

# Notas dplyr

Carlos C.

2022-12-25

## 00. Acordeón

- `filter( condition )`: Permite hacer un filtrado de datos, devuelve aquellos registros que cumplan la condición impuesta
- `summarise( estadistico=function(x) )`: Muestra los estadísticos de resumen que se especifiquen para un conjunto de datos `x`
- `n()`: es una función que puede usarse dentro de `summarise()` y sirve para hacer un conteo de registros.
- `group_by( ~ )`: Podemos realizar agrupación de datos de un dataframe de acuerdo a las categorías de la variable `~`. Una buena práctica es usar `ungroup()` para que los resultados no aparezcan “agrupados”
- `mutate( nombre=operacion )`: Permite crear una nueva columna con un `nombre` dado y generada a partir de cierta `operacion`
- `arrange( ~ )`: Realiza un ordenamiento de los datos de acuerdo a la variable `~` (por ejemplo, mes). Se puede emplear una segunda variable que sirva como criterio de desempate.
- `inner_join( ~ )`: Permite realizar un inner join con la información de la tabla `~`, mismo razonamiento que en SQL. Si las “variables clave” no coinciden en nombre, se debe especificar con el argumento `by=c(a,b)` cuáles son el nombre de las variables.
- `rename()`: Realiza un cambio de nombre a una columna ya creada de un dataframe. La sintaxis que sigue es `nombre_nuevo=nombre_viejo`
- `select()`: Permite seleccionar algunas variables (columnas) de un dataframe para trabajar exclusivamente con ellas. Es posible cambiar el nombre de las columnas con esta función. Además, permite argumentos como `starts_with("a")`, `ends_with("e")` y `contains("palabra")`