

# Scatter plot

Biểu đồ phân tán dùng để trực quan hóa mối tương quan giữa 2 biến, tính phân cụm của dữ liệu Trong thư viện matplotlib, scatter có thể được vẽ theo 2 cách:

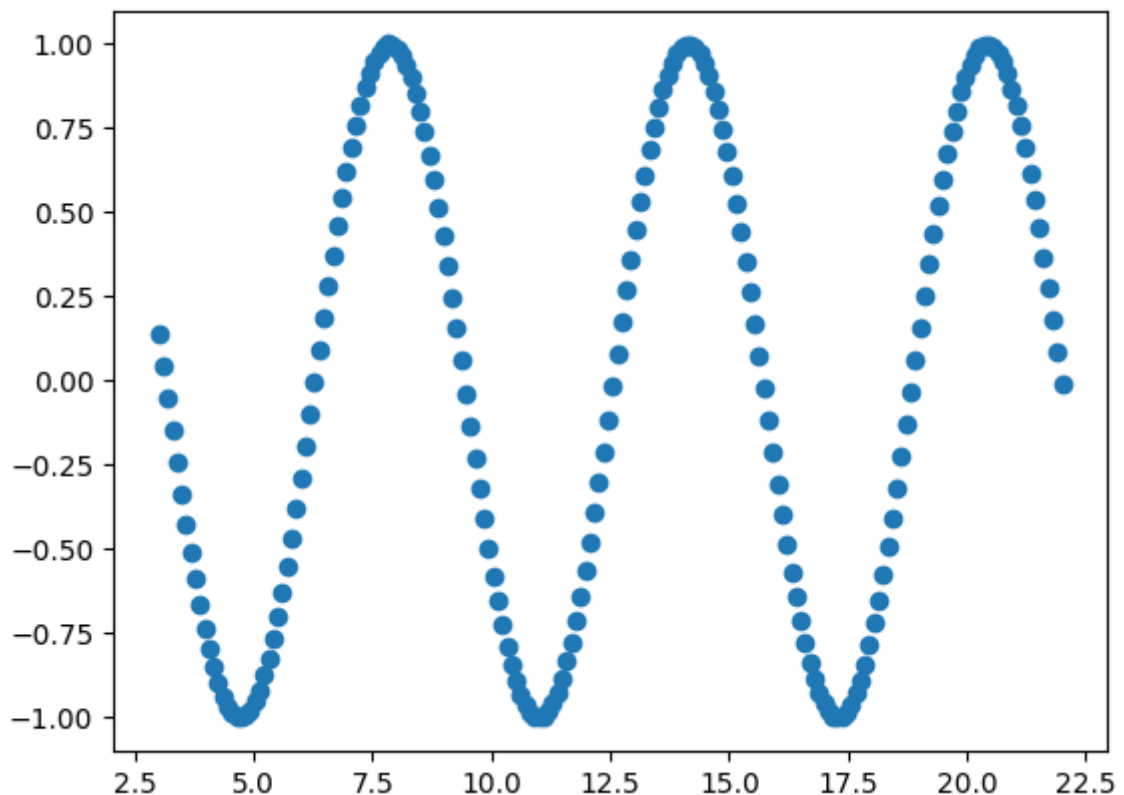
- Hàm plot()
- Hàm scatter()

## Sử dụng hàm plot()

```
In [3]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

x = np.linspace(3, 22, 198)
fig, axes = plt.subplots(dpi=100)
axes.plot(x, np.sin(x), "o") # sử dụng marker là chấm
```

Out[3]: [matplotlib.lines.Line2D at 0x7fb11c7662e0]



## Sử dụng hàm scatter()

Hàm scatter() sử dụng tương tự như hàm plot(), ngoài ra có thêm các tham số để điều chỉnh đồ thị:

- x, y: các mảng 1 chiều chứa giá trị tọa độ các điểm dữ liệu (tương tự như hàm plot)
- s: số vô hướng hoặc mảng để điều chỉnh kích thước marker
- c: màu của marker

- marker: hình dạng marker
- cmp: color map của marker, sử dụng trong trường hợp muốn tinh chỉnh màu của marker sâu hơn
- alpha: sử dụng khi muốn tạo hiệu ứng trong/mờ của biểu đồ

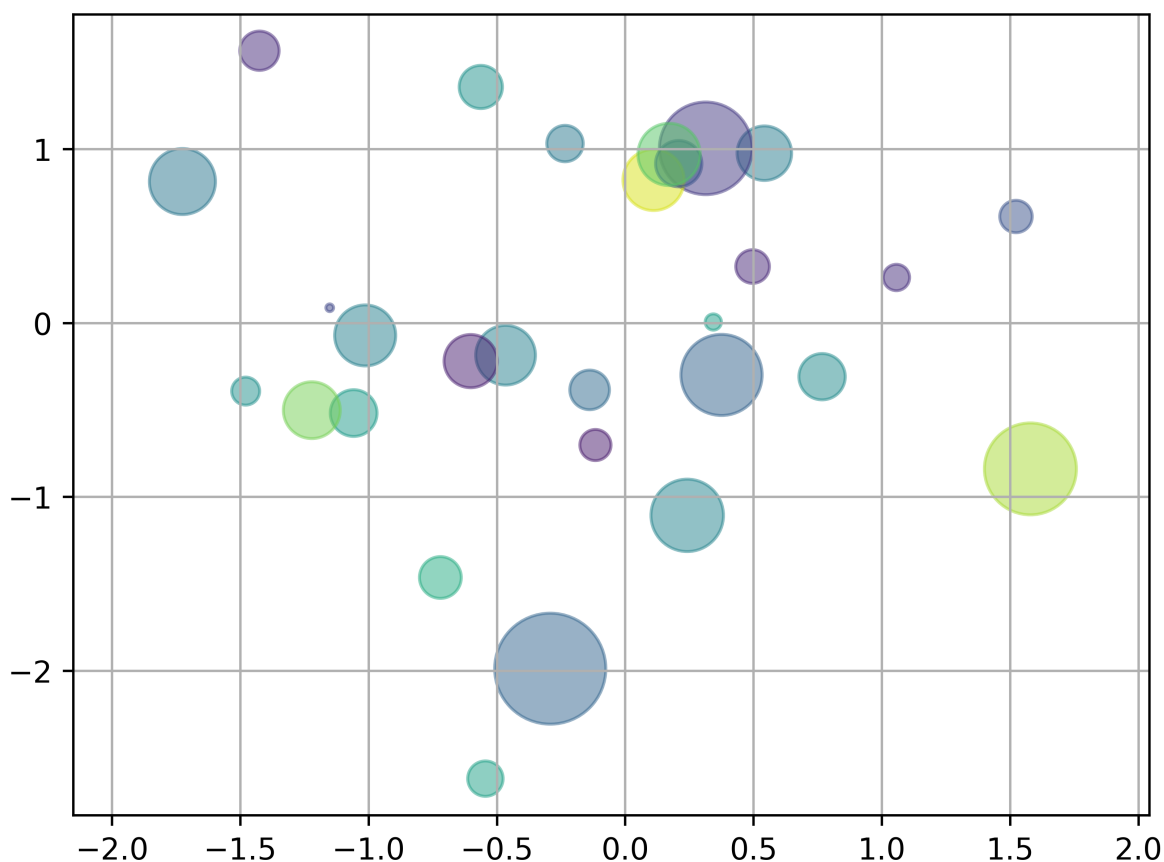
```
In [4]: import numpy as np
import matplotlib.pyplot as plt
import pandas as pd

# Test thử hàm scatter với 50 điểm dữ liệu
fig, axes = plt.subplots(dpi=600)
random_state = np.random.RandomState(42)
x = random_state.randn(50)
y = random_state.randn(50)
colors = random_state.randn(50)
sizes = random_state.randn(50) * 500
axes.grid()
axes.scatter(x, y, c=colors, s=sizes, alpha=0.5)
```

/Users/ducva/anaconda3/envs/jupyter-notebook/lib/python3.8/site-packages/matplotlib/collections.py:981: RuntimeWarning: invalid value encountered in sqrt

```
scale = np.sqrt(self._sizes) * dpi / 72.0 * self._factor
```

Out[4]: <matplotlib.collections.PathCollection at 0x7fb106fa6040>



In [ ]: