

Big Data 

Matplotlib

Enjoy your data analysis

강사 김은영



학습목표

1. 시각화 할 수 있는 차트의 종류를 알 수 있다.
2. Matplotlib을 이용하여 데이터를 시각화 할 수 있다.

데이터 시각화란?

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

Matplotlib

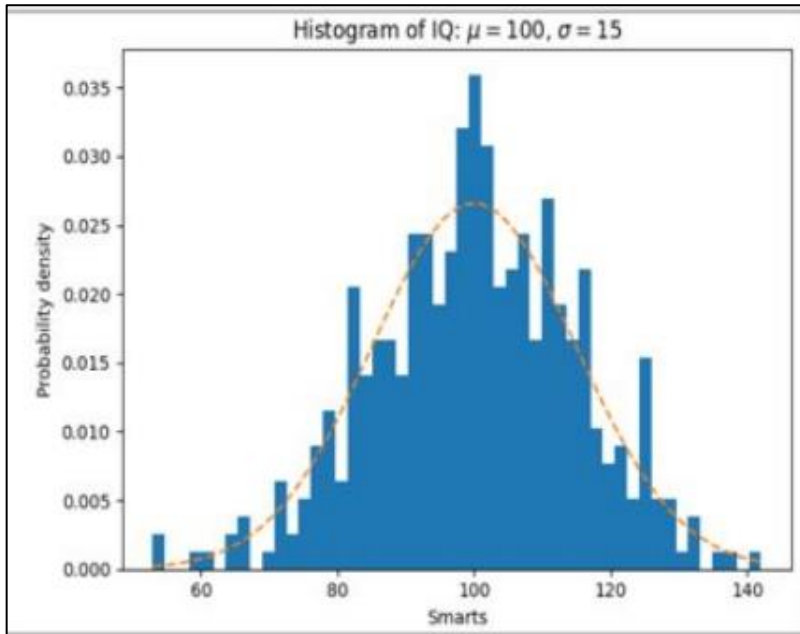
Enjoy your data analysis

기종	년월	상권 코드	상권 코드	중 유동인	남성 유동	여성 유동	연령대_10	연령대_20	연령대_30	연령대_40	연령대_50	연령대_60	시간대_1	시간대_2	시간대_3	시간대_4	시간대_5	시간대_6
1	201810	1744	회무정로1	78736	36725	42011	3479	23194	21866	13353	9456	7388	14050	12060	11338	12975	18614	9699
2	201810	1743	회무정로3길	208177	105723	102453	16490	115942	27240	17922	16829	13755	27389	26504	35324	38543	54123	26294
3	201810	1742	회무정로2길	120515	62362	58153	11978	70973	13749	9723	8428	5663	16687	13564	20086	22201	32313	15664
4	201810	1741	회무정로23길	21067	11373	9693	269	4446	4489	3702	4482	3678	4414	4037	3231	2788	4004	2592
5	201810	1740	회무정로14길	18377	10004	8374	488	6372	3099	2829	2907	2682	3903	3025	2531	2559	4000	2359
6	201810	1739	회무정로11길	8062	4490	3572	335	3180	1319	983	1105	1139	1184	1108	1156	1231	2150	1233
7	201810	1738	회무정로10길	14772	8462	6309	1487	7638	1913	1479	1332	922	1727	1307	2023	2579	4598	2538
8	201810	1737	후암로57길	36888	20949	15939	733	10550	9529	6793	5586	3697	1800	6476	9424	7942	8592	2654
9	201810	1736	후암로4길	32179	16016	16163	2236	7500	7644	5470	4615	4714	4410	7462	4792	4685	6371	4460
10	201810	1735	후암로35길	14054	7353	6700	547	2998	3177	2754	2456	2121	3143	2077	2495	1939	2952	1447
11	201810	1734	후암로30길	14143	7753	6390	477	3202	3411	2862	2405	1786	2436	2122	2499	2494	3133	1458
12	201810	1733	후암로23길	11459	5787	5671	489	2340	2469	2310	1961	1889	2632	1842	1839	1548	2342	1257
13	201810	1732	후암로13길	13264	7141	6123	1106	2798	2938	2834	1922	1666	1647	2301	2601	2188	3146	1380
14	201810	1731	효창원로9	103291	49963	53328	7518	26029	19906	18653	18073	13112	17226	21545	16503	16456	19809	11752
15	201810	1730	효창원로9	45433	12194	33238	5117	21255	4464	4912	5364	4321	4500	7658	12153	11273	7399	2450
16	201810	1729	효창원로9	72294	28879	43416	4166	24262	13195	11349	11701	7621	15225	12812	14202	13455	11532	5067
17	201810	1728	효창원로8	25508	13481	12027	2153	4902	5830	5059	4009	3555	3976	5403	4520	4230	4792	2586
18	201810	1727	효창원로7	30280	13283	16997	2663	8800	5299	5114	4581	3823	4283	5727	5979	5473	5916	2900
19	201810	1726	효창원로6	18139	8503	9636	1083	3844	3716	3603	2929	2963	3775	3933	3142	2745	2904	1641
20	201810	1725	효창원로6	19930	10979	8950	1099	4554	4518	3112	3441	3207	2288	4766	4002	3545	3629	1700
21	201810	1724	효창원로3	41896	20979	20917	1830	7885	9506	7883	7780	7012	7700	7392	5978	6536	9074	5215
22	201810	1723	효창원로74길	25161	13463	11698	810	4886	6402	5826	4069	3168	3977	5046	4972	4460	4494	2211
23	201810	1722	효창원로68길	19906	11006	8900	174	3609	5839	4435	3167	2682	1612	3903	4821	3916	4153	1500
24	201810	1721	효창원로67길	51149	26749	24400	737	11648	13996	11012	7255	6502	6136	10622	10760	9693	9580	4358
25	201810	1720	효창원로40길	16914	9673	7241	3273	2786	2975	3143	2518	2218	1573	3337	4068	3159	3251	1526
26	201810	1719	효창원로34길	34159	17027	17132	2600	8567	6604	5305	5344	5739	3352	5740	7418	6835	7517	3297
27	201810	1718	효창원로29길	55361	27610	27751	4470	14961	11662	9477	7842	6949	7793	9443	10925	9972	11510	5719
28	201810	1717	회나무로4	96187	46335	49853	2196	43305	25298	12893	7849	4646	19159	8808	10698	13581	25777	18165
29	201810	1716	회나무로1	32032	16130	15901	50	9712	10360	5115	3811	2984	3430	5199	5958	6654	7355	3435

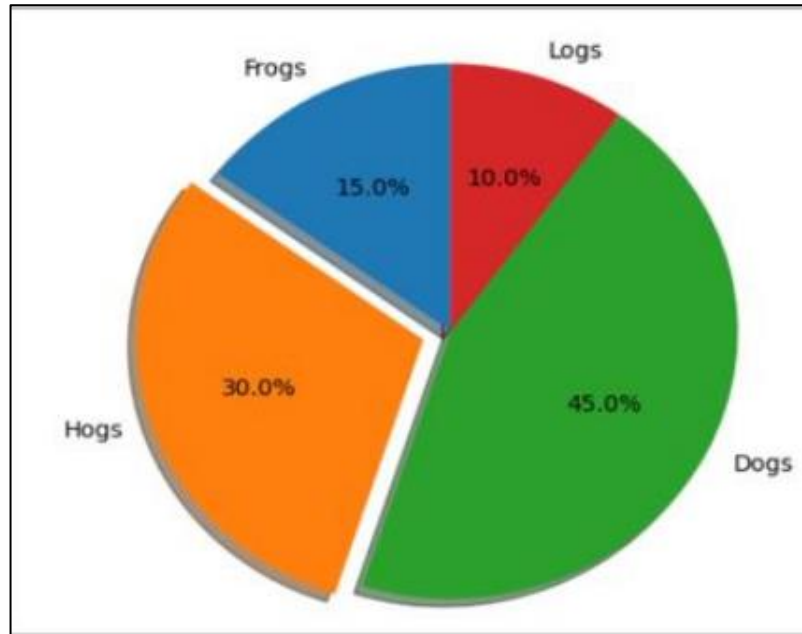
서울시 우리마을가게 상권분석서비스(상권-추정유동인구)



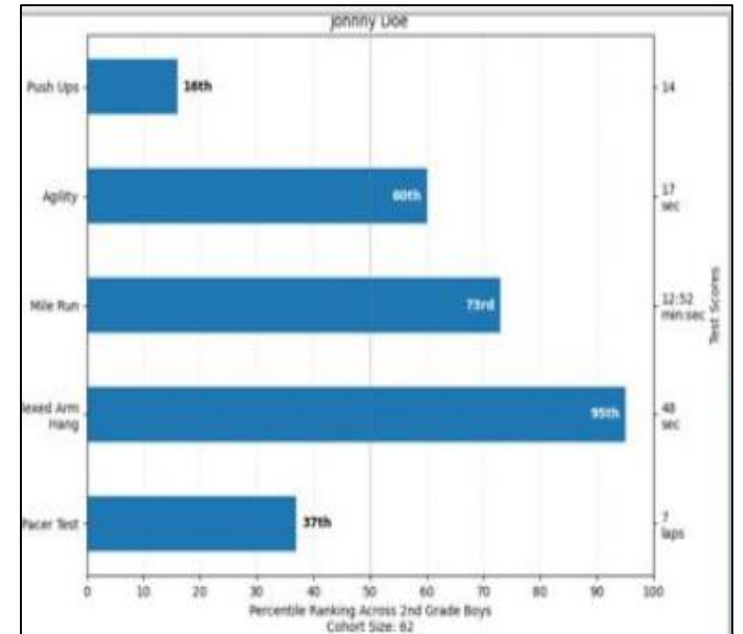
차트의 종류 - 1차원



Histogram

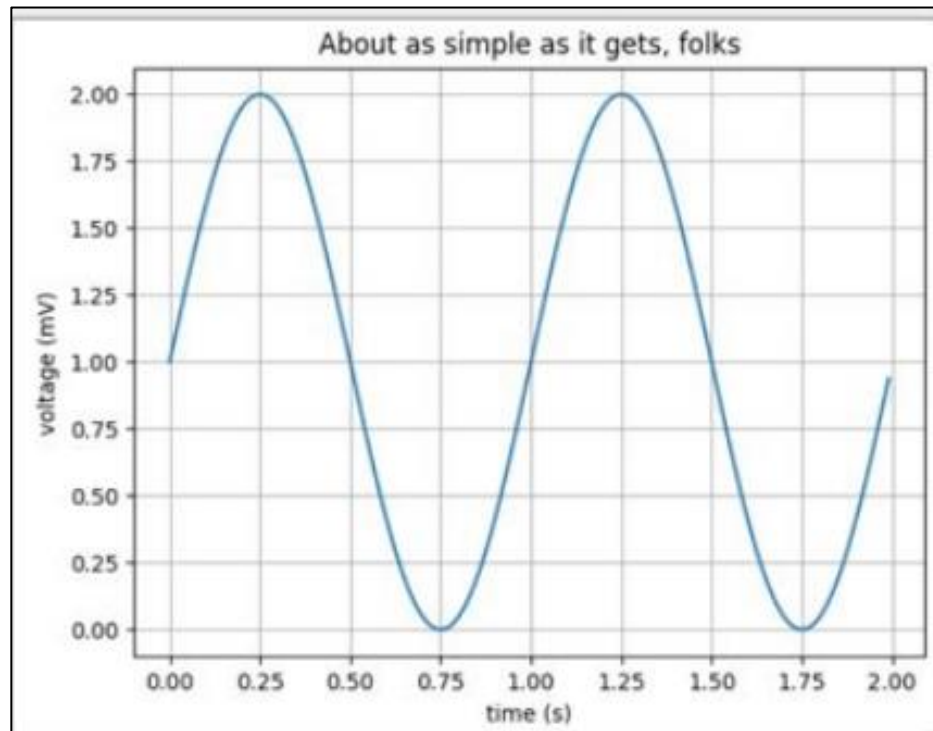


Pie Chart

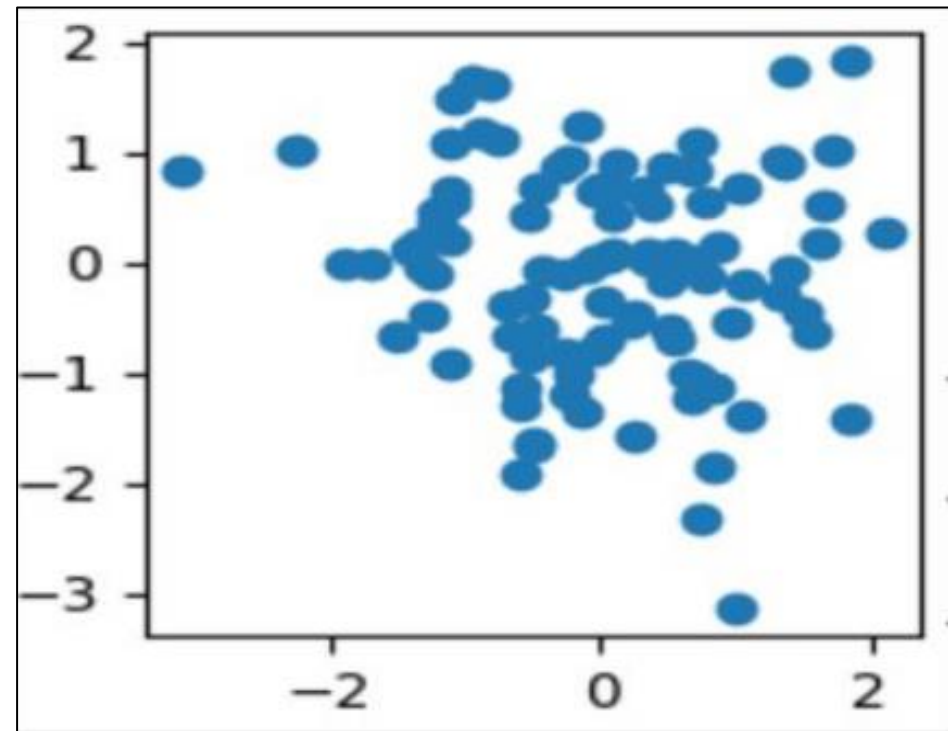


Bar Chart

차트의 종류 - 2차원

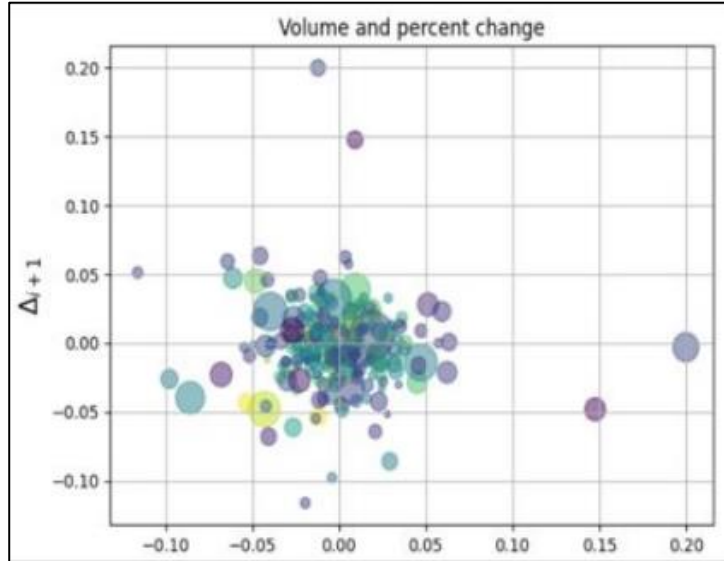


Line Plot

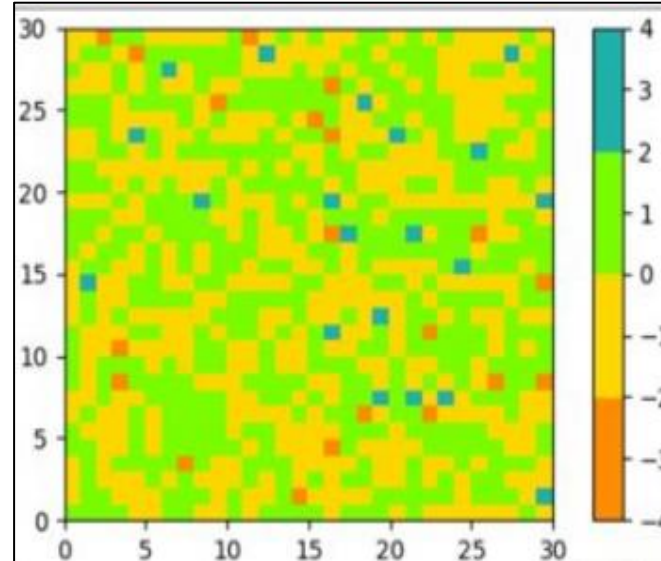


Scatter Chart

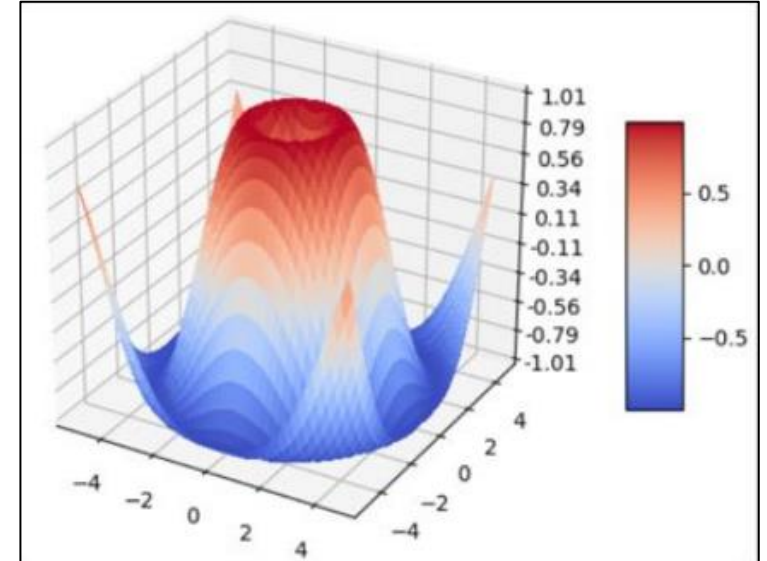
차트의 종류 - 3차원



Bubble Chart



Color Map



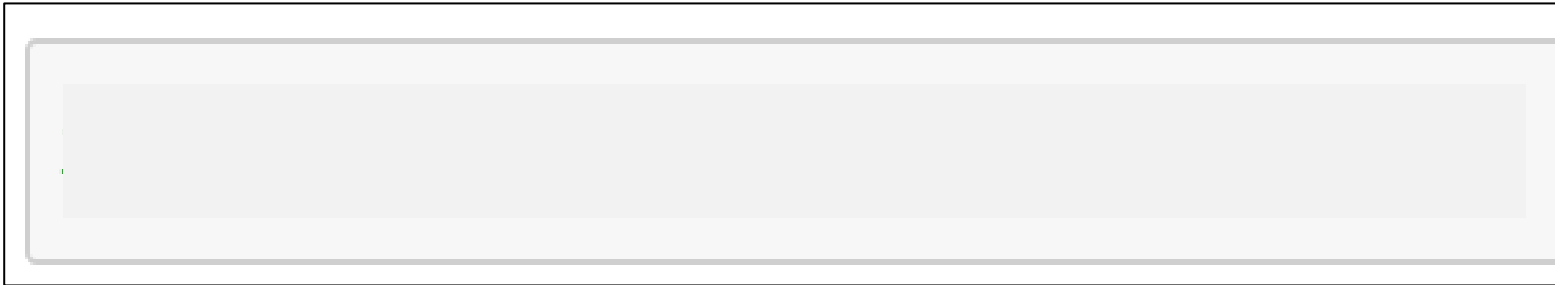
Surface Plot

matplotlib

- Python에서 데이터를 관리하고 시각화 할 수 있는 대표적인 패키지
- **pyplot**과 **pylab**이란 sub패키지

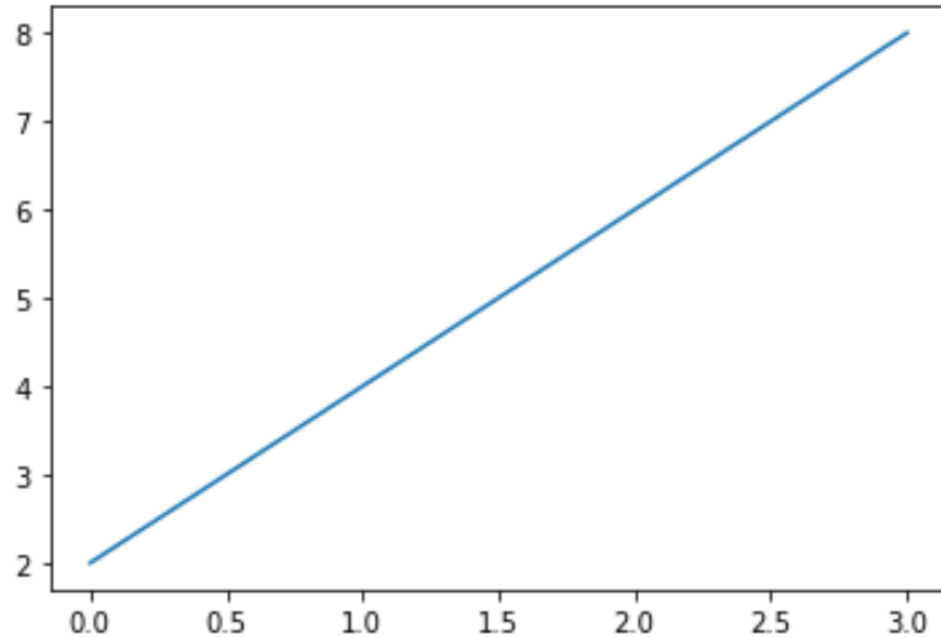
	pyplot	pylab
기능	시각화	시각화 + Numpy
특징	비대화형 (간단한 정보만 입력)	대화형 (비교적 많은 정보를 입력)

Matplotlib 사용하기



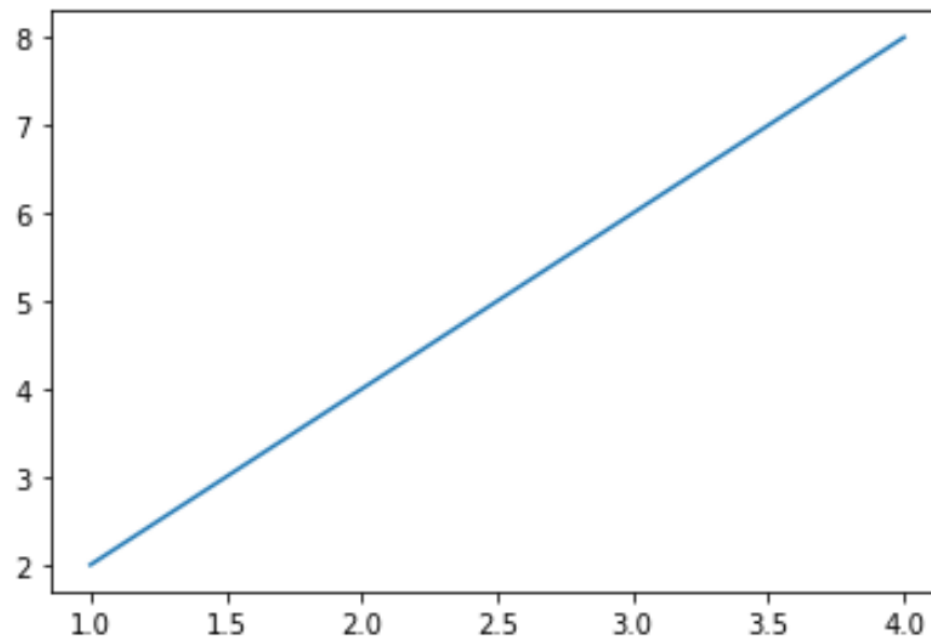
Line plot 그리기

```
y = [2,4,6,8]  
plt.plot(y)  
plt.show()
```



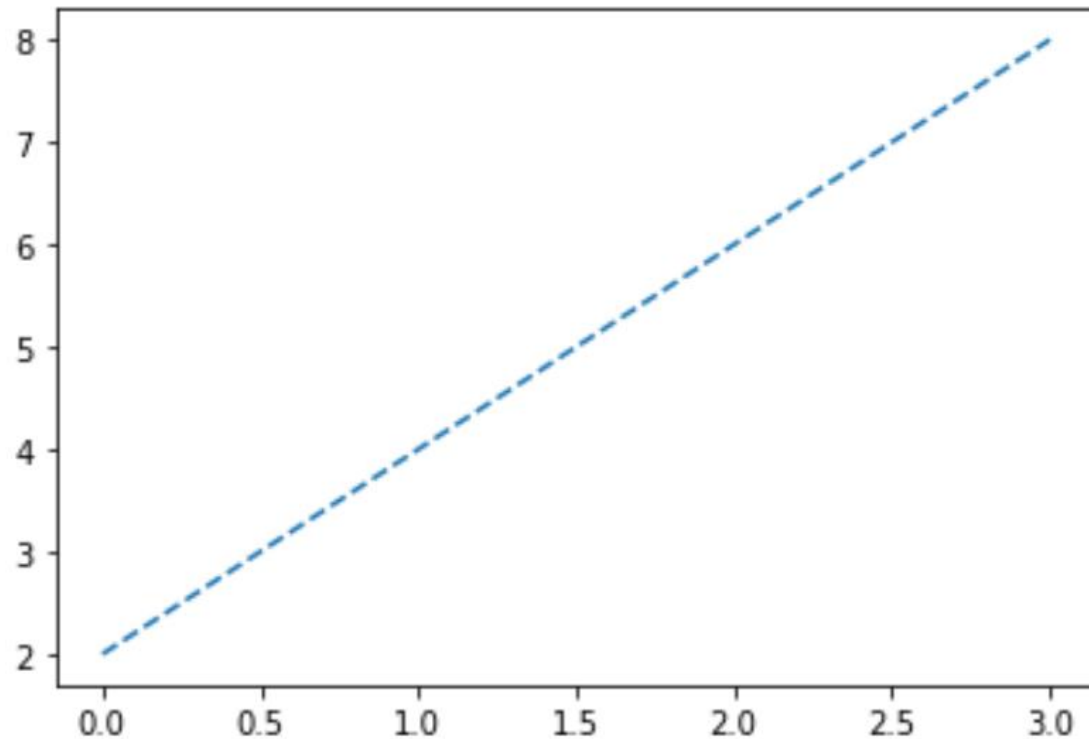
Line plot 그리기

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



Line style 설정하기

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

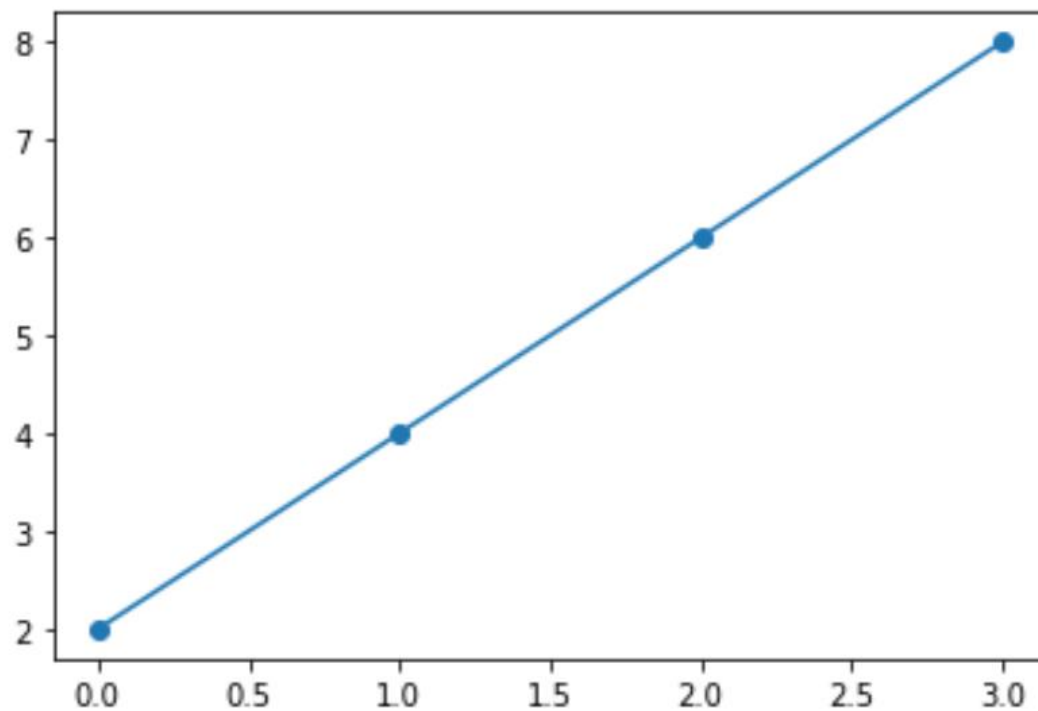


Line style 설정하기

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

Marker 찍기

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

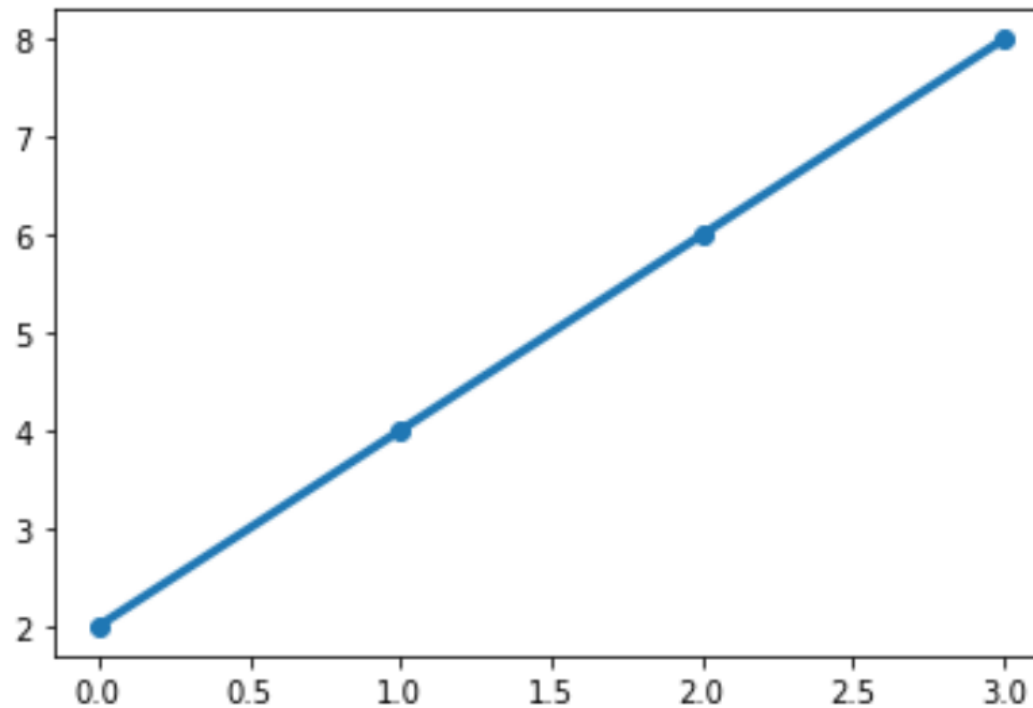


Marker 종류

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

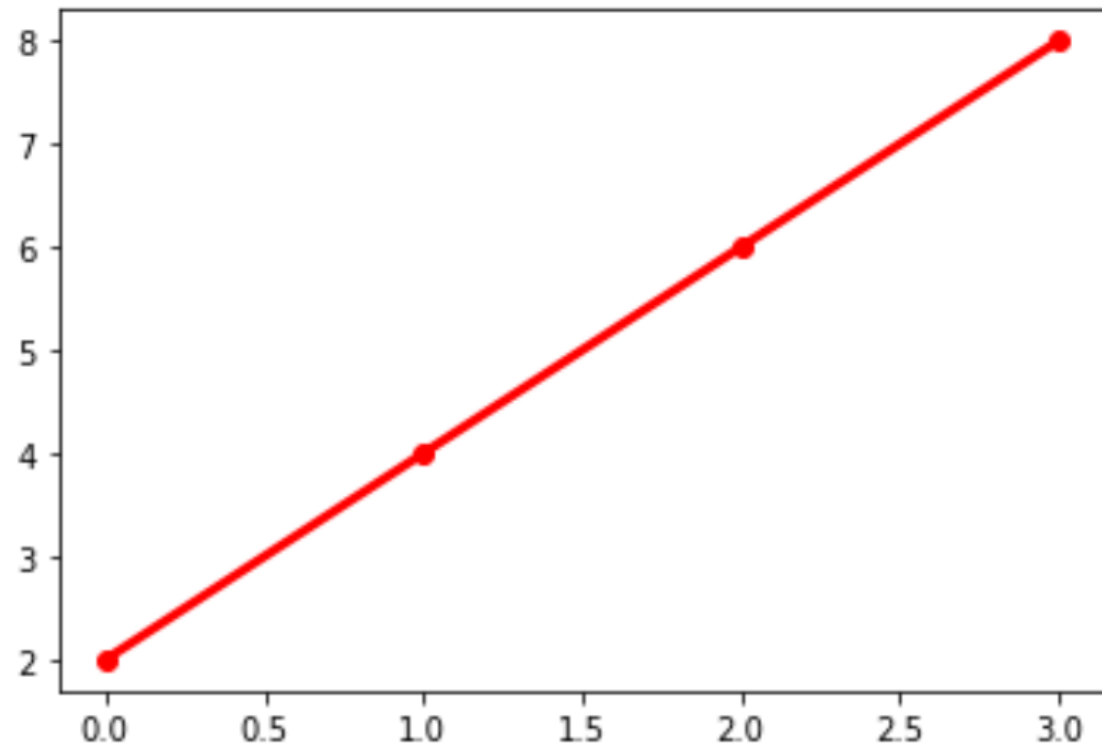
Line width

```
plt.plot(y, marker = 'o', lw = 3)  
plt.show()
```



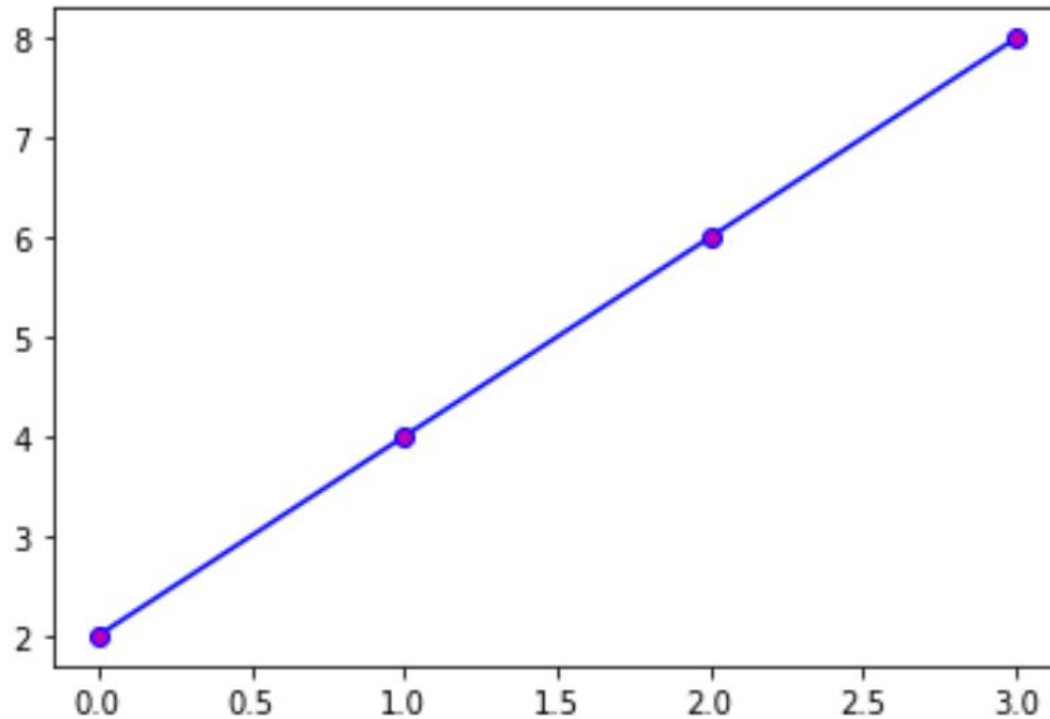
Line color

```
plt.plot(y, marker = 'o', lw = 3, c = 'r')  
plt.show()
```



Marker color

```
plt.plot(y, marker='o', color='b', markerfacecolor='m')  
plt.show()
```



Line color 종류

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

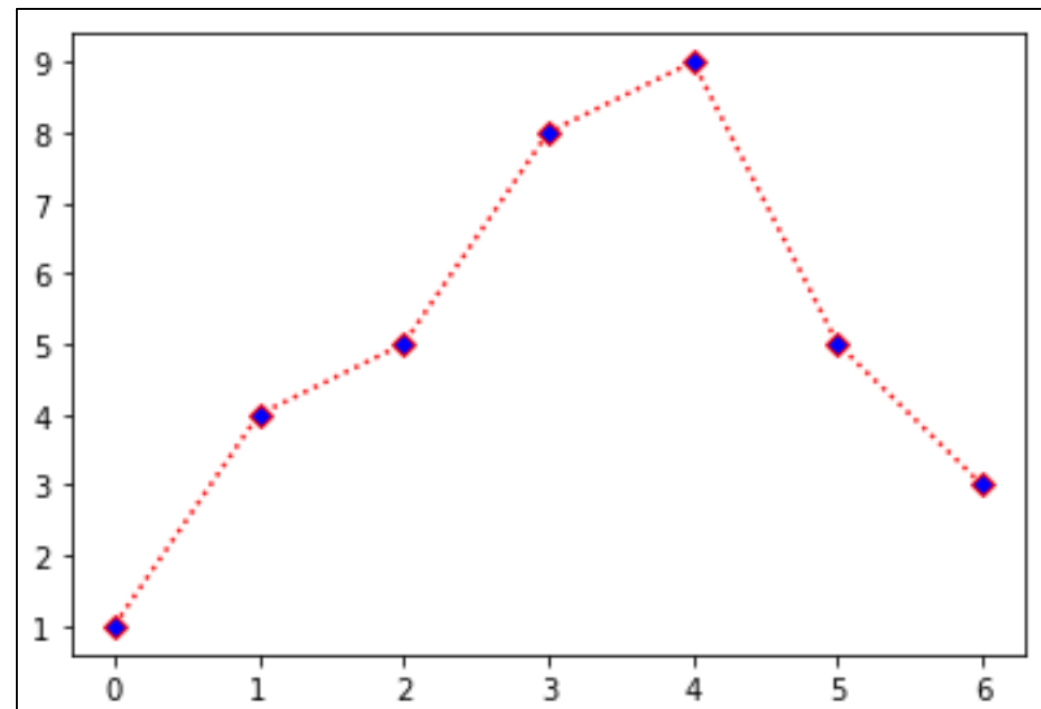
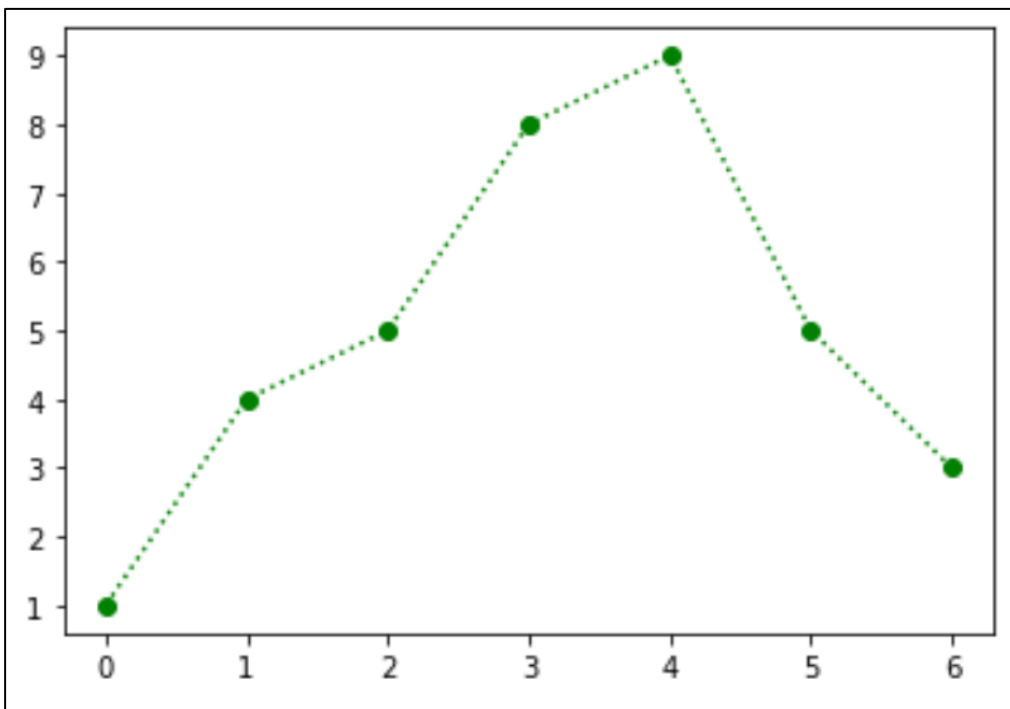
그 외 스타일 옵션들

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)
```

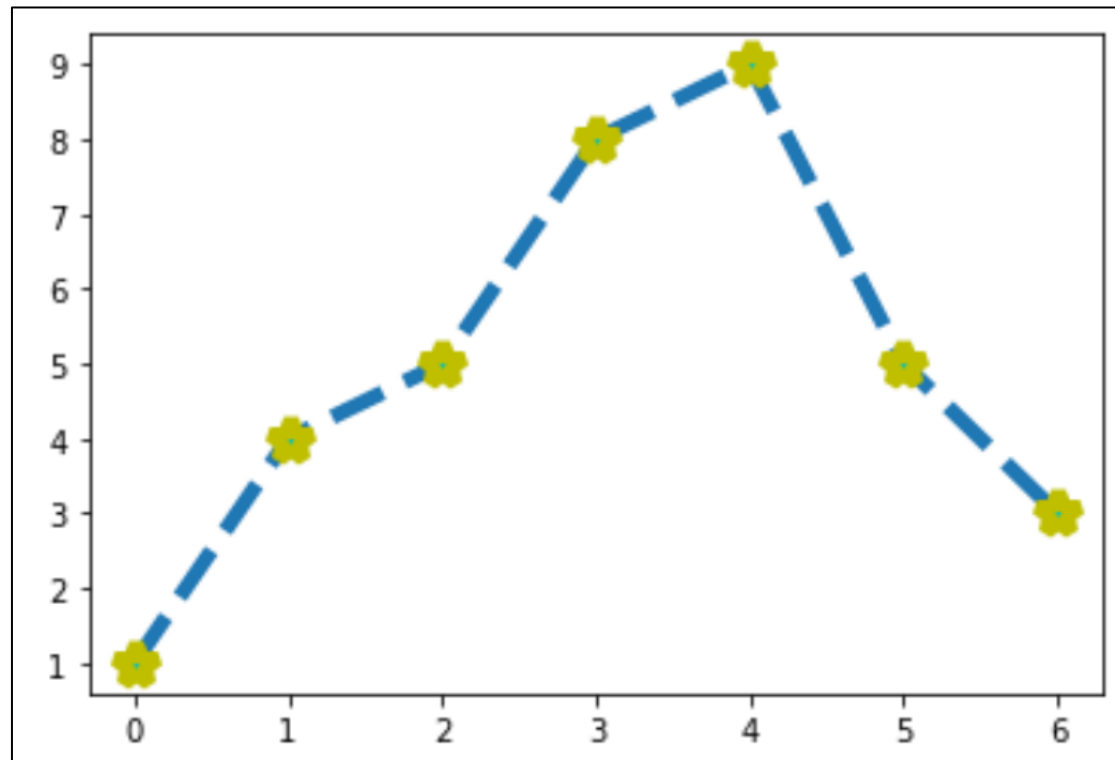
```
y = [1,4,5,8,9,5,3]
```



[연습문제]

```
x = np.arange(7)
```

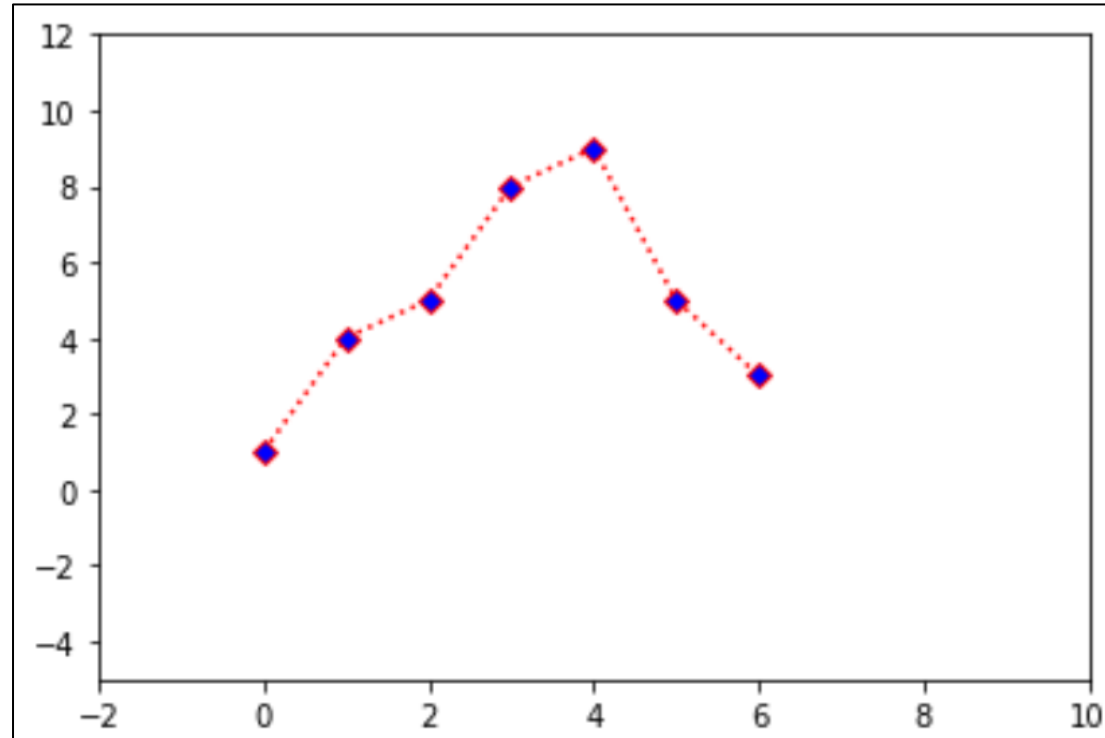
```
y = [1,4,5,8,9,5,3]
```



그래프 옵션 설정 - 그림 범위 지정

```
plt.xlim(-2,10)
```

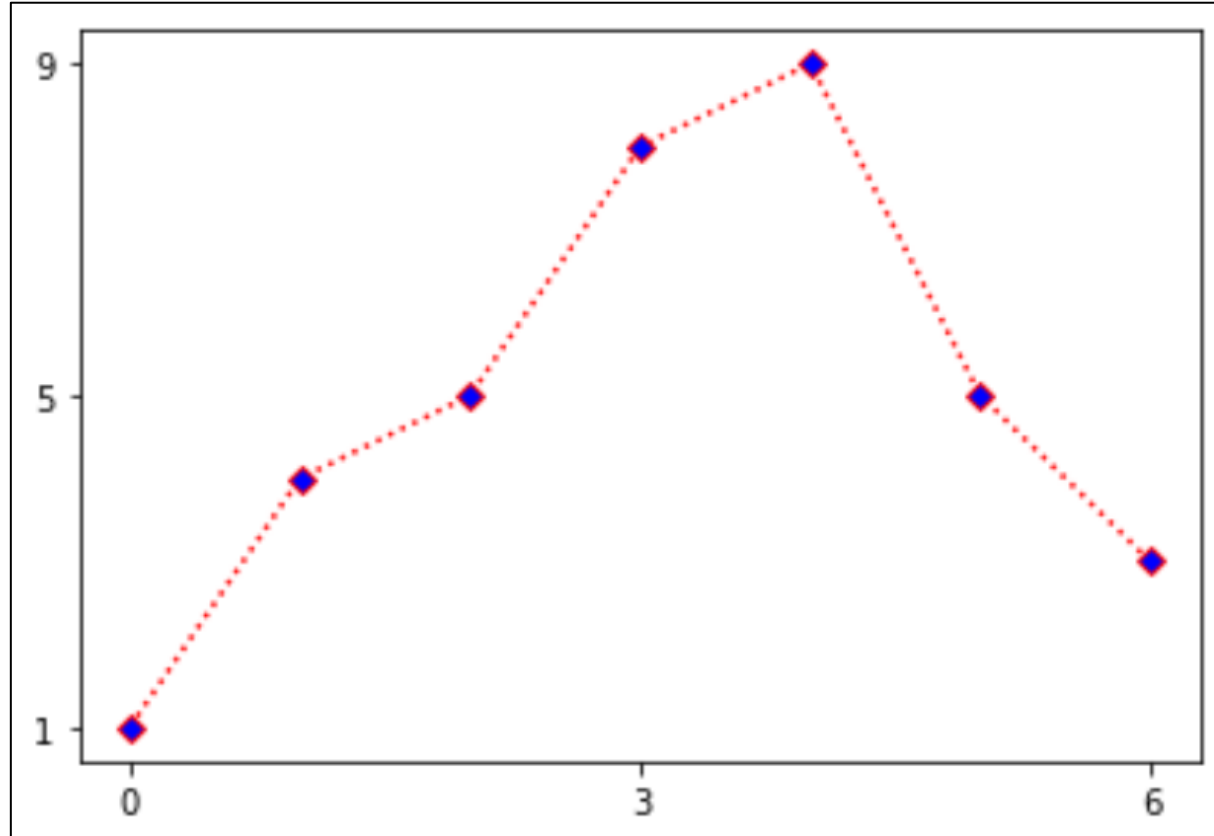
```
plt.ylim(-5,12)
```



틱 설정

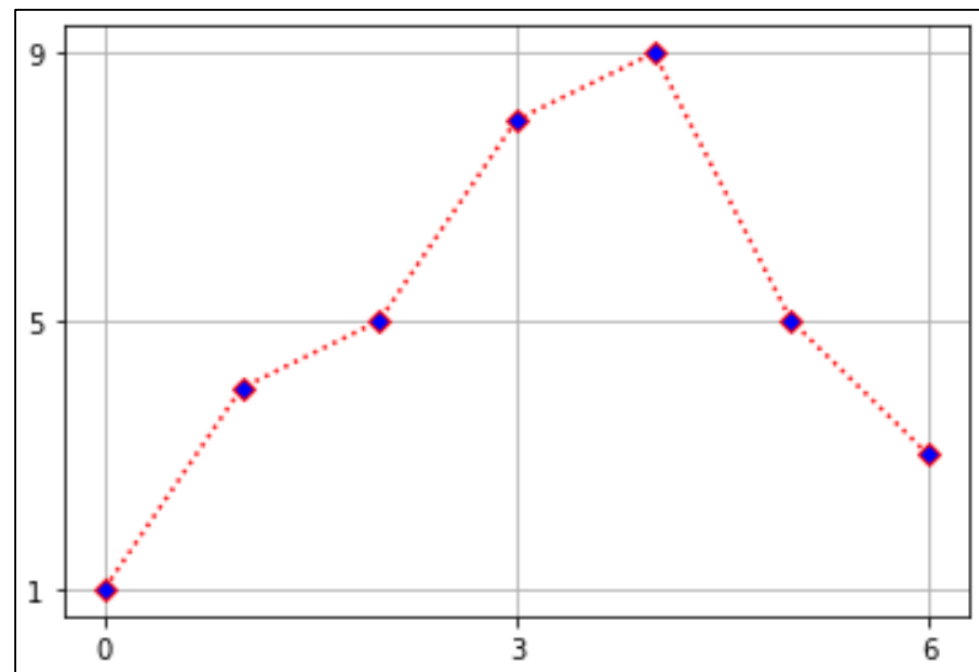
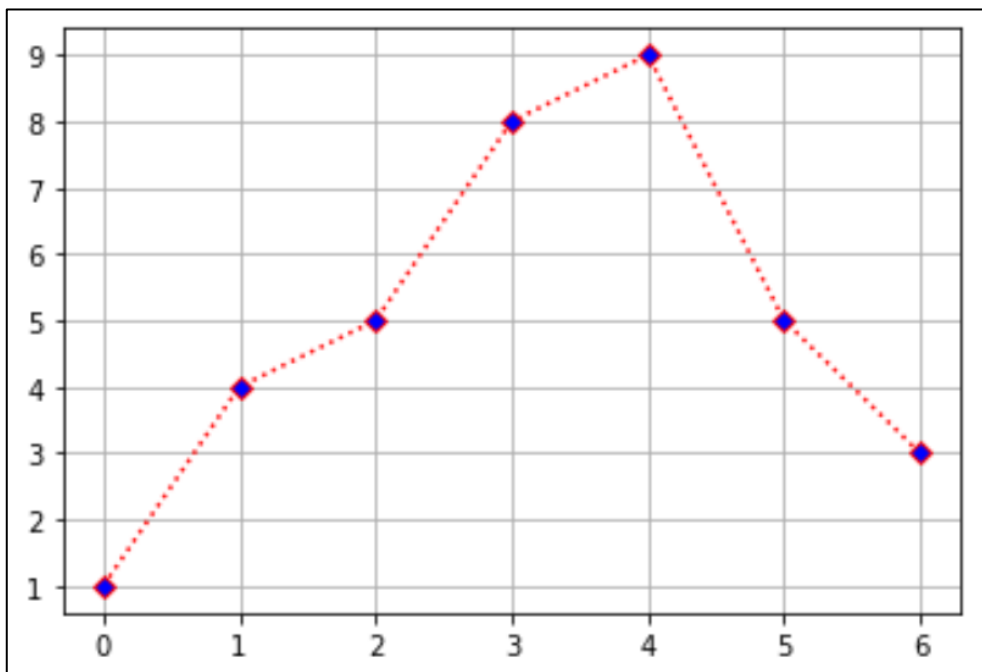
```
plt.xticks([0,3,6])
```

```
plt.yticks([1,5,9])
```



그리드 그리기

`plt.grid()`

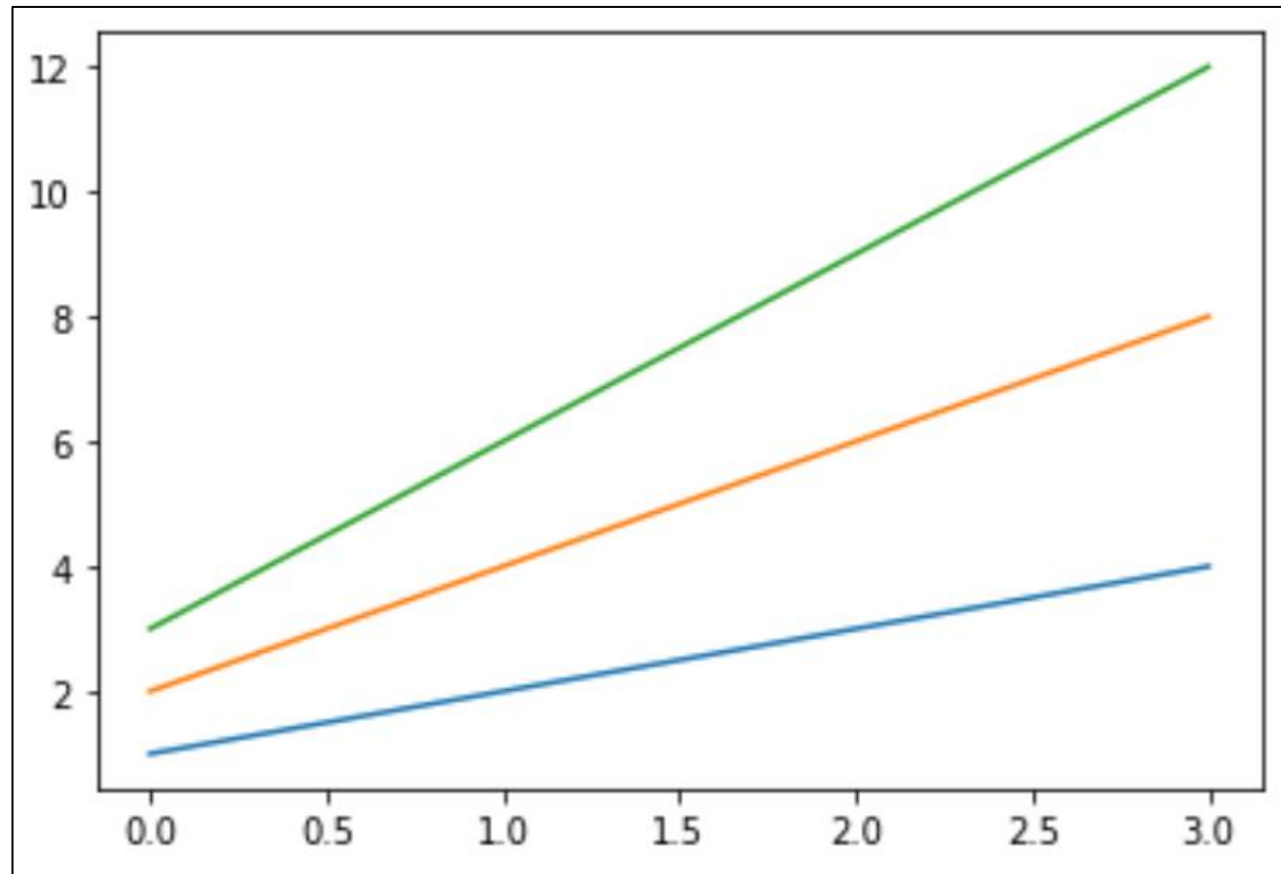


여러 개 plot 그리기

$x = [1, 2, 3, 4]$

$y = [2, 4, 6, 8]$

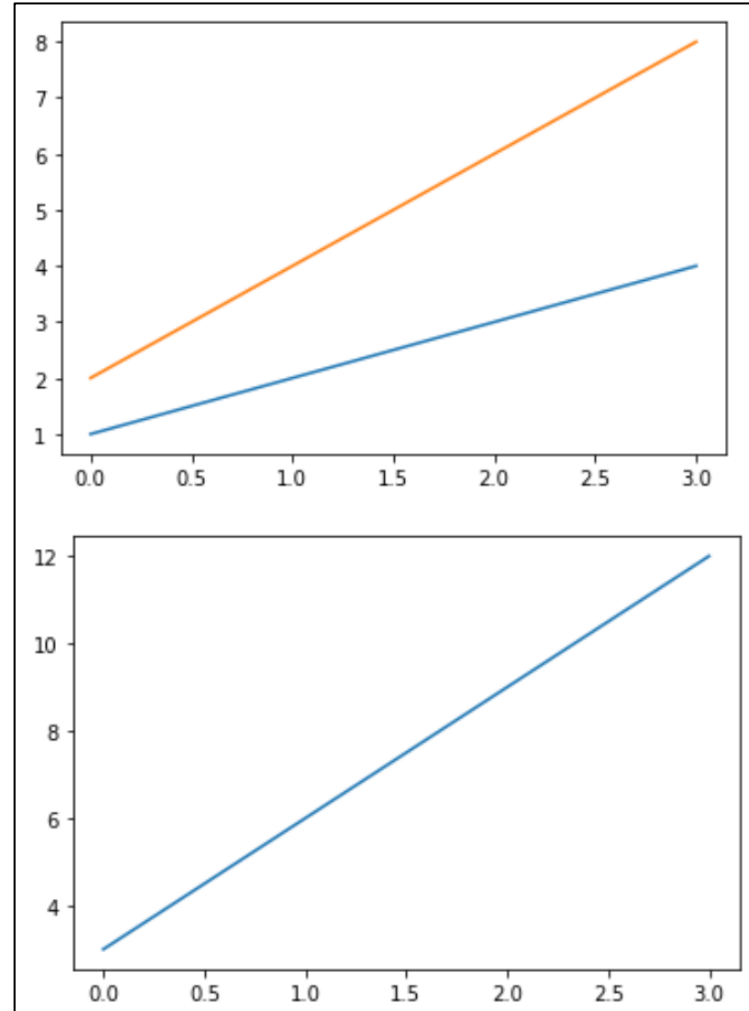
$z = [3, 6, 9, 12]$



여러 개 plot 그리기

```
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()
```

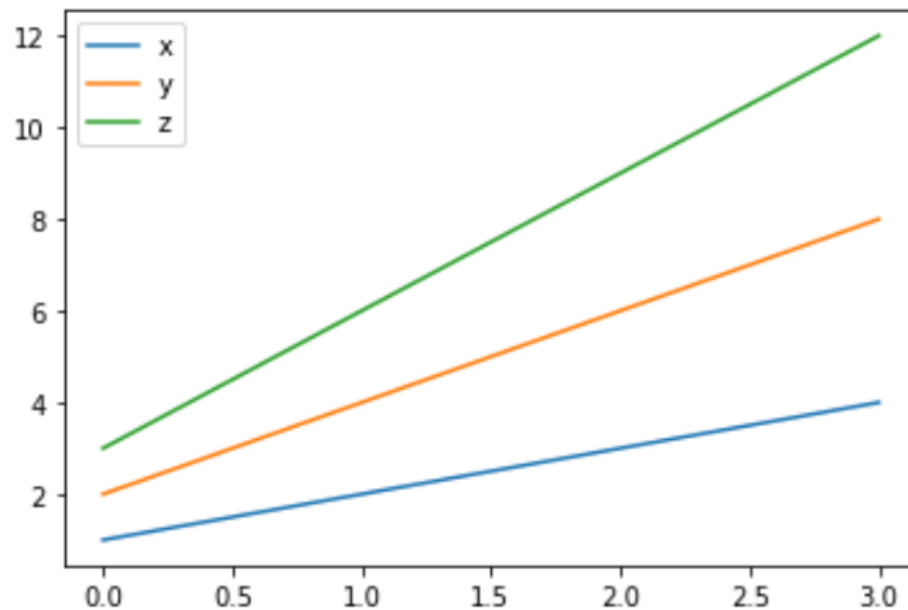


차트 옵션

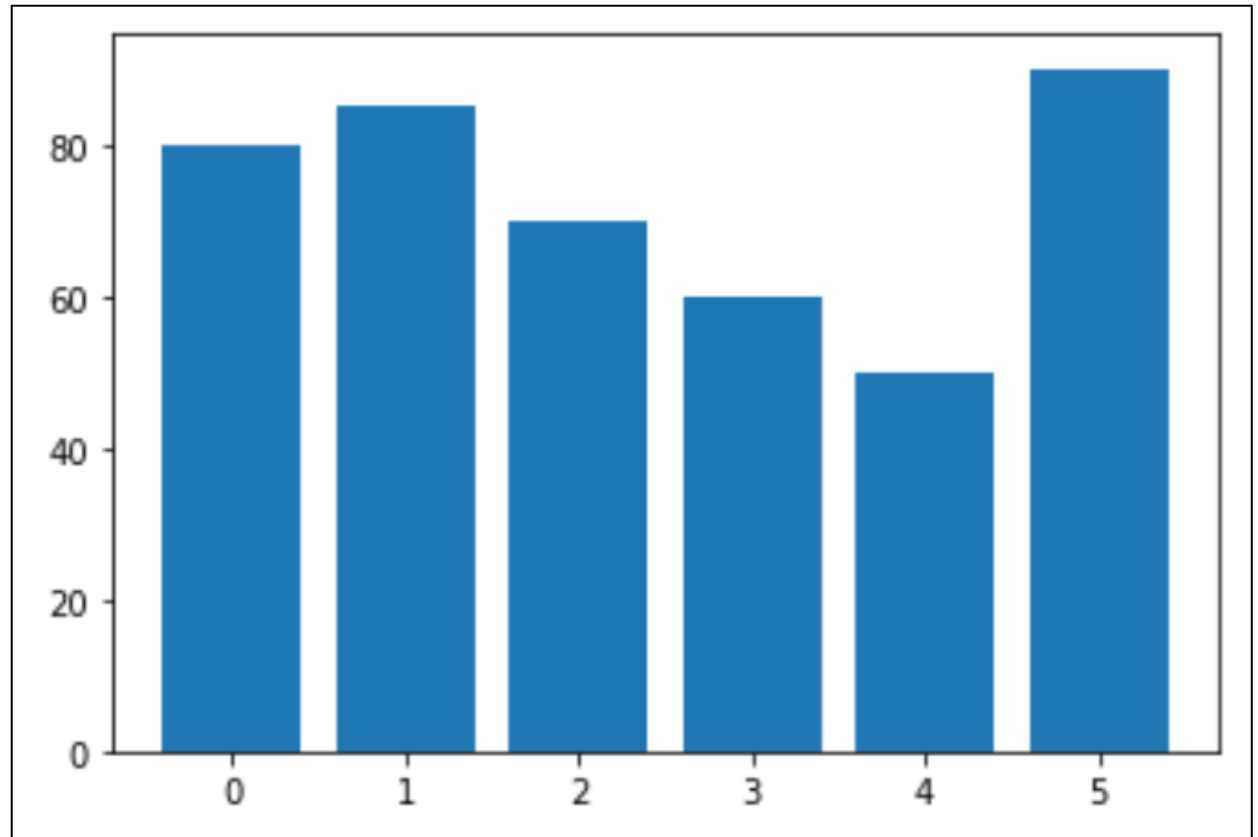
plot 옵션	정의
xlim, ylim	x 축 범위, y축 범위
grid	격자눈금
legend	범례
xlabel, ylabel	x축 타이틀, y축 타이틀
title	그래프 제목
xticks, yticks	x축 눈금 조정, y축 눈금 조정

Bar chart 그리기

`x = [0,1,2,3,4,5]`

`y = [80,85,70,60,50,90]`

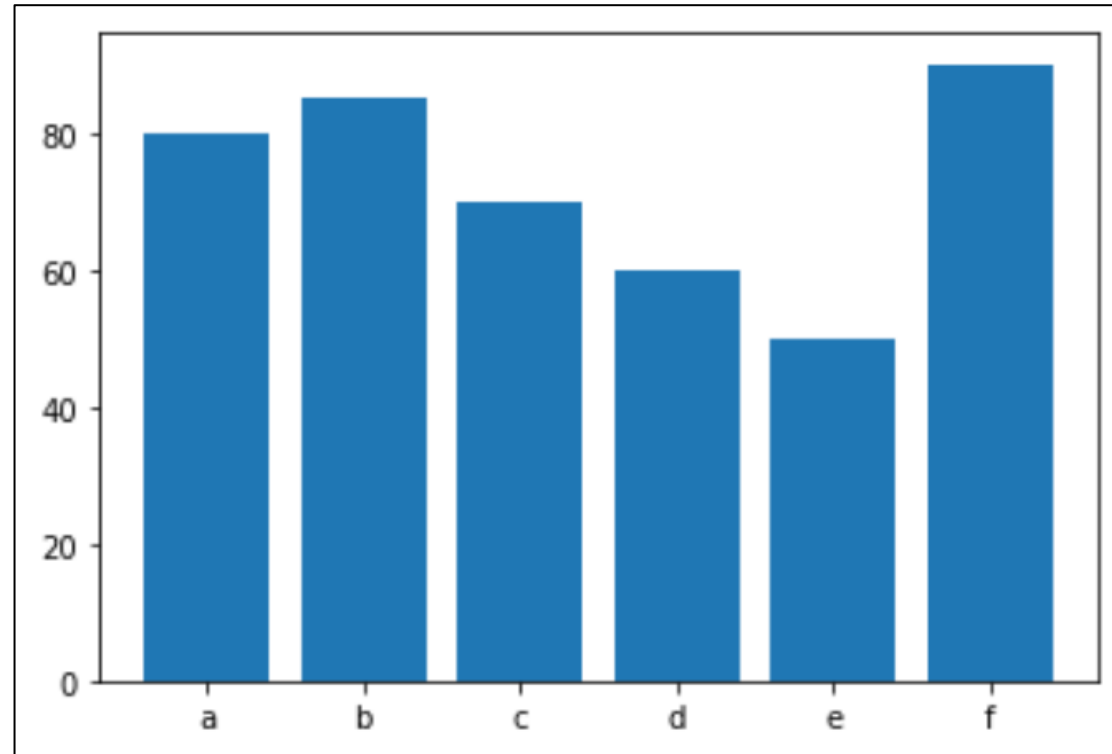
`plot.bar(x,y)`



[연습문제]

x축의 값은 a,b,c,d,e,f 이고

y축의 값은 80,85,70,60,50,90인 막대 그래프를 그리시오

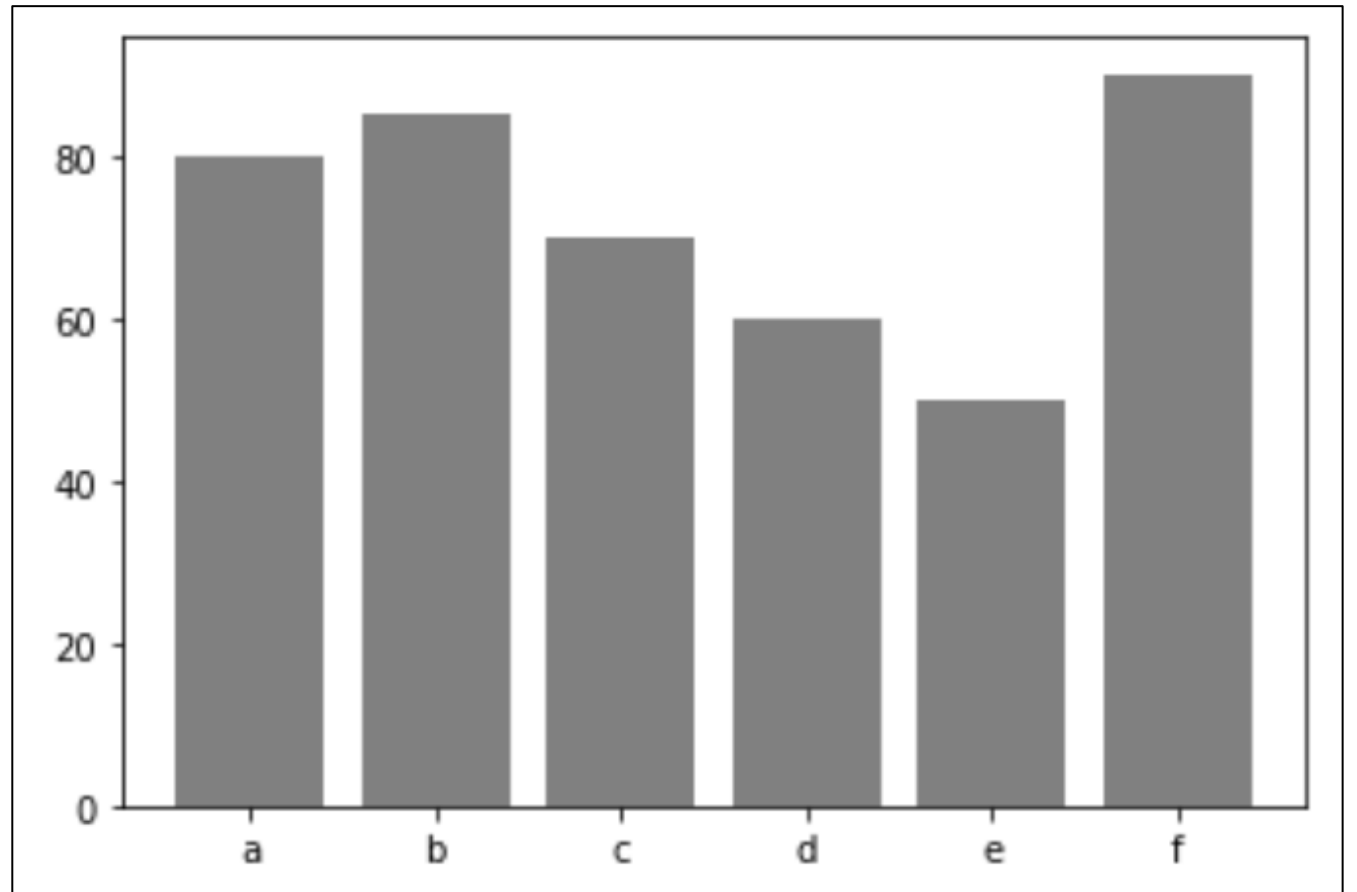


Bar 색상 변경하기

```
x = ['a','b','c','d','e','f']
```

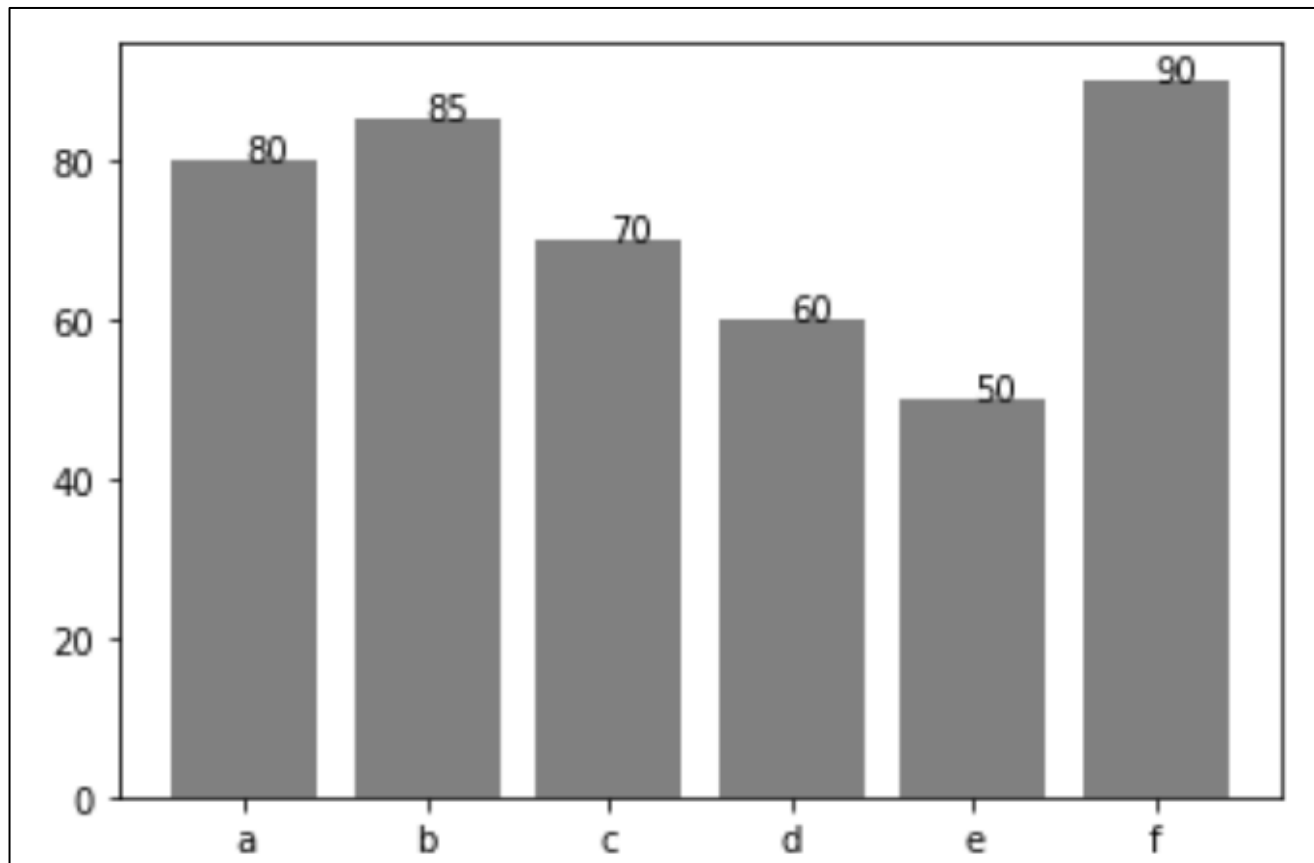
```
y = [80,85,70,60,50,90]
```

```
plot.bar(x,y, color='gray')
```



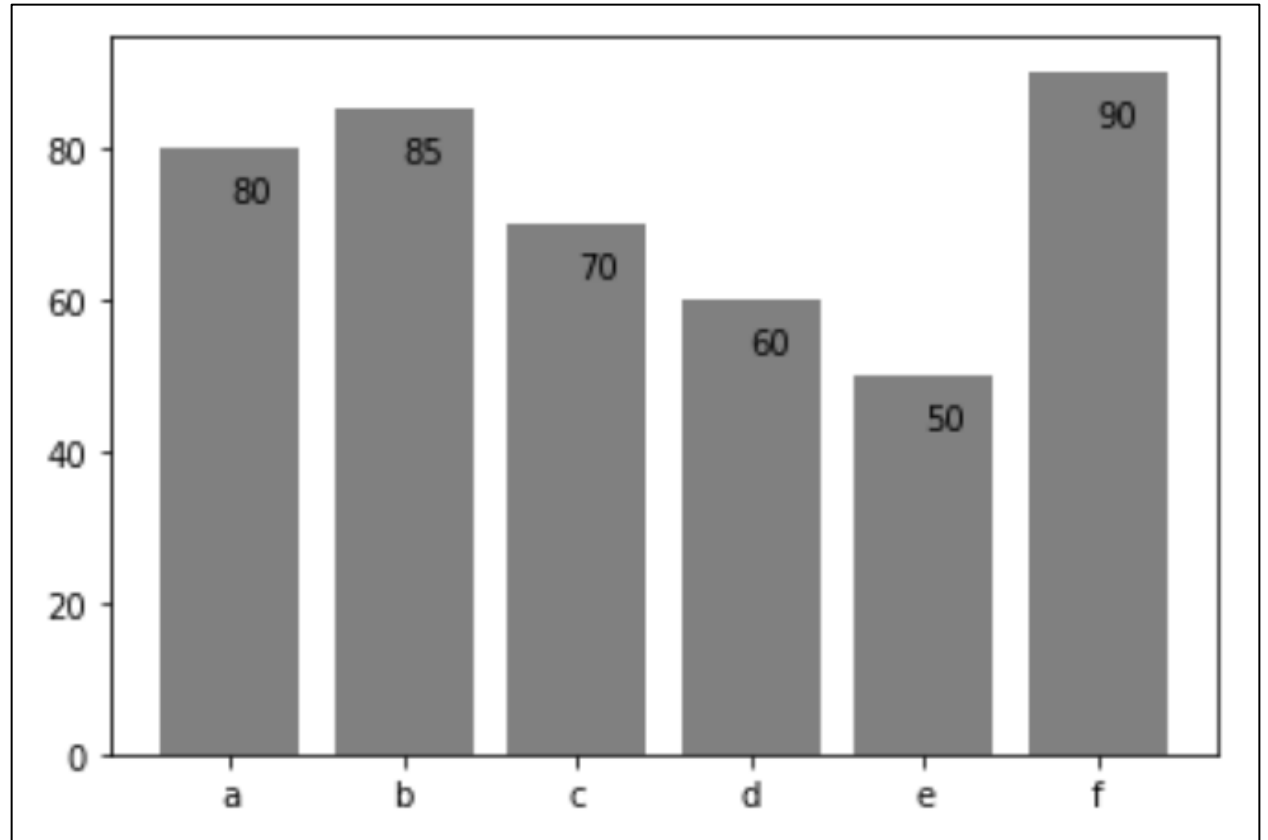
Bar 위에 text 쓰기

```
x = ['a','b','c','d','e','f']  
y = [80,85,70,60,50,90]  
plot.bar(x,y, color='gray')  
  
for i in range(6):  
    plt.text(x[i], y[i], f'{y[i]}')
```



Bar 내부에 text 쓰기

```
x = ['a','b','c','d','e','f']  
y = [80,85,70,60,50,90]  
plot.bar(x,y, color='gray')  
  
for i in range(6):  
    plt.text(x[i], y[i]-7, f'{y[i]}')
```



차트에 한글 표현하기

```
from matplotlib import font_manager, rc
# 시스템에 설치된 Font 리스트 확인
font_list = font_manager.findSystemFonts()
font_list

'C:\\Windows\\Fonts\\verdanaz.ttf',
'C:\\Windows\\Fonts\\ELEPHNT.TTF',
'C:\\windows\\Fonts\\segoeuib.ttf',
'C:\\Windows\\Fonts\\MAIAN.TTF',
'C:\\windows\\Fonts\\HPSimplified_BdIt.ttf',
'C:\\Windows\\Fonts\\phagspa.ttf',
'C:\\Windows\\Fonts\\HANBatangExt.ttf',

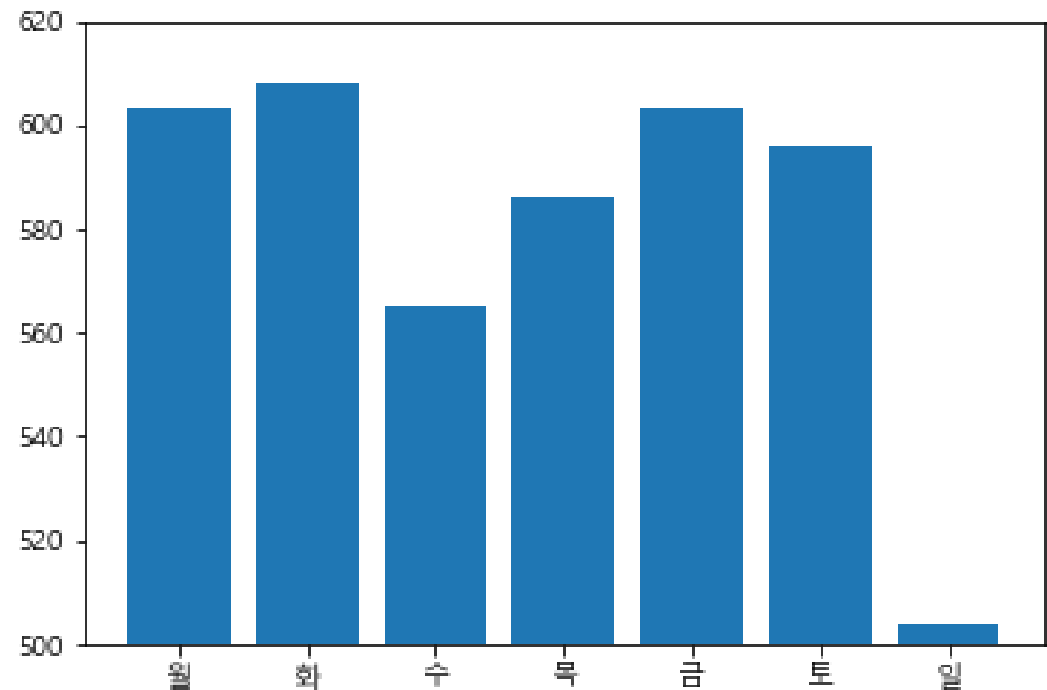
# font_managet 객체를 이용하여 폰트 파일의 이름 확인
font_name = font_manager.FontProperties(fname = 'c:/Windows/Fonts/malgun.ttf').get_name()

# 폰트 설정
rc('font', family = font_name)
```

[공공데이터 활용 데이터 분석 & 시각화]

전국사망교통사고 2017 을 시각화 해보자

1. 요일별 사망교통사고 시각화



1) 데이터 로드하기

```
import pandas as pd
```

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

※ 데이터 크기, 정보 확인하기

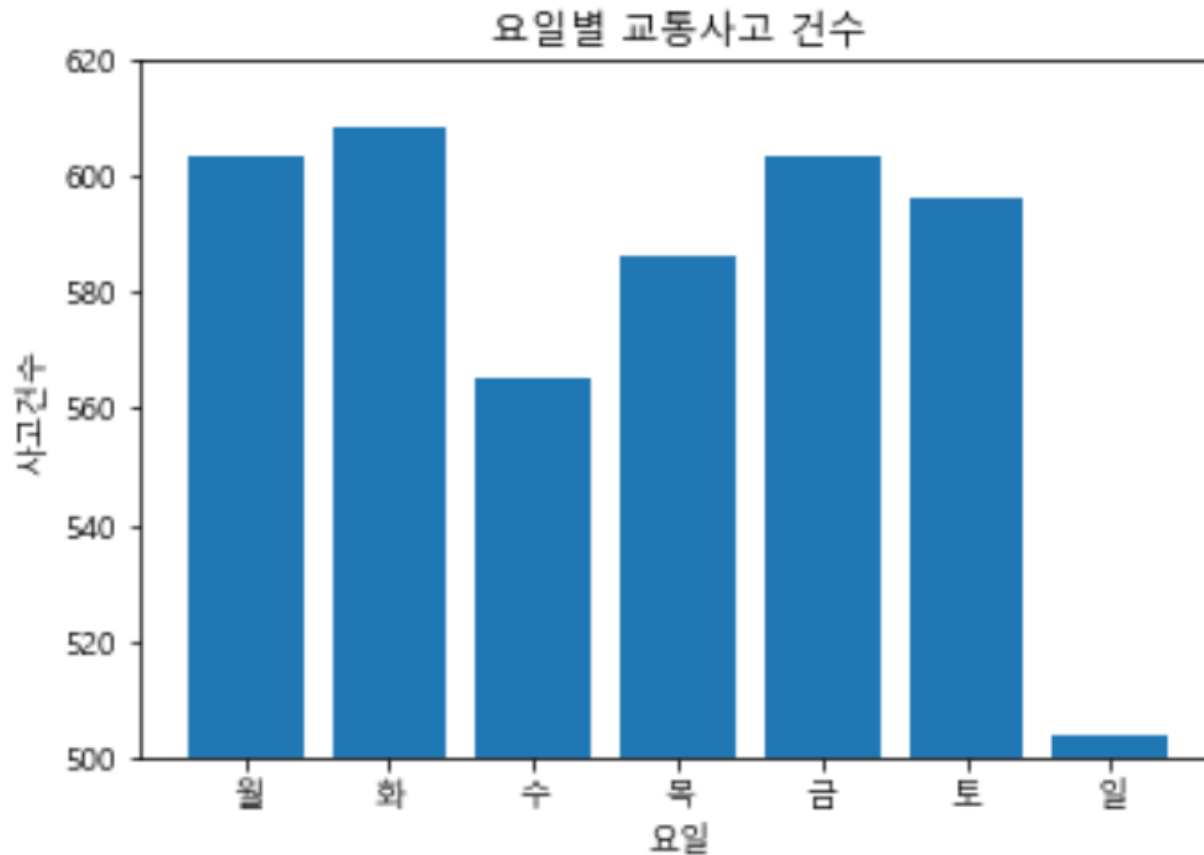
2) 요일별 사고 건수 count 하기

```
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

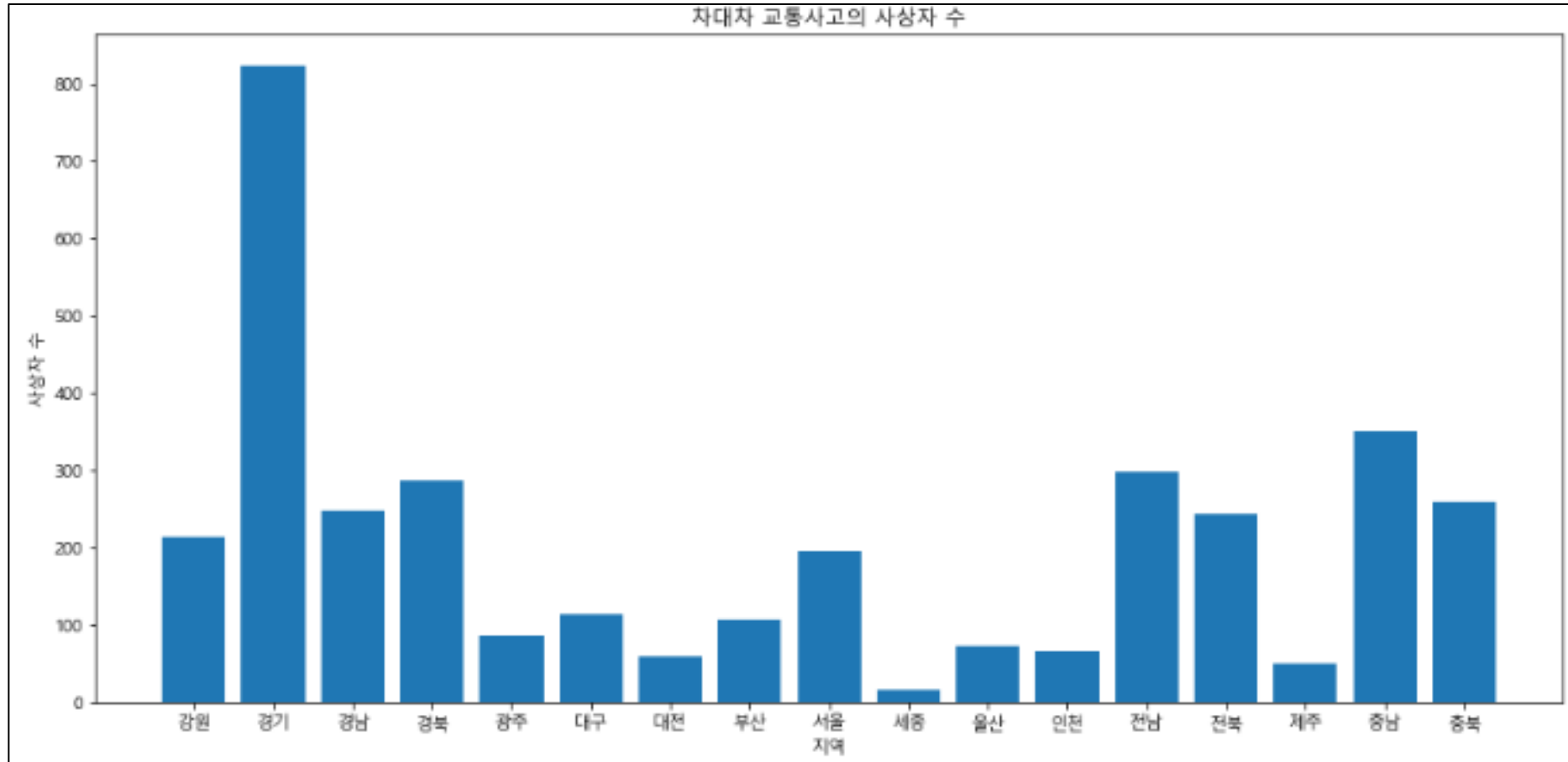
Name: 요일, dtype: int64

3) Bar chart 그리기 - 그래프 속성 추가



xlim, ylim 으로 x, y축 범위 설정
xlabel, ylabel x, y축을 의미하는 제목 설정
title : 전체 그래프 제목 설정

2. '차대차' 사건 중 사상자 수가 많은 발생지 시도를 알아보고 시각화 하기



- 1) 사고유형_대분류 중 '차대차'인 행만 출력
- 2) '발생지시도'에 있는 각각의 데이터 횟수 확인
- 3) 2)번 확인 후 '사상자수', '발생지시도' 컬럼 인덱싱

	사상자수	발생지시도
0	2	전남
2	2	충남
8	3	제주
9	7	충남
17	2	경북
...
4050	1	경남
4055	6	광주
4058	1	경북
4060	4	경남
4063	11	경기

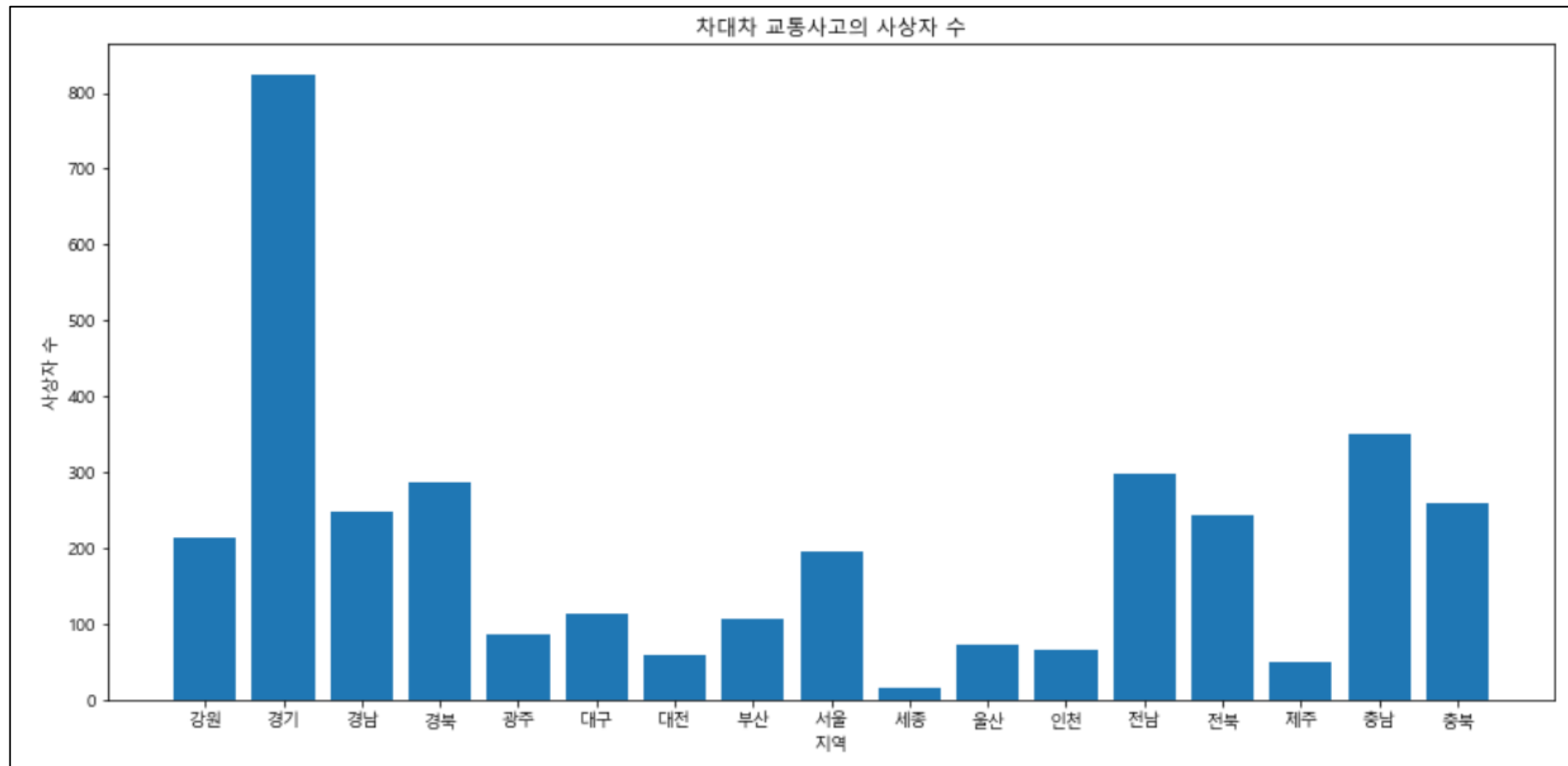
1642 rows × 2 columns

```
경기      330
충남      163
경북      156
경남      146
전남      134
전북      133
서울      112
충북       95
강원       80
대구       67
부산       53
광주       42
인천       41
제주       27
대전       27
울산       26
세종       10
Name: 발생지시도, dtype: int64
```

4) 각 발생지시도별로 묶어주기 위해 groupby 적용 후 총합 구하기

사상자수	
발생지시도	
강원	214
경기	824
경남	248
경북	287
광주	87
대구	115
대전	61
부산	107
서울	197
세종	17
울산	73
인천	66
전남	298
전북	244
제주	50
충남	351
충북	261

5) Bar chart로 결과 시각화



Big Data 

수고하셨습니다.

Enjoy your data analysis

