

데이터분석입문

Lecture 08. 인구 데이터 분석하기

동양미래대학교
인공지능소프트웨어학과
강 환수

- ❖ 01. 우리 동네 인구 구조 시각화하기
- ❖ 02. 인구 구조를 다양한 형태로 시각화하기
- ❖ 03. 우리 동네 인구 구조를 원그래프로 나타내기
- ❖ 04. 우리 동네 인구 구조를 산점도로 나타내기

01. 우리 동네 인구 구조 시각화하기

- 02. 인구 구조를 다양한 형태로 시각화하기
- 03. 우리 동네 인구 구조를 원그래프로 나타내기
- 04. 우리 동네 인구 구조를 산점도로 나타내기

❖ 우리 동네 인구 구조 시각화하기 (1/10)

- 인구 데이터 파일(age.csv)을 읽어와서 한 줄씩 출력하는 코드를 작성합니다.

```
import csv

f = open('age.csv', encoding='cp949')
data = csv.reader(f)

for row in data:
    print(row)

f.close()
```

❖ 우리 동네 인구 구조 시각화하기 (2/10)

- in 연산자를 사용하여 우리 동네 이름이 포함된 지역명 찾기

```
print('신도림' in '서울특별시 구로구 신도림동(1153051000)')  
print('1153' in '서울특별시 구로구 신도림동(1153051000)')  
print('() ' in '서울특별시 구로구 신도림동(1153051000)')
```

True
True
False

- "A in B"는 A가 B 안에 존재하면 참(True), 존재하지 않으면 거짓(False)이 됩니다.

❖ 우리 동네 인구 구조 시각화하기 (3/10)

- in 연산자를 활용하여 우리 동네 인구수 데이터 출력하기

```
import csv

f = open('age.csv', encoding='cp949')
data = csv.reader(f)

for row in data:
    if '신도림' in row[0]:
        print(row)

f.close()
```

```
['서울특별시 구로구 신도림동(1153051000)', '35,964', '35,964', '274', '270', '319', '304', '321', '377', '426', '403', '374', '391', '408', '384', '370', '400', '354', '322', '332', '311', '328', '353', '359', '422', '355', '370', '429', '451', '456', '441', '454', '486', '501', '472', '561', '619', '636', '612', '619', '664', '701', '687', '760', '691', '712', '637', '628', '602', '626', '636', '663', '600', '616', '579', '567', '511', '462', '487', '509', '472', '426', '479', '477', '484', '453', '457', '412', '370', '450', '368', '340', '338', '229', '278', '249', '248', '265', '161', '141', '134', '135', '154', '121', '107', '97', '94', '71', '69', '58', '53', '35', '30', '23', '14', '16', '13', '11', '8', '8', '4', '4', '2', '4']
```

❖ 우리 동네 인구 구조 시각화하기 (4/10)

- 인구수 데이터를 저장하기에 앞서 헤더(Header)를 살펴봅니다.

```
import csv

f = open('age.csv', encoding='cp949')
data = csv.reader(f)

header = next(data)
print(header)

# for row in data:
#     if '신도림' in row[0]:
#         print(row)

f.close()
```

인구수 데이터는
row[3]부터 저장되어 있겠군요.

'행정구역', '2021년11월_계_총인구수', '2021년11월_계_연령구간인구수', '2021년11월_계_0세', '2021년11월_계_1세', '2021년11월_계_2세', '2021년11월_계_3세', '2021년11월_계_4세', '2021년11월_계_5세', '2021년11월_계_6세', '2021년11월_계_7세', '2021년11월_계_8세', '2021년11월_계_9세', '2021년11월_계_10세', '2021년11월_계_11세', '2021년11월_계_12세', '2021년11월_계_13세', '2021년11월_계_14세', '2021년11월_계_15세', '2021년11월_계_16세', '2021년11월_계_17세', '2021년11월_계_18세', '2021년11월_계_19세', '2021년11월_계_20세', '2021년11월_계_21세', '2021년11월_계_22세', '2021년11월_계_23세', '2021년11월_계_24세', '2021년11월_계_25세', '2021년11월_계_26세', '2021년11월_계_27세', '2021년11월_계_28세', '2021년11월_계_29세', '2021년11월_계_30세', '2021년11월_계_31세', '2021년11월_계_32세', '2021년11월_계_33세', '2021년11월_계_34세', '2021년11월_계_35세', '2021년11월_계_36세', '2021년11월_계_37세', '2021년11월_계_38세', '2021년11월_계_39세', '2021년11월_계_40세', '2021년11월_계_41세', '2021년11월_계_42세', '2021년11월_계_43세', '2021년11월_계_44세', '2021년11월_계_45세', '2021년11월_계_46세', '2021년11월_계_47세', '2021년11월_계_48세', '2021년11월_계_49세', '2021년11월_계_50세', '2021년11월_계_51세', '2021년11월_계_52세', '2021년11월_계_53세', '2021년11월_계_54세', '2021년11월_계_55세', '2021년11월_계_56세', '2021년11월_계_57세', '2021년11월_계_58세', '2021년11월_계_59세', '2021년11월_계_60세', '2021년11월_계_61세', '2021년11월_계_62세', '2021년11월_계_63세', '2021년11월_계_64세', '2021년11월_계_65세', '2021년11월_계_66세', '2021년11월_계_67세', '2021년11월_계_68세', '2021년11월_계_69세', '2021년11월_계_70세', '2021년11월_계_71세', '2021년11월_계_72세', '2021년11월_계_73세', '2021년11월_계_74세', '2021년11월_계_75세', '2021년11월_계_76세', '2021년11월_계_77세', '2021년11월_계_78세', '2021년11월_계_79세', '2021년11월_계_80세', '2021년11월_계_81세', '2021년11월_계_82세', '2021년11월_계_83세', '2021년11월_계_84세', '2021년11월_계_85세', '2021년11월_계_86세', '2021년11월_계_87세', '2021년11월_계_88세', '2021년11월_계_89세', '2021년11월_계_90세', '2021년11월_계_91세', '2021년11월_계_92세', '2021년11월_계_93세', '2021년11월_계_94세', '2021년11월_계_95세', '2021년11월_계_96세', '2021년11월_계_97세', '2021년11월_계_98세', '2021년11월_계_99세', '2021년11월_계_100세'

❖ 우리 동네 인구 구조 시각화하기 (5/10)

- 우리 동네의 연령별 인구수 출력하기

```
import csv

f = open('age.csv', encoding='cp949')
data = csv.reader(f)

header = next(data)

for row in data:
    if '신도림' in row[0]:
        for i in row[3:]:
            print(i)

f.close()
```


❖ 우리 동네 인구 구조 시각화하기 (6/10)

- 우리 동네의 연령별 인구수 데이터를 리스트에 저장하기

```
import csv

f = open('age.csv', encoding='cp949')
data = csv.reader(f)

header = next(data)

result = []
for row in data:
    if '신도림' in row[0]:
        for i in row[3:]:
            result.append(i)

f.close()
print(result)
```

문자열(String) 자료형이군요.
시각화를 위해서는 정수(Integer) 자료형으로
변환할 필요가 있습니다.

```
['274', '270', '319', '304', '321', '377', '426', '403', '374', '391', '408', '384', '370', '400', '354', '322', '332', '311', '328', '353', '359', '422', '355', '370', '429', '451', '456', '441', '454', '486', '501', '472', '561', '619', '636', '612', '619', '664', '701', '687', '760', '691', '712', '637', '628', '602', '626', '636', '663', '600', '616', '579', '567', '511', '462', '487', '509', '472', '426', '479', '477', '484', '453', '457', '412', '370', '450', '368', '340', '338', '229', '278', '249', '248', '265', '161', '141', '134', '135', '154', '121', '107', '97', '94', '71', '69', '58', '53', '35', '30', '23', '14', '16', '13', '11', '8', '8', '4', '4', '2', '4']
```

❖ 우리 동네 인구 구조 시각화하기 (7/10)

- 문자열 자료형을 정수 자료형으로 변환하기

```
import csv

f = open('age.csv', encoding='cp949')
data = csv.reader(f)

header = next(data)

result = []
for row in data:
    if '신도림' in row[0]:
        for i in row[3:]:
            result.append(int(i))

f.close()
print(result)
```

작은 따옴표 (")가 사라졌습니다.

[274, 270, 319, 304, 321, 377, 426, 403, 374, 391, 408, 384, 370, 400, 354, 322, 332, 311, 328, 353, 359, 422, 355, 370, 429, 451, 456, 441, 454, 486, 501, 472, 561, 619, 636, 612, 619, 664, 701, 687, 760, 691, 712, 637, 628, 602, 626, 636, 663, 600, 616, 579, 567, 511, 462, 487, 509, 472, 426, 479, 477, 484, 453, 457, 412, 370, 450, 368, 340, 338, 229, 278, 249, 248, 265, 161, 141, 134, 135, 154, 121, 107, 97, 94, 71, 69, 58, 53, 35, 30, 23, 14, 16, 13, 11, 8, 8, 4, 4, 2, 4]

❖ 우리 동네 인구 구조 시각화하기 (8/10)

- 우리 동네의 연령별 인구수 데이터를 시각화하기

```
import csv
import matplotlib.pyplot as plt

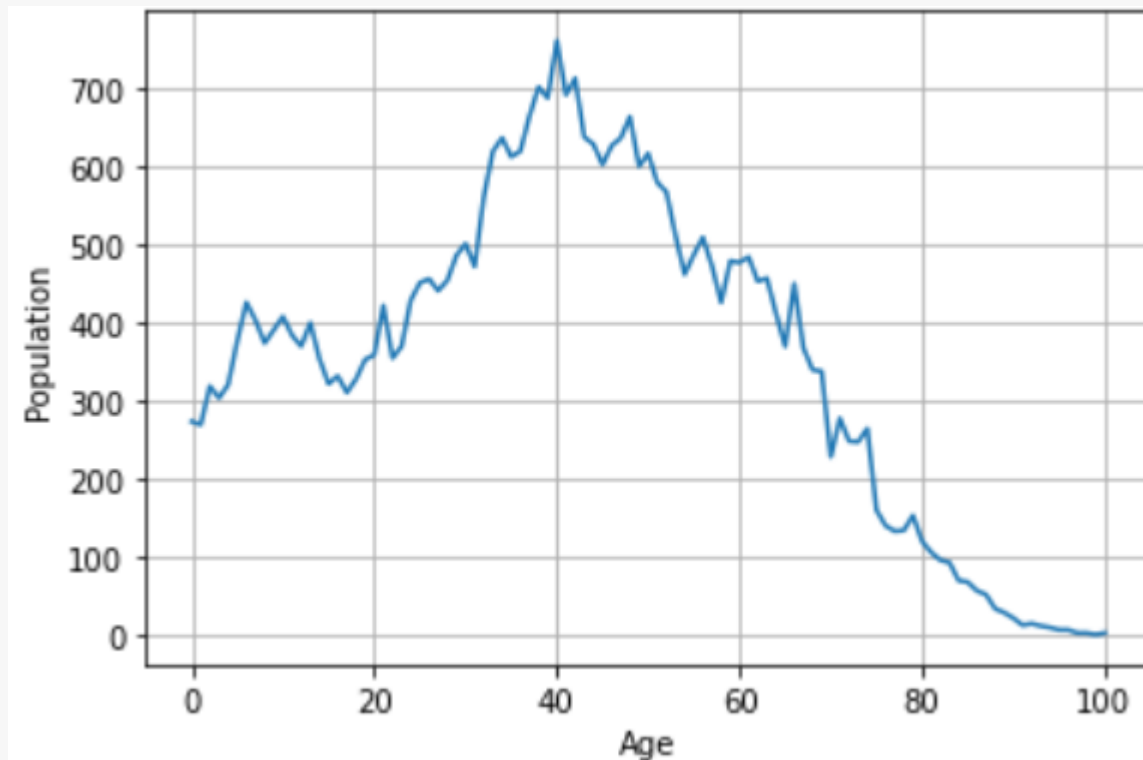
f = open('age.csv')
data = csv.reader(f)

header = next(data)

result = []
for row in data:
    if '신도림' in row[0]:
        for i in row[3:]:
            result.append(int(i))

f.close()

plt.plot(result)
plt.xlabel('Age')
plt.ylabel('Population')
plt.grid(True)
plt.show()
```



❖ 우리 동네 인구 구조 시각화하기 (9/10)

- 지역명을 입력 받아서 연령별 인구수 데이터를 시각화하기

```
import csv
import matplotlib.pyplot as plt

f = open('age_0922.csv')
data = csv.reader(f)

header = next(data)

name = input('인구 구조가 알고 싶은 지역의(읍면동 단위)을 입력해 주세요: ')

result = []
for row in data:
    if name in row[0]:
        for i in row[3:]:
            result.append(int(i))

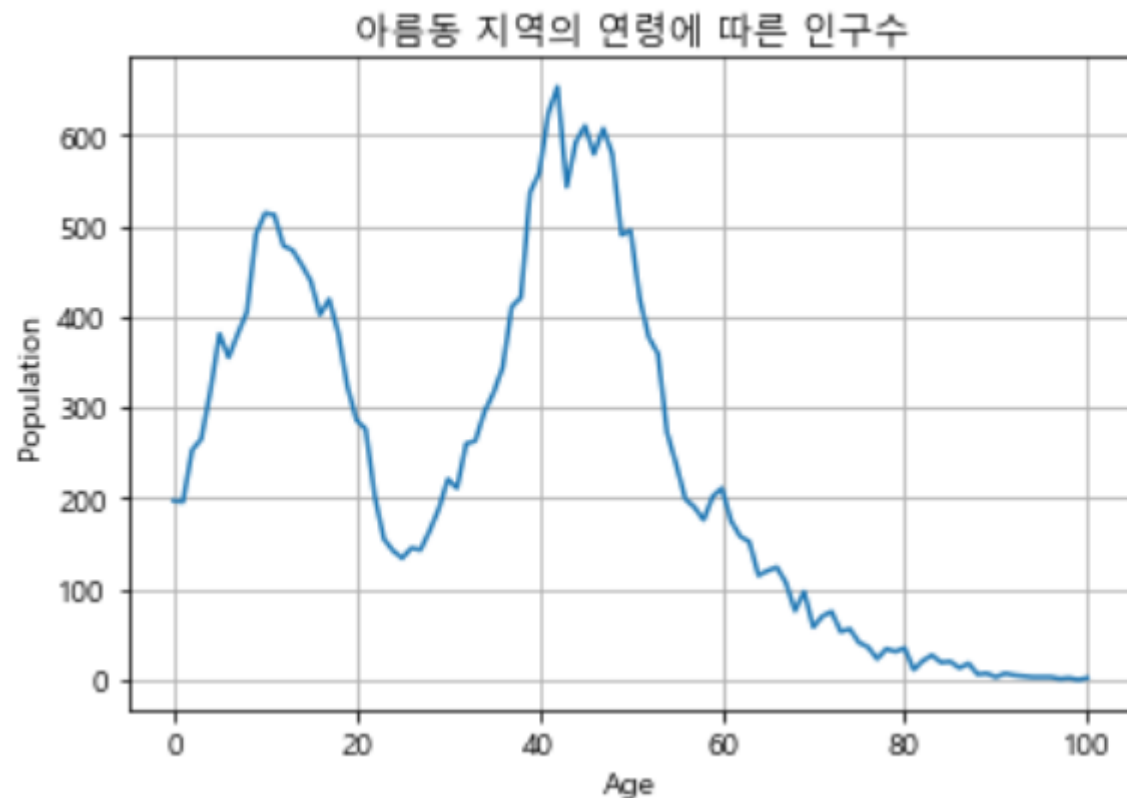
f.close()

plt.rc('font', family='Malgun Gothic')
plt.plot(result)
plt.title(name + ' 지역의 연령에 따른 인구수')
plt.xlabel('Age')
plt.ylabel('Population')
plt.grid(True)
plt.show()
```

❖ 우리 동네 인구 구조 시각화하기 (10/10)

- 지역명을 입력 받아서 연령별 인구수 데이터를 시각화하기

인구 구조가 알고 싶은 지역의 이름(읍면동 단위)을 입력해 주세요: 아름동



02. 인구 구조를 다양한 형태로 시각화하기

- 01. 우리 동네 인구 구조 시각화하기
- 03. 우리 동네 인구 구조를 원그래프로 나타내기
- 04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ① 막대그래프 그리기 (1/2)

- 막대그래프를 활용하여 우리 동네의 연령별 인구수 데이터 시각화하기

```
import matplotlib.pyplot as plt
import csv

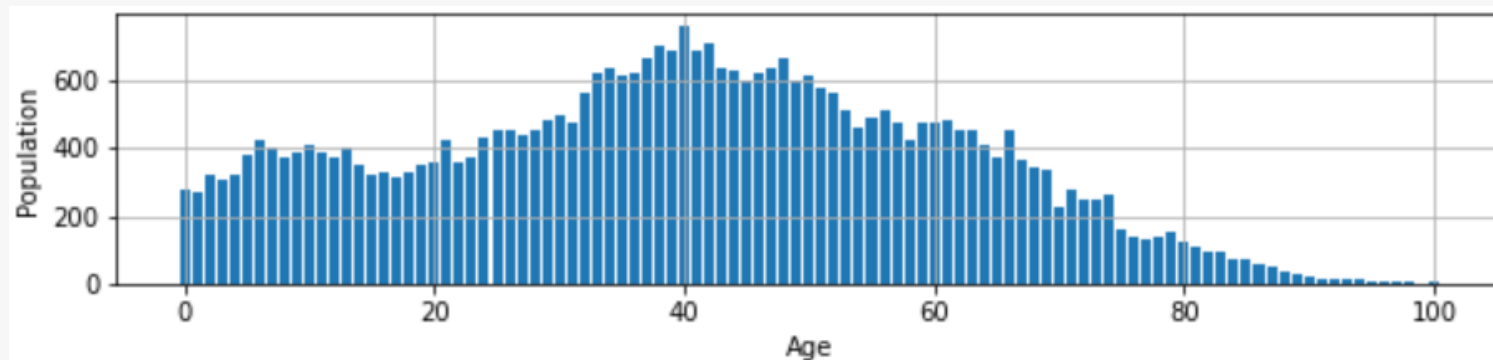
f = open('age.csv', encoding='cp949')
data = csv.reader(f)

header = next(data)
result = []

for row in data:
    if '신도림' in row[0]:
        for i in row[3:]:
            result.append(int(i))

f.close()
plt.figure(figsize=(10, 2))
plt.bar(range(len(result)), result)
plt.xlabel('Age')
plt.ylabel('Population')
plt.grid(True)
plt.show()
```

막대그래프를 수직이 아닌 수평 방향으로
그릴 수는 없을까요?



❖ ① 막대그래프 그리기 (2/2)

- 수평 막대그래프를 활용하여 우리 동네의 연령별 인구수 데이터 시각화하기
- `barh()` 함수

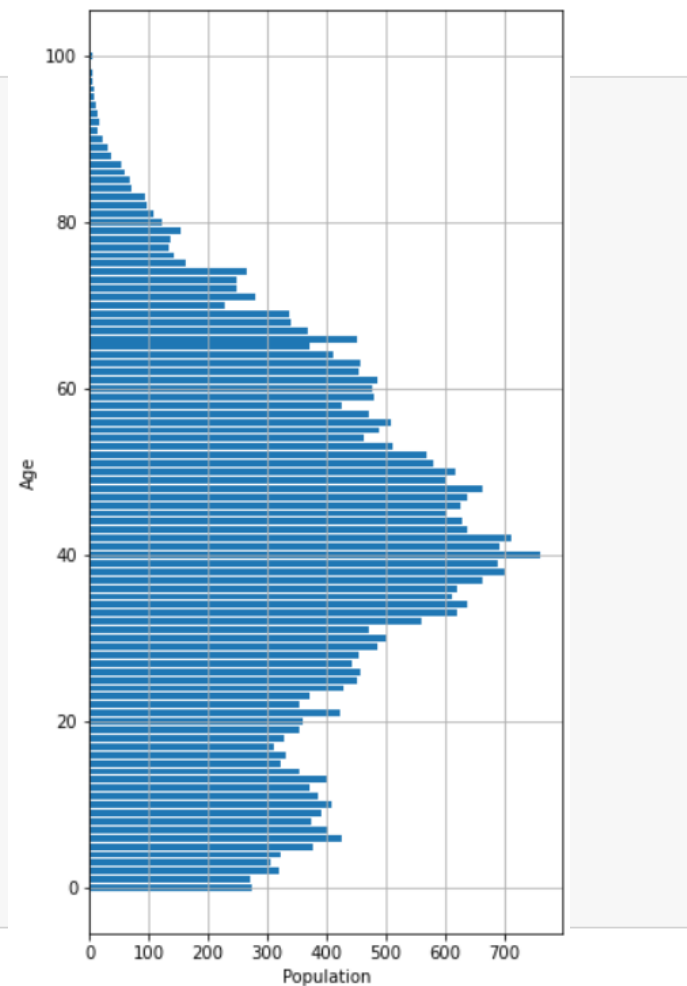
```
import matplotlib.pyplot as plt
import csv

f = open('age.csv', encoding='cp949')
data = csv.reader(f)

header = next(data)
result = []

for row in data:
    if '신도림' in row[0]:
        for i in row[3:]:
            result.append(int(i))

f.close()
plt.figure(figsize=(5, 10))
plt.barh(range(len(result)), result)
plt.xlabel('Population')
plt.ylabel('Age')
plt.grid(True)
plt.show()
```



❖ ② 항아리 모양 그래프 그리기 (1/5)

- 수평 막대그래프를 활용하여 성별 인구수 데이터 시각화하기

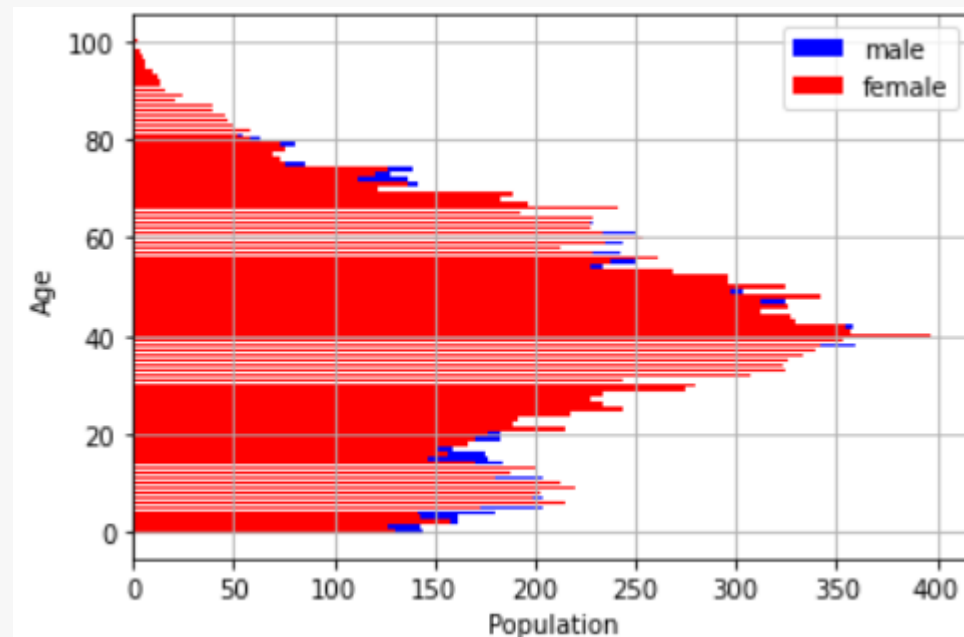
```
import matplotlib.pyplot as plt
import csv

f = open('gender.csv', encoding='cp949')
data = csv.reader(f)

header = next(data)
male = []
female = []

for row in data:
    if '신도림' in row[0]:
        for i in row[3:104]:
            male.append(int(i))
        for i in row[106:207]: # row[106:]이라고 적어도 됩니다.
            female.append(int(i))

f.close()
plt.barh(range(len(male)), male, color='b', label='male')
plt.barh(range(len(female)), female, color='r', label='female')
plt.xlabel('Population')
plt.ylabel('Age')
plt.grid(True)
plt.legend()
plt.show()
```



그래프가 겹쳐서
알아보기가 어렵습니다.

❖ ② 항아리 모양 그래프 그리기 (2/5)

- 수평 막대그래프를 활용하여 성별 인구수 데이터 시각화하기

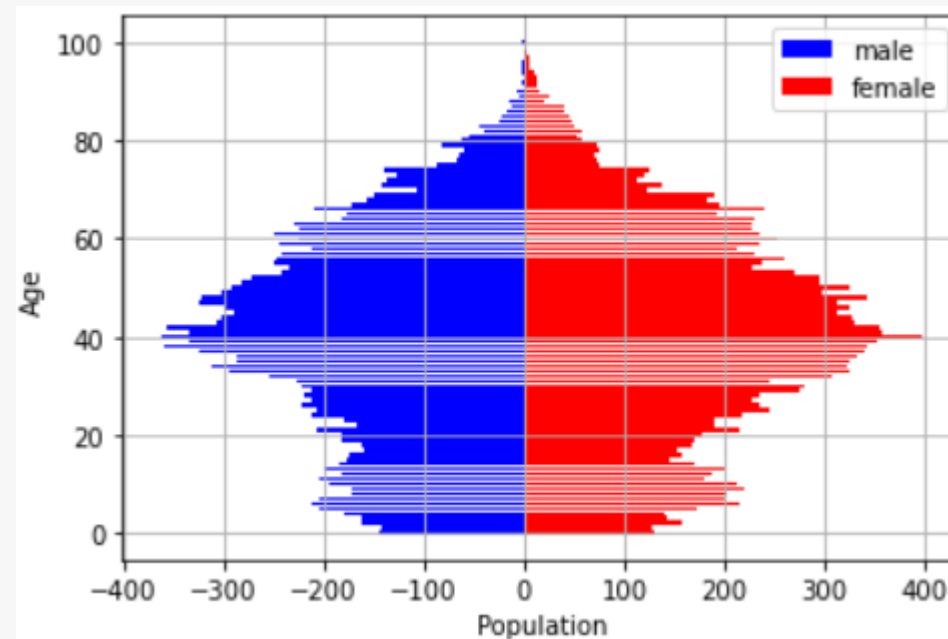
```
import matplotlib.pyplot as plt
import csv

f = open('gender.csv', encoding='cp949')
data = csv.reader(f)

header = next(data)
male = []
female = []

for row in data:
    if '신도림' in row[0]:
        for i in row[3:104]:
            male.append(-int(i))
        for i in row[106:207]: # row[106:]이라고 적어도 됩니다.
            female.append(int(i))

f.close()
plt.barh(range(len(male)), male, color='b', label='male')
plt.barh(range(len(female)), female, color='r', label='female')
plt.xlabel('Population')
plt.ylabel('Age')
plt.grid(True)
plt.legend()
plt.show()
```



남성 데이터를
음수로 변경하였습니다.

❖ ② 항아리 모양 그래프 그리기 (3/5)

- 한글 폰트 사용하기

- ◆ Malgun Gothic은 "맑은 고딕"입니다.

- ◆ 만약 macOS 운영체제를 사용하고 있다면 "AppleGothic"이라고 쓰세요.

```
plt.rc('font', family='Malgun Gothic')  
plt.title('한글 제목')
```

- 한글 폰트 사용시 마이너스 부호 표현하기

```
plt.rcParams['axes.unicode_minus'] = False
```

❖ ② 항아리 모양 그래프 그리기 (4/5)

- 찾고 싶은 지역의 이름을 입력 받아 성별 인구수 데이터 시각화하기

```
import matplotlib.pyplot as plt
import csv

f = open('gender.csv', encoding='cp949')
data = csv.reader(f)

header = next(data)
male = []
female = []

name = input('찾고 싶은 지역의 이름을 알려주세요: ')

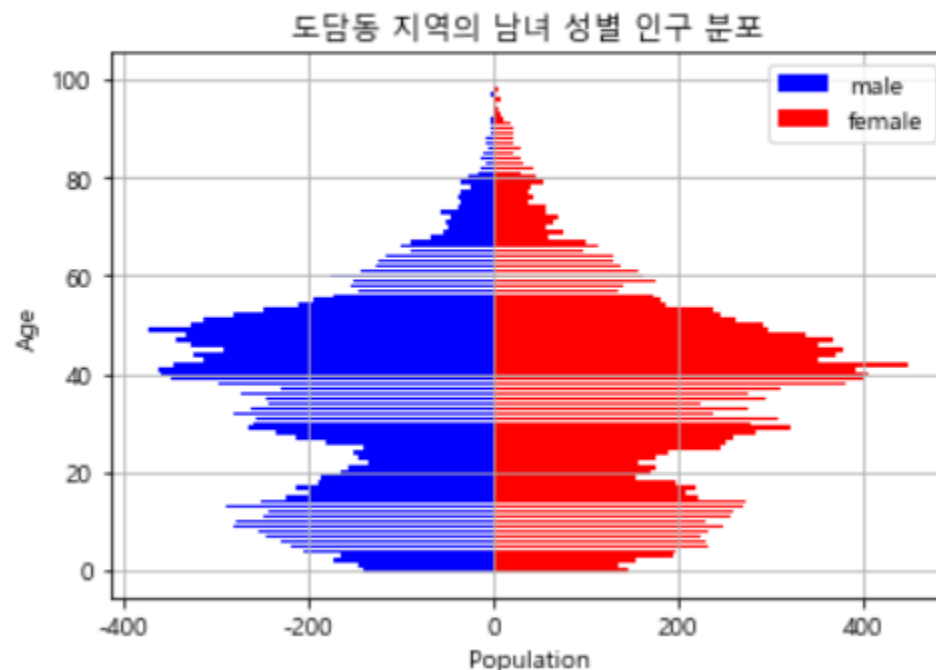
for row in data:
    if name in row[0]:
        for i in row[3:104]:
            male.append(-int(i))
        for i in row[106:207]: # row[106:]이라고 적어도 됩니다.
            female.append(int(i))

f.close()
plt.rc('font', family='Malgun Gothic')
plt.rcParams['axes.unicode_minus'] = False
plt.title(name + ' 지역의 남녀 성별 인구 분포')
plt.barh(range(len(male)), male, color='b', label='male')
plt.barh(range(len(female)), female, color='r', label='female')
plt.xlabel('Population')
plt.ylabel('Age')
plt.grid(True)
plt.legend()
plt.show()
```

❖ ② 항아리 모양 그래프 그리기 (5/5)

- 찾고 싶은 지역의 이름을 입력 받아 성별 인구수 데이터 시각화하기(실행결과)

찾고 싶은 지역의 이름을 알려주세요: 도담동



03. 우리 동네 인구 구조를 원그래프로 나타내기

- 01. 우리 동네 인구 구조 시각화하기
- 02. 인구 구조를 다양한 형태로 시각화하기
- 04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ① 제주도에는 여성의 비율이 더 높을까? (1/8)

예로부터 제주도를 돌, 바람, 여성이 많다 하여 삼다도(三多島)라고 불렀다고 합니다.

정말로 그럴까요?

정말인지 확인해보기 위해서 제주도의 성별 인구 비율을 비교해보겠습니다.

❖ ① 제주도에는 여성의 비율이 더 높을까? (2/8)

- 제주특별자치도의 성별 분포를 항아리 모양 그래프로 표현하기

```
import matplotlib.pyplot as plt
import csv

f = open('gender.csv', encoding='cp949')
data = csv.reader(f)

header = next(data)
male = []
female = []

name = input('찾고 싶은 지역의 이름을 알려주세요: ')

for row in data:
    if name in row[0]:
        for i in row[3:104]:
            male.append(-int(i))
        for i in row[106:207]: # row[106:]이라고 적어도 됩니다.
            female.append(int(i))

f.close()
plt.rc('font', family='Malgun Gothic')
plt.rcParams['axes.unicode_minus'] = False
plt.title(name + ' 지역의 남녀 성별 인구 분포')
plt.barh(range(len(male)), male, color='b', label='male')
plt.barh(range(len(female)), female, color='r', label='female')
plt.xlabel('Population')
plt.ylabel('Age')
plt.grid(True)
plt.legend()
plt.show()
```


03. 우리 동네 인구 구조를 원그래프로 나타내기

❖ ① 제주도에는 여성의 비율이 더 높을까? (3/8)

- 제주특별자치도의 성별 분포를 항아리 모양 그래프로 표현하기(실행결과 및 원인분석)

찾고 싶은 지역의 이름을 알려주세요: 제주특별자치도

```
ValueError                                Traceback (most recent call last)
~\AppData\Local\Temp\ipykernel_15000\3042654285.py in <module>
     14     if name in row[0]:
     15         for i in row[3:104]:
----> 16             male.append(-int(i))
     17         for i in row[106:207]: # row[106:]이라고 적어도 됩니다.
     18             female.append(int(i))
```

ValueError: invalid literal for int() with base 10: '1,934'

ValueError가 발생하였습니다.
문자열 '1,934'에서 콤마 (,) 때문에
문제가 발생한 것으로 예상됩니다.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
3818	제주특별자치도 (5000000000)	339,256	339,256	1,934	2,134	2,352	2,640	2,916	3,250	3,561	3,523	3,628	4,043	3,861
3819	제주특별자치도 제주시 (5011000000)	246,637	246,637	1,482	1,597	1,769	2,039	2,181	2,459	2,661	2,672	2,739	2,994	2,827
3820	제주특별자치도 제주시 한림읍(5011025000)	10,914	10,914	39	56	42	46	65	83	107	98	99	106	95
3821	제주특별자치도 제주시 애월읍(5011025300)	19,234	19,234	87	117	118	139	193	197	215	220	248	272	239
3822	제주특별자치도 제주시 구좌읍(5011025600)	7,904	7,904	21	33	37	29	44	45	41	63	56	79	77
3823	제주특별자치도 제주시 조천읍(5011025900)	13,048	13,048	69	73	73	104	96	124	152	131	142	172	155
3824	제주특별자치도 제주시 한경면(5011031000)	4,664	4,664	19	15	19	26	19	31	36	39	43	34	36
3825	제주특별자치도 제주시 추자면(5011032000)	860	860	2	2	4	2	1	2	1	4	6	4	7
3826	제주특별자치도 제주시 우도면(5011033000)	887	887	6	3	8	6	2	5	2	1	4	8	6
3827	제주특별자치도 제주시 일도1동(5011051000)	1,284	1,284	2	1	5	2	6	8	1	6	7	6	4
3828	제주특별자치도 제주시 일도2동(5011052000)	15,980	15,980	95	85	98	113	136	148	159	172	159	176	206
3829	제주특별자치도 제주시 이도1동(5011053000)	3,755	3,755	22	15	14	21	25	18	39	31	39	32	32
3830	제주특별자치도 제주시 이도2동(5011054000)	24,252	24,252	148	139	171	190	212	257	283	267	271	307	307
3831	제주특별자치도 제주시 삼도1동(5011055000)	6,626	6,626	30	35	46	42	37	64	64	57	64	81	66
3832	제주특별자치도 제주시 삼도2동(5011056000)	4,095	4,095	22	24	11	35	19	24	20	29	29	28	34
3833	제주특별자치도 제주시 용담1동(5011057000)	2,535	2,535	12	15	12	15	12	27	10	21	20	20	26

❖ ① 제주도에는 여성의 비율이 더 높을까? (4/8)

- replace() 함수를 활용하여 제주특별자치도의 성별 분포를 항아리 모양 그래프로 표현하기

```
import matplotlib.pyplot as plt
import csv

f = open('gender.csv', encoding='cp949')
data = csv.reader(f)

header = next(data)
male = []
female = []

name = input('찾고 싶은 지역의 이름을 알려주세요: ')

for row in data:
    if name in row[0]:
        for i in row[3:104]:
            i = i.replace(',', '')
            male.append(-int(i))
        for i in row[106:207]: # row[106:]이라고 적어도 됩니다.
            i = i.replace(',', '')
            female.append(int(i))

f.close()

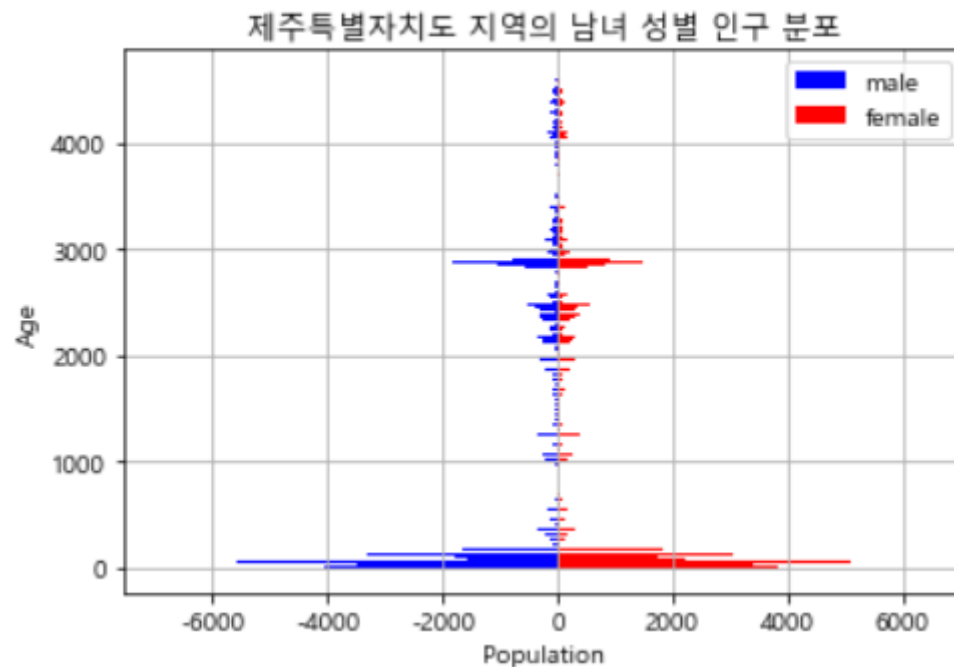
plt.rc('font', family='Malgun Gothic')
plt.rcParams['axes.unicode_minus'] = False
plt.title(name + ' 지역의 남녀 성별 인구 분포')
plt.barh(range(len(male)), male, color='b', label='male')
plt.barh(range(len(female)), female, color='r', label='female')
plt.xlabel('Population')
plt.ylabel('Age')
plt.grid(True)
plt.legend()
plt.show()
```

replace() 함수를 활용하면
문자열에서 특정 문자를 제거할 수 있습니다.
replace('특정 문자', '') → 특정 문자를 제거

❖ ① 제주도에는 여성의 비율이 더 높을까? (5/8)

- replace() 함수를 활용하여 제주특별자치도의 성별 분포를 향아리 모양 그래프로 표현하기(실행결과)

찾고 싶은 지역의 이름을 알려주세요: 제주특별자치도



ValueError는 발생하지 않았지만,
y축 Age 부분의 값이 엄청 많네요.

03. 우리 동네 인구 구조를 원그래프로 나타내기

❖ ① 제주도에는 여성의 비율이 더 높을까? (6/8)

- replace() 함수를 활용하여 제주특별자치도의 성별 분포를 향아리 모양 그래프로 표현하기(원인분석)

gender - Excel

파열, 삽입, 페이지 레이아웃, 수식, 데이터, 검토, 보기, 개발 도구, ACROBAT, 수월할 작업을 알려 주세요.

파일, 홈, 삽입, 페이지 레이아웃, 수식, 데이터, 검토, 보기, 개발 도구, ACROBAT, 수월할 작업을 알려 주세요.

찾기 및 바꾸기

찾기(D) 바꾸기(B)

찾을 내용(N): 제주특별자치도

모두 찾기(F) 다음 찾기(N) 닫기

통합 문서 시트 이름 셀 값

gender.csv gender \$A\$3823 제주특별자치도 (5000000000)

gender.csv gender \$A\$3824 제주특별자치도 제주시 (5011000000)

gender.csv gender \$A\$3825 제주특별자치도 제주시 한림읍(5011025000)

gender.csv gender \$A\$3826 제주특별자치도 제주시 애월읍(5011025300)

gender.csv gender \$A\$3827 제주특별자치도 제주시 구좌읍(5011025600)

gender.csv gender \$A\$3828 제주특별자치도 제주시 조천읍(5011025900)

gender.csv gender \$A\$3829 제주특별자치도 제주시 한경면(5011031000)

gender.csv gender \$A\$3830 제주특별자치도 제주시 추자면(5011032000)

gender.csv gender \$A\$3831 제주특별자치도 제주시 우도면(5011033000)

gender.csv gender \$A\$3832 제주특별자치도 제주시 일도1동(5011051000)

gender.csv gender \$A\$3833 제주특별자치도 제주시 일도2동(5011052000)

gender.csv gender \$A\$3834 제주특별자치도 제주시 이도1동(5011053000)

gender.csv gender \$A\$3835 제주특별자치도 제주시 이도2동(5011054000)

gender.csv gender \$A\$3836 제주특별자치도 제주시 삼도1동(5011055000)

gender.csv gender \$A\$3837 제주특별자치도 제주시 삼도2동(5011056000)

gender.csv gender \$A\$3838 제주특별자치도 제주시 용담1동(5011057000)

gender.csv gender \$A\$3839 제주특별자치도 제주시 용담2동(5011058000)

gender.csv gender \$A\$3840 제주특별자치도 제주시 건입읍(5011059000)

gender.csv gender \$A\$3841 제주특별자치도 제주시 화북읍(5011060000)

gender.csv gender \$A\$3842 제주특별자치도 제주시 삼양읍(5011061000)

gender.csv gender \$A\$3843 제주특별자치도 제주시 봉개동(5011062000)

gender.csv gender \$A\$3844 제주특별자치도 제주시 아라동(5011063000)

gender.csv gender \$A\$3845 제주특별자치도 제주시 오라동(5011064000)

gender.csv gender \$A\$3846 제주특별자치도 제주시 연동(5011065000)

gender.csv gender \$A\$3847 제주특별자치도 제주시 노형동(5011066000)

gender.csv gender \$A\$3848 제주특별자치도 제주시 외도동(5011067000)

gender.csv gender \$A\$3849 제주특별자치도 제주시 이호동(5011068000)

gender.csv gender \$A\$3850 제주특별자치도 제주시 도두동(5011069000)

gender.csv gender \$A\$3851 제주특별자치도 서귀포시 (5013000000)

gender.csv gender \$A\$3852 제주특별자치도 서귀포시 대정읍(5013025000)

gender.csv gender \$A\$3853 제주특별자치도 서귀포시 남원읍(5013025300)

gender.csv gender \$A\$3854 제주특별자치도 서귀포시 성산읍(5013025900)

gender.csv gender \$A\$3855 제주특별자치도 서귀포시 안덕면(5013031000)

gender.csv gender \$A\$3856 제주특별자치도 서귀포시 표선면(5013032000)

gender.csv gender \$A\$3857 제주특별자치도 서귀포시 송산동(5013051000)

"제주특별자치도"라는 문자열을 포함한 행이 무려 46개나 있네요.



즉, 46개의 데이터가 하나의 리스트에 저장되어 그래프로 표현되었던 거군요!

❖ ① 제주도에는 여성의 비율이 더 높을까? (7/8)

- break 문을 활용하여 제주특별자치도의 성별 분포를 항아리 모양 그래프로 표현하기

```
import matplotlib.pyplot as plt
import csv

f = open('gender.csv', encoding='cp949')
data = csv.reader(f)

header = next(data)
male = []
female = []

name = input('찾고 싶은 지역의 이름을 알려주세요: ')

for row in data:
    if name in row[0]:
        for i in row[3:104]:
            i = i.replace(',', '')
            male.append(-int(i))
        for i in row[106:207]: # row[106:]이라고 적어도 됩니다.
            i = i.replace(',', '')
            female.append(int(i))
        break

f.close()

plt.rc('font', family='Malgun Gothic')
plt.rcParams['axes.unicode_minus'] = False
plt.title(name + ' 지역의 남녀 성별 인구 분포')
plt.barh(range(len(male)), male, color='b', label='male')
plt.barh(range(len(female)), female, color='r', label='female')
plt.xlabel('Population')
plt.ylabel('Age')
plt.grid(True)
plt.legend()
plt.show()
```

사용자로부터 입력 받은 내용이 포함된 데이터 중
처음 만나는 데이터만 male, female 리스트에
저장하도록 하겠습니다.

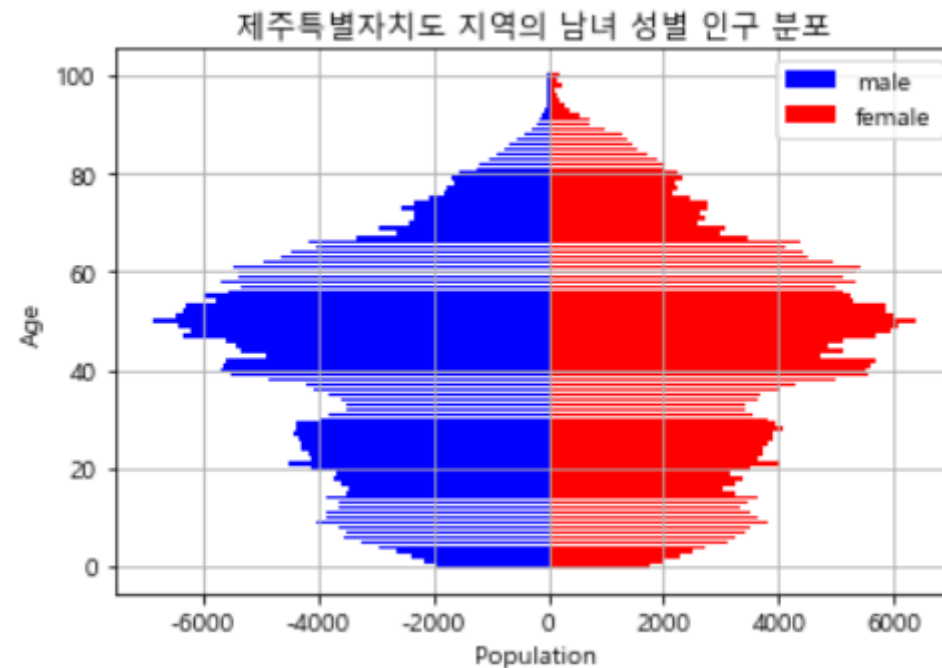


break 문을 추가하면
for 반복문을 멈출 수 있습니다.

❖ ① 제주도에는 여성의 비율이 더 높을까? (8/8)

- break 문을 활용하여 제주특별자치도의 성별 분포를 향아리 모양 그래프로 표현하기(실행결과)

찾고 싶은 지역의 이름을 알려주세요: 제주특별자치도



남성과 여성 중 어떤 성별의 비율이 높을까요?
향아리 모양의 그래프로는 알기 어려운 정보입니다.

❖ ② 제주도의 성별 인구 비율 표현하기 (1/5)

● 제주특별자치도의 성별 인구 합계 구하기

```
import csv

f = open('gender.csv', encoding = 'cp949')
data = csv.reader(f)

header = next(data)
# print(header)

name = input('궁금한 지역의 이름을 입력하세요: ')

for row in data:
    if (name in row[0]):
        male = 0
        female = 0
        for i in row[3:104]:      # 3번 방에 0세, ..., 103번 방에 100세 이상
            i = i.replace(',', '') # '1,934' --> '1934'
            male = male + int(i)
        for i in row[106:207]:    # 106번 방에 0세, ..., 206번에 100세 이상
            i = i.replace(',', '') # '1,934' --> '1934'
            female = female + int(i)
        break

f.close()

print('남성 인구 합: ', male, '명')
print('여성 인구 합: ', female, '명')
```

궁금한 지역의 이름을 입력하세요: 제주특별자치도
남성 인구 합: 339256 명
여성 인구 합: 337512 명

❖ ② 제주도의 성별 인구 비율 표현하기 (2/5)

- 제주특별자치도의 성별 인구 합계를 원그래프(Pie Chart)로 시각화하기

```
import csv
import matplotlib.pyplot as plt

f = open('gender.csv', encoding = 'cp949')
data = csv.reader(f)

header = next(data)
# print(header)

name = input('궁금한 지역의 이름을 입력하세요: ')

for row in data:
    if (name in row[0]):
        male = 0
        female = 0
        for i in row[3:104]:      # 3번 방에 0세, ..., 103번 방에 100세 이상
            i = i.replace(',', '') # '1,934' ----> '1934'
            male = male + int(i)
        for i in row[106:207]:    # 106번 방에 0세, ..., 206번에 100세 이상
            i = i.replace(',', '') # '1,934' ----> '1934'
            female = female + int(i)
        break

f.close()

labels = ['male', 'female']
colors = ['skyblue', 'pink']

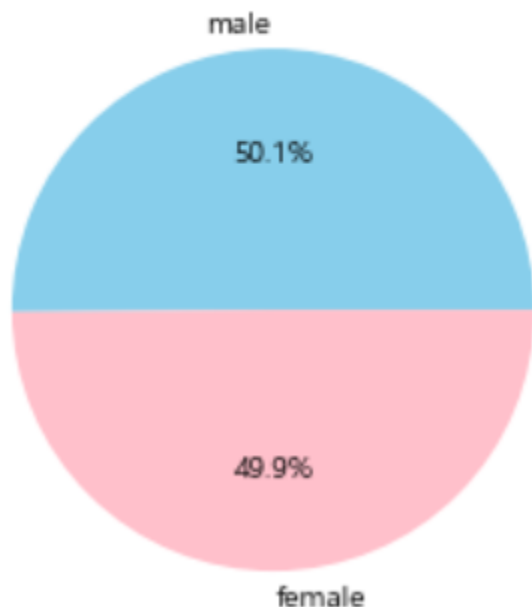
plt.rc('font', family='Malgun Gothic')
plt.title(name + ' 지역의 성별 비율')
plt.pie([male, female], labels=labels, colors=colors, autopct='%1f%%')
plt.show()
```


❖ ② 제주도의 성별 인구 비율 표현하기 (3/5)

- 제주특별자치도의 성별 인구 합계를 원그래프(Pie Chart)로 시각화하기(실행결과)

궁금한 지역의 이름을 입력하세요 : 제주특별자치도

제주특별자치도 지역의 성별 비율



원그래프의 시작 각도를 조정하려면 어떻게 해야 할까요?

❖ ② 제주도의 성별 인구 비율 표현하기 (4/5)

● 원그래프(Pie Chart)의 시작 각도 조정하기

```
import csv
import matplotlib.pyplot as plt

f = open('gender.csv', encoding = 'cp949')
data = csv.reader(f)

header = next(data)
# print(header)

name = input('궁금한 지역의 이름을 입력하세요: ')

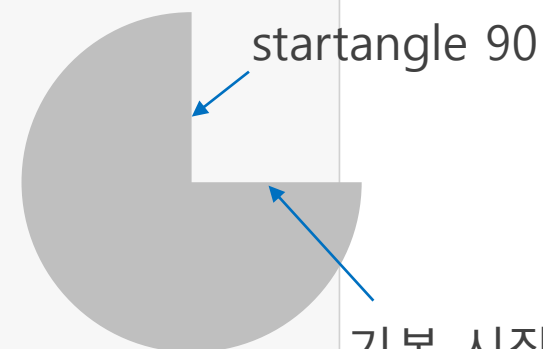
for row in data:
    if (name in row[0]):
        male = 0
        female = 0
        for i in row[3:104]:      # 3번 방에 0세, ..., 103번 방에 100세 이상
            i = i.replace(',', '') # '1,934' ----> '1934'
            male = male + int(i)
        for i in row[106:207]:    # 106번 방에 0세, ..., 206번에 100세 이상
            i = i.replace(',', '') # '1,934' ----> '1934'
            female = female + int(i)
        break

f.close()

labels = ['male', 'female']
colors = ['skyblue', 'pink']

plt.rc('font', family='Malgun Gothic')
plt.title(name + ' 지역의 성별 비율')
plt.pie([male, female], labels=labels, colors=colors, autopct='%1f%%', startangle=90)
plt.show()
```

startangle 속성에
시작 각도를 지정하여 줍니다.

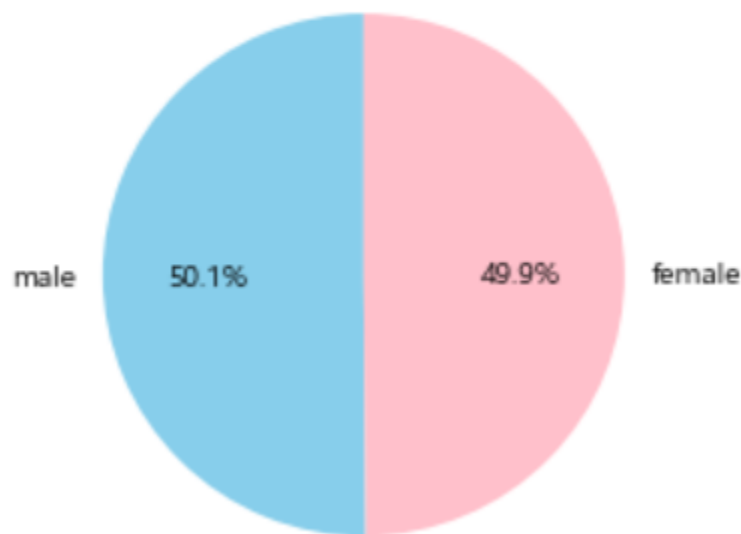


❖ ② 제주도의 성별 인구 비율 표현하기 (5/5)

- 원그래프(Pie Chart)의 시작 각도 조정하기(실행결과)

궁금한 지역의 이름을 입력하세요: 제주특별자치도

제주특별자치도 지역의 성별 비율



확인 결과 2021년 기준으로 제주도의 남녀 인구 비율은 거의 동일하네요!

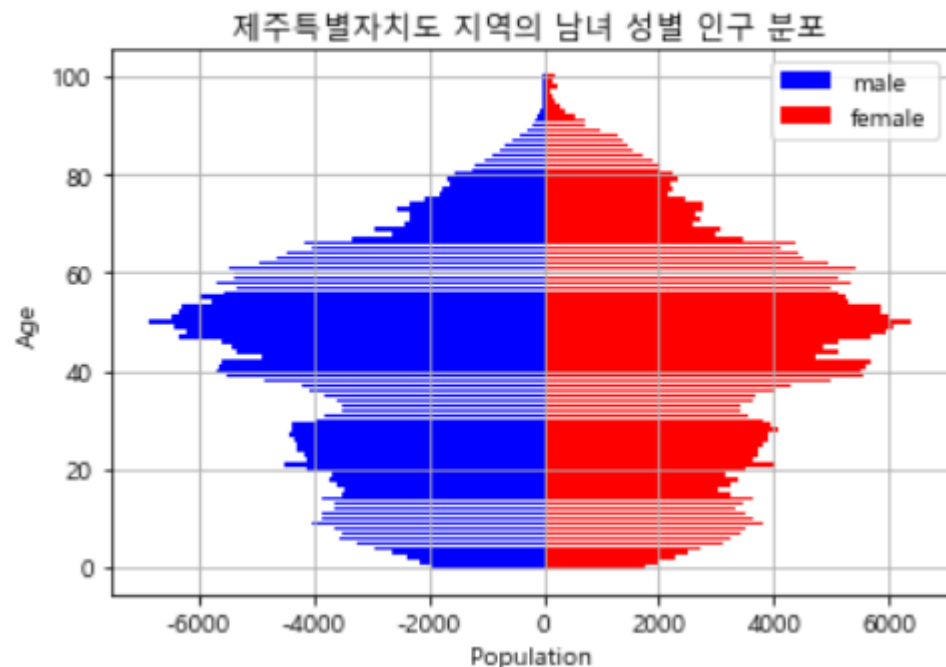
04. 우리 동네 인구 구조를 산점도로 나타내기

- 01. 우리 동네 인구 구조 시각화하기
- 02. 인구 구조를 다양한 형태로 시각화하기
- 03. 우리 동네 인구 구조를 원그래프로 나타내기

04. 우리 동네 인구 구조를 산점도로 나타내기

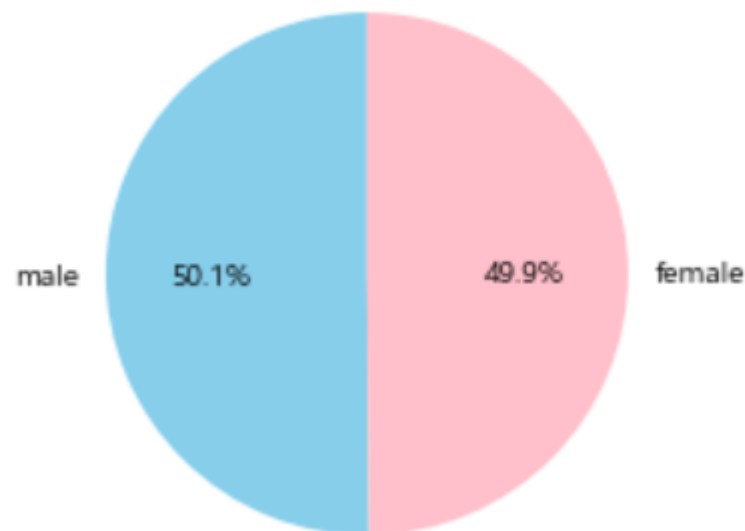
❖ 제주도의 연령대별 성별 비율은?

찾고 싶은 지역의 이름을 알려주세요: 제주특별자치도



궁금한 지역의 이름을 입력하세요: 제주특별자치도

제주특별자치도 지역의 성별 비율



어떻게 하면 연령대별 성별 비율을
지금보다 더 알아보기 쉽게 표현할 수 있을까요?

❖ ① 꺾은선 그래프로 표현하기 (1/2)

- 연령대별 성별 인구수를 꺾은선 그래프로 시각화하기

```
import csv
import matplotlib.pyplot as plt

f = open('gender.csv', encoding='cp949')
data = csv.reader(f)
next(data)

male = []
female = []

name = input('궁금한 동네를 입력해 주세요: ')
for row in data:
    if name in row[0]:
        for i in range(3, 104):
            male.append(int(row[i].replace(',', '')))
        for i in range(106, 207):
            female.append(int(row[i].replace(',', '')))
        break

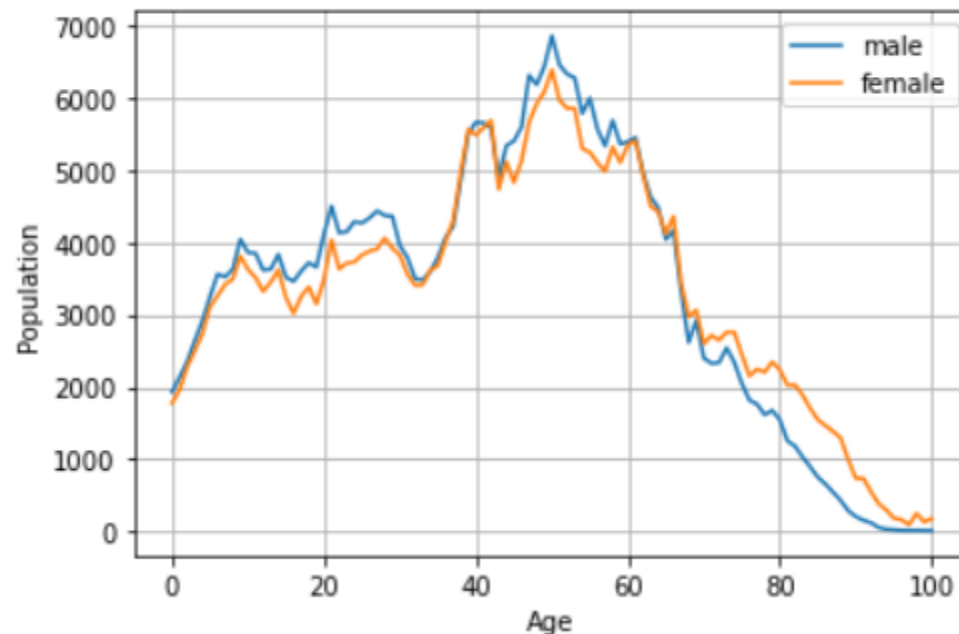
f.close()

plt.plot(male, label='male')
plt.plot(female, label='female')
plt.xlabel('Age')
plt.ylabel('Population')
plt.legend()
plt.grid(True)
plt.show()
```

❖ ① 꺾은선 그래프로 표현하기 (2/2)

- 연령대별 성별 인구수를 꺾은선 그래프로 시각화하기(실행결과)

궁금한 동네를 입력해 주세요: 제주특별자치도



연령별로 남성 인구수에서 여성 인구수를 뺀 차이를 막대그래프로 표현하면 어떨까요?

❖ ② 막대그래프로 표현하기 (1/2)

- 연령별 남성 인구수 - 여성 인구수 결과를 막대그래프로 가시화하기

```
import csv
import matplotlib.pyplot as plt

f = open('gender.csv', encoding='cp949')
data = csv.reader(f)
next(data)

diff = []

name = input('궁금한 동네를 입력해 주세요: ')
for row in data:
    if name in row[0]:
        for i in range(3, 104):
            diff.append(int(row[i].replace(',', '')) - int(row[i + 103].replace(',', '')))
        break

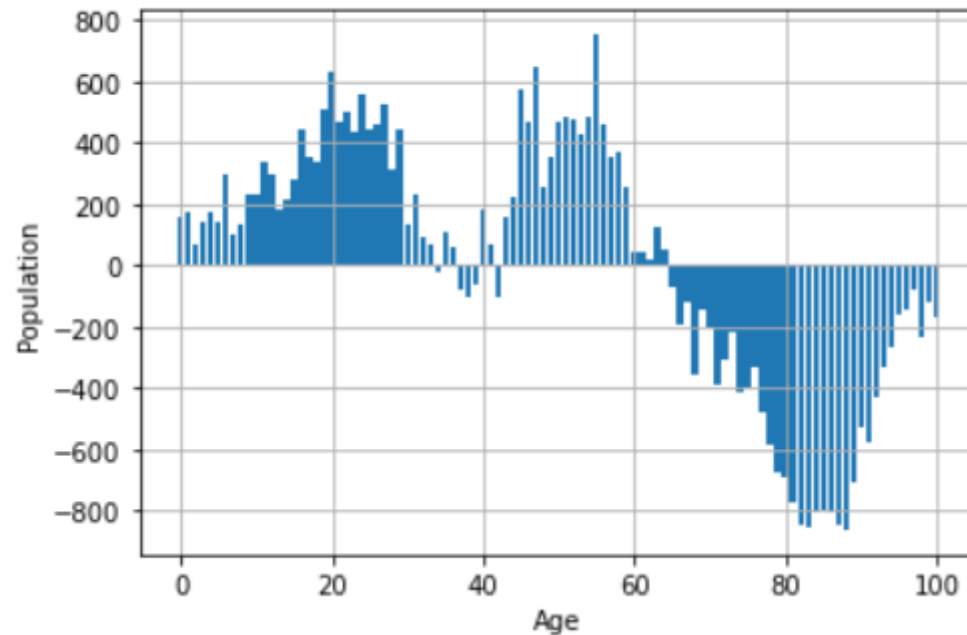
f.close()

plt.bar(range(len(diff)), diff)
plt.xlabel('Age')
plt.ylabel('Population')
plt.grid(True)
plt.show()
```


❖ ② 막대그래프로 표현하기 (2/2)

- 연령별 남성 인구수 - 여성 인구수 결과를 막대그래프로 가시화하기(실행결과)

궁금한 동네를 입력해 주세요: 제주특별자치도



같은 데이터라도 다양한 형태로 시각화하면
데이터에 숨겨진 다양한 사실을 발견할 수 있습니다.

❖ ③ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (1/5)

- 연령별 남성 및 여성 인구수를 scatter() 함수로 시각화하기

```
import csv
import matplotlib.pyplot as plt

f = open('gender.csv', encoding='cp949')
data = csv.reader(f)
next(data)

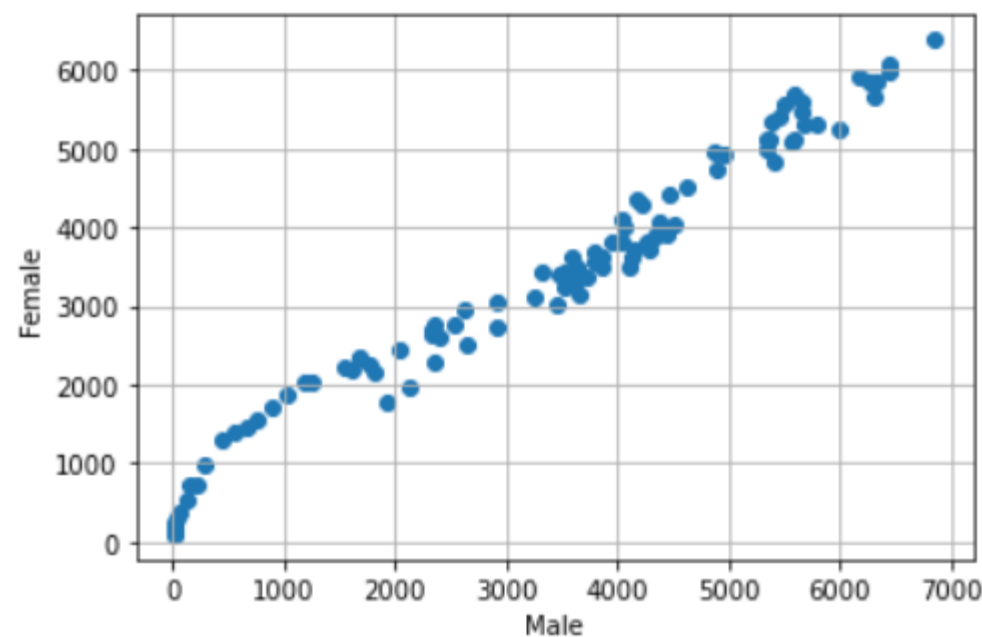
male = []
female = []

name = input('궁금한 동네를 입력해 주세요: ')
for row in data:
    if name in row[0]:
        for i in range(3, 104):
            male.append(int(row[i].replace(',', '')))
            female.append(int(row[i + 103].replace(',', '')))
        break

f.close()

plt.scatter(male, female)
plt.xlabel('Male')
plt.ylabel('Female')
plt.grid(True)
plt.show()
```

궁금한 동네를 입력해 주세요: 제주특별자치도



❖ ③ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (2/5)

- 컬러바, 추세선 추가하기

```
import csv
import matplotlib.pyplot as plt

f = open('gender.csv', encoding='cp949')
data = csv.reader(f)
next(data)

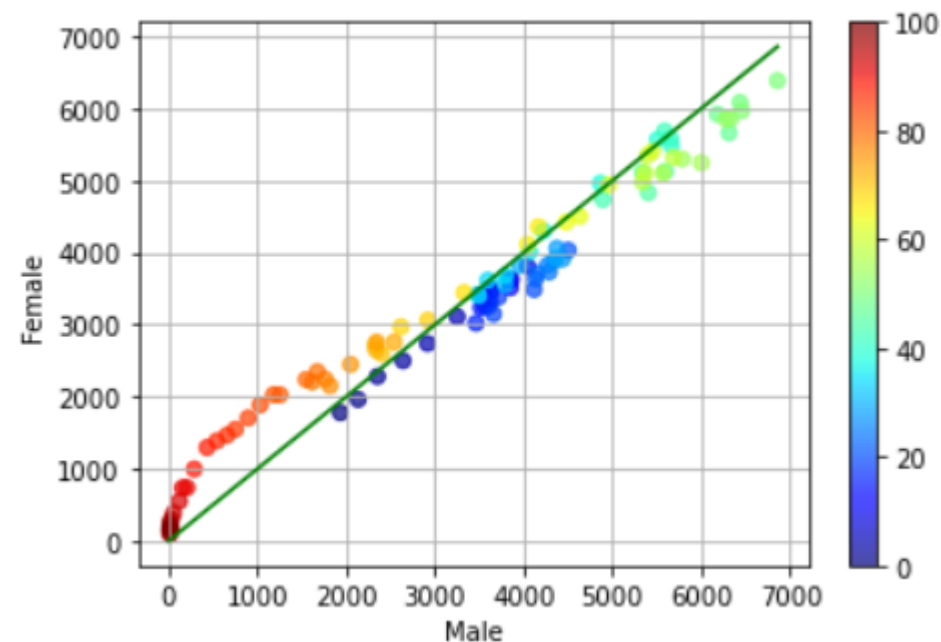
male = []
female = []

name = input('궁금한 동네를 입력해 주세요: ')
for row in data:
    if name in row[0]:
        for i in range(3, 104):
            male.append(int(row[i].replace(',', '')))
            female.append(int(row[i + 103].replace(',', '')))
        break

f.close()

plt.scatter(male, female, c=range(len(male)), cmap='jet', alpha=0.7)
plt.colorbar()
plt.plot(range(max(male)), range(max(male)), 'g')
plt.xlabel('Male')
plt.ylabel('Female')
plt.grid(True)
plt.show()
```

궁금한 동네를 입력해 주세요: 제주특별자치도



❖ ③ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (3/5)

● 버블의 크기 조정하기

```
import csv
import matplotlib.pyplot as plt
import math

f = open('gender.csv', encoding='cp949')
data = csv.reader(f)
next(data)

male = []
female = []
size = []

name = input('궁금한 동네를 입력해 주세요: ')
for row in data:
    if name in row[0]:
        for i in range(3, 104):
            male.append(int(row[i].replace(',', '')))
            female.append(int(row[i + 103].replace(',', '')))
            size.append(math.sqrt(int(row[i].replace(',', '')) + int(row[i + 103].replace(',', ''))))
        break

f.close()

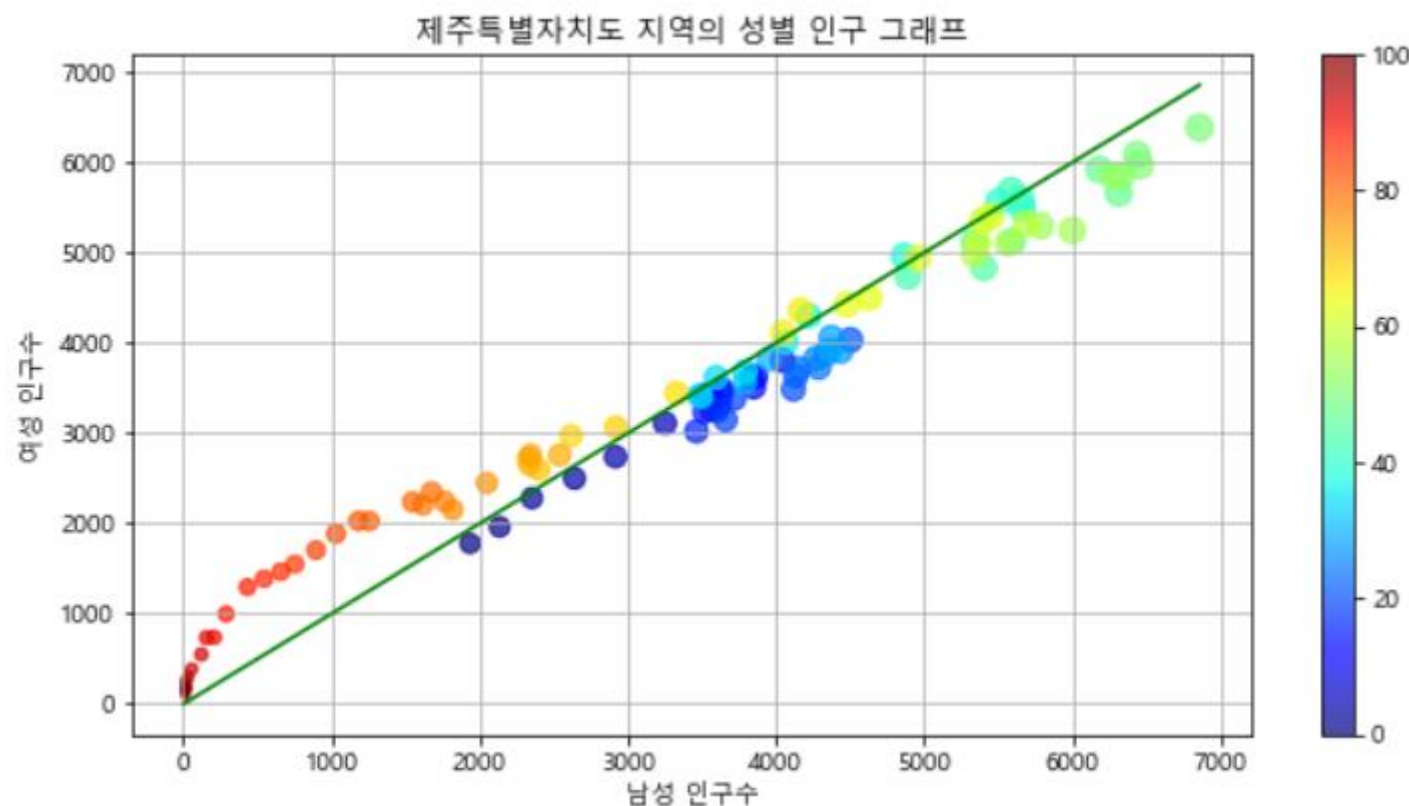
plt.rc('font', family='Malgun Gothic')
plt.figure(figsize=(10, 5))
plt.title(name + ' 지역의 성별 인구 그래프')
plt.scatter(male, female, c=range(len(male)), cmap='jet', alpha=0.7, s=size)
plt.colorbar()
plt.plot(range(max(male)), range(max(male)), 'g')
plt.xlabel('남성 인구수')
plt.ylabel('여성 인구수')
plt.grid(True)
plt.show()
```

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ③ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (4/5)

- 버블의 크기 조정하기(실행결과)

궁금한 동네를 입력해 주세요: 제주특별자치도

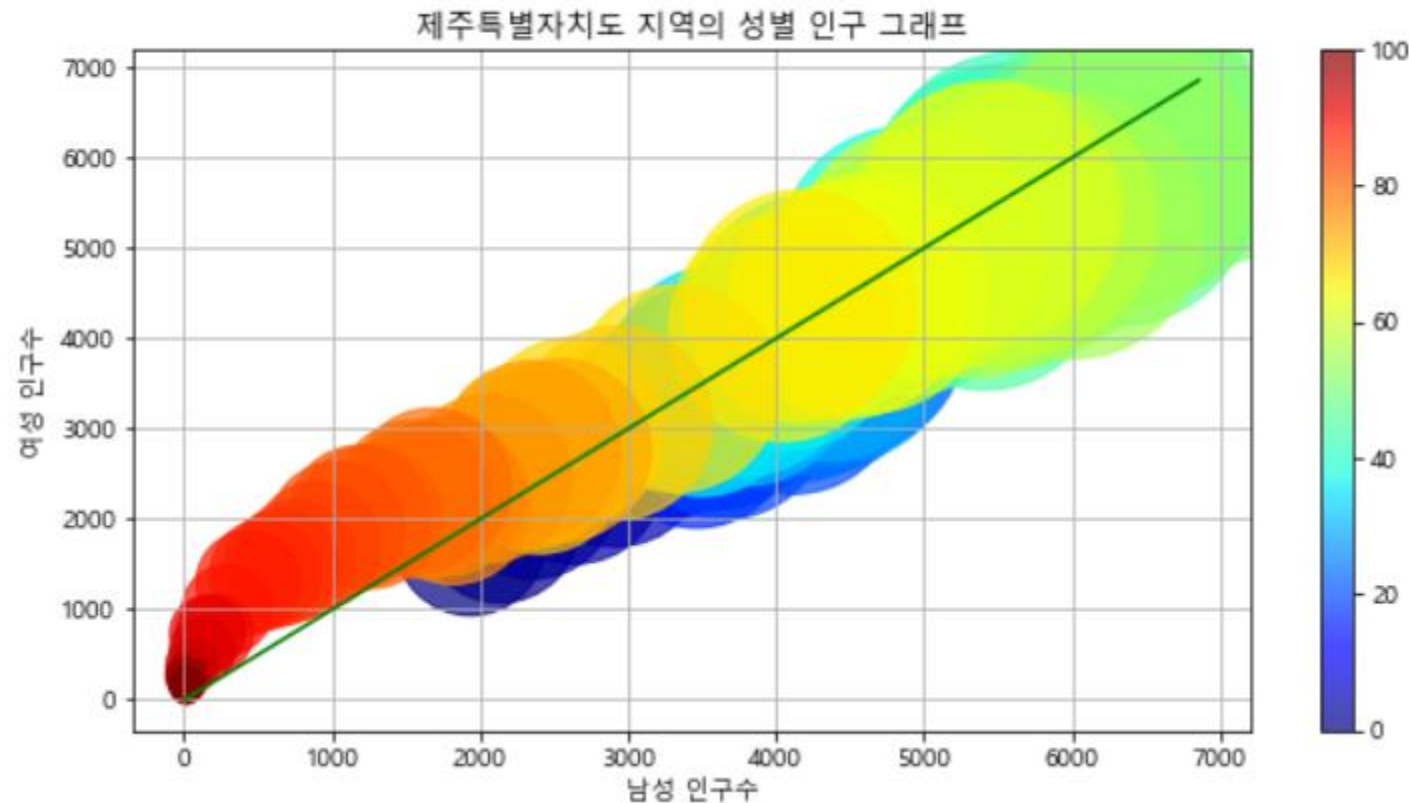


연령대 정보

❖ ③ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (5/5)

- 연령별 남성 및 여성 인구수를 scatter() 함수로 시각화하기(제공근을 사용하지 않았을 경우)

궁금한 동네를 입력해 주세요: 제주특별자치도



위의 그래프에서와 같이 s 속성값에 너무 큰 값이 지정되면 시각화 효과가 떨어집니다!

- ❖ 01. 우리 동네 인구 구조 시각화하기
- ❖ 02. 인구 구조를 다양한 형태로 시각화하기
- ❖ 03. 우리 동네 인구 구조를 원그래프로 나타내기
- ❖ 04. 우리 동네 인구 구조를 산점도로 나타내기

THANK YOU!

Q & A

- Name: 강환수
- Office: 동양미래대학교 2호관 706호 (02-2610-1941)
- E-mail: hsknag@dongyang.ac.kr
- Homepage: <https://github.com/ai7dnn/2023-DA>