

COVID-19 (coronavirus) - 2021

by (Ahmed Al-dayel)

This dataset includes information about COVID-19 (coronavirus) by Our World in Data

I have focused on the stages of development of the Coronavirus in Saudi Arabia

```
In [1]: #Import libraries
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sb
%matplotlib inline
```

```
In [2]: #read the csv (dataset)
df = pd.read_csv('owid-covid-data.csv')
```

```
In [3]: #display the dataset
df.head()
```

	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
1	AFG	Asia	Afghanistan	2020-02-25	1.0	0.0		NaN	NaN	NaN
2	AFG	Asia	Afghanistan	2020-02-26	1.0	0.0		NaN	NaN	NaN
3	AFG	Asia	Afghanistan	2020-02-27	1.0	0.0		NaN	NaN	NaN
4	AFG	Asia	Afghanistan	2020-02-28	1.0	0.0		NaN	NaN	NaN

5 rows × 11 columns

```
In [4]: #display the dataset
df.tail()
```

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed
58180	ZWE	Africa	Zimbabwe	2020-12-31	13867.0	242.0		154.429	363.0	3.0
58181	ZWE	Africa	Zimbabwe	2021-01-01	14084.0	217.0		172.000	369.0	6.0
58182	ZWE	Africa	Zimbabwe	2021-01-02	14491.0	407.0		218.286	377.0	8.0
58183	ZWE	Africa	Zimbabwe	2021-01-03	15265.0	774.0		312.571	380.0	3.0
58184	ZWE	Africa	Zimbabwe	2021-01-04	15829.0	564.0		383.000	384.0	4.0

5 rows × 11 columns

```
In [5]: #show the information about the dataset
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 58185 entries, 0 to 58184
Data columns (total 52 columns):
#   Column                                     Non-Null Count  Dtype
---  --
0   iso_code                                   57852 non-null  object
1   continent                                 57602 non-null  object
2   location                                  58185 non-null  object
3   date                                      58185 non-null  object
4   total_cases                              57626 non-null  float64
5   new_cases                                57619 non-null  float64
6   new_cases_smoothed                       56664 non-null  float64
7   total_deaths                             49393 non-null  float64
8   new_deaths                              49302 non-null  float64
9   new_deaths_smoothed                     56664 non-null  float64
10  total_cases_per_million                   57293 non-null  float64
11  new_cases_per_million                     57286 non-null  float64
12  new_cases_smoothed_per_million            56336 non-null  float64
13  total_deaths_per_million                  48983 non-null  float64
14  new_deaths_per_million                    48982 non-null  float64
15  new_deaths_smoothed_per_million           56336 non-null  float64
16  reproduction_rate                         44995 non-null  float64
17  icu_patients                             6072 non-null   float64
18  icu_patients_per_million                  6072 non-null   float64
19  hosp_patients                             6831 non-null   float64
20  hosp_patients_per_million                 6831 non-null   float64
21  weekly_icu_admissions                     543 non-null    float64
22  weekly_icu_admissions_per_million         543 non-null    float64
23  weekly_hosp_admissions                     877 non-null    float64
24  weekly_hosp_admissions_per_million        877 non-null    float64
25  new_tests                                 26968 non-null  float64
26  total_tests                               26902 non-null  float64
27  total_tests_per_thousand                  26902 non-null  float64
28  new_tests_per_thousand                    26968 non-null  float64
29  new_tests_smoothed                        30211 non-null  float64
30  new_tests_smoothed_per_thousand           30211 non-null  float64
31  positive_rate                             28509 non-null  float64
32  tests_per_case                             28043 non-null  float64
33  tests_units                               31316 non-null  object
34  total_vaccinations                        212 non-null    float64
35  total_vaccinations_per_hundred            212 non-null    float64
36  stringency_index                          52309 non-null  float64
37  population                                57852 non-null  float64
38  population_density                       56628 non-null  float64
39  median_age                                5366 non-null   float64
40  aged_65_and_over                          54722 non-null  float64
41  aged_70_and_over                          55962 non-null  float64
42  gdp_per_capita                            55382 non-null  float64
43  extreme_poverty                           37012 non-null  float64
44  cardiovascular_death_rate                 55962 non-null  float64
45  diabetes_prevalence                       56606 non-null  float64
46  female_smokers                             44295 non-null  float64
47  male_smokers                              43608 non-null  float64
48  handwashing_facilities                    27987 non-null  float64
49  hospital_beds_per_thousand                 54439 non-null  float64
50  life_expectancy                           57555 non-null  float64
51  human_development_index                   55626 non-null  float64
dtypes: float64(47), object(5)
memory usage: 23.1+ MB
```

```
In [6]: #here give us the shape of dataset
df.shape
```

```
Out[6]: (58185, 52)
```

```
In [7]: #show us the number of duplicated value
sum(df.duplicated())
```

```
Out[7]: 0
```

```
In [8]: #know the value for each column
df.describe()
```

	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	total_cases_per_million
count	5.762600e+04	57619.000000	56664.000000	4.930300e+04	49302.000000	56664.000000	57293
mean	2.766940e+05	2945.148441	2932.853233	9.356388e+03	75.182224	64.402447	472
std	2.705976e+06	26114.256152	25722.202758	7.411128e+04	546.989516	493.186865	960
min	1.000000e+00	-46076.000000	-1121.714000	1.000000e+00	-1918.000000	-232.143000	
25%	4.610000e+02	1.000000	4.143000	2.600000e+01	0.000000	0.000000	
50%	4.783000e+03	39.000000	48.143000	1.450000e+02	1.000000	0.714000	9
75%	4.662900e+04	448.000000	488.571000	1.297500e+03	11.000000	7.857000	451
max	8.567203e+07	756739.000000	646630.286000	1.853334e+06	15525.000000	11670.429000	10676

8 rows × 8 columns

```
In [9]: #here i will drop some columns, because i didn't need it
df.drop(['gdp_per_capita', 'handwashing_facilities', 'new_cases_smoothed_per_million', 'new_deaths_smoothed', 'new_cases_smoothed'], axis=1, inplace=True)
```

```
In [10]: #display after drop the coulumns
df.head()
```

	iso_code	continent	location	date	total_cases	new_cases	total_deaths	new_deaths	total_cases_per_million	new_deaths_smoothed
0	AFG	Asia	Afghanistan	2020-02-24	1.0	1.0	NaN	NaN		0.026
1	AFG	Asia	Afghanistan	2020-02-25	1.0	0.0	NaN	NaN		0.026
2	AFG	Asia	Afghanistan	2020-02-26	1.0	0.0	NaN	NaN		0.026
3	AFG	Asia	Afghanistan	2020-02-27	1.0	0.0	NaN	NaN		0.026
4	AFG	Asia	Afghanistan	2020-02-28	1.0	0.0	NaN	NaN		0.026

5 rows × 11 columns

```
In [11]: #display after drop the coulumns
df.tail()
```

	iso_code	continent	location	date	total_cases	new_cases	total_deaths	new_deaths	total_cases_per_million	new_deaths_smoothed
58180	ZWE	Africa	Zimbabwe	2020-12-31	13867.0	242.0		363.0	3.0	932.993
58181	ZWE	Africa	Zimbabwe	2021-01-01	14084.0	217.0		369.0	6.0	947.593
58182	ZWE	Africa	Zimbabwe	2021-01-02	14491.0	407.0		377.0	8.0	974.976
58183	ZWE	Africa	Zimbabwe	2021-01-03	15265.0	774.0		380.0	3.0	1027.052
58184	ZWE	Africa	Zimbabwe	2021-01-04	15829.0	564.0		384.0	4.0	1064.999

5 rows × 11 columns

```
In [12]: sa = df[['location', 'population', 'date', 'continent', 'total_cases', 'new_cases', 'new_deaths', 'total_deaths', 'icu_patients', 'total_cases_per_million', 'aged_65_and_over', 'cardiovascular_death_rate', 'diabetes_prevalence', 'female_smokers', 'male_smokers', 'hospital_beds_per_thousand', 'total_deaths_smoothed', 'new_deaths_smoothed']]
sa.rename({'icu_patients': 'critical_patients'}, axis=1, inplace=True)
```

D:\Conda\lib\site-packages\pandas\core\frame.py:4298: SettingWithCopyWarning: A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

errors=errors,

```
In [14]: saudi = sa[sa['location']=='Saudi Arabia']
```

```
In [15]: saudi.head()
```

	location	population	date	continent	total_cases	new_cases	new_deaths	total_deaths	critical_patients	total_cases_smoothed
44920	Saudi Arabia	34813867.0	2020-03-02	Asia	1.0	1.0	NaN	NaN	NaN	
44921	Saudi Arabia	34813867.0	2020-03-03	Asia	1.0	0.0	NaN	NaN	NaN	
44922	Saudi Arabia	34813867.0	2020-03-04	Asia	1.0	0.0	NaN	NaN	NaN	
44923	Saudi Arabia	34813867.0	2020-03-05	Asia	5.0	4.0	NaN	NaN	NaN	
44924	Saudi Arabia	34813867.0	2020-03-06	Asia	5.0	0.0	NaN	NaN	NaN	

5 rows × 11 columns

```
In [16]: saudi.tail()
```

	location	population	date	continent	total_cases	new_cases	new_deaths	total_deaths	critical_patients	total_cases_smoothed
45224	Saudi Arabia	34813867.0	2020-12-21	Asia	362741.0	140.0	9.0	6223.0	NaN	
45225	Saudi Arabia	34813867.0	2021-01-01	Asia	362878.0	137.0	7.0	6230.0	NaN	
45226	Saudi Arabia	34813867.0	2021-01-02	Asia	362979.0	101.0	9.0	6239.0	NaN	
45227	Saudi Arabia	34813867.0	2021-01-03	Asia	363061.0	82.0	7.0	6246.0	NaN	
45228	Saudi Arabia	34813867.0	2021-01-04	Asia	363155.0	94.0	10.0	6256.0	NaN	

5 rows × 11 columns

```
In [17]: t_saudi = saudi[['date', 'total_cases', 'total_deaths', 'new_cases', 'new_deaths']]
```

```
In [18]: total_saudi = t_saudi[t_saudi['date']=='2021-01-04']
```

```
In [19]: #here will show us the total number of cases in Saudi Arabia
total_cs = total_saudi['total_cases'].sum()
print(total_cs)
363155.0
```

```
In [20]: #here will show us the total number of deaths in Saudi Arabia
total_ds = total_saudi['total_deaths'].sum()
print(total_ds)
6256.0
```

```
In [21]: #here will show us the total number of recover cases in Saudi Arabia
recover_saudi = total_cs - total_ds
print(recover_saudi)
356899.0
```

```
In [22]: saudi['critical_patients'].sum()
```

```
Out[22]: 0.0
```

```
In [23]: #here i will drop some column, because i didn't need it
saudi.drop(columns=['critical_patients'], inplace=True)
```

D:\Conda\lib\site-packages\pandas\core\frame.py:4164: SettingWithCopyWarning: A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

errors=errors,

```
In [24]: #display after drop the coulumns
saudi.head(1)
```

	location	population	date	continent	total_cases	new_cases	new_deaths	total_deaths	total_cases_per_million	aged_65_and_over
44920	Saudi Arabia	34813867.0	2020-03-02	Asia	1.0	1.0	NaN	NaN	0.029	
45228	Saudi Arabia	34813867.0	2021-01-04	Asia	363155.0	94.0	10.0	6256.0	10431.332	

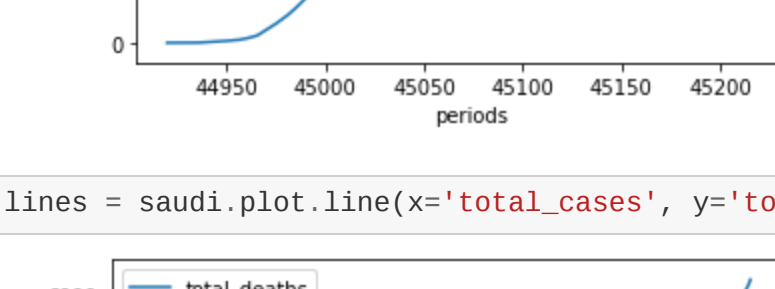
3 rows × 11 columns

```
In [26]: #here will show us the total number of cases in Saudi Arabia today
print(total_c)
94.0
```

```
In [27]: #here will show us the total number of deaths in Saudi Arabia today
total_d = total_saudi['new_deaths'].sum()
print(total_d)
10.0
```

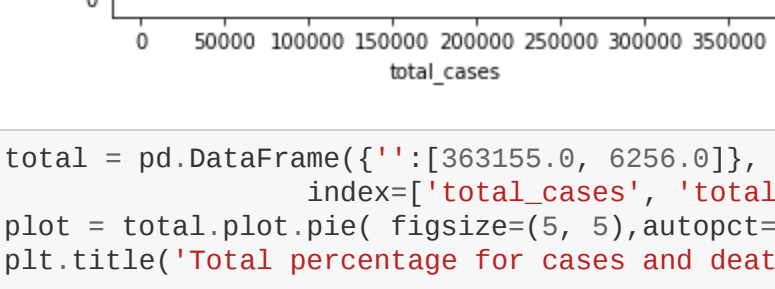
```
In [28]: saudi['total_deaths'].plot(xlabel='periods', ylabel='total_deaths');
plt.title('Stages of rising cases of deaths from corona in Saudi Arabia')
```

Text(0.5, 1.0, 'Stages of rising cases of deaths from corona in Saudi Arabia')

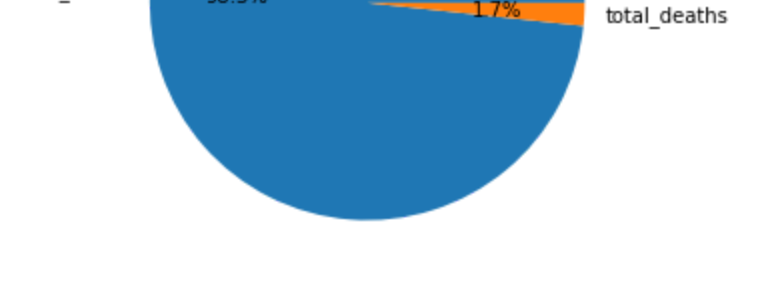


```
In [29]: saudi['total_cases'].plot(xlabel='periods', ylabel='total_cases');
plt.title('Stages of rising cases of corona in Saudi Arabia')
```

Text(0.5, 1.0, 'Stages of rising cases of corona in Saudi Arabia')

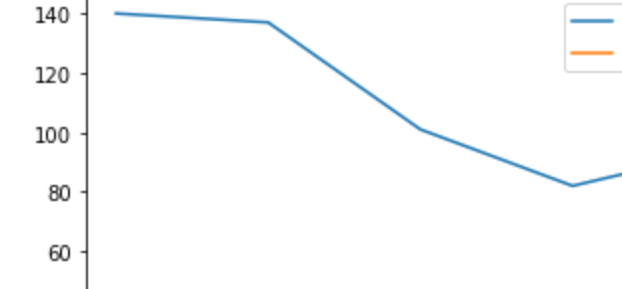


```
In [30]: lines = saudi.plot.line(x='total_cases', y='total_deaths', ylabel='total_deaths')
```



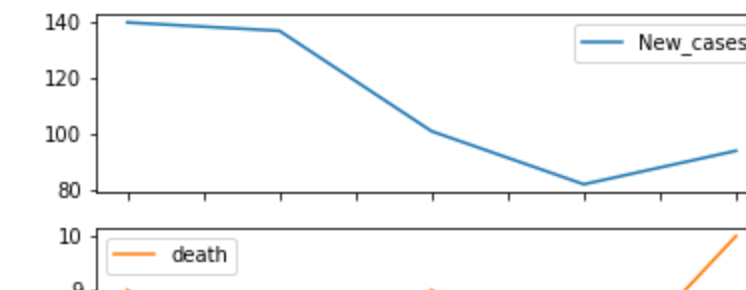
```
In [31]: total = pd.DataFrame({'total_cases': [363155.0, 6256.0]},
                             index=['total_cases', 'total_deaths'])
plot = total.plot.pie(figsize=(5, 5), autopct='%1.1f%%', subplots=True)
plt.title('Total percentage for cases and deaths in Saudi Arabia')
```

Text(0.5, 1.0, 'Total percentage for cases and deaths in Saudi Arabia')

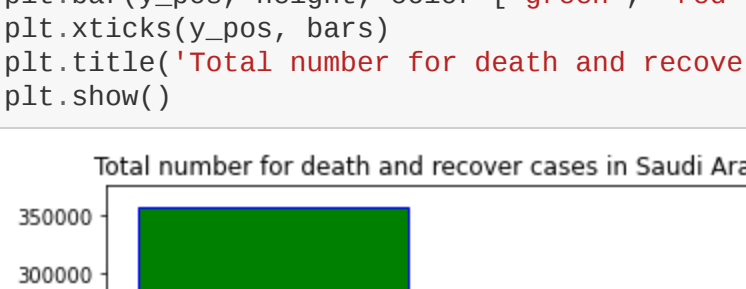


```
In [32]: df = pd.DataFrame({'New_cases': [140, 137, 101, 82, 94],
                          'death': [9, 7, 9, 7, 10]
                          }, index=['2020-12-31', '2021-1-1', '2021-1-2', '2021-1-3', '2021-1-4'])
lines = df.plot.line()
plt.title('Total number for cases and deaths in Saudi Arabia in last five days')
```

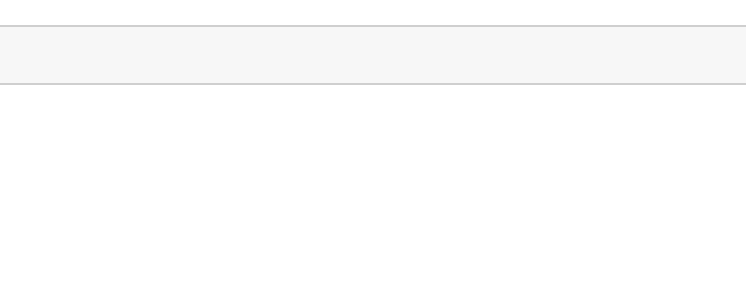
Text(0.5, 1.0, 'Total number for cases and deaths in Saudi Arabia in last five days')



```
In [33]: axes = df.plot.line(subplots=True)
type(axes)
numpy.ndarray
```



```
In [34]: height = [356899, 6256]
bars = ('Total recover', 'Total death')
y_pos = np.arange(len(bars))
plt.bar(y_pos, height, color=['green', 'red'], edgecolor='blue')
plt.xticks(y_pos, bars)
plt.title('Total number for death and recover cases in Saudi Arabia')
plt.show()
```



```
In [ ]:
```