

빅분기 실습 프로젝트

황삭기 MES 데이터 분석

데이터 시각화

01. 시각화 라이브러리

matplotlib

설치(cmd)

`pip install matplotlib`

사용

`import matplotlib.pyplot as plt`

특장점

- Python의 가장 오래된 시각화 라이브러리
- 고도로 커스터마이징 가능
- matplotlib.pyplot 서브모듈 중심 사용

```
import matplotlib.pyplot as plt
```

seaborn

`pip install seaborn`

`import seaborn as sns`

- 통계적 데이터 시각화 라이브러리
- Matplotlib 기반, 간결하고 미려한 그래프 제공
- 다양한 테마와 컬러 팔레트 제공

```
import seaborn as sns
```

02. 설치 방법

```
명령 프롬프트
Microsoft Windows [Version 10.0.22621.2428]
(c) Microsoft Corporation. All rights reserved.

C:\Users\USER>python --version
Python 3.9.7

C:\Users\USER>pip install numpy pandas openpyxl xlrd xlwt matplotlib seaborn scikit-learn scipy
Collecting numpy
  Downloading numpy-1.26.2-cp39-cp39-win_amd64.whl (15.8 MB)
    | 15.8 MB 6.8 MB/s
Collecting pandas
  Downloading pandas-2.1.3-cp39-cp39-win_amd64.whl (10.8 MB)
    | 10.8 MB ...
Collecting openpyxl
  Downloading openpyxl-3.1.2-py2.py3-none-any.whl (249 kB)
    | 249 kB ...
Collecting xlrd
  Downloading xlrd-2.0.1-py2.py3-none-any.whl (96 kB)
    | 96 kB 2.7 MB/s
Collecting xlwt
  Downloading xlwt-1.3.0-py2.py3-none-any.whl (99 kB)
    | 99 kB 6.8 MB/s
```

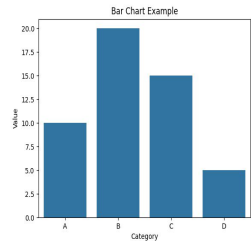
윈도우 검색창 > 명령 프롬프트
(또는 cmd)

> pip install matplotlib
seaborn

03. 기본적인 시각화 유형



막대 그래프 (Bar Chart)

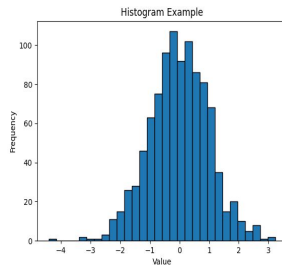


범주형 데이터 비교

x : category

y : value

히스토그램 (Histogram)



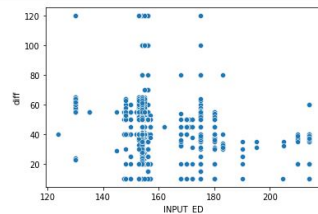
데이터 분포 확인

x : value

y : frequency

bins = 범위

산점도 (Scatter Plot)

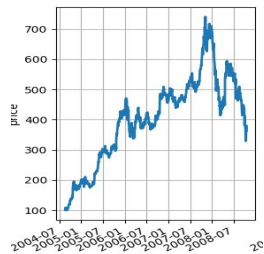


x-y 상관관계

x : value

y : value

선 그래프 (Line Chart)

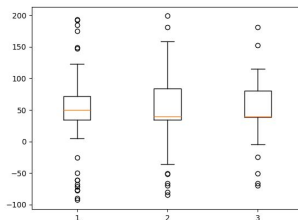


선 그래프는 시간에 따른 변화

x : value (시간 등)

y : value

박스플롯 (Box Plot)

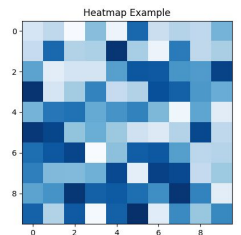


데이터의 분포와 이상치

x : columns

y : value

히트맵 (Heatmap)



행렬 형태의 데이터를 색상으로 표현

x, y 조합으로 나타나는 값

04. 시각화 그래프 용어

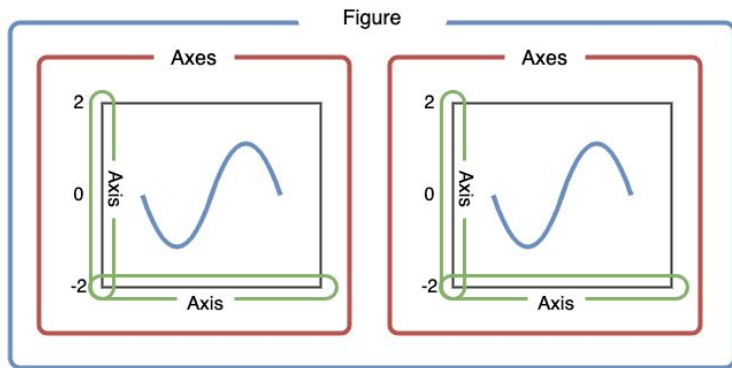
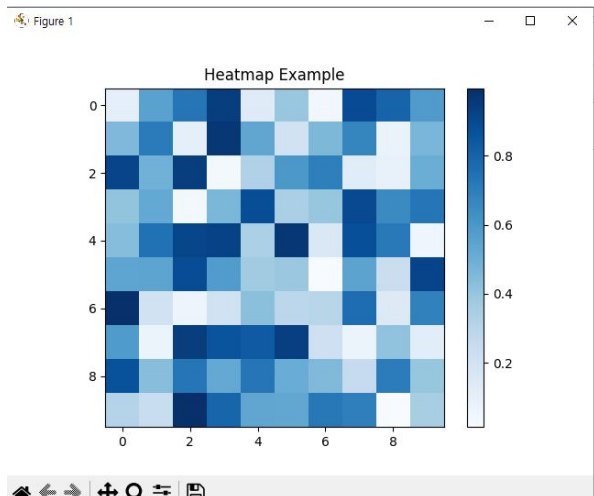


Figure	Axes (좌표축)	Axis
<p>전체 그림을 나타내는 컨테이너</p> <p>하나 이상의 Axes 포함 가능</p> <p><code>plt.figure()</code>로 생성</p> <p><code>plt.subplots()</code> 여러 플롯 배치</p>	<p>실제 플롯이 그려지는 영역</p> <p>X축, Y축, 타이틀 등 포함</p> <p><code>fig.add_subplot()</code> 또는 <code>plt.subplot()</code>로 생성</p>	<p>Axes 내의 축(X축과 Y축)</p> <p>데이터의 범위와 눈금 설정 가능</p> <p><code>ax.set_xlim()</code>, <code>ax.set_ylim()</code>으로 범위 설정</p>

05. figure와 axes

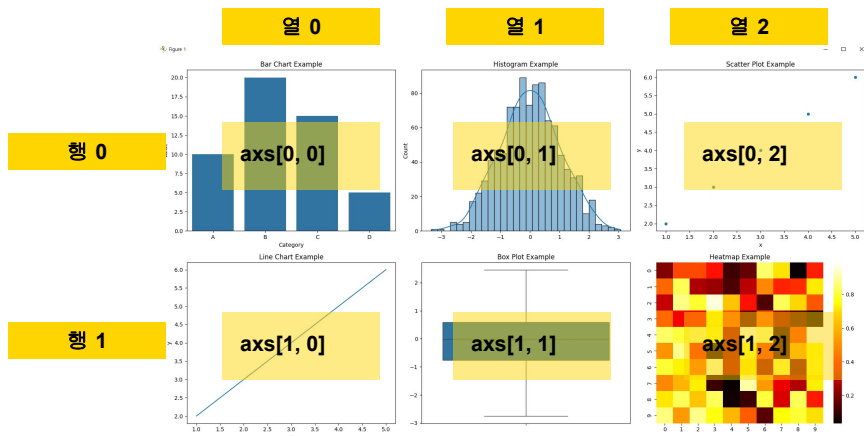
figure, axes 에 대한 설정 없이 자동으로 현재 공간,
범위에 그래프를 그림 > 1 figure, 1 axe



```
# Figure, Axes 객체 생성 없이 그래프만  
sns.heatmap(data)
```

figure, axes 에 대한 값을 설정

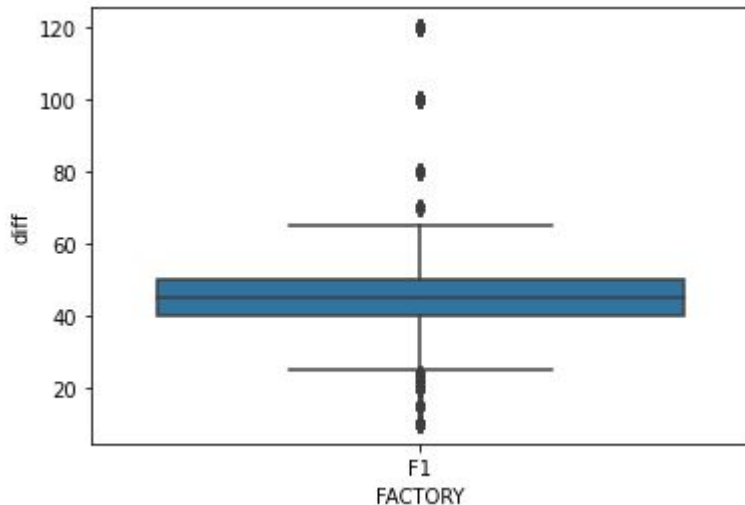
원하는 figure 객체 (=그릴 공간, 종이)에 원하는 axes
객체(= 위치, figure 내 `axs[행, 열]` 지정)



```
# Figure와 Axes 객체 생성  
fig, axs = plt.subplots(2, 3, figsize=(18, 10))
```

향후 업데이트 예정..

BOXPLOT



상자 (Box)

- **중앙값 (Median)** : 상자의 가운데에 그려진 선 → 데이터의 중앙값(50th percentile)
- **1사분위수 (Q1)** : 상자의 아래쪽 경계 → 데이터의 25th percentile
- **3사분위수 (Q3)** : 상자의 위쪽 경계 → 데이터의 75th percentile
- **상자 너비** : Q1과 Q3 사이의 거리로, 데이터의 IQR (Interquartile Range)을 나타냅니다.

수염 (Whiskers)

- 수염은 상자에서 확장된 선 → 데이터의 최소값과 최대값 범위
- 데이터가 이상치(outlier)를 포함하는 경우, 수염의 끝은 $Q1 - 1.5 * IQR$ 과 $Q3 + 1.5 * IQR$ 범위로 제한

이상치 (Outliers)

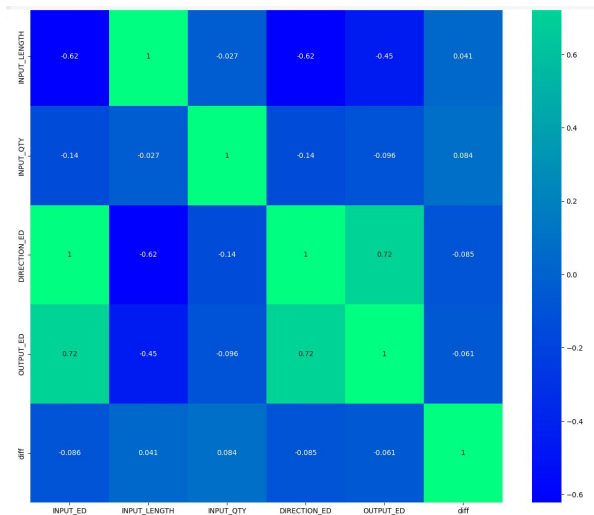
- 상자와 수염의 범위를 넘어서는 개별 데이터 포인트
- $Q1 - 1.5 * IQR$ 보다 작거나 $Q3 + 1.5 * IQR$ 보다 큰 값을 가진 데이터 포인트로, 개별 점으로 표시

IQR (Interquartile Range)

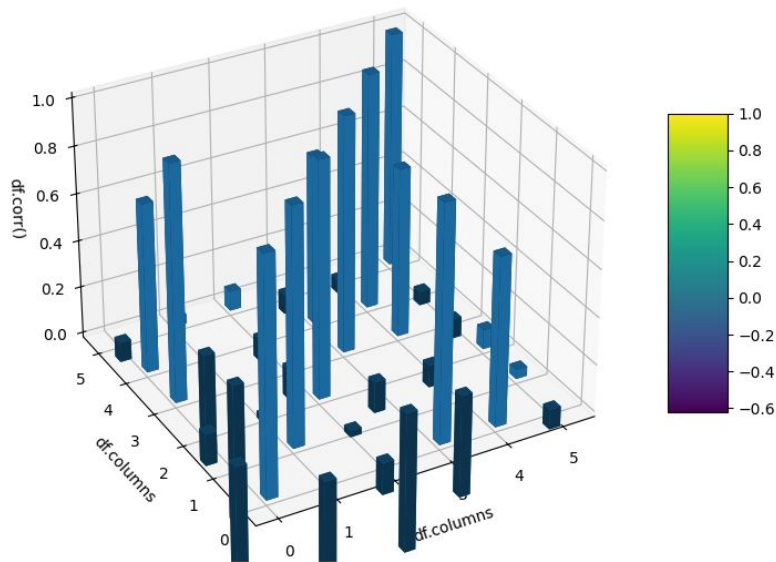
- 상자의 너비 → 데이터의 중간 50% 범위

[재미로..] interactive 3D graph _ Heatmap (df.corr())

	INPUT_ED	INPUT_LENGTH	INPUT_QTY	DIRECTION_ED	OUTPUT_ED	diff
INPUT_ED	1.00000	-0.62267	-0.13687	0.99855	0.71561	-0.08587
INPUT_LENGTH	-0.62267	1.00000	-0.02741	-0.62209	-0.45215	0.04079
INPUT_QTY	-0.13687	-0.02741	1.00000	-0.13510	-0.09600	0.08367
DIRECTION_ED	0.99855	-0.62209	-0.13510	1.00000	0.71670	-0.08464
OUTPUT_ED	0.71561	-0.45215	-0.09600	0.71670	1.00000	-0.06076
diff	-0.08587	0.04079	0.08367	-0.08464	-0.06076	1.00000



3D Surface Plot with Color Mapping



https://github.com/millim1983/bigdata_project01/blob/main/interactive_3d_test_colormapping.py