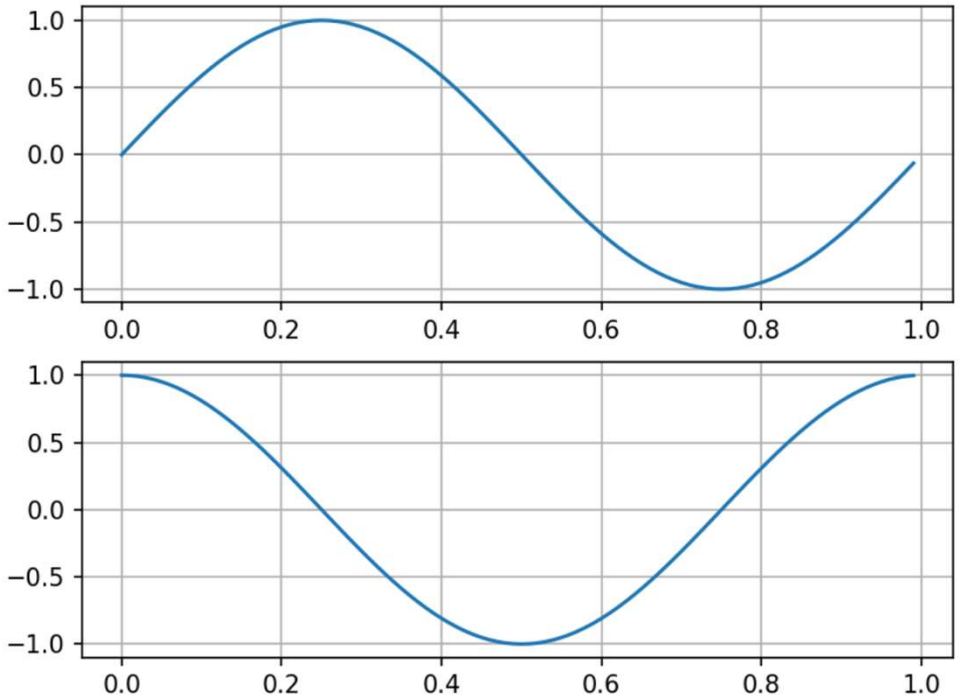
데이터 시각화

그래프

```
import numpy as np
import matplotlib.pyplot as plt
t = np.arange(0, 100) * 0.01
s = np.sin(2 * np.pi * t)
c = np.cos(2 * np.pi * t)

plt.subplot(2, 1, 1); plt.plot(t, s); plt.grid()
plt.subplot(2, 1, 2); plt.plot(t, c); plt.grid()
plt.show()
```




```
import numpy as np
import matplotlib.pyplot as plt

x = np.arange(0, 5, 0.1)
y = np.sin(x)
plt.plot(x, y)
```

https://matplotlib.org/stable/api/_as_gen/matplotlib.pyplot.html

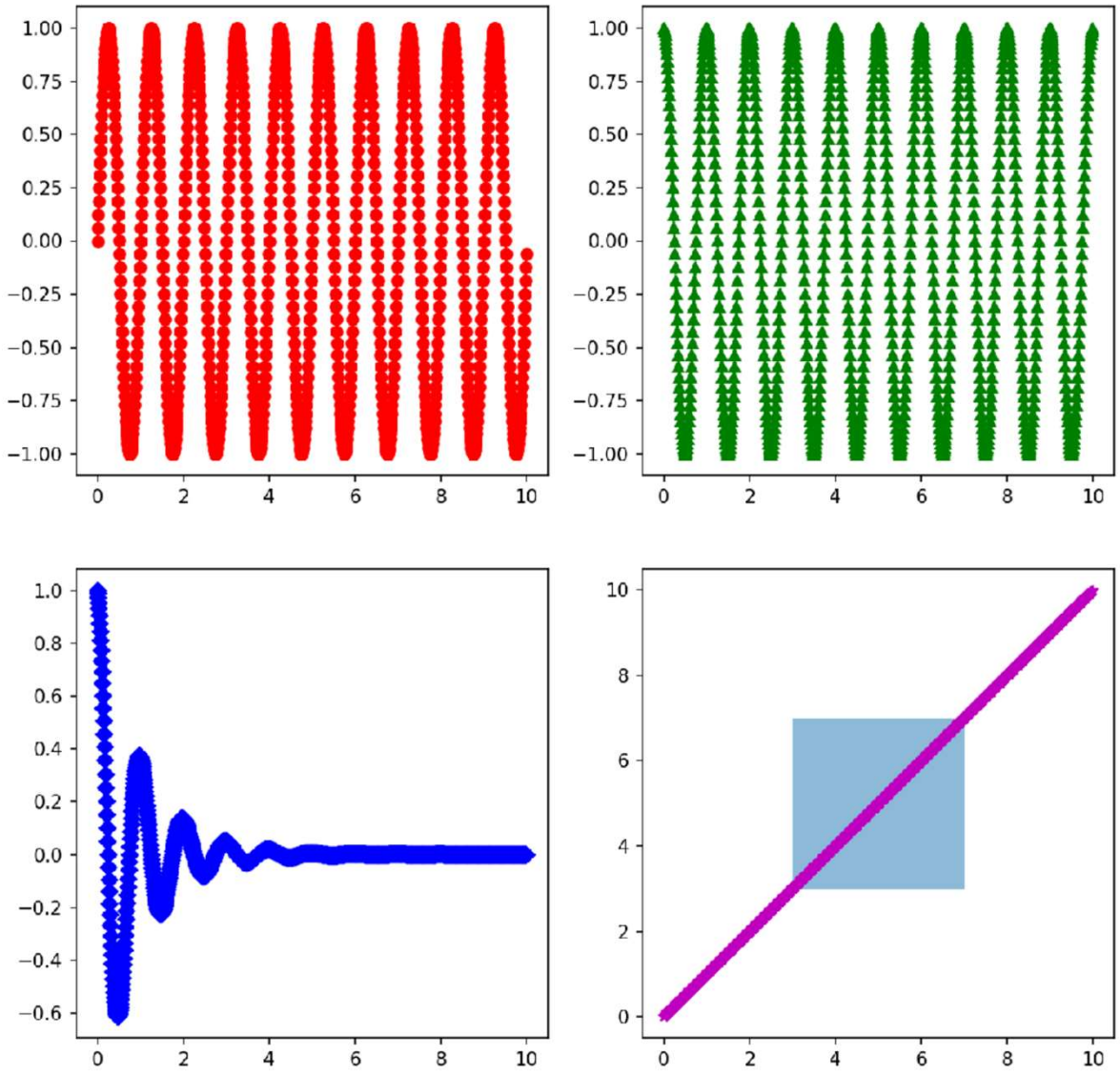
그래프 스타일

```
import numpy as np
import matplotlib.pyplot as plt

t = np.arange(0, 1000) * 0.01
a = np.sin(2 * np.pi * t)
b = np.cos(2 * np.pi * t)
c = np.cos(2 * np.pi * t) * np.exp(-t)
d = t

plt.figure(figsize=(7, 7))

plt.subplot(2, 2, 1); plt.plot(t, a, "ro-")
plt.subplot(2, 2, 2); plt.plot(t, b, "g^--")
plt.subplot(2, 2, 3); plt.plot(t, c, "bD:")
plt.subplot(2, 2, 4); plt.plot(t, d, "m*-.")
plt.fill([3, 3, 7, 7], [3, 7, 7, 3], alpha=0.5)
plt.show()
```



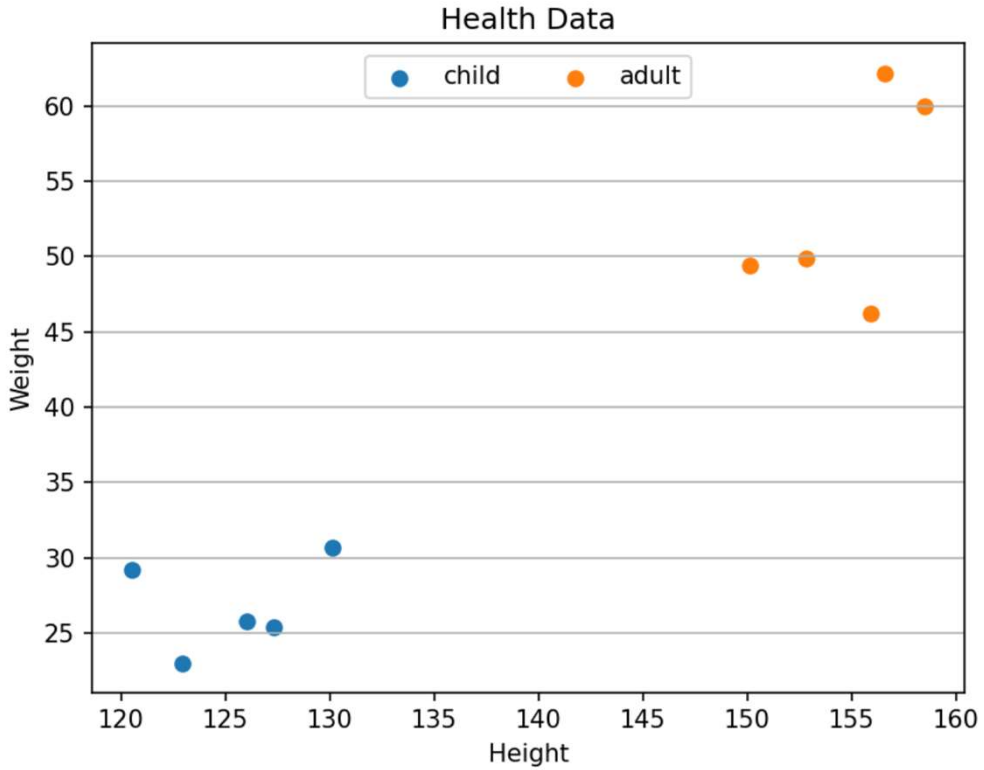
어린이, 청소년 데이터

```python
w = []; h = []; t = []
with open("data/health.csv", "r") as file:
    lines = file.readlines()[1:]
    for line in lines:
        a, b, c = line.strip().split(",")
        h.append(float(a))  # 키
        w.append(float(b))  # 몸무게
        t.append(int(c))    # 어린이/청소년
data = [[x, y] for x, y in zip(h, w)]            # 리스트 생성 [키, 몸무게]
data3 = [[x, y, z] for x, y, z in zip(h, w, t)]  # 리스트 생성 [키, 몸무게, 정답]
ch_h = [x for x, y, z in data3 if z == 1]        # 어린이 키
ch_w = [y for x, y, z in data3 if z == 1]        # 어린이 몸무게
ad_h = [x for x, y, z in data3 if z == 0]        # 청소년 키
ad_w = [y for x, y, z in data3 if z == 0]        # 청소년 몸무게
```

```
H,W,T
130.1,30.7,1
120.5,29.2,1
127.3,25.4,1
122.9,23.0,1
126.0,25.8,1
152.8,49.9,0
155.9,46.2,0
158.5,60.0,0
156.6,62.2,0
150.1,49.4,0
```

산점도

```
import matplotlib.pyplot as plt
plt.scatter(ch_h, ch_w)
plt.scatter(ad_h, ad_w)
plt.xlabel("Height")
plt.ylabel("Weight")
plt.title("Health Data")
plt.xticks()
plt.yticks()
plt.legend(["child", "adult"], ncol = 2, loc = "upper center")
plt.grid(axis = "y")
plt.show()
```



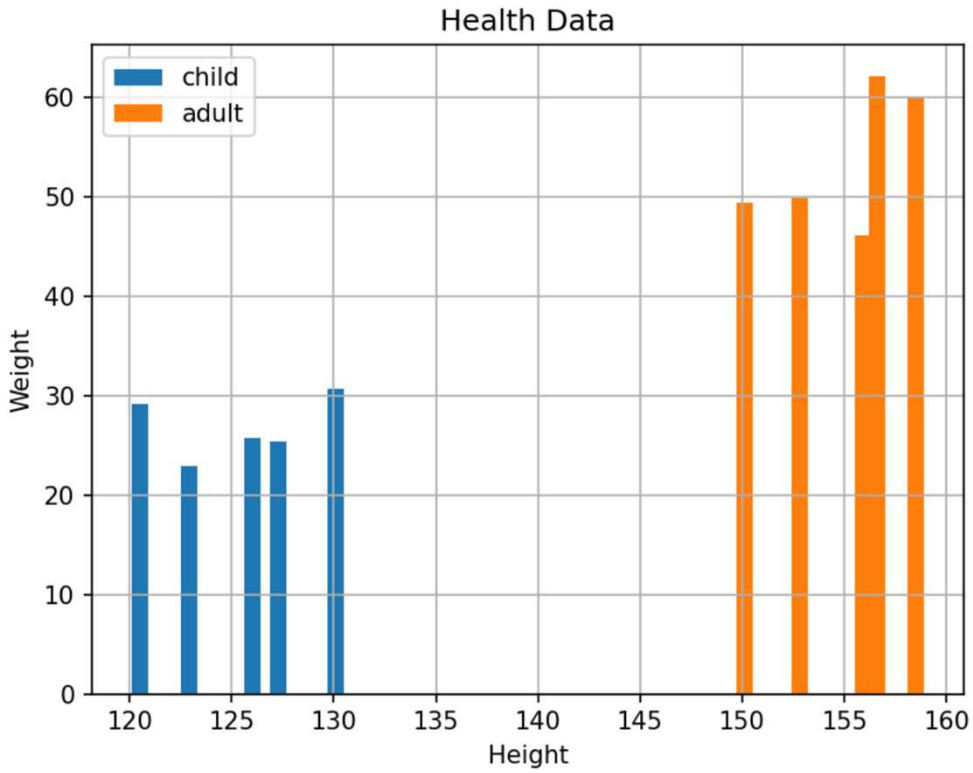

```
plt.legend(["child", "adult"], ncol = 2, loc = "higher center")
```

```
ValueError: 'higher center' is not a valid value for loc; supported values are 'best',
```

```
'upper right', 'upper left', 'lower left', 'lower right', 'right', 'center left', 'center right', 'lower center', 'upper center', 'center'
```

막대그래프

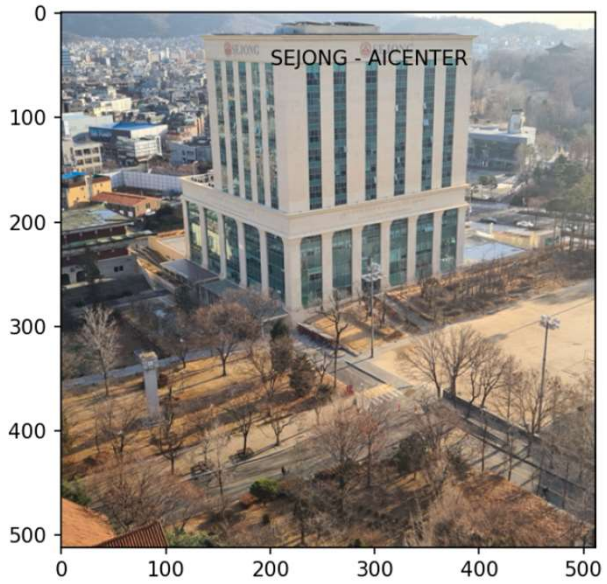
```
import matplotlib.pyplot as plt
plt.bar(ch_h, ch_w)
plt.bar(ad_h, ad_w)
plt.xlabel("Height")
plt.ylabel("Weight")
plt.title("Health Data")
plt.xticks()
plt.yticks()
plt.legend(["child", "adult"], ncol = 1, loc = "upper left")
plt.grid(axis = "both")
plt.show()
```



그림

```
import matplotlib.pyplot as plt
import matplotlib.image as img

bmp = img.imread("AICenter.bmp")
plt.text(200, 50, "SEJONG - AICENTER")
plt.imshow(bmp)
plt.savefig("AICenter_Text.png")
plt.show()
```




```
plt.savefig("AICenter_Text.bmp")
```

```
ValueError: Format 'bmp' is not supported (supported formats: eps, jpeg, jpg, pdf, pgf, png, ps, raw, rgba, svg, svgz, tif, tiff)
```

참고자료

- 지능기전공학부 최유경 교수님 자료, https://github.com/sejongresearch/2021.MachineLearning
- 코랩(Colab), https://colab.research.google.com/
- 파이썬(Python), https://www.python.org/doc/
- 사이킷런(sckit-learn), https://scikit-learn.org/stable/index.html
- 판다스(pandas), https://pandas.pydata.org/
- 맷플롯립(matplotlib), https://matplotlib.org/
- 씨본(seaborn), https://seaborn.pydata.org/
- 캐글(Kaggle), https://www.kaggle.com/
- 넘파이(numpy), https://numpy.org/doc/stable/
- 스택오퍼플러우(stackoverflow), https://stackoverflow.com/