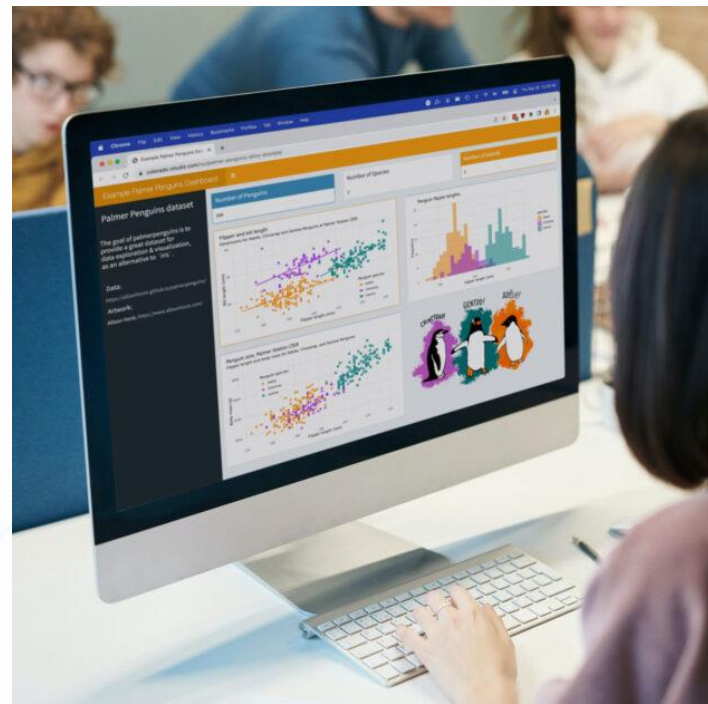


Introducción a la Estadística para el análisis político usando R

Christian Chiroque



● Sesión 1

> Riqsinakusun!

- ¿Cuál es tu nombre?
- ¿Algún tema de interés en particular dentro de la Estadística?
- ¿Cuál es tu principal pasatiempo?



> El reto



> Estadística como herramienta de gobierno

- La estadística es una rama de las matemáticas que se dedica a la recolección, análisis, interpretación y presentación de datos.
- Aunque originada como una subdisciplina matemática, la estadística ha evolucionado para convertirse en una ciencia interdisciplinaria por sí misma, siendo fundamental en diversas áreas, desde las ciencias sociales hasta la biología, la economía y la física, entre otras.
- El término "*estadística*" proviene de la palabra italiana "*statista*", que se refiere a un estadista o alguien involucrado en asuntos de estado. La estadística, en sus orígenes, estaba relacionada con la recopilación de datos para el gobierno y la descripción de estados o comunidades (a mediados del XVIII, Gottfried Achenwall, un jurista y economista alemán, adopta y populariza el término "*statistik*" en Alemania. En este contexto, se refiere a análisis comparativos de diferentes países).



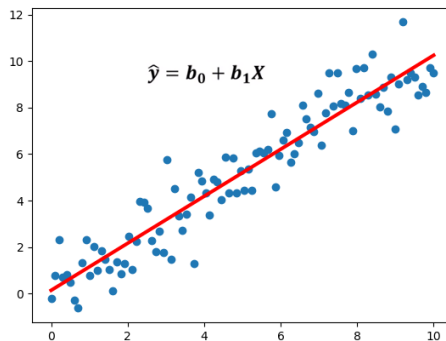
> ¿Variar “estadísticas”?



Estadística descriptiva

Análisis, organización y presentación de datos.

Nos interesa descubrir patrones con los datos que YA TENEMOS.



Estadística Inferencial

Se ocupa de la inferencia y la predicción a partir de datos.

Estudio de una población a través de una muestra.

 OpenAI



Y luego qué?

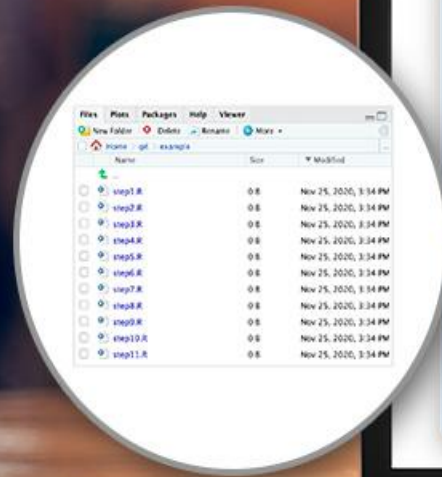
Estadística (computacional) como base para comprender la nueva tecnología.

> ¿Puedo usar ChatGpt?

Claro que sí.

La inteligencia artificial generativa es una gran ayuda para los programadores y aquellos que utilizamos R como lenguaje no somos la excepción.

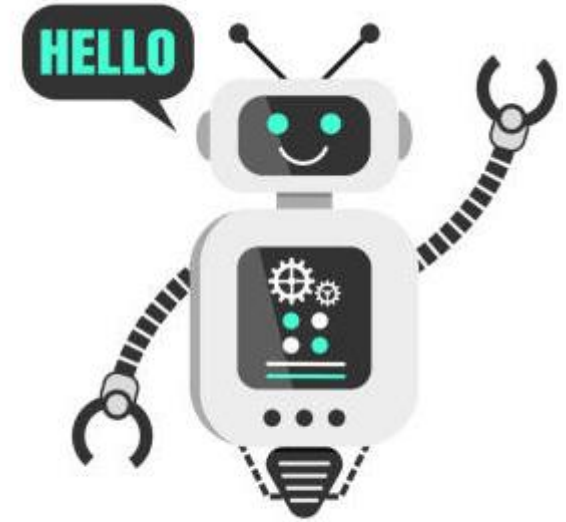




El R como herramienta para el análisis estadístico

¿Por qué un lenguaje de programación?

- ❑ Un lenguaje de programación, como R, es una forma de comunicarnos con la computadora para decirle qué queremos que haga, usando un conjunto de reglas y comandos específicos.
- ❑ Es como un idioma que la computadora entiende y que nos permite realizar tareas, análisis o crear programas.





- ☐ Es un **lenguaje de programación y software libre**, lo que significa que es gratuito y su código es accesible para todos.
- ☐ **Especializado** en análisis estadístico y visualización de datos.
- ☐ Posee una **comunidad activa** que desarrolla y comparte constantemente paquetes adicionales, ampliando sus capacidades.
- ☐ **Compatible** con diversas plataformas y sistemas operativos.
- ☐ La primera versión de R fue lanzada el 29 de febrero de **2000**. Sin embargo, es importante mencionar que el desarrollo de R comenzó mucho antes, en 1993, como una implementación alternativa del lenguaje S, que existía desde 1976.

Un lenguaje orientado a objetos (POO)

- ❑ Trabajamos con objetos.
- ❑ Los objetos tienen clases.
- ❑ Las clases definen a los objetos (atributos, propiedades y comportamientos que puede tener) y lo que podemos hacer con ellos.



Construyendo algo nuevo...



CAJAS DE
HERRAMIENTAS
(*Packages*)

CAJA DE HERRAMIENTAS
ELEGIDA
(*library*)

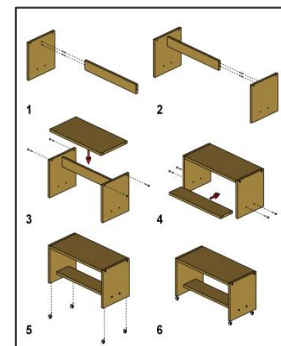


HERRAMIENTAS A
UTILIZAR
(*Functions*)

OBJETO
(*Data*)



MESA DE TRABAJO
(*R Console*)



INSTRUCTIVO
(*Script*)

CREAMOS UN NUEVO PROYECTO

SCRIPT
(espacio para escribir instrucciones)
"Manual"

```
1 #SI USAS EL sav (del SPSS)
2 library(foreign) #necesario para que funcione read.spss
3 ejemplo<- read.spss("ejemplo.sav",use.value.labels=TRUE, max.value.labels=Inf,
4
5 #SI USAS CSV (grabado desde EXCEL)
6 ejemplo<-read.csv("ejemplo.csv")
7
8 #SI USAS rda (formato propio de R - igual que Rdata)
9 load("ejemplo.rda")
```

CONSOLA
(espacio para escribir instrucciones)

```
> setwd("C:/Users/user/Desktop/PRUEBA R")
> |
```

> (prompt)

Significa que R espera líneas de comando

Historial de objetos almacenados en la memoria:
Environment is empty

- **Importar datos.**
- **Limpiar historial**

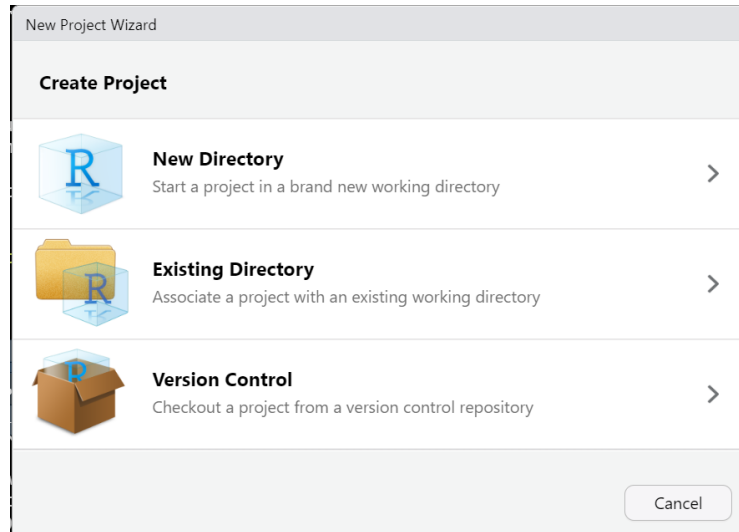
Sección Multifunción:

- **FILES:** Permite ver los archivos del directorio
- **PLOTS:** Los gráficos creados
- **PACKAGES:** Los paquetes disponibles
- **HELP:** IMPORTANTE! R Studio brinda tutoriales para guiar la instalación y uso de paquetes.

Por qué utilizar un Proyecto en R?

Un proyecto en R es una forma de organizar y agrupar todos los archivos relacionados con un análisis específico en un solo lugar. Al usarlo, se garantiza que el código se ejecute correctamente, evitando errores al acceder a datos o scripts.

Usar proyectos en R permite mantener un trabajo ordenado y claro. Ayuda a separar diferentes análisis, facilitando la reproducibilidad y evitando confusiones al mezclar archivos de diferentes tareas. Es esencial para una gestión eficiente de los análisis en R.



Los objetos indivisibles en el R



Numeric

Un tipo de dato que representa números, ya sean enteros o decimales.

5



Character

Un tipo de dato que representa texto o cadenas de caracteres.

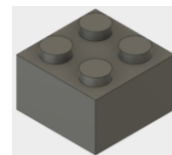
“hola”



Logical

Un tipo de dato que representa valores de verdad: verdadero (TRUE) o falso (FALSE).

TRUE
FALSE



Factor

Un tipo de dato que representa categorías o niveles dentro de un conjunto de datos.

1 A
2 B

mi_codigo.R

```
1  1.5
2
3  "hola"
4
5  TRUE
6
7  1 + 1    # Podemos utilizar R como calculadora
8
9
10
11
12
13
14
```



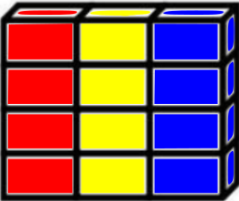

Operadores de asignación

- ❑ El signo <- indica la asignación de un valor a un nombre de variable. Ej: nombre_variable <- "Christian"
- ❑ El valor es almacenado en la memoria de la computadora.
- ❑ La asignación une el valor y la variable.
- ❑ Se recupera el valor al llamar a la variable al tipear el nombre_variable.
- ❑ Si se vuelve a asignar un valor a una variable, el anterior se elimina de la memoria.
- ❑ Recordar: También es posible asignar con el signo =, sin embargo, este tiene algunas consideraciones especiales




```
1  Numero<-1.5          # Creamos el objeto
2  Numero               # Lo "llamamos"
3
4  Caracter<-"hola"     # Creamos el objeto
5  Caracter             # Lo "llamamos"
6
7  Logico<- TRUE        # Creamos el objeto
8  Logico               # Lo "llamamos"
9
10
11
12
13
14
```

Las estructuras básicas

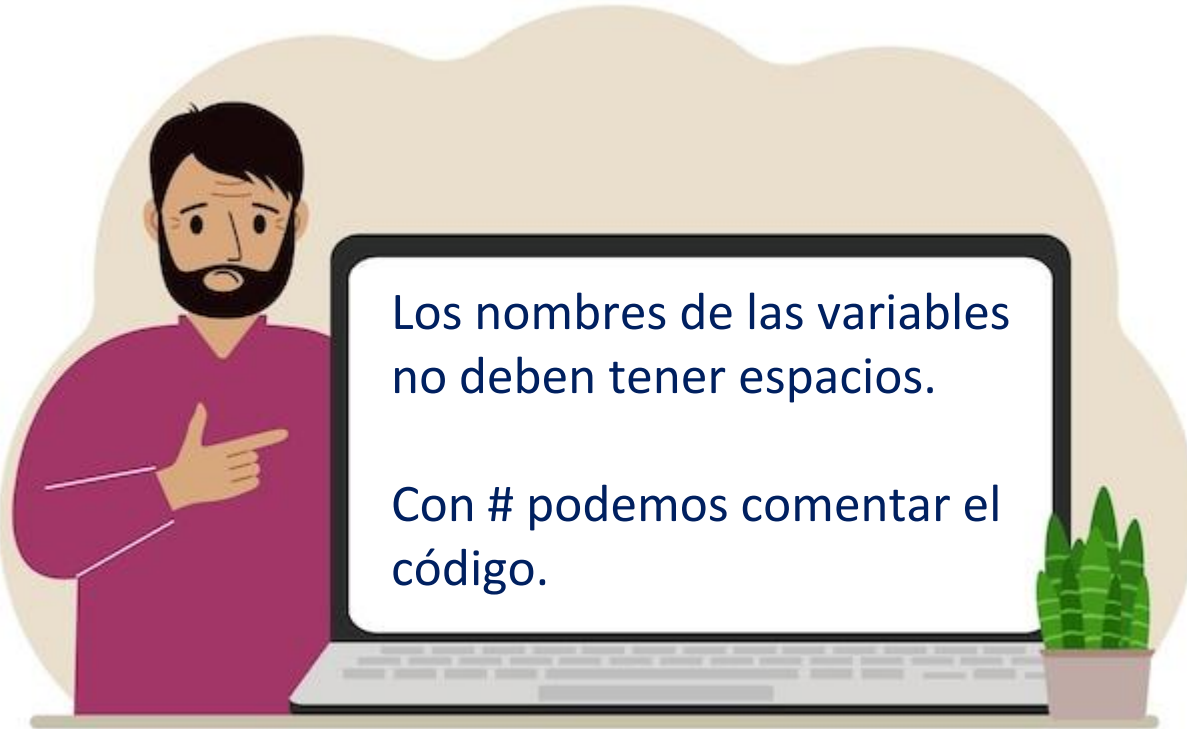
	Mismo tipo de datos	Diferentes tipos de datos
1D	vector 	lista <ul style="list-style-type: none">- vectores- matrices- data frames- listas
2D	matriz 	data frame 
nD	array 	

Los vectores

- ❑ Un vector es una colección de uno o más datos del mismo tipo.
- ❑ Tipo. Un vector tiene el mismo tipo que los datos que contiene. Si tenemos un vector que contiene datos de tipo numérico, el vector será también de tipo numérico. Los vectores son atómicos, pues sólo pueden contener datos de un sólo tipo, no es posible mezclar datos de tipos diferentes dentro de ellos.
- ❑ Largo. Es el número de elementos que contiene un vector. El largo es la única dimensión que tiene esta estructura de datos.
- ❑ NO TIENE DIMENSIÓN (dim)

	Mismo tipo de datos
1D	vector 

RECUERDA




```
1 function(objeto, argumento)
```

```
2 |  
3 |  
4 | La función es nuestra herramienta para una  
5 | determinada actividad.
```

```
6 |  
7 | Este objeto puede ser una base de datos,  
8 | un vector, una tabla, entre muchos otros.
```

```
9 |  
10 | Una función normalmente tienen más de un  
11 | argumento. Encontrarás algunas con dos o  
12 | tres, pero también unas con 10 o 20!
```

```
1  Vector_numerico<-c(1, 3, 5)           # Creamos el objeto
2  Vector_numerico                       # Lo "llamamos"
3  class(Vector_numerico)                # Clase del objeto
4
5  Vector_character<-c("hola", "chau")  # Creamos el objeto
6  Vector_character                      # Lo "llamamos"
7  class(Vector_character)               # Clase del objeto
8
9  Vector_logical<- c(TRUE, FALSE)      # Creamos el objeto
10 Vector_logical                        # Lo "llamamos"
11 class(Vector_logical)                 # Clase del objeto
12
13
14
```

Y qué pasó con los factores?

- ❑ Un factor es un tipo de datos específico a R. Puede ser descrito como un dato numérico representado por una etiqueta.
- ❑ Supongamos que tenemos un conjunto de datos que representan el sexo de personas encuestadas por teléfono, pero estos se encuentran capturados con los números 1 y 2. El número 1 corresponde a femenino y el 2 a masculino.
- ❑ A diferencia del carácter, el factor tiene niveles (levels)

```
> sexo <- c("M", "H", "M", "M", "M", "H", "M", "M", "H", "H");sexo
[1] "M" "H" "M" "M" "M" "H" "M"
[8] "M" "H" "H"
>
> sexo <- factor(sexo);sexo
[1] M H M M M H M M H H
Levels: H M
> sexo <- c(1,0,0,0,1,0);sexo
[1] 1 0 0 0 1 0
> sexo <- factor(sexo);sexo
[1] 1 0 0 0 1 0
Levels: 0 1
```

```
1  sexo<-c(1,2,2,1)          # Variable en el que 1 = Mujer y 2 = Hombre
2
3  sexo<-factor(sexo,        # Objeto
4                levels= 1:2, # Los niveles del factor
5                labels=c("Mujer", "Hombre"), # Las etiquetas
6                ordered = FALSE) # Debes indicar si es ordenado o no
7
8  sexo          # Lo "llamamos"
9
10 class(sexo)   # Clase del objeto
11
12
13
14
```



En el R es clave
asignar la clase de
objeto adecuada a
las variables con las
que estamos
trabajando!

¿Recuerdas las variables y tipos de variables?

Una variable es una característica, cantidad o atributo que puede ser medido o contado.



Variable cualitativa

Es una característica o atributo que representa categorías o etiquetas sin valor numérico intrínseco.



Variable cuantitativa

Es una característica que representa cantidades y puede someterse a operaciones aritméticas.

Algunos ejemplos



Variable cualitativa

Estatus Marital: Soltero, casado, divorciado, viudo.

Religión: Católica, protestante, judía, hinduista, etc.

Nacionalidad: Estadounidense, francesa, japonesa, etc.

Lengua Materna: Español, inglés, mandarín, árabe, etc.

Nivel educativo: Primaria, Secundaria, Superior.



Variable cuantitativa

Número de Hijos: 0, 1, 2, 3, ...

Ingreso Anual: Cantidad en dólares, euros, pesos, etc.

Años de Educación: Número de años completados en el sistema educativo.

Índice de Desarrollo Humano (IDH): Medida que varía entre 0 y 1, indicando el nivel de desarrollo humano de un país o región.

Edad: Número de años que una persona ha vivido.

Pregunta



Variable cualitativa

Qué tipo de objetos serán en el R??

FACTOR



Variable cuantitativa

Qué tipo de objetos serán en el R??

NUMERIC

Los data frames

- ❑ Los data frames son estructuras de datos de dos dimensiones (rectangulares) que pueden contener datos de diferentes tipos, por lo tanto, son heterogéneas. Compuesto por vectores.
- ❑ Estructura más usada para ciencia de datos.
- ❑ Mientras que en una matriz todas las celdas deben contener datos del mismo tipo, los renglones de un data frame admiten datos de distintos tipos, pero sus columnas conservan la restricción de contener datos de un sólo tipo.
- ❑ En términos generales, los renglones en un data frame representan casos, individuos u observaciones, mientras que las columnas representan atributos, rasgos o variables

data frame

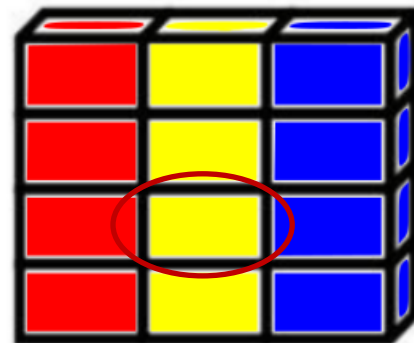


```
1  mi_df <- data.frame(Vector_numerico,      # Creamos el objeto
2                        Vector_character,
3                        Vector_logical)
4
5
6  mi_df                                     # Lo "llamamos"
7
8
9
10
11
12
13
14
```

Los Índices

- ❑ Usar índices para obtener subconjuntos es el procedimiento más universal en R, pues funciona para todas las estructuras de datos.
- ❑ Un índice en R representa una posición.
- ❑ Cuando usamos índices le pedimos a R que extraiga de una estructura los datos que se encuentran en una o varias posiciones específicas dentro de ella
- ❑ Podemos usar los [] o el símbolo \$ (con este último el elemento extraído siempre será una columna)

data frame



Ubicación:
dataframe[3,2]

mi_codigo.R

```
1  mi_df[3,2]      # Muestrame el elemento que está en la ubicación:  
2                  FILA 3, COLUMNA 2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14
```


Los paquetes en el R

- ❑ En R, los paquetes son colecciones de funciones, conjuntos de datos y documentación relacionada que amplían las capacidades del lenguaje base.
- ❑ Son herramientas creadas por la comunidad para resolver tareas específicas o simplificar ciertos procesos, permitiendo a los usuarios beneficiarse de soluciones previamente desarrolladas y compartidas.
- ❑ Para instalar un paquete debemos colocar `install.packages("nombredelpaquete")`.
- ❑ Para abrir un paquete y utilizar sus funciones debemos colocar `library(nombredelpaquete)`.



Importación

- ❑ Importar en R se refiere al proceso de cargar datos externos, provenientes de archivos como CSV, Excel, SPSS, entre otros, hacia el entorno de R para su análisis y manipulación.
- ❑ Al importar, se convierten estos datos en estructuras reconocibles por R, como data frames o tibbles, permitiendo a los usuarios trabajar con ellos utilizando las funciones y herramientas del lenguaje.
- ❑ Una de las librerías más usadas en `rio`



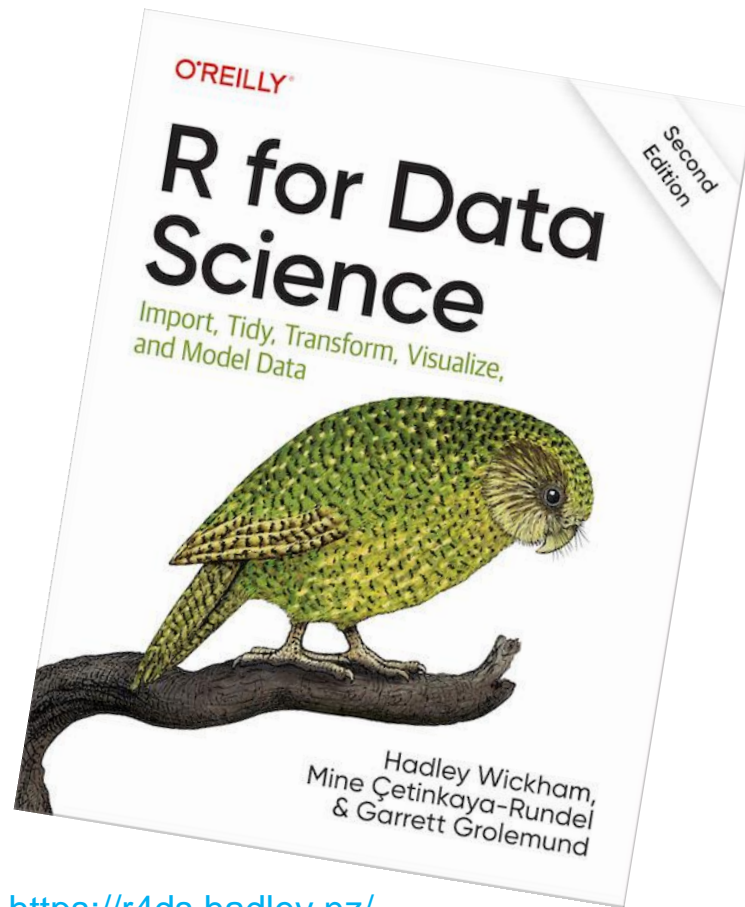
```
1  install.packages("rio")      # Instalamos el paquete
2  library(rio)                 # Abrimos el paquetev
3
4  data<-import("nombredelarchivo") # Abrimos el archivo con "import"
5
6  head(data) # Visualizamos los 10 primeros valores
7
8
9
10
11
12
13
14
```

Exploremos algunas funciones básicas

- ☐ summary()
- ☐ str()
- ☐ names()
- ☐ mean()
- ☐ median()
- ☐ sd()
- ☐ length()
- ☐ sum()
- ☐ min()
- ☐ max()
- ☐ head() y tail()

Lectura sugerida

- ❑ "R for Data Science" (2da Edición) es un libro escrito por Hadley Wickham, Mine Cetinkaya-Runder y Garrett Golemund que sirve como guía para aprender a usar R con el objetivo de realizar ciencia de datos. El texto aborda todo el proceso de análisis, desde la importación y limpieza de datos hasta la visualización y modelado, utilizando principalmente el conjunto de paquetes conocido como tidyverse.
- ❑ Este libro se destaca por su enfoque práctico, proporcionando múltiples ejemplos y ejercicios, y es ampliamente reconocido como una de las referencias esenciales para quienes desean aprender o profundizar en la ciencia de datos con R.



<https://r4ds.hadley.nz/>