

In [2]: %run Imports.ipynb

## Introduction

In this report, the pandemic test, case, recovery, death, and vaccination totals are covered. This report uses time-series modeling to determine the progression of the pandemic.

Dataset of country pandemic totals.

```
In [3]: df=Json('df1: Covid19 Totals')
df.web_scrape('https://www.worldometers.info/coronavirus/?referer=app', ['0', 'Country', 'Total_Cases', 'New_Cases', 'Total_Deaths', 'New_Deaths', 'Total_Recovered',
'1', 'Active_Cases', 'Serious_Critical', 'Tot_Cases_per_1m_Pop', 'Tot_Deaths_per_1m_Pop',
'Total_Tests', 'Tot_Tests_per_1m_Pop', 'Population', 'Unknown', 'Continent', 'Unknown', 'Unknown'])
df1=json_storage['df1: Covid19 Totals'][1]
df1
```

	0	Country	Total_Cases	New_Cases	Total_Deaths	New_Deaths	Total_Recovered	1	Active_Cases	Serious_Critical	Tot_Cases_per_1m_Pop	Tot_Deaths_per_1m_Pop	Total_Tests	Tot_Tests_per_1m_Pop	Popula		
0	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	N	
1	\nNorth America\n	35,865,701	+73,533	819,241	+1,910	27,575,037	+91,550	7,471,423	15,365								
2	\nSouth America\n	21,150,579	+118,206	552,021	+3,995	18,882,100	+116,117	1,716,458	21,109								
3	\nAsia\n	28,490,944	+182,032	428,059	+1,434	26,093,254	+112,944	1,969,631	25,421								
4	\nEurope\n	39,609,225	+223,855	913,171	+3,921	27,982,751	+117,852	10,713,303	31,419								
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
233	Total:	39,609,225	+223,855	913,171	+3,921	27,982,751	+117,852	10,713,303	31,419								
234	Total:	4,251,595	+12,213	113,036	+235	3,814,105	+8,308	324,454	3,037								
235	Total:	57,040	+400	1,142	+4	34,567	+21	21,331	6								
236	Total:	721		15		706		0	0								
237	Total:	129,425,805	+610,239	2,826,685	+11,499	104,382,520	+446,792	22,216,600	96,357		16,604.1		362.6				

238 rows × 19 columns

Dataset of USA state pandemic totals.

```
In [4]: df=Json('df2: USA Covid19 Totals')
df.web_scrape('https://www.worldometers.info/coronavirus/country/us/', ['0', 'Country', 'Total_Cases', 'New_Cases', 'Total_Deaths', 'New_Deaths', 'Total_Recovered',
'Active_Cases', 'Tot_Cases_per_1m_Pop', 'Tot_Deaths_per_1m_Pop',
'Total_Tests', 'Tot_Tests_per_1m_Pop', 'Population','1','1'])
df2=json_storage['df2: USA Covid19 Totals'][1]
df2
```

	0	Country	Total_Cases	New_Cases	Total_Deaths	New_Deaths	Total_Recovered	Active_Cases	Tot_Cases_per_1m_Pop	Tot_Deaths_per_1m_Pop	Total_Tests	Tot_Tests_per_1m_Pop	Population	1		
0	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None	None
1	USA Total	31,161,286	+63,698	565,193	+1,052	23,670,368	6,925,725		94,142		1,708	403,189,039		1,218,084		
2	1 \nCalifornia	3,668,451	\n+1,703	\n58,912	+123	\n1,940,386	\n1,669,153		92,843		1,491	\n53,786,487		1,361,262	\n39,512,223	\n[view by county] [1] \n

	0	Country	Total_Cases	New_Cases	Total_Deaths	New_Deaths	Total_Recovered	Active_Cases	Tot_Cases_per_1m_Pop	Tot_Deaths_per_1m_Pop	Total_Tests	Tot_Tests_per_1m_Pop	Population	1
3	2	\nTexas	2,792,858		\n48,479		\n2,646,224	\n98,155	96,319	1,672	\n25,536,604	880,698	\n28,995,881	\n[view by county] [1] [2] [3]
4	3	\nFlorida	2,057,735	\n+5,294	\n33,450	+84	\n1,520,607	\n503,678	95,808	1,557	\n25,289,334	1,177,467	\n21,477,737	\n[view by county] [1] [2] [3]
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
61	60	\nNavajo Nation	30,079		\n1,247		\n16,396	\n12,436			\n253,994\n			\n[1] [2] [3]
62	61	\nGrand Princess Ship	122		\n7		\n115	\n0			\n			\n[1]
63	62	\nWuhan Repatriated	3		\n		\n3	\n0			\n3\n			\n
64	63	\nDiamond Princess Ship	46		\n		\n46	\n0			\n46\n			\n
65		Total:	31,161,286	+63,698	565,193	+1,052	23,670,368	6,925,725	94,142	1,708	403,189,039	1,218,084		

66 rows × 15 columns

Dataset of country vaccination totals.

In [5]:

```
df=Json('df3: vaccinated_total')
df.csv('https://raw.githubusercontent.com/owid/covid-19-data/master/public/data/vaccinations/vaccinations.csv')
df3=json_storage['df3: vaccinated_total'][1]
df3
```

Out[5]:

	location	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hundred	people_fully_vaccinate	
0	Afghanistan	AFG	2021-02-22	0.0	0.0		NaN	NaN	NaN	0.00	0.00	
1	Afghanistan	AFG	2021-02-23		NaN	NaN	NaN	NaN	1367.0		NaN	
2	Afghanistan	AFG	2021-02-24		NaN	NaN	NaN	NaN	1367.0		NaN	
3	Afghanistan	AFG	2021-02-25		NaN	NaN	NaN	NaN	1367.0		NaN	
4	Afghanistan	AFG	2021-02-26		NaN	NaN	NaN	NaN	1367.0		NaN	
...	...	...	...	...	...	...	...	...	...	...	...	
9813	Zimbabwe	ZWE	2021-03-25	54892.0	54892.0		NaN	5488.0	2127.0	0.37	0.37	

	location	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hundred	people_fully_vaccinate
9814	Zimbabwe	ZWE	2021-03-26	61093.0	61093.0	NaN	6201.0	2772.0	0.41	0.41	
9815	Zimbabwe	ZWE	2021-03-27	65466.0	65466.0	NaN	4373.0	3322.0	0.44	0.44	
9816	Zimbabwe	ZWE	2021-03-28	68511.0	68511.0	NaN	3045.0	3683.0	0.46	0.46	
9817	Zimbabwe	ZWE	2021-03-29	69751.0	69751.0	NaN	1240.0	3780.0	0.47	0.47	

9818 rows × 12 columns

Dataset of USA state vaccinations.

```
In [6]: df=Json('df4: usa_vaccinated_total')
df.csv('vaccine_data.csv')
df4=json_storage['df4: usa_vaccinated_total'][1]
df4
```

	Location	Total COVID-19 Vaccines Delivered	Total COVID-19 Vaccines Administered	Share of Delivered Vaccines That Have Been Administered	Number of People Who Have Been Vaccinated	Share of Population Vaccinated
0	United States	154199235.0	118313818.0	0.76728	77050155.0	0.234738
1	Alabama	2141210.0	1426696.0	0.66630	936863.0	0.191072
2	Alaska	514835.0	356055.0	0.69159	214185.0	0.292784
3	Arizona	3444465.0	2760907.0	0.80155	1714739.0	0.235583
4	Arkansas	1407320.0	952919.0	0.67712	634546.0	0.210267
...	...	...	...	...	...	...
69	*Vaccines delivered* are cumulative counts of ...	NaN	NaN	NaN	NaN	NaN
70		NaN	NaN	NaN	NaN	NaN
71	*Vaccines administered* are cumulative counts ...	NaN	NaN	NaN	NaN	NaN
72		NaN	NaN	NaN	NaN	NaN
73	*Number of people who have been vaccinated* re...	NaN	NaN	NaN	NaN	NaN

74 rows × 6 columns

Dataset of country cases time-series.

```
In [7]: df=Json('df5: time_series')
df.csv('https://covid19.who.int/WHO-COVID-19-global-data.csv')
df5=json_storage['df5: time_series'][1]
df5
```

Out[7]:

	Date_reported	Country_code	Country	WHO_region	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths
0	2020-01-03	AF	Afghanistan	EMRO	0	0	0	0
1	2020-01-04	AF	Afghanistan	EMRO	0	0	0	0
2	2020-01-05	AF	Afghanistan	EMRO	0	0	0	0
3	2020-01-06	AF	Afghanistan	EMRO	0	0	0	0
4	2020-01-07	AF	Afghanistan	EMRO	0	0	0	0
...	...	...	...	...	...	...	...	...
107139	2021-03-27	ZW	Zimbabwe	AFRO	27	36805	0	1518
107140	2021-03-28	ZW	Zimbabwe	AFRO	13	36818	1	1519
107141	2021-03-29	ZW	Zimbabwe	AFRO	4	36822	1	1520
107142	2021-03-30	ZW	Zimbabwe	AFRO	17	36839	0	1520
107143	2021-03-31	ZW	Zimbabwe	AFRO	19	36858	1	1521

107144 rows × 8 columns

Dataset of USA state cases time-series.

```
In [8]: df=Json('df6: usa_time_series_cases_deaths')
df.csv('https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-states.csv')
df6=json_storage['df6: usa_time_series_cases_deaths'][1]
df6
```

Out[8]:

	date	state	fips	cases	deaths
0	2020-01-21	Washington	53	1	0
1	2020-01-22	Washington	53	1	0
2	2020-01-23	Washington	53	1	0
3	2020-01-24	Illinois	17	1	0
4	2020-01-24	Washington	53	1	0
...	...	...	...	...	...
21624	2021-03-30	Virginia	51	617941	10242
21625	2021-03-30	Washington	53	365816	5302
21626	2021-03-30	West Virginia	54	141322	2640
21627	2021-03-30	Wisconsin	55	635523	7293
21628	2021-03-30	Wyoming	56	56236	695

21629 rows × 5 columns

Dataset of government responses to covid 19.

## Data Mining

Country pandemic totals data information.

```
In [112]: df1.info(verbose=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 238 entries, 0 to 237
Data columns (total 19 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   0                237 non-null    object  
 1   Country          237 non-null    object  
 2   Total_Cases      237 non-null    object  
 3   New_Cases        237 non-null    object  
 4   Total_Deaths     237 non-null    object  
 5   New_Deaths       237 non-null    object  
 6   Total_Recovered  237 non-null    object  
 7   1                237 non-null    object  
 8   Active_Cases     237 non-null    object  
 9   Serious_Critical 237 non-null    object  
 10  Tot_Cases_per_1m_Pop 237 non-null    object  
 11  Tot_Deaths_per_1m_Pop 237 non-null    object  
 12  Total_Tests      237 non-null    object  
 13  Tot_Tests_per_1m_Pop 237 non-null    object  
 14  Population       237 non-null    object  
 15  Unknown          237 non-null    object  
 16  Continent        237 non-null    object  
 17  Unknown          237 non-null    object  
 18  Unknown          237 non-null    object  
dtypes: object(19)
memory usage: 35.5+ KB
```

USA state vaccination totals.

```
In [113]: df2.info(verbose=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 74 entries, 0 to 73
Data columns (total 6 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Location          70 non-null    object  
 1   Total COVID-19 Vaccines Delivered 60 non-null    float64 
 2   Total COVID-19 Vaccines Administered 60 non-null    float64 
 3   Share of Delivered Vaccines That Have Been Administered 60 non-null    float64 
 4   Number of People Who Have Been Vaccinated 60 non-null    float64 
 5   Share of Population Vaccinated      53 non-null    float64 
dtypes: float64(5), object(1)
memory usage: 3.6+ KB
```

Country vaccination time-series data information.

```
In [60]: df3.info(verbose=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6495 entries, 0 to 6494
Data columns (total 12 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   location         6495 non-null    object  
 1   iso_code          6131 non-null    object  
 2   date              6495 non-null    object  
 3   total_vaccinations 4229 non-null    float64 
 4   people_vaccinated 3782 non-null    float64 
 5   people_fully_vaccinated 2587 non-null    float64 
 6   daily_vaccinations_raw 3595 non-null    float64 
 7   daily_vaccinations 6299 non-null    float64
```

```
8 total_vaccinations_per_hundred      4229 non-null   float64
9 people_vaccinated_per_hundred     3782 non-null   float64
10 people_fully_vaccinated_per_hundred 2587 non-null   float64
11 daily_vaccinations_per_million    6299 non-null   float64
dtypes: float64(9), object(3)
memory usage: 609.0+ KB
```

Country cases time-series data information.

```
In [61]: df4.info(verbose=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 103306 entries, 0 to 103305
Data columns (total 8 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   Date_reported    103306 non-null   object  
 1   Country_code      102868 non-null   object  
 2   Country           103306 non-null   object  
 3   WHO_region        103306 non-null   object  
 4   New_cases         103306 non-null   int64  
 5   Cumulative_cases  103306 non-null   int64  
 6   New_deaths        103306 non-null   int64  
 7   Cumulative_deaths 103306 non-null   int64  
dtypes: int64(4), object(4)
memory usage: 6.3+ MB
```

USA state cases time-series data information.

```
In [29]: df7.info(verbose=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 20914 entries, 0 to 20913
Data columns (total 5 columns):
 #   Column   Non-Null Count  Dtype  
--- 
 0   date      20914 non-null   object  
 1   states    20914 non-null   object  
 2   cases     20914 non-null   int64  
 3   deaths    20914 non-null   int64  
 4   State      20914 non-null   object  
dtypes: int64(2), object(3)
memory usage: 817.1+ KB
```

USA state pandemic totals data information.

```
In [62]: df8.info(verbose=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 66 entries, 0 to 65
Data columns (total 15 columns):
 #   Column           Non-Null Count  Dtype  
--- 
 0   0                 65 non-null     object  
 1   Country          65 non-null     object  
 2   Total_Cases      65 non-null     object  
 3   New_Cases        65 non-null     object  
 4   Total_Deaths     65 non-null     object  
 5   New_Deaths       65 non-null     object  
 6   Total_Recovered   65 non-null     object  
 7   Active_Cases     65 non-null     object  
 8   Tot_Cases_per_1m_Pop 65 non-null     object  
 9   Tot_Deaths_per_1m_Pop 65 non-null     object  
 10  Total_Tests      65 non-null     object  
 11  Tot_Tests_per_1m_Pop 65 non-null     object
```

```
12 Population           65 non-null   object
13 1                   65 non-null   object
14 1                   64 non-null   object
dtypes: object(15)
memory usage: 7.9+ KB
```

Government responses to covid 19 data information.

```
In [63]: df9.info(verbose=True)
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 62700 entries, 0 to 62699
Data columns (total 43 columns):
 #   Column          Non-Null Count  Dtype  
--- 
 0   country         62700 non-null   object 
 1   geoid            62425 non-null   object 
 2   iso              62700 non-null   object 
 3   d                62700 non-null   object 
 4   cases             57750 non-null   float64
 5   deaths            57750 non-null   float64
 6   school            45758 non-null   float64
 7   school_local       45758 non-null   float64
 8   domestic           55275 non-null   float64
 9   domestic_local      55275 non-null   float64
 10  travel             55275 non-null   float64
 11  travel_partial      55275 non-null   float64
 12  travel_dom          55275 non-null   float64
 13  travel_dom_partial  55275 non-null   float64
 14  curf               55275 non-null   float64
 15  curf_partial        55275 non-null   float64
 16  mass                55275 non-null   float64
 17  mass_partial         55275 non-null   float64
 18  elect               14834 non-null   float64
 19  elect_partial        14834 non-null   float64
 20  sport               55275 non-null   float64
 21  sport_partial        55275 non-null   float64
 22  rest                55275 non-null   float64
 23  rest_local           55275 non-null   float64
 24  testing              55275 non-null   float64
 25  testing_narrow        55275 non-null   float64
 26  masks               51459 non-null   float64
 27  masks_partial         51459 non-null   float64
 28  surveillance          55275 non-null   float64
 29  surveillance_partial  55275 non-null   float64
 30  state                55275 non-null   float64
 31  state_partial         55275 non-null   float64
 32  cash                 54450 non-null   float64
 33  wage                 54450 non-null   float64
 34  credit               54450 non-null   float64
 35  taxc                 54450 non-null   float64
 36  taxd                 54450 non-null   float64
 37  export               54450 non-null   float64
 38  rate                 54450 non-null   float64
 39  Rigidity_Public_Health 55275 non-null   float64
 40  Economic_Measures      54450 non-null   float64
 41  population_2019        57750 non-null   float64
 42  continent             61875 non-null   object 

dtypes: float64(38), object(5)
memory usage: 20.6+ MB
```

## Data Cleaning

Cleans country pandemic totals dataframe by droping parts of dataframe that are undesired, replacing undesired values to desired values, and turning some columns into numeric datatype.

In [9]:

```
df1=df1.drop(0, axis=0)
df1=df1.drop(1, axis=0)
df1=df1.drop(2, axis=0)
df1=df1.drop(3, axis=0)
df1=df1.drop(4, axis=0)
df1=df1.drop(5, axis=0)
df1=df1.drop(6, axis=0)
df1=df1.drop(7, axis=0)
df1=df1.drop(230, axis=0)
df1=df1.drop(231, axis=0)
df1=df1.drop(232, axis=0)
df1=df1.drop(233, axis=0)
df1=df1.drop(234, axis=0)
df1=df1.drop(235, axis=0)
df1=df1.drop(236, axis=0)
df1=df1.drop(237, axis=0)
df1=df1.replace('USA', 'United States')
df1=df1.replace('UK', 'United Kingdom')
df1=df1.replace('UAE', 'United Arab Emirates')
df1=df1.replace('S. Korea', 'South Korea')
df1=df1.replace('S. Korea', 'South Korea')
df1=df1.replace('CAR', 'Central African Republic')
df1=df1.replace('CAR', 'Central African Republic')
df1=df1.replace('DRC', 'Democratic Republic of the Congo')
df1=df1.replace(' ', 0)
df1=df1.replace('N/A', 0)
df1['Total_Cases'] = df1['Total_Cases'].str.replace(",","")
df1['Total_Cases'] = pd.to_numeric(df1['Total_Cases'])
df1['New_Cases'] = df1['New_Cases'].str.replace(",","")
df1['New_Cases'] = df1['New_Cases'].str.replace("+","")
df1['New_Cases'] = pd.to_numeric(df1['New_Cases'])
df1['Total_Deaths'] = df1['Total_Deaths'].str.replace(",","")
df1['Total_Deaths'] = pd.to_numeric(df1['Total_Deaths'])
df1['New_Deaths'] = df1['New_Deaths'].str.replace("+","")
df1['New_Deaths'] = df1['New_Deaths'].str.replace(",","")
df1['New_Deaths'] = pd.to_numeric(df1['New_Deaths'])
df1['Total_Recovered'] = df1['Total_Recovered'].str.replace(",","")
df1['Total_Recovered'] = pd.to_numeric(df1['Total_Recovered'])
df1['Active_Cases'] = df1['Active_Cases'].str.replace("+","")
df1['Active_Cases'] = df1['Active_Cases'].str.replace(",","")
df1['Active_Cases'] = pd.to_numeric(df1['Active_Cases'])
df1['Serious_Critical'] = df1['Serious_Critical'].str.replace(",","")
df1['Serious_Critical'] = pd.to_numeric(df1['Serious_Critical'])
df1['Tot_Cases_per_1m_Pop'] = df1['Tot_Cases_per_1m_Pop'].str.replace(",","")
df1['Tot_Cases_per_1m_Pop'] = pd.to_numeric(df1['Tot_Cases_per_1m_Pop'])
df1['Tot_Deaths_per_1m_Pop'] = df1['Tot_Deaths_per_1m_Pop'].str.replace(",","")
df1['Tot_Deaths_per_1m_Pop'] = pd.to_numeric(df1['Tot_Deaths_per_1m_Pop'])
df1['Total_Tests'] = df1['Total_Tests'].str.replace(",","")
df1['Total_Tests'] = pd.to_numeric(df1['Total_Tests'])
df1['Tot_Tests_per_1m_Pop'] = df1['Tot_Tests_per_1m_Pop'].str.replace(",","")
df1['Tot_Tests_per_1m_Pop'] = pd.to_numeric(df1['Tot_Tests_per_1m_Pop'])
df1['Population'] = df1['Population'].str.replace(",","")
df1['Population'] = pd.to_numeric(df1['Population'])
df1=df1.replace(np.nan, 0)
df1
```

Out[9]:

0	Country	Total_Cases	New_Cases	Total_Deaths	New_Deaths	Total_Recovered	1	Active_Cases	Serious_Critical	Tot_Cases_per_1m_Pop	Tot_Deaths_per_1m_Pop	Total_Tests	Tot_Tests_per_1m_Pop	Pop
---	---------	-------------	-----------	--------------	------------	-----------------	---	--------------	------------------	----------------------	-----------------------	-------------	----------------------	-----

	0	Country	Total_Cases	New_Cases	Total_Deaths	New_Deaths	Total_Recovered	1	Active_Cases	Serious_Critical	Tot_Cases_per_1m_Pop	Tot_Deaths_per_1m_Pop	Total_Tests	Tot_Tests_per_1m_Pop	Pop	
8		World	129425805	610239.0	2826685.0	11499.0	104382520.0	+446,792	22216600.0	96357.0	16604.0	362.6	0.0	0.0	0.0000	
9	1	United States	31161286	63698.0	565193.0	1052.0	23670368.0	+83,454	6925725.0	8756.0	93733.0	1700.0	403189039.0	1212791.0	3.3244	
10	2	Brazil	12748747	84689.0	321515.0	3579.0	11169937.0	+95,454	1257295.0	8318.0	59662.0	1505.0	28600000.0	133843.0	2.1368	
11	3	India	12220669	72182.0	162960.0	458.0	11472494.0	+40,442	585215.0	8944.0	8791.0	117.0	243672940.0	175293.0	1.3900	
12	4	France	4644423	41907.0	95640.0	303.0	294638.0	+1,842	4254145.0	5053.0	71036.0	1463.0	63999096.0	978861.0	6.5387	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
225	217	Marshall Islands	4	0.0	0.0	0.0	4.0		0.0	0.0	67.0		0.0	0.0	0.0	5.9487
226	218	Samoa	3	0.0	0.0	0.0	2.0		1.0	0.0	15.0		0.0	0.0	0.0	1.9939
227	219	Vanuatu	3	0.0	0.0	0.0	1.0		2.0	0.0	10.0		0.0	470.0	1504.0	3.1244
228	220	Micronesia	1	0.0	0.0	0.0	1.0		0.0	0.0	9.0		0.0	0.0	0.0	1.1591
229	221	China	90201	11.0	4636.0	0.0	85385.0	+4	180.0	2.0	63.0	3.0	160000000.0	111163.0	1.4393	

222 rows × 19 columns

Cleans country vaccinations dataframe by replacing undesired values with desired ones, renaming undesired column names to desired ones, and storing global vaccinations total in a variable wvt.

In [10]:	df3=df3.replace(np.nan,0) df3=df3.rename(columns={'location':'Country'}) wvt=df3.loc[df3['Country']=='World']['total_vaccinations'].iloc[-1] df3																																																																																																												
Out[10]:	<table border="1"> <thead> <tr> <th></th><th>Country</th><th>iso_code</th><th>date</th><th>total_vaccinations</th><th>people_vaccinated</th><th>people_fully_vaccinated</th><th>daily_vaccinations_raw</th><th>daily_vaccinations</th><th>total_vaccinations_per_hundred</th><th>people_vaccinated_per_hundred</th><th>people_fully_vaccinate</th></tr> </thead> <tbody> <tr> <td>0</td><td>Afghanistan</td><td>AFG</td><td>2021-02-22</td><td>0.0</td><td>0.0</td><td>0.0</td><td>0.0</td><td>0.0</td><td>0.0</td><td>0.00</td><td>0.00</td></tr> <tr> <td>1</td><td>Afghanistan</td><td>AFG</td><td>2021-02-23</td><td>0.0</td><td>0.0</td><td>0.0</td><td>0.0</td><td>0.0</td><td>1367.0</td><td>0.00</td><td>0.00</td></tr> <tr> <td>2</td><td>Afghanistan</td><td>AFG</td><td>2021-02-24</td><td>0.0</td><td>0.0</td><td>0.0</td><td>0.0</td><td>0.0</td><td>1367.0</td><td>0.00</td><td>0.00</td></tr> <tr> <td>3</td><td>Afghanistan</td><td>AFG</td><td>2021-02-25</td><td>0.0</td><td>0.0</td><td>0.0</td><td>0.0</td><td>0.0</td><td>1367.0</td><td>0.00</td><td>0.00</td></tr> <tr> <td>4</td><td>Afghanistan</td><td>AFG</td><td>2021-02-26</td><td>0.0</td><td>0.0</td><td>0.0</td><td>0.0</td><td>0.0</td><td>1367.0</td><td>0.00</td><td>0.00</td></tr> <tr> <td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td><td>...</td></tr> <tr> <td>9813</td><td>Zimbabwe</td><td>ZWE</td><td>2021-03-25</td><td>54892.0</td><td>54892.0</td><td>0.0</td><td>5488.0</td><td>2127.0</td><td>0.37</td><td></td><td>0.37</td></tr> <tr> <td>9814</td><td>Zimbabwe</td><td>ZWE</td><td>2021-03-26</td><td>61093.0</td><td>61093.0</td><td>0.0</td><td>6201.0</td><td>2772.0</td><td>0.41</td><td></td><td>0.41</td></tr> </tbody> </table>		Country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hundred	people_fully_vaccinate	0	Afghanistan	AFG	2021-02-22	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00	1	Afghanistan	AFG	2021-02-23	0.0	0.0	0.0	0.0	0.0	1367.0	0.00	0.00	2	Afghanistan	AFG	2021-02-24	0.0	0.0	0.0	0.0	0.0	1367.0	0.00	0.00	3	Afghanistan	AFG	2021-02-25	0.0	0.0	0.0	0.0	0.0	1367.0	0.00	0.00	4	Afghanistan	AFG	2021-02-26	0.0	0.0	0.0	0.0	0.0	1367.0	0.00	0.00	...	...	...	...	...	...	...	...	...	...	...	...	9813	Zimbabwe	ZWE	2021-03-25	54892.0	54892.0	0.0	5488.0	2127.0	0.37		0.37	9814	Zimbabwe	ZWE	2021-03-26	61093.0	61093.0	0.0	6201.0	2772.0	0.41		0.41
	Country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hundred	people_fully_vaccinate																																																																																																		
0	Afghanistan	AFG	2021-02-22	0.0	0.0	0.0	0.0	0.0	0.0	0.00	0.00																																																																																																		
1	Afghanistan	AFG	2021-02-23	0.0	0.0	0.0	0.0	0.0	1367.0	0.00	0.00																																																																																																		
2	Afghanistan	AFG	2021-02-24	0.0	0.0	0.0	0.0	0.0	1367.0	0.00	0.00																																																																																																		
3	Afghanistan	AFG	2021-02-25	0.0	0.0	0.0	0.0	0.0	1367.0	0.00	0.00																																																																																																		
4	Afghanistan	AFG	2021-02-26	0.0	0.0	0.0	0.0	0.0	1367.0	0.00	0.00																																																																																																		
...	...	...	...	...	...	...	...	...	...	...	...																																																																																																		
9813	Zimbabwe	ZWE	2021-03-25	54892.0	54892.0	0.0	5488.0	2127.0	0.37		0.37																																																																																																		
9814	Zimbabwe	ZWE	2021-03-26	61093.0	61093.0	0.0	6201.0	2772.0	0.41		0.41																																																																																																		

	Country	iso_code	date	total_vaccinations	people_vaccinated	people_fully_vaccinated	daily_vaccinations_raw	daily_vaccinations	total_vaccinations_per_hundred	people_vaccinated_per_hundred	people_fully_vaccinate
9815	Zimbabwe	ZWE	2021-03-27	65466.0	65466.0	0.0	4373.0	3322.0	0.44	0.44	
9816	Zimbabwe	ZWE	2021-03-28	68511.0	68511.0	0.0	3045.0	3683.0	0.46	0.46	
9817	Zimbabwe	ZWE	2021-03-29	69751.0	69751.0	0.0	1240.0	3780.0	0.47	0.47	

9818 rows × 12 columns

Makes a new dataframe df\_stats that the function stats will populate with desired data.

```
In [11]: df_stats=pd.DataFrame(columns=['Country','Population','Tests', 'Tests/Population','Cases','Active_Cases','Cases/Tests','Recoveries','Recoveries/Cases','Deaths','Deaths/Cases','Vaccinations','Vaccinations/Population'])
df_stats
```

```
Out[11]: Country Population Tests Tests/Population Cases Active_Cases Cases/Tests Recoveries Recoveries/Cases Deaths Deaths/Cases Vaccinations Vaccinations/Population
```

```
In [12]: countries=sorted(df1['Country'].unique())
stats(countries, df1, 'Country', df3, df_stats)
json_storage['df_stats: Country_Stats_df']=df_stats
```

Makes a dataframe containing global pandemic totals.

```
In [13]: world_df=df_stats.loc[df_stats['Country']=='World']
df_stats=df_stats.drop(218, axis=0)
world_df=world_df.rename(columns={'Country':'Location'})
world_df['Population']=sum(df_stats['Population'])
world_df['Tests']=sum(df_stats['Tests'])
world_df['Tests/Population']=(world_df['Tests']/world_df['Population'])*100
world_df['Active_Cases']=int(world_df['Active_Cases'])
world_df['Cases/Tests']=(world_df['Cases']/world_df['Tests'])*100
world_df['Vaccinations']=int(wvt)
world_df['Vaccinations/Population']=(world_df['Vaccinations']/world_df['Population'])*100
json_storage['world_df: World_Stats_df']=world_df
```

Cleans USA State pandemic totals dataframe by droping parts of dataframe that are undesired, replacing undesired values to desired values, turning some columns into numeric datatype, storing in usa\_total variable the total population, and removing white space from the end of each state name.

```
In [14]: df2=df2.drop(0, axis=0)
df2=df2.drop(53, axis=0)
df2=df2.drop(54, axis=0)
df2=df2.drop(55, axis=0)
df2=df2.drop(56, axis=0)
df2=df2.drop(57, axis=0)
df2=df2.drop(58, axis=0)
df2=df2.drop(59, axis=0)
df2=df2.drop(60, axis=0)
df2=df2.drop(61, axis=0)
df2=df2.drop(62, axis=0)
df2=df2.drop(63, axis=0)
df2=df2.drop(64, axis=0)
```

```

df2=df2.drop(65, axis=0)
df2=df2.replace(np.nan, 0)
df2=df2.replace(' ', 0)
df2=df2.replace('N/A', 0)
df2['Country']=df2['Country'].astype('str')
df2['Country'] = df2['Country'].str.replace("\n", "")
df2['Total_Cases']=df2['Total_Cases'].astype('str')
df2['Total_Cases'] = df2['Total_Cases'].str.replace(",","",)
df2['Total_Cases'] = df2['Total_Cases'].str.replace("\n","",)
df2['Total_Cases'] = pd.to_numeric(df2['Total_Cases'])
df2['New_Cases']=df2['New_Cases'].astype('str')
df2['New_Cases'] = df2['New_Cases'].str.replace(",","",)
df2['New_Cases'] = df2['New_Cases'].str.replace("+","",)
df2['New_Cases'] = df2['New_Cases'].str.replace("\n","",)
df2['New_Cases'] = pd.to_numeric(df2['New_Cases'])
df2['Total_Deaths']=df2['Total_Deaths'].astype('str')
df2['Total_Deaths'] = df2['Total_Deaths'].str.replace(",","",)
df2['Total_Deaths'] = df2['Total_Deaths'].str.replace("\n","",)
df2['Total_Deaths'] = pd.to_numeric(df2['Total_Deaths'])
df2['New_Deaths']=df2['New_Deaths'].astype('str')
df2['New_Deaths'] = df2['New_Deaths'].str.replace("+","",)
df2['New_Deaths'] = df2['New_Deaths'].str.replace(",","",)
df2['New_Deaths'] = df2['New_Deaths'].str.replace("\n","",)
df2['New_Deaths'] = pd.to_numeric(df2['New_Deaths'])
df2['Total_Recovered']=df2['Total_Recovered'].astype('str')
df2['Total_Recovered'] = df2['Total_Recovered'].str.replace(",","",)
df2['Total_Recovered'] = df2['Total_Recovered'].str.replace("\n","",)
df2['Total_Recovered'] = df2['Total_Recovered'].str.replace("N/A","",)
df2['Total_Recovered'] = df2['Total_Recovered'].str.replace(" ",',')
df2['Total_Recovered'] = pd.to_numeric(df2['Total_Recovered'])
df2['Active_Cases']=df2['Active_Cases'].astype('str')
df2['Active_Cases'] = df2['Active_Cases'].str.replace("+","",)
df2['Active_Cases'] = df2['Active_Cases'].str.replace(",","",)
df2['Active_Cases'] = df2['Active_Cases'].str.replace("\n","",)
df2['Active_Cases'] = df2['Active_Cases'].str.replace("N/A","",)
df2['Active_Cases'] = df2['Active_Cases'].str.replace(" ",',')
df2['Active_Cases'] = pd.to_numeric(df2['Active_Cases'])
df2['Tot_Cases_per_1m_Pop']=df2['Tot_Cases_per_1m_Pop'].astype('str')
df2['Tot_Cases_per_1m_Pop'] = df2['Tot_Cases_per_1m_Pop'].str.replace(",","",)
df2['Tot_Cases_per_1m_Pop'] = df2['Tot_Cases_per_1m_Pop'].str.replace("\n","",)
df2['Tot_Cases_per_1m_Pop'] = pd.to_numeric(df2['Tot_Cases_per_1m_Pop'])
df2['Tot_Deaths_per_1m_Pop']=df2['Tot_Deaths_per_1m_Pop'].astype('str')
df2['Tot_Deaths_per_1m_Pop'] = df2['Tot_Deaths_per_1m_Pop'].str.replace(",","",)
df2['Tot_Deaths_per_1m_Pop'] = df2['Tot_Deaths_per_1m_Pop'].str.replace("\n","",)
df2['Tot_Deaths_per_1m_Pop'] = pd.to_numeric(df2['Tot_Deaths_per_1m_Pop'])
df2['Total_Tests']=df2['Total_Tests'].astype('str')
df2['Total_Tests'] = df2['Total_Tests'].str.replace(",","",)
df2['Total_Tests'] = df2['Total_Tests'].str.replace("\n","",)
df2['Total_Tests'] = pd.to_numeric(df2['Total_Tests'])
df2['Tot_Tests_per_1m_Pop']=df2['Tot_Tests_per_1m_Pop'].astype('str')
df2['Tot_Tests_per_1m_Pop'] = df2['Tot_Tests_per_1m_Pop'].str.replace(",","",)
df2['Tot_Tests_per_1m_Pop'] = df2['Tot_Tests_per_1m_Pop'].str.replace("\n","",)
df2['Tot_Tests_per_1m_Pop'] = pd.to_numeric(df2['Tot_Tests_per_1m_Pop'])
df2['Population']=df2['Population'].astype('str')
df2['Population'] = df2['Population'].str.replace(",","",)
df2['Population'] = df2['Population'].str.replace("\n","",)
df2['Population'] = pd.to_numeric(df2['Population'])
df2=df2.replace(np.nan, 0)
df2=df2.replace(' ', 0)
df2=df2.replace('N/A', 0)
usa_total=df2.iloc[0]

```

```

df2=df2.drop(1, axis=0)
df2=df2.rename(columns={'Country':'State'})
df2['State']=df2['State'].replace('District Of Columbia ','District of Columbia ')
df2['State'] = df2['State'].apply(lambda x: x.strip())
df2

```

	0	State	Total_Cases	New_Cases	Total_Deaths	New_Deaths	Total_Recovered	Active_Cases	Tot_Cases_per_1m_Pop	Tot_Deaths_per_1m_Pop	Total_Tests	Tot_Tests_per_1m_Pop	Population	
2	1	California	3668451	1703.0	58912	123	1940386.0	1669153.0	92843	1491	53786487	1361262	3951223.0	\n[view
3	2	Texas	2792858	0.0	48479	0	2646224.0	98155.0	96319	1672	25536604	880698	28995881.0	\n[view by cour
4	3	Florida	2057735	5294.0	33450	84	1520607.0	503678.0	95808	1557	25289334	1177467	21477737.0	\n[view by cour
5	4	New York	1915254	8628.0	50587	87	1124544.0	740123.0	98453	2600	44936152	2309919	19453561.0	\n[view by cour
6	5	Illinois	1244585	2592.0	23579	37	1158104.0	62902.0	98217	1861	20313050	1603010	12671821.0	\n[1]
7	6	Georgia	1059548	1807.0	19055	68	692672.0	347821.0	99793	1795	8291638	780946	10617423.0	
8	7	Pennsylvania	1030587	4223.0	25210	42	932619.0	72758.0	80502	1969	11687576	912950	12801989.0	\n[view
9	8	Ohio	1017566	1989.0	18609	0	966333.0	32624.0	87053	1592	10979647	939306	11689100.0	\n[view by cour
10	9	North Carolina	914132	1929.0	12112	25	876108.0	25912.0	87159	1155	11336690	1080911	10488084.0	\r
11	10	New Jersey	908816	3672.0	24561	75	685539.0	198716.0	102319	2765	12100950	1362384	8882190.0	\n[view
12	11	Arizona	841811	733.0	16967	26	803013.0	21831.0	115654	2331	4498406	618022	7278717.0	
13	12	Tennessee	811842	1313.0	11904	10	786959.0	12979.0	118879	1743	7219625	1057174	6829174.0	\r
14	13	Michigan	746351	7107.0	17119	11	569460.0	159772.0	74733	1714	12065261	1208114	9986857.0	\r
15	14	Indiana	686497	1044.0	13039	1	655004.0	18454.0	101972	1937	8903743	1322557	6732219.0	
16	15	Massachusetts	635580	2499.0	17185	37	586484.0	31911.0	92213	2493	18821008	2730649	6892503.0	
17	16	Virginia	618976	1035.0	10252	10	51481.0	557243.0	72518	1201	8396642	983730	8535519.0	
18	17	Wisconsin	577195	563.0	6622	10	563534.0	7039.0	99133	1137	3309569	568417	5822434.0	\n[1] [2]
19	18	Missouri	577008	698.0	9081	0	428548.0	139379.0	94015	1480	6191242	1008768	6137428.0	\n[1] [2] [3]
20	19	South Carolina	551630	961.0	9146	20	0.0	0.0	107139	1776	6756672	1312303	5148714.0	
21	20	Minnesota	519529	1648.0	6916	12	499395.0	13218.0	92121	1226	8211019	1455949	5639632.0	
22	21	Alabama	515388	408.0	10554	27	315743.0	189091.0	105113	2152	2534623	516934	4903185.0	
23	22	Colorado	462081	1763.0	6238	12	328754.0	127089.0	80240	1083	2725439	473270	5758736.0	
24	23	Louisiana	444933	519.0	10141	9	429935.0	4857.0	95709	2181	6495793	1397307	4648794.0	\n[view
25	24	Oklahoma	438364	390.0	4953	103	422816.0	10595.0	110783	1252	3783639	956196	3956971.0	
26	25	Kentucky	426876	803.0	6090	25	49678.0	371108.0	95548	1363	4949269	1107796	4467673.0	\r
27	26	Maryland	411344	1366.0	8286	13	9862.0	393196.0	68039	1371	8766709	1450078	6045680.0	
28	27	Utah	385641	514.0	2122	4	373856.0	9663.0	120289	662	4235928	1321267	3205958.0	
29	28	Iowa	379387	570.0	5743	14	332959.0	40685.0	120247	1820	4447615	1409672	3155070.0	
30	29	Washington	366134	0.0	5290	0	179498.0	181346.0	48081	695	5890924	773606	7614893.0	\n[view
31	30	Arkansas	330398	212.0	5626	10	323032.0	1740.0	109483	1864	3383520	1121186	3017804.0	\n[1]
32	31	Connecticut	310888	832.0	7886	1	271294.0	31708.0	87199	2212	7696039	2158603	3565287.0	

		State	Total_Cases	New_Cases	Total_Deaths	New_Deaths	Total_Recovered	Active_Cases	Tot_Cases_per_1m_Pop	Tot_Deaths_per_1m_Pop	Total_Tests	Tot_Tests_per_1m_Pop	Population
33	32	Mississippi	305146	288.0	7032	19	292872.0	5242.0	102530	2363	2610281	877067	2976149.0
34	33	Nevada	303762	268.0	5249	10	280217.0	18296.0	98619	1704	2950952	958053	3080156.0
35	34	Kansas	303675	0.0	4913	11	290550.0	8212.0	104237	1686	1318593	452609	2913314.0 \n[1] [2] [3] [4] [5] [6] [
36	35	Nebraska	209346	0.0	2180	0	162560.0	44606.0	108222	1127	1002240	518112	1934408.0
37	36	New Mexico	191652	275.0	3937	5	173701.0	14014.0	91401	1878	3000869	1431146	2096829.0
38	37	Idaho	180536	460.0	1962	5	100936.0	77638.0	101024	1098	1186846	664131	1787065.0
39	38	Oregon	165012	442.0	2383	2	147710.0	14919.0	39123	565	4191940	993884	4217737.0
40	39	West Virginia	141738	416.0	2676	36	132784.0	6278.0	79088	1493	2447894	1365900	1792147.0
41	40	Rhode Island	137329	564.0	2619	1	8267.0	126443.0	129634	2472	3518593	3321430	1059361.0
42	41	South Dakota	117759	264.0	1935	0	113302.0	2522.0	133112	2187	444576	502539	884659.0
43	42	Montana	104552	231.0	1437	0	102203.0	912.0	97824	1345	1197763	1120685	1068778.0
44	43	North Dakota	103091	211.0	1466	0	100485.0	1140.0	135279	1924	415765	545579	762062.0
45	44	Delaware	94802	229.0	1559	0	86673.0	6570.0	97356	1601	660836	678641	973764.0
46	45	New Hampshire	84176	462.0	1238	1	79825.0	3113.0	61907	910	1887927	1388477	1359711.0
47	46	Alaska	60333	0.0	309	0	52369.0	7655.0	82473	422	1896375	2592288	731545.0
48	47	Wyoming	56310	74.0	700	5	55132.0	478.0	97294	1209	683766	1181435	578759.0
49	48	Maine	50504	251.0	743	5	12963.0	36798.0	37571	553	2076157	1544516	1344212.0
50	49	District of Columbia	44513	100.0	1064	3	31512.0	11937.0	63072	1508	1375475	1948958	705749.0
51	50	Hawaii	29681	100.0	463	1	0.0	0.0	20963	327	1261216	890770	1415872.0
52	51	Vermont	19275	166.0	227	2	15866.0	3182.0	30890	364	1332173	2134930	623989.0

Cleans state vaccination totals by dropping undesired columns and renaming undesired column names to desired ones.

In [15]:

```
df4=df4.drop('Total COVID-19 Vaccines Delivered', axis=1)
df4=df4.drop('Share of Delivered Vaccines That Have Been Administered', axis=1)
df4=df4.drop('Number of People Who Have Been Vaccinated', axis=1)
df4=df4.drop('Share of Population Vaccinated', axis=1)
df4=df4.rename(columns={'Location':'State'})
df4=df4.rename(columns={'Total COVID-19 Vaccines Administered':'Vaccines_Administered'})
df4
```

Out[15]:

	State	Vaccines_Administered
0	United States	118313818.0
1	Alabama	1426696.0
2	Alaska	356055.0
3	Arizona	2760907.0
4	Arkansas	952919.0
...	...	...

	State	Vaccines_Administered
69	*Vaccines delivered* are cumulative counts of ...	NaN
70		NaN
71	*Vaccines administered* are cumulative counts ...	NaN
72		NaN
73	*Number of people who have been vaccinated* re...	NaN

74 rows × 2 columns

Makes a new dataframe df\_usa\_stats that the function usa\_stats will populate with desired data.

```
In [16]: df_usa_stats=pd.DataFrame(columns=['State','Population','Tests','Tests/Population','Cases','Active_Cases','Cases/Tests','Recoveries','Recoveries/Cases','Deaths','Deaths/Cases','Vaccinations','Vaccinations/Population'])
df_usa_stats
```

```
Out[16]: State Population Tests Tests/Population Cases Active_Cases Cases/Tests Recoveries Recoveries/Cases Deaths Deaths/Cases Vaccinations Vaccinations/Population
```

```
In [17]: states=sorted(df2['State'].unique())
stats(states, df2, 'State', df4, df_usa_stats)
```

Cleans country cases time-series dataframe by renaming undesired name to desired one.

```
In [18]: df5=df5.rename(columns={'Date_reported':'Date'})
df5
```

	Date	Country_code	Country	WHO_region	New_cases	Cumulative_cases	New_deaths	Cumulative_deaths
0	2020-01-03	AF	Afghanistan	EMRO	0	0	0	0
1	2020-01-04	AF	Afghanistan	EMRO	0	0	0	0
2	2020-01-05	AF	Afghanistan	EMRO	0	0	0	0
3	2020-01-06	AF	Afghanistan	EMRO	0	0	0	0
4	2020-01-07	AF	Afghanistan	EMRO	0	0	0	0
...	...	...	...	...	...	...	...	...
107139	2021-03-27	ZW	Zimbabwe	AFRO	27	36805	0	1518
107140	2021-03-28	ZW	Zimbabwe	AFRO	13	36818	1	1519
107141	2021-03-29	ZW	Zimbabwe	AFRO	4	36822	1	1520
107142	2021-03-30	ZW	Zimbabwe	AFRO	17	36839	0	1520
107143	2021-03-31	ZW	Zimbabwe	AFRO	19	36858	1	1521

107144 rows × 8 columns

Cleans USA state cases time-series dataframe by renaming state name values with initials so that coloropleth function recognizes them using map method, dropping undesired column, and renaming column to desired name.

```
In [19]: state = {
    'Alabama': 'AL',
```

```
'Alaska': 'AK',
'American Samoa': 'AS',
'Arizona': 'AZ',
'Arkansas': 'AR',
'California': 'CA',
'Colorado': 'CO',
'Connecticut': 'CT',
'Delaware': 'DE',
'District of Columbia': 'DC',
'Florida': 'FL',
'Georgia': 'GA',
'Guam': 'GU',
'Hawaii': 'HI',
'Idaho': 'ID',
'Illinois': 'IL',
'Indiana': 'IN',
'Iowa': 'IA',
'Kansas': 'KS',
'Kentucky': 'KY',
'Louisiana': 'LA',
'Maine': 'ME',
'Maryland': 'MD',
'Massachusetts': 'MA',
'Michigan': 'MI',
'Minnesota': 'MN',
'Mississippi': 'MS',
'Missouri': 'MO',
'Montana': 'MT',
'Nebraska': 'NE',
'Nevada': 'NV',
'New Hampshire': 'NH',
'New Jersey': 'NJ',
'New Mexico': 'NM',
'New York': 'NY',
'North Carolina': 'NC',
'North Dakota': 'ND',
'Northern Mariana Islands': 'MP',
'Ohio': 'OH',
'Oklahoma': 'OK',
'Oregon': 'OR',
'Pennsylvania': 'PA',
'Puerto Rico': 'PR',
'Rhode Island': 'RI',
'South Carolina': 'SC',
'South Dakota': 'SD',
'Tennessee': 'TN',
'Texas': 'TX',
'Utah': 'UT',
'Vermont': 'VT',
'Virgin Islands': 'VI',
'Virginia': 'VA',
'Washington': 'WA',
'West Virginia': 'WV',
'Wisconsin': 'WI',
'Wyoming': 'WY'}
```

```
df6['State'] = df6['state'].map(state)
df6=df6.drop('fips', axis=1)
df6=df6.rename(columns={'state':'states'})
df6
```

Out[19]:

	date	states	cases	deaths	State
0	2020-01-21	Washington	1	0	WA
1	2020-01-22	Washington	1	0	WA
2	2020-01-23	Washington	1	0	WA
3	2020-01-24	Illinois	1	0	IL
4	2020-01-24	Washington	1	0	WA
...	...	...	...	...	...
21624	2021-03-30	Virginia	617941	10242	VA
21625	2021-03-30	Washington	365816	5302	WA
21626	2021-03-30	West Virginia	141322	2640	WV
21627	2021-03-30	Wisconsin	635523	7293	WI
21628	2021-03-30	Wyoming	56236	695	WY

21629 rows × 5 columns

Set index to date column for time series data.

```
In [30]: df5.set_index('Date', inplace=True)
df3.set_index('date', inplace=True)
```

Makes a dataframe that holds the total virus cases time series.

```
In [27]: global_cases_df=df5.loc[df5['Country']=='Afghanistan']['Cumulative_cases']
for i in df5['Country'].unique():
    if i=='Afghanistan':
        continue
    else:
        dd=df5.loc[df5['Country']==i]['Cumulative_cases']
        global_cases_df+=dd
global_cases_df
```

```
Out[27]: Date
2020-01-03      0
2020-01-04      1
2020-01-05      1
2020-01-06      4
2020-01-07      4
...
2021-03-27  125863438
2021-03-28  126456023
2021-03-29  126987247
2021-03-30  127424100
2021-03-31  127933916
Name: Cumulative_cases, Length: 454, dtype: int64
```

Makes a dataframe that holds the total virus vaccination time series.

```
In [31]: global_vaccinations_df=df3.loc[df3['Country']=='World']['people_fully_vaccinated']
global_vaccinations_df
```

```
Out[31]: date
2020-12-13      0.0
2020-12-14      0.0
```

```
2020-12-15      0.0
2020-12-16      0.0
2020-12-17      0.0
...
2021-03-26  116309735.0
2021-03-27  118768359.0
2021-03-28  120832732.0
2021-03-29  124654695.0
2021-03-30  126779833.0
Name: people_fully_vaccinated, Length: 108, dtype: float64
```

Makes a sql database that contains cleaned dataframes as tables.

```
In [35]: conn = sql.connect('covid19.db')

In [93]: df_stats.to_sql('country_stats', conn)

In [94]: df_usa_stats.to_sql('usa_state_stats', conn)

In [95]: pd.read_sql('''SELECT name from sqlite_master where type="table";''', conn)
```

```
Out[95]:
name
0   country_stats
1   usa_state_stats
```

## Data Exploration

24.1% of the global population has been tested for the virus. 6.8% of the tested global tested have been diagnosed with the virus. 80.6% of global cases have recovered and 2.2% have died from the virus. 7.2% of the global population has been vaccinated.

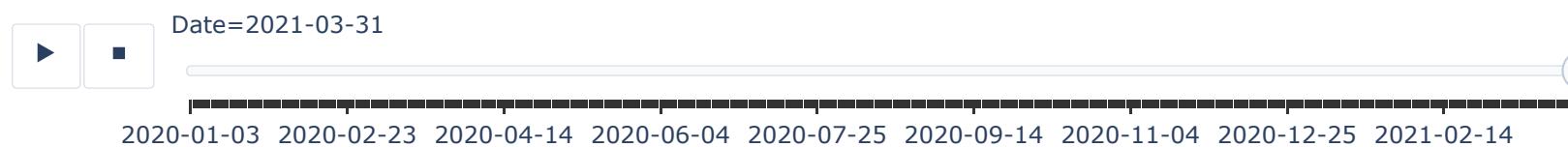
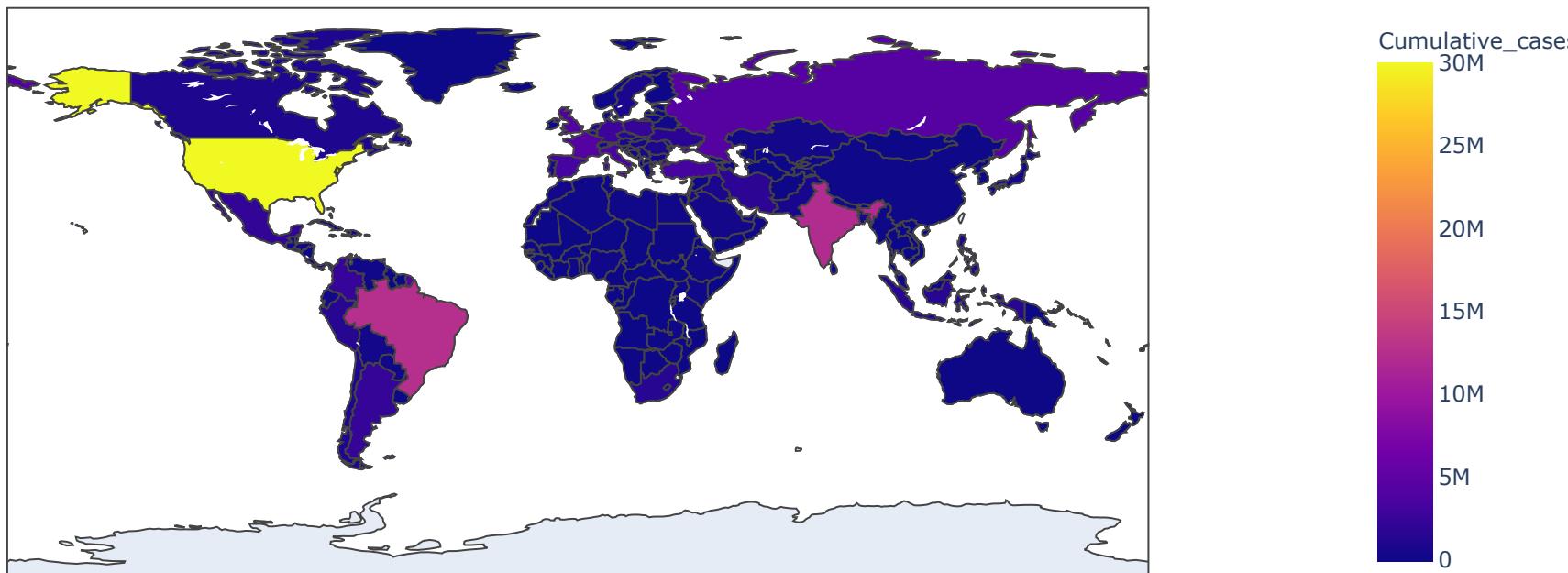
```
In [97]: world_df

Out[97]:
Location  Population  Tests  Tests/Population  Cases  Active_Cases  Cases/Tests  Recoveries  Recoveries/Cases  Deaths  Deaths/Cases  Vaccinations  Vaccinations/Population
218    World    7815306159  1886790963    24.142253  128734212    22154126    6.822919  103766203    80.604993  2813883    2.185808  564561593    7.223794
```

The United States has the most cases of the virus.

```
In [ ]:
graph1=graph()
graph1.choropleth(df5, 'Country', 'country names', 'Cumulative_cases', df5.index, 'world', 'Global spread of Covid19')
```

## Global spread of Covid19



```
In [107]: desc(df_stats['Population'])
```

```
Population
Count: 221
Mean: 35363376.28506787
Standard Deviation: 139675340.05062267
Min: 0
Q1 quantile: 628125.0
Q2 quantile: 6510483.0
Q3 quantile: 23848794.0
Max: 1439323776
```

Selects countries with larger populations than average country and fewer cases than average country.

```
In [233]: pd.read_sql('''SELECT Country, Population, Cases FROM country_stats
                  WHERE Population > (SELECT AVG(Population) FROM country_stats) AND Cases < (SELECT AVG(Cases) FROM country_stats)
                  ORDER BY Population DESC;''' , conn)
```

	Country	Population	Cases
0	China	1439323776	90190
1	Nigeria	209896709	162641
2	Japan	126188355	470175
3	Ethiopia	117057673	204521

	Country	Population	Cases
4	Egypt	103757682	200739
5	Vietnam	97987233	2594
6	Democratic Republic of the Congo	91558337	28076
7	Thailand	69930037	28821
8	Tanzania	60983198	509
9	Myanmar	54680205	142412
10	Kenya	54643416	132646
11	South Korea	51301970	102582
12	Uganda	46799838	40839
13	Sudan	44602643	29825
14	Algeria	44435162	116946
15	Afghanistan	39573734	56454
16	Morocco	37234187	495421

```
In [109]: desc(df_stats['Tests'])
```

Tests  
Count: 221  
Mean: 8537515.669683257  
Standard Deviation: 35658632.49301704  
Min: 0  
Q1 quantile: 65827.0  
Q2 quantile: 620810.0  
Q3 quantile: 3539320.0  
Max: 401946739

```
In [205]: desc(df_stats['Tests/Population'])
```

Tests/Population  
Count: 221  
Mean: 33.432348755114575  
Standard Deviation: 35.11752363793047  
Min: 0.0  
Q1 quantile: 3.126226593925527  
Q2 quantile: 18.86441799275805  
Q3 quantile: 56.7955491303749  
Max: 99.99

The United States likely has the most cases of the virus than other countries because it has tested more than other countries.

```
In [36]: pd.read_sql('''SELECT Country, Tests FROM country_stats
                    ORDER BY Tests DESC
                    LIMIT 10;''', conn)
```

	Country	Tests
0	United States	401946739
1	India	242650025
2	China	160000000

	Country	Tests
3	United Kingdom	124452321
4	Russia	119900000
5	France	63999096
6	Italy	49551436
7	Germany	48979281
8	Spain	42707830
9	Turkey	38338045

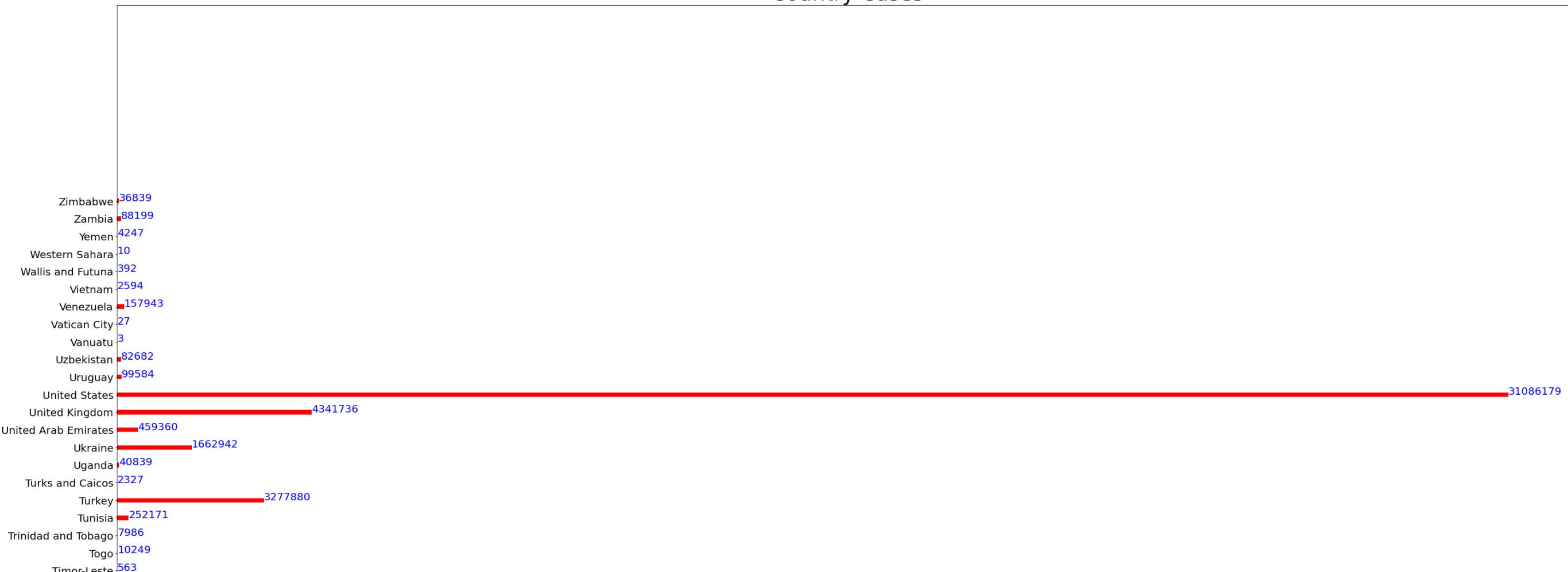
```
In [113]: desc(df_stats['Cases'])
```

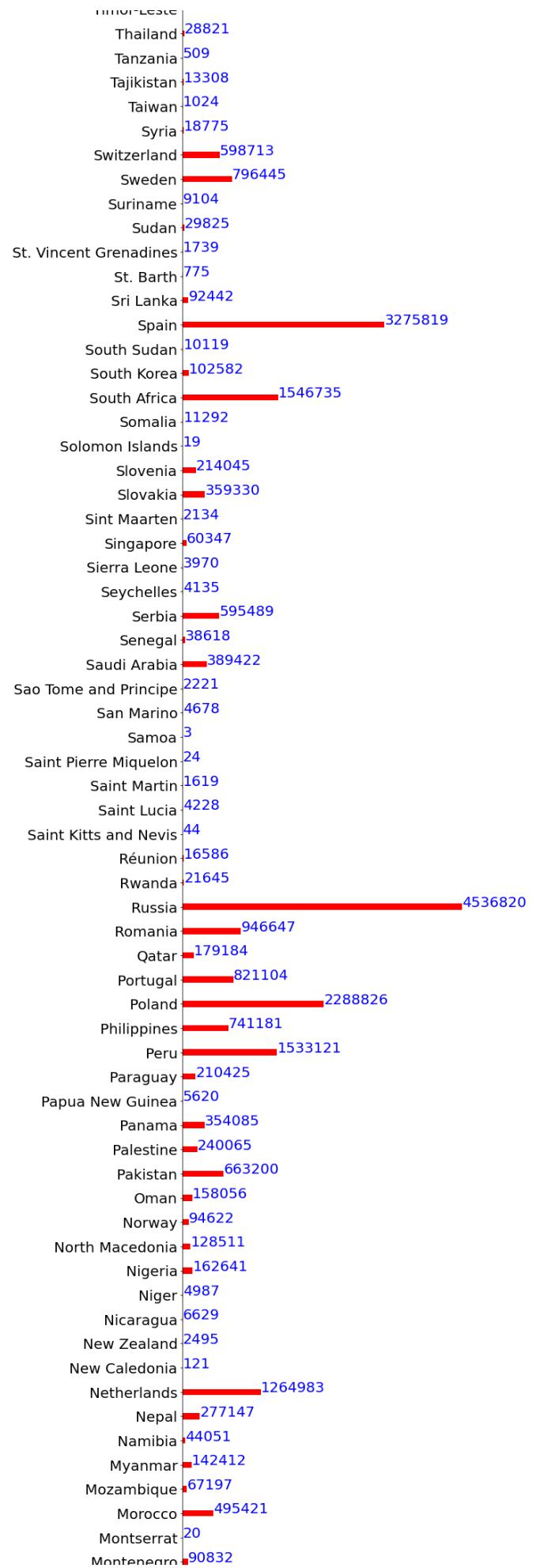
Cases  
Count: 221  
Mean: 582507.7466063348  
Standard Deviation: 2480007.4734795145  
Min: 1  
Q1 quantile: 4228.0  
Q2 quantile: 36839.0  
Q3 quantile: 252171.0  
Max: 31086179

```
In [114]: graph5=graph()  

graph5.bar_chart(df_stats['Country'], df_stats['Cases'], 'Country Cases', 'Cases', 'Country')
```

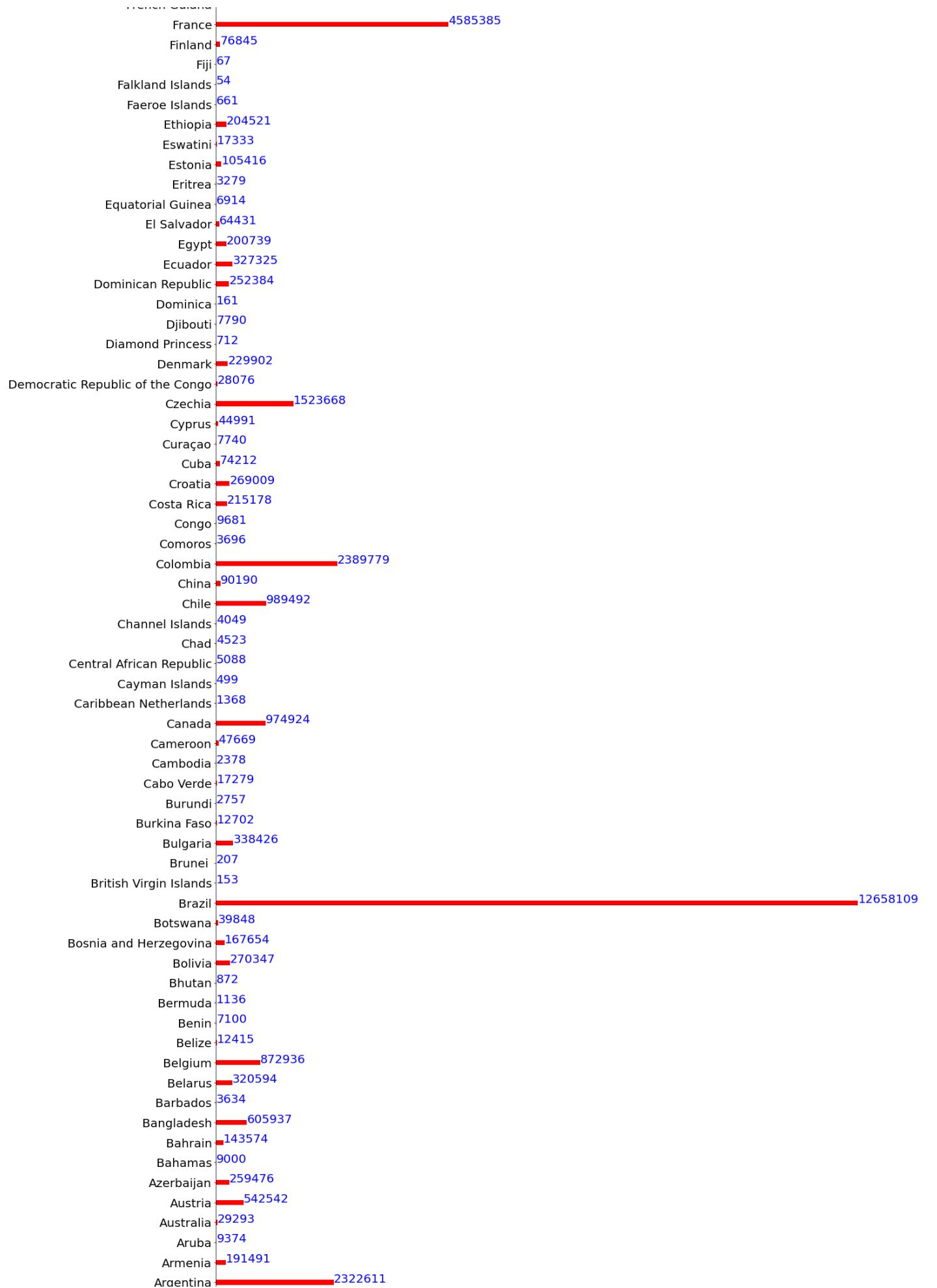
Country Cases

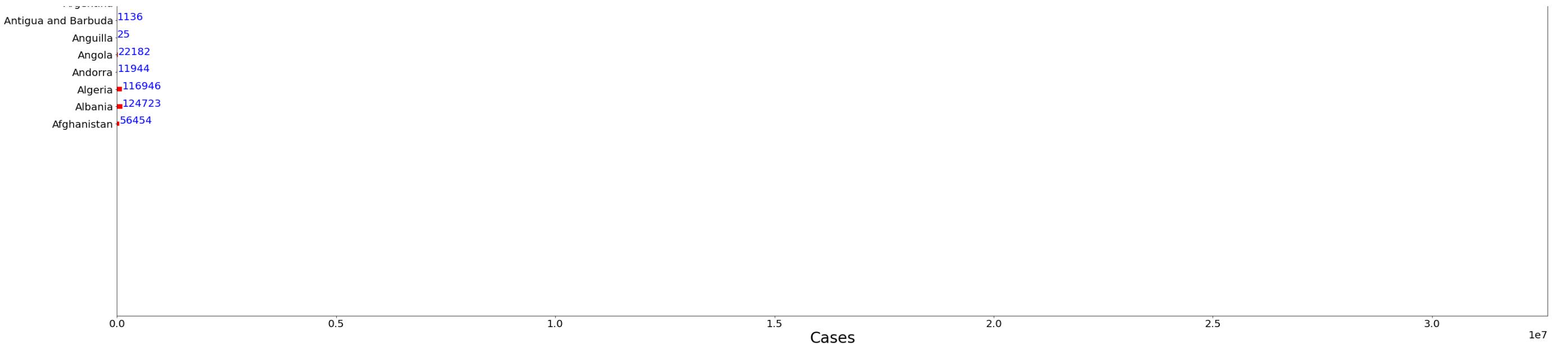




Country







```
In [115]: desc(df_stats['Active_Cases'])
```

```
Active_Cases
Count: 221
Mean: 91154.82805429865
Standard Deviation: 556770.0224245826
Min: 0
Q1 quantile: 216.0
Q2 quantile: 2548.0
Q3 quantile: 20787.0
Max: 6948028
```

```
In [216]: pd.read_sql('''SELECT Country, Active_Cases FROM country_stats
                  ORDER BY Active_Cases DESC
                  LIMIT 10;''', conn)
```

```
Out[216]:
```

	Country	Active_Cases
0	United States	6948028
1	France	4197252
2	Brazil	1371216
3	Belgium	793295
4	Italy	562832
5	India	553874
6	Poland	388235
7	United Kingdom	379848
8	Ukraine	323448
9	Russia	282382

```
In [117]: desc(df_stats['Cases/Tests'])
```

```
Cases/Tests
Count: 221
Mean: 9.311129392984794
Standard Deviation: 10.57910265675415
```

```
Min: 0.0
Q1 quantile: 1.90079008585828
Q2 quantile: 7.312139686363121
Q3 quantile: 12.840935999827781
Max: 79.51923076923076
```

```
In [119]: desc(df_stats['Recoveries'])
```

```
Recoveries
Count: 221
Mean: 469428.1583710407
Standard Deviation: 1977647.8892793
Min: 0
Q1 quantile: 2801.0
Q2 quantile: 25444.0
Q3 quantile: 213590.0
Max: 23574225
```

```
In [121]: desc(df_stats['Recoveries/Cases'])
```

```
Recoveries/Cases
Count: 221
Mean: 81.70323256192705
Standard Deviation: 22.465763813300406
Min: 0.0
Q1 quantile: 79.23560108840752
Q2 quantile: 90.94063222821896
Q3 quantile: 94.77160154100164
Max: 100.0
```

```
In [123]: desc(df_stats['Deaths'])
```

```
Deaths
Count: 221
Mean: 12732.502262443439
Standard Deviation: 49453.05612650376
Min: 0
Q1 quantile: 56.0
Q2 quantile: 536.0
Q3 quantile: 4585.0
Max: 563926
```

```
In [125]: desc(df_stats['Deaths/Cases'])
```

```
Deaths/Cases
Count: 221
Mean: 1.940678362827644
Standard Deviation: 2.384017192118266
Min: 0.0
Q1 quantile: 0.7595365375777073
Q2 quantile: 1.4604354753053639
Q3 quantile: 2.394331207421303
Max: 22.22222222222222
```

```
In [127]: desc(df_stats['Vaccinations'])
```

```
Vaccinations
Count: 221
Mean: 2553901.040723982
Standard Deviation: 13281198.043055674
Min: 0
Q1 quantile: 0.0
Q2 quantile: 37780.0
Q3 quantile: 475100.0
Max: 145812835
```

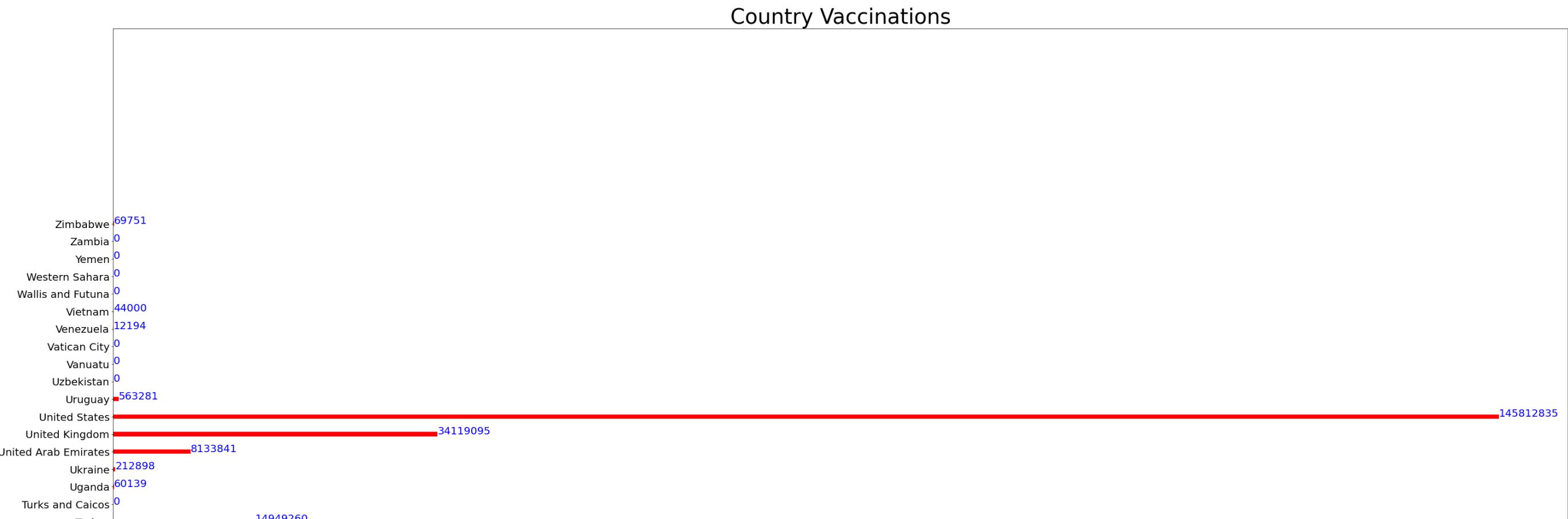
Countries that have more vaccinations than active cases.

```
In [128... pd.read_sql('''SELECT Country, Active_Cases, Vaccinations FROM country_stats  
WHERE Vaccinations>Active_Cases  
ORDER BY Vaccinations DESC;''', conn)
```

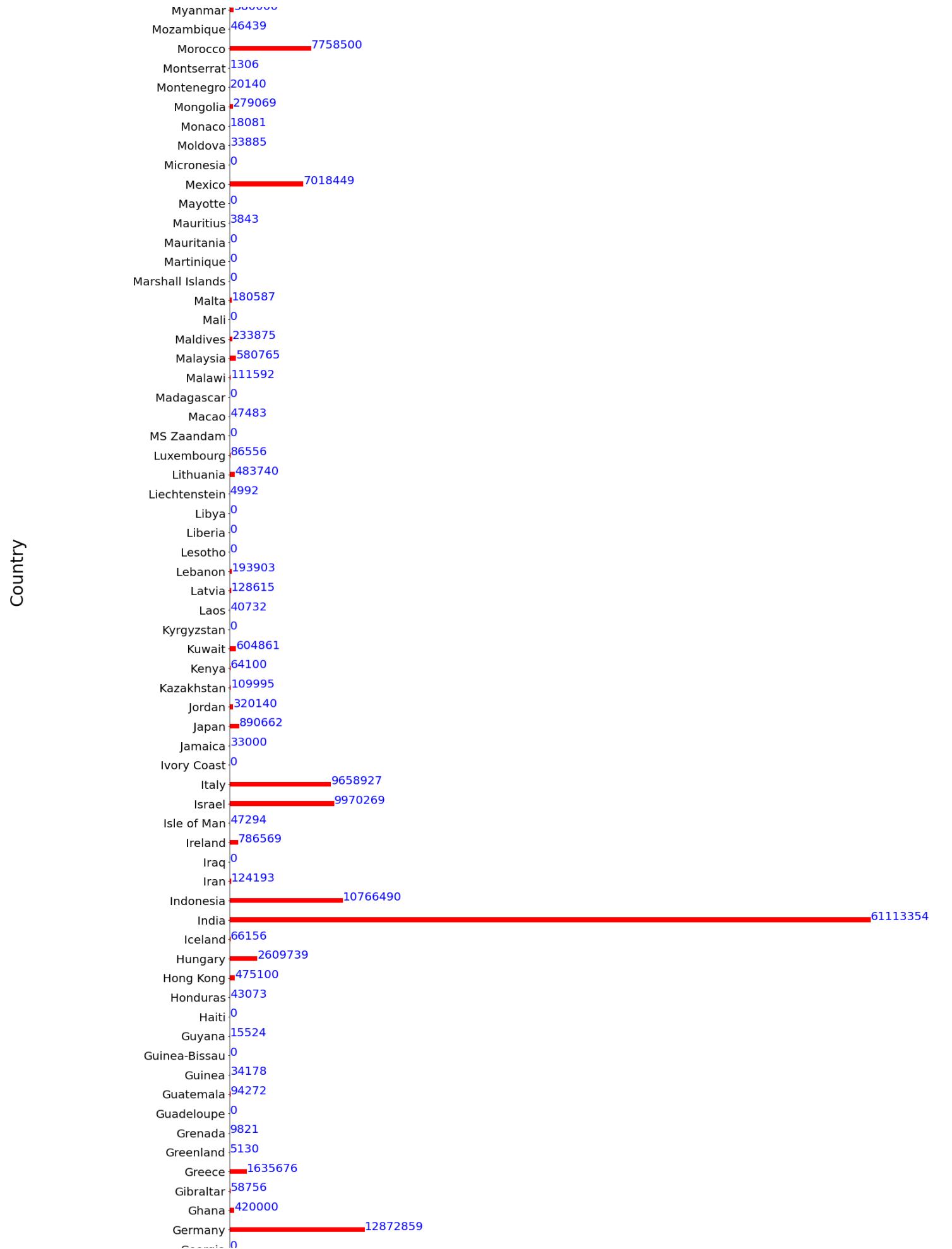
```
Out[128...  
Country Active_Cases Vaccinations  
0 United States 6948028 145812835  
1 China 173 110962000  
2 India 553874 61113354  
3 United Kingdom 379848 34119095  
4 Brazil 1371216 18082153  
... ... ...  
133 Liechtenstein 46 4992  
134 Mauritius 343 3843  
135 Falkland Islands 0 2187  
136 Montserrat 0 1306  
137 Trinidad and Tobago 245 991
```

138 rows × 3 columns

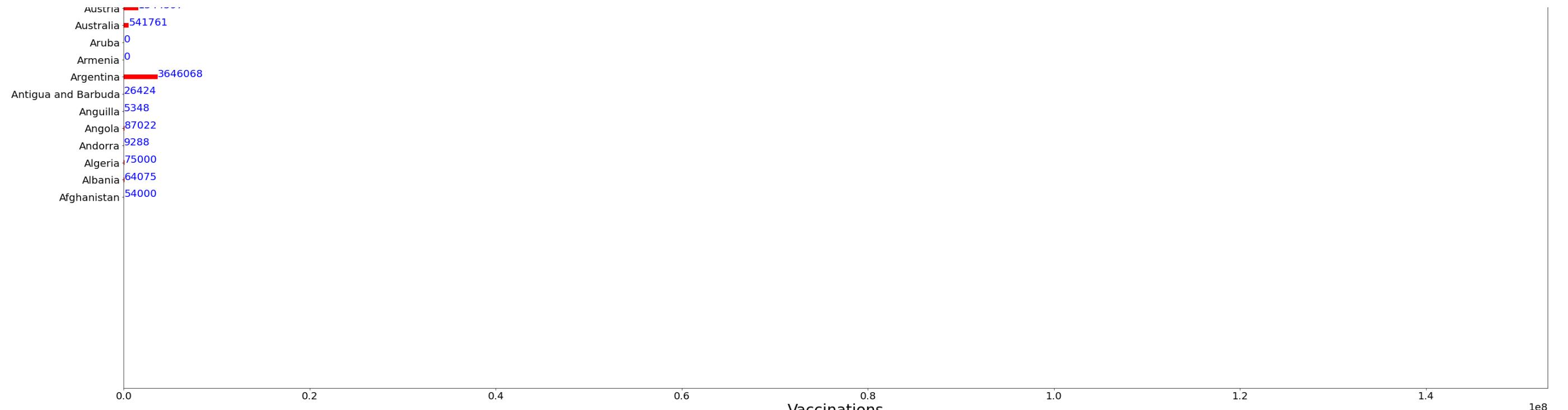
```
In [129... graph12=graph()  
graph12.bar_chart(df_stats['Country'], df_stats['Vaccinations'], 'Country Vaccinations', 'Vaccinations', 'Country')
```











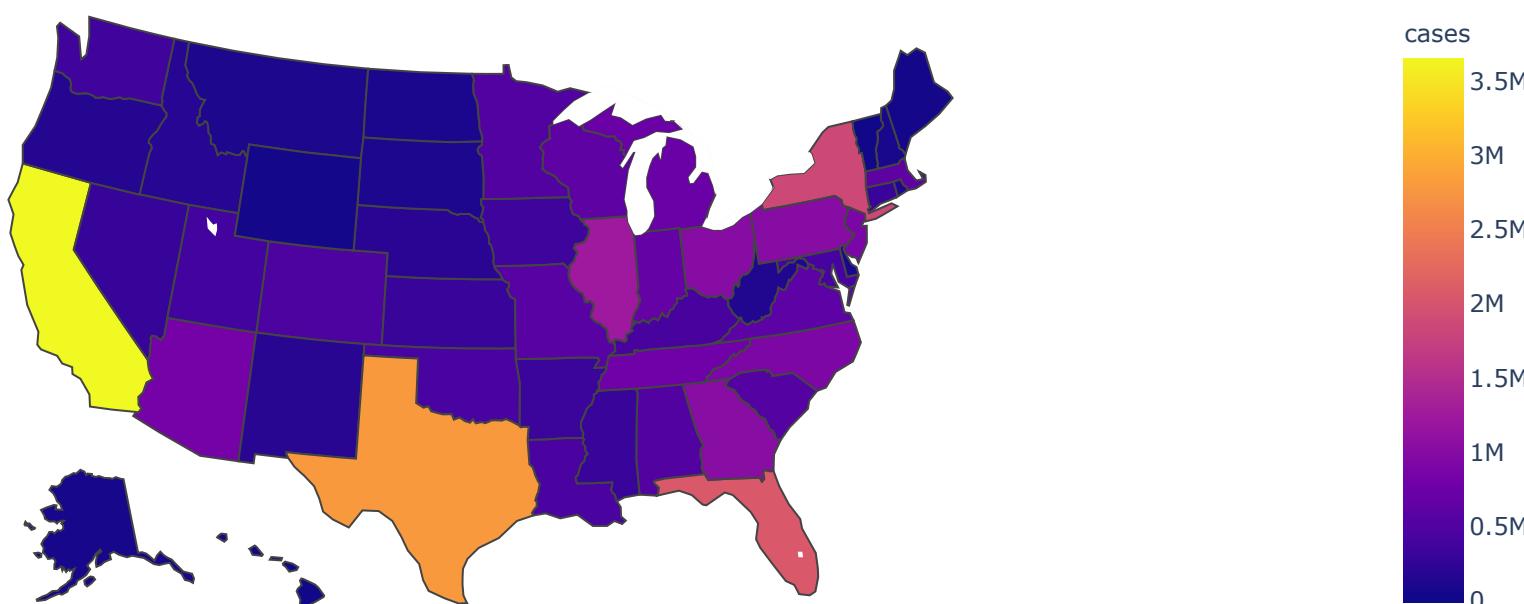
```
In [130]: desc(df_stats['Vaccinations/Population'])
```

Vaccinations/Population  
Count: 221  
Mean: 9.106300982497189  
Standard Deviation: 17.315025255455687  
Min: 0.0  
Q1 quantile: 0.0  
Q2 quantile: 0.5806679769595254  
Q3 quantile: 13.663205982033064  
Max: 99.99

```
In [132]: graph14=graph()  

graph14.choropleth(df6, 'State', 'USA-states', 'cases', 'date', 'usa', 'USA spread of Covid19')
```

USA spread of Covid19





```
In [133... desc(df_usa_stats['Population'])
```

```
Population
Count: 51
Mean: 6436069.0784313725
Standard Deviation: 7360660.467814472
Min: 578759
Q1 quantile: 1789606.0
Q2 quantile: 4467673.0
Q3 quantile: 7446805.0
Max: 39512223
```

```
In [135... desc(df_usa_stats['Tests'])
```

```
Tests
Count: 51
Mean: 7799215.019607843
Standard Deviation: 10344328.926277911
Min: 415165
Q1 quantile: 1974464.0
Q2 quantile: 4219331.0
Q3 quantile: 8555160.0
Max: 53684932
```

```
In [137... desc(df_usa_stats['Tests/Population'])
```

```
Tests/Population
Count: 51
Mean: 89.09704823192001
Standard Deviation: 17.95355830734586
Min: 45.12071132737494
Q1 quantile: 87.88821399793724
Q2 quantile: 99.99
Q3 quantile: 99.99
Max: 99.99
```

```
In [139... desc(df_usa_stats['Cases'])
```

```
Cases
Count: 51
Mean: 593752.9215686275
Standard Deviation: 696554.5233106406
Min: 19109
Q1 quantile: 152743.0
Q2 quantile: 409978.0
Q3 quantile: 712348.5
Max: 3665712
```

USA states that have more cases than the global average.

```
In [140... pd.read_sql('''SELECT State, us.Cases, AVG(c.Cases) AS AVG_Country_Cases FROM usa_state_stats AS us
    INNER JOIN country_stats AS c
    WHERE us.Cases>(SELECT AVG(Cases) FROM country_stats)''')
```

```
GROUP BY State  
ORDER BY us.Cases DESC;''' , conn)
```

Out[140...]

	State	Cases	AVG_Country_Cases
0	California	3665712	582507.746606
1	Texas	2792526	582507.746606
2	Florida	2052441	582507.746606
3	New York	1905416	582507.746606
4	Illinois	1241993	582507.746606
5	Georgia	1057741	582507.746606
6	Pennsylvania	1026364	582507.746606
7	Ohio	1015577	582507.746606
8	North Carolina	912203	582507.746606
9	New Jersey	905144	582507.746606
10	Arizona	841078	582507.746606
11	Tennessee	809692	582507.746606
12	Michigan	739244	582507.746606
13	Indiana	685453	582507.746606
14	Massachusetts	633081	582507.746606
15	Virginia	617941	582507.746606

In [142...]

```
desc(df_usa_stats['Active_Cases'])
```

Active\_Cases  
Count: 51  
Mean: 127061.0  
Standard Deviation: 273073.6189250071  
Min: 0  
Q1 quantile: 6728.0  
Q2 quantile: 22385.0  
Q3 quantile: 127955.0  
Max: 1668426

In [144...]

```
desc(df_usa_stats['Cases/Tests'])
```

Cases/Tests  
Count: 51  
Mean: 9.538101156508498  
Standard Deviation: 5.989973810731985  
Min: 1.4486194563337016  
Q1 quantile: 5.96619370943851  
Q2 quantile: 8.170816627620518  
Q3 quantile: 11.449936925406222  
Max: 26.48687885625917

In [146...]

```
desc(df_usa_stats['Recoveries'])
```

Recoveries  
Count: 51  
Mean: 444745.0  
Standard Deviation: 520181.4595664478

```
Min: 0
Q1 quantile: 93407.5
Q2 quantile: 292872.0
Q3 quantile: 619436.0
Max: 2645990
```

```
In [148... desc(df_usa_stats['Recoveries/Cases'])
```

```
Recoveries/Cases
Count: 51
Mean: 74.96929470760719
Standard Deviation: 30.26098848716949
Min: 0.0
Q1 quantile: 67.4453636906641
Q2 quantile: 90.46235058906976
Q3 quantile: 95.74451352367961
Max: 98.01744082576971
```

```
In [150... desc(df_usa_stats['Deaths'])
```

```
Deaths
Count: 51
Mean: 10758.843137254902
Standard Deviation: 12998.423240335806
Min: 225
Q1 quantile: 2147.5
Q2 quantile: 6208.0
Q3 quantile: 12562.5
Max: 58770
```

```
In [152... desc(df_usa_stats['Deaths/Cases'])
```

```
Deaths/Cases
Count: 51
Mean: 1.6933454018414336
Standard Deviation: 0.4916554344871608
Min: 0.5175012560710098
Q1 quantile: 1.4242130564196027
Q2 quantile: 1.6468785905783225
Q3 quantile: 1.9642166328700568
Max: 2.708658133793306
```

```
In [154... desc(df_usa_stats['Vaccinations'])
```

```
Vaccinations
Count: 51
Mean: 2289585.5490196077
Standard Deviation: 2534229.140847499
Min: 220735
Q1 quantile: 667738.0
Q2 quantile: 1585046.0
Q3 quantile: 2830693.5
Max: 13882984
```

```
In [156... desc(df_usa_stats['Vaccinations/Population'])
```

```
Vaccinations/Population
Count: 51
Mean: 37.3155827527672
Standard Deviation: 4.4378244287272155
Min: 27.415645020453645
Q1 quantile: 34.570600742724004
Q2 quantile: 36.95886329202409
Q3 quantile: 39.81495585780323
Max: 49.17911761044892
```

# Time Series Models

Time series assumes stationarity, a probability distribution that stays the same over time. The rolling mean, MA, and rolling standard deviation, MSD, are used to visually determine whether the data is stationary.

$$MA = \frac{1}{N} \sum_i^N p_i$$

$$MSD = \sqrt{\frac{1}{N} \sum_i^N (p_i - MA)^2}$$

The Dickey-Fuller test is a statistical test used to determine whether the data is stationary. The null hypothesis is that the data is non-stationary. If  $\gamma=0$ , then the data follows a random walk and if  $\gamma\neq 0$ , then the data is stationary.  $\varepsilon$  is white noise.

$$\Delta y_i = \alpha + \beta_i + \gamma y_{i-1} + \varepsilon_i$$

Autocorrelation function is used to determine lag correlation for an Auto Regressive model, p, and partial autocorrelation function is used to determine lag correlation for a Moving Average model, q. ACF solves for the correlation of series values with their lagged values. PACF solves for the correlation of the values after removing the effects of previous lags. ACF is covariance of residuals devided by standard deviation multiplied by standard deviation of lag.

$$cov(y_i, y_{i-k}) = \frac{\sum_i^{N-k} (y_{i+1} - \bar{y})(y_{i-k} - \bar{y})}{N - 1}$$

$$\sigma y_i = \sqrt{\frac{\sum_i^N (y_i - \bar{y})^2}{N - 1}}$$

$$ACF(y_i, y_{i-k}) = \frac{cov(y_i, y_{i-k})}{\sigma y_i * \sigma y_{i-k}}$$

$$PACF(y_i, y_{i-k}) = \frac{cov(y_i, y_{i-k} | y_{i+1} \dots y_{i-k-1})}{\sigma y_i | y_{i+1} \dots y_{i-k-1} * \sigma y_{i-k} | y_{i+1} \dots y_{i-k+1}}$$

AIC, akaike information criterion, estimates prediction error. k is the model parameters and  $\hat{L}$  is the model maximum value of the likelihood function.

$$AIC = 2k - 2 * ln(\hat{L})$$

An AR model uses lagged values to make forecasts. c is a constant, p is the number of data lags,  $\theta$  is the coefficients, y is the lags, and  $\Theta$  is an order p polynomial function of L. MA model uses residuals of previous forecasts to make new forecasts. q is the number of residual lags,  $\phi$  is the coefficients,  $\varepsilon$  is the residuals, and  $\Phi$  is an order q polynomial function of L. ARMA model adds the two models.

$$AR : y_t = c + \sum_k^p \phi_k * y_{t-k} + \varepsilon_t = \Phi(L)^p y_t + \varepsilon_t$$

$$MA : y_t = c + \sum_k^q \theta_k * \varepsilon_{t-k} + \varepsilon_t = \Theta(L)^q \varepsilon_t + \varepsilon_t$$

$$ARMA : y_t = \Phi(L)^p y_t + \Theta(L)^q \varepsilon_t + \varepsilon_t$$

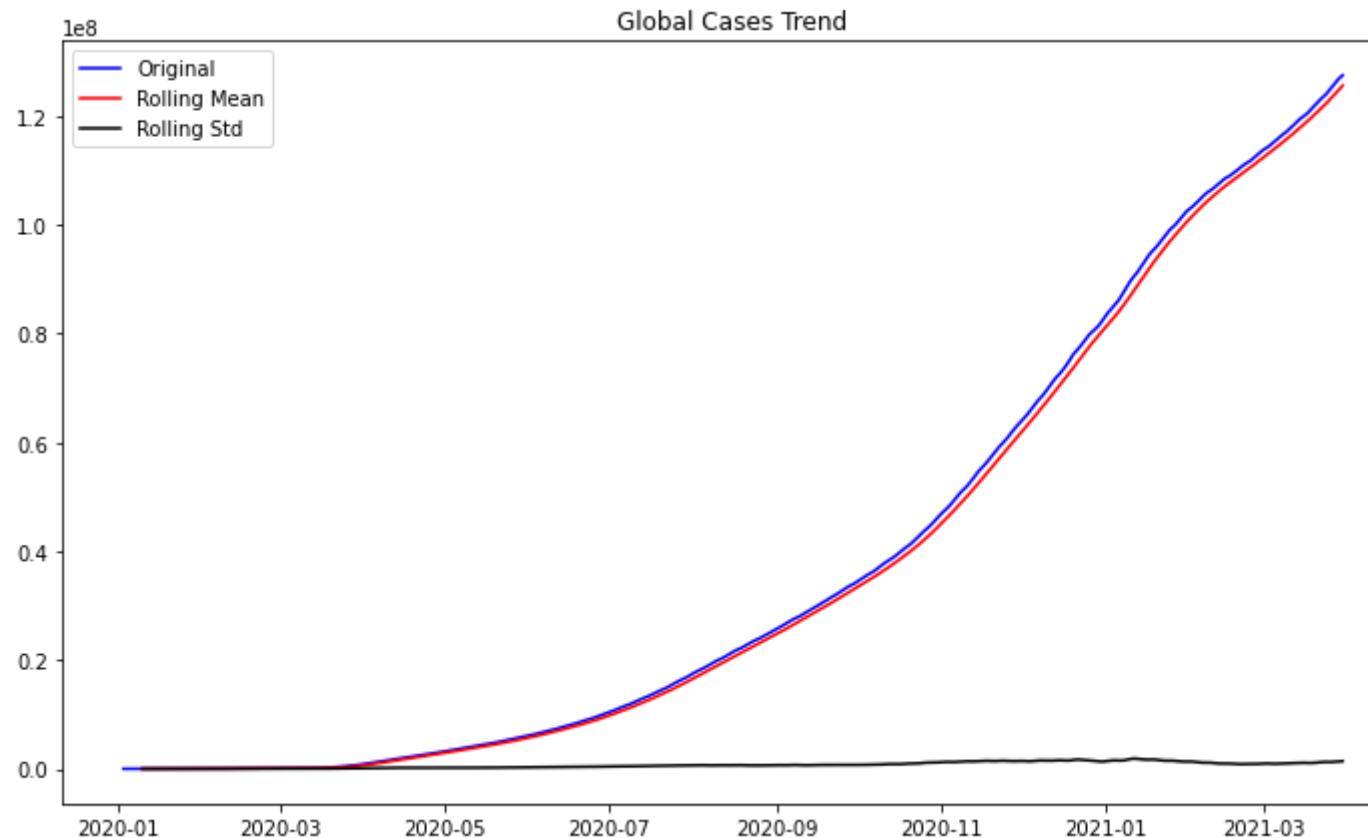
Remove trend by differencing data with lags. An ARIMA model I, integrated, component differences the data to make it stationary. A SARIMA model has a component to represent seasonality, s.

$$(1 - L)^d = y_t - y_{t-1}$$

$$ARIMA : (1 - \Phi(L)^p)(1 - L)^d y_t = (1 + \Theta(L)^q) \varepsilon_t$$

$$SARIMA : (1 - \Phi(L)^p)(1 - \Phi(L)^{ps})(1 - L)^{ds} y_t = (1 + \Theta(L)^q)(1 + \Theta(L)^{qs}) \varepsilon_t$$

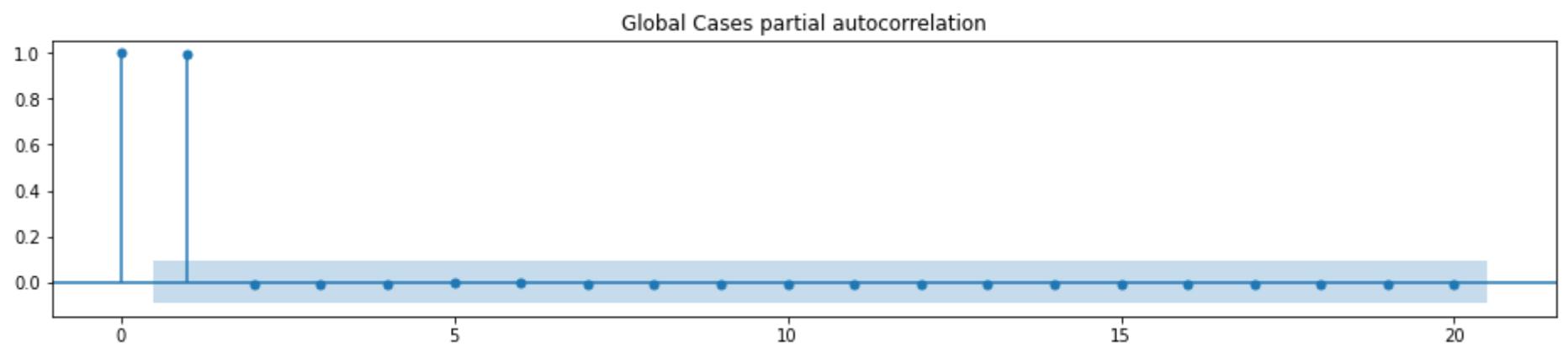
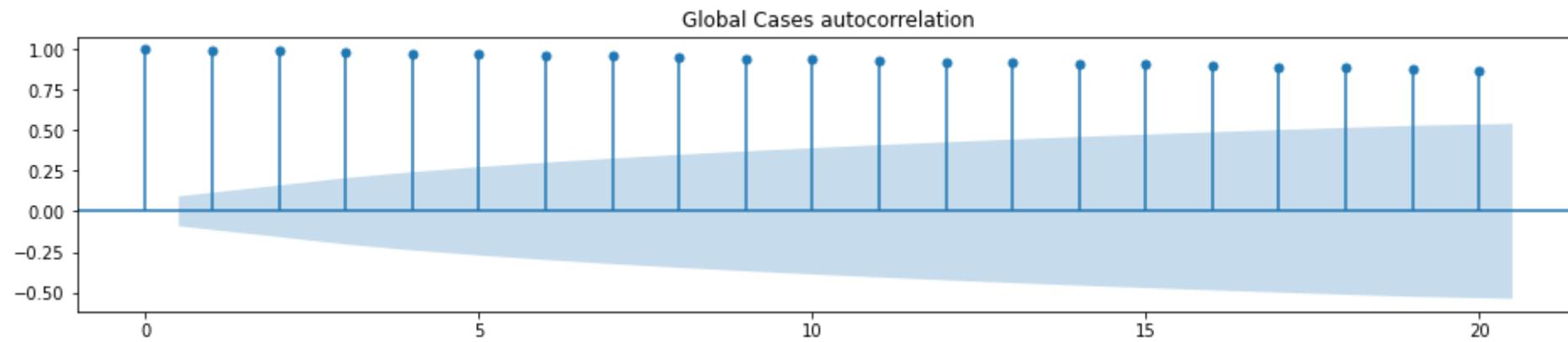
```
In [198]:  
start_new_cases={}
end_new_cases={}
forecast(global_cases_df, 8, start_new_cases, end_new_cases, 30, 20, '2021-03-30',
         '2021-04-30', 'Global Cases')
```



Global Cases Dickey-Fuller test results:

Test Statistic	0.503617
p-value	0.984983
#Lags Used	17.000000
Number of Observations Used	435.000000
Critical Value (1%)	-3.445473
Critical Value (5%)	-2.868207
Critical Value (10%)	-2.570321

dtype: float64



Global Cases AIC Scores:

```

ARIMA (0, 0, 0) x (0, 0, 0, 7)12 : AIC Calculated =17403.899242023348
ARIMA (0, 0, 0) x (0, 0, 1, 7)12 : AIC Calculated =16815.40297035946
ARIMA (0, 0, 0) x (0, 1, 0, 7)12 : AIC Calculated =14377.113223902381
ARIMA (0, 0, 0) x (0, 1, 1, 7)12 : AIC Calculated =13849.229962001154
ARIMA (0, 0, 0) x (1, 0, 0, 7)12 : AIC Calculated =13650.176776753855
ARIMA (0, 0, 0) x (1, 0, 1, 7)12 : AIC Calculated =13316.50049021042
ARIMA (0, 0, 0) x (1, 1, 0, 7)12 : AIC Calculated =12154.93594454918
ARIMA (0, 0, 0) x (1, 1, 1, 7)12 : AIC Calculated =12045.53470519615
ARIMA (0, 0, 1) x (0, 0, 0, 7)12 : AIC Calculated =17045.129572760714
ARIMA (0, 0, 1) x (0, 0, 1, 7)12 : AIC Calculated =16697.798772105678
ARIMA (0, 0, 1) x (0, 1, 0, 7)12 : AIC Calculated =14017.472418488542
ARIMA (0, 0, 1) x (0, 1, 1, 7)12 : AIC Calculated =14381.086125930502
ARIMA (0, 0, 1) x (1, 0, 0, 7)12 : AIC Calculated =16733.824253468272
ARIMA (0, 0, 1) x (1, 0, 1, 7)12 : AIC Calculated =16660.787579122243
ARIMA (0, 0, 1) x (1, 1, 0, 7)12 : AIC Calculated =13752.314567121779
ARIMA (0, 0, 1) x (1, 1, 1, 7)12 : AIC Calculated =13691.561195303377
ARIMA (0, 1, 0) x (0, 0, 0, 7)12 : AIC Calculated =12823.06146654599
ARIMA (0, 1, 0) x (0, 0, 1, 7)12 : AIC Calculated =12325.411132138272
ARIMA (0, 1, 0) x (0, 1, 0, 7)12 : AIC Calculated =10776.58589157331
ARIMA (0, 1, 0) x (0, 1, 1, 7)12 : AIC Calculated =10603.97322343279
ARIMA (0, 1, 0) x (1, 0, 0, 7)12 : AIC Calculated =10798.146526868799
ARIMA (0, 1, 0) x (1, 0, 1, 7)12 : AIC Calculated =10766.86341254904
ARIMA (0, 1, 0) x (1, 1, 0, 7)12 : AIC Calculated =10627.787681432274
ARIMA (0, 1, 0) x (1, 1, 1, 7)12 : AIC Calculated =10607.016417653387
ARIMA (0, 1, 1) x (0, 0, 0, 7)12 : AIC Calculated =12465.5307302228
ARIMA (0, 1, 1) x (0, 0, 1, 7)12 : AIC Calculated =12197.331271658437
ARIMA (0, 1, 1) x (0, 1, 0, 7)12 : AIC Calculated =10579.141363432998
ARIMA (0, 1, 1) x (0, 1, 1, 7)12 : AIC Calculated =10419.249428102168
ARIMA (0, 1, 1) x (1, 0, 0, 7)12 : AIC Calculated =12207.900672277858
ARIMA (0, 1, 1) x (1, 0, 1, 7)12 : AIC Calculated =12155.074916041805
ARIMA (0, 1, 1) x (1, 1, 0, 7)12 : AIC Calculated =10465.49039816989
ARIMA (0, 1, 1) x (1, 1, 1, 7)12 : AIC Calculated =10421.001459059305
ARIMA (1, 0, 0) x (0, 0, 0, 7)12 : AIC Calculated =12095.506402127294
ARIMA (1, 0, 0) x (0, 0, 1, 7)12 : AIC Calculated =11617.123282205253
ARIMA (1, 0, 0) x (0, 1, 0, 7)12 : AIC Calculated =10792.766057190805

```

```

ARIMA (1, 0, 0) x (0, 1, 1, 7)12 : AIC Calculated =10761.910459316105
ARIMA (1, 0, 0) x (1, 0, 0, 7)12 : AIC Calculated =11501.313830840205
ARIMA (1, 0, 0) x (1, 0, 1, 7)12 : AIC Calculated =11708.729752731037
ARIMA (1, 0, 0) x (1, 1, 0, 7)12 : AIC Calculated =10623.122098877706
ARIMA (1, 0, 0) x (1, 1, 1, 7)12 : AIC Calculated =10633.167454724146
ARIMA (1, 0, 1) x (0, 0, 0, 7)12 : AIC Calculated =11736.809039680335
ARIMA (1, 0, 1) x (0, 0, 1, 7)12 : AIC Calculated =11471.561578725867
ARIMA (1, 0, 1) x (0, 1, 0, 7)12 : AIC Calculated =10597.14431687535
ARIMA (1, 0, 1) x (0, 1, 1, 7)12 : AIC Calculated =10639.824541901668
ARIMA (1, 0, 1) x (1, 0, 0, 7)12 : AIC Calculated =11446.581812010412
ARIMA (1, 0, 1) x (1, 0, 1, 7)12 : AIC Calculated =11422.932254107087
ARIMA (1, 0, 1) x (1, 1, 0, 7)12 : AIC Calculated =10460.316968678078
ARIMA (1, 0, 1) x (1, 1, 1, 7)12 : AIC Calculated =10435.891254583414
ARIMA (1, 1, 0) x (0, 0, 0, 7)12 : AIC Calculated =10904.676245607701
ARIMA (1, 1, 0) x (0, 0, 1, 7)12 : AIC Calculated =10576.663481934826
ARIMA (1, 1, 0) x (0, 1, 0, 7)12 : AIC Calculated =10430.461050203881
ARIMA (1, 1, 0) x (0, 1, 1, 7)12 : AIC Calculated =10155.04060459443
ARIMA (1, 1, 0) x (1, 0, 0, 7)12 : AIC Calculated =10512.997016030013
ARIMA (1, 1, 0) x (1, 0, 1, 7)12 : AIC Calculated =10457.682020111657
ARIMA (1, 1, 0) x (1, 1, 0, 7)12 : AIC Calculated =10238.931717826596
ARIMA (1, 1, 0) x (1, 1, 1, 7)12 : AIC Calculated =10155.185847406947
ARIMA (1, 1, 1) x (0, 0, 0, 7)12 : AIC Calculated =10839.228119624611
ARIMA (1, 1, 1) x (0, 0, 1, 7)12 : AIC Calculated =10539.892757580637
ARIMA (1, 1, 1) x (0, 1, 0, 7)12 : AIC Calculated =10397.324362570864
ARIMA (1, 1, 1) x (0, 1, 1, 7)12 : AIC Calculated =10122.56527967197
ARIMA (1, 1, 1) x (1, 0, 0, 7)12 : AIC Calculated =10482.13939440128
ARIMA (1, 1, 1) x (1, 0, 1, 7)12 : AIC Calculated =10404.244885721017
ARIMA (1, 1, 1) x (1, 1, 0, 7)12 : AIC Calculated =10227.278224816539
ARIMA (1, 1, 1) x (1, 1, 1, 7)12 : AIC Calculated =10121.468270563942
Global Cases ARIMA:

```

#### SARIMAX Results

```

=====
Dep. Variable: Cumulative_cases No. Observations: 453
Model: SARIMAX(1, 1, 1)x(1, 1, 1, 7) Log Likelihood -5055.734
Date: Tue, 30 Mar 2021 AIC 10121.468
Time: 20:09:35 BIC 10141.856
Sample: 01-03-2020 HQIC 10129.514
- 03-30-2021

```

Covariance Type: opg

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.9659	0.014	69.479	0.000	0.939	0.993
ma.L1	-0.2191	0.028	-7.882	0.000	-0.274	-0.165
ar.S.L7	0.1211	0.038	3.208	0.001	0.047	0.195
ma.S.L7	-0.8012	0.033	-24.074	0.000	-0.866	-0.736
sigma2	9.06e+08	1.09e-11	8.33e+19	0.000	9.06e+08	9.06e+08

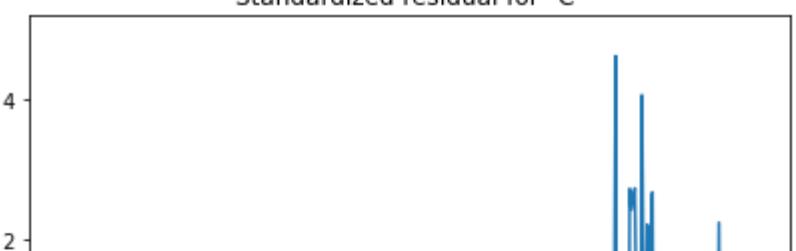
```

=====
Ljung-Box (L1) (Q): 0.06 Jarque-Bera (JB): 4635.09
Prob(Q): 0.81 Prob(JB): 0.00
Heteroskedasticity (H): 44.60 Skew: -1.02
Prob(H) (two-sided): 0.00 Kurtosis: 18.84
=====
```

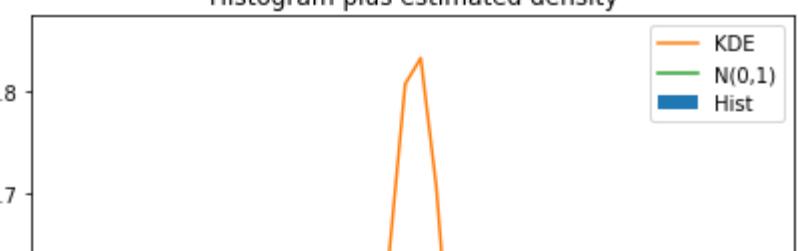
#### Warnings:

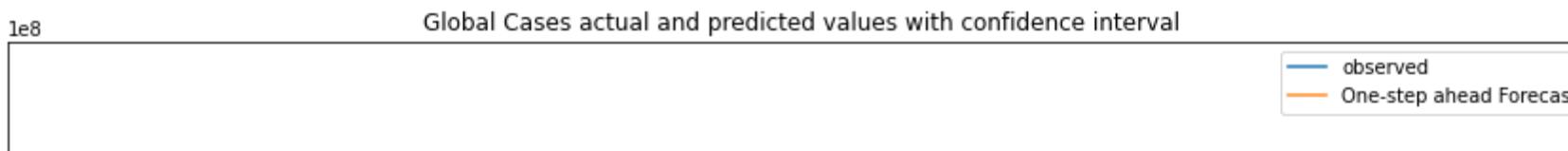
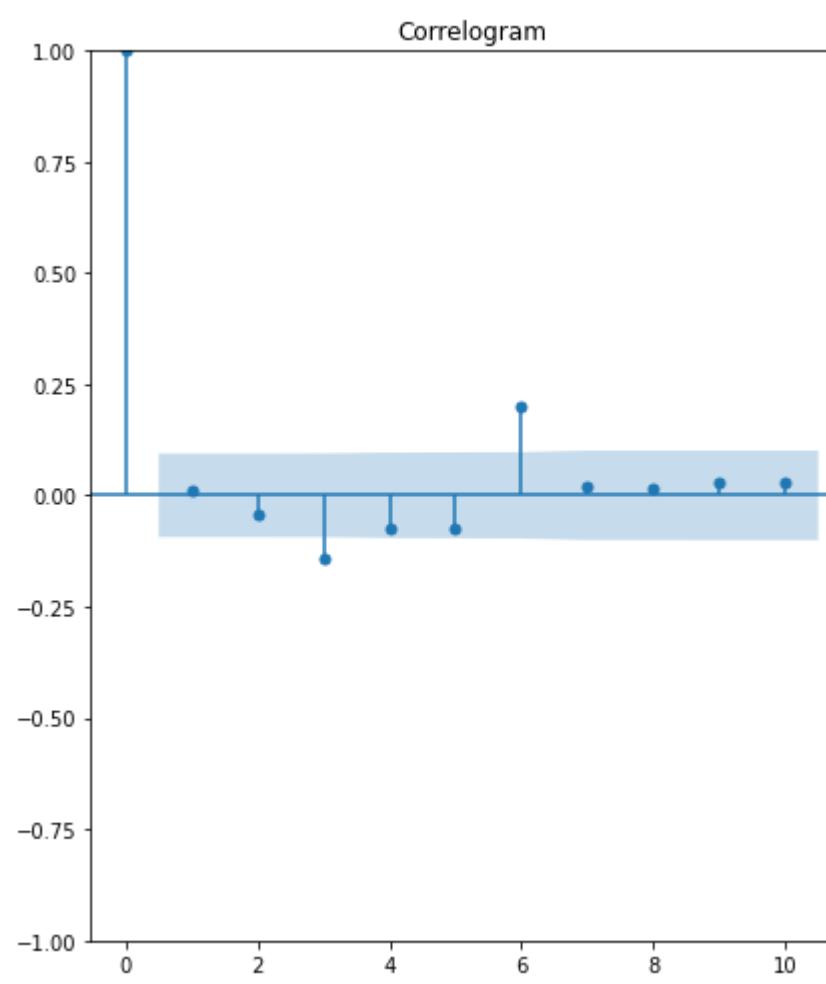
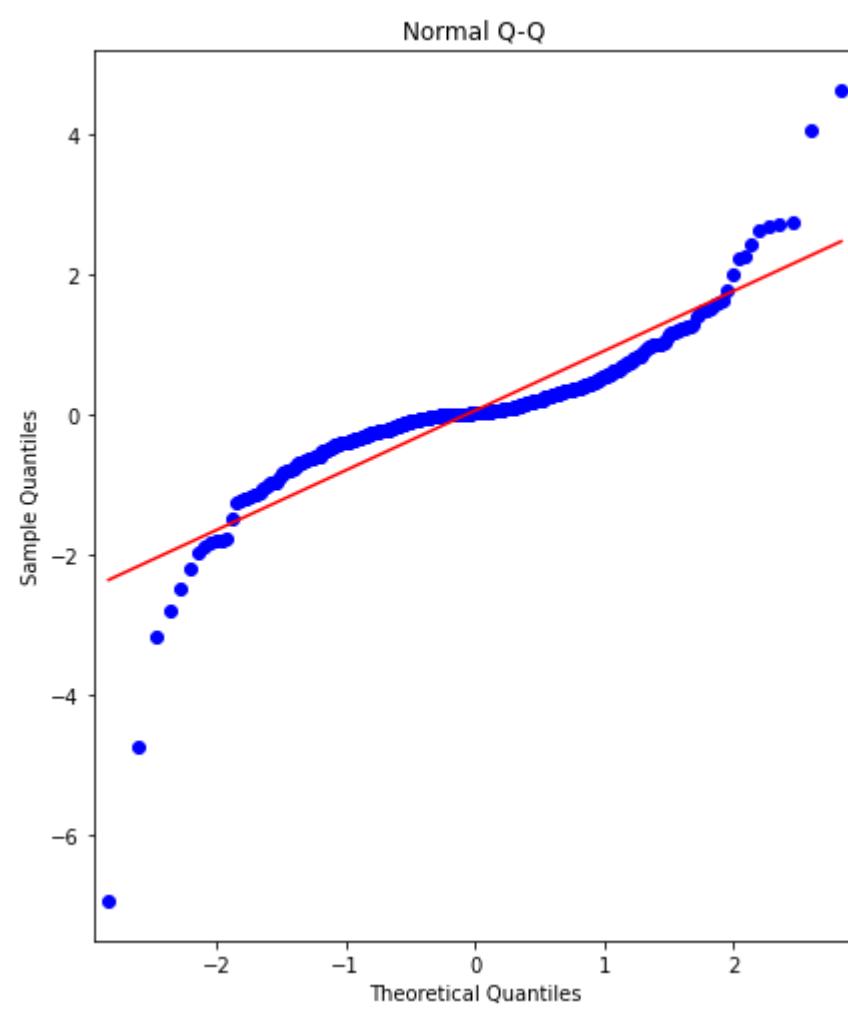
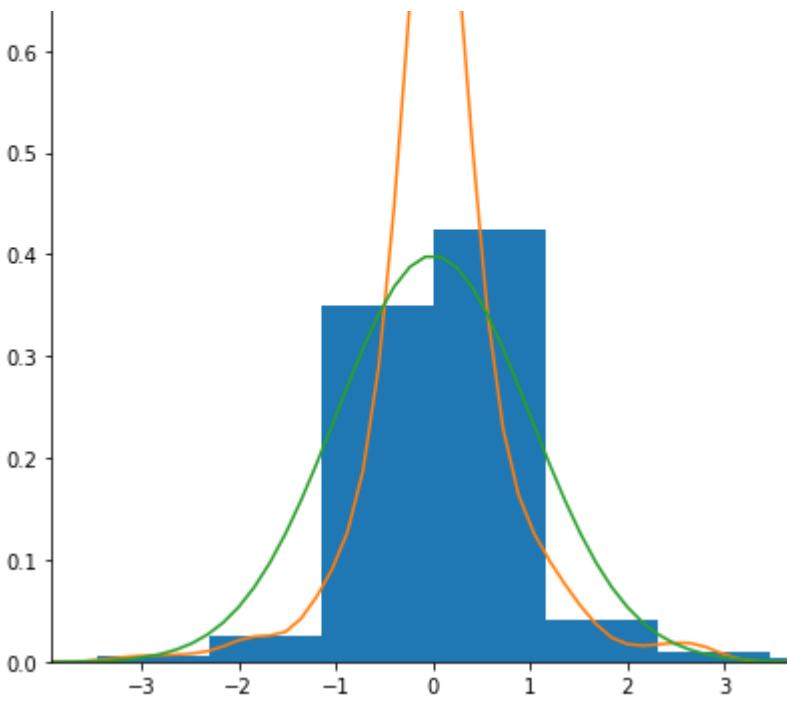
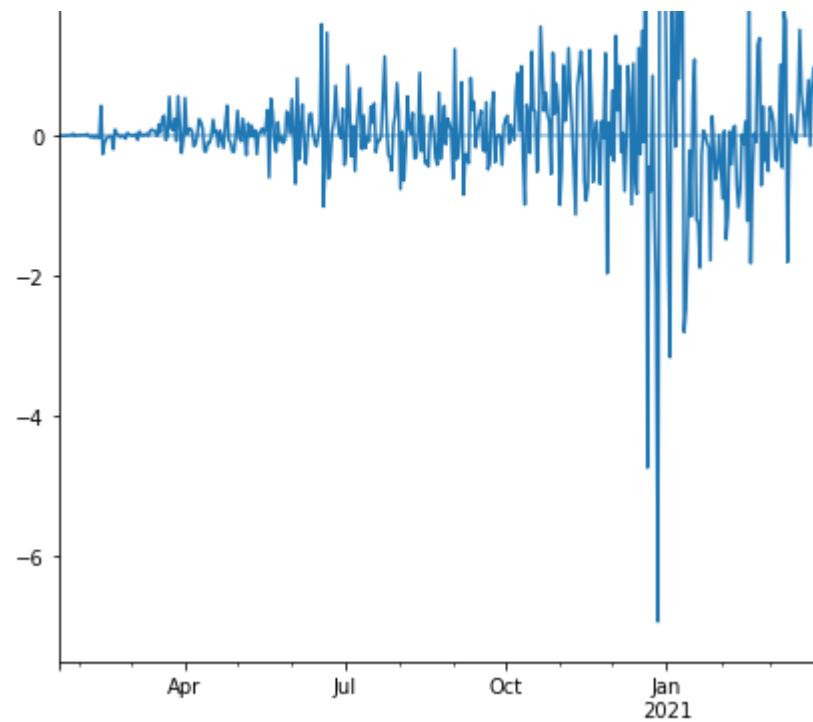
- [1] Covariance matrix calculated using the outer product of gradients (complex-step).
- [2] Covariance matrix is singular or near-singular, with condition number 1.53e+35. Standard errors may be unstable.

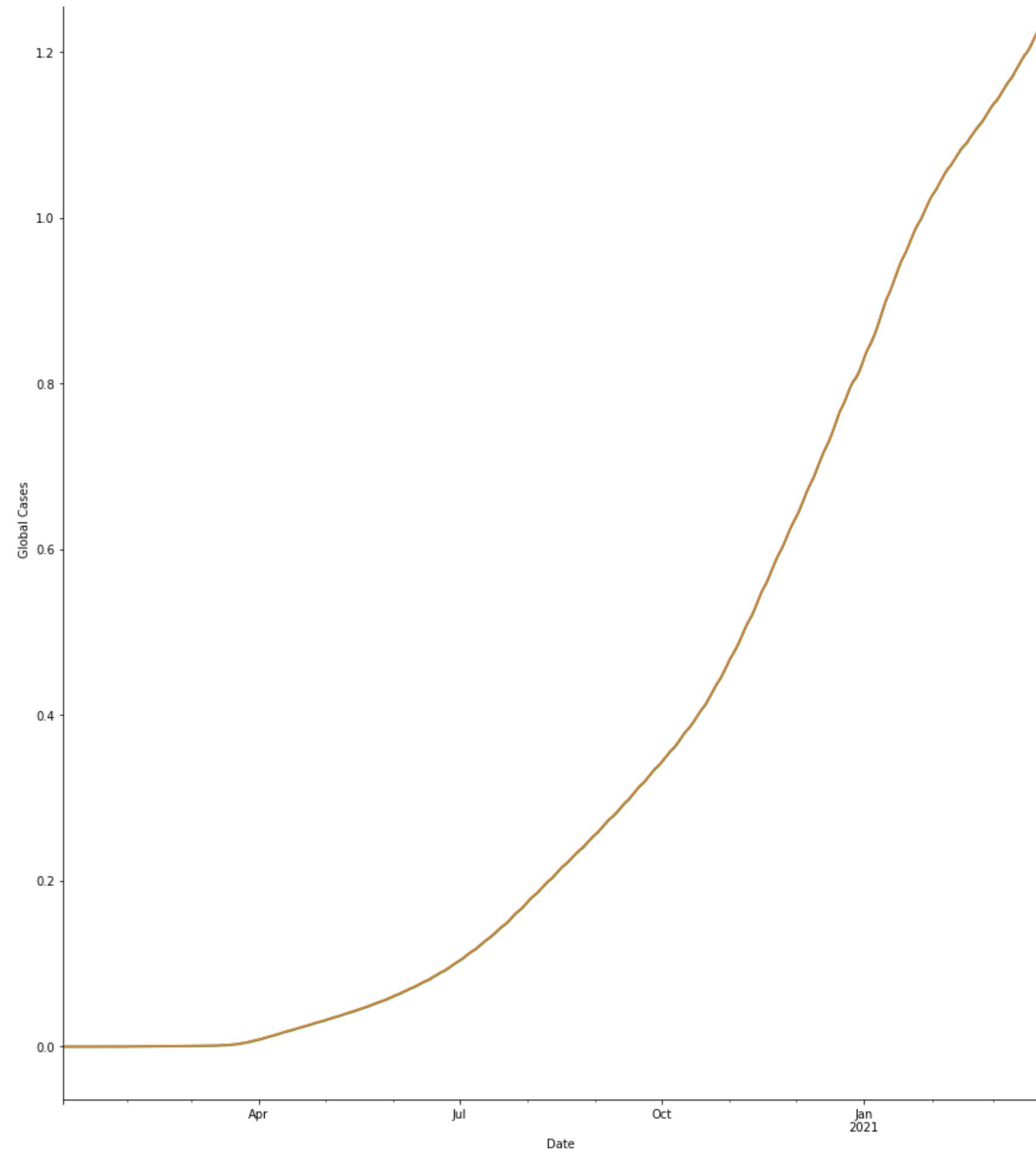
Standardized residual for "C"



Histogram plus estimated density





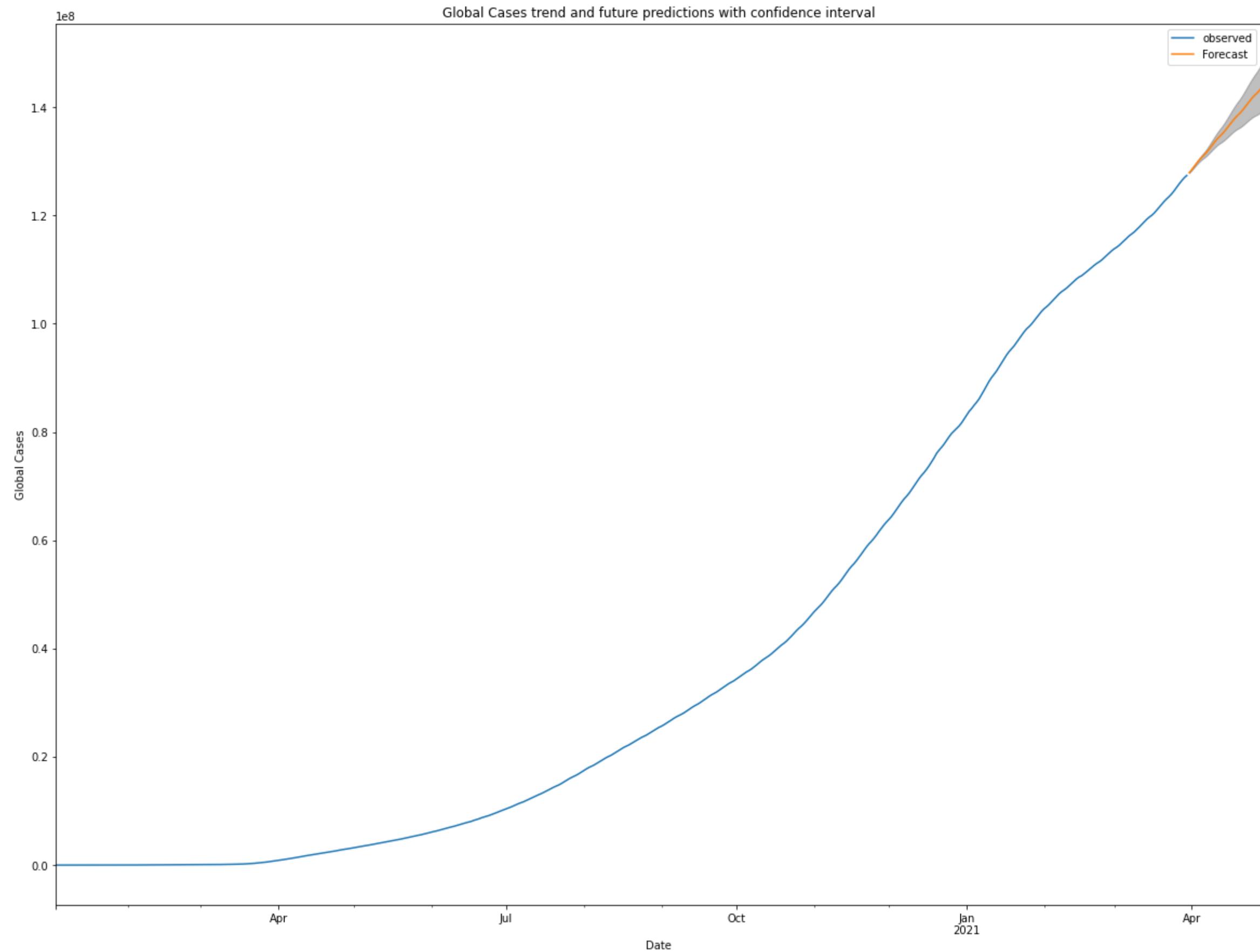


Global Cases Root Mean Squared Error:25212.24

Global Cases Forecast:

2021-03-31	1.279096e+08
2021-04-01	1.284911e+08
2021-04-02	1.290963e+08
2021-04-03	1.296977e+08
2021-04-04	1.302711e+08
2021-04-05	1.307797e+08
2021-04-06	1.312127e+08
2021-04-07	1.317172e+08
2021-04-08	1.322971e+08
2021-04-09	1.328986e+08
2021-04-10	1.334956e+08
2021-04-11	1.340667e+08
2021-04-12	1.345727e+08
2021-04-13	1.350054e+08
2021-04-14	1.355092e+08
2021-04-15	1.360882e+08
2021-04-16	1.366886e+08
2021-04-17	1.372844e+08
2021-04-18	1.378546e+08
2021-04-19	1.383596e+08
2021-04-20	1.387917e+08
2021-04-21	1.392949e+08
2021-04-22	1.398732e+08
2021-04-23	1.404730e+08
2021-04-24	1.410681e+08
2021-04-25	1.416376e+08
2021-04-26	1.421421e+08
2021-04-27	1.425737e+08
2021-04-28	1.430763e+08
2021-04-29	1.436541e+08

Freq: D, Name: predicted\_mean, dtype: float64



```
In [181]: total_cases=round(sum(start_new_cases.values()))
print('Total Global Cases:', total_cases)
```

Total Global Cases: 127405632

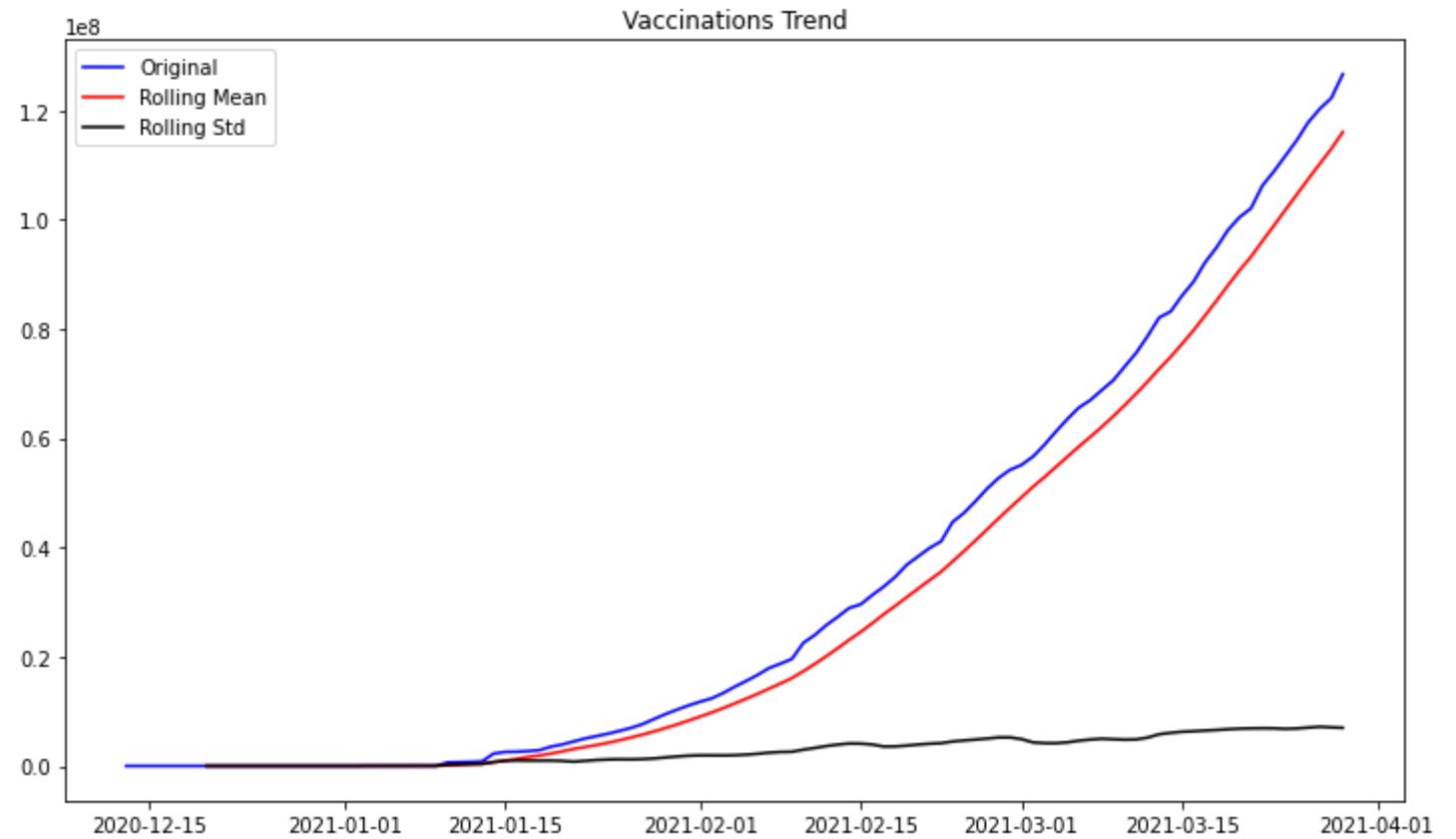
```
In [182]: forecasted_cases=round(sum(end_new_cases.values()))
print('Forecasted Global Cases:', forecasted_cases)
```

Forecasted Global Cases: 143654082

```
In [183]: percentage_change(total_cases,forecasted_cases)
```

Out[183]: 'Percentage Change: 12.75332161140255%

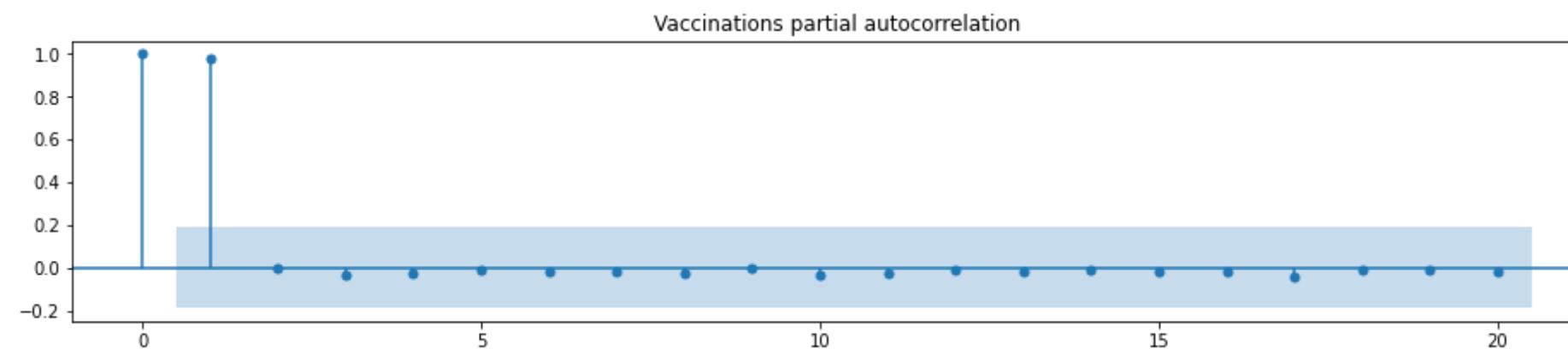
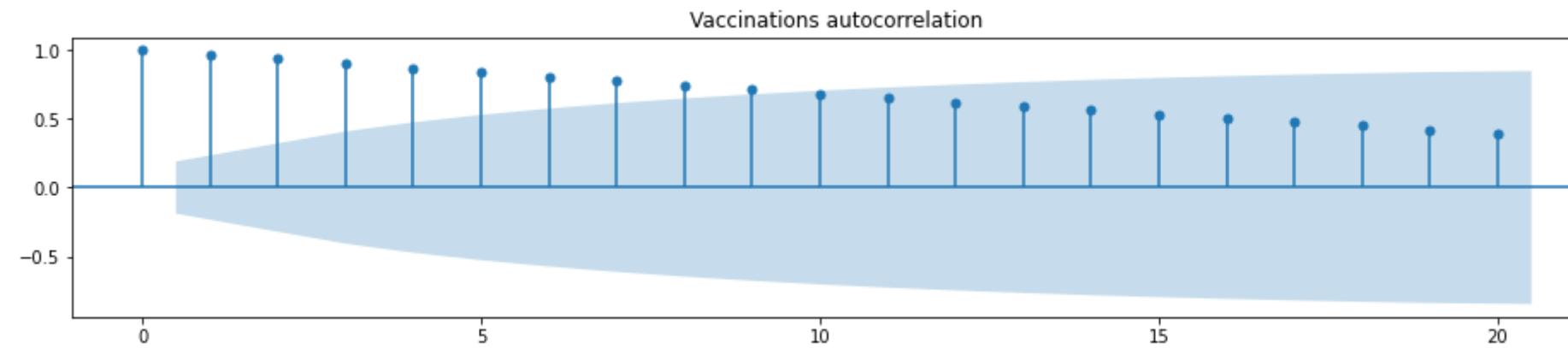
```
In [200]: start_new_vaccinations={}
end_new_vaccinations={}
forecast(global_vaccinations_df, 8, start_new_vaccinations, end_new_vaccinations, 100, 20, '2021-03-30',
        '2021-04-30', 'Vaccinations')
```



Vaccinations Dickey-Fuller test results:

Test Statistic	-0.148784
p-value	0.944357
#Lags Used	13.000000
Number of Observations Used	93.000000
Critical Value (1%)	-3.502705
Critical Value (5%)	-2.893158
Critical Value (10%)	-2.583637

dtype: float64



#### Vaccinations AIC Scores:

```

ARIMA (0, 0, 0) x (0, 0, 0, 7)12 : AIC Calculated =4062.6945152499584
ARIMA (0, 0, 0) x (0, 0, 1, 7)12 : AIC Calculated =3909.324132754518
ARIMA (0, 0, 0) x (0, 1, 0, 7)12 : AIC Calculated =3489.636233279091
ARIMA (0, 0, 0) x (0, 1, 1, 7)12 : AIC Calculated =3382.426219999726
ARIMA (0, 0, 0) x (1, 0, 0, 7)12 : AIC Calculated =3279.0291982809968
ARIMA (0, 0, 0) x (1, 0, 1, 7)12 : AIC Calculated =3213.4961322390564
ARIMA (0, 0, 0) x (1, 1, 0, 7)12 : AIC Calculated =2937.5999450237955
ARIMA (0, 0, 0) x (1, 1, 1, 7)12 : AIC Calculated =2908.486500657422
ARIMA (0, 0, 1) x (0, 0, 0, 7)12 : AIC Calculated =4135.931356968775
ARIMA (0, 0, 1) x (0, 0, 1, 7)12 : AIC Calculated =3672.964445340726
ARIMA (0, 0, 1) x (0, 1, 0, 7)12 : AIC Calculated =3408.735699880898
ARIMA (0, 0, 1) x (0, 1, 1, 7)12 : AIC Calculated =3332.3978323734805
ARIMA (0, 0, 1) x (1, 0, 0, 7)12 : AIC Calculated =3739.716430124371
ARIMA (0, 0, 1) x (1, 0, 1, 7)12 : AIC Calculated =3667.004382383141
ARIMA (0, 0, 1) x (1, 1, 0, 7)12 : AIC Calculated =3220.243922301542
ARIMA (0, 0, 1) x (1, 1, 1, 7)12 : AIC Calculated =3153.446286023534
ARIMA (0, 1, 0) x (0, 0, 0, 7)12 : AIC Calculated =3306.6590714846425
ARIMA (0, 1, 0) x (0, 0, 1, 7)12 : AIC Calculated =3043.9957464722975
ARIMA (0, 1, 0) x (0, 1, 0, 7)12 : AIC Calculated =2877.1974274122385
ARIMA (0, 1, 0) x (0, 1, 1, 7)12 : AIC Calculated =2680.1739692096808
ARIMA (0, 1, 0) x (1, 0, 0, 7)12 : AIC Calculated =2903.6258285163467
ARIMA (0, 1, 0) x (1, 0, 1, 7)12 : AIC Calculated =2869.4514253049806
ARIMA (0, 1, 0) x (1, 1, 0, 7)12 : AIC Calculated =2707.5938942512757
ARIMA (0, 1, 0) x (1, 1, 1, 7)12 : AIC Calculated =2681.9096495471676
ARIMA (0, 1, 1) x (0, 0, 0, 7)12 : AIC Calculated =3227.466246371288
ARIMA (0, 1, 1) x (0, 0, 1, 7)12 : AIC Calculated =2989.2824102109003
ARIMA (0, 1, 1) x (0, 1, 0, 7)12 : AIC Calculated =2847.954845950243
ARIMA (0, 1, 1) x (0, 1, 1, 7)12 : AIC Calculated =2650.3253090129356
ARIMA (0, 1, 1) x (1, 0, 0, 7)12 : AIC Calculated =3036.1509252229544
ARIMA (0, 1, 1) x (1, 0, 1, 7)12 : AIC Calculated =2976.3255101368723
ARIMA (0, 1, 1) x (1, 1, 0, 7)12 : AIC Calculated =2706.9865080760123
ARIMA (0, 1, 1) x (1, 1, 1, 7)12 : AIC Calculated =2652.2991780710367
ARIMA (1, 0, 0) x (0, 0, 0, 7)12 : AIC Calculated =3134.6761370697764
ARIMA (1, 0, 0) x (0, 0, 1, 7)12 : AIC Calculated =3083.9216295463784
ARIMA (1, 0, 0) x (0, 1, 0, 7)12 : AIC Calculated =2896.1191458545804

```

```

ARIMA (1, 0, 0) x (0, 1, 1, 7)12 : AIC Calculated =2858.018649047074
ARIMA (1, 0, 0) x (1, 0, 0, 7)12 : AIC Calculated =2895.430255723068
ARIMA (1, 0, 0) x (1, 0, 1, 7)12 : AIC Calculated =2896.2821222248967
ARIMA (1, 0, 0) x (1, 1, 0, 7)12 : AIC Calculated =2692.967709500886
ARIMA (1, 0, 0) x (1, 1, 1, 7)12 : AIC Calculated =2691.070461327456
ARIMA (1, 0, 1) x (0, 0, 0, 7)12 : AIC Calculated =3100.058362389429
ARIMA (1, 0, 1) x (0, 0, 1, 7)12 : AIC Calculated =3054.703340867389
ARIMA (1, 0, 1) x (0, 1, 0, 7)12 : AIC Calculated =2869.5523257820246
ARIMA (1, 0, 1) x (0, 1, 1, 7)12 : AIC Calculated =2833.8631400559684
ARIMA (1, 0, 1) x (1, 0, 0, 7)12 : AIC Calculated =2895.5285655732487
ARIMA (1, 0, 1) x (1, 0, 1, 7)12 : AIC Calculated =2868.4383230413046
ARIMA (1, 0, 1) x (1, 1, 0, 7)12 : AIC Calculated =2695.012456881652
ARIMA (1, 0, 1) x (1, 1, 1, 7)12 : AIC Calculated =2663.37878292511
ARIMA (1, 1, 0) x (0, 0, 0, 7)12 : AIC Calculated =3130.8936920301867
ARIMA (1, 1, 0) x (0, 0, 1, 7)12 : AIC Calculated =2903.8857196893064
ARIMA (1, 1, 0) x (0, 1, 0, 7)12 : AIC Calculated =2875.8183828679576
ARIMA (1, 1, 0) x (0, 1, 1, 7)12 : AIC Calculated =2678.0034468863114
ARIMA (1, 1, 0) x (1, 0, 0, 7)12 : AIC Calculated =2886.469252971434
ARIMA (1, 1, 0) x (1, 0, 1, 7)12 : AIC Calculated =2883.8502947874376
ARIMA (1, 1, 0) x (1, 1, 0, 7)12 : AIC Calculated =2678.0533461802247
ARIMA (1, 1, 0) x (1, 1, 1, 7)12 : AIC Calculated =2680.0027866635255
ARIMA (1, 1, 1) x (0, 0, 0, 7)12 : AIC Calculated =3044.5300839433175
ARIMA (1, 1, 1) x (0, 0, 1, 7)12 : AIC Calculated =2833.714951708625
ARIMA (1, 1, 1) x (0, 1, 0, 7)12 : AIC Calculated =2841.4966103238958
ARIMA (1, 1, 1) x (0, 1, 1, 7)12 : AIC Calculated =2633.0006021059266
ARIMA (1, 1, 1) x (1, 0, 0, 7)12 : AIC Calculated =2858.9092672695715
ARIMA (1, 1, 1) x (1, 0, 1, 7)12 : AIC Calculated =2832.2557867273804
ARIMA (1, 1, 1) x (1, 1, 0, 7)12 : AIC Calculated =2665.4648839673227
ARIMA (1, 1, 1) x (1, 1, 1, 7)12 : AIC Calculated =2630.715005164511
Vaccinations ARIMA:

```

#### SARIMAX Results

```

=====
Dep. Variable: people_fully_vaccinated No. Observations: 107
Model: SARIMAX(1, 1, 1)x(1, 1, 1, 7) Log Likelihood -1310.358
Date: Tue, 30 Mar 2021 AIC 2630.715
Time: 20:10:01 BIC 2643.214
Sample: 12-13-2020 HQIC 2635.755
- 03-29-2021

```

Covariance Type: opg

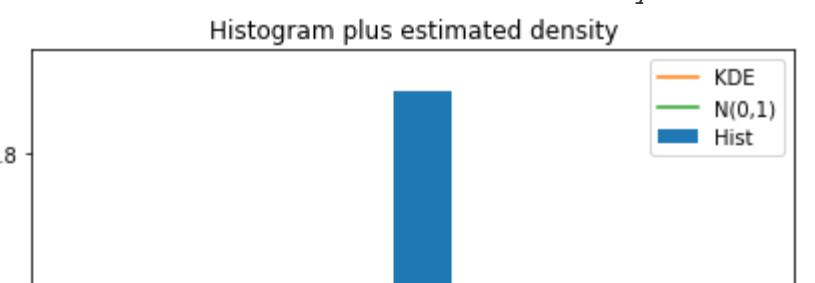
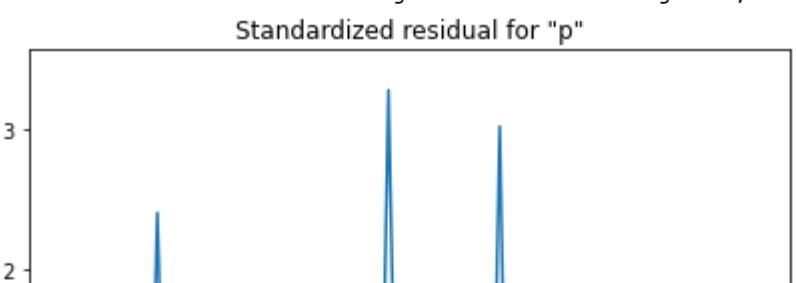
	coef	std err	z	P> z	[0.025	0.975]
ar.L1	1.0209	0.007	142.644	0.000	1.007	1.035
ma.L1	-0.9171	0.075	-12.161	0.000	-1.065	-0.769
ar.S.L7	0.5873	0.137	4.298	0.000	0.319	0.855
ma.S.L7	-1.0629	0.094	-11.290	0.000	-1.247	-0.878
sigma2	3.181e+11	4.17e-14	7.63e+24	0.000	3.18e+11	3.18e+11

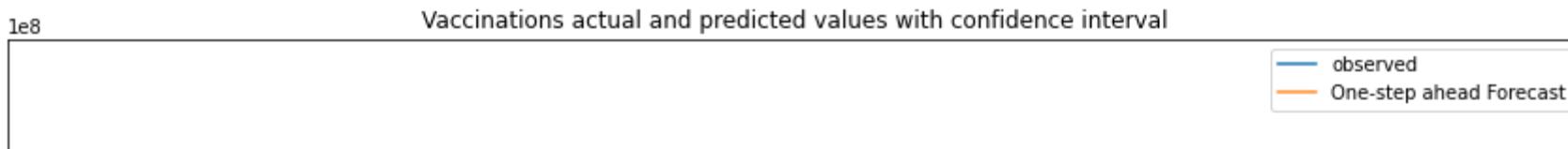
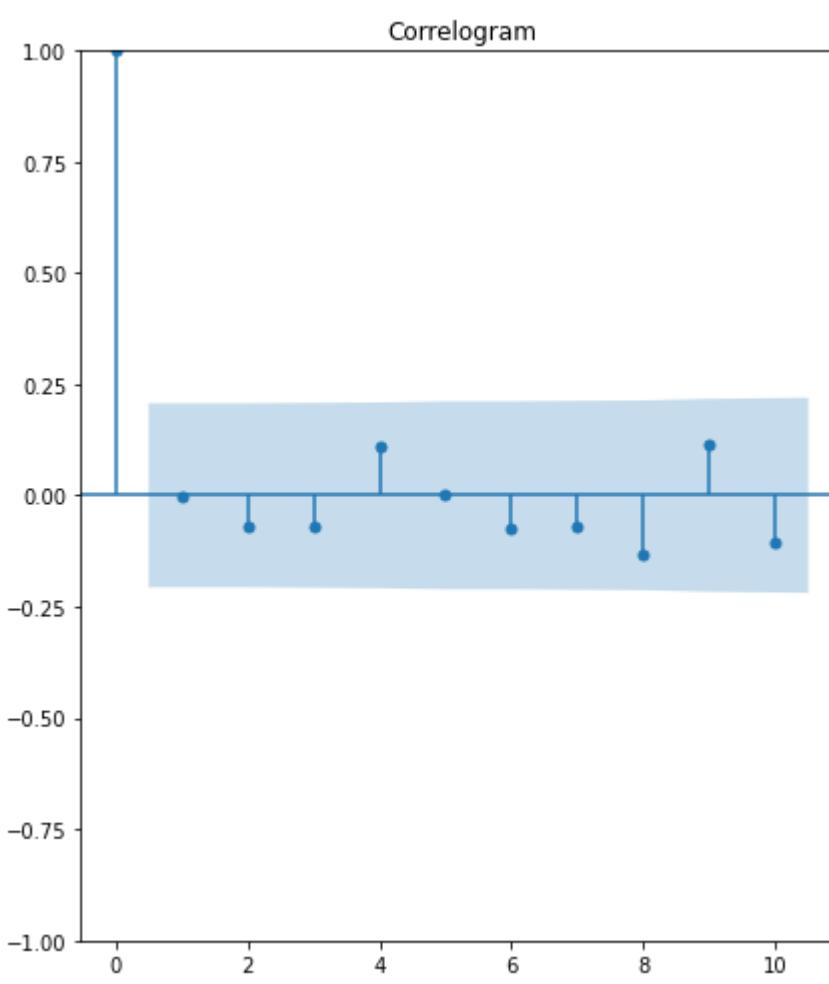
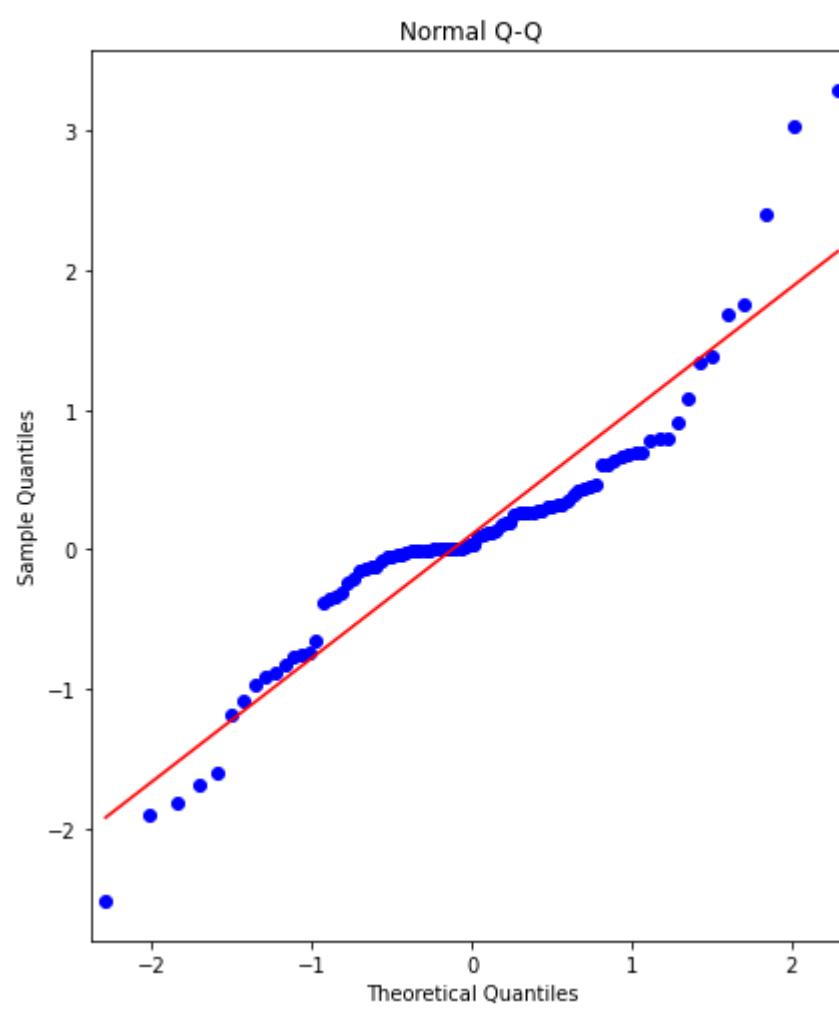
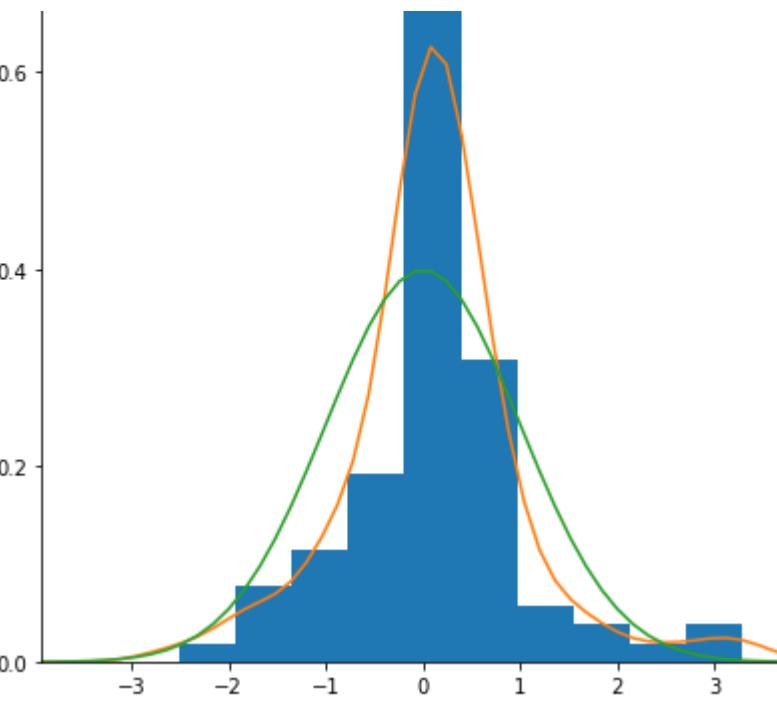
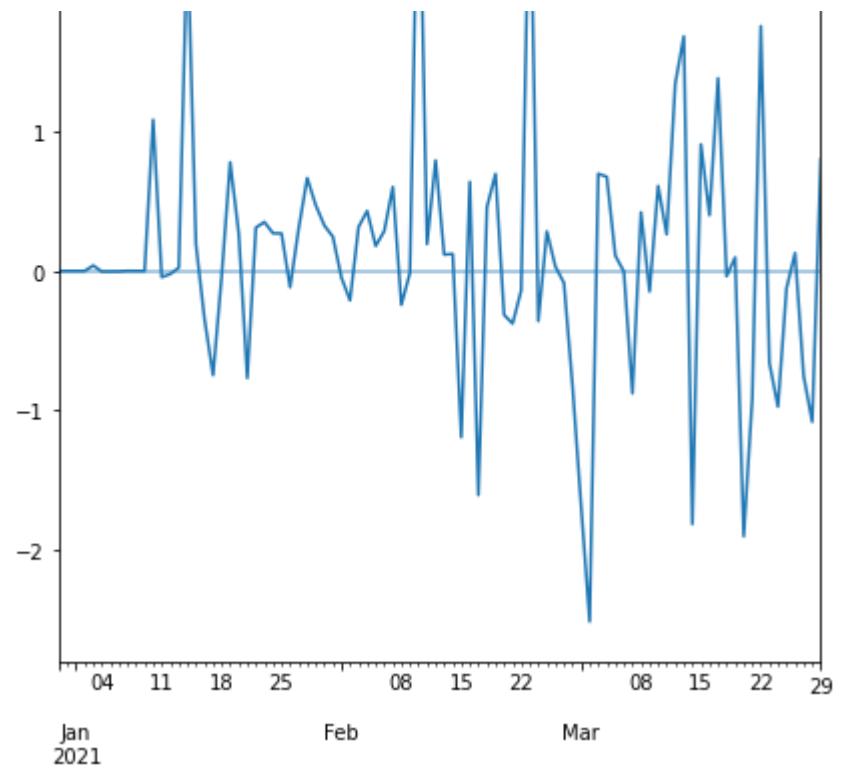
```

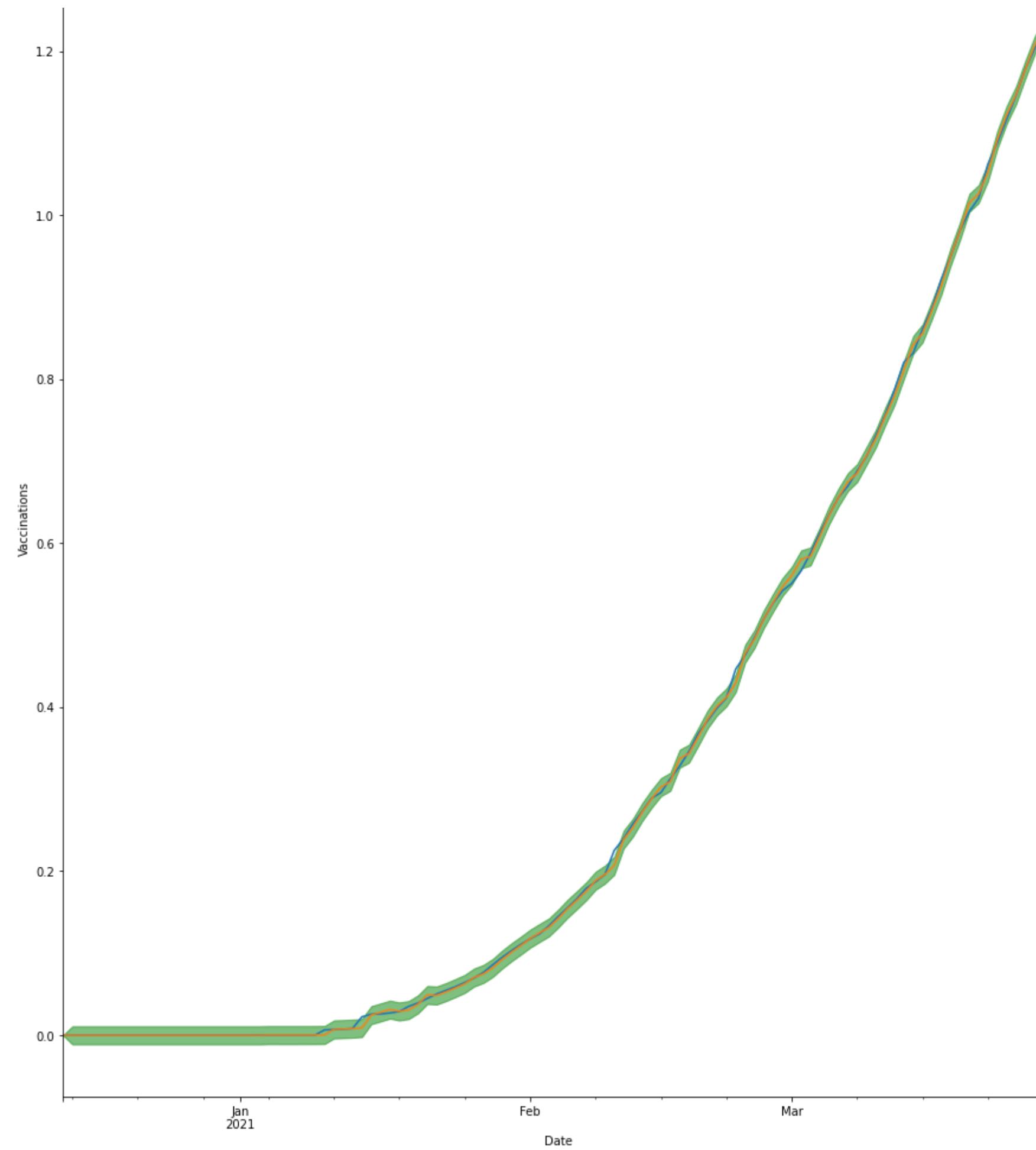
=====
Ljung-Box (L1) (Q): 0.00 Jarque-Bera (JB): 36.71
Prob(Q): 0.97 Prob(JB): 0.00
Heteroskedasticity (H): 3.47 Skew: 0.49
Prob(H) (two-sided): 0.00 Kurtosis: 5.97
=====
```

#### Warnings:

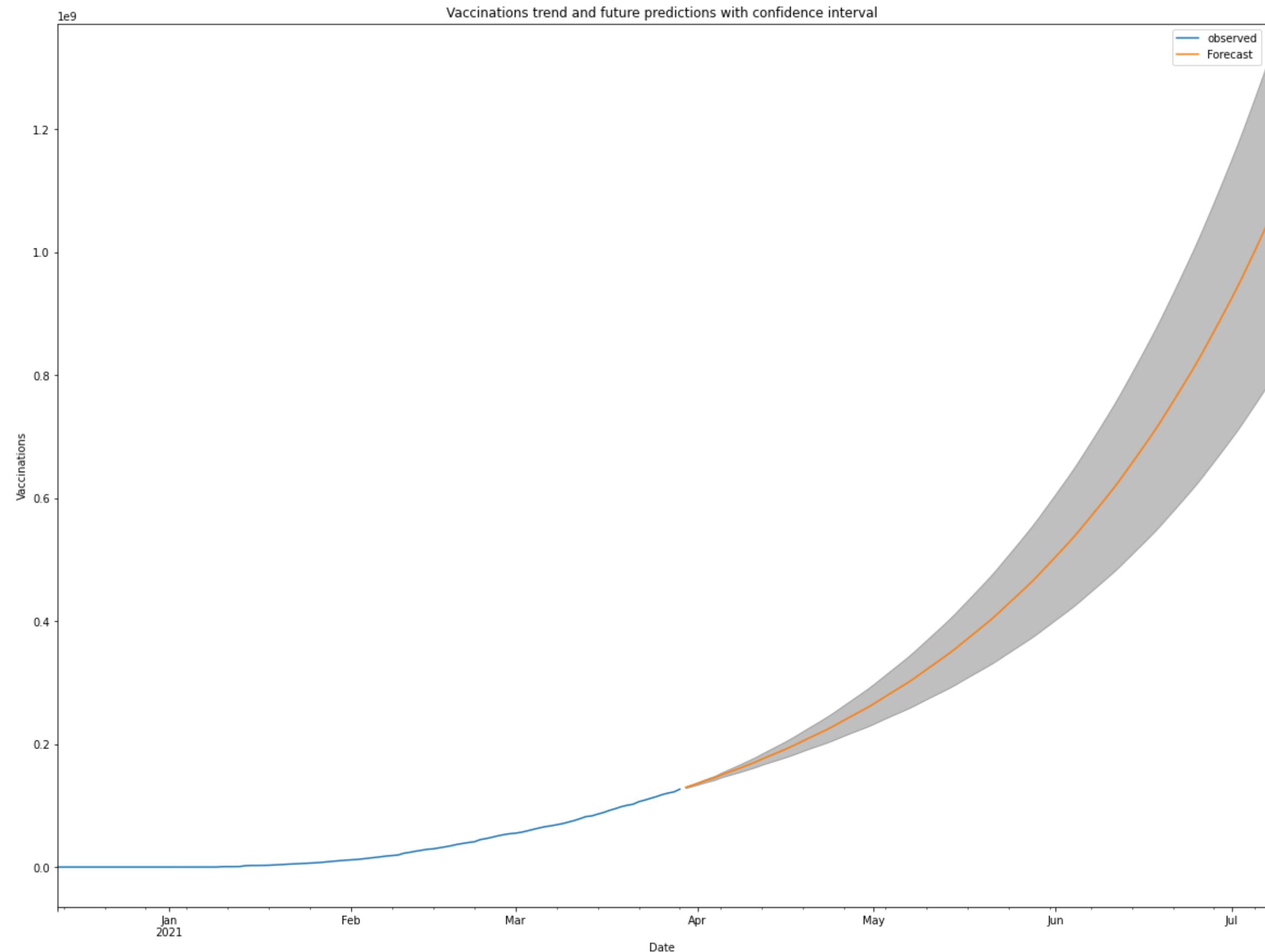
- [1] Covariance matrix calculated using the outer product of gradients (complex-step).
- [2] Covariance matrix is singular or near-singular, with condition number 8.43e+40. Standard errors may be unstable.







```
Vaccinations Root Mean Squared Error:461358.01
Vaccinations Forecast:
2021-03-30    1.296501e+08
2021-03-31    1.327731e+08
2021-04-01    1.358942e+08
2021-04-02    1.392879e+08
2021-04-03    1.423837e+08
...
2021-07-03    9.628710e+08
2021-07-04    9.829492e+08
2021-07-05    1.003047e+09
2021-07-06    1.023523e+09
2021-07-07    1.044179e+09
Freq: D, Name: predicted_mean, Length: 100, dtype: float64
```



```
In [201]: total_vaccinations=round(sum(start_new_vaccinations.values()))
print('Total Global Cases:', total_vaccinations)
```

Total Global Cases: 126616238

```
In [202]: forecasted_vaccinations=round(sum(end_new_vaccinations.values()))
print('Total Global Cases:', forecasted_vaccinations)
```

```
Total Global Cases: 1044179084
```

```
In [203]: percentage_change(total_vaccinations,forecasted_vaccinations)
```

```
Out[203]: 'Percentage Change: 724.6802309826959%
```

## Conclusion

The time-series models suggest that, over the next month, vaccinations will trend upward by 724.7% and cases of the virus will trend upward by 12.8%. The virus spread is slowing and will decrease more as more people get vaccinated.

```
In [ ]:
```