



Clase 2

Exploración de tablas

Taller de Análisis de datos I | UC | 28 de julio, 2023

👤 **José D. Conejeros** | ✉ jdconejeros@uc.cl

Guía

1. Importar
2. Explorar una tabla de datos
3. Exportar

Importar y exportar

A pesa de que el código para importar varía según formato podemos generalizarlo de la forma `read.formato()`. Por ejemplo:

```
read.table(file = "ruta/nombreakchivo.txt", header = TRUE, sep="\t", dec = ".", row.names = 1)
```

- `file=path/to/file` → Ubicación y nombre del archivo a leer.
- `header = TRUE` → Si se entregan los nombres de las variables en la primera línea.
- `sep="\t"` → Los valores en cada línea están separados por este caracter ("\t"=tabulación; "" = blanco; ";" = ,; etc.).
- `dec="."` → Separador decimal para números ("." ó ",").
- `row.names = 1` → Si la primera columna entrega el nombre de los individuos.

Para saber más de esta función pueden escribir en su consola de R para revisar la documentación:

```
?read.table # o  
help(read.table)
```

Funciones para importar datos

- `read.table()`: Lee conjuntos de datos presentados como tablas, como suele ser el caso de Estadística.
- `read.csv()`: Lee archivos delimitados por coma.
- `read.csv2()`: Lee archivos delimitados por punto y coma (común en países donde se utiliza , como separador decimal)
- `read.delim()`: Lee archivos con cualquier delimitador
- `read.delim2()`: Igual que el anterior, pero se usa cuando los números tienen comas como decimales en lugar de puntos.
- `read.ftable()`: Lee tablas de contingencia (frecuencias).

Hay otras librerías que permiten importar otros formatos de datos. Por ejemplo, el paquete `readxl` tiene una función que facilita la importación de tablas de excel: `readxl::read_excel()`

También está la manera manual desde las opciones de R

File > Import Dataset > Form...

Exploración

Lo primero que uno realiza en cual proyecto de análisis de datos es explorar sus datos:

- Vista panorámica de los datos: columnas, filas, estructura
- Vista previa de la base de datos
- Exploración de etiquetas
- Estadísticos descriptivos
- Valores fuera de rango o missing
- Imputaciones de datos

Exploración

Algunas funciones útiles para la exploración de datos son:

- `head(df, k)`: Muestra los primeros k registros.

```
head(mtcars)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225	105	2.76	3.460	20.22	1	0	3	1

- `dim(df)`: Filas y columnas de un objeto.

```
dim(mtcars)
```

```
[1] 32 11
```

Exploración

- `tail(df,k)`: Muestra los últimos k registros.

```
tail(mtcars)
```

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Porsche 914-2	26.0	4	120.3	91	4.43	2.140	16.7	0	1	5	2
Lotus Europa	30.4	4	95.1	113	3.77	1.513	16.9	1	1	5	2
Ford Pantera L	15.8	8	351.0	264	4.22	3.170	14.5	0	1	5	4
Ferrari Dino	19.7	6	145.0	175	3.62	2.770	15.5	0	1	5	6
Maserati Bora	15.0	8	301.0	335	3.54	3.570	14.6	0	1	5	8
Volvo 142E	21.4	4	121.0	109	4.11	2.780	18.6	1	1	4	2

- `length(df)`: Número de objetos dentro del objeto df.

```
length(mtcars)
```

```
[1] 11
```

Exploración

Algunas funciones útiles para la exploración de datos son:

- `str(df)`: Estructura de la base de datos df.

```
str(mtcars)
```

```
'data.frame':   32 obs. of  11 variables:
 $ mpg  : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
 $ cyl  : num   6 6 4 6 8 6 8 4 4 6 ...
 $ disp: num  160 160 108 258 360 ...
 $ hp   : num  110 110 93 110 175 105 245 62 95 123 ...
 $ drat: num   3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
 $ wt   : num   2.62 2.88 2.32 3.21 3.44 ...
 $ qsec: num   16.5 17 18.6 19.4 17 ...
 $ vs   : num   0 0 1 1 0 1 0 1 1 1 ...
 $ am   : num   1 1 1 0 0 0 0 0 0 0 ...
 $ gear: num   4 4 4 3 3 3 3 4 4 4 ...
 $ carb: num   4 4 1 1 2 1 4 2 2 4 ...
```

- `class(df)`: Naturaleza o clase del objeto df.

```
class(mtcars)
```

Exploración

Algunas funciones útiles para la exploración de datos son:

- `names(df)`: Nombres del objeto df.

```
names(mtcars)
```

```
[1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"  
[11] "carb"
```

En el taller puede encontrar más funciones

Exportar bases de datos

Exportar datos es darle un formato a una fuente interna R para que sea legible por otro lenguaje o procesador.

Para eso uno debe conocer el formato de datos que quiere importar: `.dta`, `.xlsx`, `.csv`, `.txt`, `.rds`, `.RData`, `.xpt`, `.mat`, `.mtp`, etc...

A pesa de que el código para importar varía según formato podemos generalizarlo de la forma `write.formato()`. Por ejemplo:

```
write.table(x=data, file = "ruta/nombreakchivo.txt", sep=" ", dec = ".", row.names = TRUE
```

- `x`: corresponde al objeto de R (marco de datos) que queremos exportar
- `file=path/to/file` → Ubicación y nombre del archivo a leer.
- `sep="\t"` → Los valores en cada línea están separados por este caracter ("`\t`"=tabulación; "" = blanco; ";" = ,; etc.).
- `dec="."` → Separador decimal para números ("." ó ",").
- `row.names = TRUE` → Si devuelve el nombre de las filas.

Para saber más de esta función pueden escribir en su consola de R para revisar la documentación:

```
?write.table; help(write.table)
```

Funciones para exportar datos

- `write.table()`: Exporta conjuntos de datos presentados como tablas, como suele ser el caso de Estadística.
- `write.csv()`: Exporta archivos delimitados por coma.
- `write.csv2()`: Exporta archivos delimitados por punto y coma (común en países donde se utiliza , como separador decimal)
- `write.delim()`: Exporta archivos con cualquier delimitador
- `write.ftable()`: Exporta tablas de contingencia (frecuencias).

Hay otras librerías que permiten importar otros formatos de datos. Por ejemplo, el paquete `writexl` tiene una función que facilita la exportación a tablas de excel: `writexl::write_xlsx()`

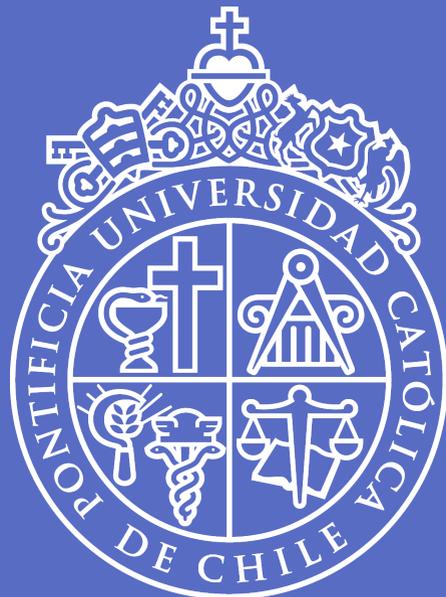
Referencias

Wickham, H., & Grolemund, G. (2016). R for data science: import, tidy, transform, visualize, and model data. " O'Reilly Media, Inc.". Recurso en línea: <https://r4ds.hadley.nz/>

Urdinez, F., & Cruz, A. (2020). R for Political Data Science: A Practical Guide. CRC Press. Cap. 1 al 4. Recurso en línea en español: <https://arcruz0.github.io/libroadp/>

Posit Cheatsheets ("hojas de trucos"): <https://posit.co/resources/cheatsheets/?type=posit-cheatsheets/>

Página oficial de Tidyverse: <https://www.tidyverse.org/>



Clase 2

Exploración de tablas

28 de julio, 2023

 **José D. Conejeros** |  jdconejeros@uc.cl |  JDConejeros