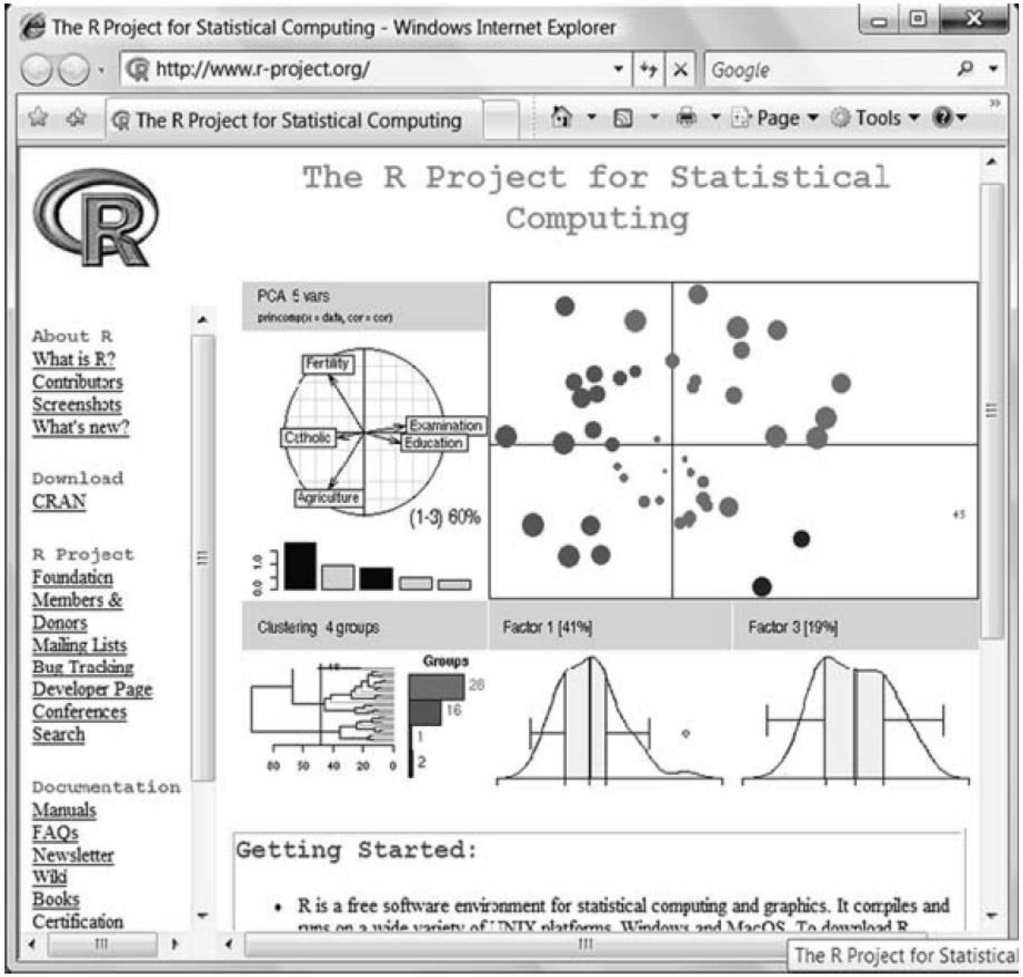
Introduction à R

L3PRO 2017

Pourquoi R?

Evolution de S (années 70), souple, open source, complet, multi-plateforme, énorme communauté…



The R Project for Statistical Computing

Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To download R

www.r-project.org

Pourquoi R?

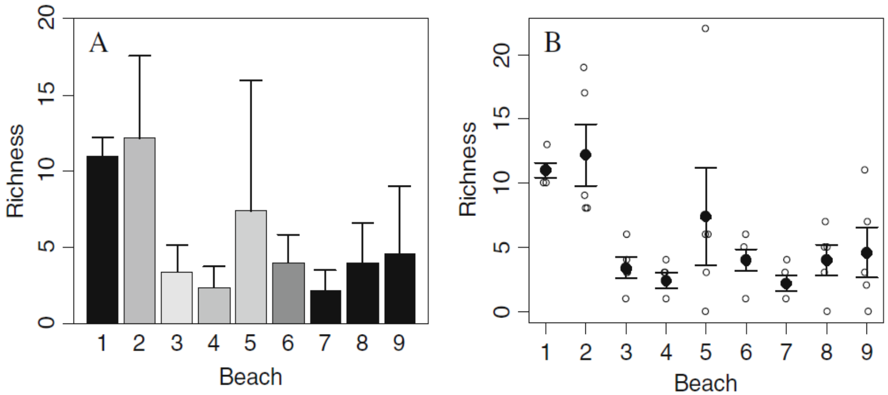


Fig. 7.3 **A**: Bar chart showing the benthic data. Mean values are represented by the *bars* with a *vertical line* showing standard deviations. The colours were changed to greyscale during the printing process. **B**: Strip chart for the raw data. The mean value per beach is plotted as a *filled dot*, and the *lines* represent the mean +/– the standard error

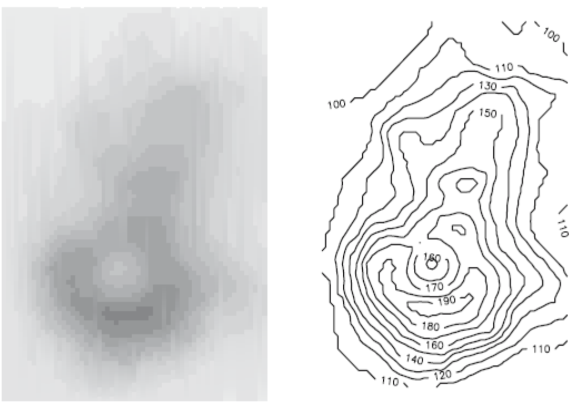


Fig. 5.1. Image plot and contour plot representations of Maunga Whau from the standard R `volcano` data set, for the same elevation class intervals (rotated to put north at the top)

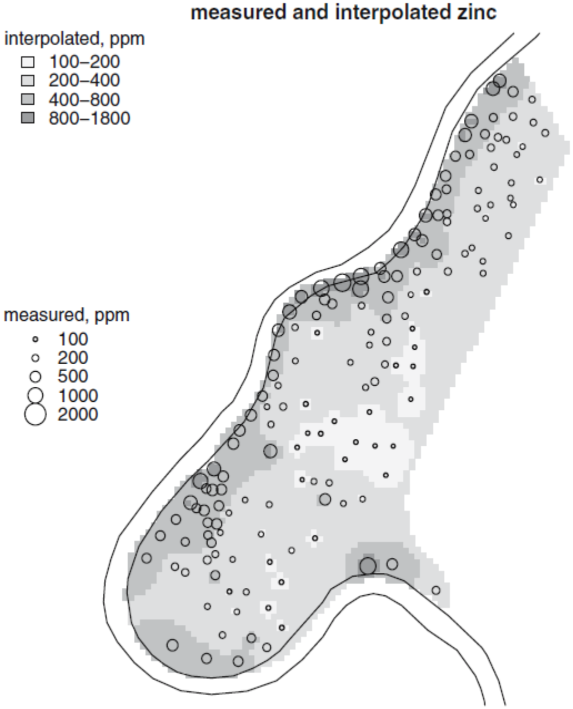
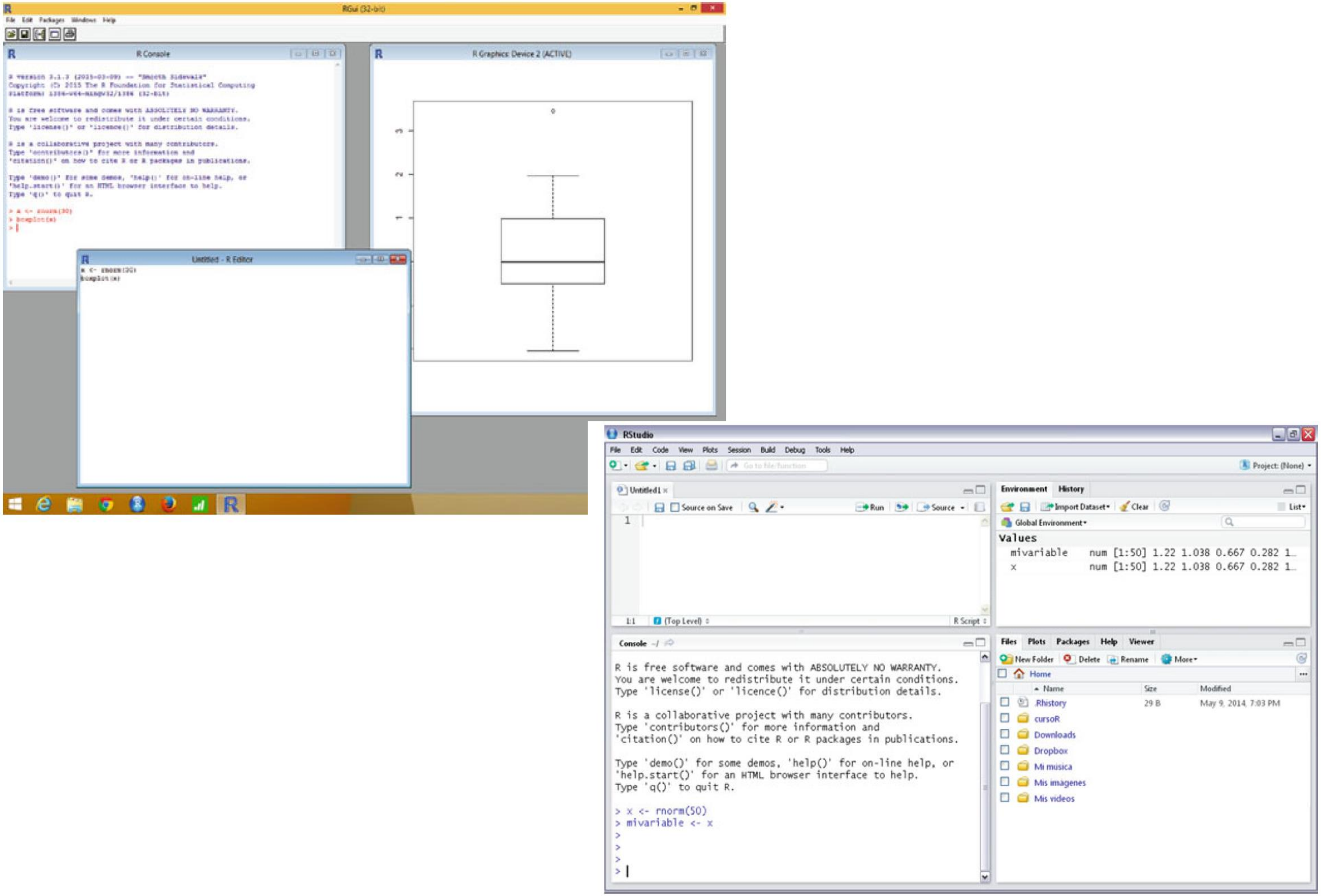


Fig. 3.8. Sample data points for zinc (ppm) plotted over an interpolated image, with symbol area proportional to measured concentration

Pourquoi R+Rstudio



```
R version 3.1.3 (2015-03-09) -- "Smooth Sidewalk"
Copyright (C) 2015 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.

> x <- rnorm(30)
> boxplot(x)
> 
```

```
x <- rnorm(30)
boxplot(x)
```

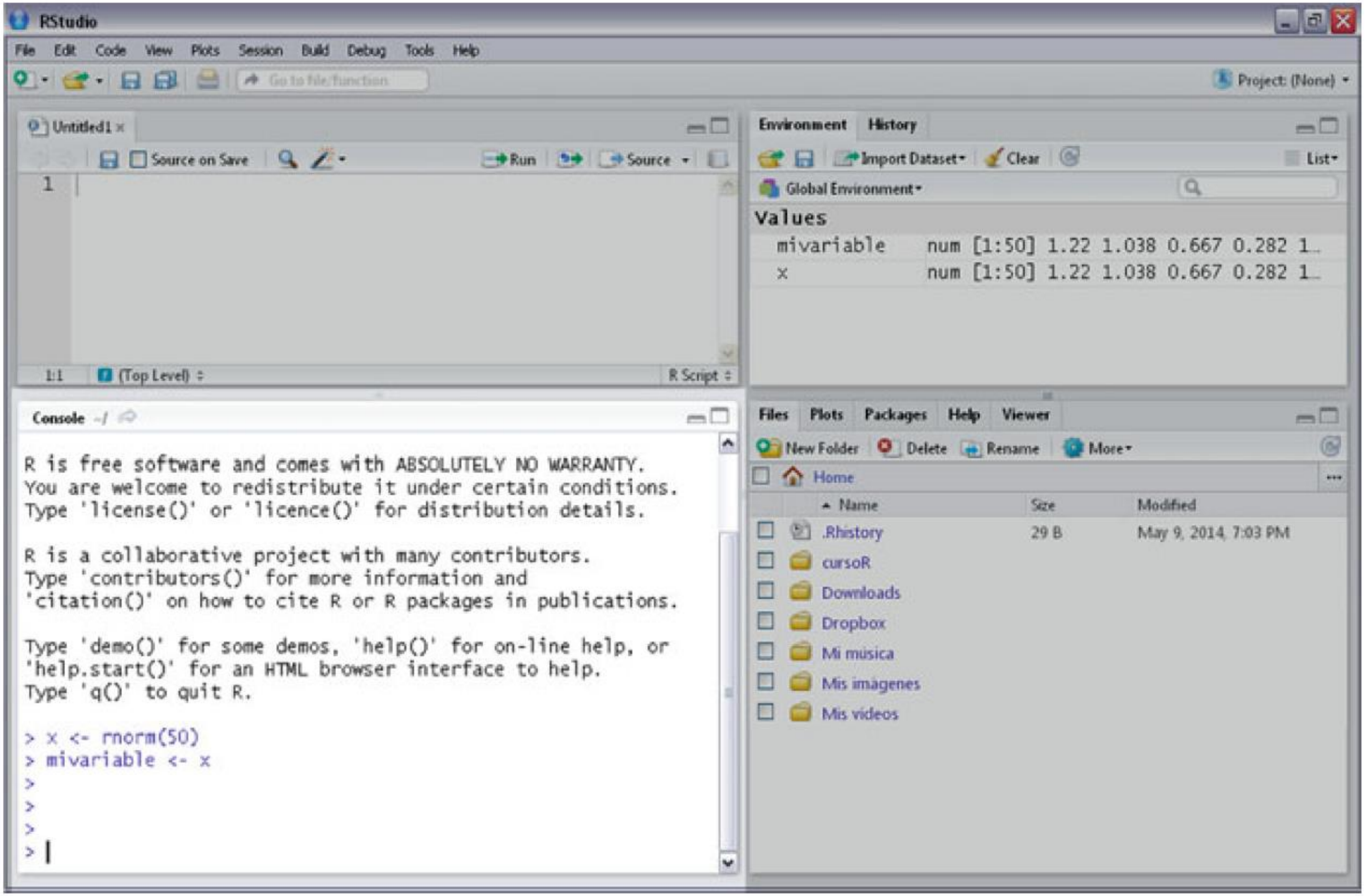
```
Values
  mivariable    num [1:50] 1.22 1.038 0.667 0.282 1…
  x             num [1:50] 1.22 1.038 0.667 0.282 1…
```

Name	Size	Modified
☐ .Rhistory	29 B	May 9, 2014, 7:03 PM
☐ cursoR		
☐ Downloads		
☐ Dropbox		
☐ Mi musica		
☐ Mis imagenes		
☐ Mis videos		

```
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.

> x <- rnorm(50)
> mivariable <- x
>
>
>
> |
```

www.r-project.org

La console

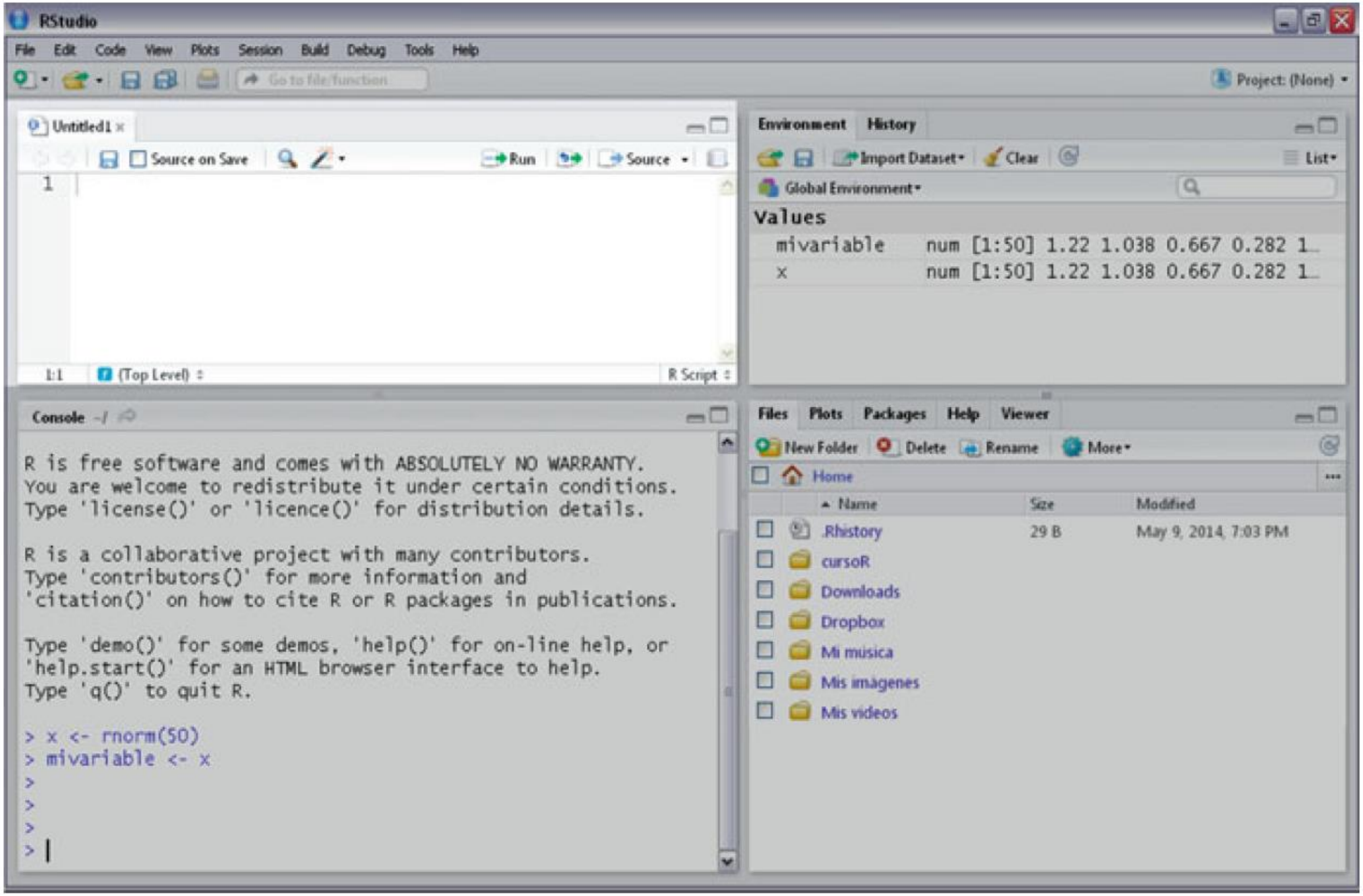


```
R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

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.

> x <- rnorm(50)
> mivariable <- x
>
>
>
> |
```

```
Values
  mivariable     num [1:50] 1.22 1.038 0.667 0.282 1...
  x              num [1:50] 1.22 1.038 0.667 0.282 1...
```

☐	Name	Size	Modified
☐	.Rhistory	29 B	May 9, 2014, 7:03 PM
☐	cursoR		
☐	Downloads		
☐	Dropbox		
☐	Mi música		
☐	Mis imágenes		
☐	Mis videos		

Le script

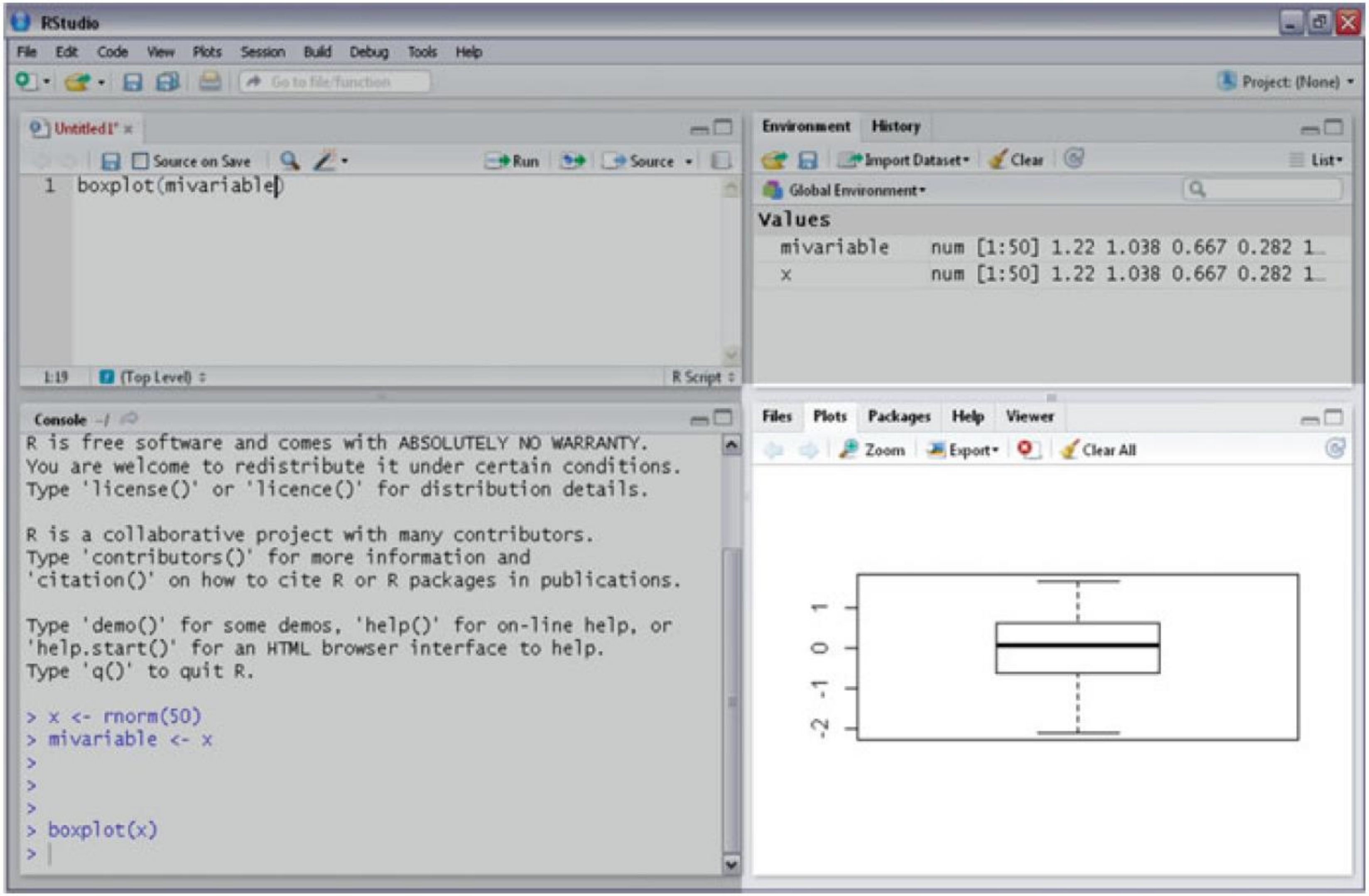


```
Values
  mivariable      num [1:50] 1.22 1.038 0.667 0.282 1…
  x               num [1:50] 1.22 1.038 0.667 0.282 1…
```

```
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.

> x <- rnorm(50)
> mivariable <- x
>
>
>
> |
```

☐	Name	Size	Modified
☐	.Rhistory	29 B	May 9, 2014, 7:03 PM
☐	cursoR		
☐	Downloads		
☐	Dropbox		
☐	Mi musica		
☐	Mis imagenes		
☐	Mis videos		

La fenêtre Plot

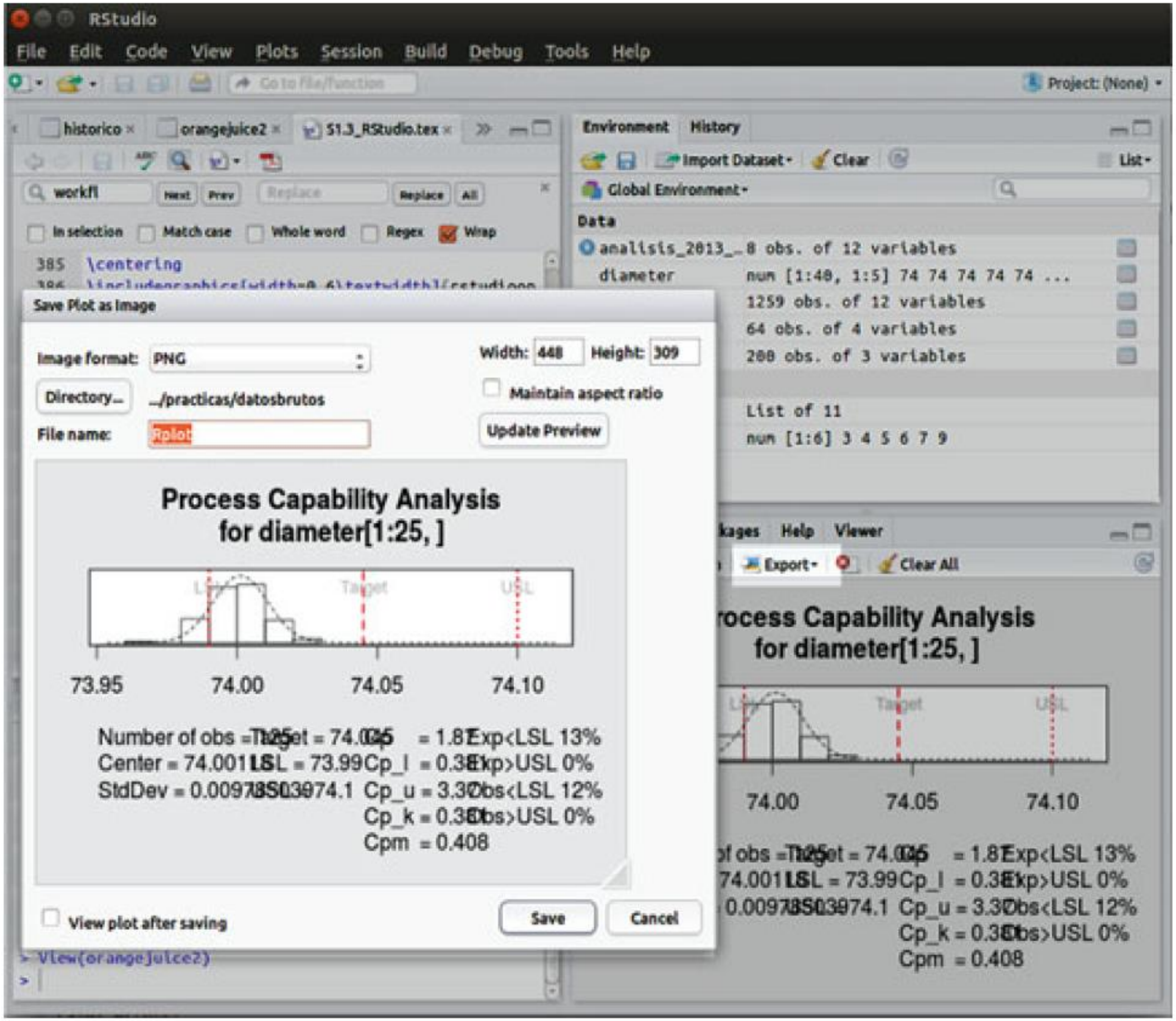


```
boxplot(mivariable)
```

```
Values
  mivariable      num [1:50] 1.22 1.038 0.667 0.282 1...
  x               num [1:50] 1.22 1.038 0.667 0.282 1...
```

```
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.

> x <- rnorm(50)
> mivariable <- x
>
>
>
> boxplot(x)
>
```

Sauver une image



workfl

☐ In selection ☐ Match case ☐ Whole word ☐ Regex ☑ Wrap

```
385 \centering
386 \includegraphics[width=0.6\textwidth]{rstudioop
```

Data

```
analisis_2013_...  8 obs. of 12 variables
diameter           num [1:40, 1:5] 74 74 74 74 74 ...
                   1259 obs. of 12 variables
                   64 obs. of 4 variables
                   200 obs. of 3 variables

                   List of 11
                   num [1:6] 3 4 5 6 7 9
```

Save Plot as Image

Image format: PNG

Width: 448 Height: 309

☐ Maintain aspect ratio

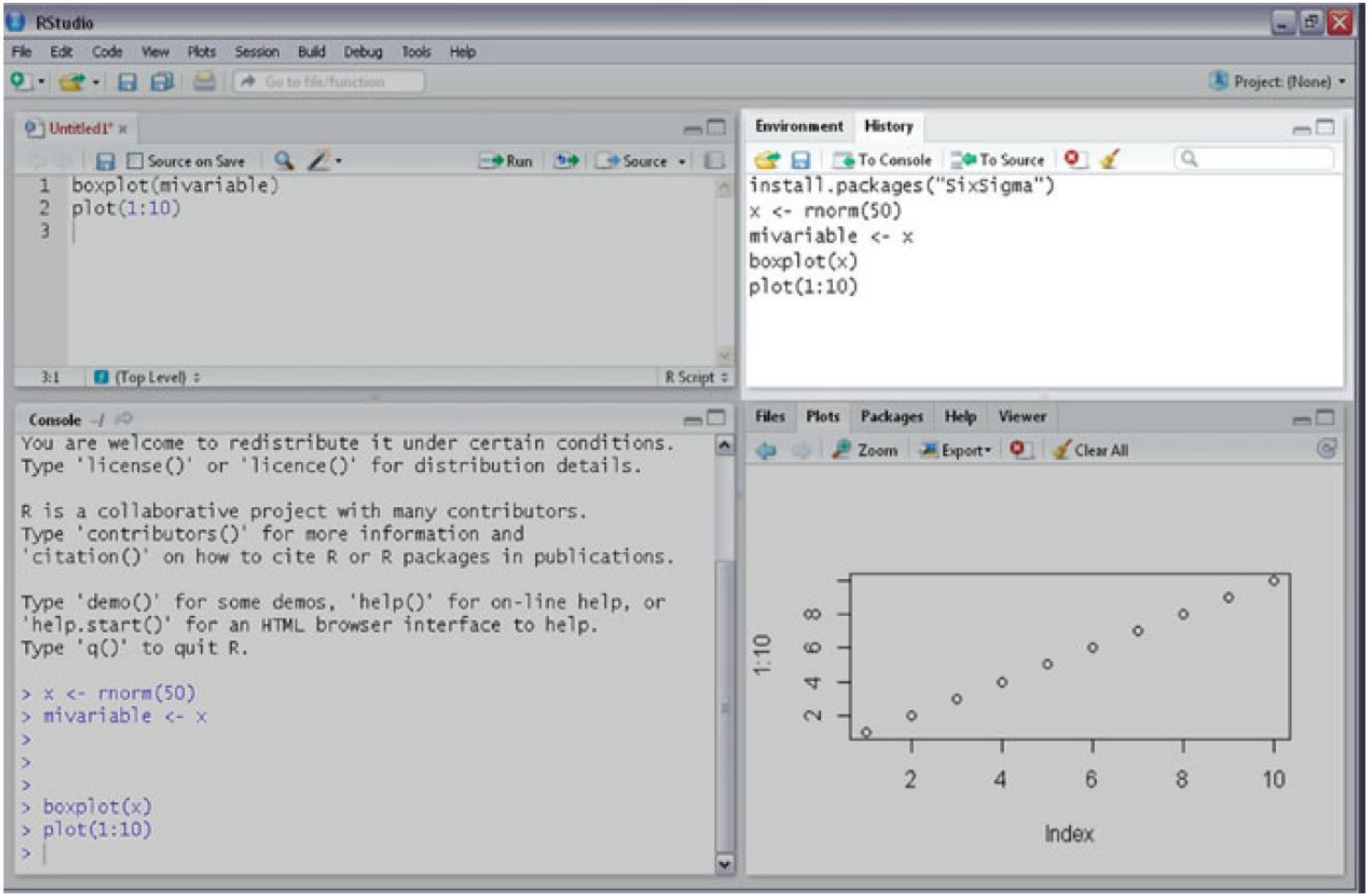
Directory... ../practicas/datosbrutos

File name: Rplot

☐ View plot after saving

```
> View(orangejuice2)
>
```

L'historique

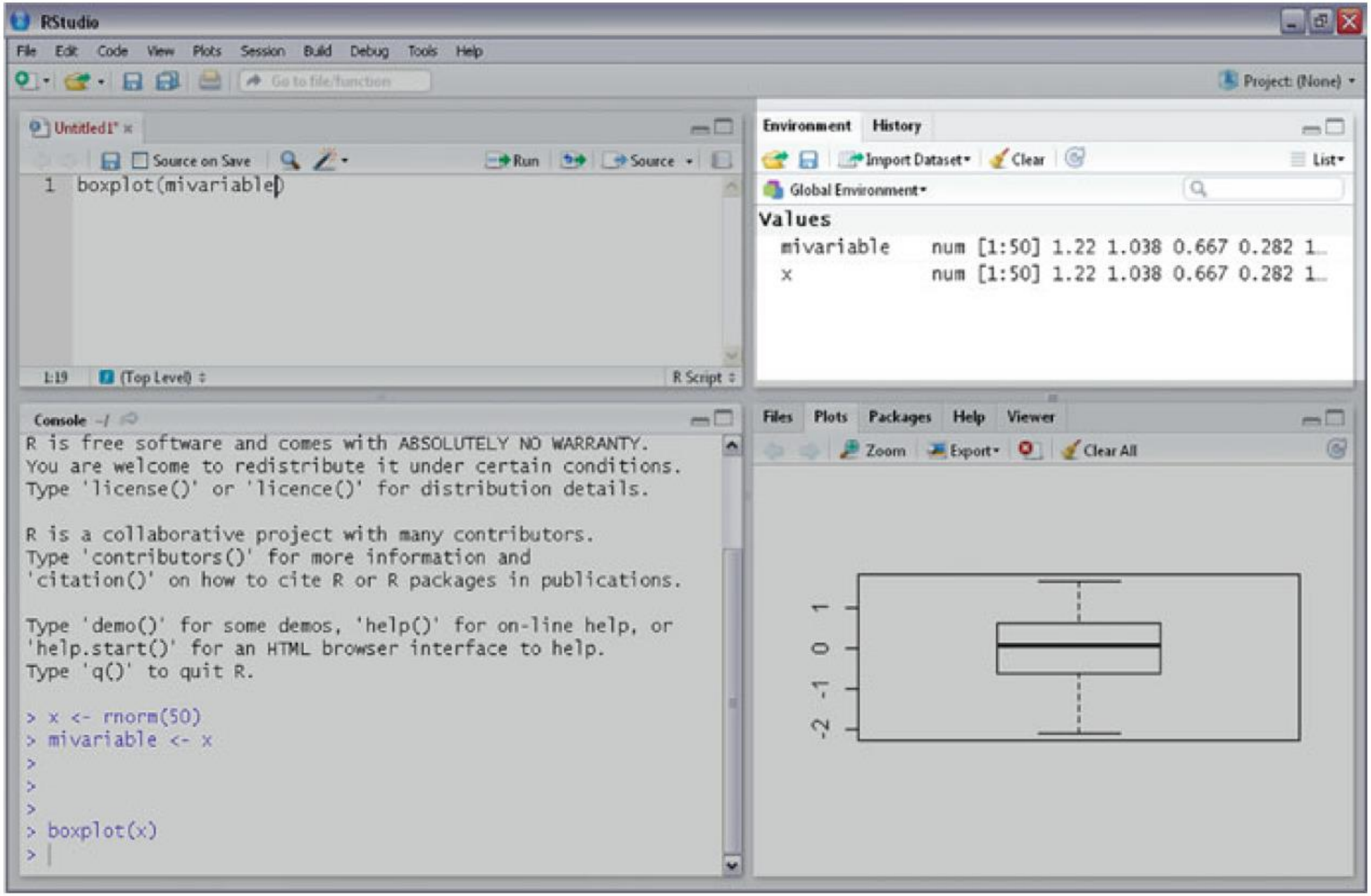


```
boxplot(mivariable)
plot(1:10)
```

```
install.packages("SixSigma")
x <- rnorm(50)
mivariable <- x
boxplot(x)
plot(1:10)
```

```
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.

> x <- rnorm(50)
> mivariable <- x
>
>
>
> boxplot(x)
> plot(1:10)
>
```

Les variables

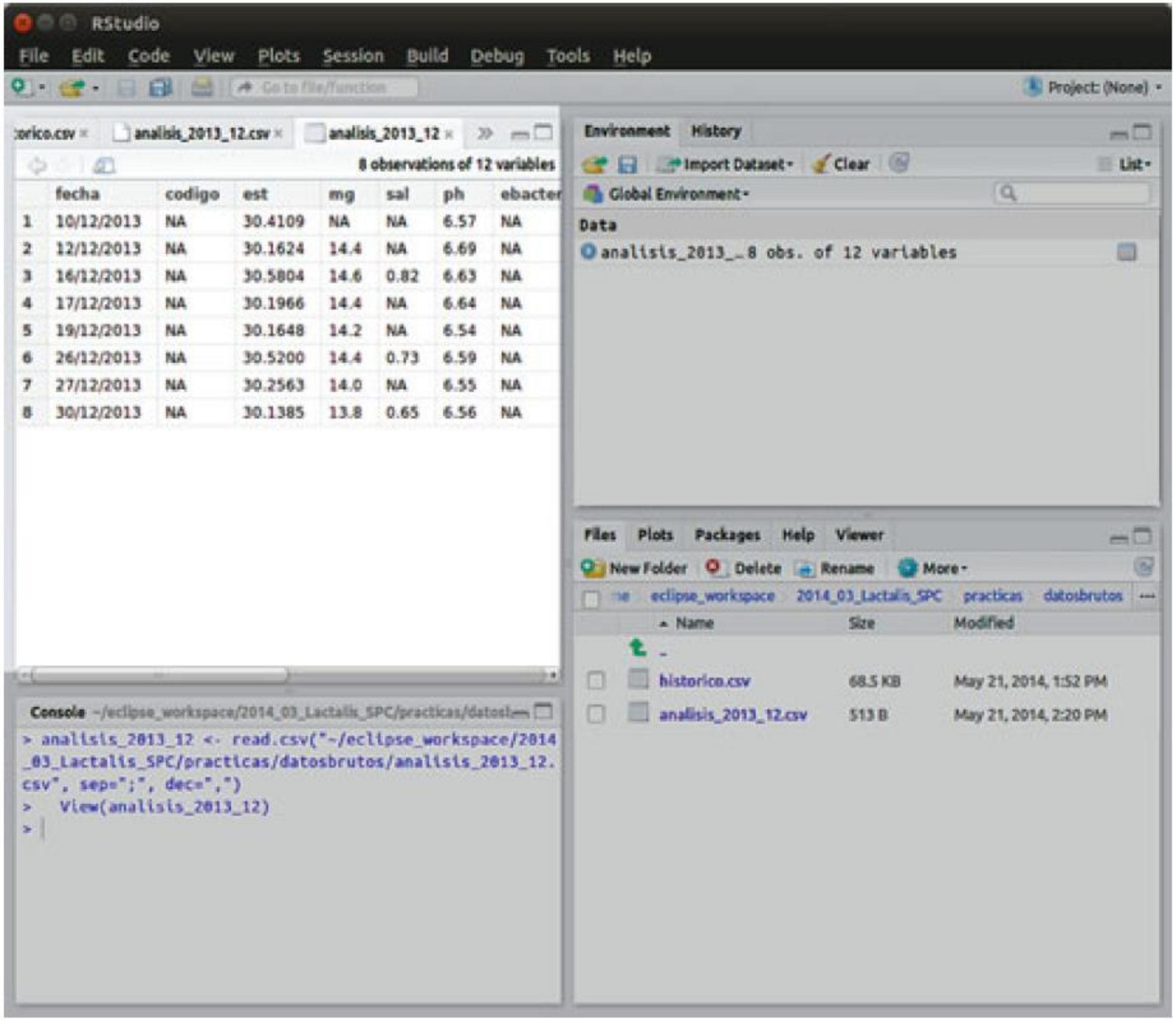


```
boxplot(mivariable)
```

```
Values
  mivariable     num [1:50] 1.22 1.038 0.667 0.282 1...
  x              num [1:50] 1.22 1.038 0.667 0.282 1...
```

```
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.

> x <- rnorm(50)
> mivariable <- x
>
>
>
> boxplot(x)
>
```

Les variables


RStudio
File Edit Code View Plots Session Build Debug Tools Help
Go to file/function
Project: (None)
orico.csv
analisis_2013_12.csv
analisis_2013_12
8 observations of 12 variables
Environment History
Import Dataset
Clear
List
Global Environment
Data
Files Plots Packages Help Viewer
New Folder Delete Rename More
eclipse_workspace 2014_03_Lactalis_SPC practicas datosbrutos
Console ~/eclipse_workspace/2014_03_Lactalis_SPC/practicas/datosb

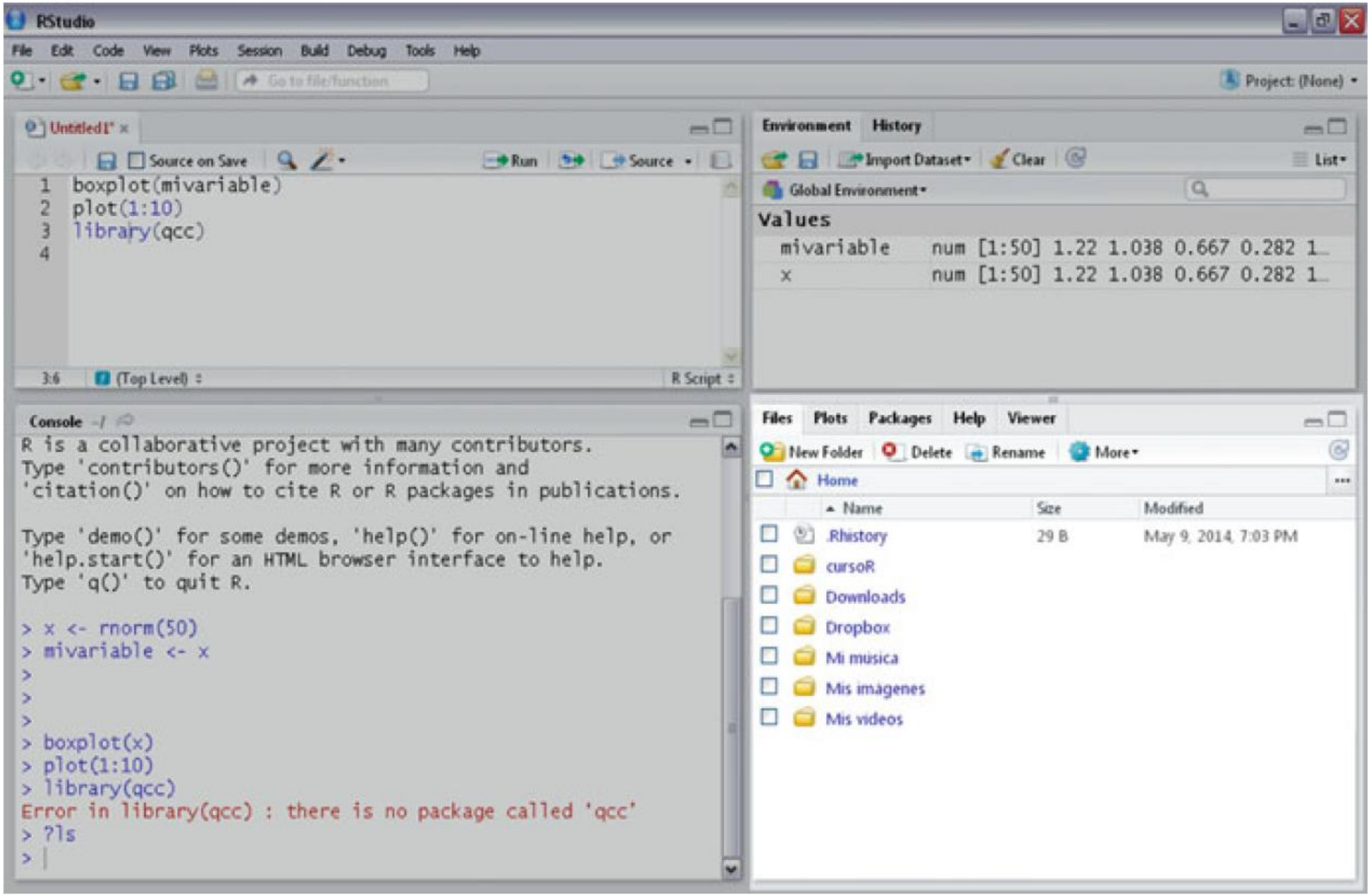
	fecha	codigo	est	mg	sal	ph	ebacter
1	10/12/2013	NA	30.4109	NA	NA	6.57	NA
2	12/12/2013	NA	30.1624	14.4	NA	6.69	NA
3	16/12/2013	NA	30.5804	14.6	0.82	6.63	NA
4	17/12/2013	NA	30.1966	14.4	NA	6.64	NA
5	19/12/2013	NA	30.1648	14.2	NA	6.54	NA
6	26/12/2013	NA	30.5200	14.4	0.73	6.59	NA
7	27/12/2013	NA	30.2563	14.0	NA	6.55	NA
8	30/12/2013	NA	30.1385	13.8	0.65	6.56	NA

```
analisis_2013_...8 obs. of 12 variables
```

Name	Size	Modified
..		
historico.csv	68.5 KB	May 21, 2014, 1:52 PM
analisis_2013_12.csv	513 B	May 21, 2014, 2:20 PM

```
> analisis_2013_12 <- read.csv("~/eclipse_workspace/2014
_03_Lactalis_SPC/practicas/datosbrutos/analisis_2013_12.
csv", sep=";", dec=",")
>   View(analisis_2013_12)
>
```

Le répertoire de travail



```
boxplot(mivariable)
plot(1:10)
library(qcc)
```

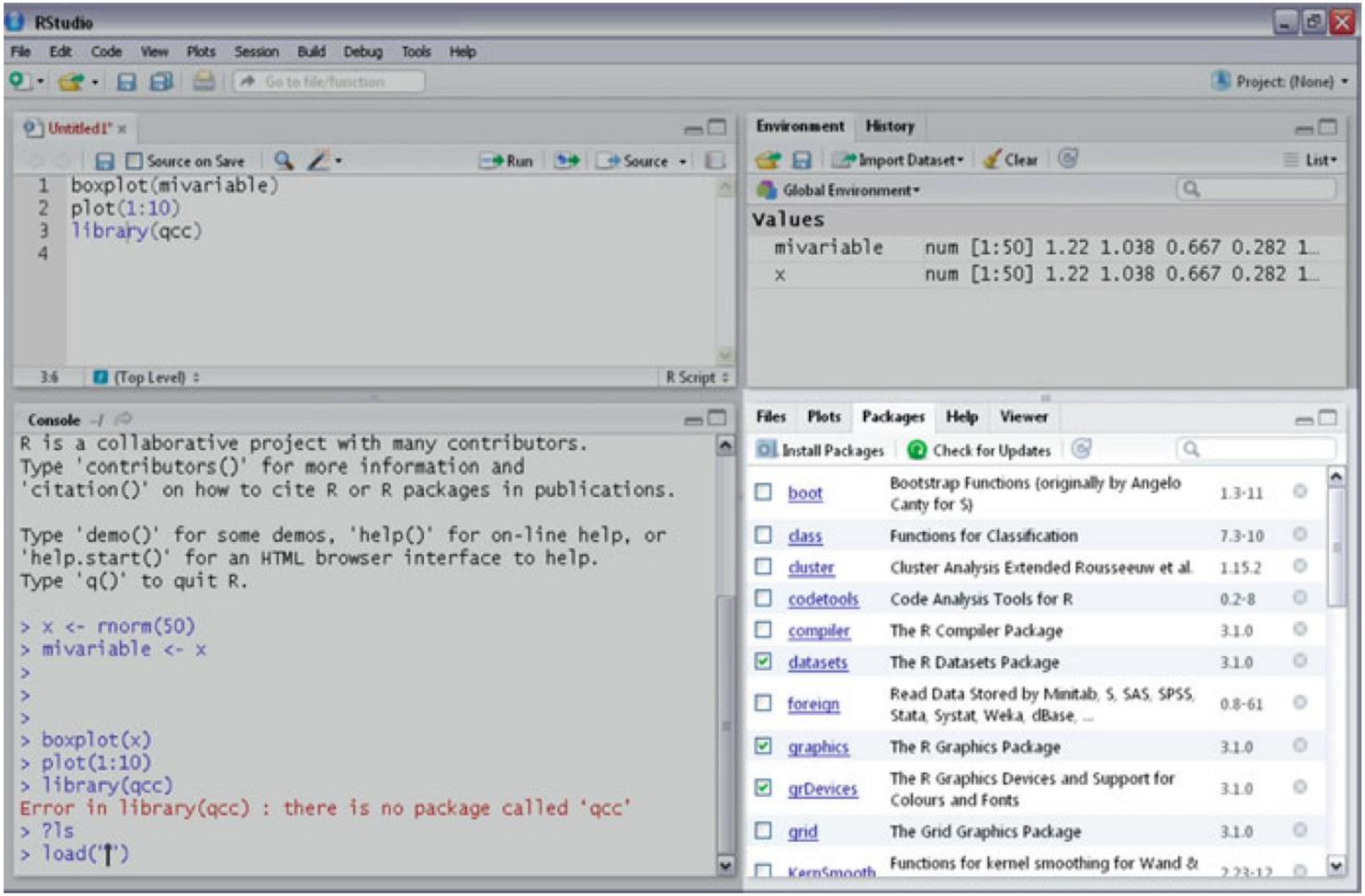
```
Values
  mivariable    num [1:50] 1.22 1.038 0.667 0.282 1…
  x             num [1:50] 1.22 1.038 0.667 0.282 1…
```

```
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.

> x <- rnorm(50)
> mivariable <- x
>
>
>
> boxplot(x)
> plot(1:10)
> library(qcc)
Error in library(qcc) : there is no package called 'qcc'
> ?ls
> 
```

▲ Name	Size	Modified
□ .Rhistory	29 B	May 9, 2014, 7:03 PM
□ cursoR		
□ Downloads		
□ Dropbox		
□ Mi musica		
□ Mis imagenes		
□ Mis videos		

```
## Correct:
setwd("C:/myscripts")
setwd("C:\\myscripts")

## Incorrect:
setwd("C:\myscripts")
```

Les packages



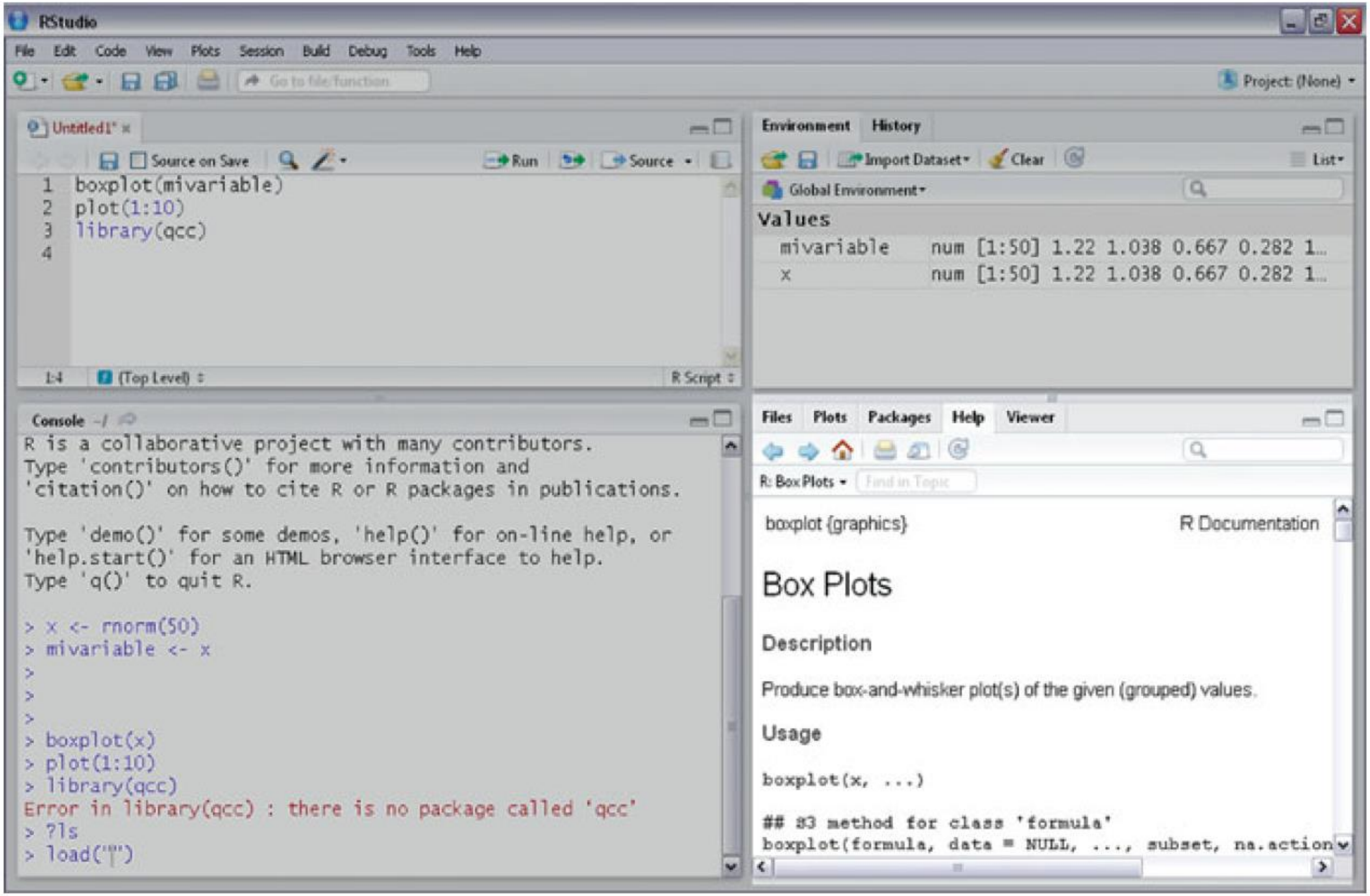
```
boxplot(mivariable)
plot(1:10)
library(qcc)
```

```
Values
  mivariable     num [1:50] 1.22 1.038 0.667 0.282 1...
  x              num [1:50] 1.22 1.038 0.667 0.282 1...
```

```
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.

> x <- rnorm(50)
> mivariable <- x
>
>
>
> boxplot(x)
> plot(1:10)
> library(qcc)
Error in library(qcc) : there is no package called ‘qcc’
> ?ls
> load("")
```

□	boot	Bootstrap Functions (originally by Angelo Canty for S)	1.3-11	○
□	class	Functions for Classification	7.3-10	○
□	cluster	Cluster Analysis Extended Rousseeuw et al.	1.15.2	○
□	codetools	Code Analysis Tools for R	0.2-8	○
□	compiler	The R Compiler Package	3.1.0	○
☑	datasets	The R Datasets Package	3.1.0	○
□	foreign	Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase, ...	0.8-61	○
☑	graphics	The R Graphics Package	3.1.0	○
☑	grDevices	The R Graphics Devices and Support for Colours and Fonts	3.1.0	○
□	grid	The Grid Graphics Package	3.1.0	○
□	KernSmooth	Functions for kernel smoothing for Wand &	2.23-12	○

Les packages



```
boxplot(mivariable)
plot(1:10)
library(qcc)
```

```
Values
  mivariable    num [1:50] 1.22 1.038 0.667 0.282 1...
  x             num [1:50] 1.22 1.038 0.667 0.282 1...
```

```
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.

> x <- rnorm(50)
> mivariable <- x
>
>
>
> boxplot(x)
> plot(1:10)
> library(qcc)
Error in library(qcc) : there is no package called 'qcc'
> ?ls
> load("")
```

boxplot {graphics} R Documentation

Box Plots

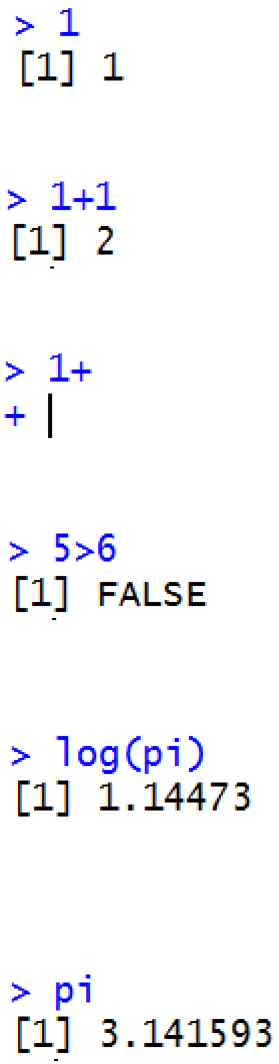
Description

Produce box-and-whisker plot(s) of the given (grouped) values.

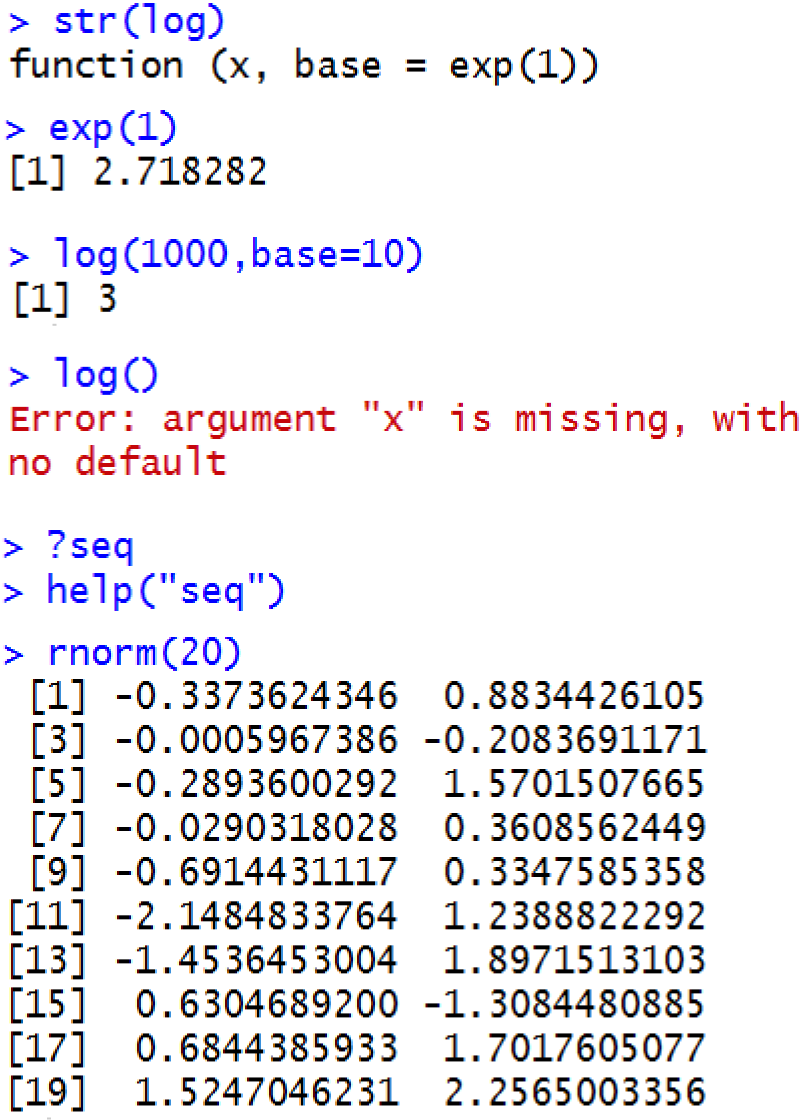
Usage

```
boxplot(x, ...)

## S3 method for class 'formula'
boxplot(formula, data = NULL, ..., subset, na.action
```

Les expressions – quelques exemples

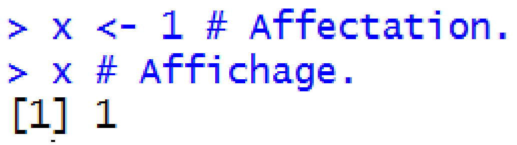


```
> 1
[1] 1

> 1+1
[1] 2

> 1+
+ |

> 5>6
[1] FALSE

> log(pi)
[1] 1.14473

> pi
[1] 3.141593
```

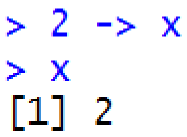


```
> str(log)
function (x, base = exp(1))

> exp(1)
[1] 2.718282

> log(1000,base=10)
[1] 3

> log()
Error: argument "x" is missing, with
no default

> ?seq
> help("seq")

> rnorm(20)
 [1] -0.3373624346  0.8834426105
 [3] -0.0005967386 -0.2083691171
 [5] -0.2893600292  1.5701507665
 [7] -0.0290318028  0.3608562449
 [9] -0.6914431117  0.3347585358
[11] -2.1484833764  1.2388822292
[13] -1.4536453004  1.8971513103
[15]  0.6304689200 -1.3084480885
[17]  0.6844385933  1.7017605077
[19]  1.5247046231  2.2565003356
```

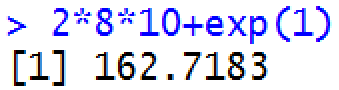
Les expressions – quelques exemples



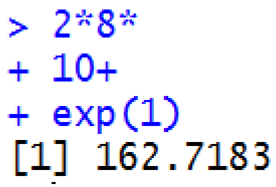
```
> x <- 1 # Affectation.
> x # Affichage.
[1] 1
```



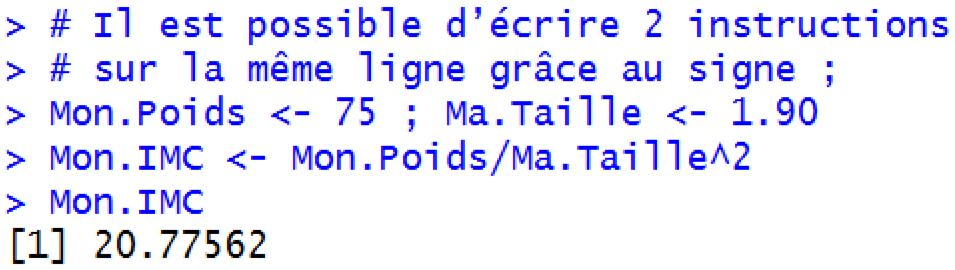
```
> 2 -> x
> x
[1] 2
```



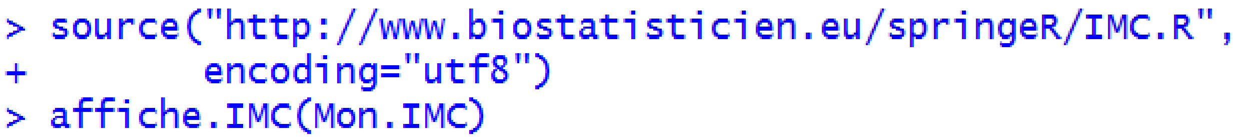
```
> 2*8*10+exp(1)
[1] 162.7183
```



```
> 2*8*
+ 10+
+ exp(1)
[1] 162.7183
```



```
> # Il est possible d’écrire 2 instructions
> # sur la même ligne grâce au signe ;
> Mon.Poids <- 75 ; Ma.Taille <- 1.90
> Mon.IMC <- Mon.Poids/Ma.Taille^2
> Mon.IMC
[1] 20.77562
```

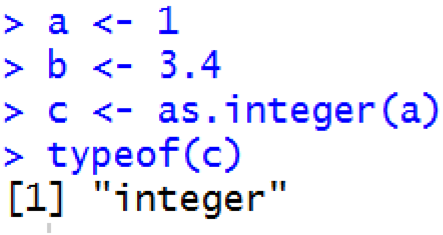


```
> source("http://www.biostatisticien.eu/springeR/IMC.R",
+         encoding="utf8")
> affiche.IMC(Mon.IMC)
```

Les principaux types

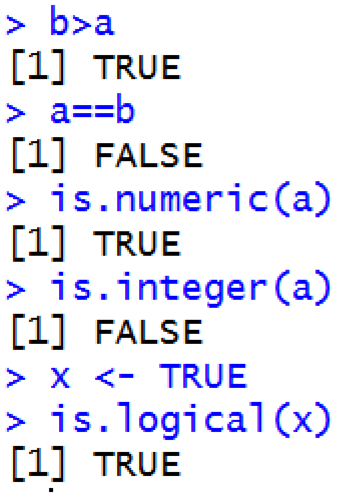
Type numerique (*numeric*)

Il y a deux types numeriques : les entiers (integer) et les reels (real ou double).

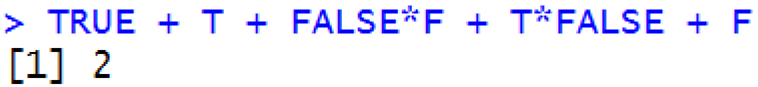


```
> a <- 1
> b <- 3.4
> c <- as.integer(a)
> typeof(c)
[1] "integer"
```

Type booleen ou logique (*logical*)



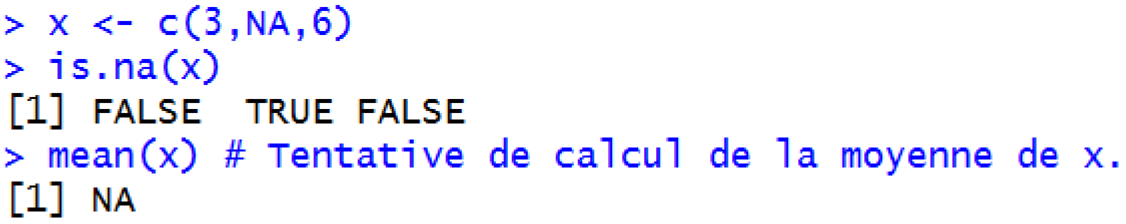
```
> b>a
[1] TRUE
> a==b
[1] FALSE
> is.numeric(a)
[1] TRUE
> is.integer(a)
[1] FALSE
> x <- TRUE
> is.logical(x)
[1] TRUE
```



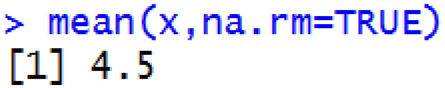
```
> TRUE + T + FALSE*F + T*FALSE + F
[1] 2
```

Les principaux types

Données manquantes (*NA*)

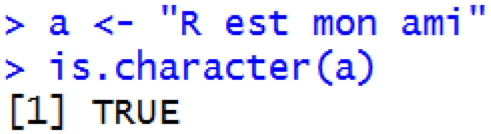


```
> x <- c(3,NA,6)
> is.na(x)
[1] FALSE  TRUE FALSE
> mean(x) # Tentative de calcul de la moyenne de x.
[1] NA
```



```
> mean(x,na.rm=TRUE)
[1] 4.5
```

Type chaînes de caracteres (*character*)

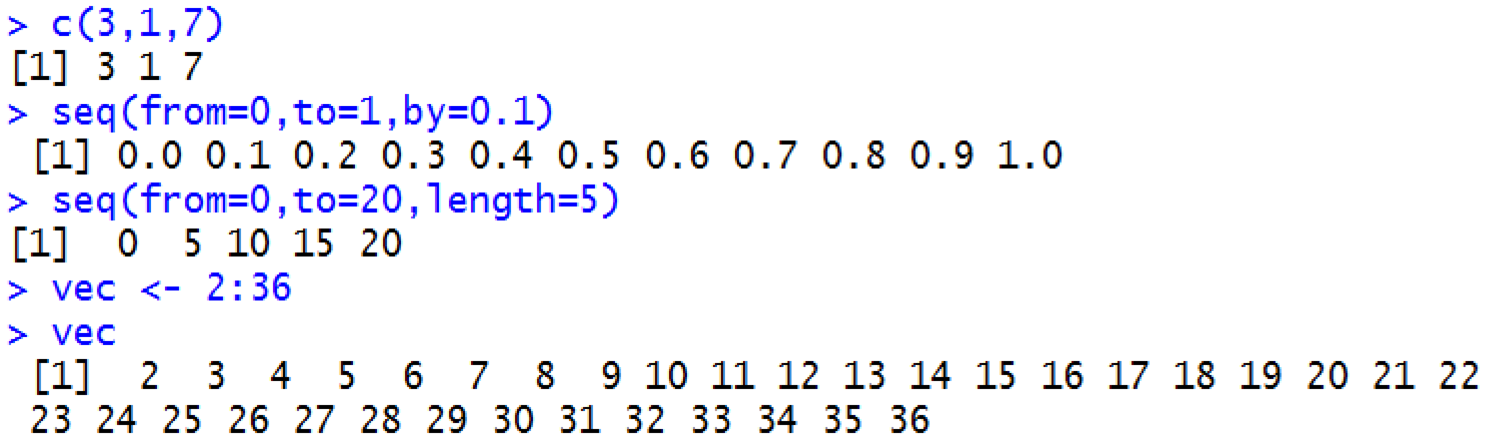


```
> a <- "R est mon ami"
> is.character(a)
[1] TRUE
```

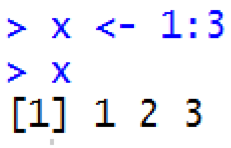
La structure des données

Les vecteurs (*vector*)

Cette structure de données est la plus simple. Elle représente une suite de données de même type.



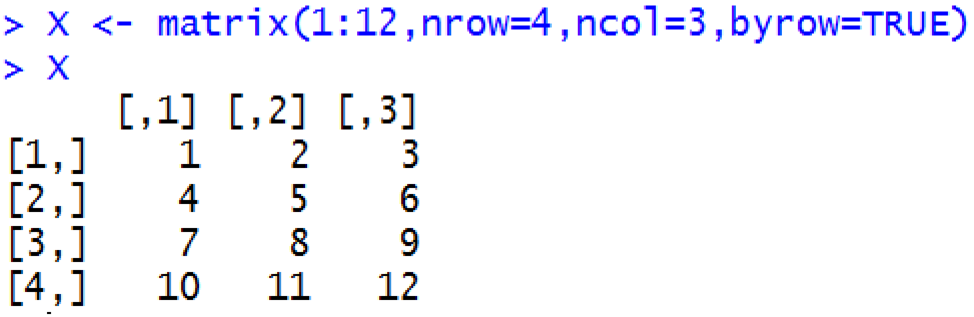
```
> c(3,1,7)
[1] 3 1 7
> seq(from=0,to=1,by=0.1)
 [1] 0.0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1.0
> seq(from=0,to=20,length=5)
[1]  0  5 10 15 20
> vec <- 2:36
> vec
 [1]  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22
23 24 25 26 27 28 29 30 31 32 33 34 35 36
```



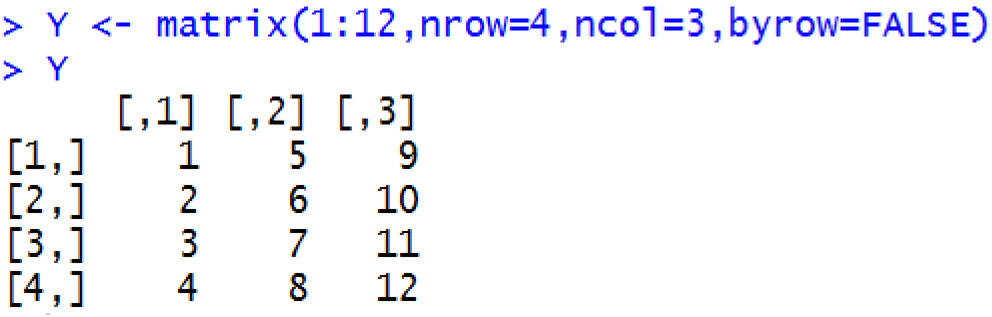
```
> x <- 1:3
> x
[1] 1 2 3
```

Les matrices (*matrix*) et les tableaux (*arrays*)

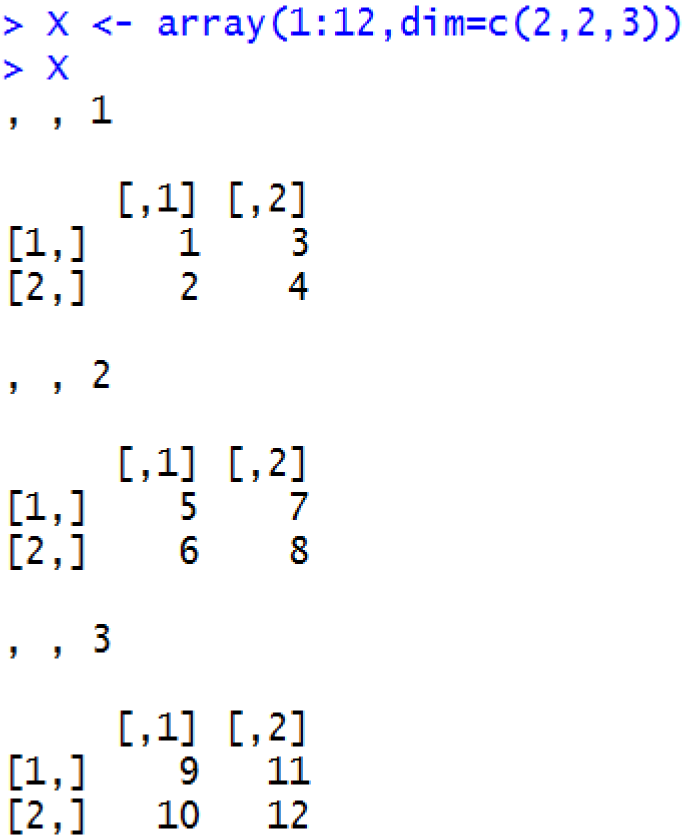
Généralisation de la notion de vecteur



```
> X <- matrix(1:12,nrow=4,ncol=3,byrow=TRUE)
> X
     [,1] [,2] [,3]
[1,]    1    2    3
[2,]    4    5    6
[3,]    7    8    9
[4,]   10   11   12
```



```
> Y <- matrix(1:12,nrow=4,ncol=3,byrow=FALSE)
> Y
     [,1] [,2] [,3]
[1,]    1    5    9
[2,]    2    6   10
[3,]    3    7   11
[4,]    4    8   12
```

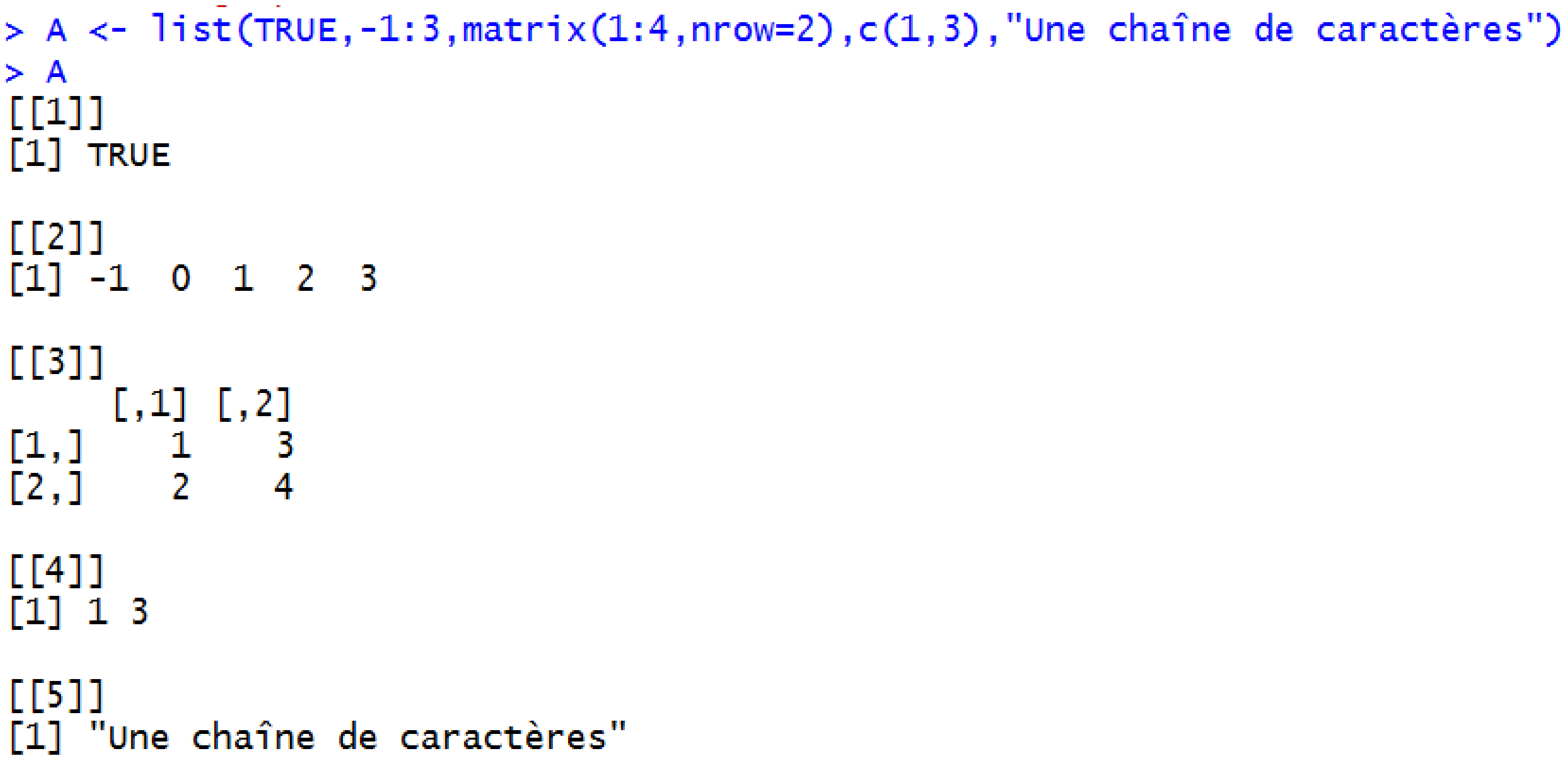


```
> X <- array(1:12,dim=c(2,2,3))
> X
, , 1

     [,1] [,2]
[1,]    1    3
[2,]    2    4

, , 2

     [,1] [,2]
[1,]    5    7
[2,]    6    8

, , 3

     [,1] [,2]
[1,]    9   11
[2,]   10   12
```

La structure des données

Les listes (*list*)

les listes permettent de regrouper dans une même structure des données de types différents

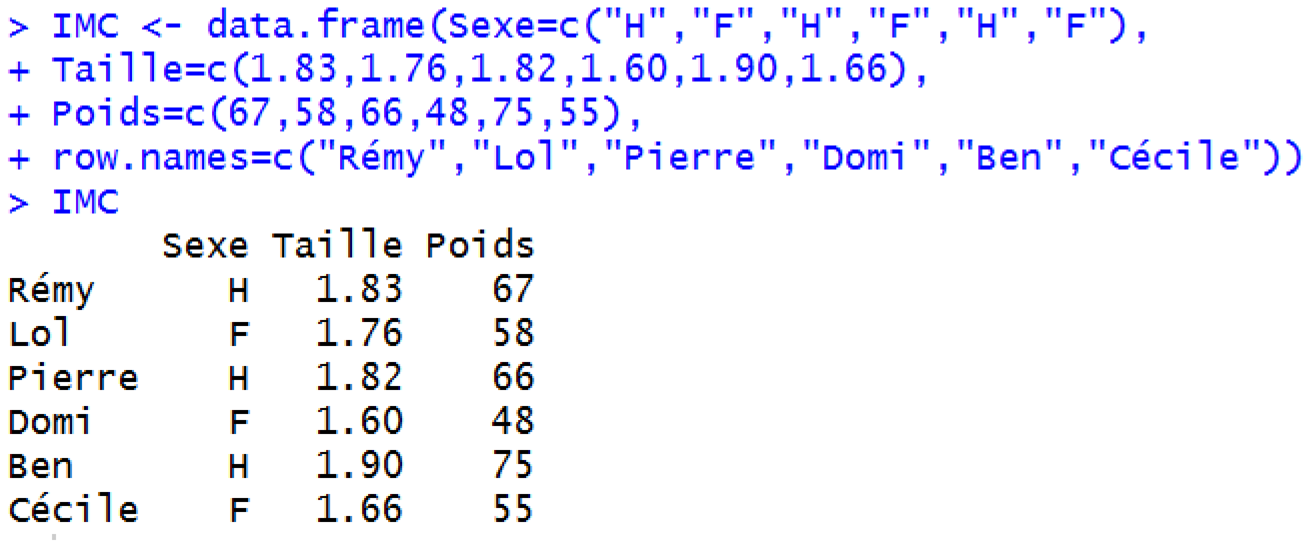


```
> A <- list(TRUE,-1:3,matrix(1:4,nrow=2),c(1,3),"Une chaîne de caractères")
> A
[[1]]
[1] TRUE

[[2]]
[1] -1  0  1  2  3

[[3]]
     [,1] [,2]
[1,]    1    3
[2,]    2    4

[[4]]
[1] 1 3

[[5]]
[1] "Une chaîne de caractères"
```

La structure des données

Le tableau individus / variables (*data.frame*)

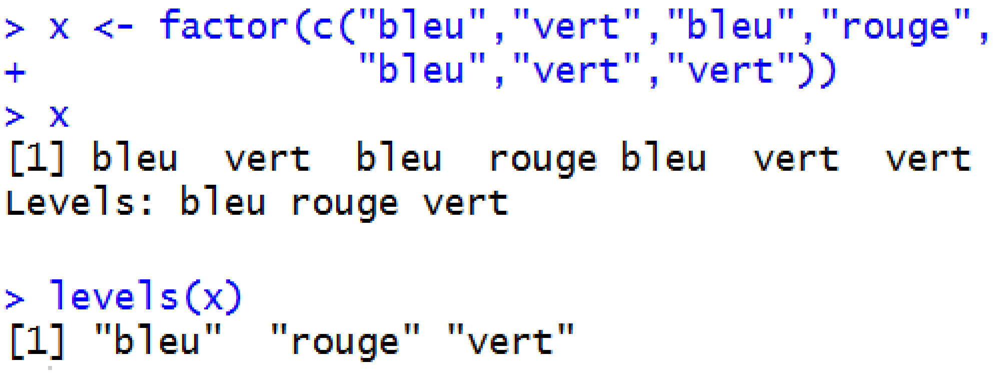
Chaque colonne représente une variable particulière dont tous les éléments sont du même type



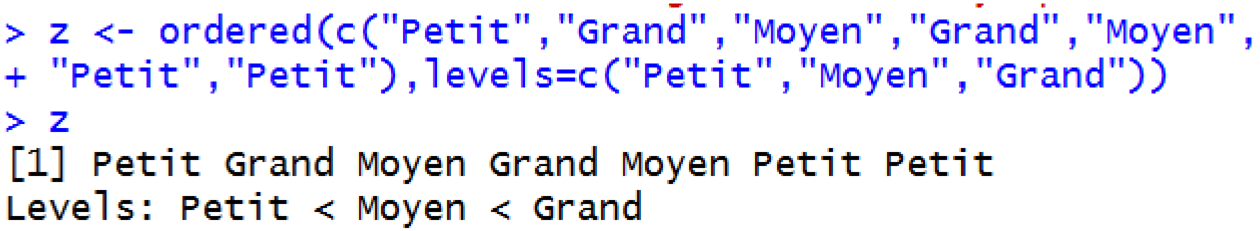
```
> IMC <- data.frame(Sexe=c("H","F","H","F","H","F"),
+ Taille=c(1.83,1.76,1.82,1.60,1.90,1.66),
+ Poids=c(67,58,66,48,75,55),
+ row.names=c("Rémy","Lol","Pierre","Domi","Ben","Cécile"))
> IMC
       Sexe Taille Poids
Rémy      H   1.83    67
Lol       F   1.76    58
Pierre    H   1.82    66
Domi      F   1.60    48
Ben       H   1.90    75
Cécile    F   1.66    55
```

La structure des données

Les facteurs (*factor*) et les variables ordinales (*ordered*)



```
> x <- factor(c("bleu","vert","bleu","rouge",
+                   "bleu","vert","vert"))
> x
[1] bleu  vert  bleu  rouge bleu  vert  vert
Levels: bleu rouge vert

> levels(x)
[1] "bleu"  "rouge" "vert"
```



```
> z <- ordered(c("Petit","Grand","Moyen","Grand","Moyen",
+ "Petit","Petit"),levels=c("Petit","Moyen","Grand"))
> z
[1] Petit Grand Moyen Grand Moyen Petit Petit
Levels: Petit < Moyen < Grand
```

Les erreurs à ne pas commettre.

Attention à la syntaxe!

Fermer les parenthèses

Chaînes de caractères avec « »

Manque de parenthèses dans les fonctions

Mauvais arguments dans les fonctions

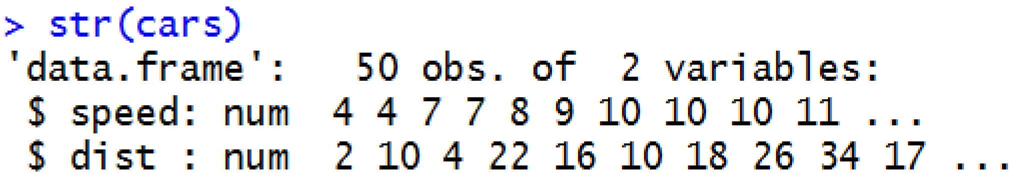
Longueurs incompatibles

Mauvaise syntaxe

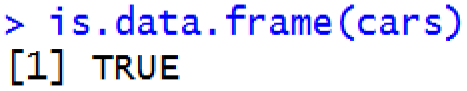
- Les arguments doivent être séparés par une virgule
- Les conditions sont entre parenthèses
- Le point est le séparateur numérique
- Expressions sur une ligne ou séparées par un point virgule

Manipulation de base des données

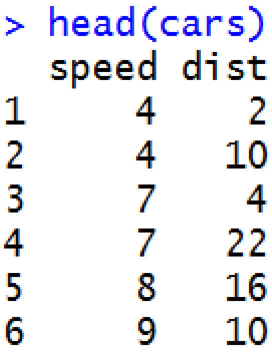
Utilisons le jeu de données cars du package `dataset`



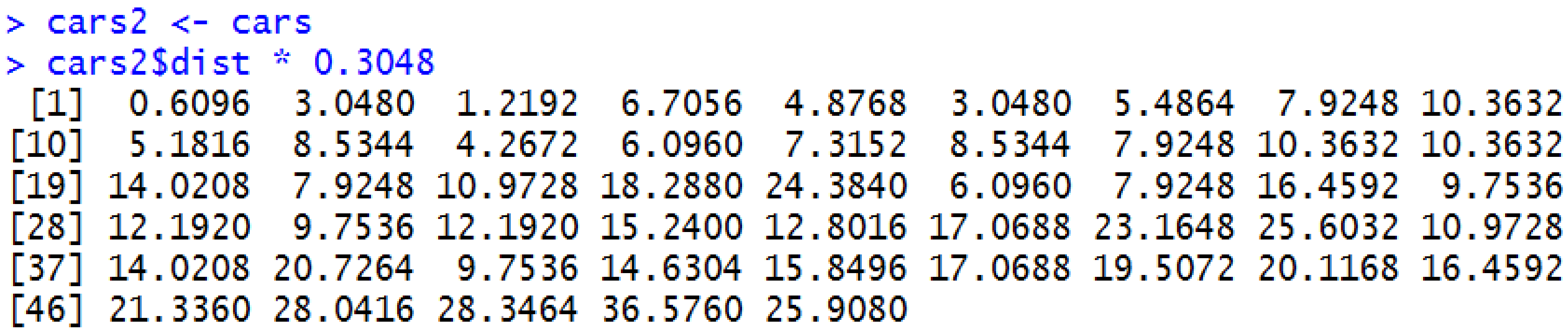
```
> str(cars)
'data.frame':   50 obs. of  2 variables:
 $ speed: num  4 4 7 7 8 9 10 10 10 11 ...
 $ dist : num  2 10 4 22 16 10 18 26 34 17 ...
```



```
> is.data.frame(cars)
[1] TRUE
```



```
> head(cars)
  speed dist
1     4    2
2     4   10
3     7    4
4     7   22
5     8   16
6     9   10
```



```
> cars2 <- cars
> cars2$dist * 0.3048
 [1]  0.6096  3.0480  1.2192  6.7056  4.8768  3.0480  5.4864  7.9248 10.3632
[10]  5.1816  8.5344  4.2672  6.0960  7.3152  8.5344  7.9248 10.3632 10.3632
[19] 14.0208  7.9248 10.9728 18.2880 24.3840  6.0960  7.9248 16.4592  9.7536
[28] 12.1920  9.7536 12.1920 15.2400 12.8016 17.0688 23.1648 25.6032 10.9728
[37] 14.0208 20.7264  9.7536 14.6304 15.8496 17.0688 19.5072 20.1168 16.4592
[46] 21.3360 28.0416 28.3464 36.5760 25.9080
```

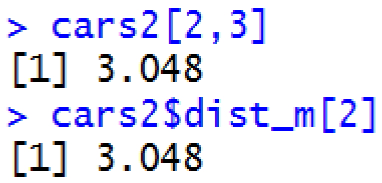
Manipulation de base des données

cars2$dist équivalent à

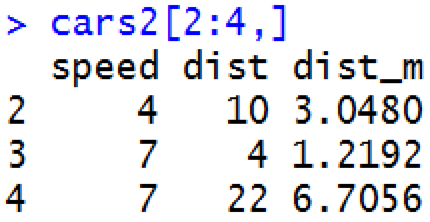
— cars2[["dist"]]

— cars2[, "dist"]

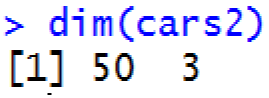
— cars2[,2]



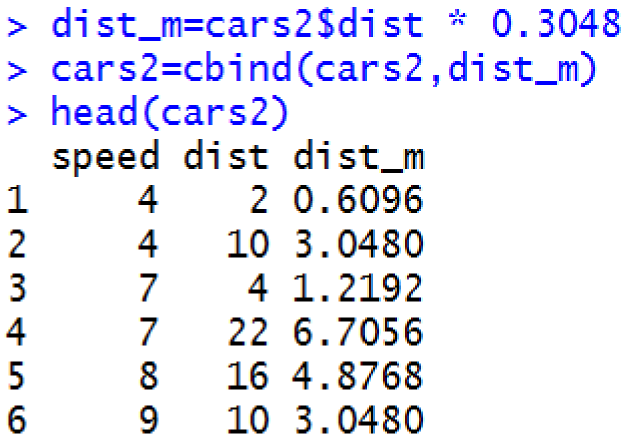
```
> cars2[2,3]
[1] 3.048
> cars2$dist_m[2]
[1] 3.048
```



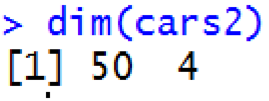
```
> cars2[2:4,]
  speed dist dist_m
2     4   10 3.0480
3     7    4 1.2192
4     7   22 6.7056
```



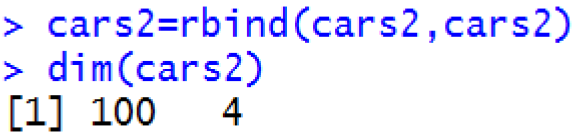
```
> dim(cars2)
[1] 50  3
```



```
> dist_m=cars2$dist * 0.3048
> cars2=cbind(cars2,dist_m)
> head(cars2)
  speed dist dist_m
1     4    2 0.6096
2     4   10 3.0480
3     7    4 1.2192
4     7   22 6.7056
5     8   16 4.8768
6     9   10 3.0480
```

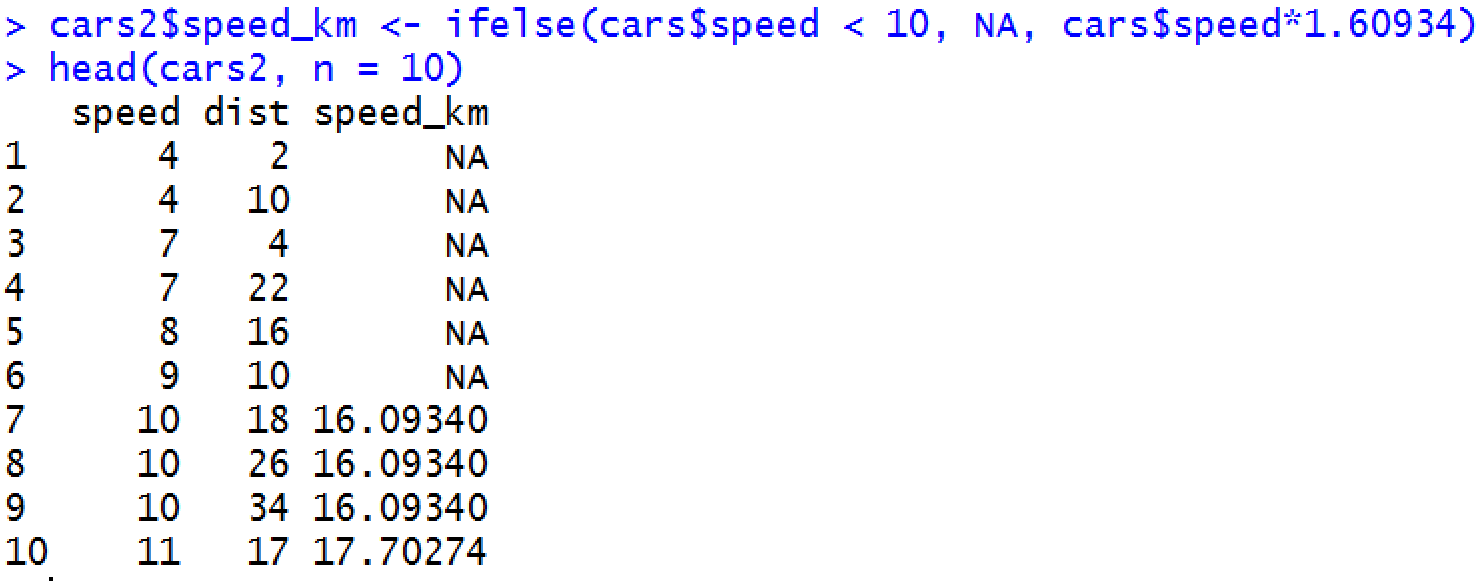


```
> dim(cars2)
[1] 50  4
```



```
> cars2=rbind(cars2,cars2)
> dim(cars2)
[1] 100   4
```

Manipulation de base des données



```
> cars2$speed_km <- ifelse(cars$speed < 10, NA, cars$speed*1.60934)
> head(cars2, n = 10)
   speed dist speed_km
1      4    2       NA
2      4   10       NA
3      7    4       NA
4      7   22       NA
5      8   16       NA
6      9   10       NA
7     10   18 16.09340
8     10   26 16.09340
9     10   34 16.09340
10    11   17 17.70274
```

Les conditions

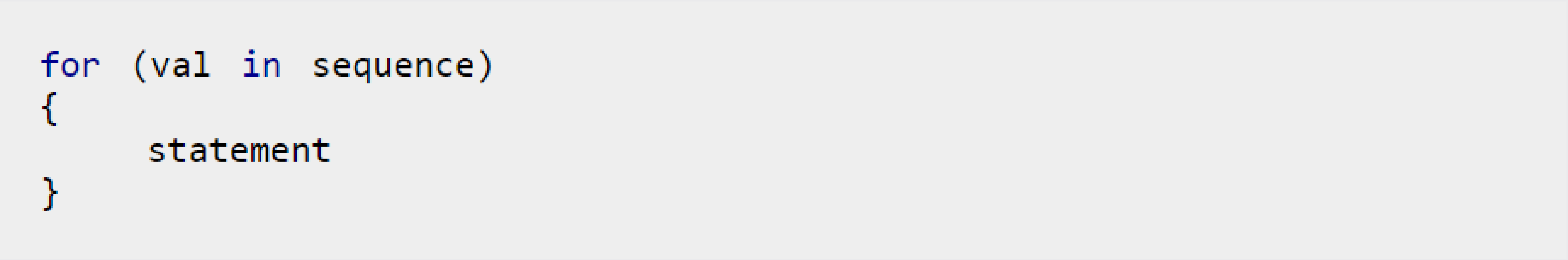
```
if (test_expression) {
 statement
}
```

```
x <- 5
if(x > 0){
     print("Positive number")
}
```

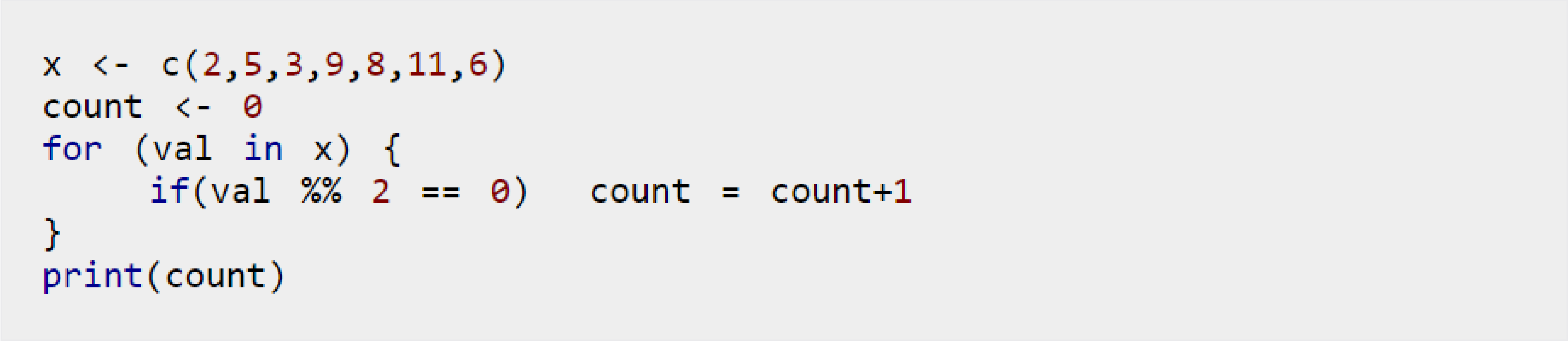
```
x <- -5
if(x > 0){
      print("Non-negative number")
} else {
      print("Negative number")
}
```

```
x <- 0
if (x < 0) {
    print("Negative number")
} else if (x > 0) {
    print("Positive number")
} else print("Zero")
```

Les boucles (for)

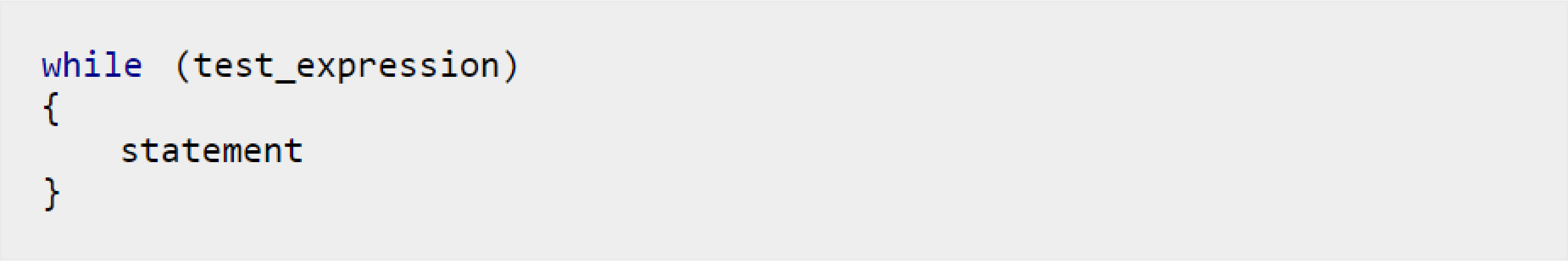


```
for (val in sequence)
{
      statement
}
```

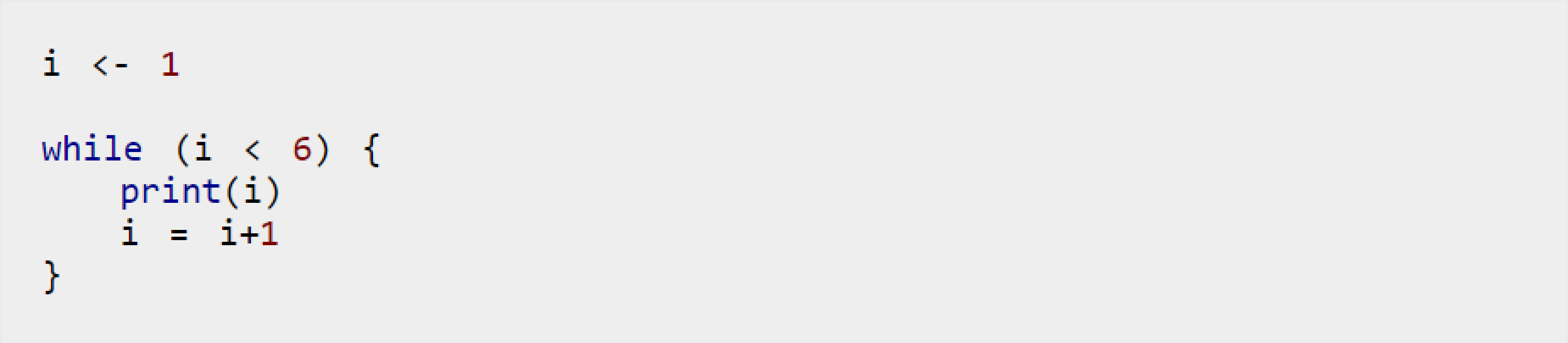


```
x <- c(2,5,3,9,8,11,6)
count <- 0
for (val in x) {
      if(val %% 2 == 0)  count = count+1
}
print(count)
```

Les boucles (while)



```
while (test_expression)
{
    statement
}
```



```
i <- 1

while (i < 6) {
    print(i)
    i = i+1
}
```