

```
In [1]: import pandas as pd
c=pd.read_csv("https://covid.ourworldindata.org/data/owid-covid-data.csv")
c.head()
```

```
Out[1]:
```

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
0	AFG	Asia	Afghanistan	2020-02-24	1.0	1.0	NaN	NaN	NaN	NaN
1	AFG	Asia	Afghanistan	2020-02-25	1.0	0.0	NaN	NaN	NaN	NaN
2	AFG	Asia	Afghanistan	2020-02-26	1.0	0.0	NaN	NaN	NaN	NaN
3	AFG	Asia	Afghanistan	2020-02-27	1.0	0.0	NaN	NaN	NaN	NaN
4	AFG	Asia	Afghanistan	2020-02-28	1.0	0.0	NaN	NaN	NaN	NaN

5 rows × 11 columns

```
In [26]: import pandas as pd
df = pd.read_csv("https://covid.ourworldindata.org/data/owid-covid-data.csv", index_col='date', parse_dates = True)
df.head()
```

```
Out[26]:
```

	iso_code	continent	location	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
2020-02-24	AFG	Asia	Afghanistan	1.0	1.0	NaN	NaN	NaN	NaN
2020-02-25	AFG	Asia	Afghanistan	1.0	0.0	NaN	NaN	NaN	NaN
2020-02-26	AFG	Asia	Afghanistan	1.0	0.0	NaN	NaN	NaN	NaN
2020-02-27	AFG	Asia	Afghanistan	1.0	0.0	NaN	NaN	NaN	NaN
2020-02-28	AFG	Asia	Afghanistan	1.0	0.0	NaN	NaN	NaN	NaN

5 rows × 10 columns

```
In [5]: df1 = df
df1['date'] = pd.to_datetime(df1['date'])
df1['date'] = df1['date'].dt.strftime('%m/%d/%Y')
df1 = df1.fillna('-')
df1.head()
```

```
Out[5]:
```

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
0	AFG	Asia	Afghanistan	02/24/2020	1	1	-	-	-	-
1	AFG	Asia	Afghanistan	02/25/2020	1	0	-	-	-	-
2	AFG	Asia	Afghanistan	02/26/2020	1	0	-	-	-	-
3	AFG	Asia	Afghanistan	02/27/2020	1	0	-	-	-	-
4	AFG	Asia	Afghanistan	02/28/2020	1	0	-	-	-	-

5 rows × 11 columns

```
In [7]: #Active cases around the world
top = df[df['date'] == df['date'].max()]
world = top.groupby('location')['total_cases', 'new_cases', 'total_deaths', 'new_deaths'].sum()
world.reset_index()
world.head()
```

```
Out[7]:
```

	location	total_cases	new_cases	total_deaths	new_deaths
0	Afghanistan	51526.0	121.0	2191.0	10.0
1	Africa	2760459.0	29789.0	65468.0	699.0
2	Albania	58316.0	589.0	1181.0	7.0
3	Algeria	99610.0	299.0	2756.0	5.0
4	Andorra	8049.0	66.0	84.0	0.0

Now as this a very big data we will focus on the dataset of united states. As a result we will filter out all other data

```
In [10]: #Time series plot for USA
dat_USA = c[c['iso_code']=="USA"]
dat_USA.shape
dat_USA.head()
```

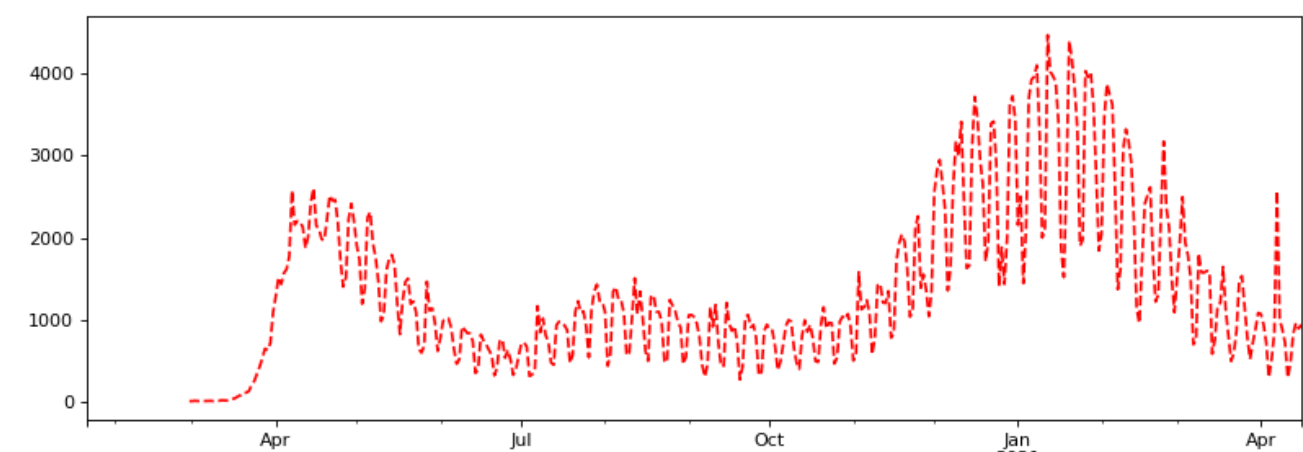
```
Out[10]:
```

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
77811	USA	North America	United States	2020-01-22	1.0	NaN	NaN	NaN	NaN	NaN
77812	USA	North America	United States	2020-01-23	1.0	0.0	NaN	NaN	NaN	NaN
77813	USA	North America	United States	2020-01-24	2.0	1.0	NaN	NaN	NaN	NaN
77814	USA	North America	United States	2020-01-25	2.0	0.0	NaN	NaN	NaN	NaN
77815	USA	North America	United States	2020-01-26	5.0	3.0	NaN	NaN	NaN	NaN

5 rows × 11 columns

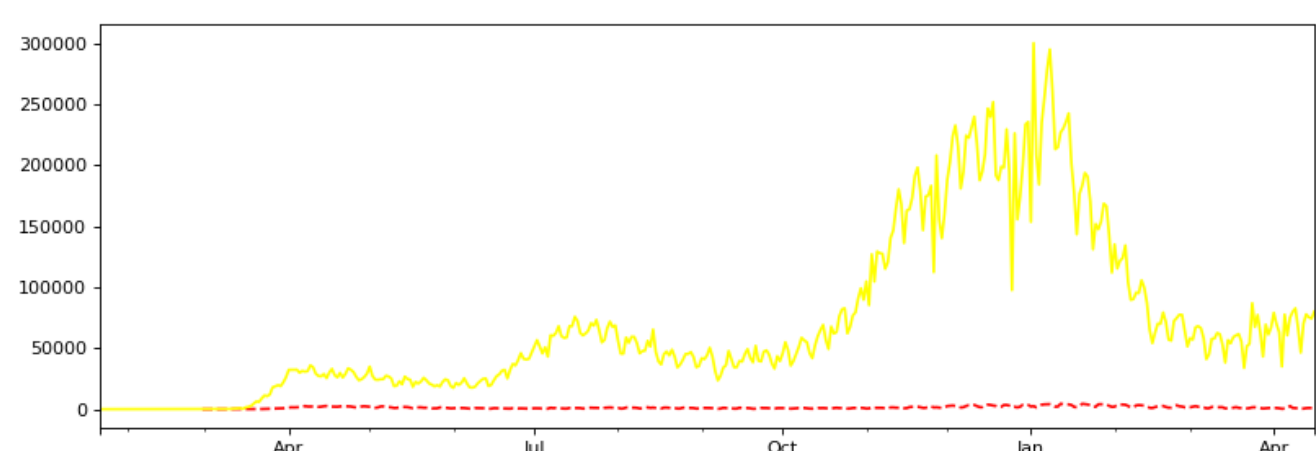
```
In [28]: %matplotlib notebook
#Time series plot for USA
dat_USA2 = df[df['iso_code']=="USA"]
#dat_USA2.head()

dat_USA2['new_deaths'].plot(figsize=(12,4),ls='--', c='red')
```



```
Out[28]: <matplotlib.axes._subplots.AxesSubplot at 0x1cbe0d32ec8>
```

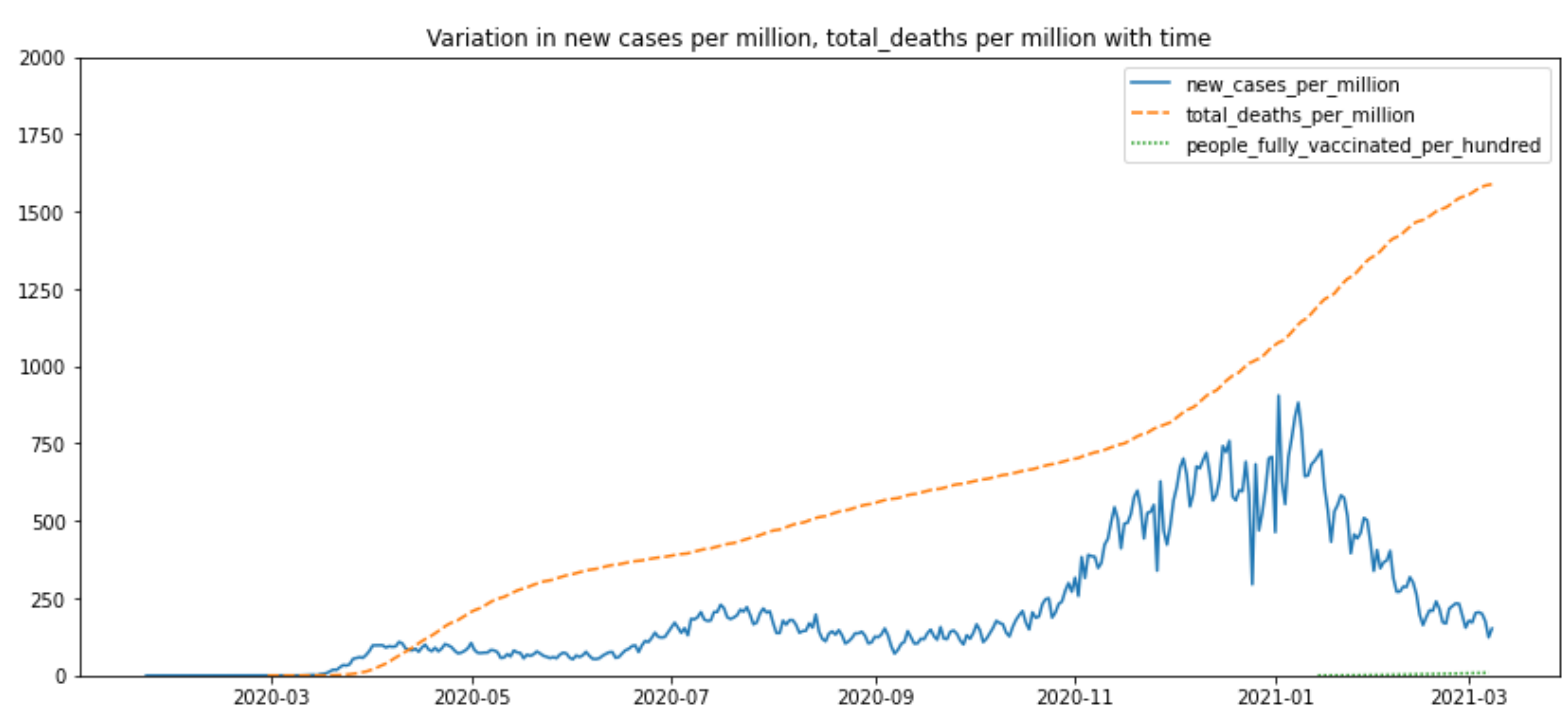
```
In [32]: dat_USA2['new_deaths'].plot(figsize=(12,4),ls='--', c='red')
dat_USA2['new_cases'].plot(figsize=(12,4),ls='solid', c='yellow')
```



```
Out[32]: <matplotlib.axes._subplots.AxesSubplot at 0x1cbe2b2f808>
```

```
In [53]: import seaborn as sns
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings("ignore")
dat_USA_index.dropna()
plt.figure(figsize=(14,6))
plt.ylim([0,2000])
plt.title("Variation in new cases per million, total deaths per million with time")
sns.lineplot(data=dat_USA_index[['new_cases_per_million', 'total_deaths_per_million', 'people_fully_vaccinated_per_hundred']])
```

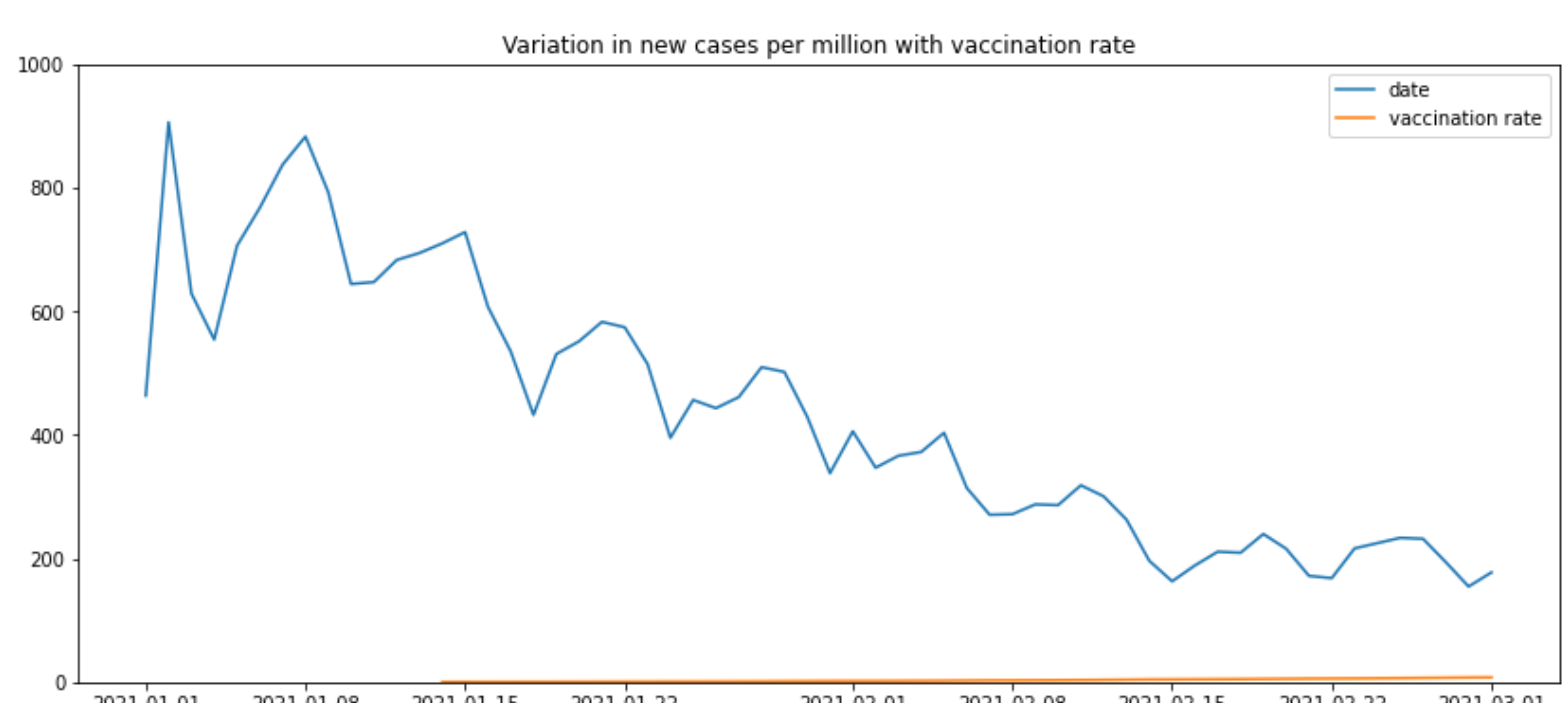
```
Out[53]: <matplotlib.axes._subplots.AxesSubplot at 0x2c1616c23c8>
```



Also we can see that during 2021 people started to get vaccinated and the death rate started to decrease.

```
In [80]: import matplotlib.dates as mdates
from datetime import datetime
plt.figure(figsize=(14,6))
plt.ylim([0,1000])
plt.title("Variation in new cases per million with vaccination rate")
sns.lineplot(data=dat_USA_index.loc['2021-01-01':'2021-03-01', 'new_cases_per_million'], label = 'date')
sns.lineplot(data=dat_USA_index.loc['2021-01-01':'2021-03-01', 'people_fully_vaccinated_per_hundred'], label = 'vaccination rate')
```

```
Out[80]: <matplotlib.axes._subplots.AxesSubplot at 0x2c160ebc588>
```



We can see a decline in new cases with vaccination dose rate whose supply increase during 2021.