

Procesamiento de datos {dplyr}

José Luis Texcalac Sangrador

Procesamiento y visualización de datos espaciales en R



clase_r

⊕ New ▾ | 📁 Open ▾ | 💾 📄 | 🔍 Search ▾ | R 4.5.1 | 📁 clase_r ▾

CONSOLE | TERMINAL | PROBLEMS | OUTPUT | ... | + ▾ | - | × | 📄 | 📄 | ☰

R 4.5.1 started.

R version 4.5.1 (2025-06-13) -- "Great Square Root"
Copyright (C) 2025 The R Foundation for Statistical Computing
Platform: aarch64-apple-darwin20

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

>

Abra Positron y verifique que está dentro de su folder de trabajo **clase_r**

Si no está dentro de su folder entonces haga clic en la barra de menú **File/Open Folder...**

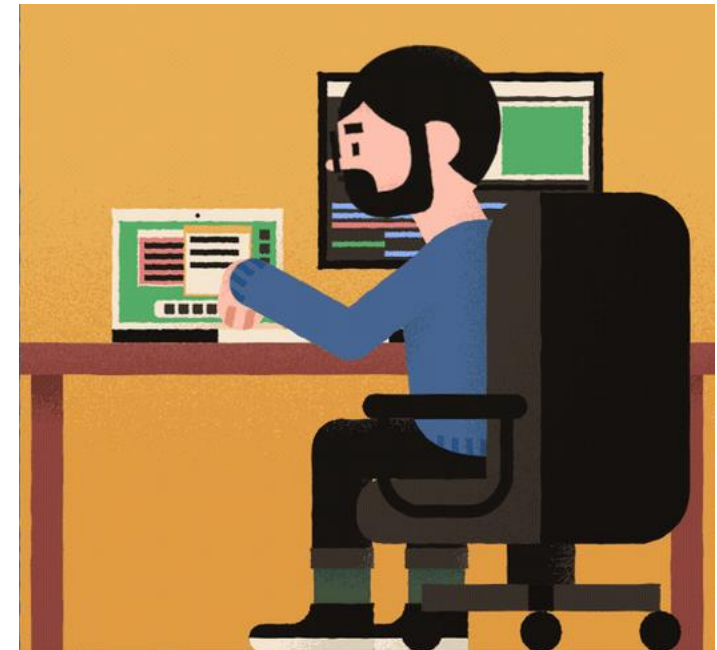


Su turno...

- Genere un nuevo script
- Agregue encabezado
- Active en su sesión la librería tidyverse

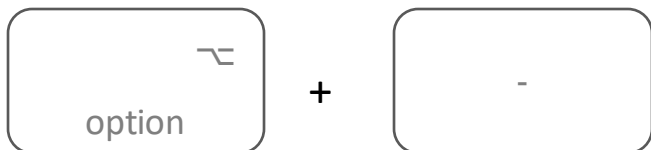
```
S04_Tex.R ×  
▶  
code > R S04_Tex.R > ...  
1 # ===== #  
2 # Script de la Sesión 04  
3 # Autor: Tex  
4 # Curso: "Procesamiento y visualización de datos espaciales en R"  
5 # Fecha: 24 de septiembre de 2025  
6 # ===== #  
7  
8 # Activo paquetes a utilizar en mi sesión  
9 library(tidyverse)  
10
```

Procesamiento de datos

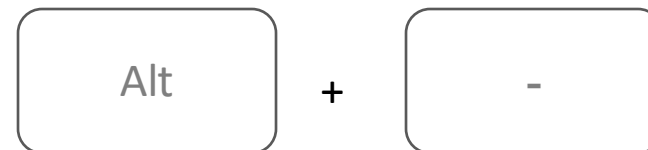


Insertar operador de asignación

<-



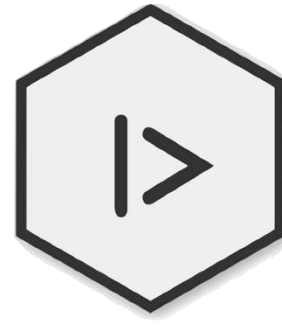
Mac



Windows



Pipe

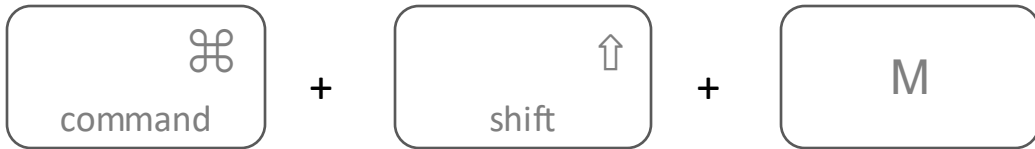


El operador `%>%` simplifica y concatena múltiples funciones (verbos)

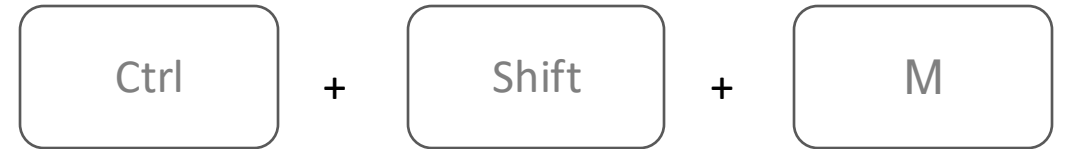
```
mall_datos %>%  
  filtro %>%  
  genero_variables %>%  
  agrupo %>%  
  paso_a_wide %>%  
  genero_variables %>%  
  selecciono_columnas
```

```
mall_datos |>  
  filtro |>  
  genero_variables |>  
  agrupo |>  
  paso_a_wide |>  
  genero_variables |>  
  selecciono_columnas
```

Insertar pipe

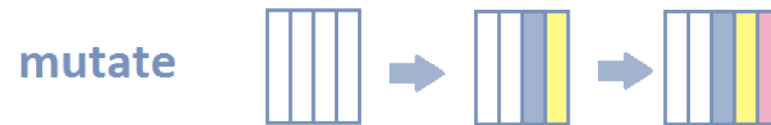
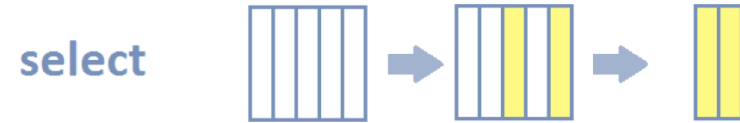


Mac



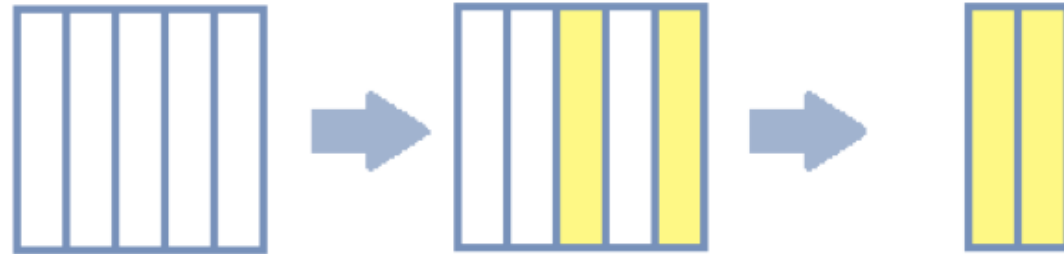
Windows

{dplyr} manipulación datos



select()

select





```
aire_tbl <- as_tibble(airquality)
```

```
> aire_tbl
# A tibble: 153 × 6
   Ozone Solar.R Wind Temp Month Day
  <int>   <int> <dbl> <int> <int> <int>
1    41    190  7.4    67     5    1
2    36    118   8     72     5    2
3    12    149 12.6    74     5    3
4    18    313 11.5    62     5    4
5    NA     NA 14.3    56     5    5
6    28     NA 14.9    66     5    6
7    23    299  8.6    65     5    7
8    19     99 13.8    59     5    8
9     8     19 20.1    61     5    9
10   NA    194  8.6    69     5   10
# i 143 more rows
# i Use `print(n = ...)` to see more rows
```

Seleccionar columnas – select()

```
> aire_tbl
# A tibble: 153 × 6
  Ozone Solar.R Wind Temp Month Day
  <int>   <int> <dbl> <int> <int> <int>
1    41    190   7.4    67     5     1
2    36    118    8     72     5     2
3    12    149  12.6    74     5     3
4    18    313  11.5    62     5     4
5    NA     NA  14.3    56     5     5
6    28     NA  14.9    66     5     6
7    23    299   8.6    65     5     7
8    19     99  13.8    59     5     8
9     8     19  20.1    61     5     9
10   NA    194   8.6    69     5    10
# i 143 more rows
# i Use `print(n = ...)` to see more rows
```

mall %>% select(...)

dataset

Argumentos de selección

Seleccionar columnas específicas

```
aire_tbl %>% select(Day, Month, Ozone)
```

```
# A tibble: 153 × 3
  Day Month Ozone
  <int> <int> <int>
1     1     5    41
2     2     5    36
3     3     5    12
4     4     5    18
5     5     5    NA
6     6     5    28
7     7     5    23
8     8     5    19
9     9     5     8
10    10     5    NA
# i 143 more rows
# i Use `print(n = ...)` to see more rows
```

Seleccionar columnas – select()

```
> aire_tbl
# A tibble: 153 × 6
  Ozone Solar.R Wind Temp Month Day
  <int>   <int> <dbl> <int> <int> <int>
1     41     190   7.4     67     5     1
2     36     118    8      72     5     2
3     12     149  12.6     74     5     3
4     18     313  11.5     62     5     4
5    NA      NA  14.3     56     5     5
6     28      NA  14.9     66     5     6
7     23     299   8.6     65     5     7
8     19      99  13.8     59     5     8
9      8      19  20.1     61     5     9
10    NA     194   8.6     69     5    10
# i 143 more rows
# i Use `print(n = ...)` to see more rows
```

mall %>% select(...)

dataset

Argumentos de selección

Seleccionar un rango de columnas

```
aire_tbl %>% select(Temp:Day)
```

```
# A tibble: 153 × 3
  Temp Month Day
  <int> <int> <int>
1     67     5     1
2     72     5     2
3     74     5     3
4     62     5     4
5     56     5     5
6     66     5     6
7     65     5     7
8     59     5     8
9     61     5     9
10    69     5    10
# i 143 more rows
# i Use `print(n = ...)` to see more rows
```

Seleccionar columnas – select()

```
> aire_tbl
# A tibble: 153 × 6
  Ozone Solar.R Wind Temp Month Day
  <int>   <int> <dbl> <int> <int> <int>
1     41     190   7.4     67     5     1
2     36     118    8      72     5     2
3     12     149  12.6     74     5     3
4     18     313  11.5     62     5     4
5    NA      NA  14.3     56     5     5
6     28      NA  14.9     66     5     6
7     23     299   8.6     65     5     7
8     19      99  13.8     59     5     8
9      8      19  20.1     61     5     9
10    NA     194   8.6     69     5    10
# i 143 more rows
# i Use `print(n = ...)` to see more rows
```

mall %>% select(...)

dataset

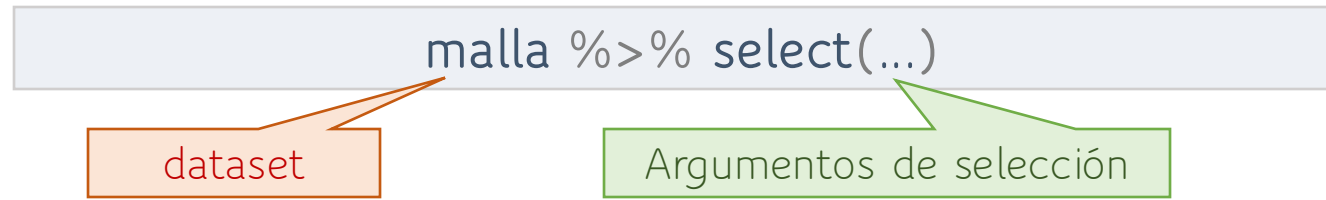
Argumentos de selección

Seleccionar columnas a excluir

```
aire_tbl %>% select(-c(Temp, Wind))
```

```
# A tibble: 153 × 4
  Ozone Solar.R Month Day
  <int>   <int> <int> <int>
1     41     190     5     1
2     36     118     5     2
3     12     149     5     3
4     18     313     5     4
5    NA      NA     5     5
6     28      NA     5     6
7     23     299     5     7
8     19      99     5     8
9      8      19     5     9
10    NA     194     5    10
# i 143 more rows
# i Use `print(n = ...)` to see more rows
```

Seleccionar columnas – `select()`



`starts_with`: Comienzan con este texto

`ends_with`: Terminan con este texto

`contains`: Contienen este texto

`matches`: Coinciden con esta expresión regular

`num_range`: Están en estas posiciones

`one_of`: alguna de estas

`everything`: Todas

filter()

filter



filter()

mall %>% filter(...)

dataset

Argumentos de filtrado

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



filter()

tabla %>% filter(...)

dataset

Argumentos de filtrado

```
> aire_tbl
```

```
# A tibble: 153 × 6
```

	Ozone	Solar.R	Wind	Temp	Month	Day
	<int>	<int>	<dbl>	<int>	<int>	<int>
1	41	190	7.4	67	5	1
2	36	118	8	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6
7	23	299	8.6	65	5	7
8	19	99	13.8	59	5	8
9	8	19	20.1	61	5	9
10	NA	194	8.6	69	5	10

```
# i 143 more rows
```

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

Filtro temperatura por arriba de 77

```
aire_tbl %>% filter(Temp > 77)
```

```
# A tibble: 85 × 6
```

	Ozone	Solar.R	Wind	Temp	Month	Day
	<int>	<int>	<dbl>	<int>	<int>	<int>
1	45	252	14.9	81	5	29
2	115	223	5.7	79	5	30
3	NA	286	8.6	78	6	1
4	NA	186	9.2	84	6	4
5	NA	220	8.6	85	6	5
6	NA	264	14.3	79	6	6
7	29	127	9.7	82	6	7
8	NA	273	6.9	87	6	8
9	71	291	13.8	90	6	9
10	39	323	11.5	87	6	10

```
# i 75 more rows
```

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

filter()

tabla %>% filter(...)

dataset

Argumentos de filtrado

Filtro temperatura por arriba de 77 y mes 5

```
aire_tbl %>% filter(Temp > 77 & Month == 5)
```

```
> aire_tbl
```

```
# A tibble: 153 × 6
```

	Ozone	Solar.R	Wind	Temp	Month	Day
	<int>	<int>	<dbl>	<int>	<int>	<int>
1	41	190	7.4	67	5	1
2	36	118	8	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6
7	23	299	8.6	65	5	7
8	19	99	13.8	59	5	8
9	8	19	20.1	61	5	9
10	NA	194	8.6	69	5	10

```
# i 143 more rows
```

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

```
# A tibble: 2 × 6
```

	Ozone	Solar.R	Wind	Temp	Month	Day
	<int>	<int>	<dbl>	<int>	<int>	<int>
1	45	252	14.9	81	5	29
2	115	223	5.7	79	5	30

filter()

tabla %>% filter(...)

dataset

Argumentos de filtrado

```
> aire_tbl
```

```
# A tibble: 153 × 6
```

	Ozone	Solar.R	Wind	Temp	Month	Day
	<int>	<int>	<dbl>	<int>	<int>	<int>
1	41	190	7.4	67	5	1
2	36	118	8	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6
7	23	299	8.6	65	5	7
8	19	99	13.8	59	5	8
9	8	19	20.1	61	5	9
10	NA	194	8.6	69	5	10

```
# i 143 more rows
```

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

Filtro valores de **Ozono** distintos de **NA**

```
aire_tbl %>% filter(!is.na(Ozone))
```

```
# A tibble: 116 × 6
```

	Ozone	Solar.R	Wind	Temp	Month	Day
	<int>	<int>	<dbl>	<int>	<int>	<int>
1	41	190	7.4	67	5	1
2	36	118	8	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	28	NA	14.9	66	5	6
6	23	299	8.6	65	5	7
7	19	99	13.8	59	5	8
8	8	19	20.1	61	5	9
9	7	NA	6.9	74	5	11
10	16	256	9.7	69	5	12

```
# i 106 more rows
```

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

filter()

mallá %>% filter(...)

dataset

Argumentos de filtrado

```
> aire_tbl
```

```
# A tibble: 153 × 6
```

	Ozone	Solar.R	Wind	Temp	Month	Day
	<int>	<int>	<dbl>	<int>	<int>	<int>
1	41	190	7.4	67	5	1
2	36	118	8	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6
7	23	299	8.6	65	5	7
8	19	99	13.8	59	5	8
9	8	19	20.1	61	5	9
10	NA	194	8.6	69	5	10

```
# i 143 more rows
```

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

Filtro valores por arriba del promedio de Temp y que sean de los días 2, 7 o 11

```
aire_tbl %>% filter(Temp > mean(Temp) & Day %in% c(2, 7, 11))
```

```
# A tibble: 11 × 6
```

	Ozone	Solar.R	Wind	Temp	Month	Day
	<int>	<int>	<dbl>	<int>	<int>	<int>
1	29	127	9.7	82	6	7
2	NA	259	10.9	93	6	11
3	49	248	9.2	85	7	2
4	77	276	5.1	88	7	7
5	NA	139	8.6	82	7	11
6	9	24	13.8	81	8	2
7	122	255	4	89	8	7
8	NA	137	11.5	86	8	11
9	78	197	5.1	92	9	2
10	20	252	10.9	80	9	7
11	44	236	14.9	81	9	11

mutate()

mutate



Generar variables – mutate()

mall %>% mutate(...)

dataset

Argumentos de creación de variables

```
> aire_tbl
```

```
# A tibble: 153 × 6
```

	Ozone	Solar.R	Wind	Temp	Month	Day
	<int>	<int>	<dbl>	<int>	<int>	<int>
1	41	190	7.4	67	5	1
2	36	118	8	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6
7	23	299	8.6	65	5	7
8	19	99	13.8	59	5	8
9	8	19	20.1	61	5	9
10	NA	194	8.6	69	5	10

```
# i 143 more rows
```

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

Genero columna con el logaritmo de Ozono

```
aire_tbl %>% mutate(o3_log = log(Ozone))
```

```
# A tibble: 153 × 7
```

	Ozone	Solar.R	Wind	Temp	Month	Day	o3_log
	<int>	<int>	<dbl>	<int>	<int>	<int>	<dbl>
1	41	190	7.4	67	5	1	3.71
2	36	118	8	72	5	2	3.58
3	12	149	12.6	74	5	3	2.48
4	18	313	11.5	62	5	4	2.89
5	NA	NA	14.3	56	5	5	NA
6	28	NA	14.9	66	5	6	3.33
7	23	299	8.6	65	5	7	3.14
8	19	99	13.8	59	5	8	2.94
9	8	19	20.1	61	5	9	2.08
10	NA	194	8.6	69	5	10	NA

```
# i 143 more rows
```

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

Generar variables – mutate()

mall %>% mutate(...)

dataset

Argumentos de creación de variables

```
> aire_tbl
```

```
# A tibble: 153 × 6
```

	Ozone	Solar.R	Wind	Temp	Month	Day
	<int>	<int>	<dbl>	<int>	<int>	<int>
1	41	190	7.4	67	5	1
2	36	118	8	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6
7	23	299	8.6	65	5	7
8	19	99	13.8	59	5	8
9	8	19	20.1	61	5	9
10	NA	194	8.6	69	5	10

```
# i 143 more rows
```

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

Genero copia de columna Month

```
aire_tbl %>% mutate(mes = Month)
```

```
# A tibble: 153 × 7
```

	Ozone	Solar.R	Wind	Temp	Month	Day	mes
	<int>	<int>	<dbl>	<int>	<int>	<int>	<int>
1	41	190	7.4	67	5	1	5
2	36	118	8	72	5	2	5
3	12	149	12.6	74	5	3	5
4	18	313	11.5	62	5	4	5
5	NA	NA	14.3	56	5	5	5
6	28	NA	14.9	66	5	6	5
7	23	299	8.6	65	5	7	5
8	19	99	13.8	59	5	8	5
9	8	19	20.1	61	5	9	5
10	NA	194	8.6	69	5	10	5

```
# i 143 more rows
```

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

Generar variables – mutate()

mall %>% mutate(...)

dataset

Argumentos de creación de variables

```
> aire_tbl
```

```
# A tibble: 153 × 6
```

	Ozone	Solar.R	Wind	Temp	Month	Day
	<int>	<int>	<dbl>	<int>	<int>	<int>
1	41	190	7.4	67	5	1
2	36	118	8	72	5	2
3	12	149	12.6	74	5	3
4	18	313	11.5	62	5	4
5	NA	NA	14.3	56	5	5
6	28	NA	14.9	66	5	6
7	23	299	8.6	65	5	7
8	19	99	13.8	59	5	8
9	8	19	20.1	61	5	9
10	NA	194	8.6	69	5	10

```
# i 143 more rows
```

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

Genero nueva columna que concatena Month y Day

```
aire_tbl %>% mutate(dia_mes = paste(Day, Month, sep = "-"))
```

```
# A tibble: 153 × 7
```

	Ozone	Solar.R	Wind	Temp	Month	Day	mes
	<int>	<int>	<dbl>	<int>	<int>	<int>	<int>
1	41	190	7.4	67	5	1	5
2	36	118	8	72	5	2	5
3	12	149	12.6	74	5	3	5
4	18	313	11.5	62	5	4	5
5	NA	NA	14.3	56	5	5	5
6	28	NA	14.9	66	5	6	5
7	23	299	8.6	65	5	7	5
8	19	99	13.8	59	5	8	5
9	8	19	20.1	61	5	9	5
10	NA	194	8.6	69	5	10	5

```
# i 143 more rows
```

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




Su turno...

- Queremos una malla de datos que contenga sólo a las columnas **Day**, **Month**, **Ozone** y **Solar.R**
- Queremos aquellos días en los que la concentración de ozono excedió o fue igual al promedio del periodo
- Queremos que los datos sean sólo del mes 7
- Requerimos transformar a logarítmica la variable de ozono

