

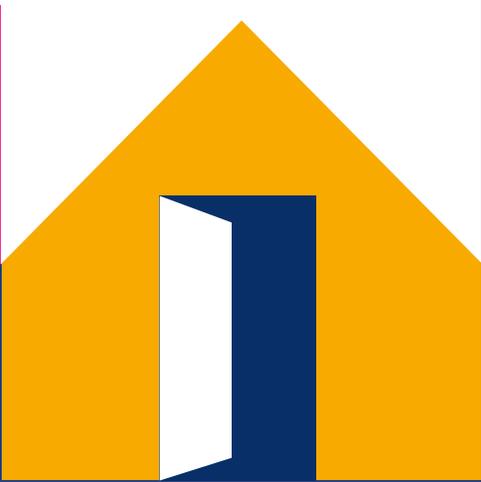


Introducción a *dataframe* y *tidyverse*

Curso: R básico



05.



Introducción a dataframes y *ti* *dyverse*

Introducción a dataframes y *tidyverse*

Como mencionamos, los ***data frames*** son la estructura de datos a la que nos enfrentaremos más comúnmente y en la que se enfoca este curso. Para trabajar con estos, el paquete ***tidyverse*** es de gran ayuda.

¿Qué es tidyverse? Es una serie de paquetes para el trabajo con *data frames* que funcionan en armonía, pues fueron programados con la misma filosofía y diseño de datos.

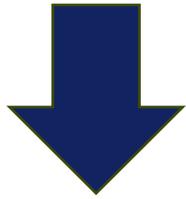
Incluye paquetes como:



Nota: la implementación de los *data frames* en el *tidyverse* es llamada *tibble*



Instalemos ***tidyverse*** y aprovechemos de instalar **guaguas** que nos proveerá datos del número de personas nacidas con cada nombre en cada año:



```
install.packages('tidyverse')  
install.packages('guaguas')
```

Luego los cargamos en nuestra sesión de la siguiente forma:

```
library(tidyverse)  
library(guaguas)
```



Carguemos la base en nuestro ambiente de trabajo:

```
datos = guaguas
```

Algunas exploraciones básicas:

Dimensiones:

```
dim(datos)  
## [1] 858782      5
```

N° de filas:

```
nrow(datos)  
## [1] 858782
```

N° de columnas:

```
ncol(datos)  
## [1] 5
```





Nombres de las columnas:

```
names(datos)
```

```
## [1] "anio"      "nombre"    "sexo"      "n"         "proporcion"
```

Tipos de datos y mirada inicial:

```
str(datos)
```

```
## tibble [858,782 x 5] (S3: tbl_df/tbl/data.frame)
## $ anio      : num [1:858782] 1920 1920 1920 1920 1920 1920 1920 1920 1920 1920
## ...
## $ nombre    : chr [1:858782] "María" "José" "Juan" "Luis" ...
## $ sexo      : chr [1:858782] "F" "M" "M" "M" ...
## $ n         : num [1:858782] 2130 984 636 631 426 340 326 289 277 269 ...
## $ proporcion: num [1:858782] 0.1045 0.0483 0.0312 0.031 0.0209 ...
```





Valores únicos de una variable:

```
distinct(datos, sexo)
```

```
## # A tibble: 3 x 1
##   sexo
##   <chr>
## 1 F
## 2 M
## 3 I
```

Conteo de valores únicos de una variable:

```
count(datos, sexo)
```

```
## # A tibble: 3 x 2
##   sexo      n
##   <chr> <int>
## 1 F    531038
## 2 I      318
## 3 M    327426
```



¡Muchas gracias!





www.ine.gob.cl