

# 코로나 데이터 분석

## 학습 목표

- 코로나 데이터와 백신 접종률 데이터를 수집하여 지금까지 내용을 나라별로 분석해 본다.
- 데이터 분석을 통해 pandas에 대해 좀 더 깊이 알아본다.

## 데이터가 수집된 사이트

- bloomberg.com
  - <https://www.bloomberg.com/graphics/covid-vaccine-tracker-global-distribution/>  
(<https://www.bloomberg.com/graphics/covid-vaccine-tracker-global-distribution/>).
- 코로나 19(COVID-19) 실시간 상황판
  - <https://coronaboard.kr/> (<https://coronaboard.kr/>).

In [1]:



```
from IPython.display import display, Image
import os, warnings
import re
warnings.filterwarnings(action='ignore')
```

## 01 파일 불러오기

In [2]:

```
os.listdir(os.getcwd())
```

Out[2]:

```
['.git',  
'idea',  
'ipynb_checkpoints',  
'01_dataScraping_corona01_today.ipynb',  
'01_dataScraping_corona02.py',  
'01_dataScraping_corona02_all.ipynb',  
'02_dataScraping_bloomberg.ipynb',  
'02_dataScraping_bloomberg.py',  
'03_corona_vaccine_merge.ipynb',  
'03_corona_vaccine_merge.py',  
'04_analysis_corr.ipynb',  
'05_people_request.ipynb',  
'06_corona_analysis.ipynb',  
'20210921_00_datamerge.csv',  
'20210921_00_datamerge.xlsx',  
'20210929_chrome_driver_01.ipynb',  
'2021_0911_13',  
'2021_0911_13.zip',  
'2021_0920_00',  
'chromedriver_91.exe',  
'data',  
'html_pdf',  
'project01_01',  
'README.md',  
'WEB_sel_project11_naverNews_corona_0731.ipynb']
```

In [12]:

```
import pandas as pd  
  
corona = pd.read_csv("./data/20210930_00_datamerge.csv")  
corona.shape
```

Out[12]:

```
(178, 24)
```

## Learn 01. 보이지 않는 전체 내용을 표시해 보기

In [13]:

```
print( pd.options.display.max_rows )  
print( pd.options.display.max_columns )
```

```
500  
500
```

In [14]:

```
pd.options.display.max_columns = 500  
pd.options.display.max_rows = 500
```

In [15]:



```
corona.head()
```

Out [15]:

	한글 (국가명)	영문 (블룸버그)	1차 접종	2차 접종	발생 률순위	2차 접종 률순위	확진자비율 (인구)	일별접종 수	치명 (%)	완치 (%)	발생률	인구수
0	미국	U.S.	64.4	55.8	13.0	52.5	0.133098	701422.0	1.6	76.1	133098	331002651
1	인도	India	46.7	16.8	103.0	119.0	0.024434	7178901.0	1.3	97.8	24434	1380004385
2	브라질	Brazil	71.3	41.9	29.0	70.5	0.100592	1317163.0	2.8	95.3	100592	212559417
3	영국	U.K.	73.0	67.1	20.0	24.0	0.113959	63126.0	1.8	80.9	113959	67886011
4	러시아	Russia	32.6	28.5	77.0	89.0	0.051305	214966.0	2.8	88.9	51305	145934462

In [16]:



```
corona.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 178 entries, 0 to 177
Data columns (total 24 columns):
 #   Column                Non-Null Count  Dtype  
---  --
 0   한글(국가명)          178 non-null    object  
 1   영문(블룸버그)        178 non-null    object  
 2   1차접종                178 non-null    float64  
 3   2차접종                178 non-null    float64  
 4   발생률순위            178 non-null    float64  
 5   2차접종률순위         178 non-null    float64  
 6   확진자비율(인구)     178 non-null    float64  
 7   일별접종수            178 non-null    float64  
 8   치명(%)              178 non-null    object  
 9   완치(%)              175 non-null    float64  
10   발생률                178 non-null    int64  
11   인구수                178 non-null    int64  
12   치료중합계            175 non-null    float64  
13   치료중1일             145 non-null    float64  
14   위중증합계            131 non-null    float64  
15   위중증1일             54 non-null     object  
16   확진자합계            178 non-null    int64  
17   확진자1일             146 non-null    float64  
18   사망자합계            178 non-null    int64  
19   사망자1일             114 non-null    float64  
20   완치합계              175 non-null    float64  
21   완치1일               130 non-null    float64  
22   백신접종수          178 non-null    float64  
23   백신보유율(인구)    178 non-null    float64  
dtypes: float64(16), int64(4), object(4)
memory usage: 33.5+ KB
```

**가설 : 백신 접종은 코로나 감염에 효과가 있다.**

## 접종률(인구)와 발생률의 관계

In [17]:



```
import seaborn as sns
import matplotlib.pyplot as plt
from matplotlib import font_manager, rc
import matplotlib
import platform
```

## 한글 표시하기

In [18]:

```
path = "C:/Windows/Fonts/malgun.ttf"
if platform.system() == "Windows":
    font_name = font_manager.FontProperties(fname=path).get_name()
    rc('font', family=font_name)
elif platform.system()=="Darwin":
    rc('font', family='AppleGothic')
else:
    print("Unknown System")

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

In [19]:

```
corona.columns
```

Out[19]:

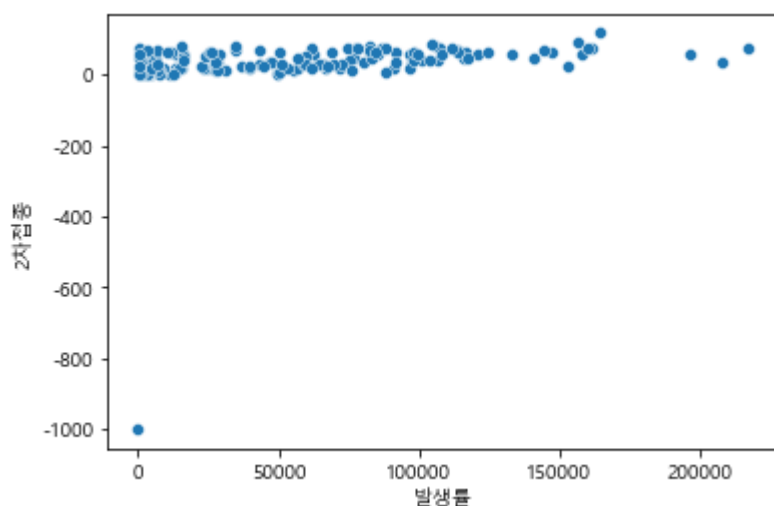
```
Index(['한글(국가명)', '영문(블룸버그)', '1차접종', '2차접종', '발생률순위', '2차접  
종률순위', '확진자비율(인구)',  
      '일별접종수', '치명(%)', '완치(%)', '발생률', '인구수', '치료중합계', '치료중  
1일', '위중증합계',  
      '위중증1일', '확진자합계', '확진자1일', '사망자합계', '사망자1일', '완치합  
계', '완치1일', '백신접종수',  
      '백신보유율(인구)'],  
      dtype='object')
```

In [20]:

```
sns.scatterplot(x='발생률', y='2차접종', data=corona)
```

Out[20]:

<AxesSubplot: xlabel='발생률', ylabel='2차접종'>



## 이상치 제거

In [21]:

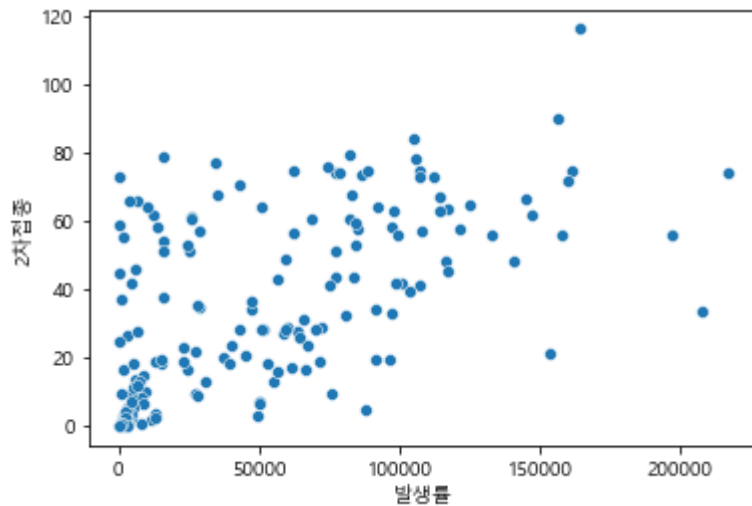
```
corona = corona[corona['2차접종'] > 0]
```

In [22]:

```
sns.scatterplot(x='발생률', y='2차접종', data=corona)
```

Out[22]:

<AxesSubplot: xlabel='발생률', ylabel='2차접종'>

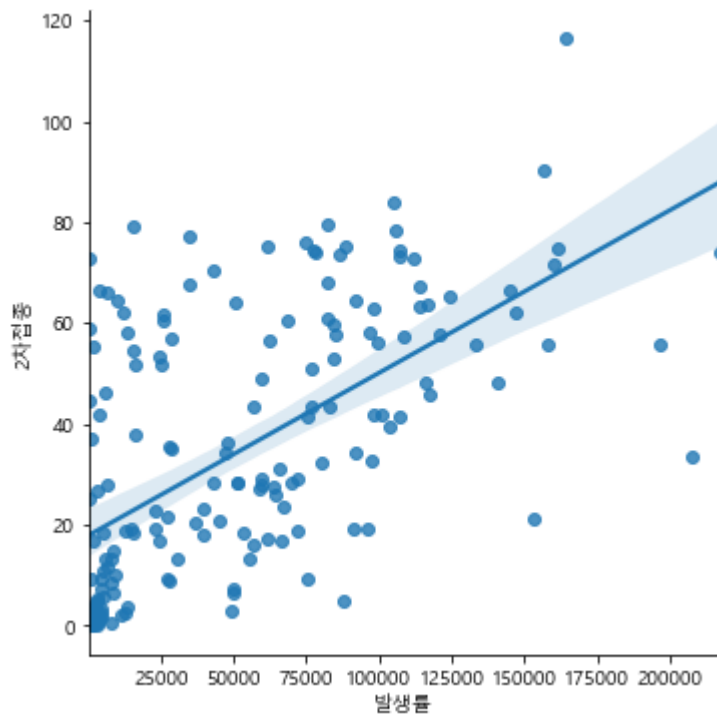


In [23]:

```
sns.lmplot(x='발생률', y='2차접종', data=corona)
```

Out[23]:

<seaborn.axisgrid.FacetGrid at 0x2638ef396a0>



In [24]:



```
corona[['발생률', '2차접종']].corr()
```

Out[24]:

	발생률	2차접종
발생률	1.000000	0.597028
2차접종	0.597028	1.000000

확인 결과 나라별 백신 접종률과 인구당 확진자 비율은 0.582의 양의 상관관계를 갖는다.

- 백신을 접종하면 할수록 인구당 확진자 비율은 양의 관계로 늘어난다.

백신 발생률과 사망률(인구당)은 어떠한가?

In [25]:



```
corona.head()
```

Out[25]:

	한글 (국가명)	영문 (블룸버그)	1차 접종	2차 접종	발생 률순 위	2차 접종 률순 위	확진자비 율(인구)	일별접종 수	치 명 (%)	완치 (%)	발생률	인구수
0	미국	U.S.	64.4	55.8	13.0	52.5	0.133098	701422.0	1.6	76.1	133098	331002651
1	인도	India	46.7	16.8	103.0	119.0	0.024434	7178901.0	1.3	97.8	24434	1380004385
2	브라질	Brazil	71.3	41.9	29.0	70.5	0.100592	1317163.0	2.8	95.3	100592	212559417
3	영국	U.K.	73.0	67.1	20.0	24.0	0.113959	63126.0	1.8	80.9	113959	67886011
4	러시아	Russia	32.6	28.5	77.0	89.0	0.051305	214966.0	2.8	88.9	51305	145934462

In [26]:



```
### 사망률 = 사망자수/인구수
corona['사망률'] = corona['사망자합계'] / corona['인구수']
corona.head(15)
```

Out[26]:

	한글 (국가명)	영문(블룸버그)	1차 접종	2차 접종	발생 률순 위	2차 접종 률순 위	확진자비 율(인구)	일별접종 수	치 명 (%)	완치 (%)	발생률	인구수
0	미국	U.S.	64.4	55.8	13.0	52.5	0.133098	701422.0	1.6	76.1	133098	331002651
1	인도	India	46.7	16.8	103.0	119.0	0.024434	7178901.0	1.3	97.8	24434	1380004385
2	브라질	Brazil	71.3	41.9	29.0	70.5	0.100592	1317163.0	2.8	95.3	100592	212559417
3	영국	U.K.	73.0	67.1	20.0	24.0	0.113959	63126.0	1.8	80.9	113959	67886011
4	러시아	Russia	32.6	28.5	77.0	89.0	0.051305	214966.0	2.8	88.9	51305	145934462
5	터키	Turkey	64.5	53.0	44.0	57.0	0.084132	381129.0	0.9	92.4	84132	84339067
6	프랑스	France	77.4	74.6	23.0	12.0	0.107278	229555.0	1.7	96.4	107278	65273511
7	이란	Iran	35.5	16.8	62.0	119.0	0.066350	1368407.0	2.2	89.9	66350	83992949
8	아르헨티나	Argentina	65.7	48.3	18.0	62.0	0.116245	222650.0	2.2	97.4	116245	45195774
9	콜롬비아	Colombia	51.4	32.8	33.0	84.0	0.097368	169329.0	2.5	96.8	97368	50882891
10	스페인	Spain	81.0	78.5	26.0	6.0	0.105956	72816.0	1.7	95.8	105956	46754778
11	이탈리아	Italy	74.5	74.5	51.0	13.0	0.077157	211345.0	2.8	95.1	77157	60461826
12	독일	Germany	67.9	64.2	79.0	31.0	0.050384	194817.0	2.2	94.5	50384	83783942



	한글 (국가명)	영문(블룸버그)	1차 접종	2차 접종	발생 률순위	2차 접종 률순위	확진자비율(인구)	일별접종수	치명 (%)	완치 (%)	발생률	인구수
13	인도네시아	Indonesia	32.9	18.4	111.0	114.5	0.015404	1916012.0	3.4	95.7	15404	273523615
14	멕시코	Mexico	49.8	35.0	95.0	80.0	0.028275	536693.0	7.6	82.3	28275	128932753

가설 : 백신은 중증과 사망자 감소에 효과가 있다.

## 2-1 백신 접종률과 사망률의 관계는 어떠한가?

In [27]:



```
corona[['사망률', '2차접종']].corr()
```

Out [27]:

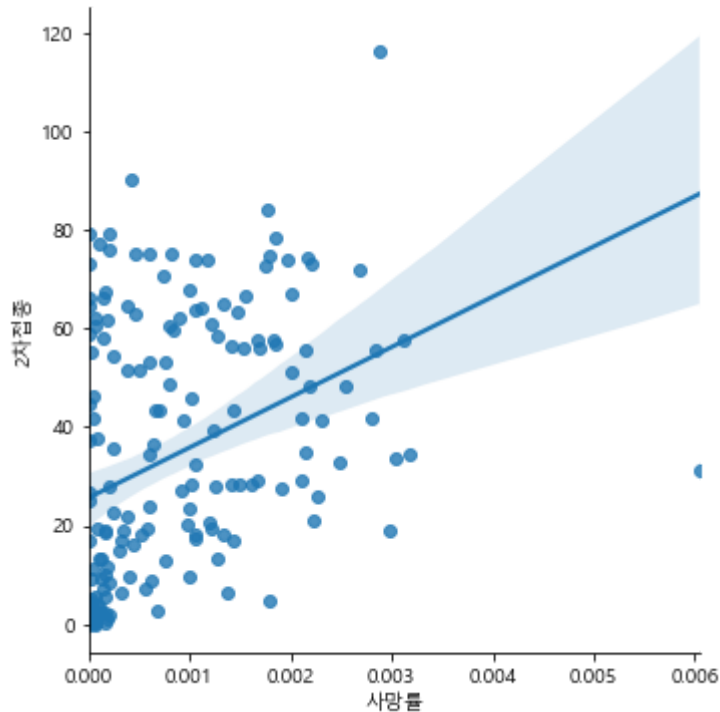
	사망률	2차접종
사망률	1.000000	0.355118
2차접종	0.355118	1.000000

In [28]:

```
sns.lmplot(x='사망률', y='2차접종', data=corona)
```

Out[28]:

<seaborn.axisgrid.FacetGrid at 0x2638efc8d30>



백신 접종률과 사망률은 0.324의 양의 상관관계를 갖는다.

- 데이터로 봤을 때, 이 부분에 대한 확인이 필요해 보인다.

2-2 백신 접종률과 중증 감소의 관계는 어떠한가?

In [29]:



```
corona.head(3)
```

Out[29]:

치명 %)	완치 (%)	발생률	인구수	치료중합 계	치료중1 일	위중증 합계	위중 증1일	확진자합 계	확진자1 일	사망자 합계	.
.6	76.1	133098	331002651	9823512.0	-14871.0	21296.0	(-523)	44055692	113357.0	711222	;
.3	97.8	24434	1380004385	284976.0	-4806.0	8944.0	NaN	33718984	4743.0	447828	
.8	95.3	100592	212559417	403027.0	-7475.0	8318.0	NaN	21381790	15395.0	595520	

## Learn02. Pandas 문자열 처리

- `[],str.strip()` : 앞뒤 공백을 제거
- `[],str.lstrip()` : 앞 공백을 제거
- `[],str.rstrip()` : 뒤 공백을 제거
- `[],str.split('구분자', n=1, expand=True)` : 구분자를 기준으로 n개로 나눈다. `expand=True`(여러컬럼), `False`(하나의 컬럼)

In [32]:



```
### 사망률 = 사망자수/인구수
corona['위중증률'] = corona['위중증합계'] / corona['인구수']
corona.head(15)
```

Out[32]:

	한글 (국가명)	영문(블룸버그)	1차 접종	2차 접종	발생 률순 위	2차 접종 률순 위	확진자비 율(인구)	일별접종 수	치 명 (%)	완치 (%)	발생률	인구수
0	미국	U.S.	64.4	55.8	13.0	52.5	0.133098	701422.0	1.6	76.1	133098	331002651
1	인도	India	46.7	16.8	103.0	119.0	0.024434	7178901.0	1.3	97.8	24434	1380004385
2	브라질	Brazil	71.3	41.9	29.0	70.5	0.100592	1317163.0	2.8	95.3	100592	212559417
3	영국	U.K.	73.0	67.1	20.0	24.0	0.113959	63126.0	1.8	80.9	113959	67886011
4	러시아	Russia	32.6	28.5	77.0	89.0	0.051305	214966.0	2.8	88.9	51305	145934462
5	터키	Turkey	64.5	53.0	44.0	57.0	0.084132	381129.0	0.9	92.4	84132	84339067
6	프랑스	France	77.4	74.6	23.0	12.0	0.107278	229555.0	1.7	96.4	107278	65273511
7	이란	Iran	35.5	16.8	62.0	119.0	0.066350	1368407.0	2.2	89.9	66350	83992949
8	아르헨티나	Argentina	65.7	48.3	18.0	62.0	0.116245	222650.0	2.2	97.4	116245	45195774
9	콜롬비아	Colombia	51.4	32.8	33.0	84.0	0.097368	169329.0	2.5	96.8	97368	50882891
10	스페인	Spain	81.0	78.5	26.0	6.0	0.105956	72816.0	1.7	95.8	105956	46754778
11	이탈리아	Italy	74.5	74.5	51.0	13.0	0.077157	211345.0	2.8	95.1	77157	60461826
12	독일	Germany	67.9	64.2	79.0	31.0	0.050384	194817.0	2.2	94.5	50384	83783942

	한글 (국가명)	영문(블룸버그)	1차 접종	2차 접종	발생 률순위	2차 접종 률순위	확진자비율(인구)	일별접종수	치명 (%)	완치 (%)	발생률	인구수
13	인도네시아	Indonesia	32.9	18.4	111.0	114.5	0.015404	1916012.0	3.4	95.7	15404	273523615
14	멕시코	Mexico	49.8	35.0	95.0	80.0	0.028275	536693.0	7.6	82.3	28275	128932753

## 에러 발생

- `TypeError: unsupported operand type(s) for /: 'str' and 'int'`

## Learn 03. 데이터 타입 변경하기

- `df.astype(dtype)` : 모든 열의 데이터 타입을 변경
- `df.astype( { '컬럼명': dtype } )` : 특정 컬럼의 데이터 타입을 변경
  - str, int, float, category 등
  - int8, int16, int32, int64와 같이 변경 가능

In [33]:



```
corona.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 177 entries, 0 to 176
Data columns (total 26 columns):
 #   Column                Non-Null Count  Dtype  
---  --
 0   한글(국가명)          177 non-null    object  
 1   영문(블룸버그)        177 non-null    object  
 2   1차접종                177 non-null    float64  
 3   2차접종                177 non-null    float64  
 4   발생률순위            177 non-null    float64  
 5   2차접종률순위         177 non-null    float64  
 6   확진자비율(인구)     177 non-null    float64  
 7   일별접종수            177 non-null    float64  
 8   치명(%)               177 non-null    object  
 9   완치(%)               174 non-null    float64  
10   발생률                177 non-null    int64  
11   인구수                177 non-null    int64  
12   치료중합계            174 non-null    float64  
13   치료중1일             145 non-null    float64  
14   위중증합계            131 non-null    float64  
15   위중증1일             54 non-null     object  
16   확진자합계            177 non-null    int64  
17   확진자1일            146 non-null    float64  
18   사망자합계            177 non-null    int64  
19   사망자1일            114 non-null    float64  
20   완치합계              174 non-null    float64  
21   완치1일              130 non-null    float64  
22   백신접종수          177 non-null    float64  
23   백신보유율(인구)    177 non-null    float64  
24   사망률                177 non-null    float64  
25   위중증률              131 non-null    float64  
dtypes: float64(18), int64(4), object(4)
memory usage: 37.3+ KB
```

In [34]:



```
corona['위중증합계'] = corona['위중증합계'].astype('int64')
```

```
-----
ValueError                                Traceback (most recent call last)
<ipython-input-34-3623f67dcdd0> in <module>
----> 1 corona['위중증합계'] = corona['위중증합계'].astype('int64')

~Wanaconda3\lib\site-packages\pandas\core\generic.py in astype(self, dtype,
copy, errors)
    5875         else:
    5876             # else, only a single dtype is given
-> 5877             new_data = self._mgr.astype(dtype=dtype, copy=copy, errors=e
rrors)
    5878             return self._constructor(new_data).__finalize__(self, method=
"astype")
    5879

~Wanaconda3\lib\site-packages\pandas\core\internals\managers.py in astype(s
elf, dtype, copy, errors)
    629         self, dtype, copy: bool = False, errors: str = "raise"
    630     ) -> "BlockManager":
--> 631         return self.apply("astype", dtype=dtype, copy=copy, errors=erro
rs)
    632
    633     def convert(

~Wanaconda3\lib\site-packages\pandas\core\internals\managers.py in apply(se
lf, f, align_keys, ignore_failures, **kwargs)
    425         applied = b.apply(f, **kwargs)
    426         else:
--> 427             applied = getattr(b, f)(**kwargs)
    428         except (TypeError, NotImplementedError):
    429             if not ignore_failures:

~Wanaconda3\lib\site-packages\pandas\core\internals\blocks.py in astype(sel
f, dtype, copy, errors)
    671         vals1d = values.ravel()
    672         try:
--> 673             values = astype_nansafe(vals1d, dtype, copy=True)
    674         except (ValueError, TypeError):
    675             # e.g. astype_nansafe can fail on object-dtype of str
ings

~Wanaconda3\lib\site-packages\pandas\core\dtypes\cast.py in astype_nansafe(a
rr, dtype, copy, skipna)
   1066
   1067         if not np.isfinite(arr).all():
-> 1068             raise ValueError("Cannot convert non-finite values (NA
or inf) to integer")
   1069
   1070         elif is_object_dtype(arr):
```

ValueError: Cannot convert non-finite values (NA or inf) to integer

에러 발생

- ValueError: cannot convert float NaN to integer

In [35]:



```
corona['위중증합계'].unique()
```

Out[35]:

```
array([2.1296e+04, 8.9440e+03, 8.3180e+03, 8.3100e+02, 2.3000e+03,
        6.3300e+02, 1.6090e+03, 6.0160e+03, 1.1840e+03, 5.4200e+02,
        7.7100e+02, 4.5900e+02, 1.4350e+03,          nan, 4.7980e+03,
        1.7400e+02, 3.1700e+03, 1.7700e+02, 9.8500e+02, 1.0210e+03,
        1.5500e+02, 4.9700e+02, 1.0020e+03, 3.8000e+01, 3.8600e+02,
        7.7600e+02, 3.3240e+03, 1.4080e+03, 6.5900e+02, 3.9480e+03,
        2.0600e+02, 1.3200e+03, 3.7000e+01, 6.8000e+01, 2.4300e+02,
        4.9800e+02, 2.2100e+02, 4.2100e+02, 1.8600e+02, 8.1000e+01,
        4.8000e+02, 2.2600e+02, 2.9500e+02, 3.2200e+02, 2.0000e+02,
        5.0000e+00, 2.2400e+02, 4.4900e+02, 7.5900e+02, 2.2000e+02,
        6.1000e+01, 2.1000e+01, 1.0000e+01, 1.1900e+02, 9.9000e+01,
        1.3000e+01, 6.3000e+01, 6.8100e+02, 4.1800e+02, 1.6000e+01,
        7.8000e+02, 2.1400e+02, 3.3100e+02, 2.0000e+01, 9.0000e+01,
        1.9200e+02, 7.4000e+01, 9.2000e+01, 1.5000e+01, 2.5000e+01,
        1.1000e+01, 1.4000e+01, 2.3000e+01, 1.3100e+02, 1.0000e+00,
        3.0000e+00, 1.8000e+01, 1.1240e+03, 3.2000e+01, 1.2000e+01,
        4.2000e+01, 2.6800e+02, 3.3000e+01, 1.2700e+02, 3.0100e+02,
        3.0000e+01, 1.5200e+02, 8.0000e+00, 7.0000e+00, 6.7000e+01,
        1.7000e+01, 2.2000e+01, 3.1000e+01, 2.4000e+01, 2.6000e+01,
        4.0000e+00, 3.6000e+01, 2.0000e+00])
```

## Learn 04. 어떻게 NaN을 체크할 것인가?



```
corona.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 177 entries, 0 to 176
Data columns (total 26 columns):
 #   Column                Non-Null Count  Dtype  
---  --
 0   한글(국가명)          177 non-null    object  
 1   영문(블룸버그)        177 non-null    object  
 2   1차접종                177 non-null    float64  
 3   2차접종                177 non-null    float64  
 4   발생률순위            177 non-null    float64  
 5   2차접종률순위         177 non-null    float64  
 6   확진자비율(인구)     177 non-null    float64  
 7   일별접종수            177 non-null    float64  
 8   치명(%)              177 non-null    object  
 9   완치(%)              174 non-null    float64  
10   발생률                177 non-null    int64  
11   인구수                177 non-null    int64  
12   치료중합계            174 non-null    float64  
13   치료중1일             145 non-null    float64  
14   위중증합계            131 non-null    float64  
15   위중증1일             54 non-null    object  
16   확진자합계            177 non-null    int64  
17   확진자1일             146 non-null    float64  
18   사망자합계            177 non-null    int64  
19   사망자1일             114 non-null    float64  
20   완치합계              174 non-null    float64  
21   완치1일               130 non-null    float64  
22   백신접종수           177 non-null    float64  
23   백신보유율(인구)    177 non-null    float64  
24   사망률                177 non-null    float64  
25   위중증률              131 non-null    float64  
dtypes: float64(18), int64(4), object(4)
memory usage: 37.3+ KB
```

## 위중증\_수가 isnull()인 것을 살펴보기

In [46]:



```
corona[ corona['위중증합계'].isnull() ].head(10)
```

Out[46]:

	한글 (국가명)	영문(블룸버그)	1차 접종	2차 접종	발생 률 순위	2차 접종 순위	확진자비율(인구)	일별접종수	치명 (%)	완치 (%)	발생률	인구수
13	인도네시아	Indonesia	32.9	18.4	111.0	114.5	0.015404	1916012.0	3.4	95.7	15404	273523615
41	네팔	Nepal	23.9	21.7	99.0	103.0	0.027256	186712.0	1.4	96.3	27256	29136808
42	베트남	Vietnam	32.1	8.6	123.0	135.0	0.008007	621033.0	2.5	74.9	8007	97338579
44	아랍에미리트	UAE	85.9	76.1	56.0	8.0	0.074388	59157.0	0.3	99.0	74388	9890402
48	조지아	Georgia	26.2	21.1	9.0	104.0	0.153232	12798.0	1.5	95.3	153232	3989167
51	벨라루스	Belarus	19.8	16.0	74.0	121.0	0.056734	26145.0	0.8	96.5	56734	9449323
53	스리랑카	Sri Lanka	66.0	53.3	104.0	56.0	0.024075	167003.0	2.5	88.5	24075	21413249
57	아제르바이잔	Azerbaijan	47.2	36.4	83.0	78.0	0.047479	53286.0	1.3	94.1	47479	10139177
59	미얀마	Myanmar	10.5	6.6	122.0	138.0	0.008474	163440.0	3.8	90.1	8474	54409800
70	리비아	Libya	20.7	2.8	82.0	149.0	0.049375	11651.0	1.4	76.2	49375	6871292

In [38]:

```
corona['위중증합계'].isnull().sum()
```

Out[38]:

46

In [39]:

```
corona['위중증합계'].isna().sum()
```

Out[39]:

46

(의문) 왜 47개국은 위중증 통계가 없을까?

위중증\_수2를 만들고, 위중증수가 없는 것은 -99999로 만들자.

In [40]:

```
corona['위중증합계2'] = corona['위중증합계'].copy()
```

In [41]:

```
corona.loc[ corona['위중증합계2'].isnull(), "위중증합계2"] = -99999
corona[ corona['위중증합계2'].isnull() ]
```

Out[41]:

한글 (국가명)	영문 (블루버그)	1차 접종	2차 접종	발생률 순위	2차 접종률 순위	확진 자비율 (인구)	일별 접종수	치명 (%)	완치 (%)	발생 률	인구 수	치 료 중 합 계	치 료 중 1 일	위 중 증 합 계	위 중 증 1 일	확 진 자 합 계	확 진 자 1 일	사 망 자 합 계	사 망 자 1 일	완 치 합 계	완 치 1 일	백 신 접 종 수
-------------	--------------	----------	----------	-----------	-----------------	-------------------	-----------	-----------	-----------	---------	---------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	-----------------------	------------------	------------------	-----------------------

In [43]:

```
corona['위중증합계2'] = corona['위중증합계2'].astype('int64')
```

In [45]:



```
### 사망률 = 사망자수/인구수  
corona['위중증률'] = corona['위중증합계2'] / corona['인구수']  
corona.head(10)
```

Out[45]:

	한글 (국가명)	영문(블룸버그)	1차 접종	2차 접종	발생 률순 위	2차 접종 률순 위	확진자비 율(인구)	일별접종 수	치 명 (%)	완치 (%)	발생률	인구수
0	미국	U.S.	64.4	55.8	13.0	52.5	0.133098	701422.0	1.6	76.1	133098	331002651
1	인도	India	46.7	16.8	103.0	119.0	0.024434	7178901.0	1.3	97.8	24434	1380004385
2	브라질	Brazil	71.3	41.9	29.0	70.5	0.100592	1317163.0	2.8	95.3	100592	212559417
3	영국	U.K.	73.0	67.1	20.0	24.0	0.113959	63126.0	1.8	80.9	113959	67886011
4	러시아	Russia	32.6	28.5	77.0	89.0	0.051305	214966.0	2.8	88.9	51305	145934462
5	터키	Turkey	64.5	53.0	44.0	57.0	0.084132	381129.0	0.9	92.4	84132	84339067
6	프랑스	France	77.4	74.6	23.0	12.0	0.107278	229555.0	1.7	96.4	107278	65273511
7	이란	Iran	35.5	16.8	62.0	119.0	0.066350	1368407.0	2.2	89.9	66350	83992949
8	아르헨티나	Argentina	65.7	48.3	18.0	62.0	0.116245	222650.0	2.2	97.4	116245	45195774
9	콜롬비아	Colombia	51.4	32.8	33.0	84.0	0.097368	169329.0	2.5	96.8	97368	50882891

In [48]:



```
corona.columns
```

Out[48]:

```
Index(['한글(국가명)', '영문(블룸버그)', '1차접종', '2차접종', '발생률순위', '2차접  
종률순위', '확진자비율(인구)',  
      '일별접종수', '치명(%)', '완치(%)', '발생률', '인구수', '치료중합계', '치료중  
1일', '위중증합계',  
      '위중증1일', '확진자합계', '확진자1일', '사망자합계', '사망자1일', '완치합  
계', '완치1일', '백신접종수',  
      '백신보유율(인구)', '사망률', '위중증률', '위중증합계2'],  
      dtype='object')
```

In [53]:



```
corona_sel = ['한글(국가명)', '발생률', '사망률', '위중증률', '인구수', '확진자비율(인구)',  
              '1차접종', '2차접종', '확진자합계', '위중증합계', '치료중합계', '위중증합계2',  
              '치료중1일', '위중증1일', '확진자1일', ]  
corona_part = corona[corona_sel]  
corona_part.head(10)
```

Out[53]:

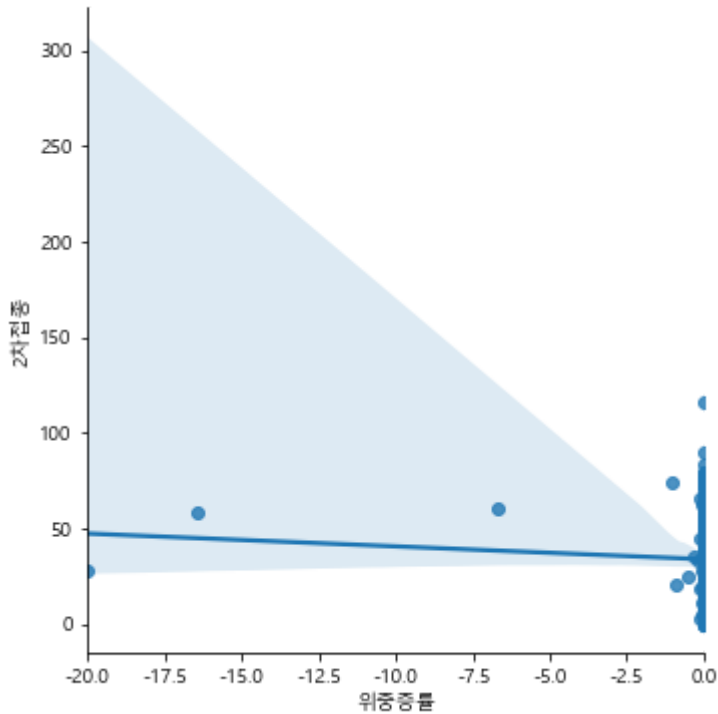
	한글 (국가명)	발생률	사망률	위중증률	인구수	확진자비율(인구)	1차 접종	2차 접종	확진자합계	위중증 합계	치료
0	미국	133098	0.002149	0.000064	331002651	0.133098	64.4	55.8	44055692	21296.0	98236
1	인도	24434	0.000325	0.000006	1380004385	0.024434	46.7	16.8	33718984	8944.0	2846
2	브라질	100592	0.002802	0.000039	212559417	0.100592	71.3	41.9	21381790	8318.0	4036
3	영국	113959	0.002009	0.000012	67886011	0.113959	73.0	67.1	7736235	831.0	13446
4	러시아	51305	0.001414	0.000016	145934462	0.051305	32.6	28.5	7487138	2300.0	6266
5	터키	84132	0.000754	0.000008	84339067	0.084132	64.5	53.0	7095580	633.0	4796
6	프랑스	107278	0.001787	0.000025	65273511	0.107278	77.4	74.6	7002393	1609.0	1356
7	이란	66350	0.001431	0.000072	83992949	0.066350	35.5	16.8	5572962	6016.0	4436
8	아르헨티나	116245	0.002545	0.000026	45195774	0.116245	65.7	48.3	5253765	1184.0	2366
9	콜롬비아	97368	0.002481	0.000011	50882891	0.097368	51.4	32.8	4954376	542.0	3266

In [54]:

```
sns.lmplot(x='위중증률', y='2차접종', data=corona_part)
```

Out[54]:

<seaborn.axisgrid.FacetGrid at 0x2638f440fd0>



In [55]:

```
corona_part[['위중증률', '2차접종']].corr()
```

Out[55]:

	위중증률	2차접종
위중증률	1.000000	-0.050268
2차접종	-0.050268	1.000000

**결과 확인 : 인구당 접종률과 위중증률은 거의 관계가 없다.**

**오류가 있을까?**

- 오류: -99999 데이터가 있었기에 정확하지 않을 수 있다.

In [56]:

```
corona_part['위중증합계'].notnull().sum()
```

Out[56]:

131

In [57]:

```
corona_part = corona_part[ corona_part['위중증합계'].notnull() ]  
corona_part['위중증합계'].isnull().sum()
```

Out[57]:

0

In [58]:

```
corona_part[['위중증률', '2차접종']].corr()
```

Out[58]:

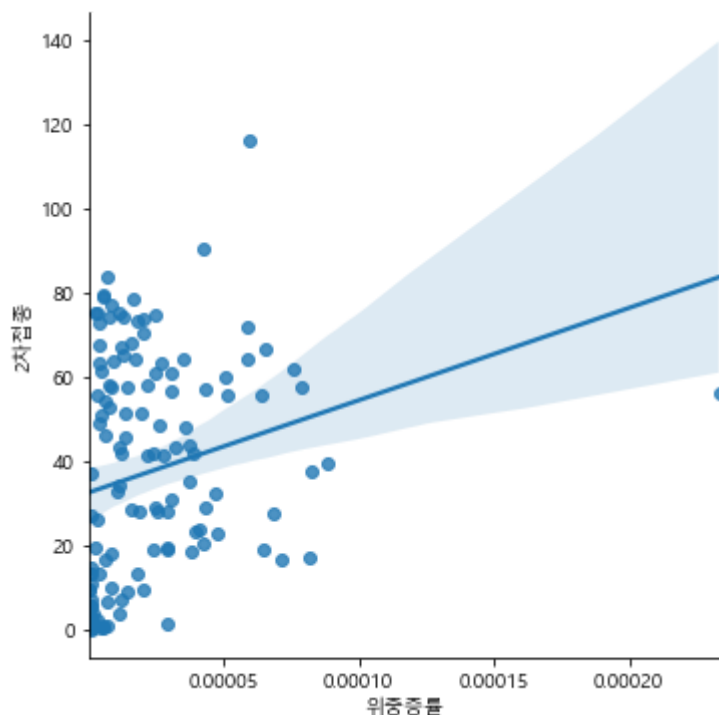
	위중증률	2차접종
위중증률	1.000000	0.231575
2차접종	0.231575	1.000000

In [59]:

```
sns.lmplot(x='위중증률', y='2차접종', data=corona_part)
```

Out[59]:

<seaborn.axisgrid.FacetGrid at 0x2638f4a4ee0>





확인 : 데이터가 있는 131개국의 인구당 접종률과 위증증률은 0.2755로 양의 상관관계를 갖는다.

In [61]:

```
corona.head(10)
```

Out[61]:

	한글 (국가명)	영문(블룸버그)	1차 접종	2차 접종	발생 률순 위	2차 접종 률순 위	확진자비 율(인구)	일별접종 수	치 명 (%)	완치 (%)	발생률	인구수
0	미국	U.S.	64.4	55.8	13.0	52.5	0.133098	701422.0	1.6	76.1	133098	331002651
1	인도	India	46.7	16.8	103.0	119.0	0.024434	7178901.0	1.3	97.8	24434	1380004385
2	브라질	Brazil	71.3	41.9	29.0	70.5	0.100592	1317163.0	2.8	95.3	100592	212559417
3	영국	U.K.	73.0	67.1	20.0	24.0	0.113959	63126.0	1.8	80.9	113959	67886011
4	러시아	Russia	32.6	28.5	77.0	89.0	0.051305	214966.0	2.8	88.9	51305	145934462
5	터키	Turkey	64.5	53.0	44.0	57.0	0.084132	381129.0	0.9	92.4	84132	84339067
6	프랑스	France	77.4	74.6	23.0	12.0	0.107278	229555.0	1.7	96.4	107278	65273511
7	이란	Iran	35.5	16.8	62.0	119.0	0.066350	1368407.0	2.2	89.9	66350	83992949
8	아르헨티나	Argentina	65.7	48.3	18.0	62.0	0.116245	222650.0	2.2	97.4	116245	45195774
9	콜롬비아	Colombia	51.4	32.8	33.0	84.0	0.097368	169329.0	2.5	96.8	97368	50882891

실습과제 : 인구 100만 이상의 국가에 대해서 다시 수행해 보자.

- 접종률과 인구당 발생률 관계 확인
- 접종률과 사망률과 중증률 관계 확인

의문 : 그렇다면 전체 국가가 아닌 30개의 국가의 데이터를 확인해 보면 어떨까?

## 02. 접종률이 높은 30개의 국가의 데이터를 확인해 보기

In [62]:



```
corona_top30 = corona[corona['2차접종률순위']<=30]
corona_top30.shape
```

Out [62]:

(30, 27)

In [63]:



corona\_top30

6	프랑스	France	77.4	74.6	23.0	12.0	0.107278	229555.0	1.7	96.4	107278	65273511	135537.0
10	스페인	Spain	81.0	78.5	26.0	6.0	0.105956	72816.0	1.7	95.8	105956	46754778	122061.0
11	이탈리아	Italy	74.5	74.5	51.0	13.0	0.077157	211345.0	2.8	95.1	77157	60461826	98872.0
24	칠레	Chile	78.5	73.8	41.0	16.0	0.086492	109701.0	2.3	97.3	86492	19116201	6944.0
25	캐나다	Canada	76.2	70.6	86.0	21.0	0.042813	105929.0	1.7	95.5	42813	37742154	44777.0
30	벨기에	Belgium	74.8	73.2	24.0	17.0	0.107012	22185.0	2.1	92.9	107012	11589623	62296.0

## Learn 05. 데이터의 인덱스를 다시 재설정하기

- `df.reset_index(inplace=True)`

In [64]:



```
corona_top30.reset_index(inplace=True)
corona_top30
```

10	65	아일랜드	Ireland	76.1	74.1	50.0	14.5	0.078419	9025.0	1.3	88.1	78419	4937786	4
11	68	덴마크	Denmark	76.7	75.1	67.0	9.5	0.061871	4355.0	0.7	97.9	61871	5792202	4
12	75	몽골	Mongolia	68.3	64.5	36.0	29.5	0.091949	16573.0	0.4	96.6	91949	3278290	4
13	78	바레인	Bahrain	78.4	75.0	5.0	11.0	0.161571	3999.0	0.5	99.2	161571	1701575	4

## Learn 06. 데이터를 정렬시키기 2가지

- `sort_index()` : 인덱스를 기준으로 정렬
- `sort_values(by=[], ascending=False, axis=0,1, inplace=True or False)` : 컬럼의 값을 기준으로 데이터 정렬
  - `by` : 정렬할 열 선택
  - `axis` : 0 : index, 1: columns

In [65]:



```
corona_top30.columns
```

Out [65]:

```
Index(['index', '한글(국가명)', '영문(블룸버그)', '1차접종', '2차접종', '발생률순위', '2차접종률순위', '확진자비율(인구)', '일별접종수', '치명(%)', '완치(%)', '발생률', '인구수', '치료중합계', '치료중1일', '위중증합계', '위중증1일', '확진자합계', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일', '백신접종수', '백신보유율(인구)', '사망률', '위중증률', '위중증합계2'], dtype='object')
```

In [66]:



```
corona_sel = ['한글(국가명)', '1차접종', '2차접종', '발생률순위', '2차접종률순위', '확진자비율(인구)',  
             '발생률', '인구수', '치료중합계', '위중증합계', '확진자합계', '사망자합계',  
             '사망률', '위중증률']  
corona_top30_part = corona_top30[corona_sel]  
corona_top30_part
```

Out[66]:

	한글(국가명)	1차접종	2차접종	발생률순위	2차접종률순위	확진자비율(인구)	발생률	인구수	치료중합계	위중증합계	확진자합계
0	영국	73.0	67.1	20.0	24.0	0.113959	113959	67886011	1344510.0	831.0	7736235
1	프랑스	77.4	74.6	23.0	12.0	0.107278	107278	65273511	135537.0	1609.0	7002393
2	스페인	81.0	78.5	26.0	6.0	0.105956	105956	46754778	122061.0	771.0	4953930
3	이탈리아	74.5	74.5	51.0	13.0	0.077157	77157	60461826	98872.0	459.0	4665049
4	칠레	78.5	73.8	41.0	16.0	0.086492	86492	19116201	6944.0	386.0	1653406
5	캐나다	76.2	70.6	86.0	21.0	0.042813	42813	37742154	44777.0	776.0	1615859
6	벨기에	74.8	73.2	24.0	17.0	0.107012	107012	11589623	62296.0	206.0	1240232
7	포르투갈	86.9	84.0	27.0	3.0	0.104792	104792	10196709	30495.0	68.0	1068530
8	아랍에미레이트	85.9	76.1	56.0	8.0	0.074388	74388	9890402	5437.0	NaN	735727
9	우루과이	77.7	72.9	21.0	19.0	0.111897	111897	3473730	1332.0	13.0	388700
10	아일랜드	76.1	74.1	50.0	14.5	0.078419	78419	4937786	40804.0	63.0	387218

	한글 (국가명)	1차접 종	2차 접종	발생 률순 위	2차 접종 률순 위	확진자비 율(인구)	발생률	인구수	치료중합 계	위중증 합계	확진자 합계
11	덴마크	76.7	75.1	67.0	9.5	0.061871	61871	5792202	4939.0	16.0	358369
12	몽골	68.3	64.5	36.0	29.5	0.091949	91949	3278290	8848.0	192.0	301434
13	바레인	78.4	75.0	5.0	11.0	0.161571	161571	1701575	733.0	5.0	274925
14	카타르	84.4	79.4	48.0	4.0	0.082108	82108	2881053	1282.0	15.0	236558
15	노르웨이	77.2	67.6	91.0	23.0	0.034733	34733	5421241	98482.0	23.0	188295
16	캄보디아	79.6	66.0	127.0	27.0	0.006679	6679	16718965	7149.0	NaN	111673
17	중국	78.6	73.0	176.0	18.0	0.000067	67	1439323776	949.0	NaN	96106
18	싱가포르	82.1	79.2	110.0	5.0	0.015687	15687	5850342	15469.0	30.0	91775
19	몰디브	105.3	90.3	8.0	2.0	0.156470	156470	540544	1579.0	23.0	84579
20	룩셈부르크	68.1	65.1	14.0	28.0	0.124602	124602	625978	1131.0	8.0	77998
21	세이셸	80.5	74.1	1.0	14.5	0.217058	217058	98347	329.0	NaN	21347
22	아루바	72.5	66.6	11.0	25.0	0.144690	144690	106766	241.0	7.0	15448
23	아이슬란드	78.8	77.2	92.0	7.0	0.034462	34462	341243	355.0	3.0	11760
24	맨섬	77.0	75.1	39.0	9.5	0.088319	88319	85033	363.0	1.0	7510

	한글 (국가명)	1차접 종	2차 접종	발생 률 순위	2차 접종 률 순위	확진자비 율(인구)	발생률	인구수	치료중합 계	위중증 합계	확진자 합계
25	지브롤터	117.4	116.4	4.0	1.0	0.164198	164198	33691	62.0	2.0	5532
26	산마리노	71.8	71.8	6.0	20.0	0.159972	159972	33931	27.0	2.0	5428
27	버뮤다	69.6	68.0	46.0	22.0	0.082613	82613	62278	1470.0	1.0	5145
28	부탄	77.1	66.3	141.0	26.0	0.003371	3371	771608	5.0	NaN	2601
29	그린란드	124.5	64.5	119.0	29.5	0.010058	10058	56770	130.0	2.0	571

In [67]:



```
# 접종률_순위 을 기준으로 정렬, 내림차순
corona_top30_part_sorted = corona_top30_part.sort_values(by=['2차접종률순위'], ascending=False)
```

In [68]:



corona\_top30\_part\_sorted

Out [68]:

	한글 (국가명)	1차접 종	2차 접종	발생 률순 위	2차 접종 률순 위	확진자비 율(인구)	발생률	인구수	치료중합 계	위중증 합계	확진자 합계
25	지브롤터	117.4	116.4	4.0	1.0	0.164198	164198	33691	62.0	2.0	5532
19	몰디브	105.3	90.3	8.0	2.0	0.156470	156470	540544	1579.0	23.0	84579
7	포르투갈	86.9	84.0	27.0	3.0	0.104792	104792	10196709	30495.0	68.0	1068530
14	카타르	84.4	79.4	48.0	4.0	0.082108	82108	2881053	1282.0	15.0	236558
18	싱가포르	82.1	79.2	110.0	5.0	0.015687	15687	5850342	15469.0	30.0	91775
2	스페인	81.0	78.5	26.0	6.0	0.105956	105956	46754778	122061.0	771.0	4953930
23	아이슬란드	78.8	77.2	92.0	7.0	0.034462	34462	341243	355.0	3.0	11760
8	아랍에미레이트	85.9	76.1	56.0	8.0	0.074388	74388	9890402	5437.0	NaN	735727
24	맨섬	77.0	75.1	39.0	9.5	0.088319	88319	85033	363.0	1.0	7510
11	덴마크	76.7	75.1	67.0	9.5	0.061871	61871	5792202	4939.0	16.0	358369
13	바레인	78.4	75.0	5.0	11.0	0.161571	161571	1701575	733.0	5.0	274925

	한글 (국가명)	1차접 종	2차 접종	발생 률순 위	2차 접종 률순 위	확진자비 율(인구)	발생률	인구수	치료중합 계	위중증 합계	확진자 합계
1	프랑스	77.4	74.6	23.0	12.0	0.107278	107278	65273511	135537.0	1609.0	7002393
3	이탈리아	74.5	74.5	51.0	13.0	0.077157	77157	60461826	98872.0	459.0	4665049
10	아일랜드	76.1	74.1	50.0	14.5	0.078419	78419	4937786	40804.0	63.0	387218
21	세이셸	80.5	74.1	1.0	14.5	0.217058	217058	98347	329.0	NaN	21347
4	칠레	78.5	73.8	41.0	16.0	0.086492	86492	19116201	6944.0	386.0	1653406
6	벨기에	74.8	73.2	24.0	17.0	0.107012	107012	11589623	62296.0	206.0	1240232
17	중국	78.6	73.0	176.0	18.0	0.000067	67	1439323776	949.0	NaN	96106
9	우루과이	77.7	72.9	21.0	19.0	0.111897	111897	3473730	1332.0	13.0	388700
26	산마리노	71.8	71.8	6.0	20.0	0.159972	159972	33931	27.0	2.0	5428
5	캐나다	76.2	70.6	86.0	21.0	0.042813	42813	37742154	44777.0	776.0	1615859
27	버뮤다	69.6	68.0	46.0	22.0	0.082613	82613	62278	1470.0	1.0	5145
15	노르웨이	77.2	67.6	91.0	23.0	0.034733	34733	5421241	98482.0	23.0	188295
0	영국	73.0	67.1	20.0	24.0	0.113959	113959	67886011	1344510.0	831.0	7736235
22	아루바	72.5	66.6	11.0	25.0	0.144690	144690	106766	241.0	7.0	15448
28	부탄	77.1	66.3	141.0	26.0	0.003371	3371	771608	5.0	NaN	2601



	한글 (국가명)	1차접 종	2차 접종	발생 률순 위	2차 접종 률순 위	확진자비 율(인구)	발생률	인구수	치료중합 계	위중증 합계	확진자 합계
16	캄보디아	79.6	66.0	127.0	27.0	0.006679	6679	16718965	7149.0	NaN	111673
20	룩셈부르크	68.1	65.1	14.0	28.0	0.124602	124602	625978	1131.0	8.0	77998
12	몽골	68.3	64.5	36.0	29.5	0.091949	91949	3278290	8848.0	192.0	301434
29	그린란	124.5	64.5	119.0	29.5	0.010058	10058	56770	130.0	2.0	571

In [69]:



```
corona_top30_part_sorted.reset_index(inplace=True)
```

In [70]:



```
corona_top30_part_sorted
```

Out[70]:

	index	한글 (국가명)	1차접 종	2차 접종	발생 률순 위	2차 접종 률순 위	확진자비 율(인구)	발생률	인구수	치료증합 계	위중증 합계	사망 수
0	25	지브롤터	117.4	116.4	4.0	1.0	0.164198	164198	33691	62.0	2.0	0
1	19	몰디브	105.3	90.3	8.0	2.0	0.156470	156470	540544	1579.0	23.0	8
2	7	포르투갈	86.9	84.0	27.0	3.0	0.104792	104792	10196709	30495.0	68.0	106
3	14	카타르	84.4	79.4	48.0	4.0	0.082108	82108	2881053	1282.0	15.0	20
4	18	싱가포르	82.1	79.2	110.0	5.0	0.015687	15687	5850342	15469.0	30.0	9
5	2	스페인	81.0	78.5	26.0	6.0	0.105956	105956	46754778	122061.0	771.0	496
6	23	아이슬란드	78.8	77.2	92.0	7.0	0.034462	34462	341243	355.0	3.0	0
7	8	아랍에미리트	85.9	76.1	56.0	8.0	0.074388	74388	9890402	5437.0	NaN	70
8	24	맨섬	77.0	75.1	39.0	9.5	0.088319	88319	85033	363.0	1.0	0
9	11	덴마크	76.7	75.1	67.0	9.5	0.061871	61871	5792202	4939.0	16.0	36
10	13	바레인	78.4	75.0	5.0	11.0	0.161571	161571	1701575	733.0	5.0	27

index	한글 (국가명)	1차접 종	2차 접종	발생 률 순 위	2차 접종 률 순 위	확진자비 율(인구)	발생률	인구수	치료중합 계	위중증 합계	사망 률
11	프랑스	77.4	74.6	23.0	12.0	0.107278	107278	65273511	135537.0	1609.0	700
12	이탈리아	74.5	74.5	51.0	13.0	0.077157	77157	60461826	98872.0	459.0	460
13	아일랜드	76.1	74.1	50.0	14.5	0.078419	78419	4937786	40804.0	63.0	38
14	세이셸	80.5	74.1	1.0	14.5	0.217058	217058	98347	329.0	NaN	2
15	칠레	78.5	73.8	41.0	16.0	0.086492	86492	19116201	6944.0	386.0	165
16	벨기에	74.8	73.2	24.0	17.0	0.107012	107012	11589623	62296.0	206.0	124
17	중국	78.6	73.0	176.0	18.0	0.000067	67	1439323776	949.0	NaN	9
18	우루과이	77.7	72.9	21.0	19.0	0.111897	111897	3473730	1332.0	13.0	38
19	산마리노	71.8	71.8	6.0	20.0	0.159972	159972	33931	27.0	2.0	
20	캐나다	76.2	70.6	86.0	21.0	0.042813	42813	37742154	44777.0	776.0	161
21	버뮤다	69.6	68.0	46.0	22.0	0.082613	82613	62278	1470.0	1.0	
22	노르웨이	77.2	67.6	91.0	23.0	0.034733	34733	5421241	98482.0	23.0	18
23	영국	73.0	67.1	20.0	24.0	0.113959	113959	67886011	1344510.0	831.0	773
24	아루바	72.5	66.6	11.0	25.0	0.144690	144690	106766	241.0	7.0	1
25	부탄	77.1	66.3	141.0	26.0	0.003371	3371	771608	5.0	NaN	

index	한글 (국가명)	1차접종	2차접종	발생률 순위	2차 접종률 순위	확진자비율 (인구)	발생률	인구수	치료중합계	위중증 합계	사망률
26	16 캄보디아	79.6	66.0	127.0	27.0	0.006679	6679	16718965	7149.0	NaN	1.0
27	20 룩셈부르크	68.1	65.1	14.0	28.0	0.124602	124602	625978	1131.0	8.0	7.0
28	12 몽골	68.3	64.5	36.0	29.5	0.091949	91949	3278290	8848.0	192.0	30.0
29	29 그린란드	124.5	64.5	119.0	29.5	0.010058	10058	56770	130.0	2.0	0.0

상관관계 확인하기

In [71]:

```
corona_top30_part_sorted[['2차접종', '확진자비율(인구)']].corr()
```

Out[71]:

	2차접종	확진자비율(인구)
2차접종	1.00000	0.35383
확진자비율(인구)	0.35383	1.00000

In [72]:

```
corona_top30_part_sorted[['2차접종', '사망률']].corr()
```

Out[72]:

	2차접종	사망률
2차접종	1.000000	0.333328
사망률	0.333328	1.000000

In [73]:



```
corona_top30_part_sorted[['2차접종', '위중증률']].corr()
```

Out[73]:

	2차접종	위중증률
2차접종	1.000000	0.027912
위중증률	0.027912	1.000000

접종률 상위 30개국의 데이터 확인 결과 큰 연관관계가 보이지 않는다.

실습 : 위중증 데이터는 처리가 안되었다. 이를 처리해 보자.

### 03. 200만 이상의 국가에 대해서 확인

In [74]:



```
corona_200m = corona[corona['인구수']>=2000000]  
corona_200m.shape
```

Out[74]:

(136, 27)

In [75]:



```
corona_200m.reset_index(inplace=True)
corona_200m.head(10)
```

Out[75]:

index		한글 (국가명)	영문(블룸버그)	1차 접종	2차 접종	발생 률순위	2차 접종 률순위	확진자비율(인구)	일별접종수	치명 (%)	완치 (%)	발생률	
0	0	미국	U.S.	64.4	55.8	13.0	52.5	0.133098	701422.0	1.6	76.1	133098	331
1	1	인도	India	46.7	16.8	103.0	119.0	0.024434	7178901.0	1.3	97.8	24434	1380
2	2	브라질	Brazil	71.3	41.9	29.0	70.5	0.100592	1317163.0	2.8	95.3	100592	212
3	3	영국	U.K.	73.0	67.1	20.0	24.0	0.113959	63126.0	1.8	80.9	113959	67
4	4	러시아	Russia	32.6	28.5	77.0	89.0	0.051305	214966.0	2.8	88.9	51305	145
5	5	터키	Turkey	64.5	53.0	44.0	57.0	0.084132	381129.0	0.9	92.4	84132	84
6	6	프랑스	France	77.4	74.6	23.0	12.0	0.107278	229555.0	1.7	96.4	107278	65
7	7	이란	Iran	35.5	16.8	62.0	119.0	0.066350	1368407.0	2.2	89.9	66350	83
8	8	아르헨티나	Argentina	65.7	48.3	18.0	62.0	0.116245	222650.0	2.2	97.4	116245	45
9	9	콜롬비아	Colombia	51.4	32.8	33.0	84.0	0.097368	169329.0	2.5	96.8	97368	50

In [76]:



```
corona_200m.columns
```

Out[76]:

```
Index(['index', '한글(국가명)', '영문(블룸버그)', '1차접종', '2차접종', '발생률순위', '2차접종률순위',  
      '확진자비율(인구)', '일별접종수', '치명(%)', '완치(%)', '발생률', '인구수',  
      '치료중합계', '치료중1일',  
      '위중증합계', '위중증1일', '확진자합계', '확진자1일', '사망자합계', '사망자1일', '완치합계', '완치1일',  
      '백신접종수', '백신보유율(인구)', '사망률', '위중증률', '위중증합계2'],  
      dtype='object')
```

In [77]:



```
corona_sel = ['한글(국가명)', '1차접종', '2차접종', '발생률순위', '2차접종률순위', '확진자비율(인구)',  
              '발생률', '인구수', '치료중합계', '위중증합계', '확진자합계', '사망자합계',  
              '사망률', '위중증률']  
corona_200m_part = corona_200m[corona_sel]  
corona_200m_part.head(10)
```

Out[77]:

	한글(국가명)	1차접종	2차접종	발생률순위	2차접종률순위	확진자비율(인구)	발생률	인구수	치료중합계	위중증합계	확진자합계
0	미국	64.4	55.8	13.0	52.5	0.133098	133098	331002651	9823512.0	21296.0	44055692
1	인도	46.7	16.8	103.0	119.0	0.024434	24434	1380004385	284976.0	8944.0	33718984
2	브라질	71.3	41.9	29.0	70.5	0.100592	100592	212559417	403027.0	8318.0	21381790
3	영국	73.0	67.1	20.0	24.0	0.113959	113959	67886011	1344510.0	831.0	7736235
4	러시아	32.6	28.5	77.0	89.0	0.051305	51305	145934462	626809.0	2300.0	7487138
5	터키	64.5	53.0	44.0	57.0	0.084132	84132	84339067	479198.0	633.0	7095580
6	프랑스	77.4	74.6	23.0	12.0	0.107278	107278	65273511	135537.0	1609.0	7002393
7	이란	35.5	16.8	62.0	119.0	0.066350	66350	83992949	443129.0	6016.0	5572962
8	아르헨티나	65.7	48.3	18.0	62.0	0.116245	116245	45195774	23643.0	1184.0	5253765
9	콜롬비아	51.4	32.8	33.0	84.0	0.097368	97368	50882891	32182.0	542.0	4954376



In [78]:



```
corona_200m_part_sorted = corona_200m_part.sort_values(by=['2차접종'], ascending=False)
corona_200m_part_sorted.head(10)
```

Out[78]:

	한글 (국가명)	1차 접종	2차 접종	발생 률순위	2차 접종 률순위	확진자비 율(인구)	발생률	인구수	치료중합 계	위중증 합계	확진자 합계	사망 합
33	포르투갈	86.9	84.0	27.0	3.0	0.104792	104792	10196709	30495.0	68.0	1068530	179
80	카타르	84.4	79.4	48.0	4.0	0.082108	82108	2881053	1282.0	15.0	236558	6
102	싱가포르	82.1	79.2	110.0	5.0	0.015687	15687	5850342	15469.0	30.0	91775	
10	스페인	81.0	78.5	26.0	6.0	0.105956	105956	46754778	122061.0	771.0	4953930	863
44	아랍에미리트	85.9	76.1	56.0	8.0	0.074388	74388	9890402	5437.0	NaN	735727	20
68	덴마크	76.7	75.1	67.0	9.5	0.061871	61871	5792202	4939.0	16.0	358369	26
6	프랑스	77.4	74.6	23.0	12.0	0.107278	107278	65273511	135537.0	1609.0	7002393	1166
11	이탈리아	74.5	74.5	51.0	13.0	0.077157	77157	60461826	98872.0	459.0	4665049	1308
65	아일랜드	76.1	74.1	50.0	14.5	0.078419	78419	4937786	40804.0	63.0	387218	52
24	칠레	78.5	73.8	41.0	16.0	0.086492	86492	19116201	6944.0	386.0	1653406	374

상관관계 확인하기

In [79]:



```
corona_200m_part_sorted[['2차접종', '확진자비율(인구)']].corr()
```

Out[79]:

	2차접종	확진자비율(인구)
2차접종	1.000000	0.609693
확진자비율(인구)	0.609693	1.000000

In [80]:



```
corona_200m_part_sorted[['2차접종', '사망률']].corr()
```

Out[80]:

	2차접종	사망률
2차접종	1.000000	0.401412
사망률	0.401412	1.000000

In [81]:



```
corona_200m_part_sorted[['2차접종', '위중증률']].corr()
```

Out[81]:

	2차접종	위중증률
2차접종	1.000000	0.158042
위중증률	0.158042	1.000000

## 결과 확인

01. 인구 200만 이상의 국가의 백신 접종률과 인구당 확진자 비율은 양의 상관관계의 관계를 갖는다.
02. 인구 200만 이상의 국가의 백신 접종률과 사망률은 양의 상관관계의 관계를 갖는다.
03. 인구 200만 이상의 국가의 백신 접종률과 위중증률은 양의 상관관계의 관계를 갖는다.(관계가 크지 않음)