- **About Dataset**

- CO2 emissions (metric tons per capita)

- Global CO2 Emissions from 1960 to 2018.

- **Tasks**

- Country with highest/lowest emissions
- Data Visualisation (world map)

import library

```
In [2]:   import matplotlib.pyplot as plt
          import seaborn as sns
          import plotly.express as px
          import numpy as np # linear algebra
          import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
          import os
          import warnings
          warnings.filterwarnings("ignore")
```

Matplotlib is building the font cache; this may take a moment.

Read the dataset

```
In [4]:   df = pd.read_csv('CO2_Emissions_1960-2018.csv', index_col='Country Name')
```

shape of the dataset

```
In [8]:   df.shape
```

```
Out[8]:   (266, 59)
```

```
In [10]:  df1=df.transpose()
          df1.index=pd.to_datetime(df1.index).year
          df1.head()
```

Out[10]:

Country Name	Aruba	Africa Eastern and Southern	Afghanistan	Africa Western and Central	Angola	Albania	Andorra
1960	204.631696	0.906060	0.046057	0.090880	0.100835	1.258195	NaN	0.6
1961	208.837879	0.922474	0.053589	0.095283	0.082204	1.374186	NaN	0.6
1962	226.081890	0.930816	0.073721	0.096612	0.210533	1.439956	NaN	0.7
1963	214.785217	0.940570	0.074161	0.112376	0.202739	1.181681	NaN	0.8
1964	207.626699	0.996033	0.086174	0.133258	0.213562	1.111742	NaN	0.9

5 rows × 266 columns

In [12]:
```python
px.line(data_frame=df1,x=df1.index, y=['World'],markers=True, title='World carbo
```

World carbon emission 1960 - 2018
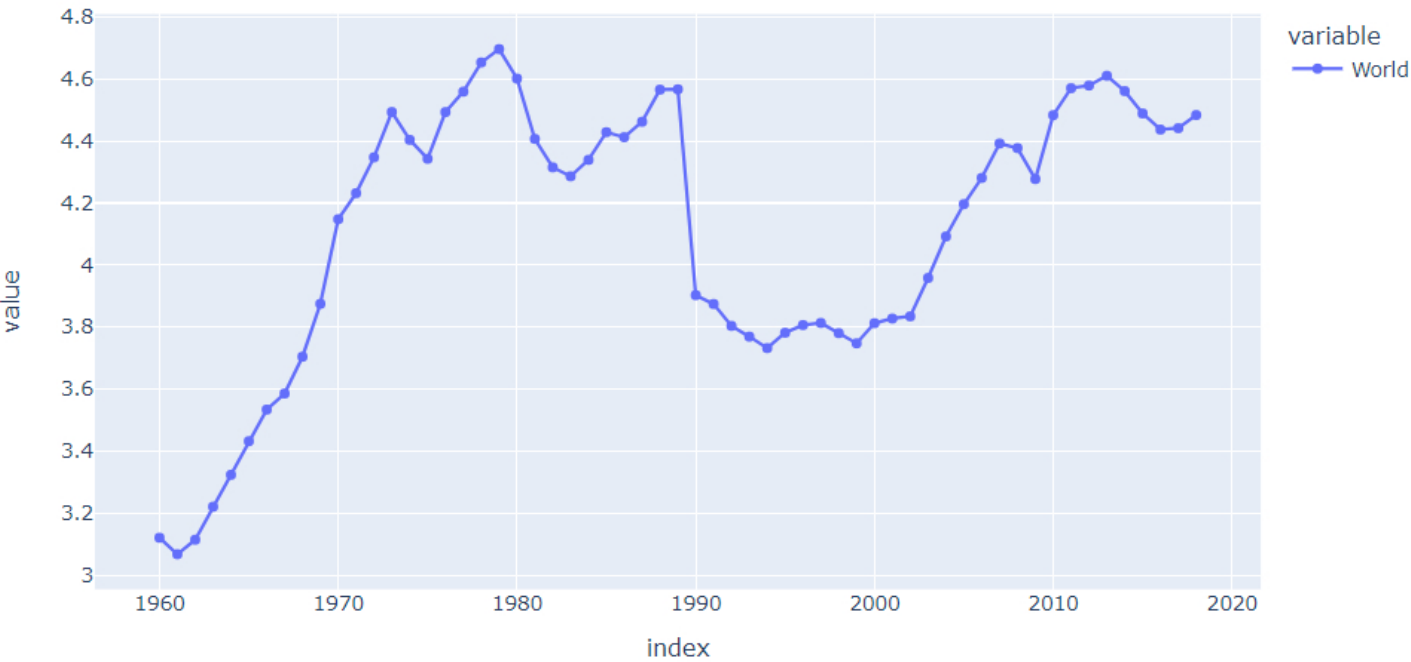
During the 1990s emissions were mainly reduced due to structural changes in the European economies from heavy manufacturing towards more service-based economies and a change from coal to a higher share of gas in the energy mix4.

In [14]:
```python
df.head()
```

Out[14]:

		1960	1961	1962	1963	1964	1965
	Country Name						
	Aruba	204.631696	208.837879	226.081890	214.785217	207.626699	185.213644
	Africa Eastern and Southern	0.906060	0.922474	0.930816	0.940570	0.996033	1.047280
	Afghanistan	0.046057	0.053589	0.073721	0.074161	0.086174	0.101285
	Africa Western and Central	0.090880	0.095283	0.096612	0.112376	0.133258	0.184803
	Angola	0.100835	0.082204	0.210533	0.202739	0.213562	0.205891

5 rows × 59 columns
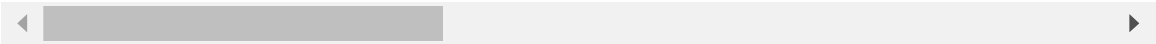
In [15]:
```python
df.reset_index(level=0, inplace=True)
```

In [16]:
```python
df2=df.melt(id_vars=['Country Name'], var_name='Year').sort_values(by=['Year'])
```
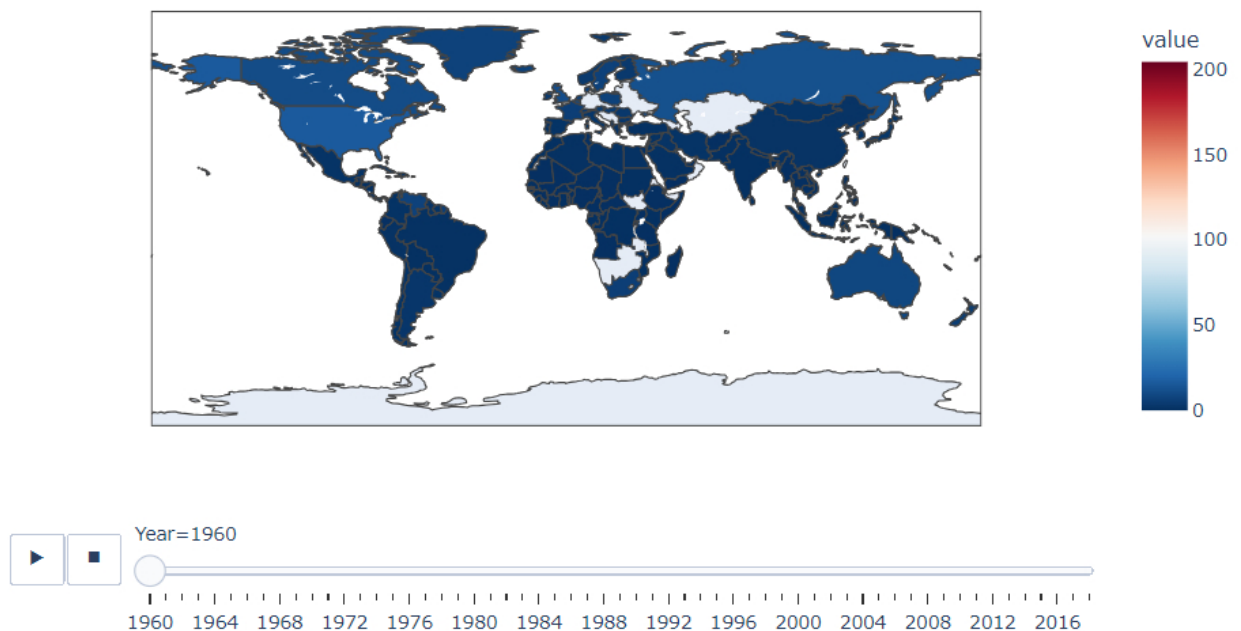
In [17]:
```python
df2.head()
```

Out[17]:

	Country Name	Year	value
0	Aruba	1960	204.631696
169	Malaysia	1960	0.437450
170	North America	1960	15.529995
171	Namibia	1960	NaN
172	New Caledonia	1960	10.908165

In [18]:
```python
px.choropleth(data_frame=df2[df2['Country Name']!='World'], locationmode='countr
```

CO2 Emission by Countries by Year


```
In [45]:   !pip install raceplotly
```

```
Defaulting to user installation because normal site-packages is not writeable
Looking in links: /usr/share/pip-wheels
Requirement already satisfied: raceplotly in ./.local/lib/python3.9/site-packages
(0.1.7)
Requirement already satisfied: plotly in /opt/conda/envs/anaconda-2022.05-py39/li
b/python3.9/site-packages (from raceplotly) (5.6.0)
Requirement already satisfied: pandas in /opt/conda/envs/anaconda-2022.05-py39/li
b/python3.9/site-packages (from raceplotly) (1.4.2)
Requirement already satisfied: numpy in /opt/conda/envs/anaconda-2022.05-py39/li
b/python3.9/site-packages (from raceplotly) (1.21.5)
Requirement already satisfied: python-dateutil>=2.8.1 in /opt/conda/envs/anaconda
-2022.05-py39/lib/python3.9/site-packages (from pandas->raceplotly) (2.8.2)
Requirement already satisfied: pytz>=2020.1 in /opt/conda/envs/anaconda-2022.05-p
y39/lib/python3.9/site-packages (from pandas->raceplotly) (2021.3)
Requirement already satisfied: six>=1.5 in /opt/conda/envs/anaconda-2022.05-py39/
lib/python3.9/site-packages (from python-dateutil>=2.8.1->pandas->raceplotly) (1.
16.0)
Requirement already satisfied: tenacity>=6.2.0 in /opt/conda/envs/anaconda-2022.0
5-py39/lib/python3.9/site-packages (from plotly->raceplotly) (8.0.1)
```
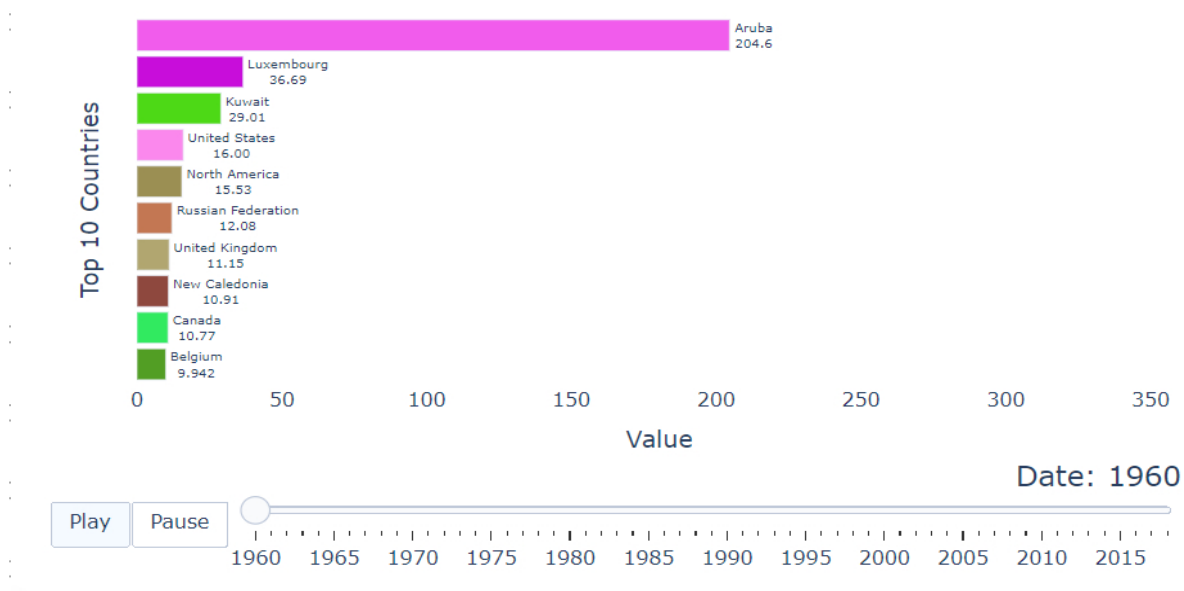
Carbon Emission Race by Country (Since 1960 to 2018)

```
In [71]:   from raceplotly.plots import barplot
           race_plot=barplot(df=df2[df2['Country Name']!='World'],item_column='Country Name
           race_plot.plot(title='Carbon Emission Race by Country (Since 1960 to 2018)', val
```

Carbon Emission Race by Country (Since 1960 to 2018)


```
In [49]: country_col=df2[df2['Country Name']!='World'].groupby(by=['Country Name']).max()
         country_col.extend(['India','China'])
```

```
In [51]: sample_mean=df2[df2['Country Name'].isin(country_col)]['value'].mean()
         pop_mean = df2['value'].mean()
```

```
In [53]: plt.figure(figsize=(15, 8))

         ax = sns.kdeplot(data=df2[df2['Country Name'].isin(country_col)],
                          x=df2['value'],
                          hue='Country Name',
                          fill=True,
                          linewidth=3,
                          edgecolor='white',
                          alpha=0.7)

         ax = plt.axvline(x=sample_mean)
         ax = plt.axvline(x=pop_mean)

         ax = plt.text(x=sample_mean, y=0.015, s='Sample mean')
         ax = plt.text(x=pop_mean, y=0.015, s='Population mean')

         plt.title('Kernel Density Plot with Sample and Population Means')
         plt.xlim(-20, 60)
         plt.ylim(0.00, 0.04)

         plt.show()
```
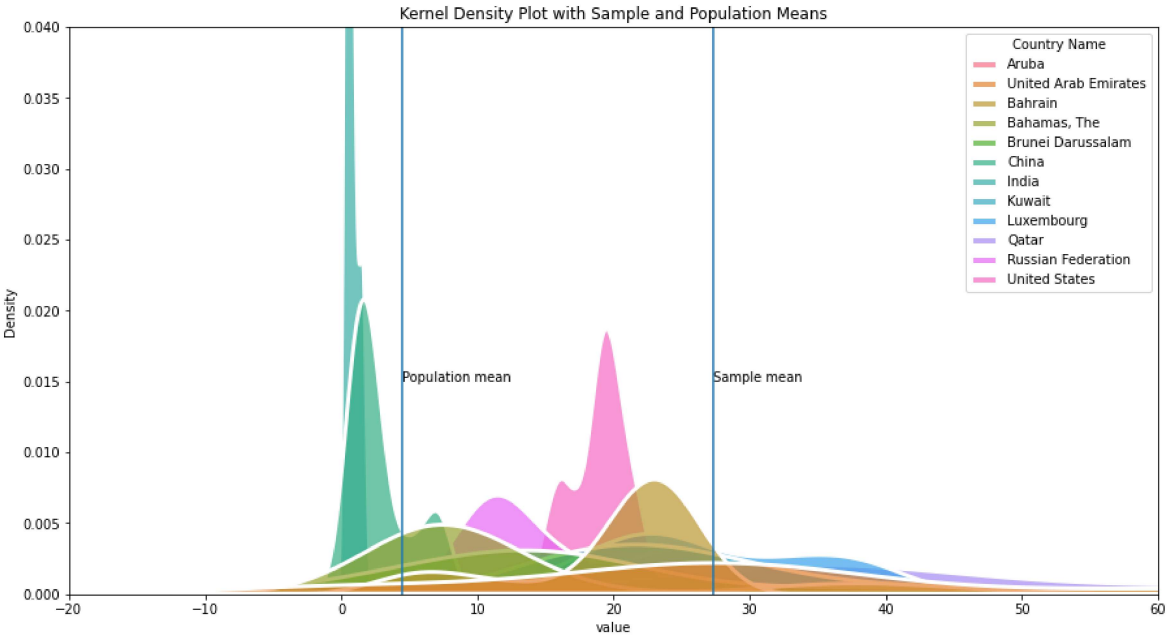
Kernel Density Plot with Sample and Population Means

In [54]: `df2`

Out[54]:

	Country Name	Year	value
0	Aruba	1960	204.631696
169	Malaysia	1960	0.437450
170	North America	1960	15.529995
171	Namibia	1960	NaN
172	New Caledonia	1960	10.908165
...	...	...	...
15522	Guyana	2018	3.132193
15523	High income	2018	10.254533
15524	Hong Kong SAR, China	2018	NaN
15510	Georgia	2018	2.538542
15693	Zimbabwe	2018	0.849793

15694 rows × 3 columns

Variation of Carbon emission by top 12 countries

In [58]:
```python
px.box( data_frame=df2[df2['Country Name'].isin(country_col)],
        x='value',
        y='Country Name',
        color='Country Name',
        title='Variation of Carbon emission by top 12 countries')
```

Variation of Carbon emission by top 12 countries
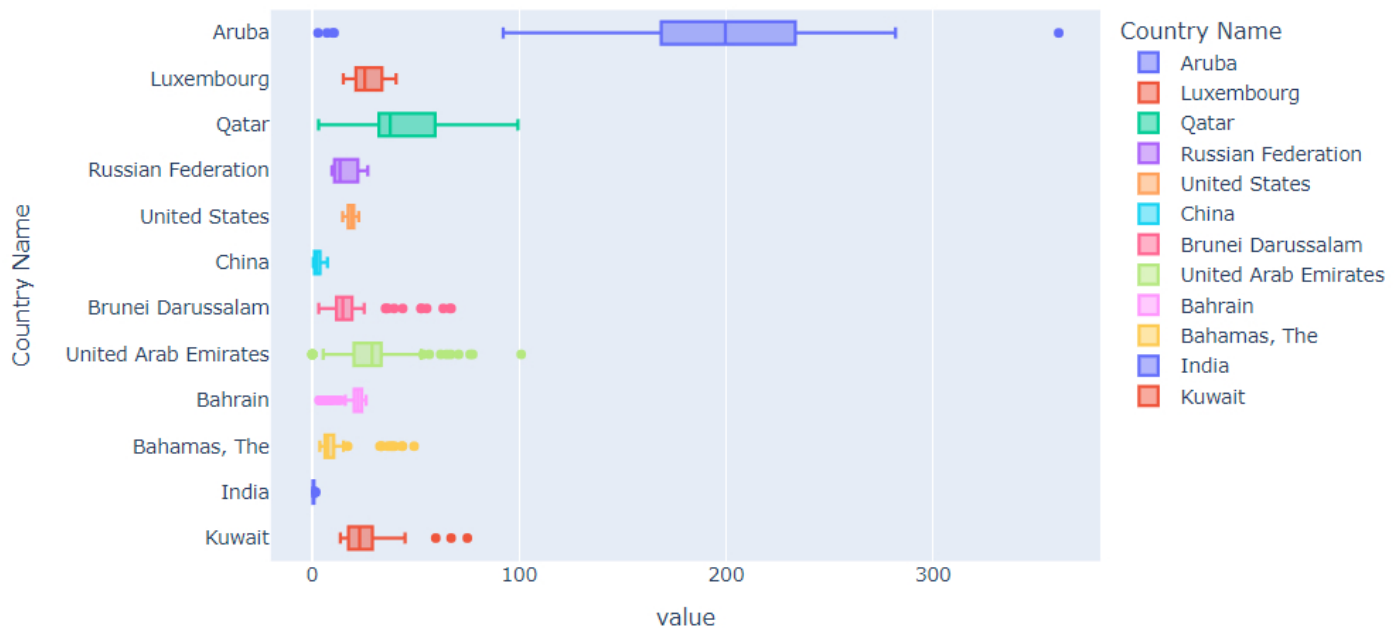

```
In [60]: df2['decade']=1
         df2['decade'][df2['Year'].between('1960','1970')]='1960-1970'
         df2['decade'][df2['Year'].between('1971','1980')]='1971-1980'
         df2['decade'][df2['Year'].between('1981','1990')]='1981-1990'
         df2['decade'][df2['Year'].between('1991','2000')]='1991-2000'
         df2['decade'][df2['Year'].between('2001','2010')]='2001-2010'
         df2['decade'][df2['Year'].between('2011','2020')]='2011-2018'
```

Variation of Carbon emission by top 12 countries
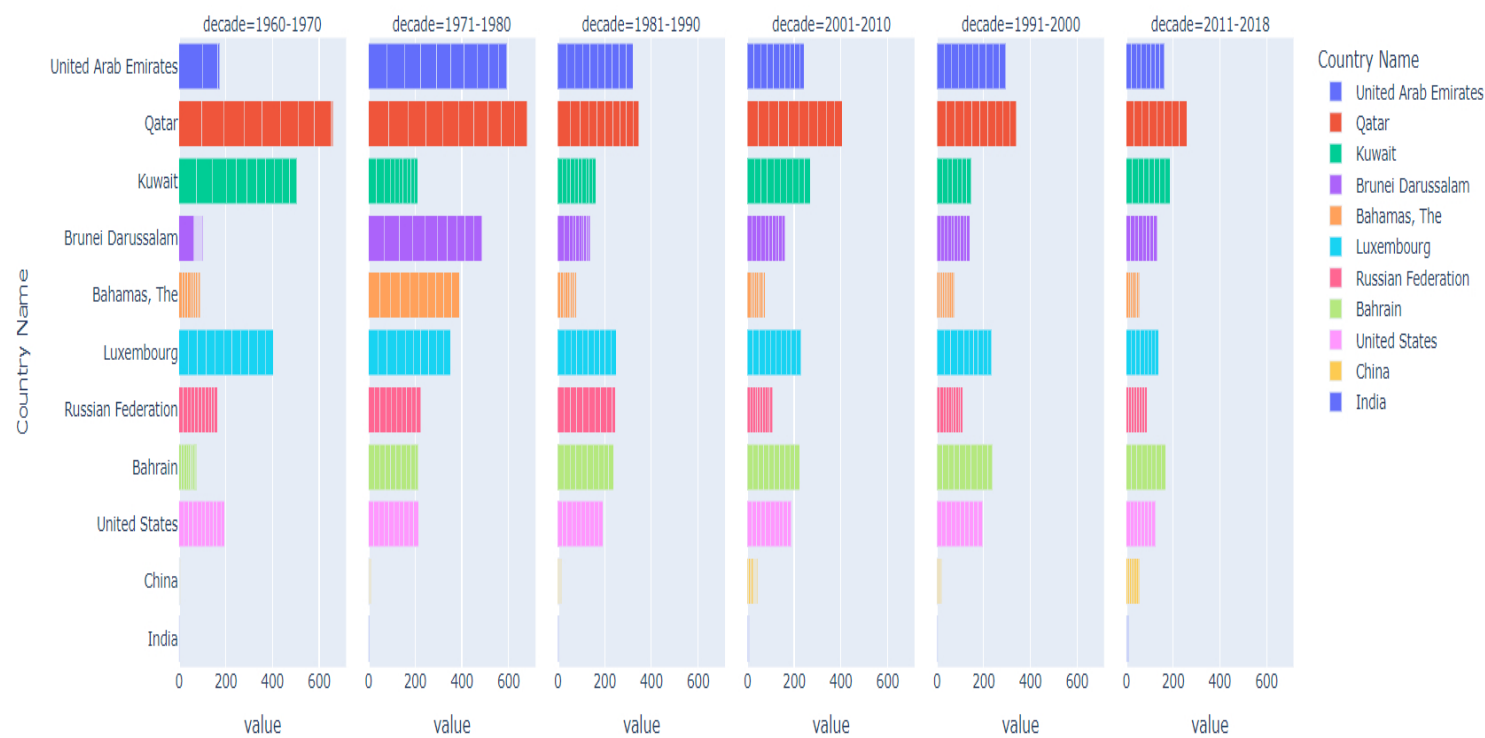
```
In [63]: px.box(data_frame=df2[df2['Country Name'].isin(country_col)],
               x='value',
               y='Country Name',
               color='Country Name',
               title='Variation of Carbon emission by top 12 countries', facet_col='deca
```

Variation of Carbon emission by top 12 countries

As we can see the country Aruba contributed high carbon emssion from 1960 - 1990, which shows high figure in the chart. after 1990, there is no much carbon emission from Aruba, in this case, let us remove Aruba to clearly view other countries data.

```
In [66]:  country_col.remove('Aruba')
```

```
In [68]:  px.bar(data_frame=df2[df2['Country Name']
                                .isin(country_col)]
                            .sort_values(by=['value'],ascending=False),
                            x='value',
                            y='Country Name',
                            facet_col='decade',
                            color='Country Name',
                            barmode='stack' )
```

Refrence

- https://www.kaggle.com/datasets/kkhandekar/co2-emissions-1960-2018
- https://plotly.com/python/plotly-express/
- https://seaborn.pydata.org/
- https://pandas.pydata.org/

In [ ]: