

# **Estadística aplicada**

## **Gráficos y Visualización de datos**

**Clase 02**

**CONQUERBLOCKS**

# Temario de la clase

- La importancia de la visualización de datos
- Tipos de gráficos más comunes
- Ejemplos prácticos en Seaborn
- Buenas prácticas de visualización

# Importancia de la representación gráfica

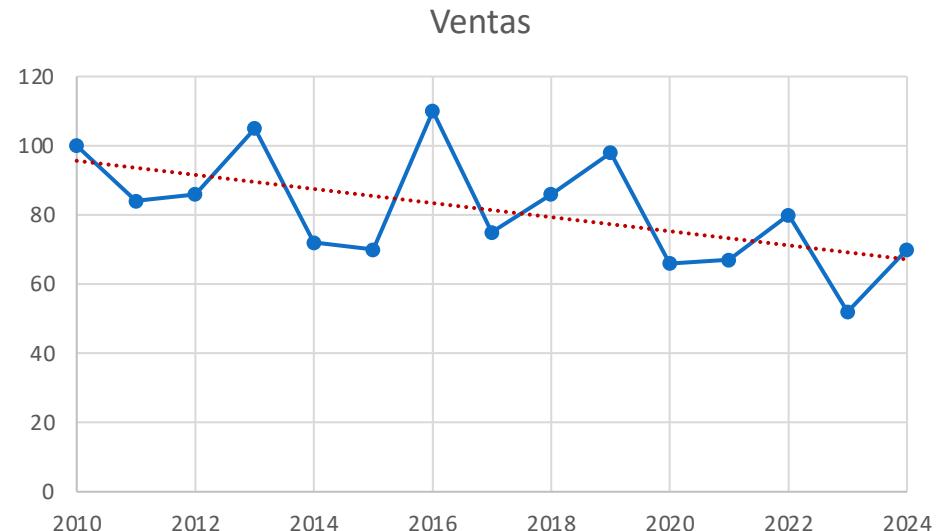
## ¿Por qué es importante aprender a visualizar datos?

La visualización de datos juega un papel crucial tanto en el análisis de datos como en la inteligencia artificial (IA) por varias razones:

1. La visualización ayuda a explorar los datos y entender sus características principales. Mediante gráficos, es posible identificar patrones, tendencias y relaciones entre variables.
2. Los gráficos facilitan la identificación de valores atípicos o anomalías que pueden afectar los resultados de un análisis.

## Evolución de las ventas de una empresa

Año	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Ventas	100	84	86	105	72	70	110	75	86	98	66	67	80	52	70



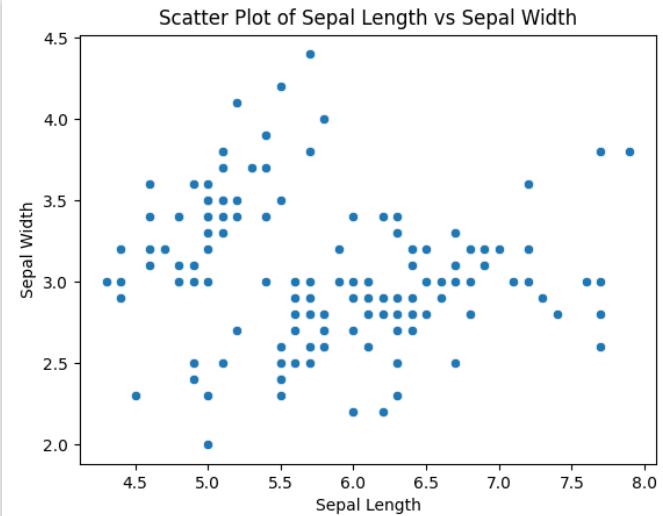
# Representación gráfica de datos: Seaborn

## ¿Qué es seaborn?

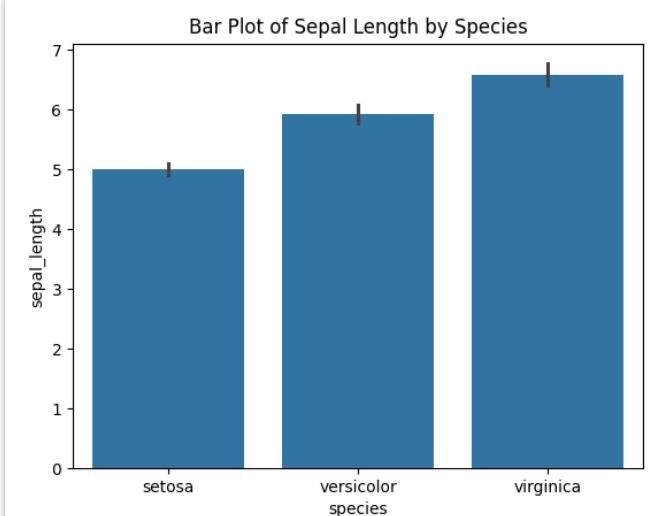
Seaborn es una **librería de visualización de datos en Python** que se basa en Matplotlib. Proporciona una interfaz de alto nivel para dibujar **gráficos estadísticos atractivos y con estilo**.

FÁCIL Y RÁPIDA DE USAR

# Tipos de gráficos útiles para análisis de datos



Gráficos de dispersión



Gráficos de barras

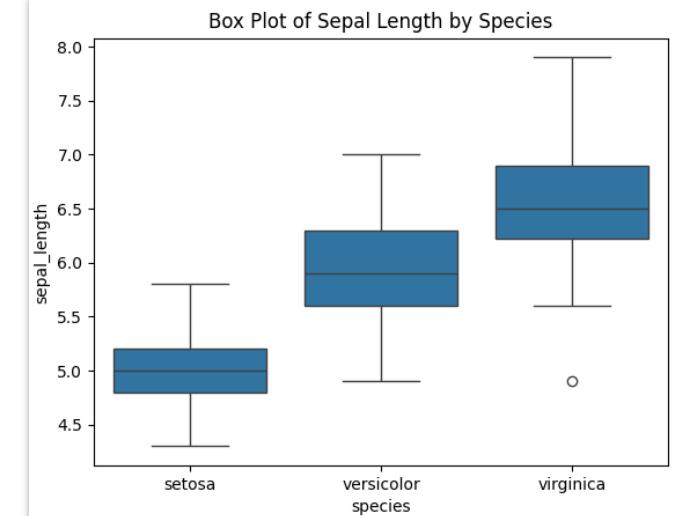
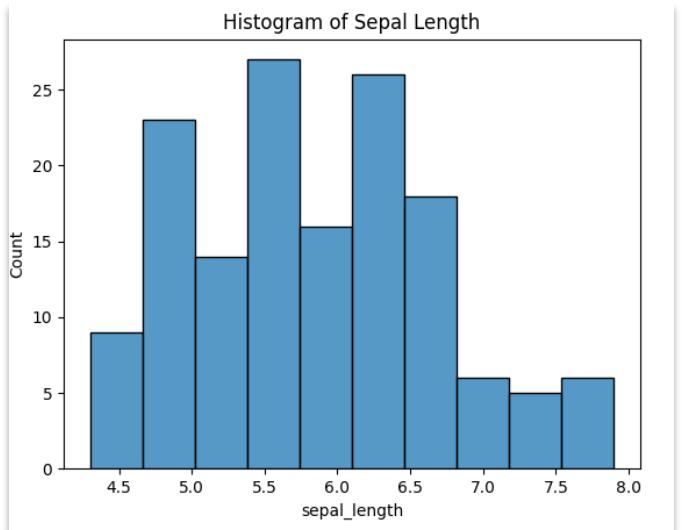


Gráfico de cajas



Histograma

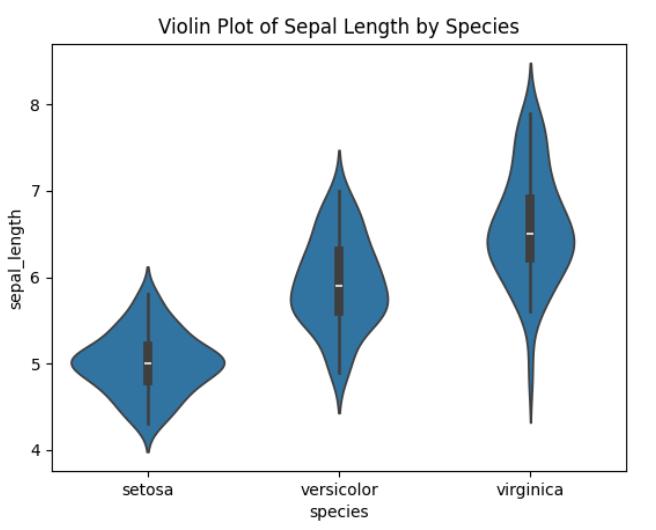
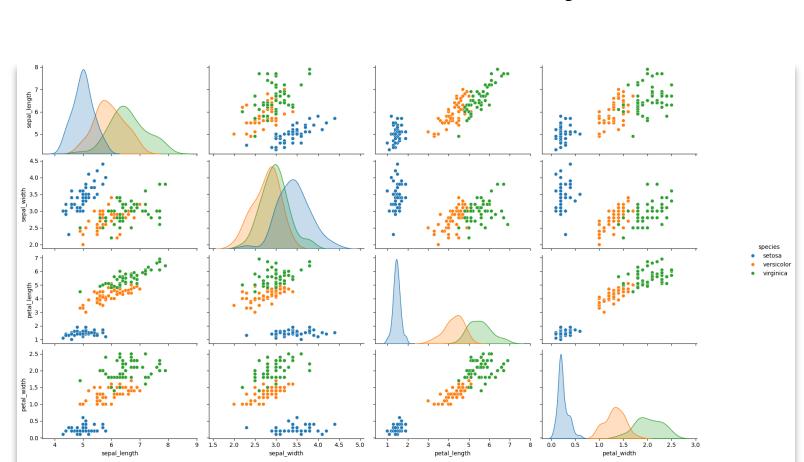


Gráfico de violín  
**CÓNQUER BLOCKS**



Mapas de calor de correlaciones

# Explicación del dataset

## Data set Iris

El conjunto de datos "Iris" es uno de los datasets más conocidos y utilizados en el ámbito de la estadística y el aprendizaje automático.

### ¿Qué contiene?

Contiene 150 observaciones de **iris** (un tipo de flor) de tres especies diferentes: **Iris setosa**, **Iris versicolor** e **Iris virginica**. Cada observación incluye cuatro características medidas de cada flor y la especie a la que pertenece.

### Columnas del dataset

**Sepal Length (cm)**: Longitud del sépalo en centímetros.

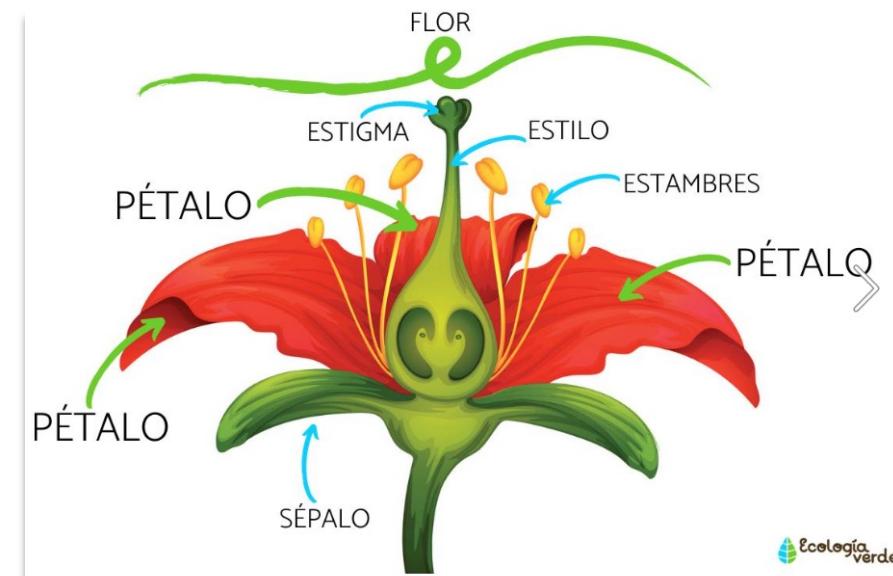
**Sepal Width (cm)**: Ancho del sépalo en centímetros.

**Petal Length (cm)**: Longitud del pétalo en centímetros.

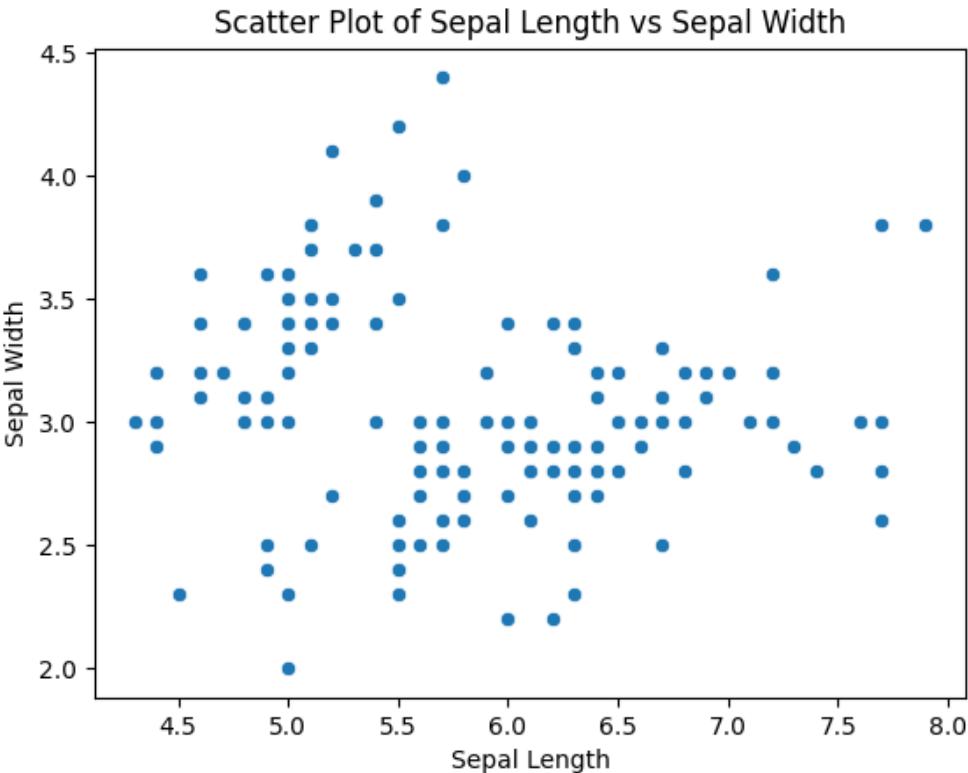
**Petal Width (cm)**: Ancho del pétalo en centímetros.

**Species**: Especie de la flor, que puede ser una de las tres:

- setosa
- versicolor
- virginica



# Gráficos de dispersión



## ¿Qué representan?

Muestra la relación entre dos variables continuas mediante puntos en un plano cartesiano, X Y

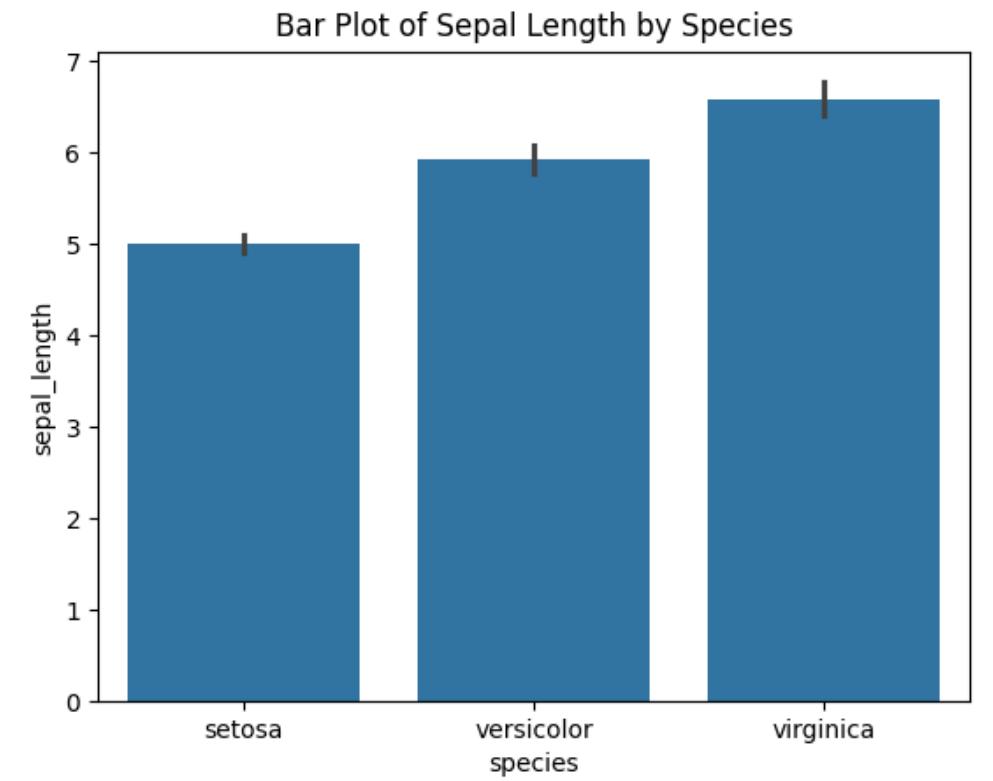
## ¿Para qué se usan?

Se usa para identificar y analizar relaciones, patrones y tendencias entre dos variables.

## ¿Qué conclusiones se pueden sacar?

Se pueden identificar correlaciones, patrones, tendencias y detectar valores atípicos.

# Gráficos de Barras



## ¿Qué representan?

Representa datos categóricos con barras rectangulares, donde la longitud de cada barra es proporcional al valor que representa.

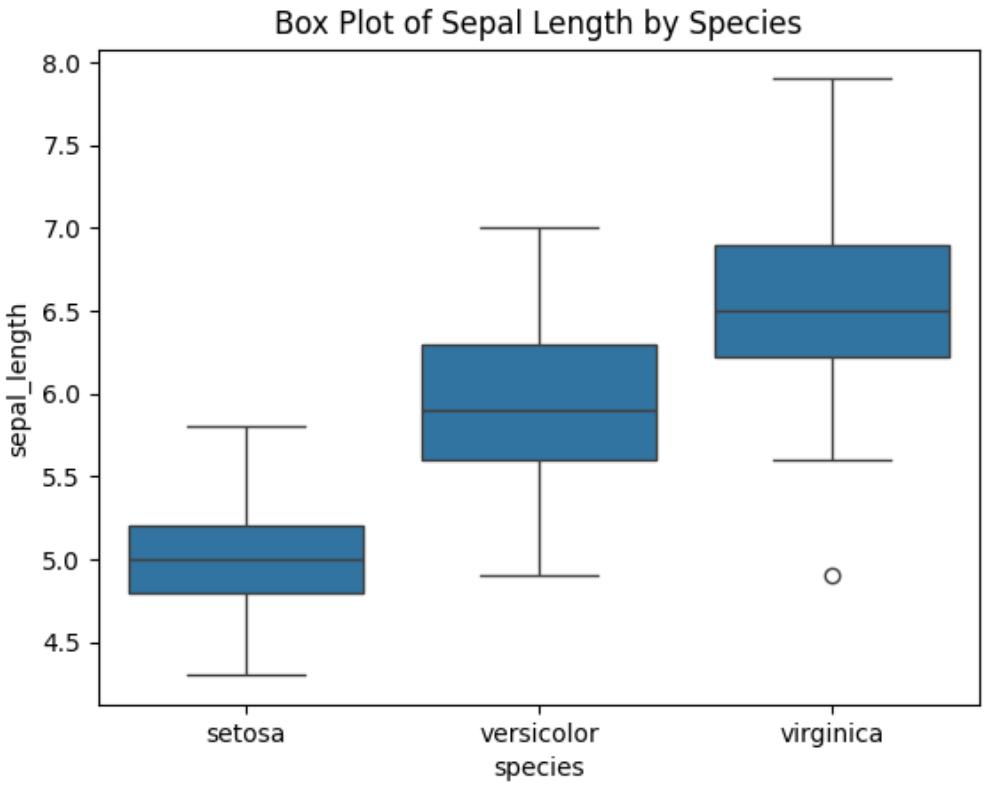
## ¿Para qué se usan?

Se usa para comparar diferentes categorías entre sí.

## ¿Qué conclusiones se pueden sacar?

Permite ver fácilmente las diferencias y comparaciones entre categorías, así como identificar las categorías con valores más altos o más bajos.

# Gráficos de Cajas



## ¿Qué representan?

Representa la distribución de una variable continua mostrando la mediana, los cuartiles y los valores atípicos.

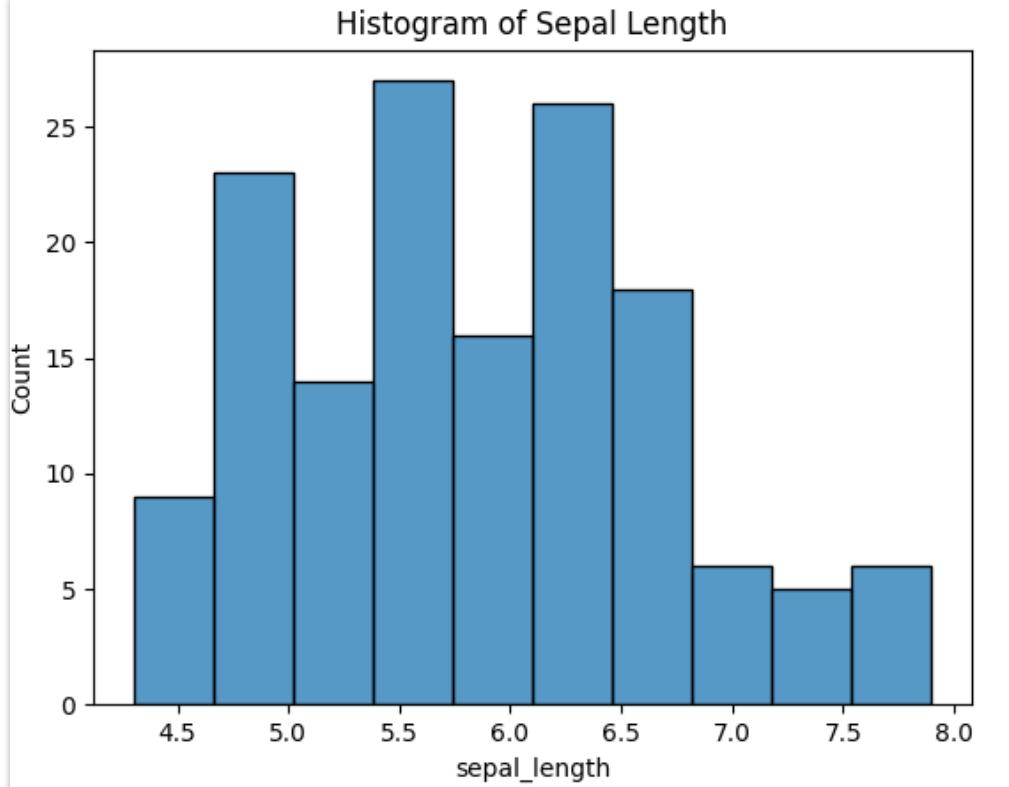
## ¿Para qué se usan?

Se usa para resumir la distribución de datos y detectar valores atípicos.

## ¿Qué conclusiones se pueden sacar?

Se puede ver la mediana, la dispersión, la presencia de valores atípicos, y comparar la distribución entre diferentes grupos.

# Gráficos de Histograma



## ¿Qué representan?

Muestra la distribución de una variable continua dividiendo el rango de valores en intervalos y contando la frecuencia de los valores en cada intervalo.

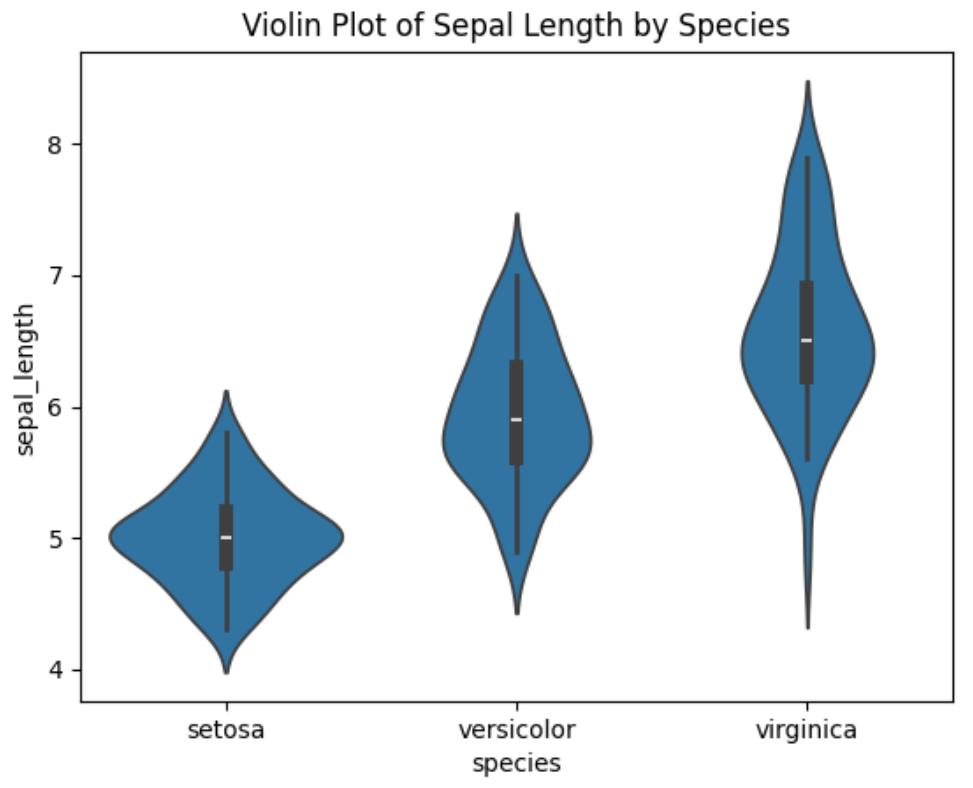
## ¿Para qué se usan?

Se usa para entender la distribución y frecuencia de una variable continua

## ¿Qué conclusiones se pueden sacar?

Se pueden identificar patrones de distribución, como la simetría, la asimetría, la presencia de múltiples picos (multimodalidad) y detectar valores atípicos.

# Gráficos de Violín



## ¿Qué representan?

Combina el box plot con un gráfico de densidad para mostrar la distribución de una variable continua.

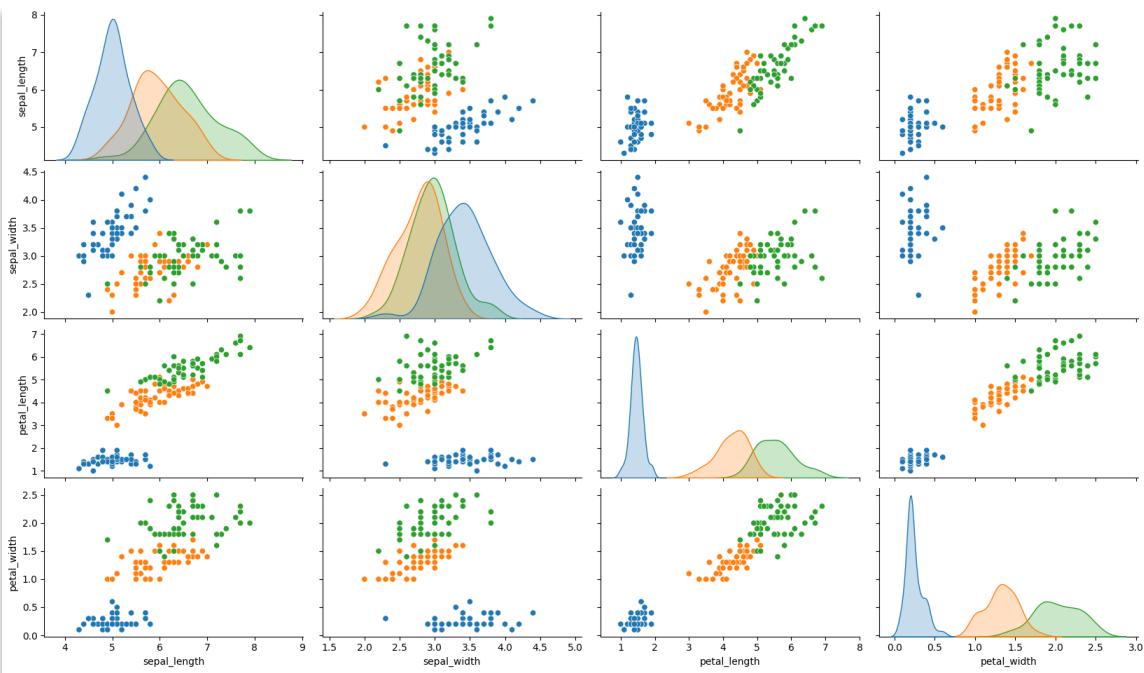
## ¿Para qué se usan?

Se usa para visualizar la distribución y la densidad de una variable, comparando varias categorías o grupos.

## ¿Qué conclusiones se pueden sacar?

Proporciona una visión más detallada de la distribución de los datos, permitiendo comparar la forma de la distribución entre diferentes grupos o categorías.

# Gráficos de Correlación (Heatmap)



## ¿Qué representan?

Muestra la relación entre múltiples variables en una matriz, donde los valores son representados con colores.

## ¿Para qué se usan?

Se usa para identificar y visualizar relaciones y patrones de correlación entre múltiples variables.

## ¿Qué conclusiones se pueden sacar?

Se puede ver qué variables están altamente correlacionadas positiva o negativamente, lo que es útil para la selección de características y la identificación de relaciones entre variables.

# Explicación del dataset

## Data set Iris

Contiene 150 observaciones de **iris** (un tipo de flor) de tres especies diferentes: **Iris setosa**, **Iris versicolor** e **Iris virginica**. Cada observación incluye cuatro características medidas de cada flor y la especie a la que pertenece.

## Columnas del dataset

**Sepal Length (cm)**: Longitud del sépalo en centímetros.

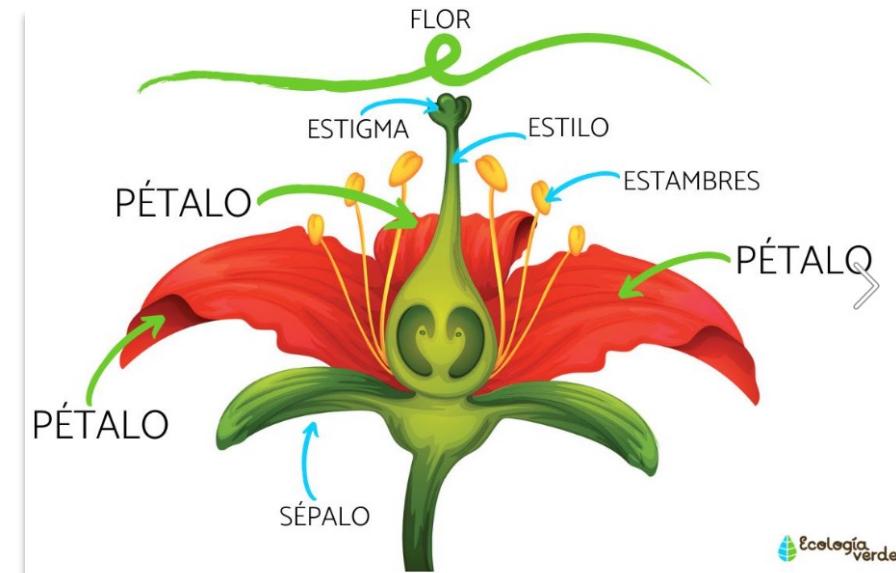
**Sepal Width (cm)**: Ancho del sépalo en centímetros.

**Petal Length (cm)**: Longitud del pétalo en centímetros.

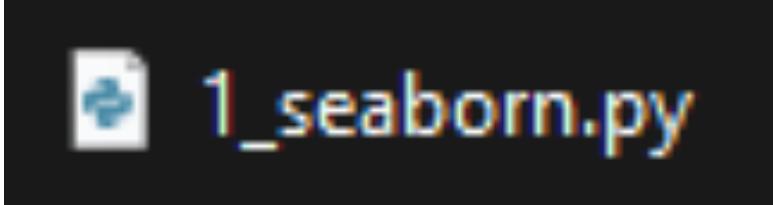
**Petal Width (cm)**: Ancho del pétalo en centímetros.

**Species**: Especie de la flor, que puede ser una de las tres:

- setosa
- versicolor
- virginica



# Programación de las visualizaciones



# Principios de un buen gráfico

La regla de oro

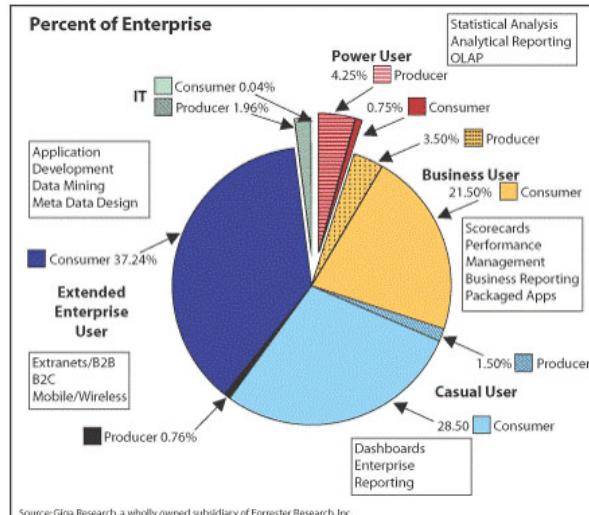


1. Claridad y Simplicidad
2. Escalas Apropriadas
3. Colores y Etiquetas
4. Narrativa y Contexto

# Principios de un buen gráfico

## 1.-Claridad y simplicidad

- **Evitar la Sobrecarga:** Los gráficos deben ser claros y no deben incluir demasiada información. Evita el uso excesivo de elementos visuales como colores, líneas o símbolos que puedan distraer al espectador.
- **Título Descriptivo:** Cada gráfico debe tener un título claro y descriptivo que explique qué se está mostrando.
- **Leyendas y Etiquetas:** Usa leyendas y etiquetas de manera adecuada para que cada elemento del gráfico sea comprensible. Asegúrate de que las etiquetas sean legibles y estén colocadas de manera que no confundan al espectador.

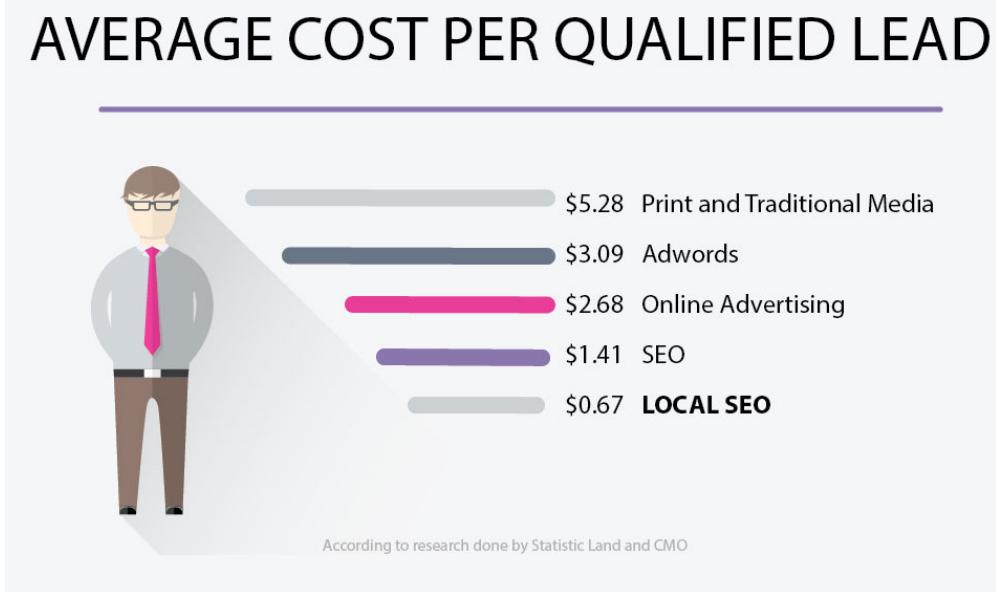


Ejemplo de un gráfico sobrecargado

# Principios de un buen gráfico

## 2.-Escala apropiada

- **Ejes Proporcionados:** Los ejes deben estar escalados correctamente para no distorsionar la interpretación de los datos. Por ejemplo, si muestras un crecimiento de ventas en un gráfico de líneas, asegúrate de que el eje Y no esté comprimido para exagerar las diferencias.
- **Intervalos Consistentes:** Usa intervalos consistentes en los ejes para facilitar la comparación. Evita los cambios bruscos de escala dentro del mismo gráfico.

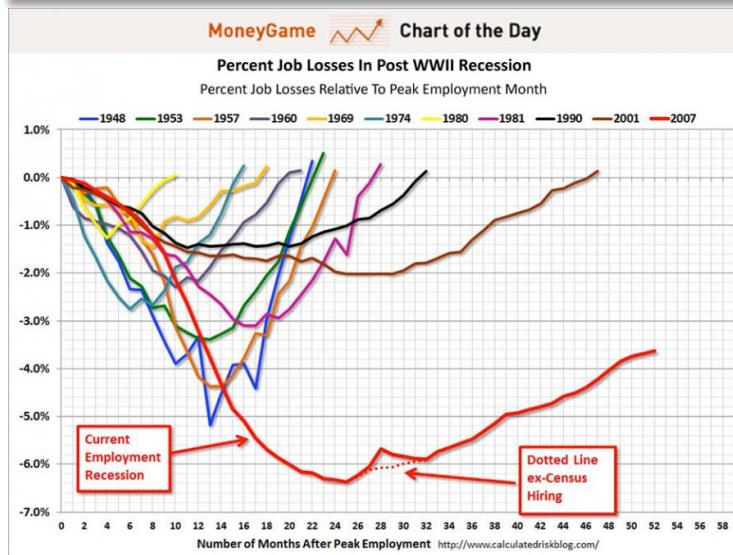


Ejemplo de escalas inadecuadas

# Principios de un buen gráfico

## 3.-Colores y Etiquetas

- **Uso Estratégico del Color:** Utiliza colores para diferenciar categorías o destacar puntos importantes, pero sin exagerar. Demasiados colores pueden confundir al espectador.
- **Contraste y Legibilidad:** Asegúrate de que haya suficiente contraste entre los colores utilizados para que todos los elementos sean fácilmente distinguibles.
- **Etiquetas Claras:** Las etiquetas deben ser precisas y proporcionar la información necesaria sin ser demasiado extensas. Incluir etiquetas directamente en el gráfico en lugar de una leyenda separada puede ayudar a la claridad.



Ejemplo de un gráfico sobrecargado, con muchos colores y difícilmente leíble

# Principios de un buen gráfico

## 4.- Narrativa y contexto

- Contexto Adicional: Proporciona el contexto necesario para entender el gráfico. Esto puede incluir notas, referencias o un breve resumen de lo que muestra el gráfico.
- Historias con Datos: Los gráficos deben contar una historia clara. Ayuda a tus espectadores a comprender el significado detrás de los datos, no solo presentándolos sino interpretándolos.

# Limitaciones y precauciones al usar gráficos

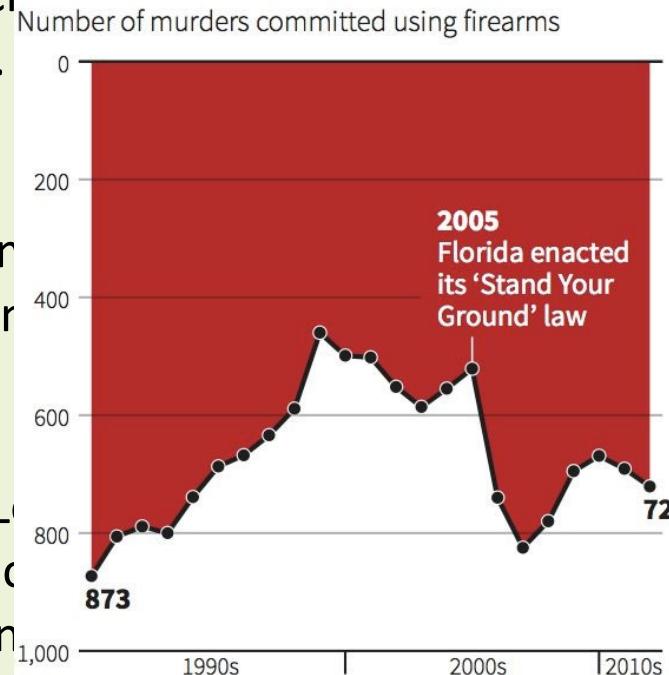


Los gráficos son herramientas muy potentes para explicar las diferentes realidades. Sin embargo, también estos tienen limitaciones.

## Sesgo y manipulación

- **Selección de Datos:** Selección de datos que llevan a interpretaciones engañosas. ¿Por qué?
- **Manipulación del Eje Y:** Cambios aparentes en los datos. Siempre comienza en cero cuando sea posible.
- **Gráficos 3D y Perspectiva:** Los gráficos 3D y perspectiva distorsionan la interpretación. Los gráficos 3D son difíciles de leer con precisión.

### Gun deaths in Florida



Source: Florida Department of Law Enforcement

C. Chan 16/02/2014

Incluir en un gráfico puede llevar a confusiones sobre qué datos se están incluyendo.

Exagerar o minimizar las diferencias entre los datos. Siempre considera empezar los ejes en cero.

Los gráficos 3D y perspectiva suelen ser visualmente atractivos pero a menudo pueden hacer que los datos sean difíciles de leer con precisión.

# Limitaciones y precauciones al usar gráficos



Los gráficos son herramientas muy potentes para explicar las diferentes realidades. Sin embargo, también estos tienen limitaciones.

## Precisión

- **Muestreo Insuficiente:** Si los datos representados en un gráfico provienen de un muestreo insuficiente, las conclusiones pueden no ser fiables. Asegúrate de que los datos sean representativos.
- **Agregación de Datos:** Los gráficos suelen presentar promedios sin desglosarlos, lo que puede ocultar datos importantes. Mostrar todos los datos es una buena práctica para garantizar la integridad de los datos.
- **Errores de Muestreo:** Los errores de muestreo se refieren a los errores de muestreo y representación. Es importante estar consciente de las limitaciones de los datos utilizados en los gráficos.



# Limitaciones y precauciones al usar gráficos



Los gráficos son herramientas muy potentes para explicar las diferentes realidades. Sin embargo, también estos tienen limitaciones.

## Contexto y Comparaciones Inadecuadas

- **Falta de Contexto:** Los gráficos que no proporcionan suficiente contexto pueden llevar a interpretaciones erróneas. Asegúrate de que los espectadores comprendan de dónde vienen los datos y en qué contexto deben ser interpretados.
- **Comparaciones Inadecuadas:** Comparar datos que no son comparables puede llevar a conclusiones erróneas. Asegúrate de que las comparaciones sean válidas y estén bien fundamentadas.