

9. 데이터 시각화 작업



이번시간에 학습할 내용은..

분석된 결과를 도식화하여 보여주는 작업도
중요하죠? 파이썬에서 시각화 작업에 가장
많이 사용되는 matplotlib에 대해 학습합니다.



I. matplotlib

- 파이썬 프로그램에서 일반적으로 많이 사용하는 시각화 모듈

```
In [1]: from matplotlib import pyplot
```

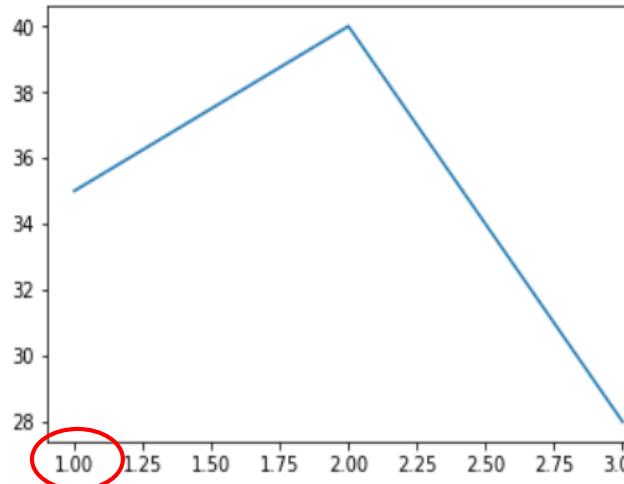
그래프 작성과 관련한 함수는 pyplot 모듈에 포함되어 있으므로 import!

matplotlib로 초간단 그래프 작성하기 -plot()

```
In [2]: pyplot.plot([1,2,3],[35,40,28])  
pyplot.show()
```

2D모형 선 그래프를 정해진 위치에 작성

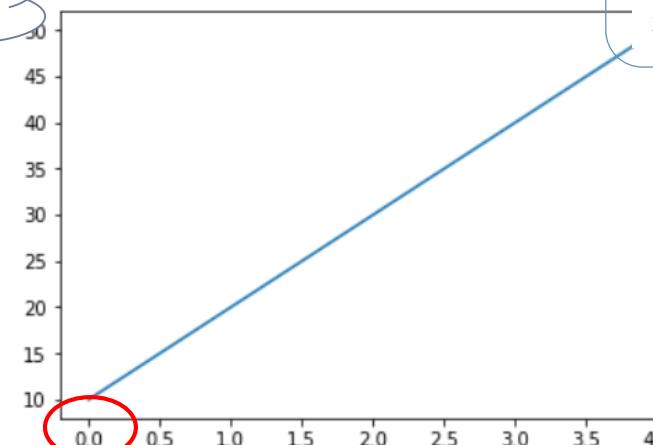
화면에 그래프 표시



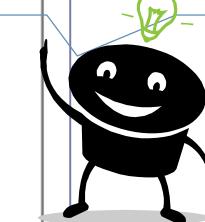
이상하다? x축의 값은 어떻게 할당되는거지?



```
In [1]: import matplotlib.pyplot as plt  
  
In [2]: plt.plot([10, 20, 30, 40, 50])  
plt.show()
```



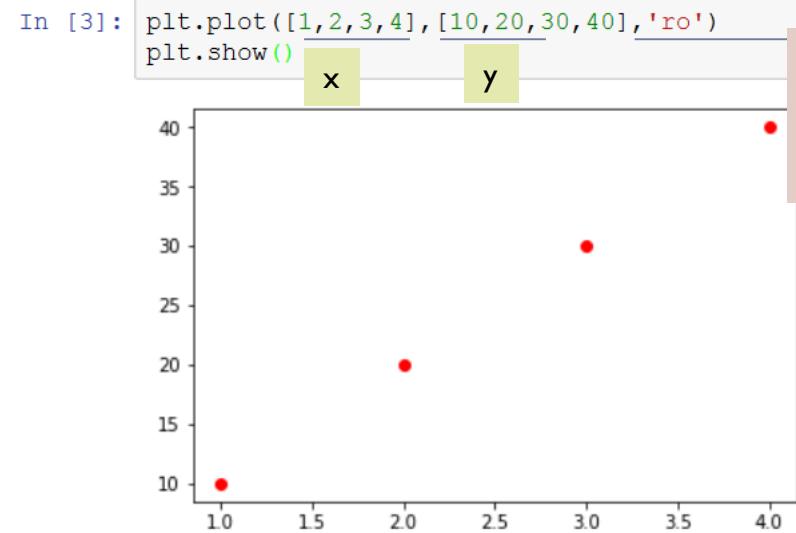
y축의 데이터 개수에 따라 x축은 자동으로 0부터 y축 개수만큼 할당돼요. 첫번째 차트의 plot함수에서 [1,2,3] 이렇게 x값을 지정하면 x축이 1부터 값이 시작되잖아요.



plot() 함수의 기본 옵션 – 색상, 선 스타일, 강조 문자 표시



plot() 함수는 꺽은선 그래프형태를 기본값으로 해요. 선과 점에 옵션값을 주면 나름 눈에 들어오는 그래프를 작성할 수 있어요.



옵션정보
linestyle=' 선스타일 지정
color=' 색상 지정
marker=' 강조를 위한 마커스타일 설정

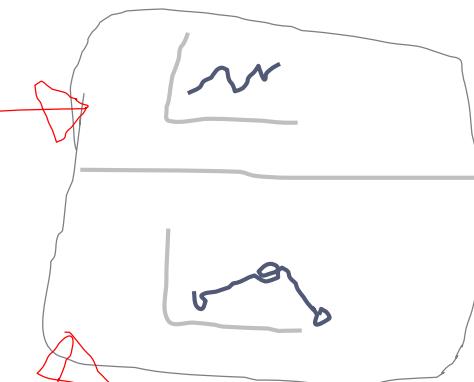
하나의 화면에 여러 개의 그래프 그려보기-figure & subplot

- 그래프는 figure 객체내에 존재함

how to step 1. figure 함수 통하여 객체 생성

```
In [3]: fig=plt.figure()
```

fig 객체는 empty 상태이므로 이를 실행하면 빈 공간만 생겨요.



how to step 2. add_subplot 메서드 사용하여 생성할 그래프 개수만큼 subplot 생성

```
In [13]: ax1=fig.add_subplot(2,1,1)  
ax2=fig.add_subplot(2,1,2)
```

화면에 2*1의 그래프 (즉, 2개) 개수를 생성하고
첫번째 그래프를 선택

how to step 3. 두 그래프에 들어갈 plot함수 작성하여 fig 객체 실행

```
In [22]: ax1.plot(np.random.randn(100))  
ax2.plot(np.random.randn(200).cumsum())  
fig.show()
```

난수값을 할당하고 fig 객체를 실행하면 차트가 나타나게 돼요.

이 부분은 생각보다 쉽게 결과가 나타나지 않아서
여러가지 고려할 사항이 많아요. ㅠㅠ
차트가 화면에 나타나지 않아서 고민이라는...
warning 메시지가 나타난다는...그래서, 또 고민...





matplotlib 모듈은 기존에 2D 차원의 시각화 도구를 제공, 최근에는 3D 차원의 시각화 도구를 제공
Jupyter에서는 인터랙티브한 기능을 제공하여 그래프의 확대보기, 회전을 가능하게 함.
따라서, Jupyter에서 이러한 설정을 먼저 실행해야 함.

```
In [17]: %matplotlib notebook
```

```
In [18]: import matplotlib.pyplot as plt  
import numpy as np
```

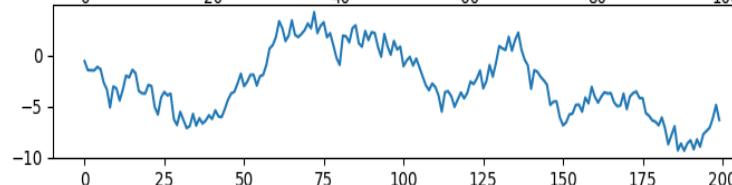
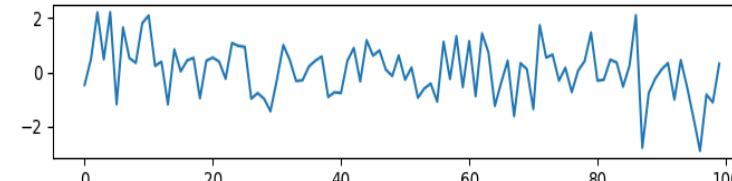
```
In [21]: fig=plt.figure()  
ax1=fig.add_subplot(2,1,1)  
ax2=fig.add_subplot(2,1,2)
```



%matplotlib inline
으로 설정하고 여러분 코딩을 해
보세요. 어떤 차이가 발생하는지..



Figure 1



pan/zoom

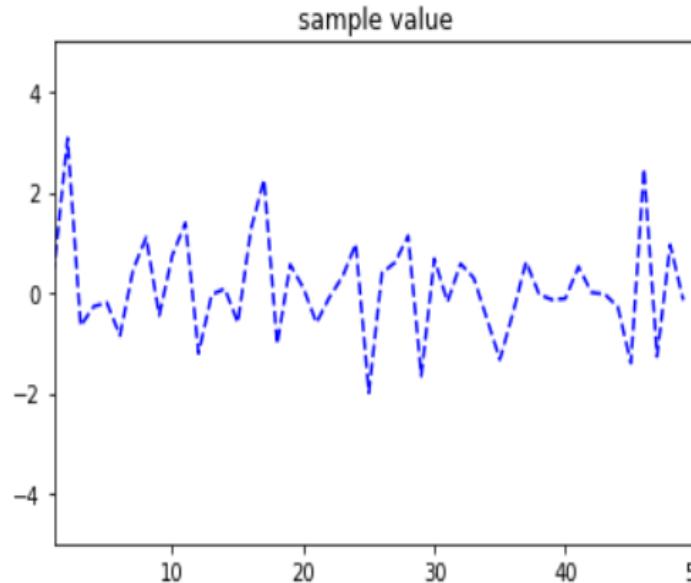
```
In [22]: ax1.plot(np.random.randn(100))  
ax2.plot(np.random.randn(200).cumsum())  
fig.show()
```



x,y축 값 범위 지정하는 set_xlim(), set_ ylim() 함수

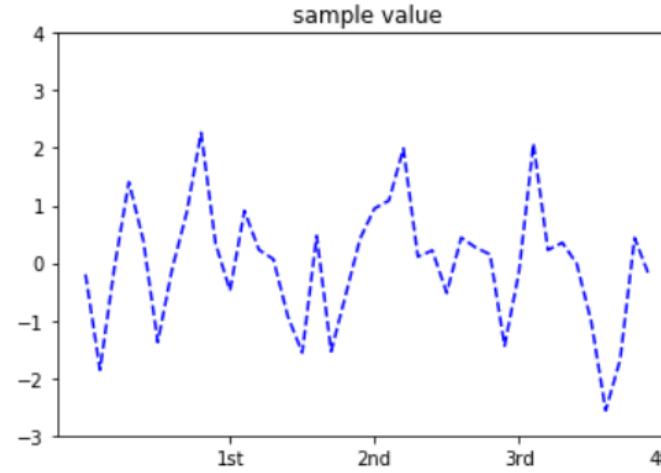
```
In [13]: fig=plt.figure()  
a1=fig.add_subplot(111)  
a1.plot(randn(50), 'b--')  
a1.set_xlim([1,50])  
a1.set_ ylim([-5,5])  
a1.set_title('sample value')
```

```
Out[13]: <matplotlib.text.Text at 0x928dd68>
```

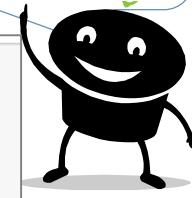


```
In [16]: fig=plt.figure()  
a1=fig.add_subplot(111)  
a1.plot(randn(40), 'b--')  
a1.set_xticks([10,20,30,40])  
a1.set_x ticklabels(['1st','2nd','3rd','4th'])  
a1.set_ ylim([-3,4])  
a1.set_title('sample value')
```

```
Out[16]: <matplotlib.text.Text at 0xa41fb38>
```



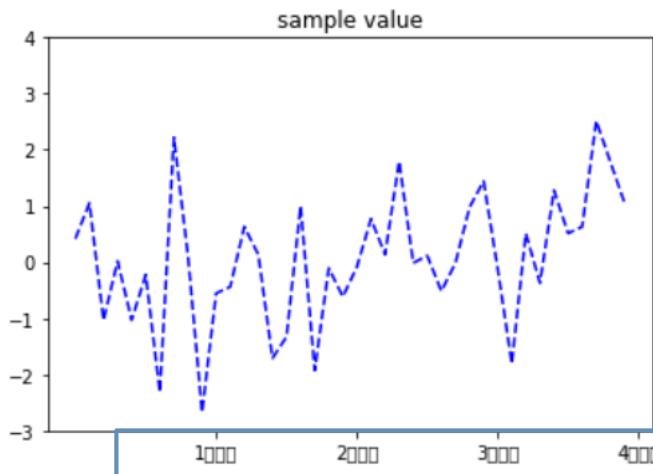
눈금의 범위를 지정하려면
set_xticks() 함수 이용
눈금이름 지정시 set_x ticklabels 이
용해요



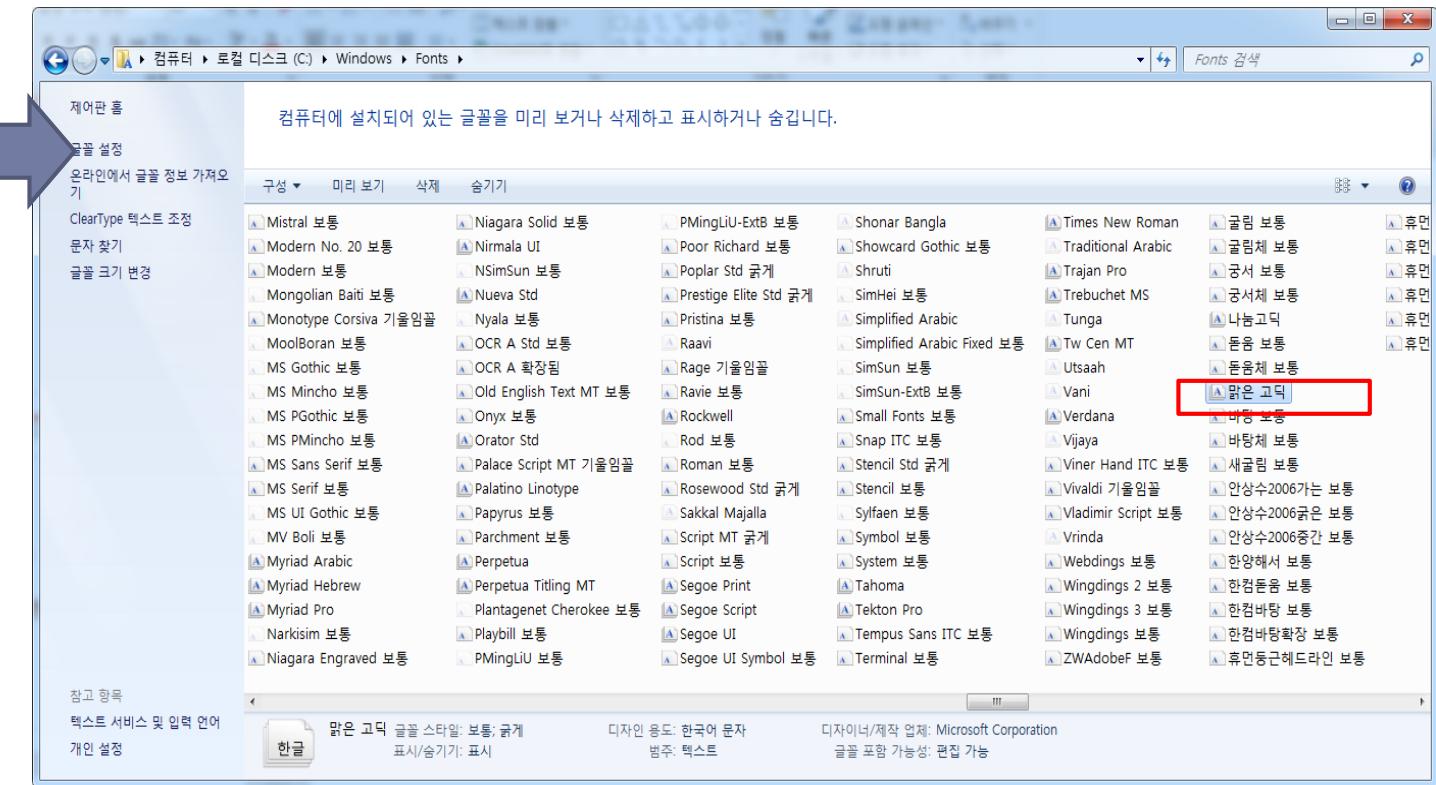
참고사항) 그래프에 한글을 표시하고자 할 경우 →한글깨짐문제발생

```
In [17]: fig=plt.figure()
a1=fig.add_subplot(111)
a1.plot(randn(40), 'b--')
a1.set_xticks([10, 20, 30, 40])
a1.set_xticklabels(['1사분기', '2사분기', '3사분기', '4사분기'])
a1.set_xlim([-3, 4])
a1.set_title('sample value')
```

```
Out[17]: <matplotlib.text.Text at 0xa4a0eb8>
```

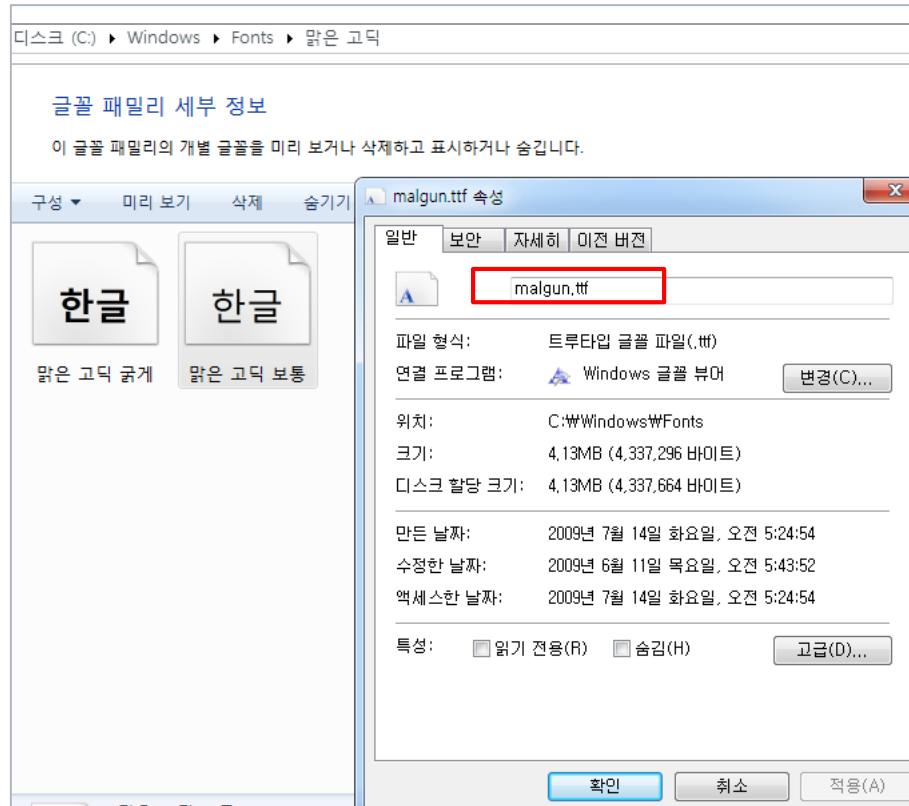


how to step 1. 윈도우 탐색기를 실행하여 그림처럼 windows 폴더에서 글자체가 설치된 font 폴더를 열어준다음 '맑은고딕' 을 더블클릭해요.

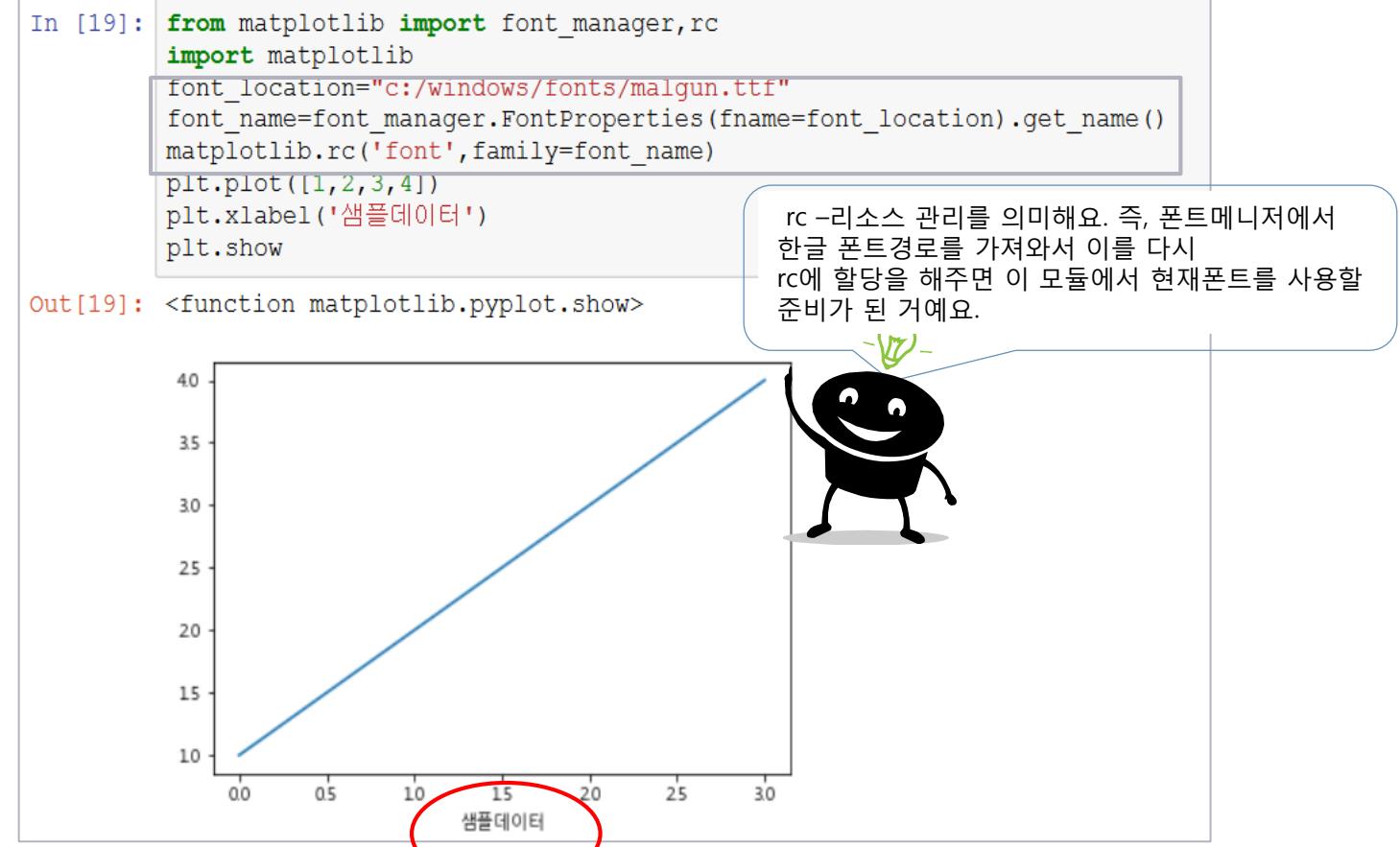




how to step 2. 글꼴에서 마우스 오른쪽 단추 눌러서 –”속성” 클릭하면 이렇게 해당 글꼴의 파일명이 표시돼요. 파일경로를 기억해요.



how to step 3. 그래프를 작성하면서 한글이 표시되려면 위에서 지정한 파일경로와 해당 파일명을 matplotlib의 font_manager 모듈에 설정해줘야 해요. 그림처럼 코딩해보아요.



3) 그래프 저장하기

- figure 객체의 savefig 메소드 이용하여 파일로 저장
- 그래프 파일의 종류는 파일 확장자를 지정하면서 결정됨.

how to step 1. savefig 메소드 이용하여 파일명 필요하다면 해상도, 배경색상, 그래프 이외 공백을 제거하는 옵션을 설정할 수 있어요.

```
In [20]: plt.savefig('s_graph.png', dpi=350) png 형식, 해상도 350으로 저장  
<matplotlib.figure.Figure at 0x939bf98>
```



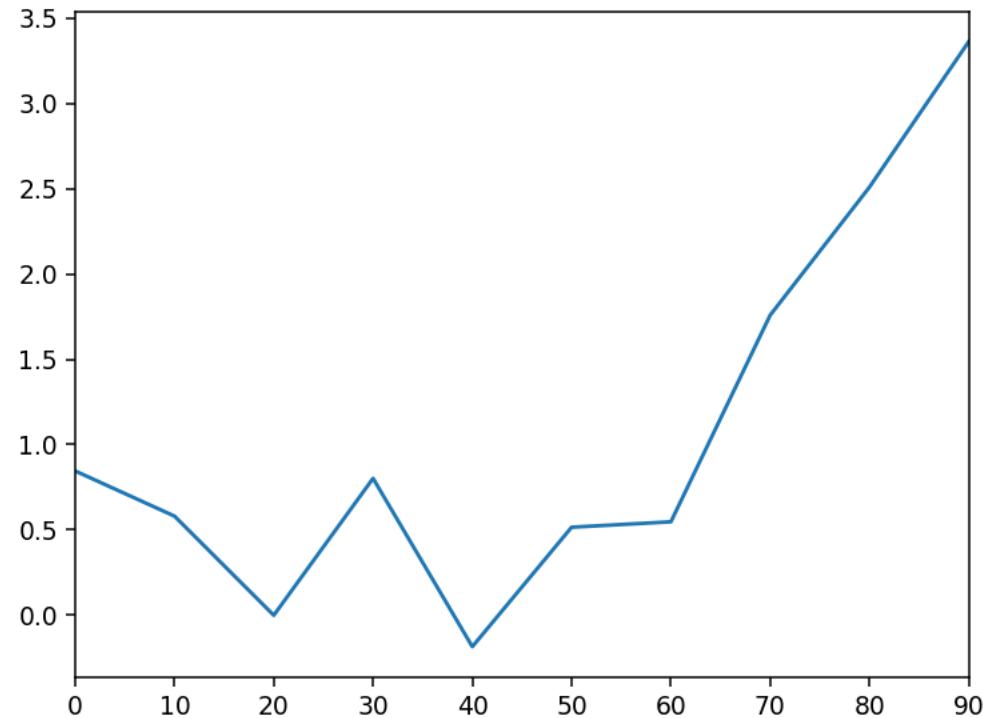
◆ pandas + matplotlib

- pandas에서 처리된 행, 열, 그룹 정보 등을 이용하여 시각화 도표 작성이 가능

I. Series 객체를 통해 그래프 작성 → Series.plot

```
In [12]: import numpy as np
a=pd.Series(np.random.randn(10).cumsum(), index=np.arange(0,100,10))
a.plot()
```

Figure 1



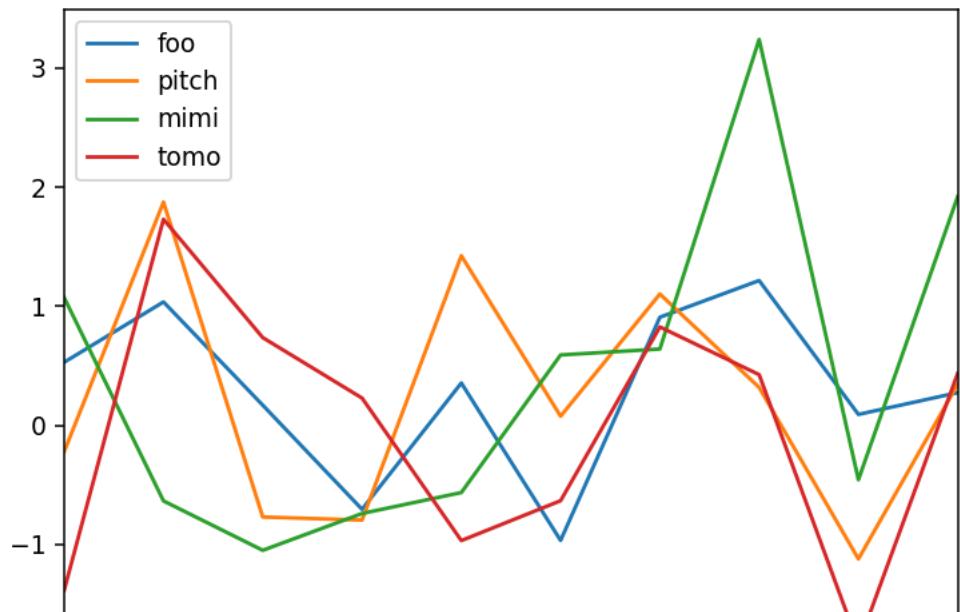
* Series.plot 메소드 인자

인자	설명
label	범례이름
style	Matplotlib의 color, linestyle에 전달할 값
alpha	그래프 투명도 설정
kind	그래프 종류(line-선, bar-막대, barh-수평막대)
Xticks/yticks	X,y 축으로 사용할 값
Xlim/ylim	X,y 축한계값
grid	그리드 표시여부

2. 데이터 프레임 객체에서 그래프 작성- DataFrame.plot

```
In [15]: df=pd.DataFrame(np.random.randn(10,4),columns=['foo','pitch','mimi','tomo'],  
index=np.arange(0,100,10))  
df.plot()
```

Figure 2



* DataFrame.plot 메소드 인자

인자	설명
Subplots	데이터프레임 컬럼을 독립된 서브플롯에 그림
figsize	생성될 그래프의 크기를 튜플로 지정
title	그래프 제목을 문자열로 지정
legend	서브플롯의 범례를 추가
Sort_columns	컬럼을 알파벳 순서로 작성