



R programming

Club Bioinfo - Institut Jacques Monod

Leslie REGAD et Gaëlle LELANDAIS

Mails :

leslie.regad@univ-paris-diderot.fr ;

gaelle.lelandais@univ-paris-diderot.fr



General Concepts

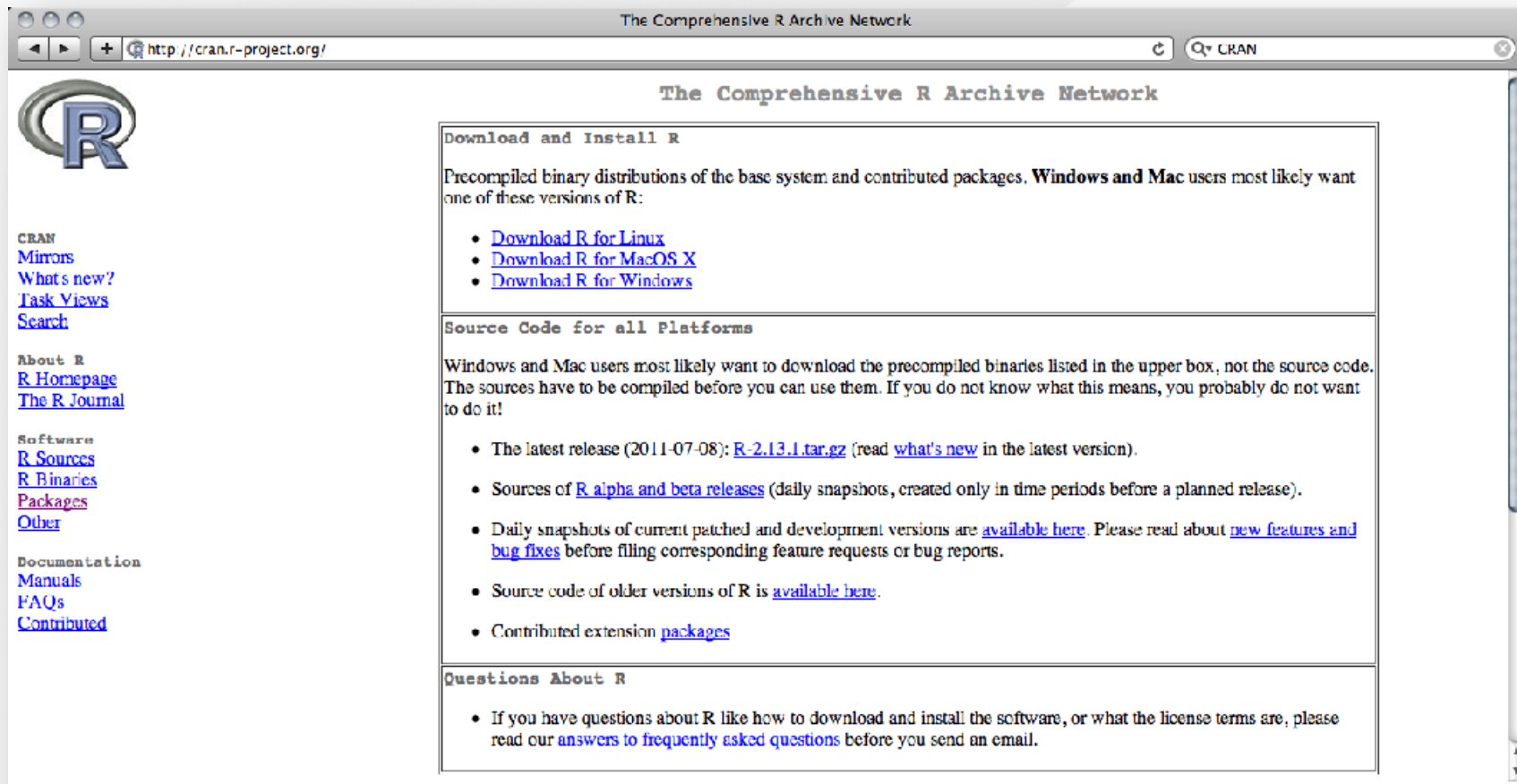
Section 1

What is R ?

- ◉ A freely available language and environment
 - > Statistical computing
 - > Graphics
- ◉ Supplementary packages
 - > Bioinformatics
 - > Data mining
 - > Database management, etc.
- ◉ Compatible with other programming languages

Reference Website: CRAN

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



The screenshot shows a web browser window titled "The Comprehensive R Archive Network". The address bar contains "http://cran.r-project.org/" and the search bar contains "CRAN". The website content includes the R logo, a navigation menu on the left, and three main sections: "Download and Install R", "Source Code for all Platforms", and "Questions About R".

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

About R
[R Homepage](#)
[The R Journal](#)

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

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

The Comprehensive R Archive Network

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:

- [Download R for Linux](#)
- [Download R for MacOS X](#)
- [Download R for Windows](#)

Source Code for all Platforms

Windows and Mac users most likely want to download 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 (2011-07-08): [R-2.13.1.tar.gz](#) (read [what's new](#) in the latest version).
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

Questions About R

- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

Start / Quit

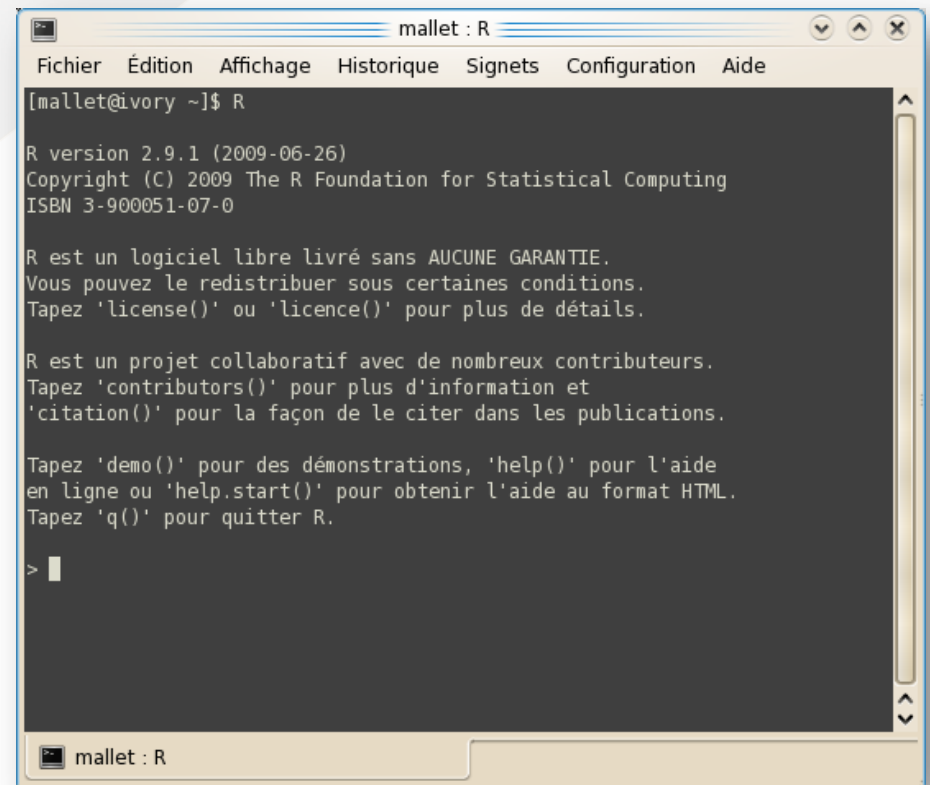
- Type « R » (in a Linux terminal) or run the program (Wind. or Mac)
- A new command line starts with the symbol « > »

Example :

```
> 1+1/4 #[press enter]
[1] 1.25
> (1+1)/4
[1] 0.5
> 21.3*2
[1] 42.6
```

- To quit the session:

```
> q()
Save workspace image? [y/n/c]
# [type « n » for « no »]
```



```
mallet : R
Fichier  Édition  Affichage  Historique  Signets  Configuration  Aide
[mallet@ivory ~]$ R
R version 2.9.1 (2009-06-26)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R est un logiciel libre livré sans AUCUNE GARANTIE.
Vous pouvez le redistribuer sous certaines conditions.
Tapez 'license()' ou 'licence()' pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Tapez 'contributors()' pour plus d'information et
'citation()' pour la façon de le citer dans les publications.

Tapez 'demo()' pour des démonstrations, 'help()' pour l'aide
en ligne ou 'help.start()' pour obtenir l'aide au format HTML.
Tapez 'q()' pour quitter R.

> |
```

Functions Available for ...

- ⊙ Performing mathematical calculations
 - > *sin()*, *cos()*, *tan()*, etc.
- ⊙ Working with numbers
 - > *round()*, *floor()*, *ceiling()*, etc.
- ⊙ Working with characters (strings)
 - > *nchar()*
 - > *toupper()*, *tolower()*, etc.
- ⊙ Performing statistical analyses
 - > *mean()*, *median()*, *var()*, *sd()*, *summary()*, etc.
 - > *sample()*, *rnorm()*, etc.

Getting Help

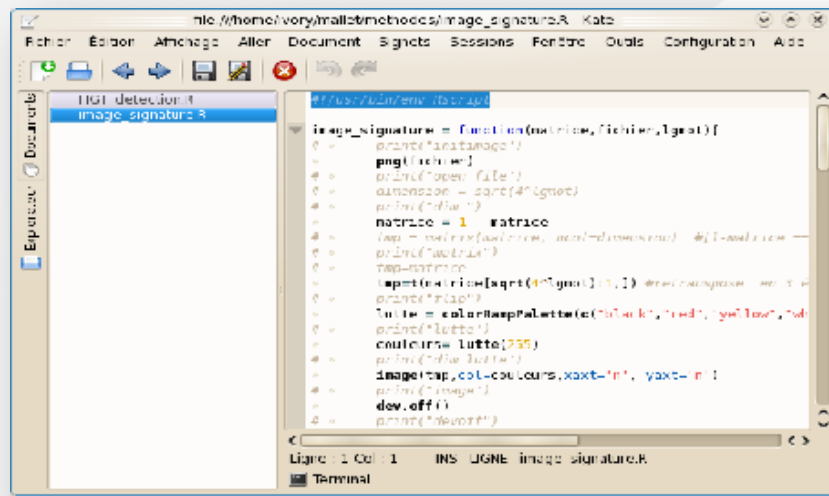
- Ask for help
 - > `? FunctionName ; help(FunctionName); help.start();`

```
> ?mean ; help(mean) ; help.search(« mean »)
```

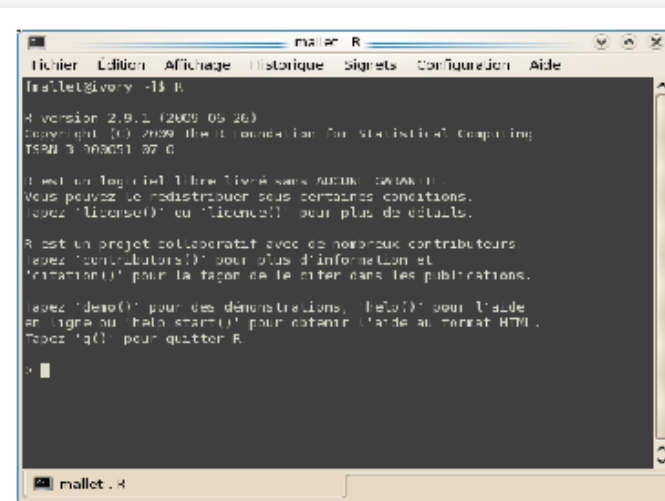
- Help page is divided into different sections
 - > *Description* → What is the purpose of the function?
 - > *Usage* → How to use the function?
 - > *Arguments* → What are the parameters of the function?
 - > *Details* → Technical information
 - > *Value* → What are the results returned by the function?
 - > *See also* → Are there similar functions?
 - > **Example** → **Short tutorial of the function**

Command Line Backup

- Choose a text editor
 - > Linux: *gedit*, *kate* or *emacs*, etc.
 - > Windows: Notepad++, TinnR
 - > Mac: *gedit*, R editor
- Copy the command lines written in the text editor and paste them in a R terminal
 - > Don't forget to save your text file



```
#/usr/bin/env Rscript
image_signature = function(matrice, fichier, lgat){
  print("signature")
  png(fichier)
  print(paste("file"))
  dimension = dim(matrice)
  print("dim")
  matrice = 1 - matrice
  img = as.matrix(apply(matrice, MARGIN=2, FUN=function(x) {
    print("matrice")
    TWO=matrix(
      apply(x, MARGIN=1, FUN=function(y) {
        print("img")
        title = colorRampPalette(c("black", "red", "yellow", "red", "black"))(length(y))
        print("lutte")
        couleur = lutte[255]
        print("couleur")
        imageTIF, col=couleur, xaxt="n", yaxt="n",
        print("image")
        dev.off()
        print("devoff")
      }
    )
  }
  )
  )
}
```



```
maller R
maller@kali:~$ R
R version 2.9.1 (2009-05-26)
Copyright (C) 2009 The R Foundation for Statistical Computing
ISBN 3-900051-87-0

R est un logiciel libre livré sans aucune garantie.
Vous pouvez le redistribuer sous certaines conditions.
Voyez "licence()" ou "licence()" pour plus de détails.

R est un projet collaboratif avec de nombreux contributeurs.
Voyez "contributors()" pour plus d'information et
"citation()" pour la façon de le citer dans les publications.

Voyez "demo()" pour des démonstrations, "help()" pour l'aide
en ligne ou "help.start()" pour obtenir l'aide au format HTML.
Voyez "q()" pour quitter R.
```


Run All the Command Lines Stored in a Text File

- To execute R scripts (command lines are successively run)
 - > `source(« ScriptR.R »)`
- Lines can be ignored (to insert comments)
 - > Symbole « # » at the beginning of lines

Text file (named « ScriptR.R ») :

```
# Example of a R script
print("my first script")

# Addition of two numbers
a = 10 + 4
print(a)
```

R software :

```
> source("ScriptR.R")
[1] "my first script"
[1] 14
```



🎯 Practical session # 1

(R installation and command line backup)



Definition of R Objects

Section 2

Getting Start with R Variables

- Variable assignment

```
> a = 1      # new R versions
> a <- 1     # old R versions
```

- Reading / Use

```
> a          # call of the variable
[1] 1        # the content of the variable is displayed
> a + 1      # change of the variable "on the flight"
[1] 2
```

- Note

- > An index is printed between characters « [] »

From Variables to Objects

- Different types of variables
 - > Integers, floats
 - > Strings
- Recommendations regarding the names of variables
 - > Case sensitive (uppercases or lowercases)
 - > Must not begin with a number or a special character (!, ;, _*)
 - > Several names are already used by the R software (*q*, *T*, *F*, *D*, *l*, *var*, *mean*, for instance)
- Combination of variables into objects
 - > Vectors
 - > Tables (matrices, data frames)
 - > Lists, etc.

Vectors

- Definition

- > Succession of elements (or variables) that have the same type (integers or floats, strings, etc.)

```
> vect = c(1, 4, 5, 6, 57)
> vect
[1] 1 4 5 6 57
> vect2 = c('a', 'k', 'm', 'p')
> vect2
[1] 'a' 'k' 'm' 'p'
```

- Special values

- > *NA* : Missing value (*Not Available*)
- > *NaN* : *Not a Number*
- > *-Inf/Inf* : Infinite symbol(+ or -)

Use of Vectors

- Functions to create a vector
 - > `c()`, `x:y`, `seq()`, `rep()`, `append()`, etc.
- Functions to manipulate a vector
 - > `data.class()`
 - > `length()`
 - > `sort()`, etc.

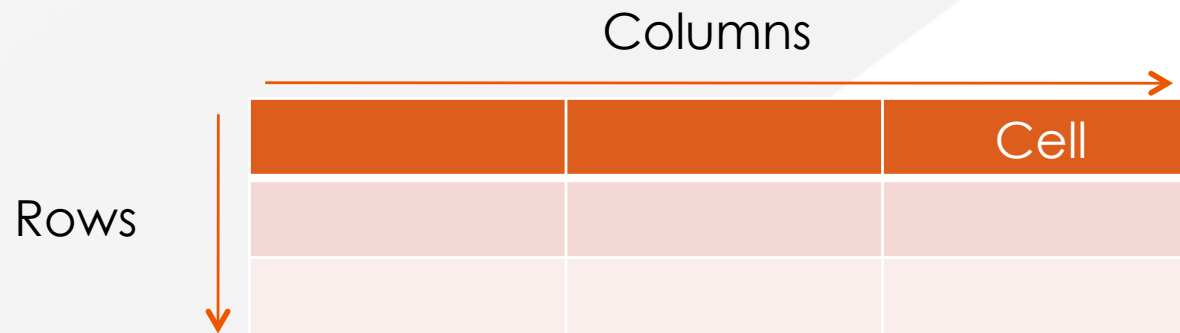
```
> c(1,2,3)
[1] 1 2 3
> 1:3
[1] 1 2 3
> seq(1,3)
[1] 1 2 3
> rep(1,3)
[1] 1 1 1
```

```
> data.class(c(1,2,3))
[1] "numeric"
> data.class(c("A","B","C"))
[1] "character"
> length(c(1,2,3))
[1] 3
> sort(c(4,5,2))
[1] 2 4 5
```

Tables : matrix and data frame

- Definition

- > Set of elements stored in a two dimensional table



- Matrix

- > Unique type for the elements

- Data frame

- > Columns can stored elements with different types (integers or floats, strings, etc.)

Use of Tables

- Functions to create a table
 - > *matrix()*, *data.frame()*, *cbind()*, *rbind()*, etc.
- Functions to manipulate a table
 - > *data.class()*
 - > *dim()*
 - > etc.

```
> matrix(0,nrow=2,ncol=2)
      [,1] [,2]
[1,]    0    0
[2,]    0    0
> dim(matrix(0,nrow=2,ncol=2))
[1] 2 2
```

```
> cbind(c(1,2),c(3,4))
      [,1] [,2]
[1,]    1    3
[2,]    2    4
> rbind(c(1,2),c(3,4))
      [,1] [,2]
[1,]    1    2
[2,]    3    4
```


Reading and Writing Files

Section 3

Reading a Text File

- ◉ Aim
 - > Import in R a dataset written in a text file
- ◉ Available functions
 - > `scan()`
 - > `read.table()`, `read.csv()`
 - > `readLines()`

Field separator character

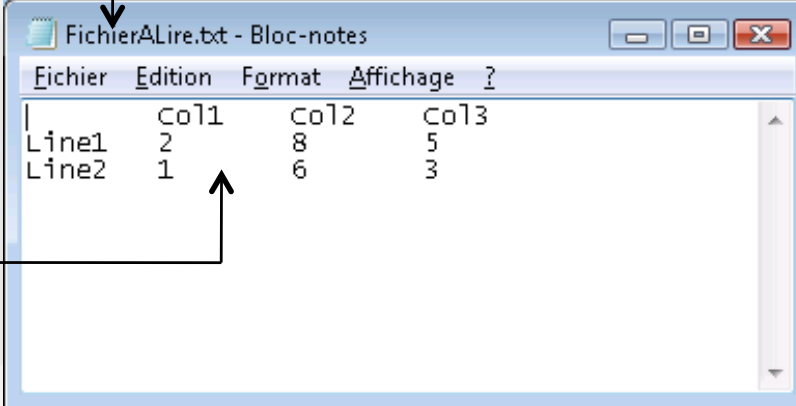


	Col 1	Col 2	...	Col n
Ligne 1	Val 1	Val 2	...	Val n
Ligne 2	Val 1	Val 2		Val n
...				
Ligne n	Val 1	Val 2		Val n

The Function « read.table() »

Name of the file to read
« FichierALire.txt »

Columns are
separated with a
« tab »



A screenshot of a Notepad window titled "FichierALire.txt - Bloc-notes". The window displays a table with three columns: "col1", "col2", and "col3". The first column contains "Line1" and "Line2". The second column contains "2" and "1". The third column contains "8" and "6". The fourth column contains "5" and "3". The columns are separated by tabs. An arrow points from the text "Name of the file to read « FichierALire.txt »" to the title bar of the Notepad window. Another arrow points from the text "Columns are separated with a « tab »" to the space between the first and second columns in the table.

	col1	col2	col3
Line1	2	8	5
Line2	1	6	3

```
> read.table("FichierALire.txt")
      col1 col2 col3
Line1    2    8    5
Line2    1    6    3
```

Writing a Text File

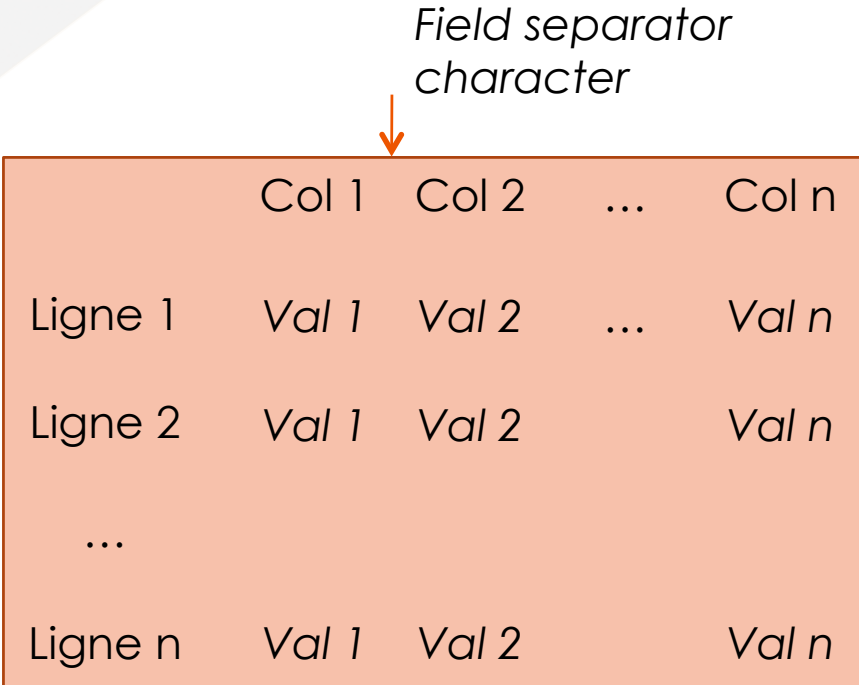
◉ Aim

- > Export a set of results obtained in R in a text file
- > This file will be opened with other softwares (Excel or OpenOffice for instance)

◉ Available functions

- > `cat()`
- > `write()`
- > `write.table()`

Field separator character



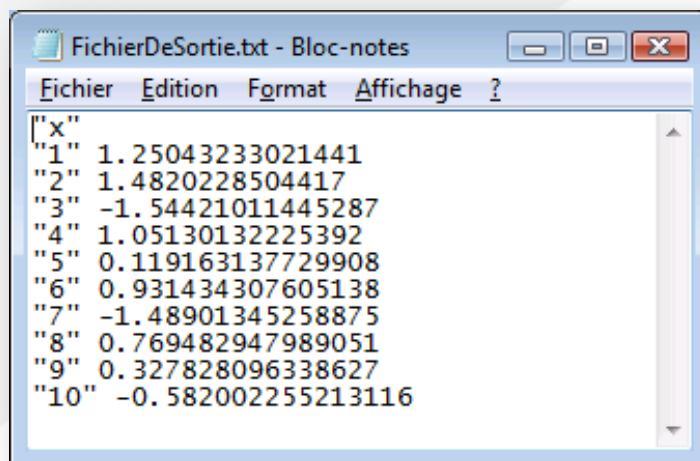
	Col 1	Col 2	...	Col n
Ligne 1	Val 1	Val 2	...	Val n
Ligne 2	Val 1	Val 2		Val n
...				
Ligne n	Val 1	Val 2		Val n

The Function « write.table() »

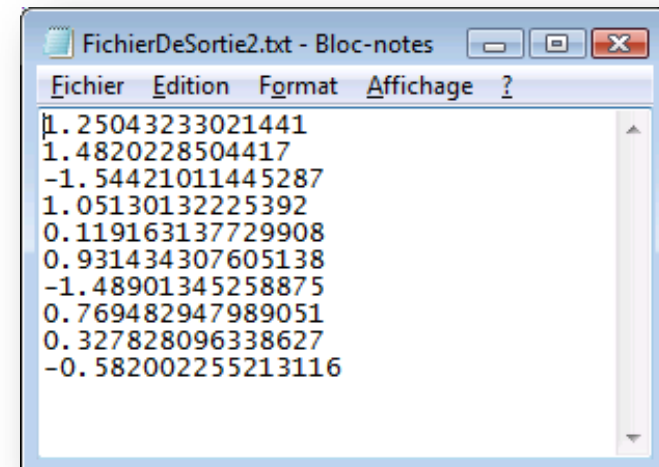
◉ Example

- > Create a vector with 10 random numbers (according to a normal distribution), and save the results into a text file

```
> data = rnorm(10)
> write.table(data, file = "FichierDeSortie.txt")
> write.table(data, file = "FichierDeSortie2.txt", row.names = F, col.names = F)
```



```
FichierDeSortie.txt - Bloc-notes
Fichier  Edition  Format  Affichage  ?
["x"
"1" 1.25043233021441
"2" 1.4820228504417
"3" -1.54421011445287
"4" 1.05130132225392
"5" 0.119163137729908
"6" 0.931434307605138
"7" -1.48901345258875
"8" 0.769482947989051
"9" 0.327828096338627
"10" -0.582002255213116
```



```
FichierDeSortie2.txt - Bloc-notes
Fichier  Edition  Format  Affichage  ?
1.25043233021441
1.4820228504417
-1.54421011445287
1.05130132225392
0.119163137729908
0.931434307605138
-1.48901345258875
0.769482947989051
0.327828096338627
-0.582002255213116
```



◎ Practical session #2

- > Exercises 4, 5, 6
- > Exercise 41

◎ Discussion: Your needs with R?