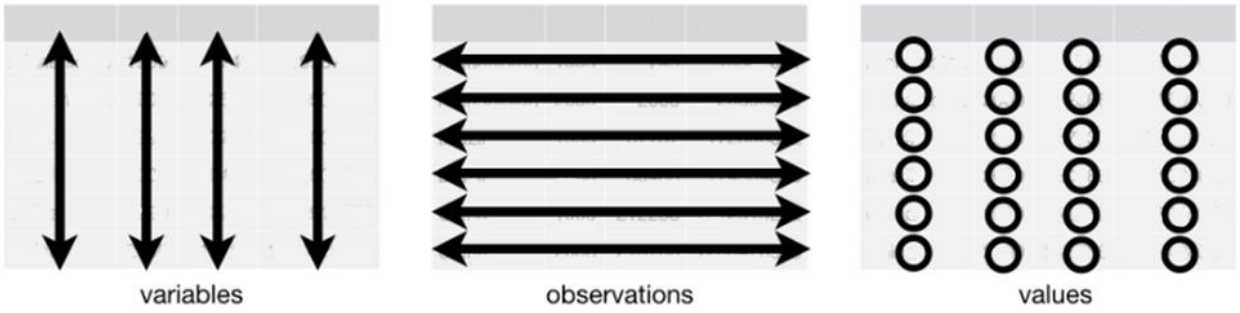
Conceptos Clase 5:

- Datos ordenados (tidy data)
- Tidyverse
    - Librería tidyr: gather(); spread(); separate(); unite()
    - Librería dplyr: filter(); arrange(); select(); mutate(); summarise() – group_by()
    - Librería magrittr: %>%

Resumen Clase 5:

Datos ordenados

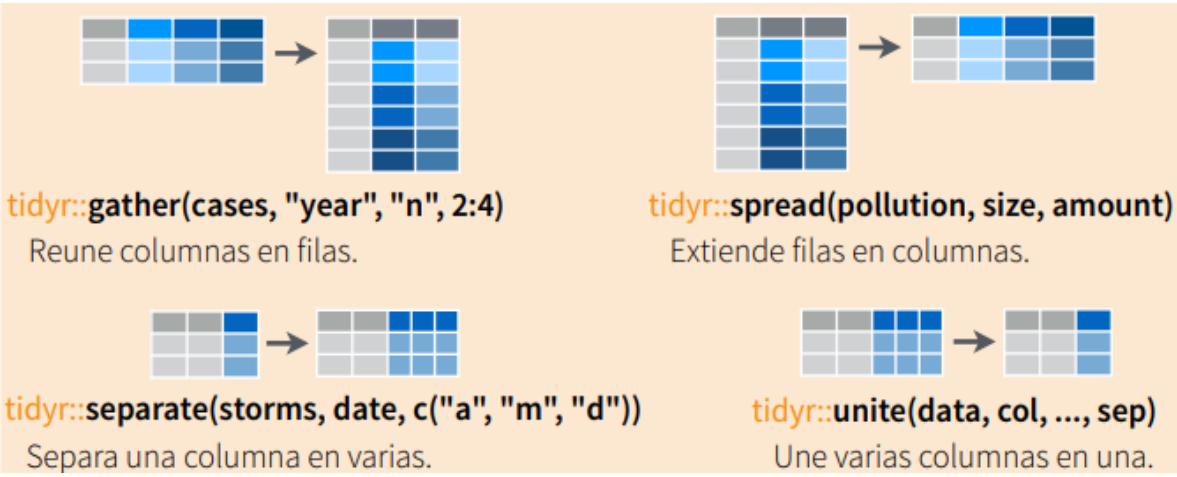
- Cada columna es una variable
- Cada fila es una observación
- Cada valor tiene su propia celda



Tidyverse (*https://www.tidyverse.org/*)

El ***tidyverse*** es una colección de paquetes R diseñados para la ciencia de datos. Todos los paquetes comparten un diseño, filosofía, gramática y estructuras de datos subyacente.

- tidyr



- dplyr

filter(.data, …**)** Extrae filas que cumplen criterios lógicos. También **filter_()**. *filter(iris, Sepal.Length > 7)*

arrange(.data, …**)** Ordena filas por valores de una columna (bajo a alto), usa con **desc()** para ordenar de alto a bajo.
arrange(mtcars, mpg)
arrange(mtcars, desc(mpg))

select(.data, …**)**
Selecciona columnas por nombre o funciones de ayuda. También **select_if()**
select(iris, Sepal.Length, Species)

mutate(.data, …**)**
Calcula columna(s) nueva(s).
mutate(mtcars, gpm = 1/mpg)

summarise(.data, …)
Calcula cuadro de resúmenes. También **summarise_**().
summarise(mtcars, avg = mean(mpg))

Usa **group_by()** para crear una copia "agrupada" de un cuadro. Funciones dplyr manipularán cada "grupo" por separado para luego combinar los resultados.

```
mtcars %>%
group_by(cyl) %>%
summarise(avg = mean(mpg))
```

- magrittr

dplyr::**%>%**

Pasa el objeto a la izquierda al primer argumento (o argumento .) de la función a la derecha creando un tubo.

`x %>% f(y)` *es lo mismo que* `f(x, y)`
`y %>% f(x, ., z)` *es lo mismo que* `f(x, y, z )`

El uso de tubos y tuberias (en Ingles: "Piping") con %>% resulta en código mas legible. Por ejemplo:

```
iris %>%
  group_by(Species) %>%
  summarise(avg = mean(Sepal.Width)) %>%
  arrange(avg)
```