

Fundamentos de ciencia de datos con R - Módulo 1

Clase 2: Lecturas y exportación de bases de datos - librerías - objetos de R

CEPAL - Unidad de Estadísticas Sociales

2025-10-28

Carga e importación de librerías

Antes de trabajar con datos en R, es necesario cargar las librerías, que son conjuntos de funciones ya creadas que nos facilitan el análisis.

- ▶ Instalar una librería (`install.packages()`) se hace solo una vez.
- ▶ Cargarla (`library()`) se debe hacer cada vez que abrimos R o RStudio.
- ▶ Algunas librerías importantes para análisis de datos son: `tidyverse`, `dplyr`, `ggplot2`, `readr`, `readxl`, entre otras.

Carga e importación de librerías

```
# Instalar (solo la primera vez):  
# install.packages("tidyverse")  
  
library(tidyverse)    # Incluye dplyr, ggplot2, readr, etc.
```

¿Por qué es importante?

Porque sin cargar las librerías, R no reconoce funciones como `filter()`, `ggplot()`, `read_csv()`, etc.

Carga e importación de base de datos

Antes de analizar, debemos importar o leer los datos y traerlos a R. Esto significa que R toma la información que está guardada en un archivo externo (como Excel, CSV, RDS, SPSS, etc.) y la convierte en un objeto dentro del entorno de trabajo.

Una vez los datos están cargados en un objeto (por ejemplo, un data frame llamado datos), podemos visualizarlos, limpiarlos, transformarlos, hacer gráficos o aplicar modelos estadísticos. Es decir, importar los datos es el paso que conecta la información real con el análisis que haremos en R.

Carga e importación de base de datos

Formatos más comunes

Tipo de archivo	Librería	Ejemplo en R
CSV	readr	<pre>datos <- read_csv("data/archivo.csv")</pre>
Texto delimitado ";"	readr	<pre>datos <- read_delim("data/archivo.txt", delim = ";")</pre>
Texto delimitado tab	readr	<pre>datos <- read_tsv("data/archivo.tsv")</pre>
Excel (.xlsx)	readxl	<pre>datos <- read_excel("data/datos.xlsx")</pre>
RDS (.rds)	Base de R	<pre>datos <- readRDS("data/base.rds")</pre>
RData (.RData)	Base de R	<pre>load("data/mi_objeto.RData")</pre>

Carga e importación de base de datos

Otros formatos

Tipo de archivo	Librería Ejemplo en R	
SPSS (.sav)	haven	<pre>datos <- read_sav("data/archivo.sav")</pre>
Stata (.dta)	haven	<pre>datos <- read_dta("data/archivo.dta")</pre>
JSON	jsonlite	<pre>datos <- fromJSON("data/archivo.json")</pre>
CSV grande (.csv.gz)	data.table	<pre>datos <- fread("data/archivo.csv.gz")</pre>

Carga e importación de base de datos

► Ejemplo: cargar un archivo RDS (.rds)

```
# Importar la base (ejemplo)
datos <- readRDS(
  "../Data/base_personas.rds"
) %>% as.data.frame() # readRDS es una funcion base de R
# Ver las primeras filas
head(datos[, 1:8], 3)
```

id_hogar	id_pers	parentesco	edad	sexo	etnia	area	ingreso
262	1	1	51	Hombre	0	1	542000.0
262	1	1	51	Hombre	0	1	536305.3
262	2	2	46	Mujer	0	1	542000.0

Objetos de R

Clases de objetos

En R, los datos no se guardan de forma suelta, sino dentro de objetos. Estos objetos pueden tener diferentes estructuras o “clases”, y conocerlas es clave para trabajar correctamente.

Objetos de R

Clases de objetos

Tipo de objeto	Descripción
Vector	Conjunto básico de valores del mismo tipo.
Factor	Vector para datos categóricos con niveles definidos.
Matrix	Arreglo rectangular de datos numéricos o de un solo tipo.
Array	Extiende la idea de matriz a más de dos dimensiones.
Data Frame	Tabla similar a Excel; cada columna puede tener distinto tipo de dato.
Tibble	Versión moderna del data frame (tidyverse), más ordenada y amigable.
List	Contenedor que puede almacenar objetos de cualquier tipo y tamaño.

Objetos de R

Clases de objetos: Vector

```
# Vector numérico
edades <- c(25, 30, 28, 40)
edades
```

```
[1] 25 30 28 40
```

```
# Vector de texto (character)
nombres <- c("Ana", "Luis", "María")
nombres
```

```
[1] "Ana"    "Luis"   "María"
```

```
# Vector lógico (TRUE / FALSE)
es_mayor_edad <- c(TRUE, TRUE, FALSE, TRUE)
es_mayor_edad
```

```
[1] TRUE TRUE FALSE TRUE
```

Objetos de R

Clases de objetos: Factor

```
sexo <- factor(c("Mujer", "Hombre", "Mujer", "Hombre"))  
levels(sexo)      # Niveles del factor
```

```
[1] "Hombre" "Mujer"
```

```
class(sexo)      # "factor"
```

```
[1] "factor"
```

Objetos de R

Clases de objetos: Matriz

```
matriz_ejemplo <- matrix(1:9, nrow = 3, ncol = 3)  
matriz_ejemplo
```

1	4	7
2	5	8
3	6	9

Objetos de R

Clases de objetos: Data Frame

```
personas <- data.frame(  
  nombre = c("Ana", "Luis", "María"),  
  edad   = c(23, 30, 28),  
  ingreso = c(1200, 1500, 1800)  
)
```

personas

nombre	edad	ingreso
Ana	23	1200
Luis	30	1500
María	28	1800

```
class(personas) # "data.frame"
```

```
[1] "data.frame"
```

Objetos de R

Clases de objetos: Tibble

```
personas_tibble <- tibble(  
  nombre = c("Ana", "Luis", "María"),  
  edad   = c(23, 30, 28),  
  ingreso = c(1200, 1500, 1800)  
)  
personas_tibble
```

nombre	edad	ingreso
Ana	23	1200
Luis	30	1500
María	28	1800

Objetos de R

Clases de objetos: Lista

```
mi_lista <- list(  
  numeros = c(1, 2, 3),  
  tabla = personas  
)
```

```
mi_lista
```

```
$numeros
```

```
[1] 1 2 3
```

```
$tabla
```

	nombre	edad	ingreso
1	Ana	23	1200
2	Luis	30	1500
3	María	28	1800