DM - CA1 - Mohamad Taha Fakharian

In this computer assignment, we're going to preprocess a dataset for covid history based on countries and display them to find some correlation between them and conclude some facts.

So let's start!

First we need to import some useful libraries for data preprocessing:

In [32]:
```
# Useful libraries
import numpy as np
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt
```

Data Preprocessing

Let's start with data preprocessing. First we need to save the dataset in a Pandas DataFrame and get a quick look at it:

In [33]:
```
covid = pd.read_csv('CA1_Dataset.csv')
covid.head()
```

Out[33]:

	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	5.0	5.0	NaN	NaN	NaN	NaN	...
1	AFG	Asia	Afghanistan	2020-02-25	5.0	0.0	NaN	NaN	NaN	NaN	...
2	AFG	Asia	Afghanistan	2020-02-26	5.0	0.0	NaN	NaN	NaN	NaN	...
3	AFG	Asia	Afghanistan	2020-02-27	5.0	0.0	NaN	NaN	NaN	NaN	...
4	AFG	Asia	Afghanistan	2020-02-28	5.0	0.0	NaN	NaN	NaN	NaN	...

5 rows × 67 columns

Q1

Let's see how many rows are NA for each column in our dataset:

In [34]:
```
pd.set_option('display.max_rows',None)
print(covid.isna().sum())
```

```
iso_code                                    0
continent                                9917
location                                    0
date                                        0
total_cases                              3030
new_cases                                3172
new_cases_smoothed                       5156
total_deaths                            20843
new_deaths                              20803
new_deaths_smoothed                     22902
total_cases_per_million                  3785
new_cases_per_million                    3927
new_cases_smoothed_per_million           5905
total_deaths_per_million                21585
new_deaths_per_million                  21545
new_deaths_smoothed_per_million         23638
reproduction_rate                       40569
icu_patients                           142246
icu_patients_per_million               142246
hosp_patients                          141072
hosp_patients_per_million              141072
weekly_icu_admissions                  160232
weekly_icu_admissions_per_million      160232
weekly_hosp_admissions                 154759
weekly_hosp_admissions_per_million     154759
new_tests                               98630
```

```
total_tests                                   96692
total_tests_per_thousand                      96692
new_tests_per_thousand                        98630
new_tests_smoothed                            81978
new_tests_smoothed_per_thousand               81978
positive_rate                                 87046
tests_per_case                                87609
tests_units                                   79655
total_vaccinations                           120658
people_vaccinated                            122844
people_fully_vaccinated                      125608
total_boosters                               148296
new_vaccinations                             128384
new_vaccinations_smoothed                     81524
total_vaccinations_per_hundred               120658
people_vaccinated_per_hundred                122844
people_fully_vaccinated_per_hundred          125608
total_boosters_per_hundred                   148296
new_vaccinations_smoothed_per_million         81524
new_people_vaccinated_smoothed                82815
new_people_vaccinated_smoothed_per_hundred    82815
stringency_index                              35774
population                                     1072
population_density                            18323
median_age                                    28378
aged_65_older                                 29866
aged_70_older                                 29114
gdp_per_capita                                27708
extreme_poverty                               74799
cardiovasc_death_rate                         29428
diabetes_prevalence                           22287
female_smokers                                60027
male_smokers                                  61476
handwashing_facilities                        97352
hospital_beds_per_thousand                    42485
life_expectancy                               11016
human_development_index                       29953
excess_mortality_cumulative_absolute         159940
excess_mortality_cumulative                  159940
excess_mortality                             159940
excess_mortality_cumulative_per_million      159940
dtype: int64
```

Let's replace them with appropriate values. As you can see, there are two non-numeric columns listed above: test_units and continent.

For continent, NA values are for rows which their locations are not really specific, like International or Africa(their records are not location-specific). For each NA value in continent's column, we'll check if its location is name of a continent or not; If it is, we'll fill NA will its location value. Otherwise, the continent is undefined and we'll fill it with UNDEFINED value.

For test_units, we had a great approach for filling NA values: For each NA value in test_units column, we check whether rows with same location have any non-NA value for test_units column: If there is any, we would fill NA with mode of that column with same location as that record. Otherwise, we would fill NA with mode of that column in total dataset. This approach can be applied for other columns too. It's a better replacement method than the traditional one, because it uses records for same location. We don't use this approach for this dataset, because it's too much time-consuming. The code for this approach is commented in the cell below. We'll use the traditional method so(mode of total dataset).

For numeric columns, there are two kind of columns: Daily stat like new cases and overall stats like population. For daily stat, since they're important in estimations and can be inaccurate if filled by mean of that column(for example, aggregation on new deaths after filling NA can be more than that country's population), we'll fill it with zero. For other stats, we first check if ratio of missing values to total records isn't that big. If it's really big(like 0.9 or even bigger), we'll fill NA with a constant like $-\infty$(because there isn't sufficient data for us to predict the missing value).Otherwise, we'll fill it with mean of that column.

Pay attention! There are other methods for filling NA like decision trees. Sufficient data and time complexity is a big problem, so methods mentioned above are good for this dataset.

Let's fill NA values with this methods:

In [35]:

```
from pandas.api.types import is_numeric_dtype
unique_continent = covid['continent'].unique()

ACCEPTABLE_MISSING_PERCENTAGE = 0.9
total = len(covid.index)

for i in covid.columns[covid.isnull().any(axis=0)]:
    if is_numeric_dtype(covid[i]):
        if i.startswith('new') or i.startswith('total'):
            covid[i].fillna(0, inplace=True)
        elif covid[i].isna().sum() / total > ACCEPTABLE_MISSING_PERCENTAGE:
            covid[i].fillna(-np.inf, inplace=True)
```

```
        else:
            covid[i].fillna(covid[i].mean(), inplace=True)
    else:
        if i == 'continent':
            covid['continent'] = covid.apply(
                lambda row: row['location'] if (pd.isna(row['continent']) and (row['location'] in unique_continer
                else ('UNDEFINED' if pd.isna(row['continent'])
                else row['continent']) , axis=1)
        else:
            covid[i].fillna(covid[i].mode()[0], inplace=True)
            # covid['tests_units'] = covid.apply(
            #      lambda row : covid['tests_units'].mode()[0] if (covid[covid['location'] == row['location']]['t
            #     else (covid[covid['location'] == row['location']]['tests_units'].mode()[0] if pd.isna(row['test
            #     else row['tests_units']), axis = 1
```

Finally, let's make sure that no more NA value is remained:

In [36]:

```
print(covid.isna().sum())
```

```
iso_code                                   0
continent                                  0
location                                   0
date                                       0
total_cases                                0
new_cases                                  0
new_cases_smoothed                         0
total_deaths                               0
new_deaths                                 0
new_deaths_smoothed                        0
total_cases_per_million                    0
new_cases_per_million                      0
new_cases_smoothed_per_million             0
total_deaths_per_million                   0
new_deaths_per_million                     0
new_deaths_smoothed_per_million            0
reproduction_rate                          0
icu_patients                               0
icu_patients_per_million                   0
hosp_patients                              0
hosp_patients_per_million                  0
weekly_icu_admissions                      0
weekly_icu_admissions_per_million          0
weekly_hosp_admissions                     0
weekly_hosp_admissions_per_million         0
new_tests                                  0
total_tests                                0
total_tests_per_thousand                   0
new_tests_per_thousand                     0
new_tests_smoothed                         0
new_tests_smoothed_per_thousand            0
positive_rate                              0
tests_per_case                             0
tests_units                                0
total_vaccinations                         0
people_vaccinated                          0
people_fully_vaccinated                    0
total_boosters                             0
new_vaccinations                           0
new_vaccinations_smoothed                  0
total_vaccinations_per_hundred             0
people_vaccinated_per_hundred              0
people_fully_vaccinated_per_hundred        0
total_boosters_per_hundred                 0
new_vaccinations_smoothed_per_million      0
new_people_vaccinated_smoothed             0
new_people_vaccinated_smoothed_per_hundred 0
stringency_index                           0
population                                 0
population_density                         0
median_age                                 0
aged_65_older                              0
aged_70_older                              0
gdp_per_capita                             0
extreme_poverty                            0
cardiovasc_death_rate                      0
diabetes_prevalence                        0
female_smokers                             0
male_smokers                               0
handwashing_facilities                     0
hospital_beds_per_thousand                 0
life_expectancy                            0
human_development_index                    0
excess_mortality_cumulative_absolute       0
excess_mortality_cumulative                0
excess_mortality                           0
excess_mortality_cumulative_per_million    0
```

```
dtype: int64
```

In [37]:
```
covid.to_csv('Completed.csv')
```

Q2

Now we want to get a better sense of data for each country. Let's aggregate new_cases, new_deaths and location columns for each country by date. For aggregation of new_cases and new_deaths, we'll use sum but for population, since its value is same for rows with same location(country), we'll use min function. So let's start:

In [38]:
```
pd.reset_option('display.max_rows')
```

In [39]:
```
summarized = covid[["location", "new_cases", "new_deaths", "new_vaccinations"]]
summarized = summarized.groupby(by=["location"]).sum()

summarized = summarized.join(covid[["location", "population"]].groupby(by=["location"]).min())
summarized
```

Out[39]:

location	new_cases	new_deaths	new_vaccinations	population
Afghanistan	174081.0	7617.0	1.374200e+04	3.983543e+07
Africa	11230524.0	248668.0	5.818190e+08	1.373486e+09
Albania	271825.0	3474.0	1.415150e+06	2.872934e+06
Algeria	265079.0	6843.0	1.707860e+05	4.461663e+07
Andorra	38249.0	151.0	4.802000e+03	7.735400e+04
...	...	...	...	...
Wallis and Futuna	454.0	7.0	0.000000e+00	1.109400e+04
World	439011701.0	5946817.0	1.122516e+10	7.874966e+09
Yemen	11772.0	2135.0	0.000000e+00	3.049064e+07
Zambia	313203.0	3955.0	1.321156e+06	1.892066e+07
Zimbabwe	237509.0	5396.0	7.080893e+06	1.509217e+07

238 rows × 4 columns

In [40]:
```
summarized.to_csv('Summarized.csv')
```

Q3

To add Jalali date to dataframe, we use 'jdatetime' library. Unfortunately, jdatetime library doesn't work with Numpy and Pandas library and thus, we can't use vectorization to calculate Jalali date. Thus, we'll loop over rows and calculate Jalali date for each row:

In [41]:
```
import jdatetime
covid['date'] = pd.to_datetime(covid['date'])
covid['jalali_date'] = covid['date']

for index, row in covid.iterrows():
    y, m , d = row['date'].year, row['date'].month, row['date'].day
    covid.at[index, 'jalali_date'] = str(jdatetime.date.fromgregorian(day=int(d),month=int(m),year=int(y)))
```

In [42]:
```
covid
```

Out[42]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothec
0	AFG	Asia	Afghanistan	2020-02-24	5.0	5.0	0.000	0.0	0.0	0.00C
1	AFG	Asia	Afghanistan	2020-02-25	5.0	0.0	0.000	0.0	0.0	0.00C
2	AFG	Asia	Afghanistan	2020-02-26	5.0	0.0	0.000	0.0	0.0	0.00C
3	AFG	Asia	Afghanistan	2020-02-27	5.0	0.0	0.000	0.0	0.0	0.00C

4	AFG	Asia	Afghanistan	2020-02-28	5.0	0.0	0.000	0.0	0.0	0.000
...	...	...	...	...	...	...	...	...	...	...
165631	ZWE	Africa	Zimbabwe	2022-02-26	235803.0	336.0	368.429	5393.0	1.0	1.000
165632	ZWE	Africa	Zimbabwe	2022-02-27	235803.0	0.0	350.143	5393.0	0.0	1.000
165633	ZWE	Africa	Zimbabwe	2022-02-28	236380.0	577.0	401.286	5395.0	2.0	1.286
165634	ZWE	Africa	Zimbabwe	2022-03-01	236871.0	491.0	413.000	5395.0	0.0	1.000
165635	ZWE	Africa	Zimbabwe	2022-03-02	237503.0	632.0	416.286	5396.0	1.0	1.143

165636 rows × 68 columns

In [43]:

```
covid.to_csv('Jalali.csv')
```

Q4

To find redundant features in dataframe, we can calculate correlation between features, which is known as correlation matrix. Let's calculate correlation matrix!

In [44]:

```
covid.corr()
```

Out[44]:

	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	total_c
total_cases	1.000000	0.806677	0.827899	0.968956	0.726685	0.747316	
new_cases	0.806677	1.000000	0.976958	0.735240	0.668096	0.654269	
new_cases_smoothed	0.827899	0.976958	1.000000	0.749612	0.663865	0.673995	
total_deaths	0.968956	0.735240	0.749612	1.000000	0.781224	0.803281	
new_deaths	0.726685	0.668096	0.663865	0.781224	1.000000	0.975575	
...	...	...	...	...	...	...	
human_development_index	0.029702	0.028335	0.028472	0.030726	0.032106	0.032645	
excess_mortality_cumulative_absolute	0.802190	0.464078	0.509117	0.928284	0.493040	0.655323	
excess_mortality_cumulative	0.139572	0.090745	0.075321	0.338096	0.301056	0.254753	
excess_mortality	0.034064	0.092382	0.070259	0.110488	0.395089	0.322358	
excess_mortality_cumulative_per_million	0.228241	0.124742	0.127365	0.370520	0.193819	0.203090	

62 rows × 62 columns

as you can see, some of features are correlated and are saved redundantly. For example, look at correlation of new_cases and total_cases. Their correlation shows that using new_cases, we can calculate total_cases from that, which is clearly true. In particular, for each feature in dataframe, if that feature could be calculated using some function(like sum or max) and other features, we can drop that feature, since it's redundant feature. In this dataframe, 'total' features can be dropped, according to this explanation.

Q5

Now we'll get Iran's dataframe from original dataframe:

In [45]:

```
iran = covid[covid['location'] == 'Iran']
iran
```

Out[45]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...
71639	IRN	Asia	Iran	2020-02-19	2.0	2.0	0.000	2.0	2.0	0.000	...
71640	IRN	Asia	Iran	2020-02-20	5.0	3.0	0.000	2.0	0.0	0.000	...
71641	IRN	Asia	Iran	2020-02-21	18.0	13.0	0.000	4.0	2.0	0.000	...
				2020-							

71642	IRN	Asia	Iran	02-22	28.0	10.0	0.000	5.0	1.0	0.000	...
71643	IRN	Asia	Iran	2020-02-23	43.0	15.0	0.000	8.0	3.0	0.000	...
...	...	...	...	...	...	...	...	...	...	...	...
72377	IRN	Asia	Iran	2022-02-26	7030943.0	7039.0	15065.429	136390.0	224.0	227.429	...
72378	IRN	Asia	Iran	2022-02-27	7040467.0	9524.0	14002.143	136631.0	241.0	227.286	...
72379	IRN	Asia	Iran	2022-02-28	7051429.0	10962.0	12838.143	136838.0	207.0	223.143	...
72380	IRN	Asia	Iran	2022-03-01	7060741.0	9312.0	11015.143	137064.0	226.0	223.571	...
72381	IRN	Asia	Iran	2022-03-02	7066975.0	6234.0	9714.286	137267.0	203.0	220.143	...

743 rows × 68 columns

In [46]:

```
iran.to_csv('Iran.csv')
```

Q6

Let's extract month from date and add a month column to iran's dataframe:

In [47]:

```
iran['month'] = iran['date'].dt.month
iran
```

```
/home/taha/.local/lib/python3.7/site-packages/ipykernel_launcher.py:1: SettingWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

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

Out[47]:

	iso_code	continent	location	date	total_cases	new_cases	new_cases_smoothed	total_deaths	new_deaths	new_deaths_smoothed	...
71639	IRN	Asia	Iran	2020-02-19	2.0	2.0	0.000	2.0	2.0	0.000	...
71640	IRN	Asia	Iran	2020-02-20	5.0	3.0	0.000	2.0	0.0	0.000	...
71641	IRN	Asia	Iran	2020-02-21	18.0	13.0	0.000	4.0	2.0	0.000	...
71642	IRN	Asia	Iran	2020-02-22	28.0	10.0	0.000	5.0	1.0	0.000	...
71643	IRN	Asia	Iran	2020-02-23	43.0	15.0	0.000	8.0	3.0	0.000	...
...	...	...	...	...	...	...	...	...	...	...	...
72377	IRN	Asia	Iran	2022-02-26	7030943.0	7039.0	15065.429	136390.0	224.0	227.429	...
72378	IRN	Asia	Iran	2022-02-27	7040467.0	9524.0	14002.143	136631.0	241.0	227.286	...
72379	IRN	Asia	Iran	2022-02-28	7051429.0	10962.0	12838.143	136838.0	207.0	223.143	...
72380	IRN	Asia	Iran	2022-03-01	7060741.0	9312.0	11015.143	137064.0	226.0	223.571	...
72381	IRN	Asia	Iran	2022-03-02	7066975.0	6234.0	9714.286	137267.0	203.0	220.143	...

743 rows × 69 columns

Q7

Now let's aggregate iran's dataframe by month in 2021!

Here, we'll drop 'date' and 'jalali_date' columns for aggregation(since they're meaningless for aggregation):

In [48]:

```
iran_2021 = iran[iran['date'].dt.year == 2021]
news = []
for column in iran_2021:
    if column.startswith('new'):
        news.append(column)

news.append('month')
iran_news = iran_2021[news].copy()
news.pop()
news.append('date')
news.append('jalali_date')
iran_others = iran_2021.drop(news, axis=1)

iran_news = iran_news.groupby(by="month").sum()
iran_others = iran_others.groupby(by="month").max()
iran_summarized = iran_news.join(iran_others)
iran_summarized
```

Out[48]:

	new_cases	new_cases_smoothed	new_deaths	new_deaths_smoothed	new_cases_per_million	new_cases_smoothed_per_million	new_de
month							
1	192857.0	191987.285	2736.0	2899.717	2268.141	2257.909	
2	213170.0	208151.000	2114.0	2095.712	2507.034	2448.008	
3	254395.0	249951.002	2592.0	2575.569	2991.869	2939.607	
4	613513.0	581349.857	9093.0	8084.143	7215.358	6837.099	
5	414059.0	446317.573	8398.0	9098.854	4869.633	5249.018	
6	291421.0	286083.287	4108.0	4275.714	3427.322	3364.546	
7	666451.0	617035.143	6366.0	5891.143	7837.948	7256.777	
8	1121055.0	1108662.715	17164.0	16160.286	13184.422	13038.678	
9	594977.0	650810.142	12634.0	13715.144	6997.364	7654.000	
10	337598.0	350551.429	5875.0	6220.143	3970.399	4122.742	
11	192807.0	207642.000	3527.0	3705.283	2267.547	2442.022	
12	76956.0	83420.570	1776.0	1932.000	905.060	981.088	

12 rows × 66 columns

In [49]:

```
iran_summarized.to_csv('Iran_Summarized.csv')
```

Data Visualization

In this part, we're going to visualize dataframes calculated in previous part and conclude from them. Let's start!

We'll use 'seaborn' library for data visualization:

In [50]:

```
import seaborn as sns
```

Q1

First we want to see which countries have had best and worst performance in controlling Corona Virus. We'll use total deaths per population for the measurement of performance and use a bar chart to see this rate among all countries. Let's go!

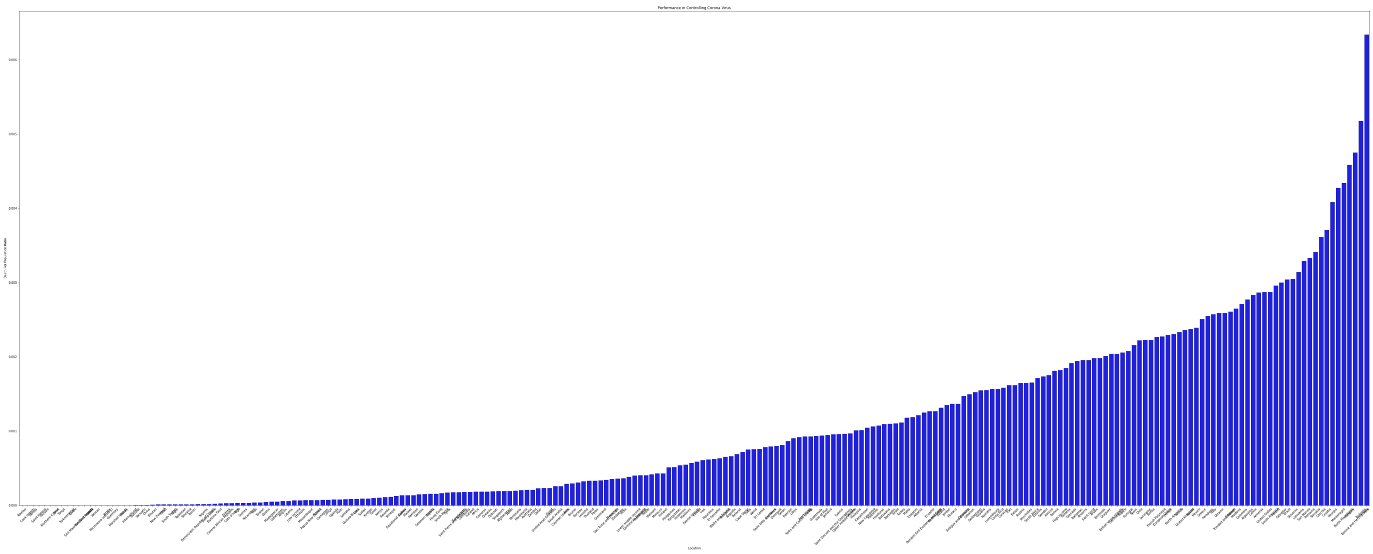
In [51]:

```
summarized_copy = summarized.copy()
summarized_copy = summarized_copy[['new_deaths', 'population']]
summarized_copy['deaths_per_population'] = summarized_copy['new_deaths'] / summarized_copy['population']
summarized_copy = summarized_copy.sort_values(by=['deaths_per_population'])
best = summarized_copy[:10]
best_over_zero = summarized_copy[summarized_copy['deaths_per_population'] > 0][:10]
worst = summarized_copy[-10:]
```

In [52]:

```
plt.figure(figsize=(80,30))
ax = sns.barplot(x=summarized_copy.index, y=summarized_copy['deaths_per_population'], color='b')
plt.xticks(rotation = 45)
ax.set_title("Performance in Controlling Corona Virus")
```

```
ax.set_xlabel('Location')
ax.set_ylabel('Death Per Population Ratio')
plt.show()
```



Let's see the results. Since for some countries, dataset may not have enough data for deaths, we get the best countries with ratio greater than zero too. In the following cells, you can see best and worst countries with their performances:

In [53]:
```
best
```

Out[53]:

	new_deaths	population	deaths_per_population
location			
Tokelau	0.0	1.368000e+03	0.0
Cook Islands	0.0	1.757200e+04	0.0
Samoa	0.0	2.001440e+05	0.0
Saint Helena	0.0	6.095000e+03	0.0
Pitcairn	0.0	4.700000e+01	0.0
Northern Cyprus	0.0	1.474690e+08	0.0
Niue	0.0	1.614000e+03	0.0
Tonga	0.0	1.067590e+05	0.0
Turkmenistan	0.0	6.117933e+06	0.0
Tuvalu	0.0	1.192500e+04	0.0

In [54]:
```
best_over_zero
```

Out[54]:

	new_deaths	population	deaths_per_population
location			
International	15.0	1.474690e+08	1.017163e-07
Burundi	38.0	1.225543e+07	3.100667e-06
Vanuatu	1.0	3.144640e+05	3.180014e-06
China	4640.0	1.444216e+09	3.212816e-06
Bhutan	6.0	7.799000e+05	7.693294e-06
New Zealand	56.0	5.126300e+06	1.092406e-05
Chad	190.0	1.691498e+07	1.123264e-05
South Sudan	137.0	1.138138e+07	1.203721e-05
Niger	307.0	2.513081e+07	1.221608e-05
Tajikistan	125.0	9.749625e+06	1.282101e-05

In [55]:
```
worst
```

Out[55]:

	new_deaths	population	deaths_per_population

location			
Slovakia	18567.0	5449270.0	0.003407
Czechia	38771.0	10724553.0	0.003615
Croatia	15122.0	4081657.0	0.003705
Georgia	16231.0	3979773.0	0.004078
Montenegro	2683.0	628051.0	0.004272
North Macedonia	9036.0	2082661.0	0.004339
Hungary	44134.0	9634162.0	0.004581
Bosnia and Herzegovina	15506.0	3263459.0	0.004751
Bulgaria	35696.0	6896655.0	0.005176
Peru	211453.0	33359415.0	0.006339

To conclude, we can see that countries with better facilities in health, have had better performance in dealing with COVID-19.

Q2

Now we want to see the effect of vaccination on number of deaths. To get a better sense, we'll sort countries by ratio of total vaccinations to population and from each quintile, we'll sample one country and finally, we'll plot a scatter plot to see the effect of vaccination on number of deaths. Let's go!

In [56]:

```
pd.set_option('display.max_rows',None)

summarized_copy = summarized.copy()
summarized_copy['first_ratio'] = (summarized_copy['new_vaccinations'] / summarized_copy['population'])
summarized_copy['second_ratio'] = (summarized_copy['new_deaths'] / summarized_copy['population'])

summarized_copy = summarized_copy.sort_values(by= 'first_ratio', ascending= False)

summarized_copy
```

Out[56]:

location	new_cases	new_deaths	new_vaccinations	population	first_ratio	second_ratio
Cuba	1070730.0	8497.0	3.065243e+07	1.131750e+07	2.708410	7.507843e-04
Gibraltar	15632.0	101.0	8.906400e+04	3.369100e+04	2.643555	2.997833e-03
Chile	3098110.0	42624.0	4.855777e+07	1.921236e+07	2.527423	2.218572e-03
Singapore	767663.0	1040.0	1.357254e+07	5.453600e+06	2.488730	1.906997e-04
Malta	71497.0	608.0	1.205547e+06	5.161000e+05	2.335879	1.178066e-03
South Korea	3691487.0	8394.0	1.192514e+08	5.130518e+07	2.324354	1.636092e-04
Denmark	2805858.0	4633.0	1.317204e+07	5.813302e+06	2.265845	7.969653e-04
Italy	12868066.0	155245.0	1.343115e+08	6.036747e+07	2.224899	2.571666e-03
Uruguay	846889.0	7005.0	7.622054e+06	3.485152e+06	2.187008	2.009955e-03
Belgium	3563841.0	30316.0	2.504046e+07	1.163233e+07	2.152660	2.606184e-03
Canada	3303284.0	36728.0	8.106477e+07	3.806791e+07	2.129478	9.648020e-04
Ireland	1315500.0	6540.0	1.056764e+07	4.982904e+06	2.120779	1.312488e-03
China	109423.0	4640.0	3.061013e+09	1.444216e+09	2.119498	3.212816e-06
France	23316777.0	138537.0	1.408496e+08	6.742200e+07	2.089075	2.054774e-03
Australia	3326214.0	5319.0	5.379783e+07	2.578822e+07	2.086140	2.062570e-04
New Zealand	166103.0	56.0	1.067335e+07	5.126300e+06	2.082077	1.092406e-05
Malaysia	3496090.0	32942.0	6.741625e+07	3.277620e+07	2.056866	1.005059e-03
Norway	1279169.0	1666.0	1.123317e+07	5.465629e+06	2.055238	3.048140e-04
Cambodia	131004.0	3032.0	3.467897e+07	1.694645e+07	2.046386	1.789166e-04
Germany	15175471.0	123431.0	1.701530e+08	8.390047e+07	2.028034	1.471160e-03
United Arab Emirates	880970.0	2301.0	2.018434e+07	9.991083e+06	2.020235	2.303054e-04
United Kingdom	18241147.0	160727.0	1.372498e+08	6.820711e+07	2.012251	2.356455e-03
Israel	3652454.0	10245.0	1.801561e+07	9.291000e+06	1.939039	1.102680e-03
Portugal	3282618.0	21111.0	1.964667e+07	1.016792e+07	1.932221	2.076235e-03
Upper middle income	118212049.0	2421471.0	4.804681e+09	2.513673e+09	1.911419	9.633199e-04

High income	236393139.0	2211971.0	2.306054e+09	1.214930e+09	1.898096	1.820657e-03
Luxembourg	186210.0	995.0	1.178840e+06	6.348140e+05	1.856985	1.567388e-03
Greece	2454429.0	25972.0	1.907120e+07	1.037075e+07	1.838941	2.504352e-03
Liechtenstein	12165.0	78.0	7.022800e+04	3.825400e+04	1.835834	2.039002e-03
Peru	3518721.0	211453.0	6.113136e+07	3.335942e+07	1.832507	6.338630e-03
European Union	110033392.0	1015772.0	8.184271e+08	4.471899e+08	1.830156	2.271456e-03
Brazil	28741413.0	650254.0	3.902730e+08	2.139934e+08	1.823761	3.038663e-03
Argentina	8912317.0	126390.0	8.251164e+07	4.560582e+07	1.809235	2.771357e-03
Switzerland	2842412.0	13269.0	1.556814e+07	8.715494e+06	1.786260	1.522461e-03
Ecuador	844764.0	22611.0	3.131789e+07	1.788847e+07	1.750730	1.263998e-03
Qatar	357583.0	670.0	5.106175e+06	2.930524e+06	1.742410	2.286281e-04
North America	93195128.0	1374621.0	1.030269e+09	5.965813e+08	1.726955	2.304164e-03
Oceania	3735739.0	7922.0	7.462766e+07	4.321995e+07	1.726695	1.832950e-04
Hong Kong	293754.0	1168.0	1.302273e+07	7.552800e+06	1.724226	1.546446e-04
South America	54222621.0	1247285.0	7.462801e+08	4.342601e+08	1.718509	2.872207e-03
Turkey	13382896.0	94837.0	1.434464e+08	8.504274e+07	1.686757	1.115169e-03
Bahrain	519584.0	1511.0	2.927653e+06	1.748295e+06	1.674576	8.642706e-04
Europe	158815550.0	1717056.0	1.247823e+09	7.489630e+08	1.666068	2.292578e-03
United States	79143715.0	954518.0	5.537498e+08	3.329151e+08	1.663336	2.867152e-03
Lithuania	913223.0	8442.0	4.449324e+06	2.689862e+06	1.654109	3.138451e-03
Saudi Arabia	746066.0	9002.0	5.737721e+07	3.534068e+07	1.623546	2.547206e-04
Czechia	3602844.0	38771.0	1.739288e+07	1.072455e+07	1.621781	3.615162e-03
Asia	117811418.0	1351250.0	7.544342e+09	4.678445e+09	1.612575	2.888246e-04
Thailand	2958176.0	23072.0	1.087844e+08	6.995084e+07	1.555156	3.298316e-04
Taiwan	20584.0	853.0	3.695129e+07	2.385501e+07	1.548995	3.575769e-05
Latvia	673218.0	5284.0	2.855540e+06	1.866934e+06	1.529535	2.830309e-03
Sri Lanka	647699.0	16267.0	3.284874e+07	2.149731e+07	1.528040	7.566995e-04
Estonia	504148.0	2268.0	1.955946e+06	1.325188e+06	1.475976	1.711455e-03
World	439011701.0	5946817.0	1.122516e+10	7.874966e+09	1.425424	7.551547e-04
Slovenia	897040.0	6327.0	2.956720e+06	2.078723e+06	1.422373	3.043696e-03
Maldives	171114.0	297.0	7.707100e+05	5.436200e+05	1.417737	5.463375e-04
Panama	756085.0	8098.0	5.934725e+06	4.381583e+06	1.354471	1.848190e-03
Spain	11139724.0	100778.0	6.310748e+07	4.674521e+07	1.350031	2.155900e-03
Mongolia	909379.0	2171.0	4.407760e+06	3.329282e+06	1.323937	6.520926e-04
Japan	5143816.0	24140.0	1.667069e+08	1.260508e+08	1.322538	1.915101e-04
Vietnam	3709481.0	40637.0	1.209025e+08	9.816883e+07	1.231578	4.139501e-04
India	42945160.0	507015.0	1.708969e+09	1.393409e+09	1.226466	3.638666e-04
Dominican Republic	575157.0	4370.0	1.331425e+07	1.095371e+07	1.215501	3.989514e-04
Lower middle income	82588725.0	1271503.0	3.974325e+09	3.330653e+09	1.193257	3.817579e-04
Curacao	39016.0	261.0	1.840890e+05	1.647960e+05	1.117072	1.583776e-03
Mexico	5521744.0	311470.0	1.451079e+08	1.302622e+08	1.113968	2.391100e-03
Bhutan	13535.0	6.0	8.620330e+05	7.799000e+05	1.105312	7.693294e-06
Isle of Man	23099.0	80.0	9.206400e+04	8.541000e+04	1.077907	9.366585e-04
Trinidad and Tobago	128691.0	3637.0	1.503347e+06	1.403374e+06	1.071238	2.591611e-03
Bolivia	893775.0	21443.0	1.258651e+07	1.183294e+07	1.063684	1.812145e-03
Aruba	33684.0	212.0	1.091530e+05	1.071950e+05	1.018266	1.977704e-03
Macao	81.0	0.0	6.627460e+05	6.583910e+05	1.006615	0.000000e+00
Azerbaijan	787367.0	9454.0	1.023182e+07	1.022334e+07	1.000829	9.247463e-04
Barbados	55543.0	316.0	2.868630e+05	2.877080e+05	0.997063	1.098336e-03
Poland	5694767.0	111864.0	3.742697e+07	3.779700e+07	0.990210	2.959600e-03
Indonesia	5630096.0	149036.0	2.724306e+08	2.763618e+08	0.985775	5.392786e-04
Kazakhstan	1450652.0	19239.0	1.870224e+07	1.899496e+07	0.984590	1.012848e-03
Russia	16353868.0	346197.0	1.434654e+08	1.459120e+08	0.983232	2.372642e-03
Colombia	6067023.0	138899.0	4.938413e+07	5.126584e+07	0.963295	2.709387e-03
Montenegro	230512.0	2683.0	5.963530e+05	6.280510e+05	0.949530	4.271946e-03

Croatia	1058453.0	15122.0	3.677004e+06	4.081657e+06	0.900861	3.704868e-03
Romania	2748777.0	63668.0	1.665058e+07	1.912777e+07	0.870492	3.328563e-03
Hungary	1793120.0	44134.0	8.276404e+06	9.634162e+06	0.859068	4.580990e-03
Greenland	11760.0	19.0	4.333600e+04	5.686800e+04	0.762045	3.341071e-04
El Salvador	156456.0	4077.0	4.886767e+06	6.518500e+06	0.749677	6.254506e-04
Kosovo	226392.0	3116.0	1.330147e+06	1.782115e+06	0.746387	1.748484e-03
Guatemala	784024.0	17021.0	1.341083e+07	1.824987e+07	0.734845	9.326643e-04
Morocco	1161290.0	16002.0	2.740605e+07	3.734479e+07	0.733866	4.284935e-04
Ukraine	5040518.0	112459.0	3.168315e+07	4.346682e+07	0.728904	2.587238e-03
Tunisia	999441.0	27824.0	8.530717e+06	1.193576e+07	0.714719	2.331145e-03
Faeroe Islands	34658.0	28.0	3.481300e+04	4.905300e+04	0.709702	5.708112e-04
Lebanon	1072537.0	10115.0	4.755512e+06	6.769151e+06	0.702527	1.494279e-03
Iceland	135748.0	66.0	2.588510e+05	3.687920e+05	0.701889	1.789627e-04
Bulgaria	1096194.0	35696.0	4.143716e+06	6.896655e+06	0.600830	5.175842e-03
Brunei	71667.0	129.0	2.627530e+05	4.415320e+05	0.595094	2.921646e-04
San Marino	14435.0	112.0	2.014100e+04	3.401000e+04	0.592208	3.293149e-03
Suriname	78353.0	1317.0	3.429360e+05	5.917980e+05	0.579482	2.225422e-03
Paraguay	642573.0	18422.0	4.105501e+06	7.219641e+06	0.568657	2.551650e-03
Bangladesh	1945108.0	29053.0	8.807564e+07	1.663035e+08	0.529608	1.746987e-04
Philippines	3663931.0	56506.0	5.494906e+07	1.110469e+08	0.494827	5.088480e-04
Albania	271825.0	3474.0	1.415150e+06	2.872934e+06	0.492580	1.209217e-03
Cyprus	326611.0	859.0	4.282040e+05	8.960050e+05	0.477904	9.587000e-04
Serbia	1916707.0	15318.0	3.257010e+06	6.871547e+06	0.473985	2.229192e-03
Zimbabwe	237509.0	5396.0	7.080893e+06	1.509217e+07	0.469177	3.575364e-04
North Macedonia	298195.0	9036.0	9.505510e+05	2.082661e+06	0.456412	4.338680e-03
Uzbekistan	236596.0	1637.0	1.512707e+07	3.393576e+07	0.445756	4.823819e-05
Seychelles	39408.0	163.0	4.286300e+04	9.891000e+04	0.433354	1.647963e-03
Africa	11230524.0	248668.0	5.818190e+08	1.373486e+09	0.423607	1.810487e-04
Jordan	1638338.0	13849.0	4.260392e+06	1.026902e+07	0.414878	1.348619e-03
Pakistan	1511754.0	30237.0	8.561347e+07	2.251999e+08	0.380167	1.342674e-04
Georgia	1616159.0	16231.0	1.413257e+06	3.979773e+06	0.355110	4.078373e-03
Moldova	502956.0	10657.0	1.305637e+06	4.024025e+06	0.324460	2.648343e-03
Nepal	977200.0	11941.0	9.492136e+06	2.967492e+07	0.319871	4.023937e-04
Palestine	648039.0	5532.0	1.655605e+06	5.222756e+06	0.316998	1.059211e-03
Kyrgyzstan	200556.0	3403.0	2.076513e+06	6.628347e+06	0.313278	5.134010e-04
Belize	56816.0	654.0	1.226920e+05	4.049150e+05	0.303007	1.615154e-03
Guernsey	0.0	0.0	1.593100e+04	6.338500e+04	0.251337	0.000000e+00
Laos	143240.0	623.0	1.802450e+06	7.379358e+06	0.244256	8.442469e-05
South Africa	3659100.0	99135.0	1.442119e+07	6.004200e+07	0.240185	1.651094e-03
Cayman Islands	19373.0	17.0	1.533900e+04	6.649800e+04	0.230669	2.556468e-04
Finland	658559.0	2381.0	1.235072e+06	5.548361e+06	0.222601	4.291357e-04
Low income	1805056.0	41804.0	1.391686e+08	6.651490e+08	0.209229	6.284907e-05
Kenya	323002.0	5640.0	1.141772e+07	5.498570e+07	0.207649	1.025721e-04
Iran	7066975.0	137267.0	1.673545e+07	8.502876e+07	0.196821	1.614360e-03
Ethiopia	468786.0	7467.0	2.131937e+07	1.178762e+08	0.180862	6.334611e-05
Jamaica	128079.0	2814.0	4.916840e+05	2.973462e+06	0.165357	9.463716e-04
Saint Lucia	22729.0	360.0	2.625500e+04	1.844010e+05	0.142380	1.952267e-03
Saint Vincent and the Grenadines	10045.0	106.0	1.569000e+04	1.112690e+05	0.141010	9.526463e-04
Saint Kitts and Nevis	5530.0	42.0	6.553000e+03	5.354600e+04	0.122381	7.843723e-04
Guinea	36397.0	440.0	1.586407e+06	1.349724e+07	0.117536	3.259926e-05
Guyana	62986.0	1222.0	8.656300e+04	7.903290e+05	0.109528	1.546192e-03
Eswatini	69211.0	1390.0	1.107650e+05	1.172369e+06	0.094480	1.185634e-03
Anguilla	2555.0	10.0	1.421000e+03	1.512500e+04	0.093950	6.611570e-04
Namibia	157275.0	4010.0	2.347090e+05	2.587344e+06	0.090714	1.549852e-03
Dominica	11142.0	57.0	6.417000e+03	7.217200e+04	0.088913	7.897800e-04

Cote d'Ivoire	81552.0	794.0	2.285727e+06	2.705363e+07	0.084489	2.934911e-05
Antigua and Barbuda	7451.0	135.0	8.107000e+03	9.872800e+04	0.082114	1.367393e-03
Zambia	313203.0	3955.0	1.321156e+06	1.892066e+07	0.069826	2.090308e-04
Rwanda	129533.0	1457.0	9.206440e+05	1.327652e+07	0.069344	1.097426e-04
Malawi	85362.0	2617.0	1.267776e+06	1.964768e+07	0.064525	1.331964e-04
Andorra	38249.0	151.0	4.802000e+03	7.735400e+04	0.062078	1.952065e-03
Fiji	63999.0	836.0	2.829500e+04	9.028990e+05	0.031338	9.259064e-04
Bosnia and Herzegovina	371553.0	15506.0	9.386200e+04	3.263459e+06	0.028762	4.751400e-03
Mauritius	71004.0	786.0	3.515300e+04	1.273428e+06	0.027605	6.172316e-04
Bahamas	33198.0	771.0	9.258000e+03	3.969140e+05	0.023325	1.942486e-03
Honduras	412753.0	10779.0	2.290260e+05	1.006299e+07	0.022759	1.071152e-03
Libya	495972.0	6279.0	1.416880e+05	6.958538e+06	0.020362	9.023447e-04
Senegal	85712.0	1960.0	3.485540e+05	1.719631e+07	0.020269	1.139780e-04
Egypt	486381.0	24132.0	2.078777e+06	1.042583e+08	0.019939	2.314635e-04
Nauru	0.0	0.0	1.680000e+02	1.087300e+04	0.015451	0.000000e+00
Uganda	163447.0	3594.0	7.200770e+05	4.712353e+07	0.015281	7.626763e-05
Nigeria	254598.0	3143.0	3.160328e+06	2.114007e+08	0.014949	1.486750e-05
Cape Verde	55889.0	402.0	7.462000e+03	5.619010e+05	0.013280	7.154285e-04
Tajikistan	17786.0	125.0	1.247450e+05	9.749625e+06	0.012795	1.282101e-05
Oman	383389.0	4246.0	6.508100e+04	5.223376e+06	0.012460	8.128842e-04
Mozambique	225096.0	2227.0	2.958120e+05	3.216304e+07	0.009197	6.924096e-05
Grenada	13707.0	216.0	1.037000e+03	1.130150e+05	0.009176	1.911251e-03
Kuwait	620980.0	2540.0	3.092700e+04	4.328553e+06	0.007145	5.868012e-04
Timor	22732.0	129.0	7.881000e+03	1.343875e+06	0.005864	9.599107e-05
Ghana	159891.0	1442.0	1.809500e+05	3.173213e+07	0.005702	4.544290e-05
Mauritania	58638.0	979.0	2.465900e+04	4.775110e+06	0.005164	2.050215e-04
Sweden	2451464.0	17611.0	4.761600e+04	1.016016e+07	0.004687	1.733339e-03
Botswana	263950.0	2619.0	1.043500e+04	2.397240e+06	0.004353	1.092506e-03
Algeria	265079.0	6843.0	1.707860e+05	4.461663e+07	0.003828	1.533733e-04
Haiti	30350.0	897.0	4.145300e+04	1.154168e+07	0.003592	7.771830e-05
French Polynesia	67660.0	641.0	8.550000e+02	2.825340e+05	0.003026	2.268753e-03
Central African Republic	14225.0	113.0	1.288700e+04	4.919987e+06	0.002619	2.296754e-05
Somalia	26351.0	1349.0	3.729200e+04	1.635950e+07	0.002280	8.245973e-05
Liberia	7510.0	294.0	1.008900e+04	5.180208e+06	0.001948	5.675448e-05
Sudan	61525.0	3910.0	4.557200e+04	4.490935e+07	0.001015	8.706427e-05
Sierra Leone	7665.0	125.0	7.981000e+03	8.141343e+06	0.000980	1.535373e-05
Guinea-Bissau	8027.0	167.0	1.658000e+03	2.015490e+06	0.000823	8.285826e-05
South Sudan	16989.0	137.0	7.489000e+03	1.138138e+07	0.000658	1.203721e-05
Gabon	47543.0	303.0	1.486000e+03	2.278829e+06	0.000652	1.329630e-04
Madagascar	63666.0	1366.0	1.354400e+04	2.842733e+07	0.000476	4.805234e-05
Democratic Republic of Congo	86039.0	1335.0	4.230300e+04	9.237799e+07	0.000458	1.445149e-05
Equatorial Guinea	15885.0	183.0	6.390000e+02	1.449891e+06	0.000441	1.262164e-04
Cameroon	119240.0	1923.0	1.029000e+04	2.722426e+07	0.000378	7.063552e-05
Afghanistan	174081.0	7617.0	1.374200e+04	3.983543e+07	0.000345	1.912117e-04
Congo	24020.0	409.0	1.486000e+03	5.657017e+06	0.000263	7.229959e-05
Iraq	2305083.0	25013.0	9.985000e+03	4.117935e+07	0.000242	6.074161e-04
Gambia	12039.0	365.0	4.290000e+02	2.486937e+06	0.000173	1.467669e-04
Myanmar	592139.0	19376.0	3.800000e+03	5.480601e+07	0.000069	3.535378e-04
Burundi	38127.0	38.0	4.600000e+02	1.225543e+07	0.000038	3.100667e-06
Northern Cyprus	0.0	0.0	1.988000e+03	1.474690e+08	0.000013	0.000000e+00
Chad	7257.0	190.0	0.000000e+00	1.691498e+07	0.000000	1.123264e-05
Angola	98746.0	1903.0	0.000000e+00	3.393361e+07	0.000000	5.608009e-05
Tonga	355.0	0.0	0.000000e+00	1.067590e+05	0.000000	0.000000e+00
Yemen	11772.0	2135.0	0.000000e+00	3.049064e+07	0.000000	7.002149e-05

Armenia	420498.0	8495.0	0.000000e+00	2.968128e+06	0.000000	2.862073e-03
Tokelau	0.0	0.0	0.000000e+00	1.368000e+03	0.000000	0.000000e+00
Togo	36808.0	272.0	0.000000e+00	8.478242e+06	0.000000	3.208212e-05
Cook Islands	2.0	0.0	0.000000e+00	1.757200e+04	0.000000	0.000000e+00
Comoros	8033.0	161.0	0.000000e+00	8.884560e+05	0.000000	1.812133e-04
Austria	2744023.0	14902.0	0.000000e+00	9.043072e+06	0.000000	1.647891e-03
Turkmenistan	0.0	0.0	0.000000e+00	6.117933e+06	0.000000	0.000000e+00
Venezuela	515582.0	5645.0	0.000000e+00	2.870495e+07	0.000000	1.966560e-04
Tuvalu	0.0	0.0	0.000000e+00	1.192500e+04	0.000000	0.000000e+00
Costa Rica	811040.0	8057.0	0.000000e+00	5.139053e+06	0.000000	1.567799e-03
Wallis and Futuna	454.0	7.0	0.000000e+00	1.109400e+04	0.000000	6.309717e-04
Burkina Faso	20751.0	375.0	0.000000e+00	2.149710e+07	0.000000	1.744422e-05
British Virgin Islands	6085.0	62.0	0.000000e+00	3.042300e+04	0.000000	2.037932e-03
Bonaire Sint Eustatius and Saba	7599.0	33.0	0.000000e+00	2.644500e+04	0.000000	1.247873e-03
Bermuda	11561.0	123.0	0.000000e+00	6.209200e+04	0.000000	1.980932e-03
Benin	26776.0	163.0	0.000000e+00	1.245103e+07	0.000000	1.309129e-05
Belarus	923432.0	6506.0	0.000000e+00	9.442867e+06	0.000000	6.889857e-04
Vanuatu	19.0	1.0	0.000000e+00	3.144640e+05	0.000000	3.180014e-06
Vatican	29.0	0.0	0.000000e+00	8.120000e+02	0.000000	0.000000e+00
Turks and Caicos Islands	5868.0	36.0	0.000000e+00	3.922600e+04	0.000000	9.177586e-04
Papua New Guinea	41351.0	638.0	0.000000e+00	9.119005e+06	0.000000	6.996377e-05
Tanzania	33620.0	798.0	0.000000e+00	6.149844e+07	0.000000	1.297594e-05
Djibouti	15547.0	189.0	0.000000e+00	1.002197e+06	0.000000	1.885857e-04
Palau	3823.0	6.0	0.000000e+00	1.817400e+04	0.000000	3.301420e-04
Niue	0.0	0.0	0.000000e+00	1.614000e+03	0.000000	0.000000e+00
Niger	8763.0	307.0	0.000000e+00	2.513081e+07	0.000000	1.221608e-05
Nicaragua	18105.0	225.0	0.000000e+00	6.702379e+06	0.000000	3.357017e-05
New Caledonia	55503.0	301.0	0.000000e+00	2.882170e+05	0.000000	1.044352e-03
Saint Helena	4.0	0.0	0.000000e+00	6.095000e+03	0.000000	0.000000e+00
Netherlands	6540294.0	21710.0	0.000000e+00	1.717309e+07	0.000000	1.264187e-03
Saint Pierre and Miquelon	1089.0	1.0	0.000000e+00	5.771000e+03	0.000000	1.732802e-04
Montserrat	164.0	2.0	0.000000e+00	4.981000e+03	0.000000	4.015258e-04
Samoa	33.0	0.0	0.000000e+00	2.001440e+05	0.000000	0.000000e+00
Monaco	9457.0	54.0	0.000000e+00	3.952000e+04	0.000000	1.366397e-03
Sao Tome and Principe	5934.0	76.0	0.000000e+00	2.233640e+05	0.000000	3.402518e-04
Micronesia (country)	1.0	0.0	0.000000e+00	1.162550e+05	0.000000	0.000000e+00
Marshall Islands	8.0	0.0	0.000000e+00	5.961800e+04	0.000000	0.000000e+00
Mali	30391.0	722.0	0.000000e+00	2.085572e+07	0.000000	3.461879e-05
Lesotho	32612.0	696.0	0.000000e+00	2.159067e+06	0.000000	3.223615e-04
Sint Maarten (Dutch part)	0.0	0.0	0.000000e+00	4.342100e+04	0.000000	0.000000e+00
Slovakia	2150666.0	18567.0	0.000000e+00	5.449270e+06	0.000000	3.407245e-03
Kiribati	2953.0	11.0	0.000000e+00	1.213880e+05	0.000000	9.061851e-05
Solomon Islands	7261.0	106.0	0.000000e+00	7.039950e+05	0.000000	1.505693e-04
Jersey	0.0	0.0	0.000000e+00	1.010730e+05	0.000000	0.000000e+00
International	721.0	15.0	0.000000e+00	1.474690e+08	0.000000	1.017163e-07
Pitcairn	0.0	0.0	0.000000e+00	4.700000e+01	0.000000	0.000000e+00
Eritrea	9705.0	103.0	0.000000e+00	3.601462e+06	0.000000	2.859950e-05
Syria	54744.0	3082.0	0.000000e+00	1.827570e+07	0.000000	1.686392e-04
Falkland Islands	115.0	0.0	0.000000e+00	3.528000e+03	0.000000	0.000000e+00

According to this dataframe, we'll choose Cuba, Sweden, Cambodia, India and Albania for visualization:

In [57]:

```
countries = ['Cuba', 'Sweden', 'Cambodia', 'India', 'Albania']
```

```
selected = summarized_copy.loc[countries]
selected
```

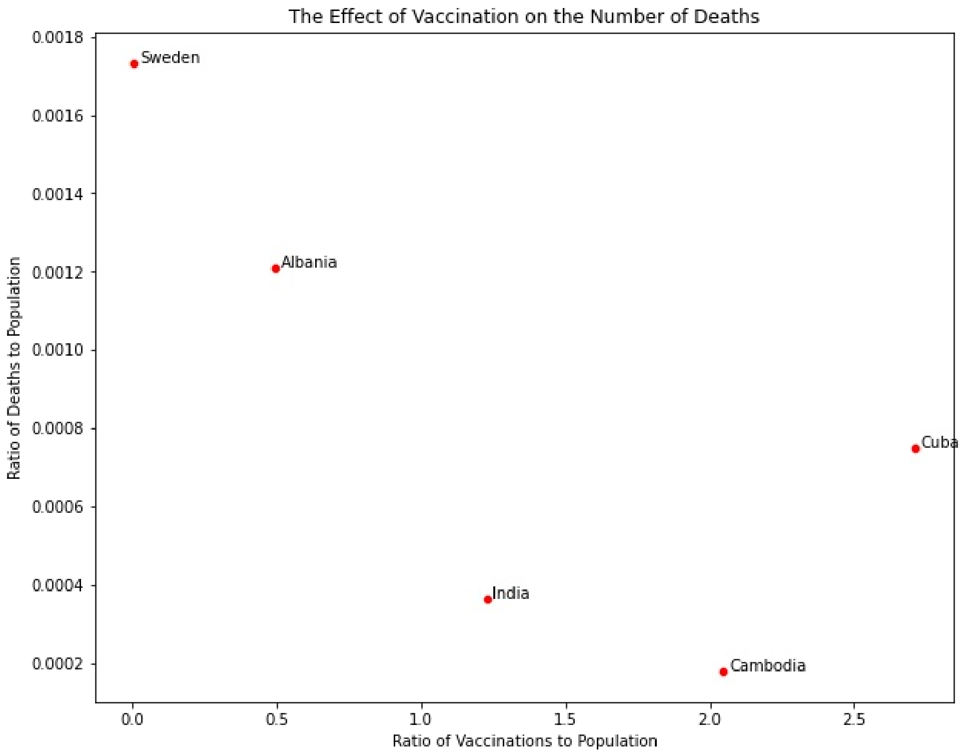
Out[57]:

location	new_cases	new_deaths	new_vaccinations	population	first_ratio	second_ratio
Cuba	1070730.0	8497.0	3.065243e+07	1.131750e+07	2.708410	0.000751
Sweden	2451464.0	17611.0	4.761600e+04	1.016016e+07	0.004687	0.001733
Cambodia	131004.0	3032.0	3.467897e+07	1.694645e+07	2.046386	0.000179
India	42945160.0	507015.0	1.708969e+09	1.393409e+09	1.226466	0.000364
Albania	271825.0	3474.0	1.415150e+06	2.872934e+06	0.492580	0.001209

For visualization, we'll use a scatter plot that shows the effect of vaccination on the number of deaths. The X-axis shows the ratio of total vaccination to population and Y-axis shows the ratio of total deaths to population. Let's see the result:

In [58]:

```
plt.figure(figsize=(10,8))
ax = sns.scatterplot(x=selected['first_ratio'], y=selected['second_ratio'], color = 'r')
for i, point in selected.iterrows():
    ax.text(point['first_ratio']+.02, point['second_ratio'], str(i))
ax.set_title("The Effect of Vaccination on the Number of Deaths")
ax.set_xlabel('Ratio of Vaccinations to Population')
ax.set_ylabel('Ratio of Deaths to Population')
plt.show()
```



As you can see, more vaccination can reduce the number of the deaths in a country in general. However, there are other factors that affect the number of deaths too. For example, Cuba has better ratio in vaccination, but worse result in ratio of controlling death.

Q3

Now we want to see speed of vaccination in different countries. We'll use total vaccinations in different months in 2021. To pick 5 countries for examination, we'll again use the previous dataframe and choose countries based on their vaccination ratio and continent.

In [59]:

```
summarized_copy
```

Out[59]:

location	new_cases	new_deaths	new_vaccinations	population	first_ratio	second_ratio
Cuba	1070730.0	8497.0	3.065243e+07	1.131750e+07	2.708410	7.507843e-04
Gibraltar	15632.0	101.0	8.906400e+04	3.369100e+04	2.643555	2.997833e-03

Chile	3098110.0	42624.0	4.855777e+07	1.921236e+07	2.527423	2.218572e-03
Singapore	767663.0	1040.0	1.357254e+07	5.453600e+06	2.488730	1.906997e-04
Malta	71497.0	608.0	1.205547e+06	5.161000e+05	2.335879	1.178066e-03
South Korea	3691487.0	8394.0	1.192514e+08	5.130518e+07	2.324354	1.636092e-04
Denmark	2805858.0	4633.0	1.317204e+07	5.813302e+06	2.265845	7.969653e-04
Italy	12868066.0	155245.0	1.343115e+08	6.036747e+07	2.224899	2.571666e-03
Uruguay	846889.0	7005.0	7.622054e+06	3.485152e+06	2.187008	2.009955e-03
Belgium	3563841.0	30316.0	2.504046e+07	1.163233e+07	2.152660	2.606184e-03
Canada	3303284.0	36728.0	8.106477e+07	3.806791e+07	2.129478	9.648020e-04
Ireland	1315500.0	6540.0	1.056764e+07	4.982904e+06	2.120779	1.312488e-03
China	109423.0	4640.0	3.061013e+09	1.444216e+09	2.119498	3.212816e-06
France	23316777.0	138537.0	1.408496e+08	6.742200e+07	2.089075	2.054774e-03
Australia	3326214.0	5319.0	5.379783e+07	2.578822e+07	2.086140	2.062570e-04
New Zealand	166103.0	56.0	1.067335e+07	5.126300e+06	2.082077	1.092406e-05
Malaysia	3496090.0	32942.0	6.741625e+07	3.277620e+07	2.056866	1.005059e-03
Norway	1279169.0	1666.0	1.123317e+07	5.465629e+06	2.055238	3.048140e-04
Cambodia	131004.0	3032.0	3.467897e+07	1.694645e+07	2.046386	1.789166e-04
Germany	15175471.0	123431.0	1.701530e+08	8.390047e+07	2.028034	1.471160e-03
United Arab Emirates	880970.0	2301.0	2.018434e+07	9.991083e+06	2.020235	2.303054e-04
United Kingdom	18241147.0	160727.0	1.372498e+08	6.820711e+07	2.012251	2.356455e-03
Israel	3652454.0	10245.0	1.801561e+07	9.291000e+06	1.939039	1.102680e-03
Portugal	3282618.0	21111.0	1.964667e+07	1.016792e+07	1.932221	2.076235e-03
Upper middle income	118212049.0	2421471.0	4.804681e+09	2.513673e+09	1.911419	9.633199e-04
High income	236393139.0	2211971.0	2.306054e+09	1.214930e+09	1.898096	1.820657e-03
Luxembourg	186210.0	995.0	1.178840e+06	6.348140e+05	1.856985	1.567388e-03
Greece	2454429.0	25972.0	1.907120e+07	1.037075e+07	1.838941	2.504352e-03
Liechtenstein	12165.0	78.0	7.022800e+04	3.825400e+04	1.835834	2.039002e-03
Peru	3518721.0	211453.0	6.113136e+07	3.335942e+07	1.832507	6.338630e-03
European Union	110033392.0	1015772.0	8.184271e+08	4.471899e+08	1.830156	2.271456e-03
Brazil	28741413.0	650254.0	3.902730e+08	2.139934e+08	1.823761	3.038663e-03
Argentina	8912317.0	126390.0	8.251164e+07	4.560582e+07	1.809235	2.771357e-03
Switzerland	2842412.0	13269.0	1.556814e+07	8.715494e+06	1.786260	1.522461e-03
Ecuador	844764.0	22611.0	3.131789e+07	1.788847e+07	1.750730	1.263998e-03
Qatar	357583.0	670.0	5.106175e+06	2.930524e+06	1.742410	2.286281e-04
North America	93195128.0	1374621.0	1.030269e+09	5.965813e+08	1.726955	2.304164e-03
Oceania	3735739.0	7922.0	7.462766e+07	4.321995e+07	1.726695	1.832950e-04
Hong Kong	293754.0	1168.0	1.302273e+07	7.552800e+06	1.724226	1.546446e-04
South America	54222621.0	1247285.0	7.462801e+08	4.342601e+08	1.718509	2.872207e-03
Turkey	13382896.0	94837.0	1.434464e+08	8.504274e+07	1.686757	1.115169e-03
Bahrain	519584.0	1511.0	2.927653e+06	1.748295e+06	1.674576	8.642706e-04
Europe	158815550.0	1717056.0	1.247823e+09	7.489630e+08	1.666068	2.292578e-03
United States	79143715.0	954518.0	5.537498e+08	3.329151e+08	1.663336	2.867152e-03
Lithuania	913223.0	8442.0	4.449324e+06	2.689862e+06	1.654109	3.138451e-03
Saudi Arabia	746066.0	9002.0	5.737721e+07	3.534068e+07	1.623546	2.547206e-04
Czechia	3602844.0	38771.0	1.739288e+07	1.072455e+07	1.621781	3.615162e-03
Asia	117811418.0	1351250.0	7.544342e+09	4.678445e+09	1.612575	2.888246e-04
Thailand	2958176.0	23072.0	1.087844e+08	6.995084e+07	1.555156	3.298316e-04
Taiwan	20584.0	853.0	3.695129e+07	2.385501e+07	1.548995	3.575769e-05
Latvia	673218.0	5284.0	2.855540e+06	1.866934e+06	1.529535	2.830309e-03
Sri Lanka	647699.0	16267.0	3.284874e+07	2.149731e+07	1.528040	7.566995e-04
Estonia	504148.0	2268.0	1.955946e+06	1.325188e+06	1.475976	1.711455e-03
World	439011701.0	5946817.0	1.122516e+10	7.874966e+09	1.425424	7.551547e-04
Slovenia	897040.0	6327.0	2.956720e+06	2.078723e+06	1.422373	3.043696e-03
Maldives	171114.0	297.0	7.707100e+05	5.436200e+05	1.417737	5.463375e-04
Panama	756085.0	8098.0	5.934725e+06	4.381583e+06	1.354471	1.848190e-03

Spain	11139724.0	100778.0	6.310748e+07	4.674521e+07	1.350031	2.155900e-03
Mongolia	909379.0	2171.0	4.407760e+06	3.329282e+06	1.323937	6.520926e-04
Japan	5143816.0	24140.0	1.667069e+08	1.260508e+08	1.322538	1.915101e-04
Vietnam	3709481.0	40637.0	1.209025e+08	9.816883e+07	1.231578	4.139501e-04
India	42945160.0	507015.0	1.708969e+09	1.393409e+09	1.226466	3.638666e-04
Dominican Republic	575157.0	4370.0	1.331425e+07	1.095371e+07	1.215501	3.989514e-04
Lower middle income	82588725.0	1271503.0	3.974325e+09	3.330653e+09	1.193257	3.817579e-04
Curacao	39016.0	261.0	1.840890e+05	1.647960e+05	1.117072	1.583776e-03
Mexico	5521744.0	311470.0	1.451079e+08	1.302622e+08	1.113968	2.391100e-03
Bhutan	13535.0	6.0	8.620330e+05	7.799000e+05	1.105312	7.693294e-06
Isle of Man	23099.0	80.0	9.206400e+04	8.541000e+04	1.077907	9.366585e-04
Trinidad and Tobago	128691.0	3637.0	1.503347e+06	1.403374e+06	1.071238	2.591611e-03
Bolivia	893775.0	21443.0	1.258651e+07	1.183294e+07	1.063684	1.812145e-03
Aruba	33684.0	212.0	1.091530e+05	1.071950e+05	1.018266	1.977704e-03
Macao	81.0	0.0	6.627460e+05	6.583910e+05	1.006615	0.000000e+00
Azerbaijan	787367.0	9454.0	1.023182e+07	1.022334e+07	1.000829	9.247463e-04
Barbados	55543.0	316.0	2.868630e+05	2.877080e+05	0.997063	1.098336e-03
Poland	5694767.0	111864.0	3.742697e+07	3.779700e+07	0.990210	2.959600e-03
Indonesia	5630096.0	149036.0	2.724306e+08	2.763618e+08	0.985775	5.392786e-04
Kazakhstan	1450652.0	19239.0	1.870224e+07	1.899496e+07	0.984590	1.012848e-03
Russia	16353868.0	346197.0	1.434654e+08	1.459120e+08	0.983232	2.372642e-03
Colombia	6067023.0	138899.0	4.938413e+07	5.126584e+07	0.963295	2.709387e-03
Montenegro	230512.0	2683.0	5.963530e+05	6.280510e+05	0.949530	4.271946e-03
Croatia	1058453.0	15122.0	3.677004e+06	4.081657e+06	0.900861	3.704868e-03
Romania	2748777.0	63668.0	1.665058e+07	1.912777e+07	0.870492	3.328563e-03
Hungary	1793120.0	44134.0	8.276404e+06	9.634162e+06	0.859068	4.580990e-03
Greenland	11760.0	19.0	4.333600e+04	5.686800e+04	0.762045	3.341071e-04
El Salvador	156456.0	4077.0	4.886767e+06	6.518500e+06	0.749677	6.254506e-04
Kosovo	226392.0	3116.0	1.330147e+06	1.782115e+06	0.746387	1.748484e-03
Guatemala	784024.0	17021.0	1.341083e+07	1.824987e+07	0.734845	9.326643e-04
Morocco	1161290.0	16002.0	2.740605e+07	3.734479e+07	0.733866	4.284935e-04
Ukraine	5040518.0	112459.0	3.168315e+07	4.346682e+07	0.728904	2.587238e-03
Tunisia	999441.0	27824.0	8.530717e+06	1.193576e+07	0.714719	2.331145e-03
Faeroe Islands	34658.0	28.0	3.481300e+04	4.905300e+04	0.709702	5.708112e-04
Lebanon	1072537.0	10115.0	4.755512e+06	6.769151e+06	0.702527	1.494279e-03
Iceland	135748.0	66.0	2.588510e+05	3.687920e+05	0.701889	1.789627e-04
Bulgaria	1096194.0	35696.0	4.143716e+06	6.896655e+06	0.600830	5.175842e-03
Brunei	71667.0	129.0	2.627530e+05	4.415320e+05	0.595094	2.921646e-04
San Marino	14435.0	112.0	2.014100e+04	3.401000e+04	0.592208	3.293149e-03
Suriname	78353.0	1317.0	3.429360e+05	5.917980e+05	0.579482	2.225422e-03
Paraguay	642573.0	18422.0	4.105501e+06	7.219641e+06	0.568657	2.551650e-03
Bangladesh	1945108.0	29053.0	8.807564e+07	1.663035e+08	0.529608	1.746987e-04
Philippines	3663931.0	56506.0	5.494906e+07	1.110469e+08	0.494827	5.088480e-04
Albania	271825.0	3474.0	1.415150e+06	2.872934e+06	0.492580	1.209217e-03
Cyprus	326611.0	859.0	4.282040e+05	8.960050e+05	0.477904	9.587000e-04
Serbia	1916707.0	15318.0	3.257010e+06	6.871547e+06	0.473985	2.229192e-03
Zimbabwe	237509.0	5396.0	7.080893e+06	1.509217e+07	0.469177	3.575364e-04
North Macedonia	298195.0	9036.0	9.505510e+05	2.082661e+06	0.456412	4.338680e-03
Uzbekistan	236596.0	1637.0	1.512707e+07	3.393576e+07	0.445756	4.823819e-05
Seychelles	39408.0	163.0	4.286300e+04	9.891000e+04	0.433354	1.647963e-03
Africa	11230524.0	248668.0	5.818190e+08	1.373486e+09	0.423607	1.810487e-04
Jordan	1638338.0	13849.0	4.260392e+06	1.026902e+07	0.414878	1.348619e-03
Pakistan	1511754.0	30237.0	8.561347e+07	2.251999e+08	0.380167	1.342674e-04
Georgia	1616159.0	16231.0	1.413257e+06	3.979773e+06	0.355110	4.078373e-03
Moldova	502956.0	10657.0	1.305637e+06	4.024025e+06	0.324460	2.648343e-03

Nepal	977200.0	11941.0	9.492136e+06	2.967492e+07	0.319871	4.023937e-04
Palestine	648039.0	5532.0	1.655605e+06	5.222756e+06	0.316998	1.059211e-03
Kyrgyzstan	200556.0	3403.0	2.076513e+06	6.628347e+06	0.313278	5.134010e-04
Belize	56816.0	654.0	1.226920e+05	4.049150e+05	0.303007	1.615154e-03
Guernsey	0.0	0.0	1.593100e+04	6.338500e+04	0.251337	0.000000e+00
Laos	143240.0	623.0	1.802450e+06	7.379358e+06	0.244256	8.442469e-05
South Africa	3659100.0	99135.0	1.442119e+07	6.004200e+07	0.240185	1.651094e-03
Cayman Islands	19373.0	17.0	1.533900e+04	6.649800e+04	0.230669	2.556468e-04
Finland	658559.0	2381.0	1.235072e+06	5.548361e+06	0.222601	4.291357e-04
Low income	1805056.0	41804.0	1.391686e+08	6.651490e+08	0.209229	6.284907e-05
Kenya	323002.0	5640.0	1.141772e+07	5.498570e+07	0.207649	1.025721e-04
Iran	7066975.0	137267.0	1.673545e+07	8.502876e+07	0.196821	1.614360e-03
Ethiopia	468786.0	7467.0	2.131937e+07	1.178762e+08	0.180862	6.334611e-05
Jamaica	128079.0	2814.0	4.916840e+05	2.973462e+06	0.165357	9.463716e-04
Saint Lucia	22729.0	360.0	2.625500e+04	1.844010e+05	0.142380	1.952267e-03
Saint Vincent and the Grenadines	10045.0	106.0	1.569000e+04	1.112690e+05	0.141010	9.526463e-04
Saint Kitts and Nevis	5530.0	42.0	6.553000e+03	5.354600e+04	0.122381	7.843723e-04
Guinea	36397.0	440.0	1.586407e+06	1.349724e+07	0.117536	3.259926e-05
Guyana	62986.0	1222.0	8.656300e+04	7.903290e+05	0.109528	1.546192e-03
Eswatini	69211.0	1390.0	1.107650e+05	1.172369e+06	0.094480	1.185634e-03
Anguilla	2555.0	10.0	1.421000e+03	1.512500e+04	0.093950	6.611570e-04
Namibia	157275.0	4010.0	2.347090e+05	2.587344e+06	0.090714	1.549852e-03
Dominica	11142.0	57.0	6.417000e+03	7.217200e+04	0.088913	7.897800e-04
Cote d'Ivoire	81552.0	794.0	2.285727e+06	2.705363e+07	0.084489	2.934911e-05
Antigua and Barbuda	7451.0	135.0	8.107000e+03	9.872800e+04	0.082114	1.367393e-03
Zambia	313203.0	3955.0	1.321156e+06	1.892066e+07	0.069826	2.090308e-04
Rwanda	129533.0	1457.0	9.206440e+05	1.327652e+07	0.069344	1.097426e-04
Malawi	85362.0	2617.0	1.267776e+06	1.964768e+07	0.064525	1.331964e-04
Andorra	38249.0	151.0	4.802000e+03	7.735400e+04	0.062078	1.952065e-03
Fiji	63999.0	836.0	2.829500e+04	9.028990e+05	0.031338	9.259064e-04
Bosnia and Herzegovina	371553.0	15506.0	9.386200e+04	3.263459e+06	0.028762	4.751400e-03
Mauritius	71004.0	786.0	3.515300e+04	1.273428e+06	0.027605	6.172316e-04
Bahamas	33198.0	771.0	9.258000e+03	3.969140e+05	0.023325	1.942486e-03
Honduras	412753.0	10779.0	2.290260e+05	1.006299e+07	0.022759	1.071152e-03
Libya	495972.0	6279.0	1.416880e+05	6.958538e+06	0.020362	9.023447e-04
Senegal	85712.0	1960.0	3.485540e+05	1.719631e+07	0.020269	1.139780e-04
Egypt	486381.0	24132.0	2.078777e+06	1.042583e+08	0.019939	2.314635e-04
Nauru	0.0	0.0	1.680000e+02	1.087300e+04	0.015451	0.000000e+00
Uganda	163447.0	3594.0	7.200770e+05	4.712353e+07	0.015281	7.626763e-05
Nigeria	254598.0	3143.0	3.160328e+06	2.114007e+08	0.014949	1.486750e-05
Cape Verde	55889.0	402.0	7.462000e+03	5.619010e+05	0.013280	7.154285e-04
Tajikistan	17786.0	125.0	1.247450e+05	9.749625e+06	0.012795	1.282101e-05
Oman	383389.0	4246.0	6.508100e+04	5.223376e+06	0.012460	8.128842e-04
Mozambique	225096.0	2227.0	2.958120e+05	3.216304e+07	0.009197	6.924096e-05
Grenada	13707.0	216.0	1.037000e+03	1.130150e+05	0.009176	1.911251e-03
Kuwait	620980.0	2540.0	3.092700e+04	4.328553e+06	0.007145	5.868012e-04
Timor	22732.0	129.0	7.881000e+03	1.343875e+06	0.005864	9.599107e-05
Ghana	159891.0	1442.0	1.809500e+05	3.173213e+07	0.005702	4.544290e-05
Mauritania	58638.0	979.0	2.465900e+04	4.775110e+06	0.005164	2.050215e-04
Sweden	2451464.0	17611.0	4.761600e+04	1.016016e+07	0.004687	1.733339e-03
Botswana	263950.0	2619.0	1.043500e+04	2.397240e+06	0.004353	1.092506e-03
Algeria	265079.0	6843.0	1.707860e+05	4.461663e+07	0.003828	1.533733e-04
Haiti	30350.0	897.0	4.145300e+04	1.154168e+07	0.003592	7.771830e-05
French Polynesia	67660.0	641.0	8.550000e+02	2.825340e+05	0.003026	2.268753e-03
Central African Republic	14225.0	113.0	1.288700e+04	4.919987e+06	0.002619	2.296754e-05

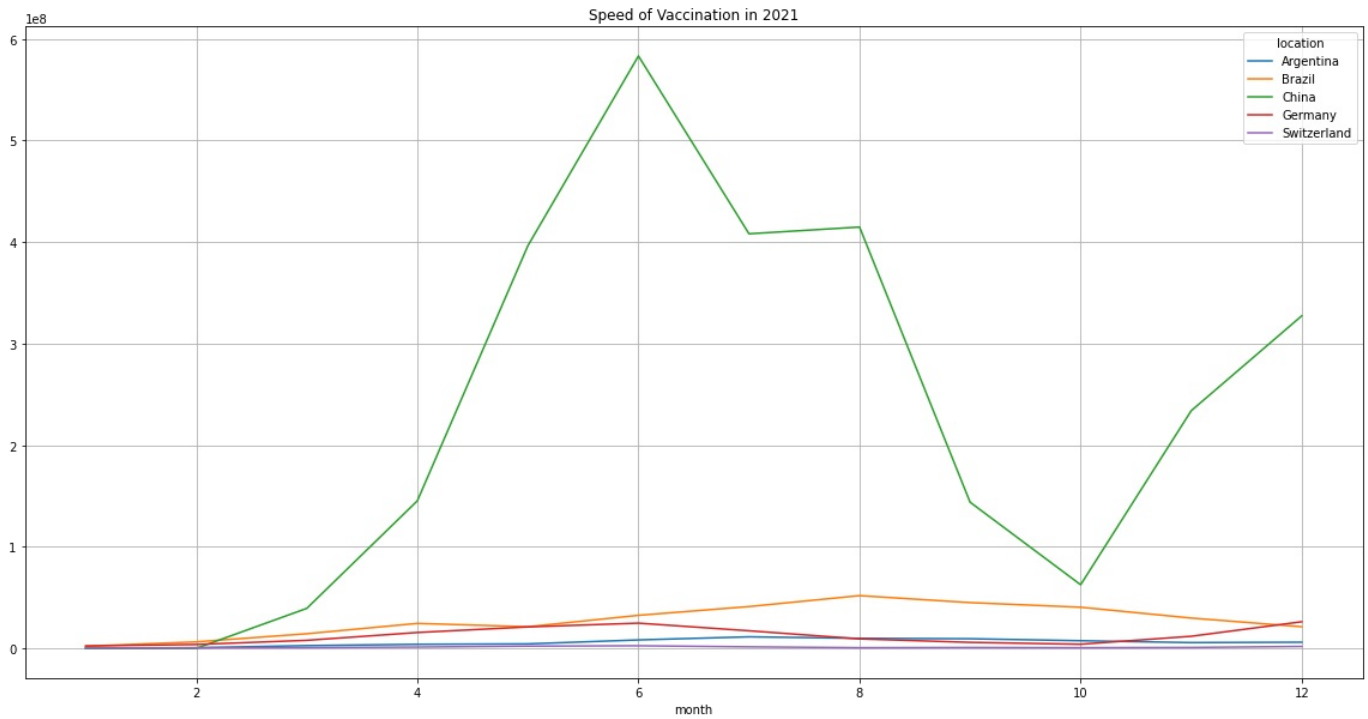
Somalia	26351.0	1349.0	3.729200e+04	1.635950e+07	0.002280	8.245973e-05
Liberia	7510.0	294.0	1.008900e+04	5.180208e+06	0.001948	5.675448e-05
Sudan	61525.0	3910.0	4.557200e+04	4.490935e+07	0.001015	8.706427e-05
Sierra Leone	7665.0	125.0	7.981000e+03	8.141343e+06	0.000980	1.535373e-05
Guinea-Bissau	8027.0	167.0	1.658000e+03	2.015490e+06	0.000823	8.285826e-05
South Sudan	16989.0	137.0	7.489000e+03	1.138138e+07	0.000658	1.203721e-05
Gabon	47543.0	303.0	1.486000e+03	2.278829e+06	0.000652	1.329630e-04
Madagascar	63666.0	1366.0	1.354400e+04	2.842733e+07	0.000476	4.805234e-05
Democratic Republic of Congo	86039.0	1335.0	4.230300e+04	9.237799e+07	0.000458	1.445149e-05
Equatorial Guinea	15885.0	183.0	6.390000e+02	1.449891e+06	0.000441	1.262164e-04
Cameroon	119240.0	1923.0	1.029000e+04	2.722426e+07	0.000378	7.063552e-05
Afghanistan	174081.0	7617.0	1.374200e+04	3.983543e+07	0.000345	1.912117e-04
Congo	24020.0	409.0	1.486000e+03	5.657017e+06	0.000263	7.229959e-05
Iraq	2305083.0	25013.0	9.985000e+03	4.117935e+07	0.000242	6.074161e-04
Gambia	12039.0	365.0	4.290000e+02	2.486937e+06	0.000173	1.467669e-04
Myanmar	592139.0	19376.0	3.800000e+03	5.480601e+07	0.000069	3.535378e-04
Burundi	38127.0	38.0	4.600000e+02	1.225543e+07	0.000038	3.100667e-06
Northern Cyprus	0.0	0.0	1.988000e+03	1.474690e+08	0.000013	0.000000e+00
Chad	7257.0	190.0	0.000000e+00	1.691498e+07	0.000000	1.123264e-05
Angola	98746.0	1903.0	0.000000e+00	3.393361e+07	0.000000	5.608009e-05
Tonga	355.0	0.0	0.000000e+00	1.067590e+05	0.000000	0.000000e+00
Yemen	11772.0	2135.0	0.000000e+00	3.049064e+07	0.000000	7.002149e-05
Armenia	420498.0	8495.0	0.000000e+00	2.968128e+06	0.000000	2.862073e-03
Tokelau	0.0	0.0	0.000000e+00	1.368000e+03	0.000000	0.000000e+00
Togo	36808.0	272.0	0.000000e+00	8.478242e+06	0.000000	3.208212e-05
Cook Islands	2.0	0.0	0.000000e+00	1.757200e+04	0.000000	0.000000e+00
Comoros	8033.0	161.0	0.000000e+00	8.884560e+05	0.000000	1.812133e-04
Austria	2744023.0	14902.0	0.000000e+00	9.043072e+06	0.000000	1.647891e-03
Turkmenistan	0.0	0.0	0.000000e+00	6.117933e+06	0.000000	0.000000e+00
Venezuela	515582.0	5645.0	0.000000e+00	2.870495e+07	0.000000	1.966560e-04
Tuvalu	0.0	0.0	0.000000e+00	1.192500e+04	0.000000	0.000000e+00
Costa Rica	811040.0	8057.0	0.000000e+00	5.139053e+06	0.000000	1.567799e-03
Wallis and Futuna	454.0	7.0	0.000000e+00	1.109400e+04	0.000000	6.309717e-04
Burkina Faso	20751.0	375.0	0.000000e+00	2.149710e+07	0.000000	1.744422e-05
British Virgin Islands	6085.0	62.0	0.000000e+00	3.042300e+04	0.000000	2.037932e-03
Bonaire Sint Eustatius and Saba	7599.0	33.0	0.000000e+00	2.644500e+04	0.000000	1.247873e-03
Bermuda	11561.0	123.0	0.000000e+00	6.209200e+04	0.000000	1.980932e-03
Benin	26776.0	163.0	0.000000e+00	1.245103e+07	0.000000	1.309129e-05
Belarus	923432.0	6506.0	0.000000e+00	9.442867e+06	0.000000	6.889857e-04
Vanuatu	19.0	1.0	0.000000e+00	3.144640e+05	0.000000	3.180014e-06
Vatican	29.0	0.0	0.000000e+00	8.120000e+02	0.000000	0.000000e+00
Turks and Caicos Islands	5868.0	36.0	0.000000e+00	3.922600e+04	0.000000	9.177586e-04
Papua New Guinea	41351.0	638.0	0.000000e+00	9.119005e+06	0.000000	6.996377e-05
Tanzania	33620.0	798.0	0.000000e+00	6.149844e+07	0.000000	1.297594e-05
Djibouti	15547.0	189.0	0.000000e+00	1.002197e+06	0.000000	1.885857e-04
Palau	3823.0	6.0	0.000000e+00	1.817400e+04	0.000000	3.301420e-04
Niue	0.0	0.0	0.000000e+00	1.614000e+03	0.000000	0.000000e+00
Niger	8763.0	307.0	0.000000e+00	2.513081e+07	0.000000	1.221608e-05
Nicaragua	18105.0	225.0	0.000000e+00	6.702379e+06	0.000000	3.357017e-05
New Caledonia	55503.0	301.0	0.000000e+00	2.882170e+05	0.000000	1.044352e-03
Saint Helena	4.0	0.0	0.000000e+00	6.095000e+03	0.000000	0.000000e+00
Netherlands	6540294.0	21710.0	0.000000e+00	1.717309e+07	0.000000	1.264187e-03
Saint Pierre and Miquelon	1089.0	1.0	0.000000e+00	5.771000e+03	0.000000	1.732802e-04
Montserrat	164.0	2.0	0.000000e+00	4.981000e+03	0.000000	4.015258e-04

Samoa	33.0	0.0	0.000000e+00	2.001440e+05	0.000000	0.000000e+00
Monaco	9457.0	54.0	0.000000e+00	3.952000e+04	0.000000	1.366397e-03
Sao Tome and Principe	5934.0	76.0	0.000000e+00	2.233640e+05	0.000000	3.402518e-04
Micronesia (country)	1.0	0.0	0.000000e+00	1.162550e+05	0.000000	0.000000e+00
Marshall Islands	8.0	0.0	0.000000e+00	5.961800e+04	0.000000	0.000000e+00
Mali	30391.0	722.0	0.000000e+00	2.085572e+07	0.000000	3.461879e-05
Lesotho	32612.0	696.0	0.000000e+00	2.159067e+06	0.000000	3.223615e-04
Sint Maarten (Dutch part)	0.0	0.0	0.000000e+00	4.342100e+04	0.000000	0.000000e+00
Slovakia	2150666.0	18567.0	0.000000e+00	5.449270e+06	0.000000	3.407245e-03
Kiribati	2953.0	11.0	0.000000e+00	1.213880e+05	0.000000	9.061851e-05
Solomon Islands	7261.0	106.0	0.000000e+00	7.039950e+05	0.000000	1.505693e-04
Jersey	0.0	0.0	0.000000e+00	1.010730e+05	0.000000	0.000000e+00
International	721.0	15.0	0.000000e+00	1.474690e+08	0.000000	1.017163e-07
Pitcairn	0.0	0.0	0.000000e+00	4.700000e+01	0.000000	0.000000e+00
Eritrea	9705.0	103.0	0.000000e+00	3.601462e+06	0.000000	2.859950e-05
Syria	54744.0	3082.0	0.000000e+00	1.827570e+07	0.000000	1.686392e-04
Falkland Islands	115.0	0.0	0.000000e+00	3.528000e+03	0.000000	0.000000e+00

Here, we'll choose China(source of COVID-19), Argentina and Brazil from South America and Germany and Switzerland from Europe to see how different are speed of vaccination between countries in same continent and different vaccination ratio. We'll use a line plot for this purpose(to see the trend in time). Let's see the result!

In [60]:

```
countries = ['China', 'Argentina', 'Brazil', 'Germany', 'Switzerland']

selected = covid[covid['location'].isin(countries)]
selected = selected[selected['date'].dt.year == 2021]
selected['month'] = selected['date'].dt.month
selected = selected.groupby(by=['location', "month"]).sum()

selected.reset_index().pivot('month','location','new_vaccinations').plot( title='Speed of Vaccination in 2021', g
plt.show()
```



As you can see, China's speed in vaccination changes dramatically over time(as the source of COVID-19) and typically has higher speed between other countries. In Europe, Germany has higher speed than Switzerland and their difference is obvious. This fact is also true for Argentina and Brazil in South America, too. Brazil has higher speed than Argentina. In general, countries with more population have higher speed in vaccination. We can also see that in some period of time, speed for countries decreases(probably because number of cases decreases at that time) and rises again(probably because of new type of COVID-19).

Q4

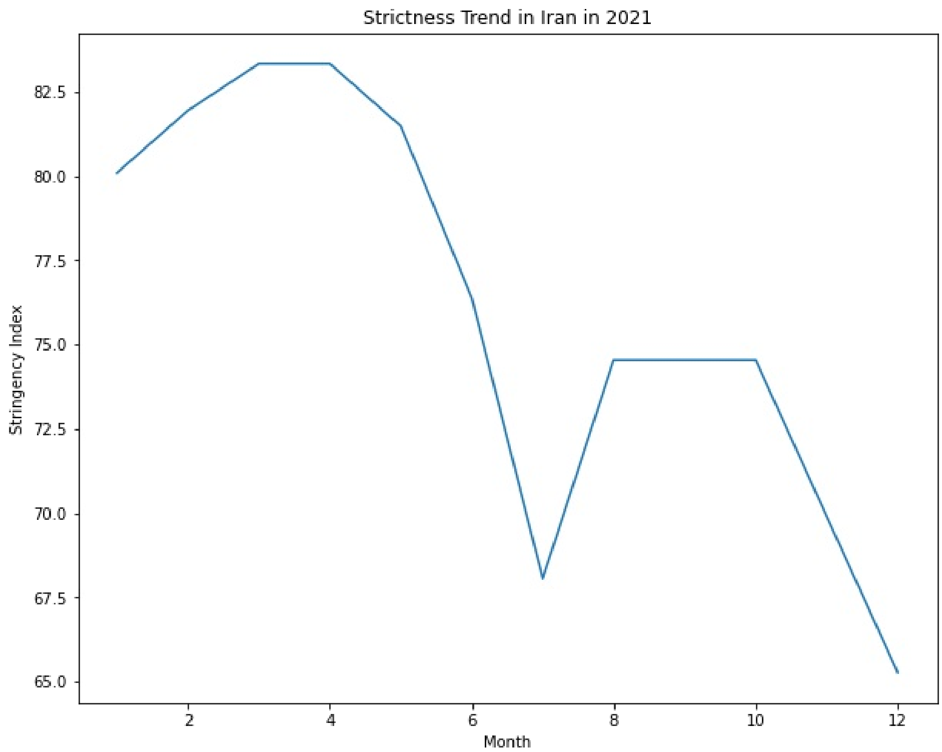
Now, we want to see strictness trend in controlling COVID-19 in Iran. To see this, we'll use the aggregated dataframe for Iran's stats, which is aggregated on 2021's months. In this aggregation, maximum stringency index(a factor for strictness) in month is used. We'll use a line plot to see the trend. Let's see the result!

In [61]:

```
pd.reset_option('display.max_rows')
```

In [62]:

```
plt.figure(figsize=(10,8))
ax = sns.lineplot(x=iran_summarized.index, y=iran_summarized['stringency_index'])
ax.set_title("Strictness Trend in Iran in 2021")
ax.set_xlabel('Month')
ax.set_ylabel('Stringency Index')
plt.show()
```



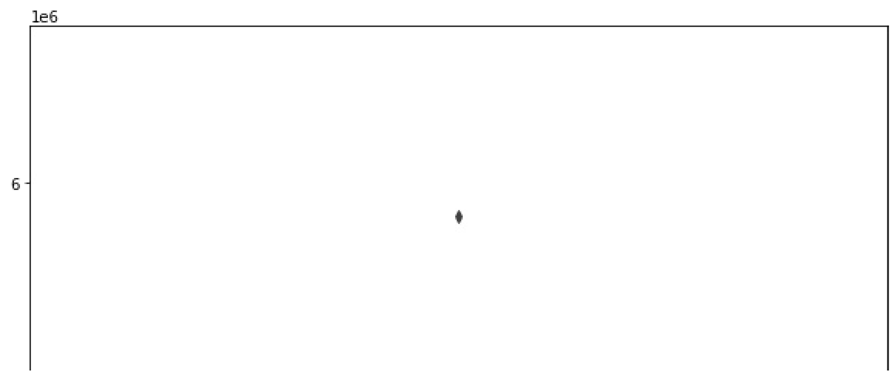
As you can see, there are some rises and falls in the trend. Highest strictness level, according to plot, is achieved in April and it falls in the next 3 months, probably because more people is vaccinated and there is no new variant of COVID-19 virus. After that, there is another rise in the next 3 month and finally, it falls till the last month of the year. One of the reasons for this fall and rises can be the spread of new variants of COVID-19. Another reason can be with reduction in strictness, there can be higher probability that virus spread in the country and number of cases increase and the government decide to increase the stringency index again.

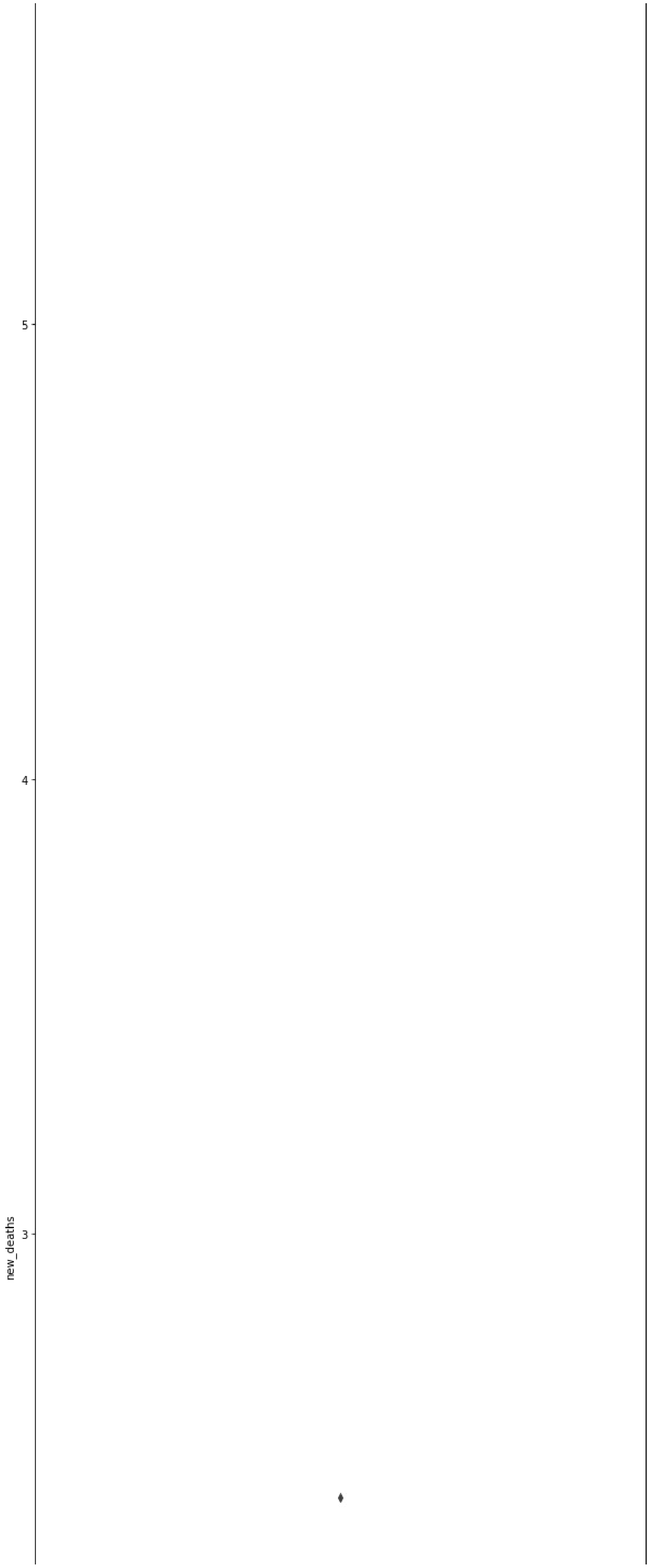
Q5

Now we'll use boxplot to see the distribution of total deaths in the dataframe. Let's see the results:

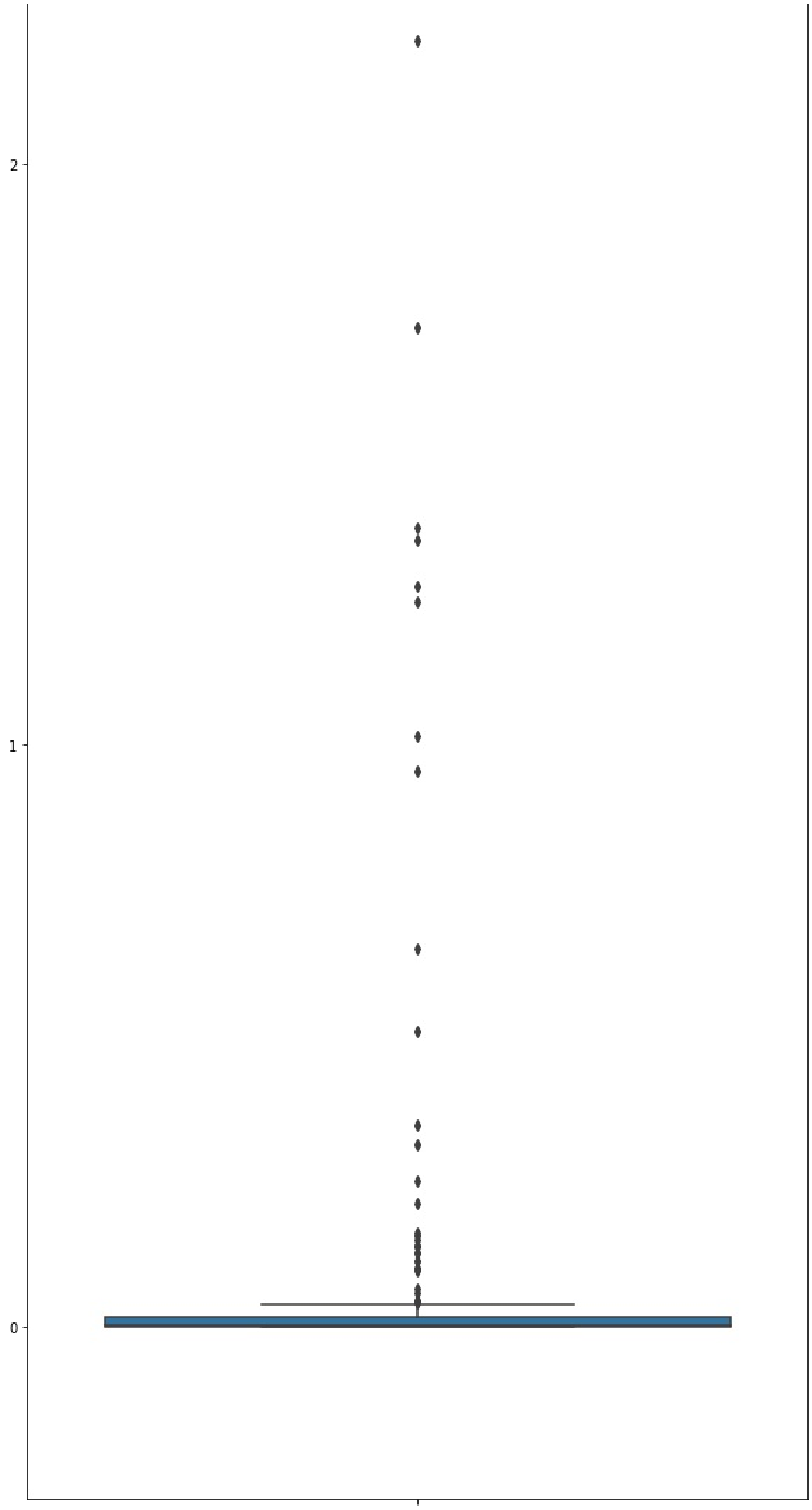
In [82]:

```
import plotly.express as px
plt.figure(figsize=(10,50))
sns.boxplot(y=summarized['new_deaths'])
plt.show()
```





5
4
3
new_deaths



As you can see, there are about 20 outliers in this box plot. To deal with with outliers, the simplest way is to ignore these countries. A better approach can be that instead of ignoring completely, choose some stats that doesn't affect estimations a lot from these outliers. Pay attention! The outliers have more population that others, which means that higher number of death is logical somehow. To see the skewness in the boxplot, first let's see the average value of deaths:

In [64]:
```
summarized['new_deaths'].mean()
```

Out[64]:
```
104214.1050420168
```

According to mean and the shape of boxplot, we can say that there is a positive skew in the plot(plot is right-skewed). To see the name of locations that are outliers, we'll see top 20 countries in the number of deaths:

In [65]:

```
summarized.sort_values(by='new_deaths', ascending = False)[:20]
```

Out[65]:

location	new_cases	new_deaths	new_vaccinations	population
World	439011701.0	5946817.0	1.122516e+10	7.874966e+09
Upper middle income	118212049.0	2421471.0	4.804681e+09	2.513673e+09
High income	236393139.0	2211971.0	2.306054e+09	1.214930e+09
Europe	158815550.0	1717056.0	1.247823e+09	7.489630e+08
North America	93195128.0	1374621.0	1.030269e+09	5.965813e+08
Asia	117811418.0	1351250.0	7.544342e+09	4.678445e+09
Lower middle income	82588725.0	1271503.0	3.974325e+09	3.330653e+09
South America	54222621.0	1247285.0	7.462801e+08	4.342601e+08
European Union	110033392.0	1015772.0	8.184271e+08	4.471899e+08
United States	79143715.0	954518.0	5.537498e+08	3.329151e+08
Brazil	28741413.0	650254.0	3.902730e+08	2.139934e+08
India	42945160.0	507015.0	1.708969e+09	1.393409e+09
Russia	16353868.0	346197.0	1.434654e+08	1.459120e+08
Mexico	5521744.0	311470.0	1.451079e+08	1.302622e+08
Africa	11230524.0	248668.0	5.818190e+08	1.373486e+09
Peru	3518721.0	211453.0	6.113136e+07	3.335942e+07
United Kingdom	18241147.0	160727.0	1.372498e+08	6.820711e+07
Italy	12868066.0	155245.0	1.343115e+08	6.036747e+07
Indonesia	5630096.0	149036.0	2.724306e+08	2.763618e+08
Colombia	6067023.0	138899.0	4.938413e+07	5.126584e+07

Q6

Now we want to see the impact of population density, median age, handwashing facilities, hospital beds per thousand and human development index on number of death and cases. To show this, we'll use a scatter plot for each pair of variables we want to find the correlation between(10 plots in total) and study the relation between them. First we need aggregate some variables for plots:

In [66]:

```
first_factors = ["population_density", "median_age", "handwashing_facilities",
                 "hospital_beds_per_thousand", "human_development_index"]
second_factors = ["new_cases", "new_deaths"]
new_summarized = covid[["location"] + first_factors]

new_summarized = new_summarized.groupby(by="location").max()
new_summarized = summarized.join(new_summarized)
new_summarized
```

Out[66]:

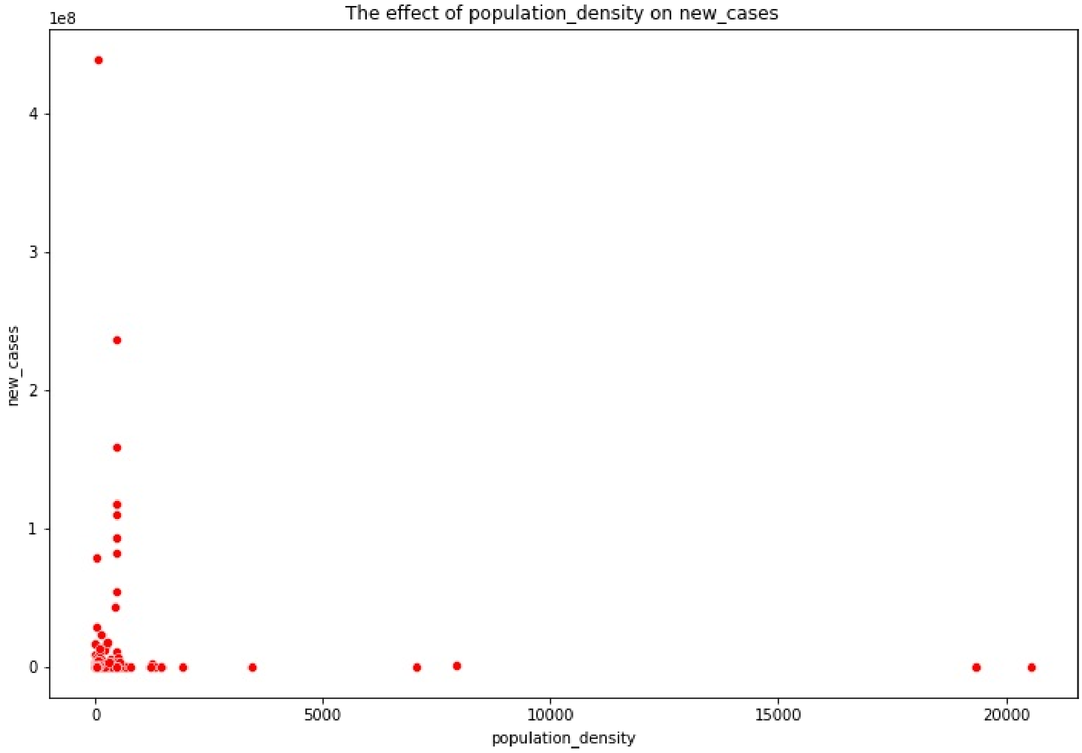
location	new_cases	new_deaths	new_vaccinations	population	population_density	median_age	handwashing_facilities	hospital_beds_p
Afghanistan	174081.0	7617.0	1.374200e+04	3.983543e+07	54.422000	18.600000	37.746000	
Africa	11230524.0	248668.0	5.818190e+08	1.373486e+09	464.408404	30.568558	50.790872	
Albania	271825.0	3474.0	1.415150e+06	2.872934e+06	104.871000	38.000000	50.790872	
Algeria	265079.0	6843.0	1.707860e+05	4.461663e+07	17.348000	29.100000	83.741000	
Andorra	38249.0	151.0	4.802000e+03	7.735400e+04	163.755000	30.568558	50.790872	
...	...	...	...	...	...	...	...	
Wallis and Futuna	454.0	7.0	0.000000e+00	1.109400e+04	464.408404	30.568558	50.790872	
World	439011701.0	5946817.0	1.122516e+10	7.874966e+09	58.045000	30.900000	60.130000	
Yemen	11772.0	2135.0	0.000000e+00	3.049064e+07	53.508000	20.300000	49.542000	
Zambia	313203.0	3955.0	1.321156e+06	1.892066e+07	22.995000	17.700000	13.938000	
Zimbabwe	237509.0	5396.0	7.080893e+06	1.509217e+07	42.729000	19.600000	36.791000	

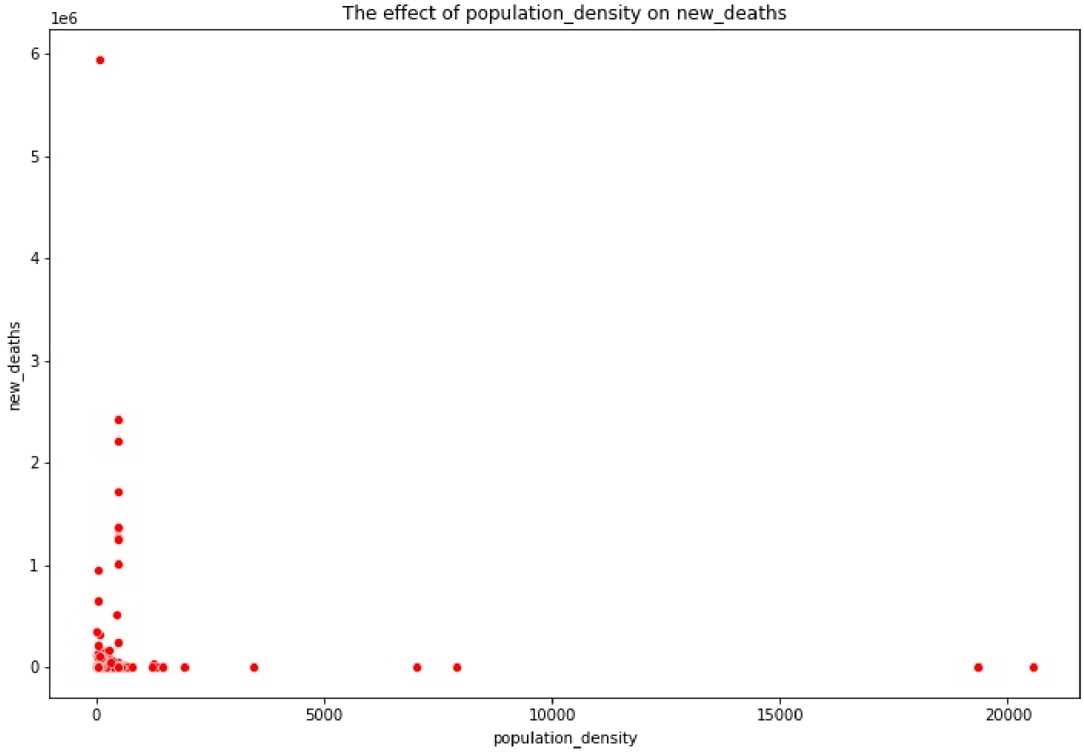
238 rows × 9 columns

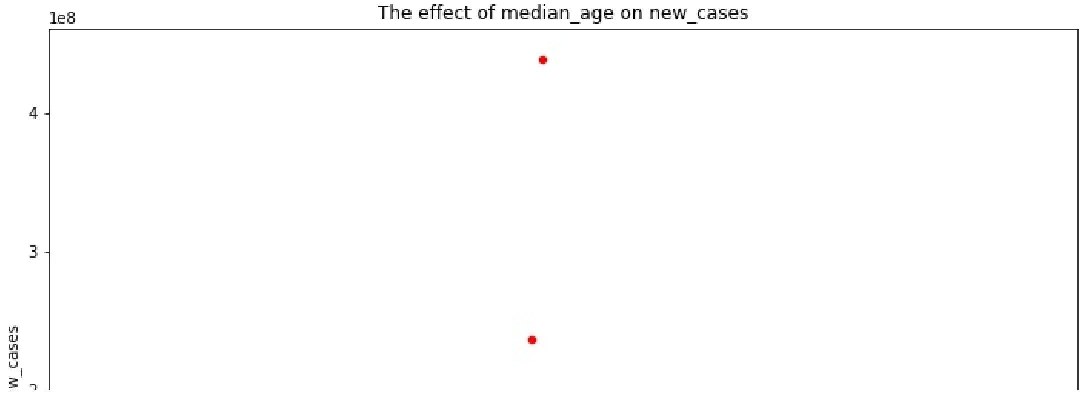
In [67]:

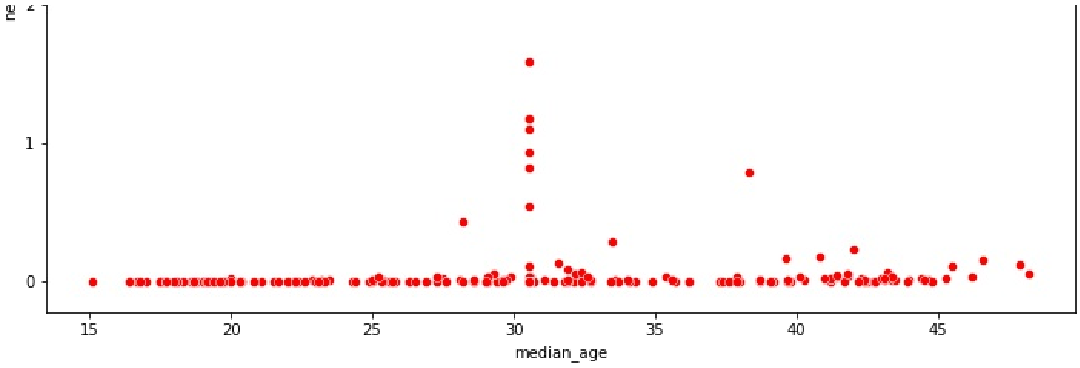
```
for first_factor in first_factors:
```

```
for first_factor in first_factors:
    for second_factor in second_factors:
        plt.figure(figsize=(12,8))
        ax = sns.scatterplot(x=new_summarized[first_factor], y=new_summarized[second_factor], color='r')
        ax.set_title('The effect of {} on {}'.format(first_factor, second_factor))
        ax.set_xlabel(first_factor)
        ax.set_ylabel(second_factor)
        plt.show()
```

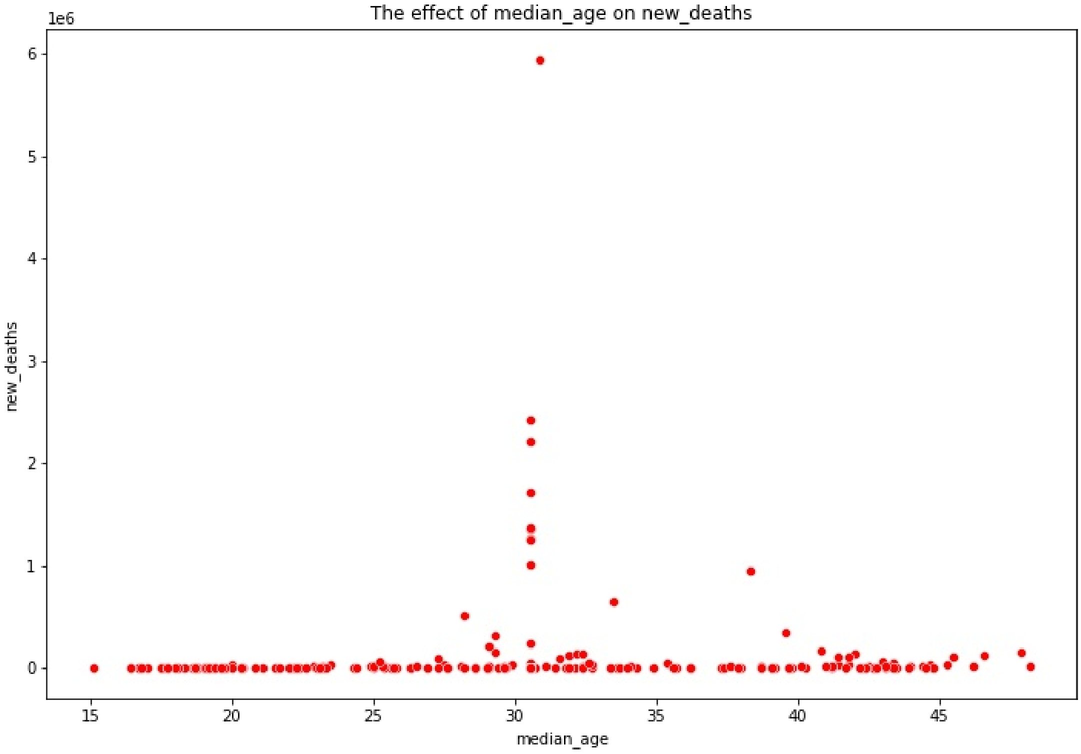




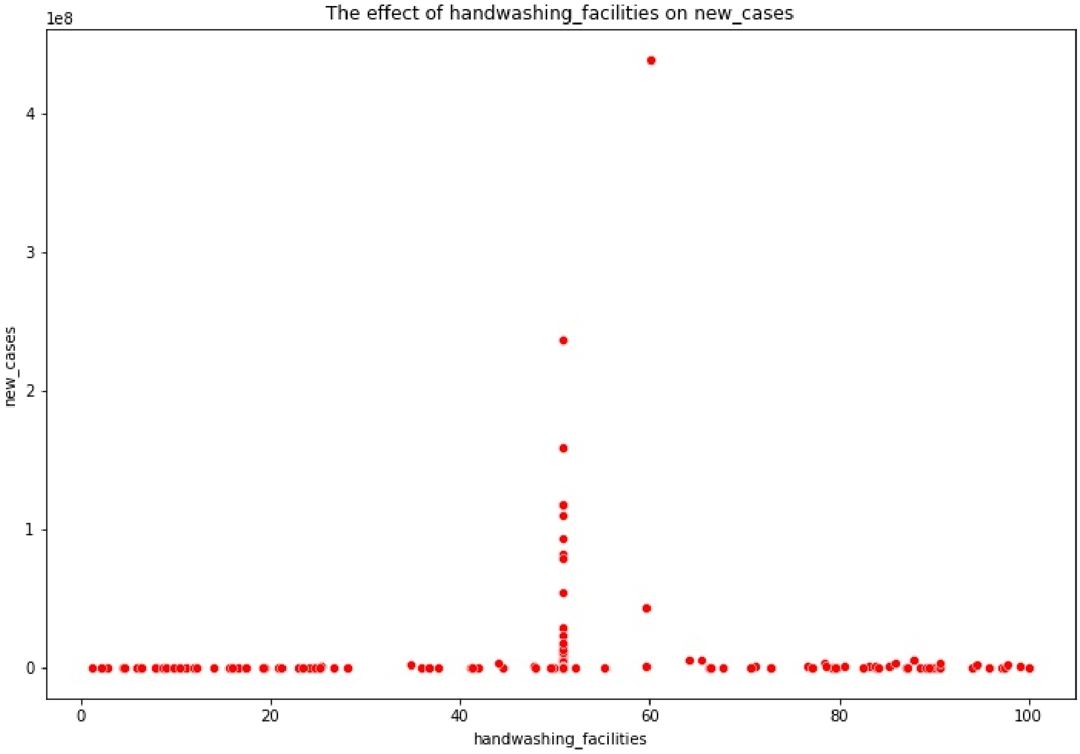




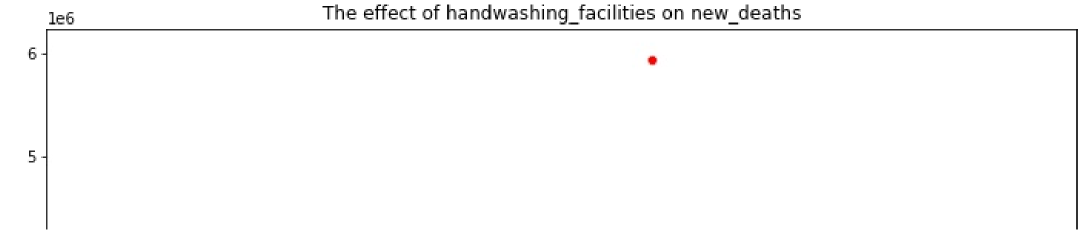
The effect of median_age on new_deaths

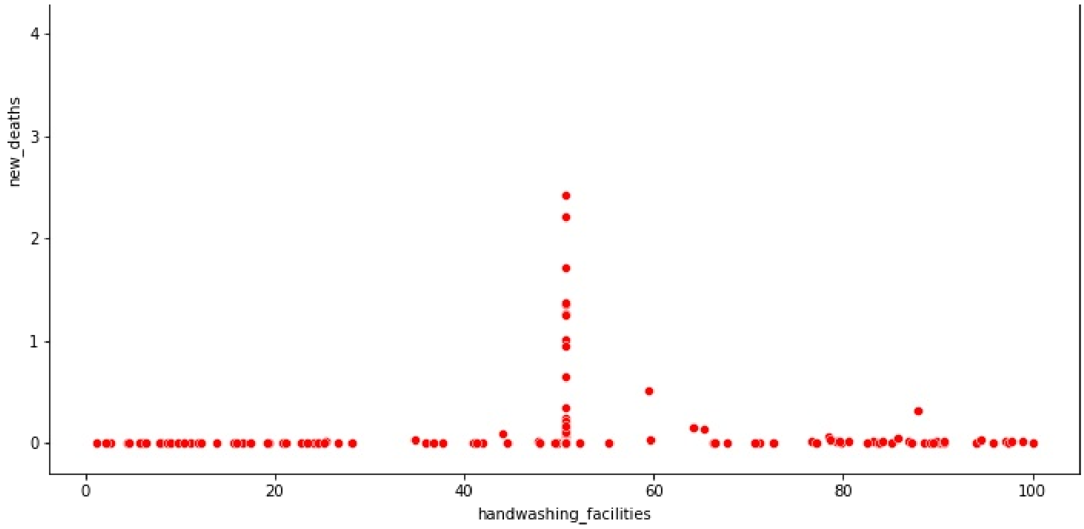


The effect of handwashing_facilities on new_cases

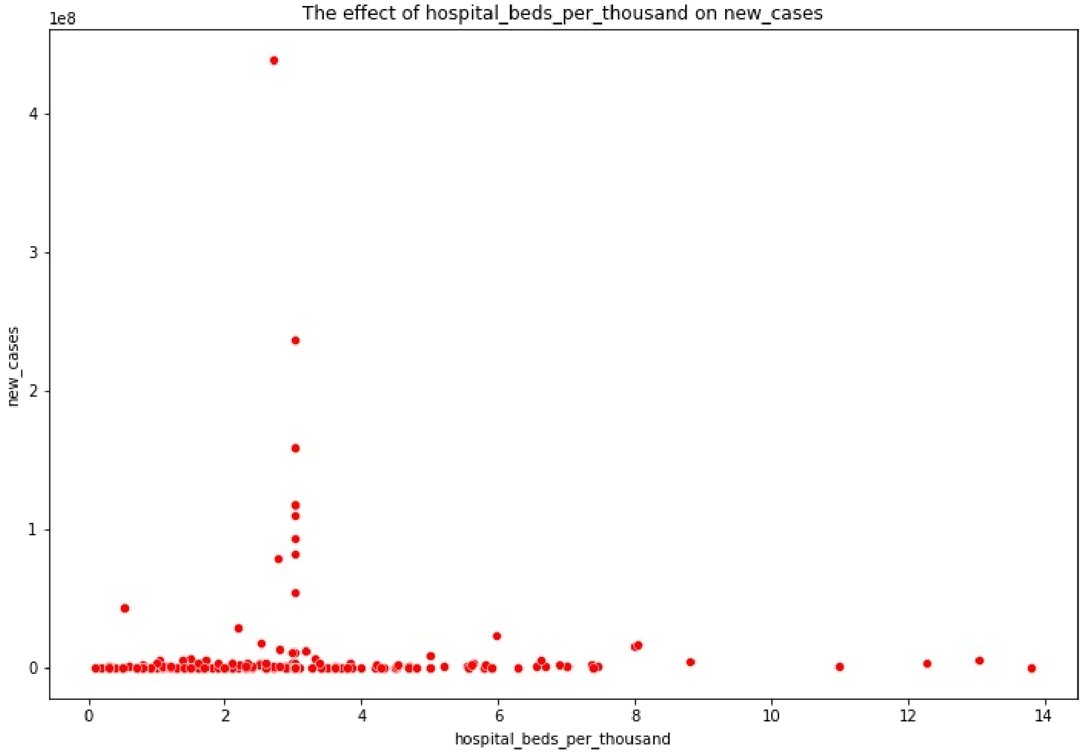


The effect of handwashing_facilities on new_deaths

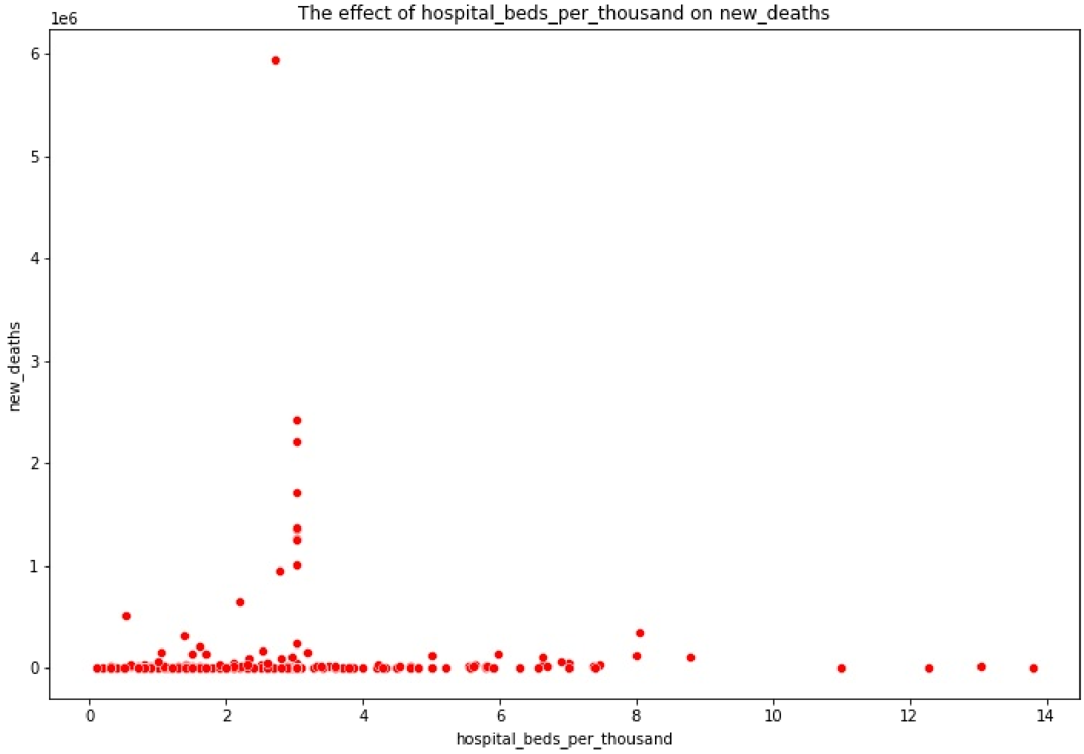




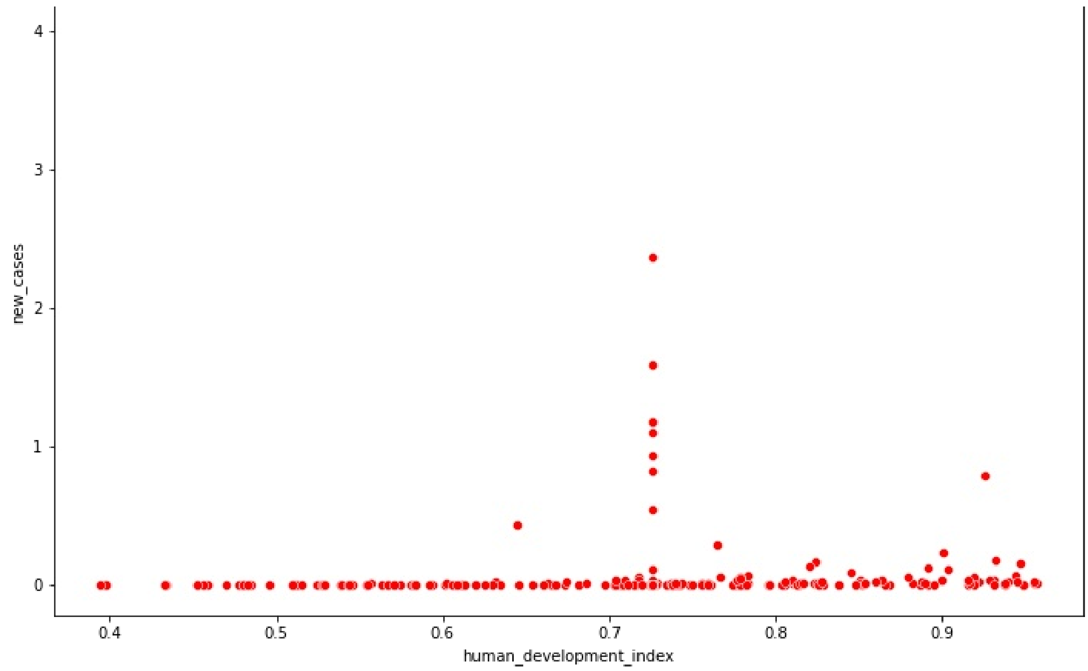
The effect of hospital_beds_per_thousand on new_cases

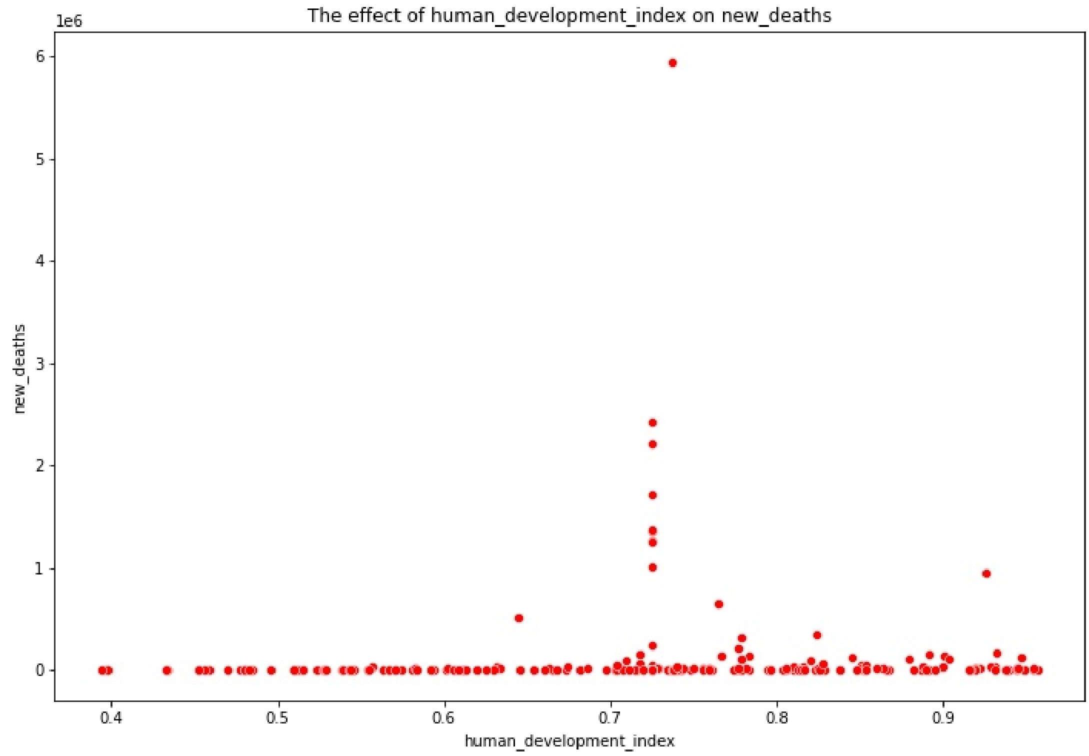


The effect of hospital_beds_per_thousand on new_deaths



The effect of human_development_index on new_cases





There are many similarities among these plots. First, we can conclude that the number of deaths and cases doesn't only depend on one factor, but depends on a number of factors together. As you can see, these numbers in almost all plots are the same for much the range of factors mentioned in the question and changes dramatically for a small part of range and after that, the normal trend continues. It means that if other factors remains unchanged, changes in mentioned factors don't really change the number of deaths and cases and only in some parts, the change is obvious. For example, if median age gets around 31, number of cases and deaths rises dramarically.

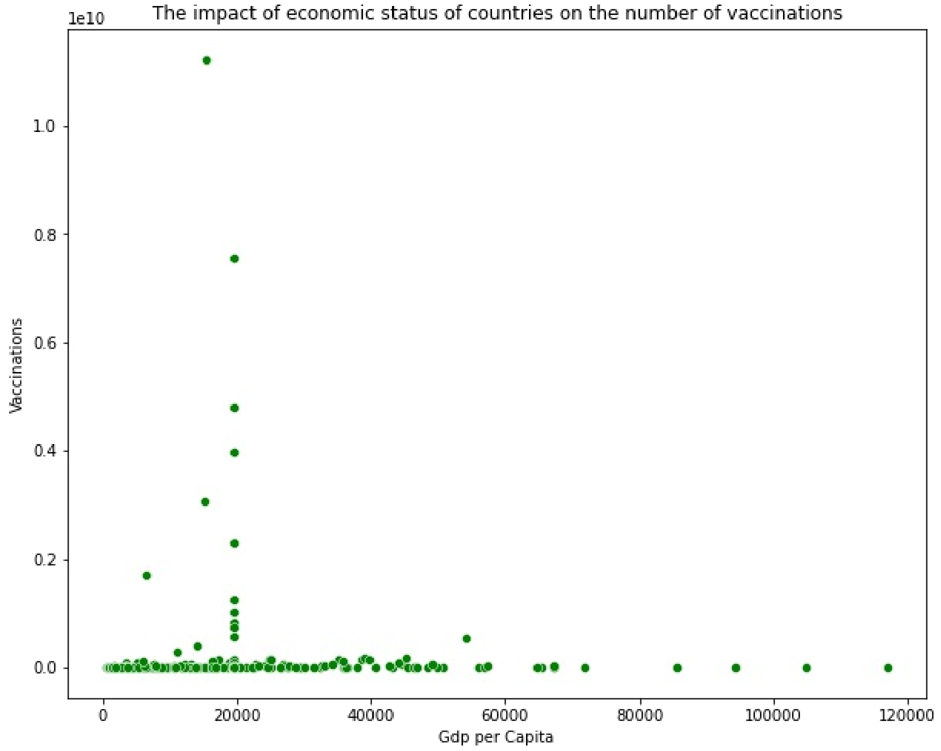
Q7

Now let's consider the effect of GDP per capita on number of vaccinations!

In [68]:

```
new_summarized = covid[["location", 'gdp_per_capita']]
new_summarized = new_summarized.groupby(by="location").max()
new_summarized = summarized.join(new_summarized)

plt.figure(figsize=(10,8))
ax = sns.scatterplot(x=new_summarized['gdp_per_capita'], y=new_summarized['new_vaccinations'], color='g')
ax.set_title("The impact of economic status of countries on the number of vaccinations")
ax.set_xlabel('Gdp per Capita')
ax.set_ylabel('Vaccinations')
```

```
plt.show()
```



Again, we can see that in most of range of GDP, total vaccinations doesn't change that much and wealth doesn't affect vaccination that much. This is because after some amount of time, most of coutries(independent of their wealth) tried to vaccinate their population with their power and help of stronger coutries and that helped vaccination speed and amount.

Q8

Finally, let's see number of cases trend over 2021 by month!

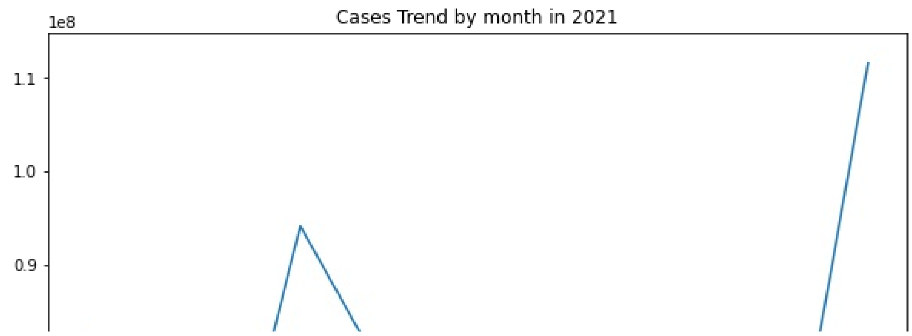
In [69]:

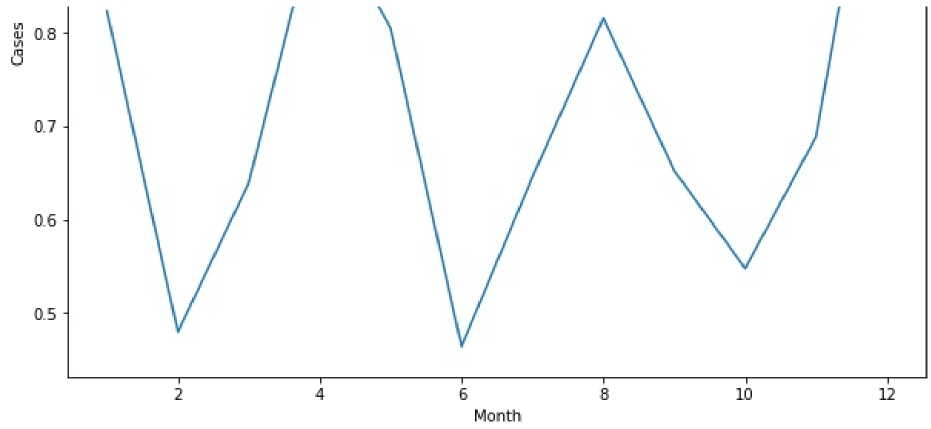
```
covid_2021 = covid[covid['date'].dt.year == 2021]
covid_2021['month'] = covid_2021['date'].dt.month
covid_2021_summarized = covid_2021[["month", "new_cases"]]
covid_2021_summarized = covid_2021_summarized.groupby(by="month").sum()
```

```
/home/taha/.local/lib/python3.7/site-packages/ipykernel_launcher.py:2: SettingWithCopyWarning:


A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

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

In [70]:

```
plt.figure(figsize=(10,8))
ax = sns.lineplot(x=covid_2021_summarized.index, y=covid_2021_summarized['new_cases'])
ax.set_title("Cases Trend by month in 2021")
ax.set_xlabel('Month')
ax.set_ylabel('Cases')
plt.show()
```





As you can see, there are many rises and falls in the plot. Once a new variant of virus spreads in the world, total number of cases rises, and scientist try to make new vaccines that is strong over this variant. Also strigency index helps coutries control new variants and this means a fall in total number of cases in world. As you can see, highest number of cases happened in December(probably because of Omicron) and least number of cases happened in June.

Bonus

Q1

Now let's see total number of deaths past 3 months over countries on world map using geopandas library!

In [71]:

```
last_3_months = covid[['location', 'new_deaths', 'population', 'date', 'iso_code']]

last_3_months = last_3_months.sort_values(by="date",ascending=True).set_index("date").last("3M")
last_3_months = last_3_months.reset_index()

last_3_months_deaths = last_3_months[["location", "new_deaths"]]
last_3_months_deaths = last_3_months_deaths.groupby(by="location").sum()
population = last_3_months[["location", "population", "iso_code"]]
population = population.groupby(by="location").max()
last_3_months = last_3_months_deaths.join(population)

last_3_months['death_per_pop'] = last_3_months['new_deaths'] / last_3_months['population']
last_3_months = last_3_months.drop(['new_deaths', 'population'], axis=1)
last_3_months = last_3_months.reset_index()
last_3_months = last_3_months.rename(columns={'location': 'name', 'iso_code' : 'CODE'})
last_3_months
```

Out[71]:

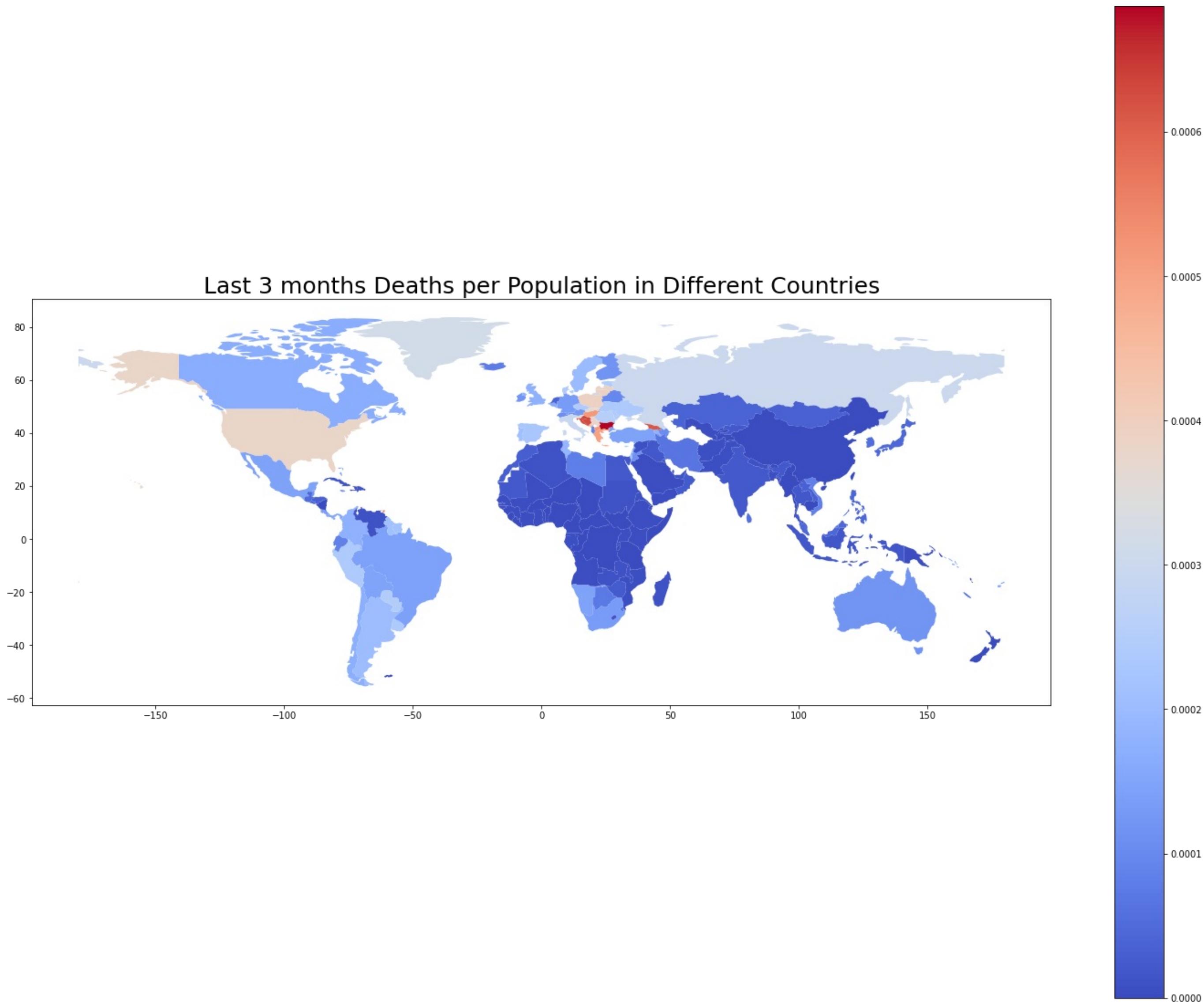
	name	CODE	death_per_pop
0	Afghanistan	AFG	0.000007
1	Africa	OWID_AFR	0.000015
2	Albania	ALB	0.000089
3	Algeria	DZA	0.000013
4	Andorra	AND	0.000142
...	...	...	...
227	Wallis and Futuna	WLF	0.000000
228	World	OWID_WRL	0.000067
229	Yemen	YEM	0.000005
230	Zambia	ZMB	0.000012
231	Zimbabwe	ZWE	0.000026

232 rows × 3 columns

In [72]:

```
import geopandas
import matplotlib.pyplot as plt

world = geopandas.read_file(geopandas.datasets.get_path('naturalearth_lowres'))
world.columns=['pop_est', 'continent', 'name', 'CODE', 'gdp_md_est', 'geometry']
merge=pd.merge(world, last_3_months, on='CODE')
```

```
merge.plot(column='death_per_pop',
           figsize=(25, 20),
           legend=True,cmap='coolwarm')
plt.title('Last 3 months Deaths per Population in Different Countries',fontsize=25)
plt.show()
```

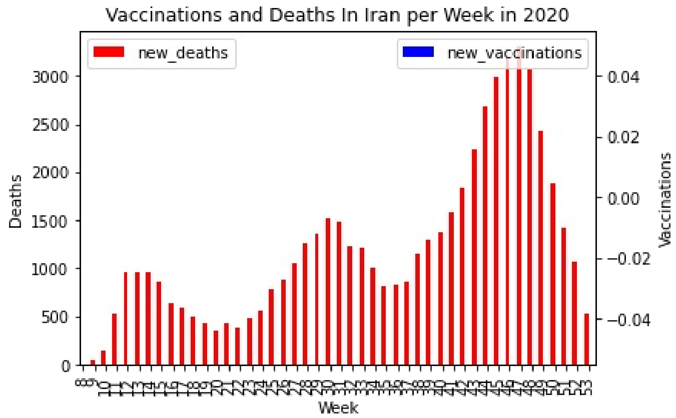


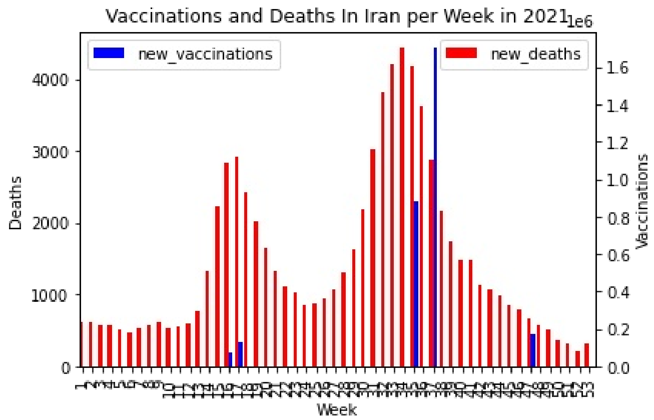
Q2

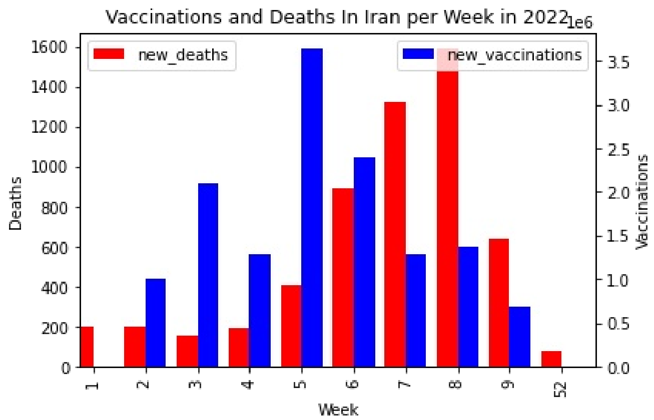
Finally, let's see number of deaths and vaccinations over weeks in years. We use a double bar chart to see the trend:

In [76]:

```
new_iran = iran.copy()
new_iran['week'] = new_iran['date'].dt.isocalendar().week

factors = ['new_deaths', 'new_vaccinations']
years = [2020, 2021, 2022]
for year in years:
    iran_year = new_iran[new_iran['date'].dt.year == year]
    iran_year_summarized = iran_year[['week'] + factors]
    iran_year_summarized = iran_year_summarized.groupby(by='week').sum()
    fig = plt.figure()

    ax = fig.add_subplot(111)
    ax2 = ax.twinx()

    width = 0.4

    iran_year_summarized.new_deaths.plot(kind='bar', color='red', ax=ax, width=width, position=1)
    iran_year_summarized.new_vaccinations.plot(kind='bar', color='blue', ax=ax2, width=width, position=0)
    ax.set_title('Vaccinations and Deaths In Iran per Week in {}'.format(year))
    ax.set_ylabel('Deaths')
    ax2.set_ylabel('Vaccinations')
    ax.set_xlabel('Week')
    ax.legend()
    ax2.legend()
    plt.show()
```







As you can see, in the first year, there is't any vaccinations in Iran(there was no vaccines available past then) and there is an ascending trend in number of deaths over weeks. In next year, as number of vaccination increases, we see a fall in number of deaths, which means that vaccination could help country control COVID-19. Finally in last year, again number of deaths increased and number of vaccinations decreased, which probably means that country couldn't get enough vaccines for the new trend.