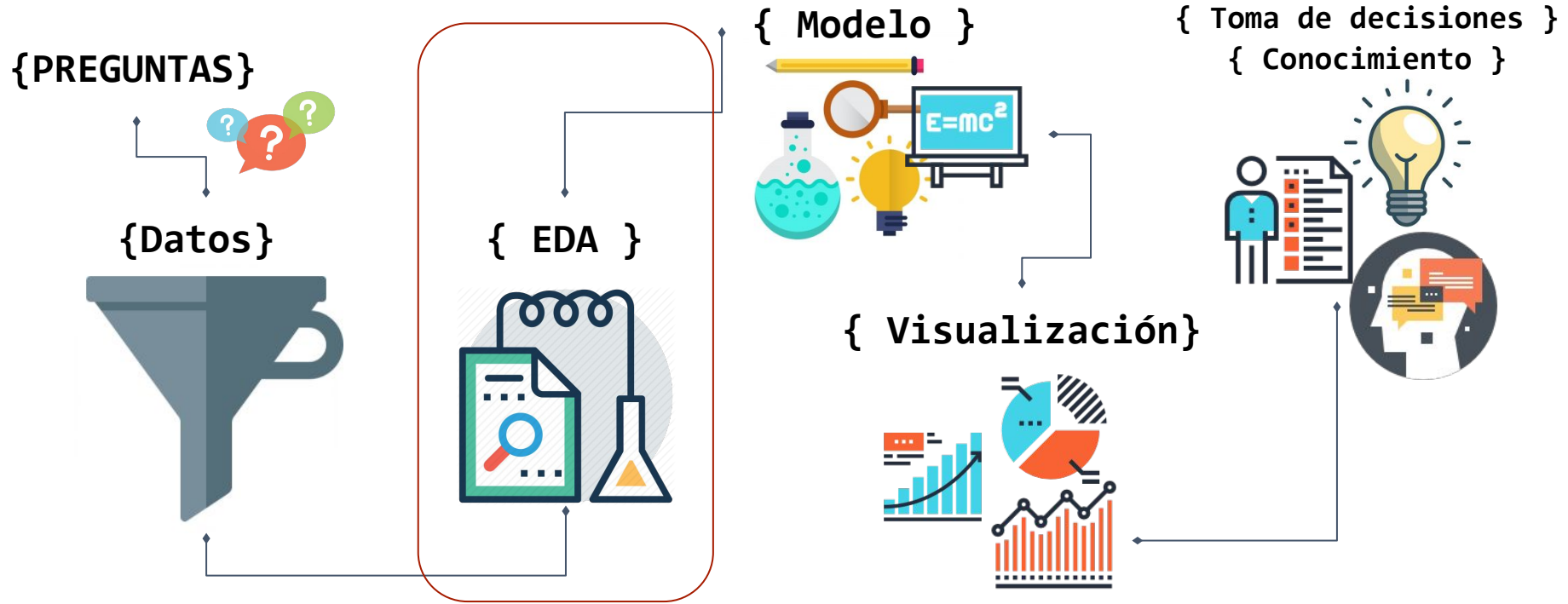
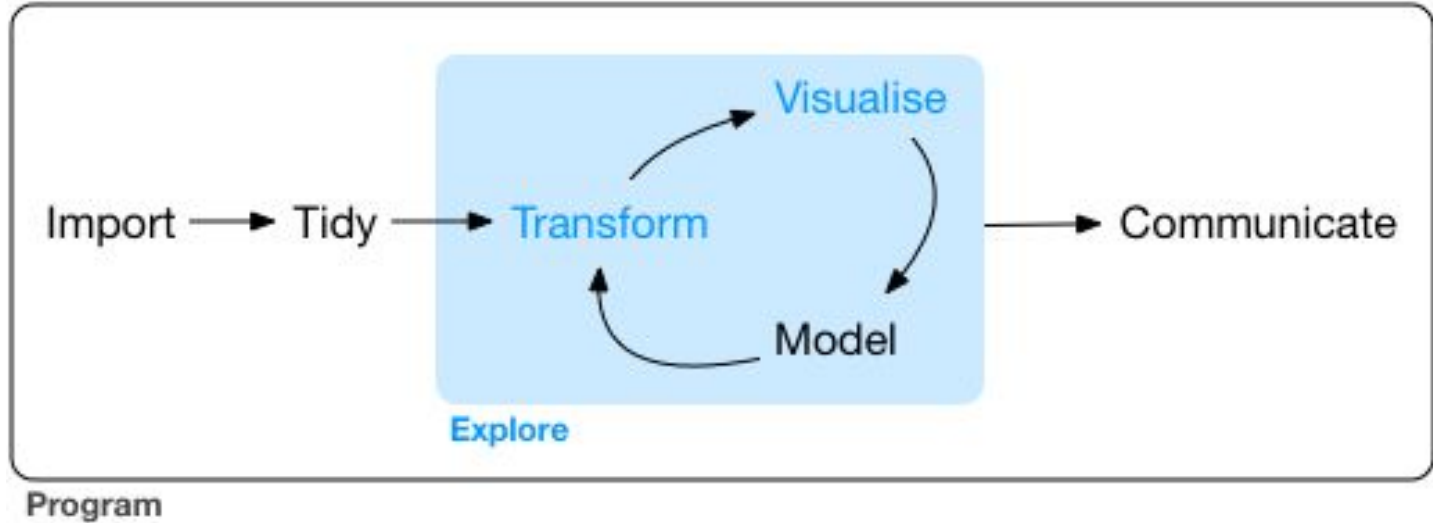


{ DataFrames }

PROCESO DE ANÁLISIS



QUE PUEDO HACER CON R?



VECTORES

Anteriormente aprendimos sobre variables que son objetos que tiene un solo dato, en el caso de los vectores son objetos que tienen varios datos pero son todos del mismo tipo.



VECTORES

También los podemos definir como **matrices de una dimensión**, que contienen en cada una de sus posiciones datos numéricos, booleanos o de tipo texto.
Para crear un vector se utiliza la función “c()” denominada combinación.

```
a= c(1,2,3,4)
```

```
b= c("a","b","c")
```

```
c= c(TRUE,FALSE,TRUE)
```

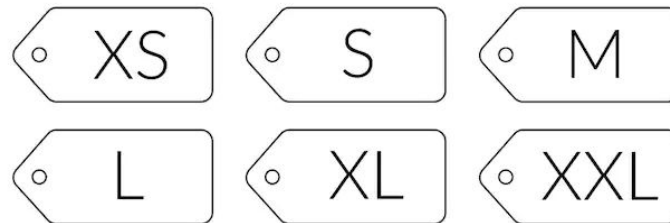


Factores en R

LOS FACTORES son una estructura de datos para trabajar con **VARIABLES CATEGÓRICAS.**



Las **VARIABLES CATEGÓRICAS** son datos que toman una cantidad finita de valores.



Variables Continuas y Discretas

CONTINUOUS

measured data, can have ∞ values within possible range.



I AM 3.1" TALL
I WEIGH 34.16 grams

DISCRETE

OBSERVATIONS CAN ONLY EXIST AT LIMITED VALUES, OFTEN COUNTS.



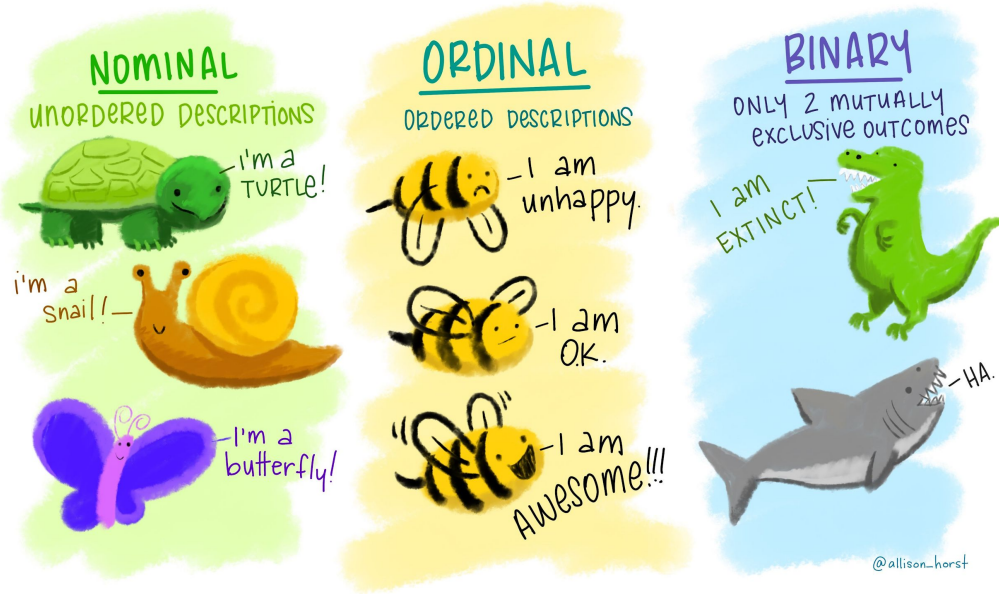
I HAVE 8 LEGS
and
4 SPOTS!

@allison_horst

https://twitter.com/allison_horst



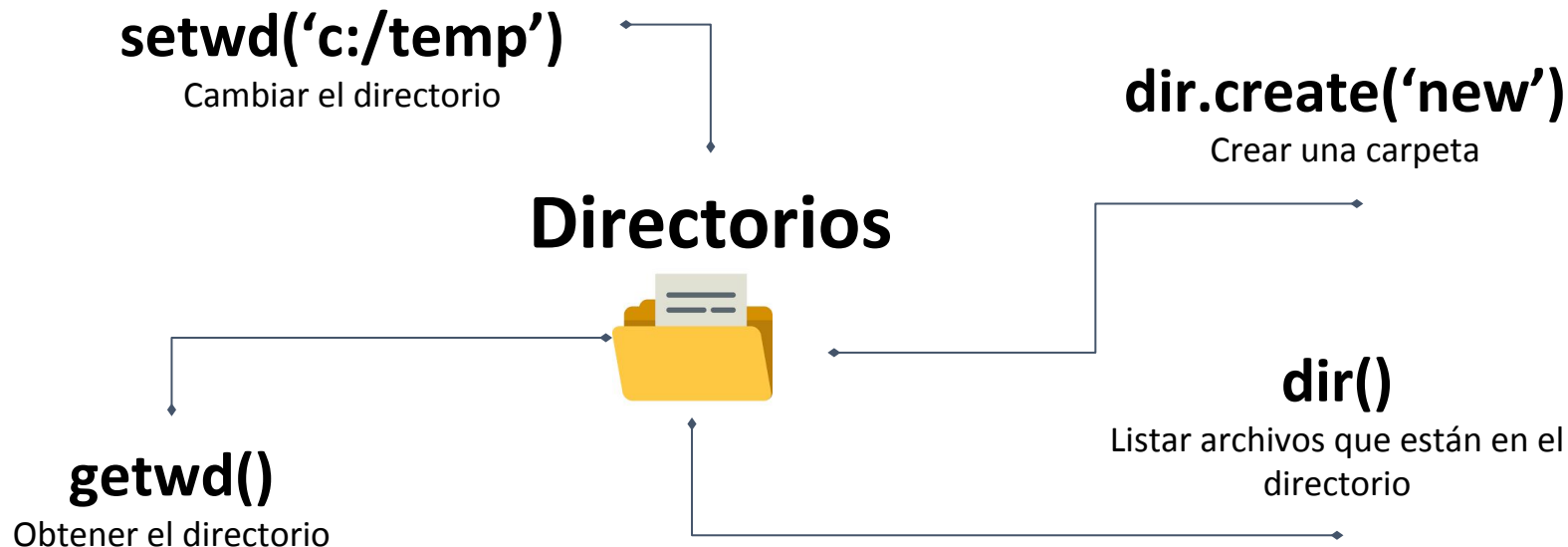
Nominal, Ordinal y Binario



https://twitter.com/allison_horst



Funciones de Directorios



Funciones de Directorios

EJERCICIO:

Crear un directorio que se tenga la siguiente estructura:

```
clase_yyyyMMddhhmmss
```

Donde:

- yyyy: es el año actual
- MM: es el mes actual
- dd: es el dia actual
- hh: es la hora actual
- mm: son los minutos actuales
- ss: son los segundos actuales



DataFrames

Un dataframe contiene las variables de un conjunto de datos como columnas y las observaciones como filas.

Los componentes son vectores de igual longitud, y la organización de los datos es similar a la de una tabla.

Ejemplo:

```
muestra_altura=c(1.85,1.50,1.63,1.90,1.75)
```

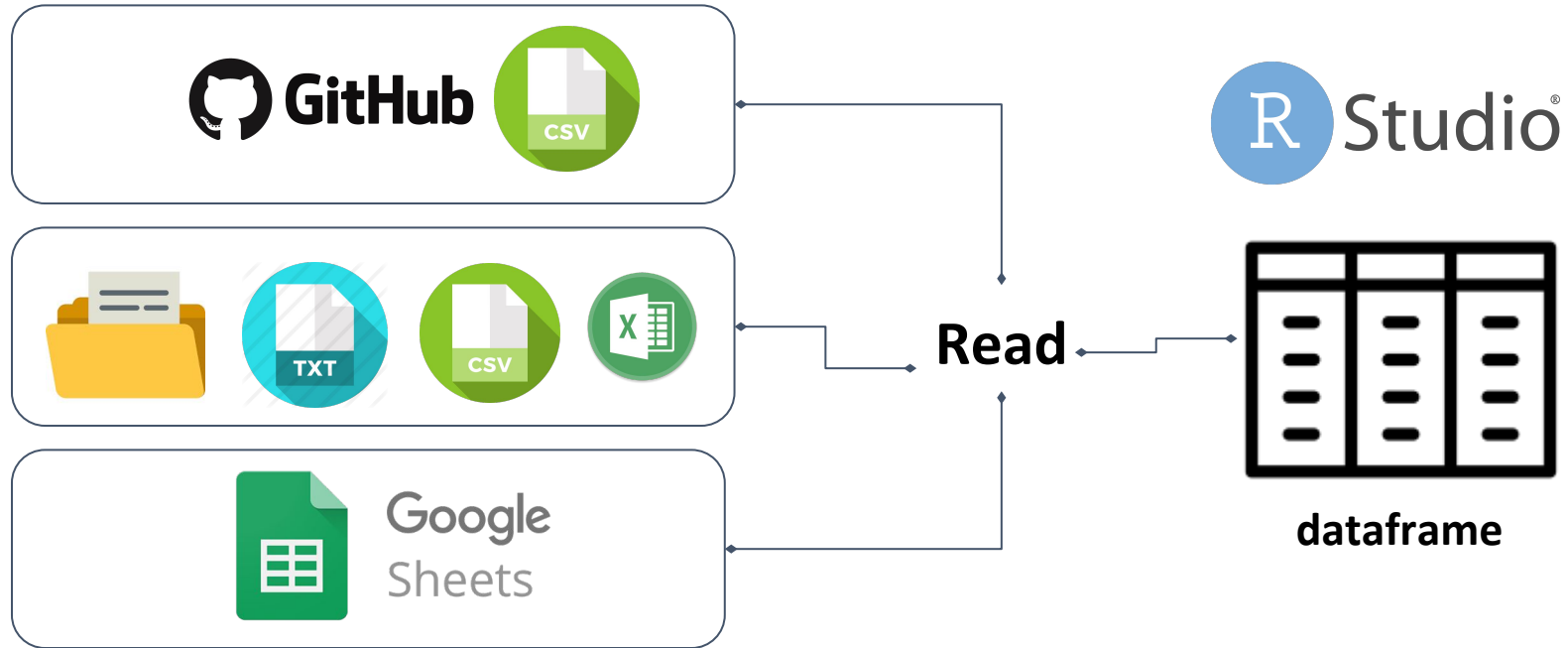
```
muestra_peso=c(100,50,65,94,70)
```

```
sexo=c('M','F','F','M','F')
```

```
muestra=data.frame(altura=muestra_altura,  
                    peso=muestra_peso,  
                    sexo)
```



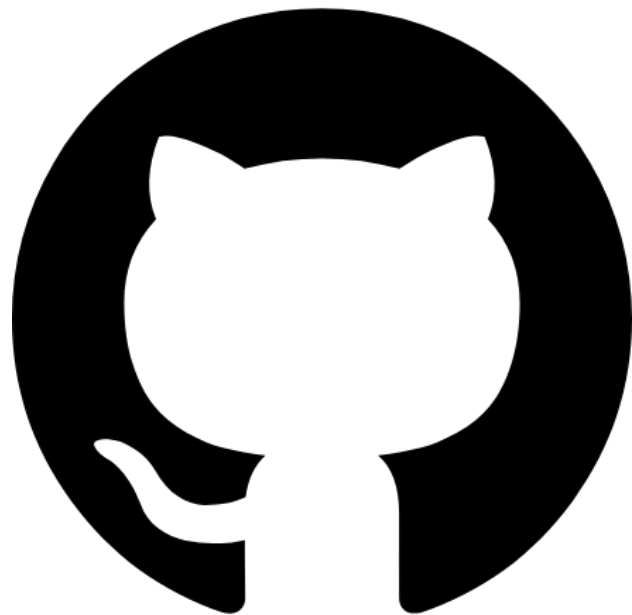
DataFrames



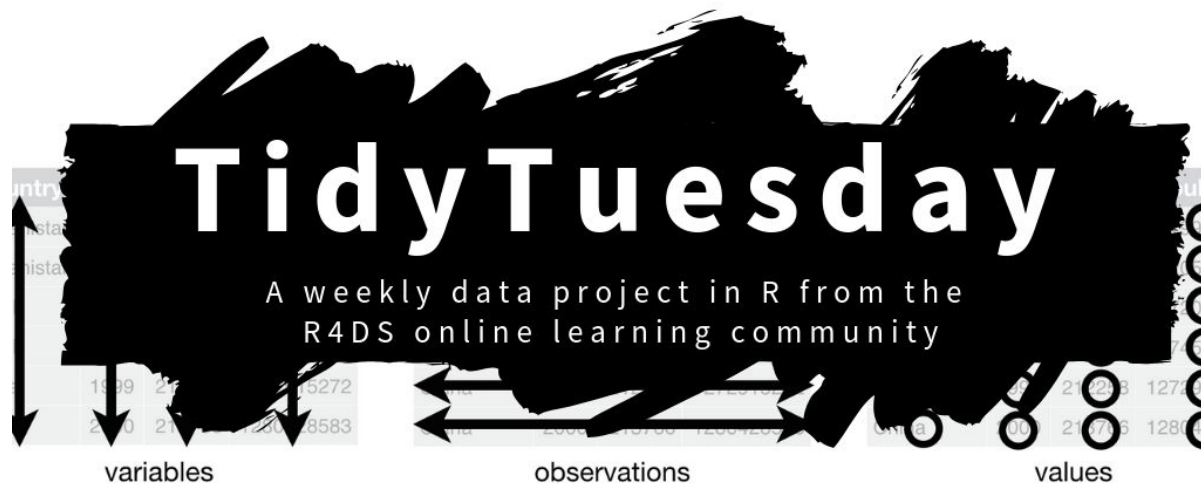
GITHUB

GitHub es un sitio web y un servicio en la nube que ayuda a los desarrolladores a almacenar y administrar su código, al igual que llevar un registro y control de cualquier cambio sobre este código.

Además permite el control de versiones, que es llevar un registro de todos los cambios en el código de un proyecto.



Importar un archivo desde github



Importar un archivo desde github



R para Ciencia de Datos @R4DS_es · 5 jun.

Esta semana en [#datosdemiercoles](#) exploraremos los datos que Airbnb liberó sobre algunas ciudades latinoamericanas. ¿Cómo ha evolucionado en ellas la oferta de este tipo de alojamiento?

 sobre el proyecto: bit.ly/datos-de-mierc...

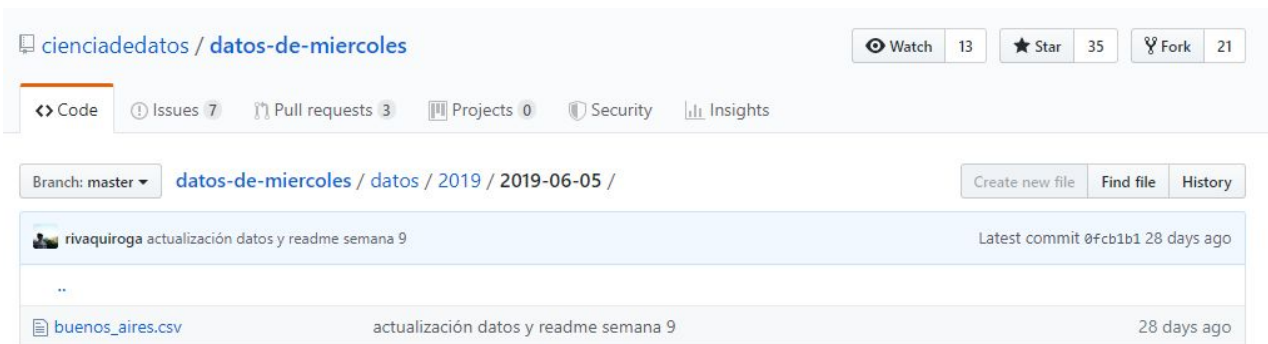
 los datos: bit.ly/datos-semana-9

[#rstatsES](#)



Importar un archivo desde github

```
buenos_aires <-  
read.csv("https://raw.githubusercontent.com/cienciad  
edatos/datos-de-miercoles/master/datos/2019/2019-06-  
05/buenos_aires.csv")
```



The screenshot shows the GitHub interface for the repository 'cienciadedatos / datos-de-miercoles'. At the top, there are buttons for 'Watch' (13), 'Star' (35), and 'Fork' (21). Below this is a navigation bar with 'Code' selected, and links for 'Issues' (7), 'Pull requests' (3), 'Projects' (0), 'Security', and 'Insights'. The main content area shows the file path 'datos-de-miercoles / datos / 2019 / 2019-06-05 /'. Below the path, there is a commit by 'rivaquioga' titled 'actualización datos y readme semana 9' with the latest commit hash '0fcb1b1' from 28 days ago. A file named 'buenos_aires.csv' is listed under this commit, also updated 28 days ago.



DataFrames Funciones

Summary es una función muy utilizada en r que permite obtener algunas medidas descriptivas

Summary de un dataframe:

```
> summary(muestra)
```

altura	peso	sexo
Min. : 1.500	Min. : 50.0	F:3
1st Qu. :1.630	1st Qu. : 65.0	M:2
Median :1.750	Median : 70.0	
Mean :1.726	Mean : 75.8	
3rd Qu. :1.850	3rd Qu. : 94.0	
Max. :1.900	Max. : 100.0	



DataFrames Funciones

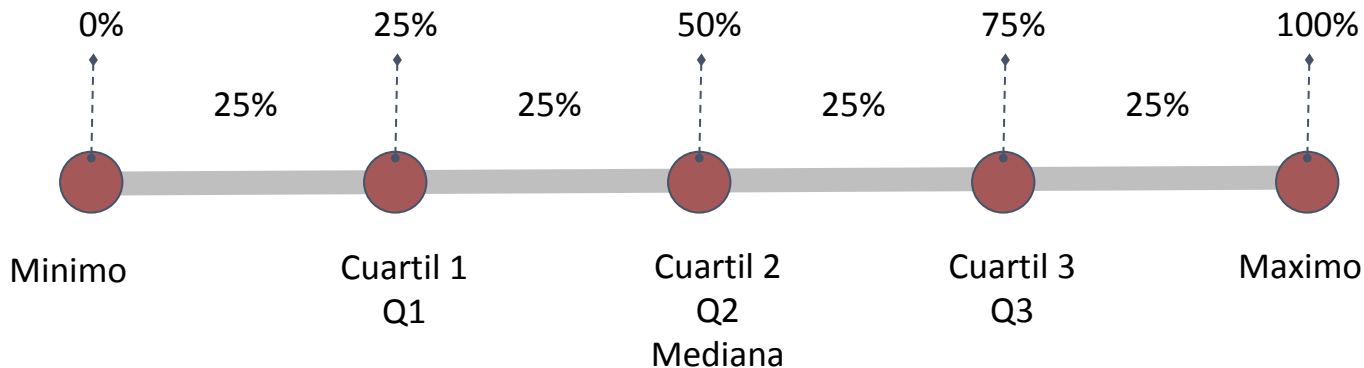
Se genera el siguiente dataframe que tiene un formato de tabla, con las columnas denominadas variables que todas tienen un tipo de dato y las filas que se denomina observaciones.

- **Número de Columnas:** `ncol(dataframe)`
- **Número de Filas:** `nrow(dataframe)`
- **Nombres de las columnas:** `names(dataframe)`



DataFrames Funciones

Cuartiles: Son valores que dividen una muestra de datos en cuatro partes iguales. Utilizando cuartiles puede evaluar rápidamente la dispersión y la tendencia central de un conjunto de dato



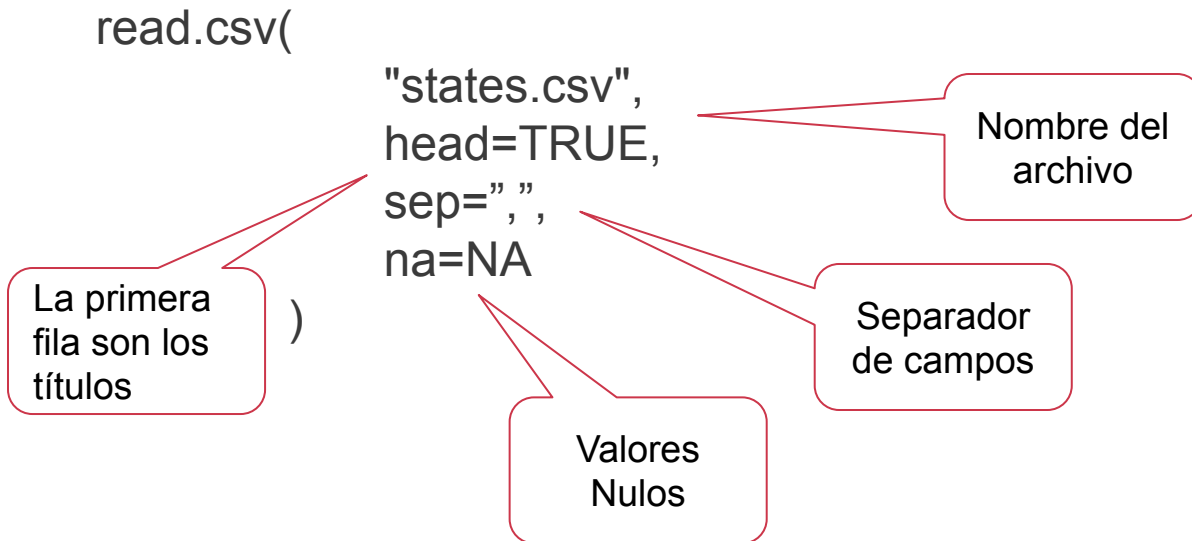
DataFrames Funciones

- 1º** Comprobar la importación de los datos
`head(data)`
- 2º** Analizar los tipos de datos importados
`str(data)`
- 3º** Obtener un análisis de cada variable
`summary(data)`
- 4º** Conocer los valores de variables categóricas
`unique(data$variable)`



Importar archivos csv y txt

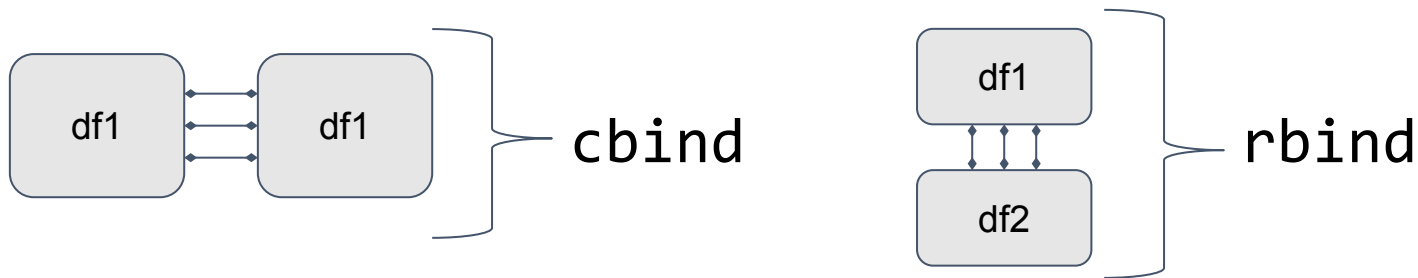
Los archivos se pueden importa utilizando la función `read.csv` o `read.table`



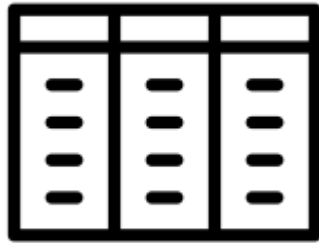
Funciones cbind y rbind

Las funciones cbind y rbind me permiten unir dos conjuntos de datos de la siguiente manera:

- cbind: Permite unir dos conjuntos de datos por columnas
- rbind: Permite unir dos conjuntos de datos por filas



DataFrames



dataframe

write



Funcion write.csv

Para exportar un dataframe a un archivo csv se utiliza la función write.csv

