Py 기본분석 : 그림 그리기

I. 그림 그리기 패키지

II. 그림 그리기

Ⅰ. 그림 그리기 패키지

- **그림 그리기 패키지로는 seaborn 및 matplotlib.pyplot 등이 있음**
- **seaborn 패키지의 주요 함수는 다음과 같음**

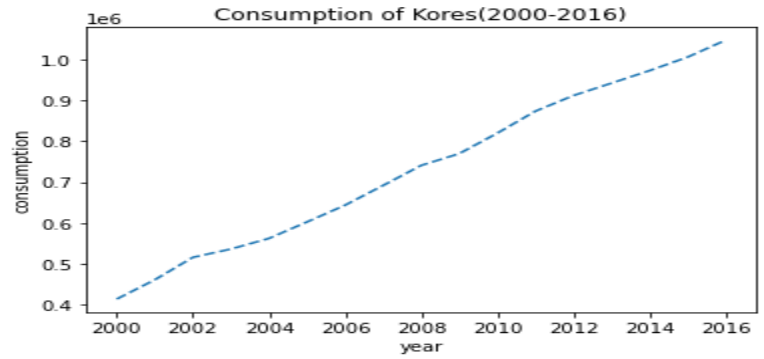
함수	그래프
sns.scatterplot()	산점도
sns.barplot()	막대그래프 – 요약표 활용
sns.counterplot()	막대그래프 – 원자료 활용
sns.lineplot()	선대그래프
sns.boxplot()	상자 그림

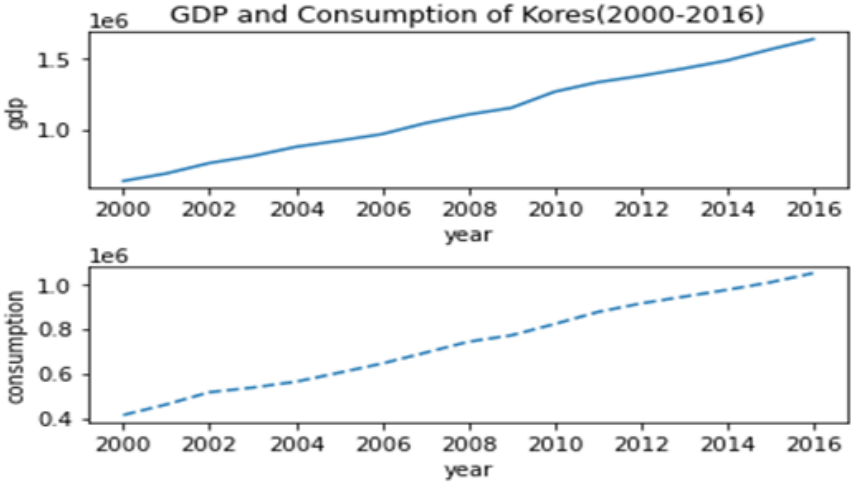
* import seaborn as sns를 제일 먼저 실행해야 함

1. 선 그래프

b1-ch3-1.py

```
import pandas as pd # pandas package를 이용
sample1 = pd.read_csv("http://kanggc.iptime.org/book/data/csv_sample1.csv")
sample1
year = sample1['year']
gdp = sample1['gdp']
consumption = sample1['consumption']


import matplotlib.pyplot as plt # matplotlib.pyplot package를 이용
import seaborn as sns # seaborn package를 이용


sns.lineplot(data=sample1, x='year', y='gdp').set(title='GDP of Kores(2000-2016)')
plt.show()
sns.lineplot(data=sample1, x='year', y='consumption', linestyle='dashed').set(title='
Consumption of Kores(2000-2016)')


#fig = plt.figure(figsize=(10,7))
#fig, axs = plt.subplots(2, 1) # 두 개의 그래프를 한 페이지에 그림
fig, axs = plt.subplots(nrows=2)
sns.lineplot(x='year', y='gdp', data=sample1, ax=axs[0]).set(title='GDP and Consum
ption of Kores(2000-2016)')
sns.lineplot(x='year', y='consumption', data=sample1, linestyle='dashed',ax=axs[1])


fig.subplots_adjust(hspace=.5) # 하위그림의 상하 간격을 조정
plt.savefig('C:/BOOK/pyBasics/pyStat/code/line.png')
```

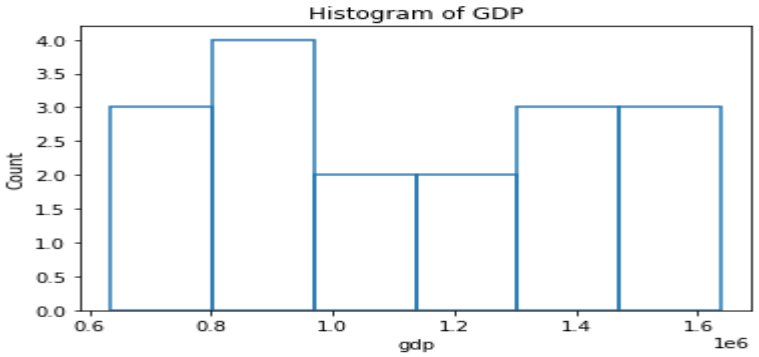


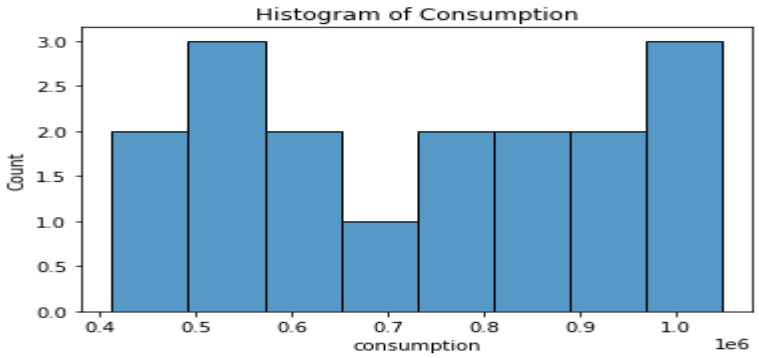


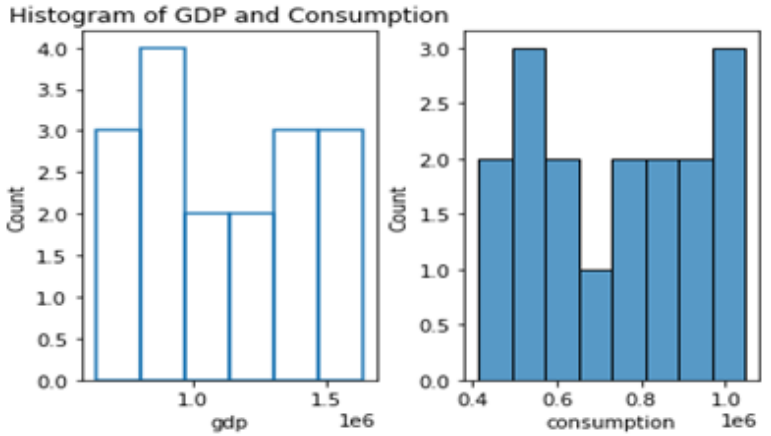
2. 히스토그램

b1-ch3-2.py

```
import pandas as pd # pandas package를 이용
sample1 = pd.read_csv("http://kanggc.iptime.org/book/data/csv_sample1.csv")
sample1
year = sample1['year']
gdp = sample1['gdp']
consumption = sample1['consumption']

import matplotlib.pyplot as plt # matplotlib.pyplot package를 이용
import seaborn as sns # seaborn package를 이용

sns.histplot(data=sample1, x='gdp', fill=False).set(title='Histogram of GDP')
plt.show()
sns.histplot(data=sample1, x='consumption', bins=8).set(title='Histogram of Consumption')

#fig = plt.figure(figsize=(10,7))
#figure, axs = plt.subplots(1,2) # 두 개의 그래프를 한 페이지에 그림
fig, axs = plt.subplots(ncols=2)
sns.histplot(x='gdp', data=sample1, fill=False,ax=axs[0]).set(title='Histogram of GDP and Con
sumption')
sns.histplot(x='consumption', data=sample1, bins=8, ax=axs[1])

fig.subplots_adjust(wspace=0.3) # 우측그림의 좌우 간격을 조정
plt.savefig('C:/BOOK/pyBasics/pyStat/code/hist.png')
```



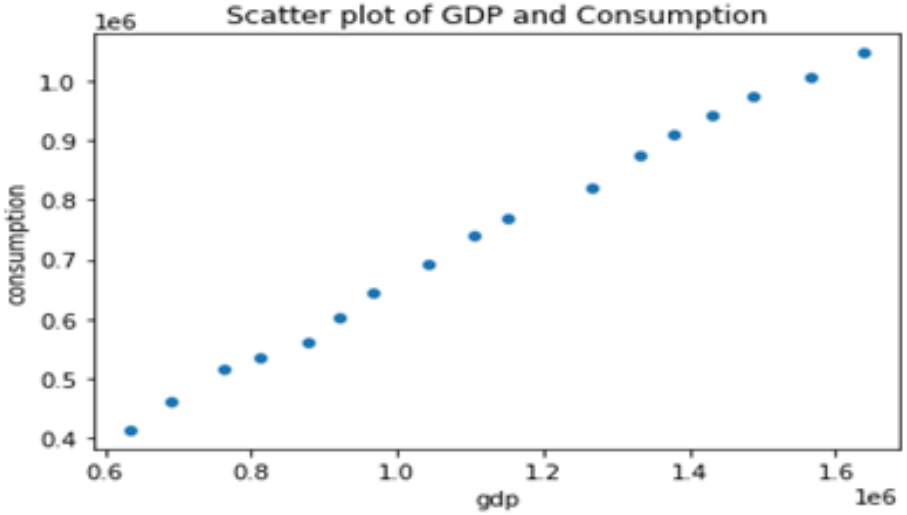




3. 산포도

b1-ch3-3.py

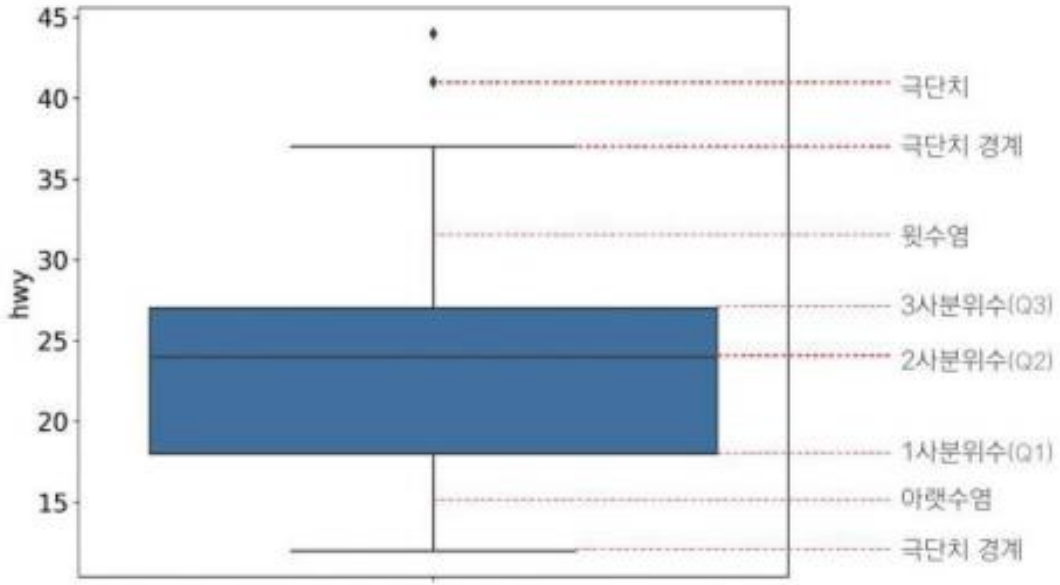
```
import pandas as pd # pandas package를 이용
sample1 = pd.read_csv("http://kanggc.iptime.org/book/data/csv_sample1.csv")
sample1
year = sample1['year']
gdp = sample1['gdp']
consumption = sample1['consumption']

import matplotlib.pyplot as plt # matplotlib.pyplot package를 이용
import seaborn as sns # seaborn packagem를 이용

sns.scatterplot(data=sample1, x='gdp', y='consumption').set(title='Scatter plot of GDP and Consumption')

#fig = plt.figure(figsize=(10,7))
plt.savefig('C:/BOOK/pyBasics/pyStat/code/scatter.png')
```



4. 상자그래프

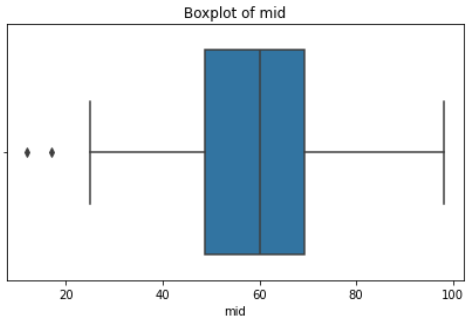
상자그림	값	설명
상자 아래 세로선	아랫수염	하위 0~25% 내에 해당하는 값
상자 밑면	1사분위수(Q1)	하위 25% 위치 값
상자 내 굵은 선	2사분위수(Q2)	하위 50% 위치 값(중앙값)
상자 윗면	3사분위수(Q3)	하위 75% 위치 값
상자 위 세로선	윗수염	하위 75~100% 내에 해당하는 값
상자 밖 가로선	극단치 경계	Q1, Q3 밖 1.5 IQR 내 최댓값
상자 밖 점 표식	극단치	Q1, Q3 밖 1.5 IQR를 벗어난 값

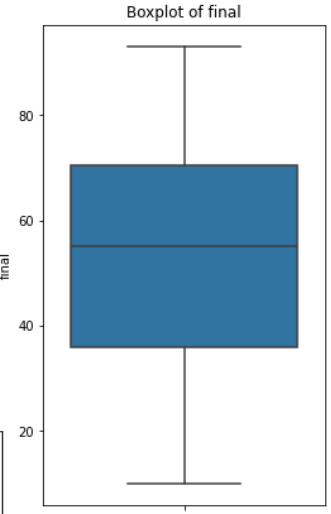
* IQR(사분위 범위)은 1사분위수와 3사분위수의 거리이며, 1.5 IQR은 IQR의 1.5배를 나타냄

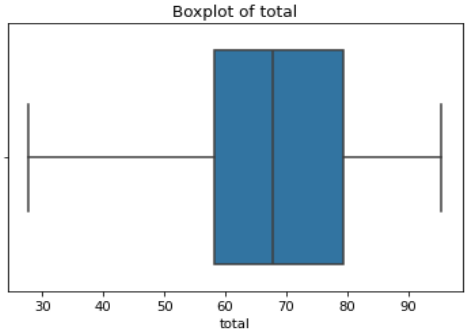


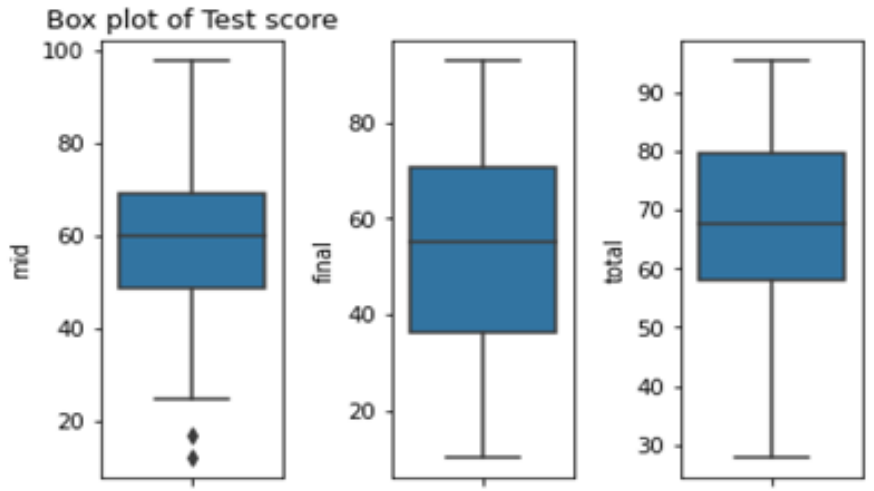
B1-ch3-4.py

```
import pandas as pd # pandas package를 이용
sample1 = pd.read_excel("http://kanggc.iptime.org/book/data/stat-1.xlsx")
sample1
sample1.info()
mid = sample1['mid']
final = sample1['final']
total = sample1['total']
grade = sample1['grade']
import matplotlib.pyplot as plt # matplotlib.pyplot package를 이용
import seaborn as sns # seaborn packagem를 이용
sns.boxplot(data=sample1, x='mid').set(title='Boxplot of mid') # 수평 상자그림
plt.show()
sns.boxplot(data=sample1, y='final').set(title='Boxplot of final') # 수직 상자그림
plt.show()
sns.boxplot(data=sample1, x='total').set(title='Boxplot of total')
plt.show()


fig = plt.figure(figsize=(12,7))
#figure, axs = plt.subplots(1,3) # 세 개의 그래프를 한 페이지에 그림
fig, axs = plt.subplots(ncols=3)
sns.boxplot(y='mid', data=sample1, ax=axs[0]).set(title='Box plot of Test score')
sns.boxplot(y='final', data=sample1, ax=axs[1])
sns.boxplot(y='total', data=sample1, ax=axs[2])


fig.subplots_adjust(wspace=0.6) # 우측그림의 좌우 간격을 조정
plt.savefig('C:/BOOK/pyBasics/pyStat/code/box.png')
```









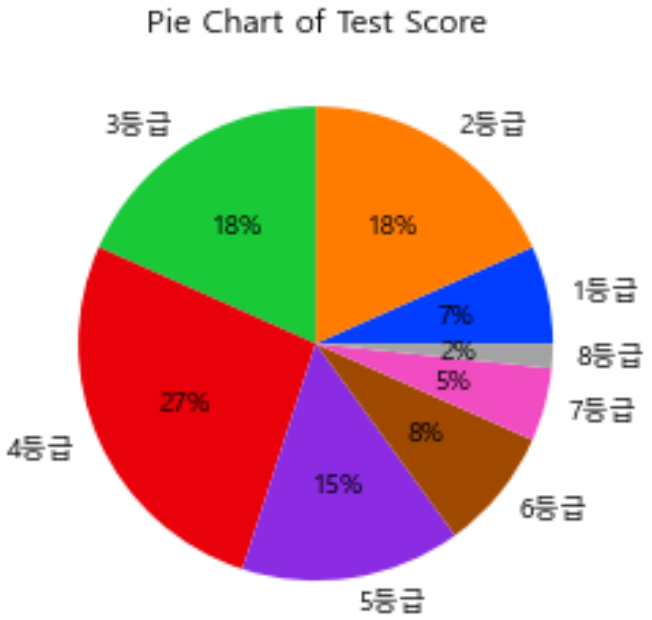
5. 원 그래프

B1-ch3-5.py

```
import pandas as pd # pandas package를 이용
sample1 = pd.read_excel("http://kanggc.iptime.org/book/data/stat-1.xlsx")
sample1
sample1.info()
mid = sample1['mid']
final = sample1['final']
total = sample1['total']
grade = sample1['grade']
#frequencies of each value in grade
#grade.value_counts()
print("Frequency of Grade is : ",f'\n{grade.value_counts()}\n')
import matplotlib.pyplot as plt # matplotlib.pyplot package를 이용
import seaborn as sns # seaborn packagem를 이용
# 한글이 깨지지 않게 하기
import matplotlib
from matplotlib import  font_manager, rc
font_path = "/Windows/Fonts/malgun.ttf"
font = font_manager.FontProperties(fname=font_path).get_name()
rc('font', family=font)
matplotlib.rcParams['axes.unicode_minus'] = False
```

B1-ch3-5.py(계속)

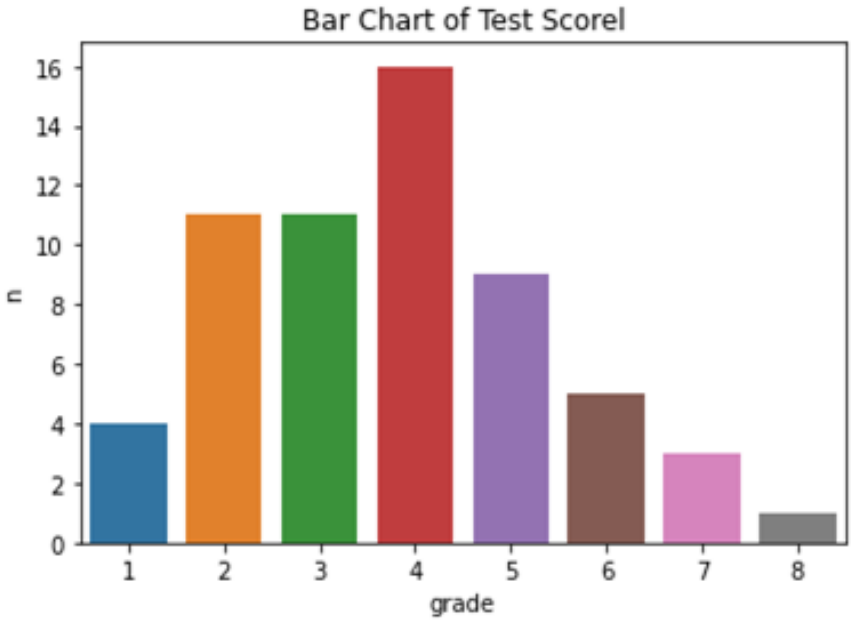
```
data = [4, 11, 11, 16, 9, 5, 3, 1]
keys = ['1등급','2등급','3등급','4등급','5등급','6등급','7등급','8등급']
# define Seaborn color palette to use
palette_color = sns.color_palette('bright')
# plotting data on chart
plt.pie(data, labels=keys, colors=palette_color, autopct='%.0f%%')
plt.title("Pie Chart of Test Score")
# displaying chart
#plt.show()
plt.savefig('C:/BOOK/pyBasics/pyStat/code/pie.png')
```



6. 막대 그래프

b1-ch3-6.py

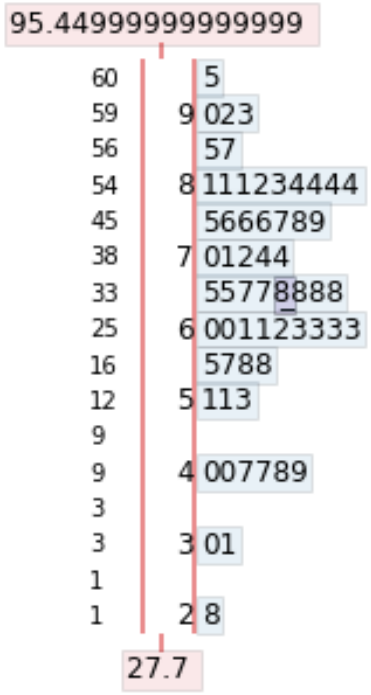
```
import pandas as pd # pandas package를 이용
sample1 = pd.read_excel("http://kanggc.iptime.org/book/data/stat-1.xlsx")
sample1


sample1.info()


#frequencies of each value in grade
df_sample1 = sample1.groupby('grade', as_index = False).agg(n=('grade', 'count'))
df_sample1


print("Frequency of Grade is : ",f'\n{df_sample1}\n')


import matplotlib.pyplot as plt # matplotlib.pyplot package를 이용
import seaborn as sns # seaborn packagem를 이용


sns.barplot(data=df_sample1, x='grade', y='n').set(title='Bar Chart of Test Scorel')


plt.savefig('C:/BOOK/pyBasics/pyStat/code/bar.png')
```



7. 잎-줄기 그래프

b1-ch3-7.R

```
import pandas as pd # pandas package를 이용
sample1 = pd.read_excel("http://kanggc.iptime.org/book/data/stat-1.xlsx")

print("Data of sample1 is : ",f'\n{sample1}\n')

mid = sample1['mid']
final = sample1['final']
total = sample1['total']
grade = sample1['grade']

import stemgraphic # stemgraphic package를 이용

#create stem-and-leaf plot
fig, ax = stemgraphic.stem_graphic(total)
```



Key: aggr|stem|leaf
∞ **9** *0* = **9.0**x10 = 90.0