



# Tidyverse (parte 2)

José Luis Texcalac Sangrador

Procesamiento y visualización de datos espaciales en R



# dplyr - manipulación datos

`mutate()`

`select()`

`filter()`

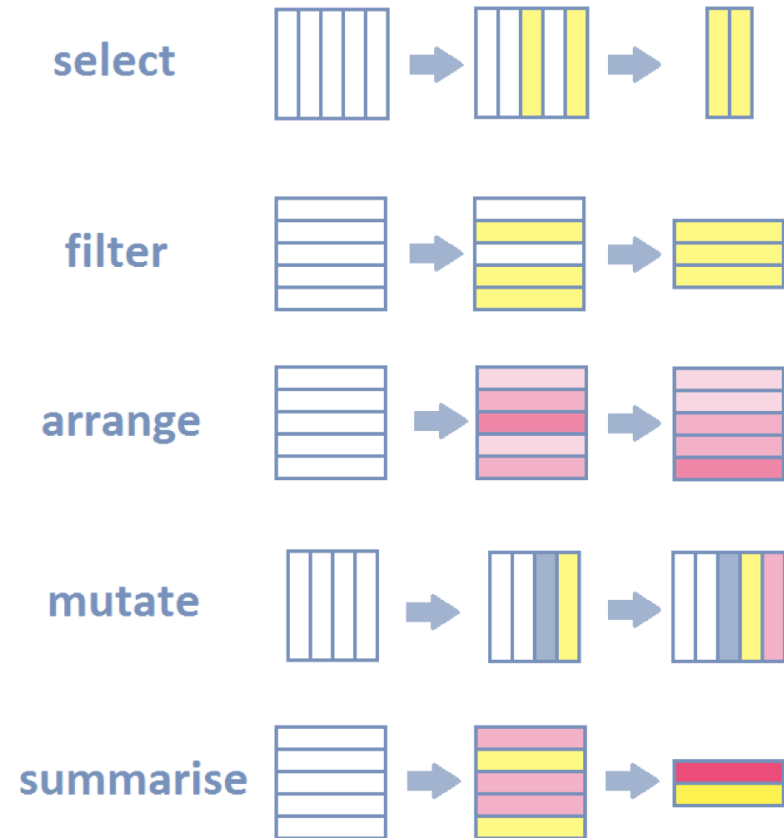
`group_by()`

`summarise()`

`arrange()`

`pivot_wide()` y `pivot_long()`

`join...`





# Malla starwars

```
> starwars
```

```
# A tibble: 87 × 14
```

	name	height	mass	hair_color	skin_color	eye_color	birth_year	sex	gender	homeworld	species	films	vehicles	starships
	<chr>	<int>	<dbl>	<chr>	<chr>	<chr>	<dbl>	<chr>	<chr>	<chr>	<chr>	<lis>	<list>	<list>
1	Luke Skywalker...	172	77	blond	fair	blue	19	male	mascu...	Tatooine	Human	<chr>	<chr>	<chr [2]>
2	C-3PO	167	75	NA	gold	yellow	112	none	mascu...	Tatooine	Droid	<chr>	<chr>	<chr [0]>
3	R2-D2	96	32	NA	white, bl...	red	33	none	mascu...	Naboo	Droid	<chr>	<chr>	<chr [0]>
4	Darth Vader	202	136	none	white	yellow	41.9	male	mascu...	Tatooine	Human	<chr>	<chr>	<chr [1]>
5	Leia Organa	150	49	brown	light	brown	19	fema...	femin...	Alderaan	Human	<chr>	<chr>	<chr [0]>
6	Owen Lars	178	120	brown, gr...	light	blue	52	male	mascu...	Tatooine	Human	<chr>	<chr>	<chr [0]>
7	Beru Whitesu...	165	75	brown	light	blue	47	fema...	femin...	Tatooine	Human	<chr>	<chr>	<chr [0]>
8	R5-D4	97	32	NA	white, red	red	NA	none	mascu...	Tatooine	Droid	<chr>	<chr>	<chr [0]>
9	Biggs Darkli...	183	84	black	light	brown	24	male	mascu...	Tatooine	Human	<chr>	<chr>	<chr [1]>
10	Obi-Wan Keno...	182	77	auburn, w...	fair	blue-gray	57	male	mascu...	Stewjon	Human	<chr>	<chr>	<chr [5]>

```
# i 77 more rows
```

```
# i Use `print(n = ...)` to see more rows
```



# distinct()

En ocasiones requerimos  
identificar los valores únicos de  
una variable

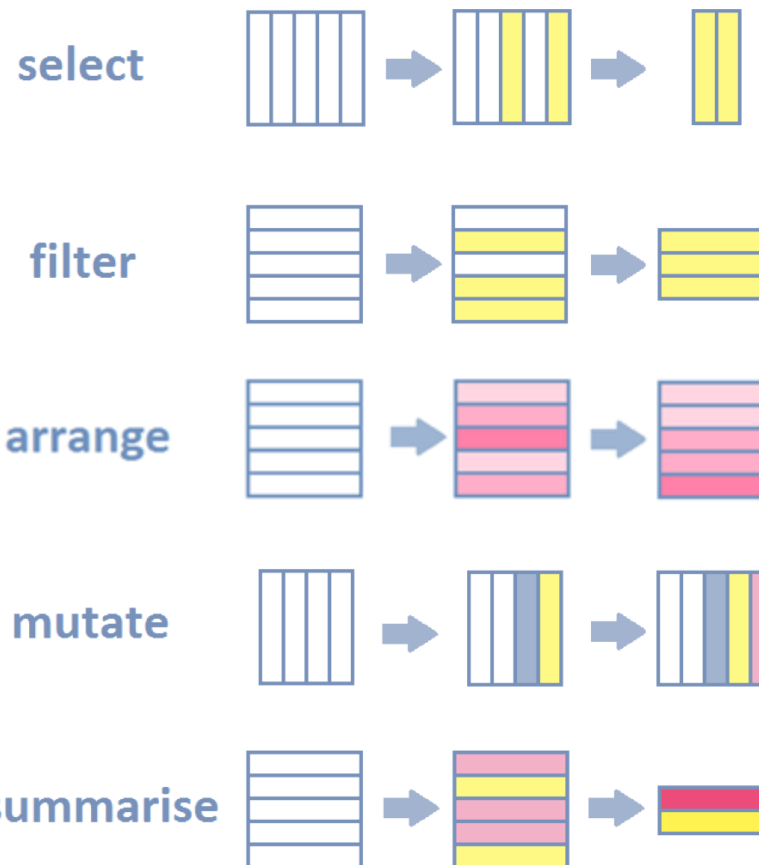
```
starwars %>%  
  distinct(eye_color)
```

```
# A tibble: 15 x 1  
  eye_color  
  <chr>  
1 blue  
2 yellow  
3 red  
4 brown  
5 blue-gray  
6 black  
7 orange  
8 hazel  
9 pink  
10 unknown  
11 red, blue  
12 gold  
13 green, yellow  
14 white  
15 dark
```



¿Cuántos tipos de especies hay en  
la malla **starwars**?

# arrange( )





# arrange()

Esta función la utilizamos para ordenar los valores de una variable.

```
starwars %>%  
  distinct(eye_color) %>%  
  arrange(eye_color)
```

```
# A tibble: 15 x :  
  eye_color  
  <chr>  
1 black  
2 blue  
3 blue-gray  
4 brown  
5 dark  
6 gold  
7 green, yellow  
8 hazel  
9 orange  
10 pink  
11 red  
12 red, blue  
13 unknown  
14 white  
15 yellow
```

# arrange()

```
starwars %>%  
  distinct(eye_color, sex)  
  arrange(eye_color)
```

```
# A tibble: 27 × 2  
  eye_color sex  
  <chr>      <chr>  
1 black      male  
2 black      female  
3 black      none  
4 blue       male  
5 blue       female  
6 blue       NA  
7 blue-gray  male  
8 brown      female  
9 brown      male  
10 brown     NA  
# i 17 more rows  
# i Use `print(n = ...)` to see more rows
```

```
starwars %>%  
  distinct(eye_color, sex) %>%  
  arrange(desc(eye_color))
```

```
# A tibble: 27 × 2  
  eye_color sex  
  <chr>      <chr>  
1 yellow     none  
2 yellow     male  
3 yellow     female  
4 white      NA  
5 unknown    male  
6 unknown    female  
7 red, blue  none  
8 red        none  
9 red        male  
10 pink       male  
# i 17 more rows  
# i Use `print(n = ...)` to see more rows
```





# Su turno...

- Trabaje con el objeto **ozono**
  - Ordene de forma descendente la columna **uax**.
    - ¿cuál fue la concentración más alta?
    - ¿de qué fecha y hora fue la concentración más alta?
  - ¿cuál es la fecha más antigua de la malla?
  - ¿qué estación o estaciones tienen la concentración más baja?





# relocate()

```
starwars %>%  
  relocate(species, .before = height)
```

```
# A tibble: 87 × 14
```

	name	species	height	mass	hair_color	skin_color
	<chr>	<chr>	<int>	<dbl>	<chr>	<chr>
1	Luke Skywalker...	Human	172	77	blond	fair
2	C-3PO	Droid	167	75	NA	gold
3	R2-D2	Droid	96	32	NA	white, bl...
4	Darth Vader	Human	202	136	none	white
5	Leia Organa	Human	150	49	brown	light
6	Owen Lars	Human	178	120	brown, gr...	light
7	Beru Whitesu...	Human	165	75	brown	light
8	R5-D4	Droid	97	32	NA	white, red
9	Biggs Darkli...	Human	183	84	black	light
10	Obi-Wan Keno...	Human	182	77	auburn, w...	fair

```
# i 77 more rows
```

```
# i 8 more variables: eye_color <chr>, birth_year <dbl>,
```

```
# sex <chr>, gender <chr>, homeworld <chr>,
```

```
# films <list>, vehicles <list>, starships <list>
```

```
# i Use `print(n = ...)` to see more rows
```

```
starwars %>%  
  relocate(name, .after = mass)
```

```
# A tibble: 87 × 14
```

	height	mass	name	hair_color	skin_color	eye_color
	<int>	<dbl>	<chr>	<chr>	<chr>	<chr>
1	172	77	Luke Skywa...	blond	fair	blue
2	167	75	C-3PO	NA	gold	yellow
3	96	32	R2-D2	NA	white, bl...	red
4	202	136	Darth Vader	none	white	yellow
5	150	49	Leia Organa	brown	light	brown
6	178	120	Owen Lars	brown, gr...	light	blue
7	165	75	Beru White...	brown	light	blue
8	97	32	R5-D4	NA	white, red	red
9	183	84	Biggs Dark...	black	light	brown
10	182	77	Obi-Wan Ke...	auburn, w...	fair	blue-gray

```
# i 77 more rows
```

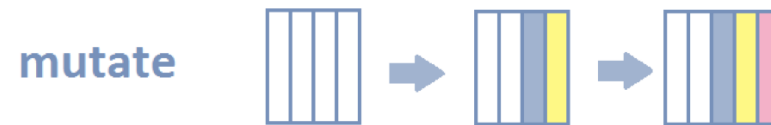
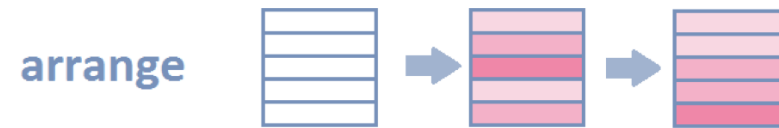
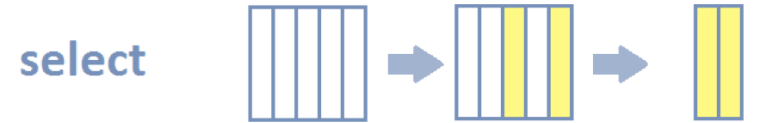
```
# i 8 more variables: birth_year <dbl>, sex <chr>,
```

```
# gender <chr>, homeworld <chr>, species <chr>,
```

```
# films <list>, vehicles <list>, starships <list>
```

```
# i Use `print(n = ...)` to see more rows
```

# filter( )



# Filtrar valores – filter( )

data %>% filter(...)

dataset

Argumentos de filtrado

$x < y$	menor qué
$x > y$	mayor qué
$x == y$	igual a
$x <= y$	menor o igual a
$x >= y$	mayor o igual a
$x != y$	diferente de
$x \%in\% y$	pertenece a
<code>is.na(x)</code>	is NA
<code>!is.na(x)</code>	Distinto de NA



# filter()

```
starwars %>%  
  filter(eye_color == "blue")
```

# A tibble: 19 x 14

name	height	mass	hair_color	skin_color	eye_color	birth_year	sex	gender	homeworld	species	films	vehicles	starships
<chr>	<int>	<dbl>	<chr>	<chr>	<chr>	<dbl>	<chr>	<chr>	<chr>	<chr>	<list>	<list>	<list>
1 Luke Skywalker	172	77	blond	fair	blue	19	male	masculine	Tatooine	Human	<chr [5]>	<chr [2]>	<chr [2]>
2 Owen Lars	178	120	brown, grey	light	blue	52	male	masculine	Tatooine	Human	<chr [3]>	<chr [0]>	<chr [0]>
3 Beru Whitesun Lars	165	75	brown	light	blue	47	female	feminine	Tatooine	Human	<chr [3]>	<chr [0]>	<chr [0]>
4 Anakin Skywalker	188	84	blond	fair	blue	41.9	male	masculine	Tatooine	Human	<chr [3]>	<chr [2]>	<chr [3]>
5 Wilhuff Tarkin	180	NA	auburn, grey	fair	blue	64	male	masculine	Eriadu	Human	<chr [2]>	<chr [0]>	<chr [0]>
6 Chewbacca	228	112	brown	unknown	blue	200	male	masculine	Kashyyyk	Wookiee	<chr [5]>	<chr [1]>	<chr [2]>
7 Jek Tono Porkins	180	110	brown	fair	blue	NA	NA	NA	Bestine IV	NA	<chr [1]>	<chr [0]>	<chr [1]>
8 Lobot	175	79	none	light	blue	37	male	masculine	Bespin	Human	<chr [1]>	<chr [0]>	<chr [0]>
9 Mon Mothma	150	NA	auburn	fair	blue	48	female	feminine	Chandriga	Human	<chr [1]>	<chr [0]>	<chr [0]>
10 Qui-Gon Jinn	193	89	brown	fair	blue	92	male	masculine	NA	Human	<chr [1]>	<chr [1]>	<chr [0]>
11 Finis Valorum	170	NA	blond	fair	blue	91	male	masculine	Coruscant	Human	<chr [1]>	<chr [0]>	<chr [0]>
12 Ric Olié	183	NA	brown	fair	blue	NA	male	masculine	Naboo	Human	<chr [1]>	<chr [0]>	<chr [1]>
13 Adi Gallia	184	50	none	dark	blue	NA	female	feminine	Coruscant	Tholothian	<chr [2]>	<chr [0]>	<chr [0]>
14 Mas Amedda	196	NA	none	blue	blue	NA	male	masculine	Champaia	Chagrian	<chr [2]>	<chr [0]>	<chr [0]>
15 Cliegg Lars	183	NA	brown	fair	blue	82	male	masculine	Tatooine	Human	<chr [1]>	<chr [0]>	<chr [0]>
16 Luminara Unduli	170	56.2	black	yellow	blue	58	female	feminine	Mirial	Mirialan	<chr [2]>	<chr [0]>	<chr [0]>
17 Barriss Offee	166	50	black	yellow	blue	40	female	feminine	Mirial	Mirialan	<chr [1]>	<chr [0]>	<chr [0]>
18 Jocasta Nu	167	NA	white	fair	blue	NA	female	feminine	Coruscant	Human	<chr [1]>	<chr [0]>	<chr [0]>
19 Tarfful	234	136	brown	brown	blue	NA	male	masculine	Kashyyyk	Wookiee	<chr [1]>	<chr [0]>	<chr [0]>



# filter()

Filtre a los personajes de piel blanca y obtenga el siguiente resultado

```
starwars %>%  
  filter(skin_color == "white") %>%  
  select(name, skin_color)
```

```
# A tibble: 2 × 2  
  name      skin_color  
  <chr>    <chr>  
1 Darth Vader white  
2 Yarael Poof white
```



```
starwars %>%
```

```
  filter(hair_color == "brown" & sex == "male") %>%
```

```
  select(name, homeworld)
```

```
# A tibble: 11 x 2
```

	name	homeworld
	<chr>	<chr>
1	Chewbacca	Kashyyyk
2	Han Solo	Corellia
3	Wedge Antilles	Corellia
4	Arvel Crynyd	NA
5	Wicket Systri Warrick	Endor
6	Qui-Gon Jinn	NA
7	Ric Olié	Naboo
8	Cliegg Lars	Tatooine
9	Tarfful	Kashyyyk
10	Raymus Antilles	Alderaan
11	Poe Dameron	NA



```
starwars %>%  
  filter(species == "human" | species == "droid") %>%  
  select(name, homeworld)  
  
# A tibble: 0 × 2  
#   i 2 variables: name <chr>, homeworld <chr>
```

¿Por qué este resultado?



```
starwars %>%  
  filter(species == "Human" | species == "Droid") %>%  
  select(name, species, homeworld)
```

```
# A tibble: 41 x 3
```

	name <chr>	species <chr>	homeworld <chr>
1	Luke Skywalker	Human	Tatooine
2	C-3PO	Droid	Tatooine
3	R2-D2	Droid	Naboo
4	Darth Vader	Human	Tatooine
5	Leia Organa	Human	Alderaan
6	Owen Lars	Human	Tatooine
7	Beru Whitesun Lars	Human	Tatooine
8	R5-D4	Droid	Tatooine
9	Biggs Darklighter	Human	Tatooine
10	Obi-Wan Kenobi	Human	Stewjon

```
# i 31 more rows  
# i Use `print(n = ...)` to see more rows
```

```
starwars %>%  
  filter(height >= 100) %>%  
  select(name, height, species, homeworld)
```

```
# A tibble: 74 × 4
```

	name <chr>	height <int>	species <chr>	homeworld <chr>
1	Luke Skywalker	172	Human	Tatooine
2	C-3PO	167	Droid	Tatooine
3	Darth Vader	202	Human	Tatooine
4	Leia Organa	150	Human	Alderaan
5	Owen Lars	178	Human	Tatooine
6	Beru Whitesun Lars	165	Human	Tatooine
7	Biggs Darklighter	183	Human	Tatooine
8	Obi-Wan Kenobi	182	Human	Stewjon
9	Anakin Skywalker	188	Human	Tatooine
10	Wilhuff Tarkin	180	Human	Eriadu

```
# i 64 more rows
```

```
# i Use `print(n = ...)` to see more rows
```

```
starwars %>%
```

```
  filter(height >= 170 & height <= 180) %>%
```

```
  select(name, height, species, homeworld)
```

```
# A tibble: 20 × 4
```

	name <chr>	height <int>	species <chr>	homeworld <chr>
1	Luke Skywalker	172	Human	Tatooine
2	Owen Lars	178	Human	Tatooine
3	Wilhuff Tarkin	180	Human	Eriadu
4	Han Solo	180	Human	Corellia
5	Greedo	173	Rodian	Rodia
6	Jabba Desilijic Tiure	175	Hutt	Nal Hutta
7	Wedge Antilles	170	Human	Corellia
8	Jek Tono Porkins	180	NA	Bestine IV
9	Palpatine	170	Human	Naboo
10	Lando Calrissian	177	Human	Socorro
11	Lobot	175	Human	Bespin
12	Ackbar	180	Mon Calamari	Mon Cala
13	Finis Valorum	170	Human	Coruscant
14	Darth Maul	175	Zabrax	Dathomir
15	Bib Fortuna	180	Twi'lek	Ryloth
16	Ayla Secura	178	Twi'lek	Ryloth
17	Eeth Koth	171	Zabrax	Iridonia
18	Luminara Unduli	170	Mirialan	Mirial
19	Shaak Ti	178	Togruta	Shili
20	Sly Moore	178	NA	Umbara

Queremos obtener una malla de datos con personajes con altura de 170 o más, que además sean humanos y con las siguientes columnas: name, height, species, homeworld.

```
starwars %>%  
  filter(height >= 170 & species == "Human") %>%  
  select(name, height, species, homeworld)
```

```
# A tibble: 24 x 4
```

	name <chr>	height <int>	species <chr>	homeworld <chr>
1	Luke Skywalker	172	Human	Tatooine
2	Darth Vader	202	Human	Tatooine
3	Owen Lars	178	Human	Tatooine
4	Biggs Darklighter	183	Human	Tatooine
5	Obi-Wan Kenobi	182	Human	Stewjon
6	Anakin Skywalker	188	Human	Tatooine
7	Wilhuff Tarkin	180	Human	Eriadu
8	Han Solo	180	Human	Corellia
9	Wedge Antilles	170	Human	Corellia
10	Palpatine	170	Human	Naboo

```
# i 14 more rows
```

```
# i Use `print(n = ...)` to see more rows
```

```
starwars %>%
  filter(is.na(height))
```

```
# A tibble: 6 × 14
```

	name	height	mass	hair_color	skin_color	eye_color	birth_year	sex	gender	homeworld
	<chr>	<int>	<dbl>	<chr>	<chr>	<chr>	<dbl>	<chr>	<chr>	<chr>
1	Arve...	NA	NA	brown	fair	brown	NA	male	mascu...	NA
2	Finn	NA	NA	black	dark	dark	NA	male	mascu...	NA
3	Rey	NA	NA	brown	light	hazel	NA	fema...	femin...	NA
4	Poe ...	NA	NA	brown	light	brown	NA	male	mascu...	NA
5	BB8	NA	NA	none	none	black	NA	none	mascu...	NA
6	Capt...	NA	NA	none	none	unknown	NA	fema...	femin...	NA

```
# i 4 more variables: species <chr>, films <list>, vehicles <list>, starships <list>
```

```
starwars %>%  
  filter(is.na(height) & sex == "male")
```

```
# A tibble: 3 × 14
```

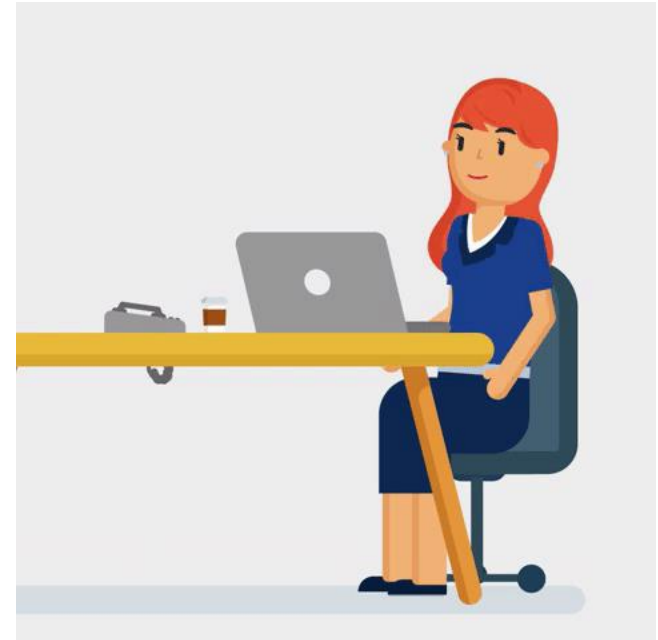
	name	height	mass	hair_color	skin_color	eye_color	birth_year	sex	gender	homeworld
	<chr>	<int>	<dbl>	<chr>	<chr>	<chr>	<dbl>	<chr>	<chr>	<chr>
1	Arve...	NA	NA	brown	fair	brown	NA	male	mascu...	NA
2	Finn	NA	NA	black	dark	dark	NA	male	mascu...	NA
3	Poe ...	NA	NA	brown	light	brown	NA	male	mascu...	NA

```
# i 4 more variables: species <chr>, films <list>, vehicles <list>, starships <list>
```



# Su turno...

- Trabaje con el objeto **ozono**
  - Posicione a las columnas **fecha**, **hora** y **uax** al inicio de la malla y ordene alfabéticamente al resto de las columnas
  - Guarde el resultado como **ozono**
- Continúe trabajando con la malla **ozono**
  - Seleccione las tres primeras columnas de la malla
  - Filtre los valores por arriba del valor de la norma horaria (90 ppb)
  - Guarde el resultado como **ozono\_nom**
- ¿Cuántos registros nos quedan?





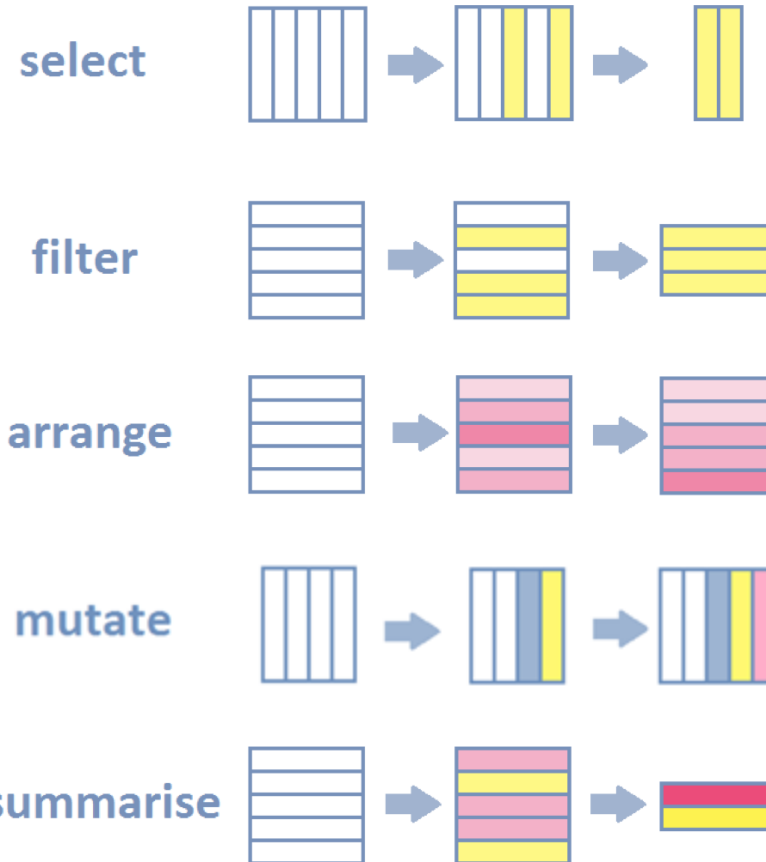
# Su turno...

- Trabaje con el objeto **ozono**
  - Mueva la columna uax a la última posición
  - Guarde el resultado como **ozono**





# mutate( )



# Generar variables – mutate()

```
mall %>% mutate(...)
```

dataset

Argumentos de creación de variables

mall

```
# A tibble: 8 × 3  
  nombre programa edad  
  <chr>   <chr>   <int>  
1 Juan    MC        30  
2 Eva     MC        25  
3 Ana     MSP       25  
4 Sofía   MSP       29  
5 Pedro   MSP       26  
6 Olivia  MC        30  
7 Mario   MSP       29  
8 David   MSP       25
```

```
mall %>%  
  mutate(edad_sqrt = sqrt(edad))
```

```
# A tibble: 8 × 4  
  nombre programa edad edad_sqrt  
  <chr>   <chr>   <dbl>   <dbl>  
1 Juan    MC        30      5.48  
2 Eva     MC        25      5  
3 Ana     MSP       25      5  
4 Sofía   MSP       29      5.39  
5 Pedro   MSP       26      5.10  
6 Olivia  MC        30      5.48  
7 Mario   MSP       29      5.39  
8 David   MSP       25      5
```

# mutate()

```
starwars %>%  
  mutate(altura = height/100,  
         desv_mass = mass - mean(mass, na.rm = TRUE),  
         log_peso = log(mass)) %>%  
  select(name, height, altura, mass, desv_mass, log_peso)
```

```
# A tibble: 87 × 5
```

	name	height	altura	mass	log_peso
	<chr>	<int>	<dbl>	<dbl>	<dbl>
1	Luke Skywalker	172	1.72	77	4.34
2	C-3P0	167	1.67	75	4.32
3	R2-D2	96	0.96	32	3.47
4	Darth Vader	202	2.02	136	4.91
5	Leia Organa	150	1.5	49	3.89
6	Owen Lars	178	1.78	120	4.79
7	Beru Whitesun Lars	165	1.65	75	4.32
8	R5-D4	97	0.97	32	3.47
9	Biggs Darklighter	183	1.83	84	4.43
10	Obi-Wan Kenobi	182	1.82	77	4.34

```
# i 77 more rows
```

```
# i Use `print(n = ...)` to see more rows
```

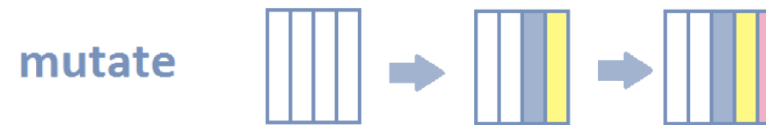
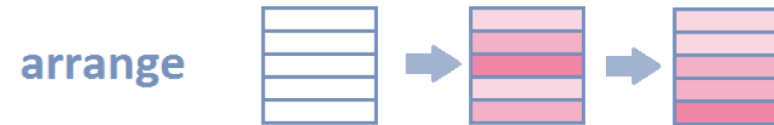
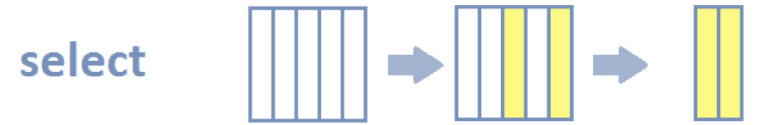


# Su turno...

- Trabaje con el objeto `ozono_nom`
  - Genere columna `uax_ppm` con la concentración de ozono en partes por billón.
  - Genere columna `uax_ugm3` con la concentración en microgramos por metro cúbico.
  - Guarde el resultado como `ozono_nom`



# summarise( )



# summarise()

```
starwars %>%  
  summarise(mean_mass = mean(mass, na.rm = TRUE),  
            sd_height = sd(height, na.rm = TRUE))
```

# A tibble: 1 × 2

	mean_mass	sd_height
	<dbl>	<dbl>
1	97.3	34.8



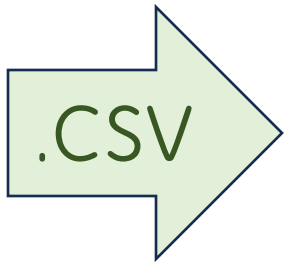
# Su turno...

- Trabaje con el objeto `ozono`
  - Genere el promedio y desviación estándar de cada columna.
  - Nombre cada columna agregando al nombre original de la columna el sufijo `_mean` para la media y `_sd` para la desviación estándar.
  - Guarde el resultado como `ozono_summary`



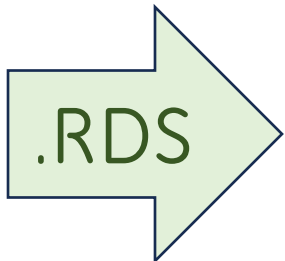


# Exportar malla de datos



```
write_excel_csv(malla, "ruta y nombre de archivo.csv")
```

```
write_excel_csv(malla, "./data/malla.csv")
```



```
write_rds(malla, "ruta y nombre de archivo.rds")
```

```
write_rds(malla, "./data/malla.rds")
```





# Su turno. ..

- Exporte el objeto `ozono`
- Guarde el archivo como `ozono_wide.rds`
- Continúe en su script
- Importe el archivo `ozono_wide.rds` y nombre al objeto como `ozono_wide`

