

Big Data

matplotlib

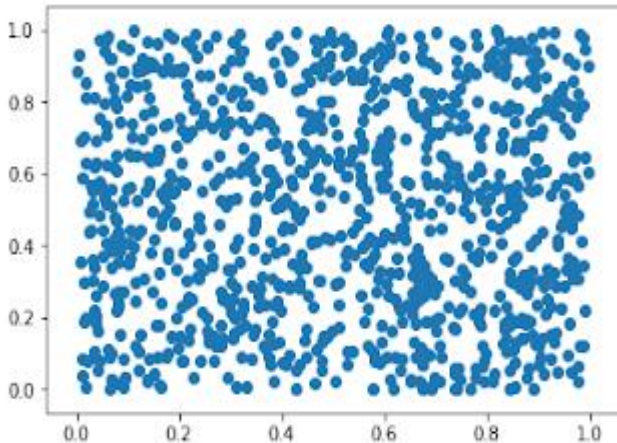
데이터 시각화란?

- 광범위하게 분산된 방대한 양의 자료를 한눈에 볼 수 있도록 도표나 차트 등으로 정리하는 것
- 시각화를 통해 데이터의 특징을 쉽게 파악 할 수 있다
- 분석 결과를 상대방에게 효과적으로 전달가능하다

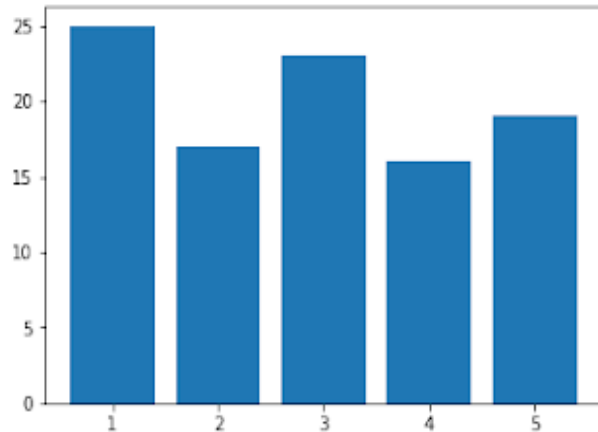
학습목표

1. 시각화 할 수 있는 차트의 종류를 안다.
2. matplotlib을 이용하여 시각화를 할 수 있다.

차트의 종류

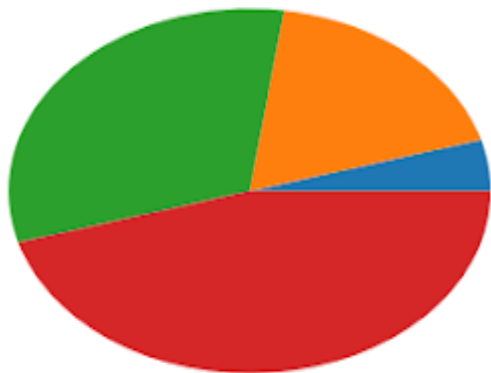


scatter plot

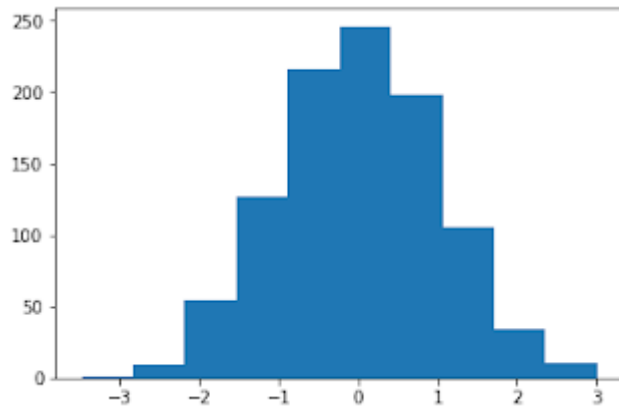


bar chart

차트의 종류



pie chart



histogram

matplotlib

- Python에서 시각화 할 수 있는 대표적인 패키지
- pyplot과 pylab이란 sub패키지가 있다.

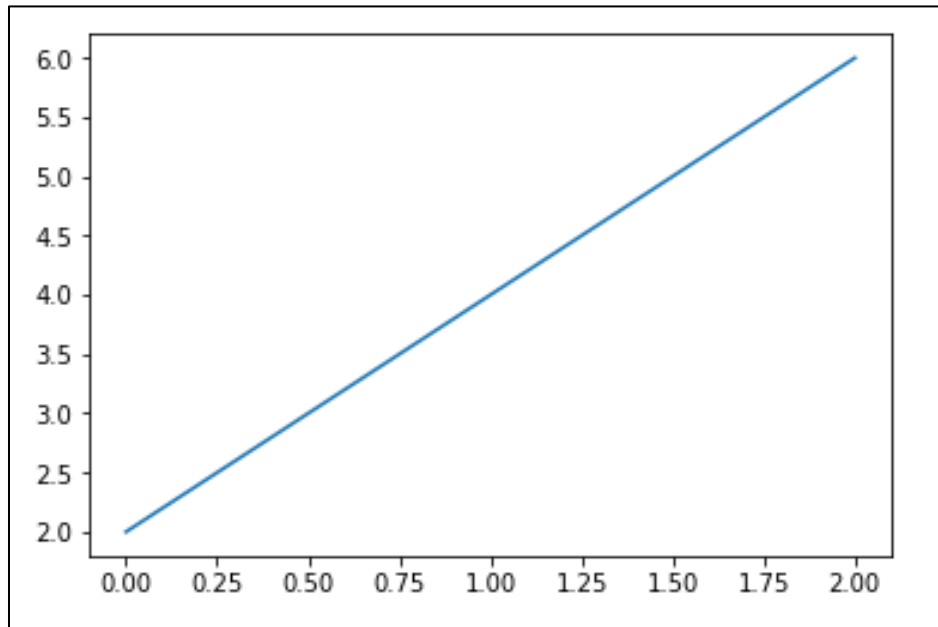
	pyplot	pylab
기능	시각화	시각화 + numpy
특징	비대화형 (간단한 정보만 입력해서 플롯이 된다.)	대화형 (비교적 많은 정보를 입력해 플롯을 요구한다.)

matplotlib 사용하기

➤ `import matplotlib.pyplot as plt`

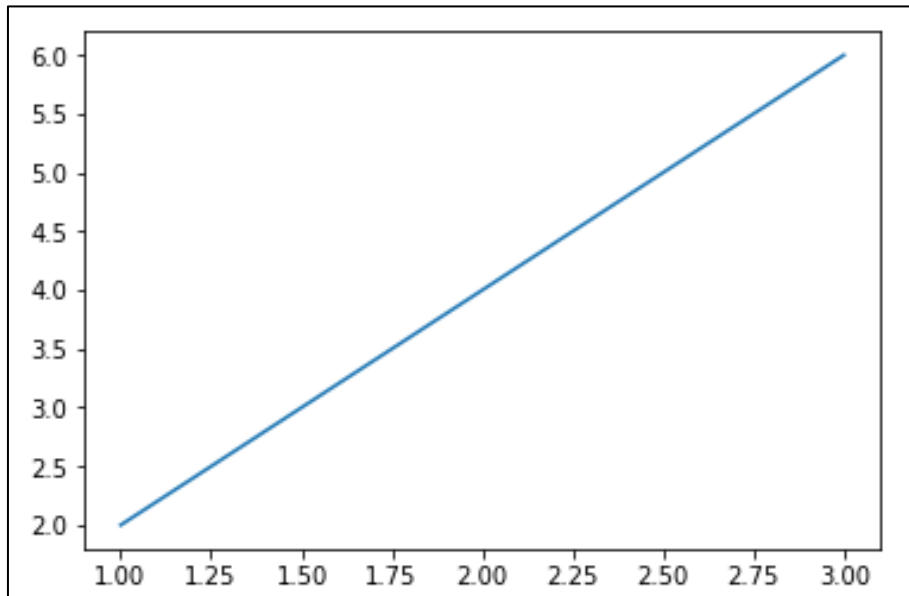
Line plot

- `y=[2,4,6]`
- `plt.plot(y)`
- `plt.show()`



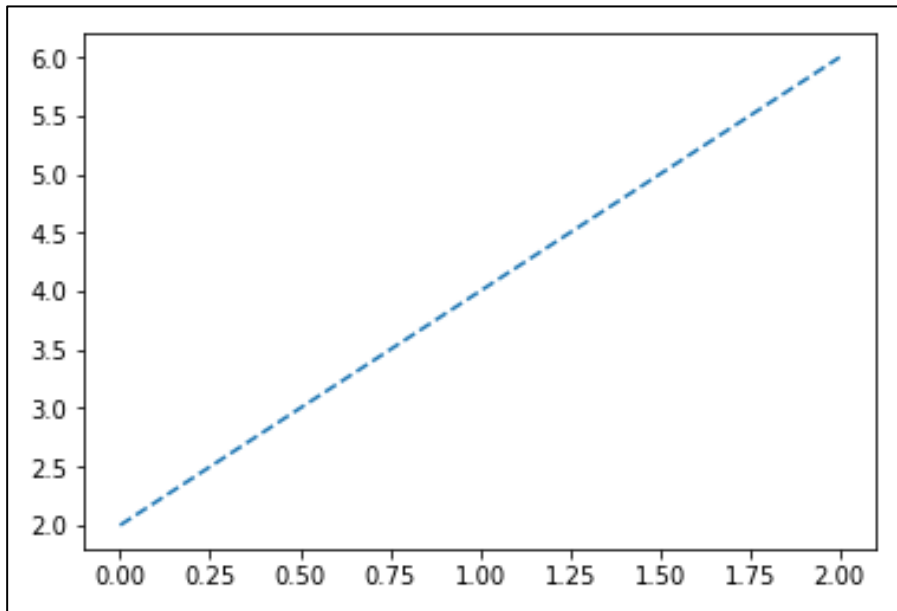
Line plot

- `x=[1,2,3]`
- `y=[2,4,6]`
- `plt.plot(x, y)`
- `plt.show()`



Line style

- `plt.plot(y, ls = '--')`
- `plt.show()`

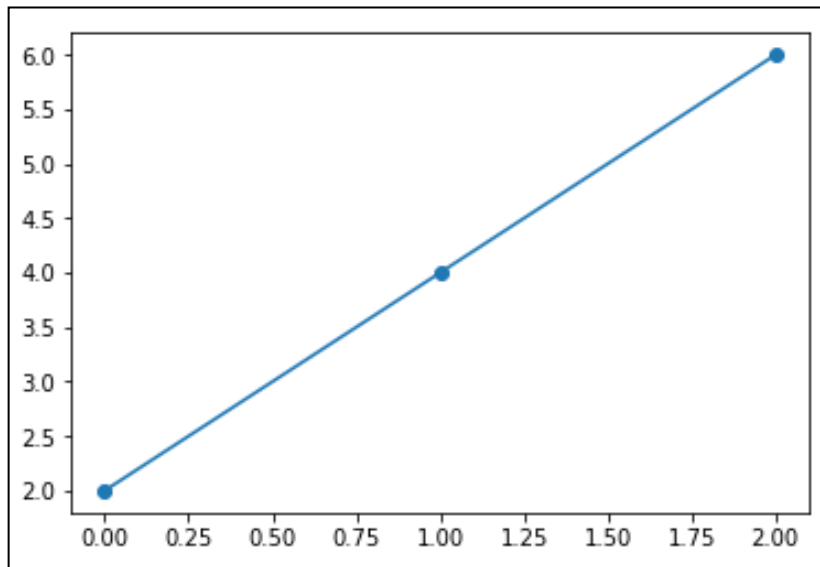


Line style

Character	Description
'-'	Solid line style
'--'	Dashed line style
'-.'	Dash-dot line style
'...'	Dotted line style

Marker

- `plt.plot(y, marker = 'o')`
- `plt.show()`

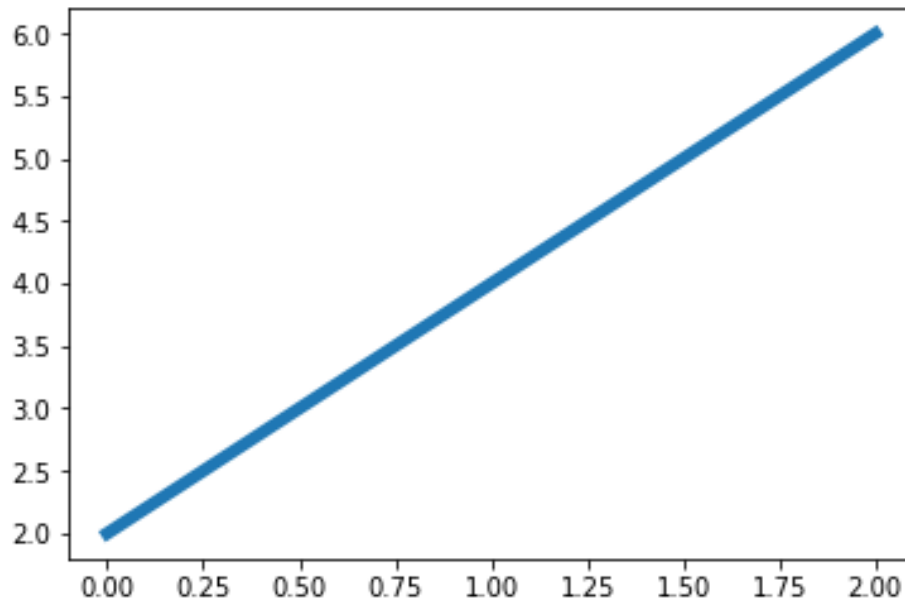


Marker

Character	Description	Character	Description
'.'	Point marker	'1'	Tri_down marker
'o'	Circle marker	'2'	Tri_up marker
'v'	Triangle_down marker	'3'	Tri_left marker
'^'	Triangle_up marker	'4'	Tri_right marker
'<'	Triangle_left marker	'*'	Star marker
'>'	Triangle_right marker	'h','H'	Hexagon1,2 marker
's'	Square marker	'+'	Plus marker
'p'	Pentagon marker	'D'	Diamond marker
' '	Vline marker	'_'	Hline marker

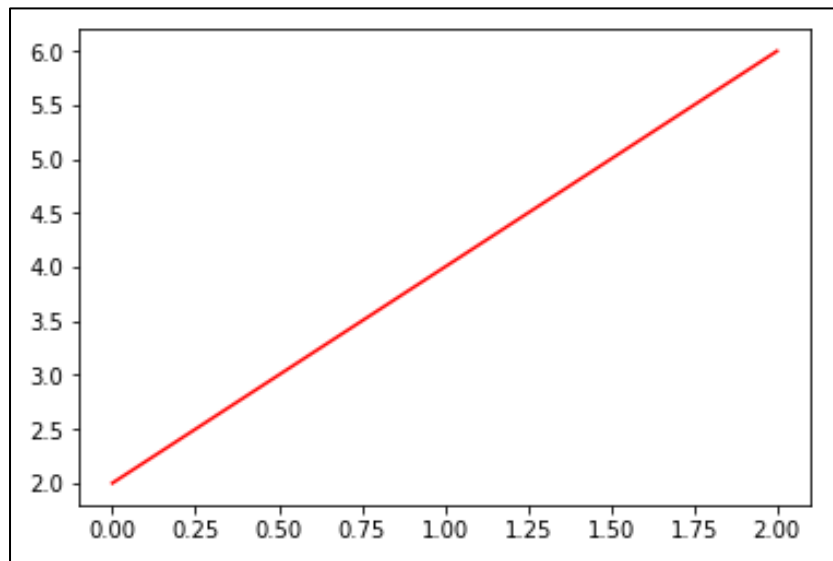
Line weight

- `plt.plot(y, lw = 5)`
- `plt.show()`



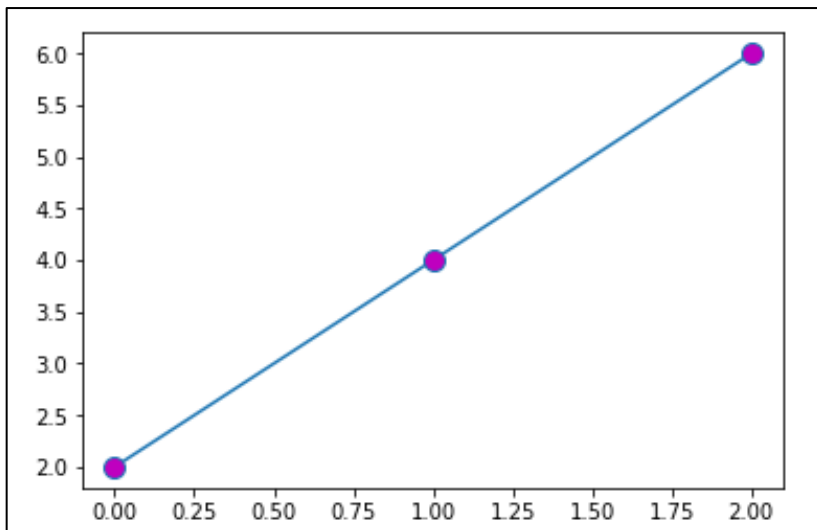
Line color

- `plt.plot(y, color = ' r ')`
- `plt.show()`



Line color

- `plt.plot(y, markerfacecolor = ' m ', markersize=10)`
- `plt.show()`



Line color

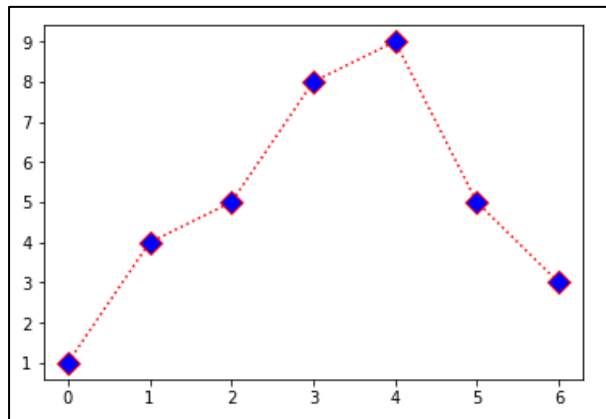
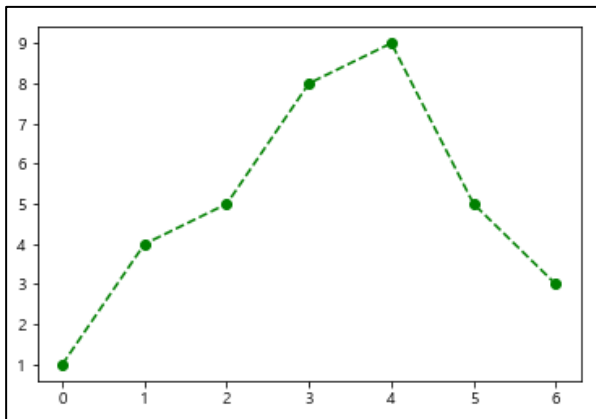
Character	Color	Character	Color
'b'	Blue	'm'	Magenta
'g'	Green	'y'	yellow
'r'	Red	'k'	Black
'c'	cyan	'w'	white

스타일 옵션들

스타일옵션종류	마커 종류정의	약자
color	선 색깔	c
linewidth	선 굵기	lw
linestyle	선 스타일	ls
marker	마커 종류	
markersize	마커 크기	ms
markeredgecolor	마커 선 색깔	mec
markeredgewidth	마커 선 굵기	mew
markerfacecolor	마커 내부 색깔	mfc

연습문제)

- `x = np.arange(7)`
- `y = [1,4,5,8,9,5,3]`



연습문제)

- `x = np.arange(7)`
- `y = [1,4,5,8,9,5,3]`
- `plt.plot(x, y, ls = '--', lw=5, marker='o', ms=15, mec='g', mew=5, mfc='r')`
- `plt.show()`

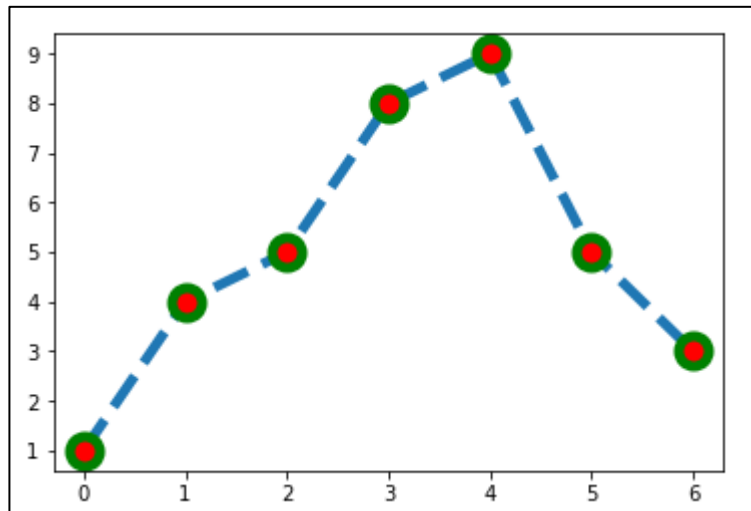
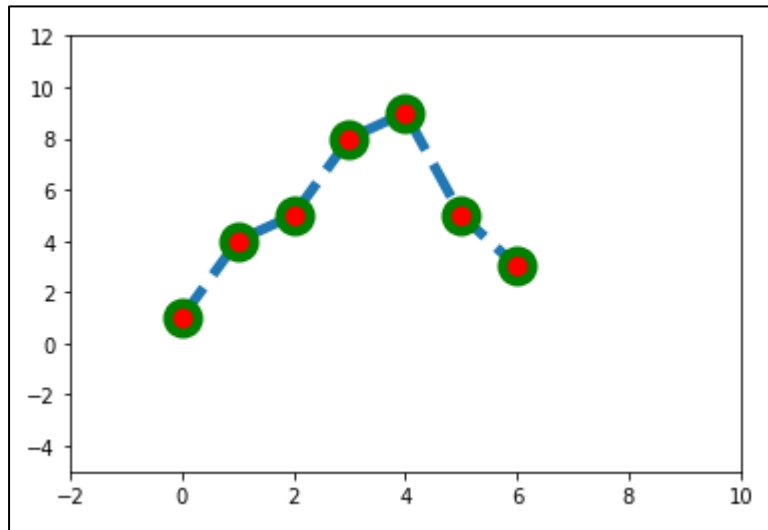


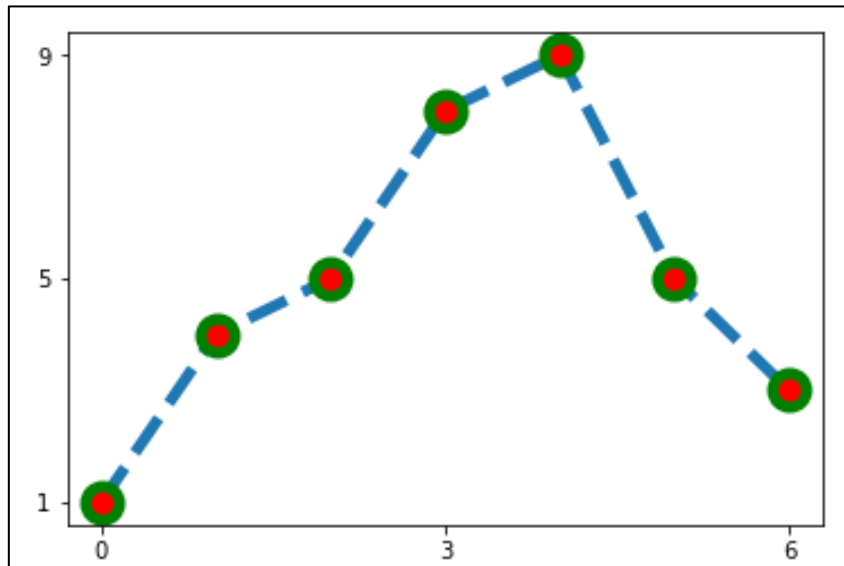
그림 범위 지정

- `plt.xlim(-2,10)`
- `plt.ylim(-5,12)`



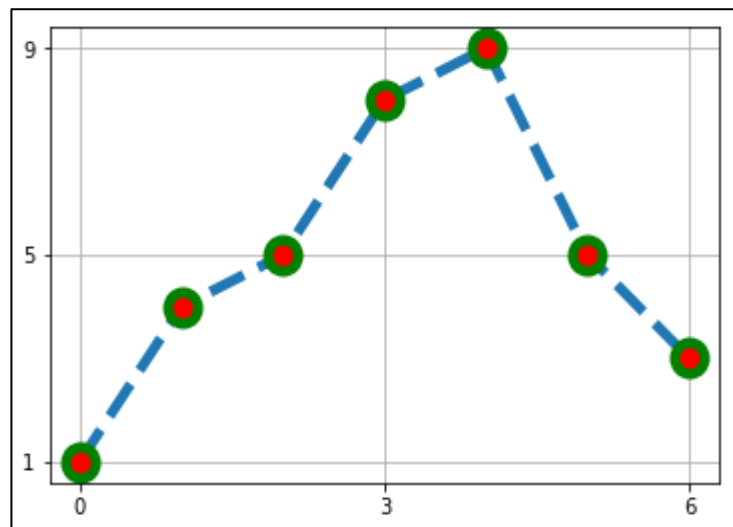
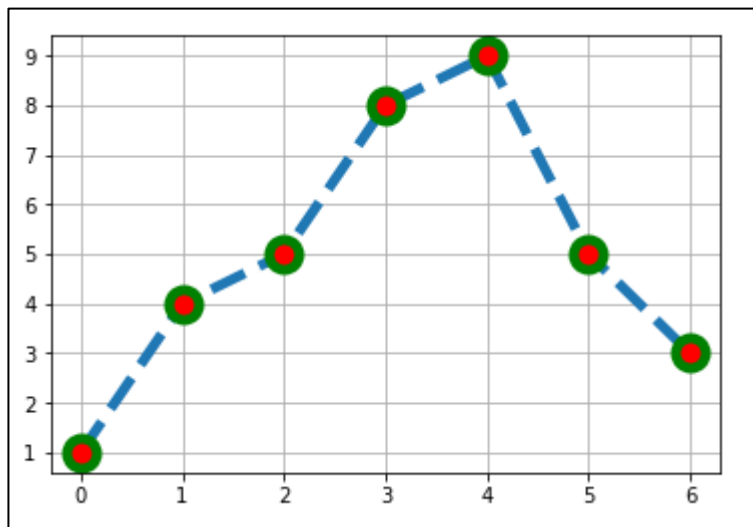
틱 설정

- `plt.xticks([0,3,6])`
- `plt.yticks([1,5,9])`



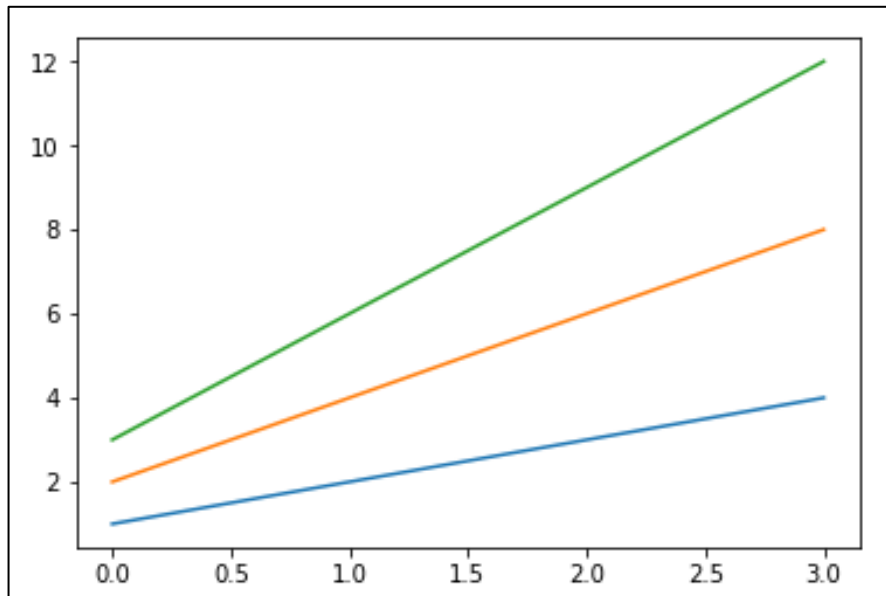
그리드

➤ `plt.grid()`



여러 개 플롯

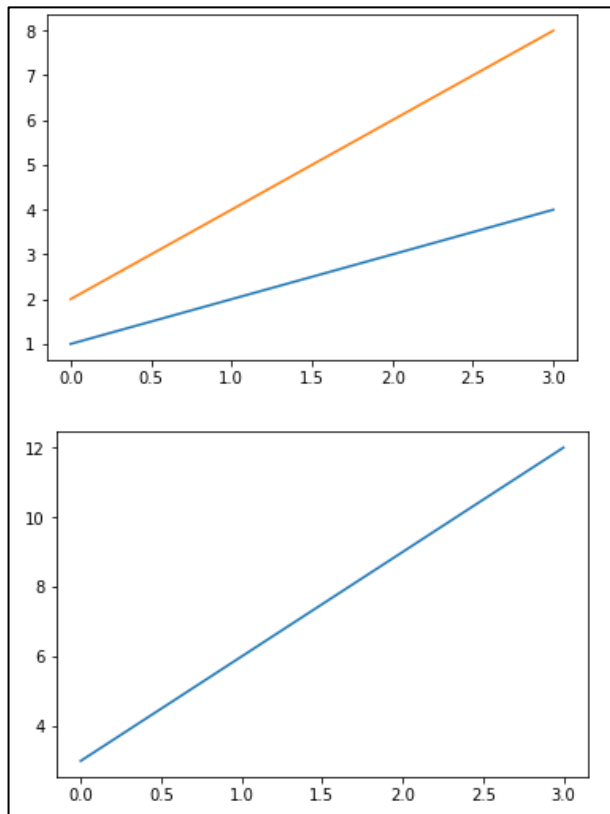
- `x = [1,2,3,4]`
- `y = [2,4,6,8]`
- `z = [3,6,9,12]`
- `plt.plot(x)`
- `plt.plot(y)`
- `plt.plot(z)`
- `plt.show()`



여러 개 플롯

- `plt.plot(x)`
- `plt.plot(y)`
- `plt.show()`

- `plt.plot(z)`
- `plt.show()`



범례 표시

- `plt.plot(x, label = 'x')`
- `plt.plot(y, label = 'y')`
- `plt.plot(z, label = 'z')`

- `plt.legend()`
- `plt.show()`

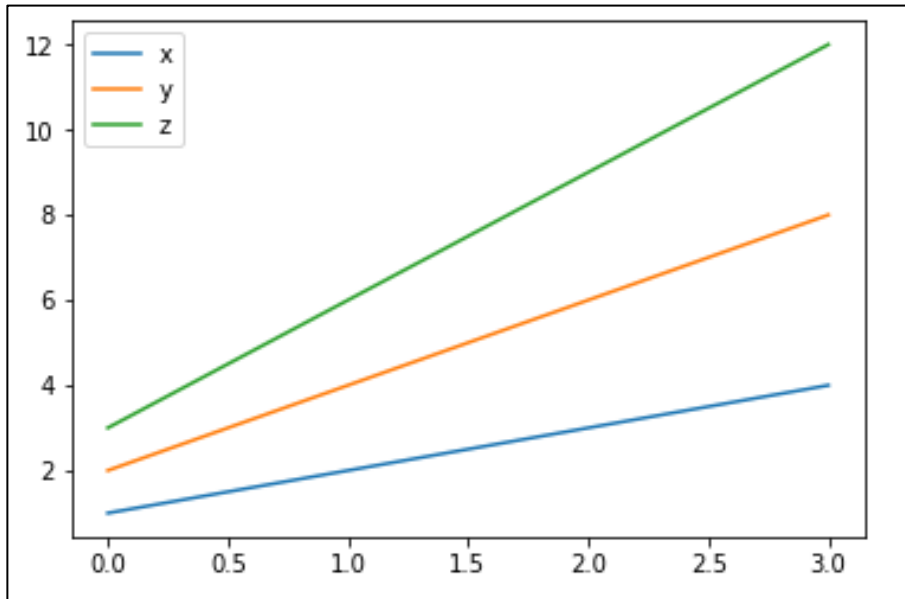


차트 옵션

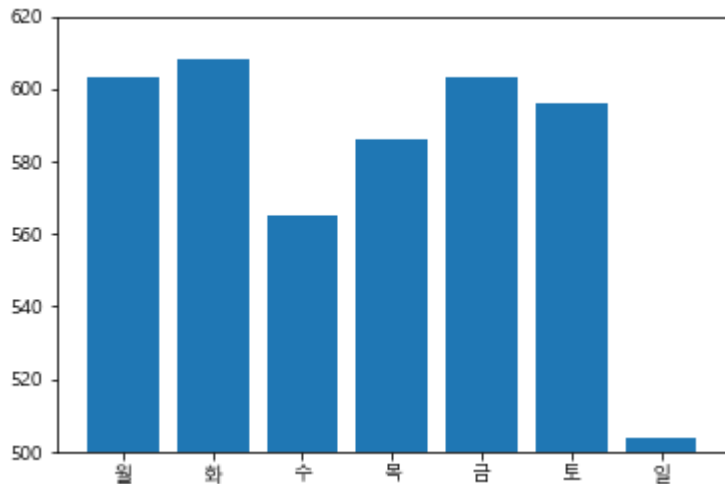
plot 옵션	정의
xlim, ylim	x 축 범위, y축 범위
grid	격자눈금
legend	범례
xlabel, ylabel	x축 타이틀, y축 타이틀
title	그래프 제목
xticks, yticks	x축 눈금 조정, y축 눈금 조정

차트 한글 보이기

```
from matplotlib import font_manager, rc  
  
font_name = font_manager.FontProperties(fname="c:/Windows/Fonts/malgun.ttf").get_name()  
rc('font', family=font_name)
```

공공데이터 : 전국사망교통사고 2017을 시각화 해보자.

➤ 요일별 사망교통사고 시각화



공공데이터 : 전국사망교통사고 2017을 시각화 해보자.

➤ 데이터 로드

```
import pandas as pd
```

```
data = pd.read_csv("Traffic_Accident_2017.csv", encoding="euc-kr")
```

공공데이터 : 전국사망교통사고 2017을 시각화 해보자.

➤ 요일별 사고 건수 count

```
1 temp = data["요일"].value_counts()
2 y = temp[["월", "화", "수", "목", "금", "토", "일"]]
3 y
```

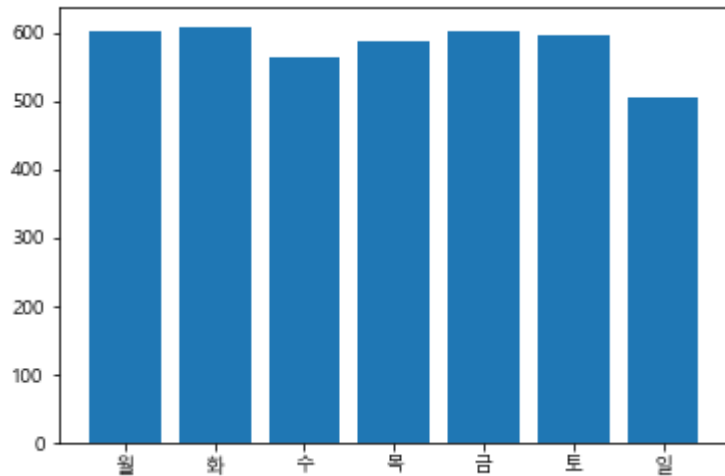
월	603
화	608
수	565
목	586
금	603
토	596
일	504

Name: 요일, dtype: int64

공공데이터 : 전국사망교통사고 2017을 시각화 해보자.

➤ bar chart 그리기

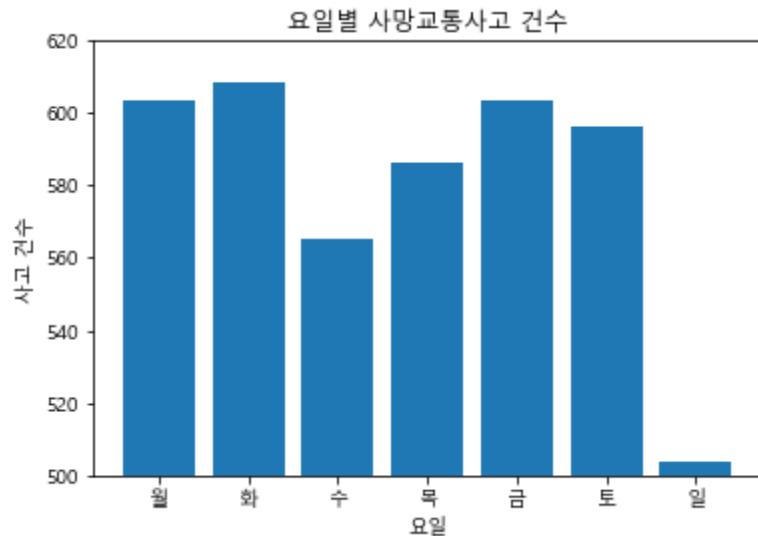
```
x = range(7)
plt.bar(x,y)
plt.xticks(x,["월","화","수","목","금","토","일"])
plt.show()
```



공공데이터 : 전국사망교통사고 2017을 시각화 해보자.

➤ bar chart 그리기

```
x = range(7)
plt.bar(x,y)
plt.xticks(x,["월","화","수","목","금","토","일"])
plt.ylim(500,620)
plt.xlabel("요일")
plt.ylabel("사고 건수")
plt.title("요일별 사망교통사고 건수")
plt.show()
```



수고하셨습니다.