



## Estructuras de datos en R (vectores y dataframes)

### Vectores atómicos

Los *vectores* son las estructuras de datos más sencillas del lenguaje, sirven para almacenar uno o más datos de un solo tipo.

Todos los vectores tienen tres propiedades que los caracterizan:

1. **Tipo.** Los más comunes son *numéricos*, *lógicos*, *carácter* y *enteros*.
2. **Longitud.** La cantidad de datos que almacena. Se conoce igual como *dimensiones*
3. **Atributos.** Serie de *metadatos* que describen a un vector

### Creando vectores

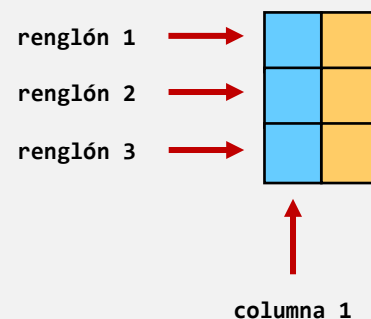
Para crear un vector, se emplea la función **c(x)** que significa concatenar, aunque hay muchas formas:

Función	Input (entrada)	Output (salida)
c(x)	c(1,3,5)	1 3 5
from:to	1:5	1 2 3 4 5
seq(...)	seq(5)	1 2 3 4 5
	seq(3,5)	3 4 5
	seq(1,5, by = 2)	1 3 5
rep(x, ...)	rep(2, times = 5)	1 1 1 1 1
	rep(1:2, times = 3)	1 2 1 2 1 2
	rep(1:3, each = 2)	1 1 2 2 3 3

Los puntos suspensivos indican una serie de argumentos distintos dentro de una función.

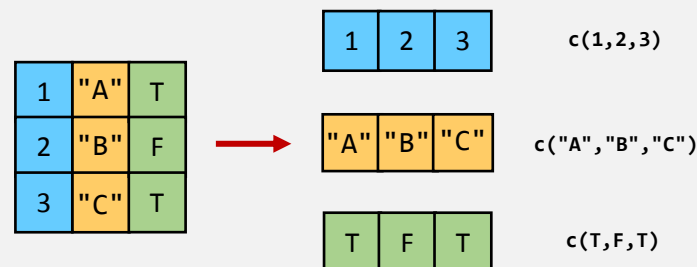
### Dataframes

Los *data frames* o *dataframes* son listas con un arreglo rectangular, tienen forma similar a una tabla y se componen de *dos dimensiones*, denotadas como *renglones* y *columnas*.



### Creando dataframes

Para crear un dataframe, se emplea la función **data.frame()**. En general, los dataframes se componen de vectores, y pueden ser de diferentes tipos (numérico, lógico, carácter, etc)



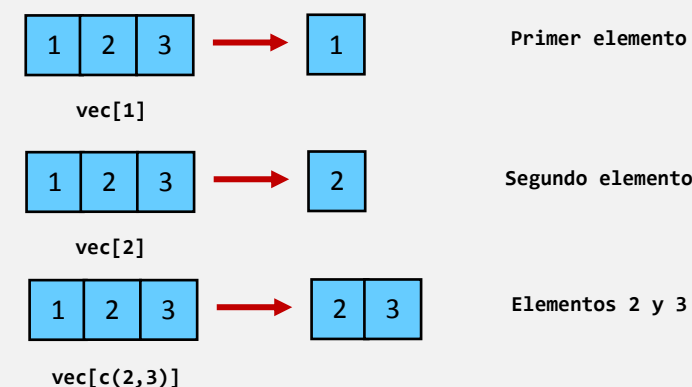
Al crear un dataframe, los vectores deben ser de la misma *longitud*.

### Extraer datos (subsetting)

El *subsetting* nos sirve para obtener datos de interés dentro de otro grupo más grande de información. En español es *subconjunto*.

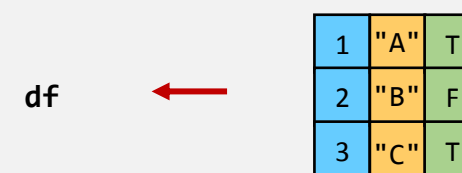
#### Subsetting de un vector

Los operadores de subconjunto (subsetting) son **[[**, **[** y **\$**. Para los vectores se emplean los corchetes simples (**[]**). Supóngase la creación de una variable llamada "vec" que contiene el 1, 2 y 3:



#### Subsetting de un dataframe

Se emplean los operadores **[**, **[[** y **\$** para extraer datos. Tomar de ejemplo el siguiente dataframe guardado en la variable "df":

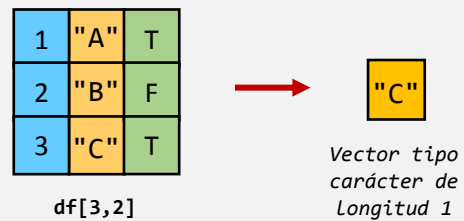
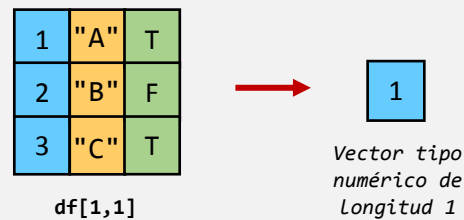




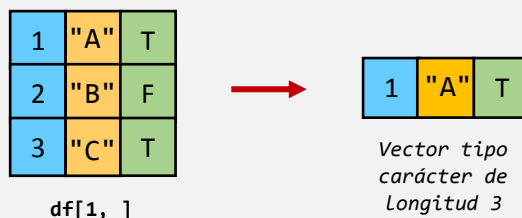
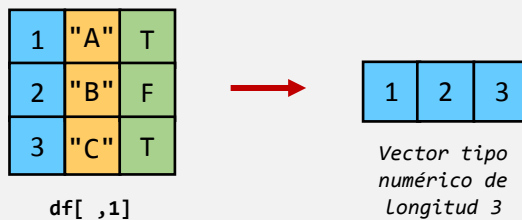
## Estructuras de datos en R (vectores y dataframes)

### Subsetting de un dataframe con [

En un dataframe, el operador debe indicar las dos dimensiones (renglón y columna), es decir: [renglón, columna]

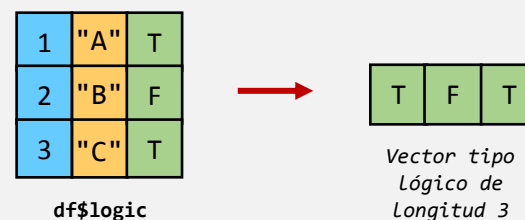
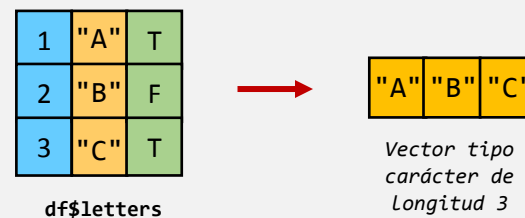
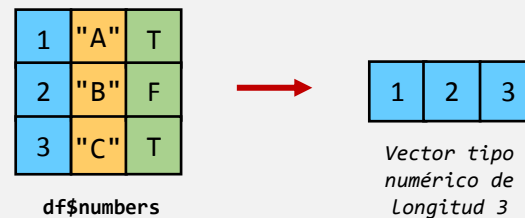
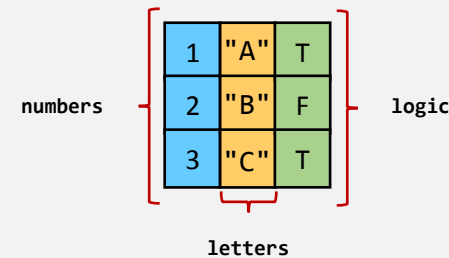


### Subsetting de renglones o columnas



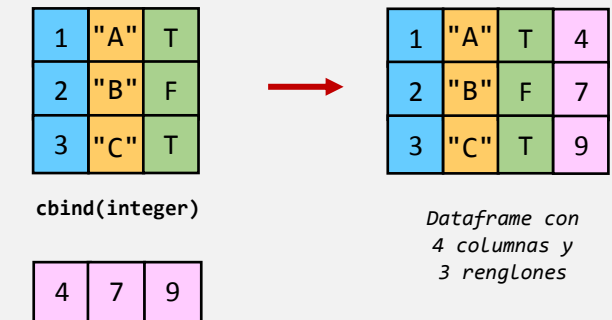
### Subsetting de un dataframe con \$

El símbolo \$ (dólar o peso) se usa para extraer columnas de manera rápida. La sintaxis es colocar el nombre del dataframe seguido del signo \$ y el nombre de la columna de interés:



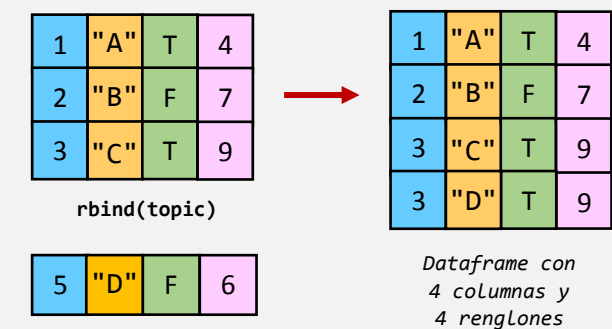
### Añadiendo columnas a un dataframe

Para añadir una columna a un dataframe se emplea la función **cbind(x)**, donde "x" es un vector de longitud similar al número de renglones del dataframe.



### Añadiendo renglones a un dataframe

Para añadir un renglón a un dataframe se emplea la función **rbind(x)**, donde "x" es un vector de longitud similar al número de columnas del dataframe.



Para saber el número de renglones se usa **nrow(x)**, columnas **ncol(x)**, y ambos **dim(x)**.