

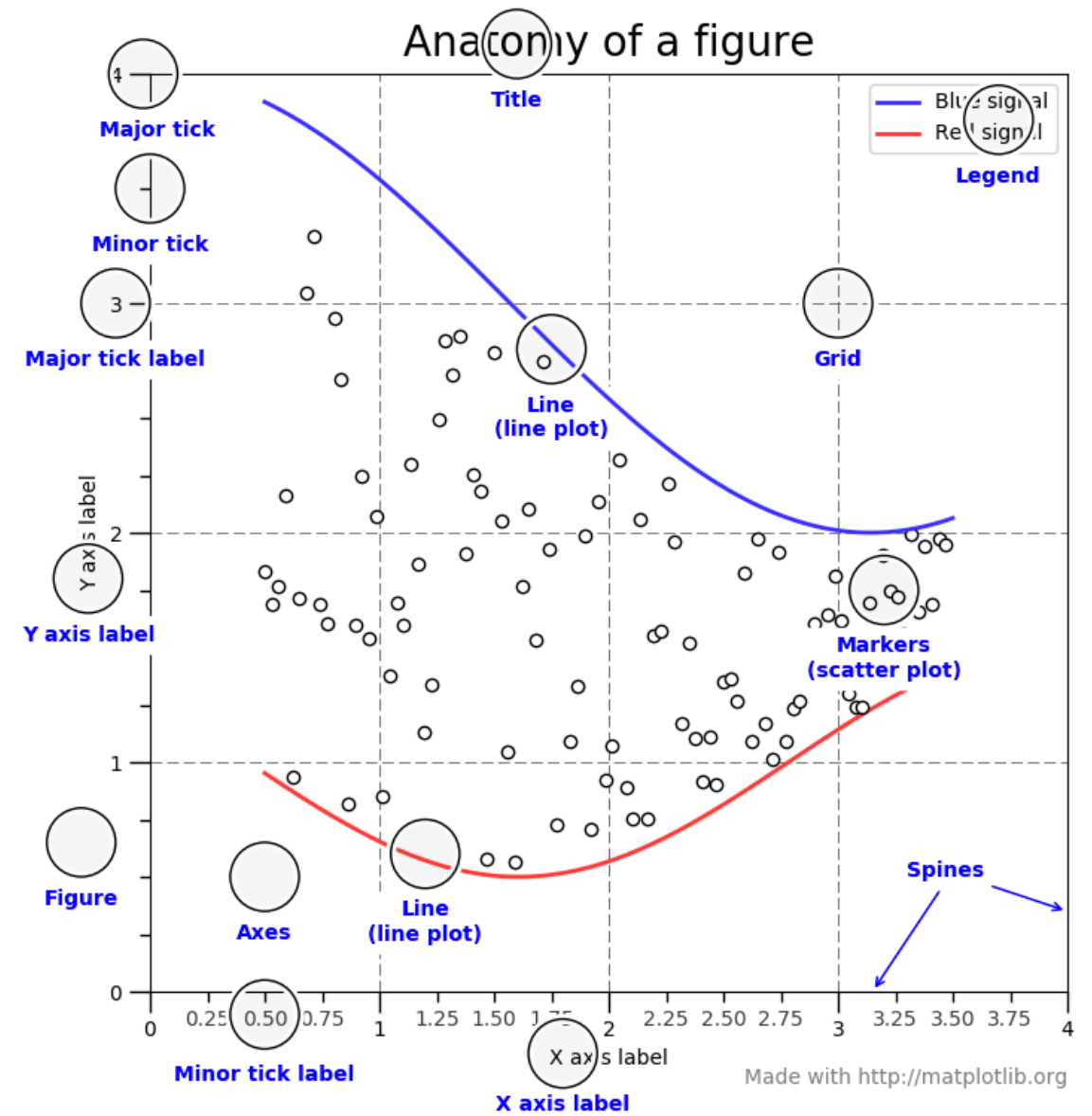
데이터 사이언스를 위한 Advanced Python

Visualization with matplotlib

2019.09.27

기본 개념

- *Figure*: 전체 그래프
- *Axes*: 그래프를 구성하는 작은 그래프
 - *Axis*
 - *Tick*
 - *Legend*



con



Made with <http://matplotlib.org>

matplotlib

라이브러리 불러오기

- 아래 코드를 이용해 *numpy*, *matplotlib*를 불러온다

```
%matplotlib inline
import numpy as np
import matplotlib.pyplot as plt
```

- 성공적으로 라이브러리가 불러와지면 아래 코드를 실행해본다

```
fig, ax = plt.subplots()

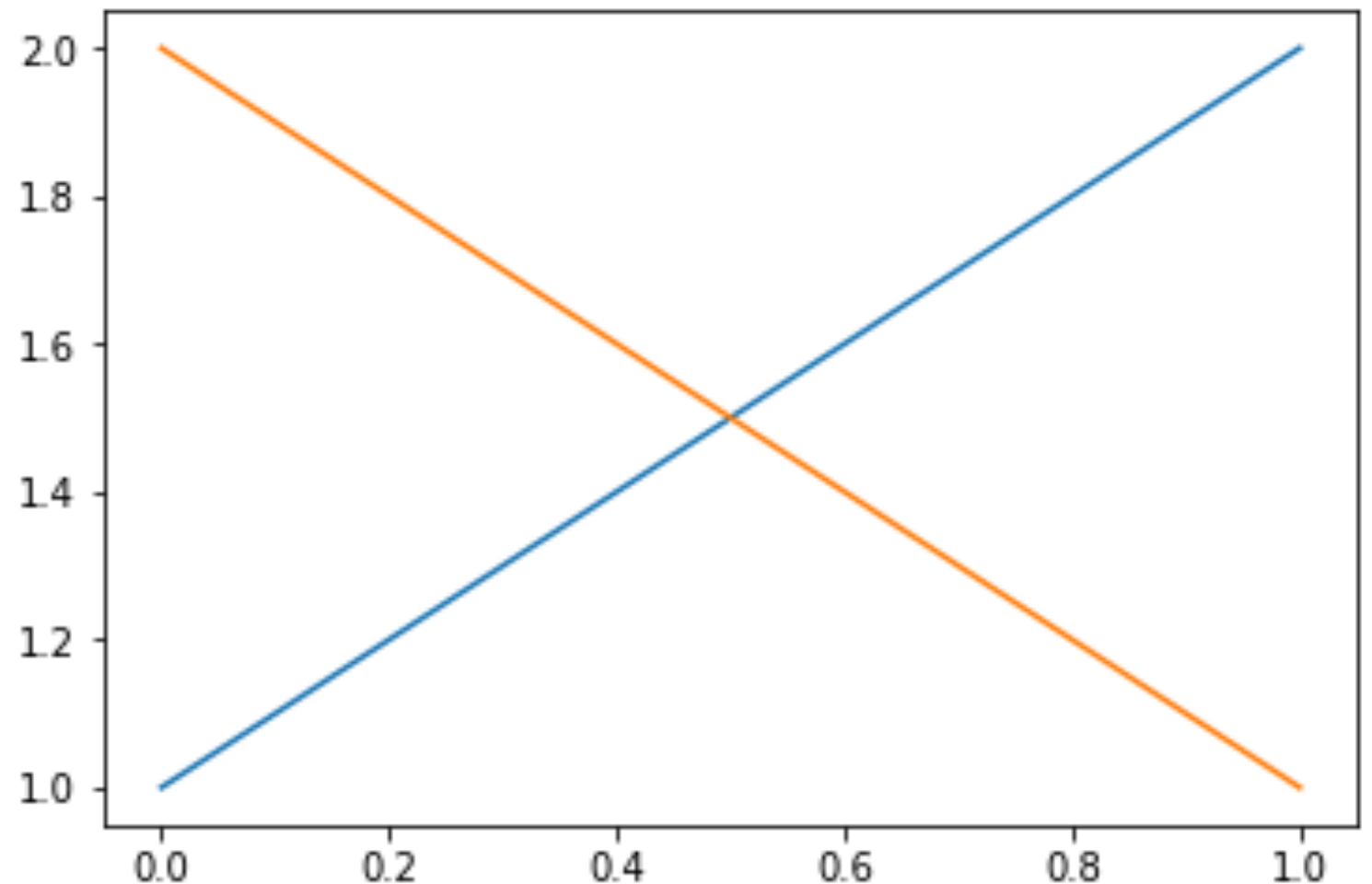
ax.plot([0, 1], [1, 2])
plt.show()
```

matplotlib

직선 그래프 그리기

- `ax.plot()` : 선 그리기
- `plt.show()` : 그래프 화면에 표시하기

```
fig, ax = plt.subplots()  
  
ax.plot([0, 1], [1, 2])  
ax.plot([0, 1], [2, 1])  
plt.show()
```



matplotlib

곡선 그래프 그리기

- `np.linspace(start, end, count)` : 구간 나누기
- `plt.show()` : 그래프 화면에 표시하기

```
fig, ax = plt.subplots()

X = np.linspace(-np.pi, np.pi, 256, endpoint=True)
S, C = np.sin(X), np.cos(X)

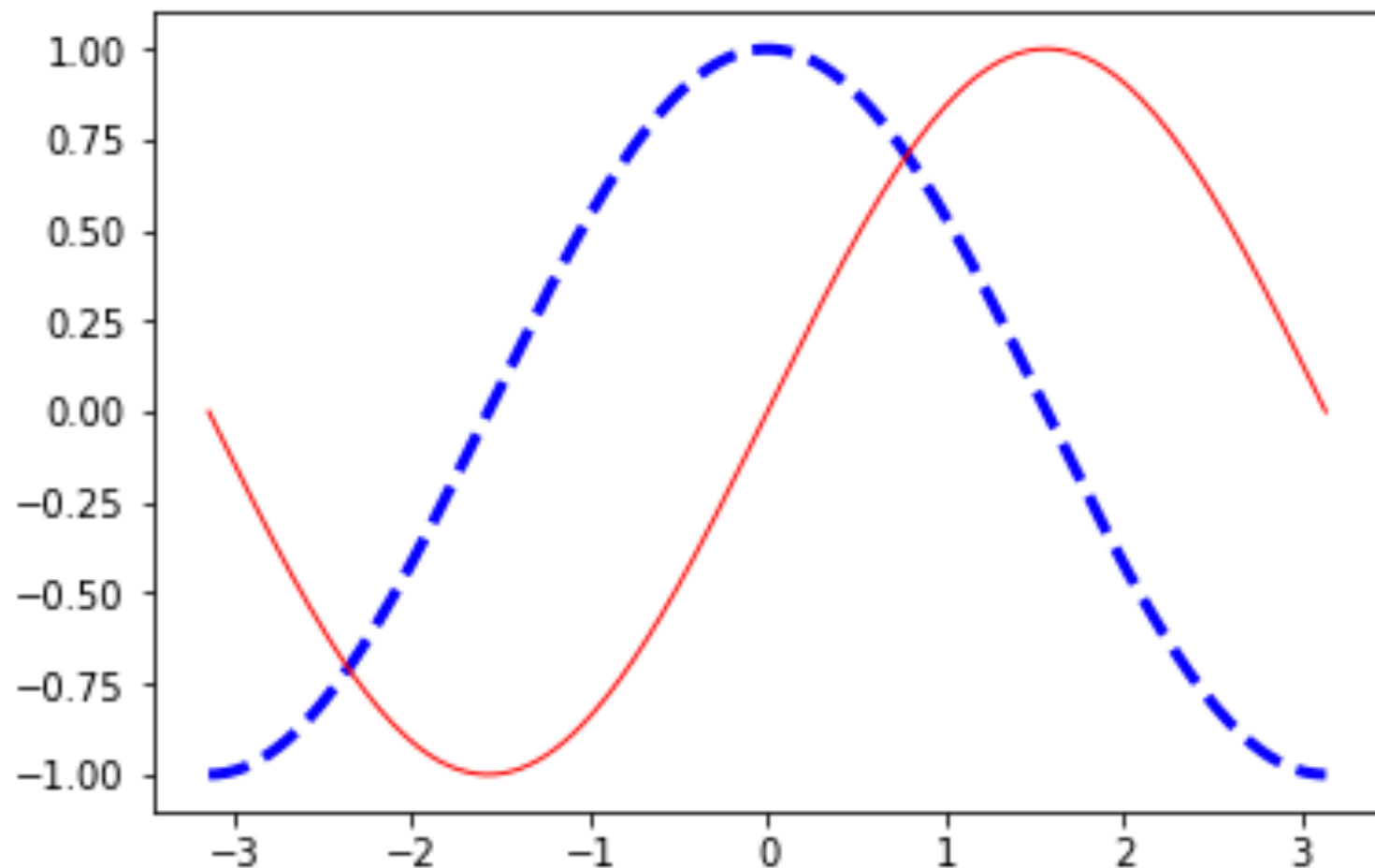
ax.plot(X, S)
ax.plot(X, C)

plt.show()
```

matplotlib

선 스타일 바꾸기

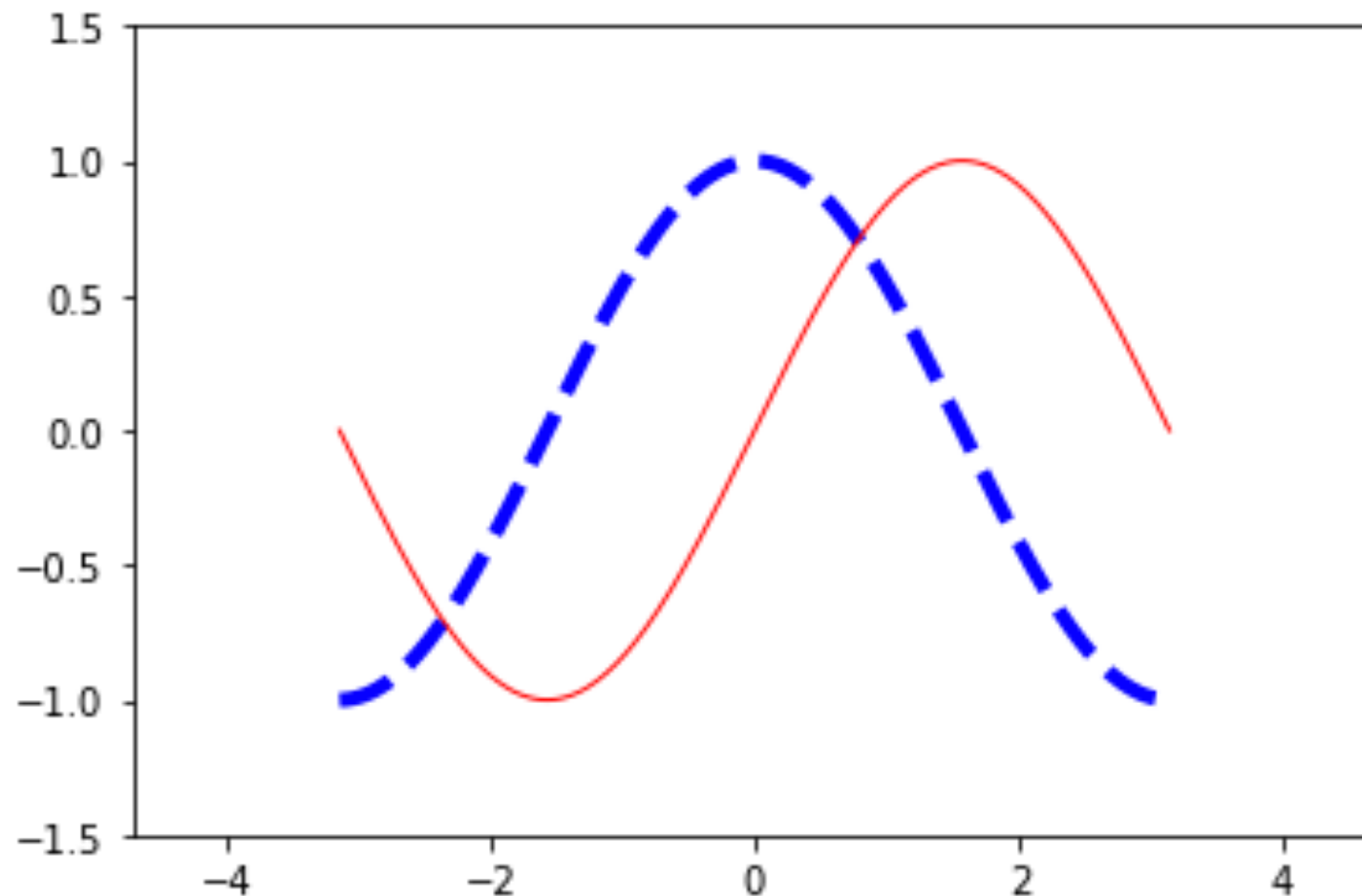
- `ax.plot(x, y, color, linewidth, linestyle)`
 - `ax.plot(X, C, color="blue", linewidth=3, linestyle="--")`
 - `ax.plot(X, S, color="red", linewidth=1, linestyle="-")`



matplotlib

그래프가 표시할 범위 설정하기

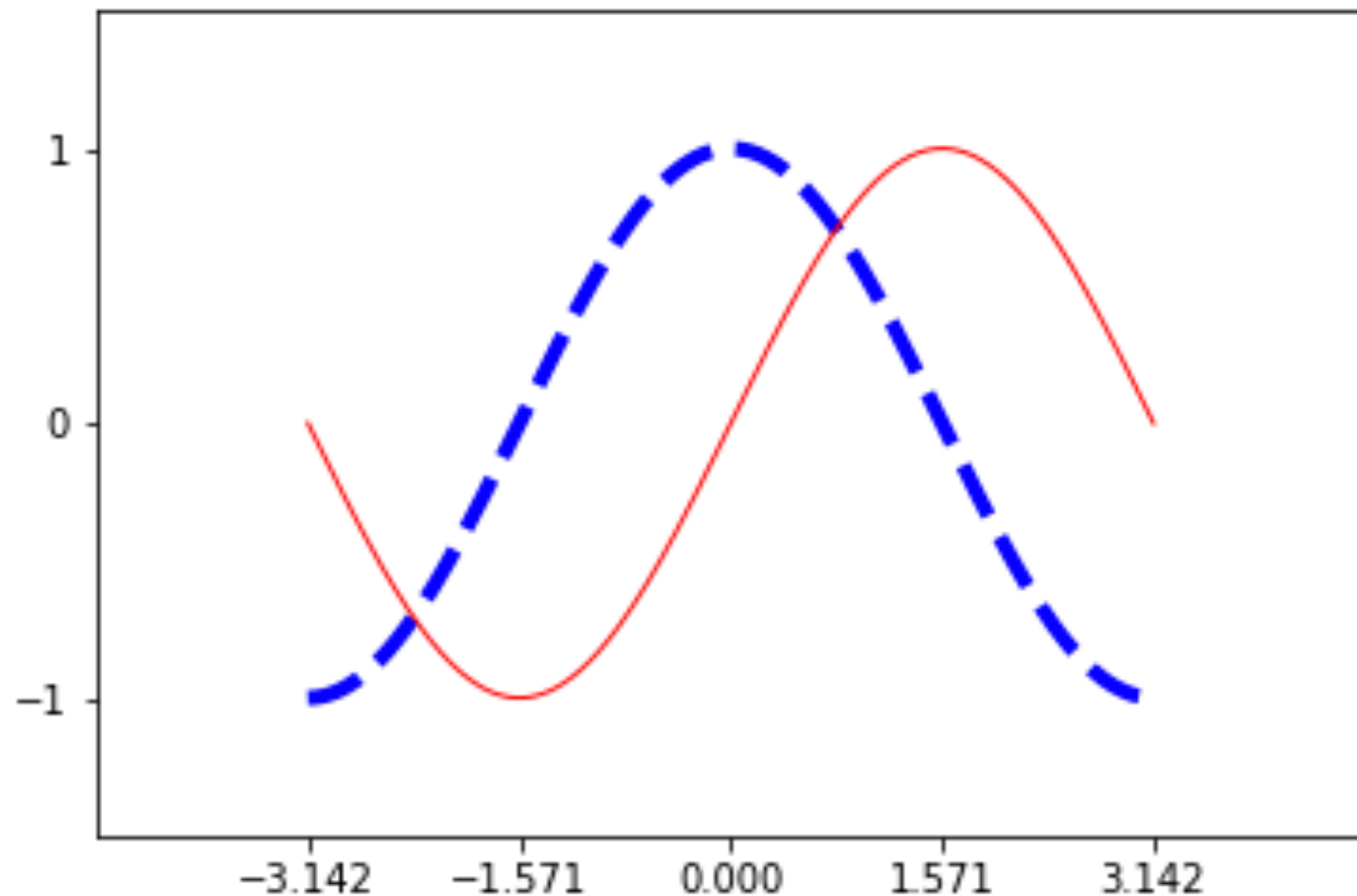
- `ax.set_xlim(x_min, x_max)`, `ax.set_ylim(y_min, y_max)`
 - `ax.set_xlim(X.min() * 1.5, X.max() * 1.5)`
 - `ax.set_ylim(C.min() * 1.5, C.max() * 1.5)`



matplotlib

가로 축, 세로 축 눈금 설정하기

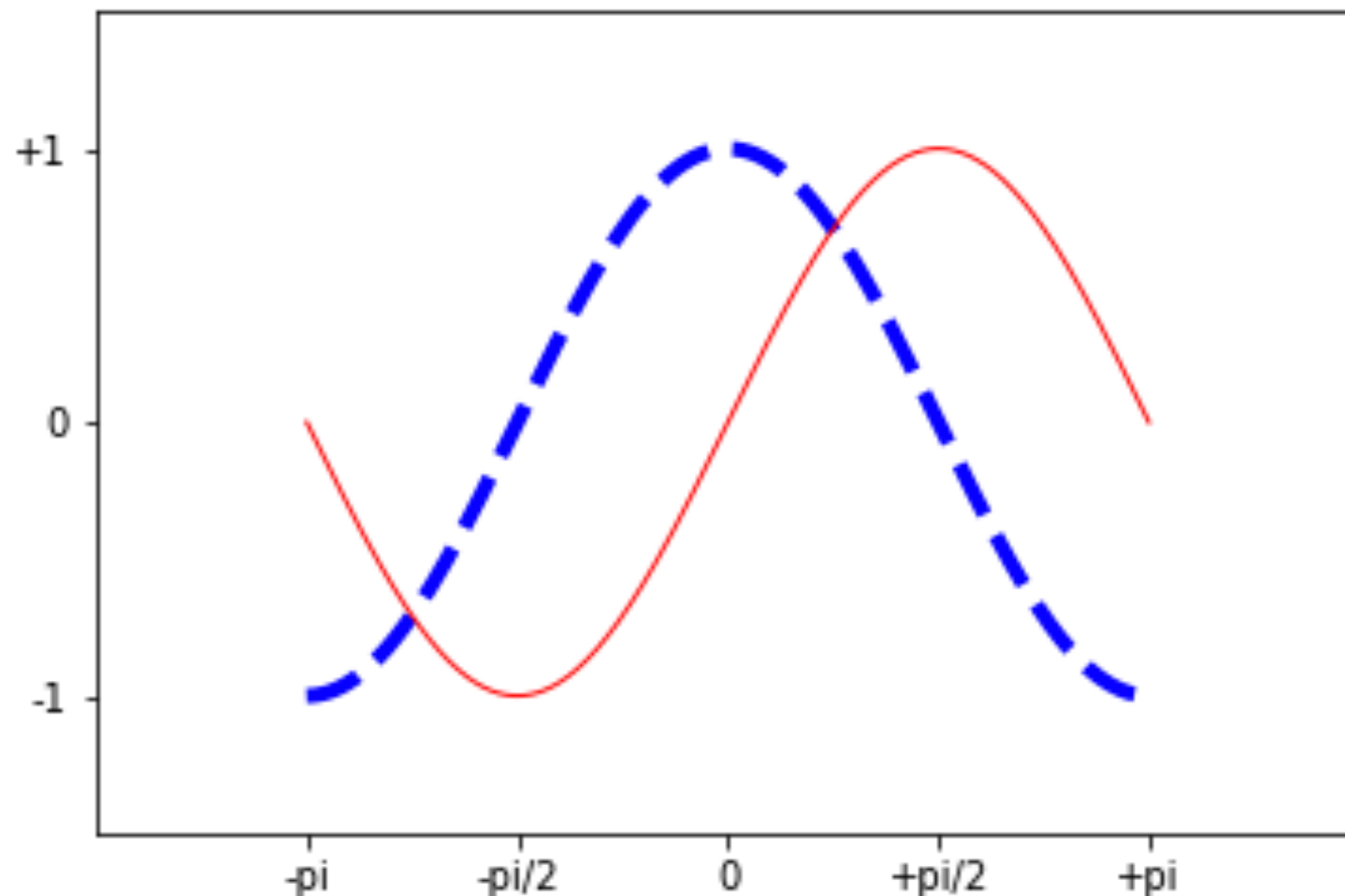
- `ax.set_xticks(list)`, `ax.set_yticks(list)`
 - `ax.set_xticks([-np.pi, -np.pi/2, 0, np.pi/2, np.pi])`
 - `ax.set_yticks([-1, 0, +1])`



matplotlib

가로 축, 세로 축 눈금에 이름 붙이기

- `ax.set_xticklabels(list)`, `ax.set_yticklabels(list)`
 - `ax.set_xticklabels(["-pi", "-pi/2", "0", "+pi/2", "+pi"])`
 - `ax.set_yticklabels(["-1", "0", "+1"])`

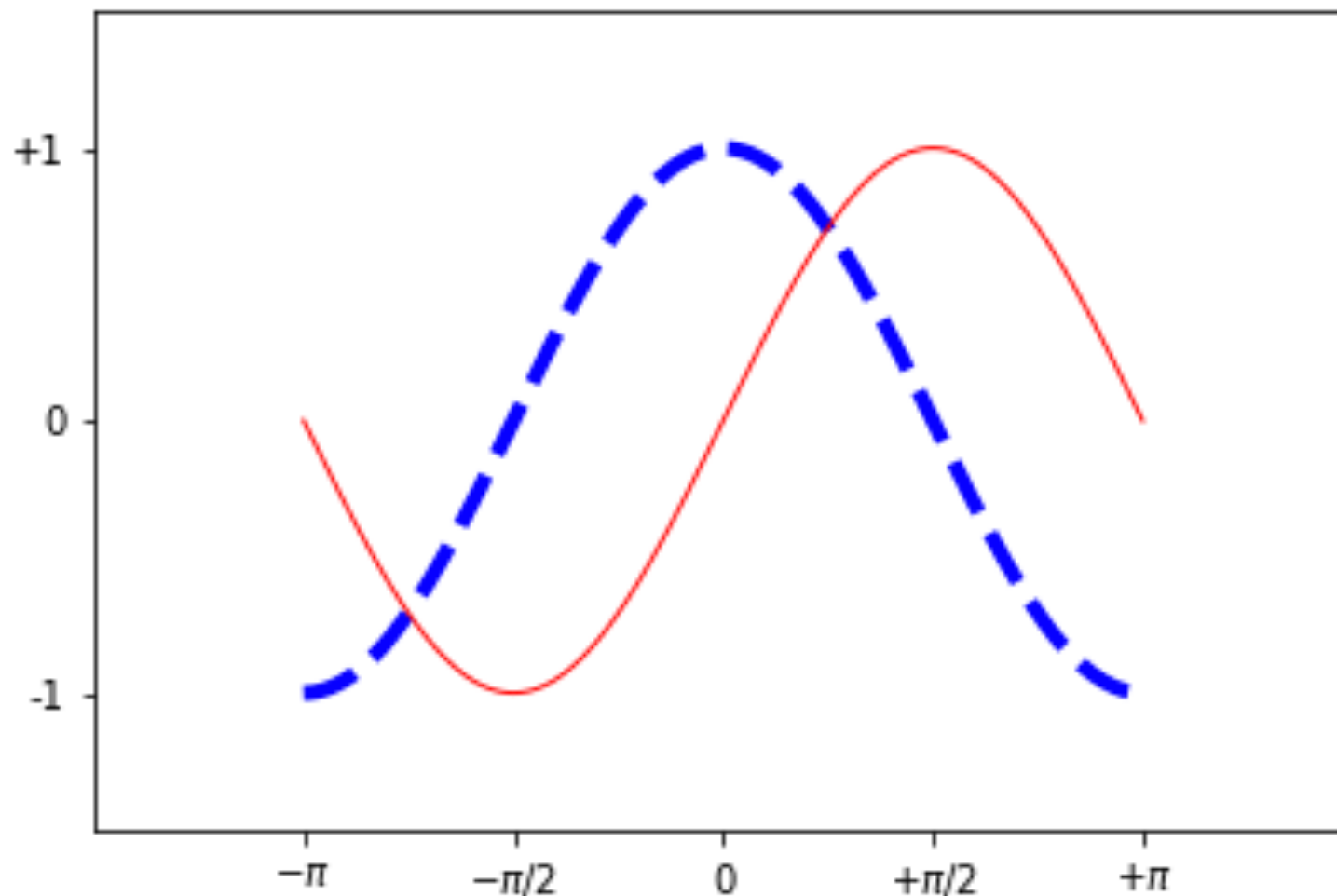


matplotlib

가로 축, 세로 축 눈금에 이름 붙이기

- 참고

- raw string을 이용하면 그래프에 수식 문자열을 사용할 수 있다
- `ax.set_xticklabels([r'$-\pi$', r'$-\pi/2$', r'0', r'$+\pi/2$', r'$+\pi$'])`



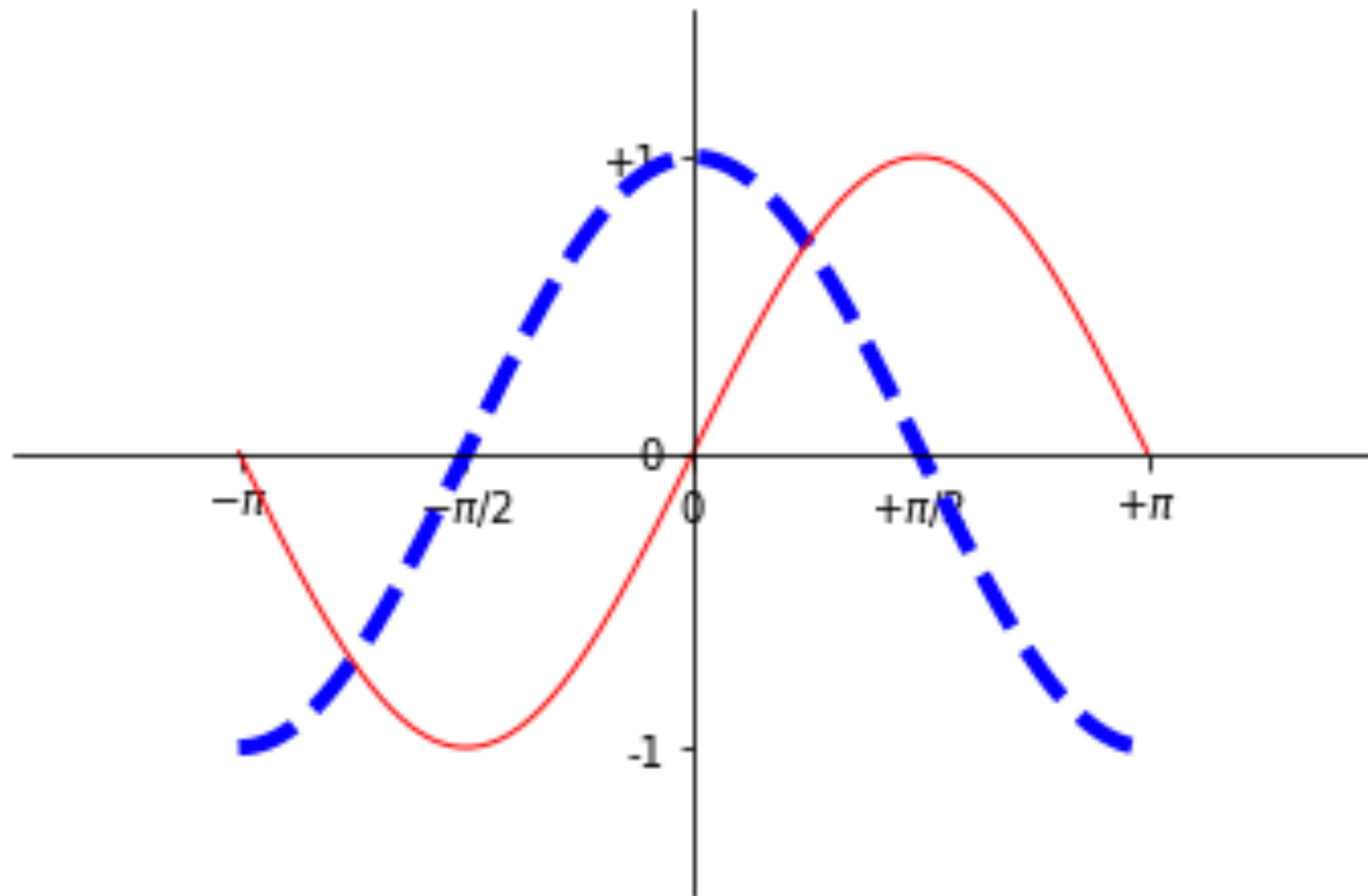
matplotlib

가로 축, 세로 축 이동하기

- `ax.spines[position].set_color(color_string)`: 축 색깔 정하기
- `ax.spines[position].set_position(location)`: 축 위치 정하기
 - `ax.spines['right'].set_color('none')`: 오른 쪽 축 표시하지 않기
 - `ax.spines['top'].set_color('none')`
 - `ax.spines['bottom'].set_position(('data',0))`: data의 0으로 이동
 - `ax.spines['left'].set_position(('data',0))`

matplotlib

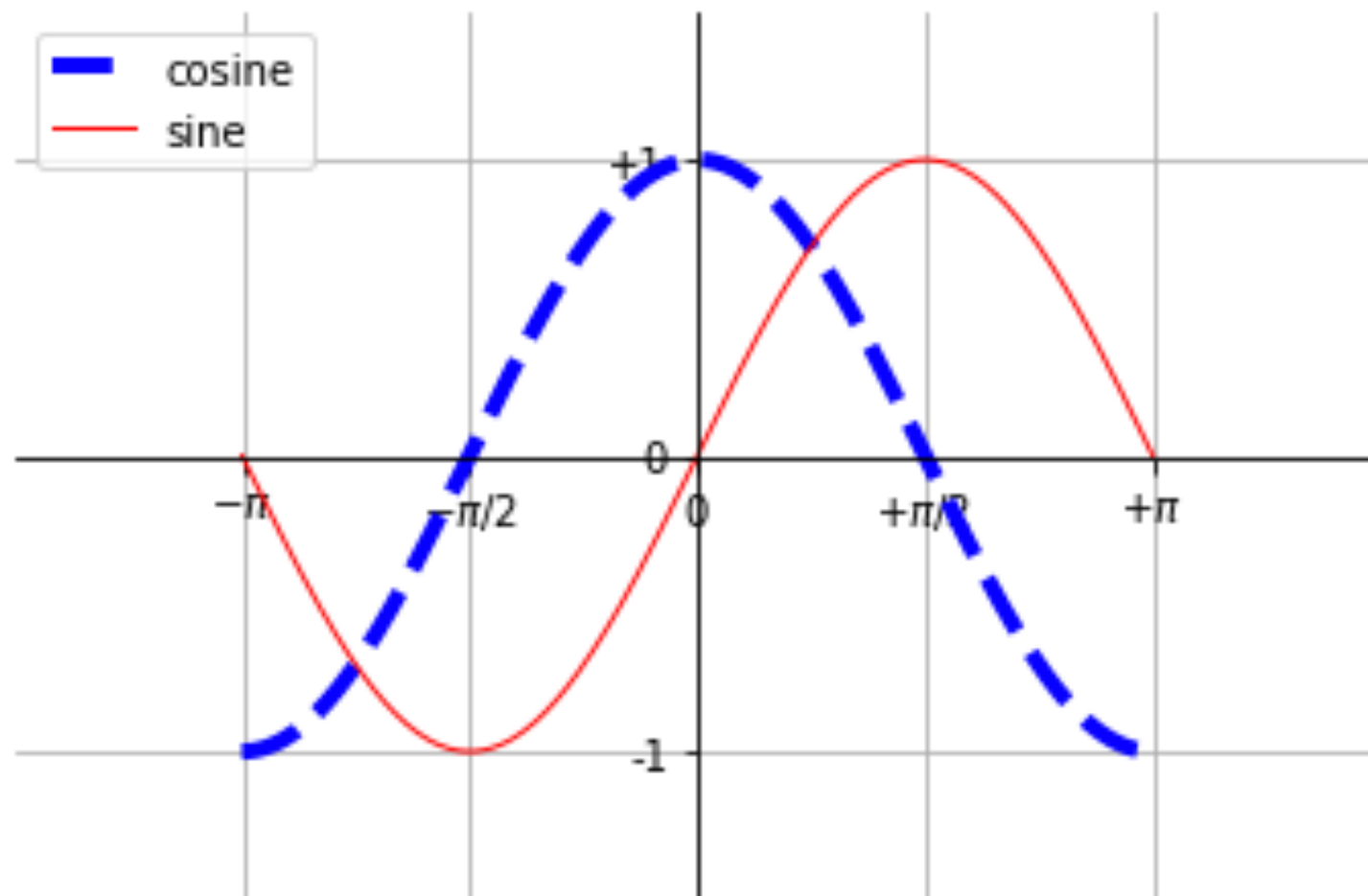
가로 축, 세로 축 이동하기



matplotlib

범례와 눈금 표시하기

- `ax.grid(True)`
- `ax.legend(loc='upper left')`
 - `ax.plot(X, C, color="blue", linewidth=4, linestyle="--", label="cosine")`



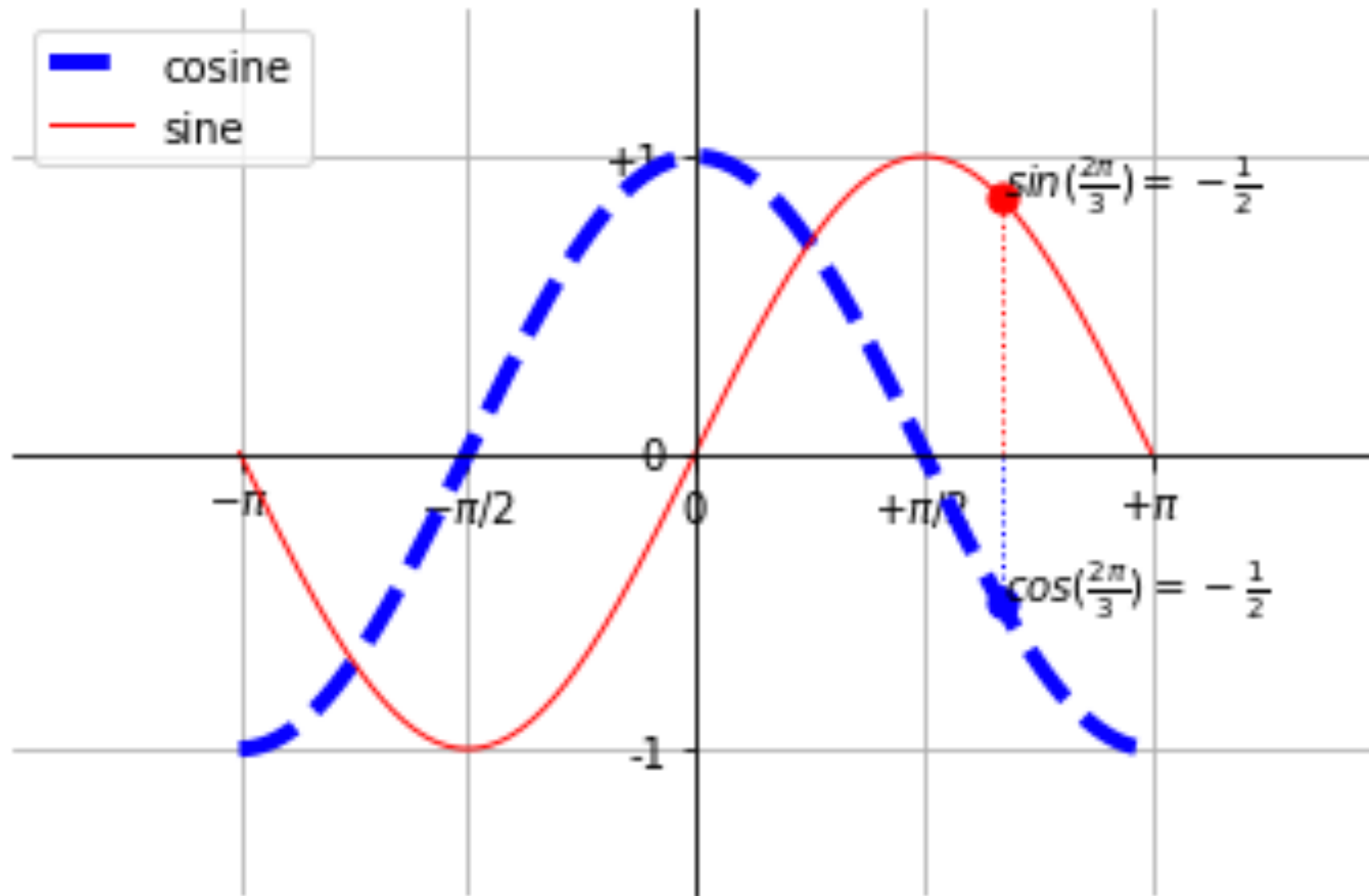
matplotlib

특정 점 강조하기

- `ax.scatter(x, y)`: 점 표시하기
- `ax.annotate(text, xy)`: 지정된 위치에 주석달기
 - `t = 2 * np.pi / 3`
 - `ax.plot([t, t], [0, np.cos(t)], color='blue', linewidth=1, linestyle=":")`
 - `ax.scatter([t,], [np.cos(t),], 50, color='blue')`
 - `ax.annotate(r'$\cos(\frac{2\pi}{3}) = -\frac{1}{2}$', xy=(t, np.cos(t)))`
 - `ax.annotate(r'$\sin(\frac{2\pi}{3}) = \frac{1}{2}$', xy=(t, np.sin(t)))`
 - `ax.plot([t, t], [0, np.sin(t)], color='red', linewidth=1, linestyle=":")`
 - `ax.scatter([t,], [np.sin(t),], 50, color='red')`

matplotlib

특정 점 강조하기



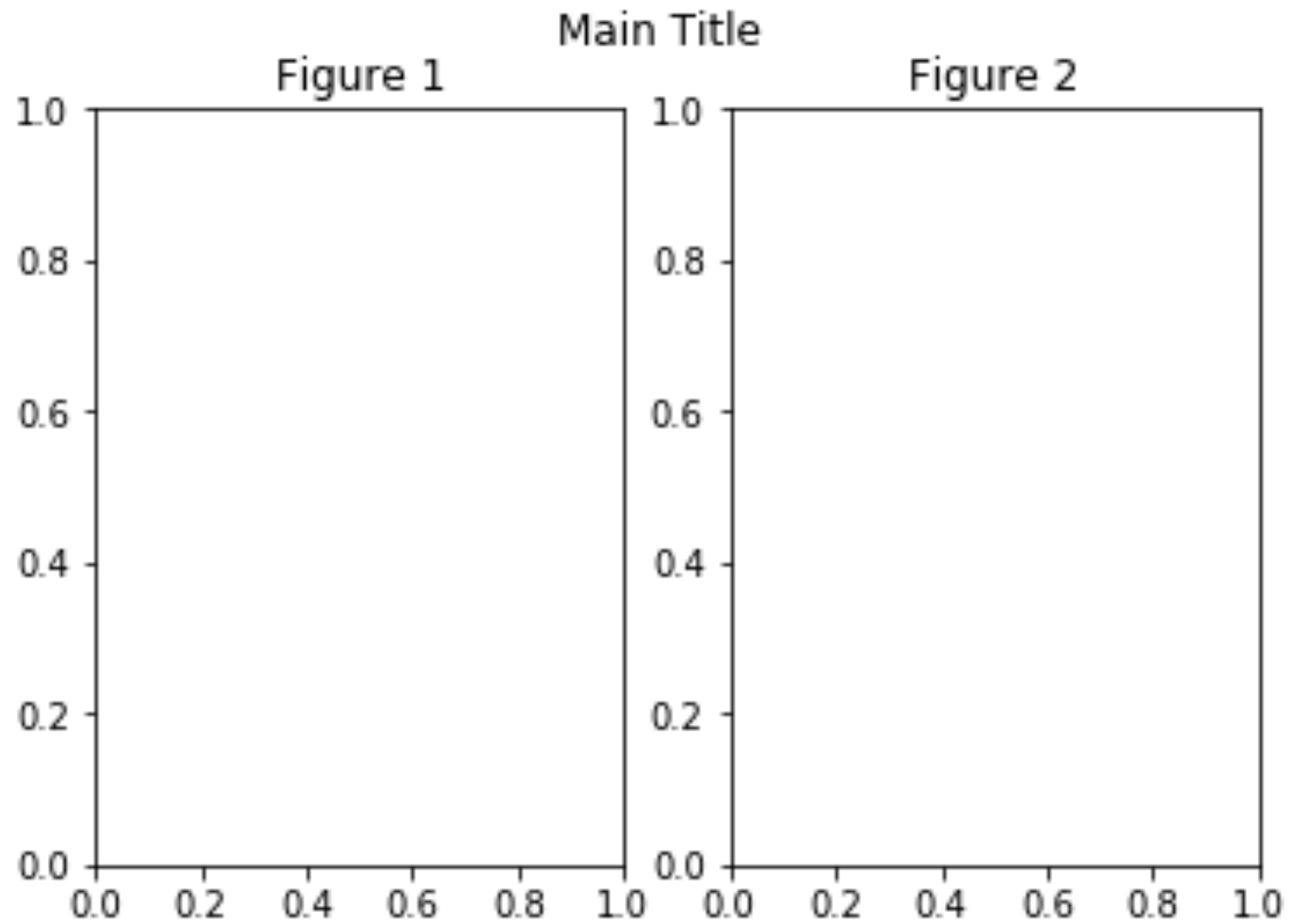
matplotlib

여러개 plot 한 번에 그리기

- `plt.subplots(nrows, ncols)`
 - `fig, axes = plt.subplots(1, 2)`
 - `fig.suptitle("Main Title")`
 - `fig.subplots_adjust(wspace=0.2) #hspace=0.5,`
- `axes[0].set_title("Figure 1 ")`
- `axes[1].set_title("Figure 2")`
- `plt.show()`

matplotlib

여러개 plot 한 번에 그리기



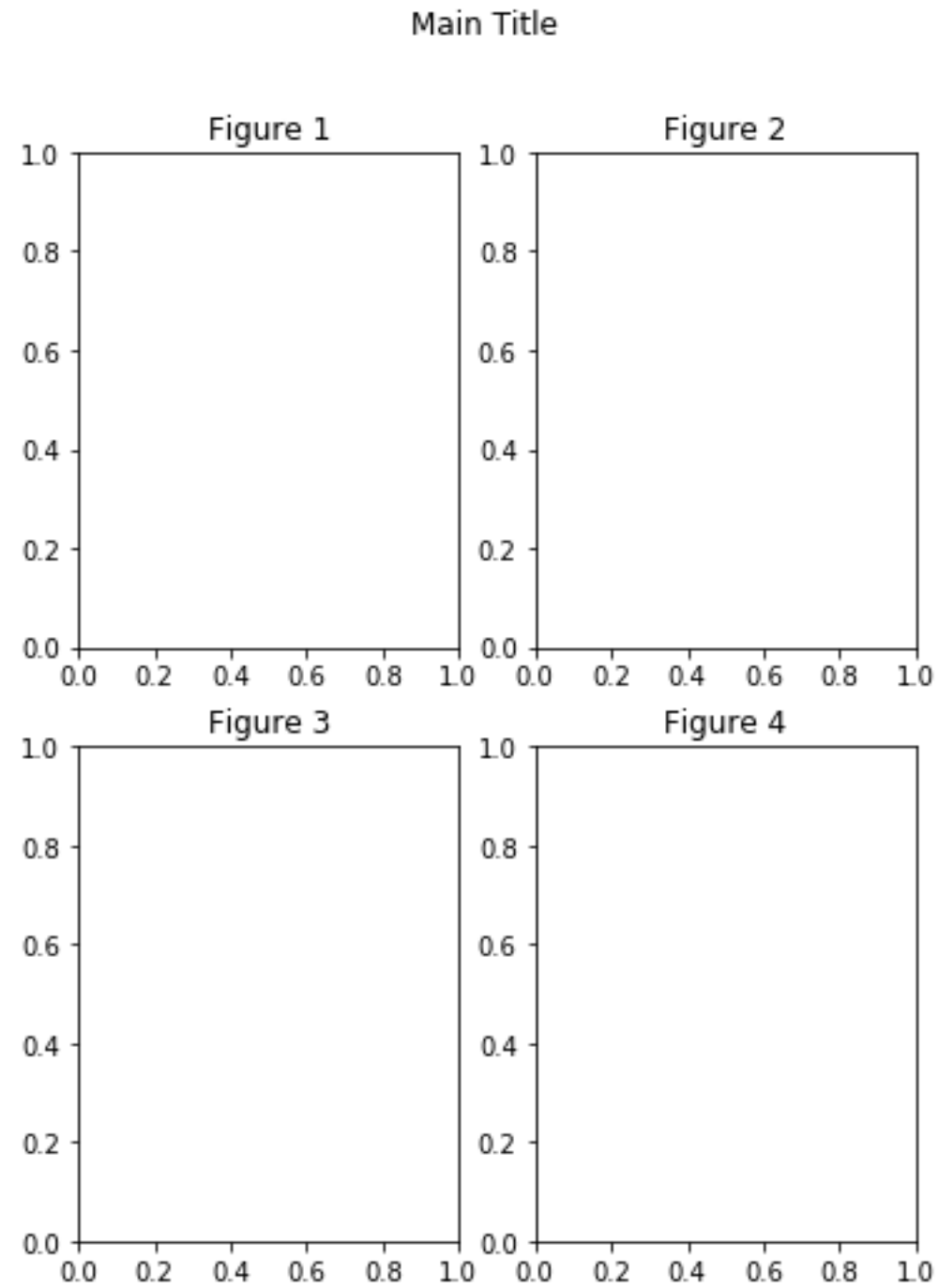
matplotlib

여러개 plot 한 번에 그리기

- `plt.subplots(nrows, ncols)`
 - `fig, axes = plt.subplots(2, 2, figsize=(6, 8))`
 - `fig.suptitle("Main Title")`
 - `axes[0][0].set_title("Figure 1")`
 - `axes[0][1].set_title("Figure 2")`
 - `axes[1][0].set_title("Figure 3")`
 - `axes[1][1].set_title("Figure 4")`
- `plt.show()`

matplotlib

여러개 plot 한 번에 그리기



matplotlib

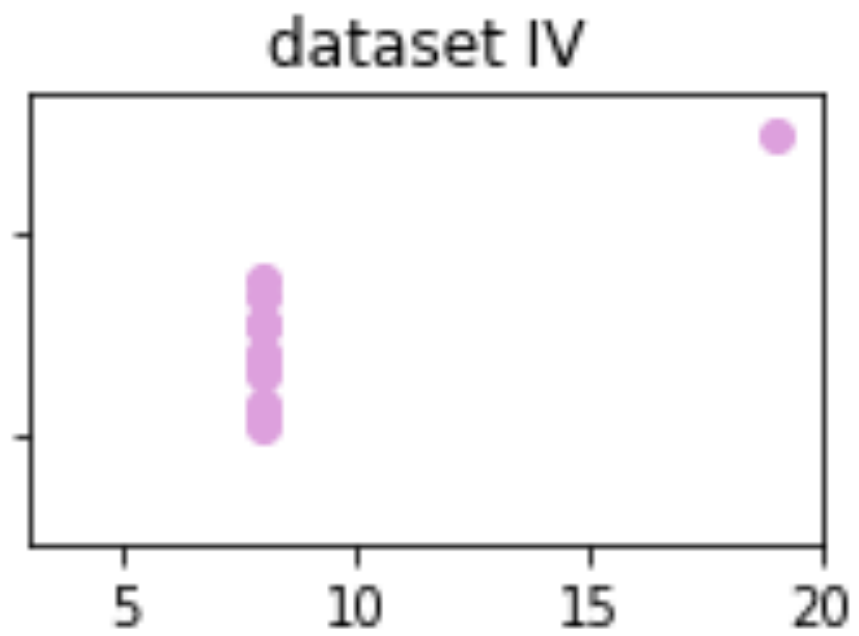
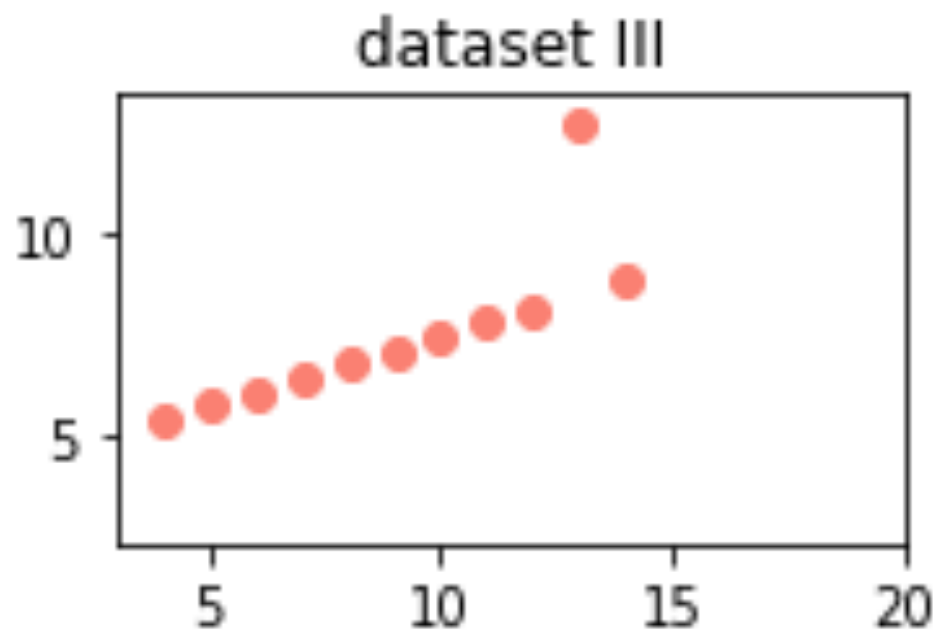
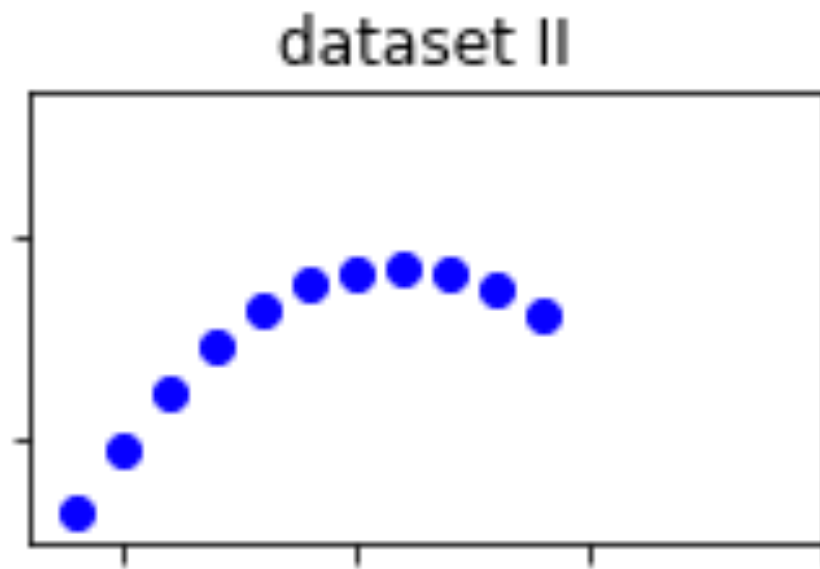
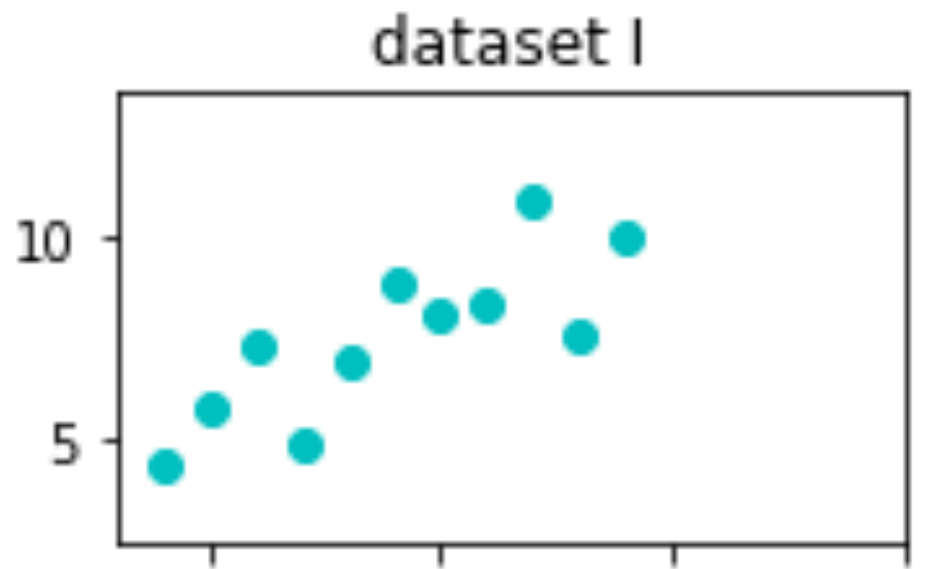
실전 연습

- `import seaborn as sns`
- `df = sns.load_dataset("anscombe")`
- `df.head(3) # dataset = ['I', 'II', 'III', 'IV']`

	dataset	x	y
0	I	10.0	8.04
1	I	8.0	6.95
2	I	13.0	7.58

matplotlib

실전 연습



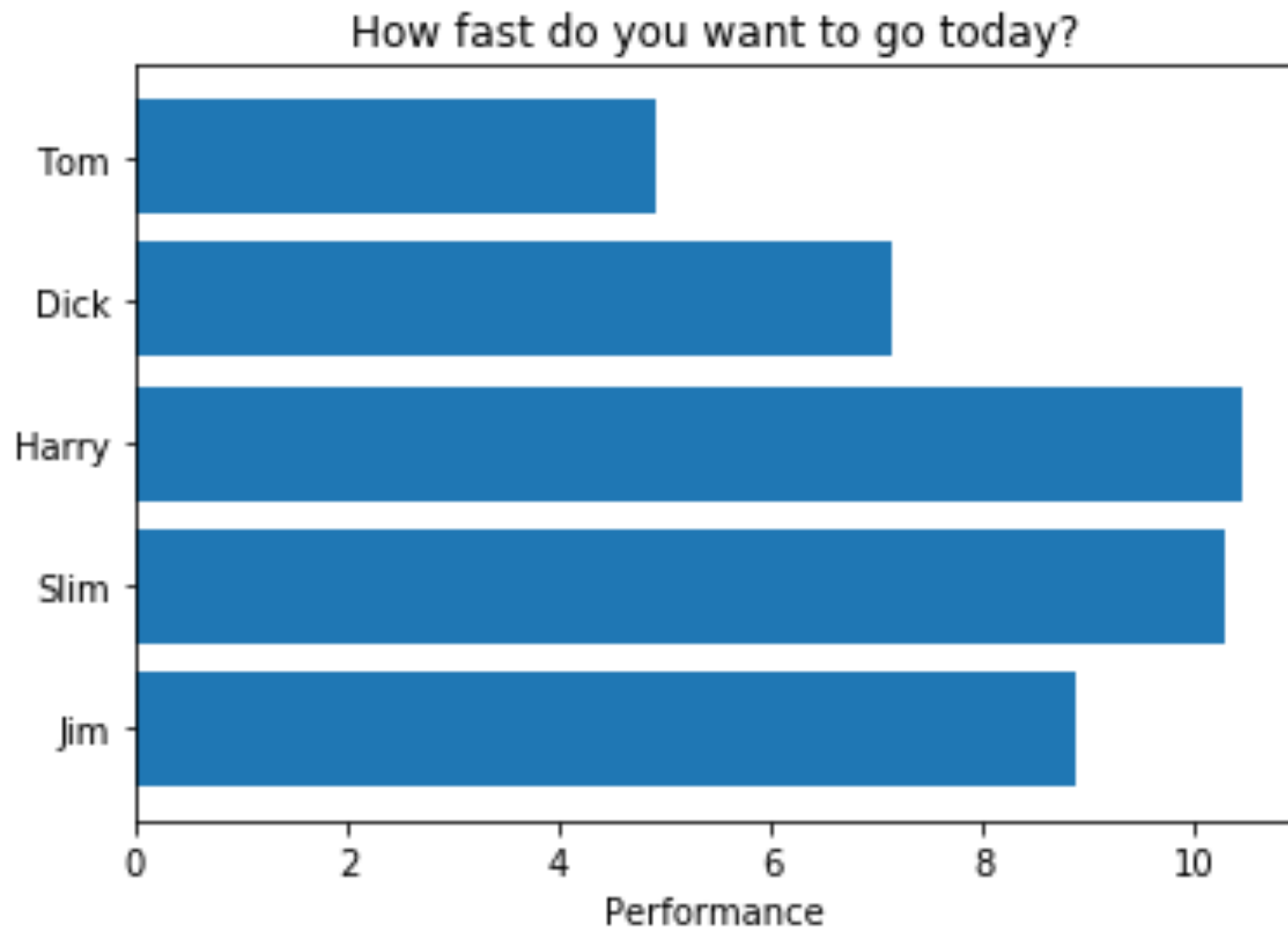
matplotlib

실전 연습

- `fig, axes = plt.subplots(2, 2, sharex=True, sharey=True)`
- `fig.subplots_adjust(hspace=0.5)`
- `axes[0][0].set_title("dataset I")`
- `axes[0][0].scatter(df.loc[df.dataset == "I", "x"], df.loc[df.dataset == "I", "y"], color="c")`
- `axes[0][1].set_title("dataset II")`
- `axes[0][1].scatter(df.loc[df.dataset == "II", "x"], df.loc[df.dataset == "II", "y"], color="b")`
- `axes[1][0].set_title("dataset III")`
- `axes[1][0].scatter(df.loc[df.dataset == "III", "x"], df.loc[df.dataset == "III", "y"], color="salmon")`
- `axes[1][1].set_title("dataset IV")`
- `axes[1][1].scatter(df.loc[df.dataset == "IV", "x"], df.loc[df.dataset == "IV", "y"], color="plum")`
- `plt.show()`

matplotlib

실전 연습



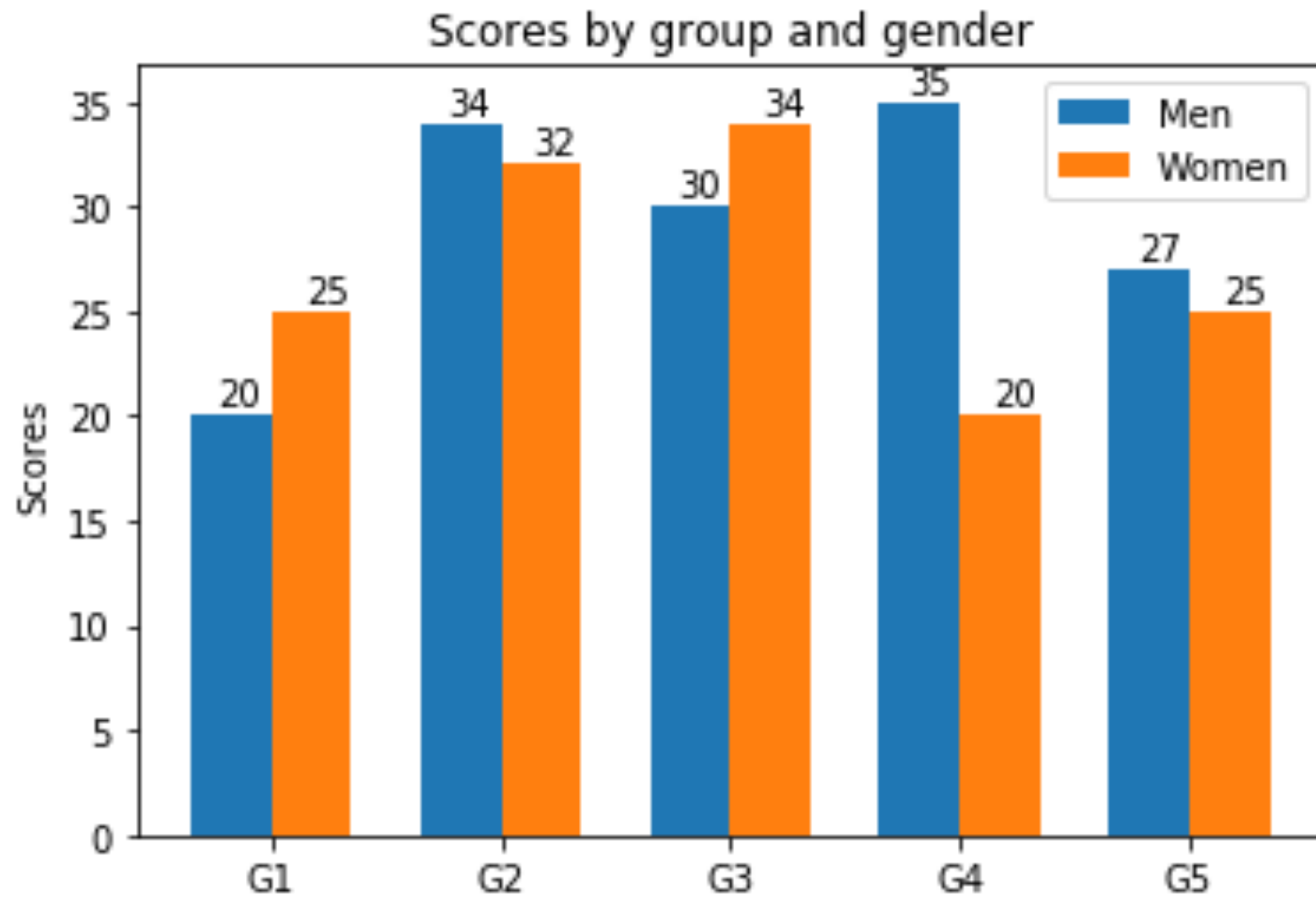
matplotlib

실전 연습

- `fig, ax = plt.subplots()`
- `# Example data`
- `people = ('Tom', 'Dick', 'Harry', 'Slim', 'Jim')`
- `y_pos = np.arange(len(people))`
- `performance = 3 + 10 * np.random.rand(len(people))`
- `ax.barh(y_pos, performance)`
- `ax.set_yticks(y_pos)`
- `ax.set_yticklabels(people)`
- `ax.invert_yaxis()` # labels read top-to-bottom
- `ax.set_xlabel('Performance')`
- `ax.set_title('How fast do you want to go today?')`
- `plt.show()`

matplotlib

실전 연습



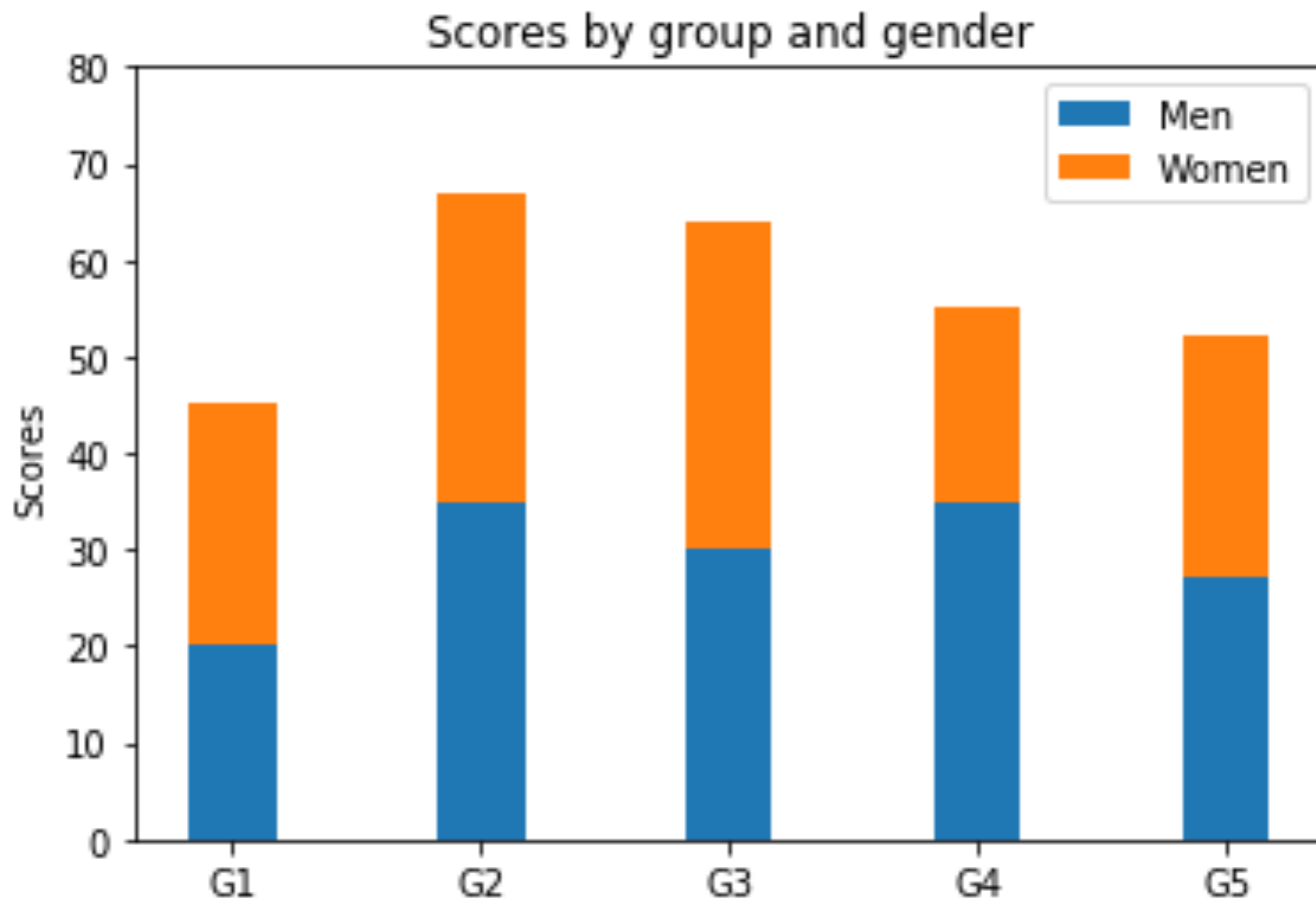
matplotlib

실전 연습

- `labels = ['G1', 'G2', 'G3', 'G4', 'G5']`
- `men_means = [20, 34, 30, 35, 27]`
- `women_means = [25, 32, 34, 20, 25]`
- `x = np.arange(len(labels)) # the label locations`
- `width = 0.35 # the width of the bars`
- `fig, ax = plt.subplots()`
- `ax.bar(x - width/2, men_means, width, label='Men')`
- `ax.bar(x + width/2, women_means, width, label='Women')`
- `ax.set_ylabel('Scores')`
- `ax.set_title('Scores by group and gender')`
- `ax.set_xticks(x)`
- `ax.set_xticklabels(labels)`
- `ax.legend()`
- `for i, label in enumerate(labels):`
 - `ax.annotate(men_means[i], xy=(x[i]-width/1.5, men_means[i]+0.5))`
 - `ax.annotate(women_means[i], xy=(x[i]+width/2.5, women_means[i]+0.5))`
- `plt.show()`

matplotlib

실전 연습



matplotlib

실전 연습

- `N = 5`
- `menMeans = (20, 35, 30, 35, 27)`
- `womenMeans = (25, 32, 34, 20, 25)`
- `x_pos = np.arange(N)` # the x locations for the groups
- `width = 0.35` # the width of the bars: can also be len(x) sequence
- `fig, ax = plt.subplots()`
- `ax.bar(x_pos, menMeans, width, label="Men")`
- `ax.bar(x_pos, womenMeans, width, bottom=menMeans, label="Women")`
- `ax.set_ylabel('Scores')`
- `ax.set_title('Scores by group and gender')`
- `ax.set_xticks(x_pos)`
- `ax.set_xticklabels(['G1', 'G2', 'G3', 'G4', 'G5'])`
- `ax.set_yticks(np.arange(0, 81, 10))`
- `ax.legend()`
- `plt.show()`

matplotlib을 넘어서

seaborn - heat map

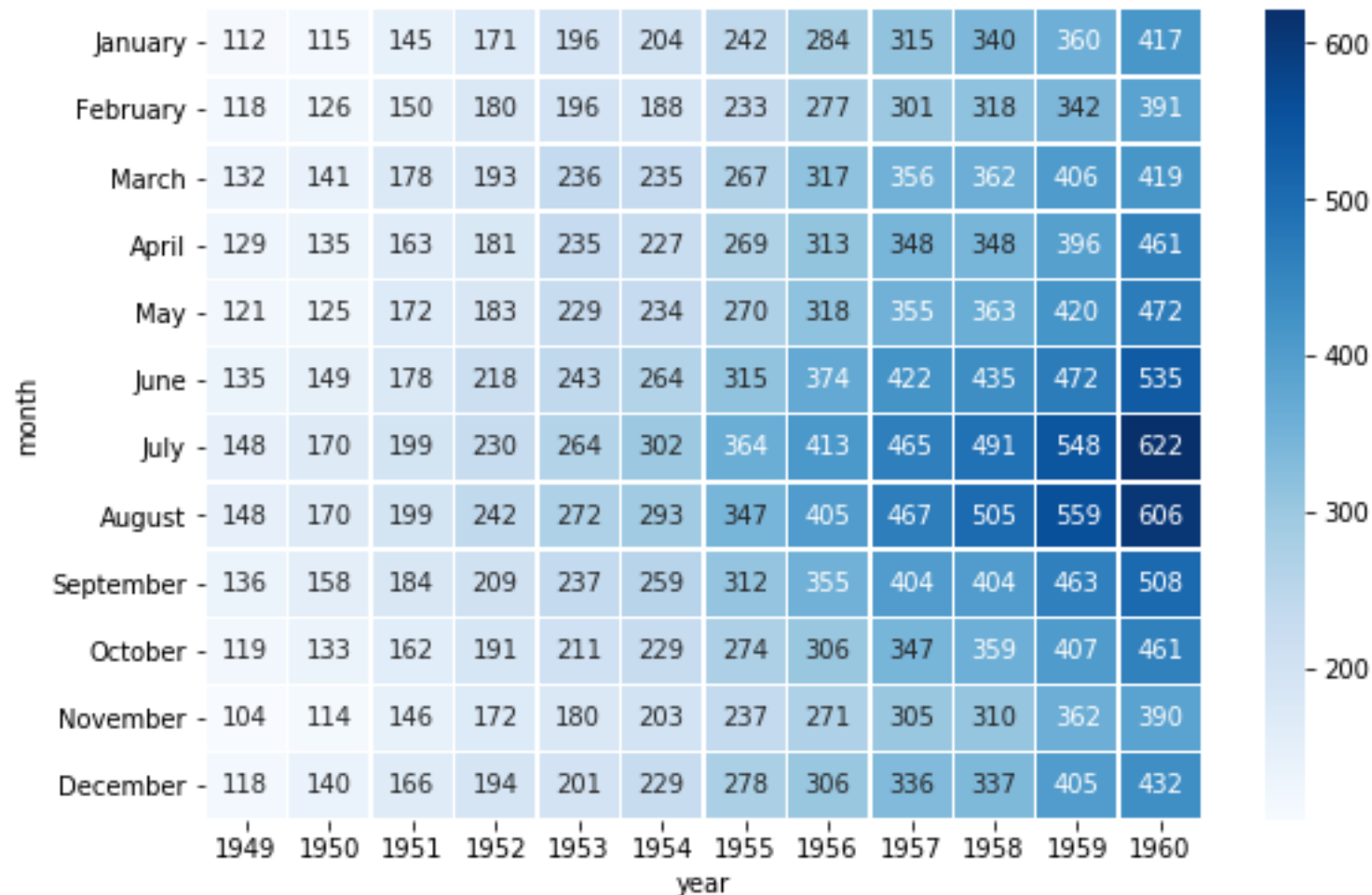
- import seaborn as sns
- flights_long = sns.load_dataset("flights")
- flights = flights_long.pivot("month", "year", "passengers")
- flights.head(3)

year	1949	1950	1951	1952	1953	1954	1955	1956	1957	1958	1959	1960
month												
January	112	115	145	171	196	204	242	284	315	340	360	417
February	118	126	150	180	196	188	233	277	301	318	342	391
March	132	141	178	193	236	235	267	317	356	362	406	419

matplotlib을 넘어서

seaborn - heat map

- `f, ax = plt.subplots(figsize=(9, 6))`
- `sns.heatmap(flights, annot=True, fmt="d", linewidths=.5, ax=ax, cmap="Blues")`
- `plt.show()`



matplotlib을 넘어서

seaborn - boxplot

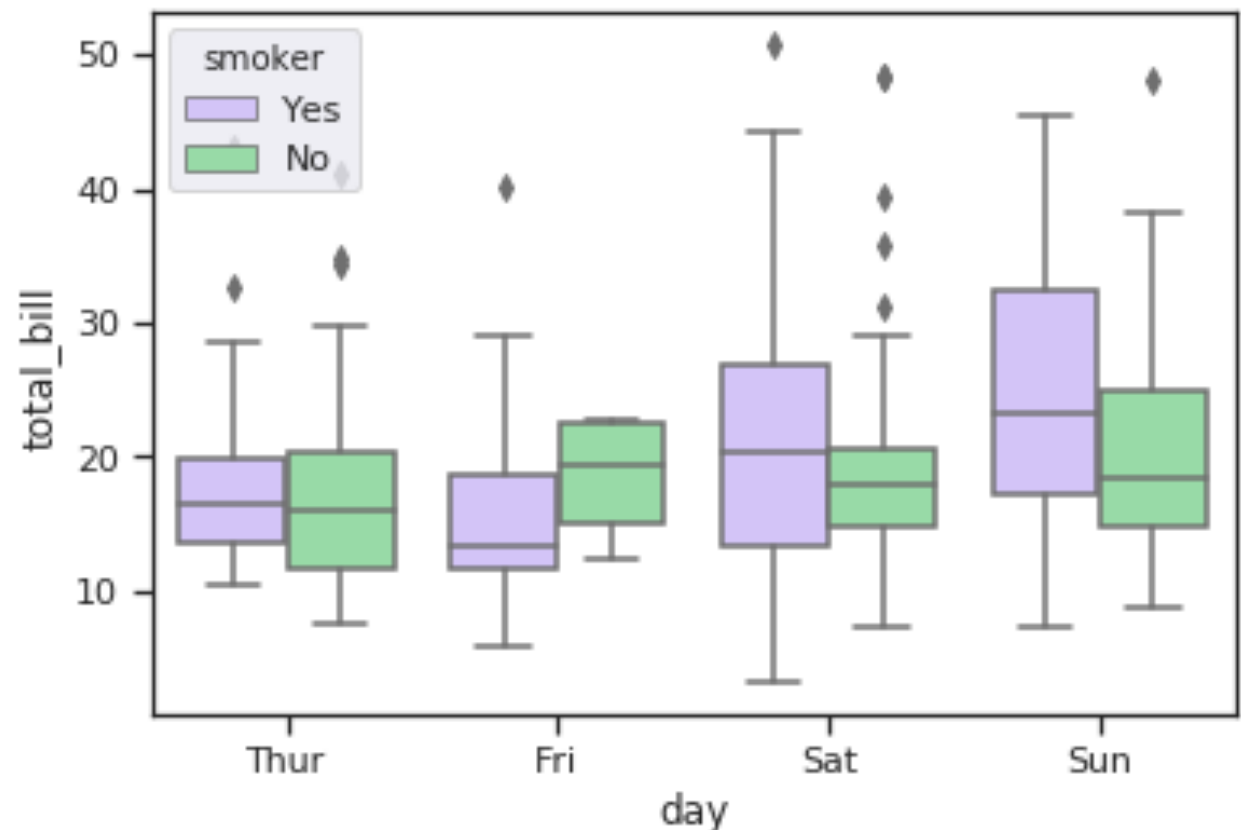
- import seaborn as sns
- tips = sns.load_dataset("tips")
- tips.head(3)

	total_bill	tip	sex	smoker	day	time	size
0	16.99	1.01	Female	No	Sun	Dinner	2
1	10.34	1.66	Male	No	Sun	Dinner	3
2	21.01	3.50	Male	No	Sun	Dinner	3

matplotlib을 넘어서

seaborn - boxplot

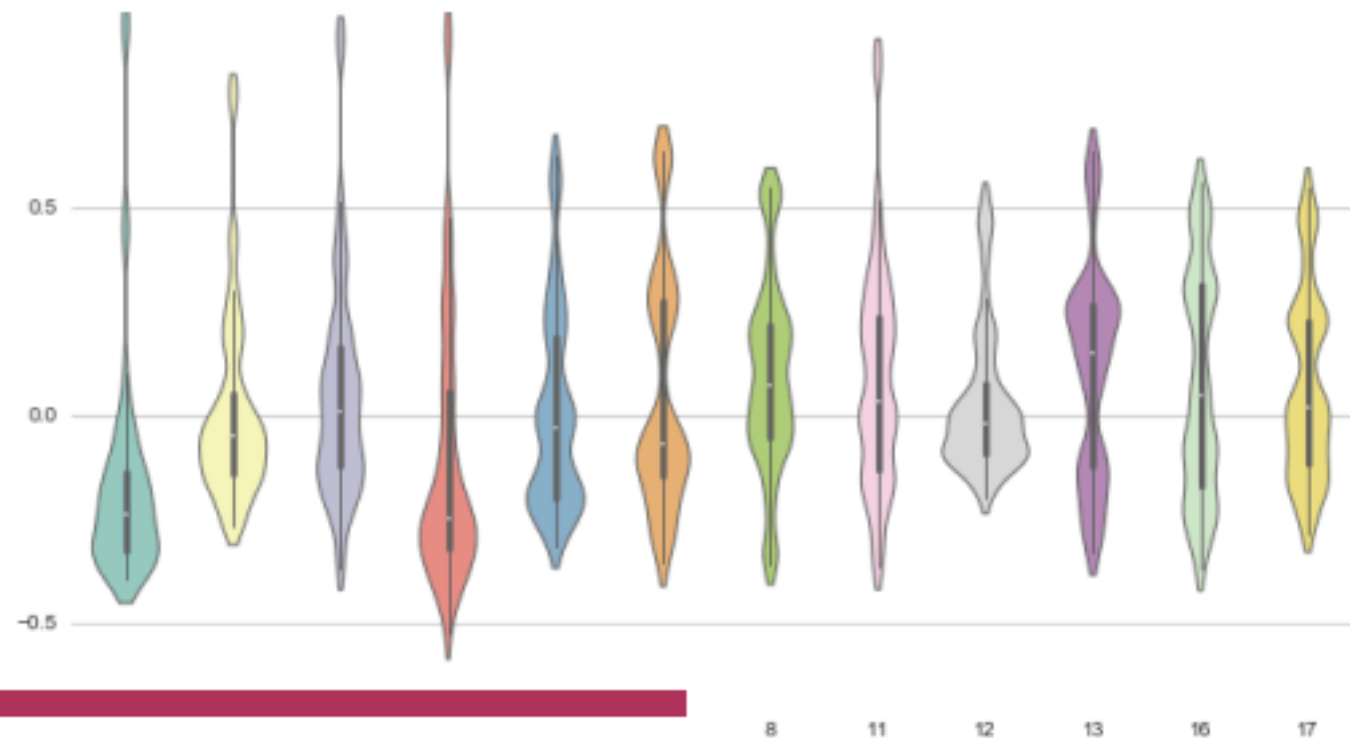
- `fig, ax = plt.subplots()`
- `sns.set(palette="pastel")`
- `sns.boxplot(x="day", y="total_bill", hue="smoker", palette=["m", "g"],`
- `data=tips, ax=ax)`
- `plt.show()`



matplotlib을 넘어서

그 밖에 다른 유용한 라이브러리

- <https://mode.com/blog/python-data-visualization-libraries>



Visualization

June 8, 2016 • 8 minute read

10 Useful Python Data Visualization Libraries for Any Discipline

MELISSA BIERLY

CONTENT MARKETING

matplotlib

연습

- 문제

- 다음 그래프를 직접 그려보자
- 색깔
 - black, blue, plum

