

Introducción a R

Matrices, dataframes y listas

Violeta Larios-Serrato

21 de octubre, 2016

Temario

- ▶ Matrices
- ▶ Dataframes
- ▶ Listas

Matrices

¿Qué es una matriz?

En R es una estructura de datos, ordenados por filas y columnas, llamada entonces una matriz de dos dimensiones.

Es una colección de elementos del mismo tipo de datos (numéricos, caracteres, o lógicos), organizados por filas y columnas.

Función

Comando:

```
matrix()
```



WELCOME TO
THE MATRIX!!!!!!

Ejemplo 1: Estructura de una matriz

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

```
##           [,1] [,2] [,3]  
## [1,]      NA  NA  NA  
## [2,]      NA  NA  NA
```

Donde:

nrow, número de filas.

ncol, número de columna.

NA, Not Available.

[fila,columna]

Argumentos separados por comas.

Ejercicio 2: Incorporación de datos a una matriz

```
matrix(1:9, byrow = TRUE, nrow = 3)
```

- ▶ El primer argumento es una colección de elementos.

```
1:9 c(1,2,4,5,6,7,8,9)
```

- ▶ El argumento **byrow** indica que la matriz será rellenada por fila, si queremos que la matriz sea llenada por columna ajustamos **byrow = FALSE**.
- ▶ El tercer argumento **nrow** indica que la matriz tendrá 3 filas.

Ejercicio 2: Incorporación de datos a una matriz

```
matrix(1:9, byrow = TRUE, nrow = 3)
```

Donde:

byrow=TRUE, es llenar por fila.

```
matrix(1:9, byrow = FALSE, nrow = 3)
```

Donde:

byrow=FALSE, es llenar por columna.

Ejercicio 3: Incorporación de datos a una matriz

```
MatrixEjemplo <- matrix (1:12, nrow = 4, byrow=T)  
  
dim(MatrixEjemplo)
```

```
## [1] 4 3
```

Donde:

byrow=T, es llenar con los elementos por fila.

dim(), es la función para obtener las dimensiones.

Ejercicio 4: Extracción de datos en una matriz

```
elemento <- MatrixEjemplo[3,2]
```

```
fila <- MatrixEjemplo[3,]
```

```
columna <- MatrixEjemplo[,2]
```

Ejercicio 5: Caso de estudio matrices

Vamos a construir una matrix con los años de publicación de algunos de los libros de Harry Potter (HP). El primer elemento es el año de publicación en inglés, y el segundo elemento es el año de publicación en español. *Fuente wikipedia

-Alohomora-

```
HPpiedra <- c(1997, 1993)
HPpazkaban <- c(1999, 2000)
HPcaliz <- c(2000, 2001)
HPprincipio <- c(2005, 2006)

cronologia <- c(HPpiedra, HPpazkaban, HPcaliz, HPprincipio)
```

Ejercicio 5: Caso de estudio matrices

Matriz de la cronología de publicación HP



Ejercicio 5: funciones utiles en matrices

```
HPmatrix <- matrix(cronologia, nrow=4)

matrixN <- rbind(HPmatrix, c(2007, 2008))

matrixN <- cbind(matrixN, c(2015, 2015, 2016, 2016, 2016))
```

Donde:

rbind, es una función para unir una fila extra.

cbind, es una función para unir una columna extra.

Dataframe

¿Qué es un dataframe?

Es una estructura, que permite el uso de diferentes tipos de datos. Combina las características de matrices y listas. Ejemplo, en una encuesta se tienen preguntas como las siguientes con sus respectivas respuestas:

Variable

Tipo de dato

¿Estas inscrito? : **si/no (lógicos)**.

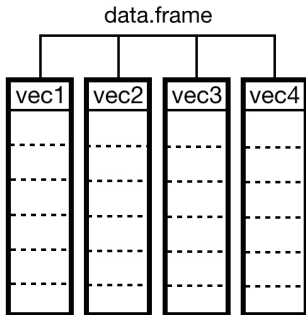
¿Año de nacimiento?: **1990, 1981, 2000, 1952 (numérico)**.

¿Nacionalidad? : **Afganistan, Brazil, China (caracteres)**.

Función

Comando

```
data.frame()
```



data.frame

country	year	cases	pop
Afghan	1999	745	19987071
Afghan	2000	2666	20595360
Brazil	1999	37737	172006362
Brazil	2000	80488	174504898
China	1999	212258	1272915272
China	2000	213766	1280428583

Son conjuntos con datos de diferente tipo.

Las variables son las columnas, o bien hay un vector en cada columna.

Las observaciones están en las filas.

Ejercicio 1: Construyendo un dataframe

```
Equipos <- c("Chiefs", "Cowboys", "Jets", "Patriots")  
SBganados <- c("0", "5", "1", "4")  
AnilloDeSB <- c(FALSE, TRUE, TRUE, TRUE)  
  
NFLdatos <- data.frame(Equipos, SBganados, AnilloDeSB)
```

Ejercicio 1: Construyendo un dataframe

```
NFLdatos <- data.frame(Equipos, SBganados, AnilloDeSB)  
NFLdatos
```

```
##      Equipos SBganados AnilloDeSB  
## 1   Chiefs          0      FALSE  
## 2 Cowboys          5       TRUE  
## 3    Jets           1       TRUE  
## 4 Patriots         4       TRUE
```


Ejercicio 3: Algunas funciones útiles para dataframe

Obteniendo estadísticas básicas de cada header.

```
summary(NFLdatos)
```

```
##      Equipos  SBganados  AnilloDeSB
## Chiefs   :1    0:1      Mode :logical
## Cowboys  :1    1:1      FALSE:1
## Jets     :1    4:1      TRUE  :3
## Patriots:1    5:1      NA's  :0
```

Donde:

summary(), es un función que brinda un resumen del objeto, en este caso muestra la frecuencia.

¿Qué es una lista?

- ▶ Los componentes de las lista no tienen por qué tener la misma cantidad de elementos.
- ▶ Una lista en R permite una variedad de objetos bajo el mismo nombre.
- ▶ Estos objetos pueden ser matrices, vectores, data frames.
- ▶ Podríamos decir que son un tipo de super datos.

Función

Para construirla usamos el comando:

```
list ()
```

Ejercicio 1: Construyendo una lista

```
vector <- 1:5  
matrix <- matrix(1:9, ncol = 3)  
dataframe <- NFLdatos  
  
lista <- list(vector, matrix, dataframe)
```

Ejercicio 1: Construyendo una lista

```
lista
```

```
## [[1]]
```

```
## [1] 1 2 3 4 5
```

```
##
```

```
## [[2]]
```

```
##      [,1] [,2] [,3]
```

```
## [1,]    1    4    7
```

```
## [2,]    2    5    8
```

```
## [3,]    3    6    9
```

```
##
```

```
## [[3]]
```

```
##      Equipos SBganados AnilloDeSB
```

```
## 1   Chiefs           0      FALSE
```

```
## 2  Cowboys           5       TRUE
```

```
## 3    Jets            1       TRUE
```

```
## 4 Patriots          4       TRUE
```

Ejercicio 2: Algunas funciones útiles para listas

Creando una lista de nombres.

```
names(lista) <- c("VectorName", "MatrixName", "DFName")
```

Donde:

names(), es una función para asignar nombres a un objeto.

Ejercicio 2: Algunas funciones útiles para listas

```
lista
```

```
## $VectorName
## [1] 1 2 3 4 5
##
## $MatrixName
##      [,1] [,2] [,3]
## [1,]    1    4    7
## [2,]    2    5    8
## [3,]    3    6    9
##
## $DFName
##      Equipos SBganados AnilloDeSB
## 1   Chiefs          0      FALSE
## 2  Cowboys          5       TRUE
## 3    Jets           1       TRUE
## 4 Patriots          4       TRUE
```

Ejercicio 3: Algunas funciones útiles para listas

Accediendo a una lista

Utilizando el nombre de la columna

```
lista$VectorName
```

```
## [1] 1 2 3 4 5
```


Ejercicio 3: Algunas funciones útiles para listas

Accediendo a una lista.

Usando un entero negativo para excluir el segundo componente

```
lista[-2]
```

```
## $VectorName
```

```
## [1] 1 2 3 4 5
```

```
##
```

```
## $DFName
```

```
##      Equipos SBganados AnilloDeSB
```

```
## 1    Chiefs          0      FALSE
```

```
## 2   Cowboys          5       TRUE
```

```
## 3      Jets          1       TRUE
```

```
## 4 Patriots          4       TRUE
```

Ejercicio 3: Algunas funciones útiles para listas

Modificando un componente de la lista.

```
lista$VectorName <- 8
```

```
lista$VectorName
```

```
## [1] 8
```

Ejercicio 3: Algunas funciones útiles para listas

Adicionando un componente una lista.

```
lista$VariableName <- FALSE
```

Ejercicio 3: Algunas funciones útiles para listas

Obtener información del objeto

```
str(lista)
```

```
## List of 4
## $ VectorName : num 8
## $ MatrixName : int [1:3, 1:3] 1 2 3 4 5 6 7 8 9
## $ DFName      : 'data.frame': 4 obs. of  3 variables:
## ..$ Equipos    : Factor w/ 4 levels "Chiefs","Cowboys",
## ..$ SBganados   : Factor w/ 4 levels "0","1","4","5": 1
## ..$ AnilloDeSB : logi [1:4] FALSE TRUE TRUE TRUE
## $ VariableName: logi FALSE
```

Donde:

str, es una función que da información de la estructura de algunos objetos.