[20]
```
import pandas as pd
data= pd.read_csv(r'C:\Users\maniv\Downloads\full_data Ramitha Mani.csv')
print(data)
```

```
            date     location  new_cases  new_deaths  total_cases  \
0     12/31/2019  Afghanistan          0           0            0
1       1/1/2020  Afghanistan          0           0            0
2       1/2/2020  Afghanistan          0           0            0
3       1/3/2020  Afghanistan          0           0            0
4       1/4/2020  Afghanistan          0           0            0
...          ...          ...        ...         ...          ...
6447   3/22/2020       Zambia          0           0            2
6448   3/23/2020       Zambia          1           0            3
6449   3/21/2020     Zimbabwe          1           0            1
6450   3/22/2020     Zimbabwe          1           0            2
6451   3/23/2020     Zimbabwe          0           0            2

      total_deaths
0                0
1                0
2                0
3                0
4                0
...            ...
6447             0
6448             0
6449             0
6450             0
6451             0

[6452 rows x 6 columns]
```

[22]
```
data.head()
```

	date	location	new_cases	new_deaths	total_cases
0	12/31/2019	Afghanistan	0	0	0
1	1/1/2020	Afghanistan	0	0	0
2	1/2/2020	Afghanistan	0	0	0
3	1/3/2020	Afghanistan	0	0	0
4	1/4/2020	Afghanistan	0	0	0

[23]
```
data.tail()
```

	date	location	new_cases	new_deaths	total_cases
6447	3/22/2020	Zambia	0	0	2
6448	3/23/2020	Zambia	1	0	3
6449	3/21/2020	Zimbabwe	1	0	1
6450	3/22/2020	Zimbabwe	1	0	2
6451	3/23/2020	Zimbabwe	0	0	2

[24]
```
data.head(3)
```

	date	location	new_cases	new_deaths	total_cases
0	12/31/2019	Afghanistan	0	0	0
1	1/1/2020	Afghanistan	0	0	0
2	1/2/2020	Afghanistan	0	0	0

[25]
```
data.dtypes
```

```
date            object
location        object
new_cases        int64
new_deaths       int64
total_cases      int64
total_deaths     int64
dtype: object
```

[37]
```
# we will convert the Time column to datatime format
# there are many options to ensure this works well with your data
data['Date'] = pd.to_datetime(data.date)
data.head()
```

	date	location	new_cases	new_deaths	total_cases

	date	location	new_cases	new_deaths	total_cases
0	12/31/2019	Afghanistan	0	0	0
1	1/1/2020	Afghanistan	0	0	0
2	1/2/2020	Afghanistan	0	0	0
3	1/3/2020	Afghanistan	0	0	0
4	1/4/2020	Afghanistan	0	0	0

[38]
```
data['Date'] = pd.to_datetime(data.date)
data.tail()
```

	date	location	new_cases	new_deaths	total_cases
6447	3/22/2020	Zambia	0	0	2
6448	3/23/2020	Zambia	1	0	3
6449	3/21/2020	Zimbabwe	1	0	1
6450	3/22/2020	Zimbabwe	1	0	2
6451	3/23/2020	Zimbabwe	0	0	2

[40]
```
data.dtypes
```

```
date                  object
location              object
new_cases              int64
new_deaths             int64
total_cases            int64
total_deaths           int64
```

```
Date            datetime64[ns]
dtype: object
```

[45]
```
data.Date.dt.hour.head()
```

```
0    0
1    0
2    0
3    0
4    0
Name: Date, dtype: int64
```

[46]
```
data.Date.dt.weekday.head()
```

```
0    1
1    2
2    3
3    4
4    5
Name: Date, dtype: int64
```

[47]
```
data.Date.dt.weekday_name.head()
```

```
0      Tuesday
1    Wednesday
2     Thursday
3       Friday
4     Saturday
Name: Date, dtype: object
```

[48]
```
data.Date.dt.dayofyear.head()
```

```
0    365
1      1
2      2
3      3
4      4
Name: Date, dtype: int64
```

[49]
```
ts = pd.to_datetime('12/31/1999')
```

[50]
```
data.loc[data.Date >= ts, :].head()
```

	date	location	new_cases	new_deaths	total_cases
0	12/31/2019	Afghanistan	0	0	0
1	1/1/2020	Afghanistan	0	0	0
2	1/2/2020	Afghanistan	0	0	0
3	1/3/2020	Afghanistan	0	0	0
4	1/4/2020	Afghanistan	0	0	0

[53]
```
data.Date.describe()
```

```
count                    6452
unique                     84
top       2020-03-23 00:00:00
freq                      181
first     2019-12-31 00:00:00
last      2020-03-23 00:00:00
Name: Date, dtype: object
```

[54]
```
data.Date.max()
```

```
Timestamp('2020-03-23 00:00:00')
```

[55]
```
data.Date.max() - data.Date.min()
```

```
Timedelta('83 days 00:00:00')
```

[56]
```
(data.Date.max() - data.Date.min()).days
```

```
83
```

[57]
```
%matplotlib inline
```

[58]
```
data['Date'] = data.Date.dt.date
```

[59]
```
data.head()
```

	date	location	new_cases	new_deaths	total_cases
0	12/31/2019	Afghanistan	0	0	0
1	1/1/2020	Afghanistan	0	0	0
2	1/2/2020	Afghanistan	0	0	0
3	1/3/2020	Afghanistan	0	0	0
4	1/4/2020	Afghanistan	0	0	0

[99]
```
# Set index
data.set_index('date', inplace=True)
print(data.head(5))
```

```
              location  new_cases  new_deaths  total_cases
total_deaths  \
date                                                      ...
12/31/2019  Afghanistan          0           0            0
0
1/1/2020    Afghanistan          0           0            0
0
1/2/2020    Afghanistan          0           0            0
```

```
0
1/3/2020    Afghanistan          0           0            0
0
1/4/2020    Afghanistan          0           0            0
0

                  Date
date
12/31/2019  2019-12-31
1/1/2020    2020-01-01
1/2/2020    2020-01-02
1/3/2020    2020-01-03
1/4/2020    2020-01-04
```

[102]
```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
```

[115]
```
data = pd.read_csv(r'C:\Users\maniv\Downloads\Coronavirus data by country.csv')
print(data)
```

```
         location  new_cases  new_deaths  total_cases  total_deaths  \
0     Afghanistan          0           0            0             0
1     Afghanistan          0           0            0             0
2     Afghanistan          0           0            0             0
3     Afghanistan          0           0            0             0
4     Afghanistan          0           0            0             0
...           ...        ...         ...          ...           ...
6447       Zambia          0           0            2             0
6448       Zambia          1           0            3             0
6449     Zimbabwe          1           0            1             0
6450     Zimbabwe          1           0            2             0
6451     Zimbabwe          0           0            2             0

      Unnamed: 5
0            NaN
1            NaN
2            NaN
3            NaN
4            NaN
...          ...
6447         NaN
6448         NaN
6449         NaN
6450         NaN
6451         NaN
```

[6452 rows x 6 columns]

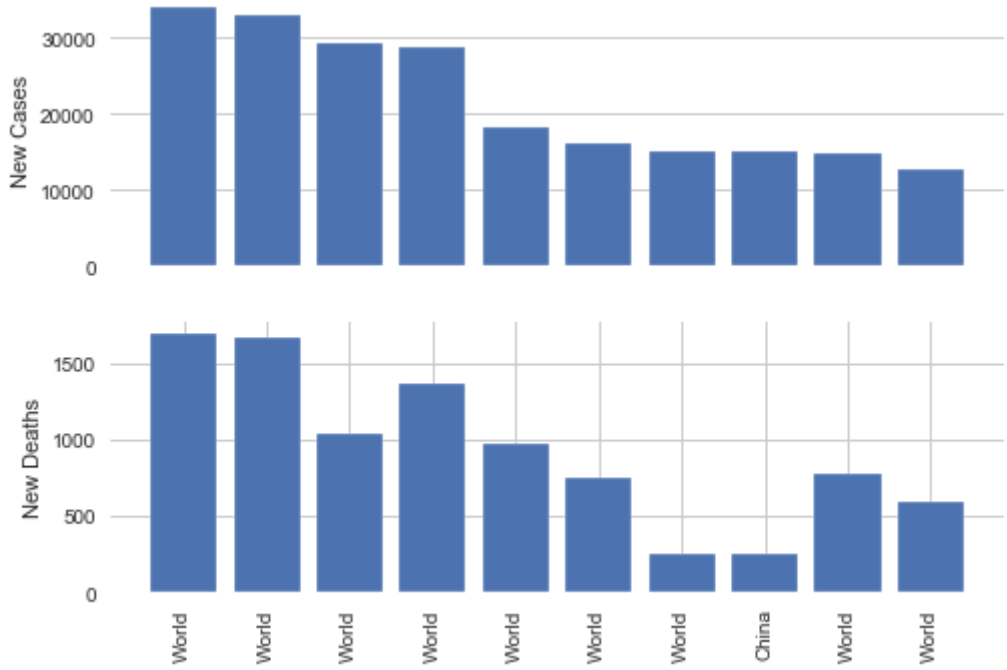
[116]
```
data.head()
```

	location	new_cases	new_deaths	total_cases	total_deaths
0	Afghanistan	0	0	0	0
1	Afghanistan	0	0	0	0
2	Afghanistan	0	0	0	0
3	Afghanistan	0	0	0	0
4	Afghanistan	0	0	0	0

[118]
```
datasort = data.sort_values('new_cases', ascending = False)
datasort = datasort.head(10)
```

[122]
```
x = range(10)
#The below code will create two plots. The parameters that
.subplot take are (row, column, no. of plots).
plt.subplot(2,1,1)
#This will create the bar graph for poulation
pop = plt.bar(x, datasort['new_cases'])
plt.ylabel('New Cases')
plt.xticks([],[])


#The below code will create the second plot.
plt.subplot(2,1,2)
#This will create the bar graph for gdp i.e gdppercapita divided
by population.
gdp =plt.bar(x, datasort['new_deaths'])
plt.ylabel('New Deaths')
plt.xticks(x, datasort['location'], rotation='vertical')
plt.show()
```

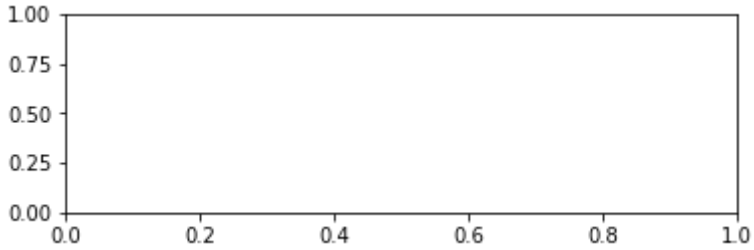


[6]
```
x = range(10)
#The below code will create two plots. The parameters that
.subplot take are (row, column, no. of plots).
plt.subplot(2,1,1)
#This will create the bar graph for poulation
pop = plt.bar(x, datasort['total_cases'])
plt.ylabel('Total Cases')
plt.xticks([],[])


#The below code will create the second plot.
plt.subplot(2,1,2)
#This will create the bar graph for gdp i.e gdppercapita divided
by population.
gdp =plt.bar(x, datasort['total_deaths'])
plt.ylabel('Total Deaths')
plt.xticks(x, datasort['location'], rotation='vertical')
plt.show()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call
last)
<ipython-input-6-df91b154d9e1> in <module>
      3 plt.subplot(2,1,1)
      4 #This will create the bar graph for poulation
----> 5 pop = plt.bar(x, datasort['total_cases'])
      6 plt.ylabel('Total Cases')
      7 plt.xticks([],[])

NameError: name 'datasort' is not defined
```

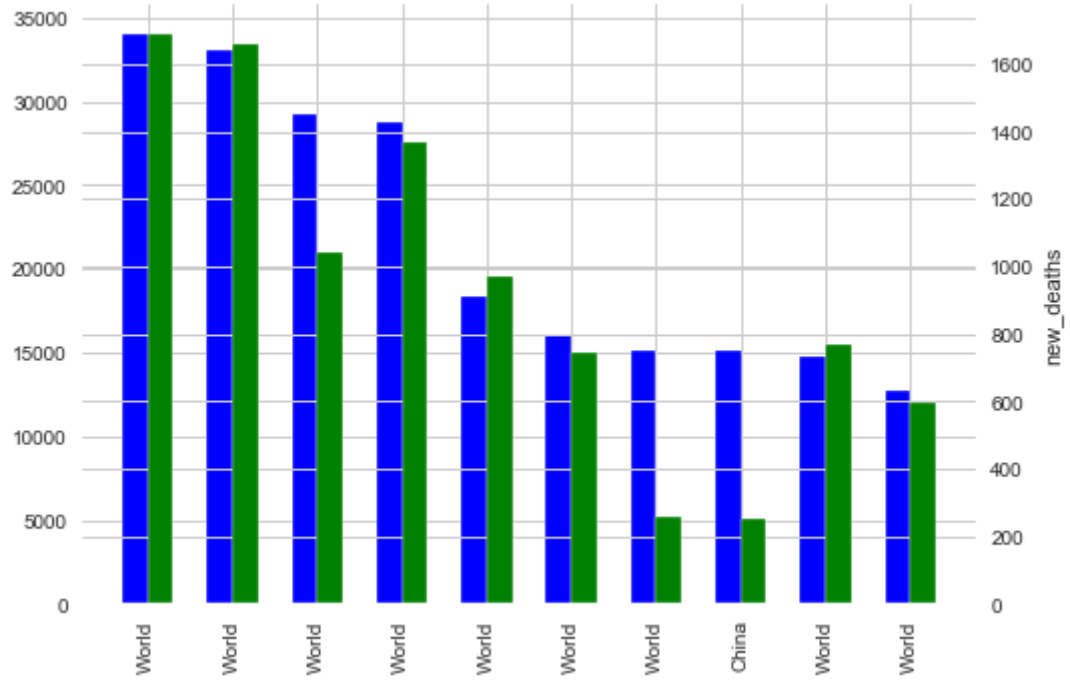


[ ]

[127]
```
x = np.arange(10)
ax1 = plt.subplot(1,1,1)
w = 0.3
#plt.xticks(), will label the bars on x axis with the respective
country names.
plt.xticks(x + w /2, datasort['location'], rotation='vertical')
pop = ax1.bar(x, datasort['new_cases'], width=w, color='b',
align='center')
#The trick is to use two different axes that share the same x
axis, we have used ax1.twinx() method.
ax2 = ax1.twinx()
#We have calculated GDP by dividing gdpPerCapita to population.
gdp =ax2.bar(x + w, datasort['new_deaths'],
width=w,color='g',align='center')
#Set the Y axis label as GDP.
plt.ylabel('new_deaths')
#To set the legend on the plot we have used plt.legend()
plt.legend([new_cases, new_deaths],['new_cases', 'new_deaths'])
#To show the plot finally we have used plt.show().
plt.show()
```

```
---------------------------------------------------------------------------
NameError                                 Traceback (most recent call
last)
<ipython-input-127-72b409b62ed8> in <module>
     12 plt.ylabel('new_deaths')
     13 #To set the legend on the plot we have used plt.legend()
---> 14 plt.legend([new_cases, new_deaths],['new_cases', 'new_deaths'])
     15 #To show the plot finally we have used plt.show().
     16 plt.show()

NameError: name 'new_deaths' is not defined
```



[7]
```
import pandas as pd
data= pd.read_csv(r'C:\Users\maniv\Downloads\Ramitha M Afghanistan.csv')
print(data)
```

```
          date     location  new_cases  new_deaths  total_cases
total_deaths
0   12/31/2019  Afghanistan          0           0            0
0
1    1/1/2020  Afghanistan          0           0            0
0
2    1/2/2020  Afghanistan          0           0            0
0
3    1/3/2020  Afghanistan          0           0            0
0
4    1/4/2020  Afghanistan          0           0            0
0
..        ...          ...        ...         ...          ...
...
69   3/19/2020  Afghanistan          0           0           22
0
70   3/20/2020  Afghanistan          0           0           22
0
71   3/21/2020  Afghanistan          2           0           24
0
72   3/22/2020  Afghanistan          0           0           24
0
73   3/23/2020  Afghanistan         10           0           34
0

[74 rows x 6 columns]
```

[8]
```
%matplotlib inline
```

[10]
```
data.head()
```

	date	location	new_cases	new_deaths	total_cases
0	12/31/2019	Afghanistan	0	0	0
1	1/1/2020	Afghanistan	0	0	0
2	1/2/2020	Afghanistan	0	0	0
3	1/3/2020	Afghanistan	0	0	0
4	1/4/2020	Afghanistan	0	0	0

[11]
```
data.tail()
```

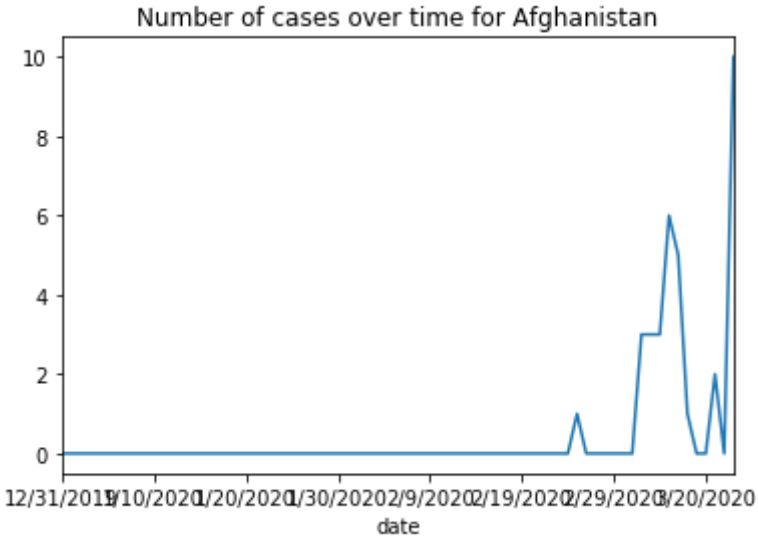
	date	location	new_cases	new_deaths	total_cases
69	3/19/2020	Afghanistan	0	0	22
70	3/20/2020	Afghanistan	0	0	22
71	3/21/2020	Afghanistan	2	0	24
72	3/22/2020	Afghanistan	0	0	24
73	3/23/2020	Afghanistan	10	0	34

[12]
```
# Set index
data.set_index('date', inplace=True)
print(data.head(5))
```

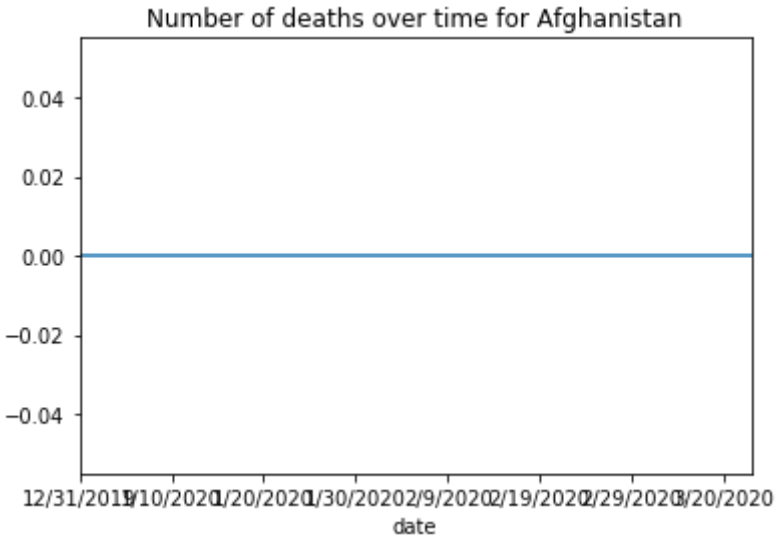
```
              location  new_cases  new_deaths  total_cases
total_deaths
date                                                      ...
12/31/2019  Afghanistan          0           0            0
0
1/1/2020    Afghanistan          0           0            0
0
1/2/2020    Afghanistan          0           0            0
0
```

```
1/3/2020    Afghanistan          0           0           0
0
1/4/2020    Afghanistan          0           0           0
0
```

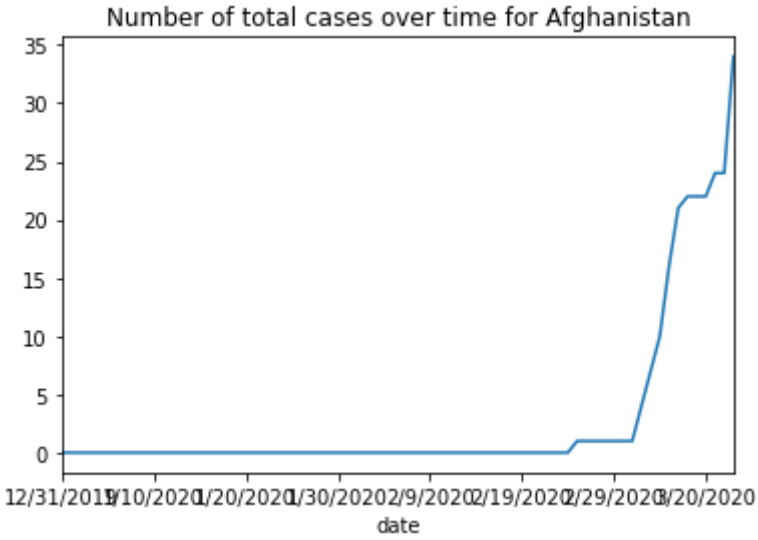
[15]
```
data.new_cases.plot(title='Number of cases over time for
Afghanistan'), data
plt.show()
```



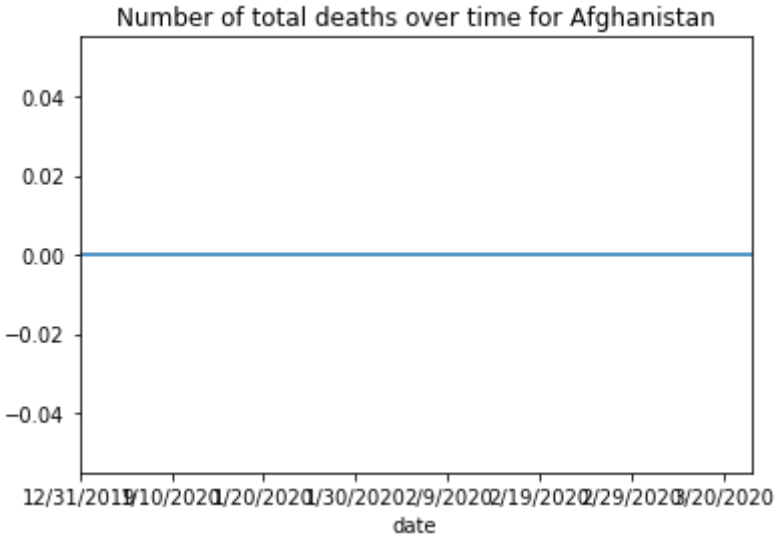
[17]
```
data.new_deaths.plot(title='Number of deaths over time for
Afghanistan'), data
plt.show()
```



[21]
```
data.total_cases.plot(title='Number of total cases over time for
Afghanistan'), data
plt.show()
```



[20]
```
data.total_deaths.plot(title='Number of total deaths over time
for Afghanistan'), data
plt.show()
```



[22]
```
import pandas as pd
data= pd.read_csv(r'C:\Users\maniv\Documents\United States
data.csv')
print(data)
```

```
          Date       Location  new_cases  new_deaths  total_cases  \
0   12/31/2019  United States          0           0            0
1    1/1/2020  United States          0           0            0
2    1/2/2020  United States          0           0            0
3    1/3/2020  United States          0           0            0
4    1/4/2020  United States          0           0            0
..        ...            ...        ...         ...          ...
74   3/14/2020  United States        511           7         2174
75   3/15/2020  United States        777          10         2951
76   3/16/2020  United States        823          12         3774
```

```
77   3/17/2020  United States        887          16         4661
78   3/18/2020  United States       1766          23         6427

    total_deaths
0              0
1              0
2              0
3              0
4              0
..           ...
74            47
75            57
76            69
77            85
78           108

[79 rows x 6 columns]
```

[23]
```
%matplotlib inline
```

[34]
```
data.head()
```

	Date	Location	new_cases	new_deaths	total_cases	tot
0	12/31/2019	World	27	0	27	0
1	1/1/2020	World	0	0	27	0
2	1/2/2020	World	0	0	27	0
3	1/3/2020	World	17	0	44	0
4	1/4/2020	World	0	0	44	0

[25]
```
data.tail()
```

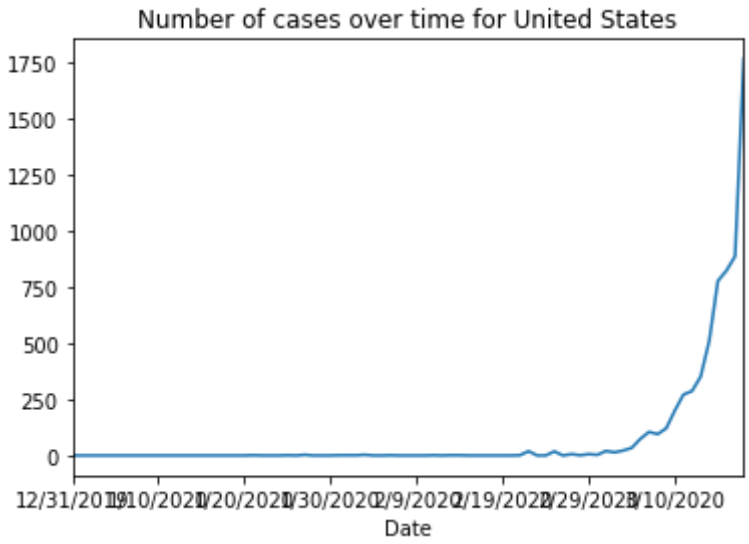
	Date	Location	new_cases	new_deaths	total_cases	tot
74	3/14/2020	United States	511	7	2174	47
75	3/15/2020	United States	777	10	2951	57

	Date	Location	new_cases	new_deaths	total_cases	tot
76	3/16/2020	United States	823	12	3774	69
77	3/17/2020	United States	887	16	4661	85
78	3/18/2020	United States	1766	23	6427	10

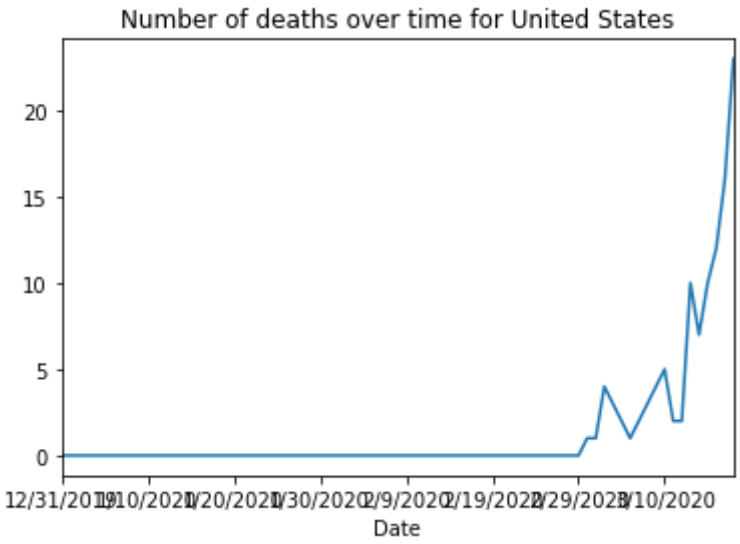
[27]
```
# Set index
data.set_index('Date', inplace=True)
print(data.head(5))
```

```
                Location  new_cases  new_deaths  total_cases
total_deaths
Date                                                        ...
12/31/2019  United States          0           0            0
0
1/1/2020    United States          0           0            0
0
1/2/2020    United States          0           0            0
0
1/3/2020    United States          0           0            0
0
1/4/2020    United States          0           0            0
0
```

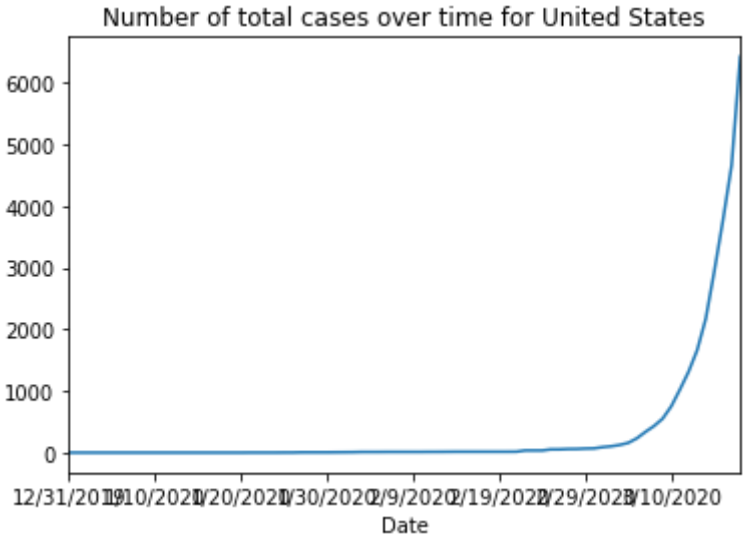
[28]
```
data.new_cases.plot(title='Number of cases over time for United
States'), data
plt.show()
```



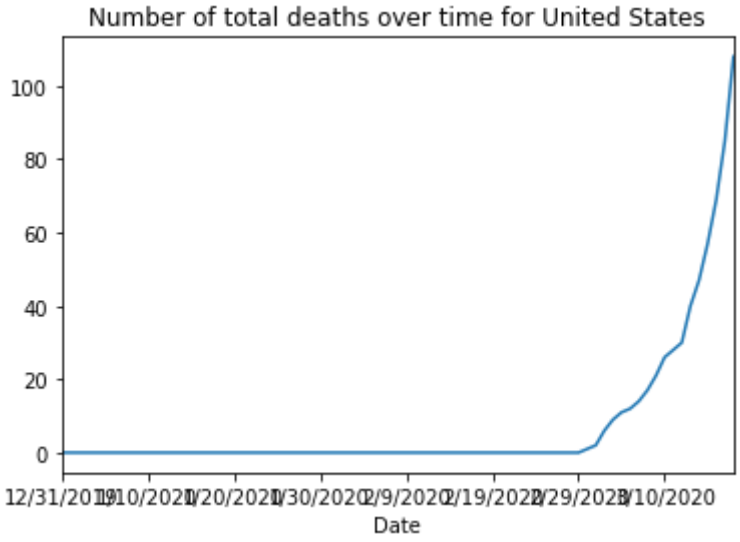
[29]
```
data.new_deaths.plot(title='Number of deaths over time for United States'), data
plt.show()
```



[30]
```
data.total_cases.plot(title='Number of total cases over time for United States'), data
plt.show()
```



[31]
```
data.total_deaths.plot(title='Number of total deaths over time for United States'), data
plt.show()
```



[36]
```
import pandas as pd
data= pd.read_csv(r'C:\Users\maniv\Documents\World data.csv')
print(data)
```

```
          Date Location  new_cases  new_deaths  total_cases
total_deaths
0   12/31/2019    World         27           0           27
0
1    1/1/2020    World          0           0           27
0
2    1/2/2020    World          0           0           27
0
3    1/3/2020    World         17           0           44
0
4    1/4/2020    World          0           0           44
0
..         ...      ...        ...         ...          ...
...
79   3/19/2020    World      18345         970       213254
8843
80   3/20/2020    World      29219        1042       242473
9885
81   3/21/2020    World      28755        1367       271228
11252
82   3/22/2020    World      34047        1690       305275
12942
83   3/23/2020    World      33032        1660       338307
14602

[84 rows x 6 columns]
```

[37]
```
%matplotlib inline
```

[38]
```
data.head()
```

	Date	Location	new_cases	new_deaths	total_cases	tot
0	12/31/2019	World	27	0	27	0
1	1/1/2020	World	0	0	27	0
2	1/2/2020	World	0	0	27	0
3	1/3/2020	World	17	0	44	0
4	1/4/2020	World	0	0	44	0

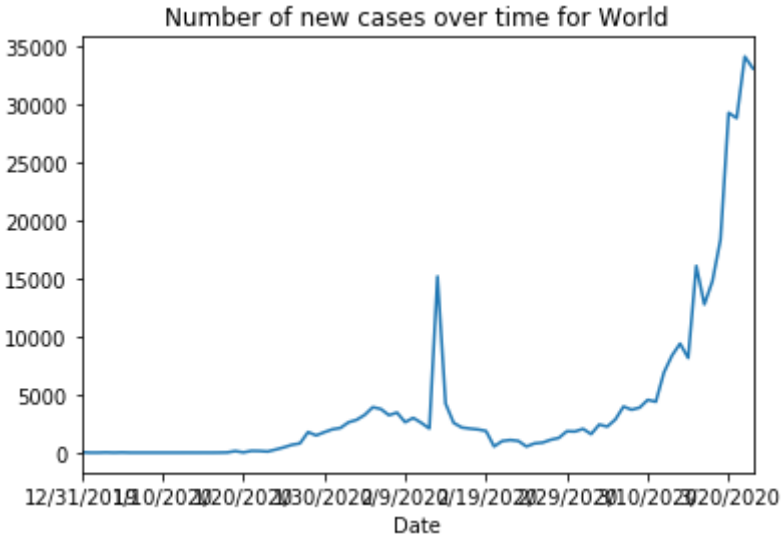
[39]
```
data.tail()
```

	Date	Location	new_cases	new_deaths	total_cases	tot
79	3/19/2020	World	18345	970	213254	88
80	3/20/2020	World	29219	1042	242473	98
81	3/21/2020	World	28755	1367	271228	11
82	3/22/2020	World	34047	1690	305275	12
83	3/23/2020	World	33032	1660	338307	14

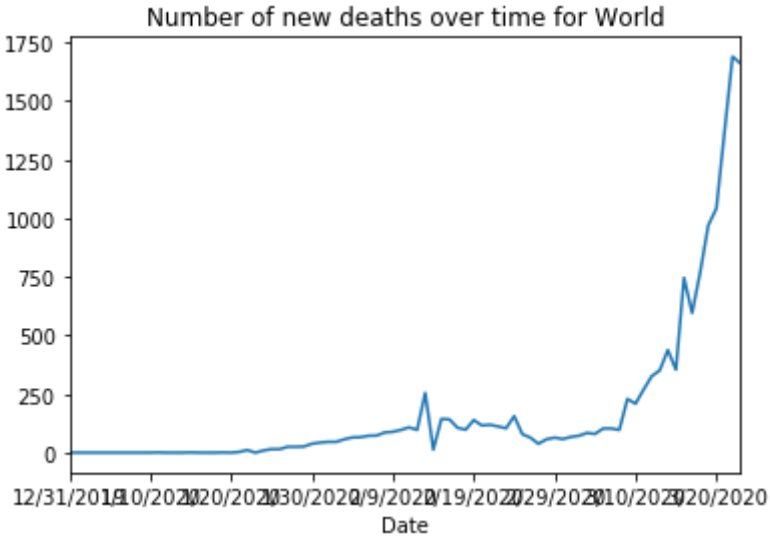
[40]
```
# Set index
data.set_index('Date', inplace=True)
print(data.head(5))
```

```
            Location  new_cases  new_deaths  total_cases  total_deaths
Date
12/31/2019     World         27           0           27             0
1/1/2020       World          0           0           27             0
1/2/2020       World          0           0           27             0
1/3/2020       World         17           0           44             0
1/4/2020       World          0           0           44             0
```

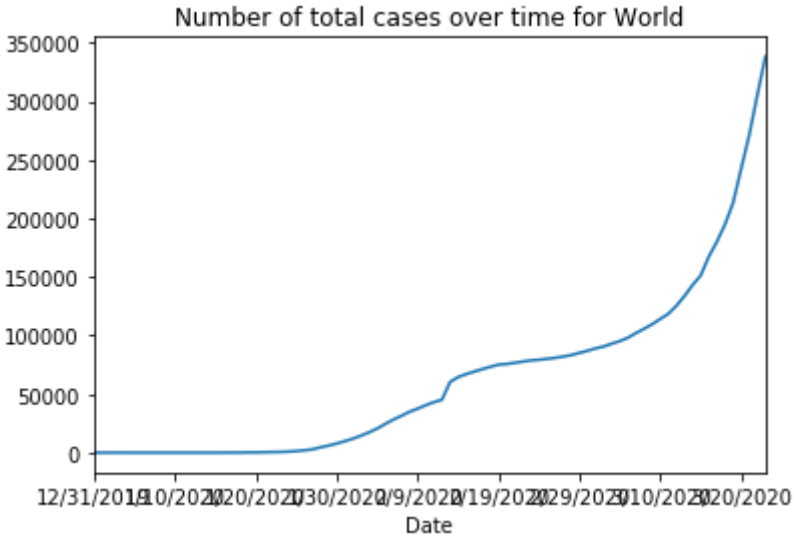
[44]
```
data.new_cases.plot(title='Number of new cases over time for
World'), data
plt.show()
```



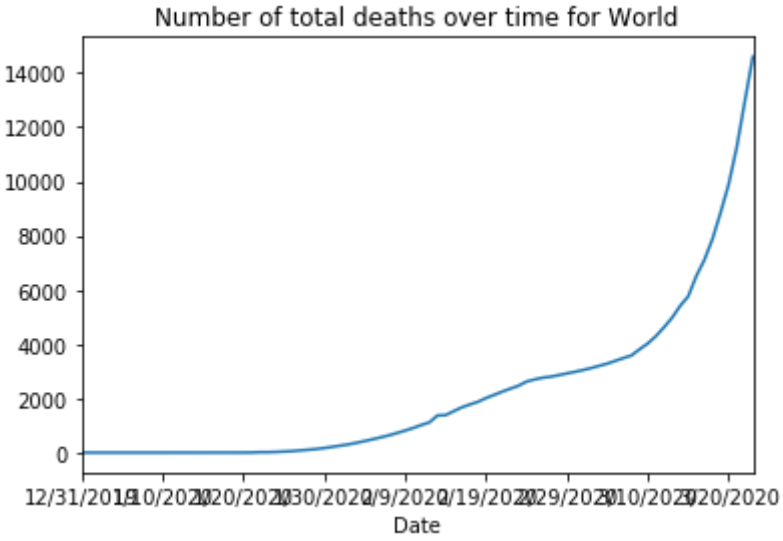
[43]
```
data.new_deaths.plot(title='Number of new deaths over time for World'), data
plt.show()
```



[45]
```
data.total_cases.plot(title='Number of total cases over time for World'), data
plt.show()
```



[47]
```
data.total_deaths.plot(title='Number of total deaths over time
for World'), data
plt.show()
```



[ ]