

# Hoja informativa: Primeros gráficos y conclusiones

## Práctica

```
# Lectura de datos de un archivo usando separadores decimales y sep

data = read_csv('file.csv', sep=';', decimal=',')

# Mostrar un histograma con n_bins (número de contenedores) y valores mínimos y máximos (min_value y max_value)

import matplotlib.pyplot as plt

data['column'].hist(bins=n_bins, range=(min_value, max_value))

plt.show()

# Mostrar un diagrama de caja

import matplotlib.pyplot as plt

data.boxplot('column')

plt.show()

# Cambiar las escalas de los ejes: x_min y x_max para el eje X, y_min e y_max para el eje Y

import matplotlib.pyplot as plt
plt.xlim(x_min, x_max)
plt.ylim(y_min, y_max)

# Descripción numérica de una columna en los datos

data['column'].describe()
```

## Teoría

**Histogramas** - gráficos que muestran con qué frecuencia ocurren los valores en un conjunto de datos en particular

**Distribución:** todos los valores posibles de una variable, así como su frecuencia

**Distribución normal:** una curva simétrica en forma de campana con valores promedio y casi-promedio que ocurren con frecuencia y aquellos más alejados que son menos comunes

**Cuartil:** divide los conjuntos de datos en dos partes separadas: un grupo menor que el cuartil y un grupo mayor que él (puede ser mayor que el 25%, el 50% o el 75% de los datos)

**Primer cuartil (Q1):** el 25% de los elementos son menores, mientras que el 75% son mayores

**Mediana (Q2):** la mitad de los elementos son menores

**Tercer cuartil (Q3):** el 75% de los elementos son menores

**Rango intercuartílico:** Q3 menos Q1

**Desviación estándar:** una medida de dispersión que indica cuánto difieren los valores de un conjunto de datos de la media; representado por sigma ( $\sigma$ )

**Descripción numérica de los datos:** la media, la mediana, la desviación estándar, el número de observaciones en el conjunto de datos, así como su distribución