

covid-data-analysis-1

July 15, 2023

```
[2]: #importing libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

df = pd.read_csv('covid1.csv')
df.head(10)
```

```
[2]: iso_code location      date  total_cases  new_cases  total_deaths  \
0    ABW    Aruba  3/13/2020         2.0        2.0         0.0
1    ABW    Aruba  3/20/2020         4.0        2.0         0.0
2    ABW    Aruba  3/24/2020        12.0        8.0         0.0
3    ABW    Aruba  3/25/2020        17.0        5.0         0.0
4    ABW    Aruba  3/26/2020        19.0        2.0         0.0
5    ABW    Aruba  3/27/2020        28.0        9.0         0.0
6    ABW    Aruba  3/28/2020        28.0        0.0         0.0
7    ABW    Aruba  3/29/2020        28.0        0.0         0.0
8    ABW    Aruba  3/30/2020       50.0       22.0         0.0
9    ABW    Aruba  4/1/2020       55.0        5.0         0.0

    new_deaths  total_cases_per_million  new_cases_per_million  \
0          0.0          18.733          18.733
1          0.0          37.465          18.733
2          0.0         112.395          74.930
3          0.0         159.227          46.831
4          0.0         177.959          18.733
5          0.0         262.256          84.296
6          0.0         262.256           0.000
7          0.0         262.256           0.000
8          0.0         468.314         206.058
9          0.0         515.145          46.831

    total_deaths_per_million  new_deaths_per_million  stringency_index  \
0              0.0              0.0              0.00
1              0.0              0.0             30.56
2              0.0              0.0             44.84
```

3	0.0	0.0	44.84
4	0.0	0.0	44.84
5	0.0	0.0	44.84
6	0.0	0.0	44.84
7	0.0	0.0	83.46
8	0.0	0.0	83.46
9	0.0	0.0	83.46

	population	population_density	median_age	aged_65_older	aged_70_older	\
0	106766.0	584.8	41.2	13.085	7.452	
1	106766.0	584.8	41.2	13.085	7.452	
2	106766.0	584.8	41.2	13.085	7.452	
3	106766.0	584.8	41.2	13.085	7.452	
4	106766.0	584.8	41.2	13.085	7.452	
5	106766.0	584.8	41.2	13.085	7.452	
6	106766.0	584.8	41.2	13.085	7.452	
7	106766.0	584.8	41.2	13.085	7.452	
8	106766.0	584.8	41.2	13.085	7.452	
9	106766.0	584.8	41.2	13.085	7.452	

	gdp_per_capita	diabetes_prevalence
0	35973.781	11.62
1	35973.781	11.62
2	35973.781	11.62
3	35973.781	11.62
4	35973.781	11.62
5	35973.781	11.62
6	35973.781	11.62
7	35973.781	11.62
8	35973.781	11.62
9	35973.781	11.62

```
[3]: df.tail(10)
```

```
[3]:
```

	iso_code	location	date	total_cases	new_cases	total_deaths	new_deaths	\
1077	AUT	Austria	NaN	NaN	NaN	NaN	NaN	
1078	AUT	Austria	NaN	NaN	NaN	NaN	NaN	
1079	AUT	Austria	NaN	NaN	NaN	NaN	NaN	
1080	AUT	Austria	NaN	NaN	NaN	NaN	NaN	
1081	AUT	Austria	NaN	NaN	NaN	NaN	NaN	
1082	AUT	Austria	NaN	NaN	NaN	NaN	NaN	
1083	AUT	Austria	NaN	NaN	NaN	NaN	NaN	
1084	AUT	Austria	NaN	NaN	NaN	NaN	NaN	
1085	AUT	Austria	NaN	NaN	NaN	NaN	NaN	
1086	AUT	Austria	NaN	NaN	NaN	NaN	NaN	

	total_cases_per_million	new_cases_per_million	\
--	-------------------------	-----------------------	---

1077	NaN	NaN
1078	NaN	NaN
1079	NaN	NaN
1080	NaN	NaN
1081	NaN	NaN
1082	NaN	NaN
1083	NaN	NaN
1084	NaN	NaN
1085	NaN	NaN
1086	NaN	NaN

	total_deaths_per_million	new_deaths_per_million	stringency_index \
1077	NaN	NaN	NaN
1078	NaN	NaN	NaN
1079	NaN	NaN	NaN
1080	NaN	NaN	NaN
1081	NaN	NaN	NaN
1082	NaN	NaN	NaN
1083	NaN	NaN	NaN
1084	NaN	NaN	NaN
1085	NaN	NaN	NaN
1086	NaN	NaN	NaN

	population	population_density	median_age	aged_65_older \
1077	NaN	NaN	NaN	NaN
1078	NaN	NaN	NaN	NaN
1079	NaN	NaN	NaN	NaN
1080	NaN	NaN	NaN	NaN
1081	NaN	NaN	NaN	NaN
1082	NaN	NaN	NaN	NaN
1083	NaN	NaN	NaN	NaN
1084	NaN	NaN	NaN	NaN
1085	NaN	NaN	NaN	NaN
1086	NaN	NaN	NaN	NaN

	aged_70_older	gdp_per_capita	diabetes_prevalence
1077	NaN	NaN	NaN
1078	NaN	NaN	NaN
1079	NaN	NaN	NaN
1080	NaN	NaN	NaN
1081	NaN	NaN	NaN
1082	NaN	NaN	NaN
1083	NaN	NaN	NaN
1084	NaN	NaN	NaN
1085	NaN	NaN	NaN
1086	NaN	NaN	NaN

```
[4]: df['location'].value_counts()
```

```
[4]: Australia          146
      United Arab Emirates 140
      Armenia           137
      Afghanistan       136
      Argentina          79
      Albania            77
      Andorra            72
      Antigua and Barbuda 66
      Angola             64
      Aruba              63
      Anguilla           59
      Austria            48
      Name: location, dtype: int64
```

```
[5]: df['median_age'].value_counts()
```

```
[5]: 37.9    146
      34.0    140
      35.7    137
      18.6    136
      31.9     79
      38.0     77
      32.1     66
      16.8     65
      41.2     63
      44.4      6
      Name: median_age, dtype: int64
```

```
[6]: df['diabetes_prevalence'].value_counts()
```

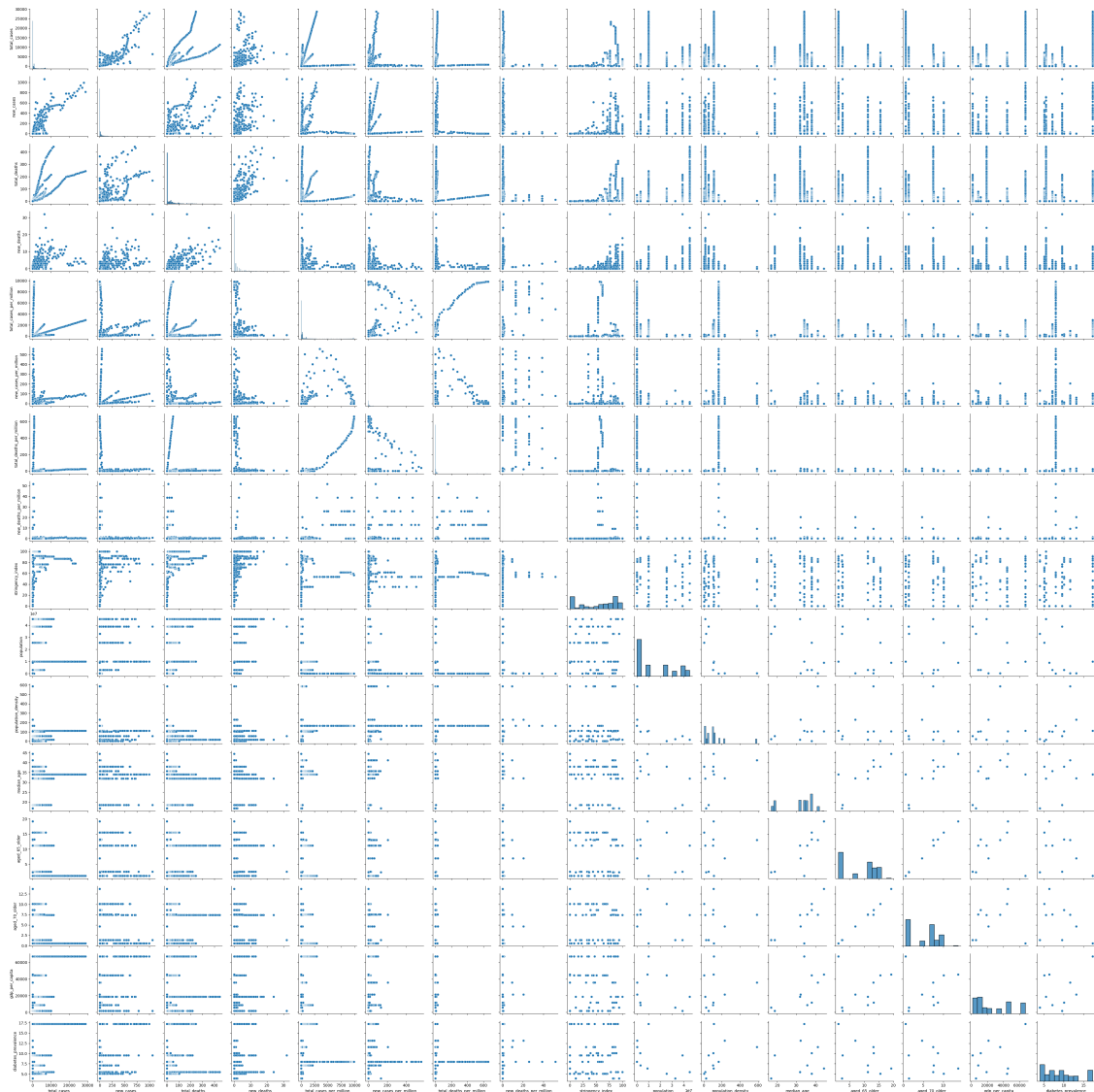
```
[6]: 5.07    146
      17.26   140
      7.11    137
      9.59    136
      5.50     79
      10.08    77
      7.97     72
      13.17    66
      3.94     65
      11.62    63
      6.35      6
      Name: diabetes_prevalence, dtype: int64
```

```
[7]: df['total_deaths'].value_counts()
```

```
[7]: 0.0      427
      2.0      107
      3.0       73
      31.0     27
      1.0      23
      ...
      208.0     1
      206.0     1
      203.0     1
      201.0     1
      102.0     1
      Name: total_deaths, Length: 163, dtype: int64
```

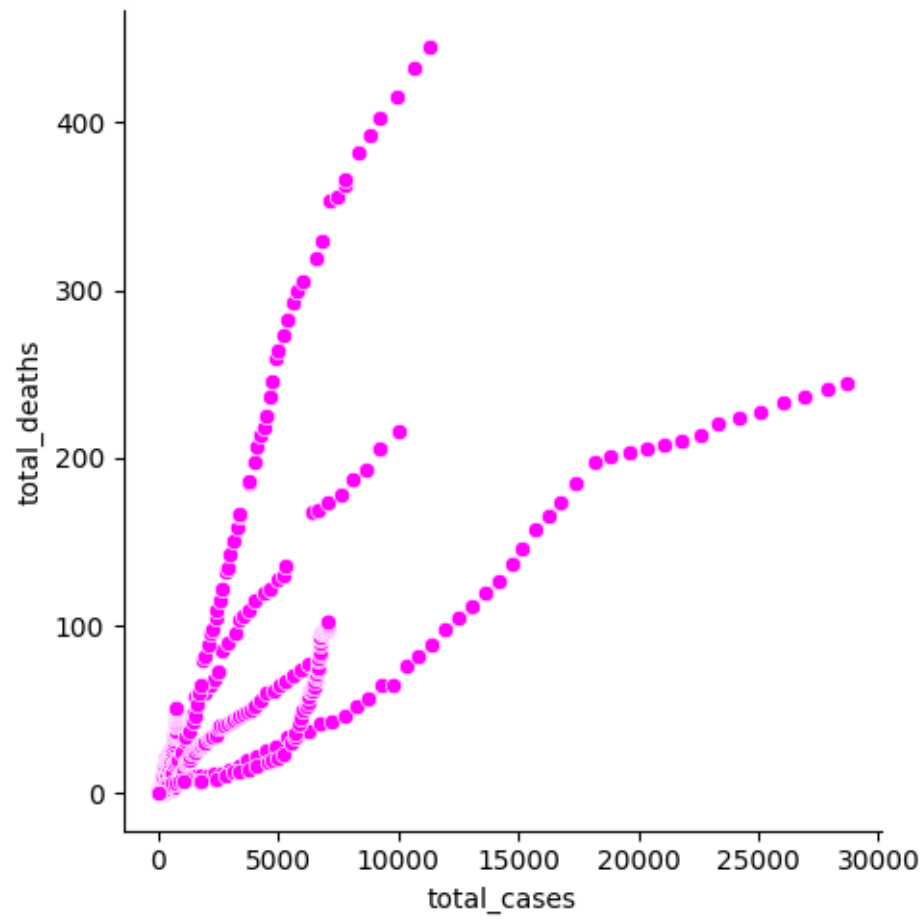
```
[8]: sns.pairplot(df)
```

```
[8]: <seaborn.axisgrid.PairGrid at 0x7dd5fdcc6dd0>
```



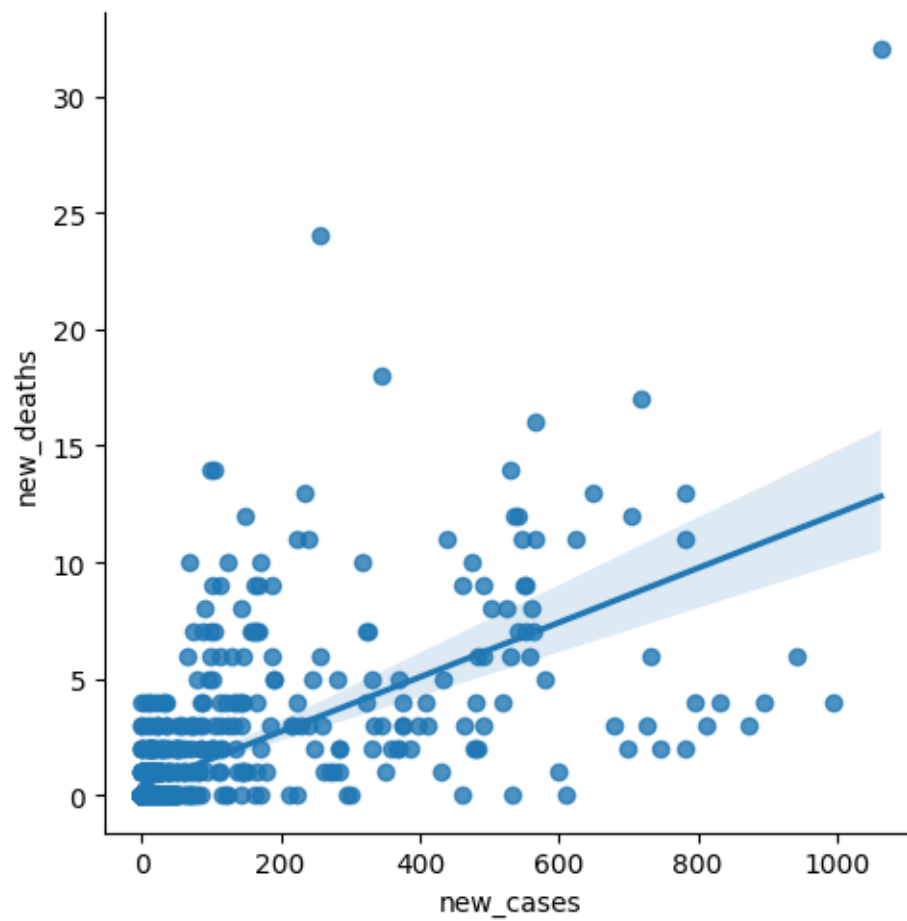
```
[20]: sns.relplot(data=df, x="total_cases", y="total_deaths",color='fuchsia')
```

```
[20]: <seaborn.axisgrid.FacetGrid at 0x7dd5e8db3760>
```



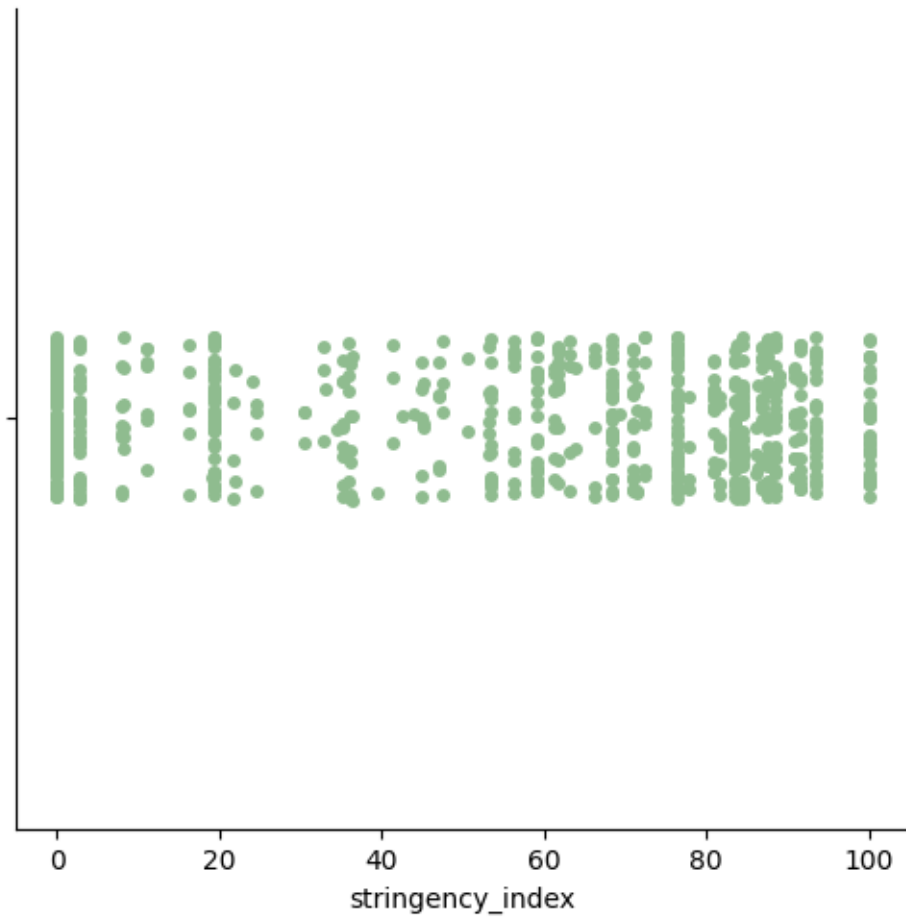
```
[11]: sns.lmplot(data=df, x="new_cases", y="new_deaths")
```

```
[11]: <seaborn.axisgrid.FacetGrid at 0x7dd5ee86b2b0>
```



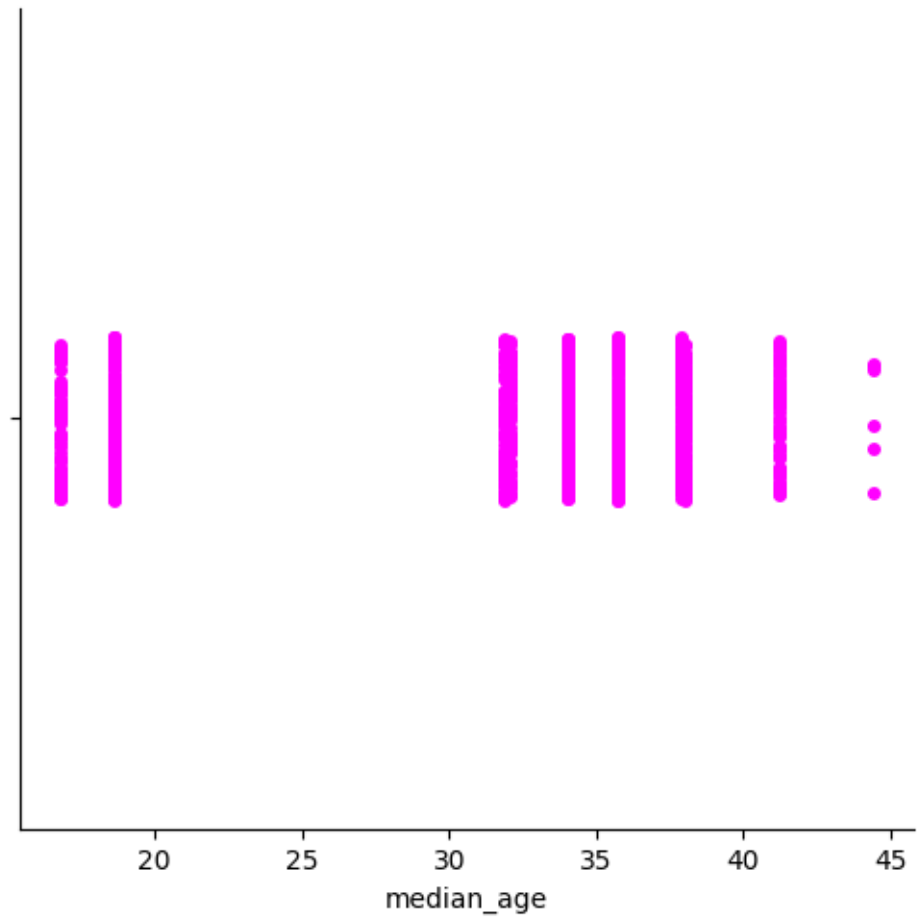
```
[12]: sns.catplot(df , x="stringency_index" , color='darkseagreen')
```

```
[12]: <seaborn.axisgrid.FacetGrid at 0x7dd5ef763730>
```

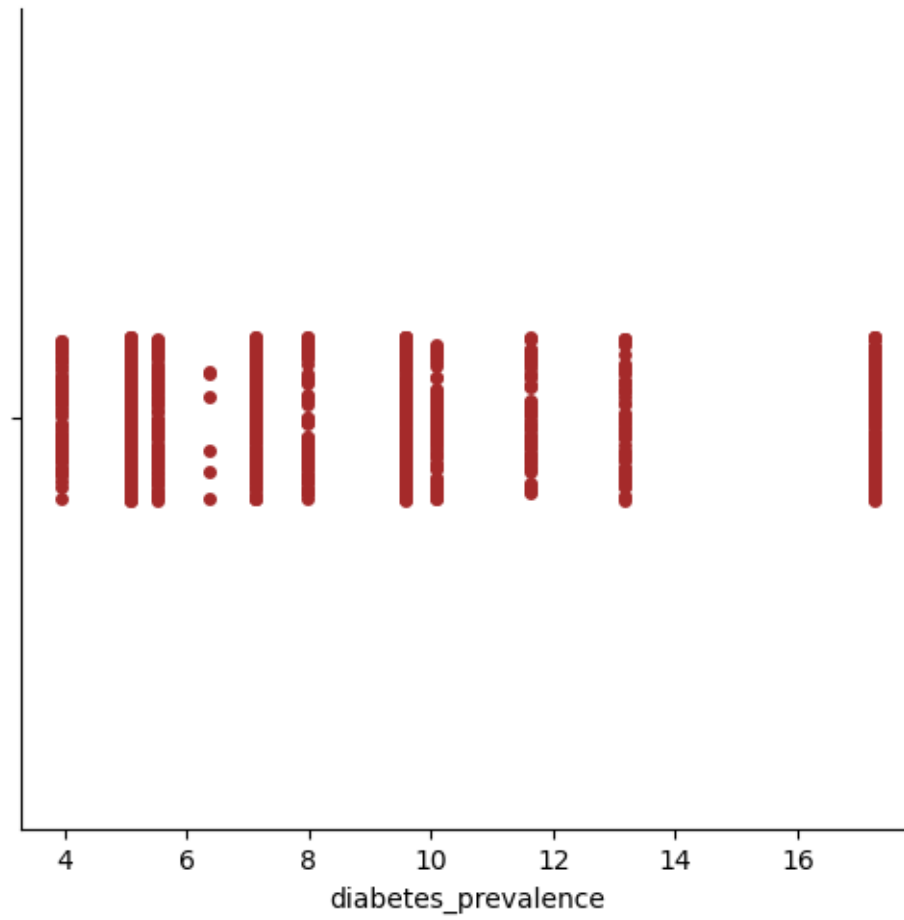
```
[21]: sns.catplot(df , x="median_age" , color='magenta')
```

```
[21]: <seaborn.axisgrid.FacetGrid at 0x7dd5e8d53d60>
```



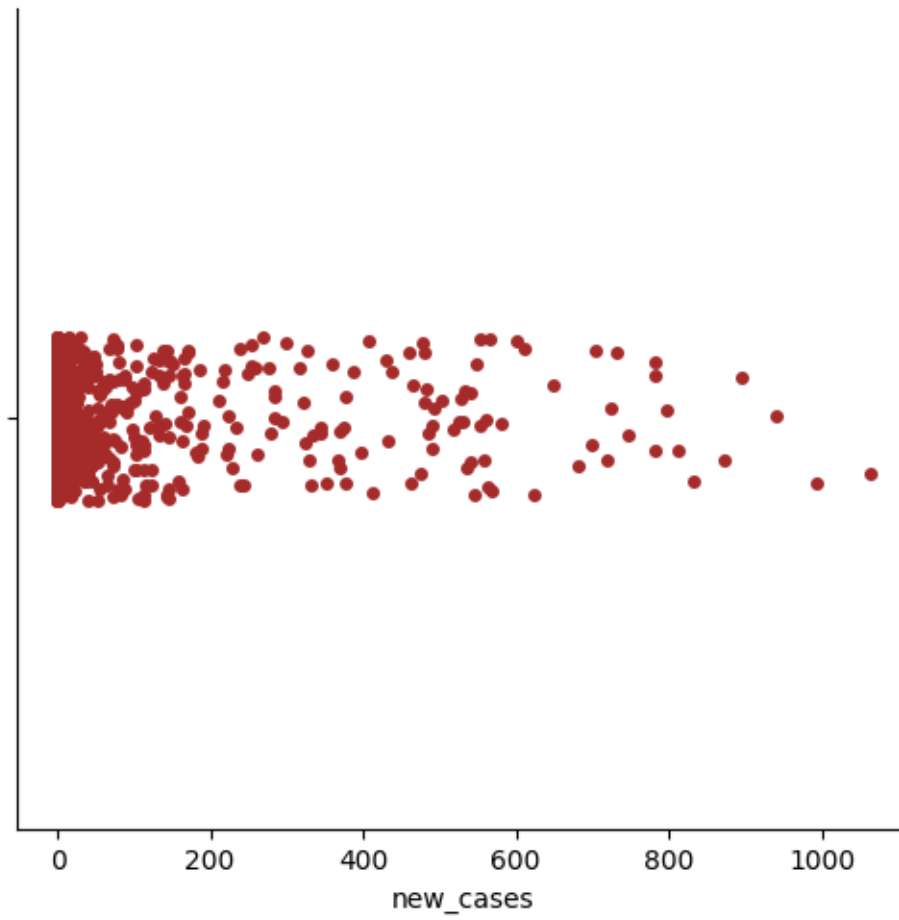
```
[18]: sns.catplot(df , x="diabetes_prevalence" , color='brown')
```

```
[18]: <seaborn.axisgrid.FacetGrid at 0x7dd5efe5a6b0>
```



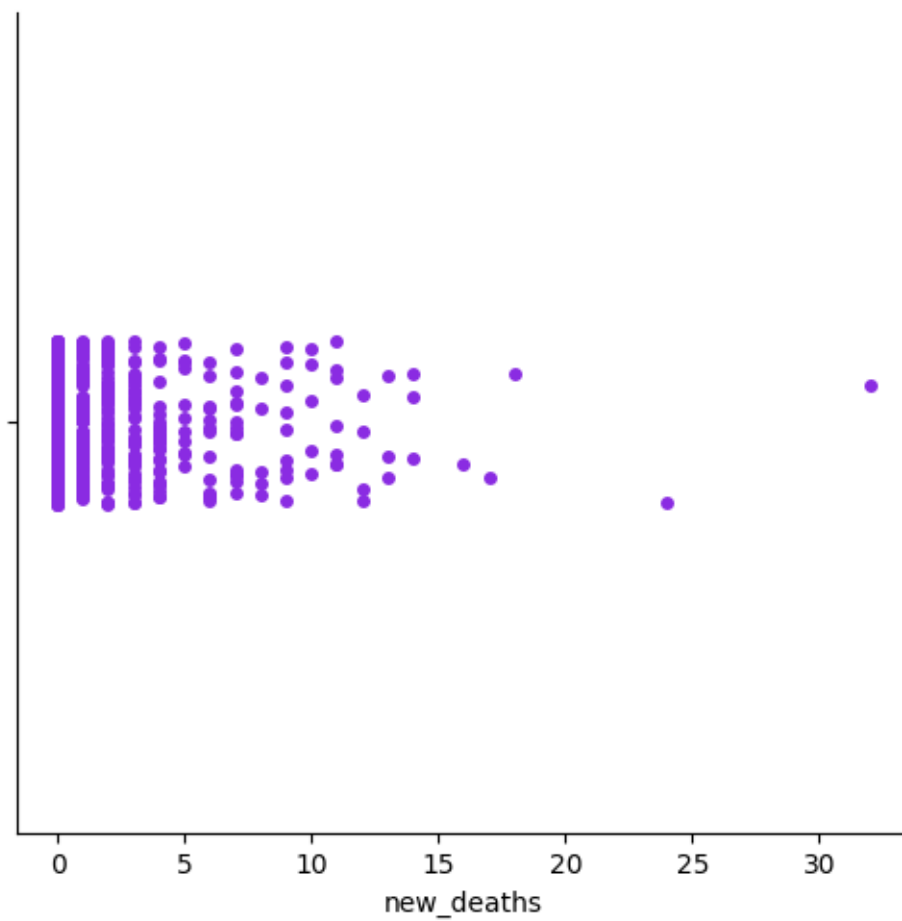
```
[15]: sns.catplot(df , x="new_cases" , color='brown')
```

```
[15]: <seaborn.axisgrid.FacetGrid at 0x7dd5ef703a00>
```



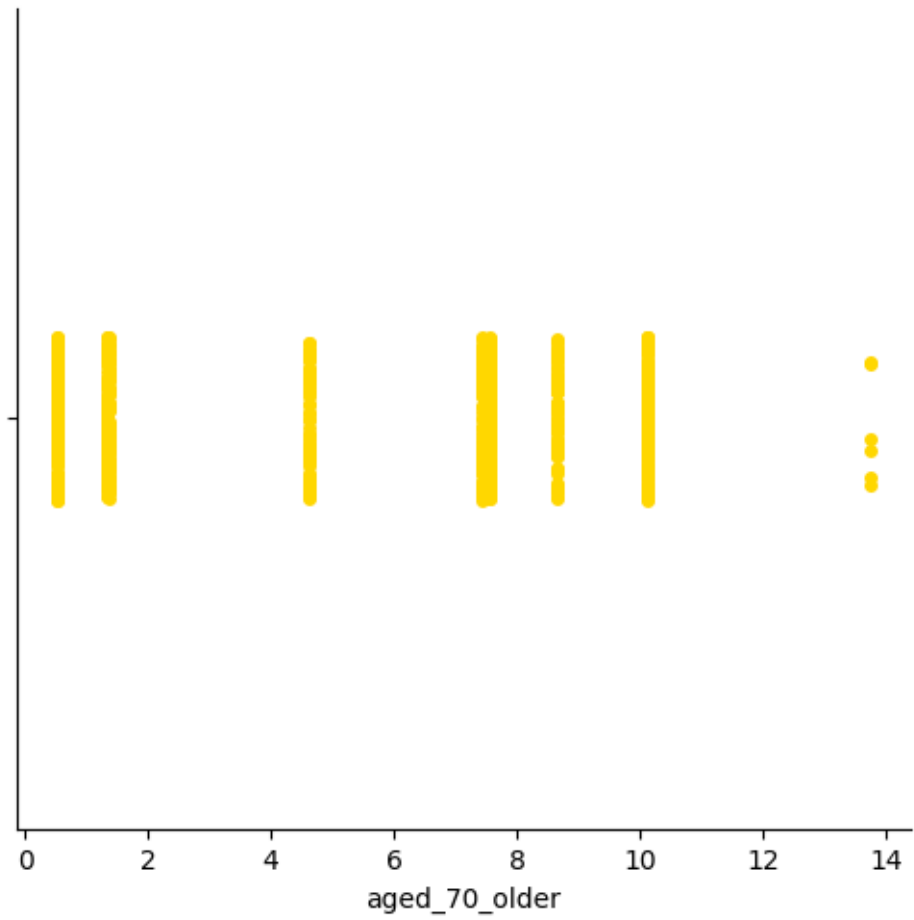
```
[22]: sns.catplot(df , x="new_deaths" , color='blueviolet')
```

```
[22]: <seaborn.axisgrid.FacetGrid at 0x7dd5e8fb6aa0>
```



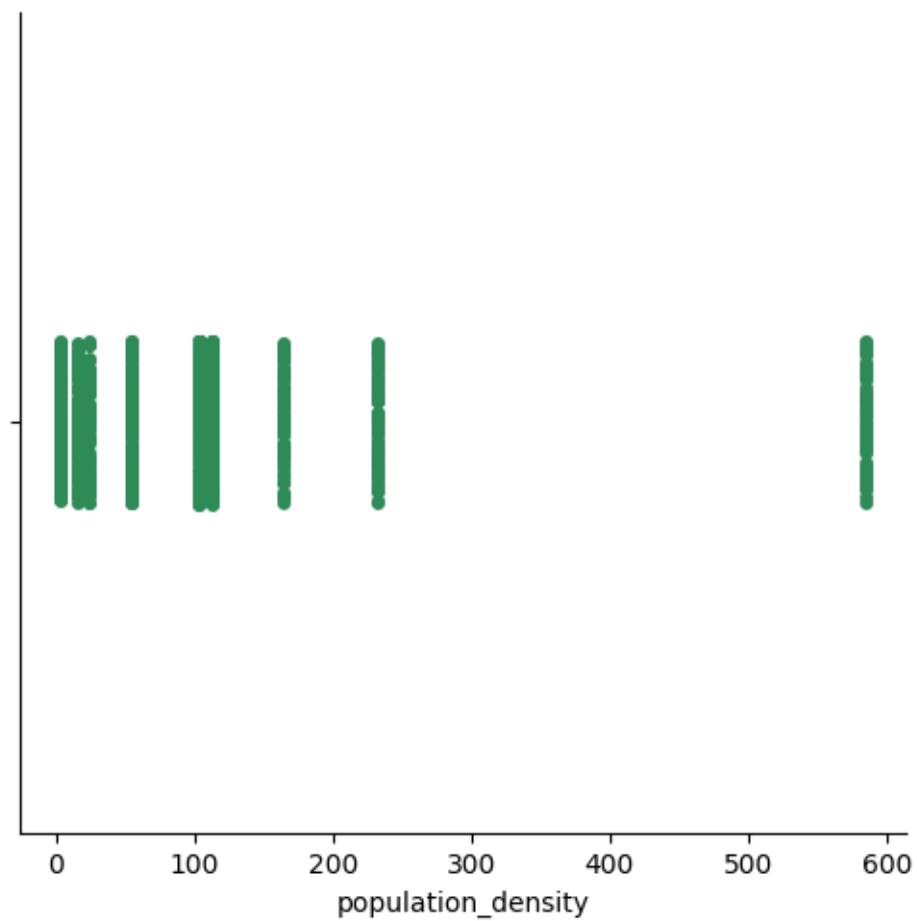
```
[23]: sns.catplot(df , x="aged_70_older" , color='gold')
```

```
[23]: <seaborn.axisgrid.FacetGrid at 0x7dd5e8b6fc40>
```



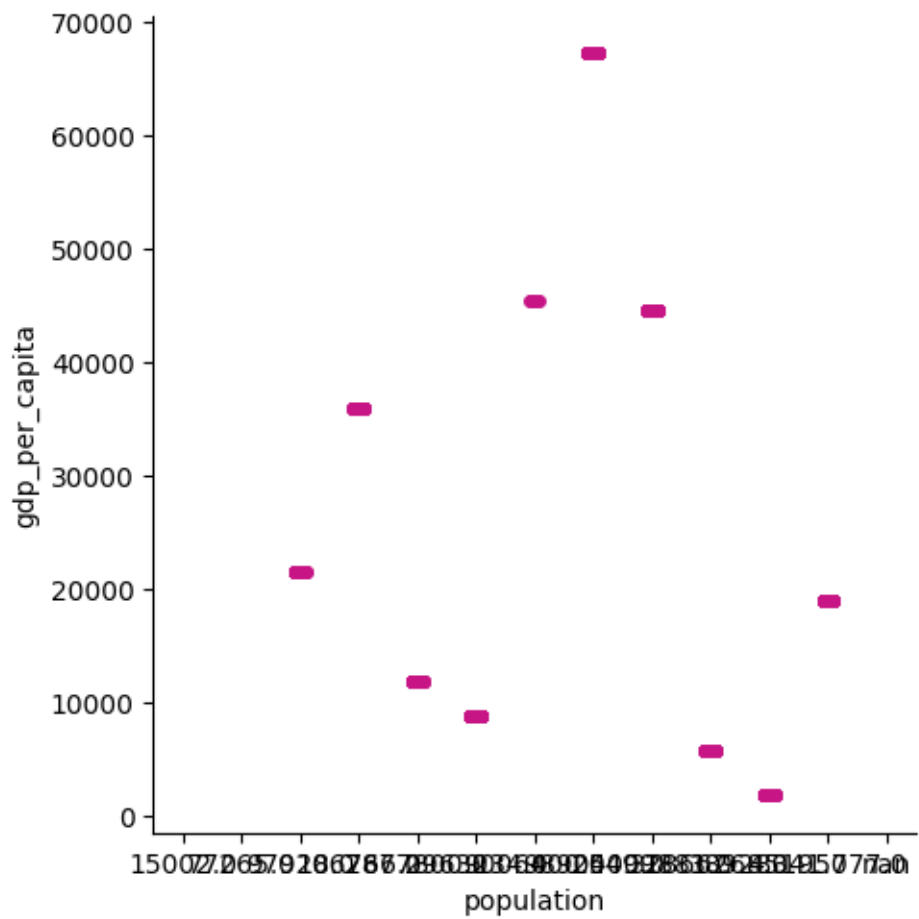
```
[24]: sns.catplot(df , x="population_density" , color='seagreen')
```

```
[24]: <seaborn.axisgrid.FacetGrid at 0x7dd5e8b02ad0>
```



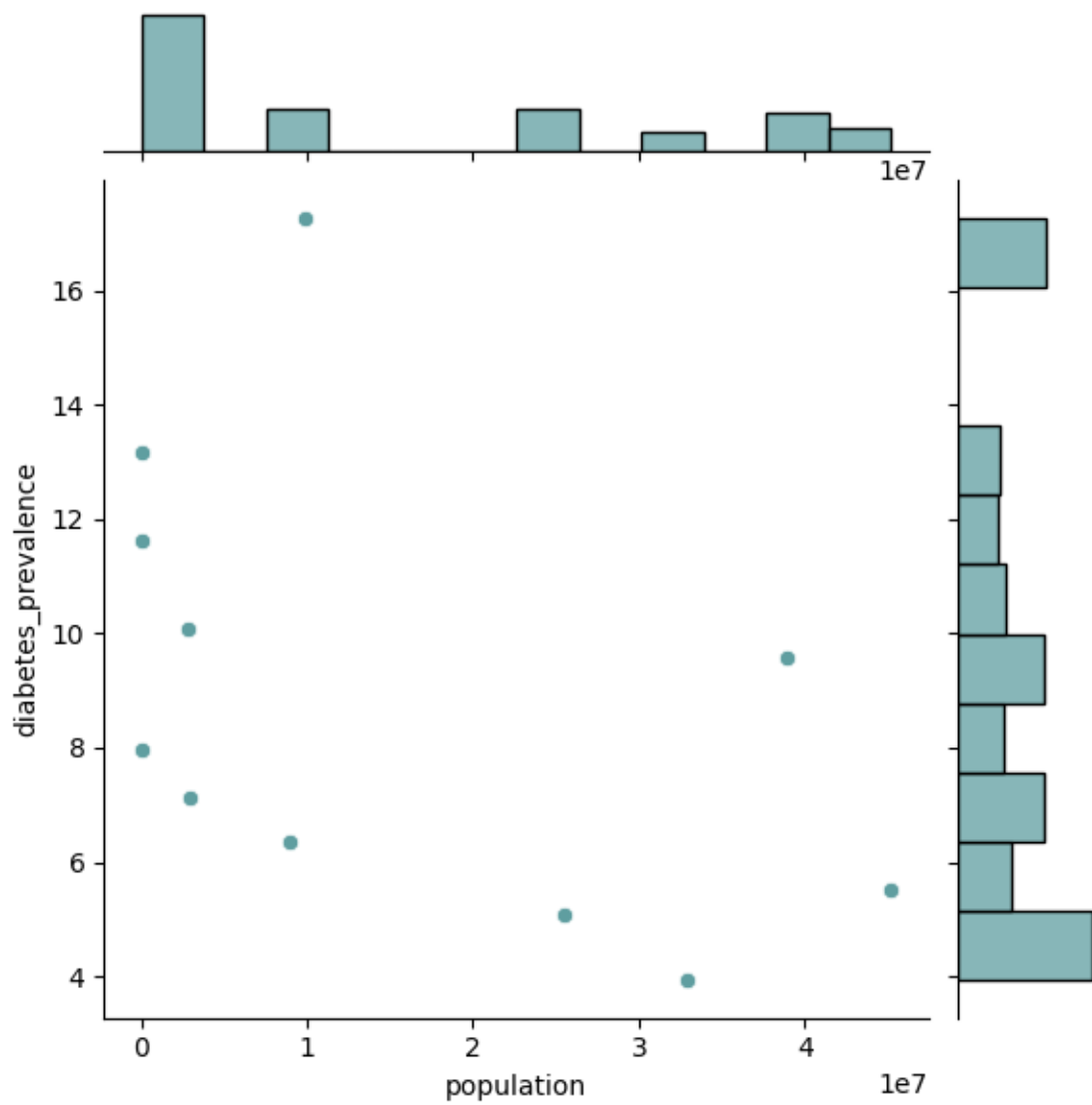
```
[25]: sns.catplot(df , x="population" , y='gdp_per_capita', color='mediumvioletred')
```

```
[25]: <seaborn.axisgrid.FacetGrid at 0x7dd5e8b8c190>
```



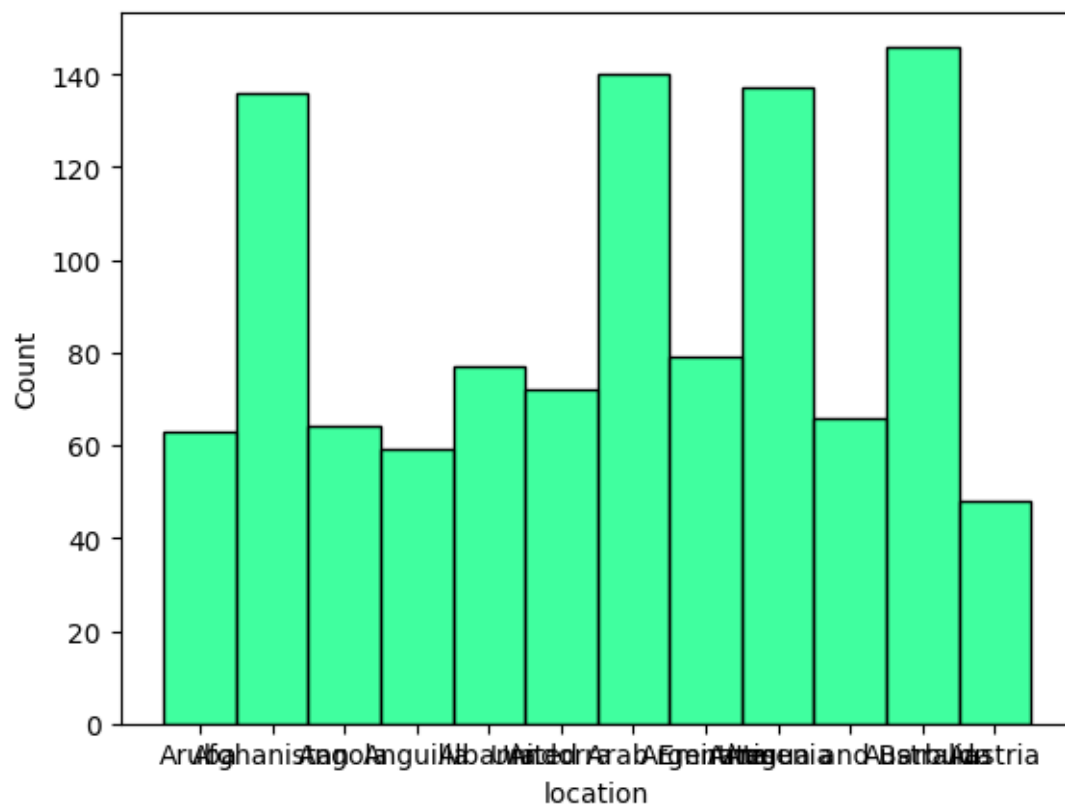
```
[26]: sns.jointplot(data = df , x = 'population' , y='diabetes_prevalence', color = 'cadetblue')
```

```
[26]: <seaborn.axisgrid.JointGrid at 0x7dd5e8fb76d0>
```

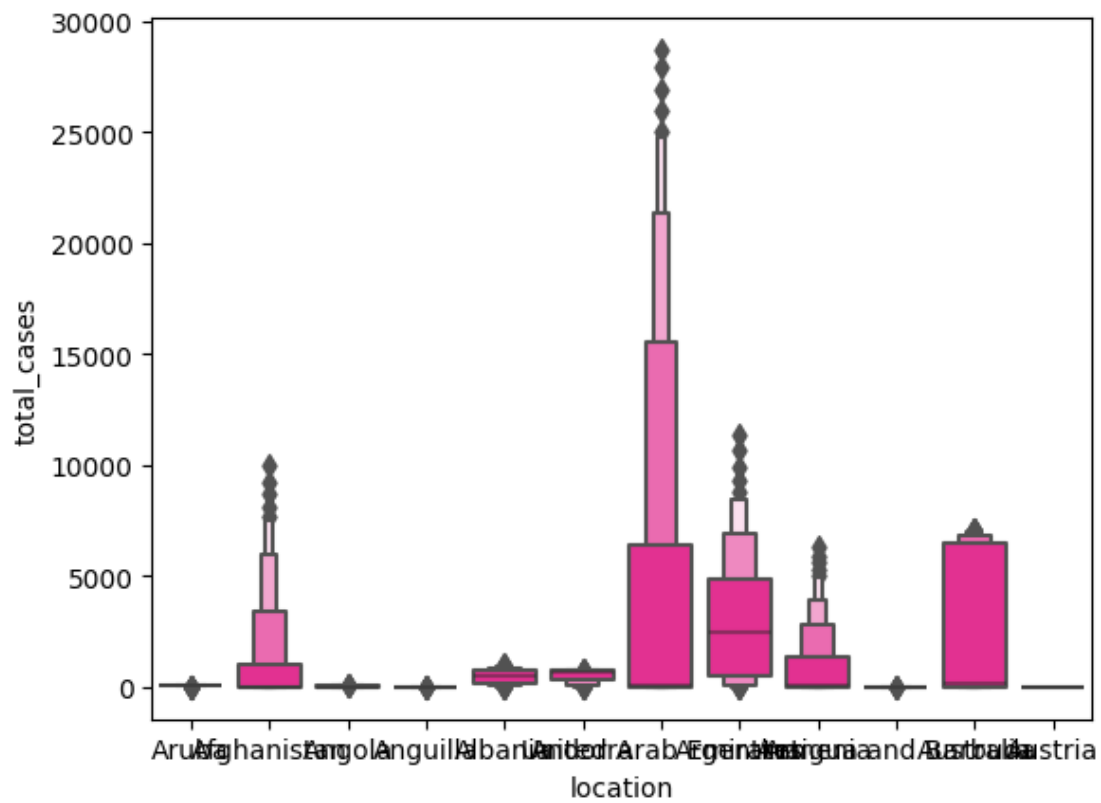
```
[31]: sns.histplot(df , x="location" , color='springgreen')
```

```
[31]: <Axes: xlabel='location', ylabel='Count'>
```



```
[32]: sns.boxenplot(df , x="location" , y='total_cases', color = 'deeppink')
```

```
[32]: <Axes: xlabel='location', ylabel='total_cases'>
```



```
[33]: sns.barplot(df , x="total_cases" , y='total_deaths' , color='green')
```

```
[33]: <Axes: xlabel='total_cases', ylabel='total_deaths'>
```

