

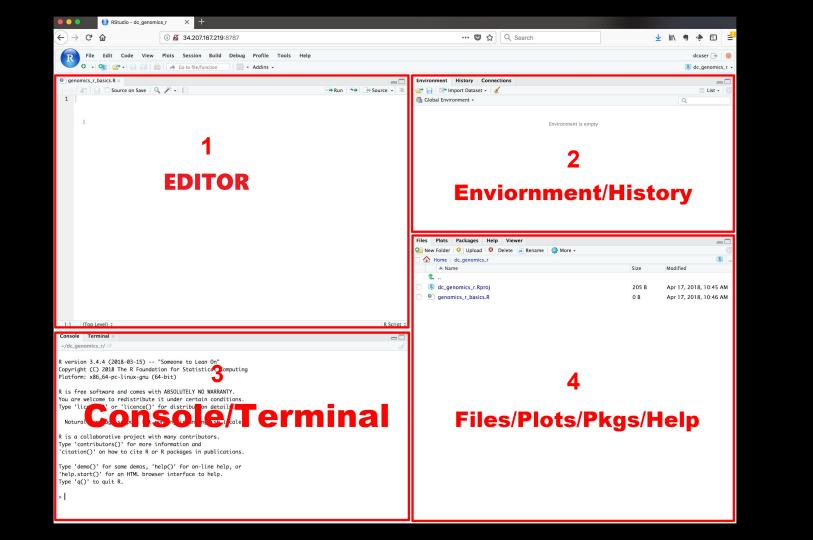
Introducción a R

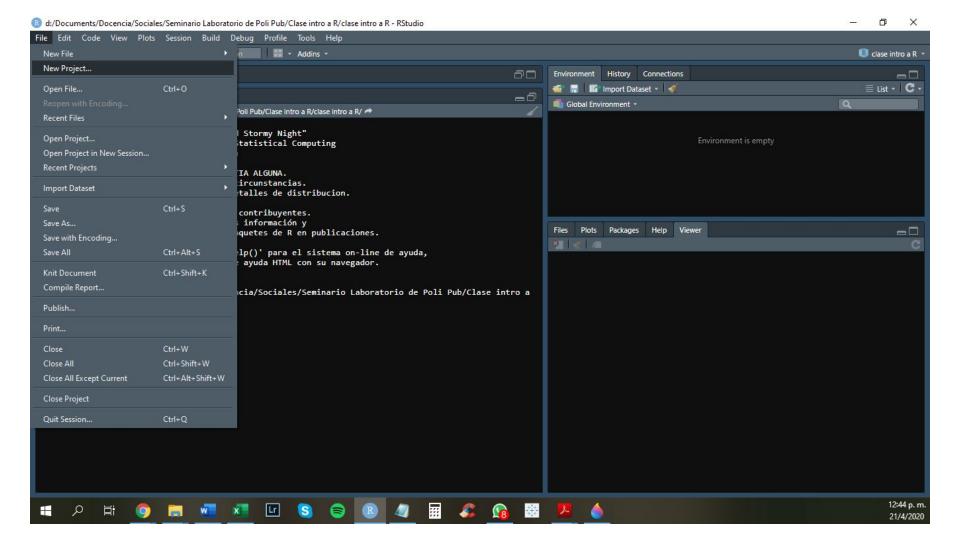
¿Qué vimos la clase pasada?

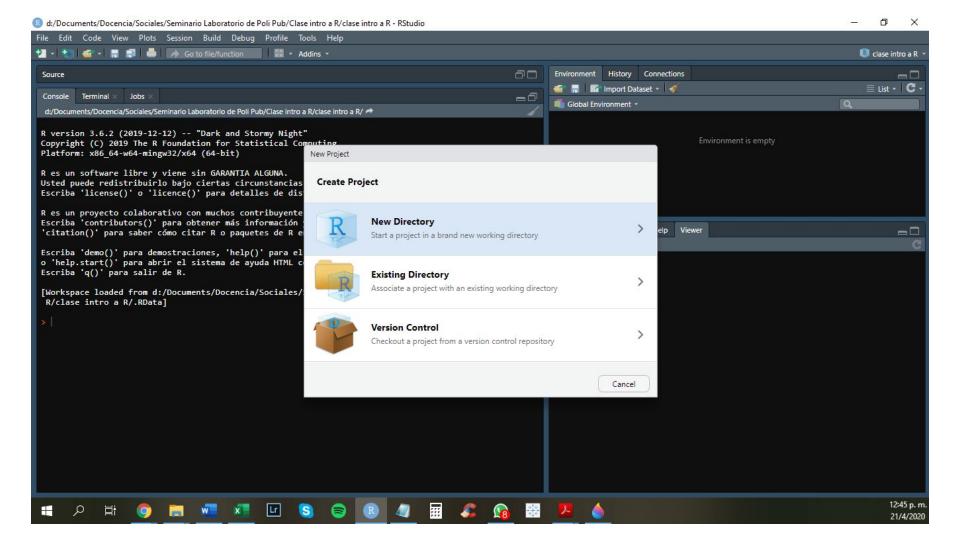
- ¿Qué es R y Rstudio?
- Descargar e instalar R y Rstudio
- Asignación de valores a variables
- "R como calculadora"

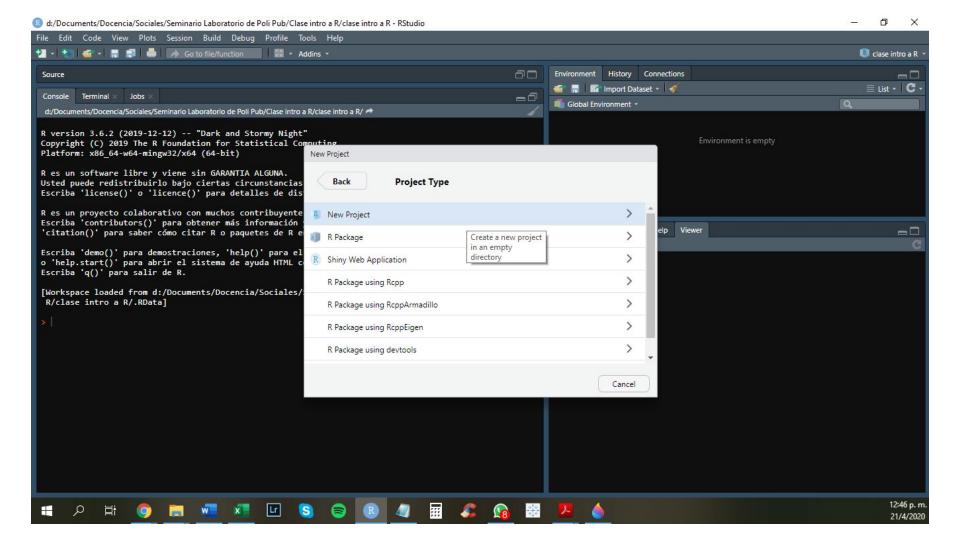
¿Qué vamos a ver hoy?

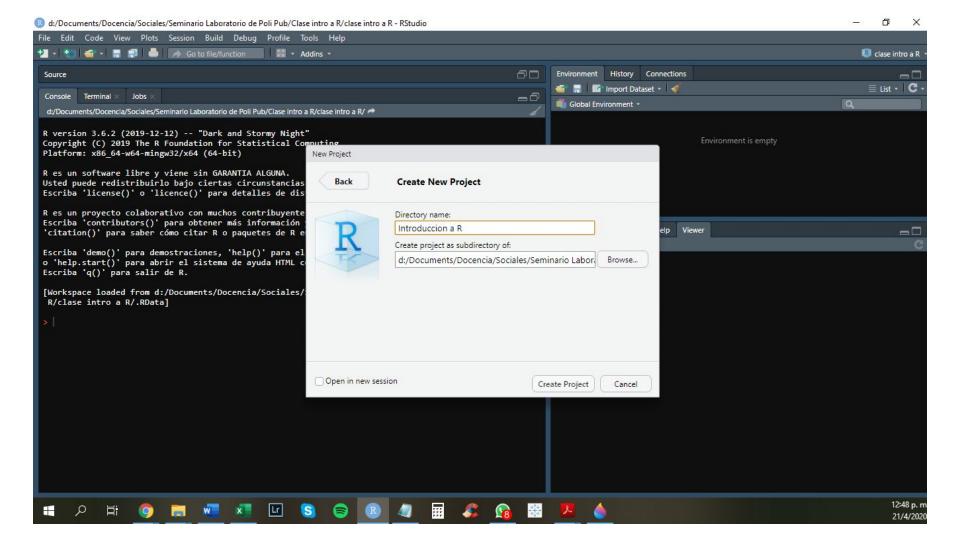
- Paneles de R
- Asignación de valores a variables
- ¿Qué es un objeto? Tipos de objetos
- Operadores lógicos
- Dificultades en la Sintaxis
- Generar un proyecto
- ¿Qué es un algoritmo?
- Pasos para trabajar en R
- Paquetes y librerías:
- Tidyverse:
 - select()
 - o filter()
 - mutate()
 - group_by()
 - summarise()
 - o pipe (%>%)





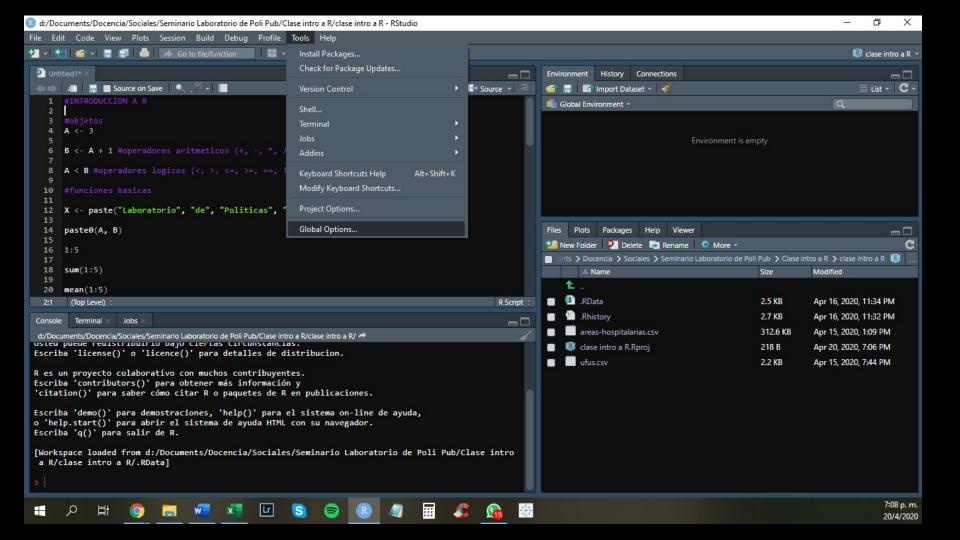


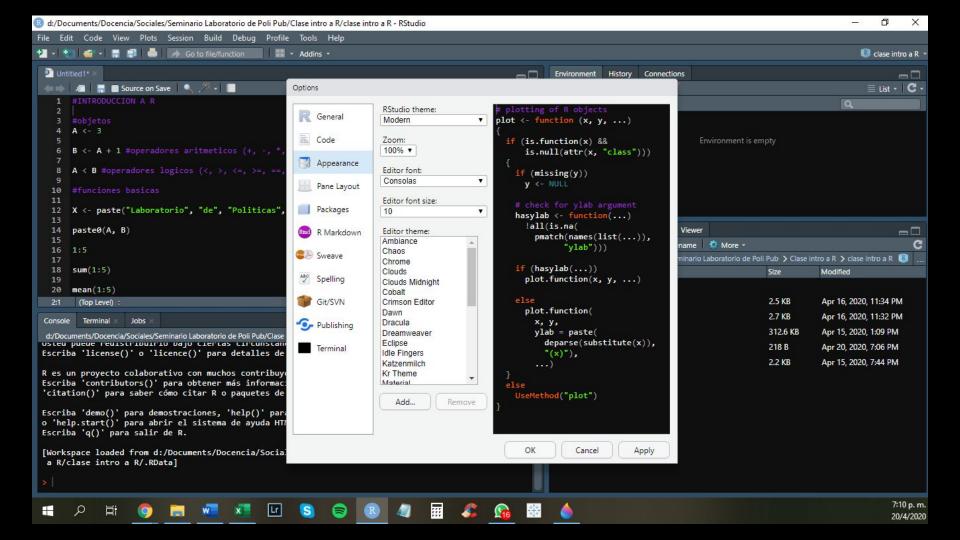






https://github.com/labpoliticasuba/



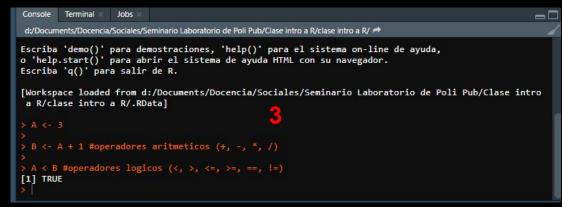


$$A < -3$$

A < B #operadores logicos (<, >, <=, >=, !=)

<u>#objetos</u>





Operadores en R

Lógico	Descripción	Condicional	Descripción
&	AND	==	IGUAL
1	OR	!=	DISTINTO
!	NOT	>	MAYOR
		·<	MENOR
		>=	MAYOR E IGUAL
		<=	MENOR E IGUAL

#funciones básicas

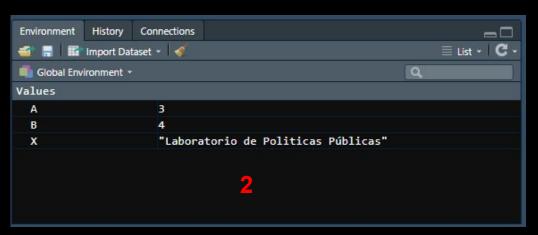
```
X <- paste("Laboratorio",
      "de", "Politicas",
      "Públicas", sep = " ")</pre>
```

```
paste0(A, B)

1:5

sum(1:5)

mean(1:5)
```



#tipos de objeto

```
class(A)
class(X)
```

```
Terminal
                     Jobs
Console
 d:/Documents/Docencia/Sociales/Seminario Laboratorio de Poli Pub/Clase intro a R/clase intro a R/
> 1:5
[1] 1 2 3 4 5
> sum(1:5)
[1] 15
> mean(1:5)
> #tipos de objeto
> class(A)
[1] "numeric"
> class(X)
    "character"
```

Numérico	1.3	numeric	
Cadena de texto	"uno"	character	
Factor	uno	factor	

TRUE

NA

NULL

integer

logical

NA

null

Ejemplo

Nombre en inglés

Tipo

Entero

Lógico

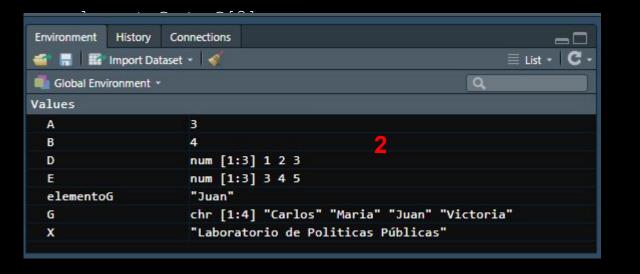
Perdido

Vacio

#vectores

```
D < -c(1,2,3)
```

$$E < - D + 2$$



#dataframe

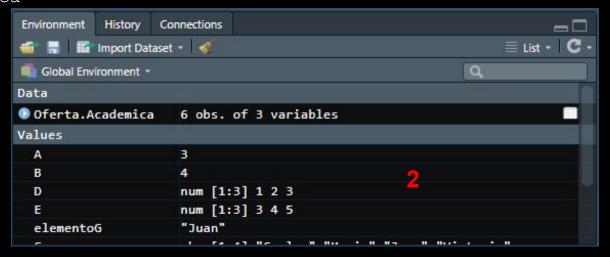
```
HORARIO <- c("7 a 9", "9 a 11", "11 a 13", "9 a 11", "11 a 13", "17 a 19")

MATERIA <- c("Fundamentos", "Teo_Pol", "Economia", "Derecho", "Historia",

"Teo_Pol")

INSCRIPTOS <- c(49,42,26,24,36,60)
```

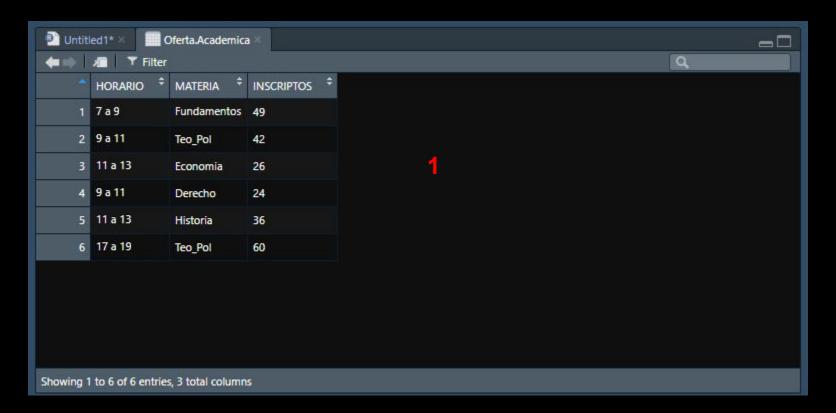
Oferta.Academica <- data.frame(HORARIO, MATERIA, INSCRIPTOS)
Oferta.Academica



#dataframe

Console	Terminal Jobs			-0
d:/Docum	ents/Docencia/Sociales/	Seminario Laboratorio de l	Poli Pub/Clase intro a R/clase intro a R/ →	
> HORARI > MATERI		ntos", "Teo_Pol",	13", "9 a 11", "11 a 13", "17 a 19") "Economia","Derecho","Historia", "Teo_Pol")	
		ata.frame(HORARIO,	MATERIA, INSCRIPTOS)	
> Oferta HORARI	Academica O MATERIA I	NSCRIPTOS		
1 7 a	9 Fundamentos	49		
2 9 a 1	1 Teo_Pol	42	2	
3 11 a 1	l3 Economia	26	J	
4 9 a 1	1 Derecho	24		
5 11 a 1	l3 Historia	36		
6 17 a 1	l9 Teo_Pol	60		

#dataframe



#SINTAXIS

(numeral) comentarios # se usa para definir el nombre de un objeto (variable o dataframe) # R reconoce la diferencia entre mayosculas, minusculas y caracteres especiales(acentos, -, ,etc) # "" (comillas) se usa para hacer referencia a valores que son caracteres # Espacios en blanco y enters son para visualizar mejor (no impacta en el # codigo) # [corchetes] y \$ son metodos para hacer referencia a objetos/columnas, # (parentesis) es para definir parametros de una función, comas, es para separar parametros dentro de una funcion

#borrar objetos

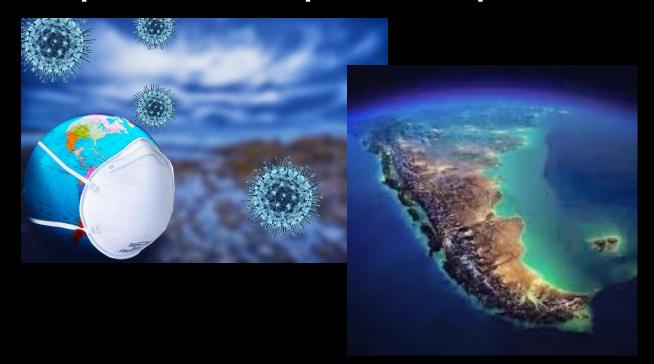
```
rm(A,
   В,
   Χ,
   D,
   Ε,
   G,
   elementoG,
   HORARIO,
   MATERIA,
   INSCRIPTOS,
   Oferta.Academica)
```

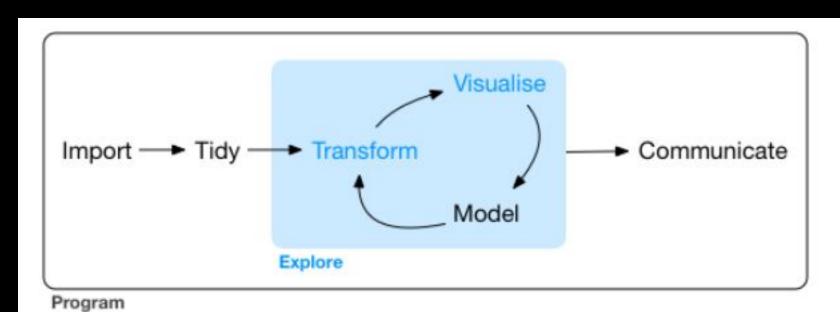
¿Qué es un algoritmo?



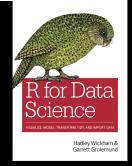
Es un conjunto de instrucciones o reglas definidas y no-ambiguas, ordenadas y finitas que permite, típicamente, solucionar un problema, realizar un cómputo, procesar datos y llevar a cabo otras tareas o actividades.

¿Cómo llegan los sistemas sanitarios de cada una de las provincias del país a esta pandemia?



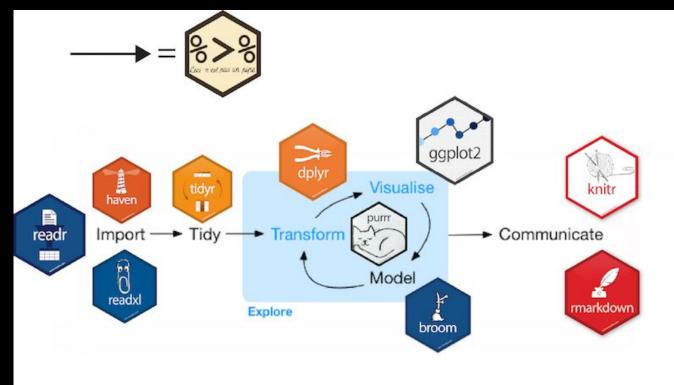


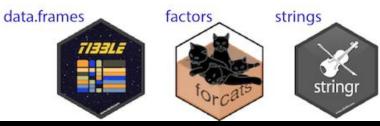
O'REILLY'





"Dad why is my sisters name Rose" "Because your mother loves Roses" "Thanks Dad" "No problem library(tidyverse)





TIDYVERSE

#INSTALAMOS PAQUETES

#es necesario instalar y activar los paquetes. es recomendable hacerlo cada vez

#que se van a usar para actualizarlos

install.packages("tidyverse")
install.packages("plotly")
#Lo activamos

library(tidyverse) library(plotly)

CSV (VALORES SEPARADOS POR COMAS)



```
usu_individual_t117: Bloc de notas
Archivo Edición Formato Ver Ayuda
CODUSU; ANO4; TRIMESTRE; NRO_HOGAR; COMPONENTE; H15; REGION; MAS_500; AGLOMERADO; PONDERA; CH03; CH04; CH05; CH06; CH07; CH06
DECINDR; IDECINDR; RDECINDR; GDECINDR; PDECINDR; ADECINDR; PONDII; V2 M; V3 M; V4 M; V5 M; V8 M; V9 M; V10 M; V11 M; V12 M; V1
"TQRMNOQURHKOMLCDEFIAH00469284";2017;1;1;4;1;43;"S";2;861;3;2;12/12/1998;18;1;4;1;2;0;2;1;;1;;1;;2;3;0;4;0;0;6
"TQRMNOQURHKOMLCDEFIAH00469284";2017;1;1;5;1;43;"S";2;861;3;2;04/01/2004;13;5;4;1;1;1;4;2;"01";1;;1;;3;3;0;3;6
"TQRMNOQURHKOMLCDEFIAH00469284";2017;1;1;6;1;43;"S";2;861;3;1;20/02/2002;15;5;4;1;1;1;4;2;"01";1;;1;;3;3;0;3;4
"TQRMNOQURHKOMLCDEFIAH00469284";2017;1;1;7;1;43;"S";2;861;3;2;03/06/2005;11;5;4;1;1;1;4;2;"00";1;;1;;3;3;0;3;6
TQRMNOQURHKOMLCDEFIAH00469284";2017;1;1;8;0;43;"S";2;861;3;1;03/10/2009;7;5;4;1;1;1;2;2;"01";1;;1;;1;;1;4;0;3;0"
"TQRMNOQURHKOMLCDEFIAH00469284";2017;1;1;9;0;43;"5";2;861;3;2;28/01/2013;4;5;4;2;1;1;1;2;"01";1;;6;;7;4;0;5;0
"TQRMNOQURHKOMLCDEFIAH00469284";2017;1;1;10;1;43;"S";2;861;4;2;;17;1;4;1;1;1;4;2;"02";1;;1;;3;3;0;3;0;0;0;0;0
"TORMNOOURHKOMLCDEFIAH00469284";2017;1;1;11;0;43;"S";2;861;5;1;06/06/2015;1;5;4;3;0;0;0;0;;1;;6;;7;4;0;5;0;0;
"TQRMNOQURHKOMLCDEFIAH00469284";2017;1;1;13;1;43;"S";2;861;4;1;22/11/1996;20;1;4;1;2;0;2;1;;1;;1;;2;1;3;0;0;0
"TQRMNOQURHKOMLCDEFIAH00469284";2017;1;1;14;0;43;"S";2;861;5;2;06/01/2016;1;5;4;3;0;0;0;0;;1;;6;;7;4;0;5;0;0;6
"TQRMNOQUTHKOMLCDEFIAH00469286";2017;1;1;1;1;43;"S";2;861;1;1;30/12/1967;49;1;1;1;2;0;4;1;;3;"30";1;;4;1;2;0;4
"TQRMNOQUTHKOMLCDEFIAH00469286";2017;1;1;2;1;43;"S";2;861;2;2;07/10/1978;38;1;4;1;2;0;2;1;;1;;1;;2;3;0;4;0;0;4
"TQRMNORWSHLMLOCDEFIAH00497956";2017;1;1;1;1;43;"S";2;514;1;2;07/07/1952;64;3;1;1;2;0;4;1;;1;;1;;4;1;3;0;0;6;6
"TQRMNOQRVHMMLNCDEFIAH00508001";2017;1;1;1;1;1;43;"S";2;503;1;1;23/07/1953;63;2;1;1;2;0;6;1;;2;;1;;6;1;1;0;0;0;6
"TQRMNOQRVHMMLNCDEFIAH00508001";2017;1;1;2;1;43;"S";2;503;2;2;23/02/1958;59;2;1;1;2;0;2;1;;2;3;0;4;0;0;4
"TQRMNOQRVHMMLNCDEFIAH00508001";2017;1;1;3;1;43;"S";2;503;3;2;11/02/1987;30;5;1;1;1;1;7;2;"06";1;;1;;5;1;3;0;6
"TQRMNORYXHKKKQCDEFIAH00508002";2017;1;1;1;1;43;"5";2;919;1;1;;76;2;12;1;2;0;4;2;"03";1;;1;;3;3;0;1;0;0;0;0;0;
"TORMNORYXHKKKQCDEFIAH00508002";2017;1;1;2;1;43;"5";2;919;2;2;;74;2;12;1;2;0;4;1;;1;;1;;4;3;0;1;0;0;0;0;0;0;0;0
"TQRMNORRUHJLKUCDEFIAH00498010";2017;1;1;1;1;43;"S";2;888;1;1;23/07/1986;30;5;1;1;1;1;7;2;"03";2;;2;;5;1;3;0;6
"TQRMNORRUHJLKUCDEFIAH00498010";2017;1;1;2;1;43;"S";2;888;8;2;02/02/1992;24;5;1;1;1;1;7;2;"03";2;;5;1;3;0;(,
                                                                  Ln 1, Col 1
                                                                                        Windows (CRLF)
```

```
#Tenemos dos opciones:
#OPCION 1 (desde internet con el siguiente link):

data <- read.csv("https://github.com/labpoliticasuba/Clases_2020/raw/master/Clase_02/data_final_clase.csv")</pre>
```

```
### descargamos el archivo,
### lo ubicamos en la carpeta de nuestro proyecto
### hacemos click derecho en la carpeta > Propiedades > copiamos y pegamos el path
### path: dirección donde se encuentra almacenado nuestro csv
```

data <- read.csv("D:/Guada/Clases/Lab_Pol_Publ/Tidyverse/data_final_clase.csv")</pre>

#OPCION 2 (desde nuestra pc):

```
# A tibble: 6 x 8
# Groups: REGION, DISTRITO, TIPO_CAMA, CASOS, MUERTOS, POBLACION [6]
 REGION DISTRITO TIPO_CAMA
                                          CASOS MUERTOS POBLACION CANTIDAD_CAMAS MEDICOS_CANTIDAD
 <fct> <fct> <fct> <chr>
                                          <int> <int>
                                                        <int>
                                                                          <db1>
                                                                                          <int>
1 CUYO
        La Rioja Aislamiento y casos leves
                                             31
                                                                                           1444
                                                          383220
                                                                           1411
2 CUYO La Rioja Casos graves
                                             31
                                                          383220
                                                                            350
                                                                                           1444
3 CUYO La Rioja Casos graves Neo
                                                                                           1444
                                             31
                                                          383220
                                                                            113
4 CUYO
        Mendoza Aislamiento y casos leves
                                             63
                                                         1949293
                                                                           3466
                                                                                           6675
5 CUYO
        Mendoza Casos graves
                                             63
                                                         1949293
                                                                                           6675
                                                                        3440
```

1949293

292

6675

63

> ### la función head() nos permite ver las primeras 10 observaciones

Mendoza Casos graves Neo

> head(data)

6 CUYO

##	la	función	dim()	nos	permite	ver	la	cantidad	de	filas	y	columnas	de	nuestro	dataset	
im(dat	a)														
72	8	3														

	### la funciór summary(data)	n summary() nos pe	ermite ver medidas	resumenes de nuest	ras variables	
	REGION	DISTRITO	TIPO_CAMA	CASOS	MUERTOS	
٨	DCENTTNA · O	Puopos Airos: 2	Longth: 72	Min . 0.00	Min . 0 000	

NOA

Class :character :12 1st Qu.: 4.75 1st Qu.: 0.000 CUYO CABA

NEA

:12 Mode :character Median : 30.50 Median : 0.000 Catamarca

Median:

:15 Chaco : 3 Mean :107.79 Mean : 4.625

Mean

PAMPEANA:18 Chubut : 3 3rd Qu.:106.50 3rd Qu.: 4.000

3rd Qu.: 1454996

PATAGONIA:15 Córdoba : 3 Max. :733.00 Max. :44.000 Max.

(Other) :54

MEDICOS_CANTIDAD

Min. : 703

1st Qu.: 1414

Median: 2345

Mean : 7254

POBLACION

1st Qu.:

164944

570254

861358

: 1853938

:17196396

CANTIDAD_CAMAS

22

188

882

: 3130 3rd Qu.: 2232

:57151

Min.

Mean

Max.

1st Qu.:

Median:

3rd Qu.: 4359

Max. :48447

FUNCIONES PARA TRABAJAR CON NUESTRAS COLUMNAS

#si queremos saber los valores unicos de una columna

unique(data\$DISTRITO)

#Si queremos ver todos los valores de nuestra columna DISTRITO: data\$DISTRITO

#si queremos ver los nombres de nuestras columnas podemos usar la funcion colnames(data)

colnames (data)

```
### funciones en Tidyverse:
```

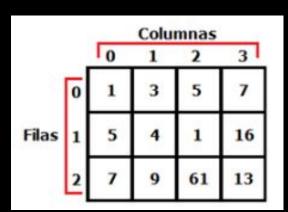
```
select()  # para seleccionar columnas

filter()  # para filtrar filas

mutate()  # para agregar o modificar valores de columnas no agrupadas

group_by()  # para agrupar filas

summarise()  # para agregar información de columnas agrupadas (calcula 1 sólo valor por grupo)
```



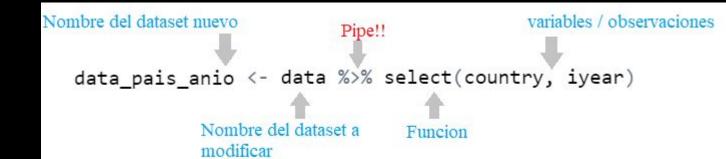
https://cran.r-project.org/web/packages/tidyverse/tidyverse.pdf

```
##### ¿Cuál es la provincia más capacitada (de la región pampeana ó que tengan más de 1.5 millones de habitantes)
#### para enfrentar la emergencia sanitaria en relación a la cantidad de
##### de médicos c/10.000 habitantes?
data_seleccionados <- select(data, REGION, DISTRITO, POBLACION, CANTIDAD_MEDICOS)</pre>
```

data_filtrado <- filter(data_seleccionados, REGION == 'PAMPEANA' | POBLACION > 1500000)

data_agrupado <- group_by(data_filtrado, DISTRITO) %>% distinct()

data_summarise <- data_agrupado %>% summarise(CANT_MEDICOS_CADA_CIENMIL = (CANTIDAD_MEDICOS/POBLACION)*10000)





```
How mornings look like for most people:
    me %>%
        wake_up() %>%
        get_out_of_bed() %>%
        get_dressed() %>%
        leave_house()
```

How my mornings look like most of the time:

leave_house(get_dressed(get_out_of_bed(wake_up(me))))

```
########## PIPE ( %>% )
```

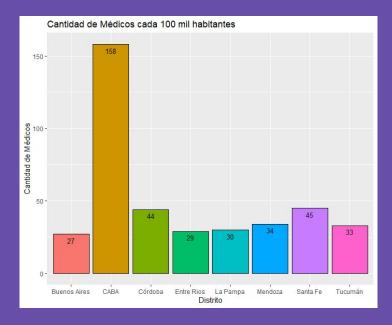
```
Ejercicio_CLASE_pipe <- data %>%
  select(REGION, DISTRITO, POBLACION, CANTIDAD_MEDICOS) %>%
  filter(REGION == 'PAMPEANA' | POBLACION > 1500000) %>%
  group_by(DISTRITO) %>%
  distinct() %>%
  summarise( CANT_MEDICOS_CADA_CIENMIL = round((CANTIDAD_MEDICOS/POBLACION)*10000),3) %>%
  arrange(desc(CANT_MEDICOS_CADA_CIENMIL))
```

```
##### GRAFIQUEMOS

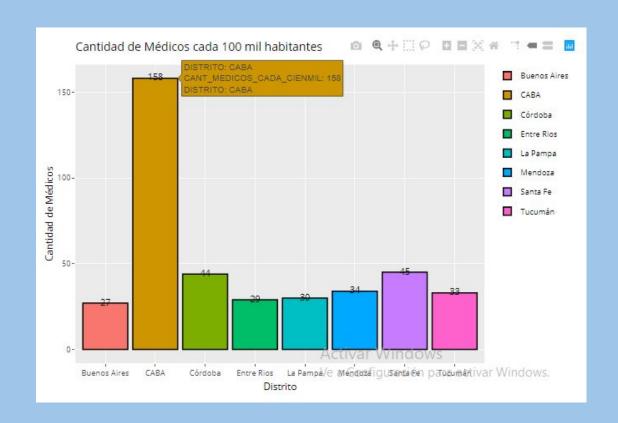
p <- ggplot(data=Ejercicio_CLASE_pipe, aes(x=DISTRITO, y=CANT_MEDICOS_CADA_CIENMIL, fill=DISTRITO)) +
    geom_bar(colour="black", stat="identity") +
    guides(fill=FALSE) +
    xlab("Distrito") +
    ylab("Cantidad de Médicos") + # Set axis labels
    ggtitle("Cantidad de Médicos cada 100 mil habitantes") +
    geom_text(aes(label=CANT_MEDICOS_CADA_CIENMIL), vjust=1.6, color="black", size=3.5)

p</pre>
```





ggplotly(p) ## si queremos hacer gráficos interactivos





- 1. ¿Cuáles son las 5 provincias con mayor cantidad de médicos por cantidad de habitantes? Vamos paso a paso:
 - a. Seleccione las columnas que vamos a utilizar para este punto en un nuevo data frame llamado "ej_1_a"
 - b. Usando el dataframe realizado en el 1.a) elabore una nueva columna llamada "medicos_x_hab" (que contenga la información CANTIDAD_MEDICOS/POBLACION). Todo esto en un nuevo dataframe que llamaremos "ej 1 b"
 - c. Elabore un group_by y seleccione las 5 provincias con mayor cantidad de casos por habitantes en un nuevo dataframe "ej 1 c"
 - d. ¿Y la que tiene menor cantidad de casos de la región de "CUYO"? (elabore un nuevo dataframe llamado ej 1 d)
- 2. a. ¿Cuál es el porcentaje de camas para casos graves (adultos) que tiene cada provincia respecto del total de camas a nivel nacional? Elabore un nuevo dataframe llamado ej_2_a
- 3. a. ¿Cuál es la región con mayor cantidad de camas para casos de aislamiento y leves- por provincia- en promedio?
 - b. De las provincias con menos de 1 millón de habitantes, ¿cuál es la que mayor cantidad de camas para aislamiento por cada 5000 habitantes ? Genere un dataframe llamado "ej_3_b"

DESAFÍO OPCIONAL:

4. a. Suponiendo que un paciente adulto infectado grave necesita en promedio 10 días para recuperarse en una cama de terapia intensiva. En el periodo de un mes (30 días)... ¿Cuántas personas graves tiene la posibilidad de atender la provincia de la consigna 3.b. en el periodo determinado sin que el sistema de salud colapse?

b. ¿Cómo sería la respuesta a la pregunta 4.a. para "Ciudad de Buenos Aires" y "Provincia de Buenos Aires"?

¿Qué vimos hoy?

- Paneles de R
- Asignación de valores a variables
- ¿Qué es un objeto? Tipos de objetos
- Operadores lógicos
- Dificultades en la Sintaxis
- Generar un proyecto
- ¿Qué es un script? ¿Qué es un algoritmo?
- Pasos para trabajar en R
- Paquetes y librerías:
- Tidyverse:
 - select()
 - filter()
 - mutate()
 - group_by()
 - summarise()
 - o pipe (%>%)

¿Qué vamos a ver la que viene?

Repaso clase anterior + dudas y preguntas

JOINS:

- primary key()
- left join()
- right_join()
- outer_join()
- full_join()

case_when()

gather() y spread()