

Clase 6: Visualización de información

FM849 - Programación Científica para Proyectos de Inteligencia Artificial (IA)

7 de enero de 2026

Visualización de información

- ▶ La información puede almacenarse en diferentes formatos: **grafos, arboles, tablas, diccionarios, entre otros.**
- ▶ En este curso nos enfocaremos en datos tabulares (tablas). Este es el formato mas simple y utilizado.
- ▶ En la clase anterior vimos como abrir un dataset tabular usando Pandas. Ahora, la idea es visualizar estos datos.

Pero antes de eso, por que nos gustaria visualizar los datos ?

- ▶ Como vimos en la clase de estadística, los datasets pueden ser bastante complejos para analizar fila por fila.
- ▶ Las tecnicas de visualización nos permiten encontrar patrones o tendencias en los datos.
- ▶ La visualización con gráficos nos permite responder a preguntas asociadas a los datos.

Matplotlib vs Seaborn

En esta clase trabajaremos con dos paquetes para visualizar datos.

- ▶ **Matplotlib:** Librería de uso general para construir todo tipo de gráficos.
 - ▶ Es muy flexible, pero requiere más código.
 - ▶ Recordar importar: `import matplotlib.pyplot as plt`
 - ▶ Las funciones de matplotlib reciben los vectores de datos a plotear como argumentos (no un `DataFrame`).
- ▶ **Seaborn:** Librería orientada al análisis de datos.
 - ▶ Está pensada para trabajar directamente con `DataFrames`.
 - ▶ Recordar importar: `import seaborn as sns`

Matplotlib vs Seaborn

En Matplotlib, los gráficos se construyen indicando:

- ▶ Directamente los **datos a graficar** (arreglos, listas o Series).
- ▶ No existe un argumento data que agrupe los datos.
- ▶ Los argumentos (x, y, etc.) corresponden a los **arreglos de valores** y no a los nombres de las columnas.

En Seaborn, los gráficos se construyen indicando:

- ▶ data: el **DataFrame** que contiene los datos.
- ▶ El resto de los argumentos (x, y, hue, etc.) son los **nombres de las columnas** de ese DataFrame que se quieren graficar.

Esto permite crear gráficos de forma simple, clara y eficiente, sin necesidad de extraer manualmente los datos como ocurre habitualmente en Matplotlib.

Gráficos de Línea

- ▶ Los ejes x e y representan los valores que adoptan las variables analizadas.
- ▶ Se usan para analizar tendencias temporales y comparación la evolución de variables.
- ▶ Las marcas son puntos que se conectan por líneas.
- ▶ La posición vertical expresa un valor cuantitativo, mientras la posición horizontal contiene las llaves ordenadas.

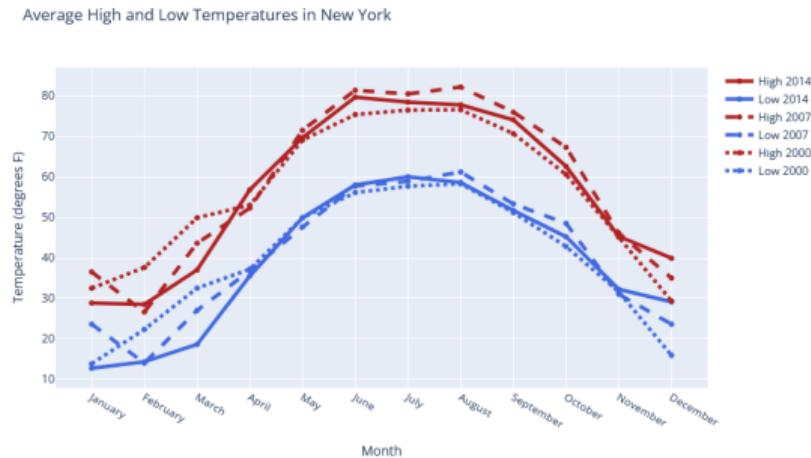


Figura 1: Gráfico de línea mostrando la evolución de la temperatura un año.

Ejemplo de gráfico de línea (Matplotlib)

Con plt.plot:

```
import matplotlib.pyplot as plt
>>> x = [0, 1, 2, 3, 4]
>>> y = [1, 3, 2, 5, 4]
>>> # plot (0,1), (1,3), ... (4,4)
>>> plt.plot(x,y)
>>> plt.ylabel("y")
>>> plt.show()
```

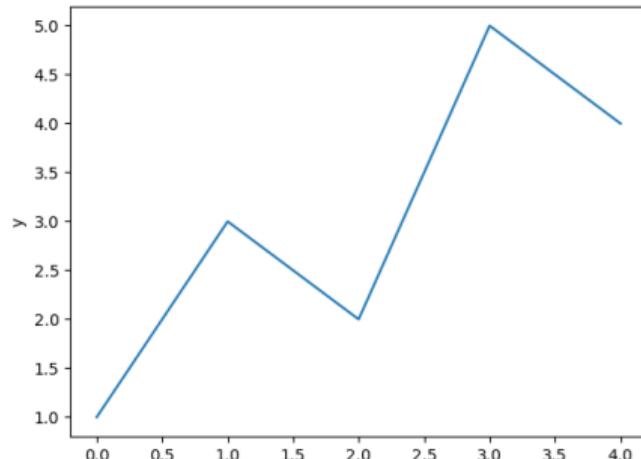


Figura 2: Gráfico de línea en matplotlib. Muestra los valores de la lista x (eje x) y lista y (eje y).

Mas ejemplos: https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.plot.html

Ejemplo de gráfico de línea (Seaborn)

Con sns.lineplot:

```
>>> import seaborn as sns  
>>> # DataFrame de vuelos  
>>> df = sns.load_dataset("flights")  
>>> sns.lineplot(data=df,  
                 x="year",  
                 y="passengers")  
>>> plt.show()
```

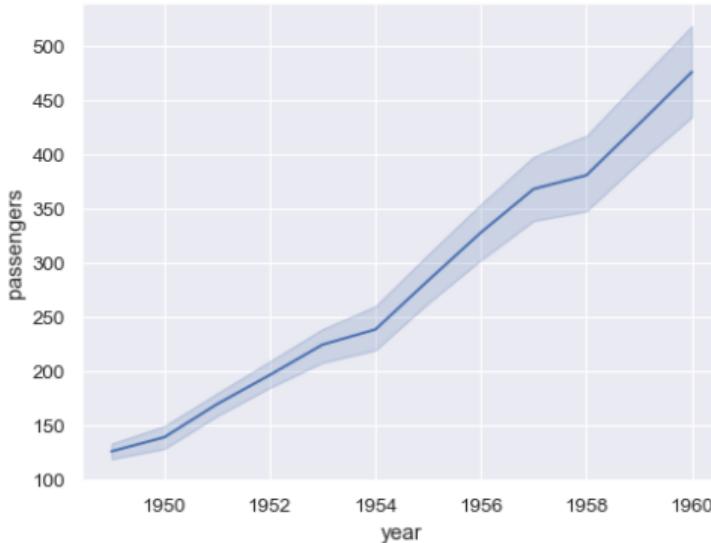


Figura 3: Gráfico de línea que muestra el numero de vuelos (eje y) según el año (eje x). El área sombreada representa un intervalo de confianza.

Mas ejemplos: <https://seaborn.pydata.org/generated/seaborn.lineplot.html>

Gráfico de Dispersion

- ▶ El gráfico de dispersión muestra la relación entre dos variables numéricas.
- ▶ Permite detectar correlaciones, identificar patrones, clusteres y encontrar outliers.
- ▶ Los canales son las posiciones: horizontal y vertical. Los ejes x e y representan los valores que adoptan las variables analizadas.

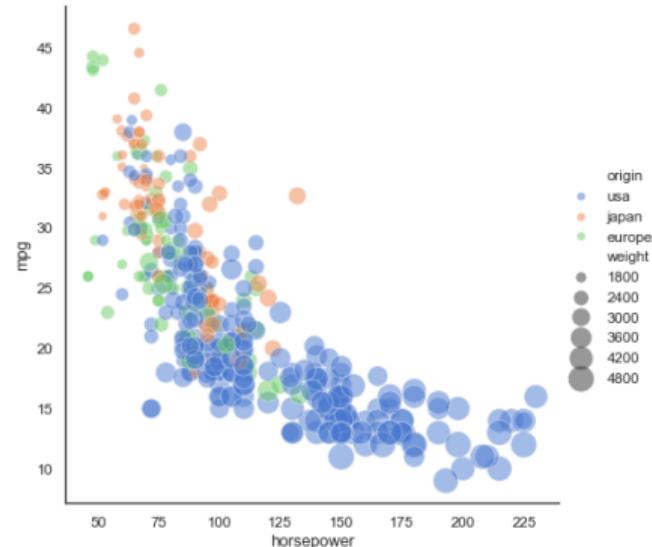


Figura 4: Gráfico de dispersión entre millas por galon (MPG, eje y) y caballos de fuerza (horsepower, eje x) en autos.

Ejemplo de gráfico de dispersión (Matplotlib)

Con plt.scatter:

```
>>> import matplotlib.pyplot as plt  
>>> x = [1, 2, 3, 4, 5]  
>>> y = [2, 1, 3, 5, 4]  
>>> # plot de (1,2), (2,1), ... , (5,4)  
>>> plt.scatter(x, y)  
>>> plt.ylabel("y")  
>>> plt.xlabel("x")  
>>> plt.show()
```

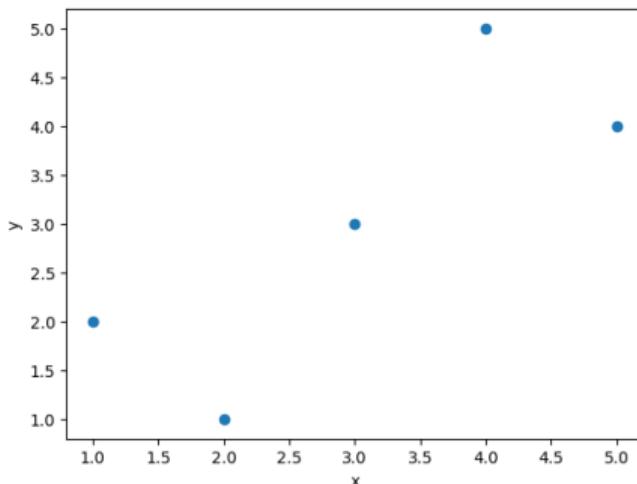


Figura 5: Gráfico de dispersión en Matplotlib.
Muestra los valores de la lista x (eje x) y lista y (eje y).

Mas ejemplos: https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.scatter.html

Ejemplo de gráfico de dispersión (Seaborn)

Con sns.scatterplot:

```
>>> import seaborn as sns  
>>> #DataFrame de propinas  
>>> df = sns.load_dataset("tips")  
>>> sns.scatterplot(data=df,  
                    x="total_bill",  
                    y="tip")  
>>> plt.show()
```

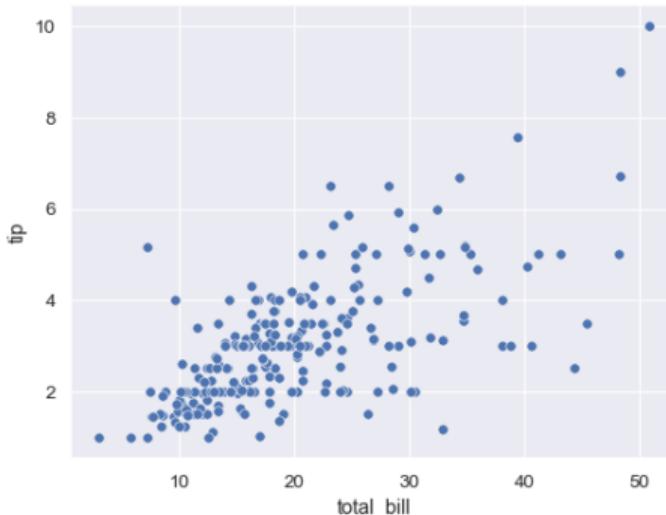


Figura 6: Gráfico de dispersión entre pago total (total bill, eje x) y propina (tip, eje y).

Mas ejemplos: <https://seaborn.pydata.org/generated/seaborn.scatterplot.html>

Grafico de Barras.

- Requiere 1 atributo categórico y 1 cuantitativo.
- Los canales que codifican información incluyen: el largo de la barra para expresar un valor cuantitativo y una separación en el espacio para representar otra categoría.
- Se puede usar para comparar categorías y encontrar casos extremos.

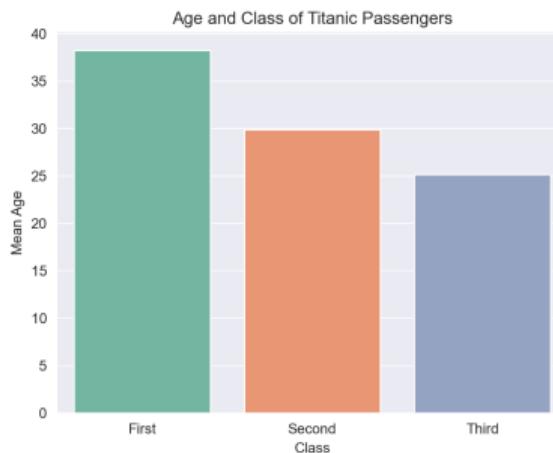


Figura 7: Edad promedio según la clase en el barco Titanic.

Ejemplo de gráfico de barras (Matplotlib)

Con plt.bar:

```
>>> import matplotlib.pyplot as plt  
>>> categorias = ["A", "B", "C", "D"]  
>>> valores = [23, 45, 12, 30]  
>>> plt.bar(categorias, valores)  
>>> plt.ylabel("Valor")  
>>> plt.xlabel("Categoría")  
>>> plt.show()
```

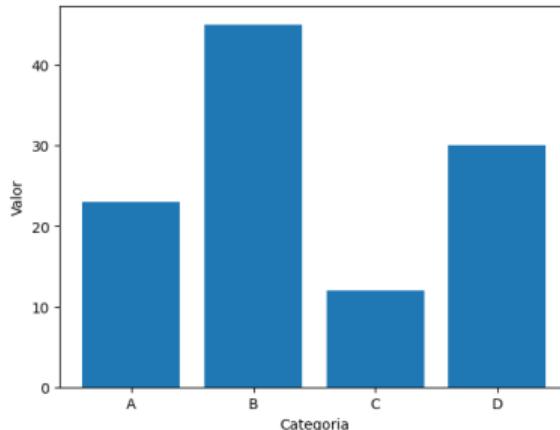


Figura 8: Gráfico de barras generado en Matplotlib. Muestra valores (eje y) para cada categoría (eje x).

Mas ejemplos: https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.bar.html

Ejemplo de gráfico de barras (Seaborn)

Con sns.barplot:

```
>>> import seaborn as sns  
>>> #DataFrame de pingüinos  
>>> df = sns.load_dataset("penguins")  
>>> sns.barplot(df,  
                 x="island",  
                 y="body_mass_g")  
>>> plt.show()
```

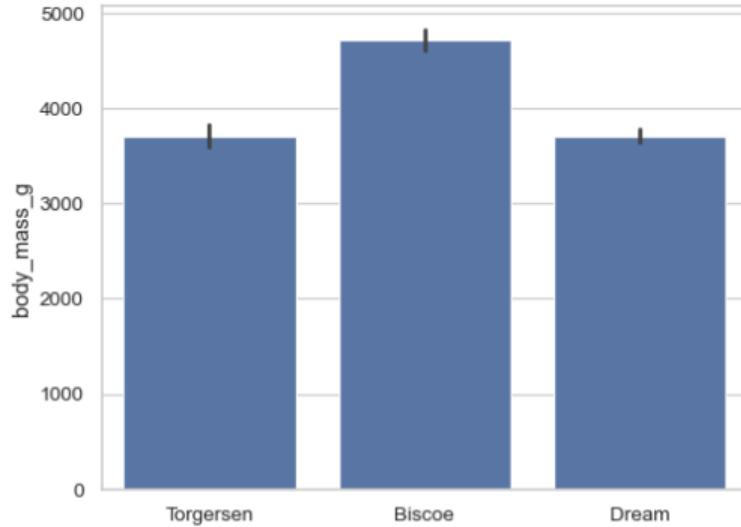


Figura 9: Gráfico de barras mostrando el peso promedio de pingüinos (variable numérica en eje y) en islas (variable categórica).

Más ejemplos: <https://seaborn.pydata.org/generated/seaborn.barplot.html>

Histograma

- ▶ Un histograma representa una variable cuantitativa agrupando los datos en intervalos.
- ▶ Cada barra corresponde a un rango de valores del eje horizontal.
- ▶ Permite analizar la forma de la distribución, su asimetría y concentración.
- ▶ El eje y muestra la frecuencia (número de observaciones) en cada intervalo.

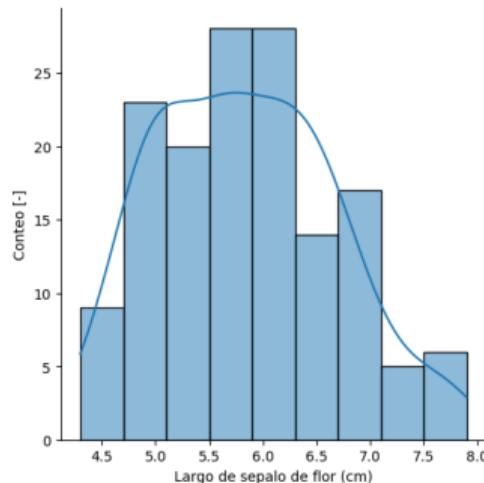


Figura 10: Histograma que muestra las distribucion del largo del sepalo en dataset Iris.

Ejemplo de histograma (Matplotlib)

Con plt.hist:

```
>>> import matplotlib.pyplot as plt  
>>> import numpy as np  
>>> temperatura = [2,33,24,25,12,20,0,23]  
>>> plt.hist(temperatura)  
>>> plt.xlabel("Temperatura [°C]")  
>>> plt.ylabel("Conteo [-]")  
>>> plt.show()
```

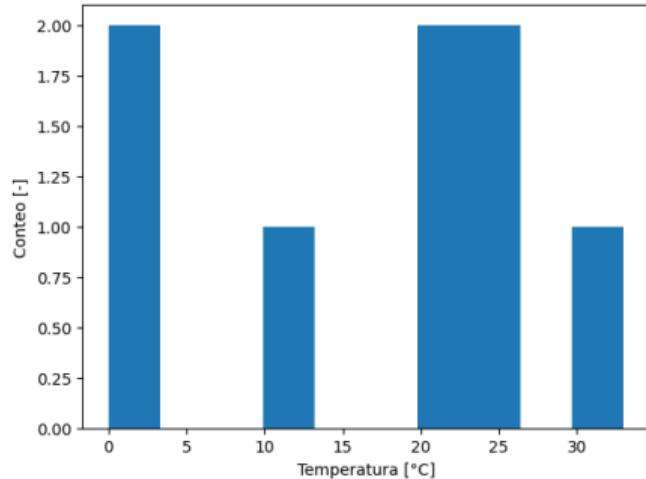


Figura 11: Histograma en matplotlib. Muestra la frecuencia (eje y) de veces que ocurre valores de temperatura (eje x).

Mas ejemplos: https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.hist.html

Ejemplo de histograma (Seaborn)

Con sns.histplot:

```
>>> import seaborn as sns  
>>> # DataFrame de pingüinos  
>>> df = sns.load_dataset("penguins")  
>>> sns.histplot(data=df,  
                 x="flipper_length_mm")  
>>> plt.show()
```

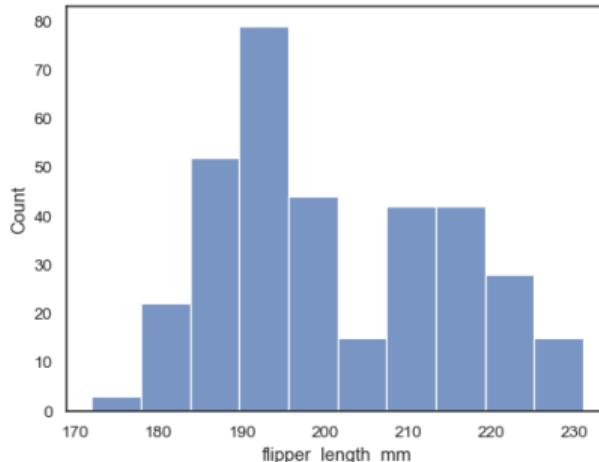


Figura 12: Histograma del largo de la aleta de pingüinos (flipper_length_mm), que muestra la frecuencia (eje y) de los distintos valores de largo de aleta observados (eje x).

Mas ejemplos: <https://seaborn.pydata.org/generated/seaborn.histplot.html>

Matplotlib: boxplot

- ▶ El boxplot resume una distribución considerando 5 valores obtenidos de la serie de datos: minimo, cuartil 1 (Q1), cuartil 2 (mediana), cuartil 3 (Q3) y maximo.
- ▶ Se usa para comparar las distribuciones entre columnas de un DataFrame en función de sus cuartiles y outliers.
- ▶ Requiere un atributo categorico (eje x) y un atributo cuantitativo (eje y).
- ▶ El principal canal de codificación es el largo de la caja (rango intercuartil) y las lineas horizontales (asociadas a outliers).

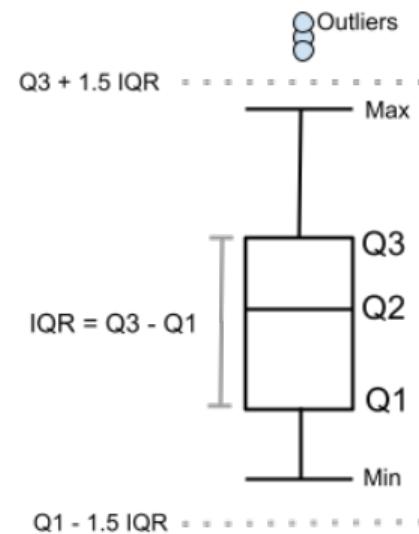


Figura 13: Componentes de boxplot. IQR es el rango intercuartil.

Matplotlib: boxplot

- ▶ El boxplot resume una distribución considerando 5 valores obtenidos de la serie de datos: minimo, cuartil 1 (Q1), cuartil 2 (mediana), cuartil 3 (Q3) y maximo.
- ▶ Se usa para comparar las distribuciones entre columnas de un DataFrame en función de sus cuartiles y outliers.
- ▶ Requiere un atributo categorico (eje x) y un atributo cuantitativo (eje y).
- ▶ El principal canal de codificación es el largo de la caja (rango intercuartil) y las lineas horizontales (asociadas a outliers).

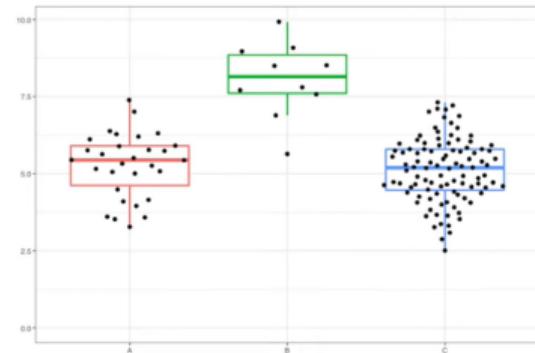


Figura 14: Ejemplo de boxplot para comparar categorias (A, B y C).

Ejemplo de boxplot en Matplotlib

Con plt.boxplot:

```
>>> import matplotlib.pyplot as plt  
>>> temperatura = [2,33,24,25,12,20,0,23,50]  
>>> plt.boxplot(temperatura)  
>>> plt.ylabel("Temperatura")  
>>> plt.show()
```

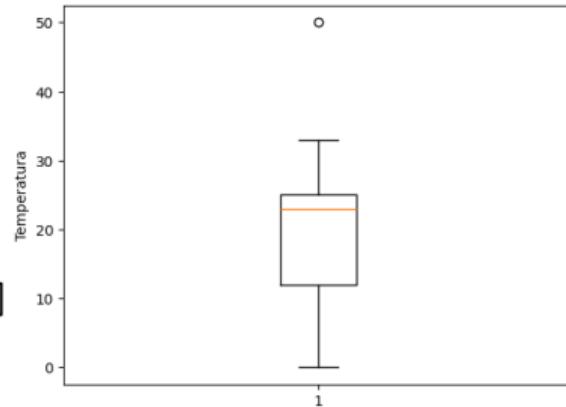


Figura 15: Boxplot de la variable temperatura (eje y), que muestra el valor mínimo, el máximo y los cuartiles primero (Q1), segundo (mediana) y tercero (Q3).

Mas ejemplos: https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.boxplot.html

Ejemplo de boxplot en Seaborn

Con sns.boxplot:

```
>>> import matplotlib.pyplot as plt  
>>> import seaborn as sns  
>>> df = sns.load_dataset('titanic')  
>>> sns.boxplot(data=df,x="class",y="age")  
>>> plt.ylabel("Edad")  
>>> plt.xlabel("Clase")  
>>> plt.show()
```

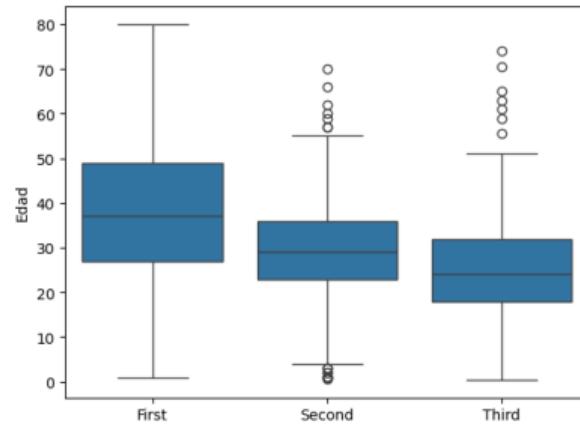


Figura 16: Boxplot de distribuciones de edades según clase en embarcación Titanic. Muestra los valores de edades mínimo, el máximo y los cuartiles primero (Q1), segundo (mediana) y tercero (Q3) separado por clase.

Mas ejemplos: <https://seaborn.pydata.org/generated/seaborn.boxplot.html>

Heatmaps

- ▶ Un heatmap es una visualización que representa valores numéricos mediante una escala de colores.
- ▶ Se utiliza para analizar patrones, concentraciones y relaciones en datos bidimensionales.
- ▶ Permite identificar rápidamente zonas de valores altos y bajos.
- ▶ Es comúnmente usado para visualizar matrices de correlación, distancias o frecuencias.
- ▶ El color es el principal canal de codificación del valor.

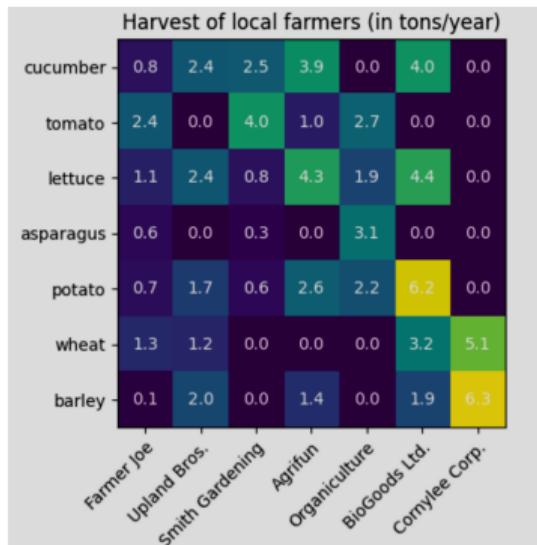


Figura 17: Ejemplo de heatmap. El eje x representa diferentes granjeros; el eje y representa verduras. El color en la matriz representa la cantidad de verduras cosechadas (toneladas/año) para combinacion de verdura y granjero.

Heatmap con Matplotlib (imshow)

Con plt.imshow:

```
>>> import matplotlib.pyplot as plt  
>>> import numpy as np  
>>> data = np.random.rand(4, 4)  
>>> plt.imshow(data, cmap="viridis")  
>>> plt.colorbar(label="Valor")  
>>> plt.xlabel("Indice columna")  
>>> plt.ylabel("Indice fila")  
>>> plt.title("Heatmap usando imshow")  
>>> plt.show()
```

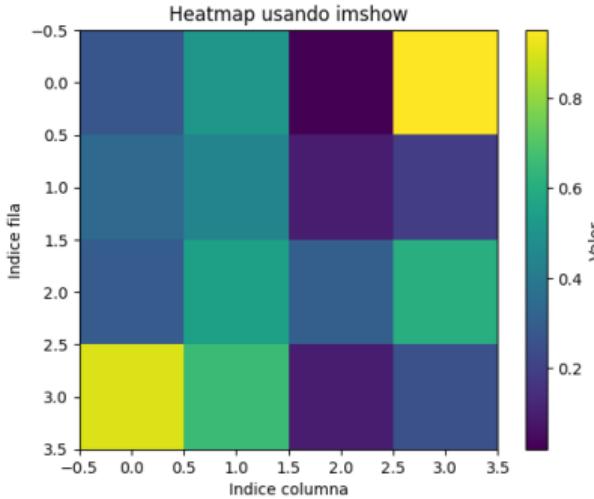


Figura 18: Heatmap generado con imshow en Matplotlib. Los colores representan los valores numéricos de una matriz de 4×4 generada aleatoriamente, donde cada celda se codifica mediante una escala de color.

Heatmap con Seaborn

Con sns.heatmap:

```
>>> import seaborn as sns  
>>> import matplotlib.pyplot as plt  
>>> df = sns.load_dataset("iris")  
>>> corr = df.corr(numeric_only=True)  
>>> sns.heatmap(corr, annot=True,  
                 cmap="coolwarm")  
>>> plt.title("Matriz de correlación")  
>>> plt.show()
```

Mas ejemplos: <https://seaborn.pydata.org/generated/seaborn.heatmap.html>

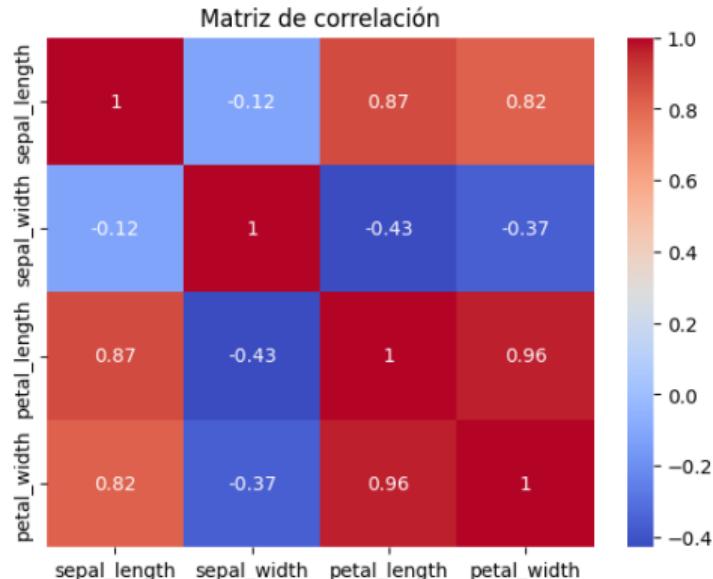


Figura 19: Heatmap de la matriz de correlación del dataset Iris, donde los colores indican la intensidad y el signo de la correlación lineal entre las variables numéricas, y los valores anotados corresponden a los coeficientes de correlación.

Vamos a ver un ejemplo práctico en Python Collab

Exploremos un dataset: <https://colab.research.google.com/drive/1Tdz4RbEasHZ69BbEusp=sharing>.



Figura 20: Pokemones de primera generación.

Matplotlib (Resumen)

```
import matplotlib.pyplot as plt # podemos importarla de esta forma
```

Función	Uso	Ejemplo
plot	Gráfico de línea	plt.plot(valores)
scatter	Gráfico de puntos	plt.scatter(valores_x, valores_y)
bar	Gráfico de barras	plt.bar(etiquetas, valores)
boxplot	Gráfico de caja	plt.boxplot(valores)
histogram	Histograma	plt.hist(valores)
imshow	Gráfico heatmap	plt.imshow(matriz)

Tabla 1: Funciones en matplotlib.

Seaborn (Resumen)

```
import seaborn as sns # podemos importarla de esta forma
```

- ▶ **lineplot**: Gráfico de que muestra la relación entre la columna 'x' e 'y' de un dataframe.
`sns.lineplot(data=dataframe, x="x", y="y")`
- ▶ **scatterplot**: Gráfico de dispersión entre columna 'x' e 'y' de un dataframe.
`sns.scatterplot(data=dataframe, x="x", y="y")`
- ▶ **histplot**: Histograma de la columna 'x' de un dataframe.
`sns.histplot(data=dataframe, x="x")`
- ▶ **boxplot**: Comparación de distribuciones de la columna '**valor**' en un valor.
`sns.boxplot(data=dataframe, x="grupo", y="valor")`
- ▶ **heatmap**: Visualización matricial donde los colores representan los valores numéricos de una tabla, comúnmente usada para analizar relaciones entre variables (por ejemplo, matrices de correlación).
`sns.heatmap(dataframe, annot=True)`

Referencias:

- ▶ Wes McKinney. (2022). Python for Data Analysis. Third Edition.
- ▶ https://seaborn.pydata.org/tutorial/function_overview.html