

서울시 청소년 정신건강 분석 데이터 분석

Auto Mobile Robot

Exported on 02/22/2024

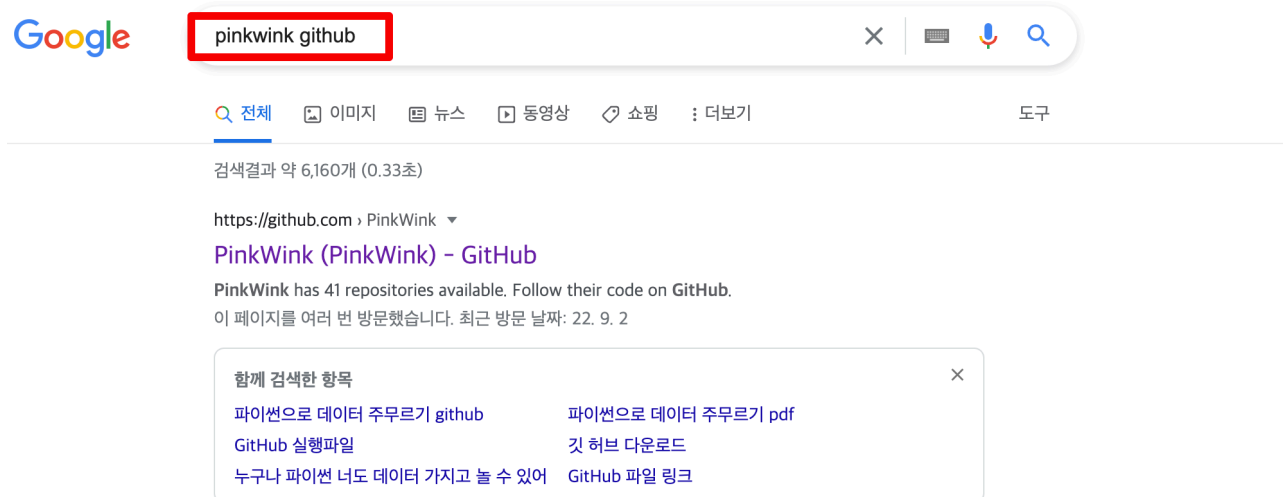
Table of Contents

1	첫 시작 - 서울시 청소년 정신건강 분석.....	4
1.1	데이터 얻으러~ pinkwink github	4
1.2	github 사이트에서 repository	4
1.3	거기서 data 폴더 선택	5
1.4	거기서 teenage_mental.xls 클릭	5
1.5	그리고 나타난 페이지에서 view_raw라는 링크에서 “주소 복사”	6
1.6	이제 jupyter notebook을 켜봅시다	6
1.7	그리고, new에 python3를 선택합니다.	6
1.8	url 이라는 변수에	7
1.9	복사한 링크를 붙여 넣기	7
1.10	SHIFT+ENTER를 누르면 실행됨	7
1.11	pandas라는 데이터 정리용 모듈을 이용해서 excel 파일을 읽어볼수 있다	7
1.12	본래 파일을 엑셀로 열어보면 이렇게 생겼다	8
1.13	그런데 읽은 결과가 마음에 들지 않는다	8
1.14	이유는	8
1.15	일단 행 범위 선택	8
1.16	데이터의 범위는 잡은듯	9
1.17	이렇게 컬럼 이름을 바꾸면 좋을 것 같다	9
2	데이터 정리하기	10
2.1	방금까지의 내용을 raw_data에 저장하자	10
2.2	loc 옵션은 뭔가를 선택할 수 있다	10
2.3	이렇게 위치를 지정하고 그 값을 조회할 수 있다	10
2.4	혹은 새로운 데이터를 만들어 볼 수 있다	11
2.5	결과는 이렇다	11
2.6	새로운 컬럼을 만들 수 있다	11
2.7	인덱스를 이렇게 바꿀 수 있다	11
3	데이터 시각화	12

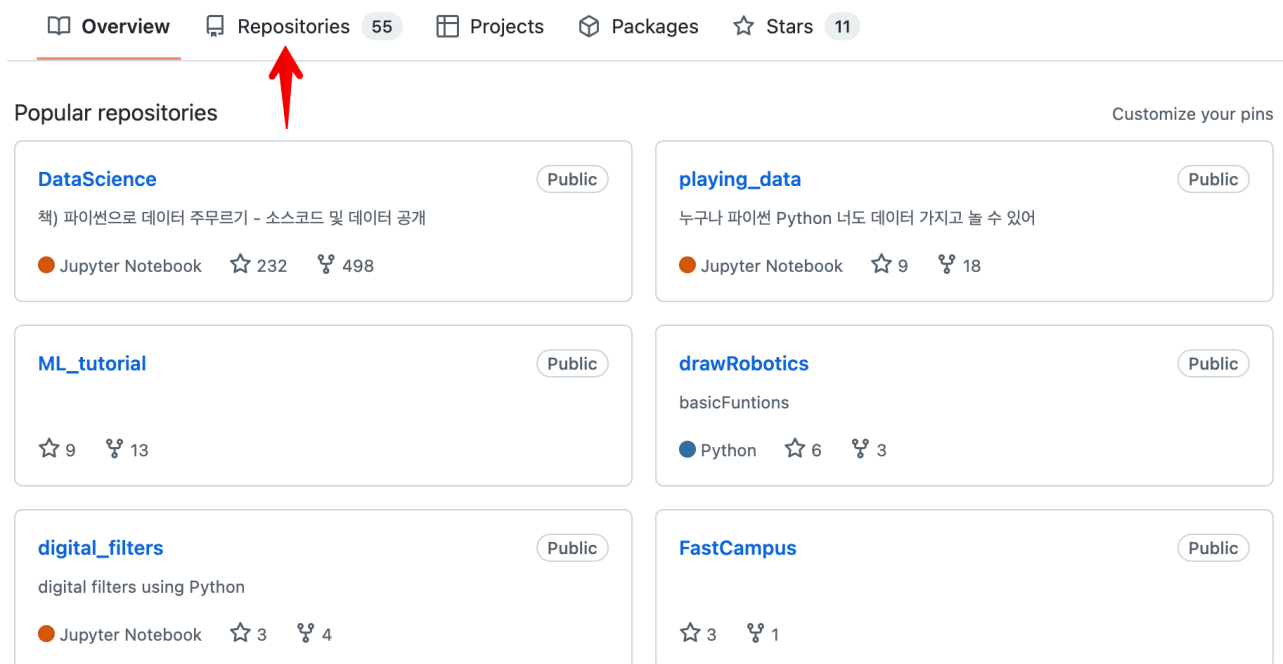
3.1	그 전에 한글 문제	12
3.2	뭘 어떻게 그릴까~	12
3.3	스트레스에 대한 그림	13
3.4	일단 요렇게 다시	14
3.5	ylabel을 없애버리자	15
3.6	요렇게~ 타이틀~	16
3.7	이렇게~	17
3.8	우울감	18
3.9	자살 생각을 시각화	19
3.10	번외편 - 이코드가 조금 어려울 수도 있을것 같지만	20
3.11	결과가 이쁘죠~	20

1 첫 시작 - 서울시 청소년 정신건강 분석

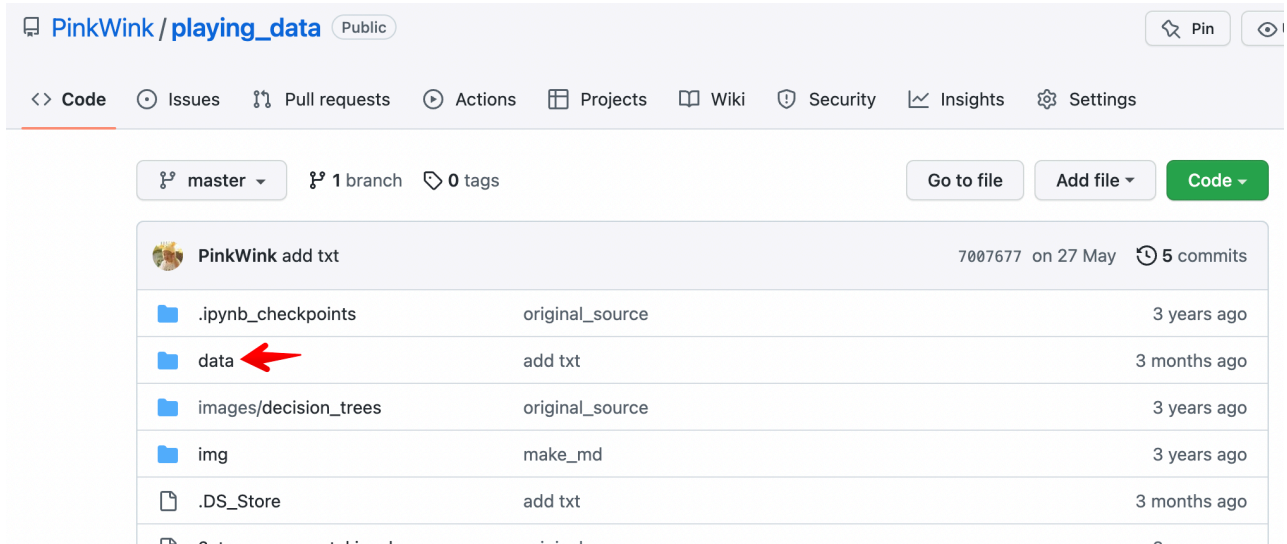
1.1 데이터 얻으러~ pinkwink github



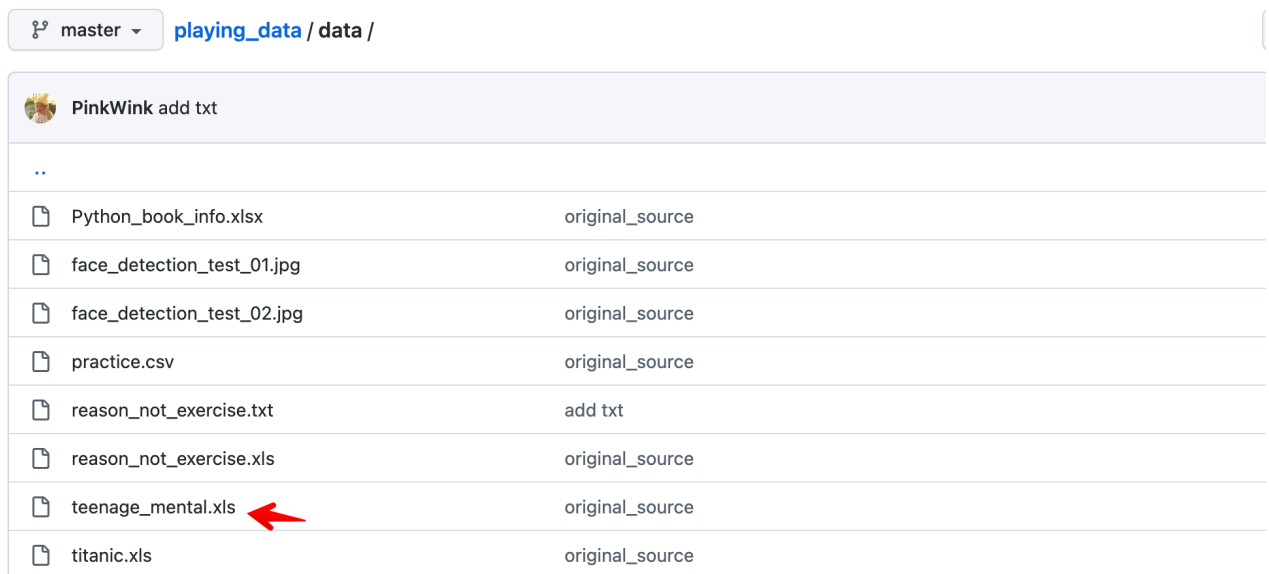
1.2 github 사이트에서 repository



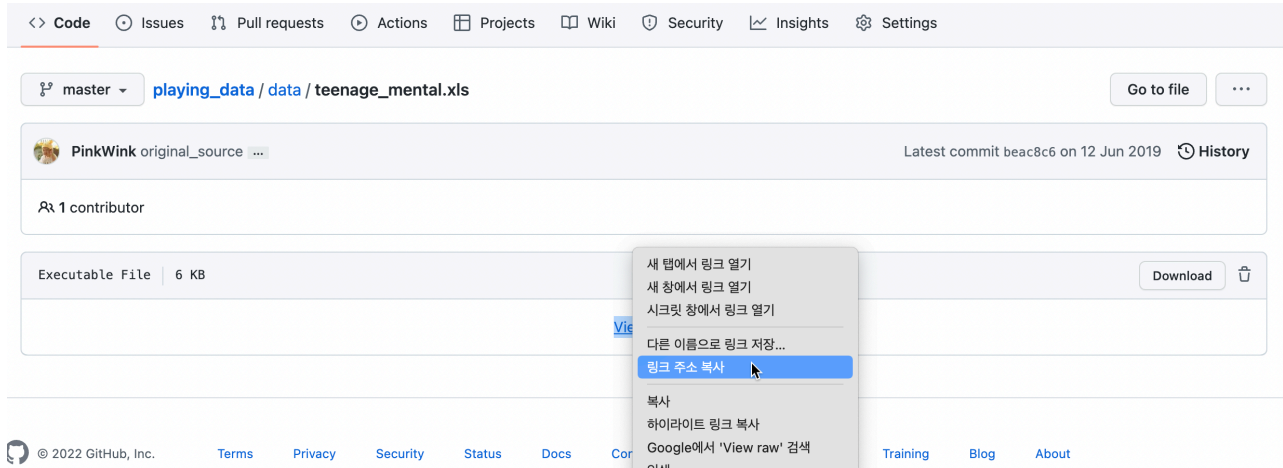
1.3 거기서 data 폴더 선택



1.4 거기서 teenage_mental.xls 클릭



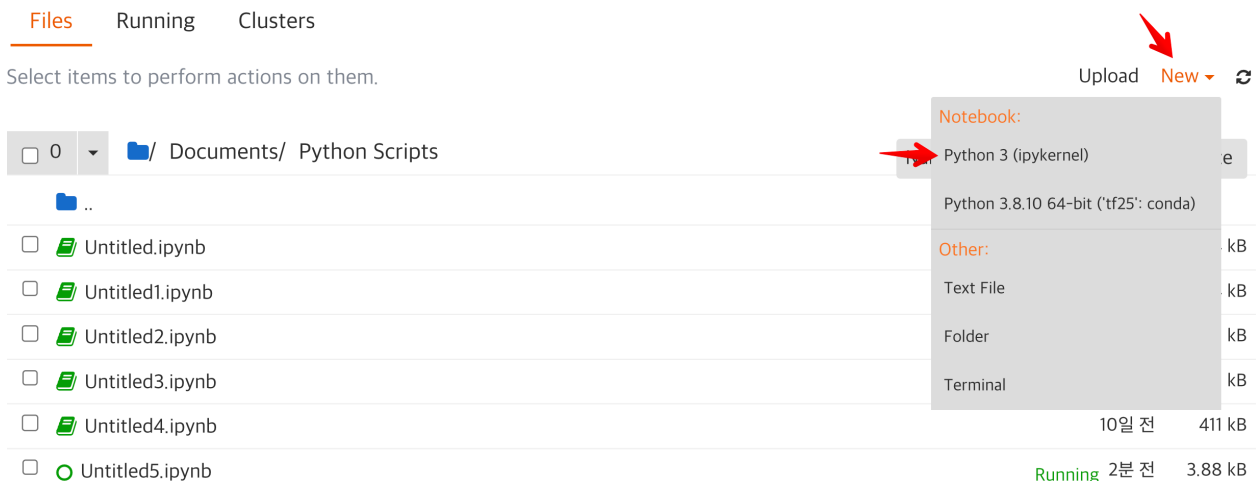
1.5 그리고 나타난 페이지에서 view_raw라는 링크에서 “주소 복사”



1.6 이제 jupyter notebook을 켜봅시다

- 윈도우 시작 버튼을 누르고 anaconda에 jupyter를 선택한 후
- 웹 브라우저는 크롬을 선택합니다.

1.7 그리고, new에 python3를 선택합니다.



1.8 url 이라는 변수에

File Edit View Insert Cell Kernel Widgets Help Trusted

```
In [ ]: url = ""
```

1.9 복사한 링크를 붙여 넣기

```
In [ ]: url = "https://github.com/PinkWink/playing_data/blob/master/data/teenage_mental.xls?raw=t"
```

```
In [ ]:
```

1.10 SHIFT+ENTER를 누르면 실행됨

```
In [1]: url = "https://github.com/PinkWink/playing_data/blob/master/data/teenage_mental.xls?raw=t"
```

```
In [ ]:
```

1.11 pandas라는 데이터 정리용 모듈을 이용해서 excel 파일을 읽어볼 수 있다

```
In [2]: import pandas as pd
pd.read_excel(url)
```

	기간	구분	스트레스 인지율	스트레스 인지율.1	스트레스 인지율.2	우울감 경험률	우울감 경험률.1	우울감 경험률.2	자살 생각률	자살 생각률.1	자살 생각률.2
0	기간	구분	전체	남학생	여학생	전체	남학생	여학생	전체	남학생	여학생
1	2018	구분	42.7	34.5	51.5	29.6	24.2	35.4	15.4	11.8	19.2

1.12 본래 파일을 엑셀로 열어보면 이렇게 생겼다

A1											
	A	B	C	D	E	F	G	H	I	J	K
1	기간	구분	스트레스 인지율			우울감 경험률			자살 생각률		
2			전체	남학생	여학생	전체	남학생	여학생	전체	남학생	여학생
3	2018	구분	42.7	34.5	51.5	29.6	24.2	35.4	15.4	11.8	19.2
4											
5											
6											
7											
8											

1.13 그런데 읽은 결과가 마음에 들지 않는다

	기간	구분	스트레스 인지율	스트레스 인지율.1	스트레스 인지율.2	우울감 경험률	우울감 경험률.1	우울감 경험률.2	자살 생각률	자살 생각률.1	자살 생각률.2
0	기간	구분	전체	남학생	여학생	전체	남학생	여학생	전체	남학생	여학생
1	2018	구분	42.7	34.5	51.5	29.6	24.2	35.4	15.4	11.8	19.2

1.14 이유는

A1											
	A	B	C	D	E	F	G	H	I	J	K
1	기간	구분	스트레스 인지율			우울감 경험률			자살 생각률		
2			전체	남학생	여학생	전체	남학생	여학생	전체	남학생	여학생
3	2018	구분	42.7	34.5	51.5	29.6	24.2	35.4	15.4	11.8	19.2
4											
5											
6											
7											
8											

1.15 일단 행 범위 선택

```
In [3]: import pandas as pd
pd.read_excel(url, header=1)
```

	기간	구분	전체	남학생	여학생	전체.1	남학생.1	여학생.1	전체.2	남학생.2	여학생.2
0	2018	구분	42.7	34.5	51.5	29.6	24.2	35.4	15.4	11.8	19.2

1.16 데이터의 범위는 잡은듯

```
In [5]: import pandas as pd
pd.read_excel(url, header=1, usecols="C:K")
```

	전체	남학생	여학생	전체.1	남학생.1	여학생.1	전체.2	남학생.2	여학생.2
0	42.7	34.5	51.5	29.6	24.2	35.4	15.4	11.8	19.2

1.17 이렇게 컬럼 이름을 바꾸면 좋을 것 같다

```
In [6]: import pandas as pd

col_names = ['스트레스', '스트레스남학생', '스트레스여학생', '우울감경험률',
             '우울남학생', '우울여학생', '자살생각률', '자살남학생', '자살여학생']
pd.read_excel(url, header=1, usecols="C:K", names=col_names)
```

	스트레스	스트레스남학생	스트레스여학생	우울감경험률	우울남학생	우울여학생	자살생각률	자살남학생	자살여학생
0	42.7	34.5	51.5	29.6	24.2	35.4	15.4	11.8	19.2

2 데이터 정리하기

2.1 방금까지의 내용을 raw_data에 저장하자

```
In [7]: import pandas as pd

col_names = ['스트레스', '스트레스남학생', '스트레스여학생', '우울감경험률',
             '우울남학생', '우울여학생', '자살생각률', '자살남학생', '자살여학생']
raw_data = pd.read_excel(url, header=1, usecols="C:K", names=col_names)
raw_data
```

	스트레스	스트레스남학생	스트레스여학생	우울감경험률	우울남학생	우울여학생	자살생각률	자살남학생	자살여학생
0	42.7	34.5	51.5	29.6	24.2	35.4	15.4	11.8	19.2

2.2 loc 옵션은 뭔가를 선택할 수 있다

```
In [3]: raw_data.loc[0]
```

```
스트레스      42.7
스트레스남학생  34.5
스트레스여학생  51.5
우울감경험률   29.6
우울남학생     24.2
우울여학생     35.4
자살생각률     15.4
자살남학생     11.8
자살여학생     19.2
Name: 0, dtype: float64
```

2.3 이렇게 위치를 지정하고 그 값을 조회할 수 있다

```
In [5]: raw_data.loc[0, '스트레스']
```

42.7

2.4 혹은 새로운 데이터를 만들어 볼 수 있다

```
In [6]: raw_data.loc[1] = 100.0 - raw_data.loc[0]
```

2.5 결과는 이렇다

```
In [7]: raw_data
```

	스트레스	스트레스남학생	스트레스여학생	우울감경험률	우울남학생	우울여학생	자살생각률	자살남학생	자살여학생
0	42.7	34.5	51.5	29.6	24.2	35.4	15.4	11.8	19.2
1	57.3	65.5	48.5	70.4	75.8	64.6	84.6	88.2	80.8

2.6 새로운 컬럼을 만들 수 있다

```
In [8]: raw_data['응답'] = ['그렇다', '아니다']
raw_data
```

	스트레스	스트레스남학생	스트레스여학생	우울감경험률	우울남학생	우울여학생	자살생각률	자살남학생	자살여학생	응답
0	42.7	34.5	51.5	29.6	24.2	35.4	15.4	11.8	19.2	그렇다
1	57.3	65.5	48.5	70.4	75.8	64.6	84.6	88.2	80.8	아니다

2.7 인덱스를 이렇게 바꿀 수 있다

```
In [10]: raw_data = raw_data.set_index('응답')
raw_data
```

	스트레스	스트레스남학생	스트레스여학생	우울감경험률	우울남학생	우울여학생	자살생각률	자살남학생	자살여학생
응답									
그렇다	42.7	34.5	51.5	29.6	24.2	35.4	15.4	11.8	19.2
아니다	57.3	65.5	48.5	70.4	75.8	64.6	84.6	88.2	80.8

3 데이터 시각화

3.1 그 전에 한글 문제

```
In [2]: import matplotlib.pyplot as plt
%matplotlib inline

from matplotlib import font_manager, rc
plt.rcParams['axes.unicode_minus'] = False

#f_path = "/Library/Fonts/AppleGothic.ttf"
f_path = "C://Windows/Fonts/malgun.ttf"
font_name = font_manager.FontProperties(fname=f_path).get_name()
rc('font', family=font_name)
```

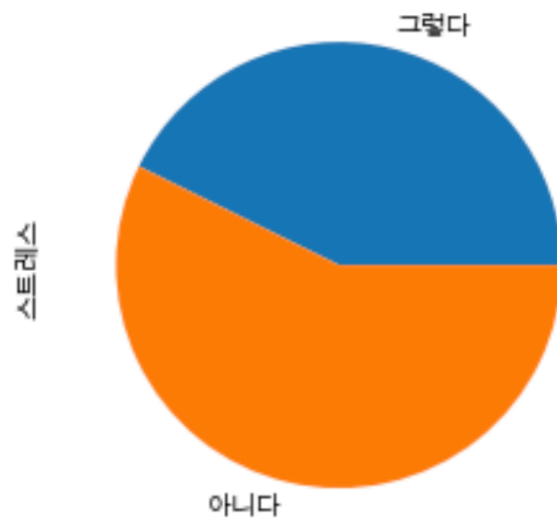
3.2 뭘 어떻게 그릴까~

```
In [15]: raw_data
```

	스트레스	스트레스남학생	스트레스여학생	우울감경험률	우울남학생	우울여학생	자살생각율	자살남학생	자살여학생
응답									
그렇다	42.7	34.5	51.5	29.6	24.2	35.4	15.4	11.8	19.2
아니다	57.3	65.5	48.5	70.4	75.8	64.6	84.6	88.2	80.8

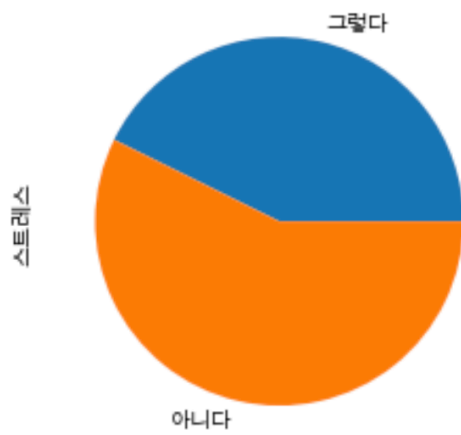
3.3 스트레스에 대한 그림

```
In [17]: raw_data['스트레스'].plot.pie();
```



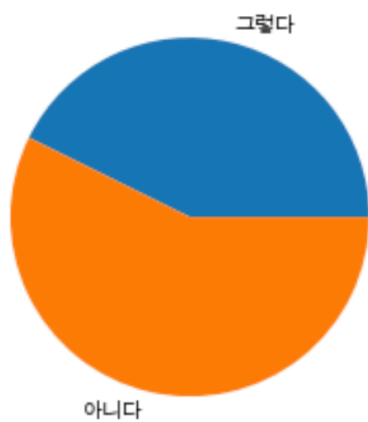
3.4 일단 이렇게 다시

```
In [18]: plt.figure()  
raw_data['스트레스'].plot.pie()  
plt.show()
```



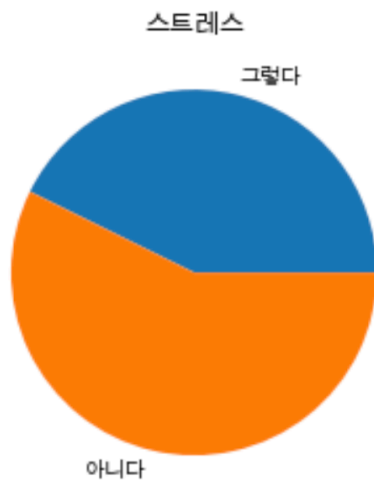
3.5 ylabel을 없애버리자

```
In [19]: plt.figure()  
raw_data['스트레스'].plot.pie()  
plt.ylabel('')  
plt.show()
```



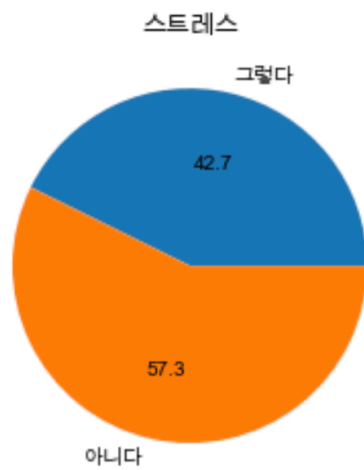
3.6 요렇게~ 타이틀~

```
In [20]: plt.figure()  
raw_data['스트레스'].plot.pie()  
plt.ylabel('')  
plt.title('스트레스')  
plt.show()
```



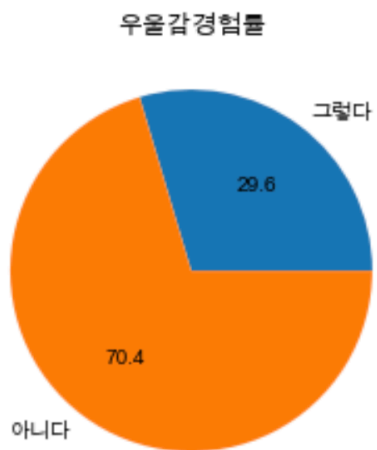
3.7 이렇게~

```
In [21]: plt.figure()  
raw_data['스트레스'].plot.pie(autopct='%1f')  
plt.ylabel('')  
plt.title('스트레스')  
plt.show()
```



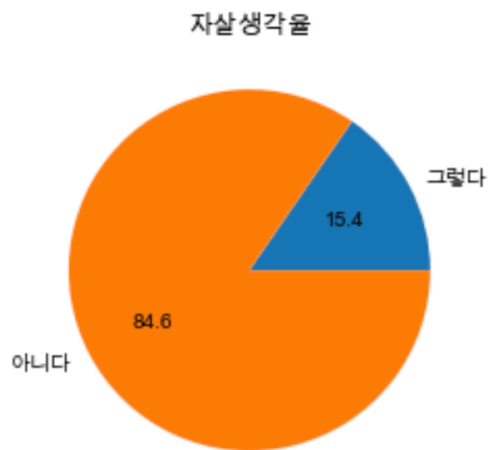
3.8 우울감

```
In [22]: plt.figure()  
raw_data['우울감경험률'].plot.pie(autopct='%.1f')  
plt.ylabel('')  
plt.title('우울감경험률')  
plt.show()
```



3.9 자살 생각을 시각화

```
In [23]: plt.figure()  
raw_data['자살생각율'].plot.pie(autopct='%.1f')  
plt.ylabel('')  
plt.title('자살생각율')  
plt.show()
```



3.10 번외편 - 이코드가 조금 어려울 수도 있을것 같지만

```
In [25]: f, ax = plt.subplots(1,3, figsize=(14,6))

raw_data['스트레스'].plot.pie(explode=[0,0.02], ax=ax[0], autopct='%.1f')
ax[0].set_title('스트레스를 받은적 있다')
ax[0].set_ylabel('')

raw_data['우울감경험률'].plot.pie(explode=[0,0.02], ax=ax[1], autopct='%.1f')
ax[1].set_title('우울증을 경험한적 있다')
ax[1].set_ylabel('')

raw_data['자살생각률'].plot.pie(explode=[0,0.02], ax=ax[2], autopct='%.1f')
ax[2].set_title('자살을 고민한적 있다')
ax[2].set_ylabel('')

plt.show()
```

3.11 결과가 이쁘죠~

