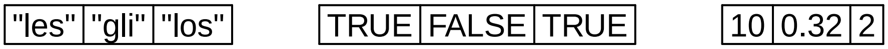
Visual R reference card (Draft)

23 janvier 2010

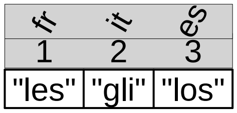
Sylvain Loiseau <sylvain.loiseau@unicaen.fr>
Université de Caen–Basse Normandie

1 Anatomy of a vector

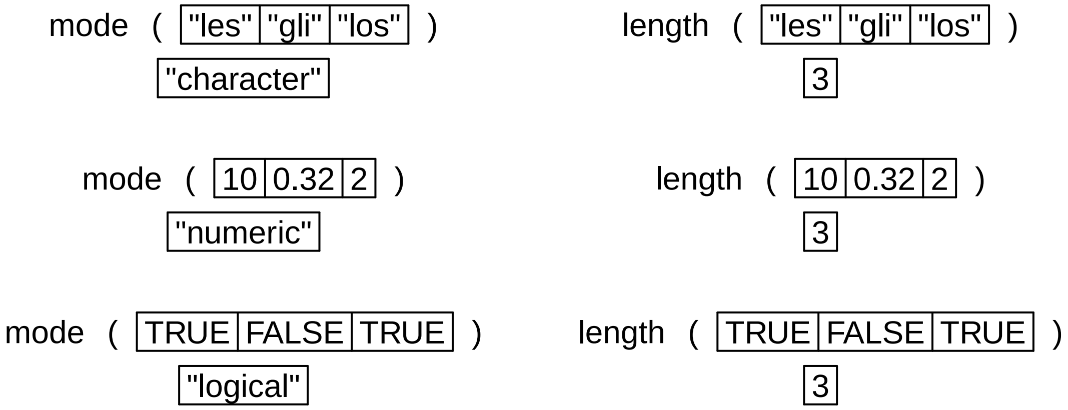
All elements of a vector have a common mode, one of character, logical, and numeric.



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



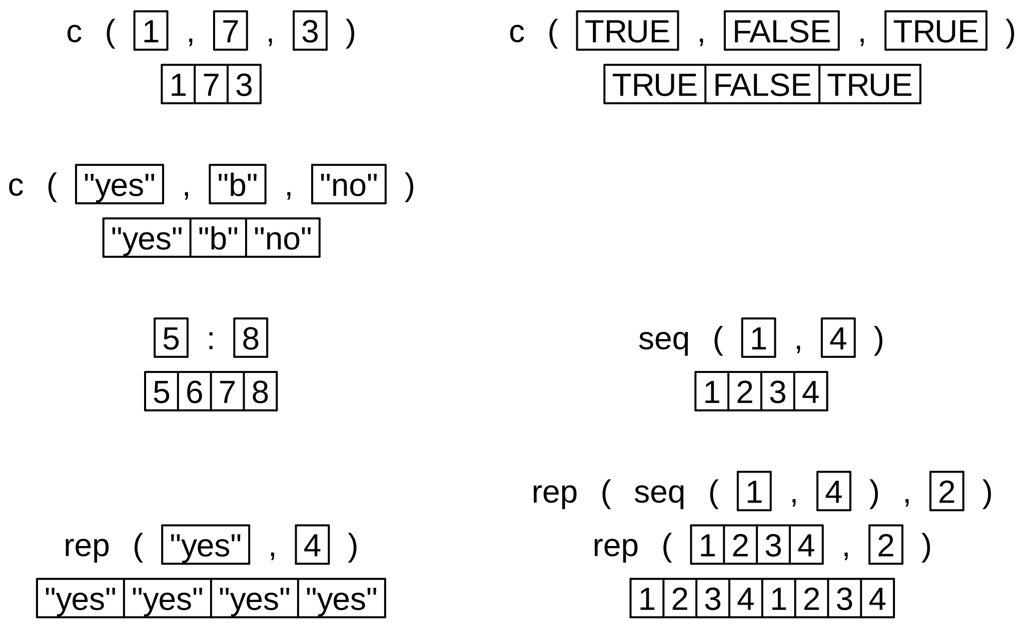
All vectors have two important properties : their mode and their length.



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

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

2 Creating a vector

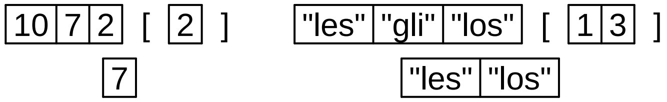


The function c() take any number of vectors of any length and any mode, but all vectors must have the same mode (see below, "Conversion"). It returns a vector of the same mode as the arguments and whose length is the sum of the length of its arguments.

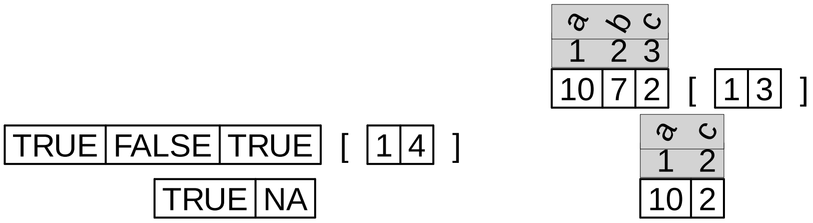
3 Extraction

A vector can be created by extracting some elements of a vector.

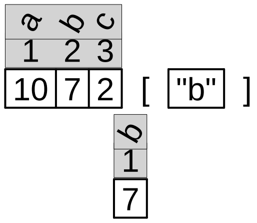
Elements to be extracted can be addressed using their index.



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.



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 !).



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))

$$\boxed{10}\boxed{7}\boxed{2}\ [\ \boxed{\text{TRUE}}\boxed{\text{FALSE}}\boxed{\text{TRUE}}\ ]$$
$$\boxed{10}\boxed{2}$$

$$\boxed{\text{"a"}}\boxed{\text{"b"}}\boxed{\text{"c"}}\boxed{\text{"d"}}\ [\ \boxed{\text{TRUE}}\boxed{\text{FALSE}}\ ]$$
$$\boxed{\text{"a"}}\boxed{\text{"c"}}$$

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

$$\boxed{\text{"a"}}\boxed{\text{"b"}}\boxed{\text{"c"}}\boxed{\text{"d"}}\ [\ \text{c}\ (\ \boxed{1}\ ,\ \boxed{1}\ ,\ \boxed{4}\ ,\ \boxed{2}\ ,\ \boxed{1}\ )\ ]$$
$$\boxed{\text{"a"}}\boxed{\text{"b"}}\boxed{\text{"c"}}\boxed{\text{"d"}}\ [\ \boxed{1}\boxed{1}\boxed{4}\boxed{2}\boxed{1}\ ]$$
$$\boxed{\text{"a"}}\boxed{\text{"a"}}\boxed{\text{"d"}}\boxed{\text{"b"}}\boxed{\text{"a"}}$$

4 Operator

Some numeric operators.

$$\boxed{5}\ +\ \boxed{6}$$
$$\boxed{11}$$

$$\boxed{5}\ -\ \boxed{6}$$
$$\boxed{-1}$$

$$\boxed{5}\ *\ \boxed{6}$$
$$\boxed{30}$$

$$\boxed{5}\ /\ \boxed{6}$$
$$\boxed{0.833333333333333}$$

$$\boxed{5}\ <\ \boxed{6}$$
$$\boxed{\text{TRUE}}$$

$$\boxed{5}\ >=\ \boxed{6}$$
$$\boxed{\text{FALSE}}$$

Some logical operators.

$$\boxed{5}\ ==\ \boxed{6}$$
$$\boxed{\text{FALSE}}$$

$$\boxed{5}\ ==\ \boxed{5}$$
$$\boxed{\text{TRUE}}$$

$$\boxed{\text{TRUE}}\ ==\ \boxed{\text{FALSE}}$$
$$\boxed{\text{FALSE}}$$

$$\boxed{\text{"oui"}}\ ==\ \boxed{\text{"non"}}$$
$$\boxed{\text{FALSE}}$$

5 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.

$$\boxed{4}\boxed{3}\boxed{8}\ +\ \boxed{32}\boxed{3}\boxed{2}$$
$$\boxed{36}\boxed{6}\boxed{10}$$

$$\boxed{4}\boxed{3}\boxed{8}\ ==\ \boxed{32}\boxed{3}\boxed{2}$$
$$\boxed{\text{FALSE}}\boxed{\text{TRUE}}\boxed{\text{FALSE}}$$

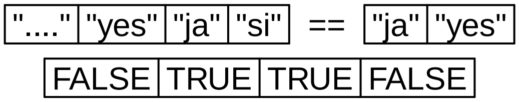
6 Recycling

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.

$$\text{c}\ (\ \boxed{1}\ ,\ \boxed{2}\ ,\ \boxed{3}\ ,\ \boxed{4}\ )\ +\ \boxed{1}$$
$$\boxed{1}\boxed{2}\boxed{3}\boxed{4}\ +\ \boxed{1}$$
$$\boxed{2}\boxed{3}\boxed{4}\boxed{5}$$

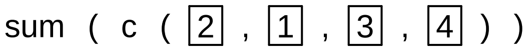
$$\boxed{5}\ :\ \boxed{8}\ >\ \boxed{6}$$
$$\boxed{5}\boxed{6}\boxed{7}\boxed{8}\ >\ \boxed{6}$$
$$\boxed{\text{FALSE}}\boxed{\text{FALSE}}\boxed{\text{TRUE}}\boxed{\text{TRUE}}$$

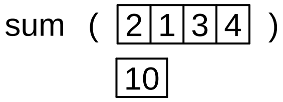
Be careful :

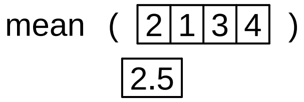


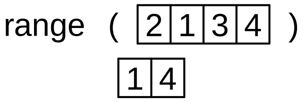
7 Some numerical functions

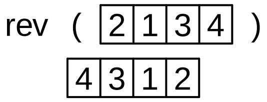
Some functions for numeric vectors.

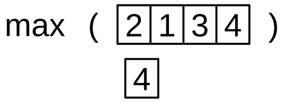


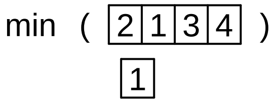


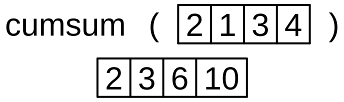








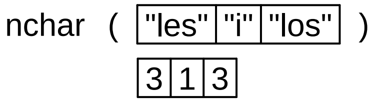




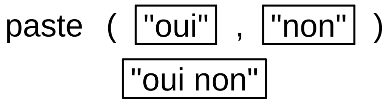
TODO : table

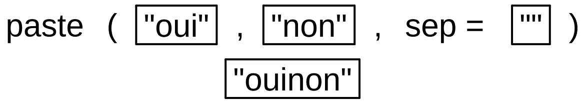
8 Some string functions

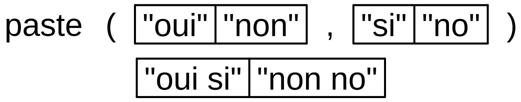
nchar() count the number of characters in all strings of a character vector.

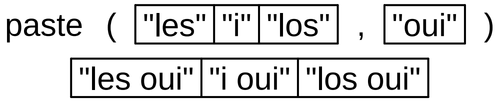


Recycling and vectorization are useful with paste(), which concatenates characters string at same index in several characters vectors :

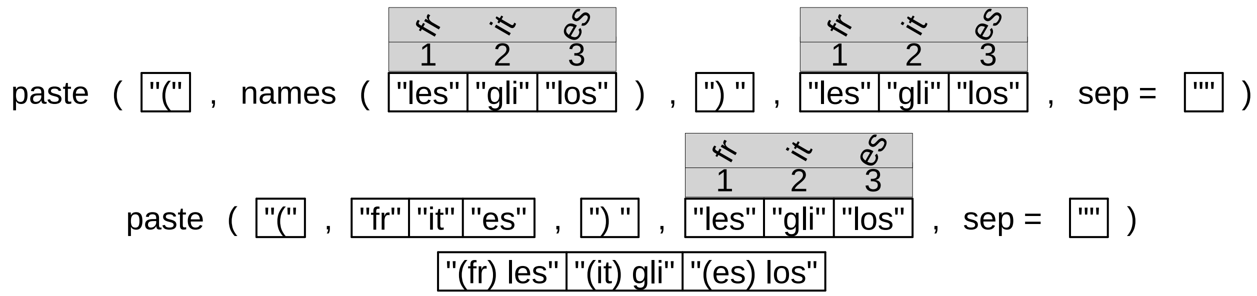







paste ("les" "i" "los" , "oui")
"les oui" "i oui" "los oui"

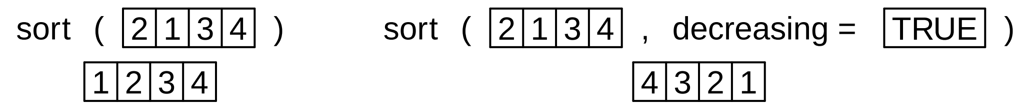
You can paste more than two vectors of characters :

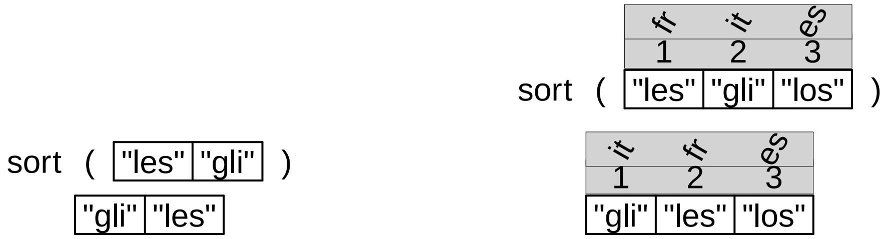


paste ("(" , names (fr it es / 1 2 3 / "les" "gli" "los") , ") " , fr it es / 1 2 3 / "les" "gli" "los" , sep = "")
paste ("(" , "fr" "it" "es" , ") " , fr it es / 1 2 3 / "les" "gli" "los" , sep = "")
"(fr) les" "(it) gli" "(es) los"

9 Sorting

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

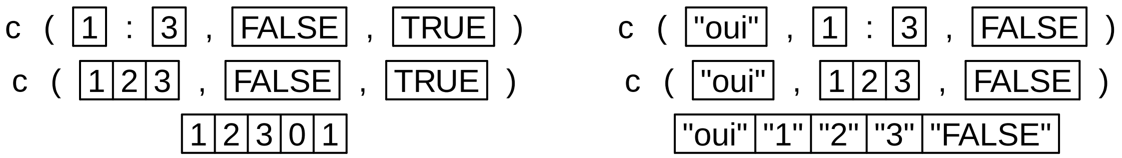




sort (fr it es / 1 2 3 / "les" "gli" "los")
it fr es / 1 2 3 / "gli" "les" "los"
sort ("les" "gli")
"gli" "les"

10 Type conversion

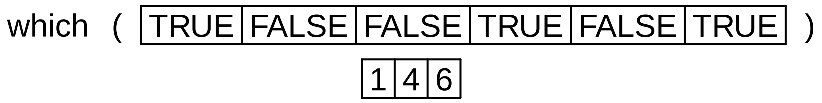
c() coerce arguments to a common mode – all elements of a vector always have a common mode. The character mode always wins. Logical always looses.

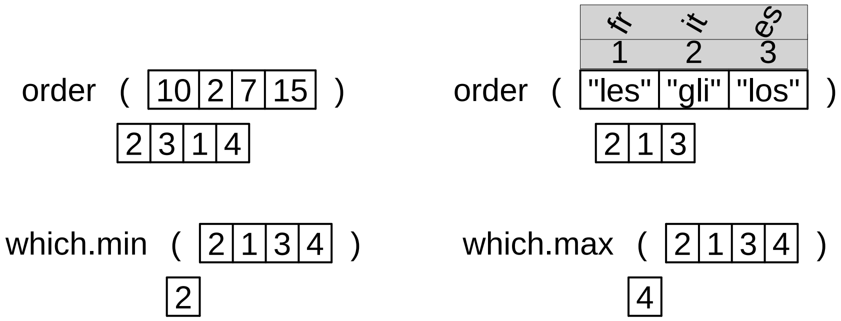


c (1 : 3 , FALSE , TRUE)
c (1 2 3 , FALSE , TRUE)
1 2 3 0 1
c ("oui" , 1 : 3 , FALSE)
c ("oui" , 1 2 3 , FALSE)
"oui" "1" "2" "3" "FALSE"

11 Index

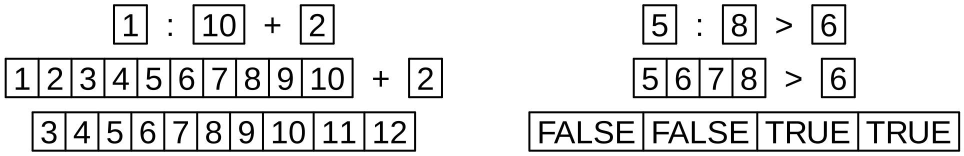
Some useful functions give index rather than the actual values.



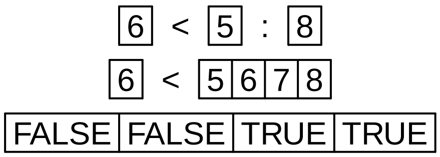


12 Precedence

Operators have precedence (see ?Syntax).
"seq" takes precedence over "+", "seq" takes precedence over logical operators...



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



13 Factor

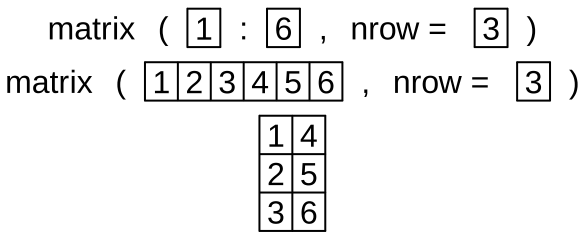
TODO

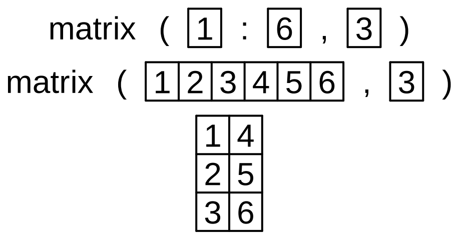
There is numerous situation where values of a vector are seen as modalities, allowing for grouping, etc. (split, rowsum, tapply, etc.)

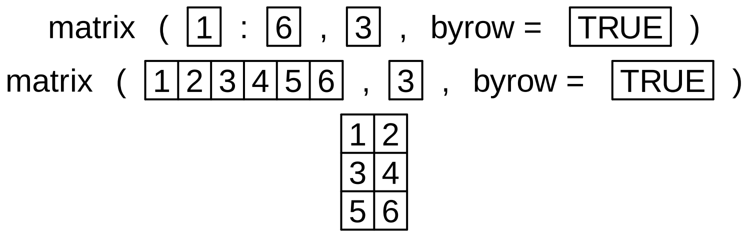
14 Matrix

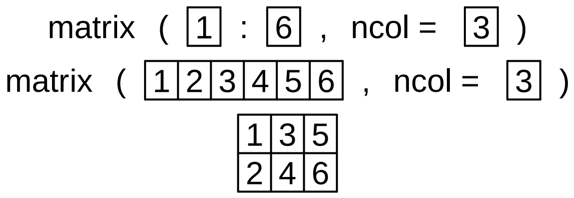
14.1 Creation

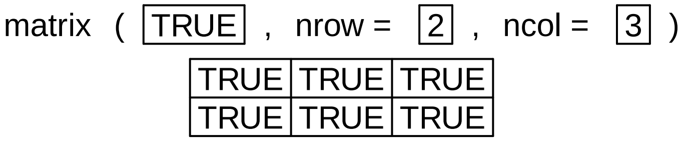
Matrix are created with the matrix() function. It takes three main arguments : a vector (any type and any length) gives the content, two vectors (numeric and length 1) give the numbers of rows and columns. If only one dimension is given, the second one is deduced from the length of the vector. If both 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



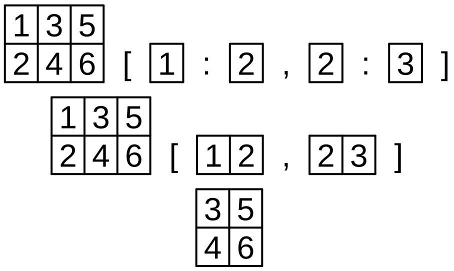

matrix (1 : 6 , 3 , byrow = TRUE)
matrix (1 2 3 4 5 6 , 3 , byrow = TRUE)
1 2
3 4
5 6

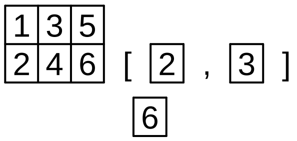

matrix (1 : 6 , ncol = 3)
matrix (1 2 3 4 5 6 , ncol = 3)
1 3 5
2 4 6

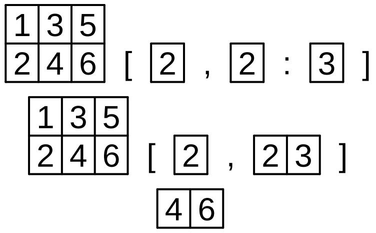


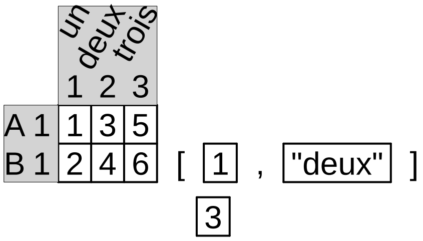
14.2 Extraction with a matrix

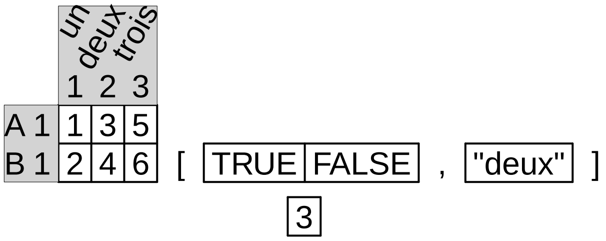
Matrices have two dimensions and you must provide extractors for each of them. You first extract the rows, then the columns.











How does extraction in a matrix preserve names ? No name is preserved if you extract a single element ; longuest dimension's names are preserved if you extract a vector of more than one element, both dimensions are preserved if you extract a sub-matrix :

$$\begin{array}{c|c|c|c|} & \text{un} & \text{deux} & \text{trois} \\ & 1 & 2 & 3 \\ \hline \text{A 1} & 1 & 3 & 5 \\ \hline \text{B 1} & 2 & 4 & 6 \\ \hline \end{array}\ [\ \boxed{2}\ ,\ \boxed{2}\ ] \quad \boxed{4}$$

$$\begin{array}{c|c|c|c|} & \text{un} & \text{deux} & \text{trois} \\ & 1 & 2 & 3 \\ \hline \text{A 1} & 1 & 3 & 5 \\ \hline \text{B 1} & 2 & 4 & 6 \\ \hline \end{array}\ [\ \boxed{1}\ ,\ \begin{array}{|c|c|}\hline 1 & 2 \\ \hline \end{array}\ ] \quad \begin{array}{|c|c|} \text{un} & \text{deux} \\ 1 & 2 \\ \hline 1 & 3 \\ \hline \end{array}$$

$$\begin{array}{c|c|c|c|} & \text{un} & \text{deux} & \text{trois} \\ & 1 & 2 & 3 \\ \hline \text{A 1} & 1 & 3 & 5 \\ \hline \text{B 1} & 2 & 4 & 6 \\ \hline \end{array}\ [\ \begin{array}{|c|c|}\hline 1 & 2 \\ \hline \end{array}\ ,\ \begin{array}{|c|c|}\hline 1 & 2 \\ \hline \end{array}\ ] \quad \begin{array}{c|c|c|} & \text{un} & \text{deux} \\ & 1 & 2 \\ \hline \text{A 1} & 1 & 3 \\ \hline \text{B 1} & 2 & 4 \\ \hline \end{array}$$

14.3 Properties

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

$$\text{rownames}\ (\ \begin{array}{c|c|c|c|} & \text{un} & \text{deux} & \text{trois} \\ & 1 & 2 & 3 \\ \hline \text{A 1} & 1 & 3 & 5 \\ \hline \text{B 1} & 2 & 4 & 6 \\ \hline \end{array}\ ) \quad \begin{array}{|c|c|}\hline \text{"A"} & \text{"B"} \\ \hline \end{array}$$

$$\text{colnames}\ (\ \begin{array}{c|c|c|c|} & \text{un} & \text{deux} & \text{trois} \\ & 1 & 2 & 3 \\ \hline \text{A 1} & 1 & 3 & 5 \\ \hline \text{B 1} & 2 & 4 & 6 \\ \hline \end{array}\ ) \quad \begin{array}{|c|c|c|}\hline \text{"un"} & \text{"deux"} & \text{"trois"} \\ \hline \end{array}$$

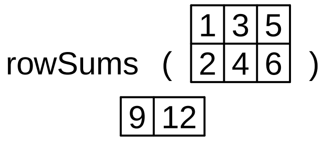
$$\text{length}\ (\ \begin{array}{|c|c|c|}\hline 1 & 3 & 5 \\ \hline 2 & 4 & 6 \\ \hline \end{array}\ ) \quad \boxed{6} \qquad \text{mode}\ (\ \begin{array}{|c|c|c|}\hline 1 & 3 & 5 \\ \hline 2 & 4 & 6 \\ \hline \end{array}\ ) \quad \boxed{\text{"numeric"}}$$

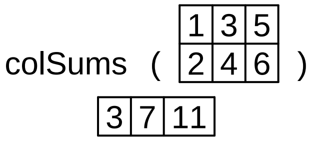
A matrix is very similar to a vector : it has a mode and a length. It has also two dimensions and, then, it has two vector of names and it takes two index vectors inside extraction operator. But you can often see a matrix as a vector. For instance, if you use only one index vector inside the extraction operator, it extracts from the underlying, column-filled vector :

$$\begin{array}{|c|c|c|}\hline 1 & 3 & 5 \\ \hline 2 & 4 & 6 \\ \hline \end{array}\ [\ \boxed{3}\ ] \quad \boxed{3}$$

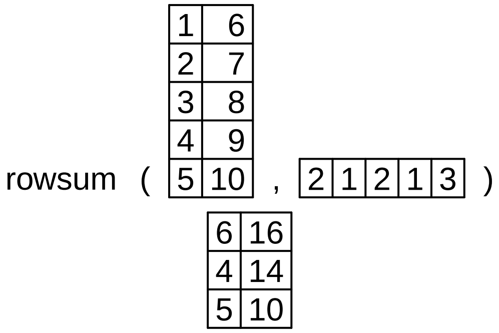
14.4 Summing a matrix

$$\text{sum}\ (\ \begin{array}{|c|c|c|}\hline 1 & 3 & 5 \\ \hline 2 & 4 & 6 \\ \hline \end{array}\ ) \quad \boxed{21}$$

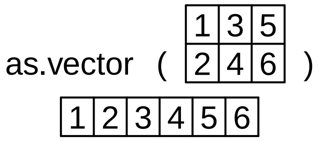


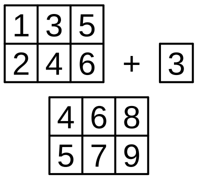


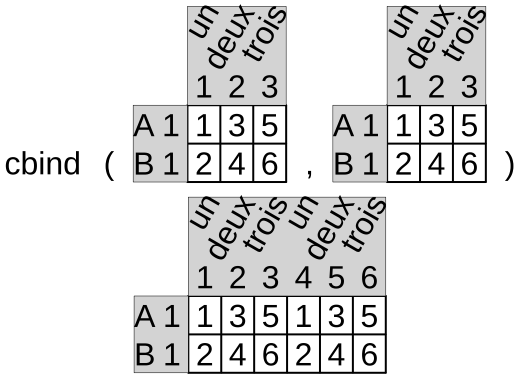
rowsum() performs a colSums on each group of rows given by the second argument.

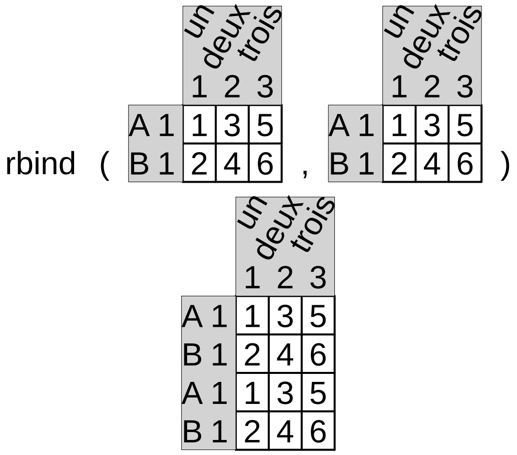


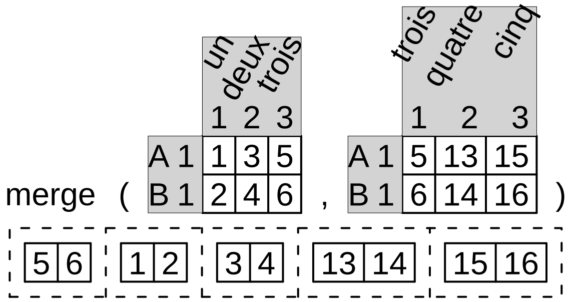
14.5 Changing

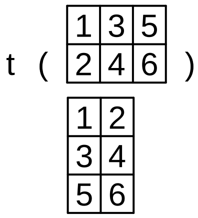










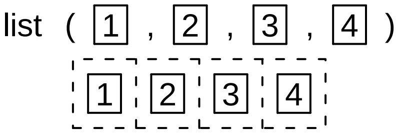


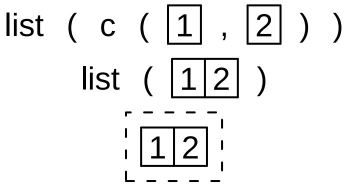
15 list

15.1 Creation, anatomy

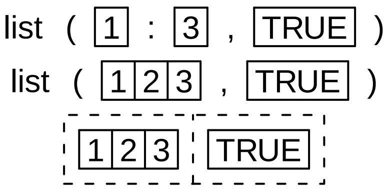
Creating a list by enumerating its components.

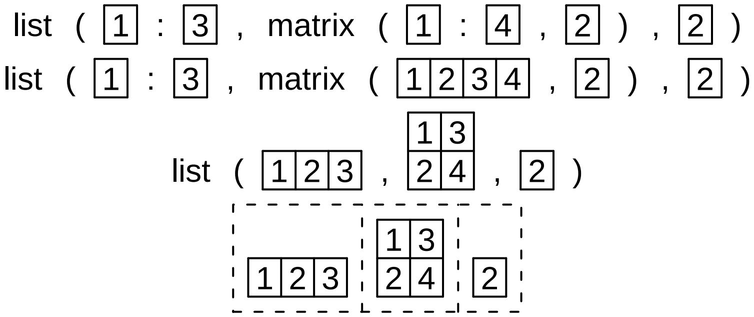
A list of length 4 : contains 4 vectors, each of length 1 (left) ; a list of length 1 : contains 1 vector of length 2 (right).



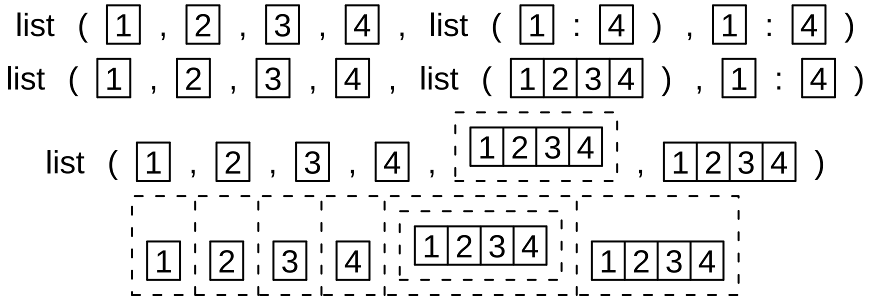


List can contain objects of different mode (left) ; it can contain objects of different dimensions (right).



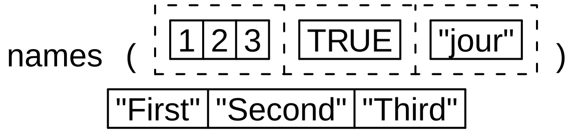

list (1 : 3 , matrix (1 : 4 , 2) , 2)
list (1 : 3 , matrix (1 2 3 4 , 2) , 2)
list (1 2 3 , 1 3 / 2 4 , 2)
1 2 3
1 3 / 2 4
2

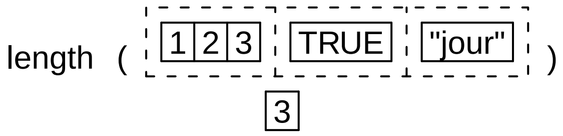
A list can be recursive : a component may be a list.

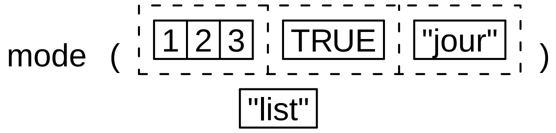

list (1 , 2 , 3 , 4 , list (1 : 4) , 1 : 4)
list (1 , 2 , 3 , 4 , list (1 2 3 4) , 1 : 4)
list (1 , 2 , 3 , 4 , 1 2 3 4 , 1 2 3 4)
1 2 3 4
1 2 3 4
1 2 3 4

15.2 Basic functions

TODO...

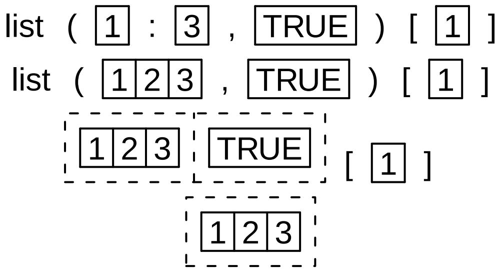


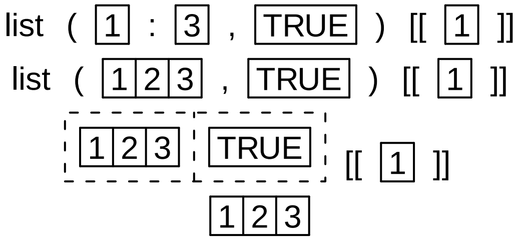




15.3 Extraction

Extracting in a list. The next two figures show the difference beween [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.

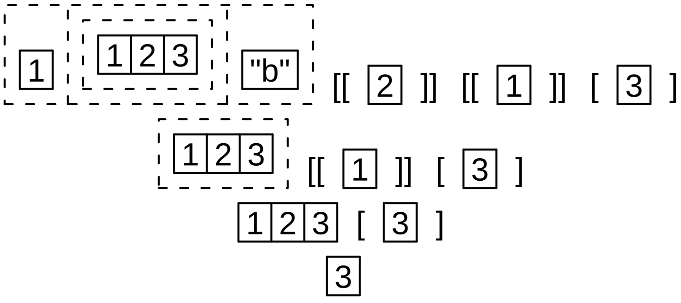




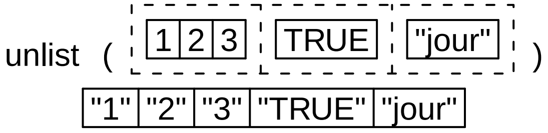
You can use vector of any length within the single-square bracket, while you can address only one element within the double-square-bracket operator, and then use only vector of length 1.

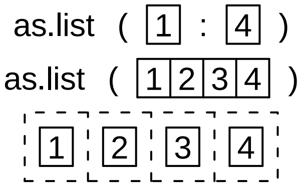
Since a list is a recursive data structure (may contain list), you can use several successive bracket operators in order to go down to the element you're interested in.

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



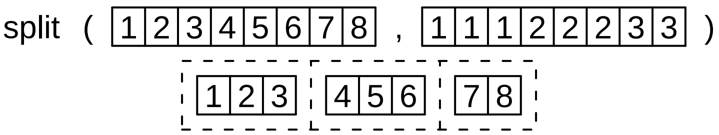
15.4 List and vector





15.5 List for expressing complex data structure

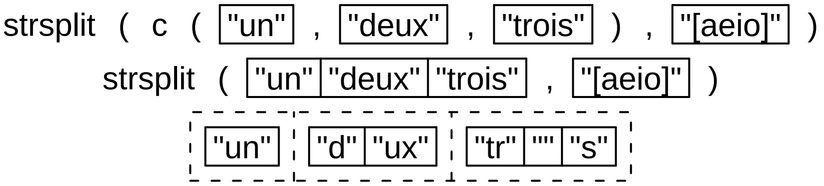
Grouping elements of a vector using a the level of a factors :

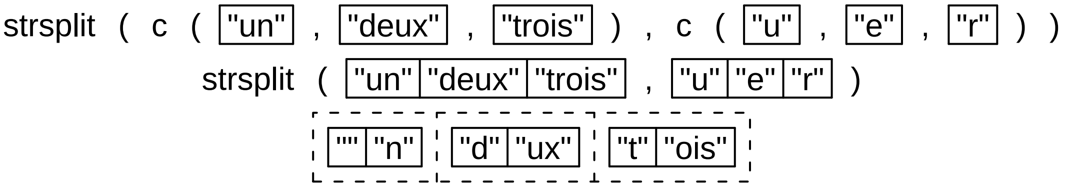


See also for instance strsplit() below.

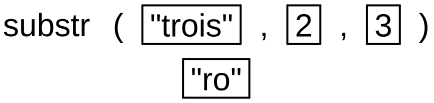
16 Regexp

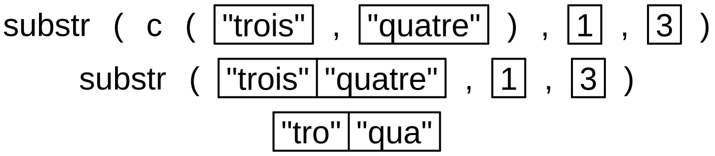
16.1 Split strings : strplit()

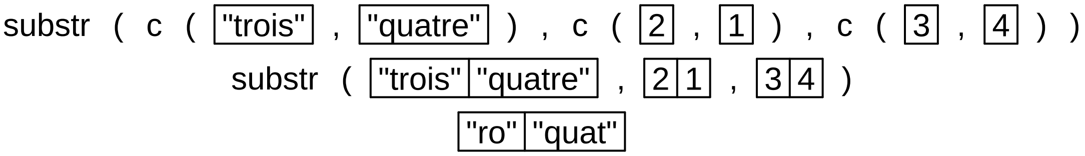

strsplit (c ("un" , "deux" , "trois") , "[aeio]")
strsplit ("un" "deux" "trois" , "[aeio]")
"un"
"d" "ux"
"tr" "" "s"


strsplit (c ("un" , "deux" , "trois") , c ("u" , "e" , "r"))
strsplit ("un" "deux" "trois" , "u" "e" "r")
"" "n"
"d" "ux"
"t" "ois"

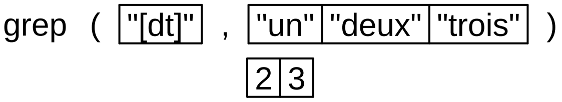
16.2 Extract sub strings : substr()

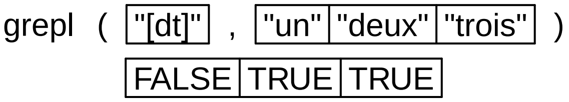



substr (c ("trois" , "quatre") , 1 , 3)
substr ("trois" "quatre" , 1 , 3)
"tro" "qua"


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()





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 : regexpr()

TODO

17 Data Frame

TODO