week8pythonassignment

May 14, 2025

[2]:
```
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

[ ]:
```
# 1.Data Loading & Exploration
```

[3]:
```
# Load the dataset
try:
    df = pd.read_csv('owid-covid-data.csv')
    print('Database loaded successfully')
except FileNotFoundError:
    print('Error: File not found. Check file path')
```

```
Database loaded successfully
```

[4]:
```
# Check for columns
print("\nThe columns in the database include:")
df.columns
```

```

The columns in the database include:
```

[4]:
```
Index(['iso_code', 'continent', 'location', 'date', 'total_cases', 'new_cases',
       'new_cases_smoothed', 'total_deaths', 'new_deaths',
       'new_deaths_smoothed', 'total_cases_per_million',
       'new_cases_per_million', 'new_cases_smoothed_per_million',
       'total_deaths_per_million', 'new_deaths_per_million',
       'new_deaths_smoothed_per_million', 'reproduction_rate', 'icu_patients',
       'icu_patients_per_million', 'hosp_patients',
       'hosp_patients_per_million', 'weekly_icu_admissions',
       'weekly_icu_admissions_per_million', 'weekly_hosp_admissions',
       'weekly_hosp_admissions_per_million', 'total_tests', 'new_tests',
       'total_tests_per_thousand', 'new_tests_per_thousand',
       'new_tests_smoothed', 'new_tests_smoothed_per_thousand',
       'positive_rate', 'tests_per_case', 'tests_units', 'total_vaccinations',
       'people_vaccinated', 'people_fully_vaccinated', 'total_boosters',
       'new_vaccinations', 'new_vaccinations_smoothed',
       'total_vaccinations_per_hundred', 'people_vaccinated_per_hundred',
       'people_fully_vaccinated_per_hundred', 'total_boosters_per_hundred',
       'new_vaccinations_smoothed_per_million',
       'new_people_vaccinated_smoothed',
       'new_people_vaccinated_smoothed_per_hundred', 'stringency_index',
       'population_density', 'median_age', 'aged_65_older', 'aged_70_older',
       'gdp_per_capita', 'extreme_poverty', 'cardiovasc_death_rate',
       'diabetes_prevalence', 'female_smokers', 'male_smokers',
       'handwashing_facilities', 'hospital_beds_per_thousand',
       'life_expectancy', 'human_development_index', 'population',
       'excess_mortality_cumulative_absolute', 'excess_mortality_cumulative',
       'excess_mortality', 'excess_mortality_cumulative_per_million'],
      dtype='object')
```

[5]:

```
# Preview rows
print("\nThese first five rows of the database include:")
df.head()
```

```

These first five rows of the database include:
```

[5]:

```
  iso_code continent     location        date  total_cases  new_cases  \
0      AFG      Asia  Afghanistan  2020-01-03          NaN        0.0
1      AFG      Asia  Afghanistan  2020-01-04          NaN        0.0
2      AFG      Asia  Afghanistan  2020-01-05          NaN        0.0
3      AFG      Asia  Afghanistan  2020-01-06          NaN        0.0
4      AFG      Asia  Afghanistan  2020-01-07          NaN        0.0

   new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  …  \
0                 NaN           NaN         0.0                  NaN  …
1                 NaN           NaN         0.0                  NaN  …
2                 NaN           NaN         0.0                  NaN  …
3                 NaN           NaN         0.0                  NaN  …
4                 NaN           NaN         0.0                  NaN  …

   male_smokers  handwashing_facilities  hospital_beds_per_thousand  \
0           NaN                  37.746                         0.5
1           NaN                  37.746                         0.5
2           NaN                  37.746                         0.5
3           NaN                  37.746                         0.5
4           NaN                  37.746                         0.5

   life_expectancy  human_development_index  population  \
0            64.83                    0.511  41128772.0
1            64.83                    0.511  41128772.0
2            64.83                    0.511  41128772.0
3            64.83                    0.511  41128772.0
4            64.83                    0.511  41128772.0
```

```
   excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
0                                   NaN                          NaN
1                                   NaN                          NaN
2                                   NaN                          NaN
3                                   NaN                          NaN
4                                   NaN                          NaN

   excess_mortality  excess_mortality_cumulative_per_million
0               NaN                                      NaN
1               NaN                                      NaN
2               NaN                                      NaN
3               NaN                                      NaN
4               NaN                                      NaN

[5 rows x 67 columns]
```

[6]:

```python
# Identify the missing values
df.isnull().sum()
```

[6]:

```
iso_code                                       0
continent                                  16665
location                                       0
date                                           0
total_cases                                37997
                                           …
population                                     0
excess_mortality_cumulative_absolute      337901
excess_mortality_cumulative               337901
excess_mortality                          337901
excess_mortality_cumulative_per_million   337901
Length: 67, dtype: int64
```

[ ]:

```python

```

[ ]:

```python
# 2. Data Cleaning
```

[7]:

```python
# Filter countries of interest (e.g., Kenya, USA, India)
countries_of_interest = ['Kenya', 'Gibraltar']
print("\nThese are the filtered countries:")
df[df['location'].isin(countries_of_interest)]
```

```

These are the filtered countries:
```

[7]:

```
       iso_code continent   location        date  total_cases  new_cases  \
116295      GIB    Europe  Gibraltar  2020-01-03          NaN        0.0
```

```
116296      GIB    Europe  Gibraltar  2020-01-04          NaN        0.0
116297      GIB    Europe  Gibraltar  2020-01-05          NaN        0.0
116298      GIB    Europe  Gibraltar  2020-01-06          NaN        0.0
116299      GIB    Europe  Gibraltar  2020-01-07          NaN        0.0
…         …       …        …         …          …        …
159167      KEN    Africa      Kenya  2023-10-14     343999.0        0.0
159168      KEN    Africa      Kenya  2023-10-15     343999.0        0.0
159169      KEN    Africa      Kenya  2023-10-16     343999.0        0.0
159170      KEN    Africa      Kenya  2023-10-17     343999.0        0.0
159171      KEN    Africa      Kenya  2023-10-18     343999.0        0.0

        new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  \
116295                 NaN           NaN         0.0                  NaN
116296                 NaN           NaN         0.0                  NaN
116297                 NaN           NaN         0.0                  NaN
116298                 NaN           NaN         0.0                  NaN
116299                 NaN           NaN         0.0                  NaN
…                    …           …         …                  …
159167               6.286        5689.0         0.0                  0.0
159168               6.286        5689.0         0.0                  0.0
159169               6.286        5689.0         0.0                  0.0
159170               0.000        5689.0         0.0                  0.0
159171               0.000        5689.0         0.0                  0.0

        …  male_smokers  handwashing_facilities  hospital_beds_per_thousand  \
116295  …           NaN                     NaN                         NaN
116296  …           NaN                     NaN                         NaN
116297  …           NaN                     NaN                         NaN
116298  …           NaN                     NaN                         NaN
116299  …           NaN                     NaN                         NaN
…     …           …                     …                         …
159167  …          20.4                  24.651                         1.4
159168  …          20.4                  24.651                         1.4
159169  …          20.4                  24.651                         1.4
159170  …          20.4                  24.651                         1.4
159171  …          20.4                  24.651                         1.4

        life_expectancy  human_development_index  population  \
116295            79.93                      NaN     32677.0
116296            79.93                      NaN     32677.0
116297            79.93                      NaN     32677.0
116298            79.93                      NaN     32677.0
116299            79.93                      NaN     32677.0
…                 …                      …         …
159167            66.70                    0.601  54027484.0
159168            66.70                    0.601  54027484.0
159169            66.70                    0.601  54027484.0
```

```
159170            66.70                    0.601  54027484.0
159171            66.70                    0.601  54027484.0

        excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
116295                                   NaN                          NaN
116296                                   NaN                          NaN
116297                                   NaN                          NaN
116298                                   NaN                          NaN
116299                                   NaN                          NaN
…                                      …                          …
159167                                   NaN                          NaN
159168                                   NaN                          NaN
159169                                   NaN                          NaN
159170                                   NaN                          NaN
159171                                   NaN                          NaN

        excess_mortality  excess_mortality_cumulative_per_million
116295               NaN                                      NaN
116296               NaN                                      NaN
116297               NaN                                      NaN
116298               NaN                                      NaN
116299               NaN                                      NaN
…                  …                                      …
159167               NaN                                      NaN
159168               NaN                                      NaN
159169               NaN                                      NaN
159170               NaN                                      NaN
159171               NaN                                      NaN

[2770 rows x 67 columns]
```

[8]:

```
# Converting date to datetime
df['date'] = pd.to_datetime(df['date'], errors='coerce')
```

[9]:

```
# Drop rows with missing dates
print("\nAfter dropping the missing dates:")
df.dropna(subset=['date'])
```

```
After dropping the missing dates:
```

[9]:

```
       iso_code continent     location       date  total_cases  new_cases  \
0           AFG      Asia  Afghanistan 2020-01-03          NaN        0.0
1           AFG      Asia  Afghanistan 2020-01-04          NaN        0.0
2           AFG      Asia  Afghanistan 2020-01-05          NaN        0.0
3           AFG      Asia  Afghanistan 2020-01-06          NaN        0.0
4           AFG      Asia  Afghanistan 2020-01-07          NaN        0.0
```

```
…         …       …          …        …          …        …
350080      ZWE    Africa     Zimbabwe 2023-10-14     265808.0        0.0
350081      ZWE    Africa     Zimbabwe 2023-10-15     265808.0        0.0
350082      ZWE    Africa     Zimbabwe 2023-10-16     265808.0        0.0
350083      ZWE    Africa     Zimbabwe 2023-10-17     265808.0        0.0
350084      ZWE    Africa     Zimbabwe 2023-10-18     265808.0        0.0

        new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  \
0                      NaN           NaN         0.0                  NaN
1                      NaN           NaN         0.0                  NaN
2                      NaN           NaN         0.0                  NaN
3                      NaN           NaN         0.0                  NaN
4                      NaN           NaN         0.0                  NaN
…                    …           …         …                  …
350080               5.286        5718.0         0.0                  0.0
350081               5.286        5718.0         0.0                  0.0
350082               5.286        5718.0         0.0                  0.0
350083               0.000        5718.0         0.0                  0.0
350084               0.000        5718.0         0.0                  0.0

        …  male_smokers  handwashing_facilities  hospital_beds_per_thousand  \
0       …           NaN                  37.746                         0.5
1       …           NaN                  37.746                         0.5
2       …           NaN                  37.746                         0.5
3       …           NaN                  37.746                         0.5
4       …           NaN                  37.746                         0.5
…     …           …                     …                          …
350080  …          30.7                  36.791                         1.7
350081  …          30.7                  36.791                         1.7
350082  …          30.7                  36.791                         1.7
350083  …          30.7                  36.791                         1.7
350084  …          30.7                  36.791                         1.7

        life_expectancy  human_development_index  population  \
0                 64.83                    0.511  41128772.0
1                 64.83                    0.511  41128772.0
2                 64.83                    0.511  41128772.0
3                 64.83                    0.511  41128772.0
4                 64.83                    0.511  41128772.0
…                 …                      …          …
350080            61.49                    0.571  16320539.0
350081            61.49                    0.571  16320539.0
350082            61.49                    0.571  16320539.0
350083            61.49                    0.571  16320539.0
350084            61.49                    0.571  16320539.0

        excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
0                                        NaN                          NaN
1                                        NaN                          NaN
2                                        NaN                          NaN
3                                        NaN                          NaN
4                                        NaN                          NaN
…                                      …                          …
350080                                   NaN                          NaN
350081                                   NaN                          NaN
350082                                   NaN                          NaN
350083                                   NaN                          NaN
350084                                   NaN                          NaN

        excess_mortality  excess_mortality_cumulative_per_million
0                    NaN                                      NaN
1                    NaN                                      NaN
2                    NaN                                      NaN
3                    NaN                                      NaN
4                    NaN                                      NaN
…                  …                                      …
350080               NaN                                      NaN
350081               NaN                                      NaN
350082               NaN                                      NaN
350083               NaN                                      NaN
350084               NaN                                      NaN

[350085 rows x 67 columns]
```

[10]:

```
# Handle missing numeric values with fillna() or interpolate().
df.fillna(0)
```

[10]:

```
       iso_code continent     location        date  total_cases  new_cases  \
0           AFG      Asia  Afghanistan 2020-01-03          0.0        0.0
1           AFG      Asia  Afghanistan 2020-01-04          0.0        0.0
2           AFG      Asia  Afghanistan 2020-01-05          0.0        0.0
3           AFG      Asia  Afghanistan 2020-01-06          0.0        0.0
4           AFG      Asia  Afghanistan 2020-01-07          0.0        0.0
…         …       …          …        …          …       …
350080      ZWE    Africa     Zimbabwe 2023-10-14     265808.0        0.0
350081      ZWE    Africa     Zimbabwe 2023-10-15     265808.0        0.0
350082      ZWE    Africa     Zimbabwe 2023-10-16     265808.0        0.0
350083      ZWE    Africa     Zimbabwe 2023-10-17     265808.0        0.0
350084      ZWE    Africa     Zimbabwe 2023-10-18     265808.0        0.0

        new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  \
0                    0.000           0.0         0.0                  0.0
1                    0.000           0.0         0.0                  0.0
2                    0.000           0.0         0.0                  0.0
```

```
3                    0.000           0.0         0.0                  0.0
4                    0.000           0.0         0.0                  0.0
…                     …           …         …                  …
350080               5.286        5718.0         0.0                  0.0
350081               5.286        5718.0         0.0                  0.0
350082               5.286        5718.0         0.0                  0.0
350083               0.000        5718.0         0.0                  0.0
350084               0.000        5718.0         0.0                  0.0

        …  male_smokers  handwashing_facilities  hospital_beds_per_thousand  \
0       …           0.0                  37.746                         0.5
1       …           0.0                  37.746                         0.5
2       …           0.0                  37.746                         0.5
3       …           0.0                  37.746                         0.5
4       …           0.0                  37.746                         0.5
…     …            …                     …                         …
350080  …          30.7                  36.791                         1.7
350081  …          30.7                  36.791                         1.7
350082  …          30.7                  36.791                         1.7
350083  …          30.7                  36.791                         1.7
350084  …          30.7                  36.791                         1.7

        life_expectancy  human_development_index  population  \
0                 64.83                    0.511  41128772.0
1                 64.83                    0.511  41128772.0
2                 64.83                    0.511  41128772.0
3                 64.83                    0.511  41128772.0
4                 64.83                    0.511  41128772.0
…                  …                      …         …
350080            61.49                    0.571  16320539.0
350081            61.49                    0.571  16320539.0
350082            61.49                    0.571  16320539.0
350083            61.49                    0.571  16320539.0
350084            61.49                    0.571  16320539.0

        excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
0                                        0.0                          0.0
1                                        0.0                          0.0
2                                        0.0                          0.0
3                                        0.0                          0.0
4                                        0.0                          0.0
…                                       …                          …
350080                                   0.0                          0.0
350081                                   0.0                          0.0
350082                                   0.0                          0.0
350083                                   0.0                          0.0
350084                                   0.0                          0.0
```

```
        excess_mortality  excess_mortality_cumulative_per_million
0                    0.0                                      0.0
1                    0.0                                      0.0
2                    0.0                                      0.0
3                    0.0                                      0.0
4                    0.0                                      0.0
…                   …                                      …
350080               0.0                                      0.0
350081               0.0                                      0.0
350082               0.0                                      0.0
350083               0.0                                      0.0
350084               0.0                                      0.0

[350085 rows x 67 columns]
```

[11]:

```python
# Handle missing numeric values with fillna() or interpolate().
df.interpolate()
```

[11]:

```
       iso_code continent     location       date  total_cases  new_cases  \
0           AFG      Asia  Afghanistan 2020-01-03          NaN        0.0
1           AFG      Asia  Afghanistan 2020-01-04          NaN        0.0
2           AFG      Asia  Afghanistan 2020-01-05          NaN        0.0
3           AFG      Asia  Afghanistan 2020-01-06          NaN        0.0
4           AFG      Asia  Afghanistan 2020-01-07          NaN        0.0
…         …       …          …        …          …        …
350080      ZWE    Africa     Zimbabwe 2023-10-14     265808.0        0.0
350081      ZWE    Africa     Zimbabwe 2023-10-15     265808.0        0.0
350082      ZWE    Africa     Zimbabwe 2023-10-16     265808.0        0.0
350083      ZWE    Africa     Zimbabwe 2023-10-17     265808.0        0.0
350084      ZWE    Africa     Zimbabwe 2023-10-18     265808.0        0.0

        new_cases_smoothed  total_deaths  new_deaths  new_deaths_smoothed  \
0                      NaN           NaN         0.0                  NaN
1                      NaN           NaN         0.0                  NaN
2                      NaN           NaN         0.0                  NaN
3                      NaN           NaN         0.0                  NaN
4                      NaN           NaN         0.0                  NaN
…                    …           …         …                  …
350080               5.286        5718.0         0.0                  0.0
350081               5.286        5718.0         0.0                  0.0
350082               5.286        5718.0         0.0                  0.0
350083               0.000        5718.0         0.0                  0.0
350084               0.000        5718.0         0.0                  0.0

        …  male_smokers  handwashing_facilities  hospital_beds_per_thousand  \
0       …           NaN                  37.746                         0.5
```

```
1       …           NaN                  37.746                         0.5
2       …           NaN                  37.746                         0.5
3       …           NaN                  37.746                         0.5
4       …           NaN                  37.746                         0.5
…     …           …                     …                         …
350080  …          30.7                  36.791                         1.7
350081  …          30.7                  36.791                         1.7
350082  …          30.7                  36.791                         1.7
350083  …          30.7                  36.791                         1.7
350084  …          30.7                  36.791                         1.7

        life_expectancy  human_development_index  population  \
0                 64.83                    0.511  41128772.0
1                 64.83                    0.511  41128772.0
2                 64.83                    0.511  41128772.0
3                 64.83                    0.511  41128772.0
4                 64.83                    0.511  41128772.0
…                 …                       …          …
350080            61.49                    0.571  16320539.0
350081            61.49                    0.571  16320539.0
350082            61.49                    0.571  16320539.0
350083            61.49                    0.571  16320539.0
350084            61.49                    0.571  16320539.0

        excess_mortality_cumulative_absolute  excess_mortality_cumulative  \
0                                        NaN                          NaN
1                                        NaN                          NaN
2                                        NaN                          NaN
3                                        NaN                          NaN
4                                        NaN                          NaN
…                                       …                          …
350080                             52794.797                         9.62
350081                             52794.797                         9.62
350082                             52794.797                         9.62
350083                             52794.797                         9.62
350084                             52794.797                         9.62

        excess_mortality  excess_mortality_cumulative_per_million
0                    NaN                                      NaN
1                    NaN                                      NaN
2                    NaN                                      NaN
3                    NaN                                      NaN
4                    NaN                                      NaN
…                   …                                      …
350080             18.41                                1501.3901
350081             18.41                                1501.3901
350082             18.41                                1501.3901
```

```
350083             18.41                                   1501.3901
350084             18.41                                   1501.3901

[350085 rows x 67 columns]
```

[ ]:

```

```

[ ]:

```python
# 3. Exploratory Data Analysis (EDA)
```

[36]:

```python
# Set the plot style
sns.set_style('darkgrid')
plt.figure(figsize=(15, 10))
```

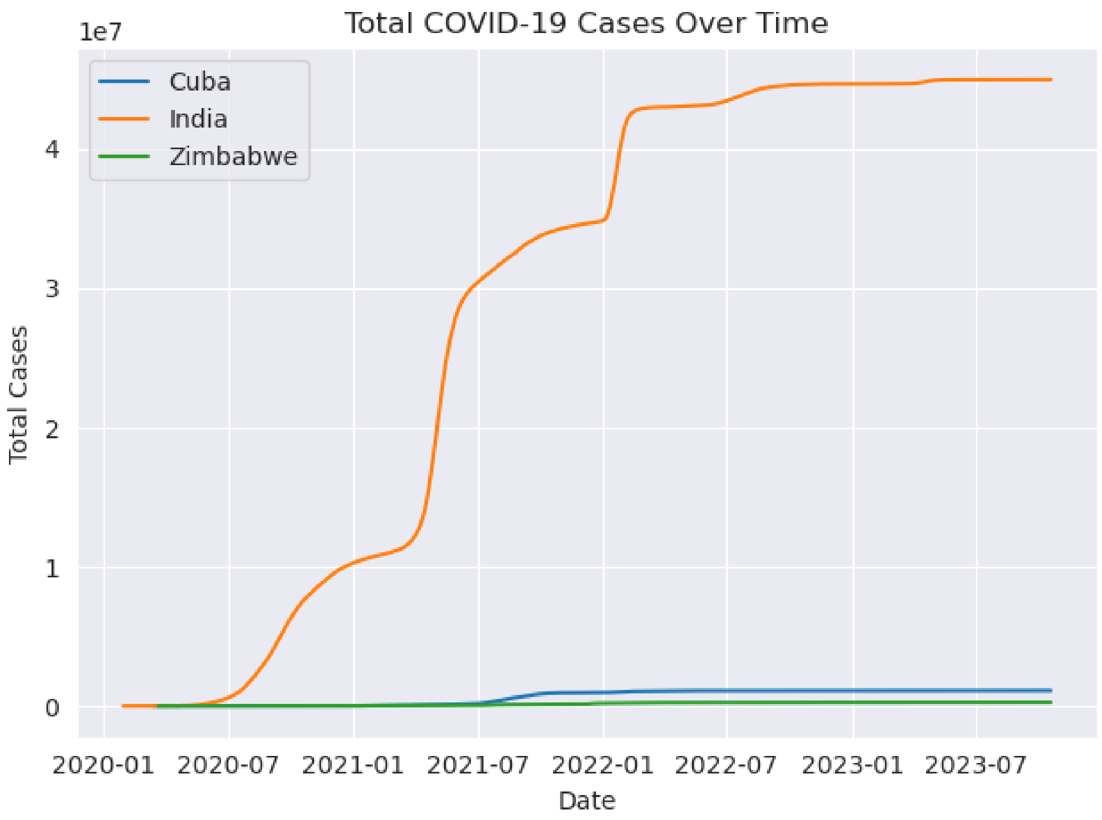
[36]:

```
<Figure size 1500x1000 with 0 Axes>
```

```
<Figure size 1500x1000 with 0 Axes>
```

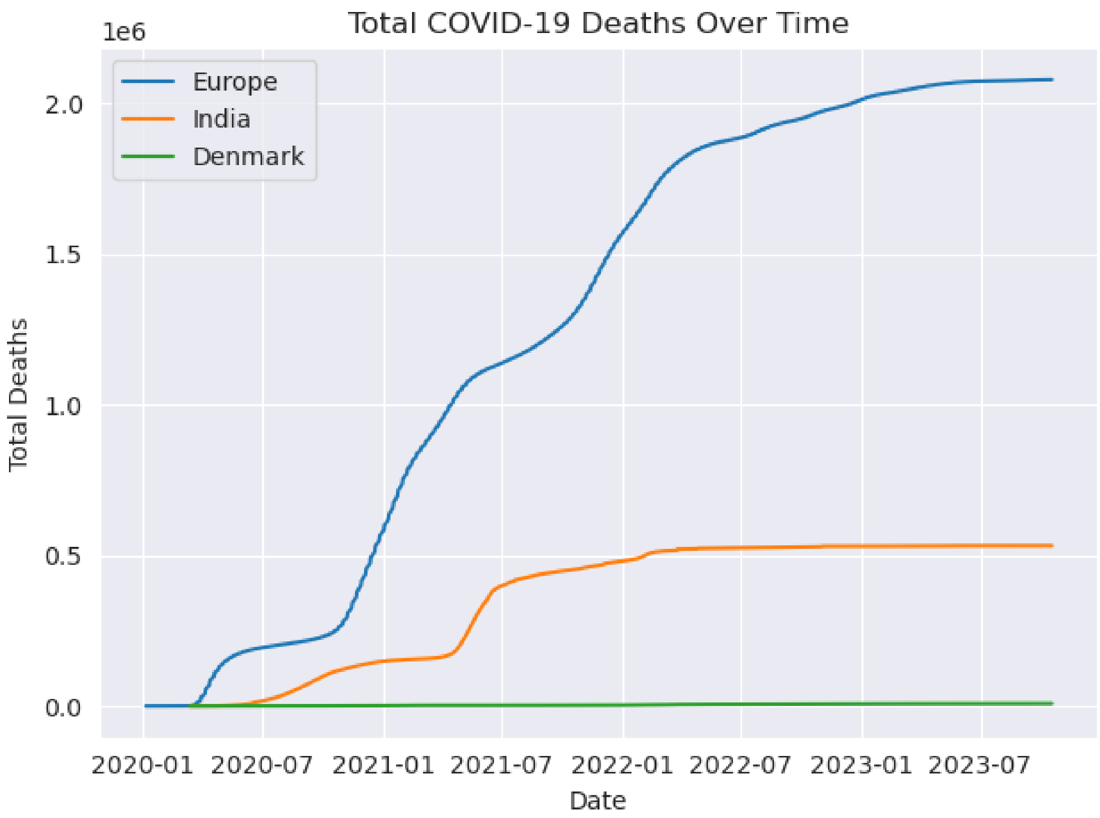
[54]:

```python
# a. Line chart: Plot total cases over time for selected countries.
# Filter data for selected countries
df_filtered = df[df['location'].isin(['Cuba', 'India', 'Zimbabwe'])].copy()

# Convert date to datetime
df_filtered['date'] = pd.to_datetime(df_filtered['date'])

# Drop rows with missing total_cases
df_filtered = df_filtered.dropna(subset=['total_cases'])

# Plot
for country in ['Cuba', 'India', 'Zimbabwe']:
    country_data = df_filtered[df_filtered['location'] == country]
    plt.plot(country_data['date'], country_data['total_cases'], label=country)

plt.title('Total COVID-19 Cases Over Time')
plt.xlabel('Date')
plt.ylabel('Total Cases')
plt.legend()
plt.tight_layout()
plt.show()
```



[55]:

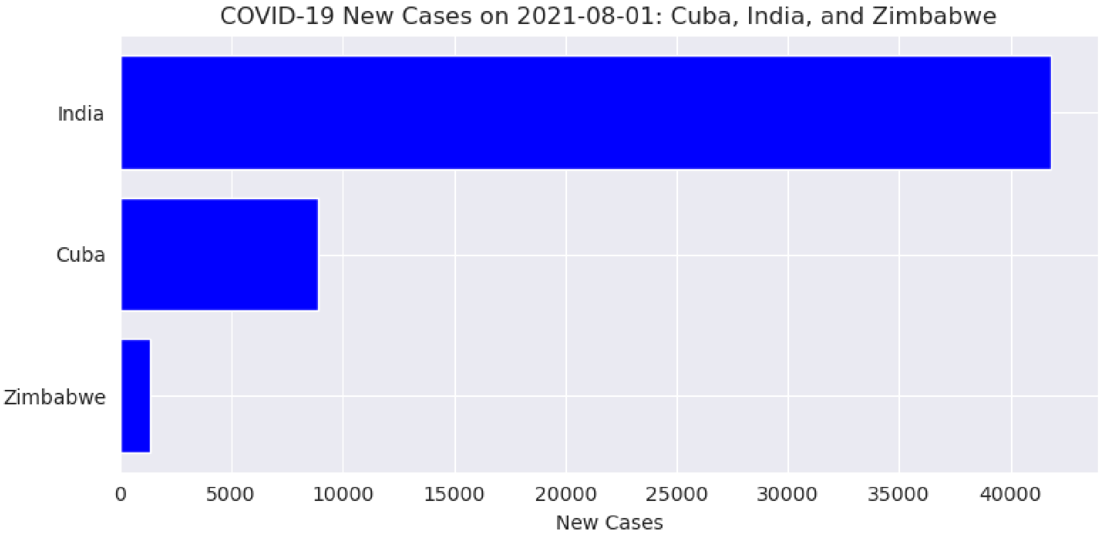
```
# b. Line chart: Plot total deaths over time

# Filter for relevant countries
df_filtered = df[df['location'].isin(['Europe', 'India', 'Denmark'])].copy()

# Ensure date is a datetime
df_filtered['date'] = pd.to_datetime(df_filtered['date'])

# Plot total deaths over time
for country in ['Europe', 'India', 'Denmark']:
    country_data = df_filtered[df_filtered['location'] == country]
    plt.plot(country_data['date'], country_data['total_deaths'], label=country)

plt.title('Total COVID-19 Deaths Over Time')
plt.xlabel('Date')
plt.ylabel('Total Deaths')
plt.legend()
plt.tight_layout()
plt.show()
```



[77]:

```
#c. Bar chart: Compare new cases between countries on a selected date.
# Convert date column to datetime
df['date'] = pd.to_datetime(df['date'])

# Choose a specific date
selected_date = '2021-08-01'

# Filter for that date and non-null new_cases
df_filtered = df[(df['date'] == selected_date) &
                 (df['location'].isin(['Cuba', 'India', 'Zimbabwe'])) &
                 (~df['new_cases'].isna())].copy()

# Sort for better visuals
df_filtered = df_filtered.sort_values('new_cases')

# Show what you're plotting
print(df_filtered[['location', 'date', 'new_cases']])

# Plot
plt.figure(figsize=(8, 4))
plt.barh(df_filtered['location'], df_filtered['new_cases'], color='blue')
plt.title(f'COVID-19 New Cases on {selected_date}: Cuba, India, and Zimbabwe')
plt.xlabel('New Cases')
plt.tight_layout()
plt.show()
```

```
        location        date  new_cases
349276  Zimbabwe  2021-08-01     1370.0
72618       Cuba  2021-08-01     8875.0
140349     India  2021-08-01    41831.0
```



[ ]:

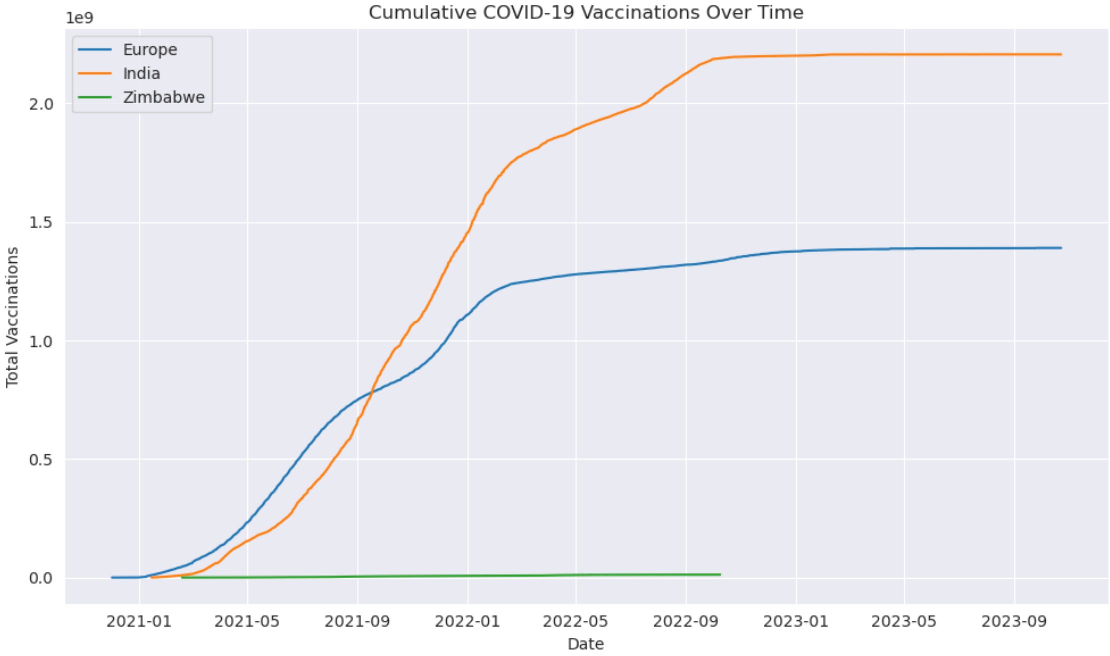
```

```

[ ]:

```
# 4. Visualizing Vaccination Progress
```

[79]:

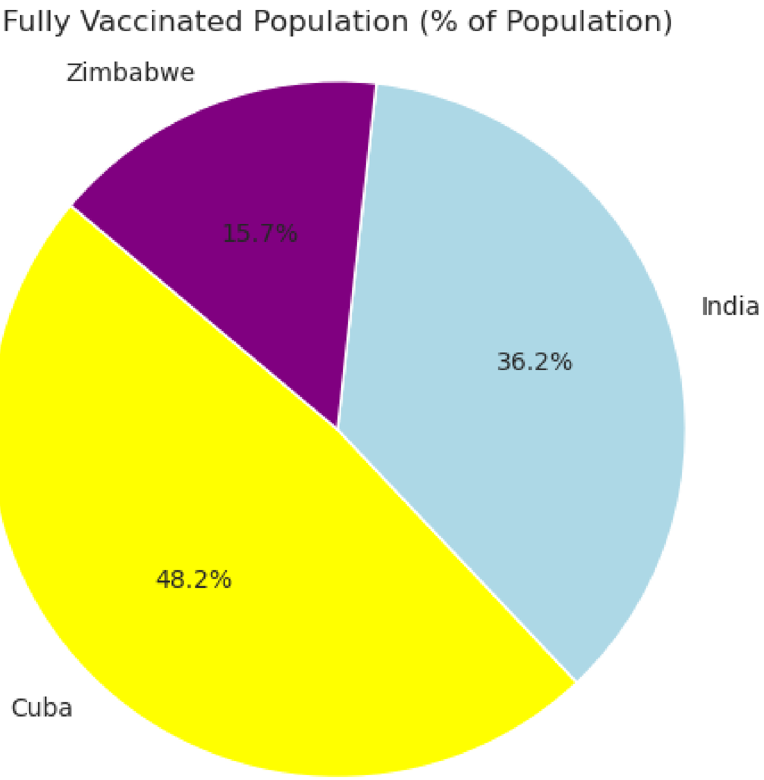
```
# a. Line chart: Plot cumulative vaccinations over time
# Ensure date column is in datetime format
df['date'] = pd.to_datetime(df['date'])

# Filter for selected countries and non-null vaccination data
selected_countries = ['Europe', 'India', 'Zimbabwe']
df_vax = df[df['location'].isin(selected_countries) & df['total_vaccinations'].
  ↪notna()].copy()

# Plot line chart
plt.figure(figsize=(10, 6))

for country in selected_countries:
    country_data = df_vax[df_vax['location'] == country]
    plt.plot(country_data['date'], country_data['total_vaccinations'],␣
  ↪label=country)

plt.title('Cumulative COVID-19 Vaccinations Over Time')
plt.xlabel('Date')
plt.ylabel('Total Vaccinations')
plt.legend()
plt.tight_layout()
plt.show()
```



[84]:

```
# c. Pie chart: Compare the vaccinated population in percentage

# Ensure the datetime format
df['date'] = pd.to_datetime(df['date'])

# Drop rows with missing vaccination data
df_vax = df[['location', 'date', 'people_fully_vaccinated_per_hundred']].
  ↪dropna()

# Get the most recent vaccination data per country
df_vax_latest = df_vax.sort_values('date').groupby('location', as_index=False).
  ↪last()

# Filter for specific countries and make a copy
```

```
df_vax_filtered = df_vax_latest[df_vax_latest['location'].isin(['Cuba',␣
 ↪'India', 'Zimbabwe'])].copy()

# Plot pie chart
plt.pie(df_vax_filtered['people_fully_vaccinated_per_hundred'],
        labels=df_vax_filtered['location'],
        autopct='%1.1f%%',
        startangle=140,
        colors=['yellow', 'lightblue', 'purple'])
plt.title('Fully Vaccinated Population (% of Population)')
plt.axis('equal')  # Equal aspect ratio ensures pie is a circle.
plt.tight_layout()
plt.show()
```



[ ]:

```
```

[ ]:

```
# 5. Build a Choropleth Map
```

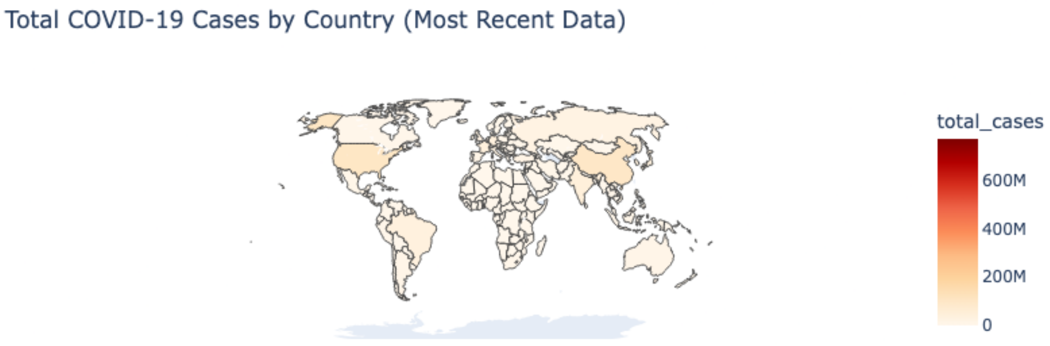
[85]:

```
import plotly.express as px

# Ensure 'date' is in datetime format
```

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

# Filter relevant columns and drop missing iso_code or total_cases
df_cases = df[['location', 'iso_code', 'date', 'total_cases']].
 ↪dropna(subset=['iso_code', 'total_cases'])

# Get the most recent total_cases per country
latest_cases = df_cases.sort_values('date').groupby('iso_code', as_index=False).
 ↪last()
```

[86]:

```
fig = px.choropleth(
    latest_cases,
    locations='iso_code',
    color='total_cases',
    hover_name='location',
    color_continuous_scale='OrRd',
    title='Total COVID-19 Cases by Country (Most Recent Data)',
    projection='natural earth'
)

fig.update_layout(geo=dict(showframe=False, showcoastlines=False))
fig.show()
```


Total COVID-19 Cases by Country (Most Recent Data)

[ ]:

```
# 6. Insights & Reporting
```

[ ]:

```
# COVID-19 Data Analysis: Cuba, India, Zimbabwe

# 1: Imports
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import plotly.express as px
import ipywidgets as widgets
from IPython.display import display
```

[ ]:

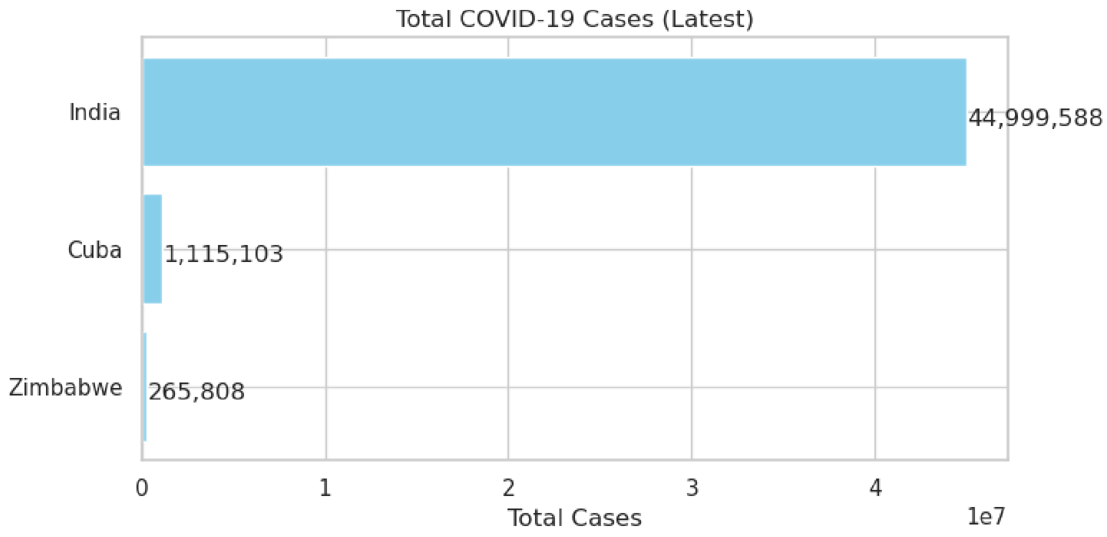
```
# 2: Load & preprocess data
df['date'] = pd.to_datetime(df['date'])
df = df[df['location'].isin(['Cuba', 'India', 'Zimbabwe'])]
```

[ ]:

```
# 3: Key Insights
1. India has the highest COVID-19 cases among the three countries analyzed.
2. Cuba demonstrates a rapid and high vaccination rate, likely due to its local␣
 ↪vaccine production efforts.
3. Zimbabwe’s lower total case count may reflect limited testing/reporting␣
 ↪capacity rather than actual low transmission.
4. Vaccination rates in Zimbabwe lag behind the other two countries.
5. Case surges in Cuba show noticeable spikes suggesting localized outbreaks or␣
 ↪policy shifts.
```

[94]:

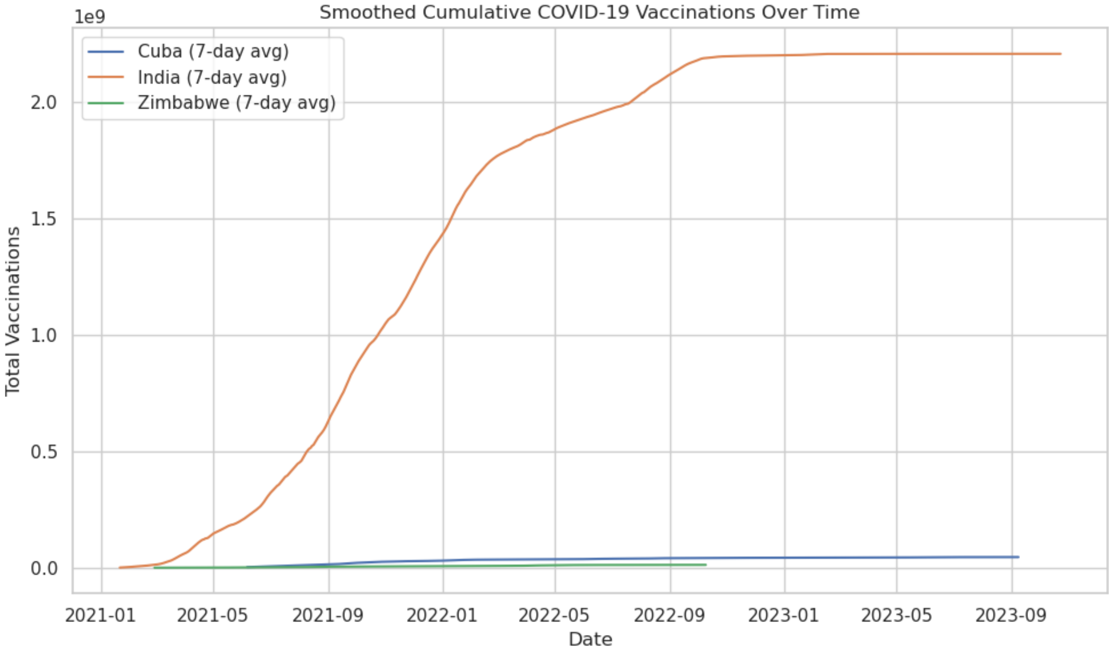
```
# 4: Bar Chart insights
#India's total cases dwarf those of Cuba and Zimbabwe. While Cuba has a␣
 ↪moderate total, Zimbabwe's numbers are significantly lower, which may␣
 ↪reflect differences in testing/reporting.
sns.set(style='whitegrid')
df_cases = df[['location', 'iso_code', 'date', 'total_cases']].dropna()
latest = df_cases.sort_values('date').groupby('iso_code', as_index=False).last()
selected = latest.sort_values('total_cases')

plt.figure(figsize=(8, 4))
bars = plt.barh(selected['location'], selected['total_cases'], color='skyblue')
plt.title('Total COVID-19 Cases (Latest)')
plt.xlabel('Total Cases')
for bar in bars:
    plt.text(bar.get_width() + 5000, bar.get_y() + 0.3, f"{int(bar.get_width()):
 ↪,}")
plt.tight_layout()
plt.show()
```

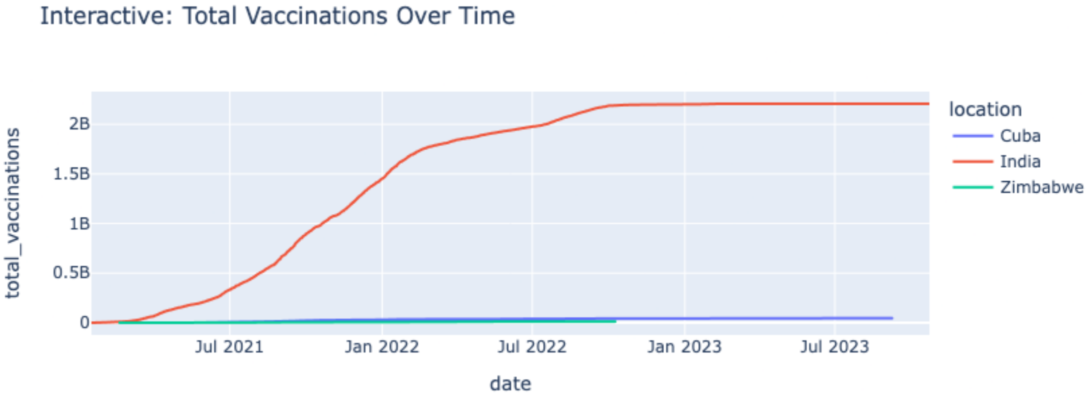


[98]:

```
# 5: Cumulative Vaccination Line Chart (Smoothed)

# Cuba shows a rapid early vaccination rollout, outpacing India and Zimbabwe.
# India's steady rise reflects scale, while Zimbabwe’s flat curve suggests␣
 ↪challenges.

df_vax = df[df['total_vaccinations'].notna()].copy()
plt.figure(figsize=(10, 6))

for country in ['Cuba', 'India', 'Zimbabwe']:
    country_data = df_vax[df_vax['location'] == country].copy()  # Ensure a␣
 ↪deep copy
    country_data.loc[:, 'smoothed'] = country_data['total_vaccinations'].
 ↪rolling(window=7).mean()
    plt.plot(country_data['date'], country_data['smoothed'], label=f"{country}␣
 ↪(7-day avg)")

plt.title('Smoothed Cumulative COVID-19 Vaccinations Over Time')
plt.xlabel('Date')
plt.ylabel('Total Vaccinations')
plt.legend()
plt.tight_layout()
plt.show()


# Interactive Plotly Line Chart
fig = px.line(df_vax, x='date', y='total_vaccinations', color='location',
              title='Interactive: Total Vaccinations Over Time')
```

```
fig.show()
```





[ ]:

```

```

[99]:

```
# 6: ipywidgets Dropdown for Daily New Cases

def plot_country(country):
    country_df = df[df['location'] == country]
    plt.figure(figsize=(8, 4))
    plt.plot(country_df['date'], country_df['new_cases'], label='New Cases')
    plt.title(f'COVID-19 New Cases in {country}')
    plt.xlabel('Date')
    plt.ylabel('New Cases')
    plt.grid(True)
    plt.tight_layout()
    plt.show()

widgets.interact(plot_country, country=widgets.Dropdown(options=['Cuba',␣
 ↪'India', 'Zimbabwe'], description='Country:'))
```

```
interactive(children=(Dropdown(description='Country:', options=('Cuba', 'India',␣
 ↪'Zimbabwe'), value='Cuba'), O…
```

[99]:
```
<function __main__.plot_country(country)>
```

[ ]:
```
Notes & Observations

1. Zimbabwe's case and vaccination data show flat lines in some periods,␣
 ↪indicating possible reporting gaps.
2. Cuba's sharp vaccination curve supports claims of domestic vaccine␣
 ↪innovation.
3. Be cautious when comparing totals directly without considering population␣
 ↪size and data completeness.
```

[ ]:
```
```