DATA WRANGLING LAB REPORT 3

Name:Tean Jin He

Matric ID : SD21063

Section: 02G

Question 1

```
#1. Interactive data visualization is the use of tools and processes to produce a visual representation of data which can be explored and analyzed directly within the visualization itself and can help to uncover insights which lead to better, data-driven decisions.
```

```
#2. Used for interactive visualization such as Qlik, Tableau, Power BI, D3.js and Plotly which can allow users to create and customize various types of charts, graphs, maps, dashboards, and other visual elements that can be manipulated and interacted with.
```

```
#3. The benefits of interactive data visualizations compare to traditional data visualization are:-
- can enable faster and easier understanding of complex data by allowing users to filter, zoom, drill-down, highlight, and compare different aspects of the data
- can empower users to do more analysis and exploration without requiring IT support or coding skills every time they have a new question or hypothesis
- can facilitate smarter and quicker decisions by providing real-time feedback and alerts based on the data changes and user actions
- can engage and captivate audiences by offering them a more immersive and participatory experience with the data
```

Question 2

In [1]:

```python
import pandas as pd
import plotly.express as px
```

In [2]:

```python
prod_df = pd.read_csv("share-electricity-renewables.csv")
prod_df
```

Out[2]:

```
           Entity  Code  Year  Renewables (% electricity)
   0  Afghanistan   AFG  2000                   65.957440
   1  Afghanistan   AFG  2001                   84.745766
   2  Afghanistan   AFG  2002                   81.159424
   3  Afghanistan   AFG  2003                   67.021280
   4  Afghanistan   AFG  2004                   62.921350
 ...          ...   ...   ...                         ...
6866     Zimbabwe   ZWE  2017                   58.503407
6867     Zimbabwe   ZWE  2018                   59.412407
6868     Zimbabwe   ZWE  2019                   55.582527
6869     Zimbabwe   ZWE  2020                   55.131580
6870     Zimbabwe   ZWE  2021                   54.975124
```

6871 rows × 4 columns

In [3]:

```
filterprod_df = prod_df[(prod_df['Year']>2007) & (prod_df['Year']<2017)]

prod_sortyear = filterprod_df.sort_values(by="Year")
prod_sortyear

```

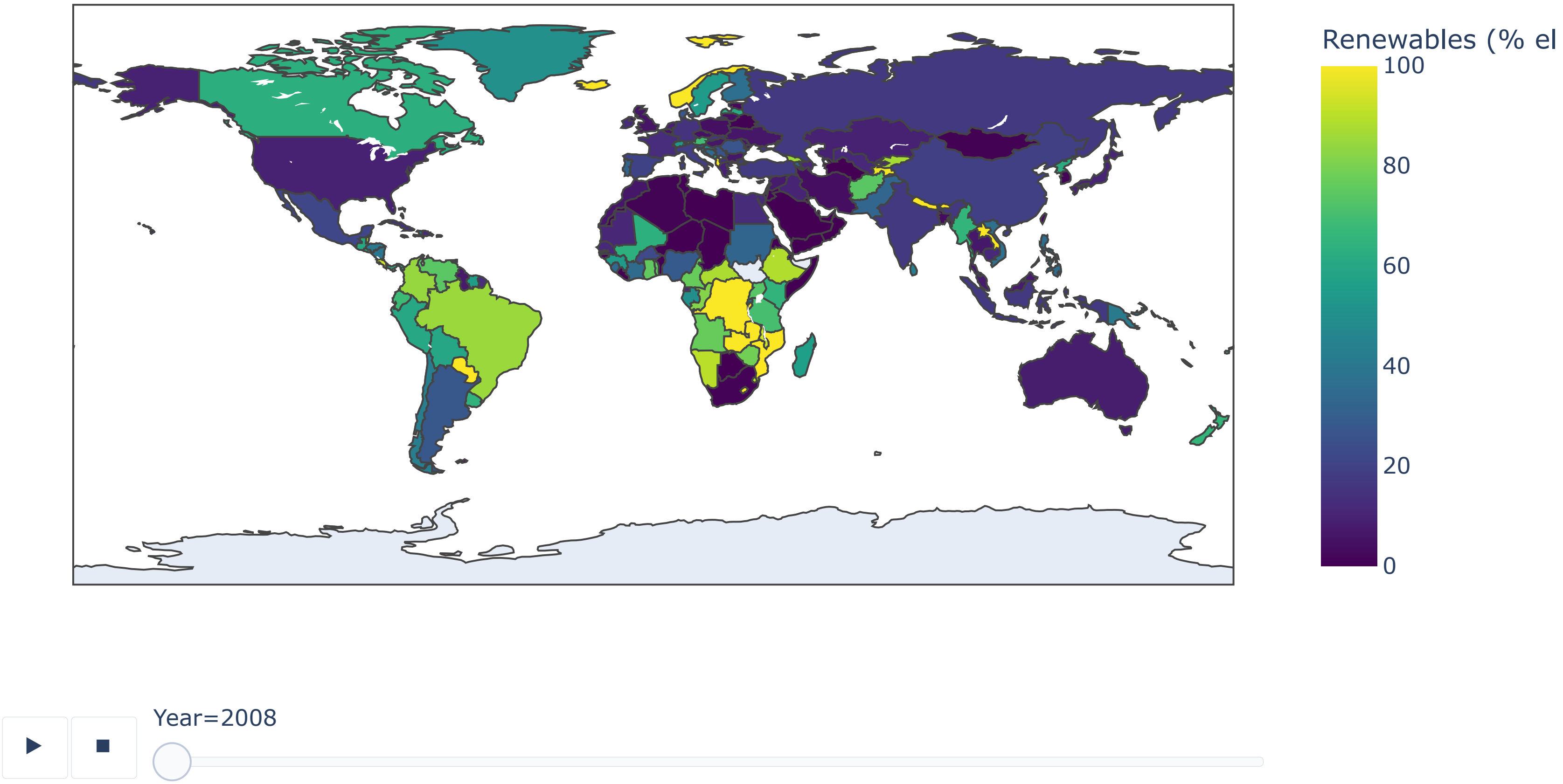
Out[3]:

```
              Entity  Code  Year  Renewables (% electricity)
   8     Afghanistan   AFG  2008                   73.972600
1383           Congo   COG  2008                   81.250000
4735  Oceania (Ember)  NaN  2008                   17.467155
4772            Oman   OMN  2008                    0.000000
1361         Comoros   COM  2008                    0.000000
 ...             ...   ...   ...                         ...
 530      Azerbaijan   AZE  2016                    8.485363
4357       Nicaragua   NIC  2016                   53.243855
1587         Czechia   CZE  2016                   11.449823
1647         Denmark   DNK  2016                   60.196400
6865        Zimbabwe   ZWE  2016                   48.680350

2183 rows × 4 columns
```

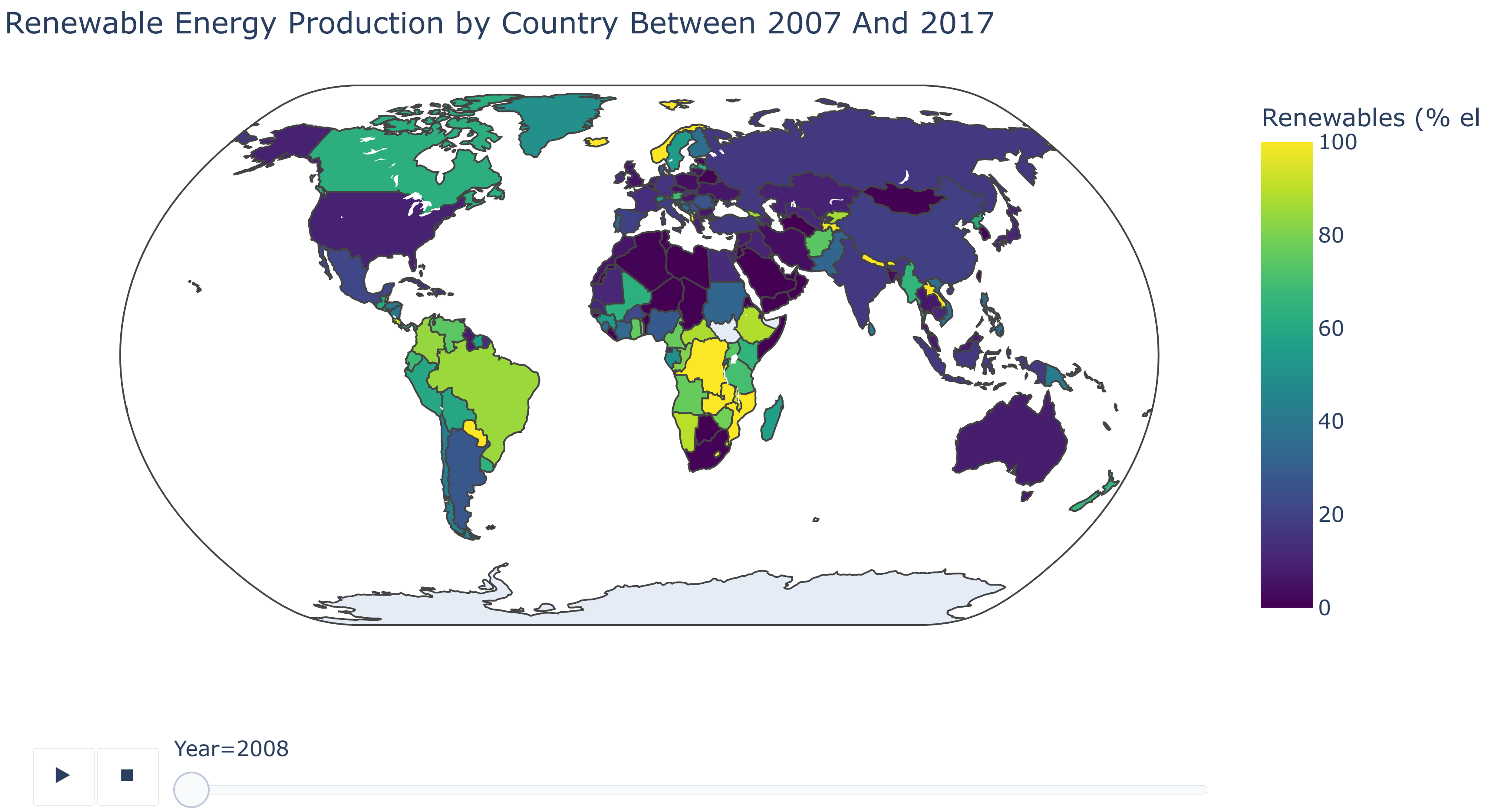
In [4]:

```
fig1 = px.choropleth(prod_sortyear,
                     locations= 'Code',
                     color= 'Renewables (% electricity)',
                     animation_frame= 'Year',
                     color_continuous_scale='viridis',
                     hover_name="Entity")
fig1
```



In [5]:

```
fig1.update_layout(title_text = 'Renewable Energy Production by Country Between 2007 And 2017',
                   geo = dict(projection_type = 'natural earth'))
fig1.show()
```


Renewable Energy Production by Country Between 2007 And 2017
Renewables (% el
100
80
60
40
20
0
Year=2008
▶
■

In [6]:

```python
cons_df = pd.read_csv("renewable-share-energy.csv")
cons_df
```

Out[6]:

```
      Entity      Code  Year  Renewables (% equivalent primary energy)
   0  Africa       NaN  1965                                  5.747495
   1  Africa       NaN  1966                                  6.122062
   2  Africa       NaN  1967                                  6.325731
   3  Africa       NaN  1968                                  7.005293
   4  Africa       NaN  1969                                  7.956088
 ...     ...       ...   ...                                       ...
5598   World  OWID_WRL  2017                                 11.355979
5599   World  OWID_WRL  2018                                 11.741059
5600   World  OWID_WRL  2019                                 12.237987
5601   World  OWID_WRL  2020                                 13.455194
5602   World  OWID_WRL  2021                                 13.470907

5603 rows × 4 columns
```

In [7]:

```python
convert_df = cons_df.melt(id_vars = ['Entity', 'Code'],
                 var_name = 'Year',
                 value_name ='Renewables_Energy_Consumption (Twh)')
convert_df
```

Out[7]:

	Entity	Code	Year	Renewables_Energy_Consumption (Twh)
0	Africa	NaN	Year	1965.000000
1	Africa	NaN	Year	1966.000000
2	Africa	NaN	Year	1967.000000
3	Africa	NaN	Year	1968.000000
4	Africa	NaN	Year	1969.000000
...	...	...	...	...
11201	World	OWID_WRL	Renewables (% equivalent primary energy)	11.355979
11202	World	OWID_WRL	Renewables (% equivalent primary energy)	11.741059
11203	World	OWID_WRL	Renewables (% equivalent primary energy)	12.237987
11204	World	OWID_WRL	Renewables (% equivalent primary energy)	13.455194
11205	World	OWID_WRL	Renewables (% equivalent primary energy)	13.470907

11206 rows × 4 columns

In [8]:

```
filtercons_df = cons_df[(cons_df['Year']>2007) & (cons_df['Year']<2017)]

cons_sortyear= filtercons_df.sort_values(by = 'Year')
cons_sortyear

```

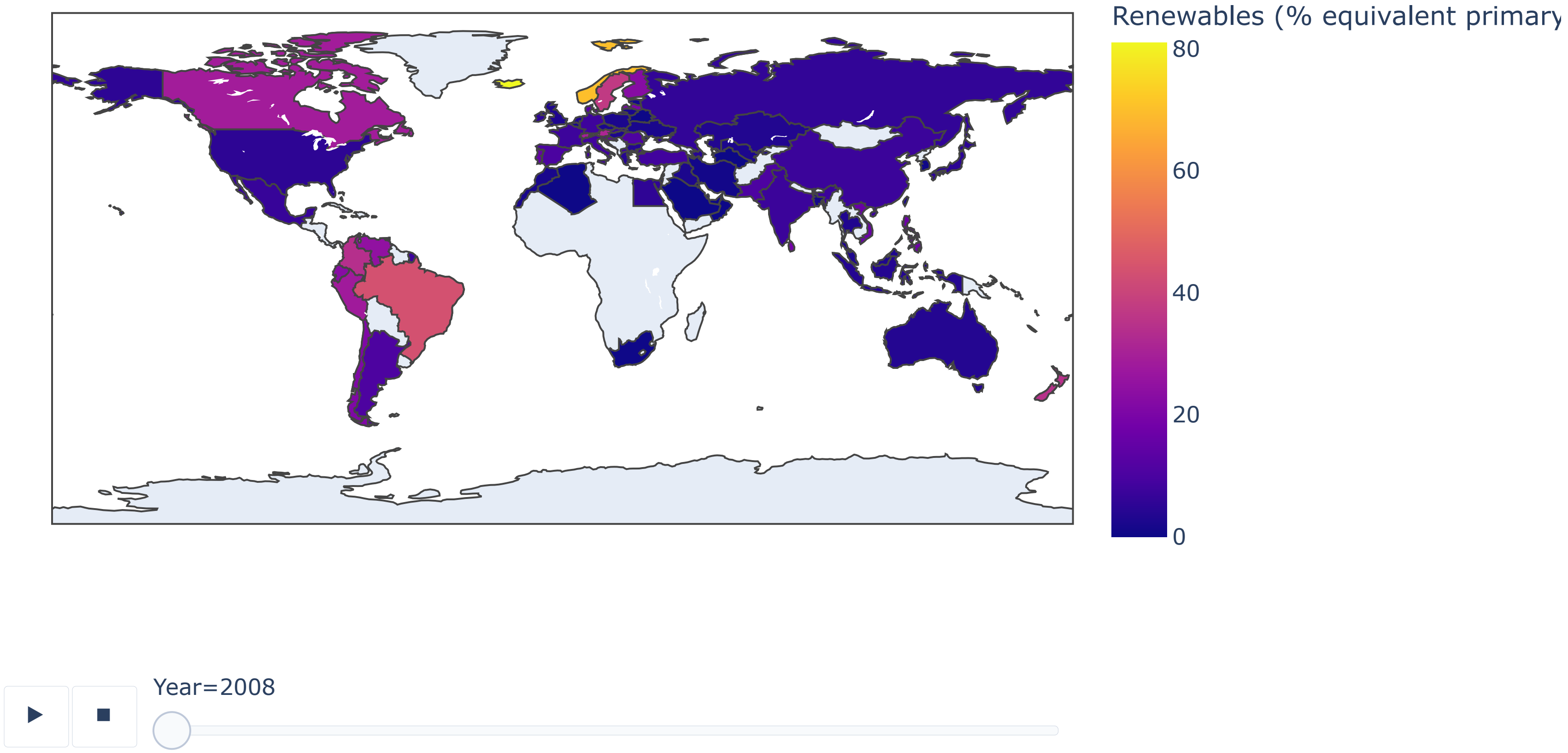
Out[8]:

	Entity	Code	Year	Renewables (% equivalent primary energy)
43	Africa	NaN	2008	6.716138
3456	North Macedonia	MKD	2008	7.438910
795	CIS (BP)	NaN	2008	5.930783
2174	India	IND	2008	7.257065
3513	Norway	NOR	2008	69.895430
...	...	...	...	...
3464	North Macedonia	MKD	2016	18.146008
746	Bulgaria	BGR	2016	9.049891
2182	India	IND	2016	7.011397
5162	United Arab Emirates	ARE	2016	0.067022
5597	World	OWID_WRL	2016	10.930195

927 rows × 4 columns

In [9]:

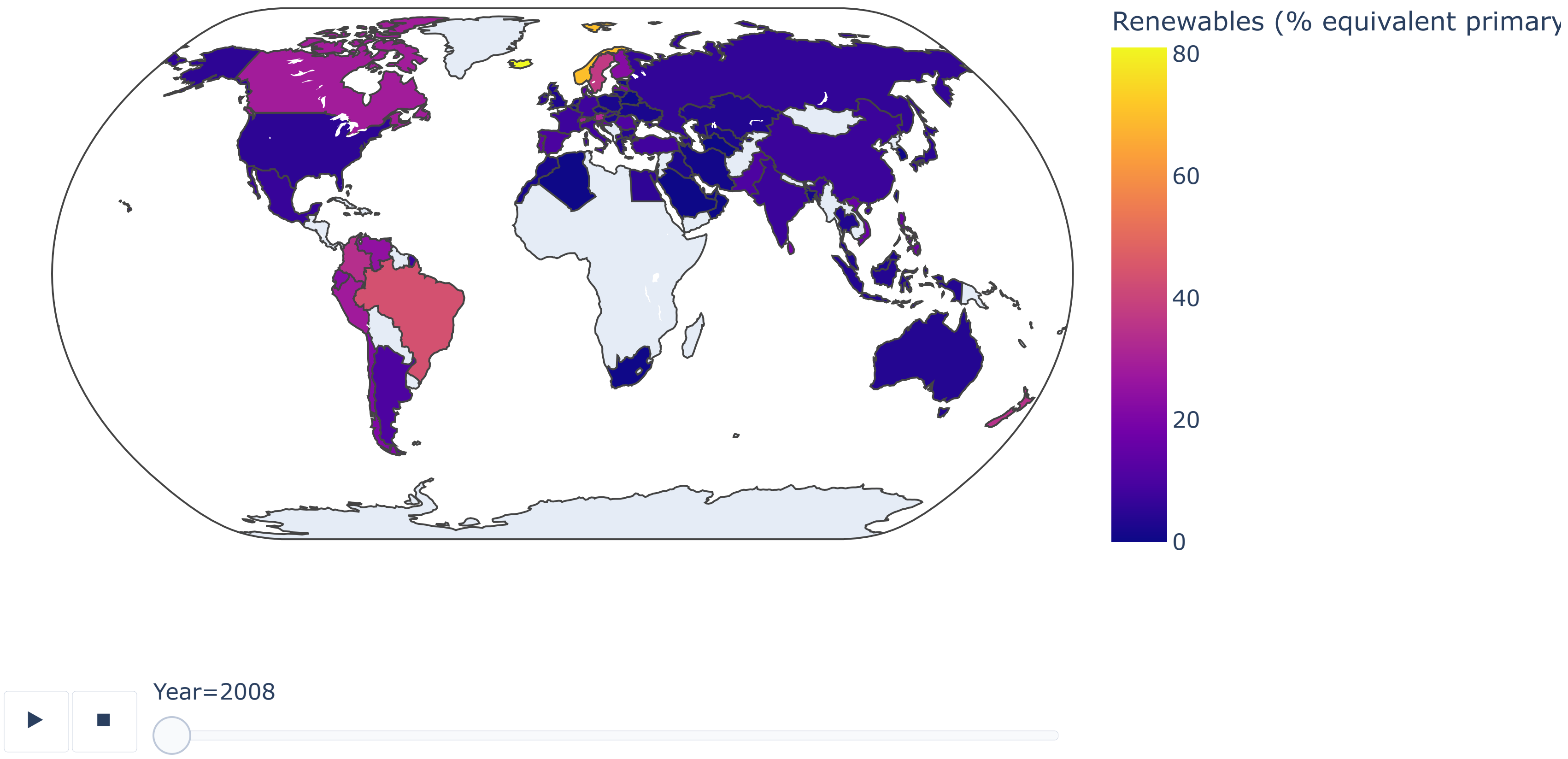
```
fig2 = px.choropleth(cons_sortyear,
                     locations= 'Code',
                     color= 'Renewables (% equivalent primary energy)',
                     animation_frame= 'Year',
                     color_continuous_scale='plasma',
                     hover_name="Entity")
fig2
```



In [10]:

```python
fig2.update_layout(title_text="Renewable Energy Consumption by County Between 2007 And 2017",
                   geo=dict(projection_type="natural earth"))
fig2.show()
```

Renewable Energy Consumption by County Between 2007 And 2017


Renewables (% equivalent primary
80
60
40
20
0
▶
■
Year=2008