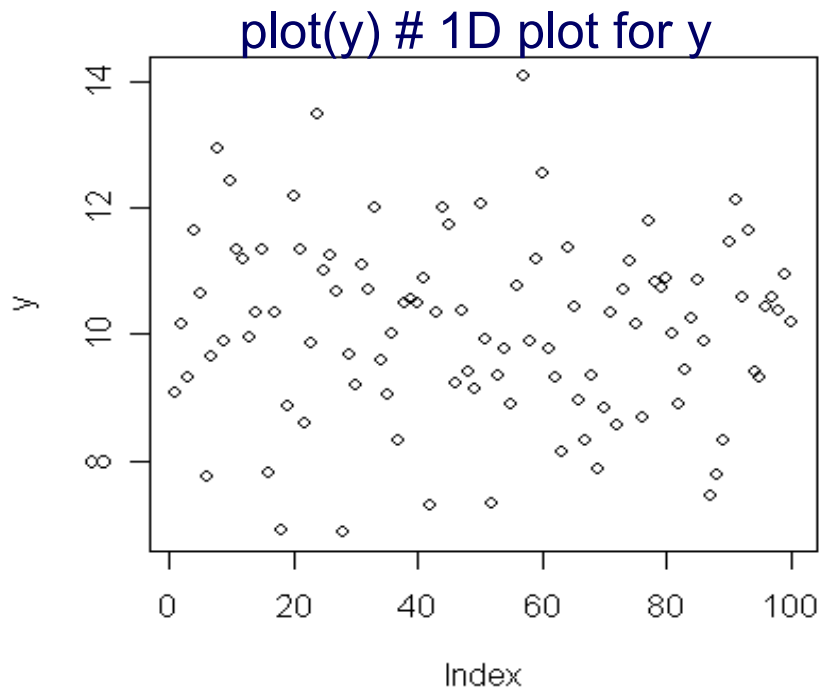
01.csv

Book1				
	A	B	C	
1	Y	X1	X2	
2	2.65168	13.80899	26.75896	
3	1.875039	17.73452	37.89857	
4	1.523964	19.89103	26.03624	
5	2.984314	15.57426	30.21754	
6	10.42309	9.293612	28.91459	
7	0.840065	8.83016	30.38578	
8	8.126936	9.615875	32.69579	



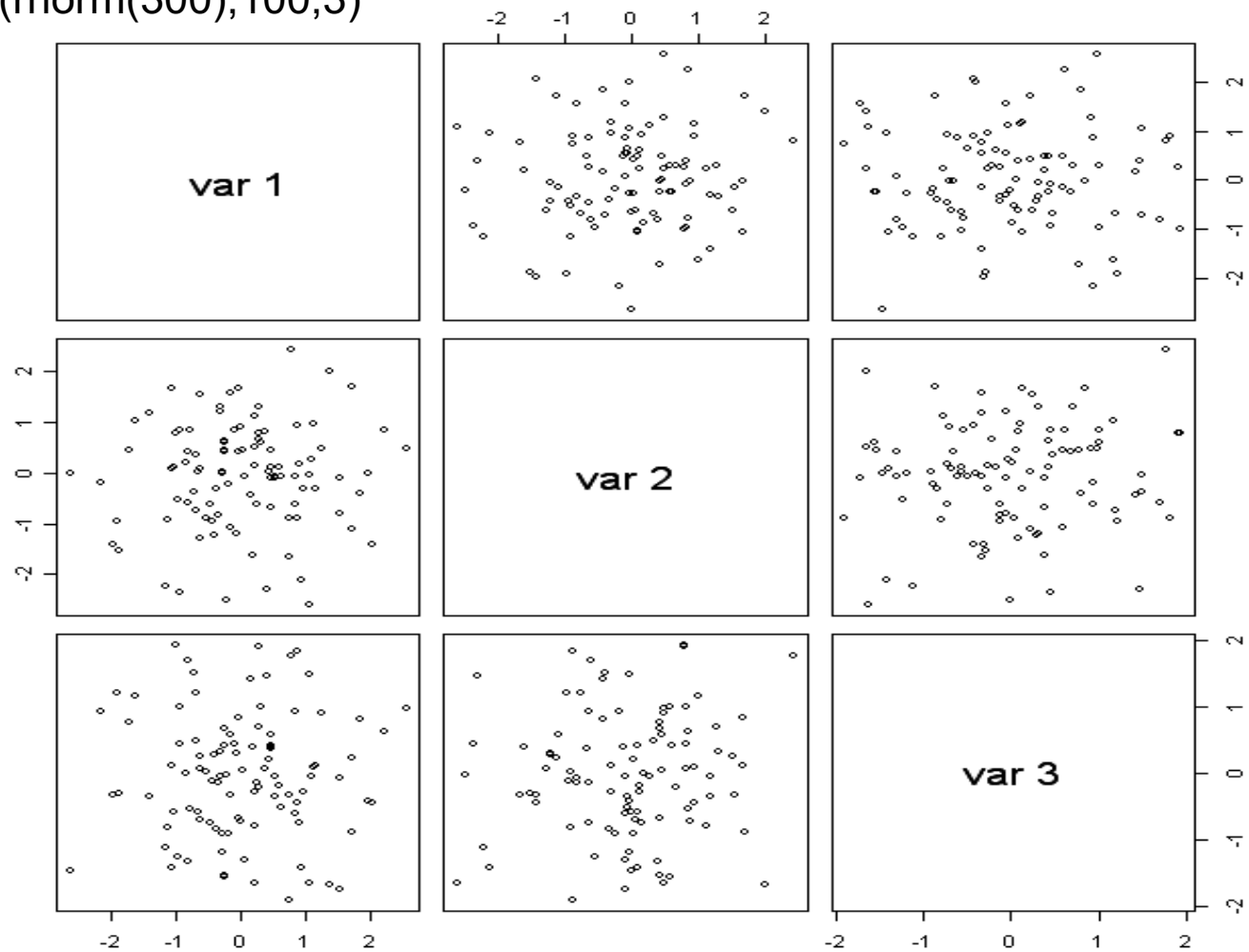
Scatter Plot in R

- Scatter plot:
 - `x = rnorm(100)` # Generated 100 $N(0,1)$
 - `y = 10 + 1.2*x + rnorm(100)` # Simulated y
 - `plot(y)` # 1D plot for y
 - `windows()` # Create a new graph device
 - `plot(x,y)` # 2D plot for x vs y
 - `title (main="Scatter Plot for x vs. y")` # Add title in plot



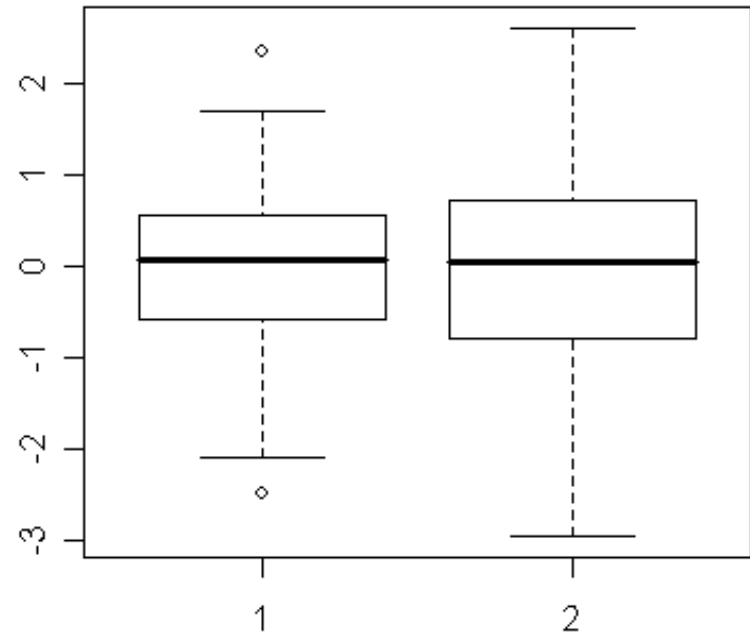
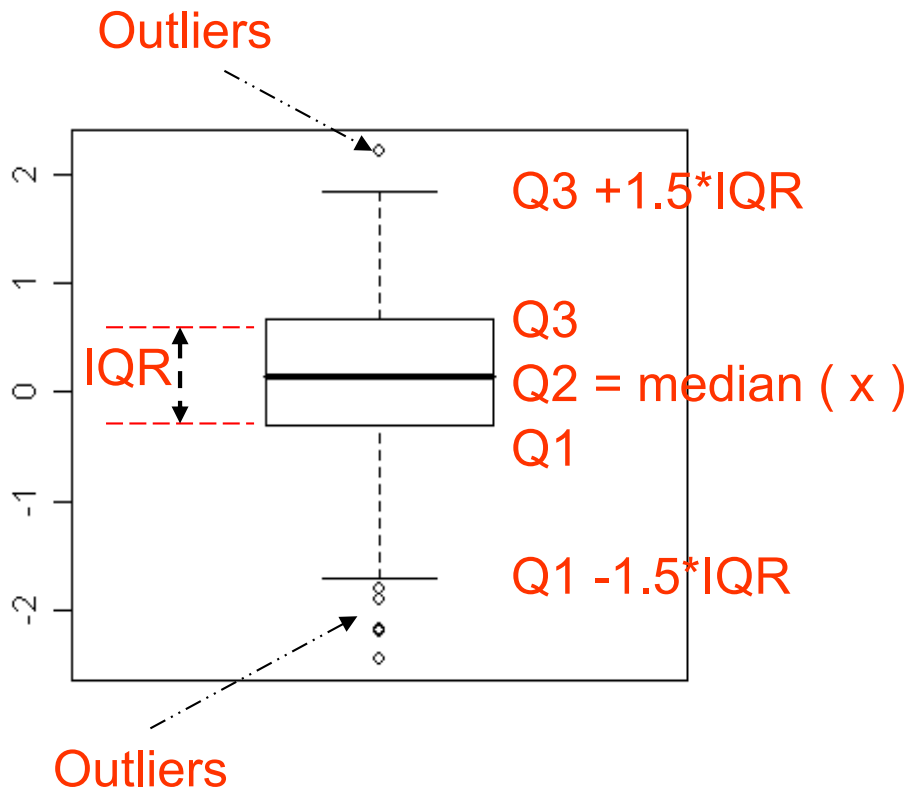
Scatterplot Matrices in R

- A matrix of scatterplots is produced.
 - Usage:
 - `pairs(x, ...)`
 - Example:
 - `X = matrix(rnorm(300),100,3)`
 - `pairs(X)`



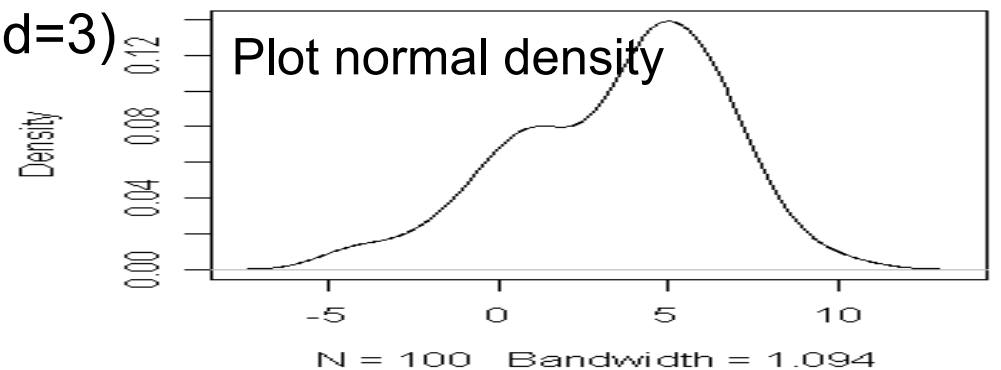
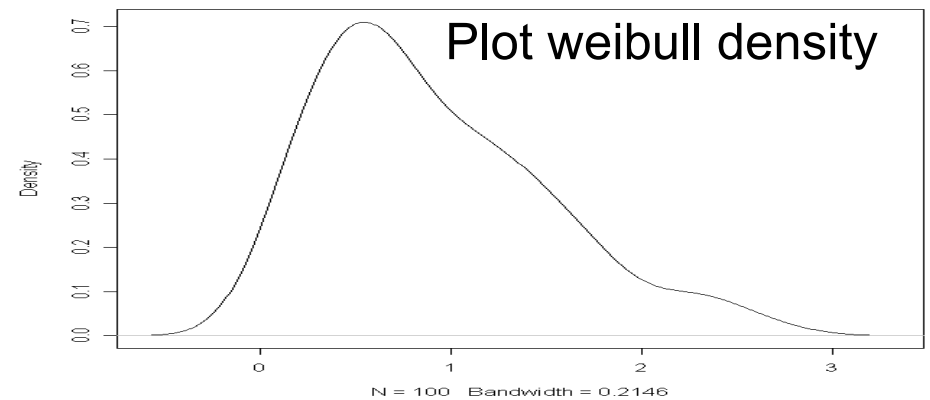
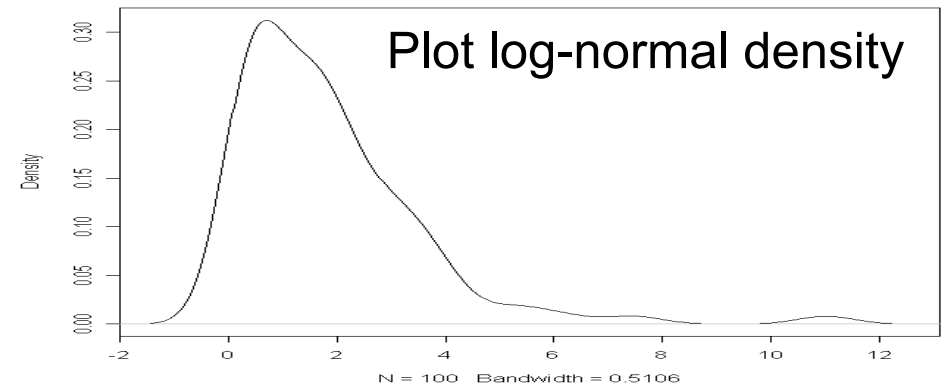
Box Plot in R

- Box plot:
 - Produce box-and-whisker plot of the given (grouped) values.
 - **Usage:** `boxplot(x, ...)`
 - **Example1:**
 - `X=rnorm(100)`
 - `boxplot(x)`
- Example2:
`x=rnorm(100); y=rnorm(100);`
`boxplot(x,y)`



Kernel Density Plot

- Kernel Density:
 - `density(x,...)`
 - Kernel Density Plot:
 - `plot(density(x,...))`
 - Examples:
 - `X= rlnorm(100,0,1)`
 - `plot(density(X))`
 - `Y= rweibull(100,1.5,1)`
 - `plot(density(Y))`
 - `x= rnorm(100, mean=3, sd=3)`
 - `plot(density(x))`



Probability Plots

- Need installation R package : [e1071](#)

The screenshot shows the R GUI interface with the following steps highlighted:

1. The 'Program Packages' menu is open, and 'Install Package...' is selected.
2. The 'CRAN mirror' dialog is open, and 'Taiwan (Taipei)' is selected.
3. The 'Packages' dialog is open, and 'e1071' is selected in the list.
4. The 'OK' button in the 'CRAN mirror' dialog is highlighted.
5. The 'OK' button in the 'Packages' dialog is highlighted.
6. The 'OK' button in the 'Packages' dialog is highlighted.

The R Console shows the following output:

```
R version 2.4.1
Copyright (C) 2005
ISBN 3-900051-07-0

R 是免費軟體，不提供任何擔保。
在某些條件下您可以將其自由散布。
用 'license()' 或 'licence()' 來獲得散布的詳細條款。

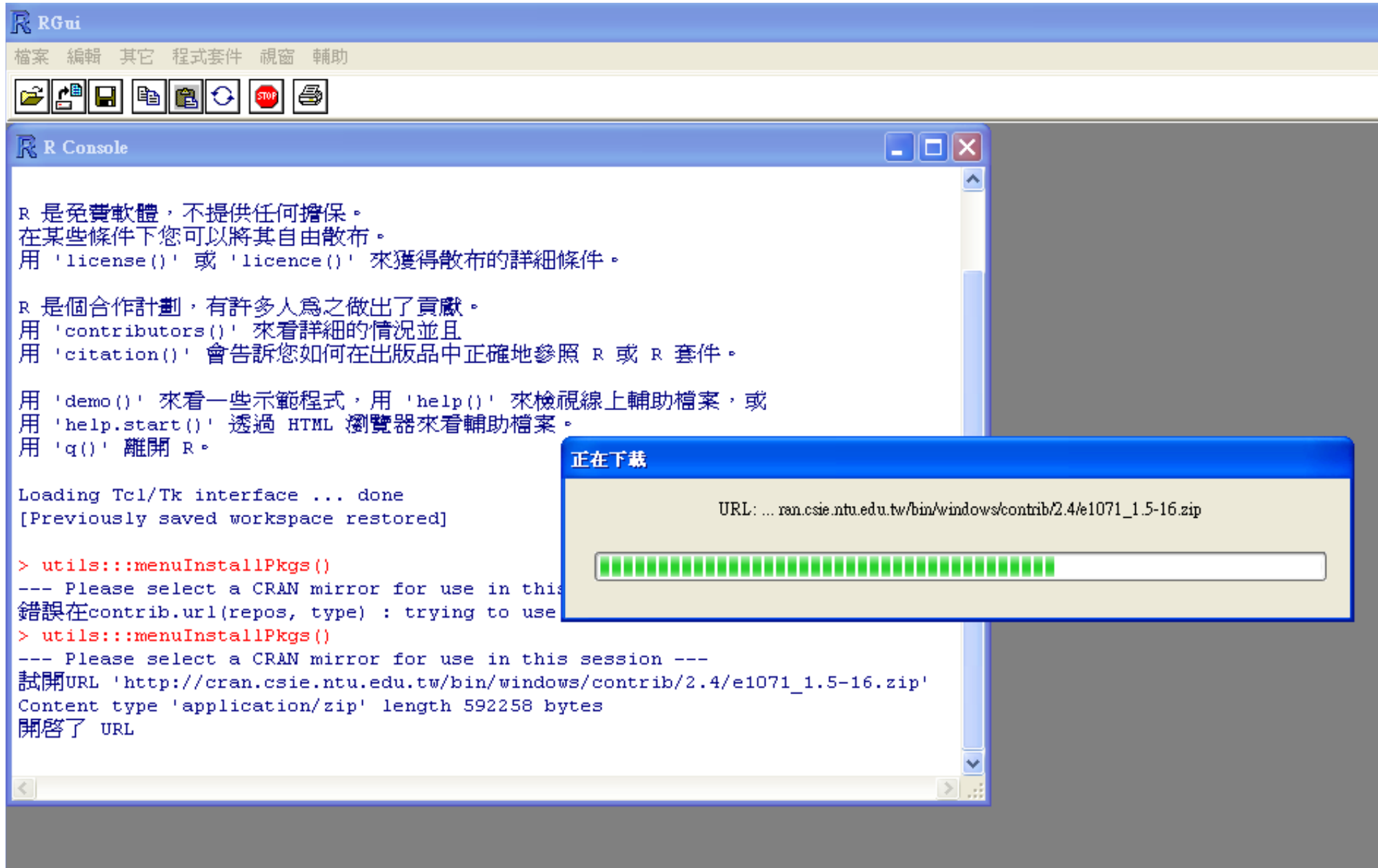
R 是個合作計劃，有許多人為之做出了貢獻。
用 'contributors()' 來看詳細的情況並且
用 'citation()' 會告訴您如何在出版品中正確地參照。

用 'demo()' 來看一些示範程式，用 'help()' 來檢視。
用 'help.start()' 透過 HTML 瀏覽器來看輔助檔案。
用 'q()' 離開 R。

Loading Tcl/Tk interface ... done
[Previously saved workspace restored]

> utils:::menuInstallPkgs()
--- Please select a CRAN mirror for use in this session
錯誤在contrib.url(repos, type) : trying to use CRAN without setting a mirror
>
```

Installing

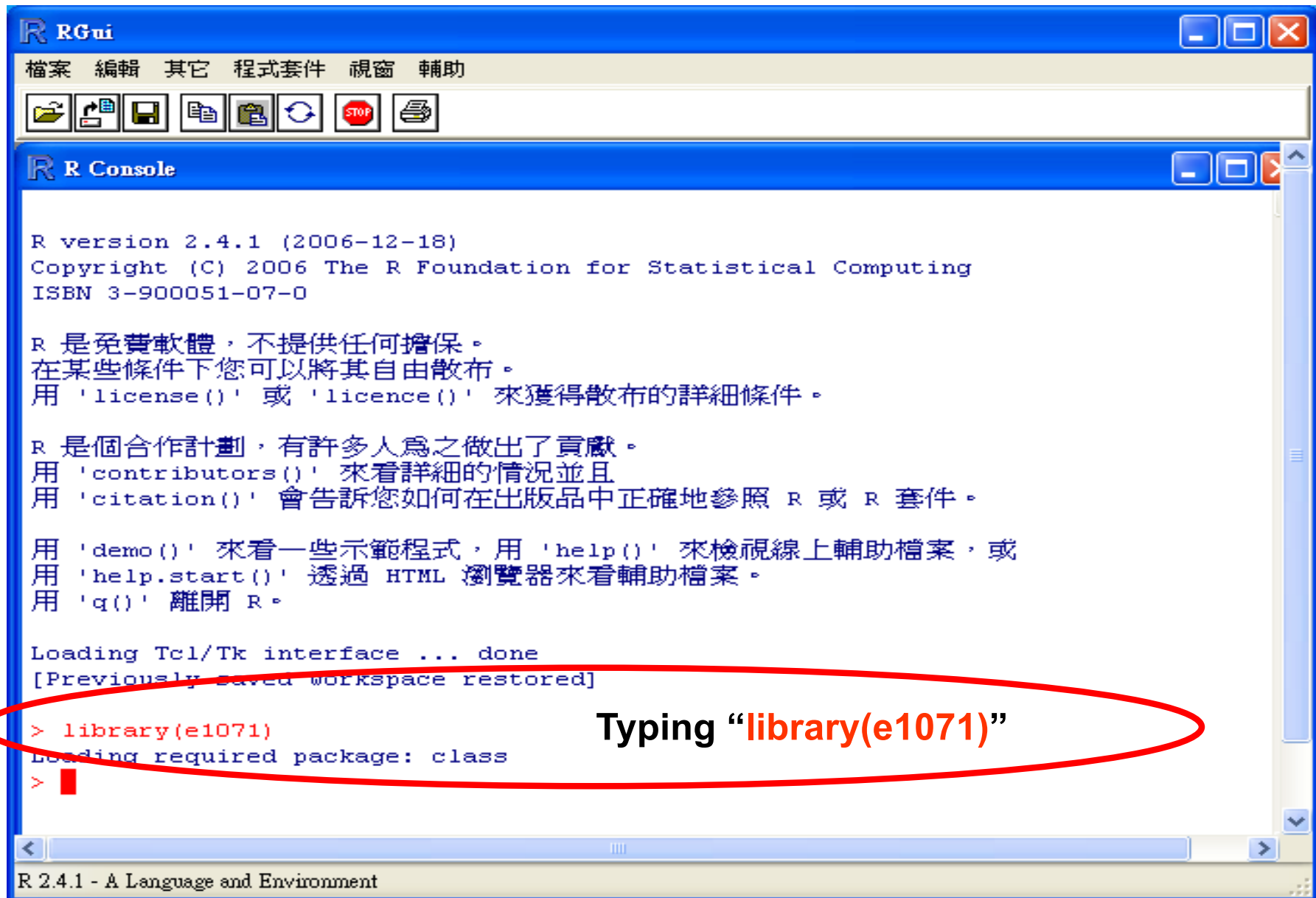


The screenshot shows the RGui interface. The R Console window displays the following text:

```
R 是免費軟體，不提供任何擔保。  
在某些條件下您可以將其自由散布。  
用 'license()' 或 'licence()' 來獲得散布的詳細條件。  
  
R 是個合作計劃，有許多人為之做出了貢獻。  
用 'contributors()' 來看詳細的情況並且  
用 'citation()' 會告訴您如何在出版品中正確地參照 R 或 R 套件。  
  
用 'demo()' 來看一些示範程式，用 'help()' 來檢視線上輔助檔案，或  
用 'help.start()' 透過 HTML 瀏覽器來看輔助檔案。  
用 'q()' 離開 R。  
  
Loading Tcl/Tk interface ... done  
[Previously saved workspace restored]  
  
> utils::menuInstallPkgs()  
--- Please select a CRAN mirror for use in this session ---  
錯誤在contrib.url(repos, type) : trying to use  
> utils::menuInstallPkgs()  
--- Please select a CRAN mirror for use in this session ---  
試開URL 'http://cran.csie.ntu.edu.tw/bin/windows/contrib/2.4/e1071_1.5-16.zip'  
Content type 'application/zip' length 592258 bytes  
開啓了 URL
```

A "正在下載" (Downloading) dialog box is overlaid on the console, showing the URL: `URL: ...ran.csie.ntu.edu.tw/bin/windows/contrib/2.4/e1071_1.5-16.zip` and a progress bar that is approximately 75% full.

Load Package e1071



```
R RGui
檔案 編輯 其它 程式套件 視窗 輔助
R R Console
R version 2.4.1 (2006-12-18)
Copyright (C) 2006 The R Foundation for Statistical Computing
ISBN 3-900051-07-0
R 是免費軟體，不提供任何擔保。
在某些條件下您可以將其自由散布。
用 'license()' 或 'licence()' 來獲得散布的詳細條件。
R 是個合作計劃，有許多人為之做出了貢獻。
用 'contributors()' 來看詳細的情況並且
用 'citation()' 會告訴您如何在出版品中正確地參照 R 或 R 套件。
用 'demo()' 來看一些示範程式，用 'help()' 來檢視線上輔助檔案，或
用 'help.start()' 透過 HTML 瀏覽器來看輔助檔案。
用 'q()' 離開 R。
Loading Tcl/Tk interface ... done
[Previously saved workspace restored]
> library(e1071)
Loading required package: class
>
```

Typing "library(e1071)"

R 2.4.1 - A Language and Environment

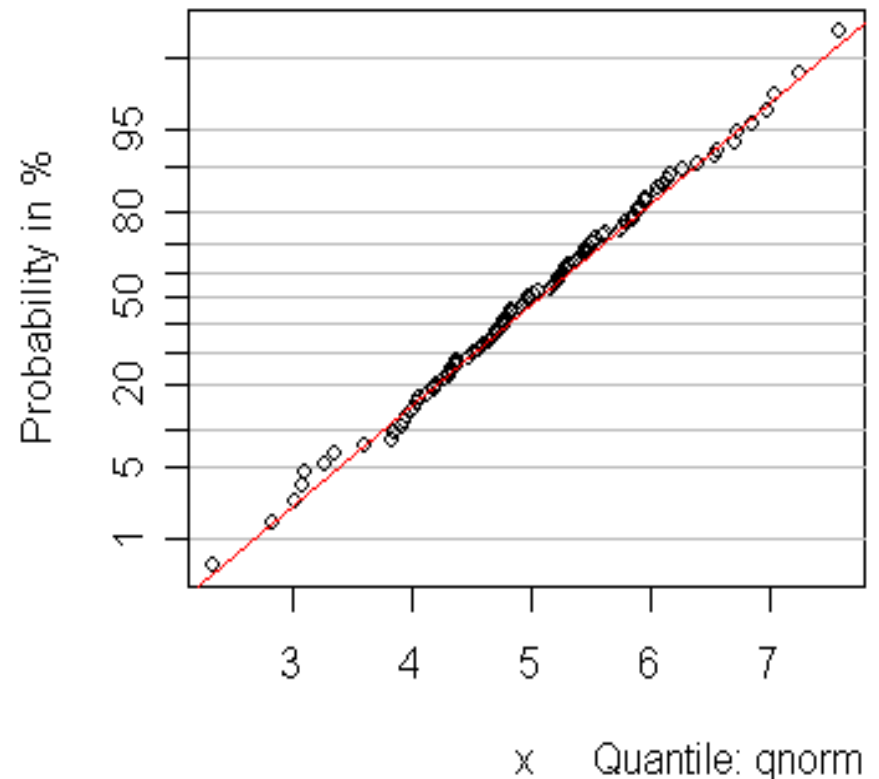
Using Probability Plot

- **Description**

- Generates a probability plot for a specified theoretical distribution, i.e., basically a qqplot where the y-axis is labeled with probabilities instead of quantiles. The function is mainly intended for teaching the concept of quantile plots.

- **Usage**

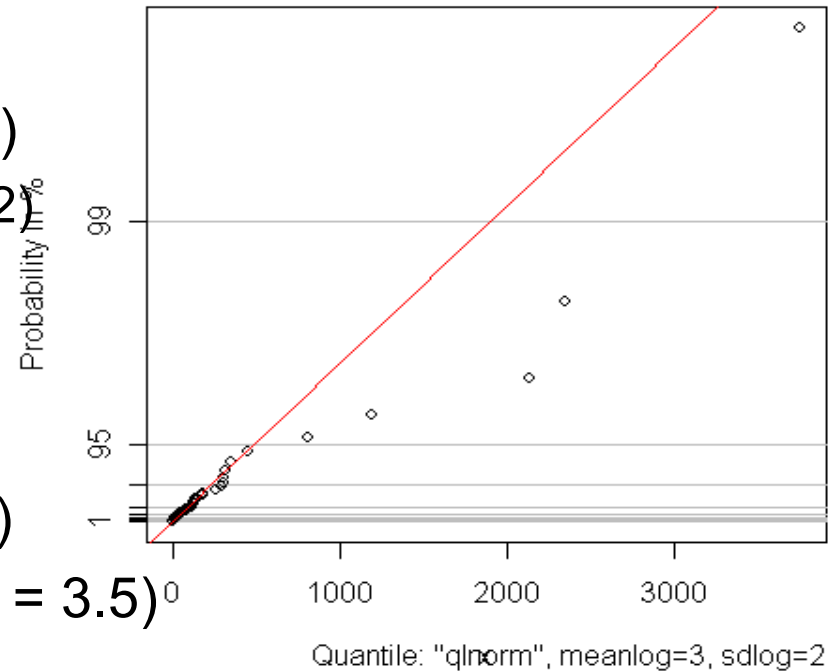
- `probplot(x,...)`
- **Example1:**
 - `x <- rnorm(100, mean=5)`
 - `probplot(x)`



More Examples

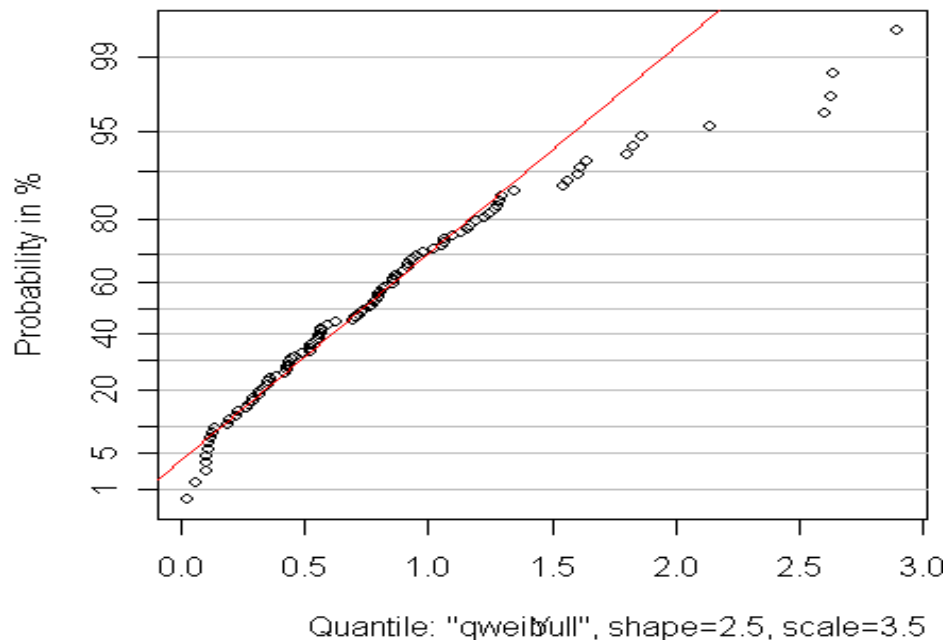
- **Example 2:**

- `x <- rlnorm(100, meanlog = 3, sdlog = 2)`
- `probplot(x, "qlnorm", meanlog = 3, sdlog = 2)`



- **Example 3:**

- `x=rweibull(100, shape=2.5, scale = 3.5)`
- `probplot(x,"qweibull", shape=2.5, scale = 3.5)`



Pareto Chart

- Load package: **library(qcc)**

- Using command:

```
pareto.chart(x, ylab = "Frequency", xlab, ylim, main,  
col = heat.colors(length(x)), ...)
```

- **Example:**

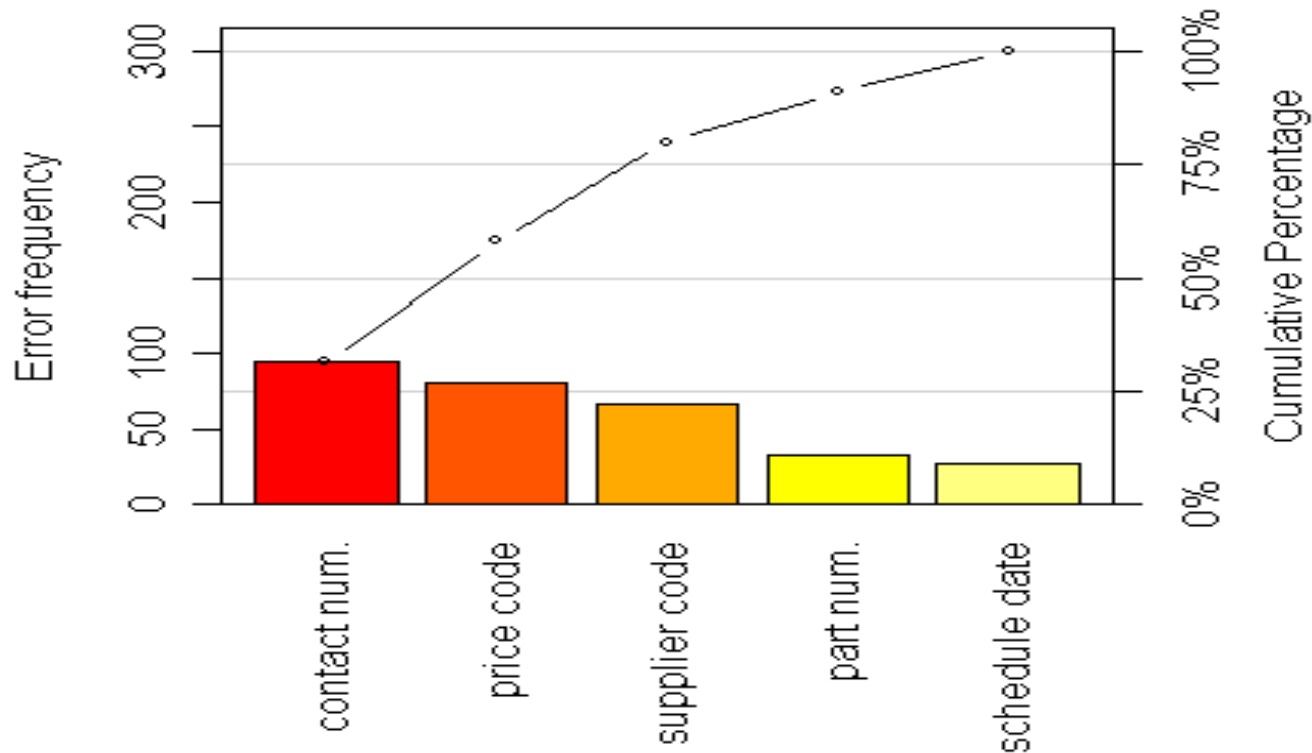
```
defect <- c(80, 27, 66, 94, 33) # Frequency  
names(defect) <- c("price code", "schedule date",  
  "supplier code", "contact num.", "part num.") # names  
pareto.chart(defect, ylab = "Error frequency")#plot
```

Pareto Chart Analysis

Pareto chart analysis for defect

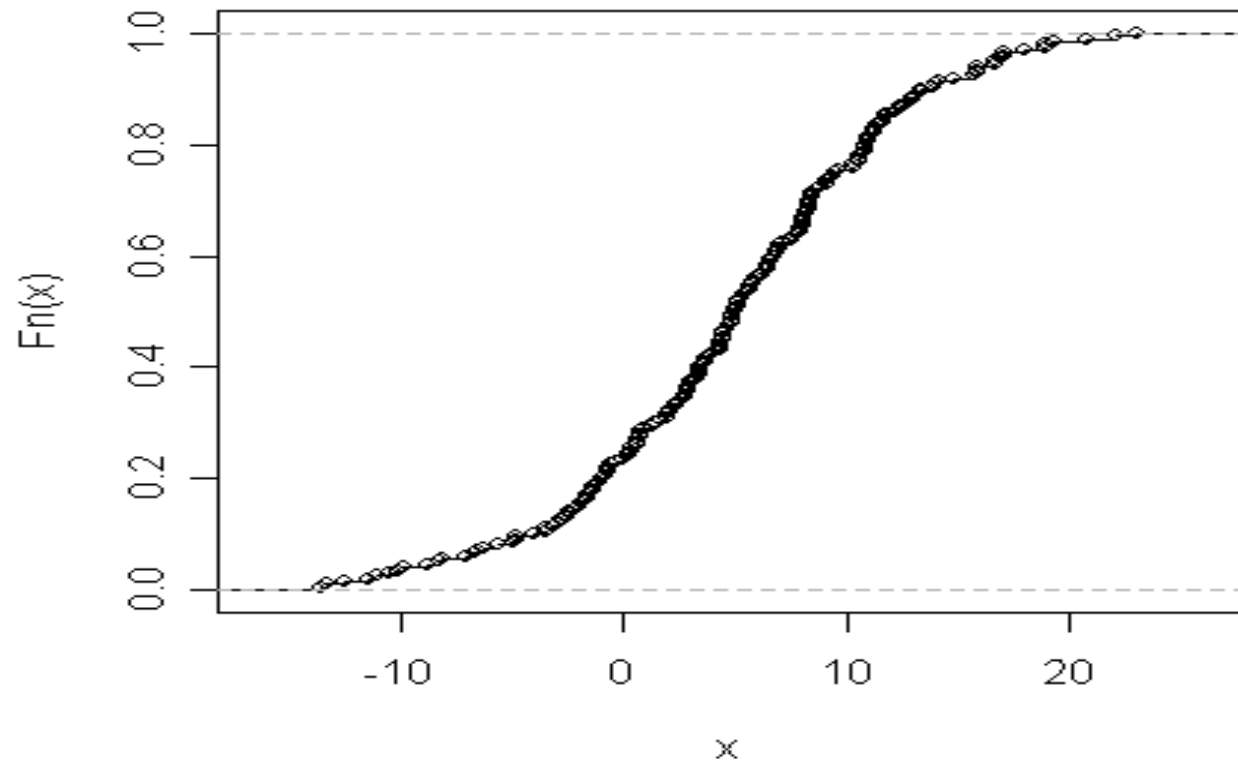
	Frequency	Cum.Freq.	Percentage	Cum.Percent.
contact num.	94	94	31.33333	31.33333
price code	80	174	26.66667	58.00000
supplier code	66	240	22.00000	80.00000
part num.	33	273	11.00000	91.00000
schedule date	27	300	9.00000	100.00000

Pareto Chart for defect



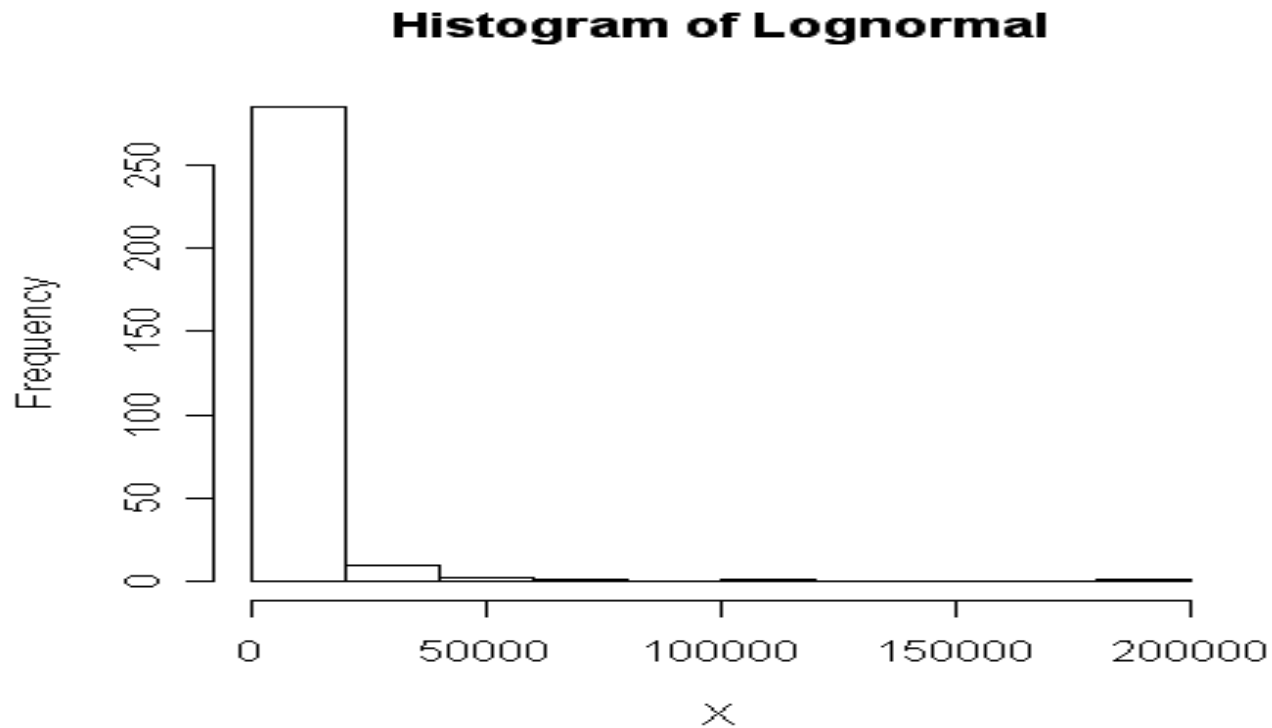
Empirical CDF

- Empirical Cumulative Distribution Function
- Usage: `ecdf(x); plot(ecdf(x),...)`
- Example:
 - `X= rnorm(200,mean=5,sd=7)`
 - `plot(ecdf(X),main= "ECDF of Normal(5,7)")`



Histogram (1)

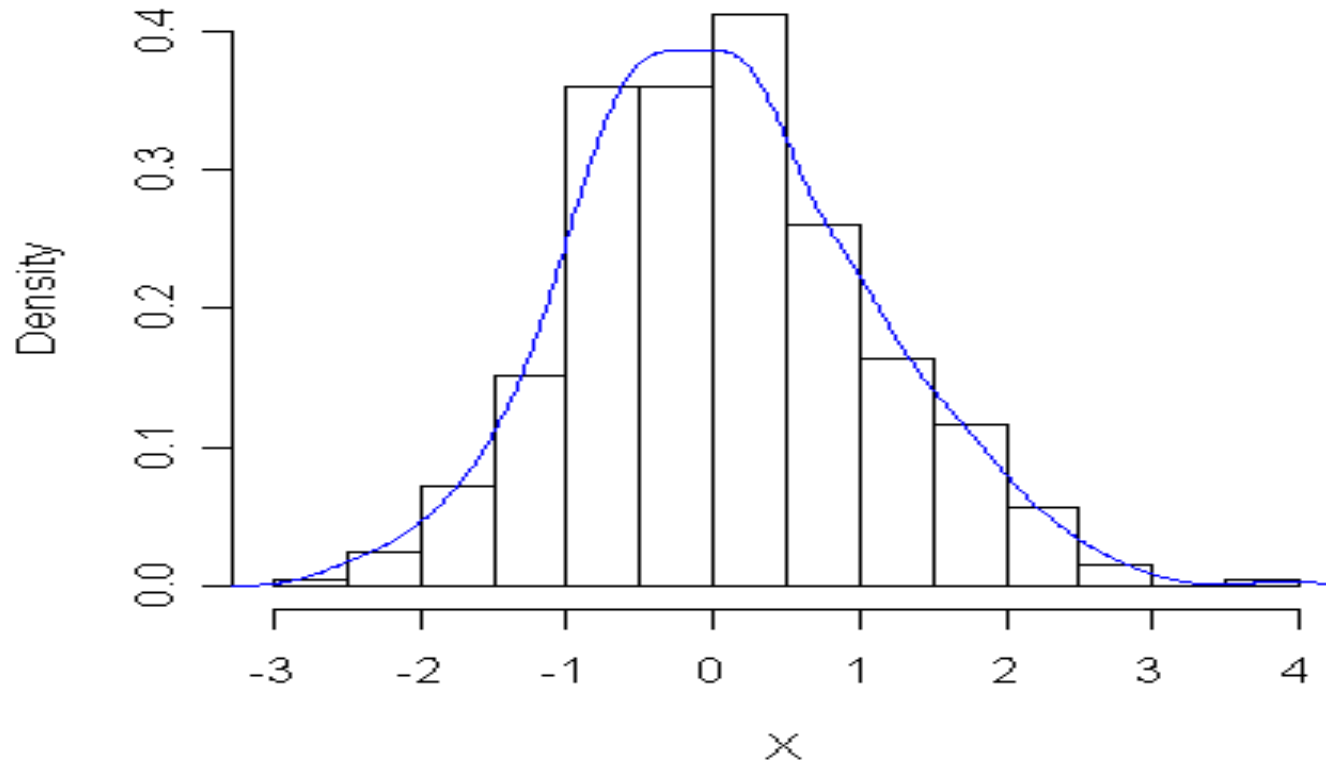
- The generic function **hist** computes a histogram of the given data values.
- Usage: **hist(x, probability = !freq, ...)**
- Example:
 - `X=rlnorm(300,lmean=5, logsd=3)`
 - `hist(X) # Using default setting`



Histogram (2)

- Example: **Density and Histogram**
 - `X=rnorm(500)`
 - `hist(X, probability=TRUE,main="Density and Histogram of Normal (0,1)")`
 - `lines(density(X),col="blue")`

Density and Histogram of Normal (0,1)



Useful Refs.

- <http://colinfay.me/intro-to-r/ordered-and-unordered-factors.html>
- https://ucdavis-bioinformatics-training.github.io/2019_August_UCD_mRNAseq_Workshop/intro2R/Intro2R
- <https://intro2r.com/install-r.html#linux-users>