In [144…
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as snr
```

In [145…
```
df=pd.read_csv("https://raw.githubusercontent.com/SR1608/Datasets/main/covid-data.csv")
```

Numbers of rows and columns in dataset

In [146…
```
df.shape
```

Out[146…
```
(57394, 49)
```

In [147…
```
df.dtypes
```

Out[147…
```
iso_code                                object
continent                               object
location                                object
date                                    object
total_cases                            float64
new_cases                              float64
new_cases_smoothed                     float64
total_deaths                           float64
new_deaths                             float64
new_deaths_smoothed                    float64
total_cases_per_million                float64
new_cases_per_million                  float64
new_cases_smoothed_per_million         float64
total_deaths_per_million               float64
new_deaths_per_million                 float64
new_deaths_smoothed_per_million        float64
reproduction_rate                      float64
icu_patients                           float64
icu_patients_per_million               float64
hosp_patients                          float64
hosp_patients_per_million              float64
weekly_icu_admissions                  float64
weekly_icu_admissions_per_million      float64
weekly_hosp_admissions                 float64
weekly_hosp_admissions_per_million     float64
total_tests                            float64
new_tests                              float64
total_tests_per_thousand               float64
new_tests_per_thousand                 float64
new_tests_smoothed                     float64
new_tests_smoothed_per_thousand        float64
tests_per_case                         float64
positive_rate                          float64
stringency_index                       float64
population                             float64
population_density                     float64
median_age                             float64
aged_65_older                          float64
aged_70_older                          float64
gdp_per_capita                         float64
extreme_poverty                        float64
cardiovasc_death_rate                  float64
diabetes_prevalence                    float64
female_smokers                         float64
male_smokers                           float64
handwashing_facilities                 float64
hospital_beds_per_thousand             float64
life_expectancy                        float64
human_development_index                float64
dtype: object
```

In [148…
```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 57394 entries, 0 to 57393
Data columns (total 49 columns):
 #   Column                                Non-Null Count  Dtype  
---  ------                                --------------  -----  
 0   iso_code                              57071 non-null  object 
 1   continent                             56748 non-null  object 
 2   location                              57394 non-null  object 
 3   date                                  57394 non-null  object 
 4   total_cases                           53758 non-null  float64
 5   new_cases                             56465 non-null  float64
 6   new_cases_smoothed                    55652 non-null  float64
 7   total_deaths                          44368 non-null  float64
 8   new_deaths                            56465 non-null  float64
 9   new_deaths_smoothed                   55652 non-null  float64
 10  total_cases_per_million               53471 non-null  float64
 11  new_cases_per_million                 56401 non-null  float64
 12  new_cases_smoothed_per_million        55587 non-null  float64
 13  total_deaths_per_million              44096 non-null  float64
 14  new_deaths_per_million                56401 non-null  float64
 15  new_deaths_smoothed_per_million       55587 non-null  float64
 16  reproduction_rate                     37696 non-null  float64
 17  icu_patients                          4490 non-null   float64
 18  icu_patients_per_million              4490 non-null   float64
 19  hosp_patients                         5005 non-null   float64
 20  hosp_patients_per_million             5005 non-null   float64
 21  weekly_icu_admissions                 357 non-null    float64
 22  weekly_icu_admissions_per_million     357 non-null    float64
 23  weekly_hosp_admissions                645 non-null    float64
 24  weekly_hosp_admissions_per_million    645 non-null    float64
 25  total_tests                           22017 non-null  float64
 26  new_tests                             21787 non-null  float64
 27  total_tests_per_thousand              22017 non-null  float64
 28  new_tests_per_thousand                21787 non-null  float64
 29  new_tests_smoothed                    24612 non-null  float64
 30  new_tests_smoothed_per_thousand       24612 non-null  float64
 31  tests_per_case                        22802 non-null  float64
 32  positive_rate                         23211 non-null  float64
 33  stringency_index                      47847 non-null  float64
 34  population                            57071 non-null  float64
 35  population_density                    54371 non-null  float64
 36  median_age                            51034 non-null  float64
 37  aged_65_older                         50265 non-null  float64
 38  aged_70_older                         50768 non-null  float64
 39  gdp_per_capita                        50367 non-null  float64
 40  extreme_poverty                       33571 non-null  float64
 41  cardiovasc_death_rate                 51013 non-null  float64
 42  diabetes_prevalence                   52881 non-null  float64
 43  female_smokers                        39669 non-null  float64
 44  male_smokers                          39156 non-null  float64
 45  handwashing_facilities                24176 non-null  float64
 46  hospital_beds_per_thousand            45936 non-null  float64
 47  life_expectancy                       56336 non-null  float64
 48  human_development_index               49247 non-null  float64
dtypes: float64(45), object(4)
memory usage: 21.5+ MB
```

In [149…
```
pd.set_option("display.max_columns",None)
df.describe(include="all").round(2)
```

Out[149…

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	t
count	57071	56748	57394	57394	53758.00	56465.00	55652.00	
unique	215	6	216	323	NaN	NaN	NaN	
top	AFG	Europe	Afghanistan	30/10/20	NaN	NaN	NaN	
freq	323	14828	323	215	NaN	NaN	NaN	
mean	NaN	NaN	NaN	NaN	167797.37	1953.58	1920.43	
std	NaN	NaN	NaN	NaN	1693037.55	18269.65	17777.39	
min	NaN	NaN	NaN	NaN	1.00	-8261.00	-552.00	
25%	NaN	NaN	NaN	NaN	180.00	0.00	0.86	
50%	NaN	NaN	NaN	NaN	2070.00	14.00	19.43	
75%	NaN	NaN	NaN	NaN	22356.75	235.00	245.29	
max	NaN	NaN	NaN	NaN	55154651.00	646281.00	584981.86	

In [150…

```
df.head()
```

Out[150…

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_de
0	AFG	Asia	Afghanistan	31/12/19	NaN	0.0	NaN	
1	AFG	Asia	Afghanistan	01/01/20	NaN	0.0	NaN	
2	AFG	Asia	Afghanistan	02/01/20	NaN	0.0	NaN	
3	AFG	Asia	Afghanistan	03/01/20	NaN	0.0	NaN	
4	AFG	Asia	Afghanistan	04/01/20	NaN	0.0	NaN	

In [151…

```
unique_locations=df["location"].nunique()
print(unique_locations)
```

```
216
```

In [152…

```
max_continent=df["continent"].value_counts().idxmax()
print(max_continent)
```

```
Europe
```

In [153…

```
Max_value_total_cases=df["total_cases"].max()
print(Max_value_total_cases)
```

```
55154651.0
```

In [154…

```
Mean_value_total_cases=df["total_cases"].sum()/len(df["total_cases"])
print(Mean_value_total_cases)
```

```
157167.1421402934
```

In [155…

```
df["total_deaths"].describe().round(2)
```

Out[155…

```
count      44368.00
mean        6858.64
std        55780.81
min            1.00
25%           13.00
50%           84.00
75%          727.00
max      1328537.00
Name: total_deaths, dtype: float64
```

In [156…

```
df.groupby("continent")['gdp_per_capita'].mean().reset_index()
```

Out[156…

	continent	gdp_per_capita
0	Africa	5606.467862
1	Asia	23228.392698
2	Europe	34460.887430
3	North America	22074.427430
4	Oceania	23950.103226
5	South America	13785.533786

In [157…

```
df.groupby("continent")["human_development_index"].mean().reset_index()
```

Out[157…

	continent	human_development_index
0	Africa	0.547803
1	Asia	0.731664
2	Europe	0.868582
3	North America	0.750815
4	Oceania	0.793431
5	South America	0.751435

In [158…

```
df=df[["continent","location","date","total_cases","total_deaths","gdp_per_capita","human_d
df
```

Out[158…

	continent	location	date	total_cases	total_deaths	gdp_per_capita	human_developn
0	Asia	Afghanistan	31/12/19	NaN	NaN	1803.987	
1	Asia	Afghanistan	01/01/20	NaN	NaN	1803.987	
2	Asia	Afghanistan	02/01/20	NaN	NaN	1803.987	
3	Asia	Afghanistan	03/01/20	NaN	NaN	1803.987	
4	Asia	Afghanistan	04/01/20	NaN	NaN	1803.987	
…	…	…	…	…	…	…	
57389	NaN	International	13/11/20	696.0	7.0	NaN	
57390	NaN	International	14/11/20	696.0	7.0	NaN	
57391	NaN	International	15/11/20	696.0	7.0	NaN	
57392	NaN	International	16/11/20	696.0	7.0	NaN	
57393	NaN	International	17/11/20	696.0	7.0	NaN	

57394 rows × 7 columns

In [159…
```
df.duplicated().sum()
```

Out[159…
```
0
```

In [160…
```
df.drop_duplicates()
```

Out[160…

	continent	location	date	total_cases	total_deaths	gdp_per_capita	human_developn
0	Asia	Afghanistan	31/12/19	NaN	NaN	1803.987	
1	Asia	Afghanistan	01/01/20	NaN	NaN	1803.987	
2	Asia	Afghanistan	02/01/20	NaN	NaN	1803.987	
3	Asia	Afghanistan	03/01/20	NaN	NaN	1803.987	
4	Asia	Afghanistan	04/01/20	NaN	NaN	1803.987	
…	…	…	…	…	…	…	
57389	NaN	International	13/11/20	696.0	7.0	NaN	
57390	NaN	International	14/11/20	696.0	7.0	NaN	
57391	NaN	International	15/11/20	696.0	7.0	NaN	
57392	NaN	International	16/11/20	696.0	7.0	NaN	
57393	NaN	International	17/11/20	696.0	7.0	NaN	

57394 rows × 7 columns

In [161…
```
df.isnull().sum()
```

Out[161…

```
continent                 646
location                    0
date                        0
total_cases              3636
total_deaths            13026
gdp_per_capita           7027
human_development_index  8147
dtype: int64
```

In [162…

```
df.dropna(subset=["continent"])
```

Out[162…

	continent	location	date	total_cases	total_deaths	gdp_per_capita	human_developm
0	Asia	Afghanistan	31/12/19	NaN	NaN	1803.987	
1	Asia	Afghanistan	01/01/20	NaN	NaN	1803.987	
2	Asia	Afghanistan	02/01/20	NaN	NaN	1803.987	
3	Asia	Afghanistan	03/01/20	NaN	NaN	1803.987	
4	Asia	Afghanistan	04/01/20	NaN	NaN	1803.987	
...	...	...	...	...	...	...	
56743	Africa	Zimbabwe	13/11/20	8696.0	255.0	1899.775	
56744	Africa	Zimbabwe	14/11/20	8765.0	257.0	1899.775	
56745	Africa	Zimbabwe	15/11/20	8786.0	257.0	1899.775	
56746	Africa	Zimbabwe	16/11/20	8786.0	257.0	1899.775	
56747	Africa	Zimbabwe	17/11/20	8897.0	257.0	1899.775	

56748 rows × 7 columns

In [163…

```
df.fillna(0)
```

Out[163…

	continent	location	date	total_cases	total_deaths	gdp_per_capita	human_developn
0	Asia	Afghanistan	31/12/19	0.0	0.0	1803.987	
1	Asia	Afghanistan	01/01/20	0.0	0.0	1803.987	
2	Asia	Afghanistan	02/01/20	0.0	0.0	1803.987	
3	Asia	Afghanistan	03/01/20	0.0	0.0	1803.987	
4	Asia	Afghanistan	04/01/20	0.0	0.0	1803.987	
...	...	...	...	...	...	...	
57389	0	International	13/11/20	696.0	7.0	0.000	
57390	0	International	14/11/20	696.0	7.0	0.000	
57391	0	International	15/11/20	696.0	7.0	0.000	
57392	0	International	16/11/20	696.0	7.0	0.000	
57393	0	International	17/11/20	696.0	7.0	0.000	

57394 rows × 7 columns

In [164…

```
df["date"]=pd.to_datetime(df["date"])
```

```
C:\Users\Dell\AppData\Local\Temp\ipykernel_16768\1303019909.py:1: UserWarning: Could not inf
er format, so each element will be parsed individually, falling back to `dateutil`. To ensur
e parsing is consistent and as-expected, please specify a format.
  df["date"]=pd.to_datetime(df["date"])
```

In [165…
```
df.head()
```

Out[165…

	continent	location	date	total_cases	total_deaths	gdp_per_capita	human_development_inde
0	Asia	Afghanistan	2019-12-31	NaN	NaN	1803.987	0.49
1	Asia	Afghanistan	2020-01-01	NaN	NaN	1803.987	0.49
2	Asia	Afghanistan	2020-02-01	NaN	NaN	1803.987	0.49
3	Asia	Afghanistan	2020-03-01	NaN	NaN	1803.987	0.49
4	Asia	Afghanistan	2020-04-01	NaN	NaN	1803.987	0.49

In [166…
```
df["month"]=pd.DatetimeIndex(df["date"]).month
df
```

Out[166…

	continent	location	date	total_cases	total_deaths	gdp_per_capita	human_developmen
0	Asia	Afghanistan	2019-12-31	NaN	NaN	1803.987	
1	Asia	Afghanistan	2020-01-01	NaN	NaN	1803.987	
2	Asia	Afghanistan	2020-02-01	NaN	NaN	1803.987	
3	Asia	Afghanistan	2020-03-01	NaN	NaN	1803.987	
4	Asia	Afghanistan	2020-04-01	NaN	NaN	1803.987	
...	...	...	...	...	...	...	
57389	NaN	International	2020-11-13	696.0	7.0	NaN	
57390	NaN	International	2020-11-14	696.0	7.0	NaN	
57391	NaN	International	2020-11-15	696.0	7.0	NaN	
57392	NaN	International	2020-11-16	696.0	7.0	NaN	
57393	NaN	International	2020-11-17	696.0	7.0	NaN	

57394 rows × 8 columns

In [167…
```
df_groupby=df.groupby("continent").max().reset_index()
print(df_groupby)
```

```
       continent                     location       date  total_cases  \
0         Africa                     Zimbabwe 2020-12-11     752269.0
1           Asia                        Yemen 2020-12-11    8874290.0
2         Europe                      Vatican 2020-12-11    1991233.0
3  North America  United States Virgin Islands 2020-12-11   11205486.0
4        Oceania             Wallis and Futuna 2020-12-11      27750.0
5  South America                    Venezuela 2020-12-11    5876464.0

   total_deaths  gdp_per_capita  human_development_index  month
0       20314.0       26382.287                    0.797     12
1      130519.0      116935.600                    0.933     12
2       52147.0       94277.965                    0.953     12
3      247220.0       54225.446                    0.926     12
4         907.0       44648.710                    0.939     12
5      166014.0       22767.037                    0.843     12
```

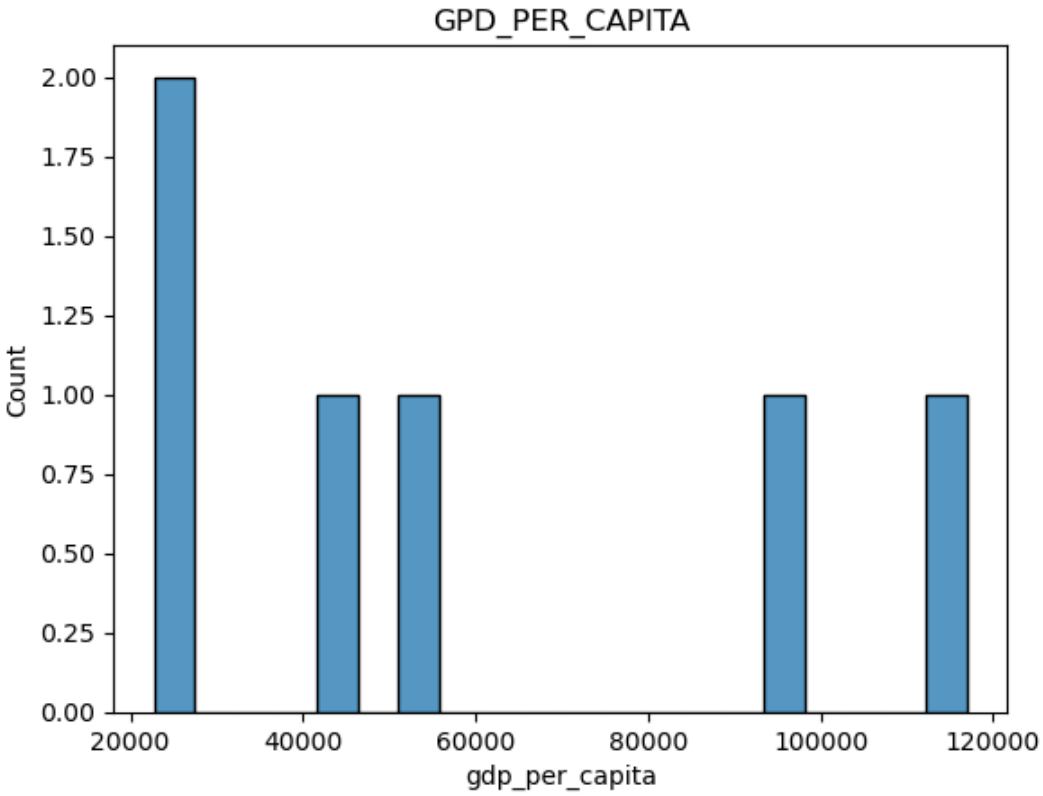
In [168…
```
df_groupby['total_deaths to total_cases_ratio']=df_groupby["total_deaths"]/df_groupby["tota
```

In [169…
```
df_groupby["total_deaths to total_cases_ratio"]*100
```

Out[169…
```
0    2.700364
1    1.470754
2    2.618830
3    2.206241
4    3.268468
5    2.825066
Name: total_deaths to total_cases_ratio, dtype: float64
```

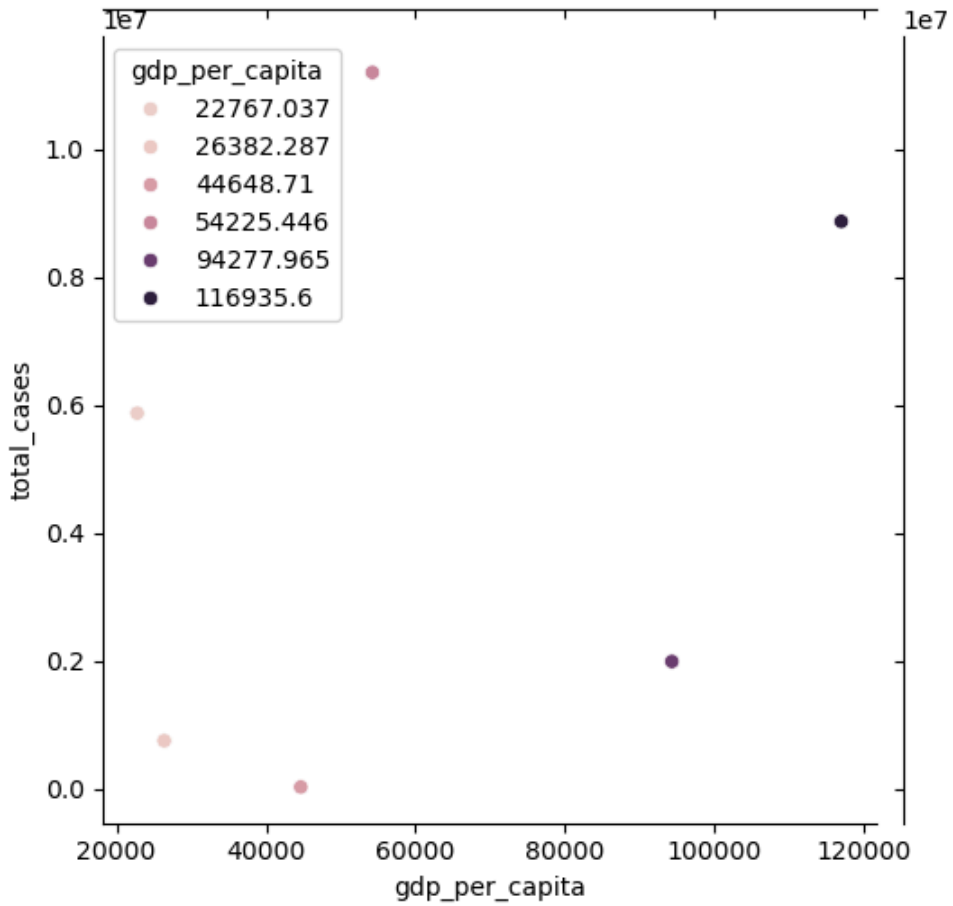
In [170…
```
snr.histplot(df_groupby["gdp_per_capita"],bins=20)
plt.title("GPD_PER_CAPITA")
```

Out[170…
```
Text(0.5, 1.0, 'GPD_PER_CAPITA')
```



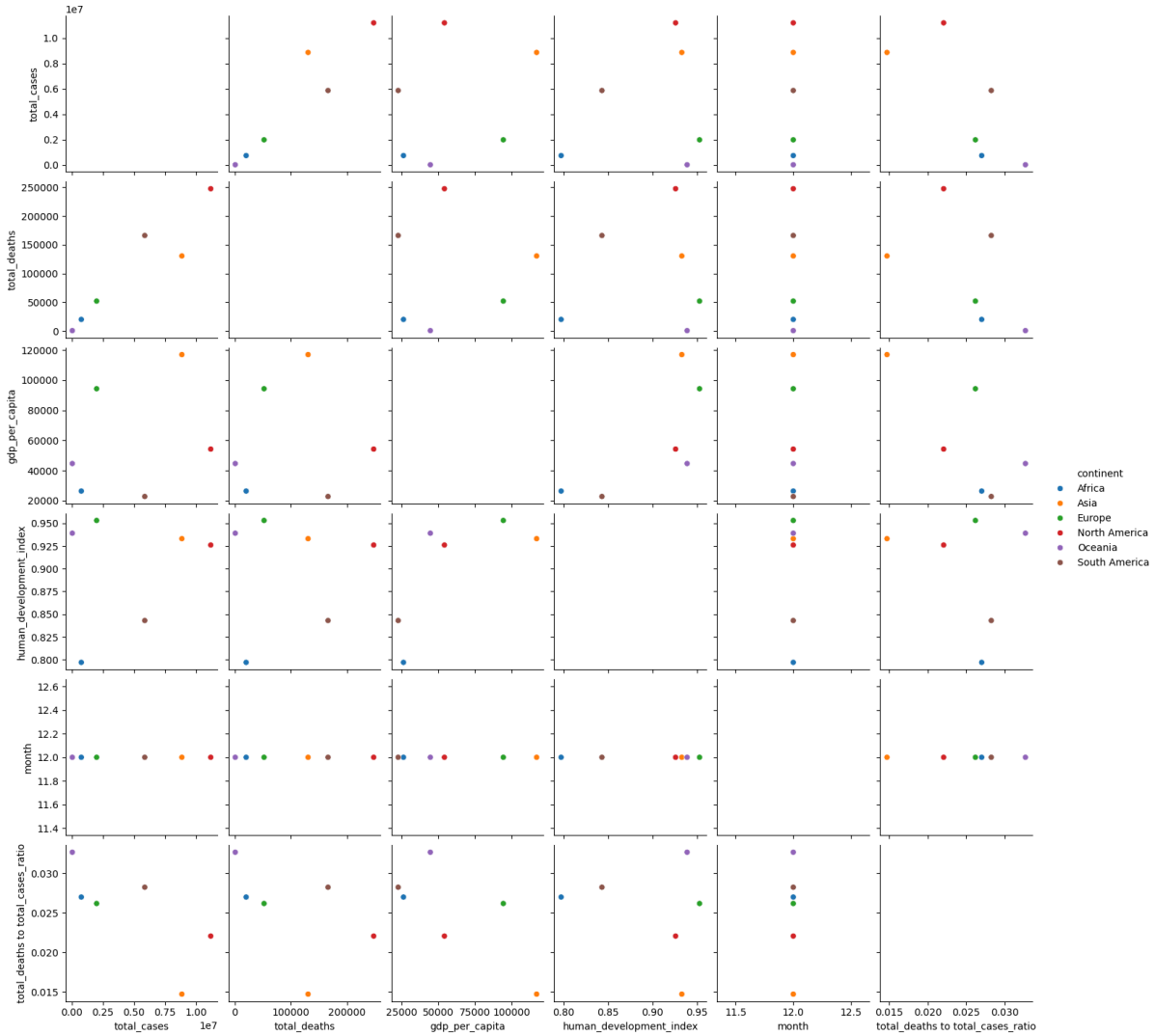
In [171…
```
snr.jointplot(x="gdp_per_capita",y="total_cases",hue="gdp_per_capita",data=df_groupby,kind=
```

Out[171…
```
<seaborn.axisgrid.JointGrid at 0x151e2c67530>
```



In [172…
```
snr.pairplot(df_groupby,hue="continent")
```

Out[172…
```
<seaborn.axisgrid.PairGrid at 0x151e29d4410>
```

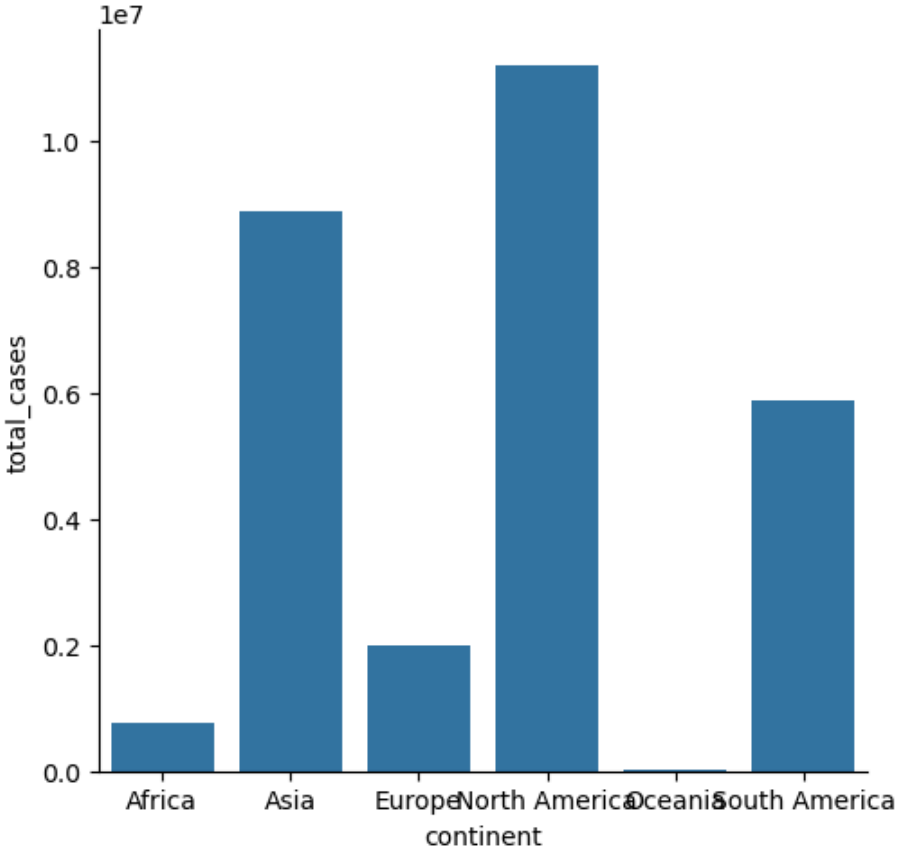


In [173…

```
snr.catplot(x="continent",y="total_cases",data=df_groupby,kind='bar')
```

Out[173…

```
<seaborn.axisgrid.FacetGrid at 0x151e5289040>
```



In [219…
```
df_groupby.to_csv("Covid_Data.csv")
```

In [221…
```
pd.read_csv("Covid_Data.csv")
```

Out[221…

	Unnamed: 0	continent	location	date	total_cases	total_deaths	gdp_per_capita	human_develo
0	0	Africa	Zimbabwe	2020-12-11	752269.0	20314.0	26382.287	
1	1	Asia	Yemen	2020-12-11	8874290.0	130519.0	116935.600	
2	2	Europe	Vatican	2020-12-11	1991233.0	52147.0	94277.965	
3	3	North America	United States Virgin Islands	2020-12-11	11205486.0	247220.0	54225.446	
4	4	Oceania	Wallis and Futuna	2020-12-11	27750.0	907.0	44648.710	
5	5	South America	Venezuela	2020-12-11	5876464.0	166014.0	22767.037	

In [ ]:

In [177…

In [ ]: