Analyzing Covid Data and World Happiness Record Data

The data spans 1/22/2020 - 1/26/2023

Summary of Analysis

Full analysis and visualizations below.

Covid death and infection rates along the time-series between the United Kingdom, USA, and Sweden are positively correlated. Peaks and troughs occur in tandem with the seasonality of covid for all three countries. This may be due to the seasonality of the covid virus and the climate similarities of the United Kingdom, Sweden and much of the continental USA and/or better access and resources for recording and reporting the data. We also see a consistent correlation and pattern between spikes in infection rate and death rates. This pattern is also consistent with India, however, not as consistent. India's data includes outliers where there are surges in death rates that do not have a corresponding surge in infection rates. This could be due to many non-lethal covid infections going undiagnosed. China's infection rate and death rate do not follow the trends seen in the other four countries. In March and April of 2022 China experienced a large increase in infection rate but there was no significant increase in the death rate. We also do not see the seasonality displayed in the other four countries. From early 2020 to late 2021 infection rates stay very close to the y-axis, virtually no infections. The death rate displayed the same trend but remained unchanged until late 2022. At the end of 2022 and into 2023 China's death rate climbed to about 60,000 in a very short period of time. It does correlate with a spike in infection rates; however, the infection rate peaked at about 80,000. This translates to roughly a 75% death rate over this period. This may be due to China's zero covid policy and strict regulations from 2020 through 2022. Civil unrest regarding China's strict regulations started in November 2022, large numbers of people gathered in protest and could have spread covid at an unprecedented rate. China also has an aging population, in 2018 more than 17% of the population was reported to be over 60 years old. Considering all variables with health facilities past capacity, the country experiencing its largest surge in infections since the outbreak, and a large percent of the population being more at risk of serious complications may have created the massive surge in death rate while leaving many nonlethal cases going unreported. This would greatly skew the calculated death rate. The cumulative average death rate for every country is within the expected range, varying from 0.87% - 1.08%. China is anomalous in comparison to the other countries. There are varying reasons and explanations as to why, regardless of the reason China is an outlier.

The covid data in relation to the World Health report initially shows a positive correlation in

varying degrees across all categories. For infection rate vs GDP we see a low moderate correlation and high statistical relevance. However, in death rate vs GDP the correlation and statistical relevance is weak. This trend continues for social support, healthy life choices, and freedom to make life choices. The difference between correlation and statistical relevance of infection rate and death rate against World Health Report data could be due to unrecorded/unreported infections. Infections that lead to death are more likely to be recorded due to the severity of the case leading to a higher necessity of hospitalization or medical help.

Loading Libraries for Analysis

In [358…

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
from scipy.stats import pearsonr as prs
print('Imported.')
```

```
Imported.
```

Importing Covid Infection Data from CSV

In [359…

```
corona_dataset_csv = pd.read_csv("Datasets/covid19_Confirmed_dataset.csv")
corona_dataset_csv.head(10)
```

Out[359]:

	Province/State	Country/Region	Lat	Long	1/22/2020	1/23/2020	1/24/2020	1/25/20
0	NaN	Afghanistan	33.93911	67.709953	0	0	0	
1	NaN	Albania	41.15330	20.168300	0	0	0	
2	NaN	Algeria	28.03390	1.659600	0	0	0	
3	NaN	Andorra	42.50630	1.521800	0	0	0	
4	NaN	Angola	-11.20270	17.873900	0	0	0	
5	NaN	Antarctica	-71.94990	23.347000	0	0	0	
6	NaN	Antigua and Barbuda	17.06080	-61.796400	0	0	0	
7	NaN	Argentina	-38.41610	-63.616700	0	0	0	
8	NaN	Armenia	40.06910	45.038200	0	0	0	
9	Australian Capital Territory	Australia	-35.47350	149.012400	0	0	0	

10 rows × 1105 columns

In [360…
```
corona_dataset_csv.shape
```

Out[360]:
```
(289, 1105)
```

Deleting Unecessary Columns

In [361…
```
corona_dataset_csv.drop(["Lat", "Long", "Province/State"], axis = 1, inplace = True)
```

In [362…
```
corona_dataset_csv.head(20)
```

Out[362]:

	Country/Region	1/22/2020	1/23/2020	1/24/2020	1/25/2020	1/26/2020	1/27/2020	1/28/2020
0	Afghanistan	0	0	0	0	0	0	0
1	Albania	0	0	0	0	0	0	0
2	Algeria	0	0	0	0	0	0	0
3	Andorra	0	0	0	0	0	0	0
4	Angola	0	0	0	0	0	0	0
5	Antarctica	0	0	0	0	0	0	0
6	Antigua and Barbuda	0	0	0	0	0	0	0
7	Argentina	0	0	0	0	0	0	0
8	Armenia	0	0	0	0	0	0	0
9	Australia	0	0	0	0	0	0	0
10	Australia	0	0	0	0	3	4	4
11	Australia	0	0	0	0	0	0	0
12	Australia	0	0	0	0	0	0	0
13	Australia	0	0	0	0	0	0	0
14	Australia	0	0	0	0	0	0	0
15	Australia	0	0	0	0	1	1	1
16	Australia	0	0	0	0	0	0	0
17	Austria	0	0	0	0	0	0	0
18	Azerbaijan	0	0	0	0	0	0	0
19	Bahamas	0	0	0	0	0	0	0

20 rows × 1102 columns

In [363…
```
corona_dataset_csv.shape
```

Out[363]:
```
(289, 1102)
```

Aggregrating Data by Country

In [364…
```
corona_dataset_aggregated = corona_dataset_csv.groupby("Country/Region").sum()
```

In [365…
```
corona_dataset_aggregated.shape
```

Out[365]:
```
(201, 1101)
```

Importing Covid Death Data from CSV

In [366…
```
corona_deaths = pd.read_csv("Datasets/covid19_deaths_dataset.csv")

corona_deaths.head()
```

Out[366]:

	Province/State	Country/Region	Lat	Long	1/22/2020	1/23/2020	1/24/2020	1/25/202
0	NaN	Afghanistan	33.93911	67.709953	0	0	0	
1	NaN	Albania	41.15330	20.168300	0	0	0	
2	NaN	Algeria	28.03390	1.659600	0	0	0	
3	NaN	Andorra	42.50630	1.521800	0	0	0	
4	NaN	Angola	-11.20270	17.873900	0	0	0	

5 rows × 1105 columns

Deleting Unecessary Columns

In [346…
```
corona_deaths.drop(["Lat", "Long", "Province/State"], axis = 1, inplace = True)
corona_deaths.head()
```

Out[346]:

	Country/Region	1/22/2020	1/23/2020	1/24/2020	1/25/2020	1/26/2020	1/27/2020	1/28/2020
0	Afghanistan	0	0	0	0	0	0	0
1	Albania	0	0	0	0	0	0	0
2	Algeria	0	0	0	0	0	0	0
3	Andorra	0	0	0	0	0	0	0
4	Angola	0	0	0	0	0	0	0

5 rows × 1102 columns

Aggregrating Data by Country

In [347…
```
corona_deaths_agg = corona_deaths.groupby("Country/Region").sum()
```

In [348…
```
corona_deaths_agg.shape
```

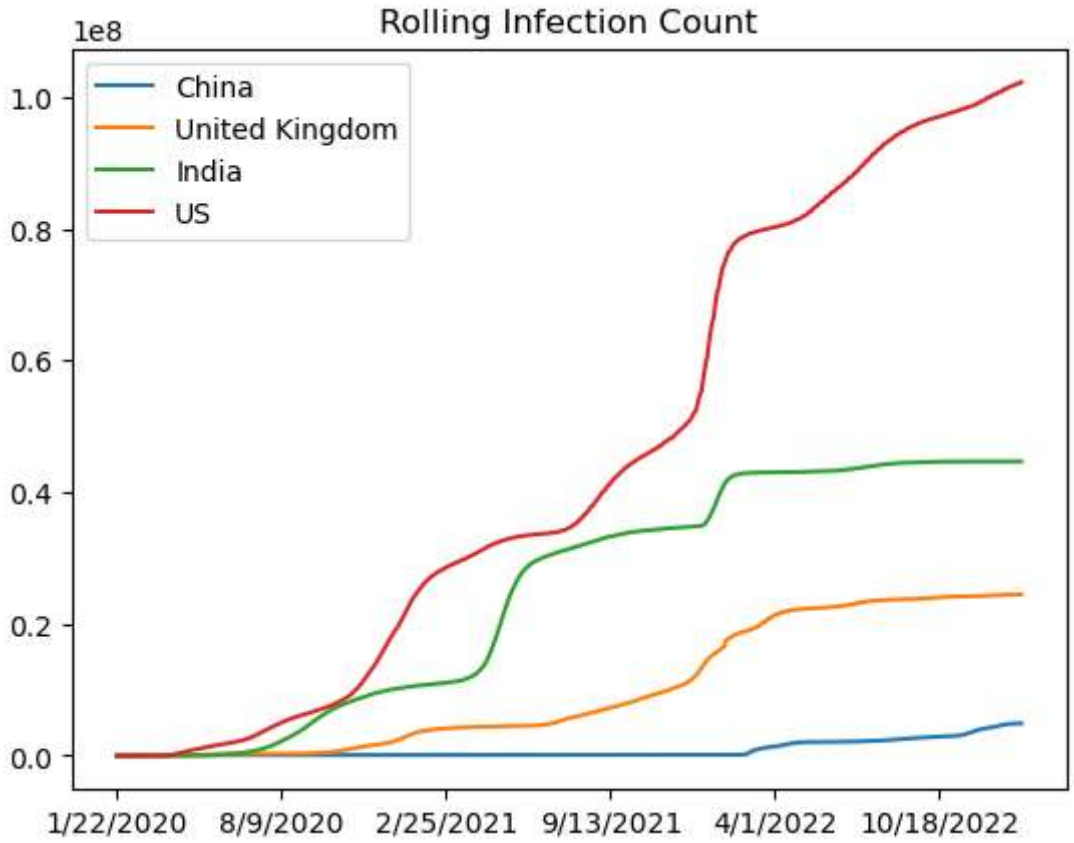
Out[348]:
```
(201, 1101)
```

Visualizing Covid Infections and Deaths

Rolling sum of covid Infections

In [349…
```
corona_dataset_aggregated.loc["China"].plot()
corona_dataset_aggregated.loc["United Kingdom"].plot()
corona_dataset_aggregated.loc["India"].plot()
corona_dataset_aggregated.loc["US"].plot()
plt.title('Rolling Infection Count')
plt.legend()
```

Out[349]:
```
<matplotlib.legend.Legend at 0x26f8a3105b0>
```

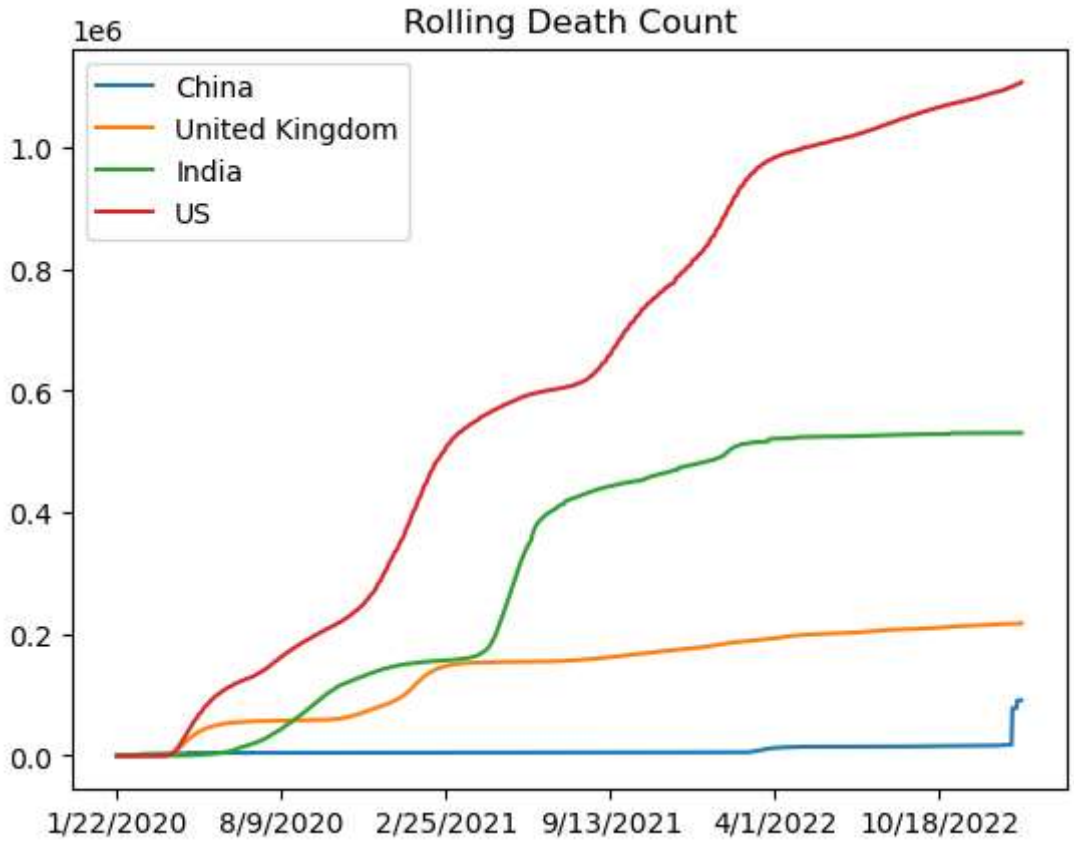


Rolling sum of covid deaths

In [350…
```
corona_deaths_agg.loc["China"].plot()
corona_deaths_agg.loc["United Kingdom"].plot()
corona_deaths_agg.loc["India"].plot()
```

```
corona_deaths_agg.loc["US"].plot()
plt.title('Rolling Death Count')
plt.legend()
```

Out[350]: <matplotlib.legend.Legend at 0x26f8a362250>

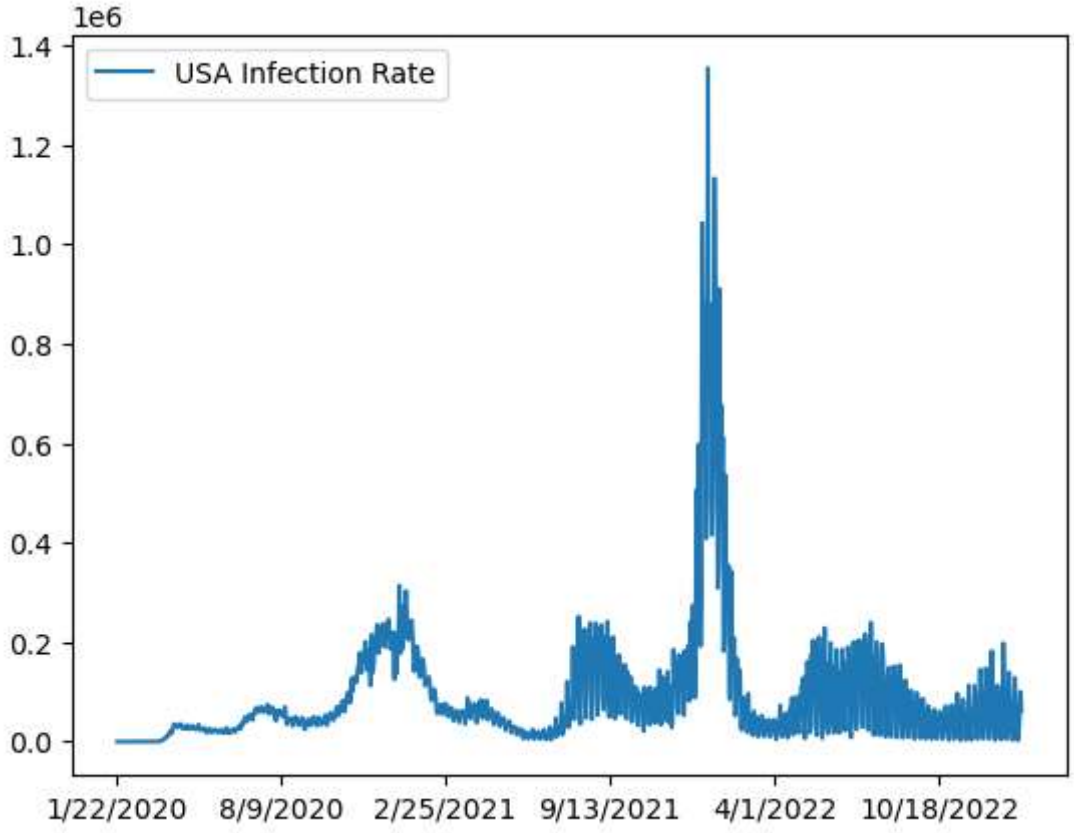


Covid infection rate (new infections within 24 hours)

In [351…

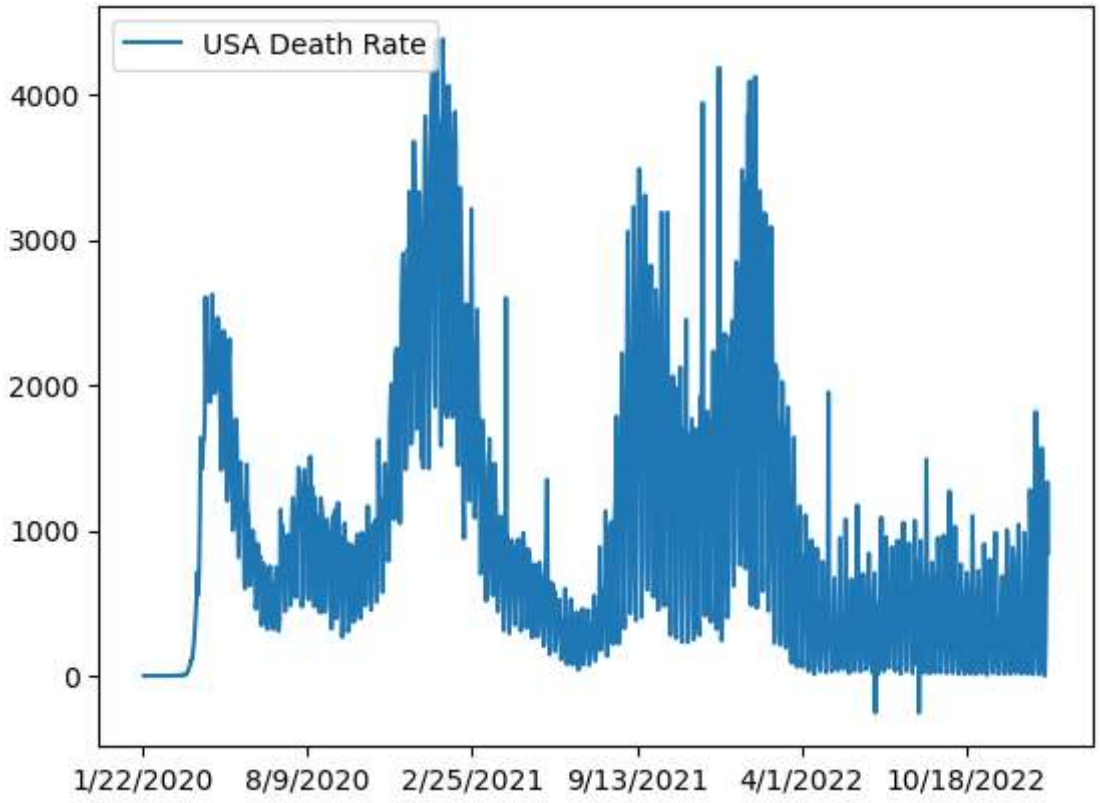
```
corona_dataset_aggregated.loc["US"].diff().plot()
plt.legend((['USA Infection Rate']), loc='upper left')
```

Out[351]: <matplotlib.legend.Legend at 0x26f8a393430>



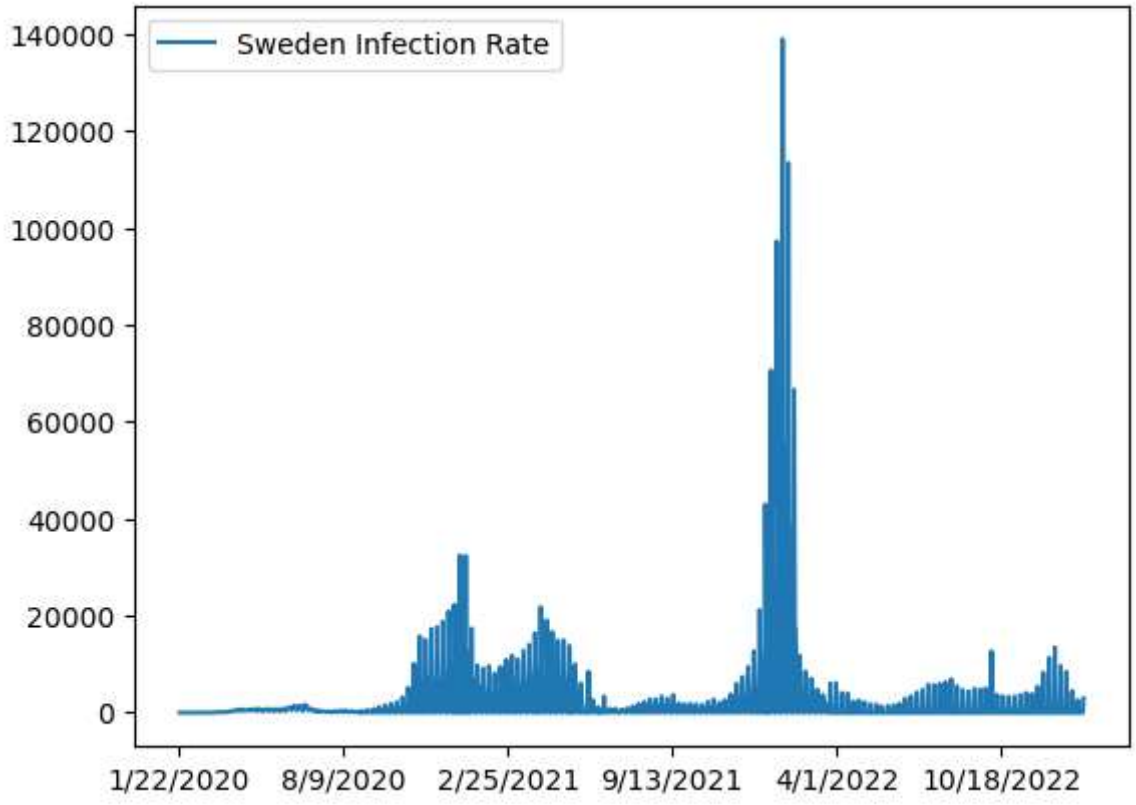
In [352…
```
corona_deaths_agg.loc["US"].diff().plot()
plt.legend(['USA Death Rate'], loc='upper left')
```

Out[352]:
```
<matplotlib.legend.Legend at 0x26f8a6b5100>
```



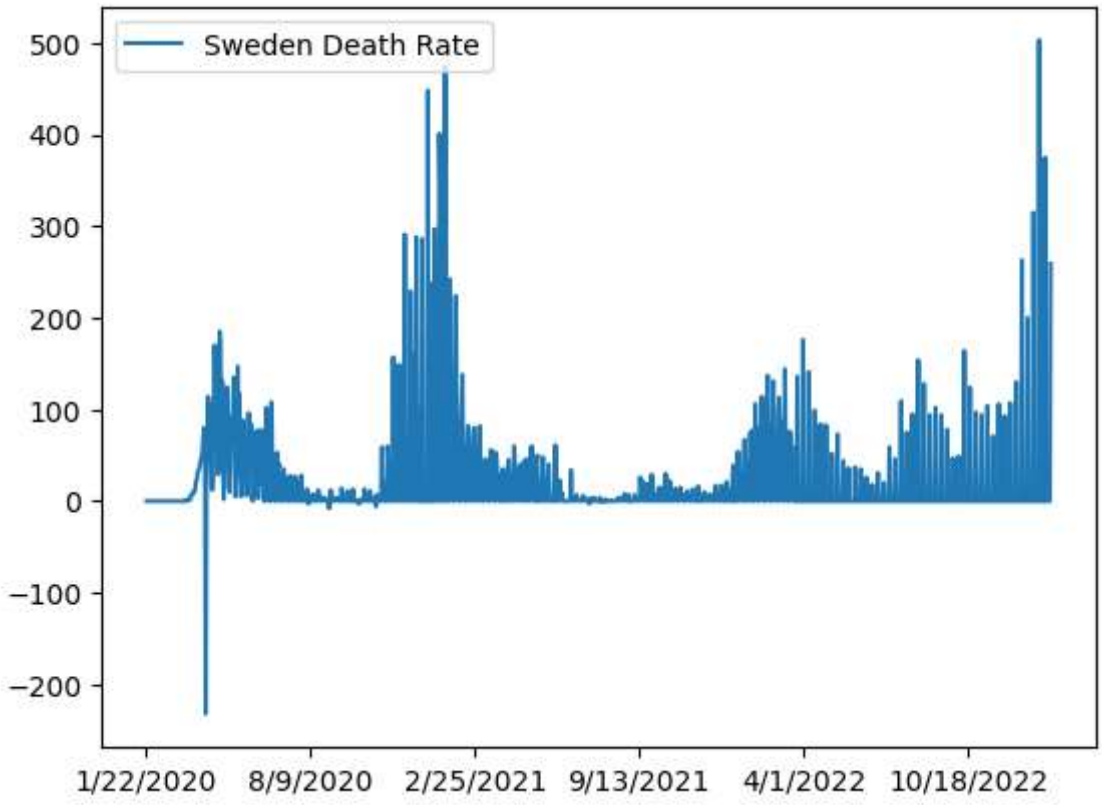
In [353…
```
corona_dataset_aggregated.loc["Sweden"].diff().plot()
plt.legend((['Sweden Infection Rate']), loc='upper left')
```

Out[353]:
```
<matplotlib.legend.Legend at 0x26f8a449520>
```



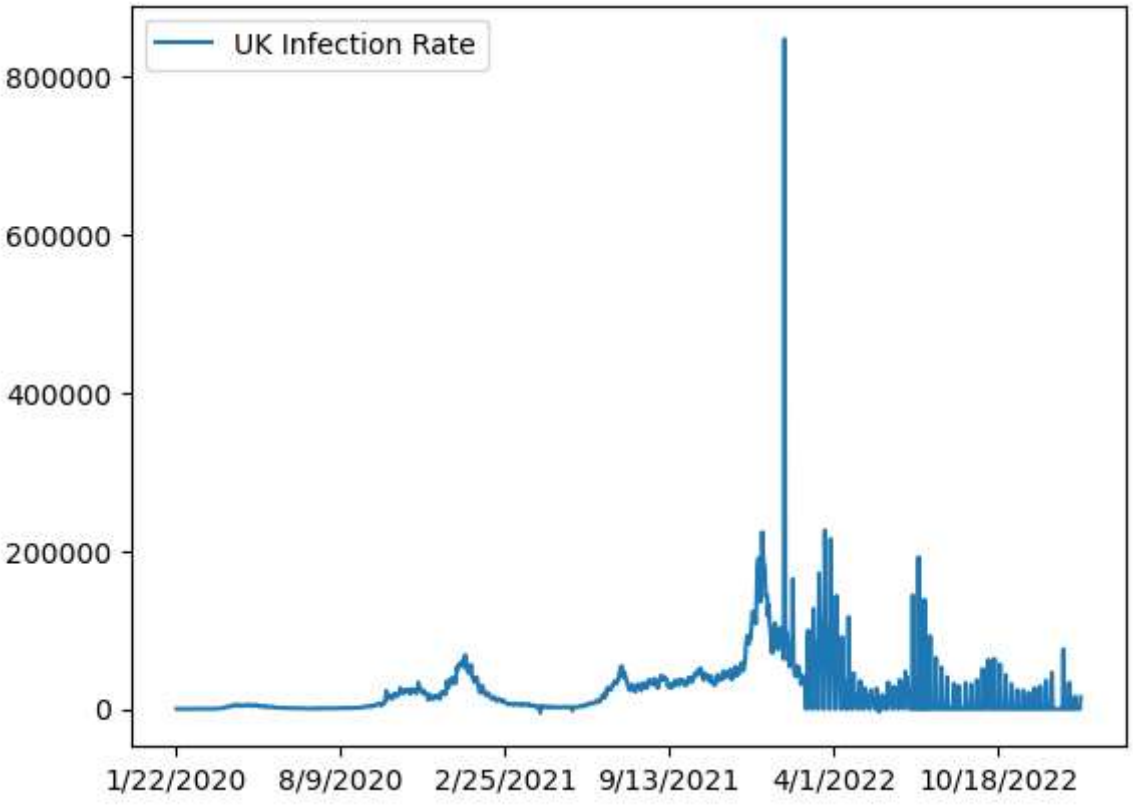
In [354…
```
corona_deaths_agg.loc["Sweden"].diff().plot()
plt.legend(['Sweden Death Rate'], loc='upper left')
```

Out[354]:
```
<matplotlib.legend.Legend at 0x26f8a3e83d0>
```



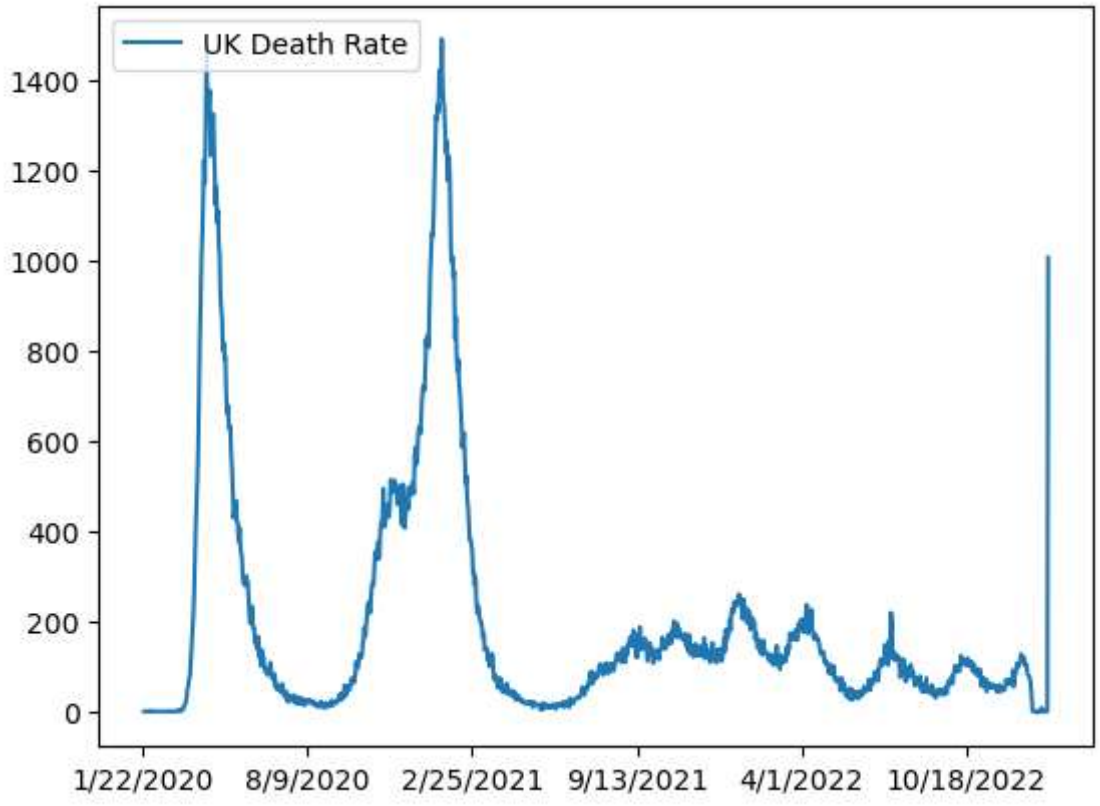
In [355…
```
corona_dataset_aggregated.loc["United Kingdom"].diff().plot()
plt.legend((['UK Infection Rate']), loc='upper left')
```

Out[355]:
```
<matplotlib.legend.Legend at 0x26f8a4cf6d0>
```



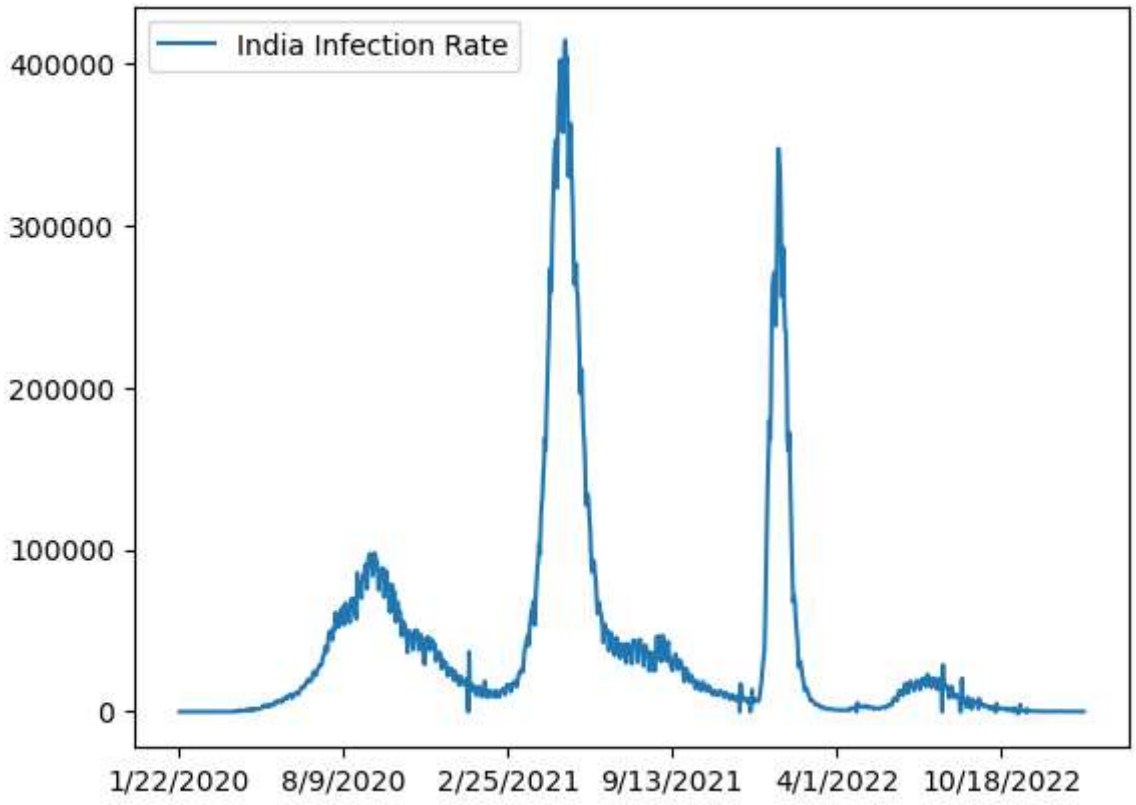
In [367…
```
corona_deaths_agg.loc["United Kingdom"].diff().plot()
plt.legend(['UK Death Rate'], loc='upper left')
```

Out[367]: <matplotlib.legend.Legend at 0x26f89c3ceb0>



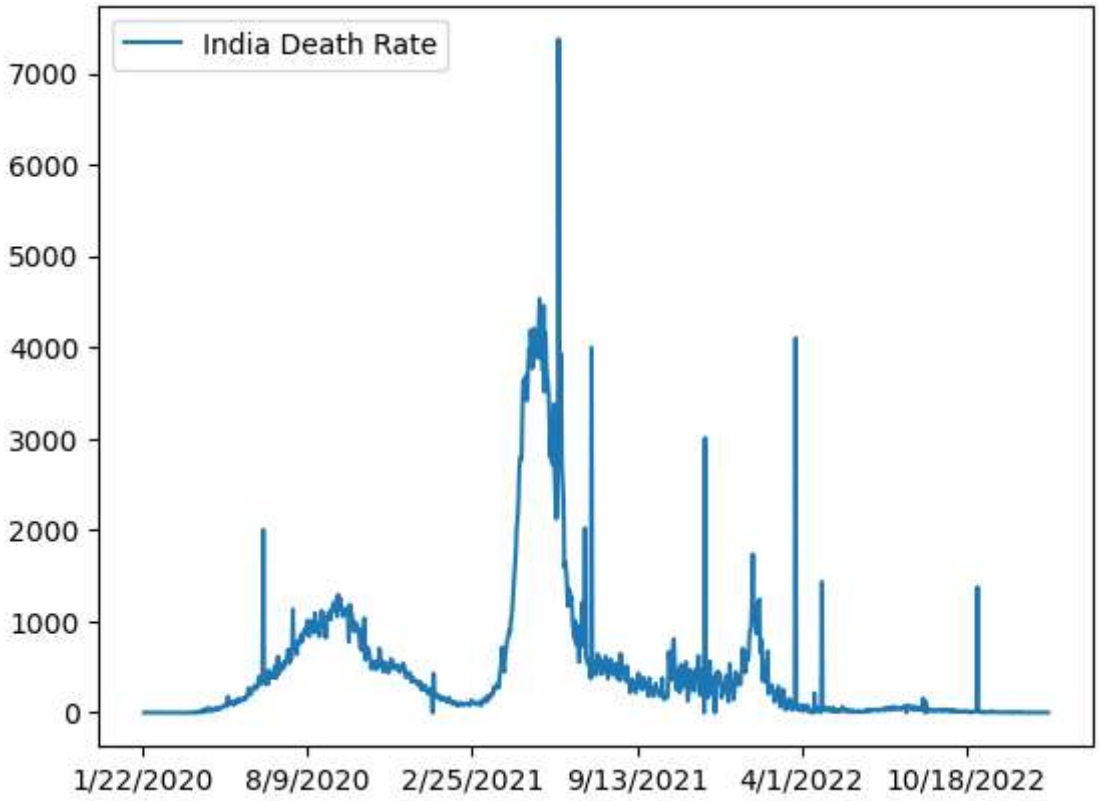
In [329…
```
corona_dataset_aggregated.loc["India"].diff().plot()
plt.legend((['India Infection Rate']), loc='upper left')
```

Out[329]: <matplotlib.legend.Legend at 0x26f89380160>



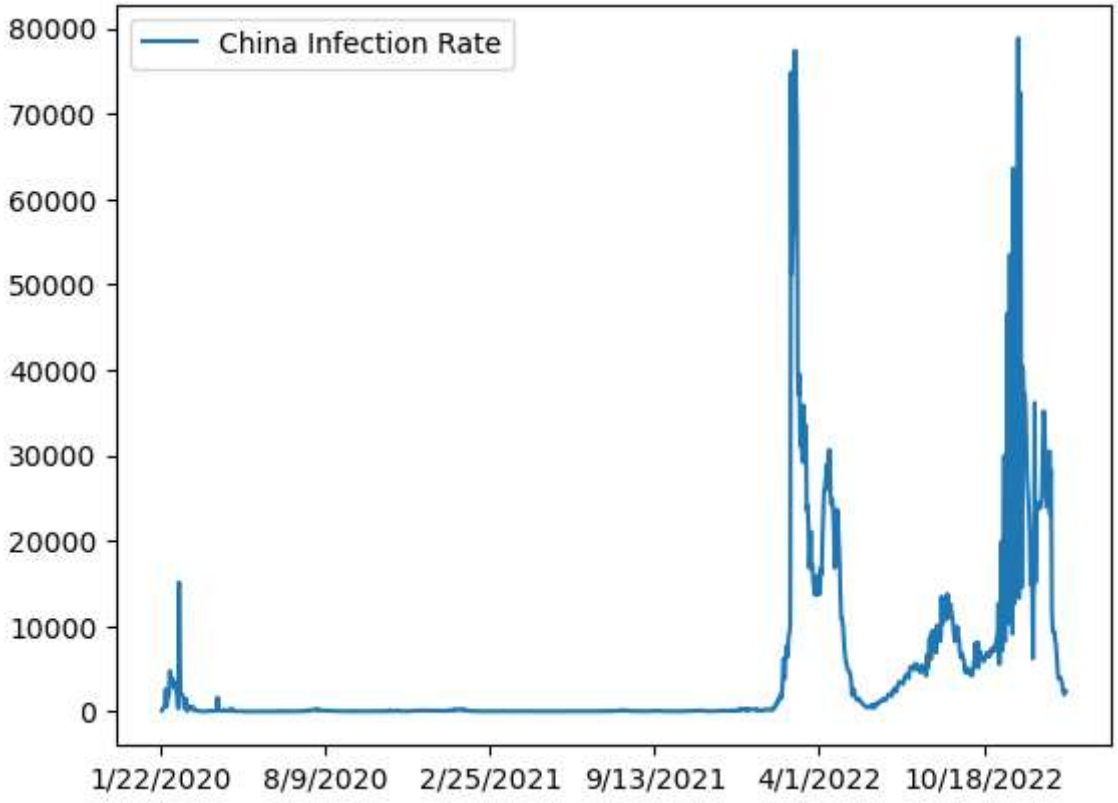
In [330…
```
corona_deaths_agg.loc["India"].diff().plot()
plt.legend(['India Death Rate'], loc='upper left')
```

Out[330]:
```
<matplotlib.legend.Legend at 0x26f892842b0>
```



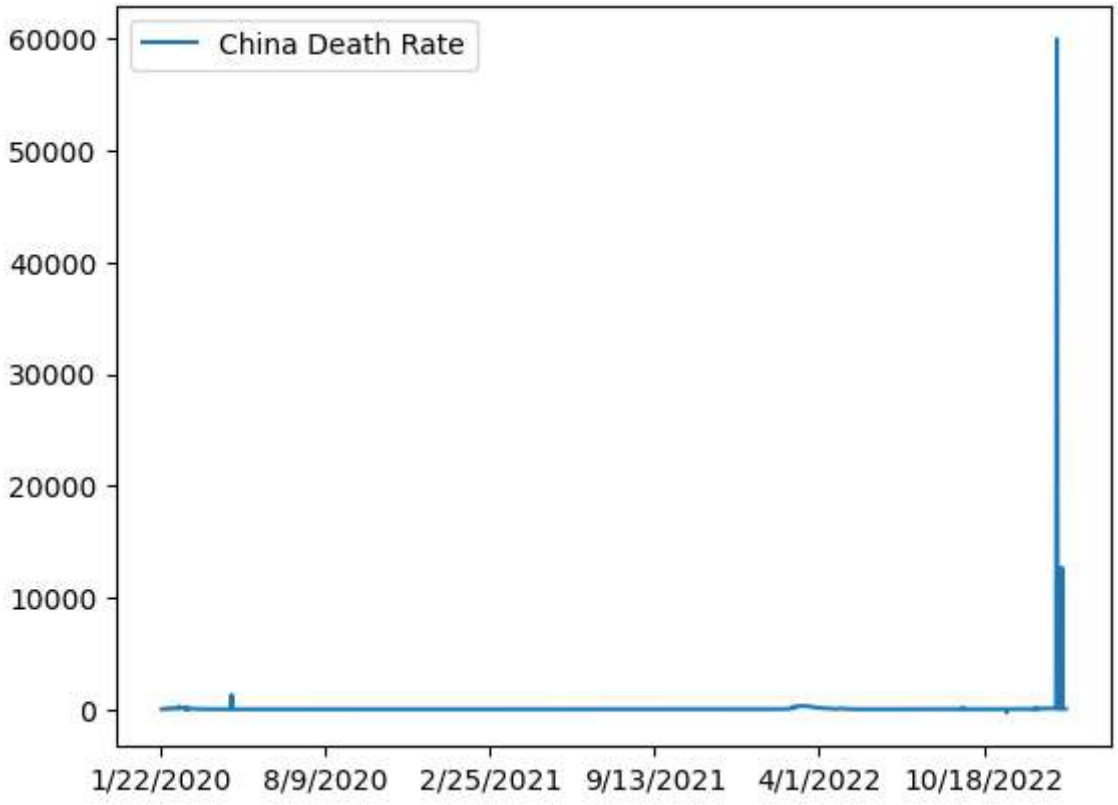
In [331…
```
corona_dataset_aggregated.loc["China"].diff().plot()
plt.legend((['China Infection Rate']), loc='upper left')
```

Out[331]:
```
<matplotlib.legend.Legend at 0x26f89605250>
```



In [332…
```
corona_deaths_agg.loc["China"].diff().plot()
plt.legend(['China Death Rate'], loc='upper left')
```

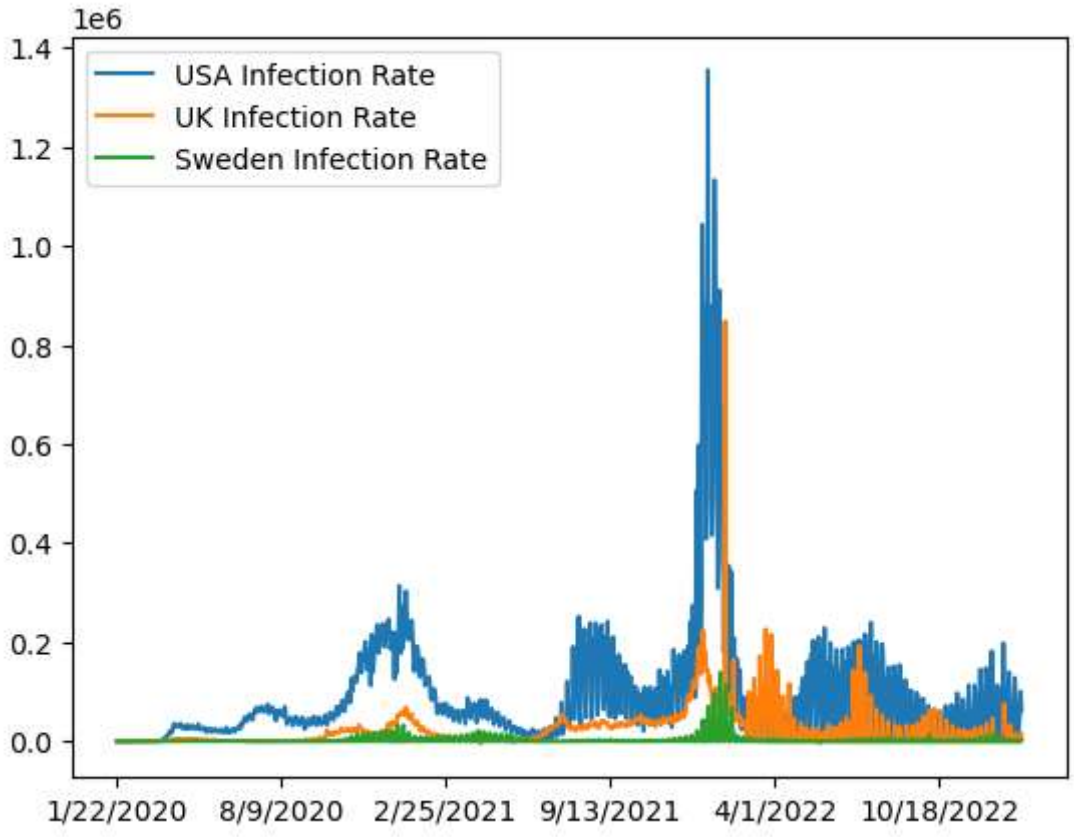
Out[332]:
```
<matplotlib.legend.Legend at 0x26f89dfdf70>
```



In [333…
```
corona_dataset_aggregated.loc["US"].diff().plot()
corona_dataset_aggregated.loc["United Kingdom"].diff().plot()
```

```
corona_dataset_aggregated.loc["Sweden"].diff().plot()
plt.legend((['Sweden Infection Rate']), loc='upper left')
plt.legend(['USA Infection Rate', 'UK Infection Rate', 'Sweden Infection Rate'], loc='
```

Out[333]: <matplotlib.legend.Legend at 0x26f89de2be0>

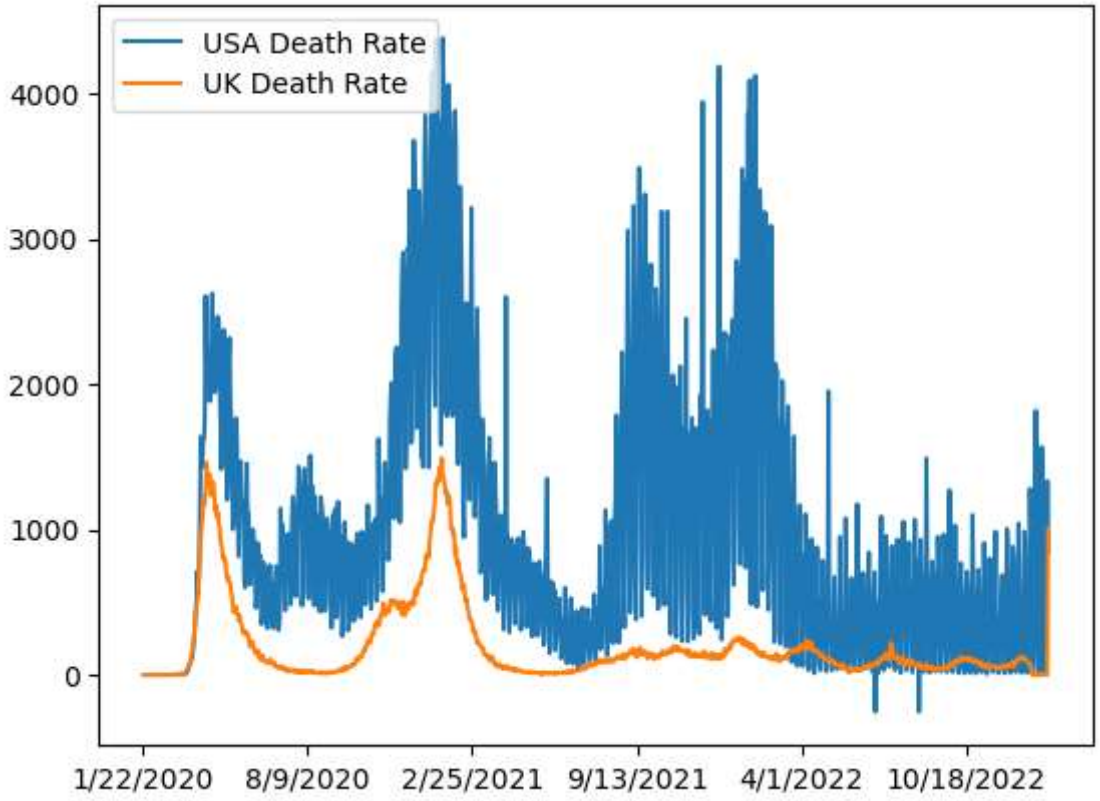


Covid death rate (new deaths within 24 hours)

In [334…

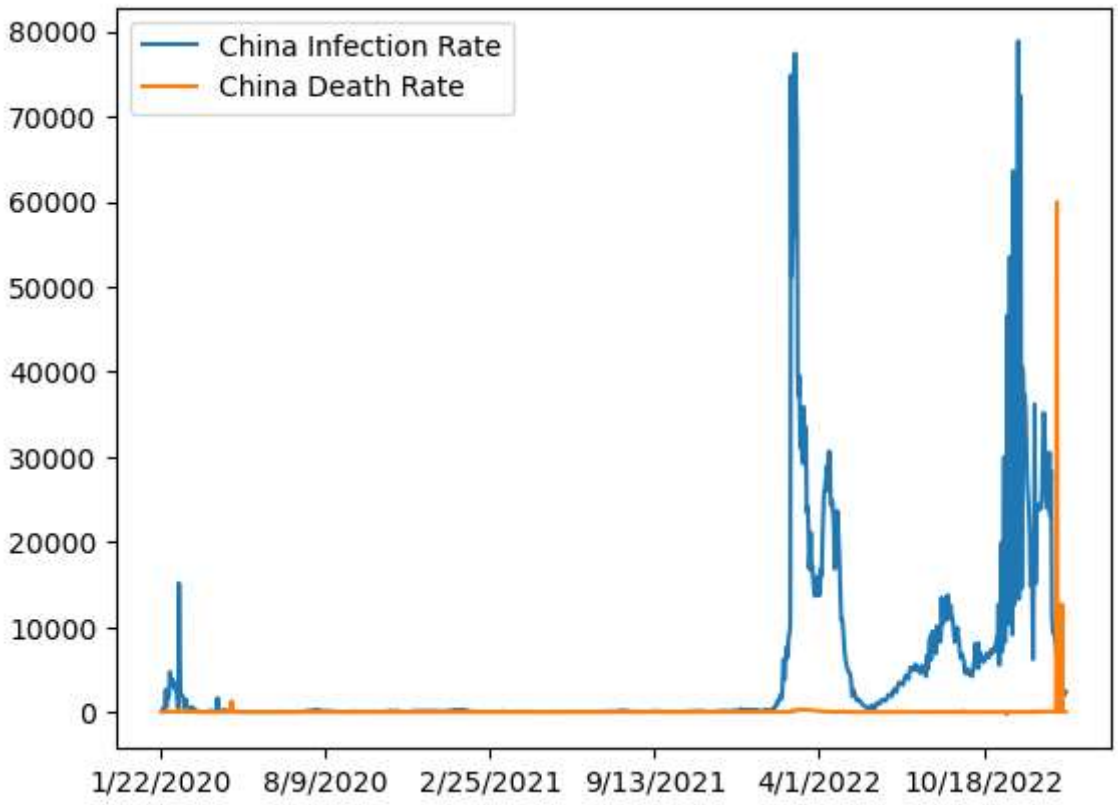
```
corona_deaths_agg.loc["US"].diff().plot()
corona_deaths_agg.loc["United Kingdom"].diff().plot()
plt.legend(['USA Death Rate', 'UK Death Rate'], loc='upper left')
```

Out[334]: <matplotlib.legend.Legend at 0x26f89b1b790>



In [335…
```
corona_dataset_aggregated.loc["China"].diff().plot()
corona_deaths_agg.loc["China"].diff().plot()
plt.legend(['China Infection Rate', 'China Death Rate'], loc='upper left')
```

Out[335]:
```
<matplotlib.legend.Legend at 0x26f89bdaac0>
```



Finding the average cumulative death rate.

In [303…
```
US_Death_Rate = corona_deaths_agg.loc["US"].max() / corona_dataset_aggregated.loc["US"

print("UsA Death Rate", US_Death_Rate)
```
```
UsA Death Rate 0.010830736866542658
```

In [302…
```
UK_Death_Rate = corona_deaths_agg.loc["United Kingdom"].max() / corona_dataset_aggrega

print("UK Death Rate", UK_Death_Rate)
```
```
UK Death Rate 0.008896404865132917
```

In [301…
```
Sweden_Death_Rate = corona_deaths_agg.loc["Sweden"].max() / corona_dataset_aggregated.

print("Sweden Death Rate", Sweden_Death_Rate)
```
```
Sweden Death Rate 0.008642793019234011
```

In [304…
```
India_Death_Rate = corona_deaths_agg.loc["India"].max() / corona_dataset_aggregated.lo

print("India Death Rate", India_Death_Rate)
```
```
India Death Rate 0.011877697308642637
```

In [305…
```
China_Death_Rate = corona_deaths_agg.loc["China"].max() / corona_dataset_aggregated.lo

print("UsA Death Rate", US_Death_Rate)
```
```
UsA Death Rate 0.010830736866542658
```

Finding max infection rate

In [220…
```
corona_dataset_aggregated.loc["China"].diff().max()
```
Out[220]:
```
78859.0
```

In [221…
```
corona_dataset_aggregated.loc["Italy"].diff().max()
```
Out[221]:
```
229122.0
```

In [222…
```
corona_dataset_aggregated.loc["Spain"].diff().max()
```
Out[222]:
```
372766.0
```

In [223…
```
corona_dataset_aggregated.loc["US"].diff().max()
```
Out[223]:
```
1354499.0
```

Finding max death rate

In [224…
```
corona_deaths_agg.loc["US"].diff().max()
```
Out[224]:
```
4381.0
```

In [225…
```
corona_deaths_agg.loc["United Kingdom"].diff().max()
```
Out[225]:
```
1492.0
```

In [226…
```
corona_deaths_agg.loc["China"].diff().max()
```
Out[226]:
```
59961.0
```

In [227…
```
corona_deaths_agg.loc["India"].diff().max()
```
Out[227]:
```
7374.0
```

Preparing Covid Data for World Happiness Analysis

Using max infection rate per country to populate into a list and adding the data as column in dataset

In [228…
```
# Creating countries variable and casting index (country) to a list
countries = list(corona_dataset_aggregated.index)
# creating empty list for max infection rate data
max_infection_rate = []
# running loop, every iteration will populate the list with max infection spike for ea
for c in countries :
    max_infection_rate.append(corona_dataset_aggregated.loc[c].diff().max())
# adding the data as a column
corona_dataset_aggregated["max infection rate"] = max_infection_rate
```

In [229…
```
corona_dataset_aggregated.head()
```

Out[229]:

Country/Region	1/22/2020	1/23/2020	1/24/2020	1/25/2020	1/26/2020	1/27/2020	1/28/2020	1/
Afghanistan	0	0	0	0	0	0	0	
Albania	0	0	0	0	0	0	0	
Algeria	0	0	0	0	0	0	0	
Andorra	0	0	0	0	0	0	0	
Angola	0	0	0	0	0	0	0	

5 rows × 1102 columns

Using max death rate per country to populate into a list and adding the data as column in dataset

In [230…
```
countries = list(corona_deaths_agg.index)
max_death_rate = []
for c in countries :
    max_death_rate.append(corona_deaths_agg.loc[c].diff().max())
corona_deaths_agg["max death rate"] = max_death_rate
```

In [231…
```
corona_deaths_agg.head()
```

Out[231]:

Country/Region	1/22/2020	1/23/2020	1/24/2020	1/25/2020	1/26/2020	1/27/2020	1/28/2020	1/
Afghanistan	0	0	0	0	0	0	0	
Albania	0	0	0	0	0	0	0	
Algeria	0	0	0	0	0	0	0	
Andorra	0	0	0	0	0	0	0	
Angola	0	0	0	0	0	0	0	

5 rows × 1102 columns

Creating new dataframes for infection rate and death rate containing countries and corresponding rate.

In [232…
```
corona_data = pd.DataFrame(corona_dataset_aggregated['max infection rate'])
```

In [233…
```
corona_data.head()
```

Out[233]:

Country/Region	max infection rate
Afghanistan	3243.0
Albania	4789.0
Algeria	2521.0
Andorra	2313.0
Angola	5035.0

In [234…
```
covid_deaths = pd.DataFrame(corona_deaths_agg['max death rate'])
covid_deaths
```

Out[234]:

Country/Region	max death rate
Afghanistan	159.0
Albania	21.0
Algeria	49.0
Andorra	7.0
Angola	30.0
…	…
West Bank and Gaza	268.0
Winter Olympics 2022	0.0
Yemen	60.0
Zambia	72.0
Zimbabwe	107.0

201 rows × 1 columns

Preparing World Happiness Data for Analysis

Importing World Happiness Report data

In [235…
```
happiness_report_csv = pd.read_csv("Datasets/worldwide_happiness_report.csv")
```

In [236…
```
happiness_report_csv.head()
```

Out[236]:

	Country name	Regional indicator	Ladder score	Standard error of ladder score	upperwhisker	lowerwhisker	Logged GDP per capita	Social support	Healt li expectan
0	Finland	Western Europe	7.842	0.032	7.904	7.780	10.775	0.954	72
1	Denmark	Western Europe	7.620	0.035	7.687	7.552	10.933	0.954	72
2	Switzerland	Western Europe	7.571	0.036	7.643	7.500	11.117	0.942	74
3	Iceland	Western Europe	7.554	0.059	7.670	7.438	10.878	0.983	73
4	Netherlands	Western Europe	7.464	0.027	7.518	7.410	10.932	0.942	72

In [237…
```
happiness_report_csv.shape
```

Out[237]:
```
(149, 20)
```

Dropping unecessary columns

In [238…
```
uneeded_col = ["Regional indicator", "Standard error of ladder score", "upperwhisker",
happiness_report_csv.drop(uneeded_col, axis=1 , inplace=True)
```

In [239…
```
happiness_report_csv.head()
```

Out[239]:

	Country name	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices
0	Finland	10.775	0.954	72.0	0.949
1	Denmark	10.933	0.954	72.7	0.946
2	Switzerland	11.117	0.942	74.4	0.919
3	Iceland	10.878	0.983	73.0	0.955
4	Netherlands	10.932	0.942	72.4	0.913

Assigning country name as indices (primary key) for joining

In [240…
```
happiness_report_csv.set_index("Country name", inplace=True)
```

In [241…
```
happiness_report_csv.head()
```

Out[241]:

Country name	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices
Finland	10.775	0.954	72.0	0.949
Denmark	10.933	0.954	72.7	0.946
Switzerland	11.117	0.942	74.4	0.919
Iceland	10.878	0.983	73.0	0.955
Netherlands	10.932	0.942	72.4	0.913

Joining World Happiness Data with Covid Infection Dataframe

In [242…
```
corona_data.head()
```

Out[242]:

Country/Region	max infection rate
Afghanistan	3243.0
Albania	4789.0
Algeria	2521.0
Andorra	2313.0
Angola	5035.0

In [243…
```
corona_data.shape
```

Out[243]:
```
(201, 1)
```

In [244…
```
happiness_report_csv.shape
```

Out[244]:
```
(149, 4)
```

Due to the difference in structure I will use an inner join

In [245…
```
data = happiness_report_csv.join(corona_data,how="inner")
data.head()
```

Out[245]:

	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	max infection rate
Finland	10.775	0.954	72.0	0.949	50889.0
Denmark	10.933	0.954	72.7	0.946	55709.0
Switzerland	11.117	0.942	74.4	0.919	89462.0
Iceland	10.878	0.983	73.0	0.955	7408.0
Netherlands	10.932	0.942	72.4	0.913	380498.0

Looking at correlation matrix

In [246…

```
data.corr()
```

Out[246]:

	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	max infection rate
Logged GDP per capita	1.000000	0.798452	0.872090	0.456510	0.322941
Social support	0.798452	1.000000	0.748850	0.486266	0.236806
Healthy life expectancy	0.872090	0.748850	1.000000	0.493911	0.344605
Freedom to make life choices	0.456510	0.486266	0.493911	1.000000	0.052039
max infection rate	0.322941	0.236806	0.344605	0.052039	1.000000

Visualizations of Results

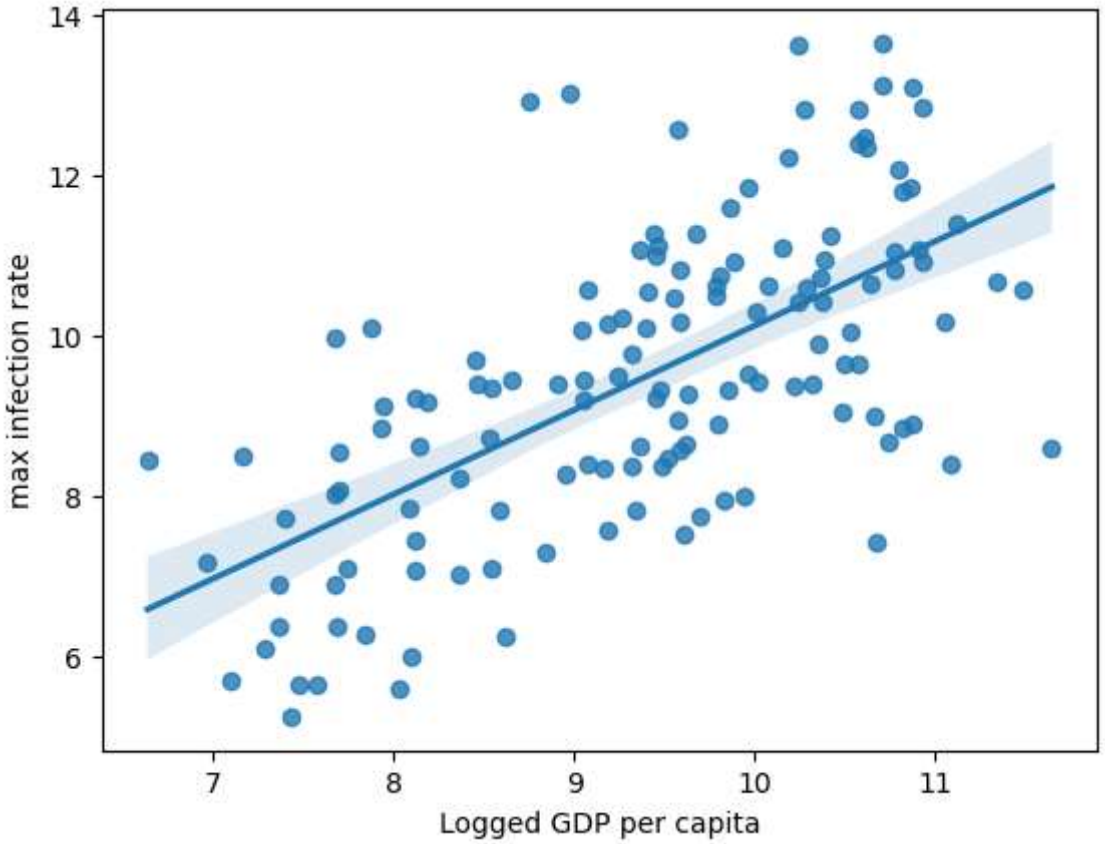
Maximum Infection Rate vs GDP

In [247…

```
x = data["Logged GDP per capita"]
y = data["max infection rate"]
#sns.scatterplot(x = x, y = np.log(y))
sns.regplot(x = x, y = np.log(y))
```

Out[247]:

```
<AxesSubplot:xlabel='Logged GDP per capita', ylabel='max infection rate'>
```



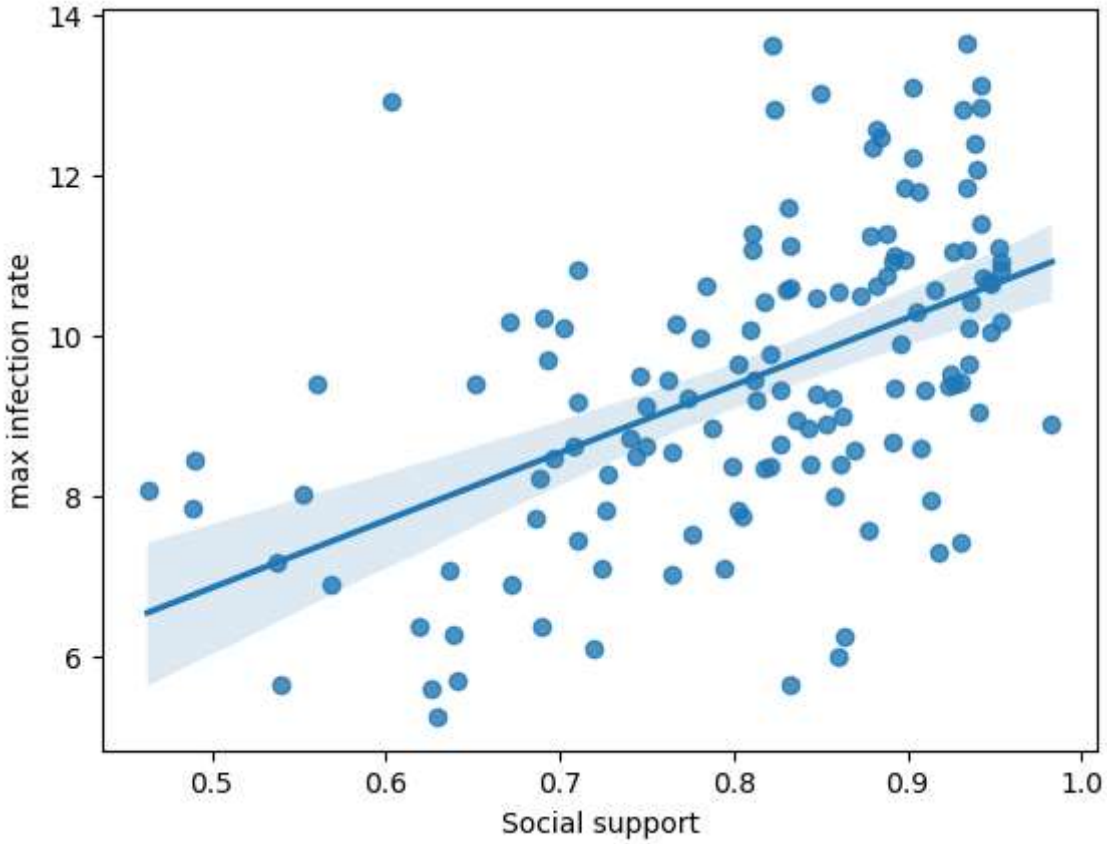
Maximum Infection rate vs Social Support

In [248…

```
x = data["Social support"]
y = data["max infection rate"]
#sns.scatterplot(x = x, y = np.log(y))
sns.regplot(x = x, y = np.log(y))
```

Out[248]:

```
<AxesSubplot:xlabel='Social support', ylabel='max infection rate'>
```



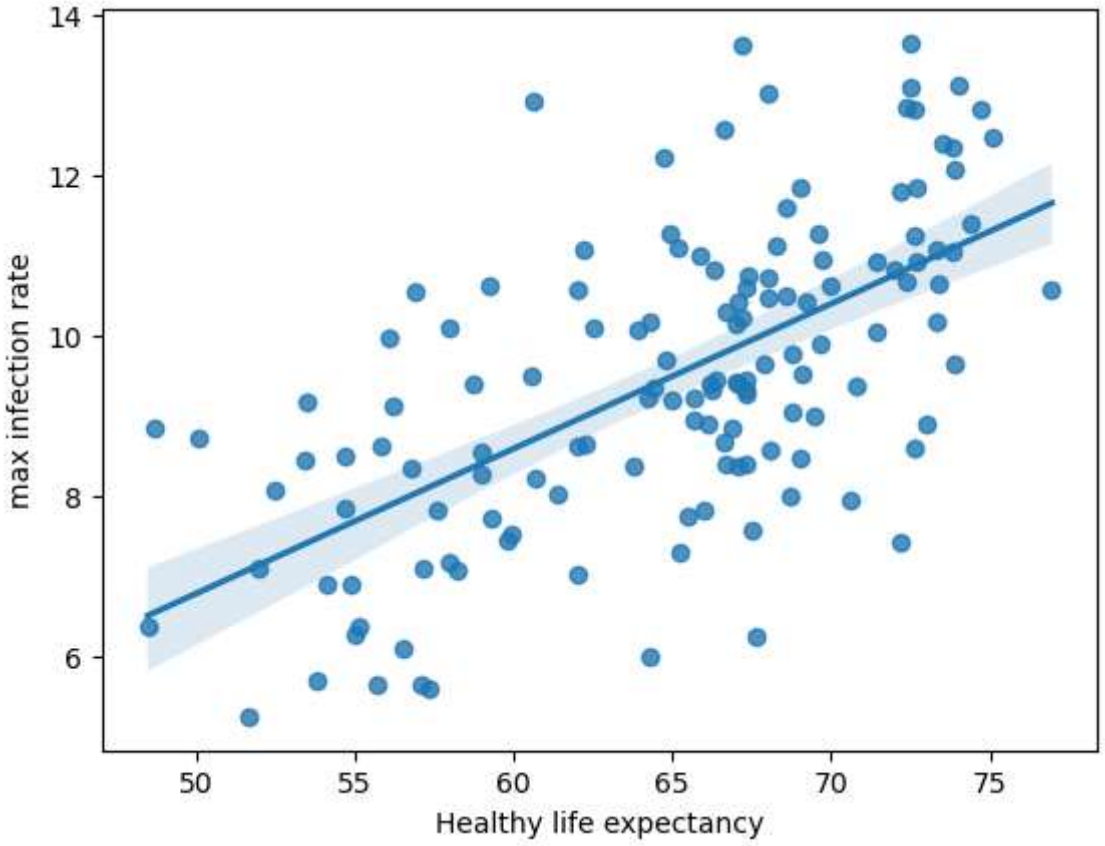
Maximum Infection Rate vs Healthy Life Expectancy

In [249…

```
x = data["Healthy life expectancy"]
y = data["max infection rate"]
#sns.scatterplot(x = x, y = np.log(y))
sns.regplot(x = x, y = np.log(y))
```

Out[249]:

```
<AxesSubplot:xlabel='Healthy life expectancy', ylabel='max infection rate'>
```



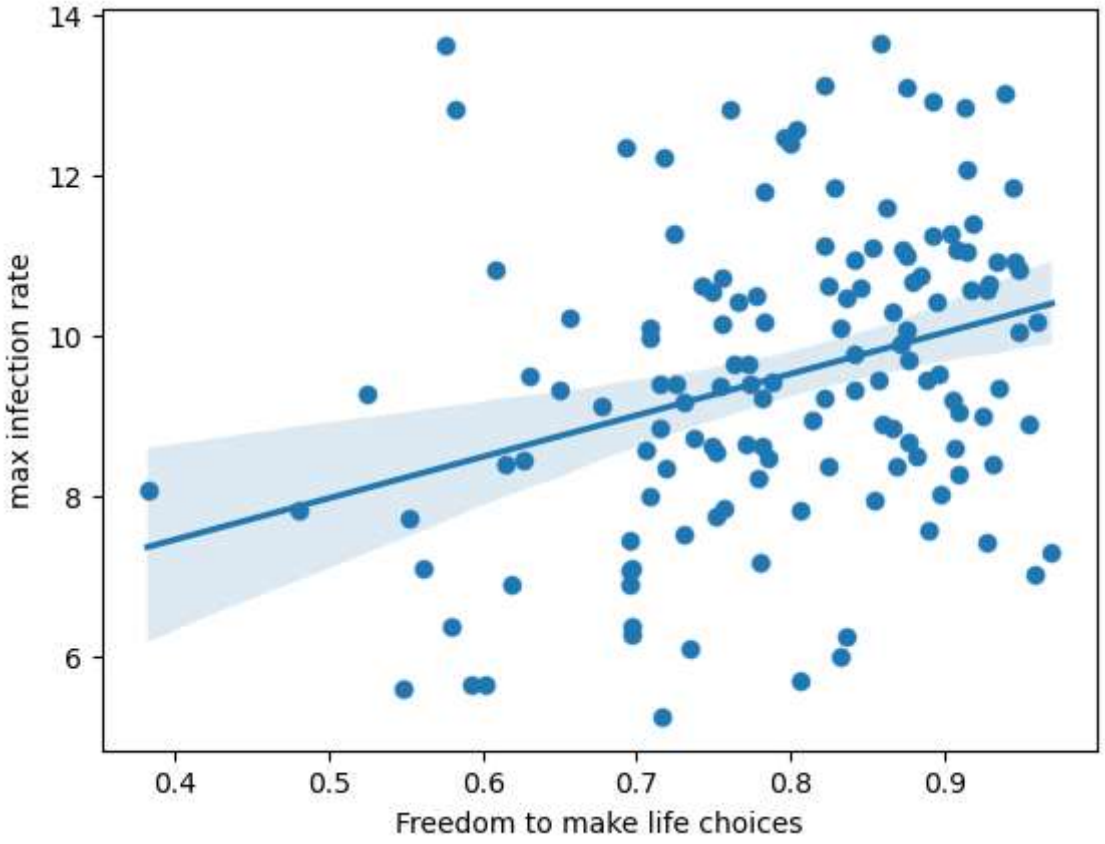
Maximum Infection Rate vs Freedome to Make Life Choices

In [250…

```
x = data["Freedom to make life choices"]
y = data["max infection rate"]
sns.scatterplot(x = x, y = np.log(y))
sns.regplot(x = x, y = np.log(y))
```

Out[250]:

```
<AxesSubplot:xlabel='Freedom to make life choices', ylabel='max infection rate'>
```



Analyzing Covid Deaths and World Happiness

Joining covid death dataframe and world happiness data

In [251…
```
data2 = happiness_report_csv.join(covid_deaths, how = "inner")
data2.head()
```

Out[251]:

	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	max death rate
Finland	10.775	0.954	72.0	0.949	330.0
Denmark	10.933	0.954	72.7	0.946	90.0
Switzerland	11.117	0.942	74.4	0.919	131.0
Iceland	10.878	0.983	73.0	0.955	34.0
Netherlands	10.932	0.942	72.4	0.913	234.0

Looking at correlation matrix

In [252…
```
data2.corr()
```

Out[252]:

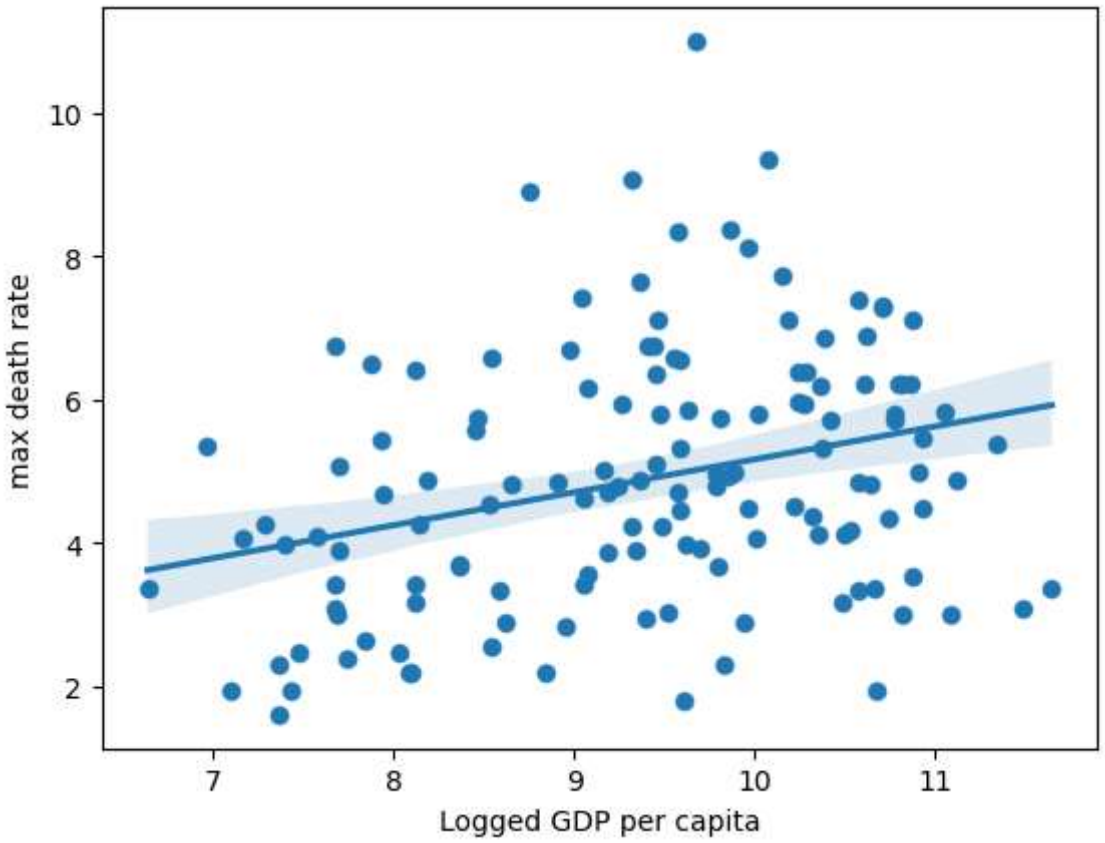
	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	max death rate
Logged GDP per capita	1.000000	0.798452	0.872090	0.456510	0.045441
Social support	0.798452	1.000000	0.748850	0.486266	0.018736
Healthy life expectancy	0.872090	0.748850	1.000000	0.493911	0.094749
Freedom to make life choices	0.456510	0.486266	0.493911	1.000000	0.099712
max death rate	0.045441	0.018736	0.094749	0.099712	1.000000

Visualizations of Results

Max Death Rate vs GDP

In [253…
```
x = data2["Logged GDP per capita"]
y = data2["max death rate"]
sns.scatterplot(x = x, y = np.log(y))
sns.regplot(x = x, y = np.log(y))
```

Out[253]:
```
<AxesSubplot:xlabel='Logged GDP per capita', ylabel='max death rate'>
```



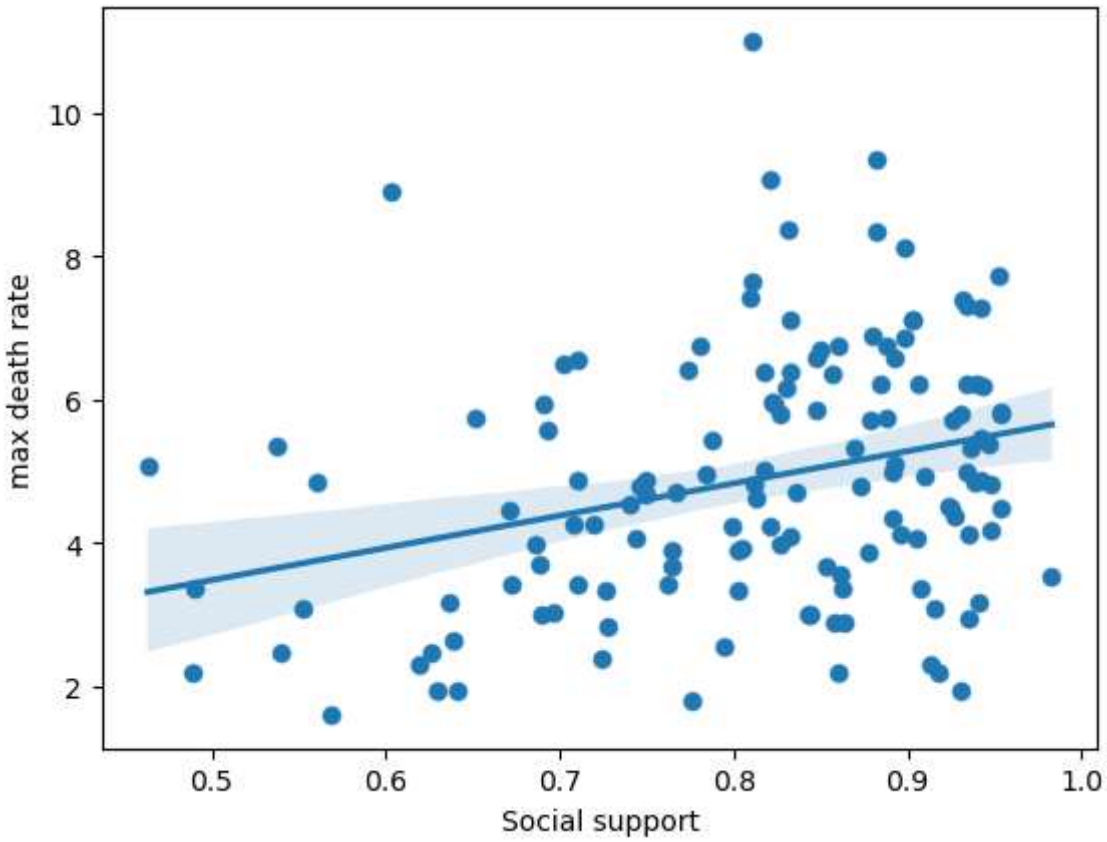
Max Death Rate vs Social Support

In [254…

```
x = data2["Social support"]
y = data2["max death rate"]
sns.scatterplot(x = x, y = np.log(y))
sns.regplot(x = x, y = np.log(y))
```

Out[254]:

```
<AxesSubplot:xlabel='Social support', ylabel='max death rate'>
```



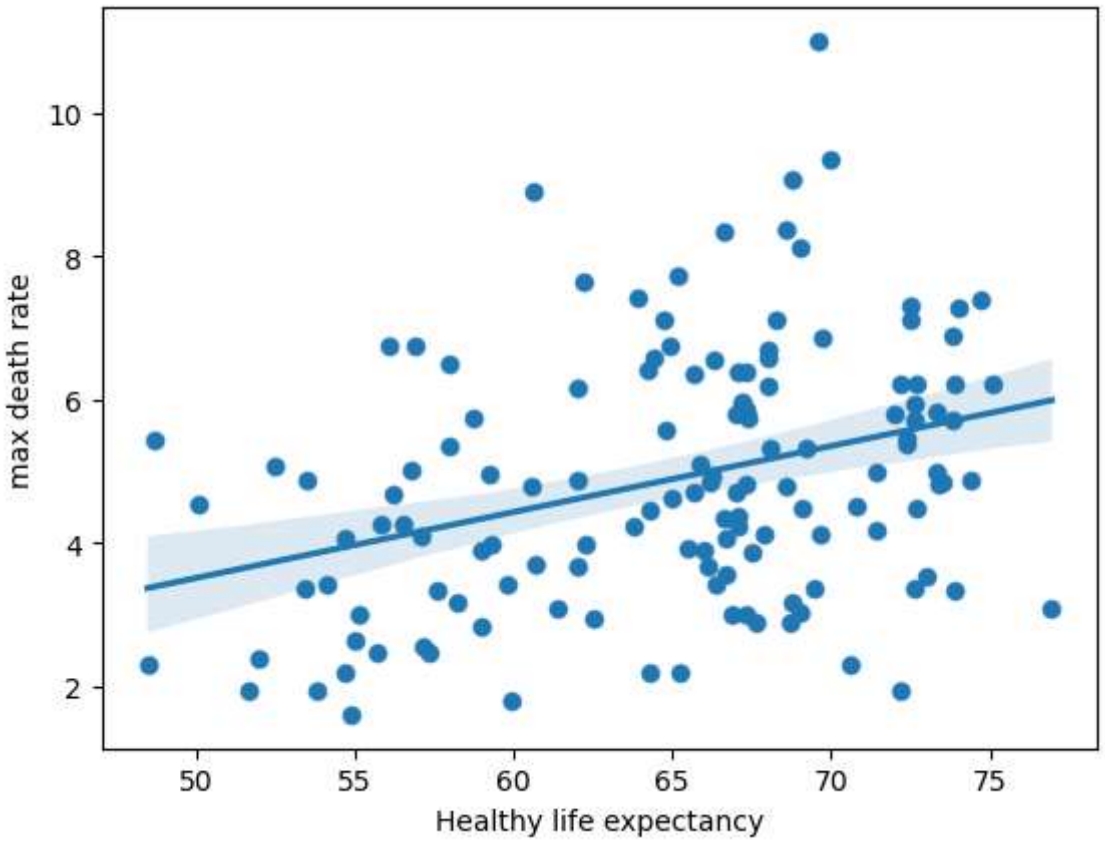
Max Death Rate vs Healthy Life Expectancy

In [255…

```
x = data2["Healthy life expectancy"]
y = data2["max death rate"]
sns.scatterplot(x = x, y = np.log(y))
sns.regplot(x = x, y = np.log(y))
```

Out[255]:

```
<AxesSubplot:xlabel='Healthy life expectancy', ylabel='max death rate'>
```



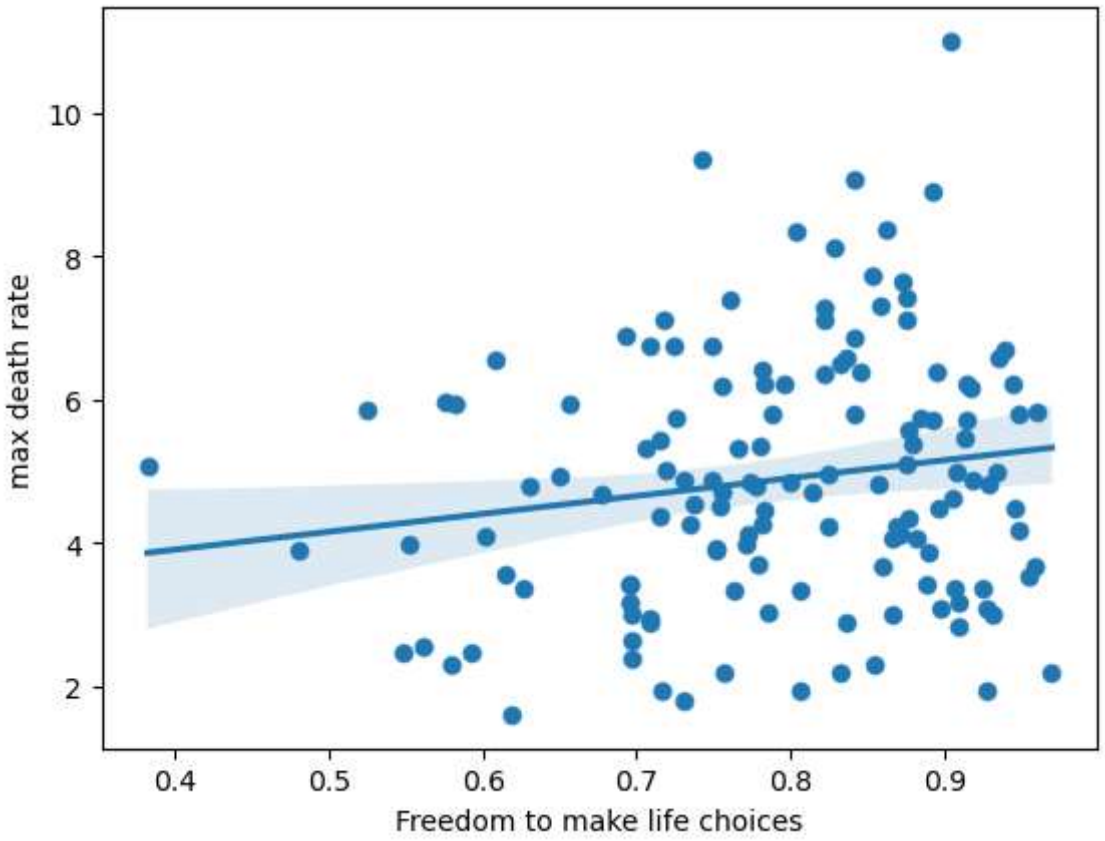
Max Death Rate vs Freedome to Make Life Choices

In [256…

```
x = data2["Freedom to make life choices"]
y = data2["max death rate"]
sns.scatterplot(x = x, y = np.log(y))
sns.regplot(x = x, y = np.log(y))
```

Out[256]:

```
<AxesSubplot:xlabel='Freedom to make life choices', ylabel='max death rate'>
```



Comparing correlation matrices between Covid infections vs World Happiness Report and Covid deaths vs World Happiness Reports

In [257…
```
data.corr()
```

Out[257]:

	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	max infection rate
Logged GDP per capita	1.000000	0.798452	0.872090	0.456510	0.322941
Social support	0.798452	1.000000	0.748850	0.486266	0.236806
Healthy life expectancy	0.872090	0.748850	1.000000	0.493911	0.344605
Freedom to make life choices	0.456510	0.486266	0.493911	1.000000	0.052039
max infection rate	0.322941	0.236806	0.344605	0.052039	1.000000

In [258…
```
data2.corr()
```

Out[258]:

	Logged GDP per capita	Social support	Healthy life expectancy	Freedom to make life choices	max death rate
Logged GDP per capita	1.000000	0.798452	0.872090	0.456510	0.045441
Social support	0.798452	1.000000	0.748850	0.486266	0.018736
Healthy life expectancy	0.872090	0.748850	1.000000	0.493911	0.094749
Freedom to make life choices	0.456510	0.486266	0.493911	1.000000	0.099712
max death rate	0.045441	0.018736	0.094749	0.099712	1.000000

Looking at the corresponding p-value to identify statistical relevance p-value),

In [259…
```
x = data["Logged GDP per capita"]
y = data["max infection rate"]
print('GDP vs Infection: ', prs(x, y))
```

```
GDP vs Infection:  PearsonRResult(statistic=0.3229409946113675, pvalue=0.000111899195
45714305)
```

In [260…
```
x = data2["Logged GDP per capita"]
y = data2["max death rate"]
print('GDP vs Death: ', prs(x, y))
```

```
GDP vs Death:  PearsonRResult(statistic=0.04544131288022931, pvalue=0.596644160106528
4)
```

In [261…
```
x = data["Social support"]
y = data["max infection rate"]
print('Social support vs Infection: ', prs(x, y))
```

```
Social support vs Infection:  PearsonRResult(statistic=0.23680585282010683, pvalue=0.
005167144209386814)
```

In [262…
```
x = data2["Social support"]
y = data2["max death rate"]
print('Social support vs Death: ', prs(x, y))
```

```
Social support vs Death:  PearsonRResult(statistic=0.018735720590074174, pvalue=0.827
3418054906269)
```

In [263…
```
x = data["Healthy life expectancy"]
y = data["max infection rate"]
print('Life Expectancy vs Infection: ', prs(x, y))
```

```
Life Expectancy vs Infection:  PearsonRResult(statistic=0.34460458042434017, pvalue=
3.490242154829875e-05)
```

In [264…
```
x = data2["Healthy life expectancy"]
y = data2["max death rate"]
```

```
print('Life Expectancy vs Death: ', prs(x, y))
```

```
Life Expectancy vs Death:  PearsonRResult(statistic=0.09474885413503667, pvalue=0.268
98122958103954)
```

In [265…

```
x = data["Freedom to make life choices"]
y = data["max infection rate"]
print('Choice Freedom vs Infection: ', prs(x, y))
```

```
Choice Freedom vs Infection:  PearsonRResult(statistic=0.05203898986738528, pvalue=0.
5444017960399976)
```

In [448…

```
x = data2["Freedom to make life choices"]
y = data2["max death rate"]
print('Choice Freedom vs Death: ', prs(x, y))
```

```
Choice Freedom vs Death:  PearsonRResult(statistic=0.09971213335953946, pvalue=0.2445
8606247896078)
```

Summary of analysis

Covid death and infection rates along the time-series between the United Kingdom, USA, and Sweden are positively correlated. Peaks and troughs occur in tandem with the seasonality of covid for all three countries. This may be due to the seasonality of the covid virus and the climate similarities of the United Kingdom, Sweden and much of the continental USA, better access and resources for recording and reporting the data, and transparency in reporting cannot be ruled out. We also see a consistent correlation and pattern between spikes in infection rate and death rates. This pattern is also consistent with India, however, not as consistent. There are outliers where some surges in death rates do not have a the corresponding surge in infection rates. This could be due to many non-lethal covid infections going undiagnosed. China's infection rate and death rate do not follow the trends seen in the other four countries. In March and April of 2022 China experienced a large increase in infection rate but there was no significant increase in the death rate. We also do not see the seasonality displayed in the other four countries. From early 2020 to late 2021 infection rates stay very close to the y-axis, virtually no infections. The death rate displayed the same trend but remained unchanged until late 2022. At the end of 2022 and into 2023 China's death rate climbed to about 60,000 in a very short period of time. It does correlate with a spike in infection rates; however, the infection rate peaked at about 80,000. This translates to roughly a 75% death rate over this period. This may be due to China's zero covid policy and strict regulations from 2020 through 2022. Civil unrest regarding China's strict regulations started in November 2022, large numbers of people gathered in protest and could have spread covid at an unprecedented rate. China also has an aging population, in 2018 more than 17% of the population was reported to be over 60 years old. Considering all variables with health facilities past capacity, the country experiencing its largest surge in infections since the outbreak, a large percent of the population being more at risk of serious complications may have created the massive surge in death rate while leaving many nonlethal cases going unreported. This would greatly skew the calculated death rate. The cumulative average death rate for every country is within the expected range, varying from 0.87% - 1.08%. China is anomalous in comparison to the other countries. There are

varying reasons and explanations as to why, regardless of the reason they are an outlier. The covid data in relation to the World Health report initially shows a positive correlation in varying degrees across all categories. For Max Infection vs GDP we see a low moderate correlation and a high statistical relevance in the p-value. However, in the Max Death Rate vs GDP the correlation is weak, and the statistical relevance is also weak. Social support vs infection rate shows a weak correlation with significant statistical relevance. Once again when social support is compared to death rate both the correlation and statistical relevance are weak. This trend continues for healthy life choices and freedom to make life choices, correlation and the statistical relevance is weak when compared to the death rate. The difference between correlation and statistical relevance of infection rate and death rate against World Health Report data could be unrecorded/unreported infections. Infections that lead to death are more likely to be recorded due to the severity of the case leading to a higher necessity of hospitalization or medical help.