

Tipos de datos

Sección 2: Tipos de datos

En esta unidad veremos los diferentes tipos de datos de R y como trabajar con ellos.

Tipo de dato	Descripción	Ejemplo
Numeric	Números decimales	1.0
Integer	Números enteros	1
Character	Cadenas de texto	'hola mundo'
Complex	Números complejos	3+2i
Logical	Verdadero o falso	TRUE / FALSE
Factor	Variable categórica	'B' < 'MBSTE' < 'STE'



Tipos de variables

- Las variables en R pueden ser números, textos, tablas, lista de números
- La manera mas común de almacenar datos en R es con Data Frames.
- Para saber el tipo de una variable podemos utilizar la función `class()`

```
> a<-2  
> class(a)  
[1] "numeric"  
> class(ls)  
[1] "function"
```

Data Frames

- Conceptualmente los Data Frames son tablas
- Las filas representan observaciones y las diferentes variables son representadas en columnas.
- La función `str()` nos muestra la estructura de un objeto

```
> library(dslabs)
> data(murders)
> str(murders)
'data.frame':  51 obs. of  5 variables:
 $ state      : chr  "Alabama" "Alaska" "Arizona" "Arkansas" ...
 $ abb        : chr  "AL" "AK" "AZ" "AR" ...
 $ region     : Factor w/ 4 levels "Northeast","South",...: 2 4 4 2 4
 4 1 2 2 2 ...
 $ population: num  4779736 710231 6392017 2915918 37253956 ...
 $ total      : num   135 19 232 93 1257 ...
> class(murders)
[1] "data.frame"
> head(murders)
```

	state	abb	region	population	total
1	Alabama	AL	South	4779736	135
2	Alaska	AK	west	710231	19

Diagram illustrating the structure of the `murders` data frame:

- The columns (state, abb, region, population, total) are labeled as **VARIABLES**.
- The rows (1, 2) are labeled as **OBSERVACIONES** (Observations).

Acceso a las variables del DataTable

- Para acceder a los valores de las variables utilizamos el símbolo \$ como \$nombreVariable:

```
> names(murders)
[1] "state"      "abb"        "region"     "population"
[5] "total"
> murders$state
[1] "Alabama"
[3] "Arizona"
[5] "California"
[7] "Connecticut"
[9] "District of Columbia"
[11] "Georgia"
[13] "Idaho"
[15] "Illinois"
```

- El orden de las entradas en nuestra variable es el mismo que el orden en el datatable original.

Vectores numéricos

- La variable `murders$population` no es solo un valor, son 51 valores.
- A este tipo de objetos se le llama vectores.

```
> length(murders$population)
[1] 51
> str(murders$population)
num [1:51] 4779736 710231 6392017 2915918 37253956 ...
```

- Los vectores numéricos son listas de uno o más números.

Vectores de caracteres

- Podemos almacenar caracteres en R

```
> #Variable numérica a con valor 1:  
> A <- 1  
> a  
[1] 1  
> #Carácter de texto a:  
> "a"  
[1] "a"
```


- Todos los elementos almacenados en un objeto character deben de ser caracteres.

```
> class(murders$state)  
[1] "character"
```

Vectores lógicos

- Son vectores de datos que pueden ser VERDADEROS o FALSOS

Operador relacional



```
> Z <- 3 == 2
> Z
[1] FALSE
> class(Z)
[1] "logical"
```


Otro tipo de datos: Factores

- Factores son utilizados para almacenar datos categorizados.

```
> class(murders$region)
[1] "factor"
> levels(murders$region)
[1] "Northeast"      "South"
[3] "North Central"  "West"
```

- Almacenar datos en categorías (niveles) maneja la memoria más eficientemente.
- En R los datos se almacenan en enteros, cada nivel se traduce de fondo en un entero que lo representa.

Versatilidad de R

- Múltiples maneras de acceder a las variables
- Podemos acceder de la manera

```
murders$population
```

- O mediante paréntesis:

```
murders[["population"]]
```

- Si en cambio queremos crear un nuevo data frame con la variable:

```
murders["population"]
```