

데이터 시각화 (2024)

데이터과학부 정진명

(jmjung@suwon.ac.kr, 글로벌경상관 918호)

7 주차

- 2~5w: **plot, scatter, bar, pie, hist, boxplot** 기본 활용
- 6w: **plot, bar, scatter**의 **color, alpha** 파라미터 활용
- 7w: **plot, bar, scatter**의 **color, alpha** 이외 **다른 스타일 파라미터 활용**

		필수 파라미터 (2~5w)	기본활용 (2~5w)	컬러 & 스타일 (6, 7w)
plot	X, Y 시각화	X, Y	fmt	color, alpha, linestyle, marker, markerevery
scatter		X, Y	s (size)	color, alpha, edgecolor, marker
bar		X, Y (height)	width, align, bottom	color, alpha, edgecolor, linestyle, hatch (중간고사 이후)
pie	분포 시각화	X	labels, explode	
hist		X	bins, density, edgecolor	
boxplot		X	positions	

plot

line style, color, marker

Line Styles

character		description
' - '	solid	solid line style
' - - '	dashed	dashed line style
' - . '	dashdot	dash-dot line style
' : '	dotted	dotted line style

Colors

The supported color abbreviations are the single letter codes

character	color
'b'	blue
'g'	green
'r'	red
'c'	cyan
'm'	magenta
'y'	yellow
'k'	black
'w'	white

https://matplotlib.org/3.1.1/api/_as_gen/matplotlib.axes.Axes.plot.html

Markers

character	description
'.'	point marker
','	pixel marker
'o'	circle marker
'v'	triangle_down marker
'^'	triangle_up marker
'<'	triangle_left marker
'>'	triangle_right marker
'1'	tri_down marker
'2'	tri_up marker
'3'	tri_left marker
'4'	tri_right marker
's'	square marker
'p'	pentagon marker
'*'	star marker
'h'	hexagon1 marker
'H'	hexagon2 marker
'+'	plus marker
'x'	x marker
'D'	diamond marker
'd'	thin_diamond marker
' '	vline marker
'_'	hline marker

plot (different line style)

```
fig=plt.figure(figsize=(12,5), dpi=100)
ax1, ax2=fig.subplots(1,2)

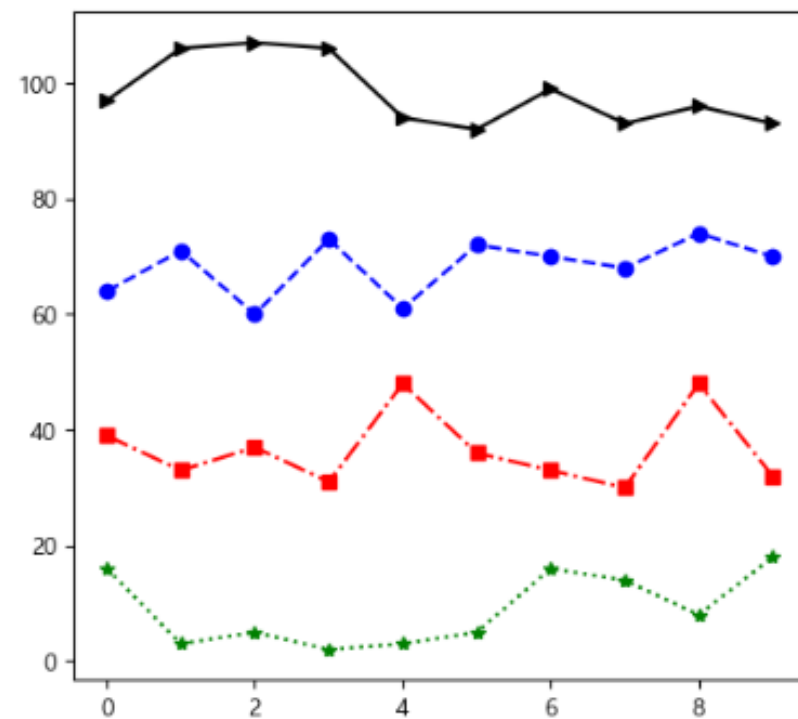
X=np.arange(10)

Y1=np.random.randint(90,110,size=10)
Y2=np.random.randint(60,80,size=10)
Y3=np.random.randint(30,50,size=10)
Y4=np.random.randint(0,20,size=10)

_ax1.plot(X, Y1, color = 'k', linestyle='solid', marker='>')
_ax1.plot(X, Y2, color = 'b', linestyle='dashed', marker='o')
_ax1.plot(X, Y3, color = 'r', linestyle='dashdot', marker='s')
_ax1.plot(X, Y4, color = 'g', linestyle='dotted', marker='*')

_ax2.plot(X, Y1, 'k->')
_ax2.plot(X, Y2, 'b--o')
_ax2.plot(X, Y3, 'r-.s')
_ax2.plot(X, Y4, 'g:*')
```

- color, linestyle, marker를 기호로 한번에 표시 (fmt parameter)
- **plot 만 가능**



markevery 파라미터

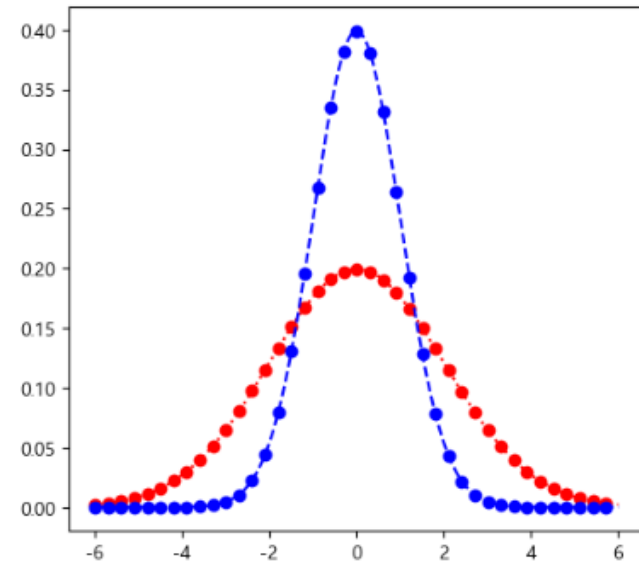
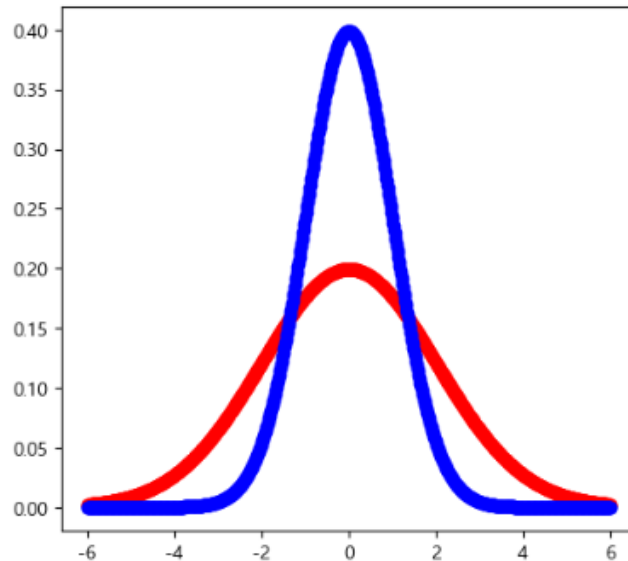
```
fig=plt.figure(figsize=(12,5), dpi=100)
ax1, ax2=fig.subplots(1,2)

def pdf(X, mu, sigma):
    a = 1/(sigma * np.sqrt(2*np.pi))
    b = -1/(2*(sigma**2))
    return a * np.exp(b * ((X - mu)** 2))

X=np.linspace(-6,6,1000)

## ax1
Y=pdf(X, 0, 2)
_=ax1.plot(X, Y,'o:r')
Y=pdf(X, 0, 1)
_=ax1.plot(X, Y, 'o--b')

## ax2
Y=pdf(X, 0, 2)
_=ax2.plot(X, Y,'o:r', markevery=25)
Y=pdf(X, 0, 1)
_=ax2.plot(X, Y, 'o--b', markevery=25)
```



scatter

실습1

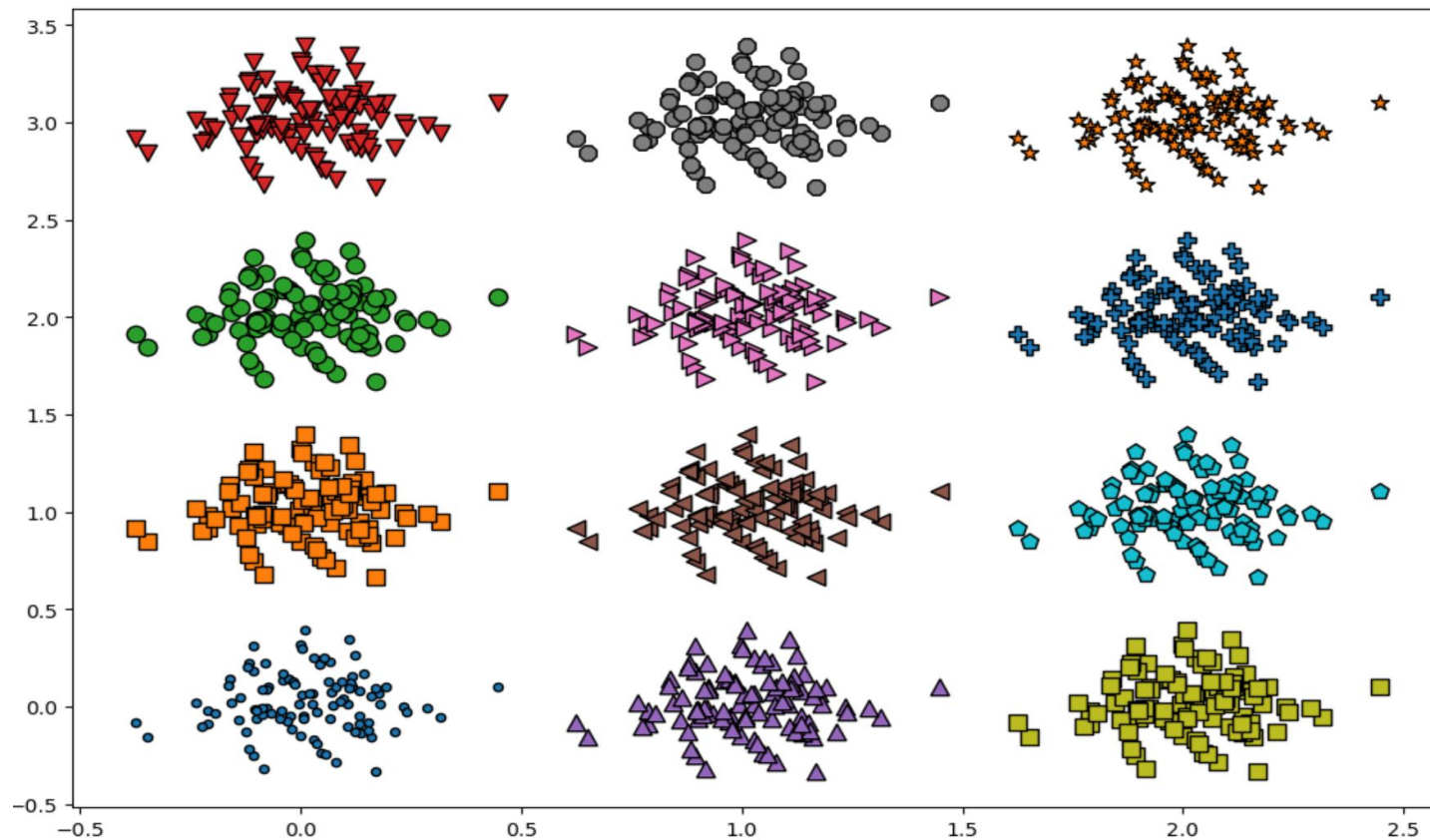
```
marks = np.array(['.', ',', 'o', 'v', '^', '<', '>', '8', 's', 'p', 'P', '*']).reshape(3,4)
marks

fig=plt.figure(figsize=(12,8), dpi=100)
ax=fig.subplots()

A=np.random.normal(0,0.15,size=(100,2))
```

주어진 A를 아래와 같이 나타내시오

- 이 중 for loop 사용
- marks matrix를 indexing으로 접근



scatter marker 사용 예제

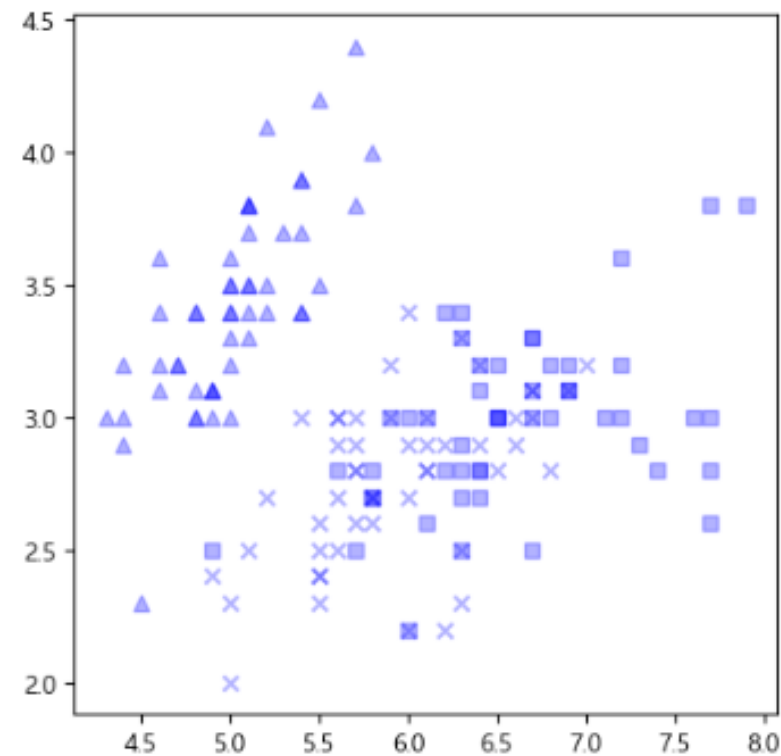
- 오른쪽의 iris data의 'sep_len', 'sep_wid' column을 x,y로 scatter 하는데, 'name' column의 꽃 이름 별로 표식 스타일을 다르게 하시오

- 1) Iris-setosa -> ^
 - 2) Iris-versicolor -> x
 - 3) Iris-virginica -> s
- (색깔은 모두 옅은 파랑색, alpha=0.3)

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

data = pd.read_table('data/dat_iris.txt', sep=',')
data
## ax1
for mark, label in zip(['^','x','s'], ['Iris-setosa', 'Iris-versicolor', 'Iris-virginica']):
    d1=data.loc[data['name']==label]
    ax1.scatter(d1['sep_len'], d1['sep_wid'], marker=mark, color='b', alpha=0.3)

## ax2
marker_dic={'Iris-setosa':'^', 'Iris-versicolor':'x', 'Iris-virginica':'s'}
for ind in data.index:
    row1=data.loc[ind]
    mark=marker_dic[row1['name']]
    _=ax2.scatter(row1['sep_len'], row1['sep_wid'], marker=mark, color='b', alpha=0.3)
```



중복된 것도 확인 가능

- marker 파라미터는 시퀀스 형태의 parameter를 받지 않는다.
(color는 가능, 다음 슬라이드 참고)

Iris scatter plot (6w 강의자료 중)

- 오른쪽의 iris data의 'sep_len', 'sep_wid' column을 x,y로 scatter 하는데, 'name' column의 꽃 이름 별로 색깔을 다르게 하시오
- 꽃 이름과 그 색깔:
 - 1) Iris-setosa -> red
 - 2) Iris-versicolor -> green
 - 3) Iris-virginica -> blue

Iris data

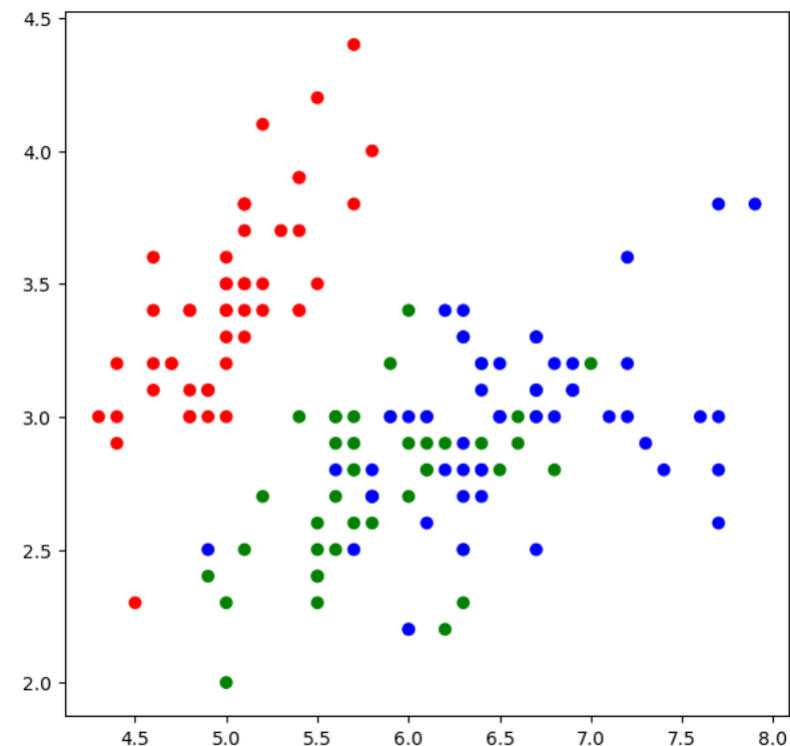
```
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
```

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

## data load
data = pd.read_table('data/dat_iris.txt', sep=',')
data.head()
color_dic={'Iris-setosa':'r','Iris-versicolor':'g','Iris-virginica':'b'}

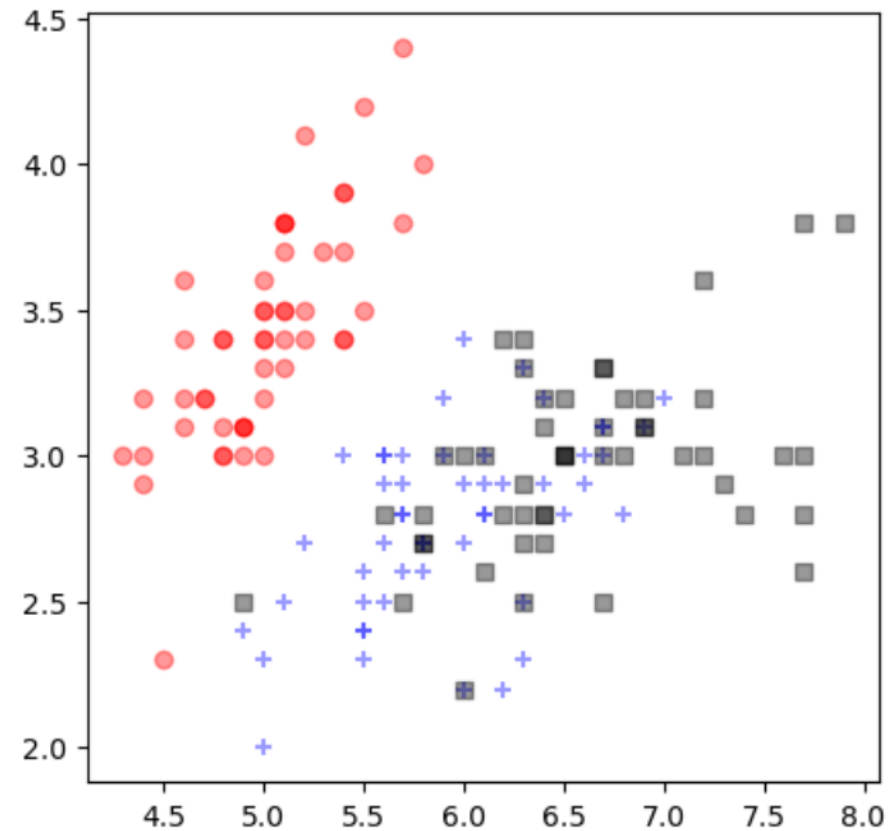
## ax1 (시퀀스 형태의 color parameter)
color=data['name'].replace(color_dic)
color
ax1.scatter(data['sep_len'], data['sep_wid'], c = color, s=10)
color parameter가 sequence자료

## ax2 (값 하나의 color parameter)
for n1 in color_dic.keys():
    d1=data.loc[data['name']==n1]
    ax2.scatter(d1['sep_len'], d1['sep_wid'], c = color_dic[n1], s=10)
color parameter가 하나의 값
```



실습2

- 이전 슬라이드 예제에서 조금 더 추가하여,
*Iris-setosa*는 빨간색 동그라미,
*Iris-versicolor*는 파란색 십자가가
*Iris-virginica*는 검은색 네모
로 그리시오 ($\alpha=0.4$)



표식 크기 제어

- 표식크기는 s 파라미터로 조절하며, 리스트를 argument로 받을 수 있다.

```
M=np.random.normal(0,1, size=(1000,2))
R=np.sum(M**2, axis=1)
M[:,2]
R[:,2]

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

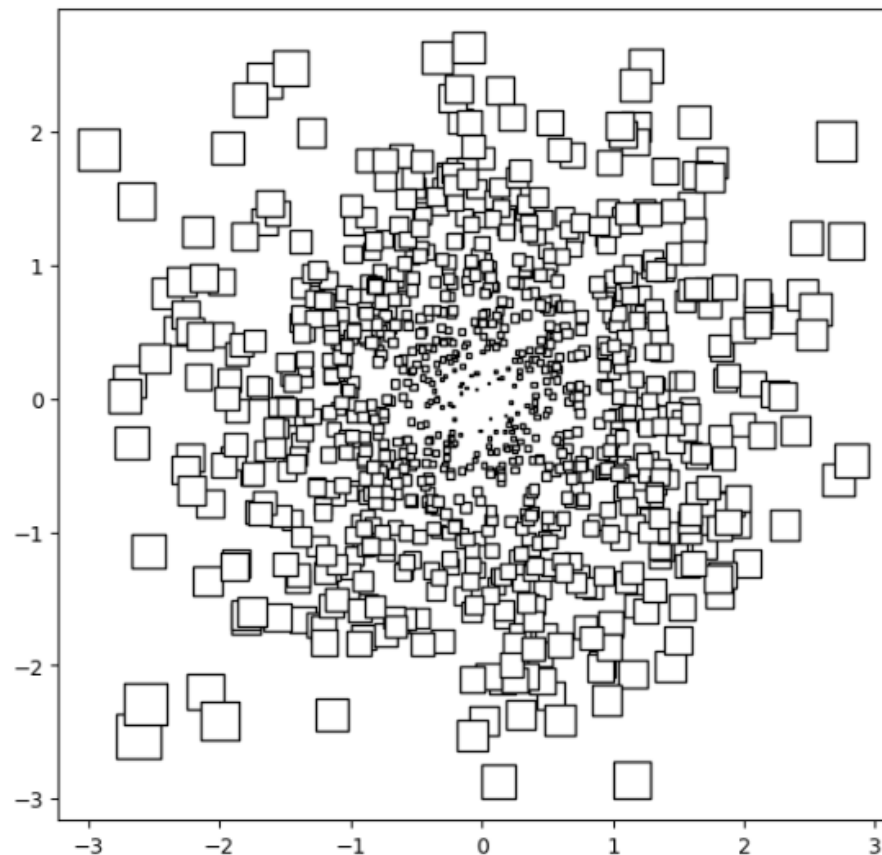
_=ax.scatter(M[:,0], M[:,1], c='w', edgecolor='k',marker='s', s=32*R)

array([[ -1.39576503,  0.02958246],
       [-0.48229261, -1.68083169]])

array([1.94903513, 3.05780132])
```

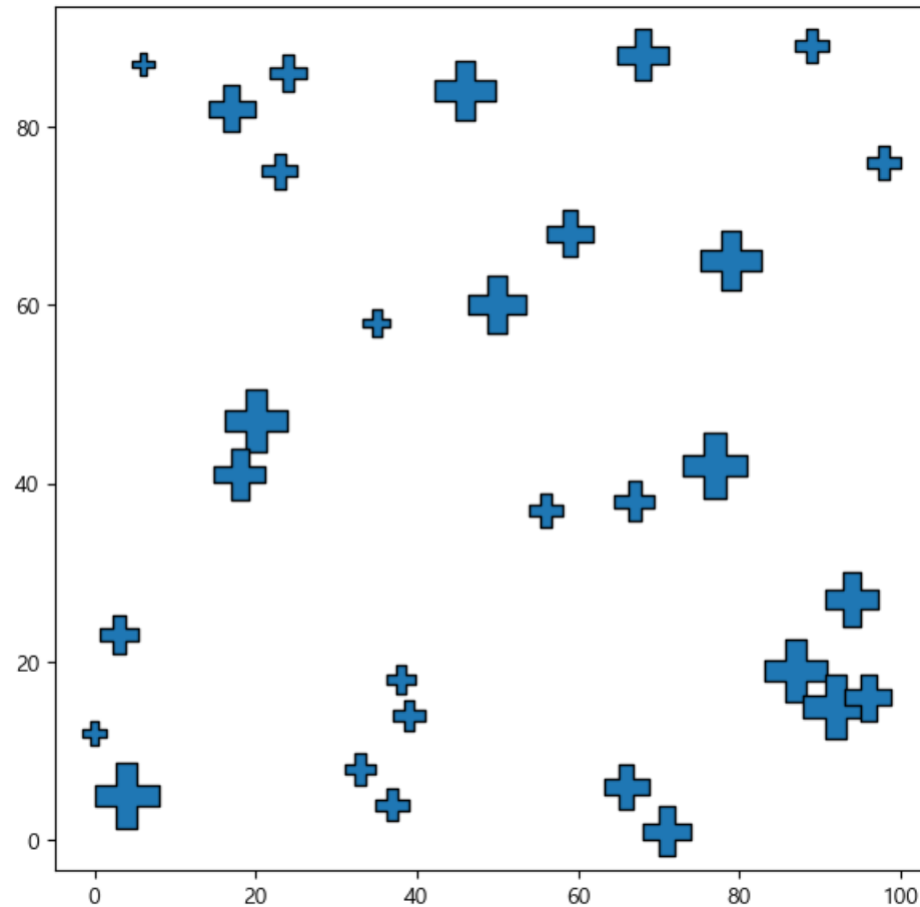
```
(-1.39576503)**2 + (0.02958246)**2
```

```
1.9490351409105524
```



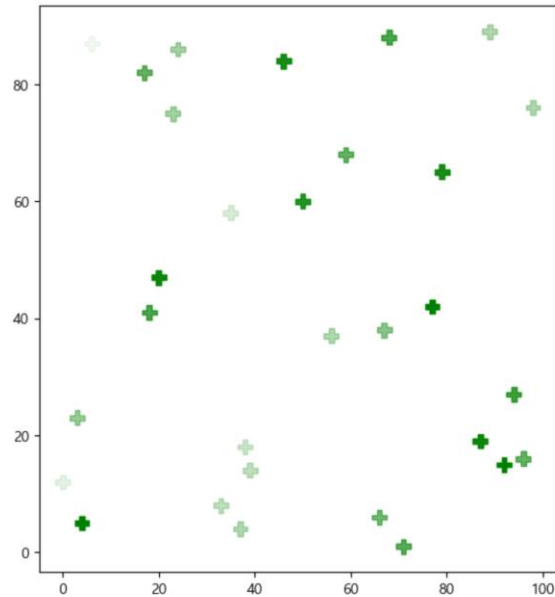
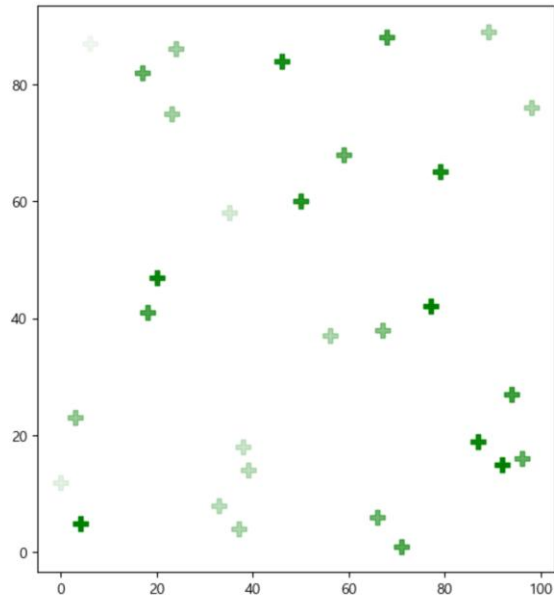
실습3

- x,y,좌표와 그 위치의 부동산 가격이 담겨져 있는 파일 xy_price.txt.를 읽어서, 각 좌표에 표식을 그리는데, 표식의 크기가 부동산 가격의 제곱에 비례하도록 하시오



실습4

- x,y,좌표와 그 위치의 부동산 가격이 담겨져 있는 파일 xy_price.txt.를 읽어서, 각 좌표에 표식을 그리는데, 표식 색깔의 농도가 부동산 가격에 비례하도록 하시오
- ax1) scatter 함수 여러 번 사용
ax2) scatter 함수 한 번 만 사용

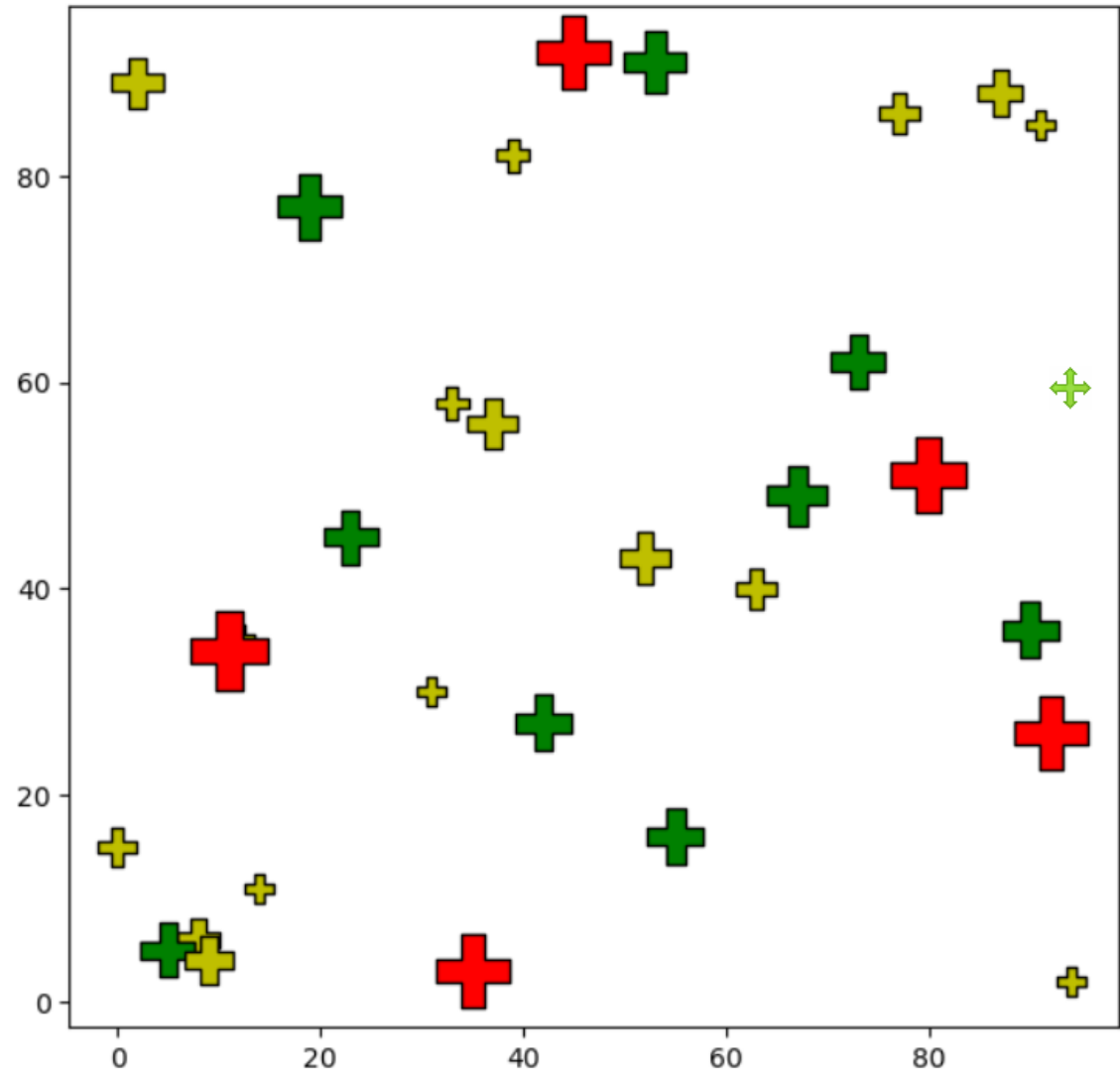


- 투명도는 alpha 파라미터로 조절하며, 리스트를 argument로 받을수 있다.

실습5

실습 3에 추가해서,
부동산 가격이
10이상 20미만이면 노랑색,
20이상 25미만이면 초록색,
25이상 30미만이면 빨간색으로 나타내시오

조건: scatter 함수 한번만 호출



style 파라미터 정리

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

Q & A

Thank you