

Procesamiento y visualización de información

# BUCEANDO R

UNTREF



# AGENDA DEL DÍA

- Presentaciones
- Motivaciones
- ¿Por qué R?
- Interfaz de R
- Objetos en R
- Operadores

# ¿Por qué R?

- Es **software libre**, es gratuito.
- Lenguaje de programación **especializado**.
- Permite **automatizar** tareas.
- Tiene una **comunidad** enorme de desarrollo.
- Es **sencillo** para comenzar.
- Permite la replicabilidad de nuestro trabajo.



# ¿Por qué R?

- R es una herramienta más, no la única.



Extracción y  
procesamiento



Modelado



Visualización

# R y Rstudio



R y Rstudio (ahora Posit) son dos softwares distintos. R es el lenguaje de programación propiamente dicho. RStudio es IDE, un programa que nos facilita trabajar con R.

# CONOCIENDO RSTUDIO

File Edit Code View Plots Session Build Debug Profile Tools Help

Go to file/function Addins

PROYECTO.R Hacinaamiento.R Scripts

```
7 library("ggrepel")
8 options(scipen = 999)
9 al<-read_sf("América/América.shp")
10 provincias<-read_sf("ign_provincia/Provincia/ign_provincia.shp")
11
12 individuo <- get_microdata(year = 2019,
13                             trimester = 4,
14                             type = "individual") %>%
15   organize_labels(type = "individual")
16
17
18 hogar <- get_microdata(year = 2019,
19                         trimester = 4,
20                         type = "hogar") %>%
21   organize_labels(type = "hogar")
22
23
24 individuo1<-individuo %>% select(CODUSU,CH06,CH04,ESTADO,CH08,PONDERA,PP07H)
25
26 hogar1 <- hogar %>%
27   select(CODUSU, AGLOMERADO, REGION,IX_TOI,III,PONDERA)
```

Código

Console terminal Jobs Consola

```
C:/Users/flavio/Desktop/Mapa escuelas/
+               trimester = 4,
+               type = "hogar") %>%
+   organize_labels(type = "hogar")
probando la URL 'https://www.indec.gob.ar/ftp/cuadros/menusuperior/eph/EPH_usu_4_Trim_2019_txt.zip'
Content type 'application/x-zip-compressed' length 3208230 bytes (3.1 MB)
downloaded 3.1 MB

> individuo <- get_microdata(year = 2019,
+                             trimester = 4,
+                             type = "individual") %>%
+   organize_labels(type = "individual")
probando la URL 'https://www.indec.gob.ar/ftp/cuadros/menusuperior/eph/EPH_usu_4_Trim_2019_txt.zip'
Content type 'application/x-zip-compressed' length 3208230 bytes (3.1 MB)
downloaded 3.1 MB

> |
```

Estado del  
procesamiento

Environment History Connections Tutorial Ambiente

Global Environment

Data	
al	53 obs. of 2 variables
hogar	19035 obs. of 88 variables
individuo	58519 obs. of 177 variables
provincias	24 obs. of 12 variables

Archivos,  
paquetes,  
visualización

Files Plots Packages Help Viewer

C:\Users\flavio\Desktop\Mapa escuelas

Name	Size	Modified
..		
.RData	136 MB	Jul 17, 2020, 6:37 PM
.Rhistory	21 KB	Jul 17, 2020, 10:21 PM
.Rproj.user		
ambito_corregido.R	4.6 KB	Jul 15, 2020, 6:03 PM
América		
Codgeo_Pais_x_loc_con_datos (1)		
conectividad(1)		
conectividad_ambito corregido.csv	18.4 MB	Jul 15, 2020, 1:05 PM
departamento		
departamentos_urp.xlsx	36.3 KB	Jul 16, 2020, 10:40 AM
ign_provincia		
Indice.R	2.6 KB	Jul 17, 2020, 4:10 PM
indice_v2.xlsx	370.3 KB	Jul 17, 2020, 4:07 PM
Localidades_alis/adas.xlsx	249.5 KB	Jul 17, 2020, 11:42 AM
Mapa_escuelas.Rproj	218 B	Jul 17, 2020, 10:21 PM

## Spaces

Your Workspace

+ New Space

## Learn

Guide

What's New

Primers

Cheat Sheets

## Help

Current System Status

Posit Community

## Info

Plans & Pricing

Terms and Conditions

File Edit Code View Plots Session Build Debug Profile Tools Help

Console Terminal Background Jobs

```
R 4.2.2 · /cloud/project/

R version 4.2.2 (2022-10-31) -- "Innocent and Trusting"
Copyright (C) 2022 The R Foundation for Statistical Computing
Platform: x86_64-pc-linux-gnu (64-bit)

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.

> |
```

Environment History Connections Tutorial

R 125 MiB Global Environment

Environment is empty

Files Plots Packages Help Viewer Presentation

New Folder New Blank File Upload Delete Rename More

Cloud > project				
	Name	Size	Modified	
	..			
<input type="checkbox"/>	.Rhistory	0 B	Mar 14, 2023, 3:01 PM	
<input type="checkbox"/>	project.Rproj	205 B	Mar 14, 2023, 3:01 PM	

# GLOSARIO R



**Correr/ejecutar:** pedir que R realice algo, en otras palabras, estamos dando una instrucción o una orden,

**Devolver:** El resultado de una ejecución que hayamos solicitado. Puede ser un gráfico, un resultado matemático, etc.

**Lllamar/invocar un objeto:** Utilizamos el nombre de un objeto creado para realizar una acción con él,

**Script:** Es el archivo donde escribimos los comandos. Lo reconocemos por la extensión **.R**,

## Objetos

Cualquier cosa que existe en R y que tiene un nombre.

- Variables
- Funciones
- Gráficos
- Tablas
- Vectores

Los objetos tienen atributos, el principal es el **NOMBRE**.

Los podemos crear, modificar y eliminar.



## OPERADORES

Operadores comparación		Operadores aritméticos		Operadores lógicos	
>	Mayor	+	Suma	!	Negación
>=	Mayor o igual a	-	Resta		O Lógico.
<	Menor	/	Cociente	&	Y Lógico
<=	Menor o igual a	*	Multiplicación	Asignación	
==	Igual	^	Potencia	<-	asignación
!=	Diferente de	%%	Módulo	=	asignación dentro de funciones

Buceando **R**



## TIPOS DE DATOS

Tipo o clase de dato	Ejemplo
character	nombres <- "Esto es un caracter"
factor	mes <- factor(months, ordered = TRUE)
numeric/integer	anio <- 2021
numeric/float	iva <- 1.21
logical	es_mayor <- TRUE

## Funciones para trabajar con tipos de datos

Consultar la clase:

- `class(x)`

Consultar si es una clase en especial:

- `is.numeric(x)`
- `is.character(x)`
- `is.logical(x)`

Modificar el tipo de dato:

- `as.numeric(x)`
- `as.character(x)`
- `as.logical(x)`
- `as.factor(x)`

# FUNCIONES



Son conjuntos de operaciones. Las funciones aceptan argumentos, es decir, especificaciones sobre cómo deben funcionar. Siempre que las llamamos algo hacen, lo veamos o no. La mayoría de las funciones que vamos a usar ya están definidas en R, otras las encontramos en las librerías, y también podemos crear nuestras propias funciones.

Ejemplo:

`mean(x= vector, na.rm =TRUE )`

Nombre de la  
función

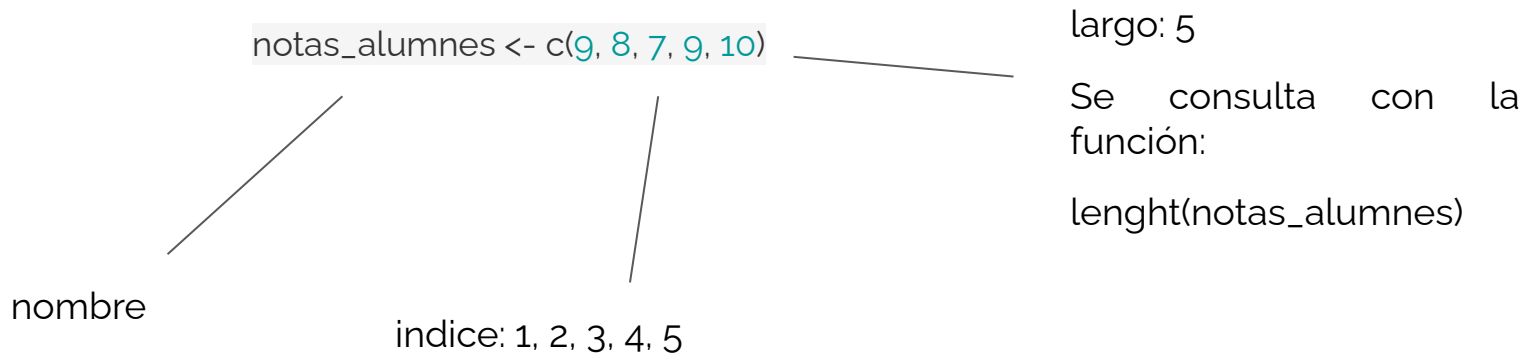
Argumento  
obligatorio

Argumento  
opcional

### Vectores

Los vectores tienen diferentes atributos. Los más importantes son:

- Nombre: el nombre con el que es definido.
- Índice: es la posición de cada elemento
- Largo: cantidad de elementos dentro del mismo



Podemos acceder a cada elemento por la posición (índice), ejemplo: notas\_alumnes[3]

## Matrices

Los atributos más importantes de las matrices son:

- Nombre: el nombre con el que es definido.
- Índice: es la posición de cada elemento
  - Hay una posición de columna
  - Y una posición de fila

```
mi_matriz <- matrix(1:9, nrow = 3, ncol = 3, byrow = TRUE)
```

nombre

índice de fila

índice de columna

[1,]	1	2	3
[2,]	4	5	6
[3,]	7	8	9

`mi_matriz[2, 3]`

Podemos acceder a los elementos internos de una matriz, indicando la dimensión de fila y columna

## Data frames

Al igual que las matrices los data frames tienen dos dimensiones:

- Nombres: el nombre con el que es definido la tabla.
  - Las columnas tienen nombres
  - Las filas pueden tener nombre también.
- Índice: es la posición de cada elemento
  - Hay una posición de columna
  - Y una posición de fila

```
notas_alumnes <- c(9, 8, 7, 9, 10)
nombre_alumnes <- c('Maria', 'Juana', 'Pedro', 'Mariana', 'Inti')
concepto_alumnes <- c('Muy bueno', 'Bueno', 'Muy bueno', 'Regular', 'Bueno')
curso <- data.frame(nombres = nombre_alumnes,
                    notas = notas_alumnes,
                    concepto= concepto_alumnes)
```



nombre

## Data frames

	nombres	notas	concepto
1	Maria	9	Muy bueno
2	Juana	8	Bueno
3	Pedro	7	Muy bueno
4	Mariana	9	Regular
5	Inti	10	Bueno

nombres e índices de filas

nombres e índices de columnas

`curso[5, 2]`

Se puede acceder a una columna entera y tratarlas como un vector a partir de concatenar el nombre de la tabla y el de la columna con el operador \$.

`curso$notas` → ahora es un vector numérico



### Listas

La principal característica es que puede contener adentro cualquier tipo de objeto, incluso listas. Sí las listas puede tener listas!!!

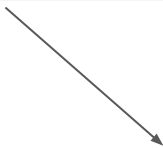
Como los otros objetos también tienen atributos como :

- Nombres: de la lista y de cada objeto dentro de la lista.
- Índice: es la posición de cada elemento
  - A la vez cada objeto también tiene sus índices.

```
mi_lista <- list(tabla= curso, matriz= mi_matriz, mi_nombre= 'Leandro', mi_edad=25)
```



nombre



```
mi_lista[[1]]
```

El primer elemento de la lista es un data frame

## TIPOS DE OBJETOS

Tipo o clase de dato	Ejemplo	Características
vectores	<pre>notas_alumnes &lt;- c(9, 8, 7, 9, 10)</pre>	Una dimensión. Acepta un solo tipo de dato.
matrices	<pre>mi_matriz &lt;- matrix(1:9, nrow = 3, ncol = 3, byrow = TRUE)</pre>	Dos dimensiones. Sólo aceptan un tipo de dato.
data frames	<pre>curso &lt;- data.frame(nombres = nombre_alumnes, notas = notas_alumnes, concepto= concepto_alumnes)</pre>	Son las famosas tablas. Dos dimensiones, y cada columna es un vector aparte. Cada columna puede tener datos diferentes.
listas	<pre>mi_lista &lt;- list(tabla= curso, matriz= mi_matriz, mi_nombre= 'Leandro', mi_edad=25)</pre>	Tienen N dimensiones.. Acepta cualquier objeto dentro del mismo. Son muy útiles para trabajar con múltiples tablas.

# Leyendo archivos



En R podemos leer todo tipo de archivos. Sólo tenemos que encontrar la función correcta y la ubicación del archivo en nuestra computadora.

```
encuesta_cc <- read.csv("datos/encc_2017.csv")
```

nombre del  
objeto de R

función para  
leer csv

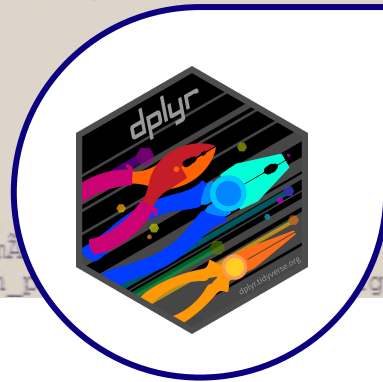
ubicación del archivo  
en mi computadora

# PAQUETES QUE VAMOS A UTILIZAR



## Encuesta Permanente de Hogares

Permite la descarga y manipulación de las bases de la EPH a través de algunas funciones.



## Tidyverse dplyr

Agrupar un conjunto de funciones dedicadas a la manipulación de datos.



## Tidyverse ggplot

Condensa una serie de funciones para representar gráficamente la información que estamos analizando



## Simple features

Paquete de funciones que permiten cargar, manipular y calcular nuevas variables con datos espaciales.