In [1]:
```python
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
```

In [2]:
```python
pd.set_option('display.max_rows', 500)
```

In [3]:
```python
coviddf = pd.read_csv('covid_19_data.csv')
```

In [4]:
```python
coviddf.head()
```

Out[4]:

	Title: Cumulative COVID-19 Cases and Deaths | KFF	Unnamed: 1	Unnamed: 2	Unnamed: 3	Unnamed: 4	Unnamed: 5	Unnamed: 6
0	Timeframe: as of 3/24 at 3:26PM	NaN	NaN	NaN	NaN	NaN	NaN
1	Location	Number of COVID-19 Cases	COVID-19 Cases per 1,000,000 Population	Deaths from COVID-19	COVID-19 Deaths per 1,000,000 Population	COVID-19 Fatality Rate	Footnotes
2	United States	29,989,272	91364	544,883	1660	0.018169264	1
3	Alabama	512,711	104567	10,487	2139	0.020454018	NaN
4	Alaska	61,611	84220	312	426	0.005064031	NaN

In [5]:
```python
coviddf = pd.read_csv('covid_19_data.csv', skiprows=2)
coviddf.head()
```

Out[5]:

	Location	Number of COVID-19 Cases	COVID-19 Cases per 1,000,000 Population	Deaths from COVID-19	COVID-19 Deaths per 1,000,000 Population	COVID-19 Fatality Rate	Footnotes
0	United States	29,989,272	91364.0	544,883	1660.0	0.018169	1.0
1	Alabama	512,711	104567.0	10,487	2139.0	0.020454	NaN
2	Alaska	61,611	84220.0	312	426.0	0.005064	NaN
3	Arizona	837,849	115109.0	16,842	2314.0	0.020101	NaN
4	Arkansas	328,946	109002.0	5,547	1838.0	0.016863	NaN

In [6]:
```python
coviddf.rename(columns={"Number of COVID-19 Cases": "Num_cases"}, inplace=True)
coviddf.head()
```

Out[6]:

	Location	Num_cases	COVID-19 Cases per 1,000,000 Population	Deaths from COVID-19	COVID-19 Deaths per 1,000,000 Population	COVID-19 Fatality Rate	Footnotes
0	United States	29,989,272	91364.0	544,883	1660.0	0.018169	1.0
1	Alabama	512,711	104567.0	10,487	2139.0	0.020454	NaN
2	Alaska	61,611	84220.0	312	426.0	0.005064	NaN
3	Arizona	837,849	115109.0	16,842	2314.0	0.020101	NaN
4	Arkansas	328,946	109002.0	5,547	1838.0	0.016863	NaN

In [7]:
```python
coviddf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 69 entries, 0 to 68
Data columns (total 7 columns):
 #   Column                                    Non-Null Count  Dtype
---  ------                                    --------------  -----
 0   Location                                  64 non-null     object
 1   Num_cases                                 57 non-null     object
 2   COVID-19 Cases per 1,000,000 Population   53 non-null     float64
 3   Deaths from COVID-19                      57 non-null     object
 4   COVID-19 Deaths per 1,000,000 Population  53 non-null     float64
 5   COVID-19 Fatality Rate                    57 non-null     float64
 6   Footnotes                                 1 non-null      float64
dtypes: float64(4), object(3)
memory usage: 3.9+ KB
```

In [8]:
```python
coviddf['Num_cases']=coviddf['Num_cases'].str.replace(',', '').astype(float)
```

In [9]:
```python
coviddf['Num_cases']
```

Out[9]:
```
0     29989272.0
1       512711.0
2        61611.0
3       837849.0
4       328946.0
5      3647140.0
6       452304.0
7       300565.0
8        92609.0
9        43595.0
10     2021656.0
11     1048173.0
12       29236.0
13      178119.0
14     1227607.0
15      680046.0
16      347202.0
17      301927.0
18      421999.0
19      442221.0
20       48972.0
21      402600.0
22      617695.0
23      706550.0
24      508541.0
25      303625.0
26      576845.0
27      103666.0
28      206939.0
29      301727.0
30       81132.0
31      877936.0
32      190064.0
33     1814662.0
34      901237.0
35      102001.0
36     1004670.0
37      435634.0
38      162016.0
39      997191.0
40      134807.0
41      543925.0
42      116370.0
43      803104.0
44     2764736.0
45        2767.0
46      382171.0
47       18028.0
48      608704.0
49      356536.0
50      138818.0
51      631301.0
52       55799.0
53           0.0
54        7789.0
55         158.0
56      104888.0
57           NaN
58           NaN
59           NaN
60           NaN
61           NaN
62           NaN
63           NaN
64           NaN
65           NaN
66           NaN
67           NaN
68           NaN
Name: Num_cases, dtype: float64
```

In [10]:
```python
coviddf['Deaths from COVID-19']=coviddf['Deaths from COVID-19'].str.replace(',', '').astype(float)
```

In [11]:
```python
coviddf['Deaths from COVID-19']
```

Out[11]:
```
0     544883.0
1      10487.0
2        312.0
3      16842.0
4       5547.0
5      57824.0
6       6074.0
7       7845.0
8       1535.0
9       1053.0
10     32850.0
11     18660.0
12       453.0
13      1951.0
14     23408.0
15     12975.0
16      5689.0
17      4806.0
18      5814.0
19     10056.0
20       731.0
21      8214.0
22     16915.0
23     16949.0
24      6866.0
25      6981.0
26      8783.0
27      1431.0
28      2165.0
29      5202.0
30      1218.0
31     24292.0
32      3903.0
33     49717.0
34     11879.0
35      1495.0
36     18381.0
37      4850.0
38      2367.0
39     24874.0
40      2606.0
41      9055.0
42      1924.0
43     11747.0
44     47683.0
45        25.0
46      2082.0
47       222.0
48     10143.0
49      5186.0
50      2619.0
51      7263.0
52       695.0
53         0.0
54       134.0
55         2.0
56      2100.0
57         NaN
58         NaN
59         NaN
60         NaN
61         NaN
62         NaN
63         NaN
64         NaN
65         NaN
66         NaN
67         NaN
68         NaN
Name: Deaths from COVID-19, dtype: float64
```

In [12]:
```python
statedf= pd.read_csv('COVID19_state.csv')
statedf.head()
```

Out[12]:

	State	Tested	Infected	Deaths	Population	Pop Density	Gini	ICU Beds	Income	GDP	...	Hospitals	Health Spending	Pollution	Med-Large Airports	Temperature	Urban	Age 0-25	Age 26-54	Age 55+	School Closure Date
0	Alaska	620170	17057	84	734002	1.2863	0.4081	119	59687	73205	...	21	11064	6.4	1	26.6	66.0	0.36	0.39	0.25	3/19/2020
1	Alabama	1356420	194892	2973	4908621	96.9221	0.4847	1533	42334	45219	...	101	7281	8.1	1	62.8	59.0	0.33	0.37	0.31	3/16/2020
2	Arkansas	1363429	113641	1985	3038999	58.4030	0.4719	732	42566	42454	...	88	7408	7.1	0	60.4	56.2	0.34	0.37	0.30	3/17/2020
3	Arizona	1792602	248139	5982	7378494	64.9550	0.4713	1559	43650	48055	...	83	6452	9.7	1	60.3	89.8	0.33	0.36	0.30	3/16/2020
4	California	18912501	930628	17672	39937489	256.3727	0.4899	7338	62586	74205	...	359	7549	12.8	9	59.4	95.0	0.33	0.40	0.26	3/19/2020

5 rows × 26 columns

In [13]:
```python
statedf.rename(columns={"State": "Location"}, inplace=True)
statedf.head()
```

Out[13]:

	Location	Tested	Infected	Deaths	Population	Pop Density	Gini	ICU Beds	Income	GDP	...	Hospitals	Health Spending	Pollution	Med-Large Airports	Temperature	Urban	Age 0-25	Age 26-54	Age 55+	School Closure Date
0	Alaska	620170	17057	84	734002	1.2863	0.4081	119	59687	73205	...	21	11064	6.4	1	26.6	66.0	0.36	0.39	0.25	3/19/2020
1	Alabama	1356420	194892	2973	4908621	96.9221	0.4847	1533	42334	45219	...	101	7281	8.1	1	62.8	59.0	0.33	0.37	0.31	3/16/2020
2	Arkansas	1363429	113641	1985	3038999	58.4030	0.4719	732	42566	42454	...	88	7408	7.1	0	60.4	56.2	0.34	0.37	0.30	3/17/2020
3	Arizona	1792602	248139	5982	7378494	64.9550	0.4713	1559	43650	48055	...	83	6452	9.7	1	60.3	89.8	0.33	0.36	0.30	3/16/2020
4	California	18912501	930628	17672	39937489	256.3727	0.4899	7338	62586	74205	...	359	7549	12.8	9	59.4	95.0	0.33	0.40	0.26	3/19/2020

5 rows × 26 columns

In [14]:
```python
statedf.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 51 entries, 0 to 50
Data columns (total 26 columns):
 #   Column               Non-Null Count  Dtype
---  ------               --------------  -----
 0   Location             51 non-null     object
 1   Tested               51 non-null     int64
 2   Infected             51 non-null     int64
 3   Deaths               51 non-null     int64
 4   Population           51 non-null     int64
 5   Pop Density          51 non-null     float64
 6   Gini                 51 non-null     float64
 7   ICU Beds             51 non-null     int64
 8   Income               51 non-null     int64
 9   GDP                  51 non-null     int64
 10  Unemployment         51 non-null     float64
 11  Sex Ratio            51 non-null     float64
 12  Smoking Rate         51 non-null     float64
 13  Flu Deaths           51 non-null     float64
 14  Respiratory Deaths   51 non-null     float64
 15  Physicians           51 non-null     int64
 16  Hospitals            51 non-null     int64
 17  Health Spending      51 non-null     int64
 18  Pollution            51 non-null     float64
 19  Med-Large Airports   51 non-null     int64
 20  Temperature          51 non-null     float64
 21  Urban                51 non-null     float64
 22  Age 0-25             51 non-null     float64
 23  Age 26-54            51 non-null     float64
 24  Age 55+              51 non-null     float64
 25  School Closure Date  48 non-null     object
dtypes: float64(13), int64(11), object(2)
memory usage: 10.5+ KB
```

In [15]:
```python
finaldf=coviddf.merge(statedf, how='outer', on='Location')

finaldf.head()
```

Out[15]:

	Location	Num_cases	COVID-19 Cases per 1,000,000 Population	Deaths from COVID-19	COVID-19 Deaths per 1,000,000 Population	COVID-19 Fatality Rate	Footnotes	Tested	Infected	Deaths	...	Hospitals	Health Spending	Pollution	Med-Large Airports	Temperature	Urban	Age 0-25	Age 26-54	Age 55+	School Closure Date
0	United States	29989272.0	91364.0	544883.0	1660.0	0.018169	1.0	NaN	NaN	NaN	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	Alabama	512711.0	104567.0	10487.0	2139.0	0.020454	NaN	1356420.0	194892.0	2973.0	...	101.0	7281.0	8.1	1.0	62.8	59.0	0.33	0.37	0.31	3/16/2020
2	Alaska	61611.0	84220.0	312.0	426.0	0.005064	NaN	620170.0	17057.0	84.0	...	21.0	11064.0	6.4	1.0	26.6	66.0	0.36	0.39	0.25	3/19/2020
3	Arizona	837849.0	115109.0	16842.0	2314.0	0.020101	NaN	1792602.0	248139.0	5982.0	...	83.0	6452.0	9.7	1.0	60.3	89.8	0.33	0.36	0.30	3/16/2020
4	Arkansas	328946.0	109002.0	5547.0	1838.0	0.016863	NaN	1363429.0	113641.0	1985.0	...	88.0	7408.0	7.1	0.0	60.4	56.2	0.34	0.37	0.30	3/17/2020

5 rows × 32 columns

In [16]:
```python
coviddf.shape
```

Out[16]: (69, 7)

In [17]:
```python
statedf.shape
```

Out[17]: (51, 26)

In [18]:
```python
finaldf.shape
```

Out[18]: (69, 32)

In [19]:
```python
finaldf['Location']
```

Out[19]:
```
0                                         United States
1                                               Alabama
2                                                Alaska
3                                               Arizona
4                                              Arkansas
5                                            California
6                                              Colorado
7                                           Connecticut
8                                              Delaware
9                                  District of Columbia
10                                              Florida
11                                              Georgia
12                                               Hawaii
13                                                Idaho
14                                             Illinois
15                                              Indiana
16                                                 Iowa
17                                               Kansas
18                                             Kentucky
19                                            Louisiana
20                                                Maine
21                                             Maryland
22                                        Massachusetts
23                                             Michigan
24                                            Minnesota
25                                          Mississippi
26                                             Missouri
27                                              Montana
28                                             Nebraska
29                                               Nevada
30                                        New Hampshire
31                                           New Jersey
32                                           New Mexico
33                                             New York
34                                       North Carolina
35                                         North Dakota
36                                                 Ohio
37                                             Oklahoma
38                                               Oregon
39                                         Pennsylvania
40                                         Rhode Island
41                                       South Carolina
42                                         South Dakota
43                                            Tennessee
44                                                Texas
45                                                 USVI
46                                                 Utah
47                                              Vermont
48                                             Virginia
49                                           Washington
50                                        West Virginia
51                                            Wisconsin
52                                              Wyoming
53                                       American Samoa
54                                                 Guam
55                             Northern Mariana Islands
56                                          Puerto Rico
57                                                  NaN
58                                                  NaN
59                                                  NaN
60                                                  NaN
61                                                  NaN
62                                                Notes
63    Data source may be updated more regularly. For...
64                                              Sources
65    Johns Hopkins University, [COVID-19 Dashboard ...
66    2019 Population data from Annual Population Es...
67                                            Footnotes
68    1. US total includes territories and the 152 c...
Name: Location, dtype: object
```

In [20]:
```python
statedf['Location']=statedf['Location'].str.strip()
```

In [21]:
```python
finaldf=coviddf.merge(statedf, how='inner', on='Location')

finaldf.head()
```

Out[21]:

	Location	Num_cases	COVID-19 Cases per 1,000,000 Population	Deaths from COVID-19	COVID-19 Deaths per 1,000,000 Population	COVID-19 Fatality Rate	Footnotes	Tested	Infected	Deaths	...	Hospitals	Health Spending	Pollution	Med-Large Airports	Temperature	Urban	Age 0-25	Age 26-54	Age 55+	School Closure Date
0	Alabama	512711.0	104567.0	10487.0	2139.0	0.020454	NaN	1356420	194892	2973	...	101	7281	8.1	1	62.8	59.0	0.33	0.37	0.31	3/16/2020
1	Alaska	61611.0	84220.0	312.0	426.0	0.005064	NaN	620170	17057	84	...	21	11064	6.4	1	26.6	66.0	0.36	0.39	0.25	3/19/2020
2	Arizona	837849.0	115109.0	16842.0	2314.0	0.020101	NaN	1792602	248139	5982	...	83	6452	9.7	1	60.3	89.8	0.33	0.36	0.30	3/16/2020
3	Arkansas	328946.0	109002.0	5547.0	1838.0	0.016863	NaN	1363429	113641	1985	...	88	7408	7.1	0	60.4	56.2	0.34	0.37	0.30	3/17/2020
4	California	3647140.0	92304.0	57824.0	1463.0	0.015855	NaN	18912501	930628	17672	...	359	7549	12.8	9	59.4	95.0	0.33	0.40	0.26	3/19/2020

5 rows × 32 columns

In [22]:
```python
finaldf['Location']
```

Out[22]:
```
0                  Alabama
1                   Alaska
2                  Arizona
3                 Arkansas
4               California
5                 Colorado
6              Connecticut
7                 Delaware
8     District of Columbia
9                  Florida
10                 Georgia
11                  Hawaii
12                   Idaho
13                Illinois
14                 Indiana
15                    Iowa
16                  Kansas
17                Kentucky
18               Louisiana
19                   Maine
20                Maryland
21           Massachusetts
22                Michigan
23               Minnesota
24             Mississippi
25                Missouri
26                 Montana
27                Nebraska
28                  Nevada
29           New Hampshire
30              New Jersey
31              New Mexico
32                New York
33          North Carolina
34            North Dakota
35                    Ohio
36                Oklahoma
37                  Oregon
38            Pennsylvania
39            Rhode Island
40          South Carolina
41            South Dakota
42               Tennessee
43                   Texas
44                    Utah
45                 Vermont
46                Virginia
47              Washington
48           West Virginia
49               Wisconsin
50                 Wyoming
Name: Location, dtype: object
```

In [23]:
```python
finaldf.head()
```

Out[23]:

	Location	Num_cases	COVID-19 Cases per 1,000,000 Population	Deaths from COVID-19	COVID-19 Deaths per 1,000,000 Population	COVID-19 Fatality Rate	Footnotes	Tested	Infected	Deaths	...	Hospitals	Health Spending	Pollution	Med-Large Airports	Temperature	Urban	Age 0-25	Age 26-54	Age 55+	School Closure Date
0	Alabama	512711.0	104567.0	10487.0	2139.0	0.020454	NaN	1356420	194892	2973	...	101	7281	8.1	1	62.8	59.0	0.33	0.37	0.31	3/16/2020
1	Alaska	61611.0	84220.0	312.0	426.0	0.005064	NaN	620170	17057	84	...	21	11064	6.4	1	26.6	66.0	0.36	0.39	0.25	3/19/2020
2	Arizona	837849.0	115109.0	16842.0	2314.0	0.020101	NaN	1792602	248139	5982	...	83	6452	9.7	1	60.3	89.8	0.33	0.36	0.30	3/16/2020
3	Arkansas	328946.0	109002.0	5547.0	1838.0	0.016863	NaN	1363429	113641	1985	...	88	7408	7.1	0	60.4	56.2	0.34	0.37	0.30	3/17/2020
4	California	3647140.0	92304.0	57824.0	1463.0	0.015855	NaN	18912501	930628	17672	...	359	7549	12.8	9	59.4	95.0	0.33	0.40	0.26	3/19/2020

5 rows × 32 columns

In [24]:
```python
finaldf.drop(["Footnotes", "COVID-19 Fatality Rate"], axis = 1, inplace = True)
```

In [25]:
```python
finaldf.head()
```

Out[25]:

	Location	Num_cases	COVID-19 Cases per 1,000,000 Population	Deaths from COVID-19	COVID-19 Deaths per 1,000,000 Population	Tested	Infected	Deaths	Population	Pop Density	...	Hospitals	Health Spending	Pollution	Med-Large Airports	Temperature	Urban	Age 0-25	Age 26-54	Age 55+	School Closure Date
0	Alabama	512711.0	104567.0	10487.0	2139.0	1356420	194892	2973	4908621	96.9221	...	101	7281	8.1	1	62.8	59.0	0.33	0.37	0.31	3/16/2020
1	Alaska	61611.0	84220.0	312.0	426.0	620170	17057	84	734002	1.2863	...	21	11064	6.4	1	26.6	66.0	0.36	0.39	0.25	3/19/2020
2	Arizona	837849.0	115109.0	16842.0	2314.0	1792602	248139	5982	7378494	64.9550	...	83	6452	9.7	1	60.3	89.8	0.33	0.36	0.30	3/16/2020
3	Arkansas	328946.0	109002.0	5547.0	1838.0	1363429	113641	1985	3038999	58.4030	...	88	7408	7.1	0	60.4	56.2	0.34	0.37	0.30	3/17/2020
4	California	3647140.0	92304.0	57824.0	1463.0	18912501	930628	17672	39937489	256.3727	...	359	7549	12.8	9	59.4	95.0	0.33	0.40	0.26	3/19/2020

5 rows × 30 columns

In [26]:
```python
finaldf.describe()
```

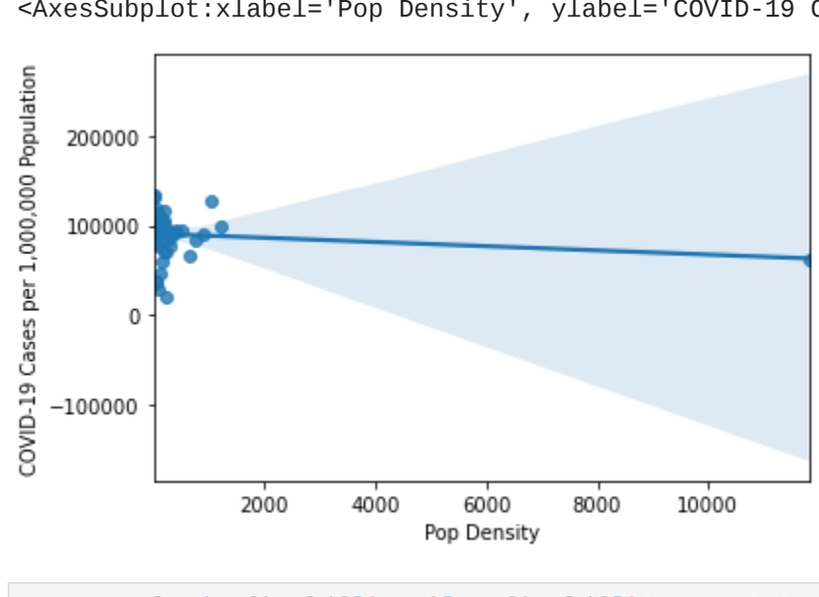
Out[26]:

	Num_cases	COVID-19 Cases per 1,000,000 Population	Deaths from COVID-19	COVID-19 Deaths per 1,000,000 Population	Tested	Infected	Deaths	Population	Pop Density	Gini	...	Physicians	Hospitals	Health Spending	Pollution	Med-Large Airports	Temperatu
count	5.100000e+01	51.000000	51.000000	51.000000	5.100000e+01	51.000000	51.000000	5.100000e+01	51.000000	51.000000	...	51.000000	51.000000	51.000000	51.000000	51.000000	51.0000
mean	5.857553e+05	90024.098039	10639.588235	1530.843137	2.904946e+06	179626.666667	4357.745098	6.496451e+06	431.560508	0.466165	...	19711.666667	101.921569	8332.156863	7.413725	1.215686	51.9999
std	6.879250e+05	24511.518654	12797.673739	591.755232	3.590449e+06	208077.881595	5637.548491	7.450657e+06	1647.225920	0.023455	...	22532.917088	88.881909	1256.751246	1.457535	1.758564	8.6279
min	1.802800e+04	20649.000000	222.000000	320.000000	1.340180e+05	2219.000000	58.000000	5.670250e+05	1.286300	0.406300	...	1172.000000	7.000000	5982.000000	4.400000	0.000000	26.6000
25%	1.504170e+05	81381.000000	2123.500000	1194.500000	8.216280e+05	46623.500000	650.500000	1.802113e+06	50.604850	0.452050	...	5656.000000	44.500000	7390.000000	6.650000	0.000000	45.3000
50%	4.026000e+05	95104.000000	6074.000000	1572.000000	1.792602e+06	120865.000000	2113.000000	4.499692e+06	108.049700	0.468000	...	12205.000000	89.000000	8107.000000	7.400000	1.000000	51.7000
75%	6.932980e+05	104102.000000	12427.000000	1894.000000	3.403572e+06	208161.000000	4983.500000	7.587794e+06	223.983100	0.479500	...	23991.500000	129.500000	9095.500000	8.150000	1.000000	58.3000
max	3.647140e+06	133849.000000	57824.000000	2735.000000	1.891250e+07	930628.000000	25838.000000	3.993749e+07	11814.541000	0.542000	...	112906.000000	523.000000	11944.000000	12.800000	9.000000	70.7000

8 rows × 28 columns

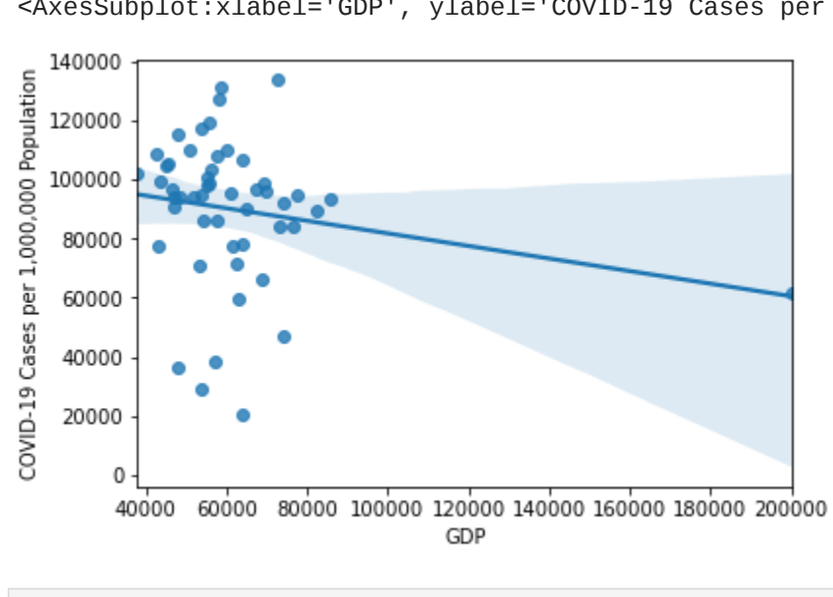
In [27]:
```python
sns.regplot(x=finaldf['Pop Density'],y=finaldf['COVID-19 Cases per 1,000,000 Population'])
```

Out[27]: <AxesSubplot:xlabel='Pop Density', ylabel='COVID-19 Cases per 1,000,000 Population'>



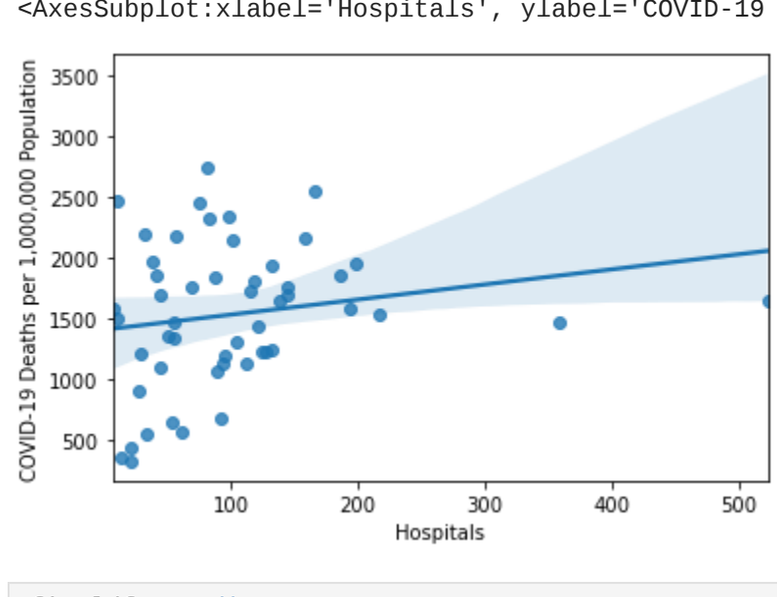
In [28]:
```python
sns.regplot(x=finaldf['GDP'],y=finaldf['COVID-19 Cases per 1,000,000 Population'])
```

Out[28]: <AxesSubplot:xlabel='GDP', ylabel='COVID-19 Cases per 1,000,000 Population'>



In [29]:
```python
sns.regplot(x=finaldf['Hospitals'],y=finaldf['COVID-19 Deaths per 1,000,000 Population'])
```

Out[29]: <AxesSubplot:xlabel='Hospitals', ylabel='COVID-19 Deaths per 1,000,000 Population'>



In [30]:
```python
finaldf.corr()
```

Out[30]:

	Num_cases	COVID-19 Cases per 1,000,000 Population	Deaths from COVID-19	COVID-19 Deaths per 1,000,000 Population	Tested	Infected	Deaths	Population	Pop Density	Gini	...	Physicians	Hospitals	Health Spending	Pollution	Med-Large Airports	Temperature	Urban	Age 0-25	A
Num_cases	1.000000	0.136232	0.958879	0.238954	0.916747	0.975152	0.849033	0.990028	-0.084996	0.388338	...	0.931321	0.880355	-0.238658	0.571439	0.914094	0.375560	0.390228	0.108069	0.1
COVID-19 Cases per 1,000,000 Population	0.136232	1.000000	0.111681	0.610667	0.060088	0.167258	0.091296	0.035472	-0.158102	0.032833	...	-0.005688	0.151686	-0.227436	0.125550	-0.042081	0.071436	0.020383	0.601388	-0.2
Deaths from COVID-19	0.958879	0.111681	1.000000	0.383514	0.931947	0.925608	0.949616	0.952162	-0.060731	0.481419	...	0.963691	0.824893	-0.132995	0.520317	0.837482	0.325377	0.420445	0.029840	0.1
COVID-19 Deaths per 1,000,000 Population	0.238954	0.610667	0.383514	1.000000	0.265757	0.238845	0.462982	0.181109	0.068205	0.569841	...	0.266642	0.186789	0.038148	0.229792	0.023077	0.268680	0.224560	0.066418	-0.0
Tested	0.916747	0.060088	0.931947	0.265757	1.000000	0.877429	0.887963	0.917761	-0.046610	0.446967	...	0.960420	0.700749	-0.053409	0.470583	0.843630	0.231321	0.431323	-0.024963	0.2
Infected	0.975152	0.167258	0.925608	0.238845	0.877429	1.000000	0.844120	0.957623	-0.087721	0.389080	...	0.873944	0.882612	-0.274788	0.510808	0.902711	0.443527	0.374364	0.127790	0.1
Deaths	0.849033	0.091296	0.949616	0.462982	0.887963	0.844120	1.000000	0.842214	-0.023913	0.522758	...	0.899385	0.692202	-0.037318	0.388125	0.725564	0.294558	0.457313	-0.041347	0.1
Population	0.990028	0.035472	0.952162	0.181109	0.917761	0.957623	0.842214	1.000000	-0.082282	0.380073	...	0.950021	0.873197	-0.219486	0.568295	0.927649	0.350538	0.405624	0.054243	0.2
Pop Density	-0.084996	-0.158102	-0.060731	0.068205	-0.046610	-0.087721	-0.023913	-0.082282	1.000000	0.506948	...	-0.032230	-0.152769	0.460614	0.251578	-0.077283	0.061071	0.326033	-0.181917	0.7
Gini	0.388338	0.032833	0.481419	0.569841	0.446967	0.389080	0.522758	0.380073	0.506948	1.000000	...	0.452432	0.291098	0.123952	0.412542	0.272112	0.469388	0.268066	-0.262721	0.4
ICU Beds	0.974746	0.075292	0.941478	0.222818	0.886886	0.962232	0.840416	0.978022	-0.079125	0.414981	...	0.928111	0.889141	-0.235379	0.569216	0.916280	0.403112	0.380334	0.053314	0.1
Income	0.175530	-0.221057	0.276819	0.066516	0.369927	0.111118	0.374086	0.216398	0.019956	0.000428	...	0.357232	-0.013090	0.456135	-0.131227	0.200799	-0.362996	0.425159	-0.308897	0.2
GDP	0.000054	-0.196773	0.046317	0.018481	0.101752	-0.031174	0.086756	0.018147	0.898326	0.390454	...	0.110092	-0.104310	0.580269	0.183789	0.017219	-0.127684	0.456731	-0.165228	0.8
Unemployment	0.018025	-0.017196	0.071412	0.206455	0.011016	-0.021688	0.059856	0.021802	0.284099	0.311418	...	0.062108	0.044494	0.247427	0.353910	0.019530	0.020485	0.044893	-0.015815	0.2
Sex Ratio	-0.177039	0.151988	-0.241952	-0.371849	-0.202278	-0.201121	-0.294413	-0.179096	-0.401153	-0.711993	...	-0.234541	-0.110359	-0.037857	-0.381696	-0.092927	-0.650249	-0.196080	0.350469	-0.2
Smoking Rate	-0.210568	0.202860	-0.204319	0.166483	-0.282789	-0.186406	-0.242660	-0.237066	-0.174545	0.075545	...	-0.262743	-0.007701	-0.003635	-0.068500	-0.258708	0.118119	-0.613050	0.045718	-0.3
Flu Deaths	-0.064482	0.183470	-0.027536	0.247967	-0.042702	-0.078609	-0.060009	-0.099613	-0.174945	0.084902	...	-0.060320	-0.008476	-0.144244	-0.084198	-0.101501	0.326962	-0.303001	0.143214	-0.2
Respiratory Deaths	-0.169970	0.319236	-0.225748	0.055056	-0.286646	-0.151578	-0.304483	-0.213398	-0.368117	-0.113053	...	-0.299129	0.022939	-0.347737	-0.035793	-0.270746	0.127467	-0.634858	0.222900	-0.4
Physicians	0.931321	-0.005688	0.963691	0.266642	0.960420	0.873944	0.899385	0.950021	-0.032230	0.452432	...	1.000000	0.770096	-0.057821	0.525119	0.851319	0.232246	0.443739	-0.047444	0.2
Hospitals	0.880355	0.151686	0.824893	0.186789	0.700749	0.882612	0.692202	0.873197	-0.152769	0.291098	...	0.770096	1.000000	-0.309285	0.483858	0.780766	0.376472	0.218999	0.244500	0.0
Health Spending	-0.238658	-0.227436	-0.132995	0.038148	-0.053409	-0.274788	-0.037318	-0.219486	0.460614	0.123952	...	-0.057821	-0.309285	1.000000	-0.238962	-0.196425	-0.491829	-0.075022	-0.514113	0.3
Pollution	0.571439	0.125550	0.520317	0.229792	0.470583	0.510808	0.388125	0.568295	0.251578	0.412542	...	0.525119	0.483858	-0.238962	1.000000	0.479477	0.404531	0.492111	0.170296	0.4
Med-Large Airports	0.914094	-0.042081	0.837482	0.023077	0.843630	0.902711	0.725564	0.927649	-0.077283	0.272112	...	0.851319	0.780766	-0.196425	0.479477	1.000000	0.370082	0.438104	0.025658	0.1
Temperature	0.375560	0.071436	0.325377	0.268680	0.231321	0.443527	0.294558	0.350538	0.061071	0.469388	...	0.232246	0.376472	-0.491829	0.404531	0.370082	1.000000	0.233079	0.070284	0.0
Urban	0.390228	0.020383	0.420445	0.224560	0.431323	0.374364	0.457313	0.405624	0.326033	0.268066	...	0.443739	0.218999	-0.075022	0.492111	0.438104	0.233079	1.000000	0.059439	0.5
Age 0-25	0.108069	0.601388	0.029840	0.066418	-0.024963	0.127790	-0.041347	0.054243	-0.181917	-0.262721	...	-0.047444	0.244500	-0.514113	0.170296	0.025658	0.070284	0.059439	1.000000	-0.1
Age 26-54	0.173281	-0.250694	0.173753	-0.043584	0.231248	0.148906	0.176299	0.204949	0.777639	0.457450	...	0.237321	0.061696	0.308054	0.404735	0.190026	0.021888	0.548735	-0.133599	1.0
Age 55+	-0.247696	-0.306562	-0.186582	-0.024459	-0.162257	-0.238493	-0.122063	-0.225115	-0.337299	-0.053329	...	-0.158354	-0.285667	0.238377	-0.409100	-0.172498	-0.070813	-0.433562	-0.753962	-0.5

28 rows × 28 columns

In [31]:
```python
from sklearn.linear_model import LinearRegression
```

In [32]:
```python
lm = LinearRegression()
```

In [33]:
```python
model = lm.fit(finaldf[['Pop Density']],finaldf['COVID-19 Cases per 1,000,000 Population'])
```

In [34]:
```python
print(f'Intercept: {model.intercept_:.3f}')
```

```
Intercept: 91039.405
```

In [35]:
```python
print (f'coefficient Pop Density: {model.coef_[0]:.3f}')
```

```
coefficient Pop Density: -2.353
```

In [36]:
```python
model = lm.fit(finaldf[['GDP']],finaldf['COVID-19 Cases per 1,000,000 Population'])
```

In [37]:
```python
print(f'Intercept: {model.intercept_:.3f}')
```

```
Intercept: 103120.638
```

In [38]:
```python
print (f'coefficient GDP: {model.coef_[0]:.3f}')
```

```
coefficient GDP: -0.213
```

In [39]:
```python
model = lm.fit(finaldf[['Hospitals']],finaldf['COVID-19 Deaths per 1,000,000 Population'])
```

In [40]:
```python
print(f'Intercept: {model.intercept_:.3f}')
```

```
Intercept: 1404.094
```

In [41]:
```python
print (f'coefficient Hospitals: {model.coef_[0]:.3f}')
```

```
coefficient Hospitals: 1.244
```

In [42]:
```python
finaldf.describe()
```

Out[42]:

	Num_cases	COVID-19 Cases per 1,000,000 Population	Deaths from COVID-19	COVID-19 Deaths per 1,000,000 Population	Tested	Infected	Deaths	Population	Pop Density	Gini	...	Physicians	Hospitals	Health Spending	Pollution	Med-Large Airports	Temperatu
count	5.100000e+01	51.000000	51.000000	51.000000	5.100000e+01	51.000000	51.000000	5.100000e+01	51.000000	51.000000	...	51.000000	51.000000	51.000000	51.000000	51.000000	51.0000
mean	5.857553e+05	90024.098039	10639.588235	1530.843137	2.904946e+06	179626.666667	4357.745098	6.496451e+06	431.560508	0.466165	...	19711.666667	101.921569	8332.156863	7.413725	1.215686	51.9999
std	6.879250e+05	24511.518654	12797.673739	591.755232	3.590449e+06	208077.881595	5637.548491	7.450657e+06	1647.225920	0.023455	...	22532.917088	88.881909	1256.751246	1.457535	1.758564	8.6279
min	1.802800e+04	20649.000000	222.000000	320.000000	1.340180e+05	2219.000000	58.000000	5.670250e+05	1.286300	0.406300	...	1172.000000	7.000000	5982.000000	4.400000	0.000000	26.6000
25%	1.504170e+05	81381.000000	2123.500000	1194.500000	8.216280e+05	46623.500000	650.500000	1.802113e+06	50.604850	0.452050	...	5656.000000	44.500000	7390.000000	6.650000	0.000000	45.3000
50%	4.026000e+05	95104.000000	6074.000000	1572.000000	1.792602e+06	120865.000000	2113.000000	4.499692e+06	108.049700	0.468000	...	12205.000000	89.000000	8107.000000	7.400000	1.000000	51.7000
75%	6.932980e+05	104102.000000	12427.000000	1894.000000	3.403572e+06	208161.000000	4983.500000	7.587794e+06	223.983100	0.479500	...	23991.500000	129.500000	9095.500000	8.150000	1.000000	58.3000
max	3.647140e+06	133849.000000	57824.000000	2735.000000	1.891250e+07	930628.000000	25838.000000	3.993749e+07	11814.541000	0.542000	...	112906.000000	523.000000	11944.000000	12.800000	9.000000	70.7000

8 rows × 28 columns

In [43]:
```python
import statsmodels.api as sm
```

In [47]:
```python
predictors= ['Pop Density','GDP','Hospitals']
outcome='COVID-19 Cases per 1,000,000 Population'
model = sm.OLS(finaldf[outcome ], finaldf[predictors].assign(const = 1))
results=model.fit()
results.summary()
```

Out[47]:

OLS Regression Results

Dep. Variable:	COVID-19 Cases per 1,000,000 Population	R-squared:	0.060
Model:	OLS	Adj. R-squared:	-0.000
Method:	Least Squares	F-statistic:	0.9968
Date:	Tue, 27 Apr 2021	Prob (F-statistic):	0.403
Time:	17:36:31	Log-Likelihood:	-585.74
No. Observations:	51	AIC:	1179.
Df Residuals:	47	BIC:	1187.
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Pop Density	2.0808	4.835	0.430	0.669	-7.646	11.807
GDP	-0.3330	0.349	-0.953	0.345	-1.036	0.370
Hospitals	38.8717	39.582	0.982	0.331	-40.756	118.500
const	1.056e+05	2.01e+04	5.264	0.000	6.53e+04	1.46e+05

Omnibus:	8.510	Durbin-Watson:	1.932
Prob(Omnibus):	0.014	Jarque-Bera (JB):	7.777
Skew:	-0.795	Prob(JB):	0.0205
Kurtosis:	4.065	Cond. No.	3.83e+05

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 3.83e+05. This might indicate that there are
strong multicollinearity or other numerical problems.

In [46]:
```python
predictors= ['Hospitals']
outcome='COVID-19 Deaths per 1,000,000 Population'
model = sm.OLS(finaldf[outcome ], finaldf[predictors].assign(const = 1))
results=model.fit()
results.summary()
```

Out[46]:

OLS Regression Results

Dep. Variable:	COVID-19 Deaths per 1,000,000 Population	R-squared:	0.035
Model:	OLS	Adj. R-squared:	0.015
Method:	Least Squares	F-statistic:	1.771
Date:	Tue, 27 Apr 2021	Prob (F-statistic):	0.189
Time:	17:36:11	Log-Likelihood:	-396.49
No. Observations:	51	AIC:	797.0
Df Residuals:	49	BIC:	800.8
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Hospitals	1.2436	0.934	1.331	0.189	-0.634	3.121
const	1404.0935	125.822	11.159	0.000	1151.246	1656.941

Omnibus:	0.819	Durbin-Watson:	1.944
Prob(Omnibus):	0.664	Jarque-Bera (JB):	0.796
Skew:	0.035	Prob(JB):	0.672
Kurtosis:	2.392	Cond. No.	206.

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

In [ ]: