

# EDA-covid

April 3, 2020

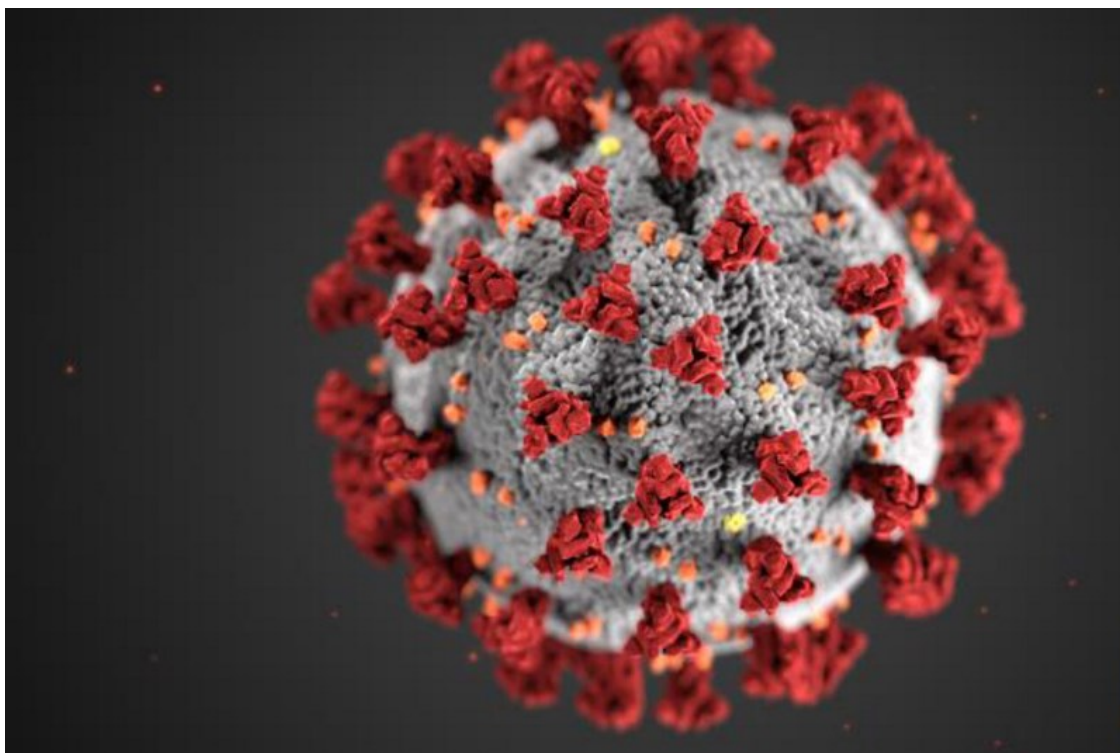
If you think the below information is helpful please UPVOTE.

Your UPVOTES keeps me boosting up to provide more insights, THANK YOU

Coronavirus disease 2019 (COVID-19) is an infectious disease caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). The disease was first identified in December 2019 in Wuhan, the capital of China's Hubei province, and has since spread globally, resulting in the ongoing 2019–20 coronavirus pandemic. Common symptoms include fever, cough, and shortness of breath. Other symptoms may include muscle pain, sputum production, diarrhea, sore throat, loss of smell, and abdominal pain. While the majority of cases result in mild symptoms, some progress to viral pneumonia and multi-organ failure. As of 3 April 2020, more than 1.03 million cases of COVID-19 have been reported in more than two hundred countries and territories, resulting in over 55 thousand deaths. More than 219 thousand people have recovered.

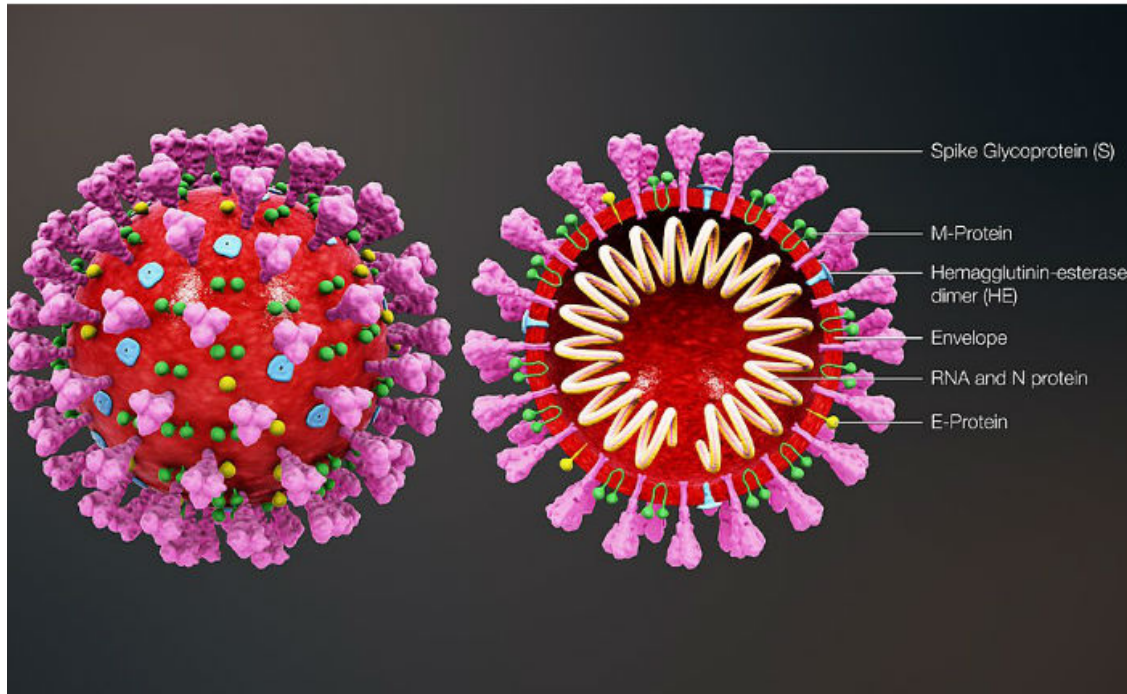
```
[1]: from IPython.display import Image  
     Image(filename='c-1.jpg')
```

[1]:



```
[2]: from IPython.display import Image
Image(filename='c-2.jpg')
```

[2]:



Preventive measures to reduce the chances of infection include staying at home, avoiding crowded places, washing hands with soap and water often and for at least 20 seconds, practicing good respiratory hygiene and avoiding touching the eyes, nose, or mouth with unwashed hands.

```
[3]: from IPython.display import Video
Video("Covid-19-Transmission-graphic-01.gif")
```

[3]: <IPython.core.display.Video object>

Imports and Datasets

Pandas - for dataset handling Numpy - Support for Pandas and calculations

Matplotlib - for visualization (Plotting graphs) plotly - for interactive plots

```
[4]: import seaborn as sns
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import warnings
```

```
warnings.filterwarnings('ignore')
import os
```

About this file

Number of covid-19 cases in India at daily leve

Context

Coronaviruses are a large family of viruses which may cause illness in animals or humans. In humans, several coronaviruses are known to cause respiratory infections ranging from the common cold to more severe diseases such as Middle East Respiratory Syndrome (MERS) and Severe Acute Respiratory Syndrome (SARS). The most recently discovered coronavirus causes coronavirus disease COVID-19 - World Health Organization

The number of new cases are increasing day by day around the world. This dataset has information from the states and union territories of India at daily level.

State level data comes from Ministry of Health & Family Welfare

Individual level data comes from covid19india

Content

COVID-19 cases at daily level is present in covid\_19\_india.csv file

Individual level details are present in IndividualDetails.csv file and is obtained from this link

Population at state level is present in population\_india\_census2011.csv file

Number of COVID-19 tests at daily level in ICMRTestingDetails.csv file

Number of hospital beds in each state in present in HospitalBedsIndia.csv file and is extracted from this link

Travel history dataset by @dheerajmpai - <https://www.kaggle.com/dheerajmpai/covidindiatravelhis>

Acknowledgements

Thanks to Indian Ministry of Health & Family Welfare for making the data available to general public.

Thanks to covid19india.org for making the individual level details and testing details available to general public.

Thanks to Wikipedia for population information.

Photo Courtesy - <https://hgis.uw.edu/virus/>

Inspiration

Looking for data based suggestions to stop / delay the spread of virus

```
[5]: covid=pd.read_csv('covid_19_india.csv')
covid.head()
```

```
[5]: Sno      Date      Time State/UnionTerritory ConfirmedIndianNational \
0    1  30/01/20  6:00 PM                Kerala                1
1    2  31/01/20  6:00 PM                Kerala                1
2    3  01/02/20  6:00 PM                Kerala                2
3    4  02/02/20  6:00 PM                Kerala                3
4    5  03/02/20  6:00 PM                Kerala                3
```

```
ConfirmedForeignNational  Cured  Deaths  Confirmed
0                        0      0      0          1
1                        0      0      0          1
2                        0      0      0          2
3                        0      0      0          3
4                        0      0      0          3
```

The above uploaded dataset consists of data which gives information about the corona virus spread, and the dataset contains columns like

serial number----> which shows the count of numbers  
Date ----->the date when the cases got registered / Date of observation  
Time -----> the time when the cases are filed / Time of observation  
State/UnionTerritory--> Name of the State / Union territory  
ConfirmedIndianNational--> Cumulative number of confirmed Indian nationals  
ConfirmedForeignNational---> Cumulative number of confirmed foreign nationals  
Cured-----> Cumulative number of cured people  
Deaths-----> Cumulative number of death cases  
Confirmed-----> Cumulative number of confirmed cases

```
[6]: covid.shape
```

```
[6]: (587, 9)
```

The data consists of 587 columns and 9 rows

```
[7]: covid.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 587 entries, 0 to 586
Data columns (total 9 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   Sno                                    587 non-null    int64
1   Date                                  587 non-null    object
2   Time                                  587 non-null    object
3   State/UnionTerritory                  587 non-null    object
4   ConfirmedIndianNational               587 non-null    object
5   ConfirmedForeignNational              587 non-null    object
6   Cured                                  587 non-null    int64
7   Deaths                                587 non-null    int64
8   Confirmed                             587 non-null    int64
```

```
dtypes: int64(4), object(5)
memory usage: 41.4+ KB
```

```
[8]: covid.columns

# The names of the columns present in the dataset
```

```
[8]: Index(['Sno', 'Date', 'Time', 'State/UnionTerritory',
          'ConfirmedIndianNational', 'ConfirmedForeignNational', 'Cured',
          'Deaths', 'Confirmed'],
          dtype='object')
```

```
[9]: covid.describe()
```

```
[9]:
```

	Sno	Cured	Deaths	Confirmed
count	587.000000	587.000000	587.000000	587.000000
mean	294.000000	1.887564	0.504259	23.301533
std	169.59658	4.771678	1.258920	42.272148
min	1.000000	0.000000	0.000000	1.000000
25%	147.50000	0.000000	0.000000	2.000000
50%	294.00000	0.000000	0.000000	7.000000
75%	440.50000	1.500000	1.000000	27.000000
max	587.00000	42.000000	13.000000	335.000000

```
[10]: covid.isnull().sum()
```

```
[10]: Sno                0
      Date                0
      Time                0
      State/UnionTerritory  0
      ConfirmedIndianNational  0
      ConfirmedForeignNational  0
      Cured                0
      Deaths                0
      Confirmed            0
      dtype: int64
```

```
[11]: covid= covid.drop(['Sno'],axis=1)
```

```
[12]: covid.head()
```

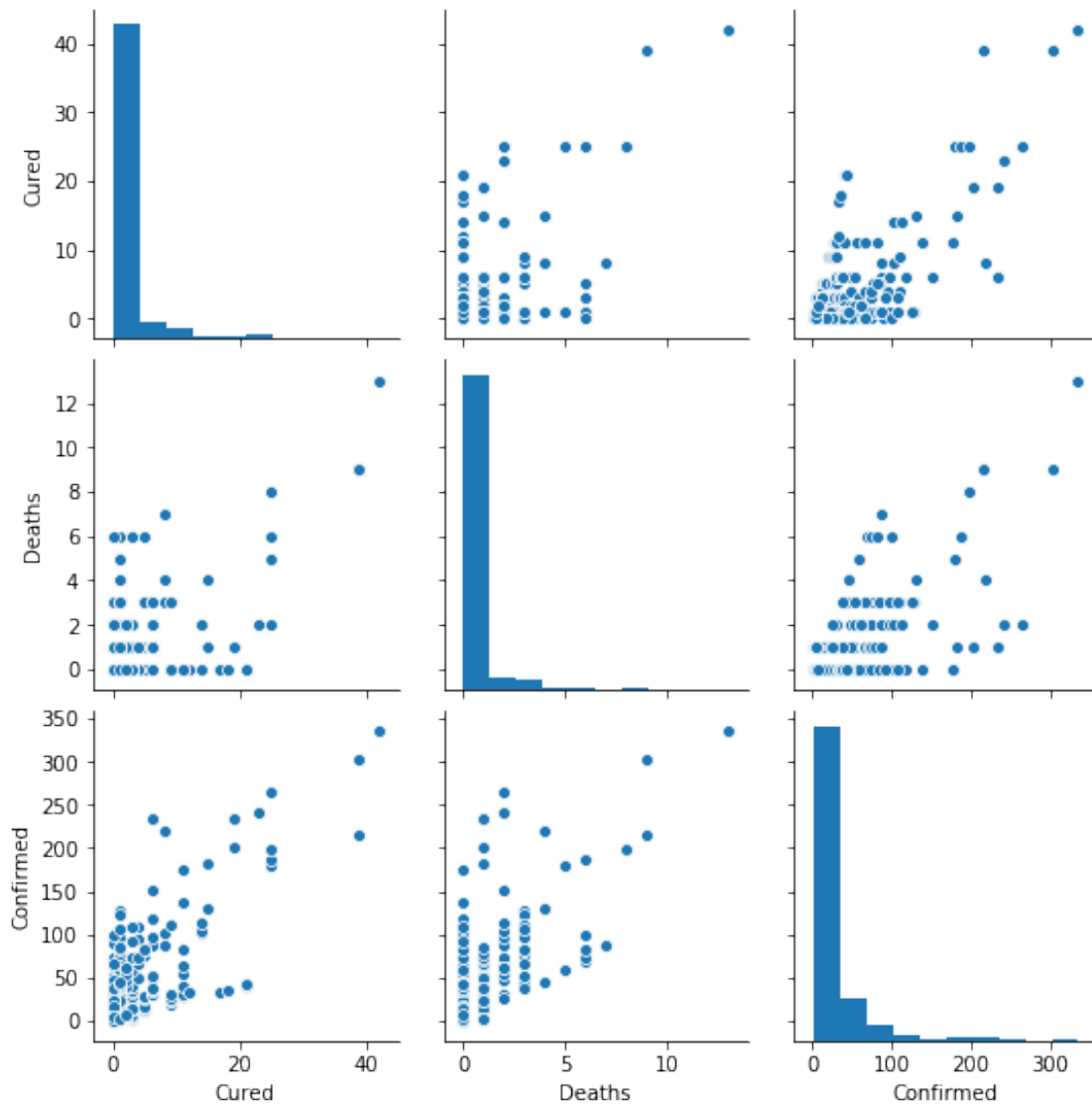
```
[12]:
```

	Date	Time	State/UnionTerritory	ConfirmedIndianNational	\
0	30/01/20	6:00 PM	Kerala	1	
1	31/01/20	6:00 PM	Kerala	1	
2	01/02/20	6:00 PM	Kerala	2	
3	02/02/20	6:00 PM	Kerala	3	
4	03/02/20	6:00 PM	Kerala	3	

	ConfirmedForeignNational	Cured	Deaths	Confirmed
0	0	0	0	1
1	0	0	0	1
2	0	0	0	2
3	0	0	0	3
4	0	0	0	3

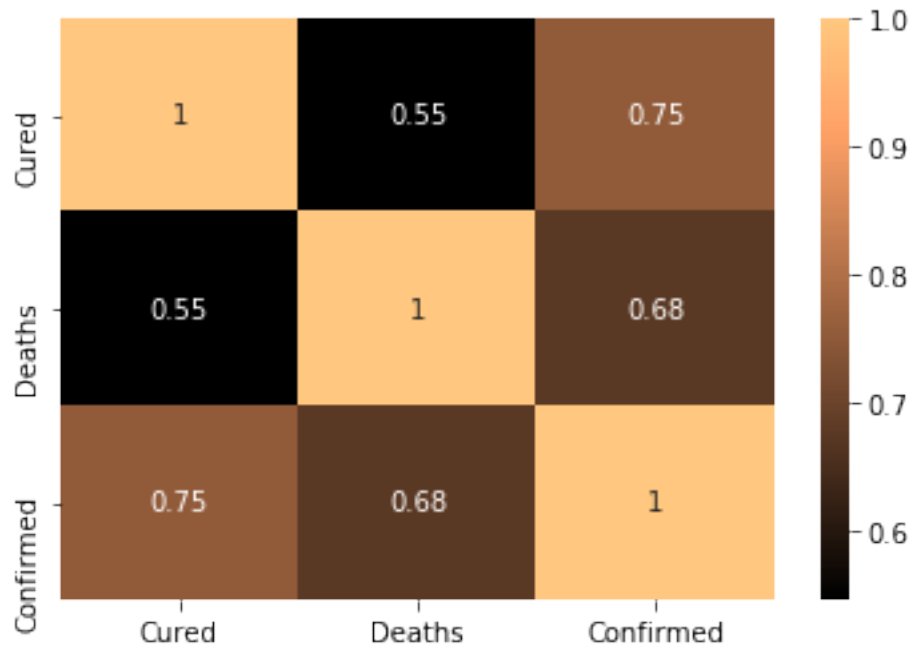
```
[13]: sns.pairplot(covid)
```

```
[13]: <seaborn.axisgrid.PairGrid at 0x19807709128>
```



```
[34]: sns.heatmap(covid.corr() , cmap = 'copper' , annot = True)
```

[34]: <matplotlib.axes.\_subplots.AxesSubplot at 0x198002cacc0>



```
[14]: group = covid.groupby('State/UnionTerritory')[ 'Confirmed', 'Cured', 'Deaths' ].  
      ↪sum().reset_index()  
group
```

```
[14]:
```

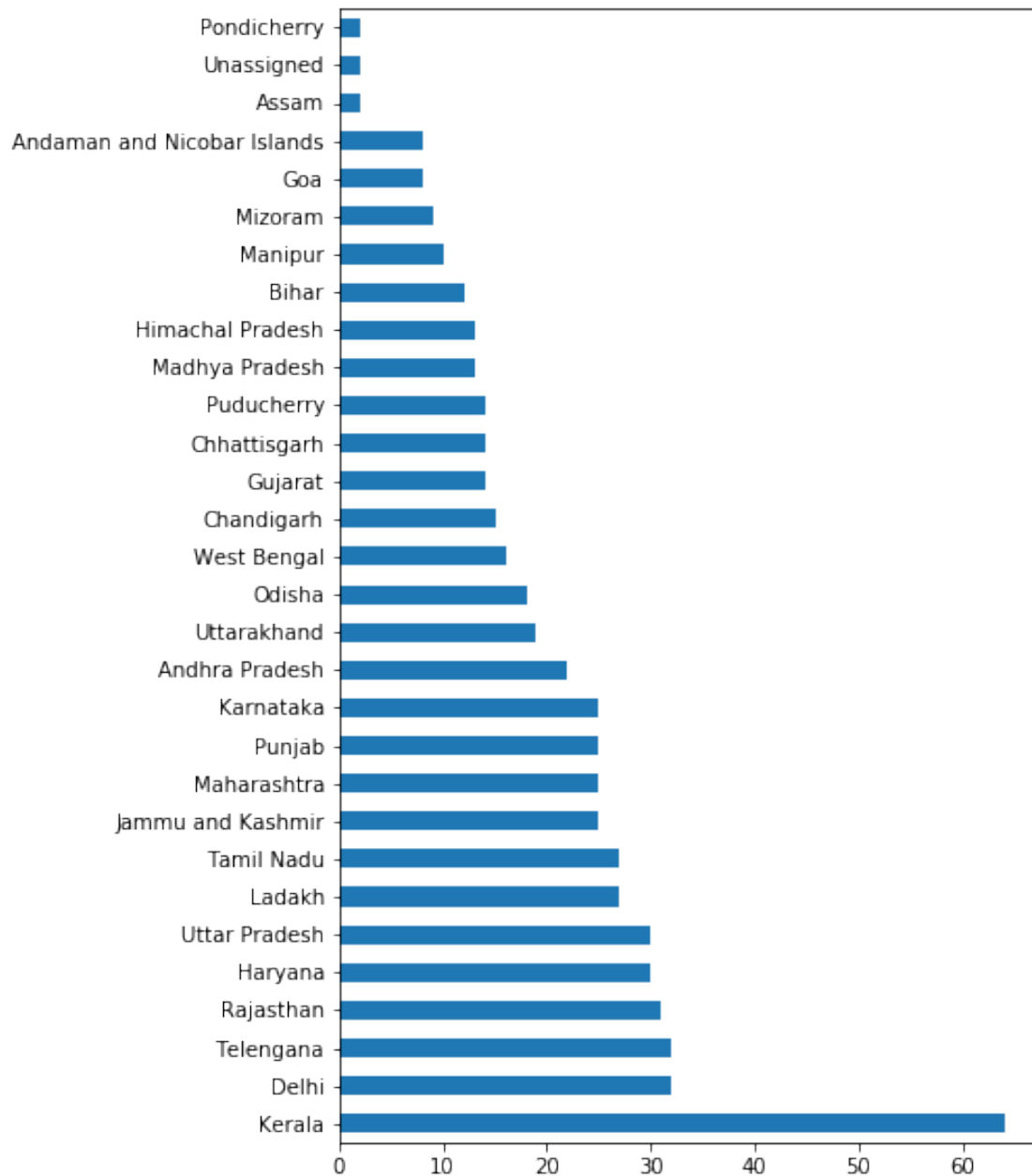
	State/UnionTerritory	Confirmed	Cured	Deaths
0	Andaman and Nicobar Islands	56	0	0
1	Andhra Pradesh	332	9	1
2	Assam	6	0	0
3	Bihar	120	0	12
4	Chandigarh	111	0	0
5	Chhattisgarh	1	0	0
6	Chhattisgarh	64	4	0
7	Delhi	967	94	27
8	Goa	34	0	0
9	Gujarat	630	18	43
10	Haryana	572	154	0
11	Himachal Pradesh	36	1	10
12	Jammu and Kashmir	352	13	12
13	Jharkhand	2	0	0
14	Karnataka	858	53	34
15	Kerala	2284	200	7
16	Ladakh	227	21	0
17	Madhya Pradesh	394	0	21

18	Maharashtra	2384	212	71
19	Manipur	10	0	0
20	Mizoram	9	0	0
21	Odisha	40	0	0
22	Pondicherry	2	0	0
23	Puducherry	18	2	0
24	Punjab	436	6	22
25	Rajasthan	820	58	0
26	Tamil Nadu	822	35	8
27	Telangana	762	19	9
28	Unassigned	84	0	0
29	Uttar Pradesh	943	187	4
30	Uttarakhand	75	10	0
31	West Bengal	227	12	15

```
[15]: covid['State/UnionTerritory'].value_counts().head(30).plot(kind='barh',
    ↪figsize=(6,10))
```

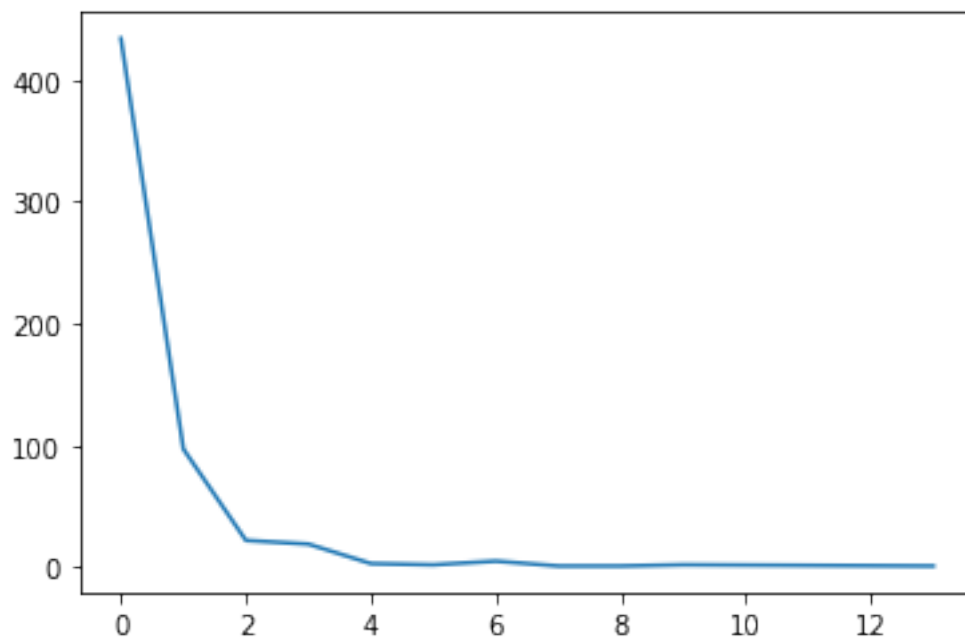
```
[15]: <matplotlib.axes._subplots.AxesSubplot at 0x198081054e0>
```





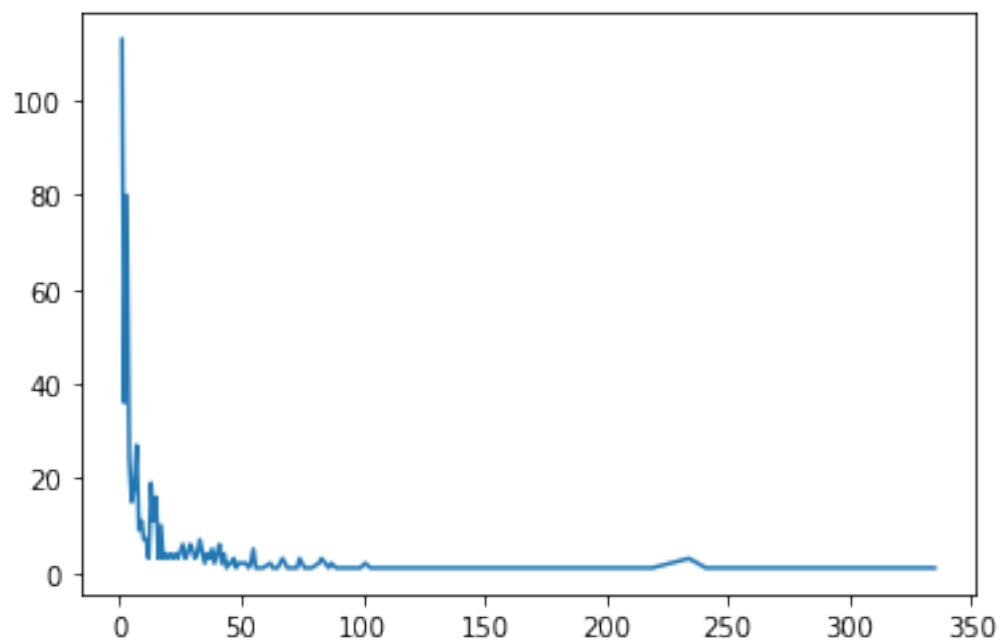
```
[16]: plt.plot(covid['Deaths'].value_counts().sort_index())
```

```
[16]: [<matplotlib.lines.Line2D at 0x19808353b00>]
```



```
[17]: plt.plot(covid['Confirmed'].value_counts().sort_index())
plt.figure(figsize=(15,4))
```

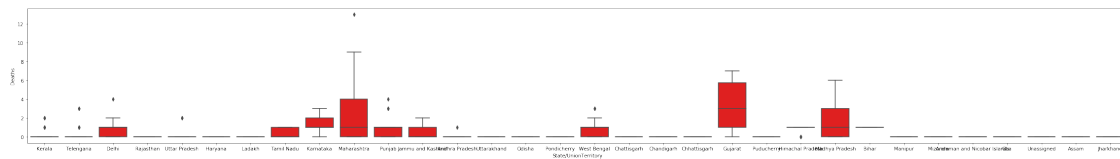
[17]: <Figure size 1080x288 with 0 Axes>



<Figure size 1080x288 with 0 Axes>

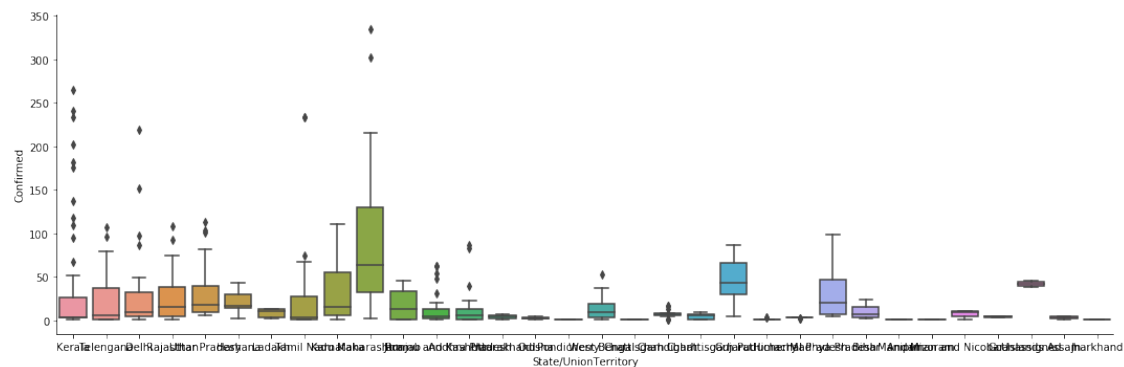
```
[18]: import seaborn as sns
plt.figure(figsize=(40,5))
sns.boxplot(
    data=covid,
    x='State/UnionTerritory',
    y='Deaths',
    color='red')
```

[18]: <matplotlib.axes.\_subplots.AxesSubplot at 0x19808232be0>



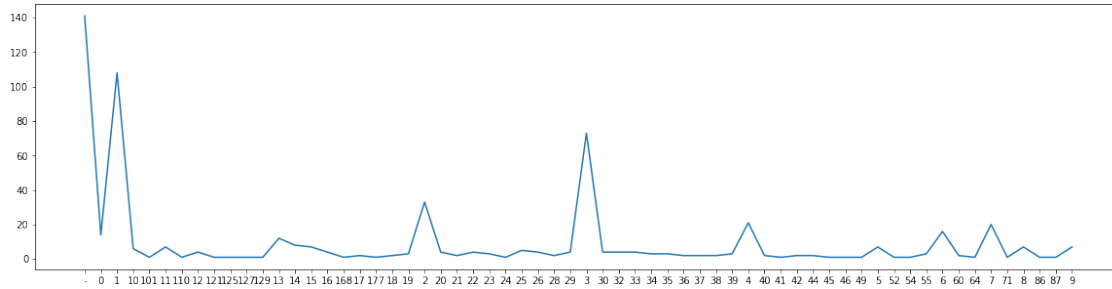
```
[19]: sns.catplot(
    data=covid,
    x='State/UnionTerritory',
    y='Confirmed',
    kind='box',
    height=5,
    aspect=3)
```

[19]: <seaborn.axisgrid.FacetGrid at 0x19808a8ca58>



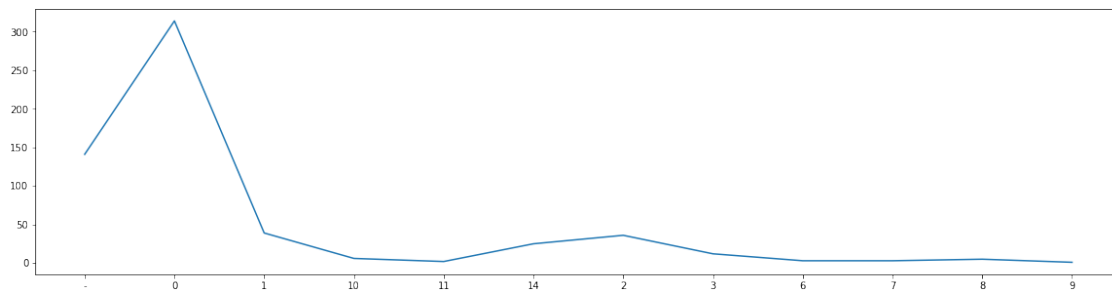
```
[20]: plt.figure(figsize=(20,5))
plt.plot(covid['ConfirmedIndianNational'].value_counts().sort_index())
```

[20]: [<matplotlib.lines.Line2D at 0x19808e10f60>]



```
[21]: plt.figure(figsize=(20,5))
plt.plot(covid['ConfirmedForeignNational'].value_counts().sort_index())
```

```
[21]: [<matplotlib.lines.Line2D at 0x198090712b0>]
```



```
[22]: fig1 = covid[['Date', 'State/UnionTerritory', 'Confirmed', 'Cured', 'Deaths']]
fig1
```

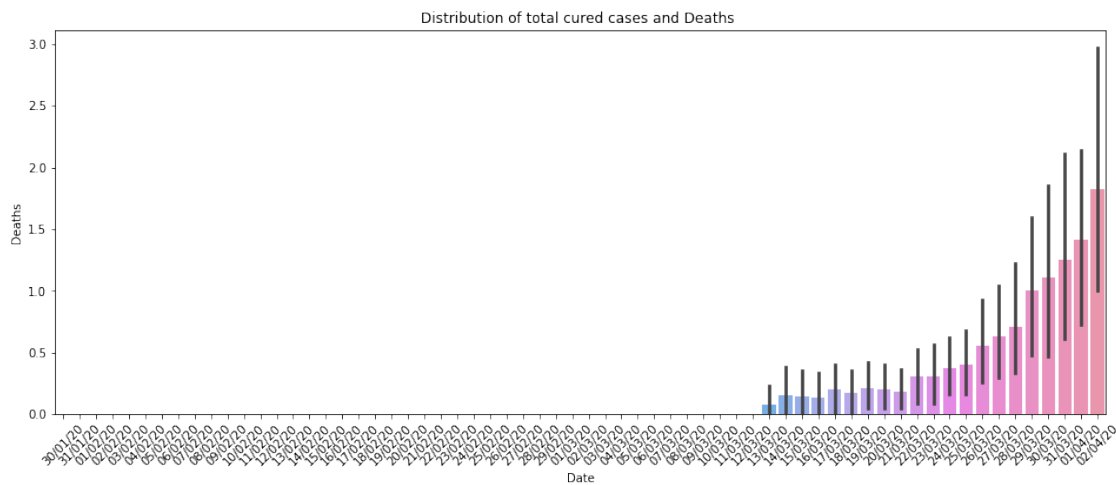
```
[22]:
```

	Date	State/UnionTerritory	Confirmed	Cured	Deaths
0	30/01/20	Kerala	1	0	0
1	31/01/20	Kerala	1	0	0
2	01/02/20	Kerala	2	0	0
3	02/02/20	Kerala	3	0	0
4	03/02/20	Kerala	3	0	0
..	...	...	...	...	...
582	02/04/20	Tamil Nadu	234	6	1
583	02/04/20	Telangana	107	1	3
584	02/04/20	Uttarakhand	7	2	0
585	02/04/20	Uttar Pradesh	113	14	2
586	02/04/20	West Bengal	53	6	3

```
[587 rows x 5 columns]
```

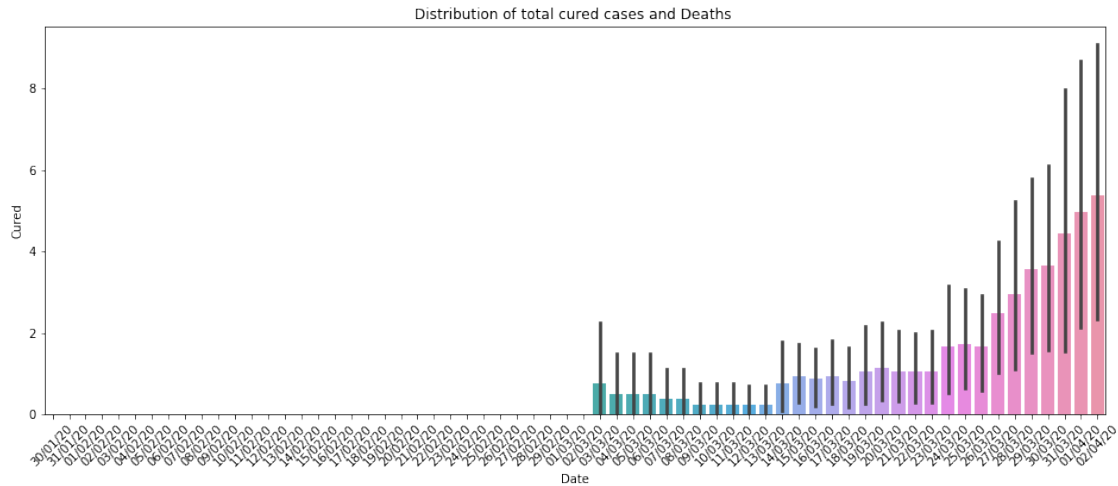
```
[31]: plt.figure(figsize=(16,6))
sns.barplot(x='Date',y='Deaths',data=covid, order=covid.Date.unique().tolist())
plt.title('Distribution of total cured cases and Deaths ')
plt.xticks(rotation=45)
```

```
[31]: (array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
        17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
        34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50,
        51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63]),
      <a list of 64 Text xticklabel objects>)
```



```
[32]: plt.figure(figsize=(16,6))
sns.barplot(x='Date',y='Cured',data=covid, order=covid.Date.unique().tolist())
plt.title('Distribution of total cured cases and Deaths ')
plt.xticks(rotation=45)
```

```
[32]: (array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11, 12, 13, 14, 15, 16,
        17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33,
        34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50,
        51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63]),
      <a list of 64 Text xticklabel objects>)
```



```
[35]: age=pd.read_csv('AgeGroupDetails.csv')
age.head()
```

```
[35]:
```

	Sno	AgeGroup	TotalCases	Percentage
0	1	0-9	22	3.18%
1	2	10-19	27	3.90%
2	3	20-29	172	24.86%
3	4	30-39	146	21.10%
4	5	40-49	112	16.18%

```
[36]: age = age.drop(['Sno'],axis=1)
```

```
[37]: age.head()
```

```
[37]:
```

	AgeGroup	TotalCases	Percentage
0	0-9	22	3.18%
1	10-19	27	3.90%
2	20-29	172	24.86%
3	30-39	146	21.10%
4	40-49	112	16.18%

```
[39]: age.isnull().sum()
```

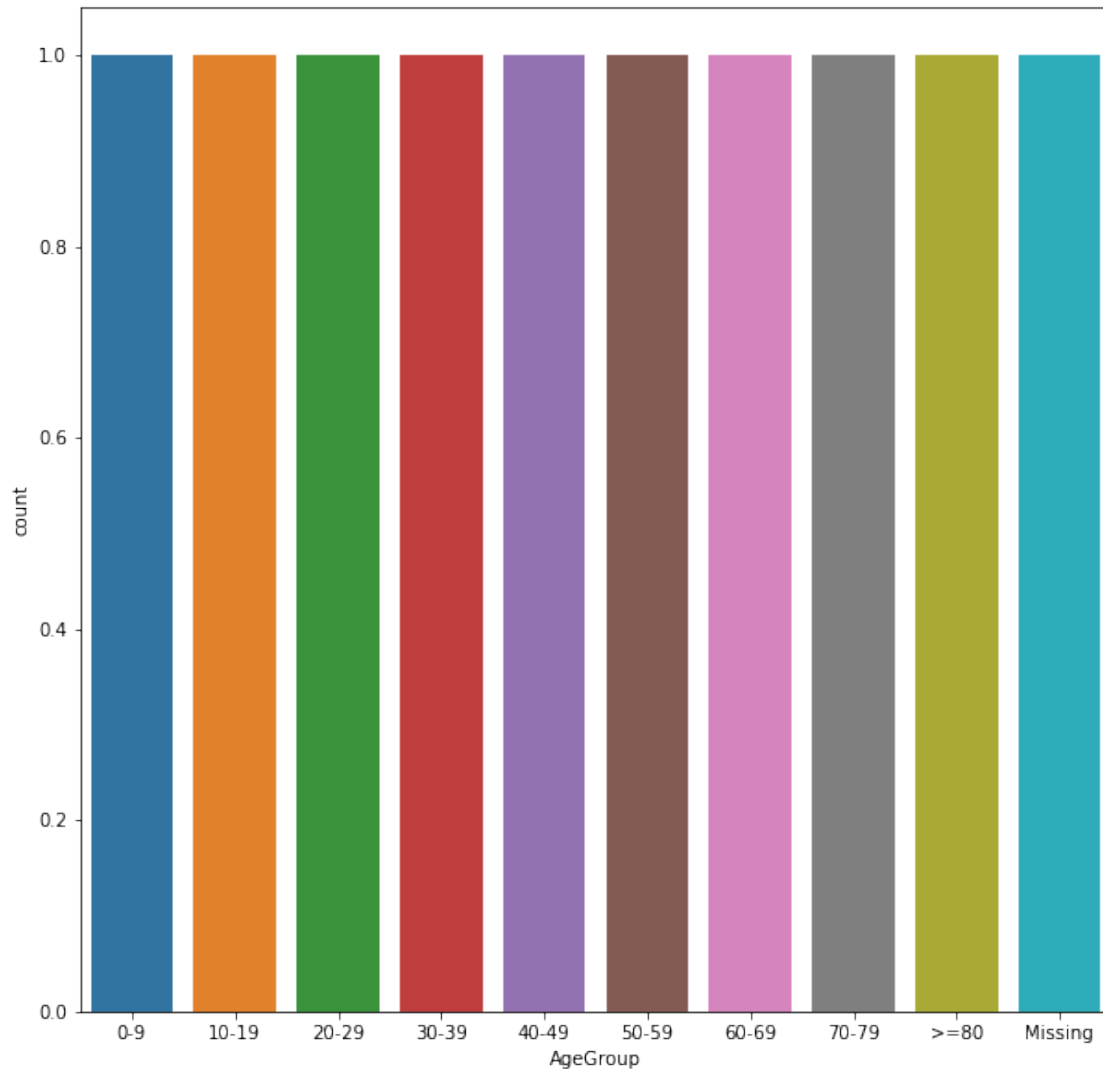
```
[39]: AgeGroup      0
TotalCases      0
Percentage      0
dtype: int64
```

```
[41]: age.shape
```

[41]: (10, 3)

```
[38]: plt.figure(figsize = (10,10))
      sns.countplot(age['AgeGroup'])
```

[38]: <matplotlib.axes.\_subplots.AxesSubplot at 0x1980b5b1a90>



```
[45]: covid.head()
```

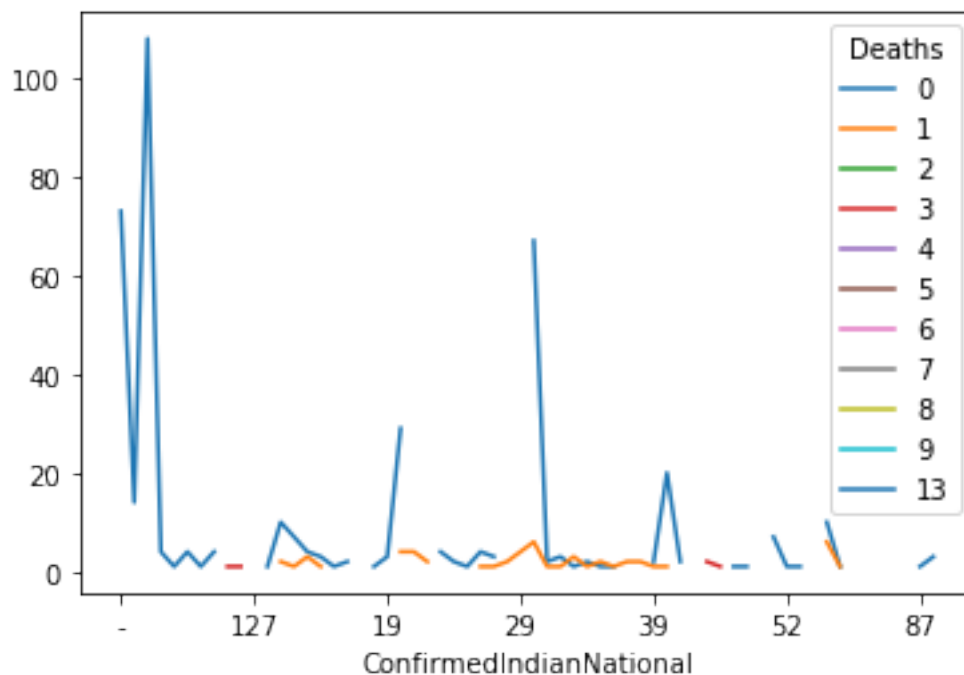
```
[45]:
```

	Date	Time	State/UnionTerritory	ConfirmedIndianNational	\
0	30/01/20	6:00 PM	Kerala	1	
1	31/01/20	6:00 PM	Kerala	1	
2	01/02/20	6:00 PM	Kerala	2	
3	02/02/20	6:00 PM	Kerala	3	

	ConfirmedForeignNational	Cured	Deaths	Confirmed
0	0	0	0	1
1	0	0	0	1
2	0	0	0	2
3	0	0	0	3
4	0	0	0	3

```
[49]: covid.groupby(by = ['ConfirmedIndianNational' , 'Deaths']).count()['State/
↳UnionTerritory'].unstack().plot()
```

```
[49]: <matplotlib.axes._subplots.AxesSubplot at 0x1980b6cdba8>
```



Keep things clean Preventative measures are your first line of defense. The best way to protect yourself from COVID-19 is to practice good hygiene and to make these CDC recommendations part of your routine:

Wash your hands often with soap and water for at least 20 seconds, especially after going to the bathroom; before eating; and after blowing your nose, coughing, or sneezing. If soap and water are not readily available, use an alcohol-based hand sanitizer with at least 60% alcohol. Always wash hands with soap and water if hands are visibly dirty. Avoid touching your eyes, nose, and mouth with unwashed hands. Avoid close contact with people who are sick. Stay home when you are sick. Cover your cough or sneeze with a tissue, then throw



the tissue in the trash and wash your hands. If you don't have a tissue, cough or sneeze into your elbow, rather than into your hands. Clean and disinfect frequently touched objects and surfaces. Visit the CDC for guidelines on how to properly wash your hands and use hand sanitizer. (Yes, there's plenty of science behind this basic habit.)

What about face masks? You may have noticed a growing number of people out and about with their faces covered, but the CDC does not currently recommend the use of face masks among the general public.

As noted above, face masks should be used by people who show COVID-19 symptoms to help prevent the spread of the disease to others. The use of face masks is also crucial for health workers and people who are taking care of someone in close settings (at home or in a health care facility).

Practice social distancing The CDC recommends maintaining a distance of approximately 6 feet from others in public places since respiratory droplets produced by coughing or sneezing do not travel more than 6 feet. While at home, remind everyone to practice everyday preventive actions -- such as washing hands and wiping down surfaces -- to help reduce the risk of getting sick. If you are symptomatic and have tested positive for COVID-19 but do not require hospitalization, you should stay in a specific ``sick room'' and away from other people in your home as much as possible and use a separate bathroom, if available. If your living space makes it difficult to keep a 6-foot distance, stay as far apart as you can and continue to practice good hygiene and wear a mask.

Restrict your travel Many states have put out the call for individuals to stay home, except to make essential trips to the grocery store, pharmacy or for some outdoor exercise.

The CDC recommends that travelers avoid all nonessential travel to China, Iran, most European countries, the United Kingdom and Ireland. Most foreign nationals who have been in one of these countries during the previous 14 days will not be allowed to enter the United States. ``For people at risk for the complications of COVID-19, such as those with underlying medical conditions or those who are older, it's prudent to avoid any long-distance travel,'' says Dr. Goldberg.

STAY SAFE STAY HOME