

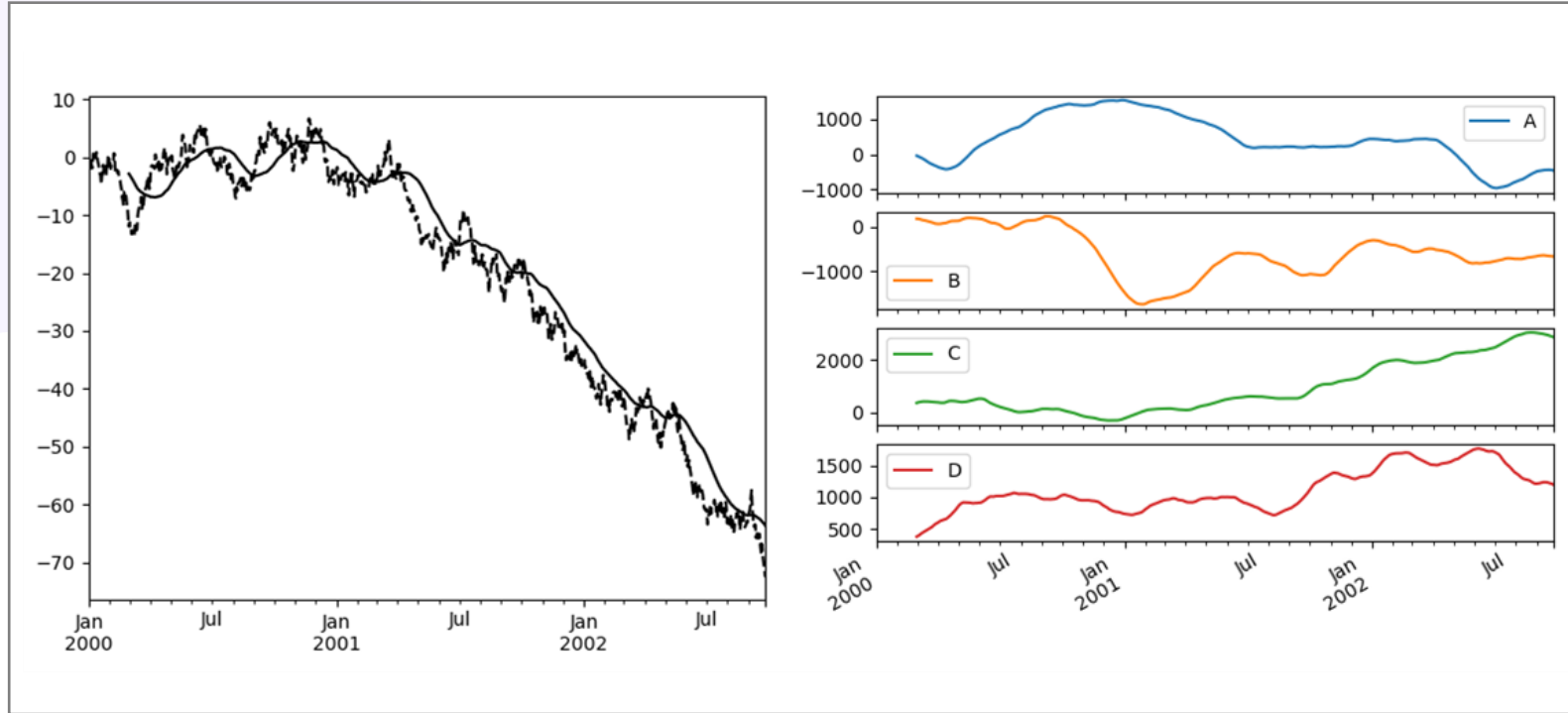


# 시계열 데이터 분석



# 판다스 Expanding and Rolling 이해하기

- 판다스 User Guide Computational Tools



<출처 : [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/computation.html](https://pandas.pydata.org/pandas-docs/stable/user_guide/computation.html)>

# 판다스 Expanding and Rolling 이해하기

## Windows

### **df.expanding()**

Return an Expanding object allowing summary functions to be applied cumulatively.

### **df.rolling(n)**

Return a Rolling object allowing summary functions to be applied to windows of length n.

01

## Expanding

- 요약 기능을 허용하는 확장 개체를 누적 적용하여 반환
  - 확장된 변형 제공
- 누적된 변경된 값의 정보 제공

# 판다스 Expanding and Rolling 이해하기

## Windows

### **df.expanding()**

Return an Expanding object allowing summary functions to be applied cumulatively.

### **df.rolling(n)**

Return a Rolling object allowing summary functions to be applied to windows of length n.

## 02

### Rolling

- 요약 기능을 허용하는 Rolling 객체를 길이 n개의 창에 적용하여 반환
- 이동 평균(Moving Average)을 구할 때 사용



# 판다스 Expanding and Rolling 이해하기

---

이론 영상 후 실습 영상 제시





# 데이터 시각화\_

## 기초 그래프

## 그리기



# 데이터 시각화 소개

The screenshot shows the pandas documentation website. The top navigation bar includes links for 'Getting started', 'User Guide', 'API reference', 'Development', and 'Release notes'. The left sidebar contains a search bar and a list of topics, with 'Visualization' highlighted. The main content area is titled 'Visualization' and explains the standard convention for referencing the matplotlib API. It includes two code blocks: one for importing matplotlib and closing all plots, and another for creating a time series and plotting it. A 'Note' box mentions that all calls to `np.random` are seeded with 123456. The right sidebar lists 'On this page' with links to various plotting topics.

pandas

Getting started User Guide API reference Development Release notes

Search the docs ...

10 minutes to pandas  
Intro to data structures  
Essential basic functionality  
IO tools (text, CSV, HDF5, ...)  
Indexing and selecting data  
MultiIndex / advanced indexing  
Merge, join, concatenate and compare  
Reshaping and pivot tables  
Working with text data  
Working with missing data  
Categorical data  
Nullable integer data type  
Nullable Boolean data type  
**Visualization**  
Computational tools  
Group by: split-apply-combine  
Time series / date functionality  
Time deltas  
Styling  
Options and settings  
Enhancing performance

## Visualization

We use the standard convention for referencing the matplotlib API:

```
In [1]: import matplotlib.pyplot as plt
In [2]: plt.close('all')
```

We provide the basics in pandas to easily create decent looking plots. See the [ecosystem](#) section for visualization libraries that go beyond the basics documented here.

**Note**

All calls to `np.random` are seeded with 123456.

### Basic plotting: `plot`

We will demonstrate the basics, see the [cookbook](#) for some advanced strategies.

The `plot` method on Series and DataFrame is just a simple wrapper around `plt.plot()`:

```
In [3]: ts = pd.Series(np.random.randn(1000),
...:                  index=pd.date_range('1/1/2000', periods=1000))
...:
In [4]: ts = ts.cumsum()
In [5]: ts.plot()
Out[5]: <matplotlib.axes._subplots.AxesSubplot at 0x7fe278bb5160>
```

On this page

- Basic plotting: `plot`
- Other plots
- Plotting with missing data
- Plotting tools
- Plot formatting
- Plotting directly with matplotlib
- Plotting backends

<출처 : [https://pandas.pydata.org/pandas-docs/stable/user\\_guide/visualization.html](https://pandas.pydata.org/pandas-docs/stable/user_guide/visualization.html)>

# 데이터 시각화 소개

- Plot

- kind : str

- 'line' : line plot (default)
    - 'bar' : vertical bar plot
    - 'barh' : horizontal bar plot
    - 'hist' : histogram
    - 'box' : boxplot
    - 'kde' : Kernel Density Estimation plot
    - 'density' : same as 'kde'
    - 'area' : area plot
    - 'pie' : pie plot
    - 'scatter' : scatter plot
    - 'hexbin' : hexbin plot





# 데이터 시각화 소개

---

이론 영상 후 실습 영상 제시



# 막대 그래프(Bar Plot)

- 막대 그래프(Bar Plot) 그리기



막대 그래프

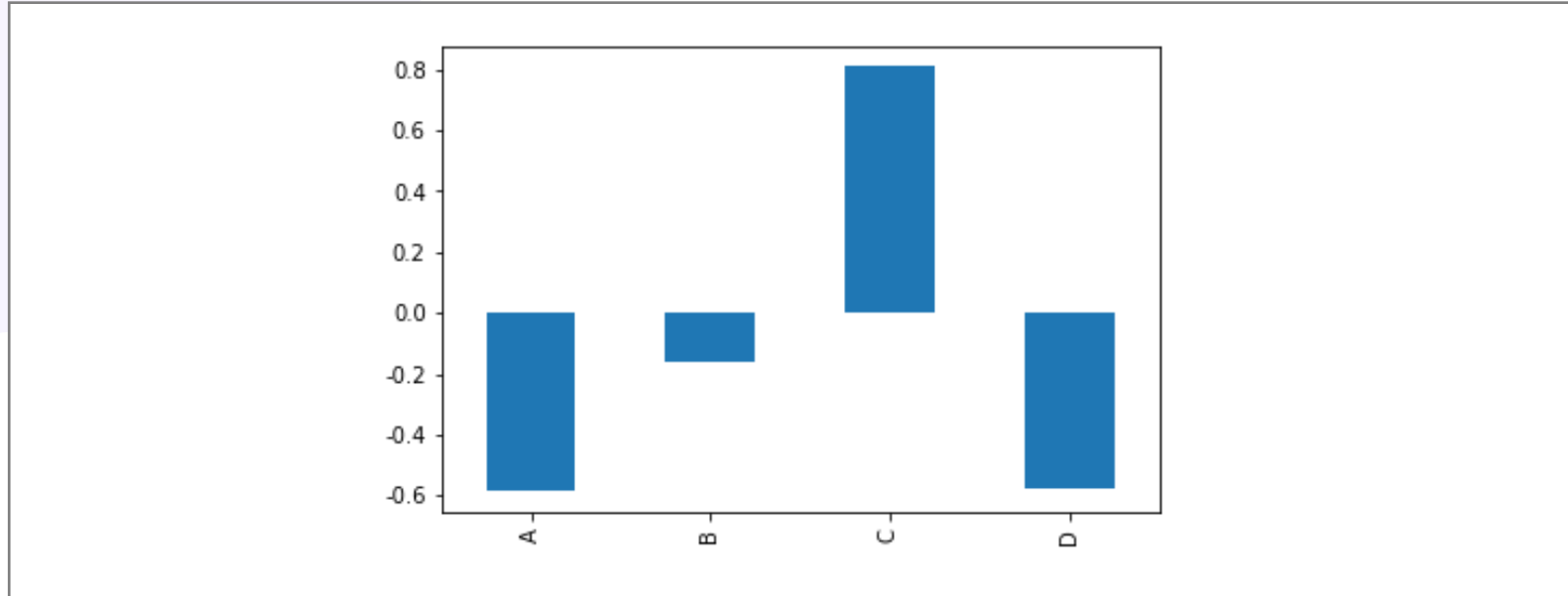
바 차트

바 그래프

- 표현 값에 비례하여 높이와 길이를 지닌 직사각형 막대로 범주형 데이터를 표현하는 차트나 그래프를 말함
- 막대는 수직으로나 수평으로 그릴 수 있으며, 수직 막대 그래프는 선 그래프 또는 라인 그래프(Line Graph)라고도 함

# 막대 그래프(Bar Plot)

- 막대 그래프(Bar Plot) 그리기

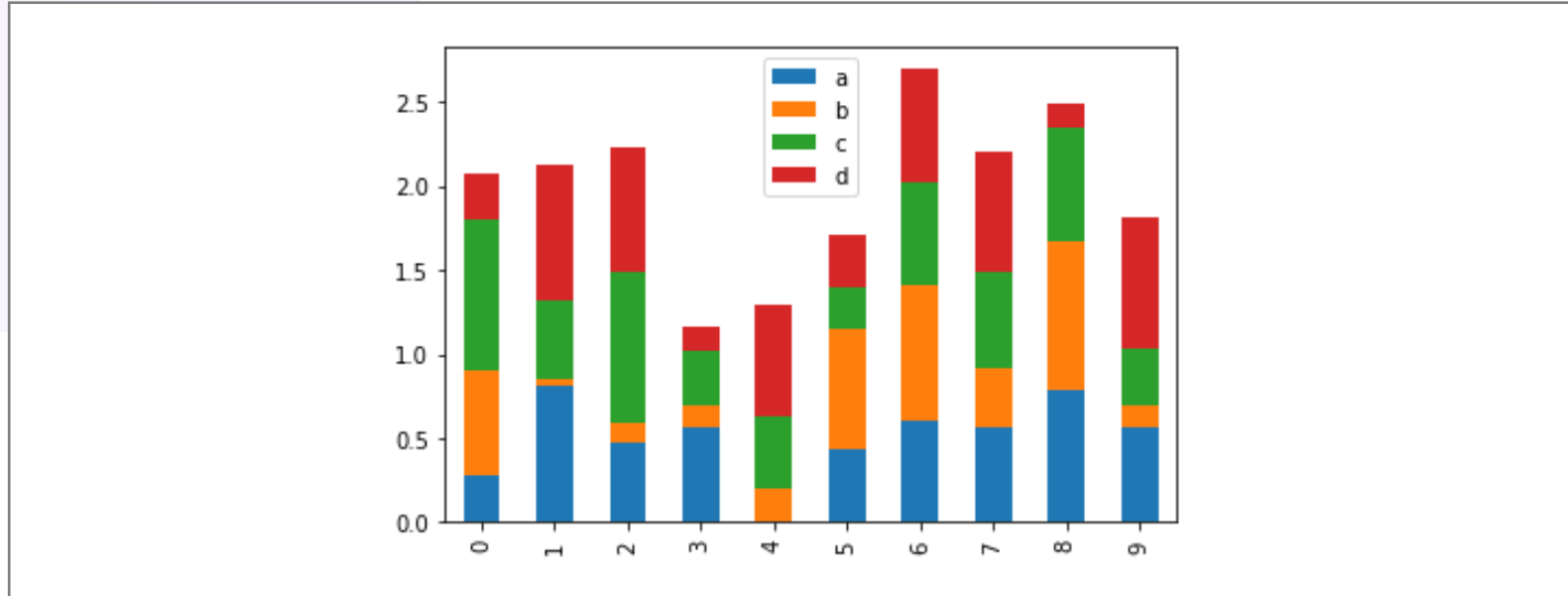


01 `df.iloc[5].plot(kind='bar'), df.iloc[5].plot.bar()`

- `iloc` : Index 번호
- 막대 그래프 그리기 위해서는 `plot(kind='bar')` 또는 `plot.bar()` 모듈 사용

# 막대 그래프(Bar Plot)

- 막대 그래프(Bar Plot) 그리기

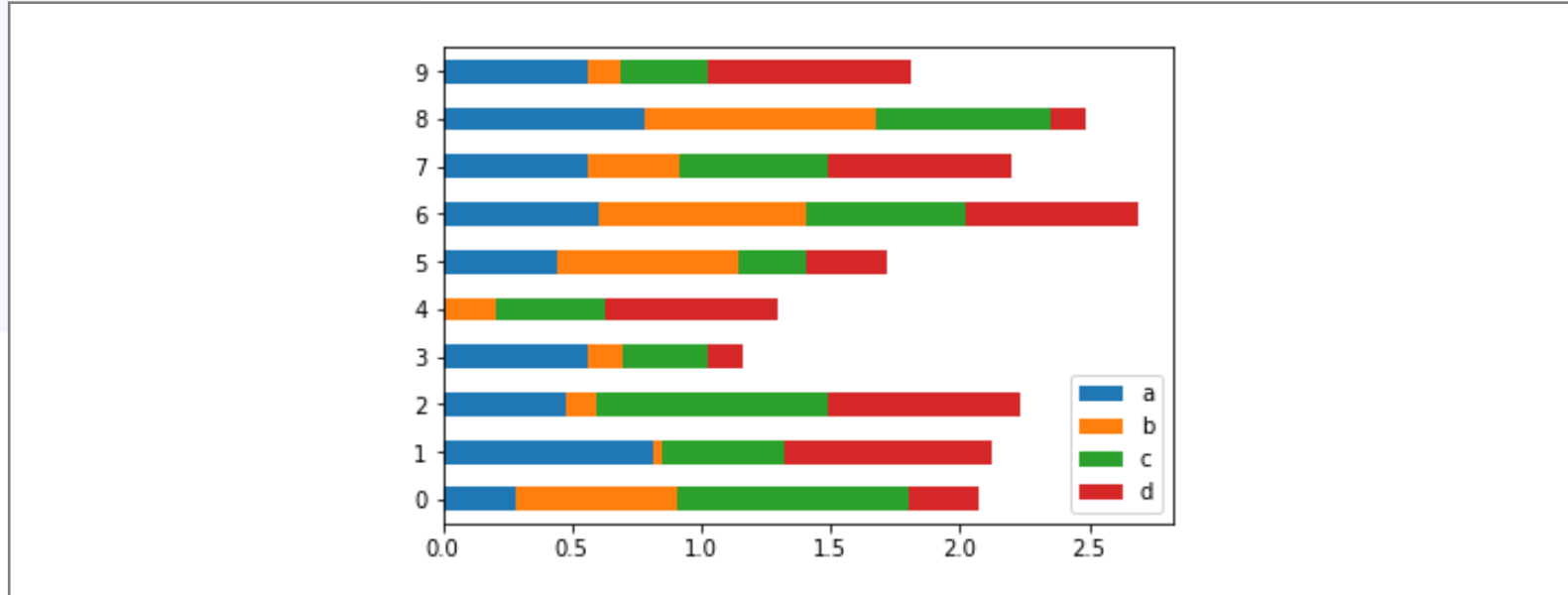


02 `df2.plot.bar(stacked=True)`

- stacked=True 옵션 : 누적한 값을 시각화

# 막대 그래프(Bar Plot)

- 막대 그래프(Bar Plot) 그리기



03

`df2.plot.barh(stacked=True)`

- barh : Bar horizontally 가로막대 그래프 그리기



# 막대 그래프(Bar Plot)

---

이론 영상 후 실습 영상 제시

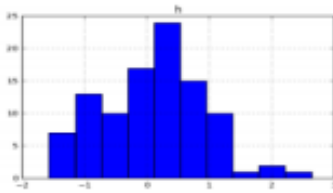


# 히스토그램과 도수분포표

## Plotting

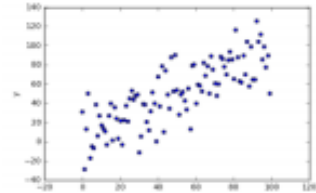
`df.plot.hist()`

Histogram for each column



`df.plot.scatter(x='w',y='h')`

Scatter chart using pairs of points



### 도수분포표

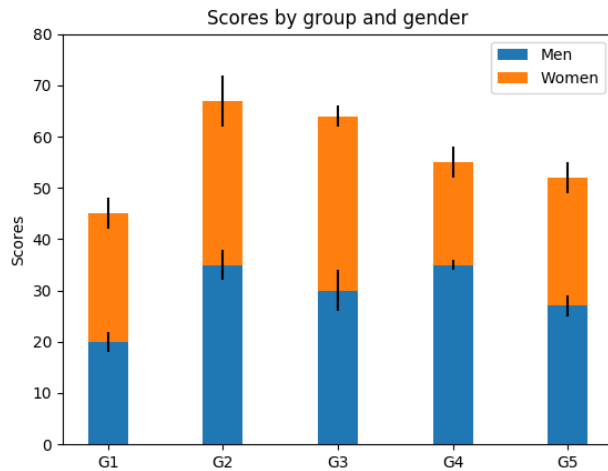
특정 구간에 속하는  
자료의 개수를 표현한 것임

### 히스토그램

도수분포표를 시각화하여  
막대 그래프로 표현 하지만  
Bar Plot과는 다름

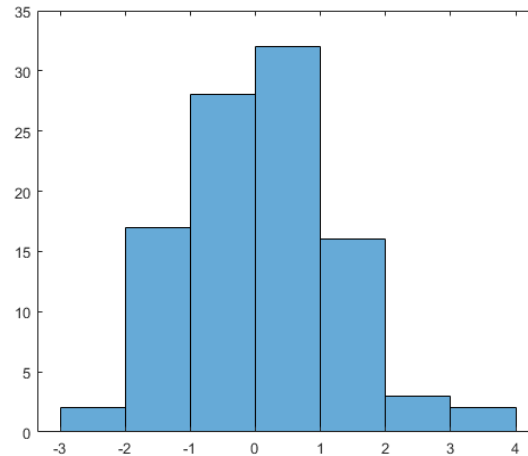
# 히스토그램과 도수분포표

## 막대 그래프(Bar Plot)



표현 값에 비례하여 높이와 길이를 지닌 직사각형 막대로 범주형 데이터를 표현하는 차트나 그래프, 합계, 평균 등의 수치를 시각화한 것임

## 히스토그램(Hist Plot)

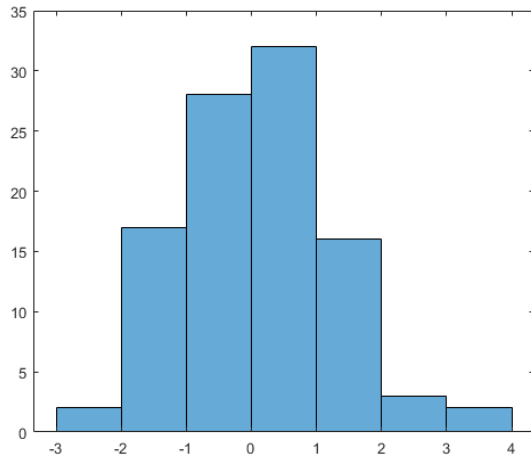


구간별 빈도수를 표현한 것임



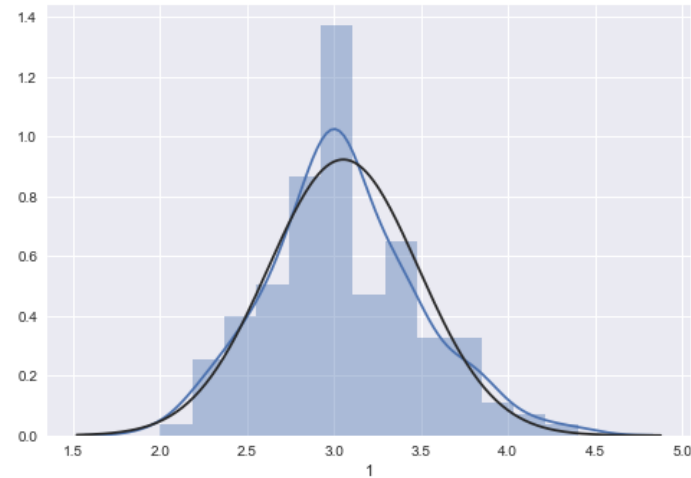
# 히스토그램과 도수분포표

히스토그램(Hist Plot)



구간별 빈도 수를 표현한 것임

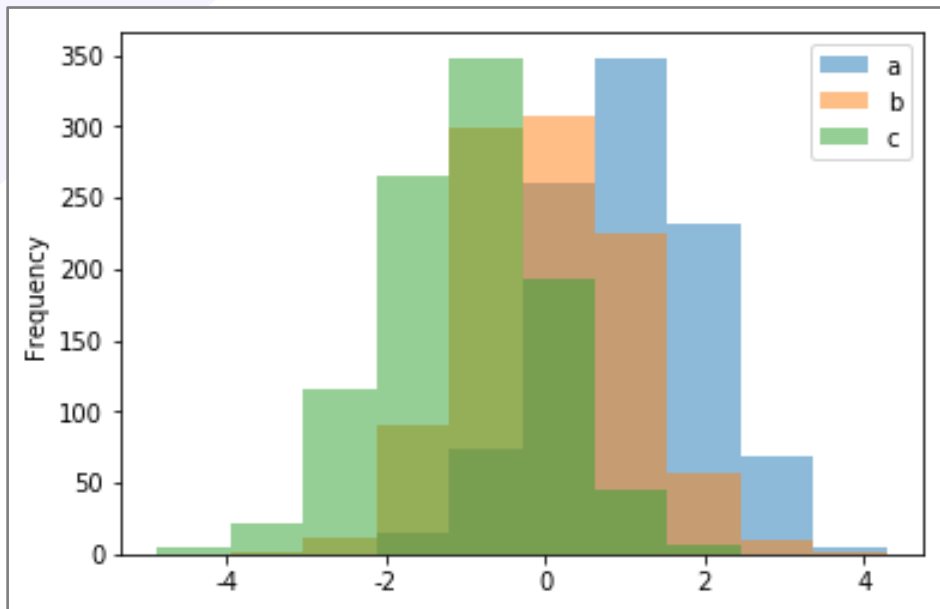
정규분포(Density Plot)



확률 밀도 함수로,  
확률 변수의 밀도를 표현한  
것임

# 히스토그램과 도수분포표

01 `df4.plot.hist(alpha=0.5)`

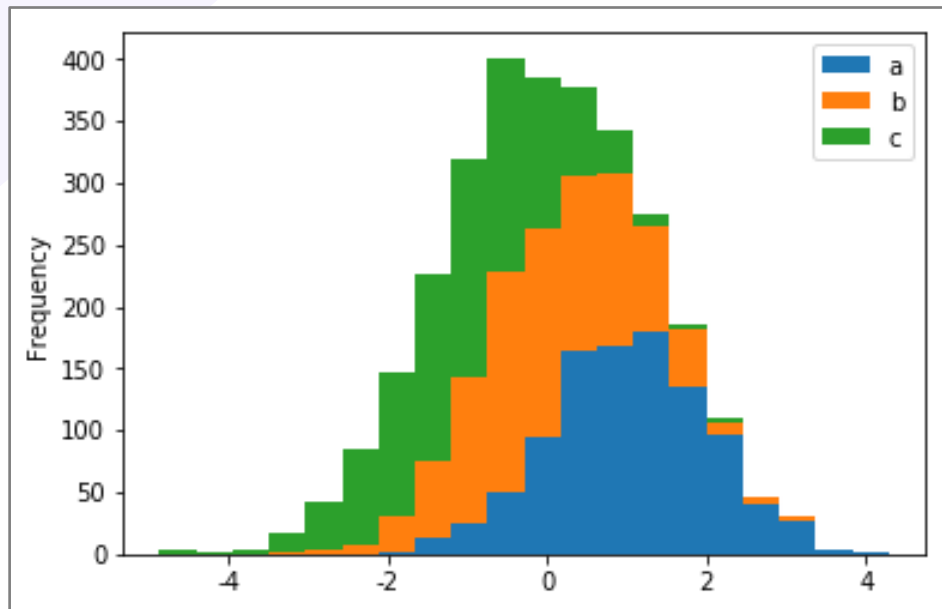


■ alpha 값 : 투명도

# 히스토그램과 도수분포표

02

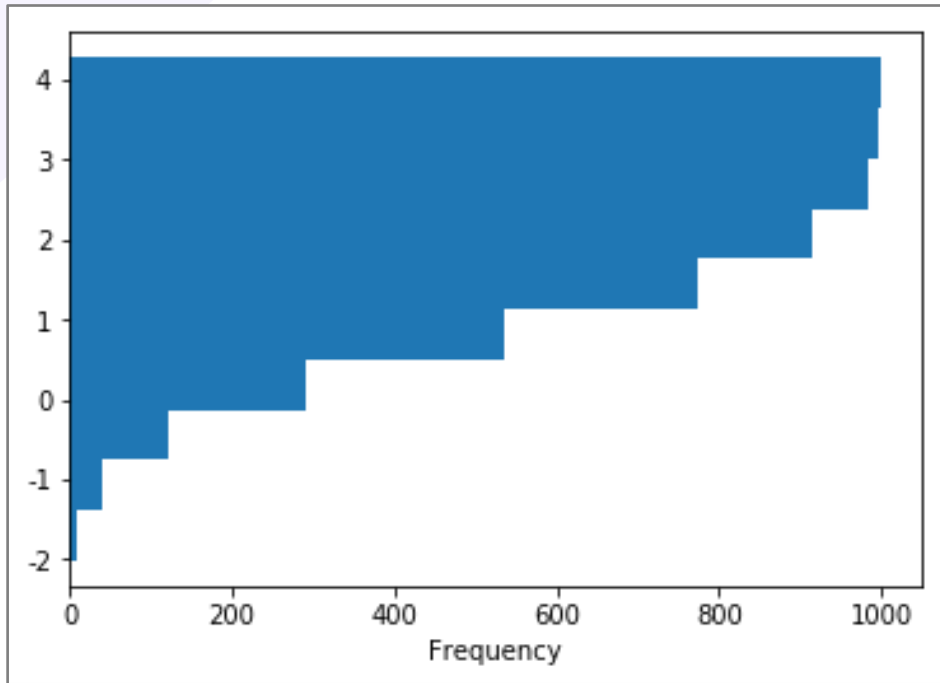
`df4.plot.hist(stacked=True, bins=20)`



- `staked=True, bins=20`
  - 누적 합, bin 값이 20개

# 히스토그램과 도수분포표

03 `df4['a'].plot.hist(orientation='horizontal', cumulative=True)`

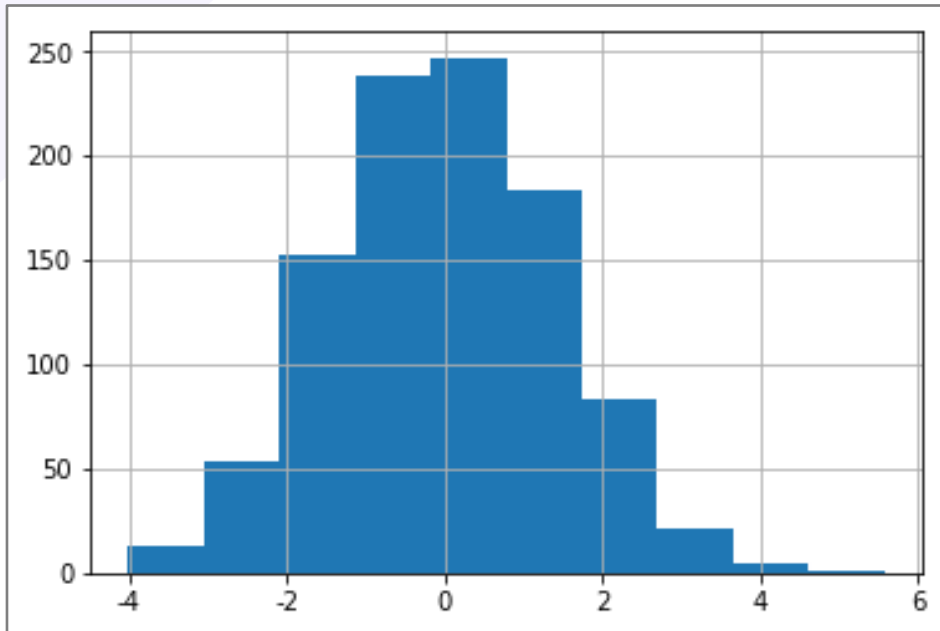


- `orientation='horizontal', cumulative=True`
- 수평 방향, 누적해서 표현



# 차분 diff 구하고 히스토그램으로 표현하기

01 `df4['a'].diff().hist()`

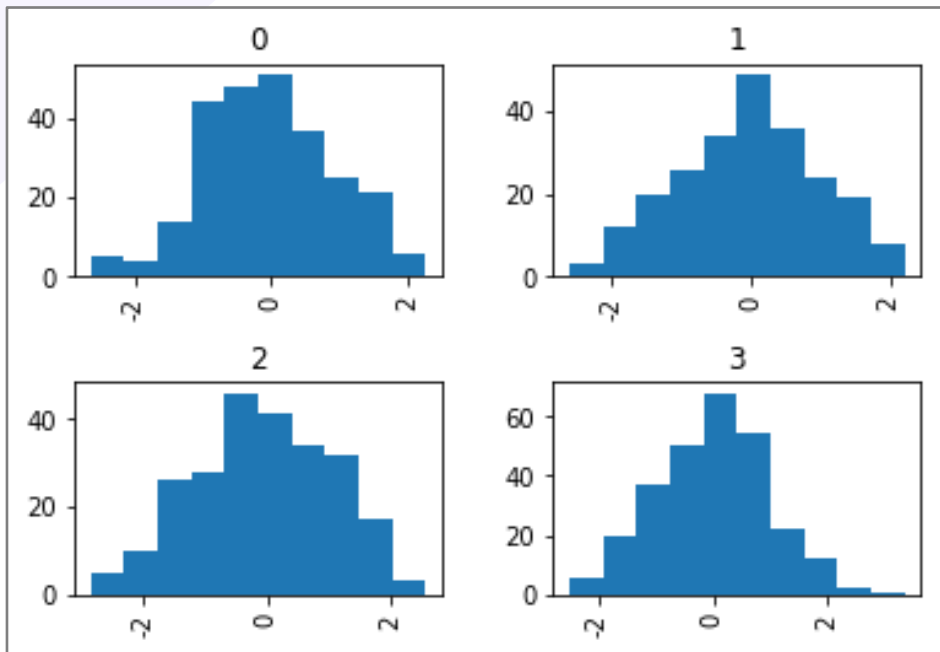


■ `diff` : 차분

# 차분 diff 구하고 히스토그램으로 표현하기

02

```
data.hist(by=np.random.randint(0,4,1000), figsize=(6,4))
```



- Category 별로 구분해서 데이터 시각화



# 차분 diff 구하고 히스토그램으로 표현하기

---

이론 영상 후 실습 영상 제시





# 데이터 시각화\_ 통계 그래프 그리기

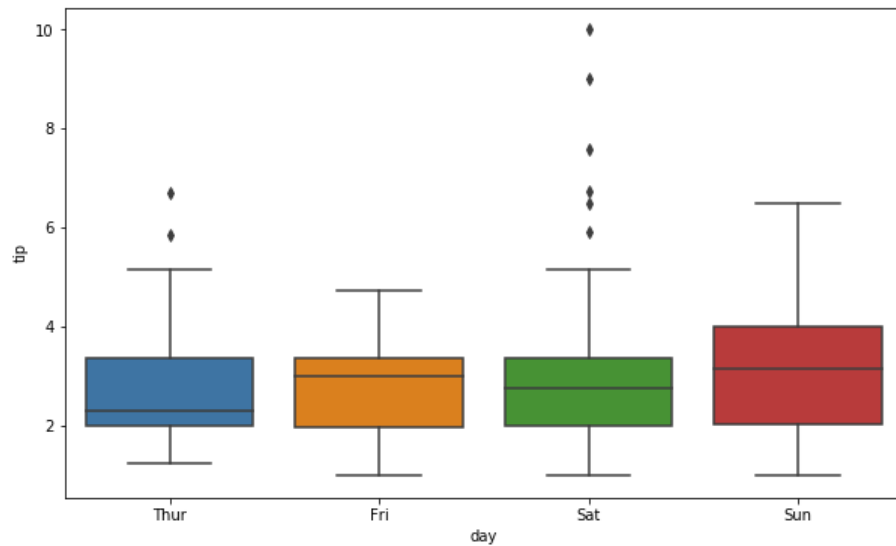




# 상자 수염(Box Plot)

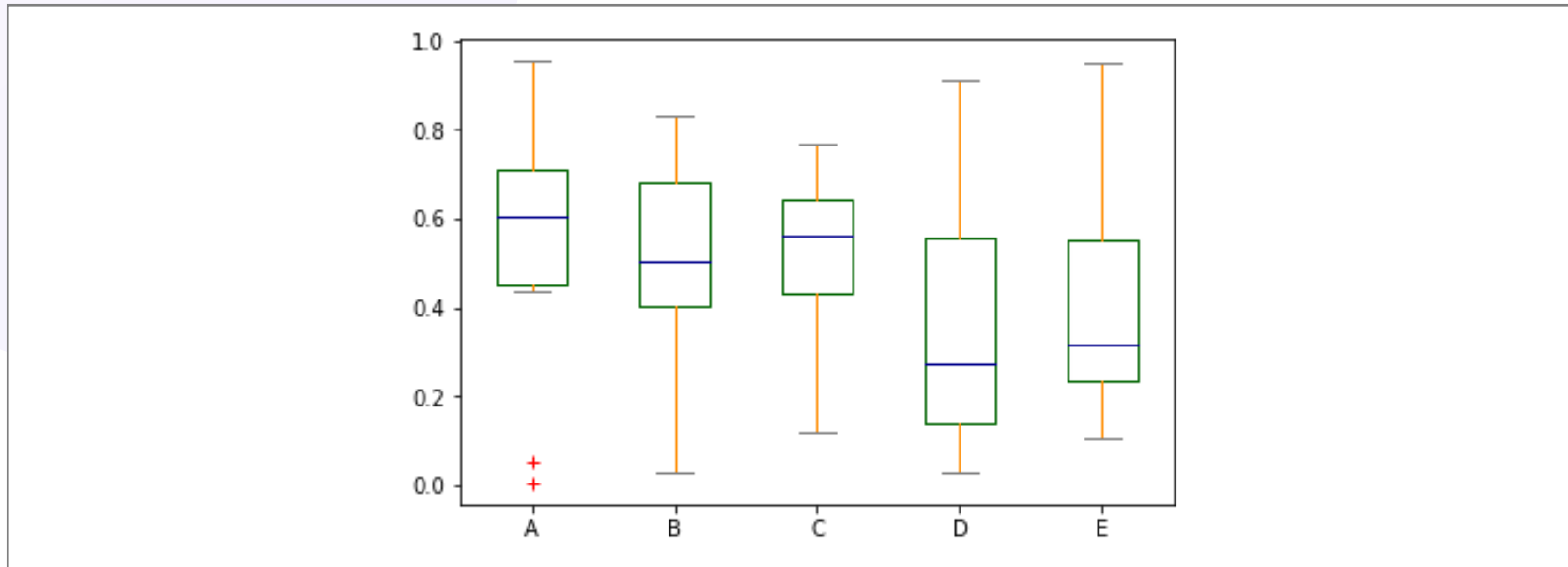
가공하지 않은 자료 그대로를 이용하여 그린 것이 아니라,  
자료로부터 얻어낸 통계량인 **5가지 요약 수치**로 그림

5가지 요약 수치란 기술통계학에서 자료의 정보를 알려주는  
아래의 다섯 가지 수치를 의미함



- 최솟값
- 제1사분위수
- 제2사분위수( ), 즉 중앙값
- 제3사분위수( )
- 최댓값

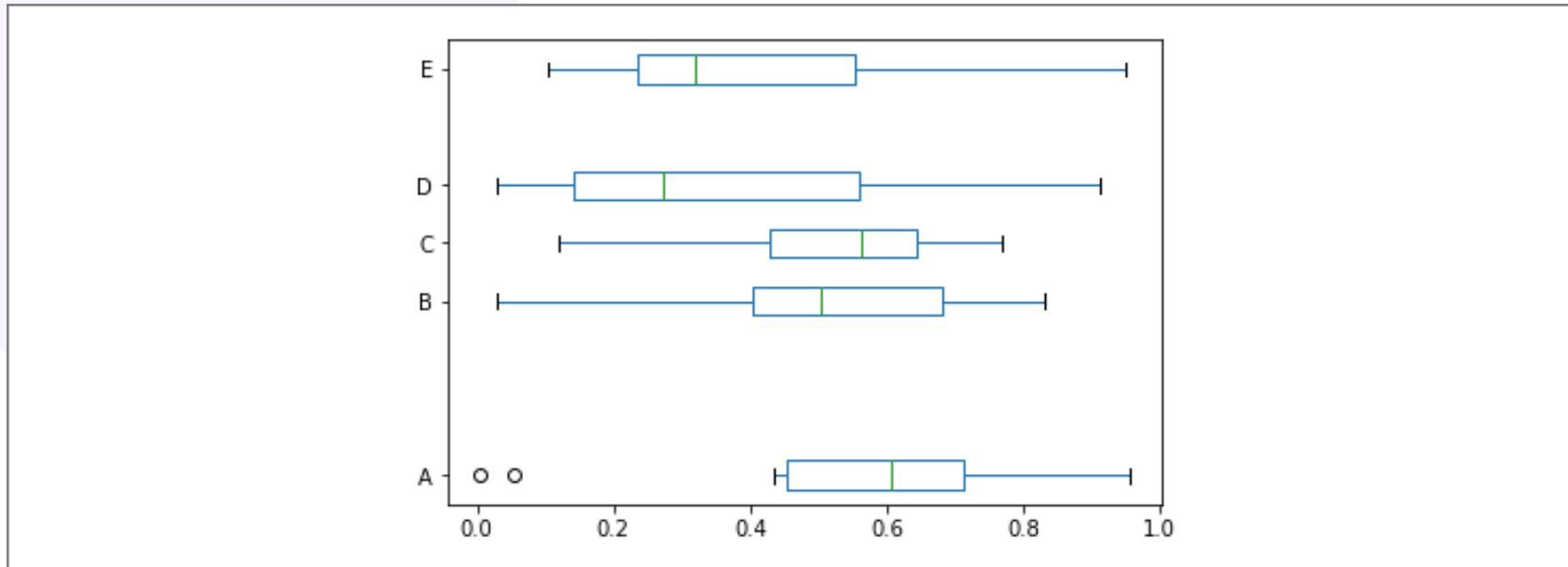
# 상자 수염(Box Plot)



01 `df.plot.box(color=color, sym='r+')`

- `sym='r+'`
  - Symbol은 빨간색 + 형태로 표현하라는 옵션

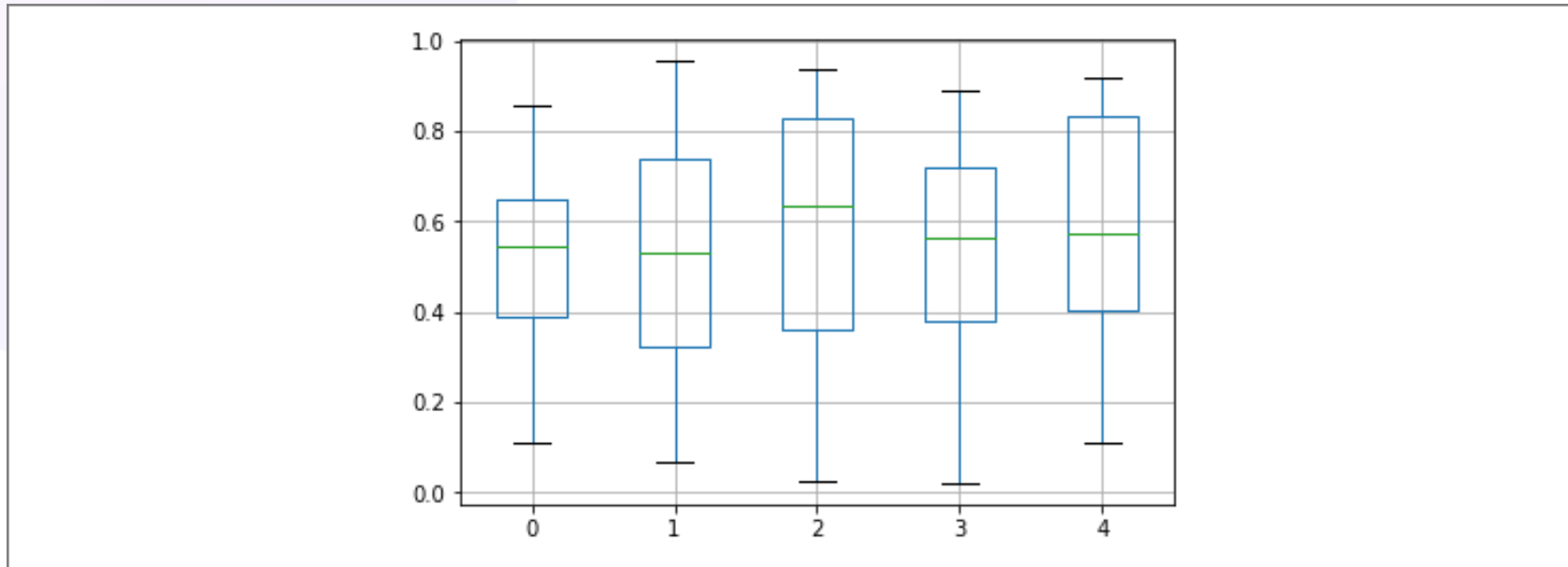
# 상자 수염(Box Plot)



02 `df.plot.box(vert=False, positions=[1, 4, 5, 6, 8])`

- `vert=False`
  - Vertical(수직 방향)이 아니게 표현

# 상자 수염(Box Plot)



03

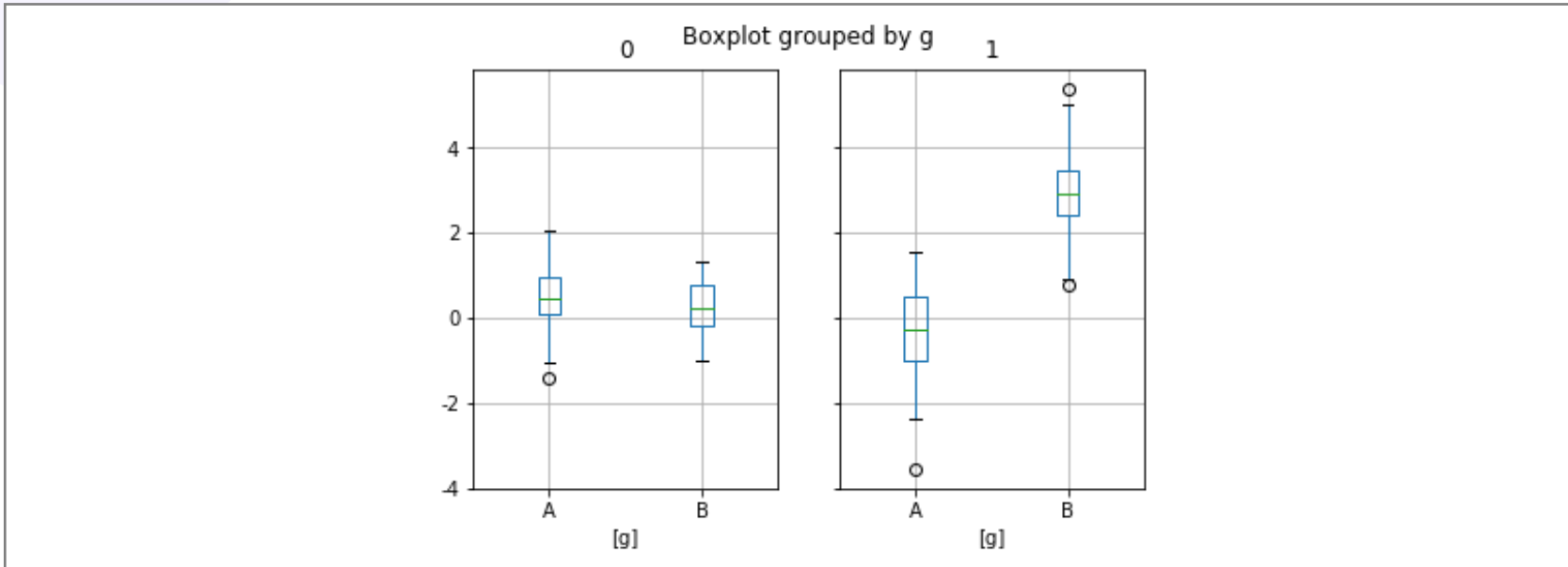
`bp = df.boxplot()`

- Grid가 나타나게 그래프를 그려줌

# 상자 수염(Box Plot)

04 `np.random.seed(1234)`

- Seed 값을 주어 난수가 계속 변하지 않게 값을 고정함



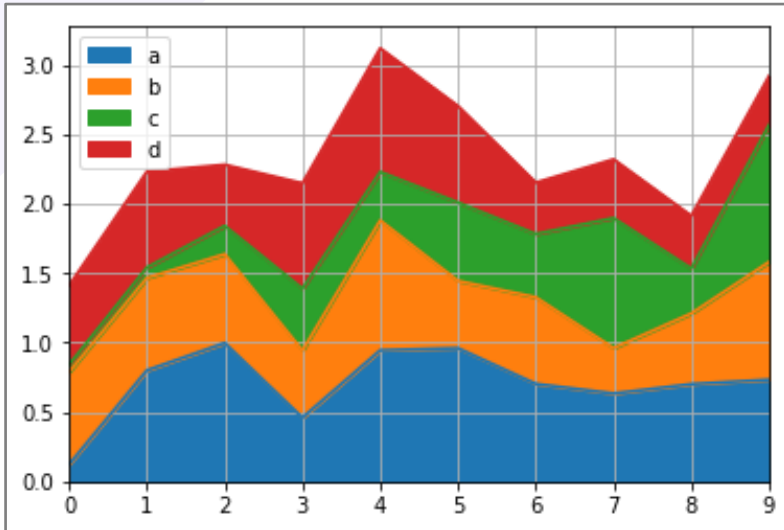
05 `Df_box.boxplot(by='g'), bp = df_box.groupby('g').boxplot()`

- 'g' 그룹별로 나누어 박스플롯 그래프를 표현



# 그래프 격자 만들기(Area Plot, Grid 옵션 활용)

## 01 df.plot.area(grid=True)

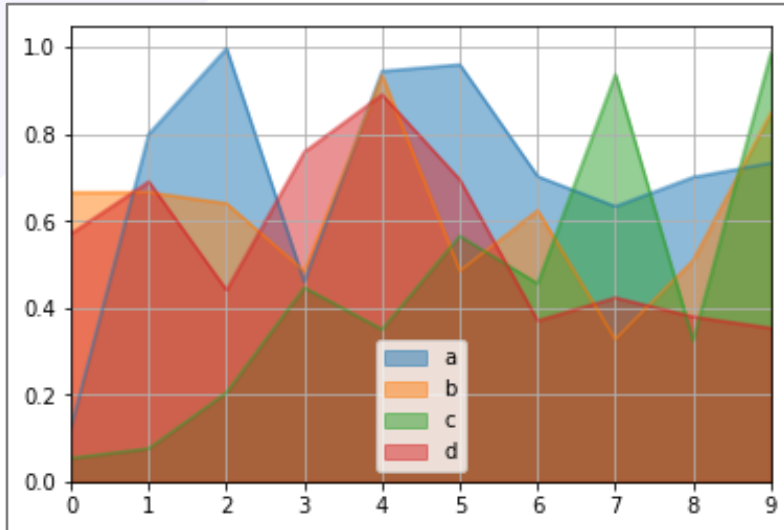


- NaN이 있으면 0으로 채움
- grid=True : 격자 표현 옵션

# 그래프 격자 만들기(Area Plot, Grid 옵션 활용)

02

`df.plot.area(stacked=False, grid=True)`



## ■ Area Plot

- `stacked=True`가 Default임



# 그래프 격자 만들기(Area Plot, Grid 옵션 활용)

이론 영상 후 실습 영상 제시





학습목차	학습평가	화면설명																
<div>평가하기</div> <div>- 학습평가</div> <div>-----</div> <div>정리하기</div> <div>- 학습정리</div>	<div>학습한 내용을 바탕으로 다음 문제를 풀어 보세요.</div> <table><tr><th>번호</th><th>문제</th><th>정답</th><th>해설</th></tr><tr><td>1</td><td>시계열 분석에 사용되는 판다스 모듈로서 요약 기능을 허용하는 Rolling 객체를 길이 n개의 창에 적용하여 반환하며 이동평균 구할 때 주로 사용하는 것은?  ① Expanding ② Rolling ③ Windows ④ Plot</td><td>2</td><td>Rolling 에 관한 설명이다.</td></tr><tr><td>2</td><td>df2.plot.bar(stacked=True)에 대한 설명 중 <b>틀린</b> 것은? ① 막대 그래프 그리기 위한 코드이다. ② stacked=True 란 누적한 값을 시각화하기 위한 옵션이다. ③ 가로막대 그래프를 그리기 위한 코드이다. ④ Bar 대신 Barh를 하면 그래프 모양이 달라진다.</td><td>3</td><td>가로막대 그래프를 그리기 위해서는 Barh 코드를 이용한다.</td></tr><tr><td>3</td><td>df2.shift(1) 판다스 명령어를 입력하면 어떠한 현상이 일어나는가? ① 도수분포표는 특정 구간에 속하는 자료의 개수를 표현한다 ② 히스토그램은 도수분포표를 시각화하여 막대 그래프로 표현한 것으로 막대 그래프의 일종이다. ③히스토그램은 구간별 빈도수를 표현한다. ④ 정규분포는 확률밀도함수이다.</td><td>2</td><td>히스토그램은 도수분포표를 시각화하여 막대 그래프로 표현한 것이나 막대 그래프와 다르다. 막대 그래프가 이산형 데이터에 사용되는 반면에 히스토그램은 연속형 데이터에 사용된다.</td></tr></table>	번호	문제	정답	해설	1	시계열 분석에 사용되는 판다스 모듈로서 요약 기능을 허용하는 Rolling 객체를 길이 n개의 창에 적용하여 반환하며 이동평균 구할 때 주로 사용하는 것은?  ① Expanding ② Rolling ③ Windows ④ Plot	2	Rolling 에 관한 설명이다.	2	df2.plot.bar(stacked=True)에 대한 설명 중 <b>틀린</b> 것은? ① 막대 그래프 그리기 위한 코드이다. ② stacked=True 란 누적한 값을 시각화하기 위한 옵션이다. ③ 가로막대 그래프를 그리기 위한 코드이다. ④ Bar 대신 Barh를 하면 그래프 모양이 달라진다.	3	가로막대 그래프를 그리기 위해서는 Barh 코드를 이용한다.	3	df2.shift(1) 판다스 명령어를 입력하면 어떠한 현상이 일어나는가? ① 도수분포표는 특정 구간에 속하는 자료의 개수를 표현한다 ② 히스토그램은 도수분포표를 시각화하여 막대 그래프로 표현한 것으로 막대 그래프의 일종이다. ③히스토그램은 구간별 빈도수를 표현한다. ④ 정규분포는 확률밀도함수이다.	2	히스토그램은 도수분포표를 시각화하여 막대 그래프로 표현한 것이나 막대 그래프와 다르다. 막대 그래프가 이산형 데이터에 사용되는 반면에 히스토그램은 연속형 데이터에 사용된다.	<div>[학습평가 페이지]</div> <div>▪ 페이지 퀴즈 컴포넌트 사용하여 페이지 개발</div>
번호	문제	정답	해설															
1	시계열 분석에 사용되는 판다스 모듈로서 요약 기능을 허용하는 Rolling 객체를 길이 n개의 창에 적용하여 반환하며 이동평균 구할 때 주로 사용하는 것은?  ① Expanding ② Rolling ③ Windows ④ Plot	2	Rolling 에 관한 설명이다.															
2	df2.plot.bar(stacked=True)에 대한 설명 중 <b>틀린</b> 것은? ① 막대 그래프 그리기 위한 코드이다. ② stacked=True 란 누적한 값을 시각화하기 위한 옵션이다. ③ 가로막대 그래프를 그리기 위한 코드이다. ④ Bar 대신 Barh를 하면 그래프 모양이 달라진다.	3	가로막대 그래프를 그리기 위해서는 Barh 코드를 이용한다.															
3	df2.shift(1) 판다스 명령어를 입력하면 어떠한 현상이 일어나는가? ① 도수분포표는 특정 구간에 속하는 자료의 개수를 표현한다 ② 히스토그램은 도수분포표를 시각화하여 막대 그래프로 표현한 것으로 막대 그래프의 일종이다. ③히스토그램은 구간별 빈도수를 표현한다. ④ 정규분포는 확률밀도함수이다.	2	히스토그램은 도수분포표를 시각화하여 막대 그래프로 표현한 것이나 막대 그래프와 다르다. 막대 그래프가 이산형 데이터에 사용되는 반면에 히스토그램은 연속형 데이터에 사용된다.															