





- ✓ Vs code 노트북 소스 활용
- ☑ 인구구조 연령별 성별 분석





- ☑ Vs code에서 노트북을 편집하고 실행할 수 있다.
- ☑ 임의의 행정구역의 인구구조 연령별 성별 분석을 위해 산점도를 그릴 수 있다.
- 산점도에서 연령별로 colorbar를 지정해 그릴 수 있다.
- ☑ 산점도에서 colorbar를 통해 연령별 분포를 이해할 수 있다.

LESSON 01

VS code 노트북 활용

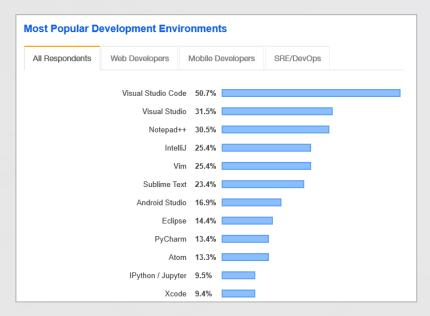




→ Vs code 인기

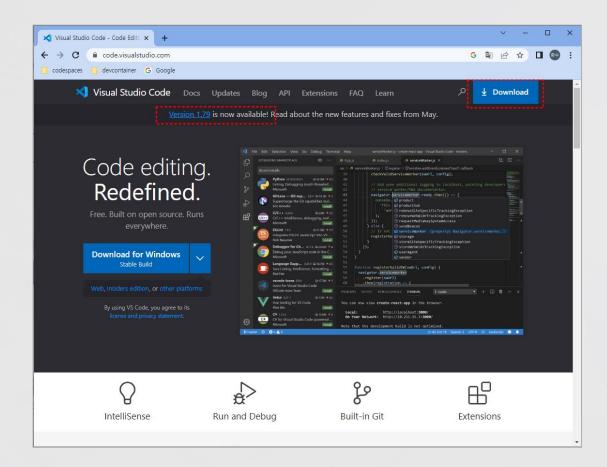


- ☑ 장점
 - ◆ 다음을 함께 코딩 가능
 - *.py, *.ipynb





→ https://code.visualstudio.com/





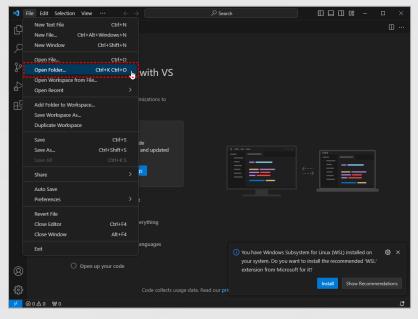
→ 이전에 설치가 있는 경우



- ☑ 이전에 설정된 모든 설정을 삭제하고자 하는 경우
- ☑ 설정 연습을 위해 설정을 삭제하고자 하는 경우
- ☑ 다음 폴더 삭제
 - + C:₩Users₩[로그인계정]₩.vscode
 - + C:₩Users₩[로그인계정]₩AppData₩Roaming₩Code



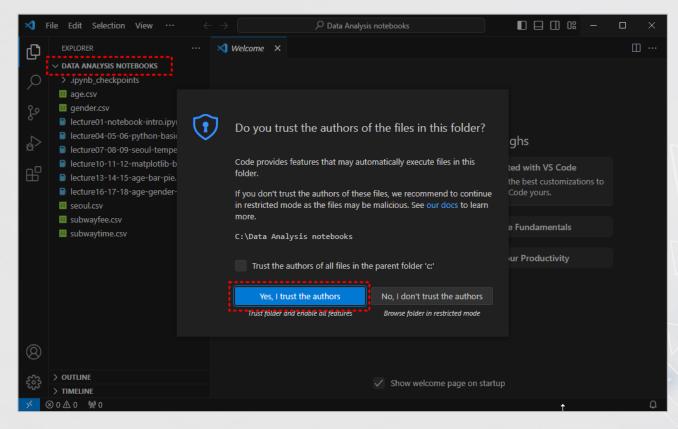
- ⊸ Vs code 실행
- ₩뉴 File | Open Folder
 - Ctrl + K Ctrl + O
 - ☑ 편집할 *.ipynb 파일이 저장될 파일 폴더 선택
 - ◆ C:₩Data Analysis notebooks





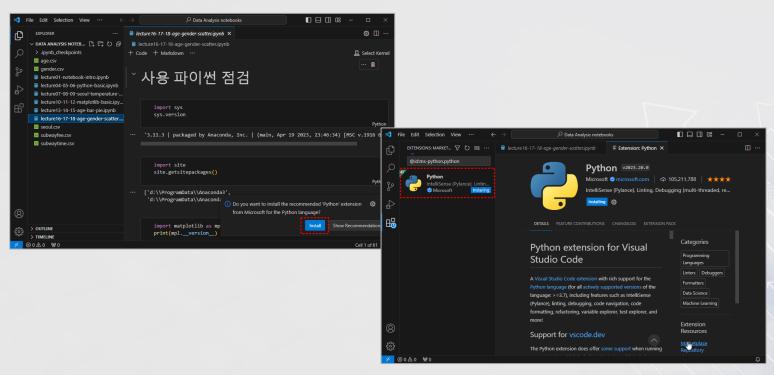
→ 폴더 열기

→ 노트북 소스 파일 생성 또는 기존 파일 열기





- ⊸ 확장 설치
- ≥ 노트북을 실행할 확장(extensions) 설치
 - ☑ 확장 python

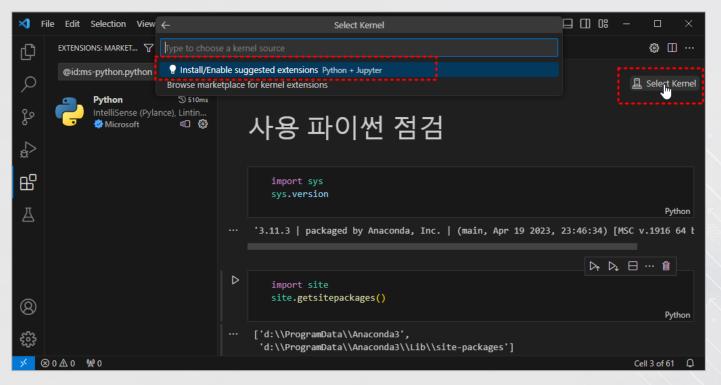




→ 커닐 선택



🔽 파일 별 선택

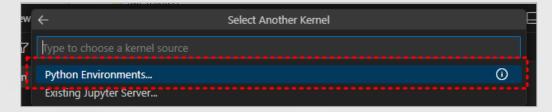




→ 커널 선택 계속



- Python Environment
- Anaconda3



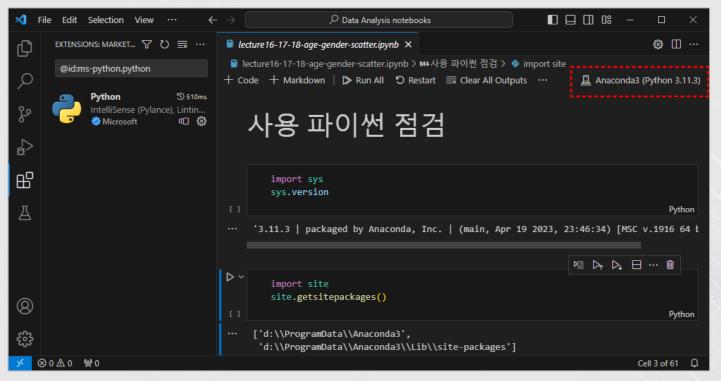
Change kernel for 'lecture16-17-18-age-gender-scatter.ipynb'

Anaconda3 (Python 3.11.3) d:\ProgramData\Anaconda3\python.exe - Currently Selected

Select Another Kernel...



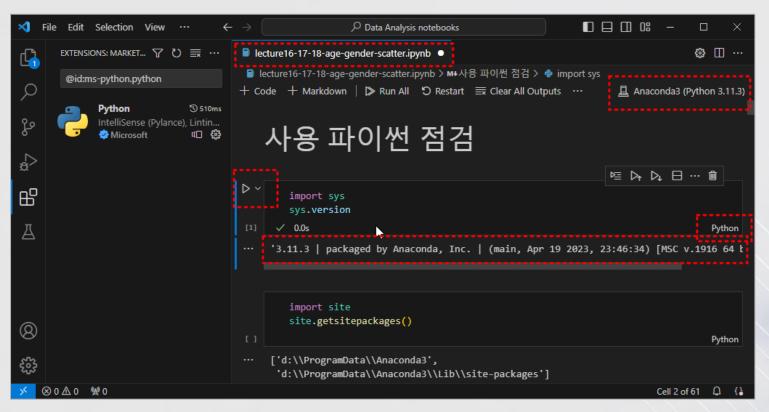
- → 파일 열때마다 커널 선택
- **반드시 Anaconda3 선택**
 - 🔽 물론 표준 파이썬에도 패키지 matplotlib, pandas, numpy 설치되었다면 가능





⊸ 노트북 실행

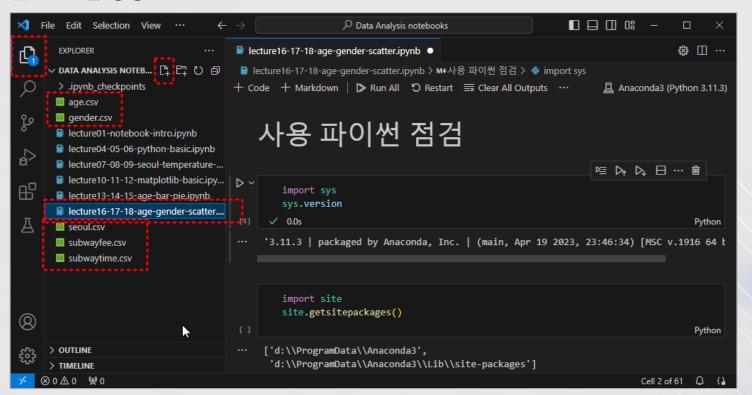
Shift + Enter



⊸ 탐색기 보기

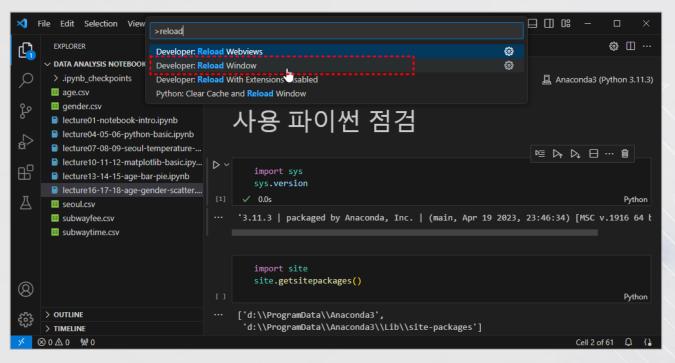


🔽 새 파일 생성





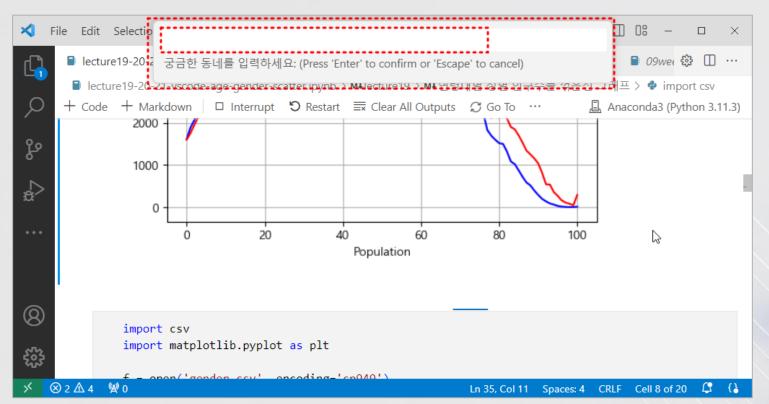
- → 문법 오류 아닌 경우
- ❤️ 문제가 생기면(주로 matplotlib에서 png 파일 오류)
 - ☑ 메뉴 View | Command Palette (Ctrl + Shift + P)
 - >reload





⊸ 표준 입력

❷ 명령 팔레트 부분에서 입력



LESSON 02

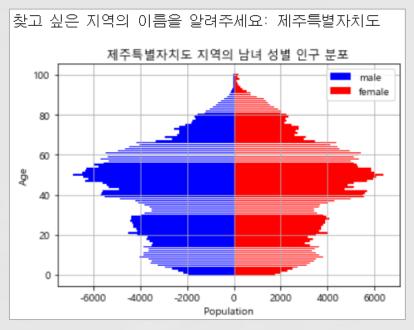
인구구조 연령별 성별 산점도 분석

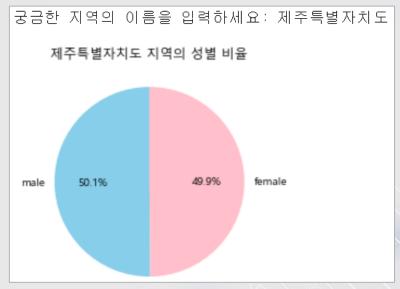


☑ 인구구조 연령별 성별 산점도 분석



→ 제주도의 연령대별 성별 비율은?





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

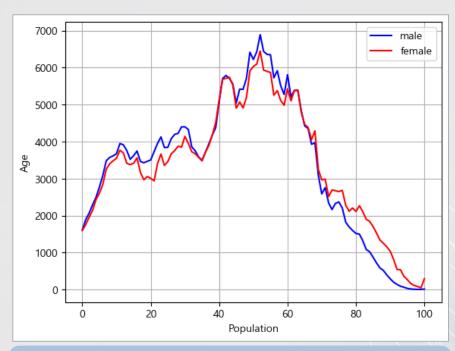


⊸ 꺾은선 그래프로 표현



▶ 연령대별 성별 인구수를 컦은선 그래프로 시각화하기

```
import csv
import matplotlib.pyplot as plt
f = open('gender.csv', encoding='cp949')
data = csv.reader(f)
header = next(data)
male = []
female = []
name = input('궁금한 동네를 입력하세요: ')
# 제주특별자치도
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.rc('font', family ='Malgun Gothic')
plt.plot(male, label='male', color='b')
plt.plot(female, label='female', color='r')
plt.xlabel('Population')
plt.vlabel('Age')
plt.grid(True)
plt.legend()
plt.show()
```



연령별로 남성 인구수에서 여성 인구수를 뺀 차이를 막대그래프로 표현하면 어떨까요?



→ 막대그래프로 표현하기 (1/2)

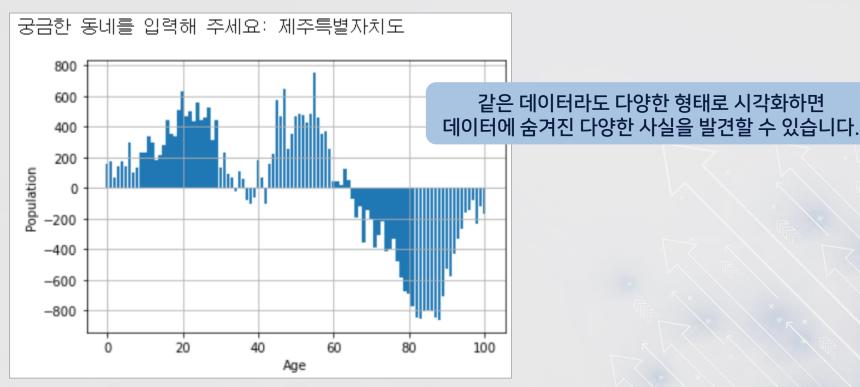


```
import csv
import matplotlib.pyplot as plt
f = open('gender.csv', encoding='cp949')
data = csv.reader(f)
next (data)
diff = []
name = input('궁금한 동네를 입력해 주세요: ')
for row in data:
    if name in row[0]:
        for i in range(3, 104):
           diff.append(int(row[i].replace(',', "")) - int(row[i + 103].replace(',', "")))
       break
f.close()
plt.bar(range(len(diff)), diff)
plt.xlabel('Age')
plt.ylabel('Population')
plt.grid(True)
plt.show()
```



⊸ 막대그래프로 표현하기 (2/2)

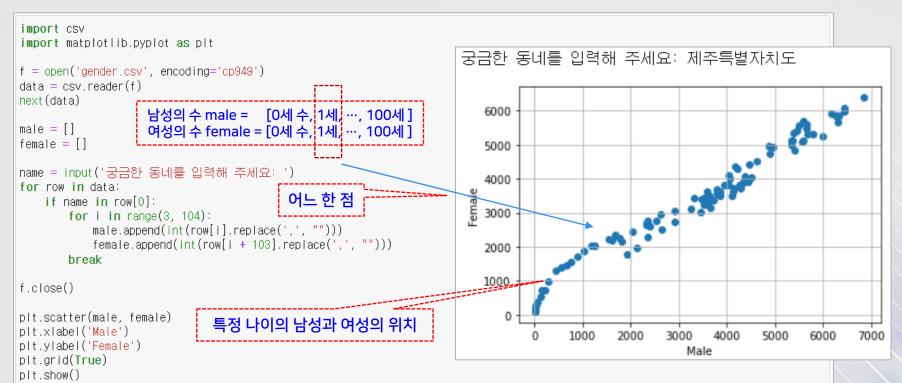
② 연령별 남성 인구수 ─ 여성 인구수 결과를 막대그래프로 가시화하기(실행결과)





→ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (1/6)

❷ 연령별 남성 및 여성 인구수를 scatter() 함수로 시각화하기





→ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (2/6)

🤪 컬러바, 추세선 추가하기

plt.show()

```
import csv
import matplotlib.pyplot as plt
f = open('gender.csv', encoding='cp949')
                                                                    궁금한 동네를 입력해 주세요: 제주특별자치도
data = csv.reader(f)
next (data)
                                                                        7000
male = []
female = []
                                                                        6000
name = input('궁금한 동네를 입력해 주세요: ')
                                                                        5000
for row in data:
   if name in row[0]:
                                                                                                                             60
                                                                        4000
                                                                     Female
       for i in range(3, 104):
          male.append(int(row[i].replace(',', "")))
                                                                        3000
          female.append(int(row[i + 103].replace('.'. "")))
                                                                                                                             40
       break
                               색상을 나이로 지정하면,
                                                                        2000
                  여성이 남성을 초과하는 연령이 노년층임을 알 수 있음
f.close()
                                                                                                                            - 20
                                                                       1000
plt.scatter(male, female, c=range(len(male)), cmap='jet', alpha=0.7)
plt.colorbar()
plt.plot(range(max(male)), range(max(male)), 'g')
                                                                                             3000 4000 5000 6000 7000
                                                                                        2000
plt.xlabel('Male')
                                                                                                Male
plt.ylabel('Female')
plt.grid(True)
```



→ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (3/6)

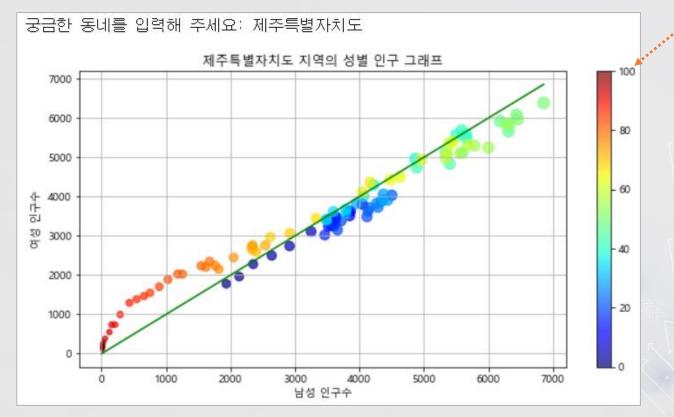


```
import csv
import matplotlib.pyplot as plt
import math
f = open('gender.csv', encoding='cp949')
data = csv.reader(f)
next (data)
male = []
female = []
size = []
name = input('궁금한 동네를 입력해 주세요: ')
for row in data:
    if name in row[0]:
        for i in range(3, 104):
            male.append(int(row[i].replace(',', "")))
           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()
```



→ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (4/6)

❤️ 버블의 크기 조정하기(실행결과)

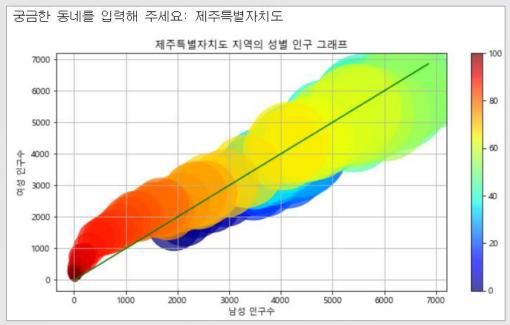


연령대 정보



→ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (5/6)





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

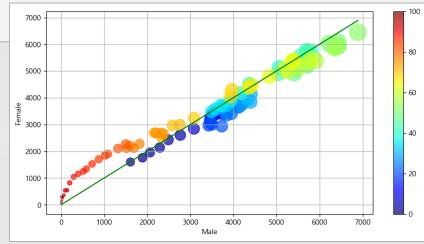


→ 제주도의 연령대별 성별 비율을 산점도로 표현하기 (6/6)

❤️ 버블의 크기를 적당히 산정

```
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((int(row[i].replace(',', '')) + int(row[i+103].replace(',', '')))/25)
            break
```

f.close()



SUMMARY

학습정긴

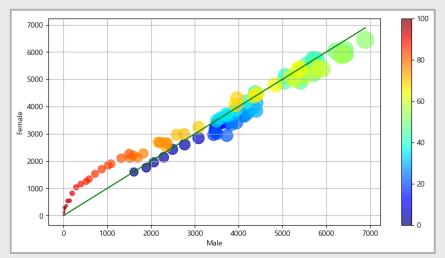




•••

Matplotlib의 scatter()

- ≫ x, y의 좌표에 마커(원) 그리기
- ≫ s=원크기, c=색상지정, cmap=색상맵지정
- >> colorbar()









•••

♥ 행정구역의 연령별 성별 인구로 꺽은선

>> 연령별 성별 차이 분석

🧿 행정구역의 연령별 성별 분석을 위한 산점도

- ▶ 각 연령별 x: 남성, y: 여성
- >> y=x의 선상 위: 여자가 다수, 선상 아래: 남자가 다수
- >> 버블 크기: 각 연령별 남녀 인구 합
- >> colorbar()의 색상이 연령 의미



