

# 데이터 시각화 (2024)

데이터과학부 정진명

([jmjung@suwon.ac.kr](mailto:jmjung@suwon.ac.kr), 글로벌경상관 918호)

# 10 주차

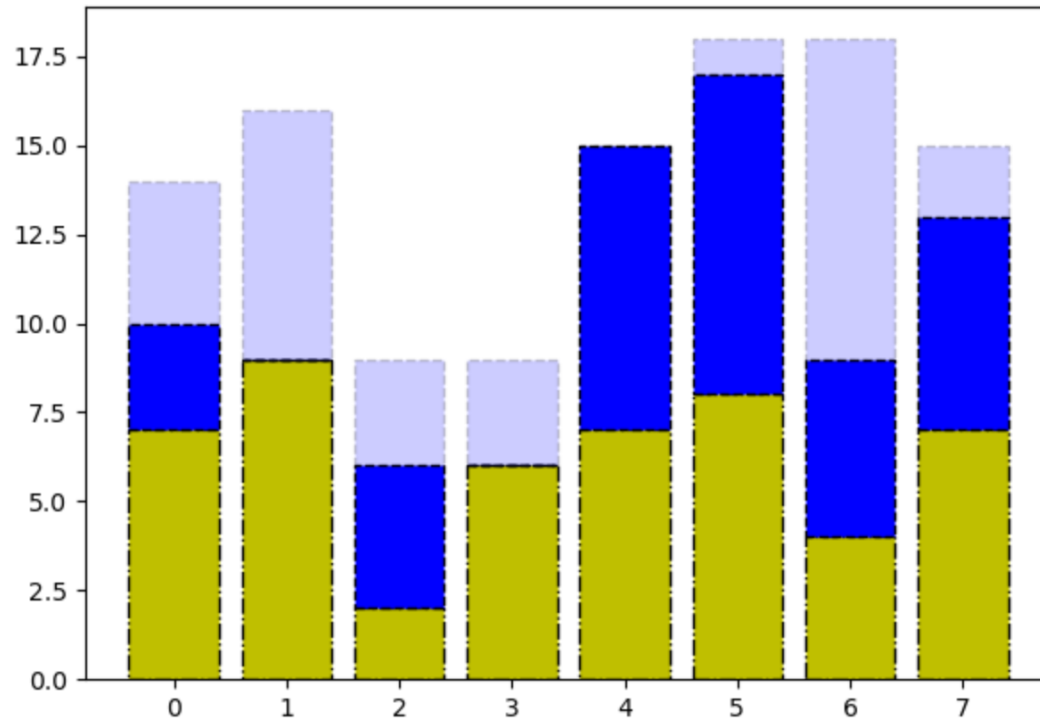
다른 스타일 파라미터 (Cont'd)  
(bar – hatch)

# bar (different style)

```
fig=plt.figure(figsize=(7,5), dpi=100)
ax=fig.subplots()
```

```
A=np.random.randint(0,10,size=8)
B=np.random.randint(0,10,size=8)
C=np.random.randint(0,10,size=8)
X=np.arange(8)
```

```
ax.bar(X, A, color='y', edgecolor='k', linestyle='-.')
ax.bar(X, B, bottom=A, color='b', edgecolor='k', linestyle='dashed')
ax.bar(X, C, bottom=A+B, color='b', edgecolor='k', linestyle='dashed', alpha=0.2)
```



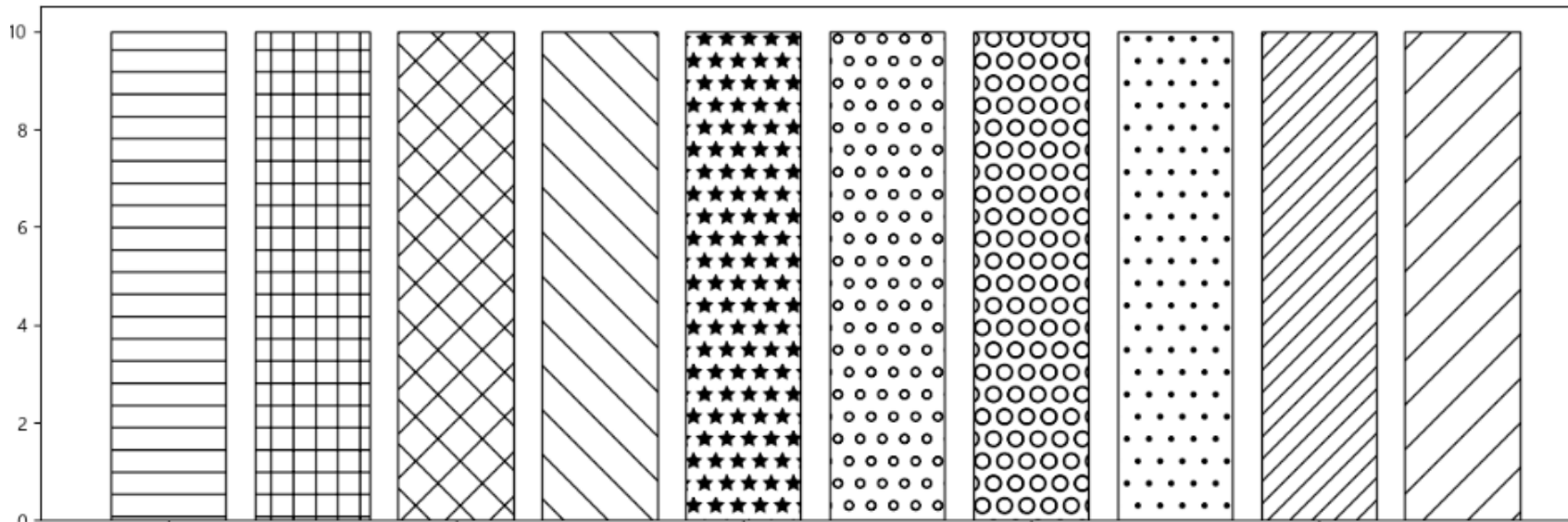
# bar (different style) - hatch

```
pats = ('-', '+', 'x', '\\', '*', 'o', '0', '.', '//', '/')

fig=plt.figure(figsize=(10,6), dpi=100)
ax1, ax2=fig.subplots(2,1)

## ax1
for ii in range(len(pats)):
    _=ax1.bar(ii,10,color='w',edgecolor='k',hatch=pats[ii])

## ax2
ax2.bar(range(10), 10, color='w',edgecolor='k',hatch=pats)
```

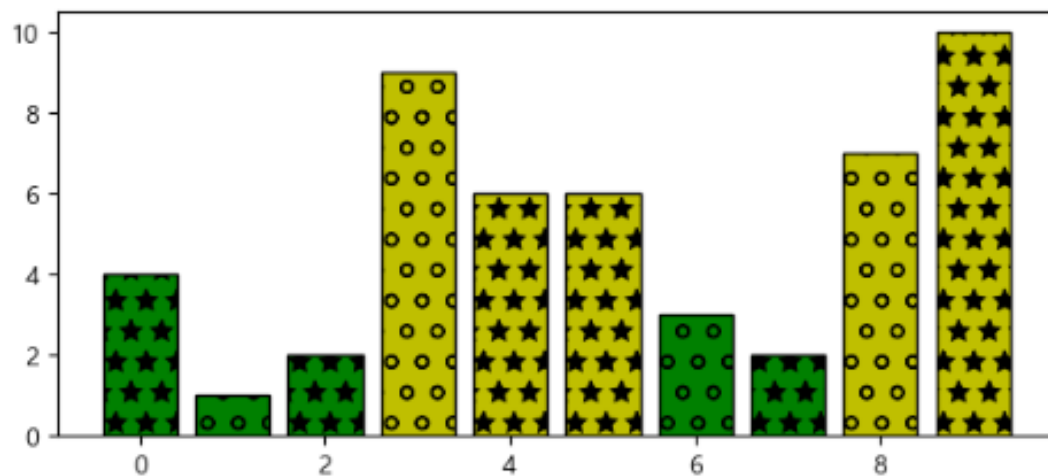


# 실습1

- 코드에 주어진 10의 값을 갖는 데이터 Y을 bar 그래프로 나타내는데,
  - 5이상은 노란색, 5미만은 초록색으로 나타내고
  - 짝수는 "별" pattern으로 홀수는 "동그라미" pattern으로 나타내시오

```
array([ 4,  1,  2,  9,  6,  6,  3,  2,  7, 10])
```

```
<BarContainer object of 10 artists>
```



hatch parameter는 리스트(시퀀스 자료형)도 argument로 가능

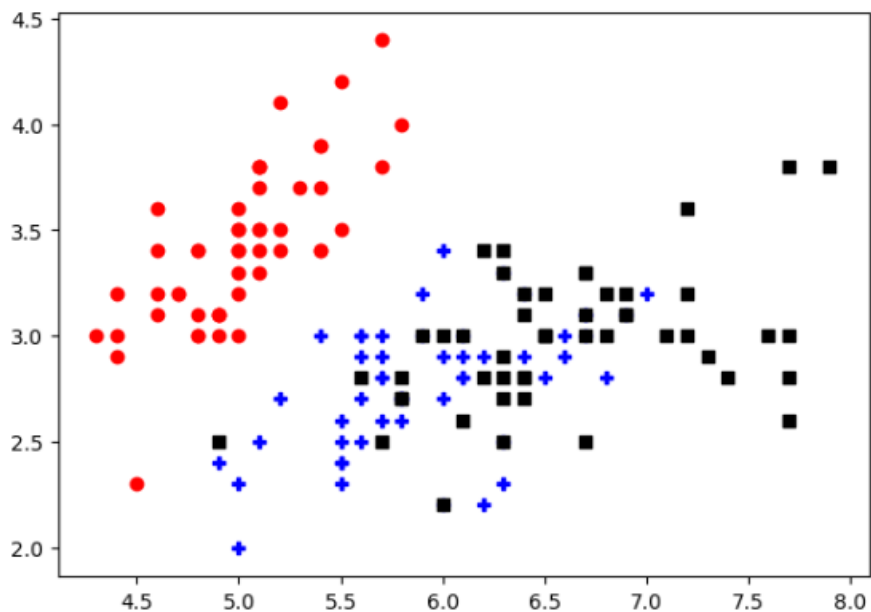
# style 파라미터 정리

Style 명	Parameter 명	시퀀스자료(ex: 리스트) 사용 가능 여부
color	color	O
size	s	O
Pattern	hatch	O
Marker	marker	X
투명도	alpha	O

- 제목 추가
- 수식 활용 (레이텍)
  - $X, Y$  label
  - $X, Y$  lim
  - $X, Y$  tick



# 이제까지 그린 그림들

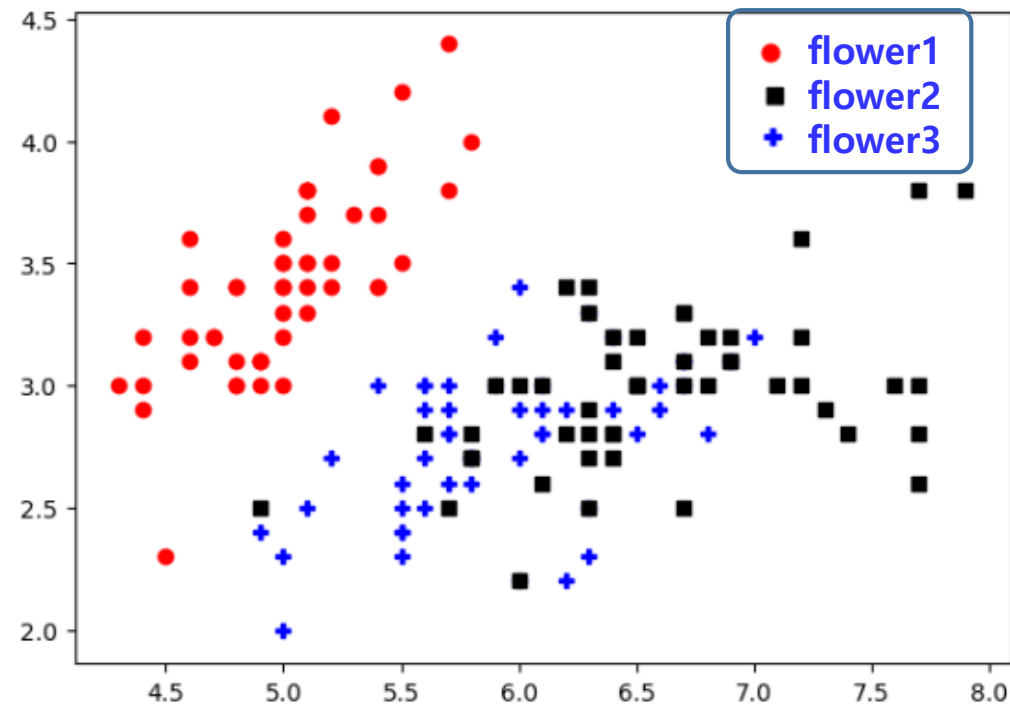


정보가 부족하다



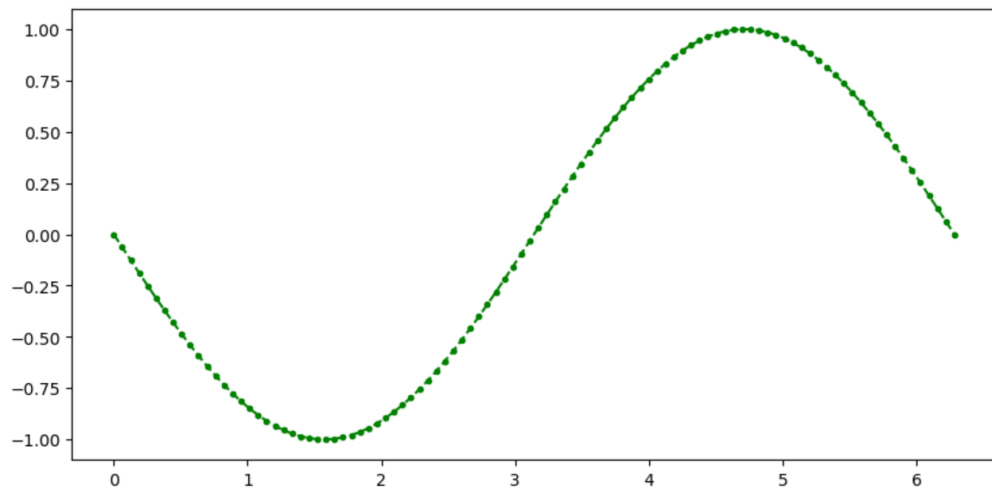
꽃잎 길이

세 종류 꽃의 꽃잎 길이와 너비의 분포

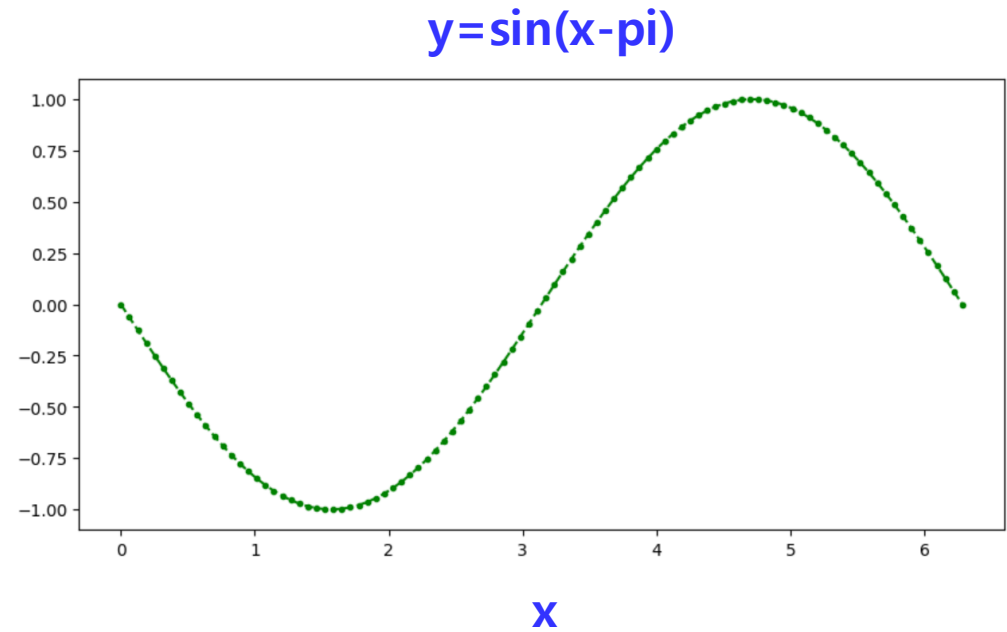


꽃잎의 길이

# 이제까지 그린 그림들

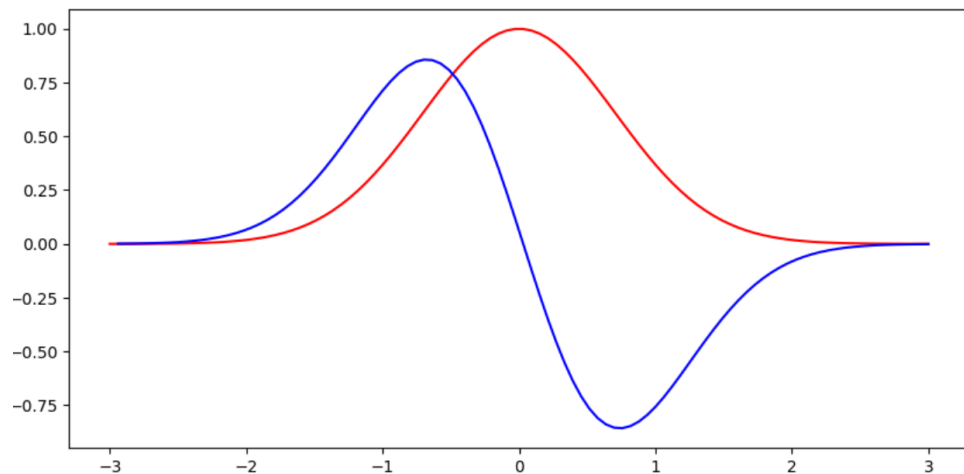


y

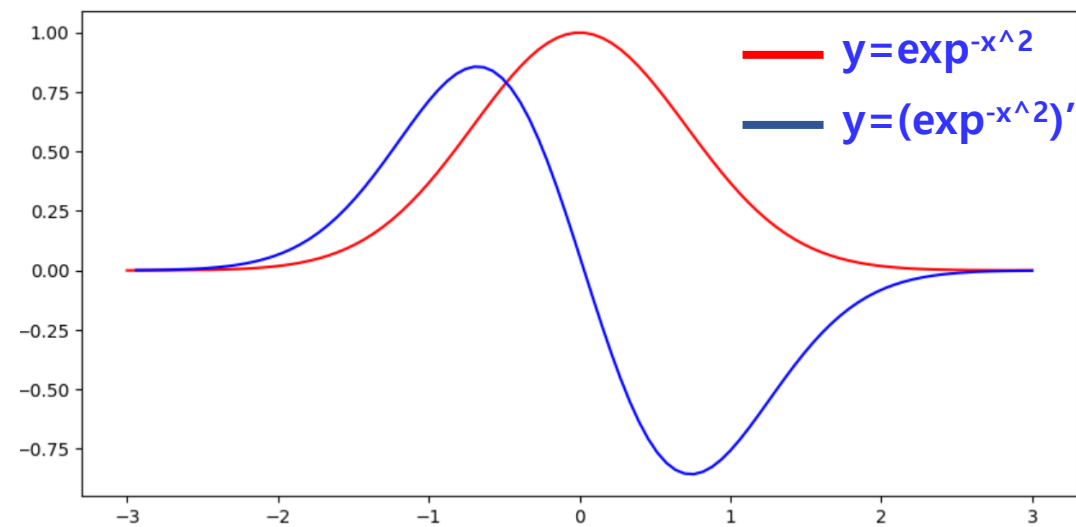


정보가 부족하다

# 이제까지 그린 그림들



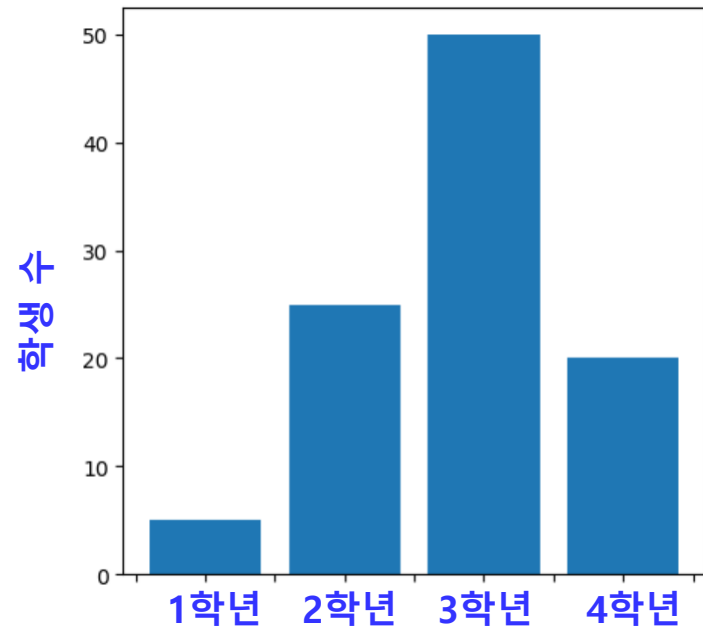
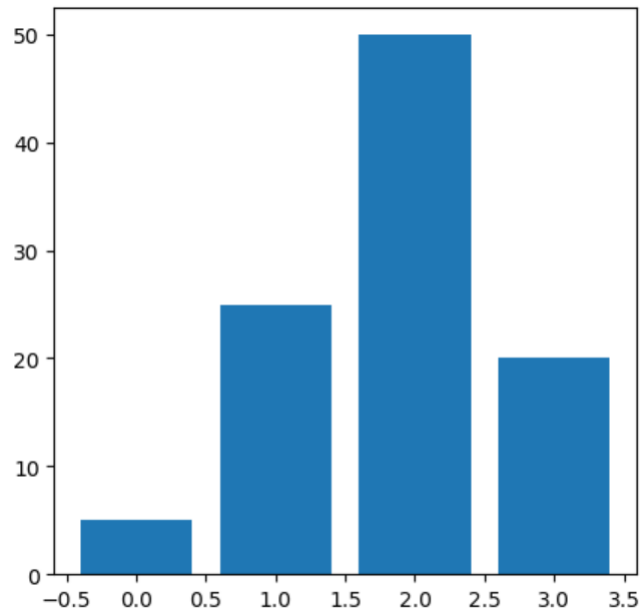
y



x

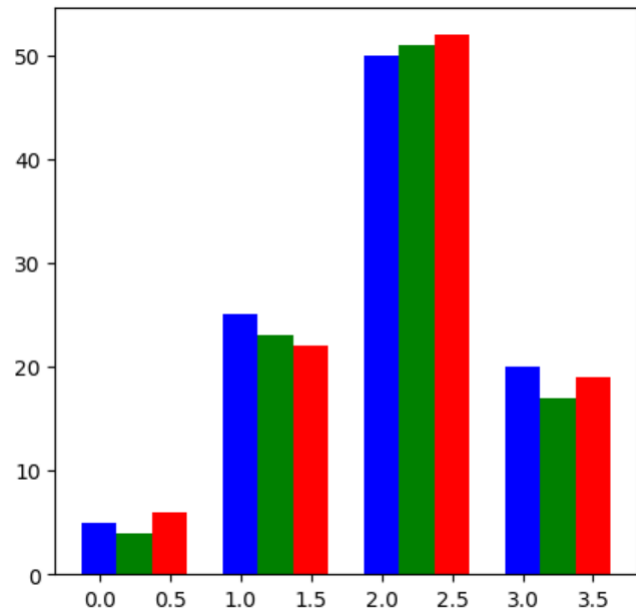
정보가 부족하다

# 이제까지 그린 그림들



정보가 부족하다

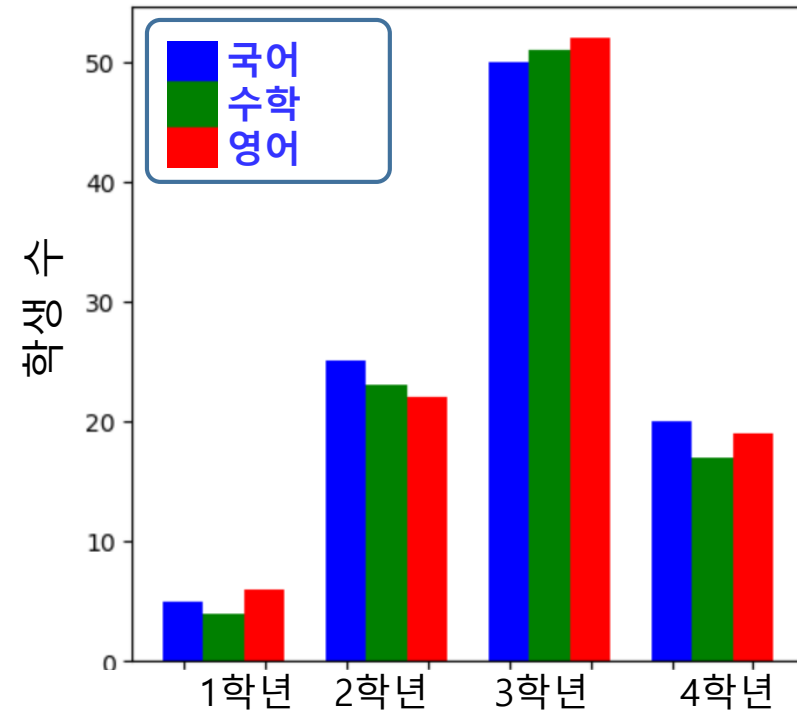
# 이제까지 그린 그림들



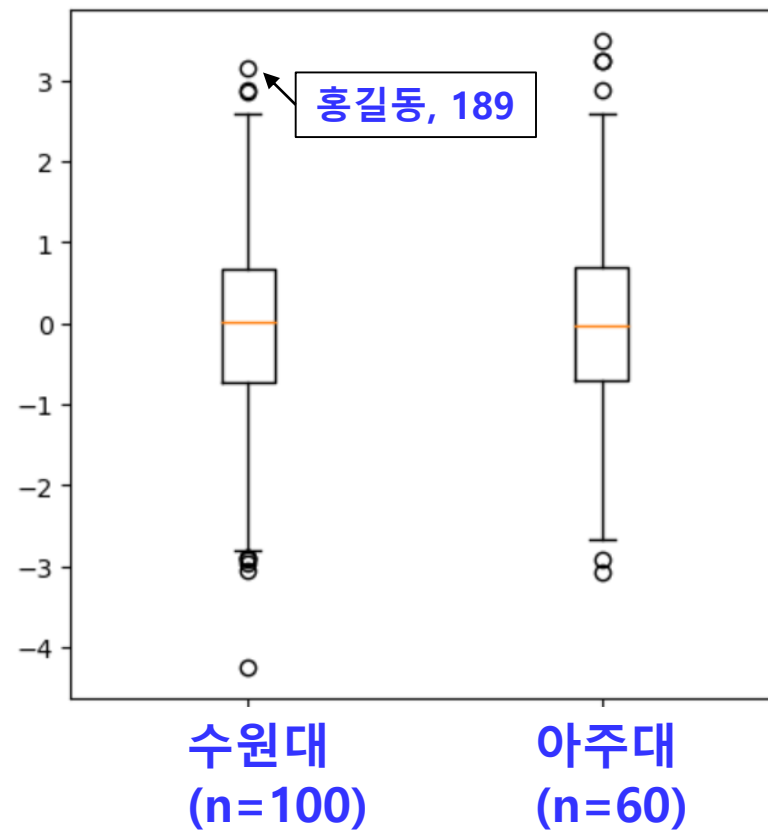
정보가 부족하다



1~4학년 국영수 평균 점수



비파리



제목 추가

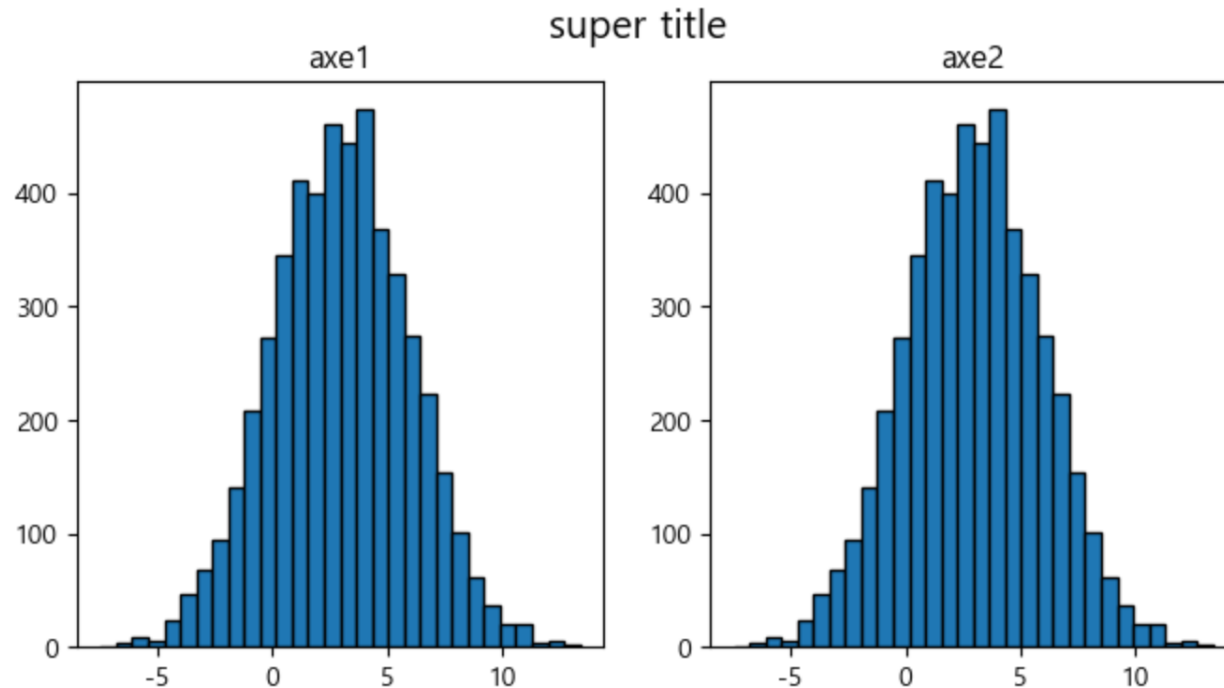
# 제목 추가 (ax.set\_title & fig.suptitle)

```
fig=plt.figure(figsize=(8,4), dpi=100)
ax1, ax2=fig.subplots(1,2)

d1=np.random.normal(3,3,size=5000)
_=ax1.hist(d1, edgecolor='k', bins=30)
_=ax1.set_title('axe1')

_=ax2.hist(d1, edgecolor='k', bins=30)
_=ax2.set_title('axe2')

_=fig.suptitle('super title', fontsize=15)
```





# 제목 추가 (set\_title)

주어진 평균, 표준편차,  
샘플 수를 활용하여  
데이터를 생성하고, 제목  
추가하기

```
# 평균이 m, 표준편차 s인 정규분포에서 n개 값 sampling 하여 히스토그램 그리기
```

```
fig=plt.figure(figsize=(4,4), dpi=100)
```

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

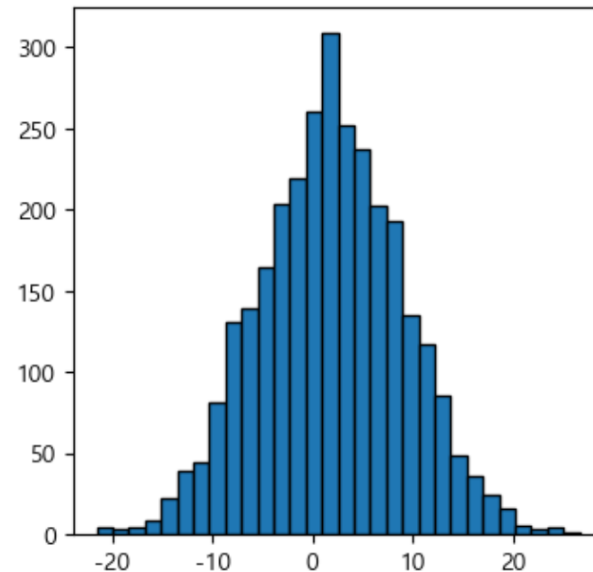
```
m,s,n=(2,7,3000)
```

```
d1=np.random.normal(m,s,size=n)
```

```
_ =ax.hist(d1, edgecolor='k', bins=30)
```

```
_ =ax.set_title('평균이 {}, 표준편차 {}인 정규분포에서\n추출한 {}개 값 히스토그램'.format(m,s,n))
```

평균이 2, 표준편차 7인 정규분포에서  
추출한 3000개 값 히스토그램



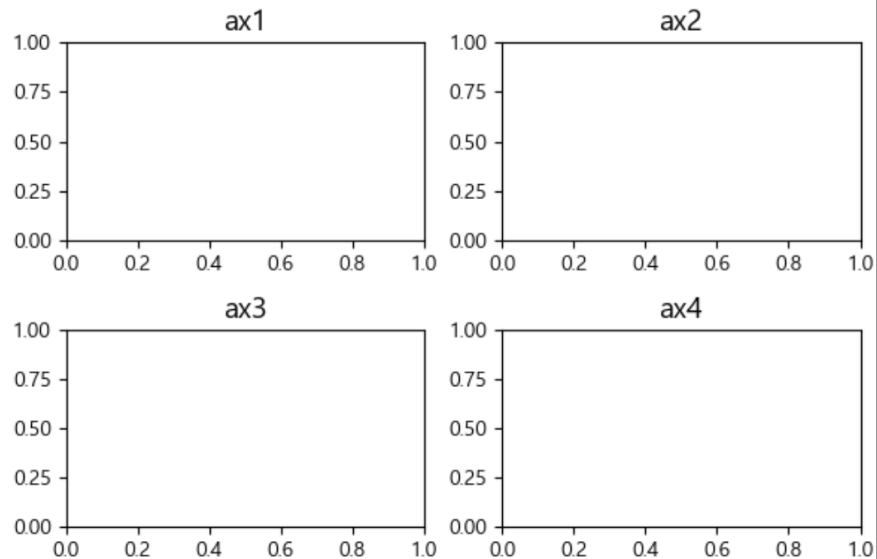
# n by m axs 에 flatten 사용하기

```
fig=plt.figure(figsize=(6,4), dpi=100)
axs=fig.subplots(2,2)
axs

_ =axs[0,0].set_title('ax1', fontsize=14)
_ =axs[0,1].set_title('ax2', fontsize=14)
_ =axs[1,0].set_title('ax3', fontsize=14)
_ =axs[1,1].set_title('ax4', fontsize=14)

fig.tight_layout()
```

```
array([[<Axes: >, <Axes: >],
       [<Axes: >, <Axes: >]], dtype=object)
```

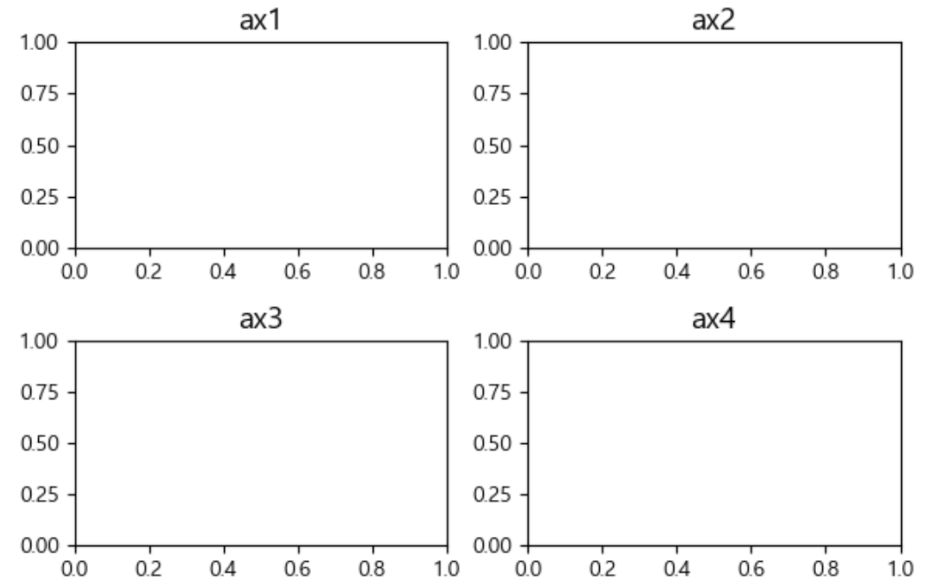


```
fig=plt.figure(figsize=(6,4), dpi=100)
axs=fig.subplots(2,2).flatten()
axs

_ =axs[0].set_title('ax1', fontsize=14)
_ =axs[1].set_title('ax2', fontsize=14)
_ =axs[2].set_title('ax3', fontsize=14)
_ =axs[3].set_title('ax4', fontsize=14)

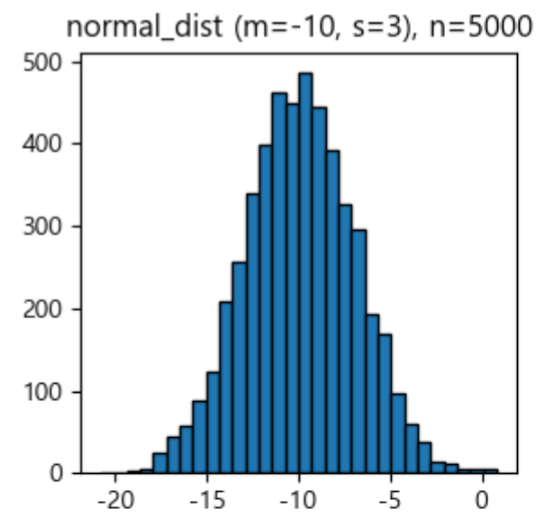
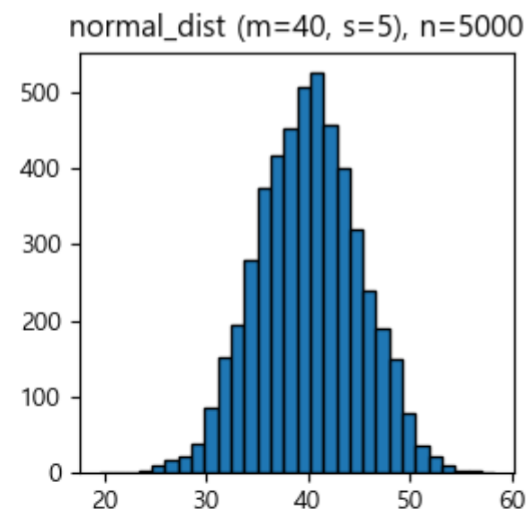
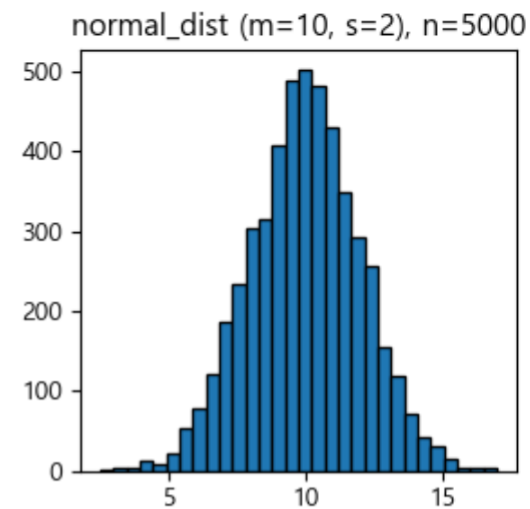
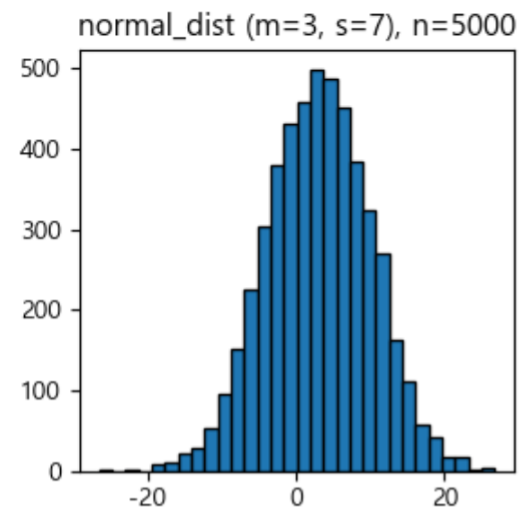
fig.tight_layout()
```

```
array([<Axes: >, <Axes: >, <Axes: >, <Axes: >], dtype=object)
```



## 실습2

- 네 개의 서로 다른 평균  $m$ , 표준편차  $s$ 인 정규분포에서 5000개 값 sampling 하여 각 axe에 히스토그램 그리기



레이텍

# 수식 제목 추가 (레이텍, LaTeX)

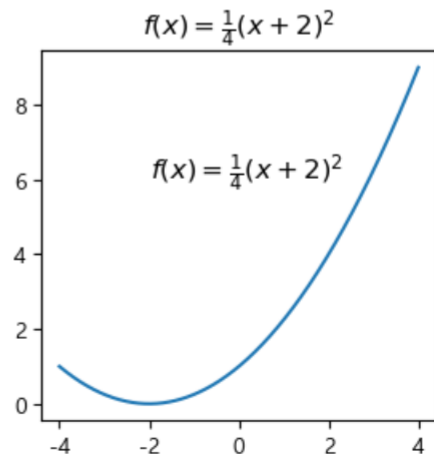
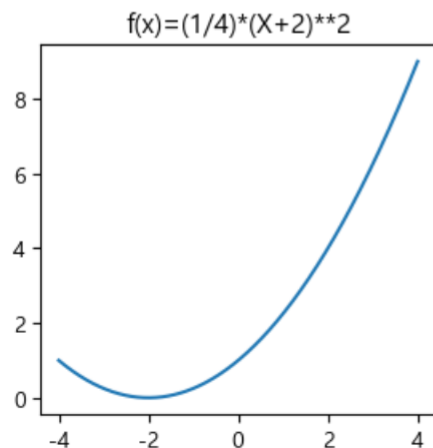
```
fig=plt.figure(figsize=(7,3), dpi=100)
ax1, ax2=fig.subplots(1,2)

X=np.linspace(-4,4,1000)
Y=(1/4)*(X+2)**2

_=ax1.plot(X,Y)
_=ax1.set_title('f(x)=(1/4)*(X+2)**2')

_=ax2.plot(X,Y)
_=ax2.set_title('$f(x)=\\frac{1}{4}(x+2)^2$')
_=ax2.text(-2,6,'$f(x)=\\frac{1}{4}(x+2)^2$', fontsize=12)
```

레이텍은 제목 뿐만 아니라,  
일반적인 문자열 표시에 사용



# 레이텍 (LaTeX) 기본 문법

## ■ 기본적인 라텍스 문법

### ■ 지수표기 방법

- $\$2x^3\$ \rightarrow 2x^3$
- $\$2^{\{3x+4\}}\$ \rightarrow 2^{3x+4}$
- $\$(2x)^3\$ \rightarrow (2x)^3$

### ■ 아래첨자 표기 방법

- $\$\log_2(x)\$ \rightarrow \log_2(x)$
- $\$\log_{\{10\}}(x)\$ \rightarrow \log_{10}(x)$

### ■ 그리스문자

- $\$\pi\$ \rightarrow \pi$
- $\$\alpha\$ \rightarrow \alpha$

(엔터 위의 역슬래쉬)

### ■ 제곱근

- $\$\sqrt{x+2}\$ \rightarrow \sqrt{x+2}$
- $\$\sqrt[3]{x+2}\$ \rightarrow \sqrt[3]{x+2}$

## ■ 분수

- $\$\frac{x^2}{x+3}\$ \rightarrow \frac{x^2}{x+3}$
- $\$f(x)=\frac{x+2}{x+3}\$ \rightarrow \underline{f(x)=\frac{x+2}{x+3}}$

(엔터 위의 역슬래쉬)

## ■ 삼각함수 및 지수함수 ([\sin](#) 사용하지 않으면 기울임꼴)

- $\$Y=\sin(X)\$ \rightarrow Y=\sin(X)$
- $\$Y=\exp(X)\$ \rightarrow \underline{Y=\exp(x)}$

라텍스 문법의 그룹 적용: 중괄호의 사용

# 레이텍 (LaTeX) 기본 문법

```
fig=plt.figure(figsize=(8,4), dpi=200)
ax=fig.subplots()

_=ax.set_title('$f(x)=\frac{1}{4}(x+2)^2$')
_=ax.text(0.1,0.1,'$2x^3$')
_=ax.text(0.1,0.3,'$2^{3x+4}$')
_=ax.text(0.1,0.5,'$(2x)^3$')

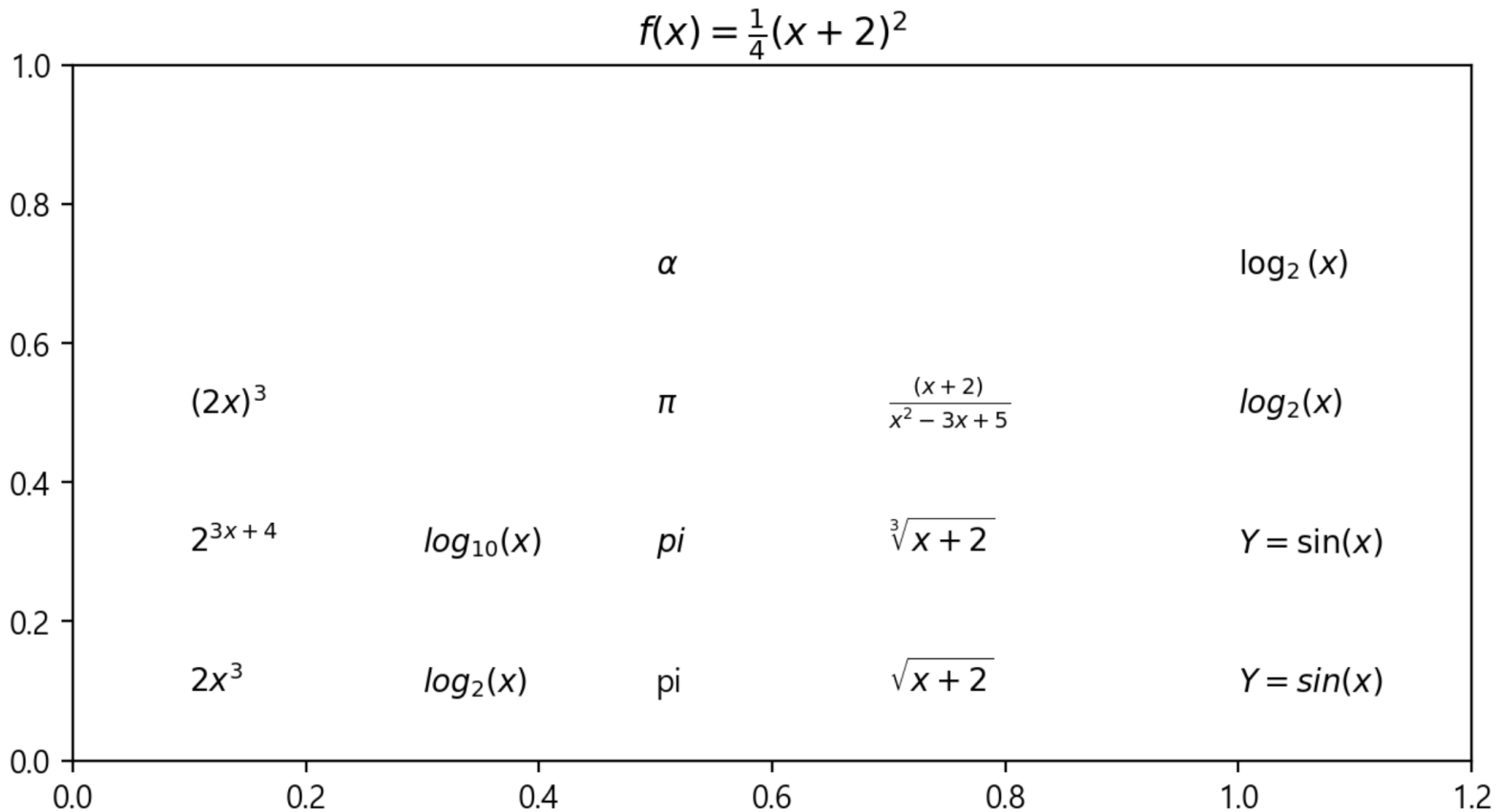
_=ax.text(0.3,0.1,'$\log_2(x)$')
_=ax.text(0.3,0.3,'$\log_{10}(x)$')

_=ax.text(0.5,0.1,'$\pi$')
_=ax.text(0.5,0.3,'$\pi$')
_=ax.text(0.5,0.5,'$\pi$')
_=ax.text(0.5,0.7,'$\alpha$')

_=ax.text(0.7,0.1,'$\sqrt{x+2}$')
_=ax.text(0.7,0.3,'$\sqrt[3]{x+2}$')
_=ax.text(0.7,0.5,'$\frac{(x+2)}{x^2-3x+5}$')

_=ax.text(1,0.1,'$Y=\sin(x)$')
_=ax.text(1,0.3,'$Y=\sin(x)$')
_=ax.text(1,0.5,'$\log_2(x)$')
_=ax.text(1,0.7,'$\log_2(x)$')

_=ax.set_xlim(0,1.2)
```

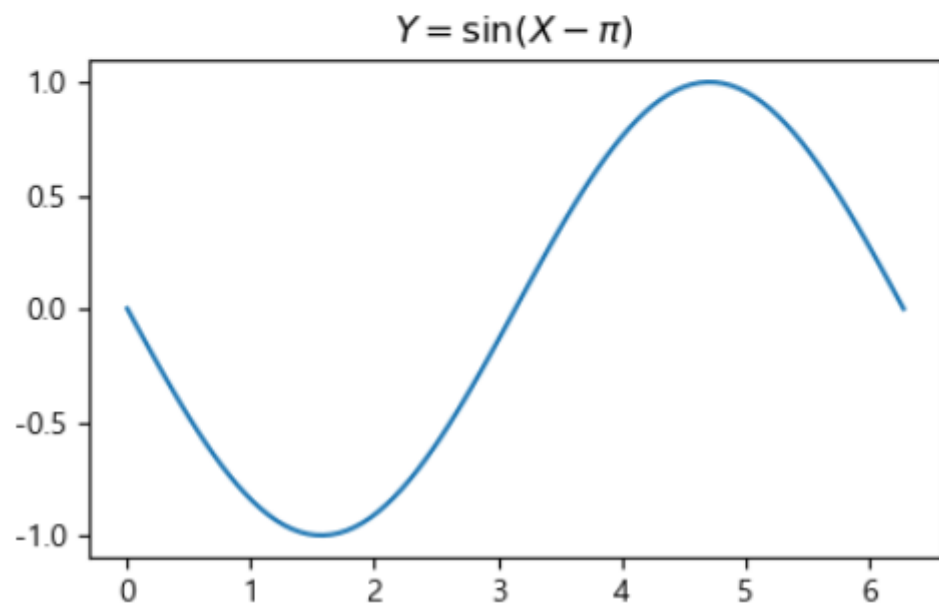


# 수식 제목 추가 (레이텍)

```
fig=plt.figure(figsize=(5,3), dpi=100)
ax=fig.subplots()

X=np.linspace(0,2*np.pi, 100)
Y=np.sin(X-np.pi)

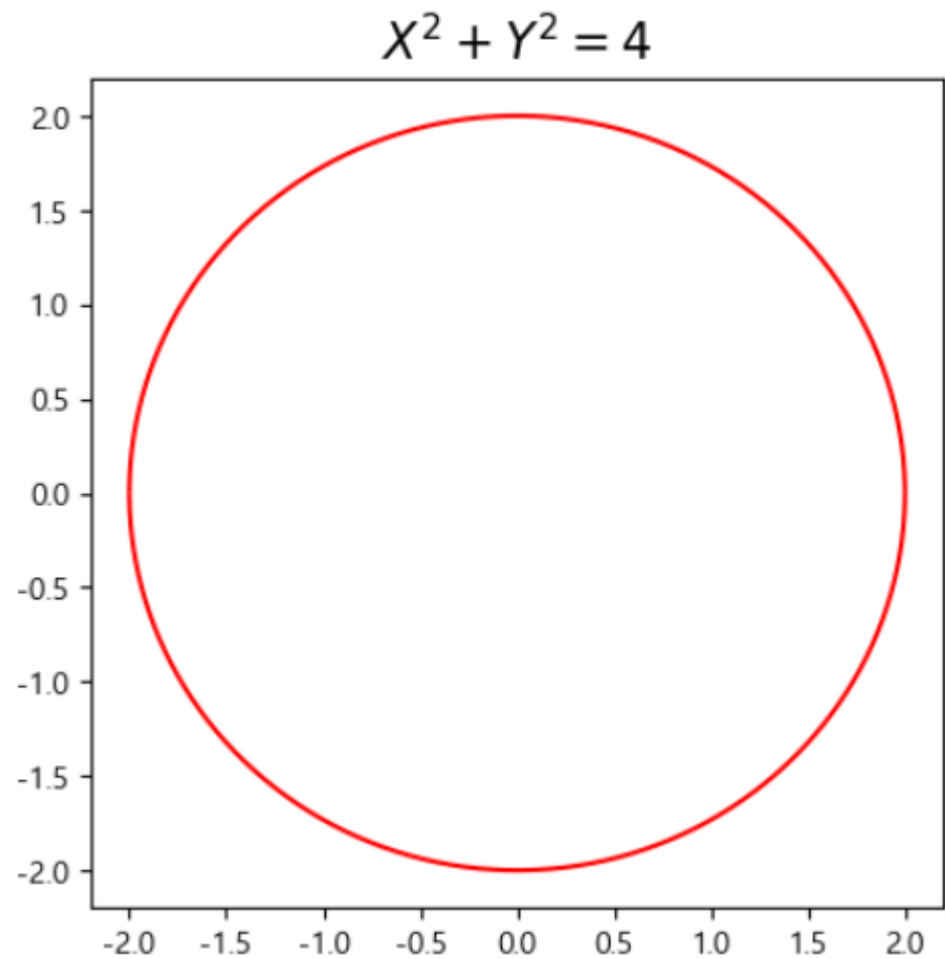
_=ax.plot(X,Y)
_=ax.set_title("$Y=\sin(X-\pi)$")
```



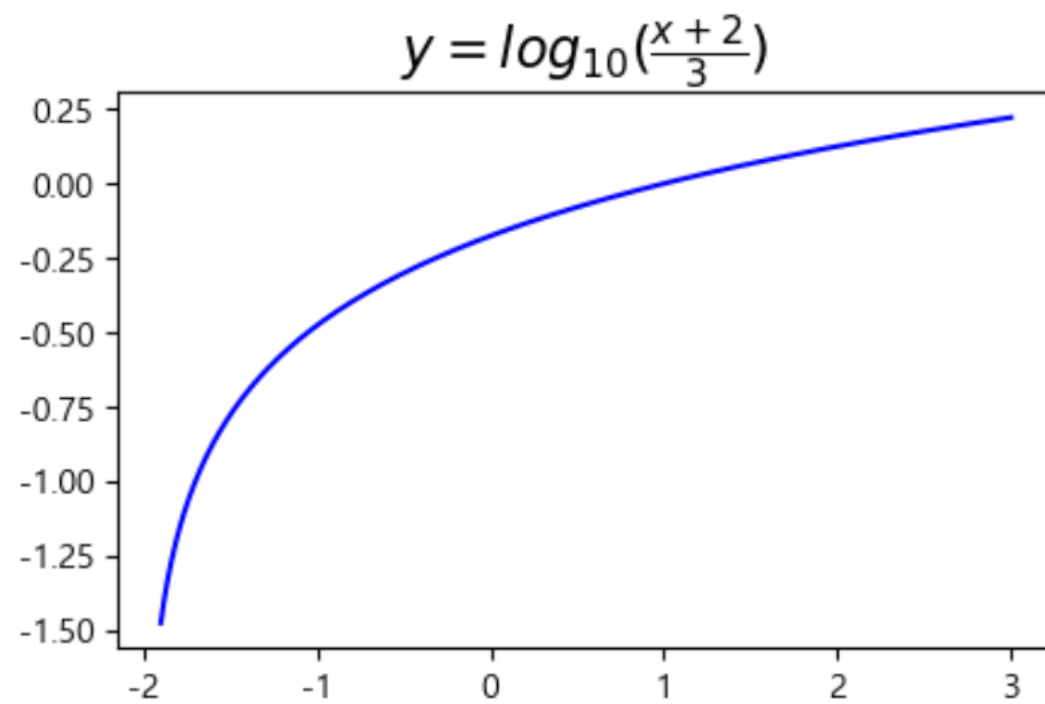


## 실습3

아래와 같은 원을 그리고 원의 방정식을 제목에 넣으시오  
(원은 아랫 반원, 윗 반원 나누어서 그리시오)



## 실습4



# 실습5 (여러 axes에 다른 title 작성)

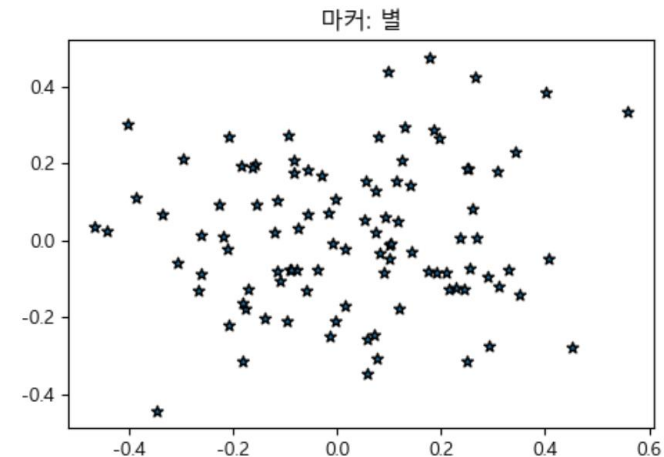
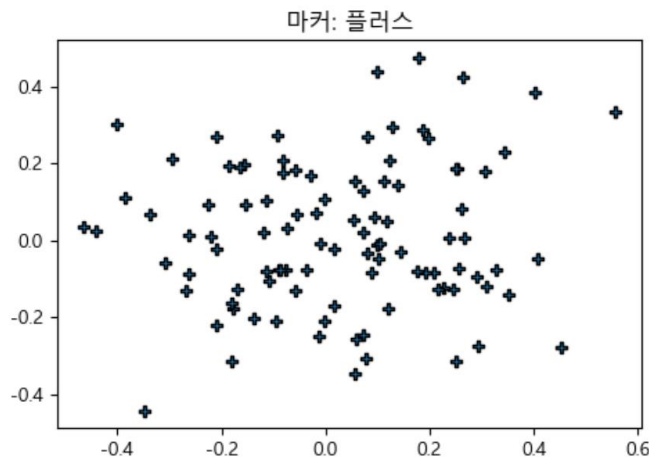
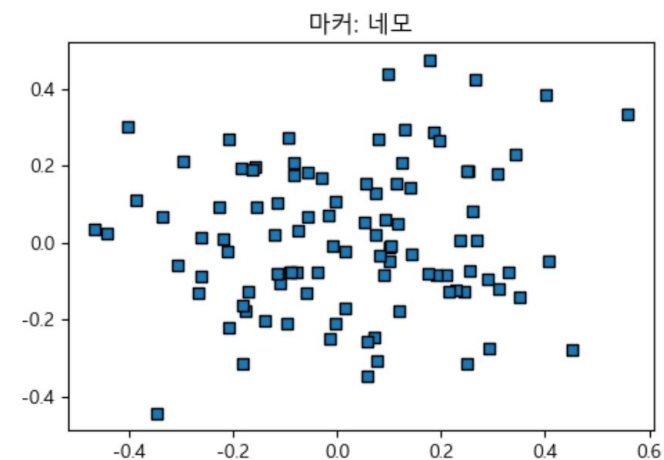
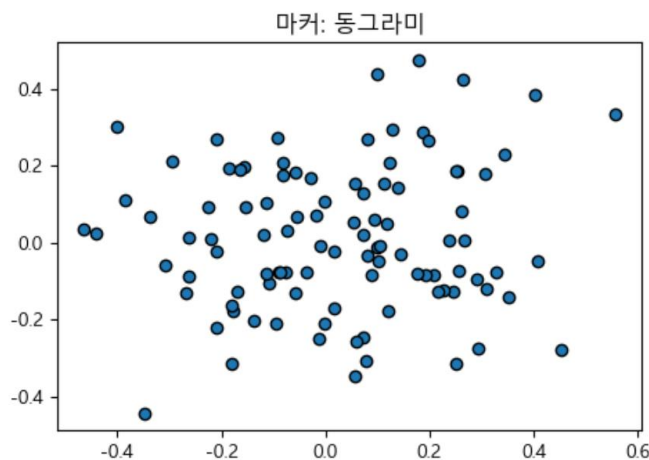
```
marks = np.array(['o', 's', 'P', '*'])
```

`np.random.normal(0,0.2,size=(100,2))`를

`marks`에 있는 각 마커들로

각 axes에 scatter 하고,

각 마커를 한글로 제목에 표시하시오



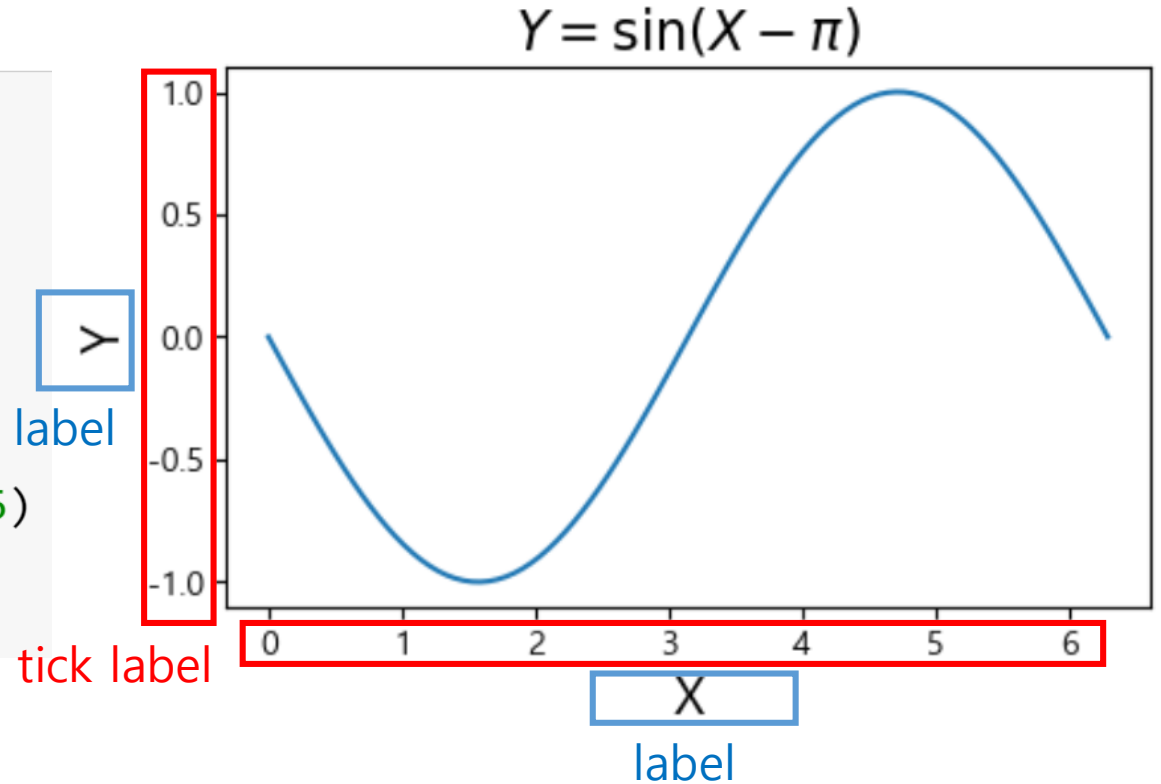
$x, y$  label

# x,y label

```
fig=plt.figure(figsize=(5,3), dpi=100)
ax1=fig.subplots()

X=np.linspace(0,2*np.pi, 100)
Y=np.sin(X-np.pi)

ax1.plot(X,Y)
ax1.set_title("$Y=\\sin(X-\\pi)$", fontsize=16)
ax1.set_xlabel('X', fontsize=16)
ax1.set_ylabel('Y', fontsize=16)
```



# 실습6

```
df1=pd.DataFrame({'A':[1,2,3,4,5],  
                  'B':[5,6,7,8,9],  
                  'C':[5,4,3,2,1],  
                  'D':[1,9,7,5,3],  
                  'E':[7,4,1,2,5]})
```

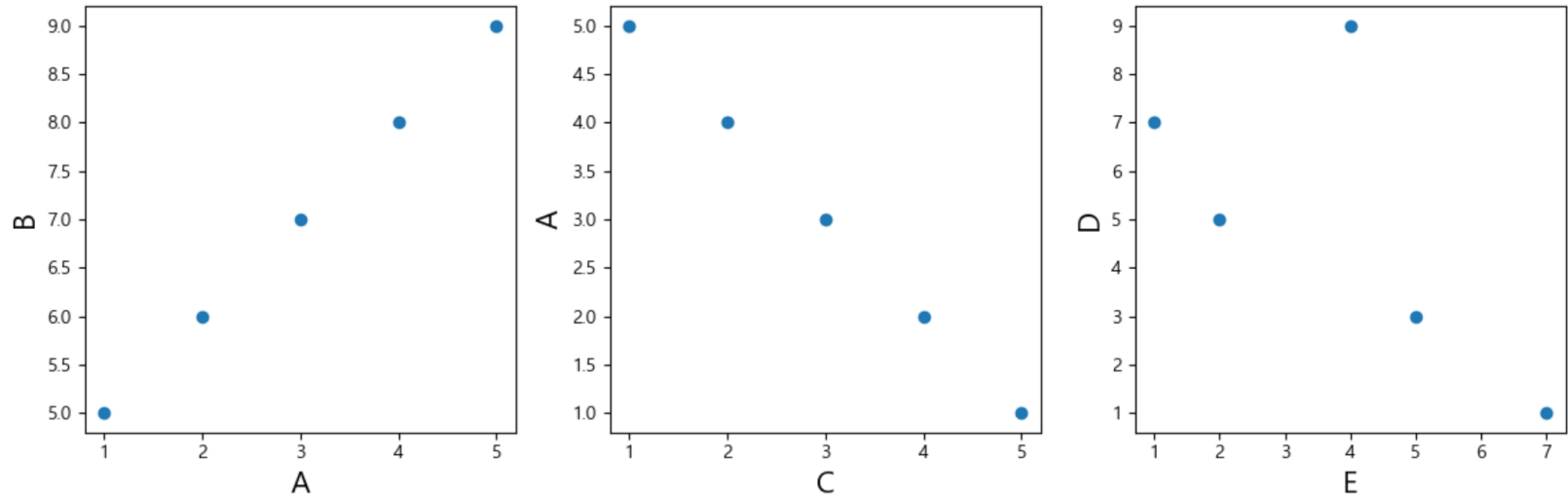
```
fig=plt.figure(figsize=(12,4), dpi=100)  
axs=fig.subplots(1,3)
```

## 코드작성

```
scatter_func('A','B',axs[0])  
scatter_func('C','A',axs[1])  
scatter_func('E','D',axs[2])
```

```
fig.tight_layout()
```

아래와 같은 그림이 나오도록 scatter\_func을 구현하시오



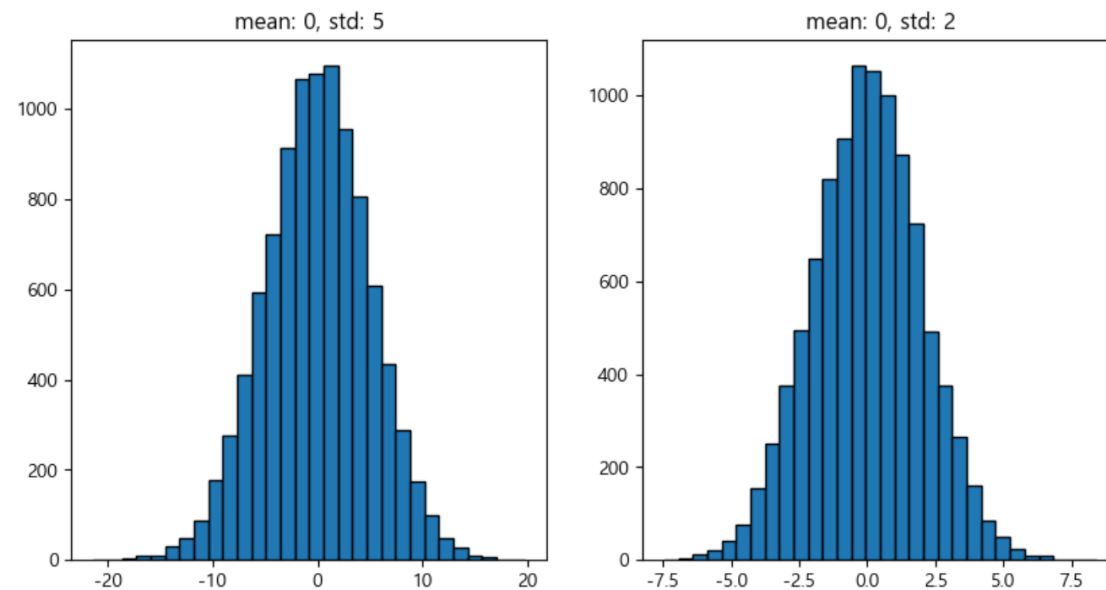
$x, y$  lim

# xlim, ylim

```
fig=plt.figure(figsize=(10,5), dpi=100)
ax1, ax2=fig.subplots(1,2)

m1=0
s1=5
data=np.random.normal(loc=m1,scale=s1,size=10000)
_=ax1.hist(data, bins=30, edgecolor='k')
_=ax1.set_title('mean: {}, std: {}'.format(m1, s1))
#_=ax1.set_xlim(-20,20)

m2=0
s2=2
data=np.random.normal(loc=m2,scale=s2,size=10000)
_=ax2.hist(data, bins=30, edgecolor='k')
_=ax2.set_title('mean: {}, std: {}'.format(m2, s2))
#_=ax2.set_xlim(-20,20)
```



x의 범위를 정확히 확인하지  
않으면, 두 분포가 같은 분  
포인 줄 착각할 수 있다.

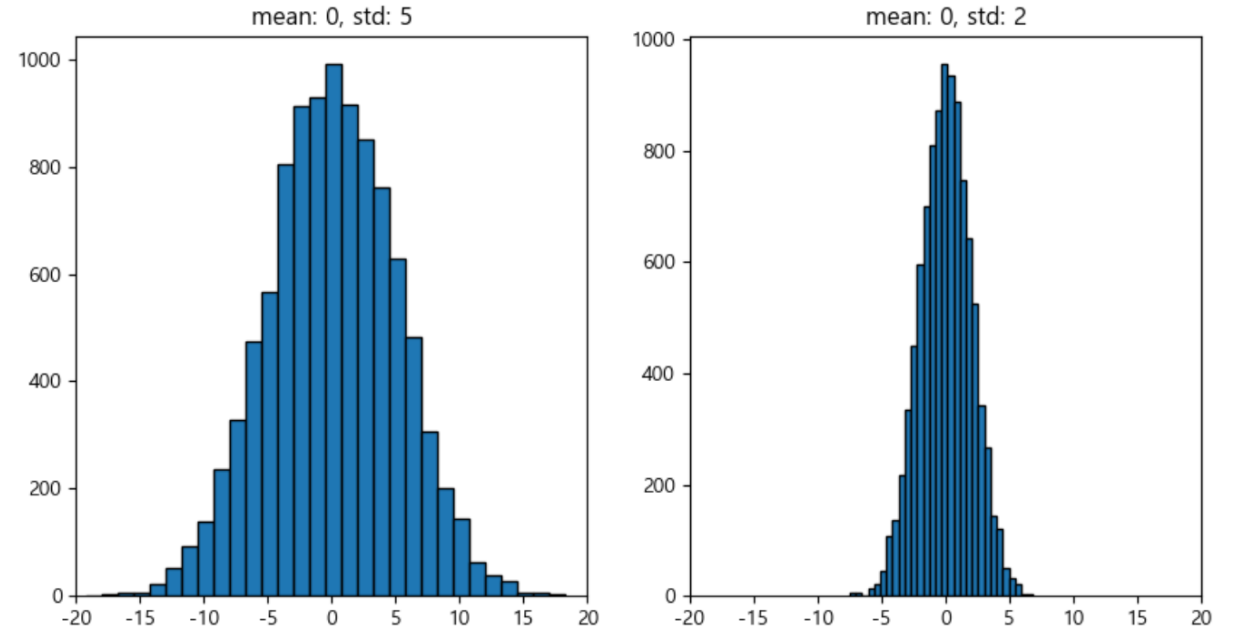


# xlim, ylim

```
fig=plt.figure(figsize=(10,5), dpi=100)
ax1, ax2=fig.subplots(1,2)

m1=0
s1=5
data=np.random.normal(loc=m1,scale=s1,size=10000)
_=ax1.hist(data, bins=30, edgecolor='k')
_=ax1.set_title('mean: {}, std: {}'.format(m1, s1))
_=ax1.set_xlim(-20,20)

m2=0
s2=2
data=np.random.normal(loc=m2,scale=s2,size=10000)
_=ax2.hist(data, bins=30, edgecolor='k')
_=ax2.set_title('mean: {}, std: {}'.format(m2, s2))
_=ax2.set_xlim(-20,20)
```



$x, y$  tick

# xtick, ytick

```
fig=plt.figure(figsize=(5,5), dpi=100)
ax=fig.subplots()

data = np.array([5,25,50,20])
X=np.arange(len(data))
_=ax.bar(X,data)
ax.get_xticks()

## 방법1
# _=ax.set_xticks(X, labels=['국','영','수','과']) # 수행안하면 이상한 tick 생성

## 방법2
_=ax.set_xticks(X)
_=ax.set_xticklabels(['국','영','수','과'])

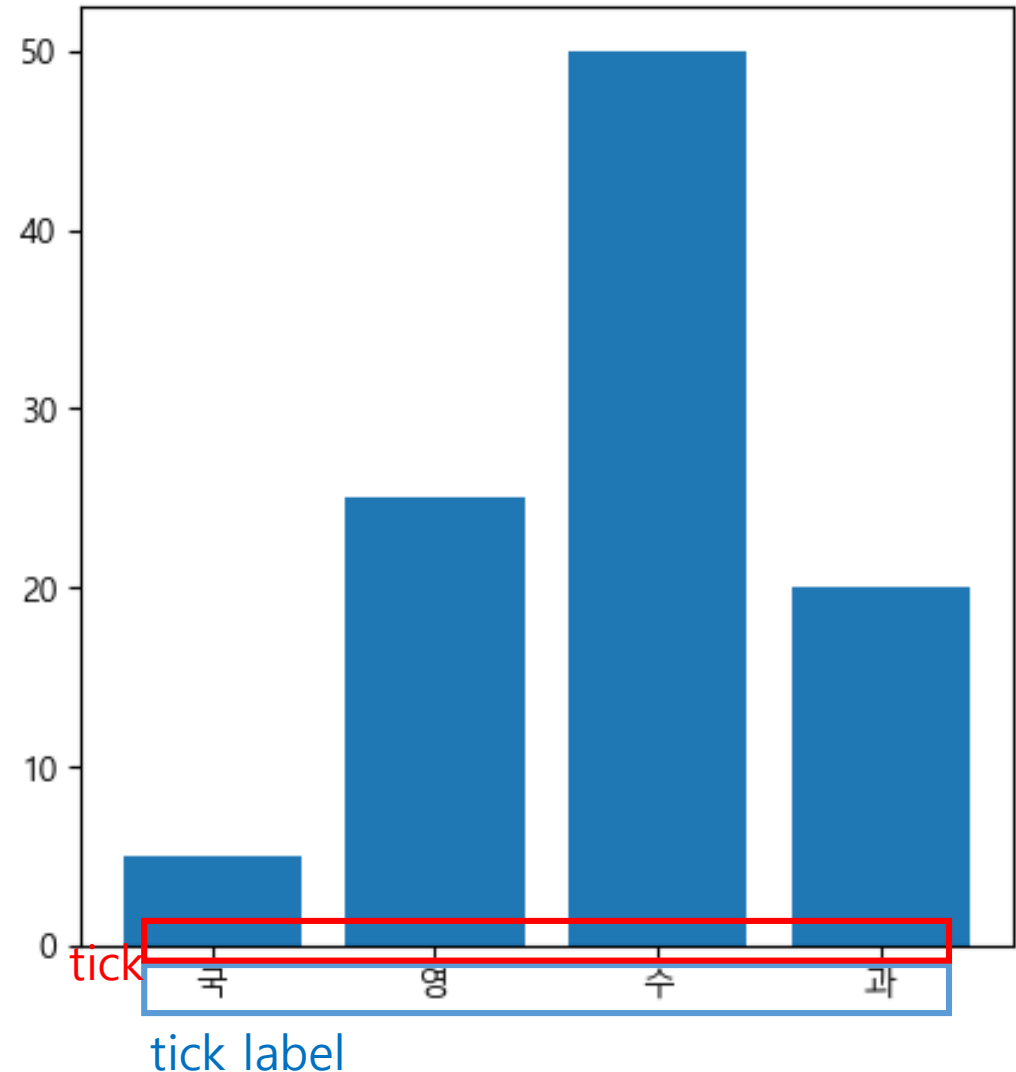
## 방법3
# _=ax.bar(['국','영','수','과'],data) # _=ax.bar(X,data) 대신

array([-1. , -0.5,  0. ,  0.5,  1. ,  1.5,  2. ,  2.5,  3. ,  3.5,  4. ])
```

axes 크기에 따라 미리정해진 일정한 간격의 tick이 생성되어 있다.  
따라서 바로 set\_xticklabels를 사용하면 이상한 결과가 나옴



1. 우선 xticks를 새롭게 설정한다.
2. 설정된 각 xtick에 label을 할당한다.  
(이때, xticks의 길이와 label의 길이는 같아야 한다.)

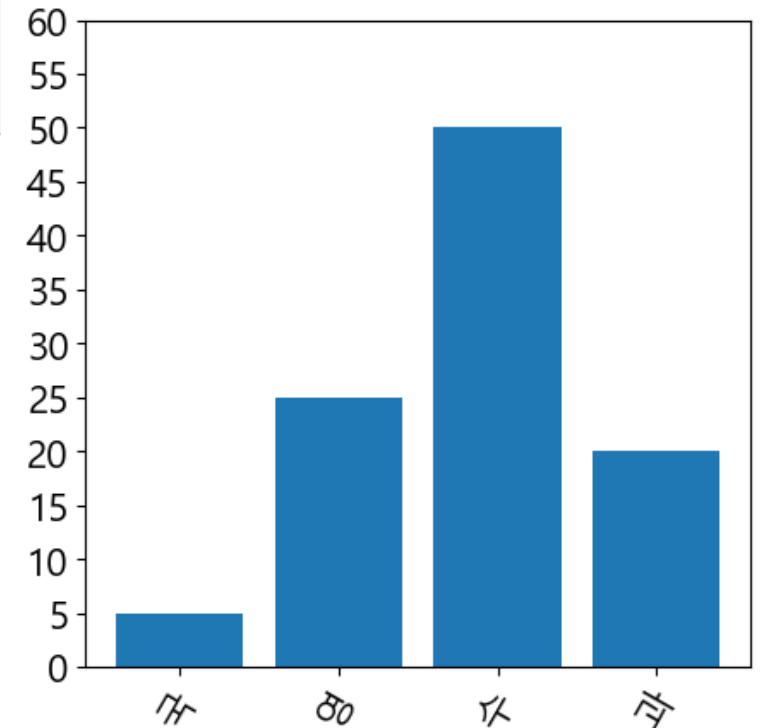


# xtick, ytick

```
fig=plt.figure(figsize=(5,5), dpi=100)
ax=fig.subplots()

data = np.array([5,25,50,20])
X=range(len(data))
_=ax.bar(X,data)
_=ax.set_xticks(X)
_=ax.set_xticklabels(['국','영','수','과'], rotation=60, fontsize=16)

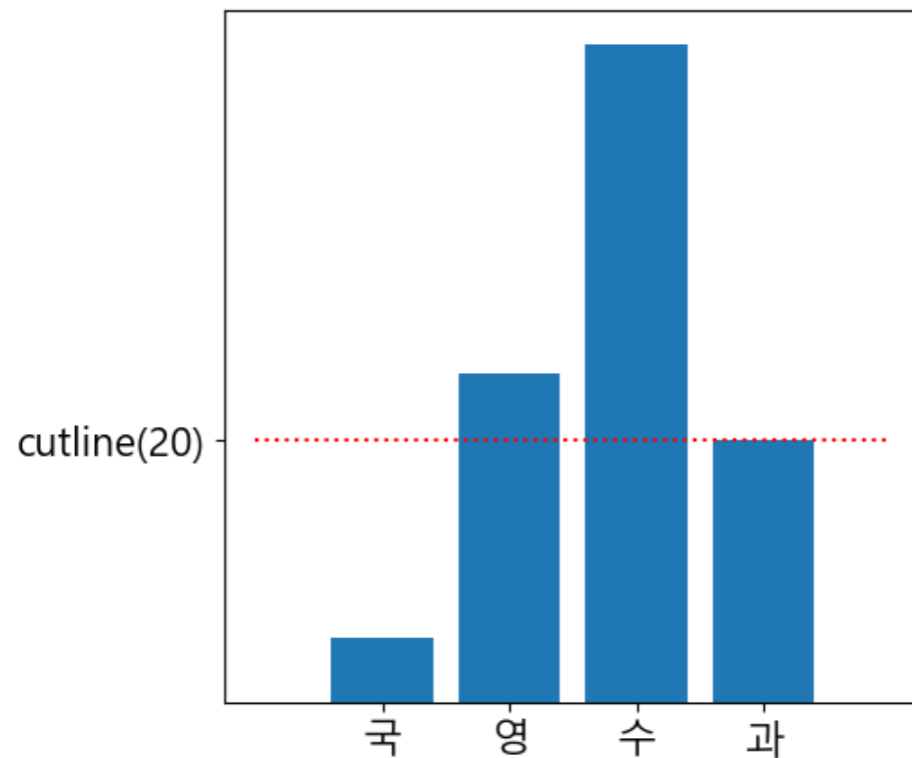
_=ax.set_yticks(range(0, 61, 5))
_=ax.set_yticklabels(range(0, 61, 5), fontsize=16)
```



# 실습7

- 주어진 data, subj, cut 정보를 이용하여 아래와 같은 그림을 그리시오

```
data = np.array([5,25,50,20])  
subj = ['국', '영', '수', '과']  
cut=20
```



# xtick, ytick

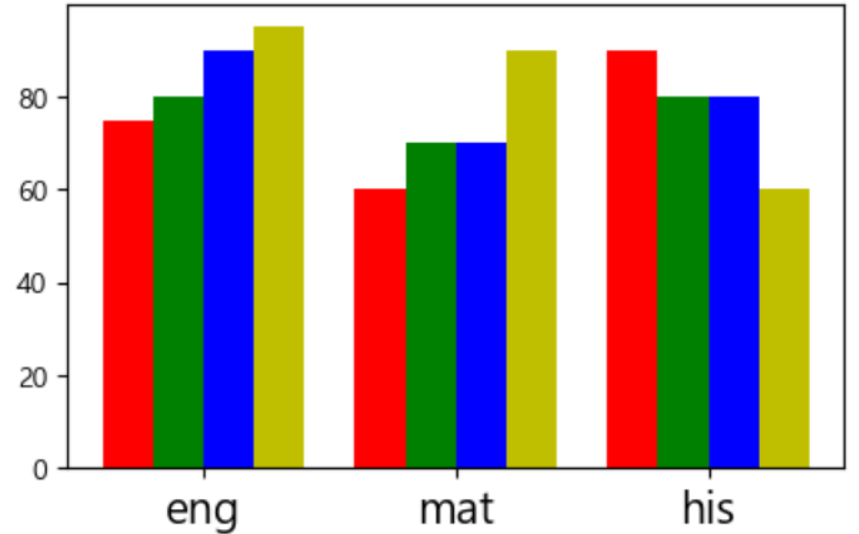
```
fig=plt.figure(figsize=(5,3), dpi=100)
ax=fig.subplots()

data = pd.read_table('data/dat_bar.txt',index_col=0,sep='\t')
data.head()

w1=0.2
X=np.arange(3)

_=ax.bar(X+0.00, data.loc['1st'], color='r',width=w1)
_=ax.bar(X+w1*1, data.loc['2nd'], color='g', width=w1)
_=ax.bar(X+w1*2, data.loc['3rd'], color='b', width=w1)
_=ax.bar(X+w1*3, data.loc['4th'], color='y', width=w1)

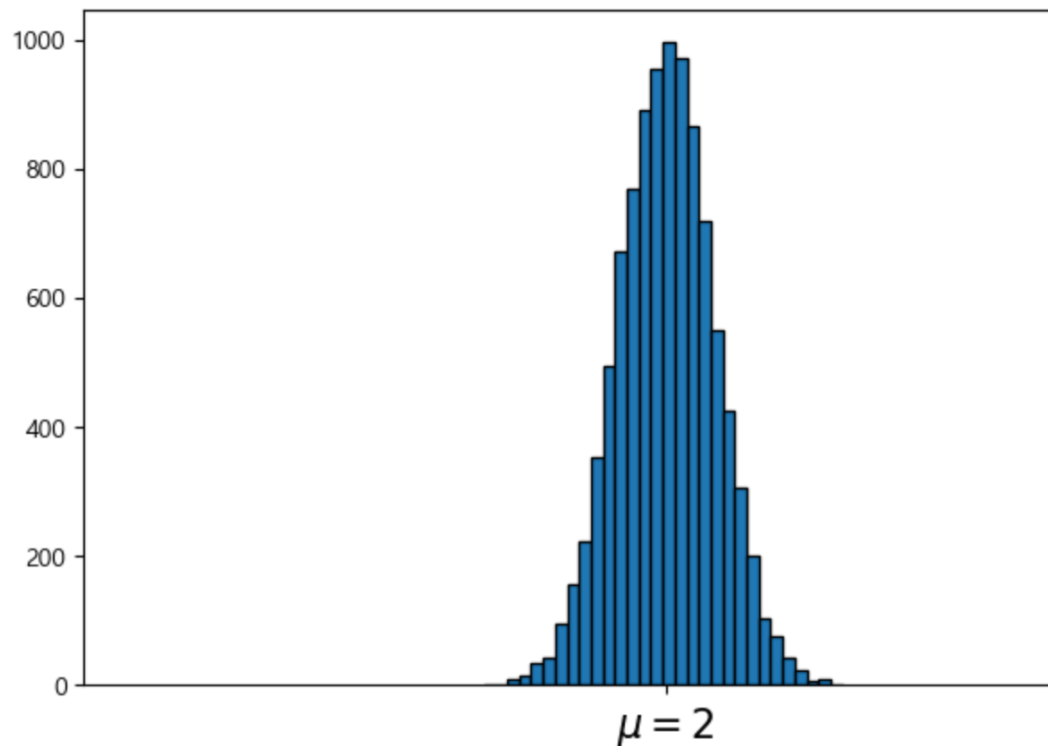
_=ax.set_xticks(X+1.5*w1, labels=data.columns, fontsize=16)
```



## 실습8

$m1=2$ ,  $s1=1$  로 주어졌을 때, (주어진 값은 변경 가능)

- 1) 평균이  $m1$ , 표준편차가  $s1$  인 정규분포에서 10,000개의 값을 sampling 한 후에 아래와 같이 histogram을 그리시오
- 2) x ticks, x ticklabels을 그림과 같이 수정하시오 ( $m$ 이 아니라 그리스문자  $\mu$ 임을 주의)
- 3) x의 limit을 -10~10으로 제한하시오



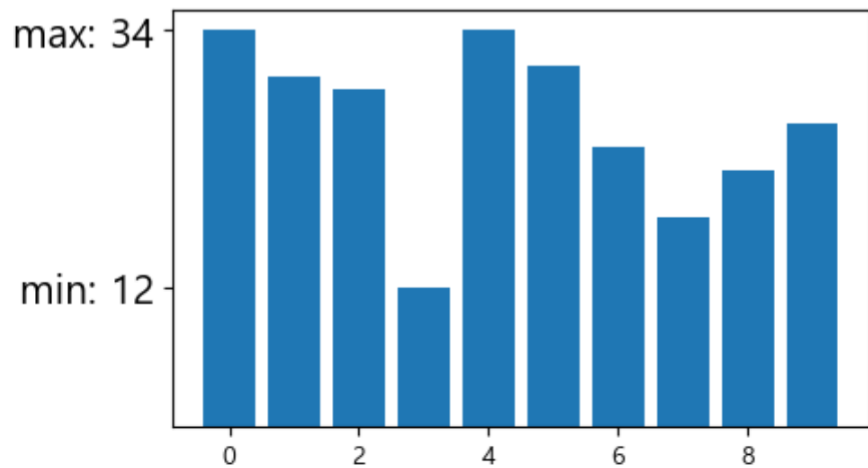
# 실습9

```
# 주어진 data를 bar그래프로 그리고,  
# data의 최대, 최소값을 y tick에 그림과 같이 표시하시오  
# h1) np.max(), np.min() 함수 사용
```

```
fig=plt.figure(figsize=(5,3), dpi=100)  
ax=fig.subplots()
```

```
data = np.random.randint(10,40,size=10)  
_=ax.bar(range(len(data)),data)
```

```
#=== 코드작성 ===#
```





## Q & A

Thank you