

# Unidad 2.1.

Diego Carmona

January 14, 2019

# ¿Qué es una base de datos?

Es la colección de información estructurada que puede ser fácilmente consultada, manipulada, y actualizada

Las bases de datos estan organizadas en filas, columnas, y tablas y suele utilizarse la indexación para facilitar el acceso a la información relevante.

[row, column]

[1, 1]	[1, 2]		
[2, 1]			
			[9, 4]

[1, 1]	[1, 2]		
[2, 1]			
			[9, 4]

[1, 1]	[1, 2]		
[2, 1]			
			[9, 4]

[1, 1]	[1, 2]		
[2, 1]			
			[9, 4]

# Tipos de bases de datos

- Flat database
- Relational database
- Distributed database
- Cloud database
- NoSql database
- Object-oriented database (ej. Graph database)

# Tipos de bases de datos

- Flat database
- Relational database
- Distributed database
- Cloud database
- NoSql database
- Object-oriented database (ej. Graph database)

# Flat database

Base sencilla que guarda toda la información en una sola tabla o lista.

- Puede contener varios campos.
- Proclive a tener datos duplicados.
- Para juntar dos flat files necesitas copiar info de A a B.
- No hay automatización entre flat files. Un cambio en A no afecta a B.
- Ofrecen funcionalidad en el almacenaje, manipulación, y acceso de información.

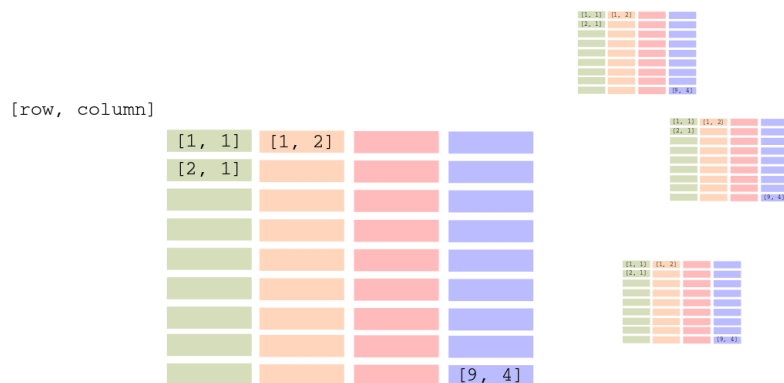
[row, column]

[1, 1]	[1, 2]		
[2, 1]			
			[9, 4]

# Relational database

Incorpora multiples tablas con métodos informáticos.

- Puede contener varios campos como las flat database.
- Resuelve el problema de tener datos duplicados por observación.
- La base A y B interactuan por lo que no necesitas copiar A en B.
- Puede haber automatización entre databases.
- Almacena, manipula, y accede a información MUCHO mas compleja y de manera eficiente.
- Al ingresar nueva información o modificar una base de datos se debe considerar la compatibilidad con las otras bases.

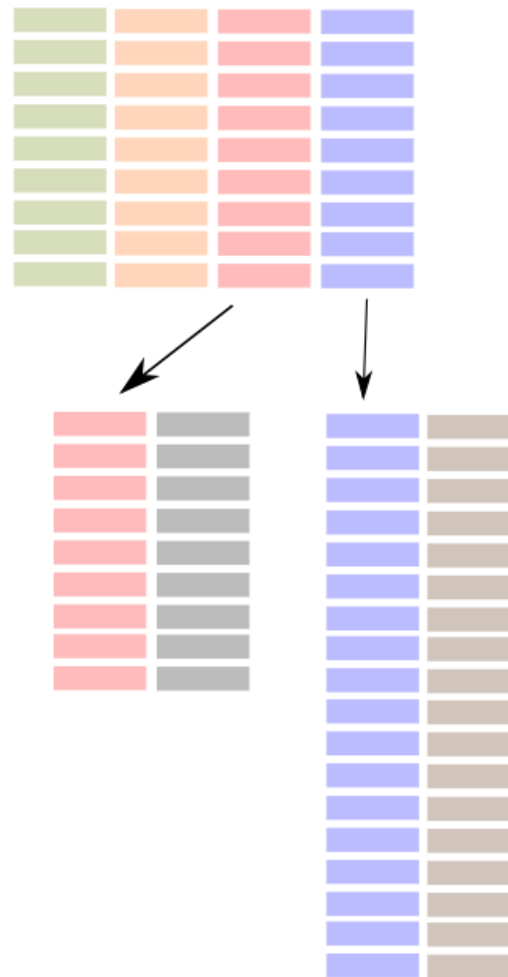


# Tipos de database relevantes para nosotros

## Flat database

[1, 1]	[1, 2]		
[2, 1]			
			[9, 4]

## Relational database

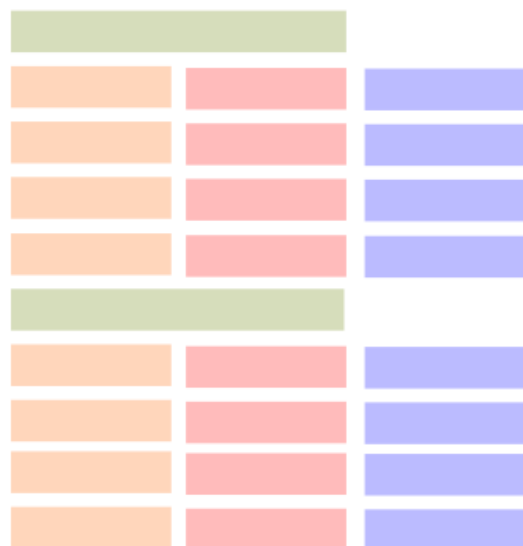


# Organización básica de datos.

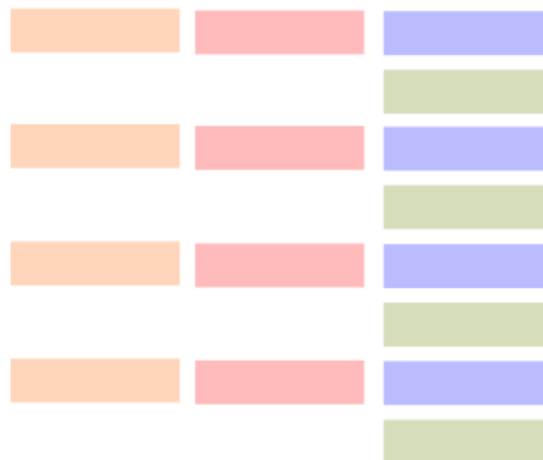
- Estructura general, objeto 2-d (i.e. filas y columnas)
- Estructura de bases de datos erróneas siguen diseños de tablas presentadas en publicaciones, o tratan de evitar redundancia de datos.

## Estructura errónea

Encabezados separados con subtablas para cada experimento



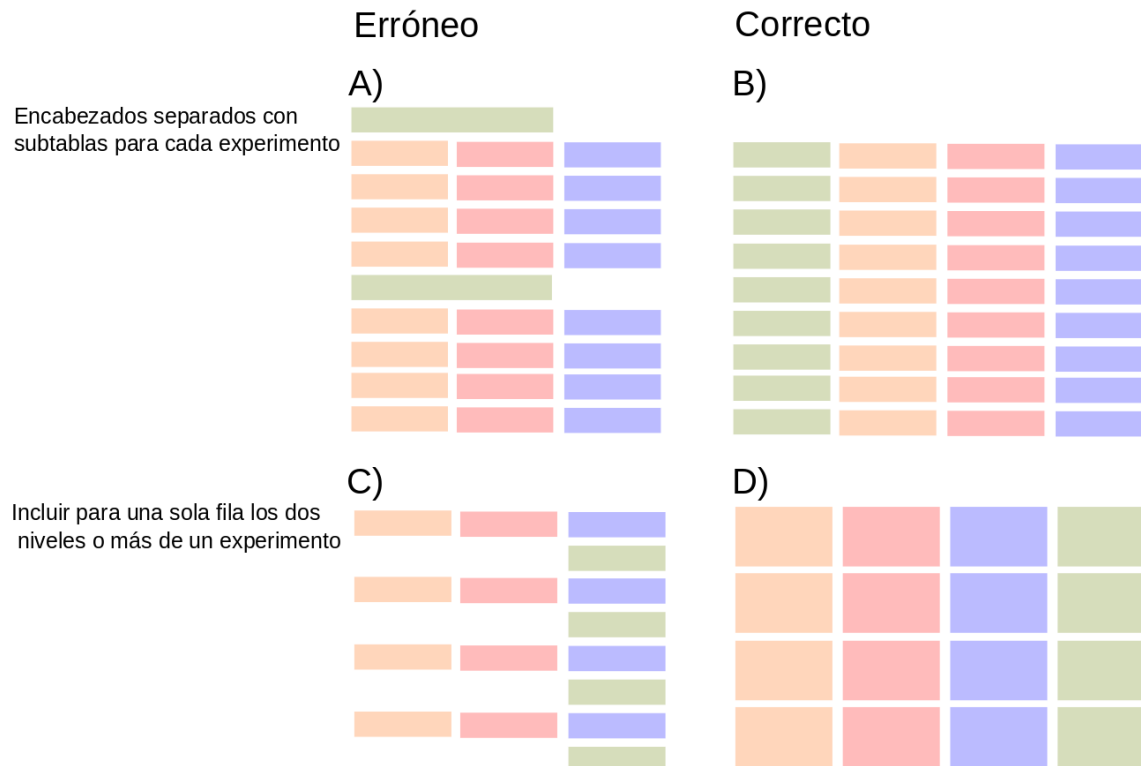
Incluir para una sola fila los dos niveles o más de un experimento





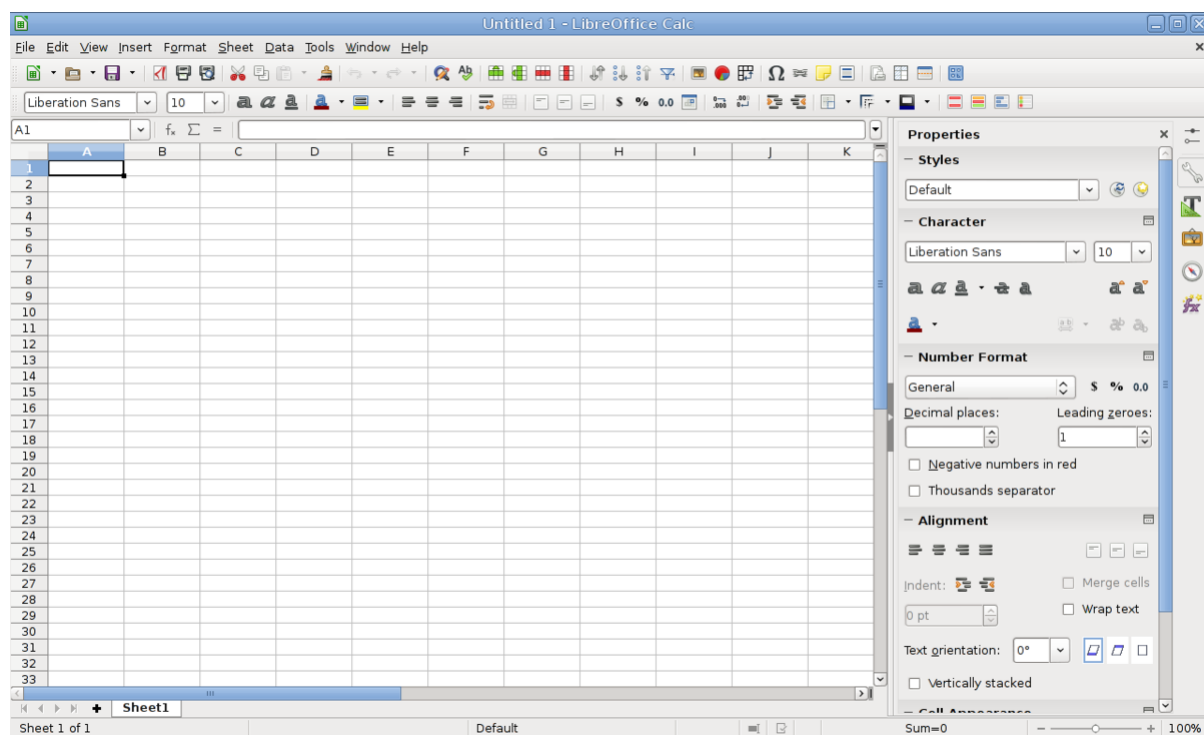
## El problema general surge cuando la información para una observación particular se encuentra en multiples filas.

Solución: la forma mas general y flexible de guardar datos en tablas es asegurándonos que cada fila tenga todo los datos necesarios para interpretar esa fila.



# Hojas de cálculo (Spreadsheets)

Falta de higiene en el uso de hojas de cálculo afecta análisis de datos.



# Recomendaciones

- Celdas deben contener valores crudos.
- No hagas cálculos o conversiones antes de meter los datos.
- Usa la hoja de cálculo para hacer estas conversiones pero manten tus datos lo menos procesados posible.
- No metas valores directos a las formulas, en cambio usa valores que estan asociados a las celdas.
- Esto ahorra tiempo y favorece reproducibilidad.
- **Al crear tu base de datos en la hoja de calculo hazlo pensando en R.**

# Hoja de Cálculo y R

# Slide with Plot

