

# Manipulación de datos

## Parte I

Christian Ballejo

03-03-2021

# Internet

Red: Aula alumnos - Clave: internet

## Diapositivas

<https://cballejo.github.io/Intro3>



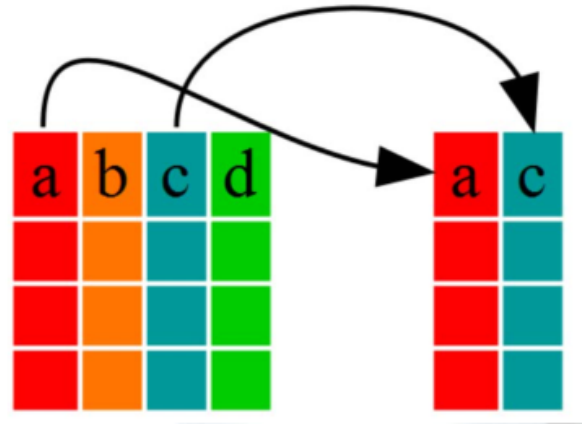
- **dplyr** es un paquete para **transformar datos** que pertenece al ecosistema **tidyverse**
- Implementa una **gramática "humana"**
- Está constituido por funciones definidas como **"verbos"**
- Hoy vamos a ver dos de esas funciones-verbos (**select** y **filter**)



# select()

La función **select()** del inglés **seleccionar**, sirve para seleccionar variables (columnas) de una tabla de datos.

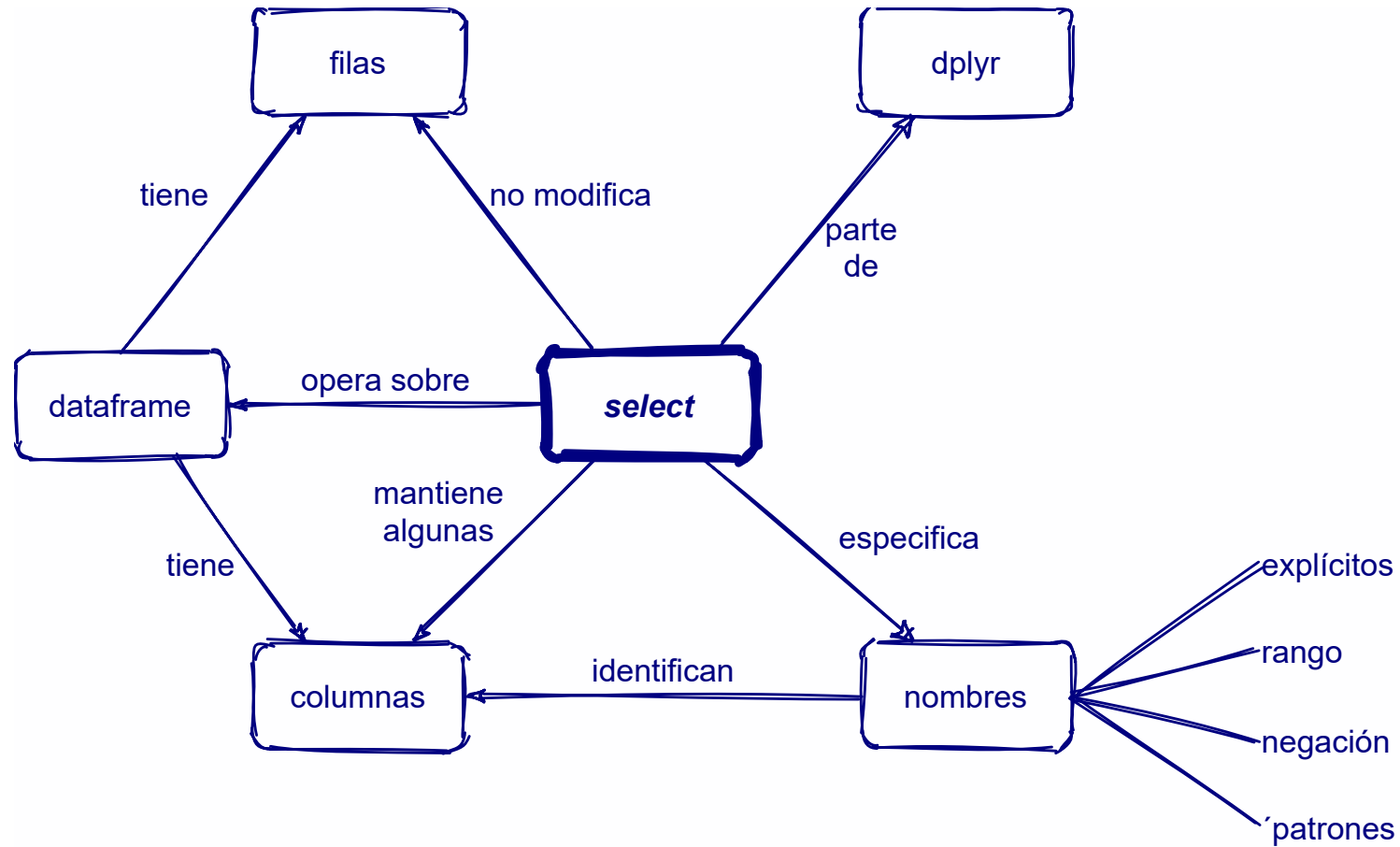
**select**(datos, a, c) ó  
datos %>% **select**(a,c)



**Abrimos el script practica3.R**

**Explicamos el funcionamiento de select() y sus variantes**

# Mapa conceptual de select()





# filter()

La función **filter()** del inglés **filtrar**, sirve para filtrar un subconjunto de observaciones (filas) de una tabla de datos a partir de una condición.

Para construir la condición se utilizan una serie de **operadores de comparación** y **operadores lógicos** similares a la de otros lenguajes de programación.

La estructura de la función puede ser cualquiera de las siguientes:

**filter**(datos, condición)

datos %>% **filter**(condición)



# Operadores de comparación

<code>x &lt; y</code>	Less than
<code>x &gt; y</code>	Greater than
<code>x == y</code>	Equal to
<code>x &lt;= y</code>	Less than or equal to
<code>x &gt;= y</code>	Greater than or equal to
<code>x != y</code>	Not equal to
<code>x %in% y</code>	Group membership
<code>is.na(x)</code>	Is NA
<code>!is.na(x)</code>	Is not NA

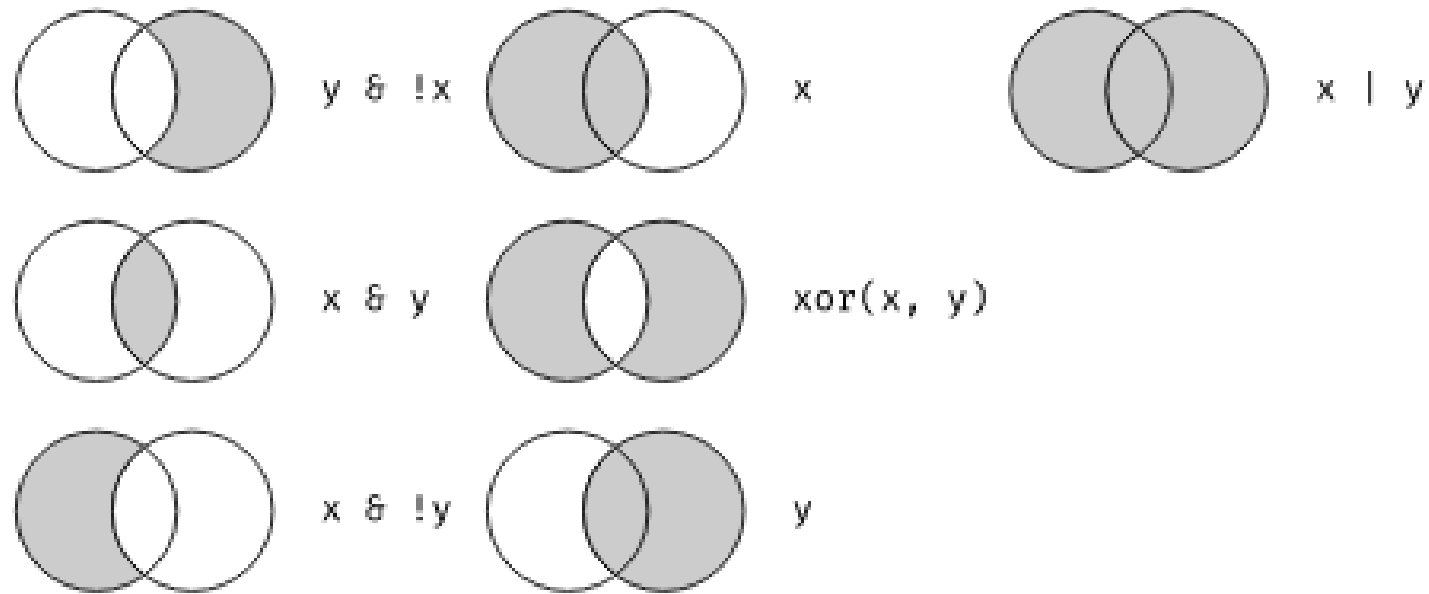




# Operadores lógicos (booleanos)

<code>a &amp; b</code>	and
<code>a   b</code>	or
<code>xor(a,b)</code>	exactly or
<code>!a</code>	not

# Operadores lógicos (booleanos)





# Tipos de datos

Tipos de datos posibles para variables con sus abreviaturas

- **int** números enteros
- **dbl** números reales
- **chr** cadena de caracteres (*texto*)
- **date** fechas
- **dtm** fechas y horas
- **lgl** valores lógicos (*TRUE* - *FALSE*)
- **fctr** factores (representan variables categóricas con valores fijos)



# Valores faltantes

- Los valores faltantes se implementan en **R** como un **valor especial NA**
- **NA** significa **Not Available** (No Disponible) y es un **"no valor"**.
- Esto hace que los valores perdidos sean "contagiosos". Cualquier operación que involucre un valor **NA** también será **NA**.

```
NA > 5
```

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

```
NA + 10
```

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

```
10 = NA
```

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

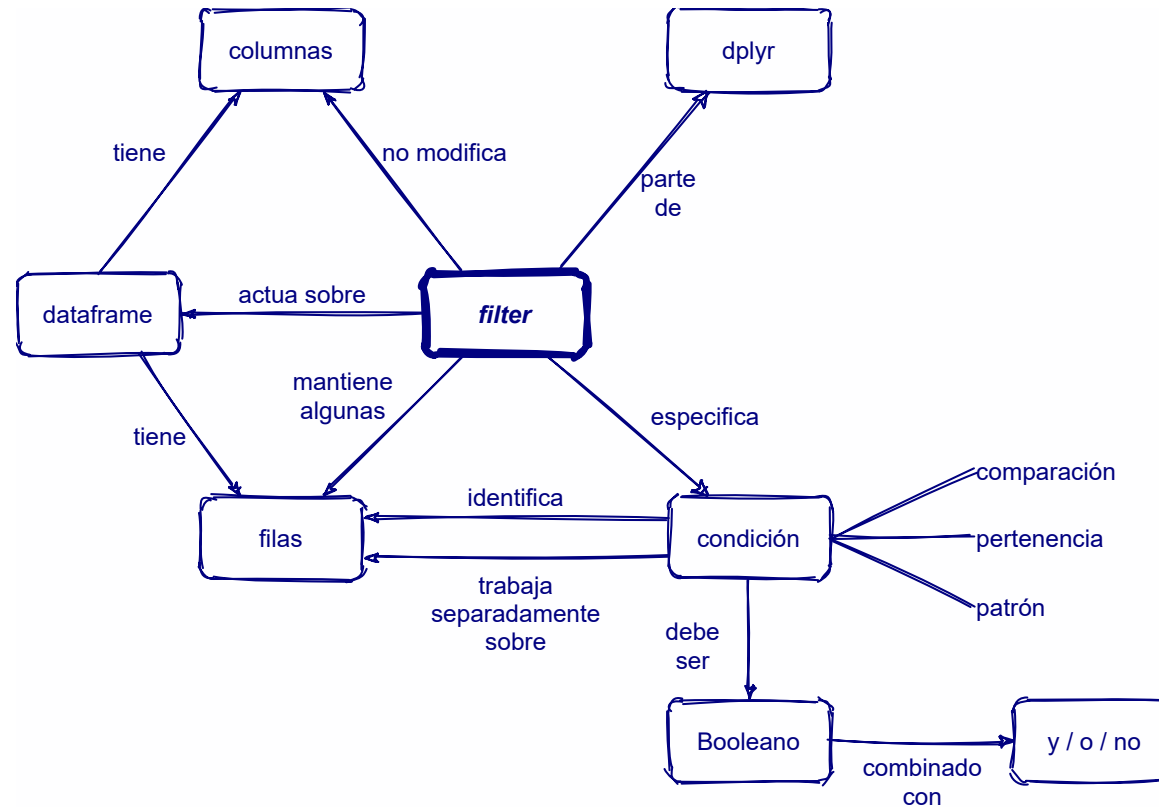
```
NA / 2
```

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

Volvemos al script practica3.R

Explicamos el funcionamiento de `filter()` y sus variantes, vinculando los subconjuntos de datos a salidas gráficas.

# Mapa conceptual de filter()





## Para profundizar...

Capítulo 5.1, 5.2 y 5.4 del libro "R para Ciencias de Datos"