



데이터 시각화 이해와 실습

Lecture 05. 인구 공공데이터

동덕여자대학교
데이터사이언스 전공
권 범

목차

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

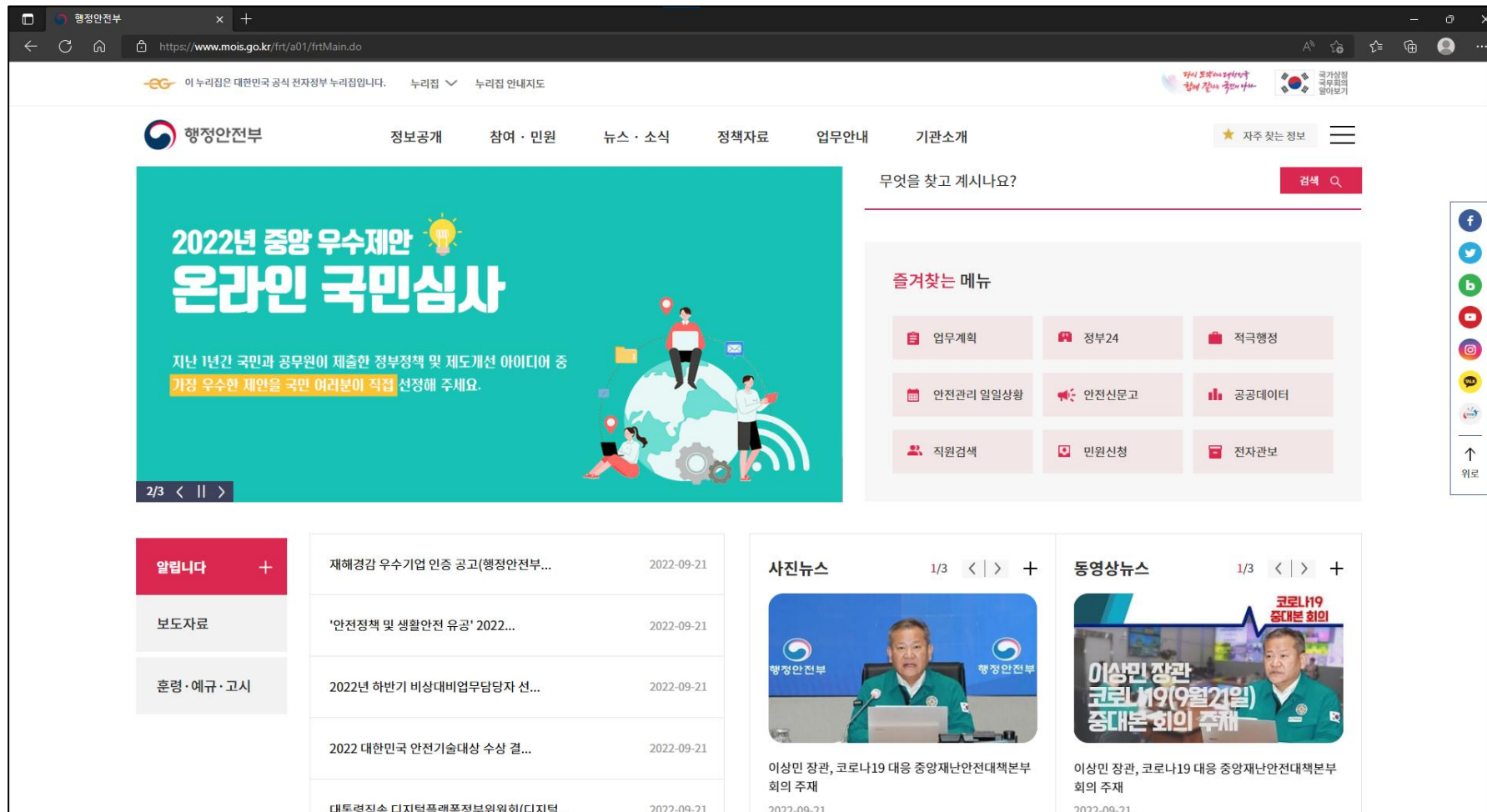
01. 우리 동네 인구 구조 시각화하기

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

01. 우리 동네 인구 구조 시각화하기

❖ ① 인구 공공데이터 내려받기 (1/8)

- 인구 데이터 수집 → 행정안전부(<https://www.mois.go.kr/>)



01. 우리 동네 인구 구조 시각화하기

❖ ① 인구 공공데이터 내려받기 (2/8)

- [정책자료] - [통계] - [주민등록 인구통계] 버튼 클릭

The screenshot shows the MOIS website interface. The top navigation bar includes '정보공개', '참여·민원', '뉴스·소식', '정책자료' (highlighted with a red box and circled '1'), '업무안내', and '기관소개'. The '정책자료' section is active, displaying a list of data categories. The '통계' (Statistics) category is highlighted with a red box and circled '2'. Within the '통계' category, the '주민등록 인구통계' (Residential Registration Population Statistics) link is highlighted with a red box and circled '3'. The bottom of the page features a '공지사항' (Notice) section with a list of recent announcements and a '사진뉴스' (Photo News) section with a video player showing a press conference.

01. 우리 동네 인구 구조 시각화하기

❖ ① 인구 공공데이터 내려받기 (3/8)

- [연령별 인구현황] 버튼 클릭

행정안전부

주민등록 인구통계

주민등록 인구 및 세대현황

통계표 | 그래프

행정구역: 전국 | 시·군·구: [선택]

등록구분: 전체

조회기간: 월간 | 연간 | 2022년 | 08월 ~ 2022년 | 08월

※ 매월 말일 작성 / 공표일시: 매월 1일 12시 이후(공표일이 주말, 공휴일인 경우에는 다음 평일에 공표)

구분: 남·여 구분 | 남·여 구성비 | 세대당인구

정렬순서: 행정기관코드 | 오름차순

검색 | 초기화

현재화면 | 전체시군구현황 | 전체읍면동현황 | csv 파일 다운로드 | xlsx 파일 다운로드

주민등록 인구 및 세대현황

행정기관	2022년 08월					
	총 인구수	세대수	세대당 인구	남자 인구수	여자 인구수	남여 비율
전국	51,558,441	23,724,829	2.17	25,699,579	25,858,862	0.99
서울특별시	9,488,454	4,472,975	2.12	4,601,295	4,887,159	0.94
부산광역시	3,331,444	1,559,462	2.14	1,627,821	1,703,623	0.96

01. 우리 동네 인구 구조 시각화하기

❖ ① 인구 공공데이터 내려받기 (4/8)

- [조회기간]: 2022년 08월 ~ 2022년 08월
- [남·여 구분] 체크 해제
- [연령 구분 단위]: 1세
- 만 연령구분: 0, 100이상
- [검색] 버튼 클릭

행정안전부

주민등록 인구통계

연령별 인구현황

통계표

그래프

행정구역: 전국, 시·군·구

등록구분: 전체

조회기간: 월간, 연간, 2022년 08월 ~ 2022년 08월

구분: ☒ 계, ☐ 남·여 구분

정렬순서: 행정기관코드, 오름차순

연령 구분 단위: 1세

만 연령구분: 0, 100이상

검색

초기화

현재화면, 전체시군구현황, 전체읍면동현황

CSV 파일 다운로드, XLSX 파일 다운로드

연령별 인구현황

행정기관

행정기관	총 인구수	연령구간인구수	0세	1세	2세	3세	4세	5세	6세

01. 우리 동네 인구 구조 시각화하기

❖ ① 인구 공공데이터 내려받기 (5/8)

- [전체읍면동현황] 체크 후, [CSV 파일 다운로드] 버튼 클릭

행정안전부

주민등록 인구통계

연령별 인구현황

통계표 | 그래프

행정구역: 전국 | 시·군·구

등록구분: 전체

조회기간: 월간 | 연간 | 2022년 | 08월 ~ 2022년 | 08월

구분: ☒ 계 ☐ 남·여 구분

정렬순서: 행정기관코드 | 오름차순

연령 구분 단위: 1세

만 연령구분: 0 | 100이상

①

○ 현재읍면 ○ 전체시군구현황 ● 전체읍면동현황

CSV 파일 다운로드

②

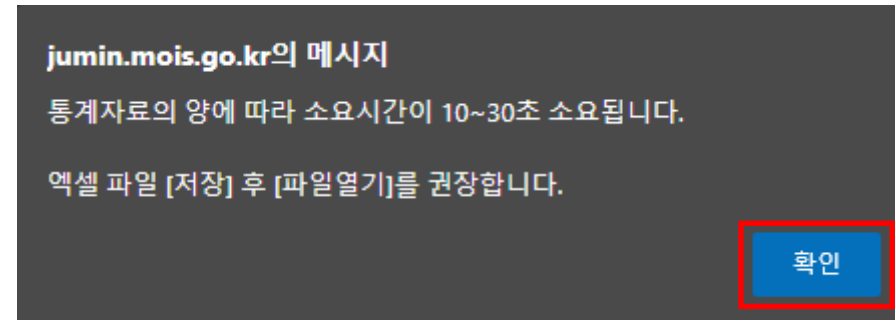
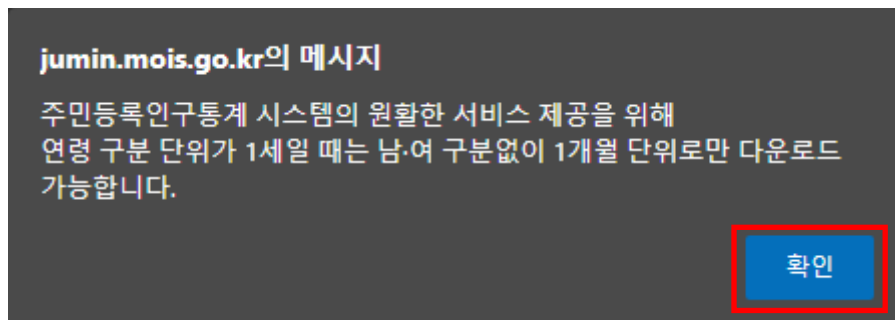
연령별 인구현황

행정기관	총 인구수	연령구간인구수	0세	1세	2세	3세	4세	5세	6세

01. 우리 동네 인구 구조 시각화하기

❖ ① 인구 공공데이터 내려받기 (6/8)

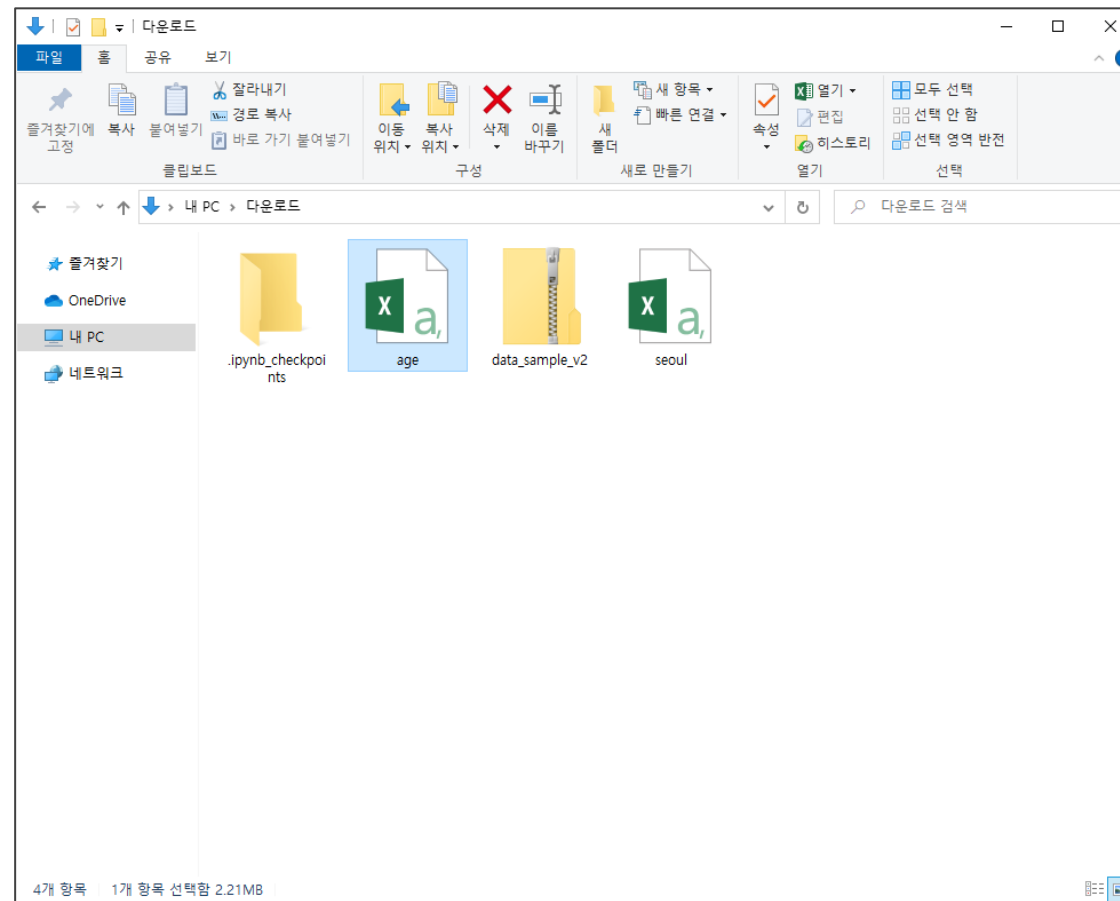
- 아래와 같이 경고 창이 뜨면, [확인] 버튼 클릭



01. 우리 동네 인구 구조 시각화하기

❖ ① 인구 공공데이터 내려받기 (7/8)

- 파일 이름을 "age.csv"로 변경함
- 파일은 다운로드(Downloads) 폴더에 저장되어 있다고 가정하고, 실습을 진행하자



01. 우리 동네 인구 구조 시각화하기

❖ ① 인구 공공데이터 내려받기 (8/8)

- age.csv 파일을 열면 전국의 읍면동, 연령별 인구를 확인할 수 있음

<

01. 우리 동네 인구 구조 시각화하기

❖ ② 인구 데이터 살펴보고 질문하기

- 어느 동네에 영유아가 가장 많을까?
- 어느 동네가 가장 고령화되었을까?
- 우리 동네에 가장 많이 살고 있는 연령은 몇 살일까?
- 내 또래 사람들이 가장 많이 사는 지역은 어디일까?

01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (1/15)

- 우리 동네의 인구 구조를 시각화 하려면, 어떤 단계(또는 절차)를 거치면 될까?

알고리즘(Algorithm) 설계하기

- ✓ Step 1) 인구 데이터 파일을 읽음
- ✓ Step 2) 전체 데이터에서 한 줄씩 반복해서 읽음
- ✓ Step 3) 우리 동네에 대한 데이터인지 확인함
- ✓ Step 4) 우리 동네일 경우 0세부터 100세 이상까지의 인구수를 순서대로 저장함
- ✓ Step 5) 저장된 연령별 인구수 데이터를 시각화함

01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (2/15)

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

```
1 import csv
2
3 f = open("age.csv", encoding="cp949")
4 data = csv.reader(f)
5
6 for row in data:
7     print(row)
8
9 f.close()
```

01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (3/15)

- in 연산자를 사용하여 우리 동네 이름이 포함된 지역명 찾기

```
1 print("신도림" in "서울특별시 구로구 신도림동(1153051000)")
2 print("1153" in "서울특별시 구로구 신도림동(1153051000)")
3 print("(" in "서울특별시 구로구 신도림동(1153051000)")
```

실행결과

```
True
True
False
```

01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (4/15)

- in 연산자를 활용하여 우리 동네 인구수 데이터 출력하기

```
1 import csv
2
3 f = open("age.csv", encoding="cp949")
4 data = csv.reader(f)
5
6 for row in data:
7     if "신도림" in row[0]:
8         print(row)
9
10 f.close()
```

실행결과

```
[ '서울특별시 구로구 신도림동(1153051000)', '35,945', '35,945', '253', '270',
  '269', '312', '291', '313', '390', '412', '386', '369', '396',
  ... (중략) ...
  '150', '124', '141', '147', '112', '93', '92', '83', '72', '75', '46', '43',
  '31', '19', '22', '6', '15', '13', '4', '8', '8', '2', '3', '2' ]
```


01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (5/15)

- 인구수 데이터를 저장하기에 앞서 헤더(Header)를 살펴 보자

```
1 import csv
2
3 f = open("age.csv", encoding="cp949")
4 data = csv.reader(f)
5
6 header = next(data)
7 print(header)
8
9 f.close()
```

인구수 데이터는
row[3]부터 저장됨

실행결과

```
[ '행정구역', '2022년08월_계_총인구수', '2022년08월_계_연령구간인구수',
  '2022년08월_계_0세', '2022년08월_계_1세', '2022년08월_계_2세',
  ... (중략) ...
  '2022년08월_계_98세', '2022년08월_계_99세', '2022년08월_계_100세 이상' ]
```

01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (6/15)

- 우리 동네의 연령별 인구수 출력하기

```
1 import csv
2
3 f = open("age.csv", encoding="cp949")
4 data = csv.reader(f)
5
6 header = next(data)
7
8 for row in data:
9     if "신도림" in row[0]:
10         for j in row[3:]:
11             print(j)
12
13 f.close()
```

01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (7/15)

- 우리 동네의 연령별 인구수 데이터를 리스트에 저장하기

```
1 import csv
2
3 f = open("age.csv", encoding="cp949")
4 data = csv.reader(f)
5
6 header = next(data)
7
8 result = []
9 for row in data:
10     if "신도림" in row[0]:
11         for j in row[3:]:
12             result.append(j)
13
14 f.close()
15
16 print(result)
```

01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (8/15)

실행결과

```
[ '253', '270', '269', '312', '291', '313', '390', '412', '386', '369', '396',  
  '404', '380', '373', '387', '324', '324', '319', '322', '331', '341', '400',  
  '394', '357', '403', '452', '479', '427', '492', '484', '533', '508', '486',  
  '626', '611', '631', '613', '610', '652', '701', '681', '735', '685', '677',  
  '644', '627', '548', '654', '635', '641', '613', '588', '588', '574', '459',  
  '475', '488', '514', '418', '448', '456', '491', '483', '445', '429', '403',  
  '413', '412', '355', '336', '312', '229', '282', '242', '248', '240', '141',  
  '150', '124', '141', '147', '112', '93', '92', '83', '72', '75', '46', '43',  
  '31', '19', '22', '6', '15', '13', '4', '8', '8', '2', '3', '2' ]
```

현재는 문자열(String) 자료형이며,
시각화를 위해서는 정수(Integer) 자료형으로 변환해야 함

01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (9/15)

- 문자열 자료형을 정수 자료형으로 변환하기

```
1 import csv
2
3 f = open("age.csv", encoding="cp949")
4 data = csv.reader(f)
5
6 header = next(data)
7
8 result = []
9 for row in data:
10     if "신도림" in row[0]:
11         for j in row[3:]:
12             result.append(int(j))
13
14 f.close()
15
16 print(result)
```

01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (10/15)

실행결과

```
[253, 270, 269, 312, 291, 313, 390, 412, 386, 369, 396, 404, 380, 373, 387, 324,
324, 319, 322, 331, 341, 400, 394, 357, 403, 452, 479, 427, 492, 484, 533, 508,
486, 626, 611, 631, 613, 610, 652, 701, 681, 735, 685, 677, 644, 627, 548, 654,
635, 641, 613, 588, 588, 574, 459, 475, 488, 514, 418, 448, 456, 491, 483, 445,
429, 403, 413, 412, 355, 336, 312, 229, 282, 242, 248, 240, 141, 150, 124, 141,
147, 112, 93, 92, 83, 72, 75, 46, 43, 31, 19, 22, 6, 15, 13, 4, 8, 8, 2, 3, 2]
```

작은 따옴표(")가 사라졌음

01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (11/15)

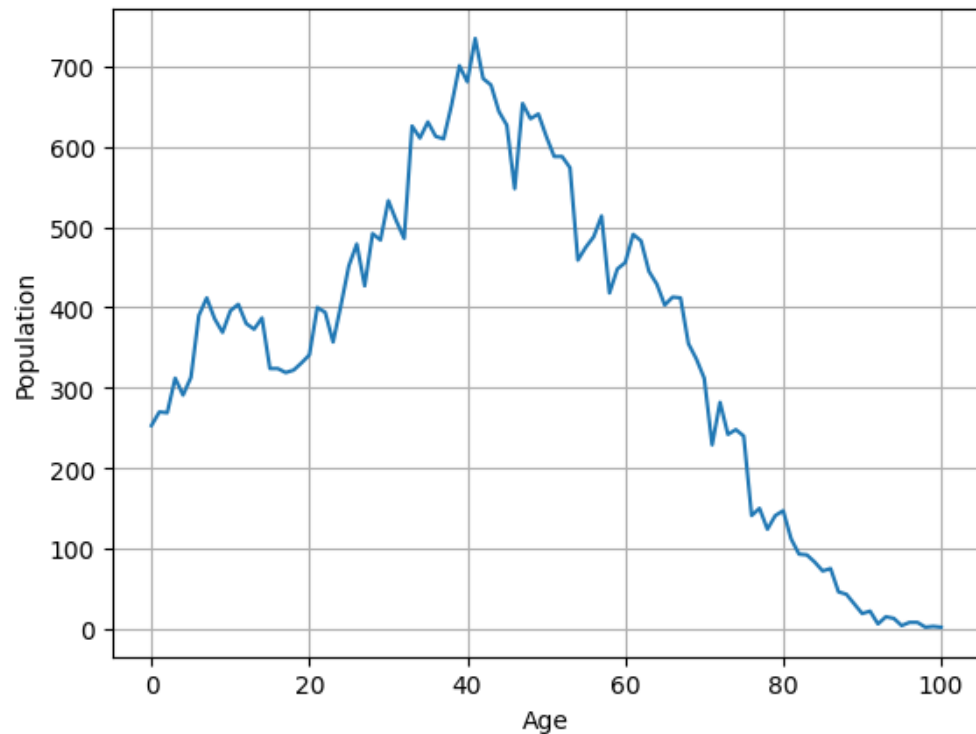
- 우리 동네의 연령별 인구수 데이터를 시각화하기

```
1 import csv
2 import matplotlib.pyplot as plt
3
4 f = open("age.csv", encoding="cp949")
5 data = csv.reader(f)
6 header = next(data)
7 result = []
8 for row in data:
9     if "신도림" in row[0]:
10         for j in row[3:]:
11             result.append(int(j))
12
13 f.close()
14 plt.figure()
15 plt.plot(result)
16 plt.xlabel("Age")
17 plt.ylabel("Population")
18 plt.grid()
19 plt.show()
```

01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (12/15)

실행결과



01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (13/15)

- 지역명을 입력 받아서 연령별 인구수 데이터를 시각화하기

```
1 import csv
2 import matplotlib.pyplot as plt
3
4 f = open("age.csv", encoding="cp949")
5 data = csv.reader(f)
6
7 header = next(data)
8
9 name = input("인구 구조가 알고 싶은 지역(읍면동 단위)을 입력하세요: ")
10
11 result = []
12 for row in data:
13     if name in row[0]:
14         for j in row[3:]:
15             result.append(int(j))
16
17 f.close()
18
19
```

01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (14/15)

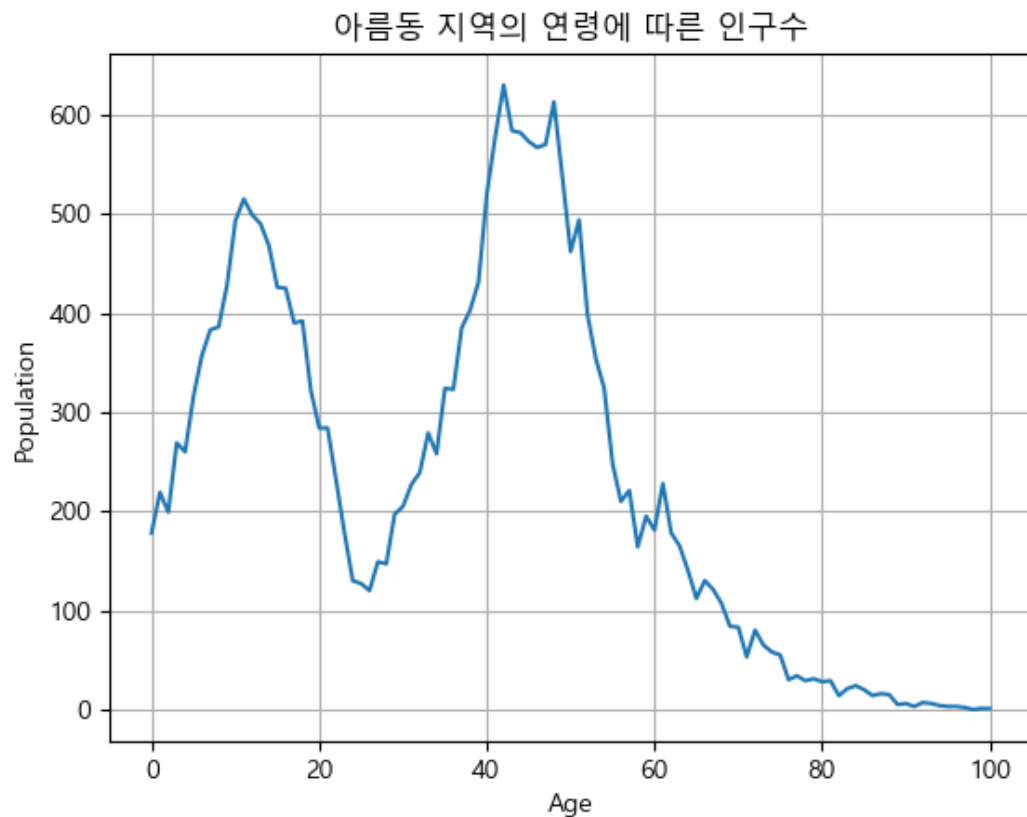
```
20 plt.rc("font", family="Malgun Gothic")
21 plt.figure()
22 plt.title(name + " 지역의 연령에 따른 인구수")
23 plt.plot(result)
24 plt.xlabel("Age")
25 plt.ylabel("Population")
26 plt.grid()
27 plt.show()
```

01. 우리 동네 인구 구조 시각화하기

❖ ③ 우리 동네 인구 구조 시각화하기 (15/15)

실행결과

인구 구조가 알고 싶은 지역(읍면동 단위)을 입력하세요: [아름동](#)



02. 인구 구조를 다양한 형태로 시각화하기

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

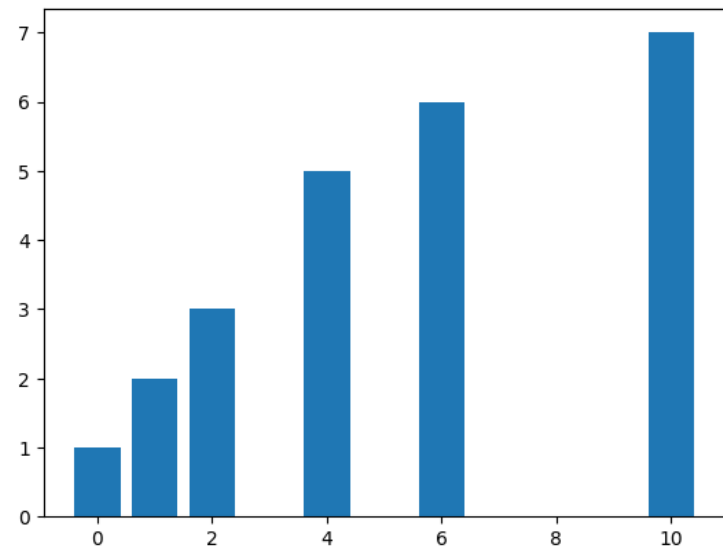
02. 인구 구조를 다양한 형태로 시각화하기

❖ ① 막대그래프(Bar Graph) 그리기 (1/6)

- 각 데이터의 크기(값)를 막대의 길이로 표현한 그래프
- `bar([막대를 표시할 위치], [막대의 길이])` 함수

```
1 import matplotlib.pyplot as plt
2 plt.figure()
3 plt.bar([0, 1, 2, 4, 6, 10], [1, 2, 3, 5, 6, 7])
4 plt.show()
```

실행결과



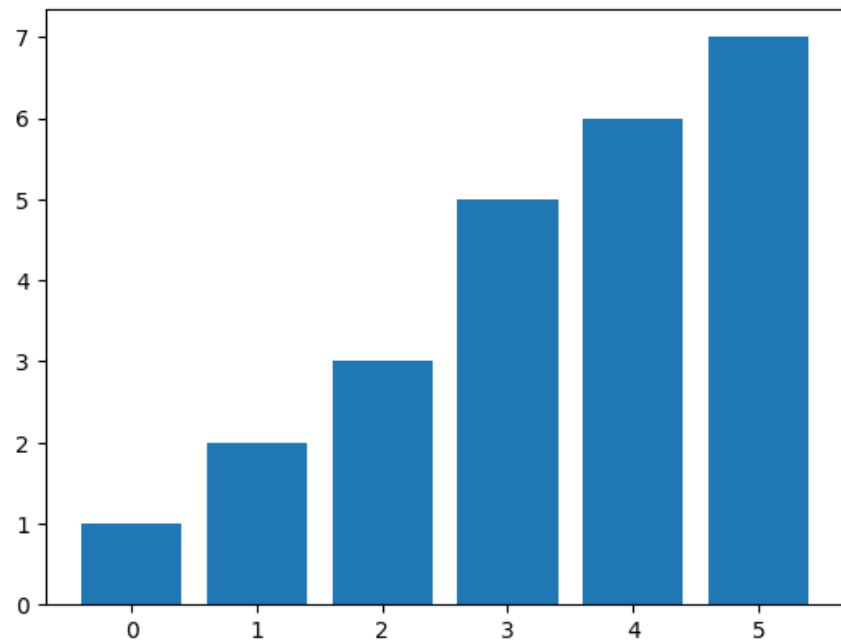
02. 인구 구조를 다양한 형태로 시각화하기

❖ ① 막대그래프(Bar Graph) 그리기 (2/6)

- range() 함수를 활용하면 막대를 표시할 위치를 쉽게 지정할 수 있음

```
1 import matplotlib.pyplot as plt
2 plt.figure()
3 plt.bar(range(0, 6, 1), [1, 2, 3, 5, 6, 7])
4 plt.show()
```

실행결과



02. 인구 구조를 다양한 형태로 시각화하기

❖ ① 막대그래프(Bar Graph) 그리기 (3/6)

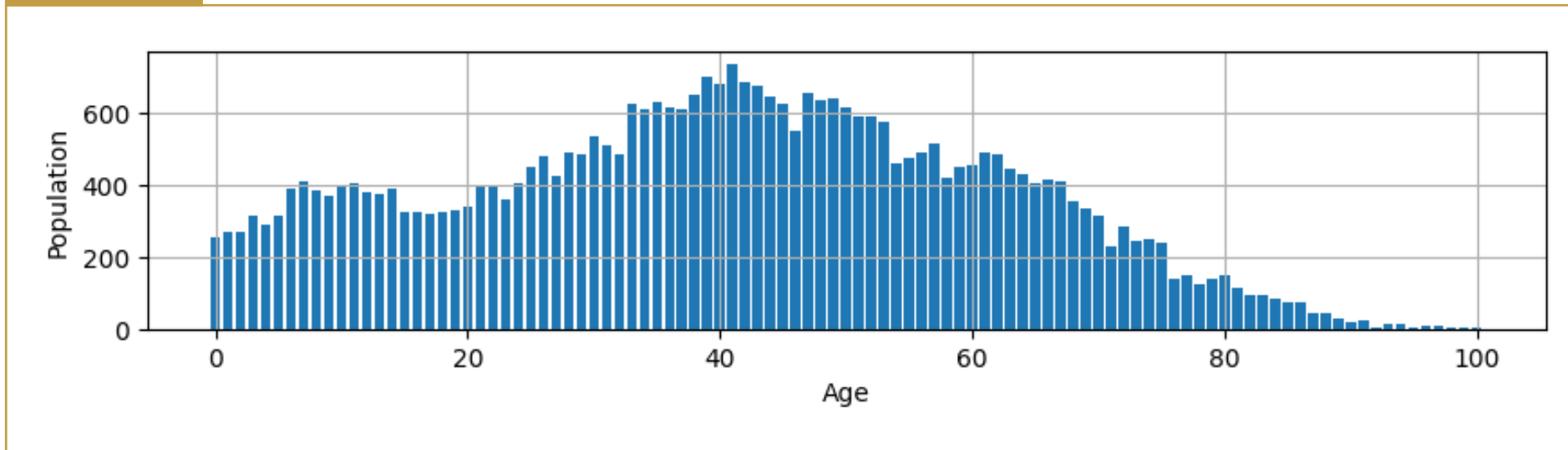
- 막대그래프를 활용하여 우리 동네의 연령별 인구수 데이터 시각화하기

```
1 import csv
2 import matplotlib.pyplot as plt
3 f = open("age.csv", encoding="cp949")
4 data = csv.reader(f)
5 header = next(data)
6 result = []
7 for row in data:
8     if "신도림" in row[0]:
9         for j in row[3:]:
10             result.append(int(j))
11
12 f.close()
13
14 plt.figure(figsize=(10, 2))
15 plt.bar(range(len(result)), result)
16 plt.xlabel("Age")
17 plt.ylabel("Population")
18 plt.grid()
19 plt.show()
```

02. 인구 구조를 다양한 형태로 시각화하기

❖ ① 막대그래프(Bar Graph) 그리기 (4/6)

실행결과



막대그래프를 수직이 아닌 수평 방향으로
그릴 수는 없을까?

02. 인구 구조를 다양한 형태로 시각화하기

❖ ① 막대그래프(Bar Graph) 그리기 (5/6)

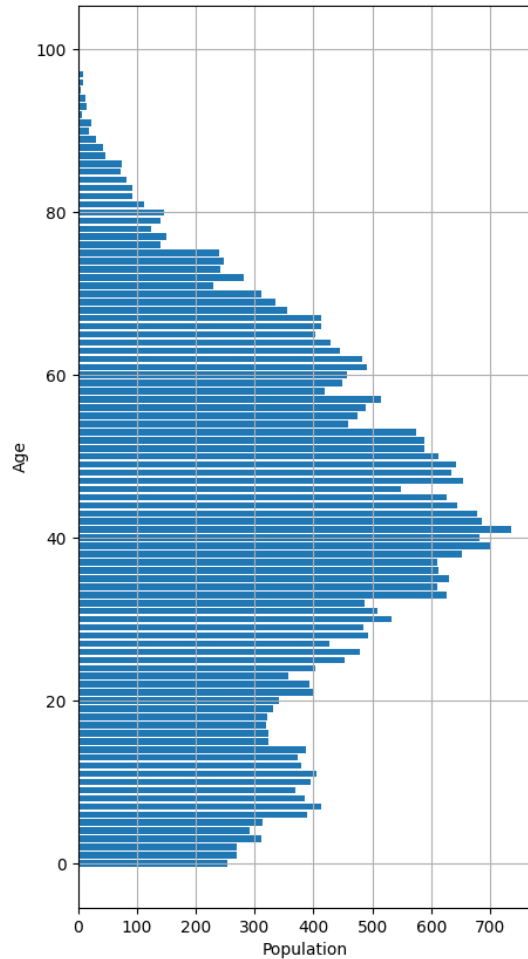
- 수평 막대그래프를 활용하여 우리 동네의 연령별 인구수 데이터 시각화하기 → `barh()` 함수

```
14 plt.figure(figsize=(5, 10))
15 plt.barh(range(len(result)), result)
16 plt.xlabel("Population")
17 plt.ylabel("Age")
18 plt.grid()
19 plt.show()
```

02. 인구 구조를 다양한 형태로 시각화하기

❖ ① 막대그래프(Bar Graph) 그리기 (6/6)

실행결과



02. 인구 구조를 다양한 형태로 시각화하기

❖ ② 남·여 인구 공공데이터 내려받기 (1/8)

- 행정안전부(<https://www.mois.go.kr/>) – [정책자료] – [통계] – [주민등록 인구통계] 버튼 클릭

The screenshot shows the homepage of the Ministry of the Interior (www.mois.go.kr). The navigation bar at the top includes links for '정보공개' (Information Disclosure), '참여·민원' (Participation·Petition), '뉴스·소식' (News·Notice), '정책자료' (Policy Data), '업무안내' (Business Guide), and '기관소개' (Institution Introduction). The '정책자료' (Policy Data) link is highlighted with a red box and a circled '1'. Below the navigation bar, there is a '정책자료' (Policy Data) section. On the left side of this section, there is a sidebar with links: '정책자료 홈', '주요업무계획', '법령정보', '전자관보', '통계' (highlighted with a red box and a circled '2'), '간행물', '정책연구보고서', and '참고자료'. The '통계' (Statistics) link is highlighted with a red box and a circled '2'. In the main content area of the '정책자료' section, there is a table of statistics. The table has two columns. The first column lists various statistics, and the second column lists the corresponding links. The link '주민등록 인구통계' (Residential Registration Population Statistics) is highlighted with a red box and a circled '3'. Below the '정책자료' section, there is a '공지사항' (Notice) section with a list of notices. To the right of the '공지사항' section, there is a '사진뉴스' (Photo News) section with a list of news items. To the right of the '사진뉴스' section, there is a '동영상뉴스' (Video News) section with a list of video news items.

02. 인구 구조를 다양한 형태로 시각화하기

❖ ② 남·여 인구 공공데이터 내려받기 (2/8)

- [연령별 인구현황] 버튼 클릭

행정안전부

주민등록 인구통계

주민등록 인구 및 세대현황

통계표 | 그래프

행정구역: 전국 | 시·군·구

등록구분: 전체

조회기간: 월간 | 연간 | 2022년 | 08월 ~ 2022년 | 08월

※ 매월 말일 작성 / 공표일시 : 매월 1일 12시 이후(공표일이 주말, 공휴일인 경우에는 다음 평일에 공표)

구분: 남·여 구분 | 남·여 구성비 | 세대당인구

정렬순서: 행정기관코드 | 오름차순

검색 | 초기화

현재화면 | 전체시군구현황 | 전체읍면동현황 | csv 파일 다운로드 | xlsx 파일 다운로드

주민등록 인구 및 세대현황

행정기관	2022년 08월					
	총 인구수	세대수	세대당 인구	남자 인구수	여자 인구수	남여 비율
전국	51,558,441	23,724,829	2.17	25,699,579	25,858,862	0.99
서울특별시	9,488,454	4,472,975	2.12	4,601,295	4,887,159	0.94
부산광역시	3,331,444	1,559,462	2.14	1,627,821	1,703,623	0.96

02. 인구 구조를 다양한 형태로 시각화하기

❖ ② 남·여 인구 공공데이터 내려받기 (3/8)

- [조회기간]: 2022년 08월 ~ 2022년 08월
- [계] 체크 해제
- [연령 구분 단위]: 1세
- 만 연령구분: 0, 100이상
- [검색] 버튼 클릭

행정안전부

주민등록 인구통계

연령별 인구현황

통계표 그래프

행정구역: 전국 시·군·구

등록구분: 전체

조회기간: 월간 연간 2022년 08월 ~ 2022년 08월

구분: 계 남·여 구분

정렬순서: 행정기관코드 으뜸차순

연령 구분 단위: 1세

만 연령구분: 0 100이상

검색 초기화

현재화면 전체시군구현황 전체읍면동현황 CSV 파일 다운로드 xlsx 파일 다운로드

연령별 인구현황

행정기관	남 인구수	연령구간인구수	0세	1세	2세	3세	4세	5세	6세
------	-------	---------	----	----	----	----	----	----	----

02. 인구 구조를 다양한 형태로 시각화하기

❖ ② 남·여 인구 공공데이터 내려받기 (4/8)

- [전체읍면동현황] 체크 후, [CSV 파일 다운로드] 버튼 클릭

행정안전부

주민등록 인구통계

연령별 인구현황

통계표 | 그래프

행정구역: 전국 | 시·군·구

등록구분: 전체

조회기간: 월간 | 연간 | 2022년 | 08월 ~ 2022년 | 08월

구분: ☐ 계 ☒ 남·여 구분

정렬순서: 행정기관코드 | 오름차순

연령 구분 단위: 1세

만 연령구분: 0 | 100이상

①

○ 현재읍면 ○ 전체시군구현황 ○ 전체읍면동현황 ○ CSV 파일 다운로드 ○ xlsx 파일 다운로드

②

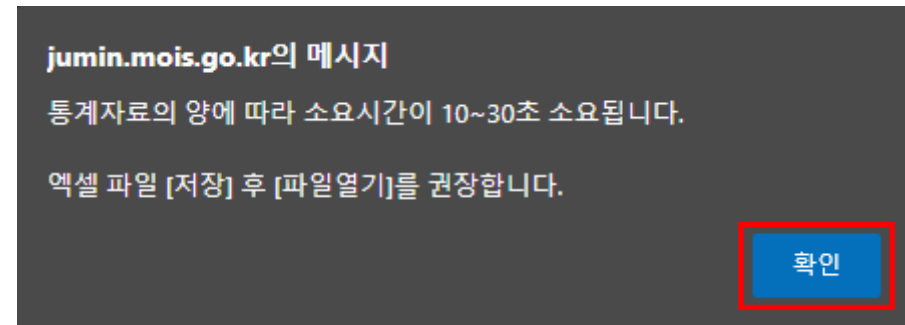
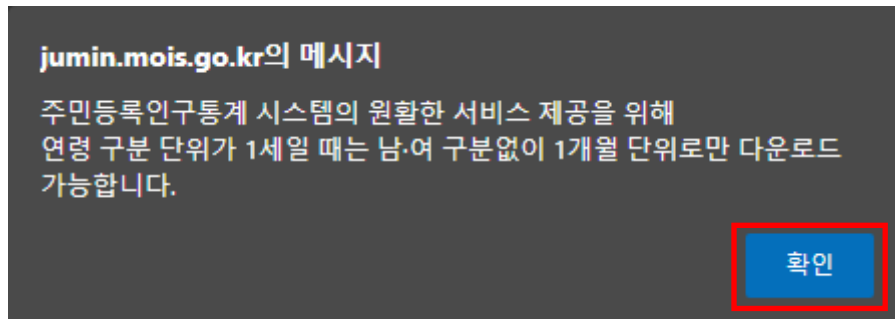
연령별 인구현황

행정기관	남 인구수	연령구간인구수	0세	1세	2세	3세	4세	5세	6세

02. 인구 구조를 다양한 형태로 시각화하기

❖ ② 남·여 인구 공공데이터 내려받기 (5/8)

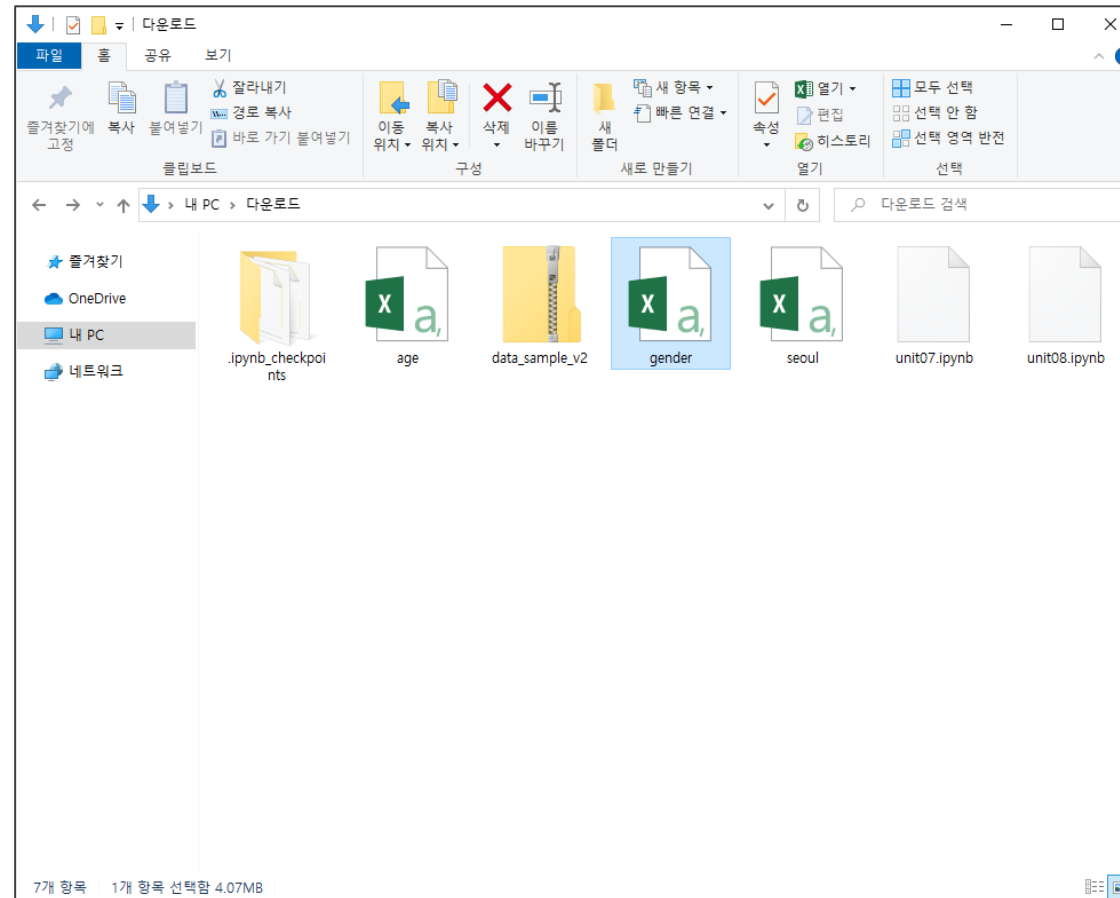
- 아래와 같이 경고 창이 뜨면, [확인] 버튼 클릭



02. 인구 구조를 다양한 형태로 시각화하기

❖ ② 남·여 인구 공공데이터 내려받기 (6/8)

- 파일 이름을 "gender.csv"로 변경함
- 파일은 다운로드(Downloads) 폴더에 저장되어 있다고 가정하고, 실습을 진행하자



02. 인구 구조를 다양한 형태로 시각화하기

❖ ② 남·여 인구 공공데이터 내려받기 (7/8)

- gender.csv 파일을 열어 보자

◆ B~CZ열: 남성 인구수 데이터

◆ DA~GY열: 여성 인구수 데이터

	CW	CX	CY	CZ	DA	DB	DC	DD	DE
1	2022년08월_남_97세	2022년08월_남_98세	2022년08월_남_99세	2022년08월_남_100세 이상	2022년08월_여_총인구수	2022년08월_여_연령구간인구수	2022년08월_여_0세	2022년08월_여_1세	2022년08월_여_2세
2	301	240	246	414	4,887,159	4,887,159	19,816	21,718	22,053
3	12	9	8	15	74,091	74,091	212	236	260
4	3	1	0	1	6,374	6,374	24	22	20
5	2	2	1	1	5,154	5,154	17	16	21
6	0	1	0	2	1,254	1,254	2	4	3
7	0	1	0	0	4,935	4,935	10	14	11
8	2	1	2	2	9,570	9,570	21	46	51
9	0	0	0	1	4,349	4,349	20	19	27
10	0	0	0	0	5,365	5,365	27	32	42
11	0	0	0	0	2,148	2,148	6	4	3
12	0	1	1	4	2,937	2,937	5	11	7
13	0	0	1	1	2,380	2,380	2	5	3
14	0	0	0	0	3,767	3,767	14	8	9
15	1	1	3	1	8,501	8,501	16	19	19
16	2	0	0	0	2,257	2,257	1	5	3
17	1	0	0	0	3,799	3,799	8	3	11
18	1	0	0	2	3,297	3,297	18	13	16
19	0	0	0	0	3,027	3,027	10	2	10
20	0	1	0	0	4,977	4,977	11	13	4
21	7	7	3	14	62,651	62,651	299	255	289
22	0	0	0	0	1,156	1,156	5	5	3
23	0	1	0	1	2,281	2,281	11	6	8

02. 인구 구조를 다양한 형태로 시각화하기

❖ ② 남·여 인구 공공데이터 내려받기 (8/8)

- 인구수 데이터와 인덱스(Index)의 관계

		남성 연령별 인구수							여성 연령별 인구수						
열 이름	지역명	총인구1	총인구2	0세	1세	(생략)	99세	100세 이상	총인구1	총인구2	0세	1세	(생략)	99세	100세 이상
인덱스	0	1	2	3	4	...	102	103	104	105	106	107	...	205	206

		남성 연령별 인구수							여성 연령별 인구수						
열 이름	지역명	총인구1	총인구2	0세	1세	(생략)	99세	100세 이상	총인구1	총인구2	0세	1세	(생략)	99세	100세 이상
인덱스	0	1	2	3	4	...	102	103	-103	-102	-101	-100	...	-2	-1

02. 인구 구조를 다양한 형태로 시각화하기

❖ ③ 항아리 모양 그래프 그리기 (1/8)

- 수평 막대그래프를 활용하여 성별 인구수 데이터 시각화하기

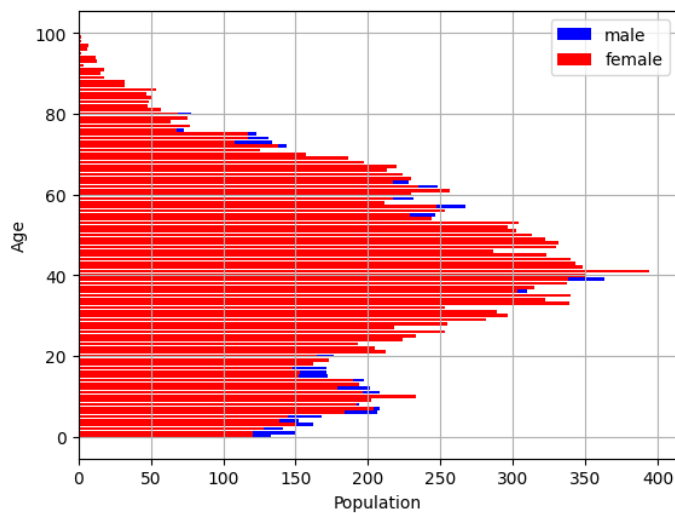
```
1 import csv
2 import matplotlib.pyplot as plt
3
4 f = open("gender.csv", encoding="cp949")
5 data = csv.reader(f)
6 header = next(data)
7
8 male = []
9 female = []
10 for row in data:
11     if "신도림" in row[0]:
12         for j in row[3:104]:
13             male.append(int(j))
14         for j in row[106:207]:           # row[106:]이라고 적어도 됨
15             female.append(int(j))
16
17 f.close()
18
19
```

02. 인구 구조를 다양한 형태로 시각화하기

❖ ③ 항아리 모양 그래프 그리기 (2/8)

```
20 plt.figure()
21 plt.barh(range(len(male)), male, color='b', label="male")
22 plt.barh(range(len(female)), female, color='r', label="female")
23 plt.xlabel("Population")
24 plt.ylabel("Age")
25 plt.grid()
26 plt.legend()
27 plt.show()
```

실행결과



그래프가 겹쳐서
알아보기가 어려움


02. 인구 구조를 다양한 형태로 시각화하기

❖ ③ 항아리 모양 그래프 그리기 (3/8)

- 수평 막대그래프를 활용하여 성별 인구수 데이터 시각화하기

```
1 import csv
2 import matplotlib.pyplot as plt
3
4 f = open("gender.csv", encoding="cp949")
5 data = csv.reader(f)
6 header = next(data)
7
8 male = []
9 female = []
10 for row in data:
11     if "신도림" in row[0]:
12         for j in row[3:104]:
13             male.append(-int(j))
14         for j in row[106:]:
15             female.append(int(j))
16
17 f.close()
18
19
```

남성 데이터를
음수로 변경함

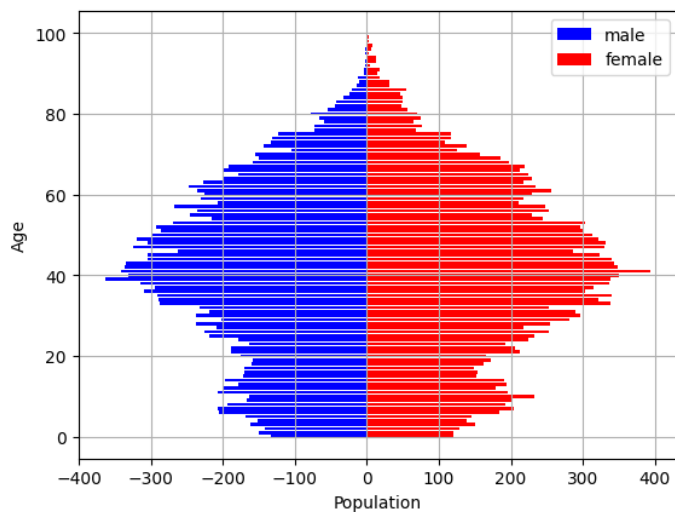


02. 인구 구조를 다양한 형태로 시각화하기

❖ ③ 항아리 모양 그래프 그리기 (4/8)

```
20 plt.figure()
21 plt.barh(range(len(male)), male, color='b', label="male")
22 plt.barh(range(len(female)), female, color='r', label="female")
23 plt.xlabel("Population")
24 plt.ylabel("Age")
25 plt.grid()
26 plt.legend()
27 plt.show()
```

실행결과



02. 인구 구조를 다양한 형태로 시각화하기

❖ ③ 항아리 모양 그래프 그리기 (5/8)

- 한글 폰트 사용하기(Windows 기준)

- ◆ Malgun Gothic은 “맑은 고딕”임

- ◆ 만약 macOS 운영체제를 사용하고 있다면, “AppleGothic”이라고 쓰면 됨

```
1 plt.rc("font", family="Malgun Gothic")  
2 plt.title("크리스마스의 기온 변화 그래프")
```

- 한글 폰트 사용시 마이너스 부호 표현하기

```
1 plt.rcParams["axes.unicode_minus"] = False
```

02. 인구 구조를 다양한 형태로 시각화하기

❖ ③ 항아리 모양 그래프 그리기 (6/8)

- 찾고 싶은 지역의 이름을 입력 받아 성별 인구수 데이터 시각화하기

```
1 import csv
2 import matplotlib.pyplot as plt
3
4 f = open("gender.csv", encoding="cp949")
5 data = csv.reader(f)
6 header = next(data)
7
8 male, female = [], []
9 name = input("찾고 싶은 지역의 이름을 알려주세요: ")
10 for row in data:
11     if name in row[0]:
12         for j in row[3:104]:
13             male.append(-int(j))
14         for j in row[106:]:
15             female.append(int(j))
16
17 f.close()
18
19
```


02. 인구 구조를 다양한 형태로 시각화하기

❖ ③ 항아리 모양 그래프 그리기 (7/8)

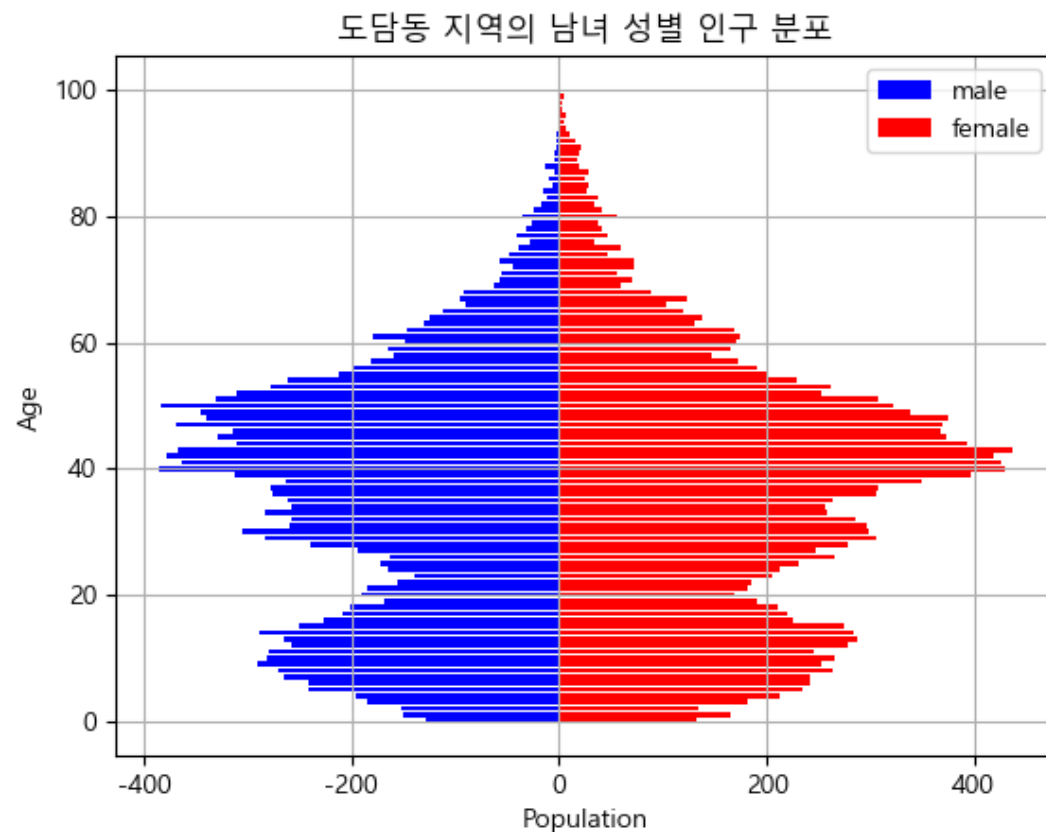
```
20 plt.rc("font", family="Malgun Gothic")
21 plt.rcParams["axes.unicode_minus"] = False
22 plt.figure()
23 plt.title(name + " 지역의 남녀 성별 인구 분포")
24 plt.barh(range(len(male)), male, color='b', label="male")
25 plt.barh(range(len(female)), female, color='r', label="female")
26 plt.xlabel("Population")
27 plt.ylabel("Age")
28 plt.legend()
29 plt.grid()
30 plt.show()
```

02. 인구 구조를 다양한 형태로 시각화하기

❖ ③ 항아리 모양 그래프 그리기 (8/8)

실행결과

찾고 싶은 지역의 이름을 알려주세요: **도담동**





03. 우리 동네 인구 구조를 파이 차트로 나타내기

pie chart

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

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ① 제주도에는 여성의 비율이 더 높을까? (1/11)

- 예로부터 제주도를 돌, 바람, 여성이 많다 하여, 삼다도(三多島)라고 불렀다고 함
- 정말로 그럴까?
- 정말인지 확인해보기 위해서 제주도의 성별 인구 비율을 비교해보자

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ① 제주도에는 여성의 비율이 더 높을까? (2/11)

- 제주특별자치도의 성별 분포를 항아리 모양 그래프로 표현하기

```
1 import csv
2 import matplotlib.pyplot as plt
3
4 f = open("gender.csv", encoding="cp949")
5 data = csv.reader(f)
6 header = next(data)
7
8 male, female = [], []
9 name = input("찾고 싶은 지역의 이름을 알려주세요: ")
10 for row in data:
11     if name in row[0]:
12         for j in row[3:104]:
13             male.append(-int(j))
14         for j in row[106:]:
15             female.append(int(j))
16
17 f.close()
18
19
```

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ① 제주도에는 여성의 비율이 더 높을까? (3/11)

```
20 plt.rc("font", family="Malgun Gothic")
21 plt.rcParams["axes.unicode_minus"] = False
22 plt.figure()
23 plt.title(name + " 지역의 남녀 성별 인구 분포")
24 plt.barh(range(len(male)), male, color='b', label="male")
25 plt.barh(range(len(female)), female, color='r', label="female")
26 plt.xlabel("Population")
27 plt.ylabel("Age")
28 plt.legend()
29 plt.grid()
30 plt.show()
```

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ① 제주도에는 여성의 비율이 더 높을까? (4/11)

실행결과

찾고 싶은 지역의 이름을 알려주세요: 제주특별자치도

```
-----  
ValueError                                Traceback (most recent call last)  
<ipython-input-4-a4a61e7316c7> in <cell line: 10>()  
    11     if name in row[0]:  
    12         for j in row[3:104]:  
--> 13             male.append(-int(j))  
    14         for j in row[106:]:  
    15             female.append(int(j))
```

ValueError: invalid literal for int() with base 10: '1,835'

실행결과 ValueError가 발생했음

문자열 '1,835'에서 콤마(,) 때문에
문제가 발생한 것으로 예상됨

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ① 제주도에는 여성의 비율이 더 높을까? (5/11)

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

```
1 import csv
2 import matplotlib.pyplot as plt
3
4 f = open("gender.csv", encoding="cp949")
5 data = csv.reader(f)
6 header = next(data)
7
8 male, female = [], []
9 name = input("찾고 싶은 지역의 이름을 알려주세요: ")
10 for row in data:
11     if name in row[0]:
12         for j in row[3:104]:
13             j = j.replace(',', '')
14             male.append(-int(j))
15         for j in row[106:]:
16             j = j.replace(',', '')
17             female.append(int(j))
18
19 f.close()
```

- ✓ replace('특정 문자', '') → 특정 문자를 제거
- ✓ replace() 함수를 활용하면 문자열에서 특정 문자를 제거할 수 있음

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ① 제주도에는 여성의 비율이 더 높을까? (6/11)

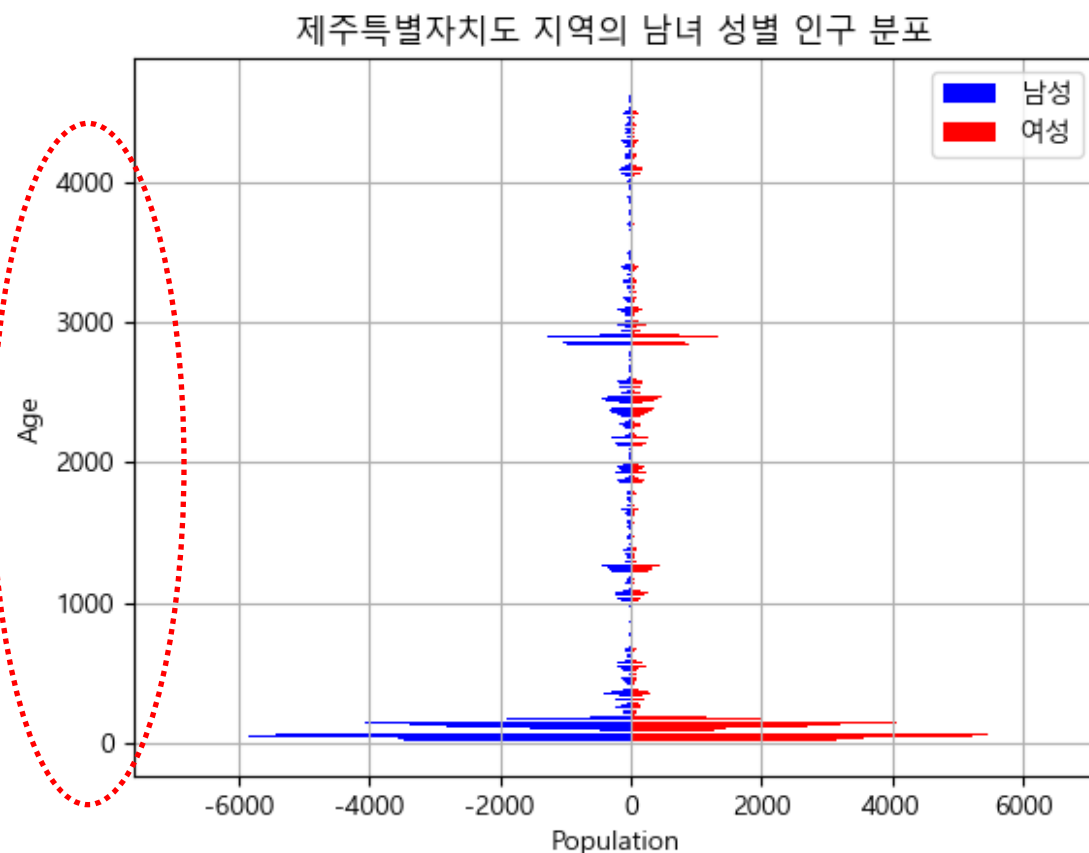
```
20 plt.rc("font", family="Malgun Gothic")
21 plt.rcParams["axes.unicode_minus"] = False
22 plt.figure()
23 plt.title(name + " 지역의 남녀 성별 인구 분포")
24 plt.barh(range(len(male)), male, color='b', label="male")
25 plt.barh(range(len(female)), female, color='r', label="female")
26 plt.xlabel("Population")
27 plt.ylabel("Age")
28 plt.legend()
29 plt.grid()
30 plt.show()
```

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ① 제주도에는 여성의 비율이 더 높을까? (7/11)

실행결과

찾고 싶은 지역의 이름을 알려주세요: 제주특별자치도



ValueError는 발생하지 않았지만,
y축 Age 부분의 값이 엄청 많음

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ① 제주도에는 여성의 비율이 더 높을까? (8/11)

- y축 Age 부분에 값이 왜 많을까?

The screenshot shows an Excel spreadsheet with a search dialog box open. The dialog box displays a list of data sources, including 'gender.csv' and 'gender'. A red box highlights the search results, and a red arrow points from a text box on the right to the highlighted area.

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 있음

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

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ① 제주도에는 여성의 비율이 더 높을까? (9/11)

- break 문을 활용하여, 제주특별자치도의 성별 분포를 항아리 모양 그래프로 표현하기

```
1 import csv
2 import matplotlib.pyplot as plt
3 f = open("gender.csv", encoding="cp949")
4 data = csv.reader(f)
5 header = next(data)
6
7 male, female = [], []
8 name = input("찾고 싶은 지역의 이름을 알려주세요: ")
9 for row in data:
10     if name in row[0]:
11         for j in row[3:104]:
12             j = j.replace(',', '')
13             male.append(-int(j))
14         for j in row[106:]:
15             j = j.replace(',', '')
16             female.append(int(j))
17         break
18
19 f.close()
```

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

break 문을 추가하면,
for 반복문을 멈출 수 있음

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ① 제주도에는 여성의 비율이 더 높을까? (10/11)

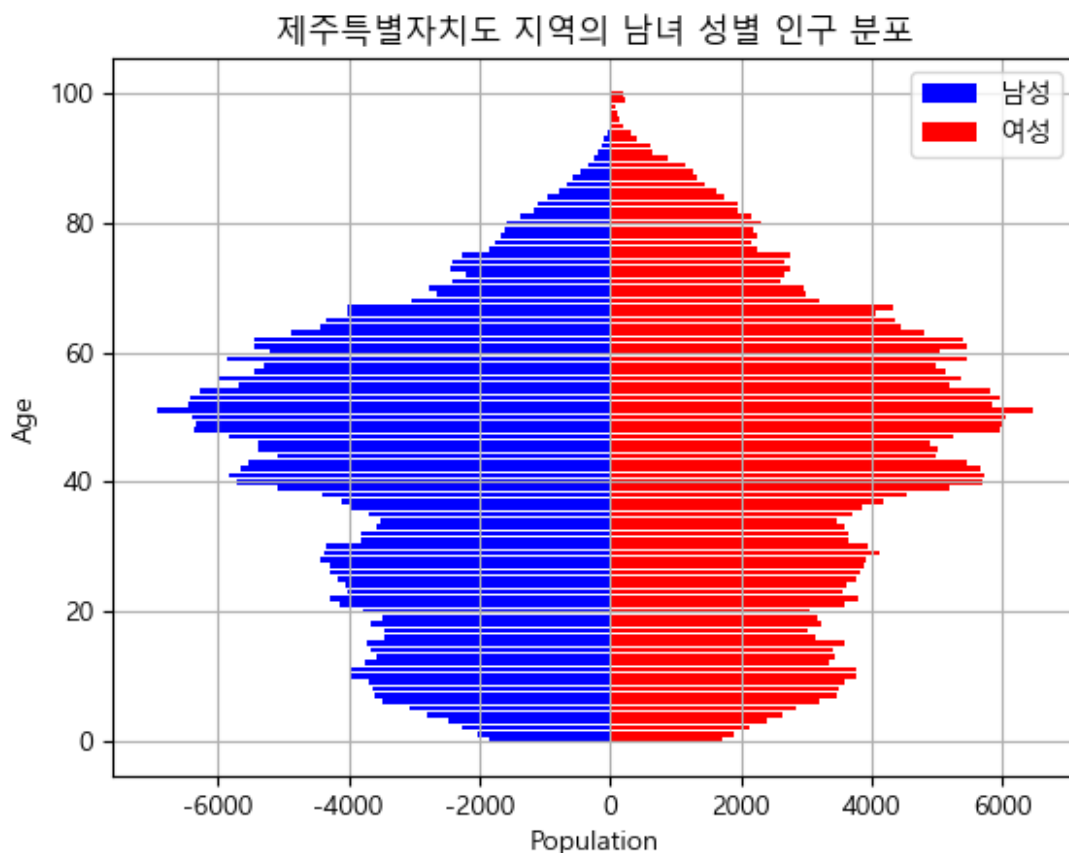
```
20 plt.rc("font", family="Malgun Gothic")
21 plt.rcParams["axes.unicode_minus"] = False
22 plt.figure()
23 plt.title(name + " 지역의 남녀 성별 인구 분포")
24 plt.barh(range(len(male)), male, color='b', label="male")
25 plt.barh(range(len(female)), female, color='r', label="female")
26 plt.xlabel("Population")
27 plt.ylabel("Age")
28 plt.legend()
29 plt.grid()
30 plt.show()
```

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ① 제주도에는 여성의 비율이 더 높을까? (11/11)

실행결과

찾고 싶은 지역의 이름을 알려주세요: 제주특별자치도



남성과 여성 중 어떤 성별의 비율이 높을까?
항아리 모양의 그래프로는 알기 어려운 정보임

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ② 혈액형 비율 표현하기 (1/9)

- 원그래프는 전체 데이터 중 특정 데이터의 비율을 보기 쉽게 표현함 → pie() 함수
- pie() 함수를 이용하면 쉽게 원그래프를 그릴 수 있음

```
1 import matplotlib.pyplot as plt
2 plt.figure()
3 plt.pie([10, 20])
4 plt.show()
```

실행결과



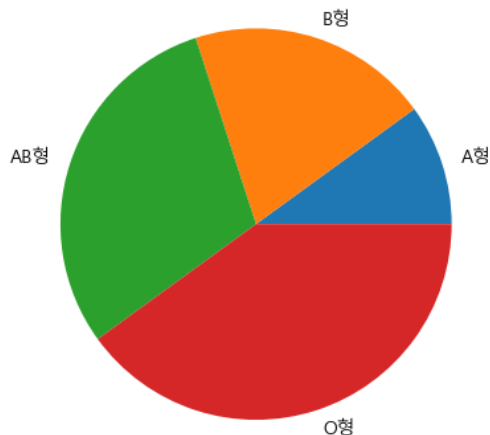
03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ② 혈액형 비율 표현하기 (2/9)

- 레이블(Label) 추가하기

```
1 import matplotlib.pyplot as plt
2 data = [10, 20, 30, 40]
3 name = ["A형", "B형", "AB형", "O형"]
4 plt.rc("font", family="Malgun Gothic")
5 plt.figure()
6 plt.pie(data, labels=name)
7 plt.show()
```

실행결과



원그래프에서는 labels에 각 항목의 이름을 입력하면
항목들의 의미를 쉽게 알아 볼 수 있음

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ② 혈액형 비율 표현하기 (3/9)

- 비율 및 범례 표시하기

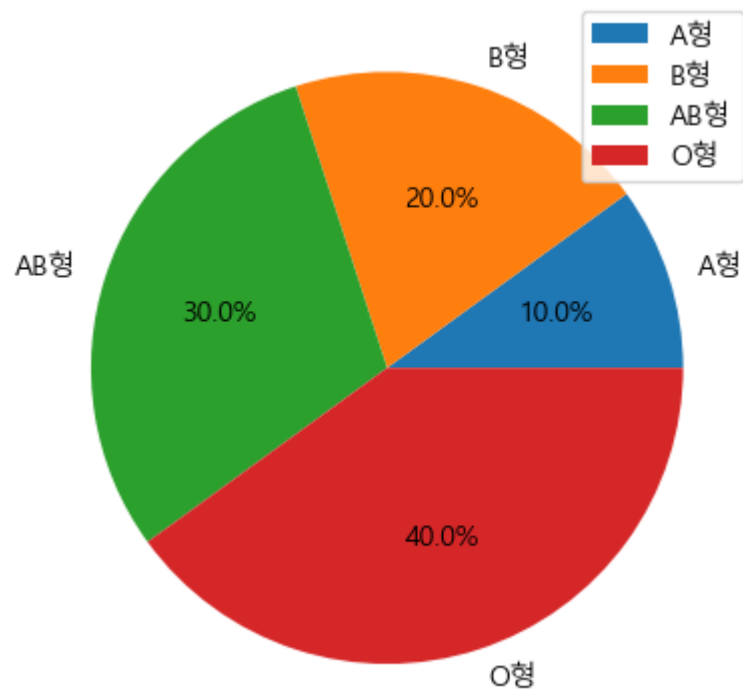
```
1 import matplotlib.pyplot as plt
2
3 data = [10, 20, 30, 40]
4 name = ["A형", "B형", "AB형", "O형"]
5
6 plt.rc("font", family="Malgun Gothic")
7 plt.figure()
8 plt.pie(data, labels=name, autopct="%.1f%%")
9 plt.legend()
10 plt.show()
```

- ✓ 각 항목의 비율을 표시하고 싶다면 autopct 속성의 값을 지정해주면 됨
- ✓ autopct는 auto percent를 의미함

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ② 혈액형 비율 표현하기 (4/9)

실행결과



'%.1f%%'의 의미

- ✓ %는 백분율로 표현하겠다는 의미임
- ✓ .1f는 소수점 아래 첫 번째 소수점까지 표현하겠다는 의미임
- ✓ %%는 표현되는 실수 뒤에 "%" 기호를 적기 위해 작성한 것임

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ② 혈액형 비율 표현하기 (5/9)

- 범례 위치 조정하기

```
1 import matplotlib.pyplot as plt
2
3 data = [10, 20, 30, 40]
4 name = ["A형", "B형", "AB형", "O형"]
5
6 plt.rc("font", family="Malgun Gothic")
7 plt.figure()
8 plt.pie(data, labels=name, autopct="%.1f%%")
9 plt.legend(loc=10)
10 plt.show()
```

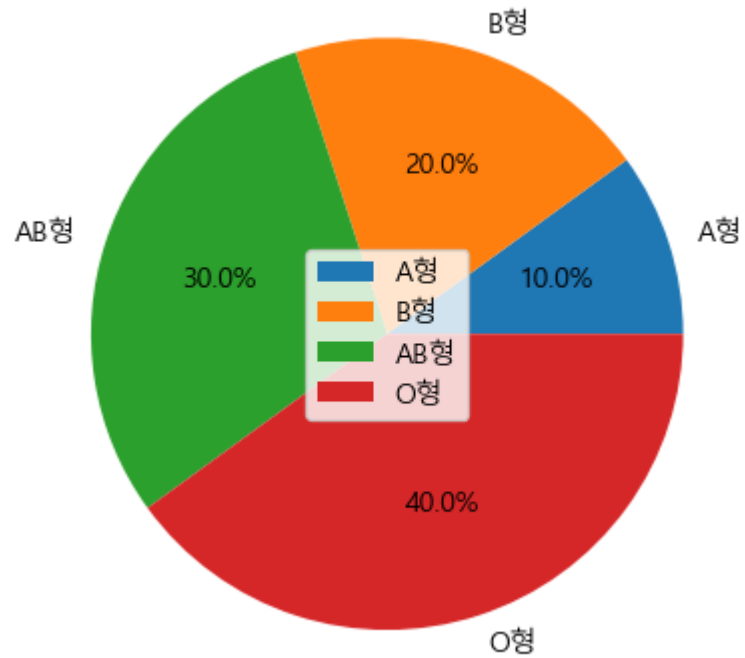
아래의 표를 참조하여 loc에 값을 입력하면,
범례의 위치를 조절할 수 있음

2	9	1
6	10	5, 7
3	8	4

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ② 혈액형 비율 표현하기 (6/9)

실행결과



03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ② 혈액형 비율 표현하기 (7/9)

- 색 및 돌출 효과 정하기

```
1 import matplotlib.pyplot as plt
2
3 data = [15, 25, 35, 45]
4 name = ["A형", "B형", "AB형", "O형"]
5 color = ["violet", "skyblue", "limegreen", "wheat"]
6
7 plt.rc("font", family="Malgun Gothic")
8 plt.figure()
9 plt.pie(data, labels=name, autopct="%.1f%%", colors=color,
10         explode=(0, 0, 0.1, 0))
11 plt.legend(loc=1)
12 plt.show()
```

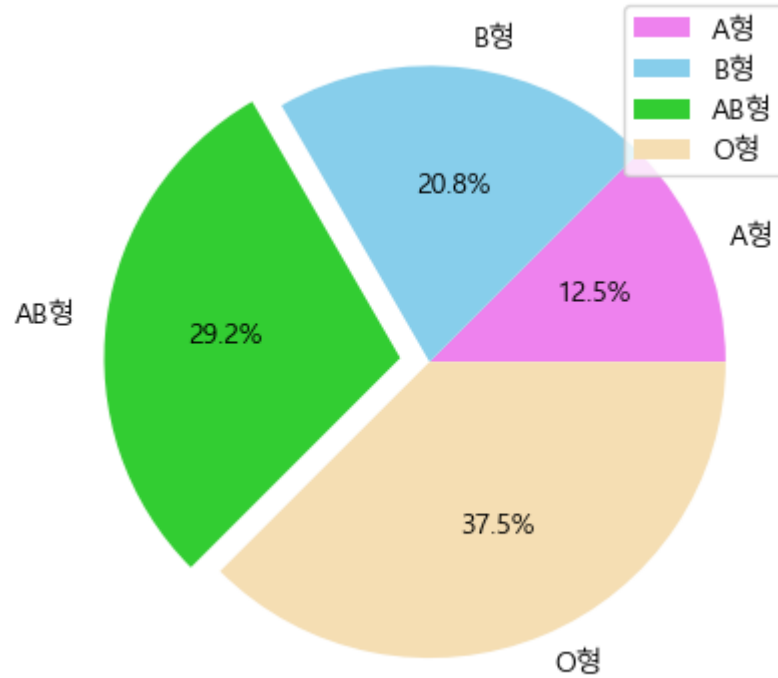
- explode
(동사) 폭발하다
(동사) 폭발시키다

- ✓ 색은 colors 속성으로 설정할 수 있음
- ✓ 돌출 효과는 explode 속성으로 설정할 수 있음

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ② 혈액형 비율 표현하기 (8/9)

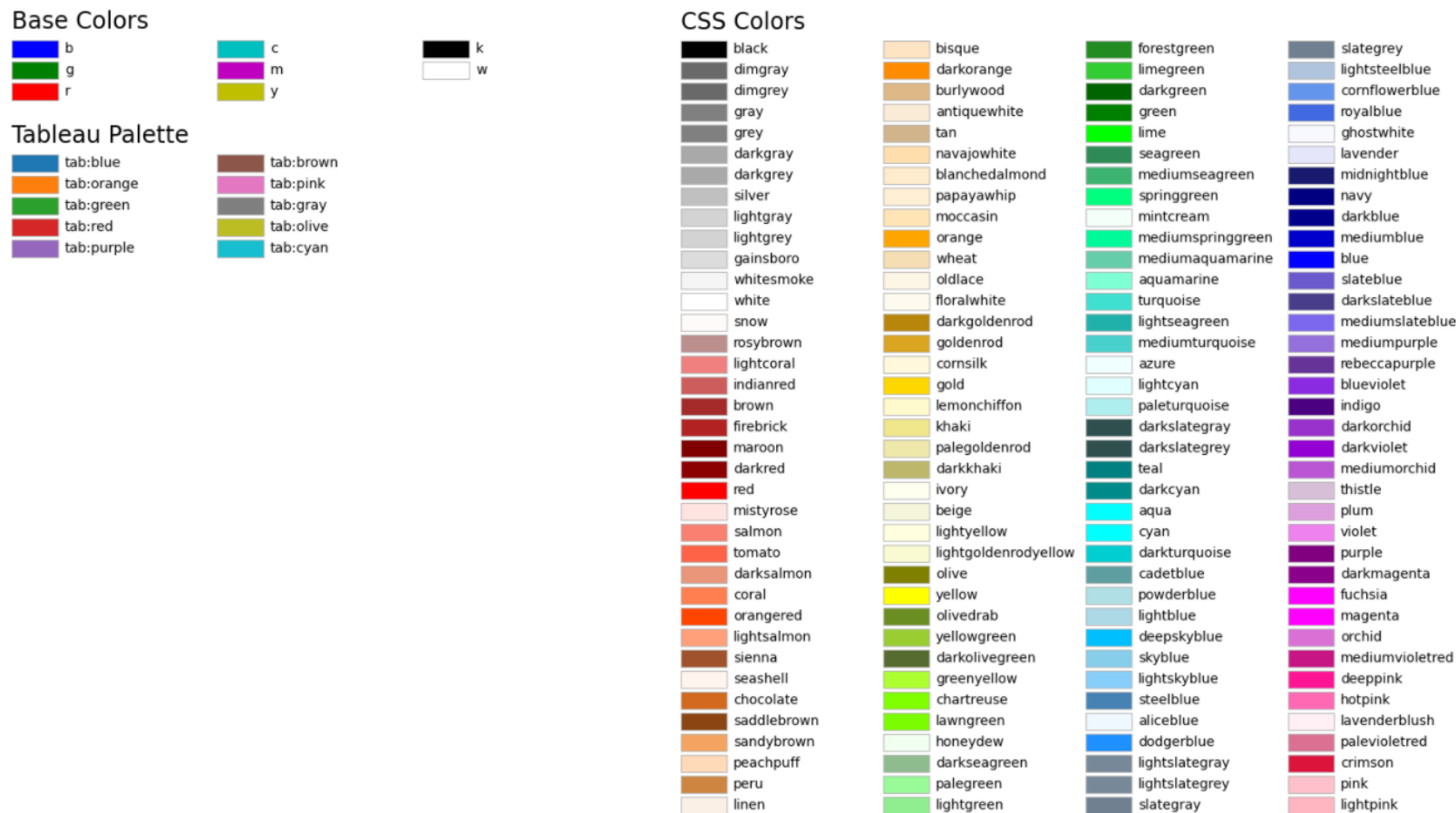
실행결과



03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ② 혈액형 비율 표현하기 (9/9)

- matplotlib 라이브러리에서 사용할 수 있는 다양한 색의 이름
- [URL] https://matplotlib.org/stable/gallery/color/named_colors.html



03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ③ 제주도의 성별 인구 비율 표현하기 (1/6)

- 제주특별자치도의 성별 인구 합계 구하기

```
1 import csv
2 import matplotlib.pyplot as plt
3 f = open("gender.csv", encoding="cp949")
4 data = csv.reader(f)
5 header = next(data)
6 name = input("찾고 싶은 지역의 이름을 알려주세요: ")
7 for row in data:
8     if name in row[0]:
9         male, female = 0, 0
10        for j in row[3:104]:           # 3번 방에 0세, ..., 103번 방에 100세 이상
11            j = j.replace(',', '')
12            male = male + int(j)
13        for j in row[106:]:           # 106번 방에 0세, ..., 206번 방에 100세 이상
14            j = j.replace(',', '')
15            female = female + int(j)
16        break
17 f.close()
18 print("남성 인구 합:", male, "명")
19 print("여성 인구 합:", female, "명")
```


03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ③ 제주도의 성별 인구 비율 표현하기 (2/6)

실행결과

찾고 싶은 지역의 이름을 알려주세요: [제주특별자치도](#)

남성 인구 합: 340105 명

여성 인구 합: 338911 명

03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ③ 제주도의 성별 인구 비율 표현하기 (3/6)

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

```
18 labels = ["male", "female"]
19 colors = ["skyblue", "pink"]
20
21 plt.rc("font", family="Malgun Gothic")
22 plt.figure()
23 plt.title(name + " 지역의 성별 비율")
24 plt.pie([male, female], labels=labels, colors=colors, autopct="%.1f%%")
25 plt.show()
```

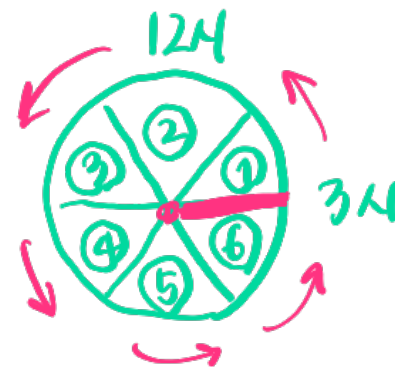
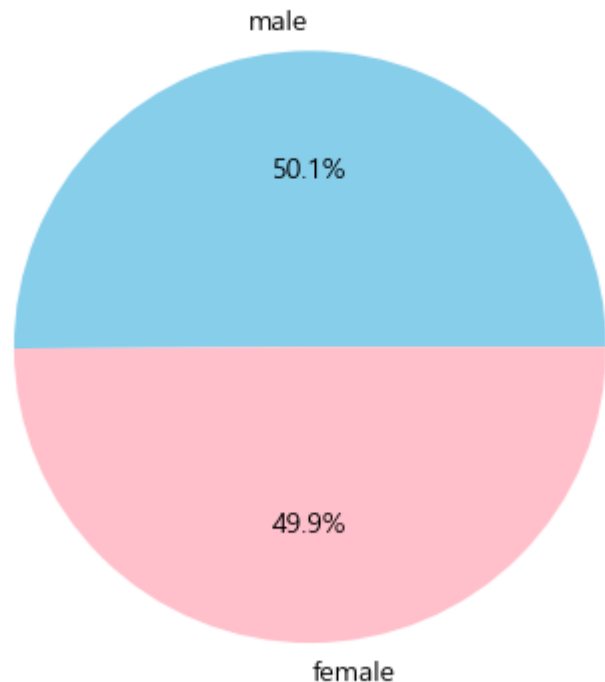
03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ③ 제주도의 성별 인구 비율 표현하기 (4/6)

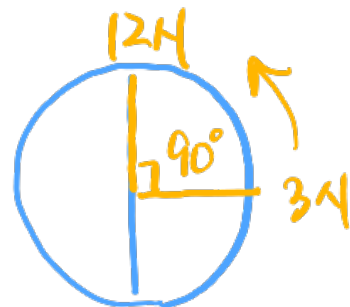
실행결과

찾고 싶은 지역의 이름을 알려주세요: 제주특별자치도

제주특별자치도 지역의 성별 비율



원그래프의 시작 각도를 조정하려면
어떻게 해야 할까요?



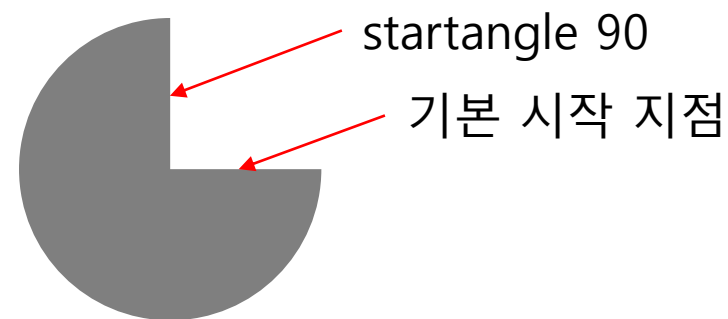
03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ③ 제주도의 성별 인구 비율 표현하기 (5/6)

- 원그래프(Pie Chart)의 시작 각도 조정하기

```
18 labels = ["male", "female"]
19 colors = ["skyblue", "pink"]
20
21 plt.rc("font", family="Malgun Gothic")
22 plt.figure()
23 plt.title(name + " 지역의 성별 비율")
24 plt.pie([male, female], labels=labels, colors=colors, autopct="%.1f%%",
25         startangle=90)
26 plt.show()
```

startangle 속성에 시작 각도를 지정하면 됨



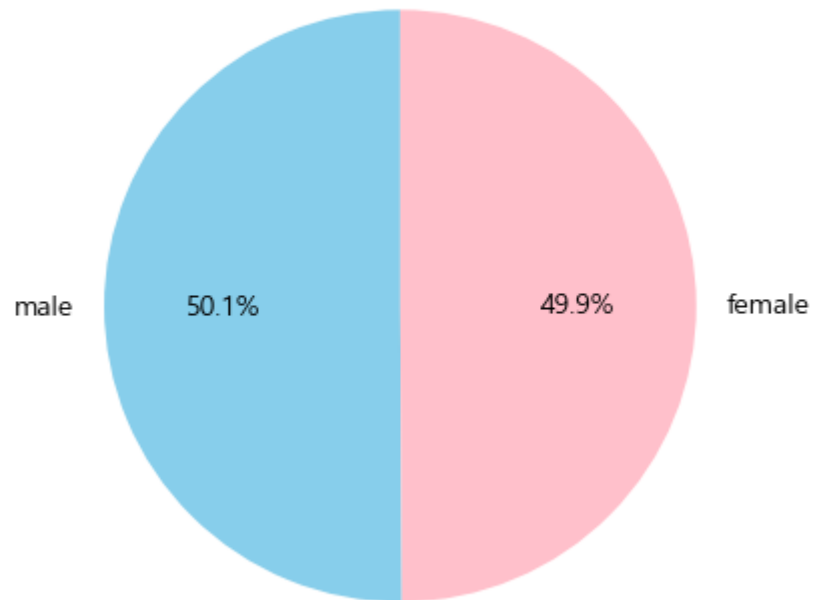
03. 우리 동네 인구 구조를 파이 차트로 나타내기

❖ ③ 제주도의 성별 인구 비율 표현하기 (6/6)

실행결과

찾고 싶은 지역의 이름을 알려주세요: 제주특별자치도

제주특별자치도 지역의 성별 비율



확인 결과 제주도의
남녀 인구 비율은 거의 동일함!

산: 산포되어 있다(=흩뿌려져 있다)

점: points

도: 그림 도

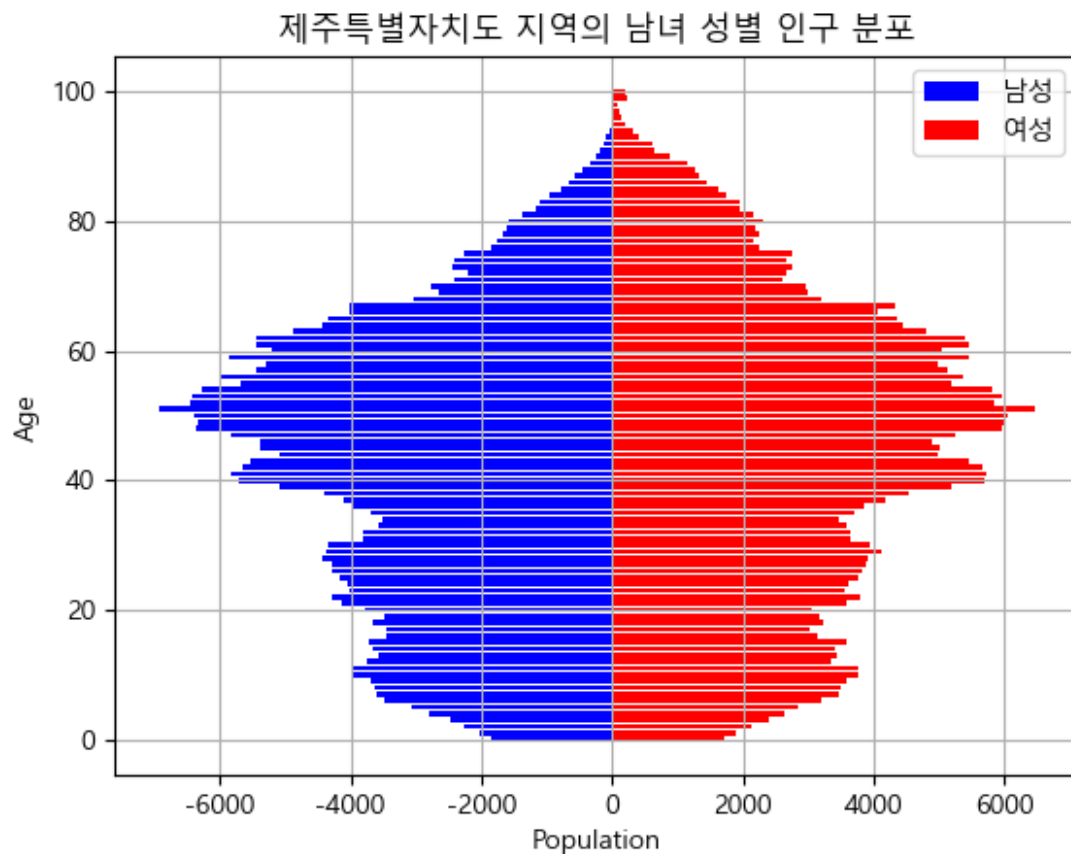
04. 우리 동네 인구 구조를 산점으로 나타내기

scatter plot

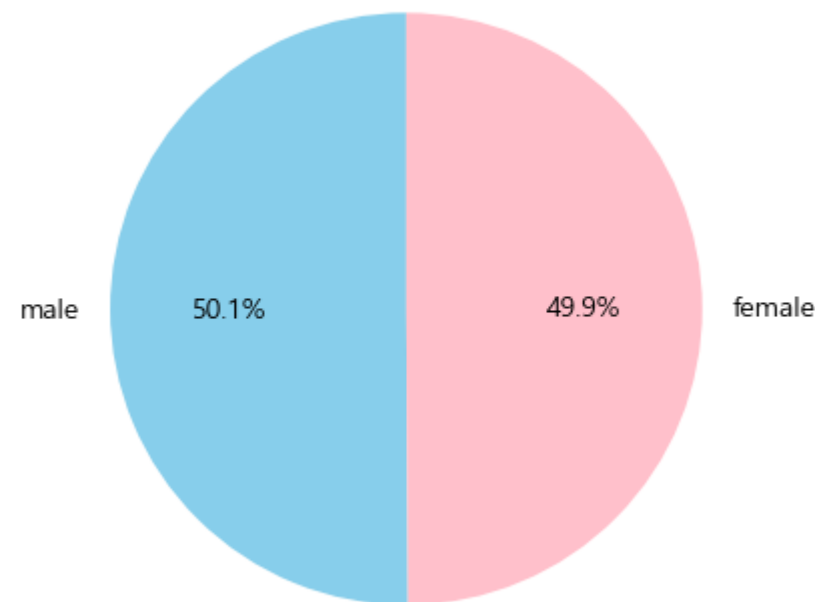
- 01. 우리 동네 인구 구조 시각화하기
- 02. 인구 구조를 다양한 형태로 시각화하기
- 03. 우리 동네 인구 구조를 파이 차트로 나타내기

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ 제주도의 연령대별 성별 비율은?



제주특별자치도 지역의 성별 비율



원 그래프로 비율을 확인함!

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

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ① 꺾은선 그래프로 표현하기 (1/3)

- 연령대별 성별 인구수를 꺾은선 그래프로 시각화하기

```
1 import csv
2 import matplotlib.pyplot as plt
3 f = open("gender.csv", encoding="cp949")
4 data = csv.reader(f)
5 header = next(data)
6
7 name = input("찾고 싶은 지역의 이름을 알려주세요: ")
8 male, female = [], []
9 for row in data:
10     if name in row[0]:
11         for j in range(3, 104, 1):
12             male.append(int(row[j].replace(',', '')))
13         for j in range(106, 207, 1):
14             female.append(int(row[j].replace(',', '')))
15         break
16
17 f.close()
18
19
```


04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ① 꺾은선 그래프로 표현하기 (2/3)

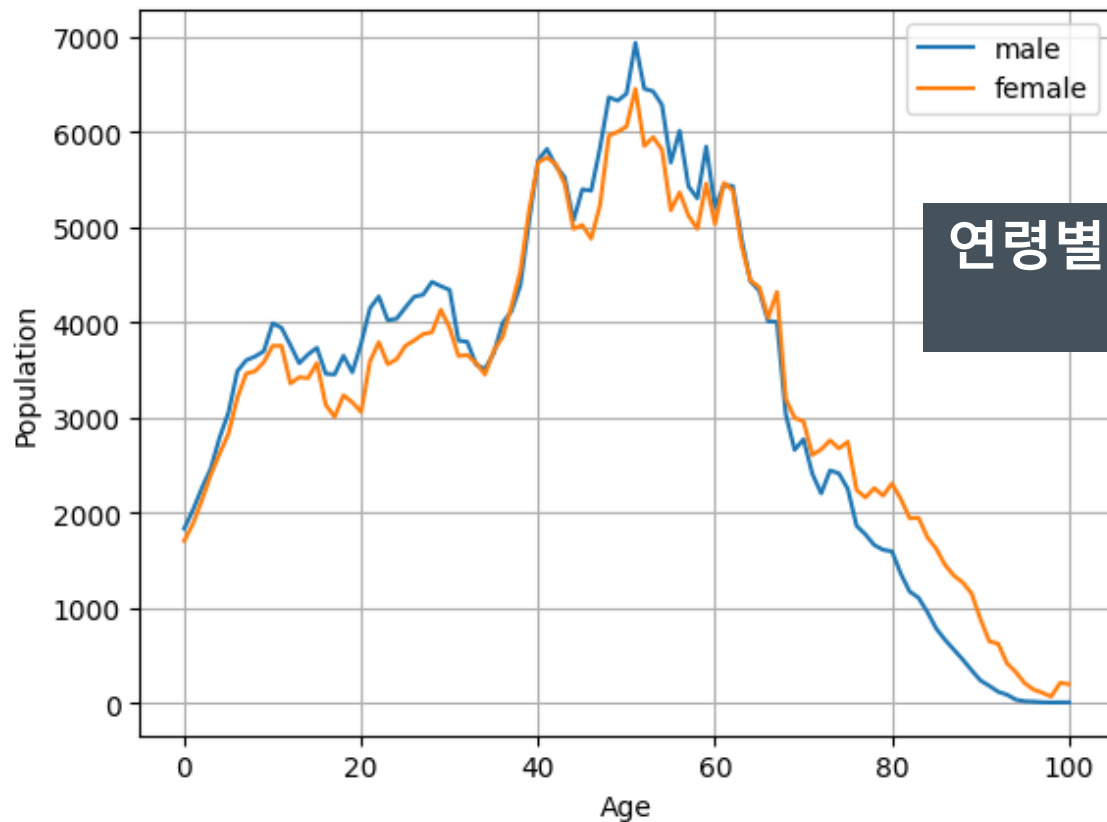
```
20 plt.figure()
21 plt.plot(male, label="male")
22 plt.plot(female, label="female")
23 plt.xlabel("Age")
24 plt.ylabel("Population")
25 plt.legend()
26 plt.grid()
27 plt.show()
```

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ① 꺾은선 그래프로 표현하기 (3/3)

실행결과

찾고 싶은 지역의 이름을 알려주세요: 제주특별자치도



연령별로 남성 인구수에서 여성 인구수를 뺀 차이를 막대그래프로 표현하면 어떨까?

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ② 막대그래프로 표현하기 (1/3)

- 연령별 '남성 인구수 - 여성 인구수' 결과를 막대그래프로 가시화하기

```
1 import csv
2 import matplotlib.pyplot as plt
3 f = open("gender.csv", encoding="cp949")
4 data = csv.reader(f)
5 header = next(data)
6
7 name = input("찾고 싶은 지역의 이름을 알려주세요: ")
8 diff = []
9 for row in data:
10     if name in row[0]:
11         for j in range(3, 104, 1):
12             diff.append(int(row[j].replace(',','')) -
13                          int(row[j+103].replace(',','')))
14         break
15
16 f.close()
17
18
19
```

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ② 막대그래프로 표현하기 (2/3)

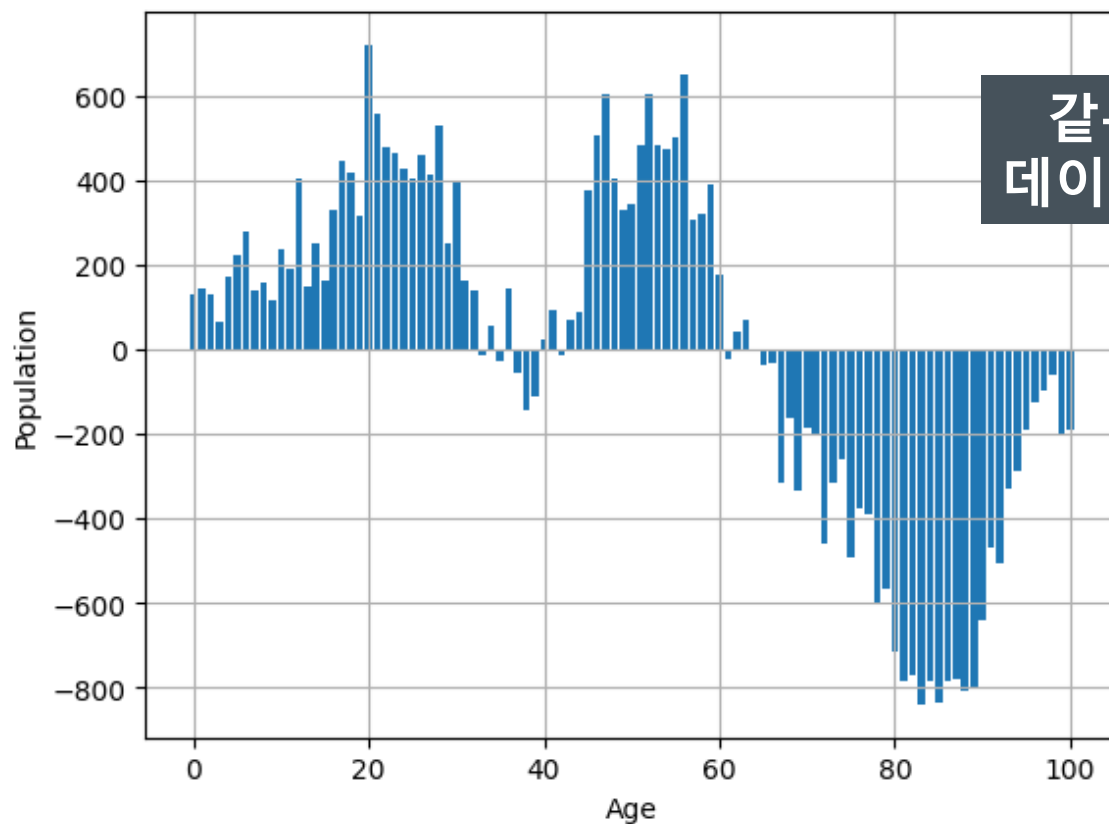
```
20 plt.figure()
21 plt.bar(range(len(diff)), diff)
22 plt.xlabel("Age")
23 plt.ylabel("Population")
25 plt.grid()
26 plt.show()
```

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ② 막대그래프로 표현하기 (3/3)

실행결과

찾고 싶은 지역의 이름을 알려주세요: 제주특별자치도



같은 데이터라도 다양한 형태로 시각화하면
데이터에 숨겨진 다양한 사실을 발견할 수 있음

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ③ 산점도(Scatter Plot)로 표현하기

- 산점도는 가로축과 세로축을 기준으로 두 요소가 서로 어떤 관계를 맺고 있는지를 파악하기 쉽게 나타낸 그래프임
- 예를 들어,
 - ◆ 가로축을 제주도 지역의 남성 인구수
 - ◆ 세로축을 제주도 지역의 여성 인구수

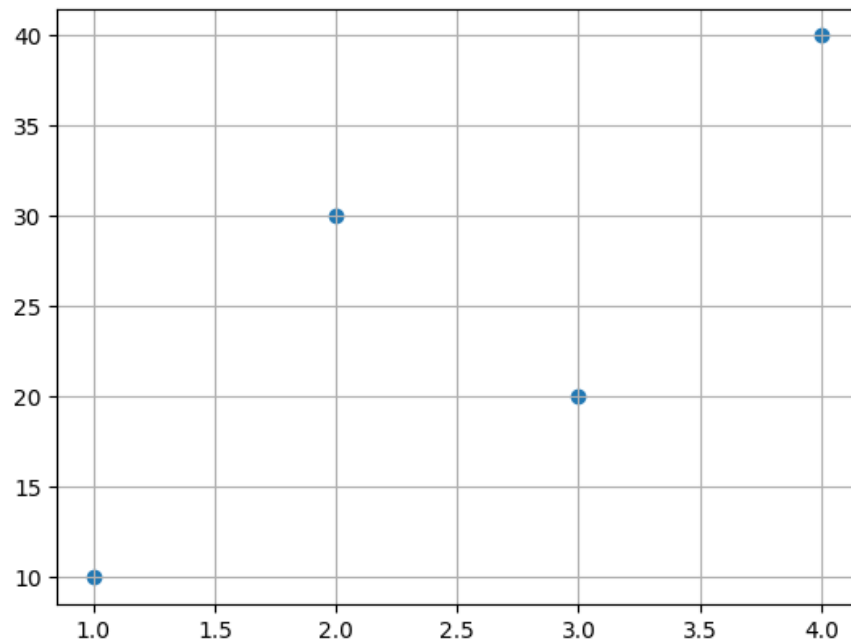
**scatter() 함수와 colorbar() 함수를 활용하여
실습을 진행해 보자**

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ④ scatter() 함수로 표현하기

```
1 import matplotlib.pyplot as plt
2 plt.figure()
3 plt.scatter([1, 2, 3, 4], [10, 30, 20, 40])
4 plt.grid()
5 plt.show()
```

실행결과



(1, 10), (2, 30), (3, 20), (4, 40) 좌표에
점이 찍히게 됨

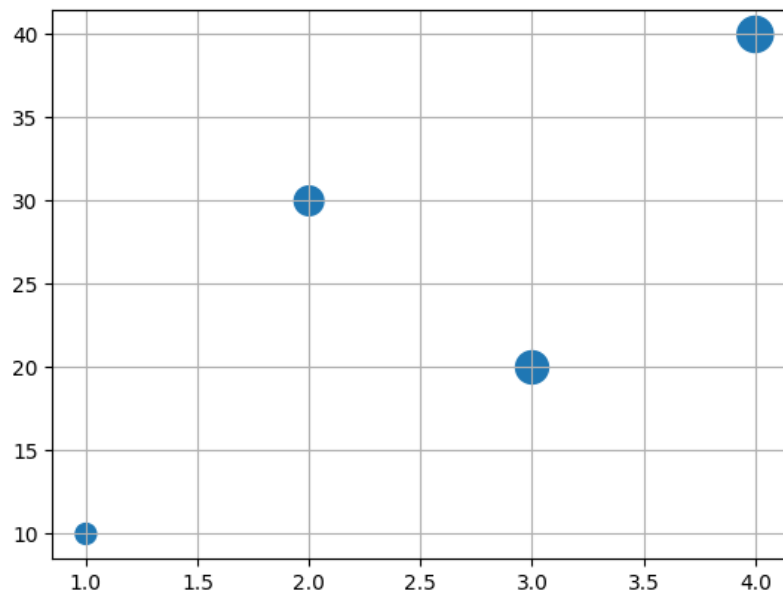
04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑤ 버블 차트(Bubble Chart)로 표현하기 (1/11)

- 점(Point)의 크기 지정하기

```
1 import matplotlib.pyplot as plt
2 plt.figure()
3 plt.scatter([1, 2, 3, 4], [10, 30, 20, 40], s=[100, 200, 250, 300])
4 plt.grid()
5 plt.show()
```

실행결과



- ✓ s 속성은 Size를 의미함
- ✓ 버블의 원하는 크기를 s 속성에 지정함

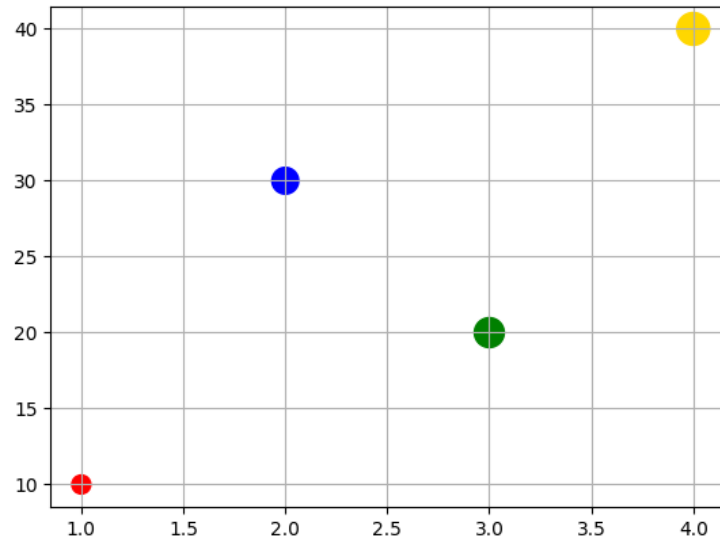
04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑤ 버블 차트(Bubble Chart)로 표현하기 (2/11)

- 버블의 색상 지정하기

```
1 import matplotlib.pyplot as plt
2 plt.figure()
3 plt.scatter([1, 2, 3, 4], [10, 30, 20, 40], s=[100, 200, 250, 300],
4             c=["red", "blue", "green", "gold"])
5 plt.grid()
6 plt.show()
```

실행결과



- ✓ c 속성은 표현하고 싶은 색상의 수를 의미함
- ✓ 각 버블의 원하는 색상을 c 속성에 지정해 줄 수 있음

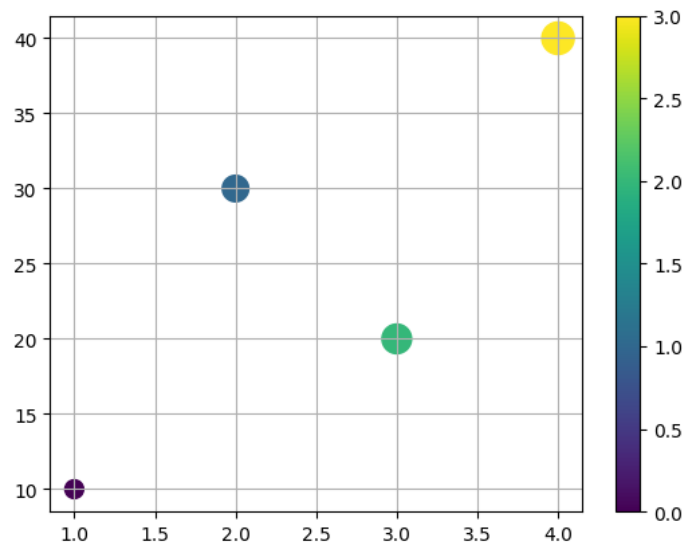
04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑤ 버블 차트(Bubble Chart)로 표현하기 (3/11)

- 컬러바(Color Bar) 추가하기

```
1 import matplotlib.pyplot as plt
2 plt.figure()
3 plt.scatter([1, 2, 3, 4], [10, 30, 20, 40], s=[100, 200, 250, 300], c=range(4))
4 plt.colorbar()
5 plt.grid()
6 plt.show()
```

실행결과



- ✓ colorbar() 함수를 사용하면 그래프 오른쪽에 컬러바가 추가됨
- ✓ 4가지 색상을 사용한다는 의미로 c=range(4)라고 수정함
- ✓ 각 데이터는 정해진 색상 개수에 따라 해당하는 컬러바의 색을 부여 받음

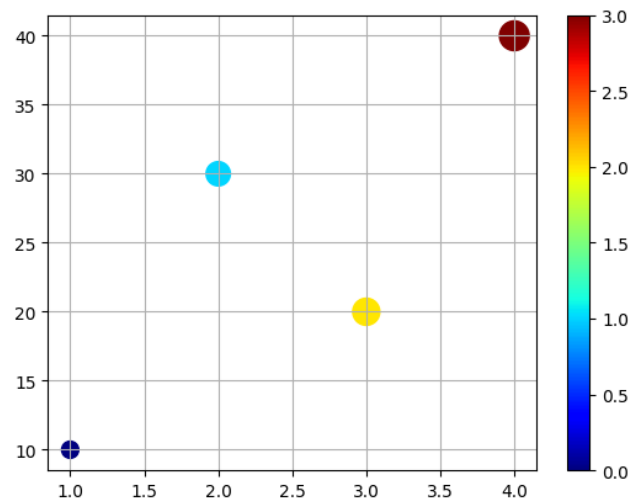
04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑤ 버블 차트(Bubble Chart)로 표현하기 (4/11)

- 컬러바에 사용될 색상 지정하기

```
1 import matplotlib.pyplot as plt
2 plt.figure()
3 plt.scatter([1, 2, 3, 4], [10, 30, 20, 40], s=[100, 200, 250, 300], c=range(4),
4             cmap="jet")
5 plt.colorbar()
6 plt.grid()
7 plt.show()
```

실행결과

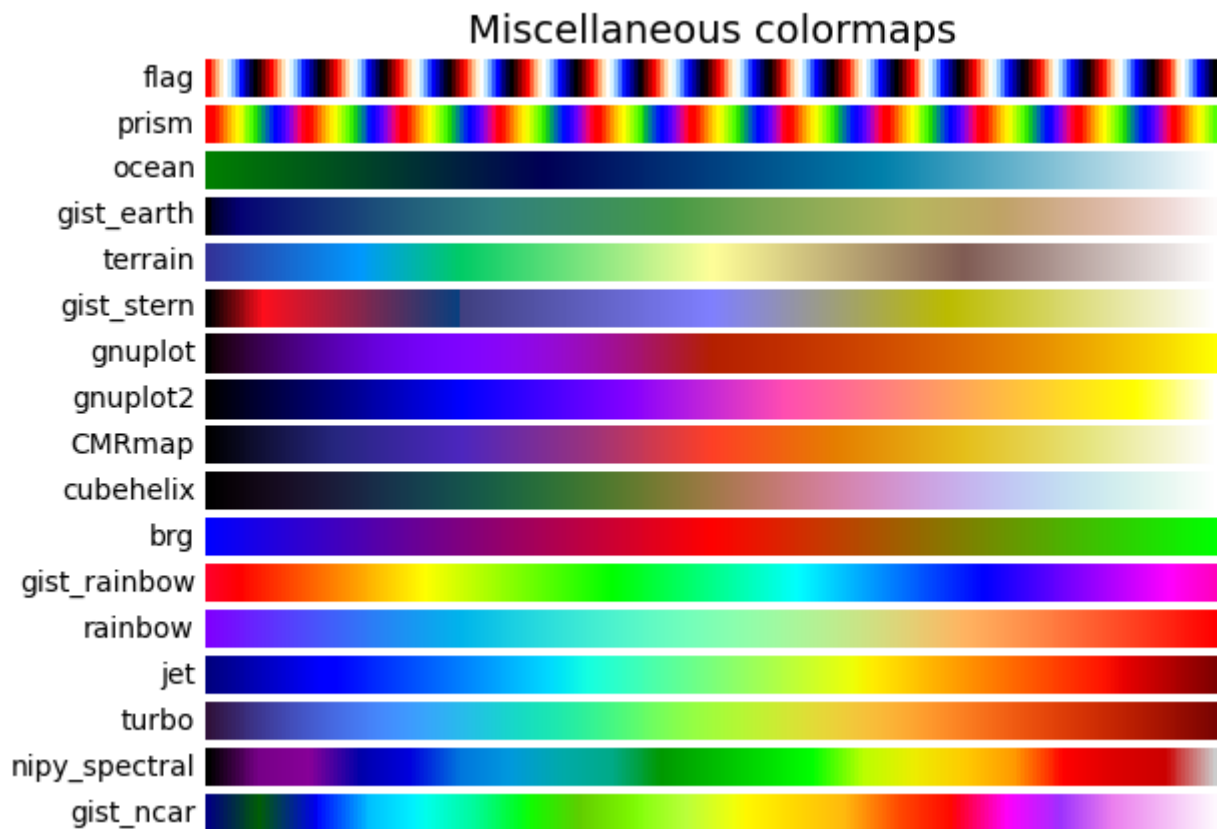


- ✓ cmap 속성은 Color Map을 의미함
- ✓ 컬러바에 사용될 색상의 종류를 지정할 수 있음

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑤ 버블 차트(Bubble Chart)로 표현하기 (5/11)

- matplotlib 라이브러리에서 사용할 수 있는 다양한 컬러맵의 종류
- [URL] <https://matplotlib.org/stable/tutorials/colors/colormaps.html>



이외에도 다양한 컬러맵이 있음

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑤ 버블 차트(Bubble Chart)로 표현하기 (6/11)

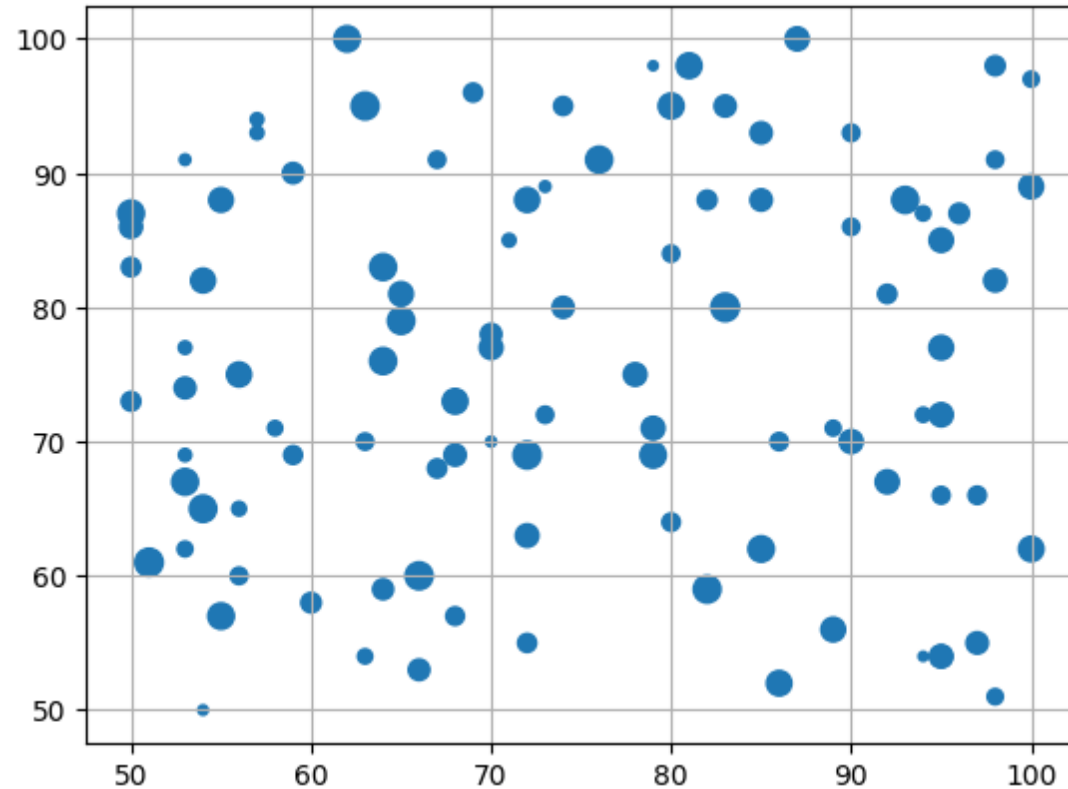
- 위치, 크기가 서로 다른 100개의 점을 산점도로 시각화하기

```
1 import matplotlib.pyplot as plt
2 import random
3
4 x = []
5 y = []
6 size = []
7 for j in range(100):
8     x.append(random.randint(50, 100))
9     y.append(random.randint(50, 100))
10    size.append(random.randint(10, 100))
11
12 plt.figure()
13 plt.scatter(x, y, s=size)
14 plt.grid()
15 plt.show()
```

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑤ 버블 차트(Bubble Chart)로 표현하기 (7/11)

실행결과



04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑤ 버블 차트(Bubble Chart)로 표현하기 (8/11)

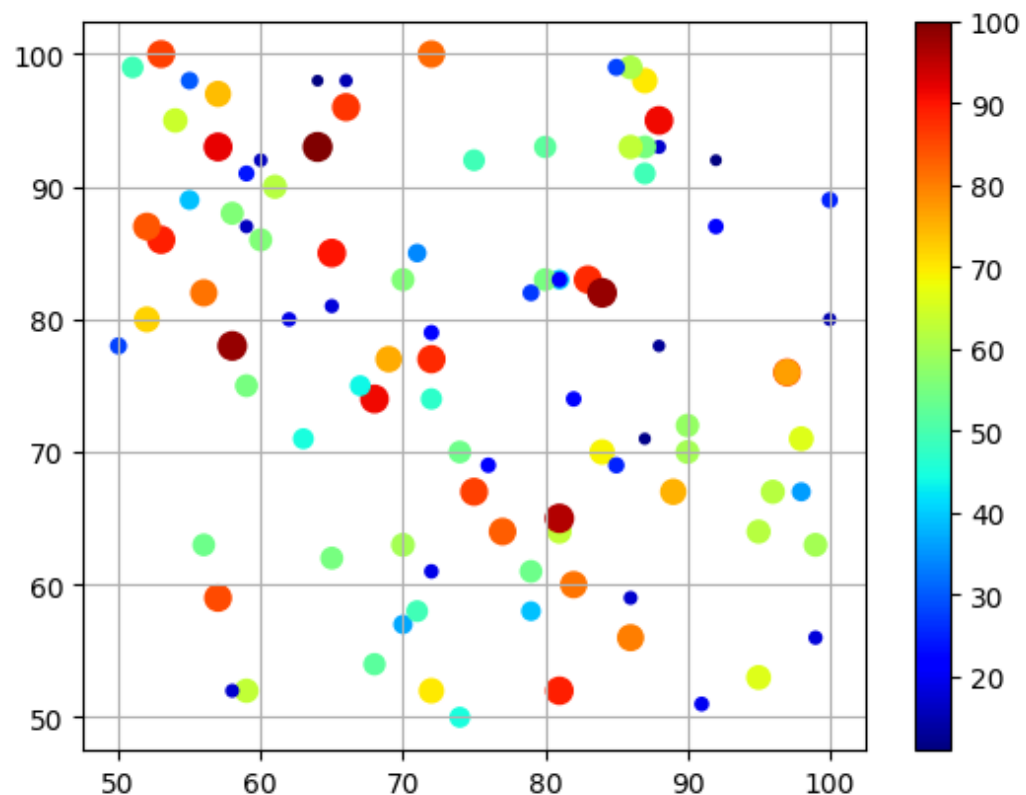
- 컬러맵, 컬러바 추가하기

```
1 import matplotlib.pyplot as plt
2 import random
3
4 x = []
5 y = []
6 size = []
7 for j in range(100):
8     x.append(random.randint(50, 100))
9     y.append(random.randint(50, 100))
10    size.append(random.randint(10, 100))
11
12 plt.figure()
13 plt.scatter(x, y, s=size, c=size, cmap="jet")
14 plt.colorbar()
15 plt.grid()
16 plt.show()
```

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑤ 버블 차트(Bubble Chart)로 표현하기 (9/11)

실행결과



작은 점들이 큰 점에 가려서 잘 안 보임.
어떻게 해야 할까?

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑤ 버블 차트(Bubble Chart)로 표현하기 (10/11)

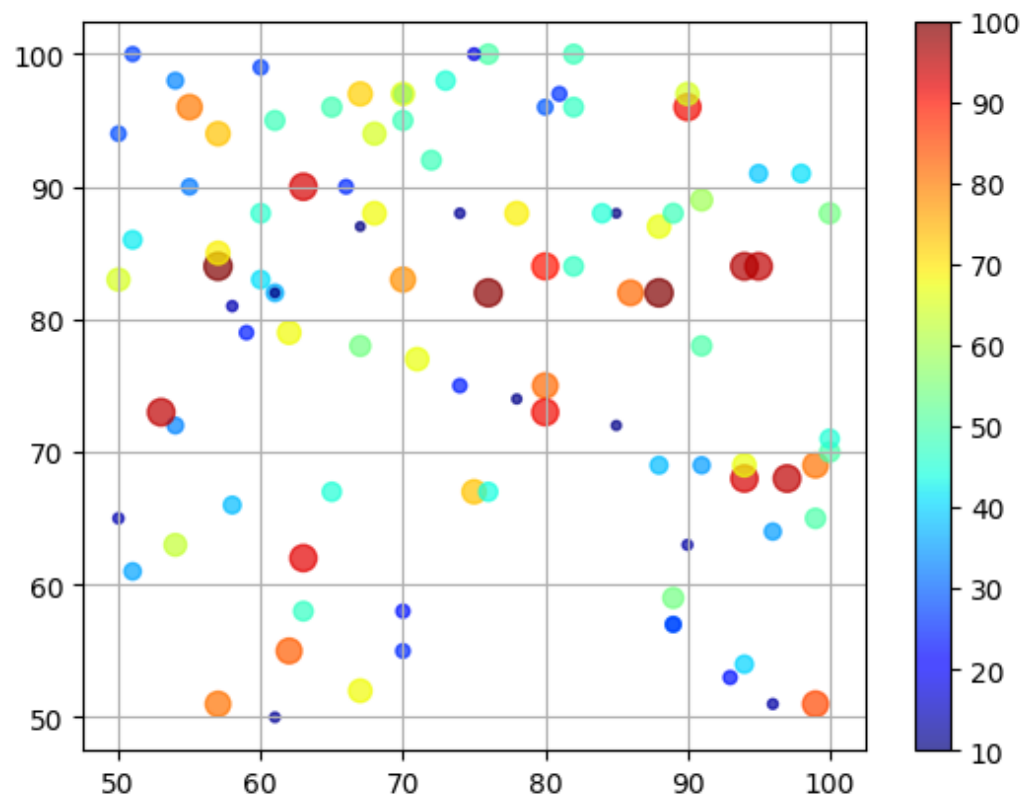
- 버블의 투명도 설정하기

```
1 import matplotlib.pyplot as plt
2 import random
3
4 x = []
5 y = []
6 size = []
7 for j in range(100):
8     x.append(random.randint(50, 100))
9     y.append(random.randint(50, 100))
10    size.append(random.randint(10, 100))
11
12 plt.figure()
13 plt.scatter(x, y, s=size, c=size, cmap="jet", alpha=0.7)
14 plt.colorbar()
15 plt.grid()
16 plt.show()
```

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑤ 버블 차트(Bubble Chart)로 표현하기 (11/11)

실행결과



04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑥ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (1/9)

- 연령별 남성 및 여성 인구수를 scatter() 함수로 시각화하기

```
1 import matplotlib.pyplot as plt
2 import random
3
4 f = open("gender.csv", encoding="cp949")
5 data = csv.reader(f)
6 next(data)
7
8 male = []
9 female = []
10 name = input("궁금한 동네를 입력해주세요: ")
11 for row in data:
12     if name in row[0]:
13         for j in range(3, 104, 1):
14             male.append(int(row[j].replace(',', '')))
15             female.append(int(row[j+103].replace(',', '')))
16         break
17
18 f.close()
19
```

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑥ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (2/9)

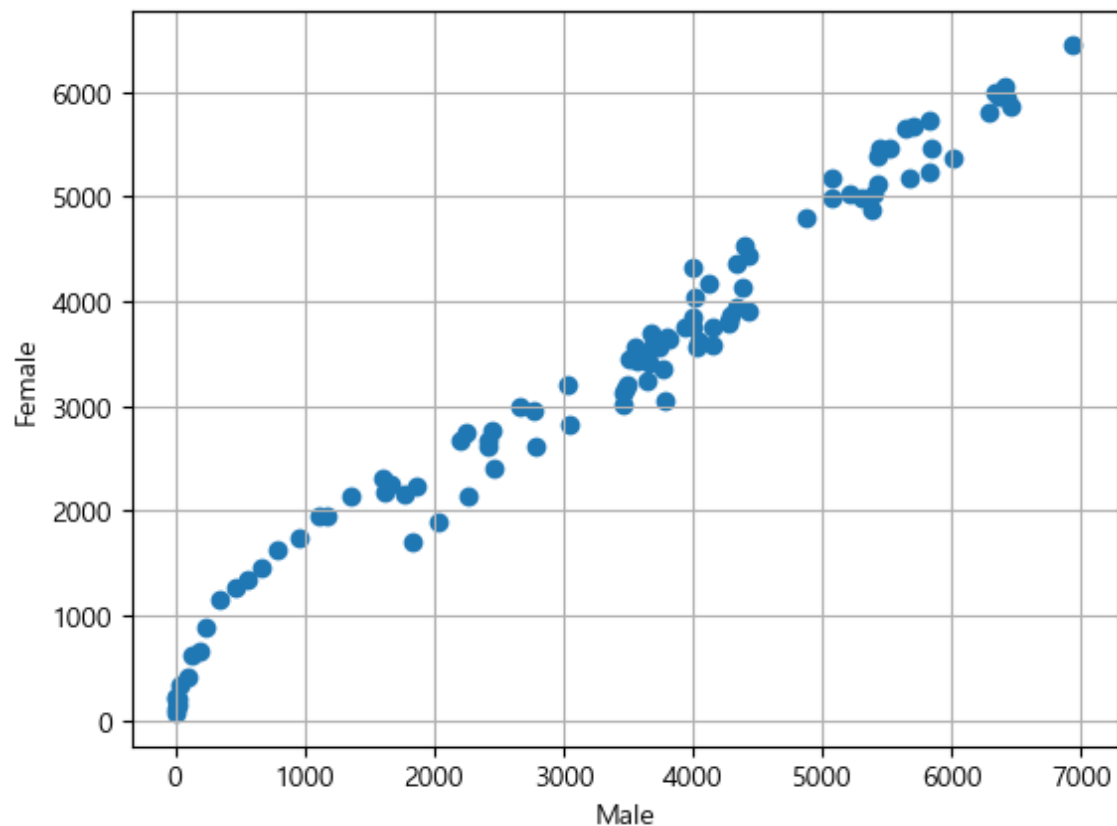
```
20 plt.figure()
21 plt.scatter(male, female)
22 plt.xlabel("Male")
23 plt.ylabel("Female")
24 plt.grid()
25 plt.show()
```

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑥ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (3/9)

실행결과

궁금한 동네를 입력해주세요: 제주특별자치도



04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑥ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (4/9)

- 컬러바, 추세선 추가하기

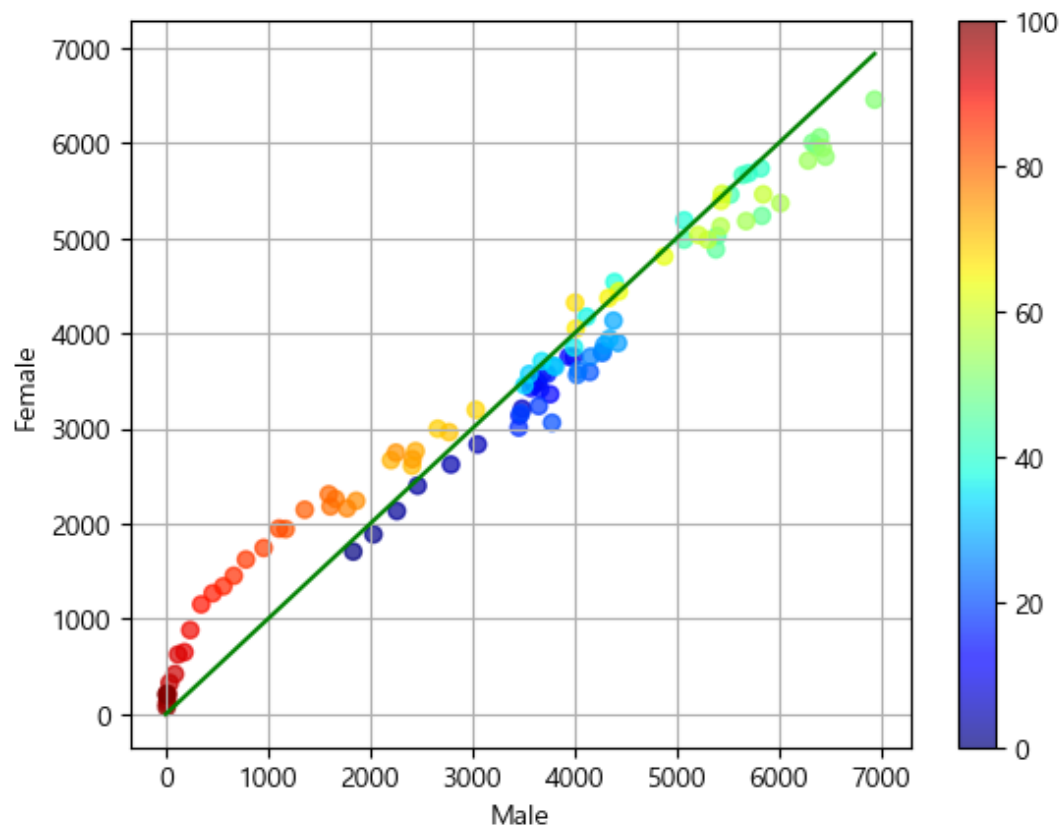
```
20 plt.figure()
21 plt.scatter(male, female, c=range(len(male)), cmap="jet", alpha=0.7)
22 plt.colorbar()
23 plt.plot(range(max(male)), range(max(male)), 'g')
24 plt.xlabel("Male")
25 plt.ylabel("Female")
26 plt.grid()
27 plt.show()
```

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑥ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (5/9)

실행결과

궁금한 동네를 입력해주세요: 제주특별자치도



04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑥ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (6/9)

- 버블의 크기 조정하기

```
1 import matplotlib.pyplot as plt
2 import random
3 import math
4 f = open("gender.csv", encoding="cp949")
5 data = csv.reader(f)
6 next(data)
7
8 male, female, size = [], [], []
9 name = input("궁금한 동네를 입력해주세요: ")
10 for row in data:
11     if name in row[0]:
12         for j in range(3, 104, 1):
13             male.append(int(row[j].replace(',', '')))
14             female.append(int(row[j+103].replace(',', '')))
15             size.append(math.sqrt(int(row[j].replace(',', '')) +
16                                   int(row[j+103].replace(',', ''))))
17         break
18
19 f.close()
```


04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑥ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (7/9)

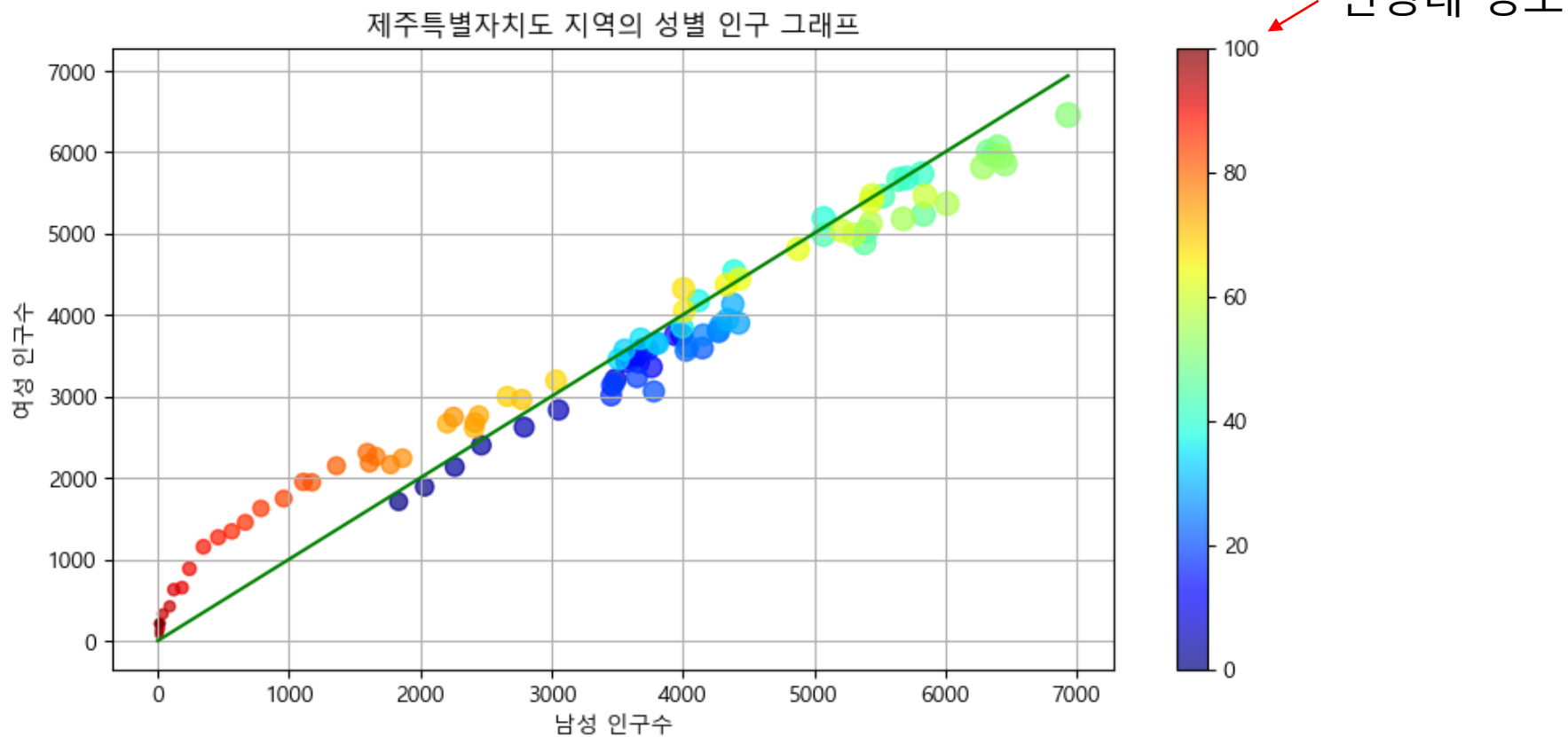
```
20 plt.rc("font", family="Malgun Gothic")
21 plt.figure(figsize=(10, 5))
22 plt.title(name + " 지역의 성별 인구 그래프")
23 plt.scatter(male, female, c=range(len(male)), cmap="jet", alpha=0.7, s=size)
24 plt.colorbar()
25 plt.plot(range(max(male)), range(max(male)), 'g')
26 plt.xlabel("남성 인구수")
27 plt.ylabel("여성 인구수")
28 plt.grid()
29 plt.show()
```

04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑥ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (8/9)

실행결과

궁금한 동네를 입력해주세요: 제주특별자치도

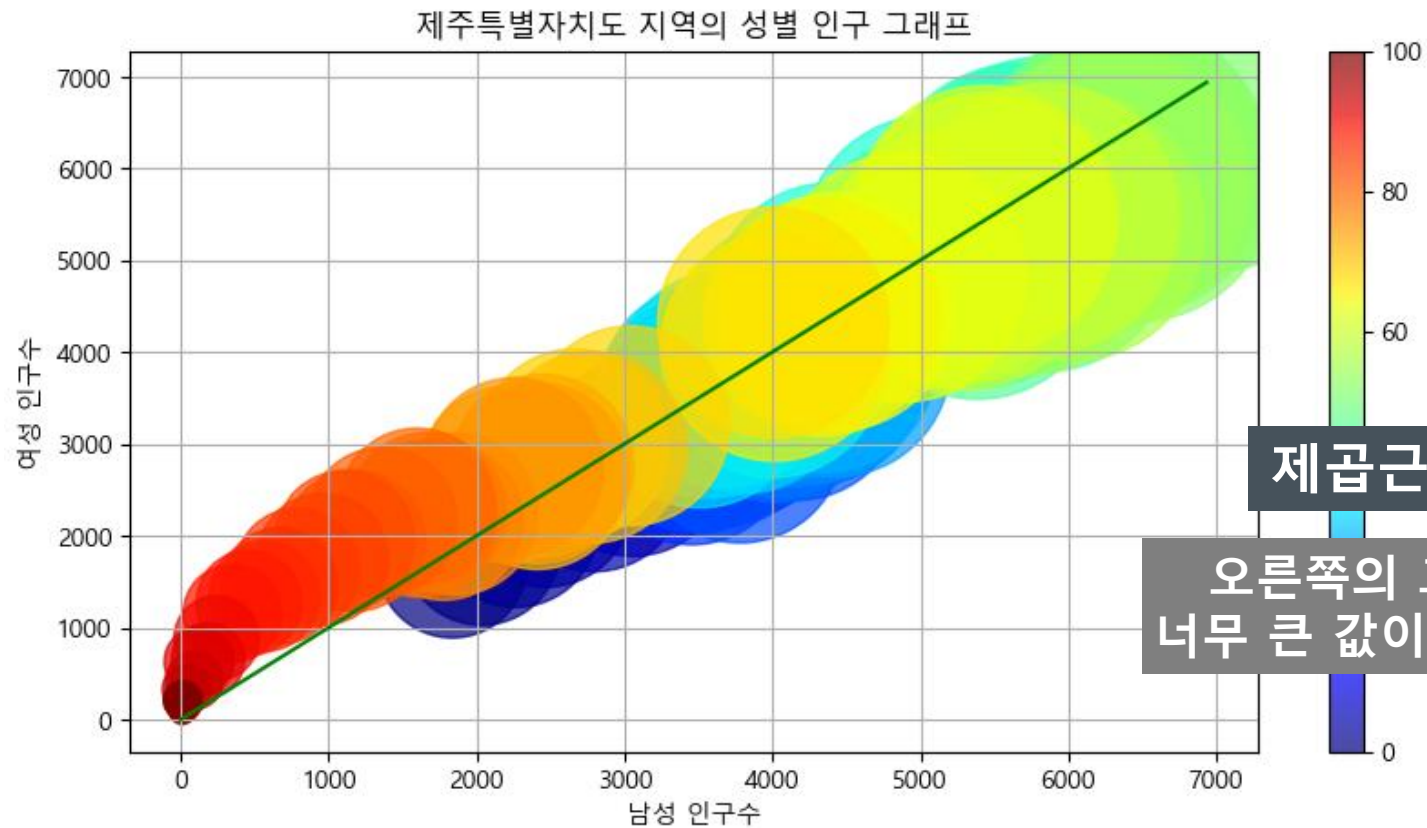


04. 우리 동네 인구 구조를 산점도로 나타내기

❖ ⑥ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (9/9)

실행결과

궁금한 동네를 입력해주세요: 제주특별자치도



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

THANK YOU!

Q & A

- Name: 권범
- Office: 동덕여자대학교 인문관 B821호
- Phone: 02-940-4752
- E-mail: bkwon@dongduk.ac.kr