

# Visual R reference card (Draft)

21 février 2010

Sylvain Loiseau <sylvain.loiseau@unicaen.fr>  
Université de Caen-Basse Normandie

## 1 Graphical conventions

In this document, evaluations of R expressions are represented graphically. For instance, the expression "c(7, 5) + 3", which create a vector and add 3 to its elements, is represented as follow :

```
c ( [7] , [5] ) + [3]
    [7] [5] + [3]
      [10] [8]
```

The first line represent the expression you typed in, the last line give the object eventually created, and each intermediary line is a step in the evaluation of the expression.

## 2 Anatomy of a vector

All elements of a vector have a common mode, one of character, logical, and numeric.

```
"les" "gli" "los"      TRUE FALSE TRUE      10 0.32 2
```

All elements of a vector have an index. They may have a name.

$\nwarrow$	$\nearrow$	$\circ$
1	2	3
"les"	"gli"	"los"

All vectors have two important properties : their mode and their length.

```
mode ( "les" "gli" "los" )      length ( "les" "gli" "los" )
      "character"                [3]
```

```
mode ( 10 0.32 2 )              length ( 10 0.32 2 )
      "numeric"                  [3]
```

```
mode ( TRUE FALSE TRUE )        length ( TRUE FALSE TRUE )
      "logical"                   [3]
```

Functions are precisely defined in terms of mode, number and length of vectors they may take as argument, and mode and length of vector they create.

1. length() take one vector of any mode and length and return a vector of numeric vector of length 1.

2. mode() take one vector of any mode and length and return a character vector of length 1.

### 3 Creating a vector

```

c ( 1 , 7 , 3 )           c ( TRUE , FALSE , TRUE )
  1 7 3                     TRUE FALSE TRUE

c ( "yes" , "b" , "no" )
  "yes" "b" "no"

      5 : 8                  seq ( 1 , 4 )
    5 6 7 8                 1 2 3 4

rep ( "yes" , 4 )          rep ( seq ( 1 , 4 ) , 2 )
"yes" "yes" "yes" "yes"   rep ( 1 2 3 4 , 2 )
                           1 2 3 4 1 2 3 4

```

The function c() take any number of vectors of any length and any mode, but all vectors must have the same mode (see below, "Conversion"). It returns a vector of the same mode as the arguments and whose length is the sum of the length of its arguments.

### 4 Extraction

A vector can be created by extracting some elements of a vector.  
Elements to be extracted can be addressed using their index.

```

10 7 2 [ 2 ]      "les" "gli" "los" [ 1 3 ]
   7                "les" "los"

```

Index are 1-based. If an index is greater than the number of element in the vector, you get "NA". If the vector has names, their are preserved.

```

      1 2 3
    10 7 2 [ 1 3 ]
      1 2
    10 2

TRUE FALSE TRUE [ 1 4 ]
  TRUE NA

```

You can also extract with character or logical vector inside the square brackets of the extraction operator. Elements of a character vector are interpreted as the names of the elements to be extracted (elements must have names!).

```

  1 2 3
10 7 2 [ "b" ]
    1
    7

```

Logical vectors must have the same length as the vector to be extracted. Elements are extracted if there is a "TRUE" value at the same position in the logical vector. (If the logical vector is shorter, it is recycled : right (see below for recycling))

10	7	2	[	TRUE	FALSE	TRUE	]	"	a	"	b	"	c	"	d	"	[	TRUE	FALSE	]	
						10	2										"	a	"	c	"

When extracting, nothing prevent you from reordering elements or extracting several times the same element :

"	a	"	b	"	c	"	d	"	[	c	(	1	,	1	,	4	,	2	,	1	)	]
"	a	"	b	"	c	"	d	"	[	1	1	4	2	1								
"	a	"	a	"	d	"	b	"	a													

## 5 Operator

Some numeric operators.

5	+	6	5	-	6	5	*	6
11			-1			30		
5	/	6	5	<	6	5	>=	6
0.8333333333333333			TRUE			FALSE		

Some logical operators.

5	==	6	5	==	5
FALSE			TRUE		
TRUE	==	FALSE	"oui"	==	"non"
FALSE			FALSE		

## 6 Vectorization

Operators – as well as many functions – may operate on vector of any length : the operation is performed on pair of elements of equal index.

4	3	8	+	32	3	2	4	3	8	==	32	3	2				
36			6			10			FALSE			TRUE			FALSE		

## 7 Recycling

In a context where vectorization is allowed, you may provide vectors of unequal length. The shorter is duplicated until its length reach the length of the longer. This is called recycling a vector.

```
c ( 1 , 2 , 3 , 4 ) + 1
1 2 3 4 + 1
2 3 4 5
```

```
5 : 8 > 6
5 6 7 8 > 6
FALSE FALSE TRUE TRUE
```

Be careful :

```
"..." "yes" "ja" "si" == "ja" "yes"
FALSE TRUE TRUE FALSE
```

## 8 Some numeric functions

Some functions for numeric vectors.

```
sum ( 2 1 3 4 )
10
```

```
mean ( 2 1 3 4 )
2.5
```

```
var ( 2 1 3 4 )
1.666666666666667
```

```
sd ( 2 1 3 4 )
1.29099444873581
```

```
max ( 2 1 3 4 )
4
```

```
min ( 2 1 3 4 )
1
```

```
range ( 2 1 3 4 )
1 4
```

```
cumsum ( 2 1 3 4 )
2 3 6 10
```

```
cumprod ( 2 1 3 4 )
2 2 6 24
```

TODO : table()

## 9 Some string functions

nchar() count the number of characters in all strings of a character vector.

```
nchar ( "les" "i" "los" )
3 1 3
```

Recycling and vectorization are useful with paste(), which concatenates characters string at same index in several characters vectors :

```
paste ( "oui" , "non" )
      "oui non"
```

```
paste ( "oui" , "non" , sep = "" )
      "ouinon"
```

```
paste ( "oui" "non" , "si" "no" )
      "oui si" "non no"
```

```
paste ( "les" "i" "los" , "oui" )
      "les oui" "i oui" "los oui"
```

You can paste more than two vectors of characters :

```
paste ( "(" , names (


| fr | it | es |
|----|----|----|
| 1  | 2  | 3  |


      "les" "gli" "los" ) , ")" ,


| fr | it | es |
|----|----|----|
| 1  | 2  | 3  |


      "les" "gli" "los" , sep = "" )
```

```
paste ( "(" , "fr" "it" "es" , ")" ,


| fr | it | es |
|----|----|----|
| 1  | 2  | 3  |


      "les" "gli" "los" , sep = "" )
      "(fr) les" "(it) gli" "(es) los"
```

## 10 Sorting

Numeric and character vectors can be sorted. Names are preserved.

```
sort ( 2 1 3 4 )
      1 2 3 4
sort ( 2 1 3 4 , decreasing = TRUE )
      4 3 2 1
```

```
sort ( "les" "gli" )
      "gli" "les"
sort (


| fr | it | es |
|----|----|----|
| 1  | 2  | 3  |


      "les" "gli" "los" )


| it | fr | es |
|----|----|----|
| 1  | 2  | 3  |


      "gli" "les" "los"
```

Flipping a vector :

```
rev ( 2 1 3 4 )
      4 3 1 2
```

## 11 Type conversion

c() coerce its arguments to a common mode – all elements of a vector always have a common mode. The character mode always wins. Logical always loses.

```
c ( 1 : 3 , FALSE , TRUE )
      1 2 3 0 1
c ( 1 2 3 , FALSE , TRUE )
      1 2 3 0 1
c ( "oui" , 1 : 3 , FALSE )
      "oui" "1" "2" "3" "FALSE"
c ( "oui" , 1 2 3 , FALSE )
      "oui" "1" "2" "3" "FALSE"
```

Many functions silently convert their arguments to the required mode. For instance, the function nchar() give the number of characters of the elements of a character vector. It makes sense only for characters string : numbers

don't have a "number of characters" themselves, but inside a convention of representation as characters strings. If the function `nchar()` receive a vector of another mode (numerical, logical), the vector is silently converted into a characters vector using `as.character()` (left). The result is identical to explicitly converting into characters, using `as.character()` (right).

```

nchar ( 3 102 14 )
1 3 2

nchar ( as.character ( 3 102 14 ) )
nchar ( "3" "102" "14" )
1 3 2

```

## 12 Index

Some useful functions give index rather than the actual values.

```

which ( TRUE FALSE FALSE TRUE FALSE TRUE )
1 4 6

```

```

order ( 10 2 7 15 )
2 3 1 4

order (
  

| $\text{rt}$ | $\text{it}$ | $\text{es}$ |
|-------------|-------------|-------------|
| 1           | 2           | 3           |
| "les"       | "gli"       | "los"       |


)
2 1 3

```

```

which.min ( 2 1 3 4 )
2

which.max ( 2 1 3 4 )
4

```

This is particularly useful in very common situations where two or more vectors are "aligned", or "synchronized". Suppose two vectors : a character vector giving forms in a corpus (left), and a numeric vector giving the total frequency for each form (right) :

```

"le" "de" "un" "a"
60 100 40 30

```

Using `max()`, you may retrieve the maximum frequency from the second vector, but you can't figure out which form have this frequency. Using `which.max()`, you're still able to extract the corresponding value in the first vector :

```

"le" "de" "un" "a" [ which.max ( 60 100 40 30 ) ]
"le" "de" "un" "a" [ 2 ]
"de"

```

Again, suppose you want to sort the row of a matrix according to the value in a column. You cannot use `sort`, since it give the actual values sorted, not the index of the row sorted. `Order` is commonly used for reordering data structure :

```


|   |   |   |
|---|---|---|
| 4 | 3 | 5 |
| 1 | 2 | 9 |
| 3 | 1 | 8 |

 [ order (


|   |   |   |
|---|---|---|
| 4 | 3 | 5 |
| 1 | 2 | 9 |
| 3 | 1 | 8 |

 [ , 

|   |
|---|
| 1 |
|---|

 ] ) , ]



|   |   |   |
|---|---|---|
| 4 | 3 | 5 |
| 1 | 2 | 9 |
| 3 | 1 | 8 |

 [ order ( 

|   |   |   |
|---|---|---|
| 4 | 1 | 3 |
|---|---|---|

 ) , ]



|   |   |   |
|---|---|---|
| 4 | 3 | 5 |
| 1 | 2 | 9 |
| 3 | 1 | 8 |

 [ 

|   |   |   |
|---|---|---|
| 2 | 3 | 1 |
|---|---|---|

 , ]



|   |   |   |
|---|---|---|
| 1 | 2 | 9 |
| 3 | 1 | 8 |
| 4 | 3 | 5 |


```

## 13 Precedence

Operators have precedence (see ?Syntax).

"seq" takes precedence over "+", "seq" takes precedence over logical operators...

```


|   |
|---|
| 1 |
|---|

 : 

|    |
|----|
| 10 |
|----|

 + 

|   |
|---|
| 2 |
|---|



|   |   |   |   |   |   |   |   |   |    |
|---|---|---|---|---|---|---|---|---|----|
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|----|

 + 

|   |
|---|
| 2 |
|---|



|   |   |   |   |   |   |   |    |    |    |
|---|---|---|---|---|---|---|----|----|----|
| 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|----|----|----|



|   |
|---|
| 5 |
|---|

 : 

|   |
|---|
| 8 |
|---|

 > 

|   |
|---|
| 6 |
|---|



|   |   |   |   |
|---|---|---|---|
| 5 | 6 | 7 | 8 |
|---|---|---|---|

 > 

|   |
|---|
| 6 |
|---|



|       |       |      |      |
|-------|-------|------|------|
| FALSE | FALSE | TRUE | TRUE |
|-------|-------|------|------|


```

The order in which the operators are written in the code does not matter!

```


|   |
|---|
| 6 |
|---|

 < 

|   |
|---|
| 5 |
|---|

 : 

|   |
|---|
| 8 |
|---|



|   |
|---|
| 6 |
|---|

 < 

|   |   |   |   |
|---|---|---|---|
| 5 | 6 | 7 | 8 |
|---|---|---|---|



|       |       |      |      |
|-------|-------|------|------|
| FALSE | FALSE | TRUE | TRUE |
|-------|-------|------|------|


```

## 14 Matrix

### 14.1 Creation

Matrix are created with the `matrix()` function. It takes three main arguments : a vector (any mode and any length) gives the content, two vectors (numeric and length 1) give the numbers of rows and columns. If only one dimension is given, the second one is deduced from the length of the vector. If both dimensions are given and the vector length do not match the number of cells, the vector is recycled to fill the matrix. The matrix is filled by column; this behavior may be changed with the option `byrow`.

```
matrix ( 1 : 6 , nrow = 3 )
matrix ( 1 2 3 4 5 6 , nrow = 3 )
```

1	4
2	5
3	6

```
matrix ( 1 : 6 , 3 )
matrix ( 1 2 3 4 5 6 , 3 )
```

1	4
2	5
3	6

```
matrix ( 1 : 6 , 3 , byrow = TRUE )
matrix ( 1 2 3 4 5 6 , 3 , byrow = TRUE )
```

1	2
3	4
5	6

```
matrix ( 1 : 6 , ncol = 3 )
matrix ( 1 2 3 4 5 6 , ncol = 3 )
```

1	3	5
2	4	6

```
matrix ( TRUE , nrow = 2 , ncol = 3 )
```

TRUE	TRUE	TRUE
TRUE	TRUE	TRUE

## 14.2 Extraction with a matrix

Matrices have two dimensions and you must provide extractors for each of them. You first extract the rows, then the columns.

```
matrix ( 1 3 5 , 2 4 6 ) [ 1 : 2 , 2 : 3 ]
```

1	3	5
2	4	6

1	3	5
2	4	6

3	5
4	6

```
matrix ( 1 3 5 , 2 4 6 ) [ 2 , 2 : 3 ]
matrix ( 1 3 5 , 2 4 6 ) [ 2 , 2 3 ]
```

1	3	5
2	4	6

1	3	5
2	4	6

4	6
---	---

```
matrix ( 1 3 5 , 2 4 6 ) [ 1 , "deux" ]
```

1	3	5
2	4	6

1	3	5
2	4	6

3
---

```
matrix ( 1 3 5 , 2 4 6 ) [ TRUE FALSE , "deux" ]
```

1	3	5
2	4	6

3
---

If you leave blank the column slot, all columns are selected (left); if you leave blank the row slot, all rows are selected (right).



Figure 1 illustrates a 2D grid structure with a 3x3 subgrid and a 2x2 subgrid. The 3x3 subgrid is labeled 'un', 'deux', 'trois' and the 2x2 subgrid is labeled '1', '2', '3'. The grid is divided into two parts: 'A' and 'B'. The first example shows a 3x3 grid with values 1, 2, 3, 4, 5, 6, 7, 8, 9. The second example shows a 3x3 grid with values 1, 2, 3, 4, 5, 6, 7, 8, 9. The third example shows a 3x3 grid with values 1, 2, 3, 4, 5, 6, 7, 8, 9.

$\text{nrow} \left( \begin{array}{|c|c|c|} \hline 1 & 3 & 5 \\ \hline 2 & 4 & 6 \\ \hline \end{array} \right)$ 
 $\text{ncol} \left( \begin{array}{|c|c|c|} \hline 1 & 3 & 5 \\ \hline 2 & 4 & 6 \\ \hline \end{array} \right)$ 
 $\text{dim} \left( \begin{array}{|c|c|c|} \hline 1 & 3 & 5 \\ \hline 2 & 4 & 6 \\ \hline \end{array} \right)$

Diagram illustrating the relationship between rownames, colnames, and a matrix:

rownames ( 

A	1	3	5
B	2	4	6

 )

colnames ( 

un	deux	trois
----	------	-------

 )

The matrix structure is shown as:

	un	deux	trois
A	1	3	5
B	2	4	6

```
length ( 

|   |   |   |
|---|---|---|
| 1 | 3 | 5 |
| 2 | 4 | 6 |

 )      mode ( 

|   |   |   |
|---|---|---|
| 1 | 3 | 5 |
| 2 | 4 | 6 |

 )
```

6
---

"numeric"
-----------

1	3	5
2	4	6

 [ 

3
---

 ]  

3
---

## 14.4 Contingency table

prop.table() compute proportion, given a matrix of frequencies. Proportion are computed either for the whole table (top), by row (middle) or by column (bottom) :

<div> <div>135</div> <div>246</div> </div>		
prop.table (    )		
0.0476190476190476	0.142857142857143	0.238095238095238
0.0952380952380952	0.19047619047619	0.285714285714286

<div> <div>135</div> <div>246</div> </div>		
prop.table (    , margin = 1 )		
0.111111111111111	0.333333333333333	0.555555555555556
0.166666666666667	0.333333333333333	0.5

<div> <div>135</div> <div>246</div> </div>		
prop.table (    , margin = 2 )		
0.333333333333333	0.428571428571429	0.454545454545455
0.666666666666667	0.571428571428571	0.545454545454545

margin.table() compute margin sum, given a matrix of frequencies. Margin are computed either for the whole table (left), for row (right) or for column (bottom).

<div> <div>435</div> <div>129</div> <div>318</div> </div>		
margin.table (    )		
36		

<div> <div>435</div> <div>129</div> <div>318</div> </div>		
margin.table (    , margin = 1 )		
12	12	12

<div> <div>435</div> <div>129</div> <div>318</div> </div>		
margin.table (    , margin = 2 )		
8	6	22

This objects with dashed lines are "lists" : see below.

## 14.5 Summing a matrix

<div> <div>135</div> <div>246</div> </div>		
sum (    )		
21		

<div> <div>135</div> <div>246</div> </div>		
rowSums (    )		
9	12	

<div> <div>135</div> <div>246</div> </div>		
colSums (    )		
3	7	11

## 14.6 Changing

as.vector ( 

1	3	5
2	4	6

 )

1	2	3	4	5	6
---	---	---	---	---	---

1	3	5
2	4	6

 + 

3
---

4	6	8
5	7	9

cbind (

		un	deux	trois			un	deux	trois	
		1	2	3			1	2	3	
A	1	1	3	5		A	1	1	3	5
B	1	2	4	6		B	1	2	4	6

)

Diagram illustrating the `rbind` function. It shows two 2x3 matrices being combined into a 4x3 matrix.

Matrix 1 (left):

A	1	1	3	5
B	1	2	4	6

Matrix 2 (right):

A	1	1	3	5
B	1	2	4	6

Resulting matrix (bottom):

A	1	1	3	5
B	1	2	4	6
A	1	1	3	5
B	1	2	4	6

t ( 

1	3	5
2	4	6

 )

1	2
3	4
5	6


15 list

## 15.1 Creation, anatomy


Creating a list by enumerating its components.

A list of length 4 : contains 4 vectors, each of length 1 (left); a list of length 1 : contains 1 vector of length 2 (right).

```
list ( 1, 2, 3, 4 )
```



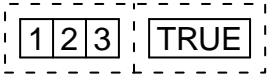
```
list ( c ( 1, 2 ) )
```



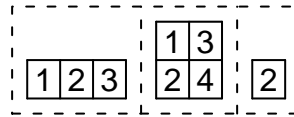
List can contain objects of different mode (left) ; it can contain objects of different dimensions (right).

```
list ( [1] : [3] , matrix ( [1] : [4] , [2] ) , [2] )
list ( [1] : [3] , matrix ( [1 2 3 4] , [2] ) , [2] )
```

```
list ( [1] : [3] , TRUE )
list ( [1 2 3] , TRUE )
```

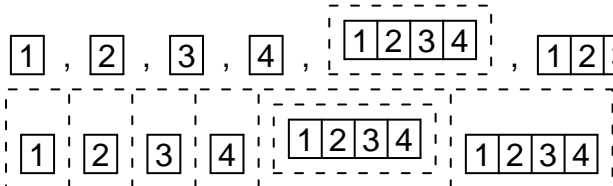


```
list ( [1 2 3] , [1 3
                2 4] , [2] )
```



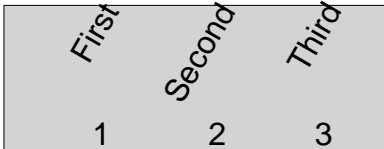
A list can be recursive : a component may be a list.

```
list ( [1] , [2] , [3] , [4] , list ( [1] : [4] ) , [1] : [4] )
list ( [1] , [2] , [3] , [4] , list ( [1 2 3 4] ) , [1] : [4] )
```




## 15.2 Basic functions


```
names ( [1 2 3] [TRUE] ["jour"] )
["First" "Second" "Third"]
```



```
length ( [1 2 3] [TRUE] ["jour"] )
[3]
```



```
mode ( [1 2 3] [TRUE] ["jour"] )
["list"]
```



## 15.3 Extraction

Extracting in a list. The next two figures show the difference between `[` and `[[` operator on list : the first create a sublist (it extracts elements, exactly as it extracts elements from a vector), while the second is completely different : it give the content of one (and only one) element of a list.

```
list ( [ 1 ] : [ 3 ] , [ TRUE ] ) [ [ 1 ] ]
list ( [ 1 2 3 ] , [ TRUE ] ) [ [ 1 ] ]
```

```
list ( [ 1 ] : [ 3 ] , [ TRUE ] ) [[ [ 1 ] ]]
list ( [ 1 2 3 ] , [ TRUE ] ) [[ [ 1 ] ]]
```

You can use vector of any length within the single-square bracket, while you can address only one element within the double-square-bracket operator, and then use only vector of length 1.

Since a list is a recursive data structure (may contain list), you can use several successive bracket operators in order to go down to the element you're interested in.

With the single-square-bracket operator you cannot walk down though the data structure.

```
[ 1 ] [ 1 2 3 ] [ "b" ] [[ [ 2 ] ]] [[ [ 1 ] ]] [ [ 3 ] ]
```

Be sure to understand the difference between "length(l[1])" and "length(l[[1]])" :

```
length ( [ 1 2 3 ] [ TRUE ] [ "jour" ] [[ [ 1 ] ] ] )
length ( [ 1 2 3 ] )
[ 3 ]
```

```
length ( [ 1 2 3 ] [ TRUE ] [ "jour" ] [ [ 1 ] ] )
length ( [ 1 2 3 ] )
[ 1 ]
```

## 15.4 List and vector

```
unlist ( [ 1 2 3 ] [ TRUE ] [ "jour" ] )
[ "1" "2" "3" "TRUE" "jour" ]
```

```
as.list ( [ 1 ] : [ 4 ] )
as.list ( [ 1 2 3 4 ] )
[ 1 ] [ 2 ] [ 3 ] [ 4 ]
```

## 15.5 List for expressing complex data structure

Grouping elements of a vector using a the level of a factors (see Factor) :

```
split ( [ 1 2 3 4 5 6 7 8 ] , [ 1 1 1 2 2 2 3 3 ] )
```

See also for instance strsplit() below.

## 16 Regexp

### 16.1 Split strings : strsplit()

```
strsplit ( c ( "un" , "deux" , "trois" ) , "[aeio]" )  
strsplit ( "un" "deux" "trois" , "[aeio]" )  
"un" "d" "ux" "tr" "" "s"  
  
strsplit ( c ( "un" , "deux" , "trois" ) , c ( "u" , "e" , "r" ) )  
strsplit ( "un" "deux" "trois" , "u" "e" "r" )  
"" "n" "d" "ux" "t" "ois"
```

### 16.2 Extract sub strings : substr()

```
substr ( "trois" , 2 , 3 )  
"ro"  
  
substr ( c ( "trois" , "quatre" ) , 1 , 3 )  
substr ( "trois" "quatre" , 1 , 3 )  
"tro" "qua"  
  
substr ( c ( "trois" , "quatre" ) , c ( 2 , 1 ) , c ( 3 , 4 ) )  
substr ( "trois" "quatre" , 2 1 , 3 4 )  
"ro" "quat"
```

### 16.3 Searching elements of a character vector with regexp : grep()

```
grep ( "[dt]" , "un" "deux" "trois" )  
2 3
```

```
grepl ( "[dt]" , "un" "deux" "trois" )  
FALSE TRUE TRUE
```

### 16.4 Substitution : sub()

```
sub ( "[ueaio]" , "v" , "un" "deux" "trois" )  
"vn" "dvux" "trvis"
```

```
gsub ( "[ueaio]" , "v" , "un" "deux" "trois" )
      "vn" "dvvx" "trvvs"
```

## 16.5 Searching substring in elements of character vector : regular expression

TODO

## 17 Factor

A factor represent a nominal random variable. It looks like a character vector, and may be created using a character vector (left). In this document, factors are represented without quotes around the values. The different values in the factor (the different modality in the random variable) may be retrieved using `levels()` (right).

```
factor ( "bleu" "rouge" "bleu" "vert" "rouge" )      levels ( bleu rouge bleu vert rouge )
      bleu rouge bleu vert rouge                    "bleu" "rouge" "vert"
```

A factor is useful for *grouping* elements. Suppose two vectors, one giving forms, and the other giving part of speech. You want to group the forms according to part of speech.

```
"pratique" "représentation" "linguistique" "sociale" "Guyane"      "nc" "nc" "adj" "adj" "npr"
```

The function `split()` groups elements, given two arguments : a vector (the elements to be grouped), and a factor (giving the group of each element). `split()` create a list, each element of the list corresponding to a group, the name of the list element corresponding to the group name.

```
split ( "pratique" "représentation" "linguistique" "sociale" "Guyane" , "nc" "nc" "adj" "adj" "npr"
```

adj	nc	npr
1	2	3
"linguistique" "sociale"	"pratique" "représentation"	"Guyane"

You may give vector of any mode as second argument to `split()` : `split()` will call `as.factor()` on this argument. `tapply()` do the same grouping, and then apply a function (it's third argument) to each group :

```
tapply ( 10 8 13 20 5 , "nc" "nc" "adj" "adj" "npr" , mean )
```

adj	nc	npr
1	2	3
16.5	9	5

`rowsum()` performs a `colSums` on each group of rows given by the second argument.

```
rowsum ( 1 6
        2 7
        3 8
        4 9
        5 10 , 2 1 2 1 3 )
```

6	16
4	14
5	10

There are numerous other functions using factor (or converting vector into factor) allowing for grouping (by(), aggregate(), etc.)

## 18 Data Frame

A data frame is a data structure for representing statistical information about a group of individuals. For each individual, you may have numerical random variable, categorical random variable, ie. different modes (numeric, character, factor).

In a data frame, each row represent an individual, and each column represent a random variable. This is like a matrix, except that the columns may have different mode. This is like a list, since it may mix vectors of different modes, but all vectors must have the same length.

From an internal representation, a data frame is a list of vectors. Thus, data frame may be created with the function data.frame() by enumerating its column-vectors. It is represented here with dotted lines, like a list, and solid lines around vector-column :

```
data.frame ( col1 = 

|   |   |   |
|---|---|---|
| 1 | 2 | 3 |
|---|---|---|

 , col2 = 

|      |        |        |
|------|--------|--------|
| "un" | "deux" | "deux" |
|------|--------|--------|

 , col3 = 

|      |      |      |
|------|------|------|
| 1001 | 1002 | 1003 |
|------|------|------|

 )
```

	col1	col2	col3
	1	2	3
1 1	1	un	1001
2 2	2	deux	1002
3 3	3	deux	1003

Like a matrix and unlike a list, a data frame may have rownames. In fact, a data frame has *always* row and column names (while matrix may have no row or column names). You can see from the previous exemple that automatic default row names have been added : a character representation of the index number of the row. Default column names are created as well for unnamed column. rownames() and colnames() may be used with data frame as with matrix.

### 18.1 Converting

You can also create a data frame from a matrix, or create a matrix from a data frame (when you convert data frame into matrix, a mode compatible with all column is used : the character mode below) :

```
as.matrix ( 

|     | col1 | col2 | col3 |
|-----|------|------|------|
|     | 1    | 2    | 3    |
| 1 1 | 1    | un   | 1001 |
| 2 2 | 2    | deux | 1002 |
| 3 3 | 3    | deux | 1003 |

 )
```

"1"	"un"	"1001"
"2"	"deux"	"1002"
"3"	"deux"	"1003"

```
as.data.frame ( 

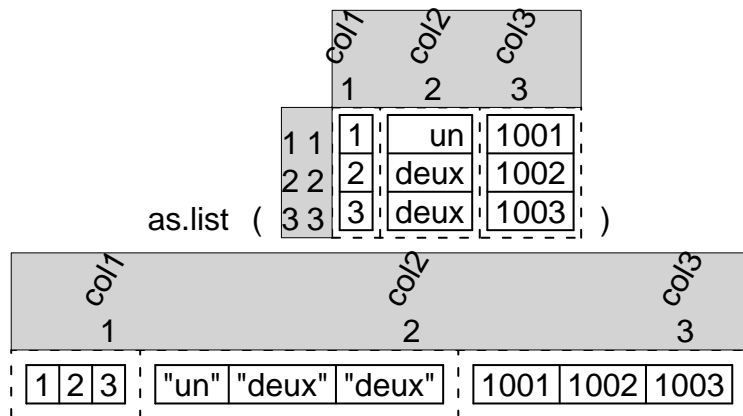
|   |   |   |
|---|---|---|
| 1 | 3 | 5 |
| 2 | 4 | 6 |

 )
```

	1	2	3
1 1	1	3	5
2 2	2	4	6

as.data.frame create default row and column names : data frame cannot be without row and column names. Row and column names are lost with as.matrix. Column names are kept with as.list() :





You may create a data frame from a list only if all list component have the same length :

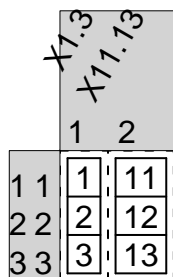
```
as.data.frame ( list ( 1 : 3 , 11 : 13 ) )
```

```
as.data.frame ( list ( 1 2 3 , 11 12 13 ) )
```

```
as.data.frame (
```

1 2 3	11 12 13
-------	----------

```
)
```



Many functions expecting a matrix will accept a data frame, silently converting it into a matrix. Similarly, many functions expecting a data frame will accept a matrix and silently convert it into a data frame.

## 18.2 Extraction

1/ with simple square bracket operator, a data frame is seen as matrix : two dimensions must be provided inside the square brackets

		col/1	col/2	col/3	
		1	2	3	
1 1	1	un	1001	[ 1 : 2 , 1 : 2 ]	
2 2	2	deux	1002		
3 3	3	deux	1003		

		col/1	col/2	col/3	
		1	2	3	
1 1	1	un	1001	[ 1 2 , 1 2 ]	
2 2	2	deux	1002		
3 3	3	deux	1003		

		col/1	col/2
		1	2
1 1	1	un	
2 2	2	deux	

However, note that row-extraction (left) and column-extraction (right) do not give the same data structure. It is a data frame (left) or a vector (right). Furthermore, in column extraction, the row name are lost (right) :

		col/1	col/2	col/3	
		1	2	3	
1 1	1	un	1001	[ 1 , ]	
2 2	2	deux	1002		
3 3	3	deux	1003		

		col/1	col/2	col/3	
		1	2	3	
1 1	1	un	1001		

		col/1	col/2	col/3	
		1	2	3	
1 1	1	un	1001	[ , 1 ]	
2 2	2	deux	1002		
3 3	3	deux	1003		
		1	2	3	

2/ with double square bracket operator, a data.frame behave like a list : one dimension must be provided.

		col/1	col/2	col/3	
		1	2	3	
1 1	1	un	1001	[[ 1 ]]	
2 2	2	deux	1002		
3 3	3	deux	1003		
		1	2	3	

		col/1	col/2	col/3	
		1	2	3	
1 1	1	un	1001	[[ "col1" ]]	
2 2	2	deux	1002		
3 3	3	deux	1003		
		1	2	3	

The "\$" operator may be used like in a list

		col1	col2	col3
		1	2	3
1	1	1	un	1001
2	2	2	deux	1002
3	3	3	deux	1003

\$ col1

1	2	3
---	---	---

### 18.3 Subsetting

To be done

### 18.4 Merging

Merge() use the columns with same name for combining two matrices or data frame :

merge (

A	1	1	3	5
	2	2	4	6
	3	3	5	6

,

B	1	5	13	15
	2	6	14	16
	3	6	14	16

)

1	1	5	1	3	13	15
	2	6	2	4	14	16
	3	6	3	5	14	16

The result is always a data frame. Columns to be used for combining may be given explicitly