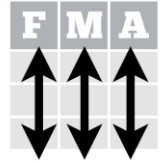


# Manipulación de tablas: : CHEAT SHEET



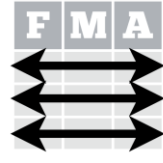
## Lo básico

### TIDYDATA



Cada variable es una columna.

&



Cada observación es una fila.

### SOBRE DPLYR



Es un paquete en R diseñado para facilitar la manipulación y análisis de datos. Se destaca por su sintaxis sencilla y eficiencia en el manejo de grandes volúmenes de información.

### USO DEL PIPE |>



Sirve para encadenar múltiples operaciones o funciones de manera secuencial. Permite pasar el resultado de una expresión como argumento a la siguiente, facilitando la lectura y escritura del código al evitar el anidamiento excesivo de funciones.

## Manipular filas

### FILTRAR CON FILTER()



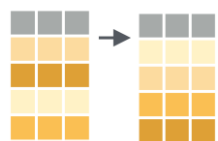
**data |> filter (condition)**

Extrae las filas que cumplen con una condición o criterio lógico.

#### Recuerda

Aquí puedes utilizar operadores lógicos como >, <, ==, entre otros!

### ORDENAR CON ARRANGE()



**data |> arrange (x)**

Ordena las filas según los valores de la columna x. Utiliza desc(x) cuando desees orden descendente.

## Manipular columnas

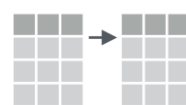
### SELECCIONAR CON SELECT()



**data |> select (x)**

Extrae la columna x como un dataframe o tibble.

### CREAR NUEVAS VARIABLES CON MUTATE()



**data |> mutate (x=fórmula)**

Crea una nueva columna x en base a una fórmula que se aplica sobre una o más variables existentes.

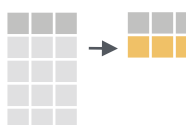
#### Recuerda

En el mutate() se aplican funciones que funcionan sobre vectores y nos devuelven vectores.



## Resumir

### APLICAR FUNCIONES A COLUMNAS



**data |> summarise (mean(x))**

Crea un estadístico de resumen a partir de los valores de una columna.

#### Recuerda

En el summarise() se aplican funciones de resumen que devuelven datos particulares.



## Agrupar casos

### GROUP\_BY()



Crear una copia "agrupada" de una tabla. Luego, cuando apliques funciones, estas manipularán cada "grupo" por separado y luego combinarán los resultados.

**data |>**

**group\_by(y) |>**

**summarise (mean(x))**

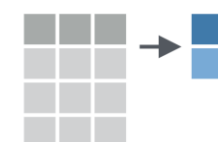
La agrupación se realiza en función de las categorías de la variable y; posteriormente, dentro de cada grupo, se calcula la media de los valores de x mediante la función mean(). Si la variable y consta de tres categorías distintas, R proporcionará tres medias correspondientes a los valores de x.

## Otras funciones



**data |> slice (5:10)**

Extrae filas dependiendo de su posición. Se utilizan índices.



**data |> count (x)**

Cuenta el número de filas en cada categoría de la variable x.

## Recodificar

### CASE\_WHEN()

```
case_when(condition-1 ~ output-value-1,
           condition-2 ~ output-value-2,
           condition-3 ~ output-value-3)
```

Formula for case 1

Formula for case 2

Formula for case 3

Es una función del paquete dplyr en R diseñada para facilitar la creación de nuevas variables o la transformación de las existentes basándose en una serie de condiciones lógicas.

La función permite especificar múltiples condiciones y los valores correspondientes que se deben asignar cuando dichas condiciones son verdaderas.

El símbolo ~ se escribe con ALT+126.

#### NOTA:

1. Cuidado con los nombres de variables y funciones: dplyr incluye varias funciones que pueden tener el mismo nombre que funciones base de R o de otros paquetes (como filter, select, etc.). Esto puede causar conflictos o comportamientos inesperados si no se especifica claramente de qué paquete se quiere usar la función.
2. Uso eficiente del pipe |>: Asegúrate de entender cómo fluyen los datos a través de las operaciones encadenadas para evitar errores lógicos.
3. Trabaja con grupos cuidadosamente: Cuando usas group\_by() para agrupar tu dataset, recuerda que muchas funciones de dplyr operarán en estos grupos de manera independiente. Esto es especialmente útil para operaciones sumarias o condicionales por grupo, pero debes recordar "desagrupar" tu data con ungroup() si ya no necesitas esta agrupación para operaciones subsiguientes, para evitar resultados inesperados.

