

Chapter. 06

데이터 이쁘게 보기: 데이터 시각화

| matplotlib 기초

FAST CAMPUS
ONLINE
데이터 탐색과 전처리 I

강사. 안길승

I matplotlib이란

- 파이썬에서 차트나 플롯으로 데이터를 **시각화**하기 위한 모듈로 사용이 간편하고 다양한 그래프를 제공해서 많이 사용됨
- matplotlib을 이용해서 그릴 수 있는 대표적인 그래프는 다음과 같음
 - 꺾은선 그래프 (line plot)
 - 산점도 (scatter plot)
 - 막대 그래프 (bar chart)
 - 히스토그램 (histogram)
 - 박스 플롯 (box plot)
- Pandas의 객체에서도 메서드로 matplotlib의 함수를 사용할 수 있음

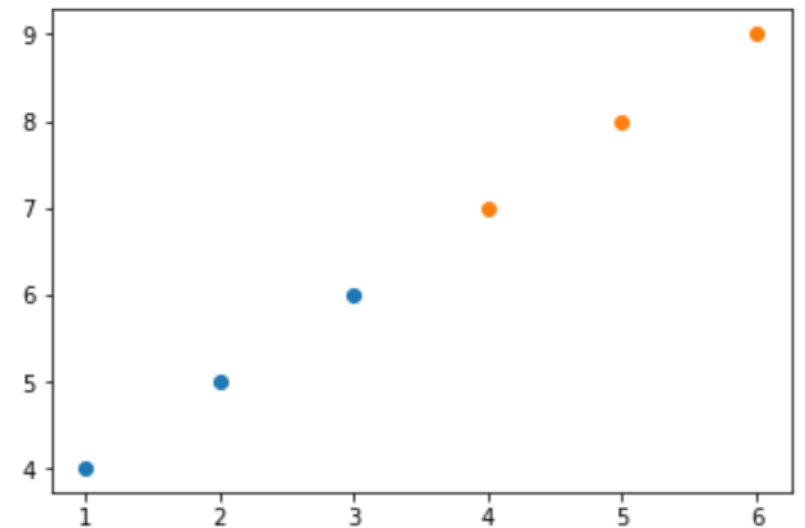
I 기초 환경 설정

- pyplot.rcParams를 이용한 폰트 설정 및 그래프 크기 설정
 - rcParams["font.family"]: 폰트 설정 (특히, 한글을 쓰는 경우에 반드시 설정 필요)
 - rcParams["font.size"]: 글씨 크기 설정
 - rcParams["figure.figsize"]: (x, y) 꼴의 튜플로 그래프의 크기를 설정

I 주피터에서 그래프 그리기

- 주피터의 매직 키워드인 `%matplotlib inline`을 이용하면 셀에 그래프를 직접 출력할 수 있음
- 주피터를 사용하지 않는 경우에는 `plt.figure()`를 사용해서 그래프 객체를 일일이 생성하고 저장해야 하지만, 주피터에서는 한 셀에서 그래프를 그릴 때마다 하나의 객체가 생성됨
 - 한 셀에서 `plt`의 함수를 사용하여 그래프를 그리면 하나의 객체에 누적해서 그래프가 생성됨
 - 그래프를 다 그린 후에는 `plt.savefig(file, dpi)`를 사용하여 그래프를 저장할 수 있음

```
1 plt.scatter(x, y) # 그래프 1
2 plt.scatter(x+3, y+3) # 그래프 2
3 plt.savefig('fig1.png', dpi = 300)
```



I 자주 사용되는 색상 설정

- 다양한 그래프에서 색상을 설정해야 하며, 자주 사용되는 색상 및 입력은 다음과 같음

색상	입력	약어
파란색	blue	b
초록색	green	g
빨간색	red	r
노란색	yellow	y
검은색	black	K

I 자주 사용되는 마커 및 선 스타일 설정

- 마커란 하나의 데이터 포인트 (샘플)을 나타내는 기호를 나타냄
- 선 스타일은 꺾은선 그래프의 선 스타일을 의미함

마커	입력
점	.
픽셀	,
원	O
별	*
더하기	+
x	X
다이아몬드	D

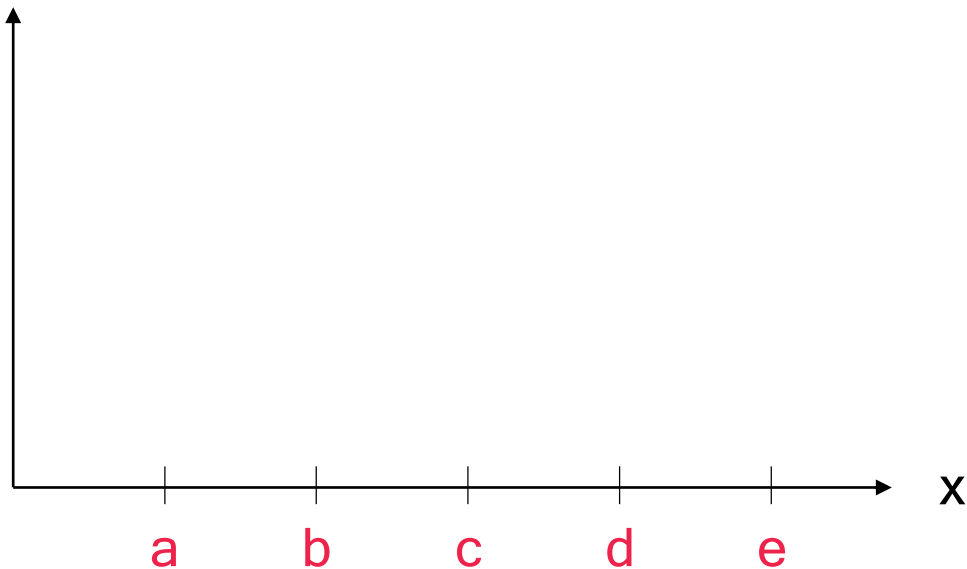
선 스타일	입력
-	실선
--	파선
-. .-	파선 - 점선
:	점선

I label 및 title 설정

- label이란 축의 이름을 의미하며, plt.xlabel과 plt.ylabel을 사용하여 정의할 수 있음
 - plt.xlabel(label, **kwargs): label (str)로 xlabel을 설정
 - plt.ylabel(label, **kwargs): label (str)로 ylabel을 설정
 - **kwargs: 텍스트의 특성을 설정하는 키워드 (예: fontproperties, fontsize, fontstyle 등)
- title이란 그래프의 이름으로, plt.title을 사용하여 정의할 수 있음
 - plt.title(label, loc): label (str)로 title을 설정 (loc: 위치 {'center', 'left', 'right'})

I tick 설정

- tick이란 그래프에서 나타내는 눈금을 의미하며, plt.xticks와 plt.yticks을 이용하여 설정할 수 있음
 - plt.xticks(ticks, labels): ticks (위치)에 labels를 입력 (x축)
 - plt.yticks(ticks, labels): ticks (위치)에 labels를 입력 (y축)
- 사용 예시: plt.xticks([1, 2, 3, 4, 5], ['a', 'b', 'c', 'd', 'e'])



legend 설정

- legend (범례)를 표시하려면 plt.legend(loc)을 사용하여 범례를 표시할 수 있음
 - loc: 범례 위치 {'upper right' (1), 'upper left' (2), 'lower left' (3), 'lower right' (4), 'right' (5), 'center' (10)}
- 단, 범례를 표시하려면 그래프를 그릴 때 label 키워드를 사용하여 범례를 정의해야 함

I 축 범위 설정

- `plt.xlim(left, right)`: x축의 범위를 (left, right)로 설정
- `plt.ylim(bottom, top)`: y축의 범위를 (bottom, top)로 설정

Chapter. 06

데이터 이쁘게 보기: 데이터 시각화

| 꺾은선 그래프와 산점도 그리기

FAST CAMPUS
ONLINE
데이터 탐색과 전처리 I

강사. 안길승

I matplotlib을 이용한 꺾은선 그래프 그리기

- pyplot.plot 함수를 사용하면 꺾은선 그래프를 그릴 수 있음
- 주요 입력
 - x, y: x, y축에 들어갈 값 (iterable한 객체여야 하며, x[i], y[i]가 하나의 점이 되므로 길이가 같아야 함)
 - linewidth: 선 두께
 - marker: 마커 종류
 - markersize: 마커 크기
 - color: 선 색상
 - linestyle: 선 스타일
 - label: 범례

I Pandas 객체의 method를 이용한 꺾은선 그래프 그리기

- DataFrame.plot() 함수를 사용하면 DataFrame을 사용한 그래프를 그릴 수 있음
- 주요 입력
 - kind: 그래프 종류 ('line'이면 선 그래프를 그림)
 - x: x축에 들어갈 컬럼명 (입력하지 않으면, index가 x축에 들어감)
 - y: y축에 들어갈 컬럼명 (목록)
 - xticks, yticks 등도 설정이 가능함 (단, pyplot을 사용해서도 설정이 가능)

I matplotlib을 이용한 산점도 그리기

- pyplot.scatter 함수를 사용하면 산점도를 그릴 수 있음
- 주요 입력
 - x, y: x, y축에 들어갈 값 (iterable한 객체여야 하며, x[i], y[i]가 하나의 점이 되므로 두 입력의 길이가 같아야 함)
 - marker: 마커 종류
 - markersize: 마커 크기
 - color: 마커 색상
 - label: 범례

I Pandas 객체의 method를 이용한 산점도 그리기

- DataFrame.plot()함수를 사용하면 DataFrame을 사용한 그래프를 손쉽게 그릴 수 있으며, 산점도 역시 그릴 수 있음
- 주요 입력
 - kind: 그래프 종류 ('scatter'면 산점도를 그림)
 - x: x축에 들어갈 컬럼명 (입력하지 않으면, index가 x축에 들어감)
 - y: y축에 들어갈 컬럼명
 - xticks, yticks 등도 설정이 가능함 (단, pyplot을 사용해서도 설정이 가능)

Chapter. 06

데이터 이쁘게 보기: 데이터 시각화

| 막대 차트 그리기

FAST CAMPUS
ONLINE
데이터 탐색과 전처리 I

강사. 안길승

I matplotlib을 이용한 막대 차트 그리기

- pyplot.bar 함수를 사용하면 막대 차트를 그릴 수 있음
- 주요 입력
 - x: 막대의 위치
 - height: 막대의 높이
 - width: 막대의 너비
 - align: 막대 정렬

I Pandas 객체의 method를 이용한 막대 차트 그리기

- DataFrame.plot()함수를 사용하면 DataFrame을 사용한 그래프를 손쉽게 그릴 수 있으며, 막대 차트 역시 그릴 수 있음
- 주요 입력
 - kind: 그래프 종류 ('bar'면 막대 그래프를 그림)
 - x: x축에 들어갈 컬럼명 (입력하지 않으면, index가 x축에 들어감)
 - y: y축에 들어갈 컬럼명 (목록)
 - xticks, yticks 등도 설정이 가능함 (단, pyplot을 사용해서도 설정이 가능)

I 다중 막대그래프 그리기

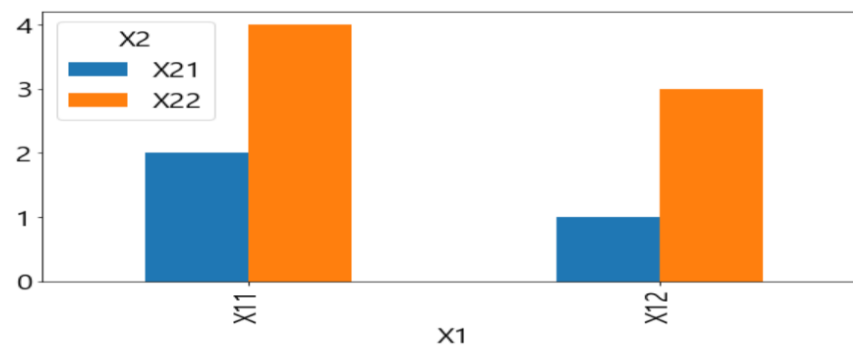
- 다중 막대그래프는 pandas의 `groupby` 혹은 `set_index`, 그리고 `unstack`, 혹은 `pivot_table`을 사용하면 쉽게 그릴 수 있음

X1	X2	Value
X11	X21	2
	X22	4
X12	X21	1
	X22	3

`unstack()`

	X21	X22
X11	2	4
X12	1	3

`plot(kind = 'bar')`



막대 그래프를 그리기 적합한 데이터
(`groupby` 혹은 `set_index` 사용)

Chapter. 06

데이터 이쁘게 보기: 데이터 시각화

| 파이 차트 그리기

FAST CAMPUS
ONLINE
데이터 탐색과 전처리 I

강사. 안길승

I matplotlib을 이용한 파이 차트 그리기

- pyplot.pie 함수를 사용하면 파이 차트를 그릴 수 있음
- 주요 입력
 - x: 각 pie의 크기
 - labels: 각 pie에 부착되는 라벨
 - labeldistance: 라벨 간 거리
 - normalize: 비율을 나타낼 것인지 여부
 - autopct: 위에 표시될 글자 형태 (예: “%1.1f%%”, “%1d%%”)
 - colors: 배열로 설정해서 각 파트의 색상을 설정 가능

I Pandas 객체의 method를 이용한 파이 차트 그리기

- DataFrame.plot()함수를 사용하면 DataFrame을 사용한 그래프를 손쉽게 그릴 수 있으며, 파이 차트 역시 그릴 수 있음
- 주요 입력
 - kind: 그래프 종류 ('pie'이면 파이 차트를 그림)
 - x: 각 pie의 크기
 - y: 각 pie에 부착되는 라벨

Chapter. 06

데이터 이쁘게 보기: 데이터 시각화

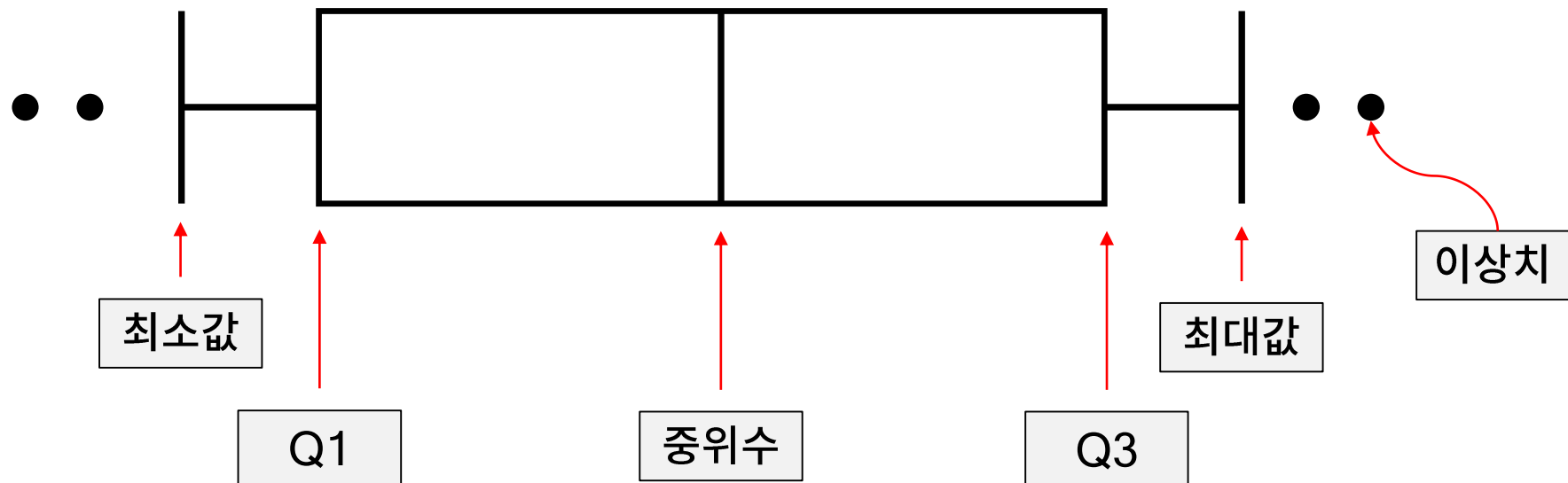
| 박스 플롯 그리기

FAST CAMPUS
ONLINE
데이터 탐색과 전처리 I

강사. 안길승

I 박스 플롯이란

- 박스 플롯이란 하나의 변수에 대한 분포를 한 눈에 보여주는 그래프임



I matplotlib을 이용한 박스 플롯 그리기

- `pyplot.boxplot` 함수를 사용하면 박스 플롯을 그릴 수 있음
- 주요 입력
 - `x`: `boxplot`을 그리기 위한 데이터 (2차원인 경우, 여러 개의 박스 플롯이 그려짐. 즉, 하나의 열이 하나의 박스 플롯으로 생성됨)

I Pandas 객체의 method를 이용한 박스 플롯 그리기

- DataFrame.boxplot()함수를 사용하면 DataFrame을 사용한 그래프를 손쉽게 그릴 수 있으며, 박스 플롯 역시 그릴 수 있음
- 주요 입력
 - column: box plot을 그릴 컬럼 목록