

Importing the needed libraries

```
In [1]: import plotly.express as px
import gc
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

import warnings
warnings.filterwarnings('ignore')
```

Adding the Dataset

```
In [2]: Dataset = pd.read_csv('PetrolDataset_june_20_2022.csv', encoding='latin-1')
```

Dataset Pre-work visualizing

```
In [3]: Dataset
Out[3]:
```

#	Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita	Price Per Gallon (USD)	Price Per Liter (USD)	Price Per Liter (QMR)
0	United States	19687287	20.30%	934.3	5.19	1.37	0.5343
1	China	12791953	13.20%	138.7	5.42	1.43	0.5577
2	India	4443000	4.60%	51.4	5.05	1.33	0.5187
3	Japan	4012877	4.10%	481.5	4.69	1.24	0.4836
4	Russia	3631287	3.70%	383.2	3.41	0.90	0.3510
...
175	Saint Pierre & Miquelon	660	0.00%	1705.1	8.27	2.19	0.8541
176	Kiribati	400	0.00%	54.5	4.09	1.08	0.412
177	Montserrat	400	0.00%	1231.1	4.57	1.21	0.4719
178	Saint Helena	70	0.00%	180.2	5.53	1.46	0.5694
179	Niue	51	0.00%	484.4	11.43	3.02	1.1778

180 rows x 8 columns

Dropping unnecessary columns

```
In [4]: #Dropping (Number, Protocol, Source, Destination) from 'trainDataset'
#Passing a value to carry the columns which will be dropped
columnsToDrop = [ '# ' ]
# Dropping the unneeded columns from axis 1 as its on the first line of the dataset
Dataset = Dataset.drop(columnsToDrop,axis=1)
```

```
In [5]: Dataset.shape
Out[5]: (180, 7)
```

```
In [6]: Dataset.head(4)
Out[6]:
```

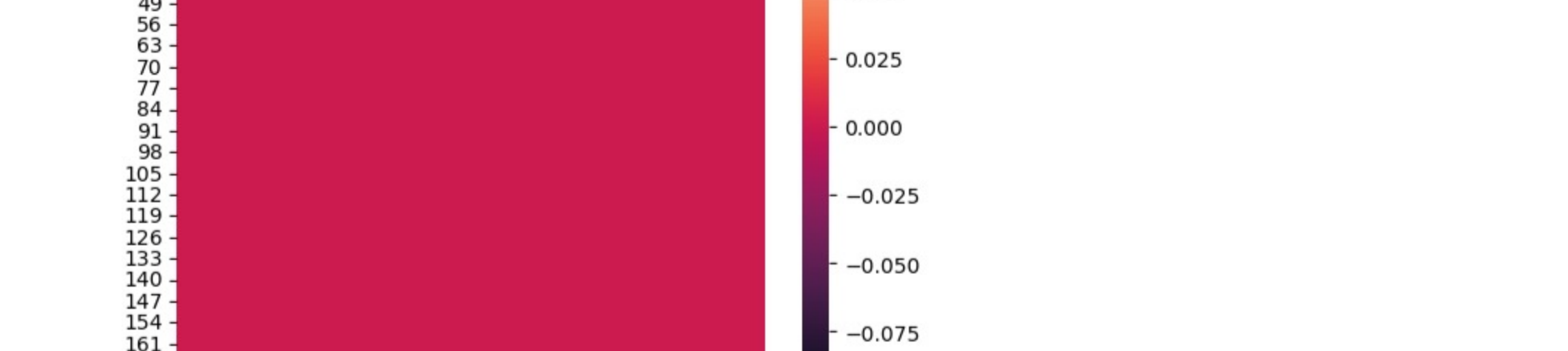
Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita	Price Per Gallon (USD)	Price Per Liter (USD)	Price Per Liter (QMR)
0 United States	19687287	20.30%	934.3	5.19	1.37	0.5343
1 China	12791953	13.20%	138.7	5.42	1.43	0.5577
2 India	4443000	4.60%	51.4	5.05	1.33	0.5187
3 Japan	4012877	4.10%	481.5	4.69	1.24	0.4836

```
In [7]: Dataset.columns
Out[7]: Index(['Country', 'Daily Oil Consumption (Barrels)', 'World Share',
'Yearly Gallons Per Capita', 'Price Per Gallon (USD)',
'Price Per Liter (USD)', 'Price Per Liter (QMR)'],
dtype='object')
```

```
In [8]: Dataset.isnull().sum()
Out[8]:
```

Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita	Price Per Gallon (USD)	Price Per Liter (USD)	Price Per Liter (QMR)
Country	0	0	0	0	0	0
Daily Oil Consumption (Barrels)	180 non-null	180 non-null	180 non-null	180 non-null	180 non-null	180 non-null
World Share	0	0	0	0	0	0
Yearly Gallons Per Capita	0	0	0	0	0	0
Price Per Gallon (USD)	0	0	0	0	0	0
Price Per Liter (USD)	0	0	0	0	0	0
Price Per Liter (QMR)	0	0	0	0	0	0

dtype: int64



```
In [10]: Dataset.info()
Out[10]:
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 180 entries, 0 to 179
Data columns (total 7 columns):
 #   Column              Non-Null Count  Dtype
---  ---
 0   Country              180 non-null    object
 1   Daily Oil Consumption (Barrels)  180 non-null    int64
 2   World Share          180 non-null    object
 3   Yearly Gallons Per Capita  180 non-null    float64
 4   Price Per Gallon (USD)  180 non-null    float64
 5   Price Per Liter (USD)    180 non-null    float64
 6   Price Per Liter (QMR)    180 non-null    float64
dtypes: float64(4), int64(1), object(2)
memory usage: 10.0v KB
```

```
In [11]: Dataset.describe().T
Out[11]:
```

	count	mean	std	min	25%	50%	75%	max
Daily Oil Consumption (Barrels)	180.0	53524.850000	1.863177e+06	51.0000	20027.00000	60886.00000	246514.00000	1.968728e+07
Yearly Gallons Per Capita	180.0	331.370000	4.378922e+02	2.2000	4.150000	177.40000	424.225000	3.679500e+03
Price Per Gallon (USD)	180.0	5.683222	4.378446e+00	0.0800	4.135000	5.27500	6.752500	5.489000e+01
Price Per Liter (USD)	180.0	1.501833	1.159937e+00	0.0200	1.09500	1.36500	1.782500	1.450000e+01
Price Per Liter (QMR)	180.0	0.585719	4.512053e-01	0.0078	0.42705	0.54405	0.695175	5.955000e+00

```
In [12]: def analyze(Dataset):
    print()
    col = Dataset.columns.values
    total = float(len(Dataset))

    print(f"Total Number of Rows:{total}")
    for col in col:
        unique = Dataset[col].unique()
        unique_count = len(unique)
        if unique_count > 10:
            print(f"() Unique Data {col}.format(col,unique_count)
        else:
            pass

analyze(Dataset)

180 Total Number of Rows
Country: 180 Unique Data
Daily Oil Consumption (Barrels): 155 Unique Data
World Share: 37 Unique Data
Yearly Gallons Per Capita: 179 Unique Data
Price Per Gallon (USD): 179 Unique Data
Price Per Liter (USD): 117 Unique Data
Price Per Liter (QMR): 117 Unique Data
```

```
In [13]: Dataset.head(4)
Out[13]:
```

Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita	Price Per Gallon (USD)	Price Per Liter (USD)	Price Per Liter (QMR)
0 United States	19687287	20.30%	934.3	5.19	1.37	0.5343
1 China	12791953	13.20%	138.7	5.42	1.43	0.5577
2 India	4443000	4.60%	51.4	5.05	1.33	0.5187
3 Japan	4012877	4.10%	481.5	4.69	1.24	0.4836

```
In [14]: Dataset.columns
Out[14]: Index(['Country', 'Daily Oil Consumption (Barrels)', 'World Share',
'Yearly Gallons Per Capita', 'Price Per Gallon (USD)',
'Price Per Liter (USD)', 'Price Per Liter (QMR)'],
dtype='object')
```

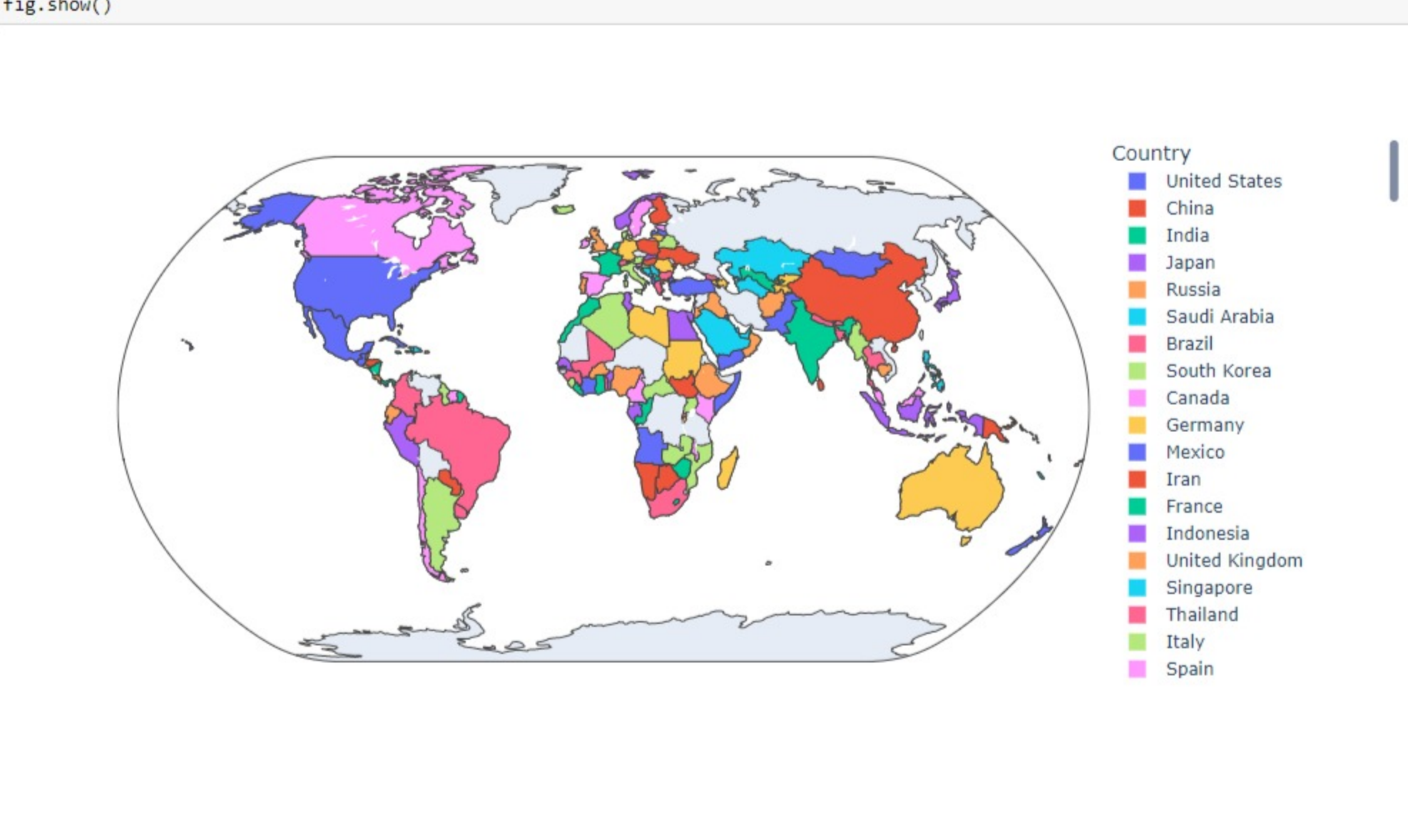
All Data Results are Plotted Out on a Worldwide Map

```
In [15]: countries = {}
for country in pycountry.countries:
    countries[country.name] = country.alpha_3

Dataset['Code'] = [countries.get(names, 'Unknown code') for names in Dataset['Country']]

fig = px.choropleth(Dataset, locations='Code',
                    hover_name='Country',
                    hover_data=Dataset.columns,
                    color='Country',
                    color_continuous_scale='Viridis',
                    projection='natural earth'
                    )

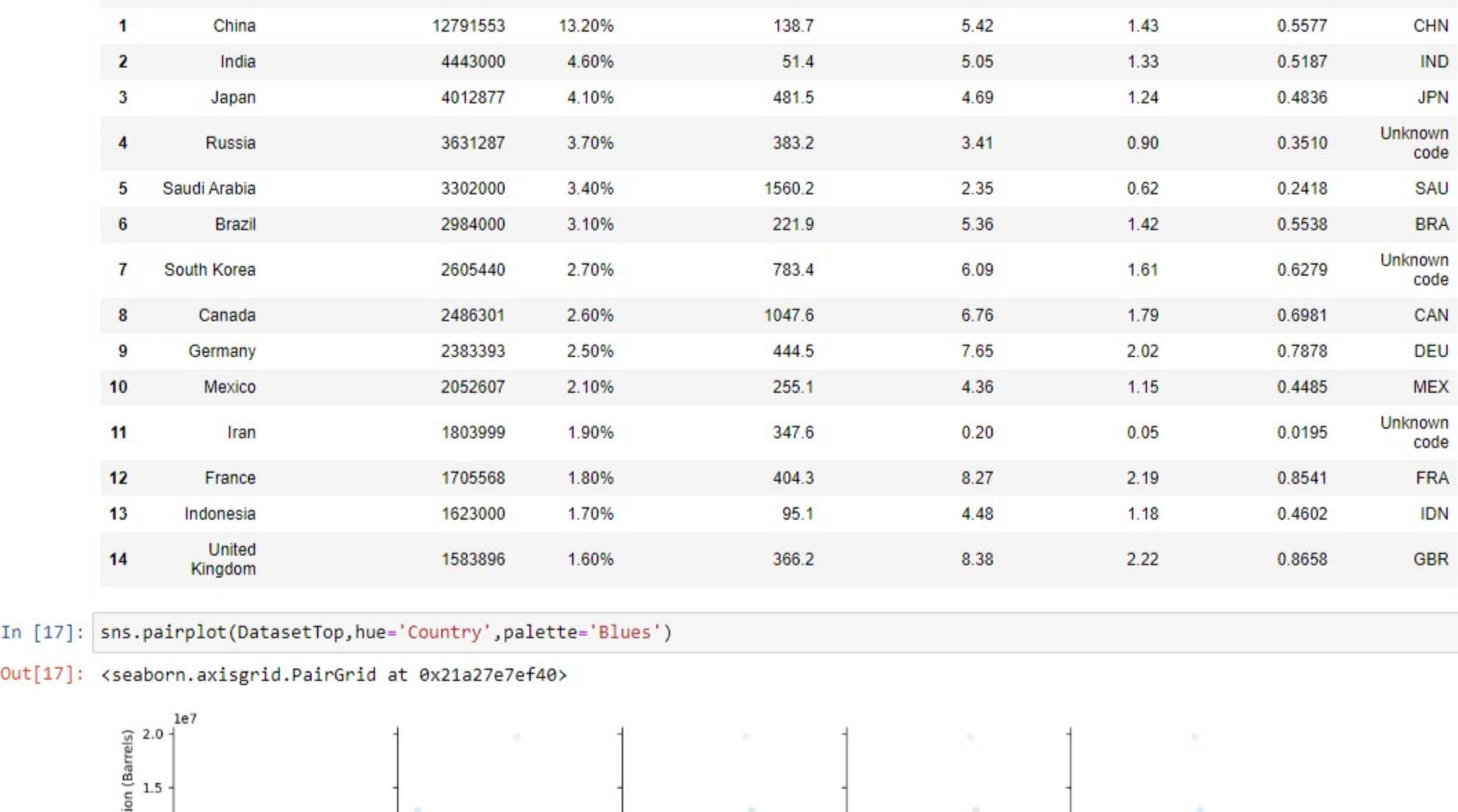
fig.show()
```



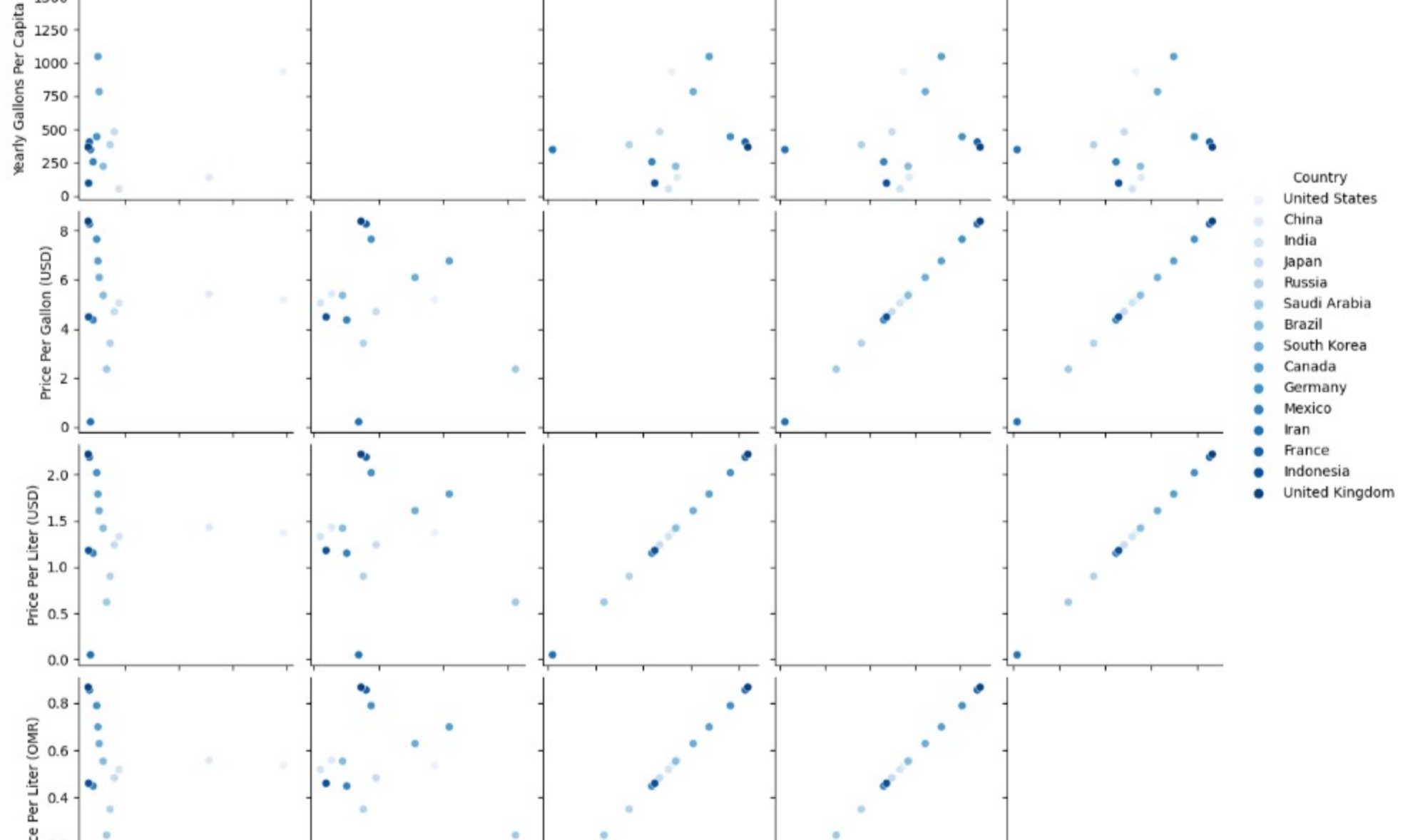
Analyzing Only the Top 15 Country

According to Daily Oil Consumption

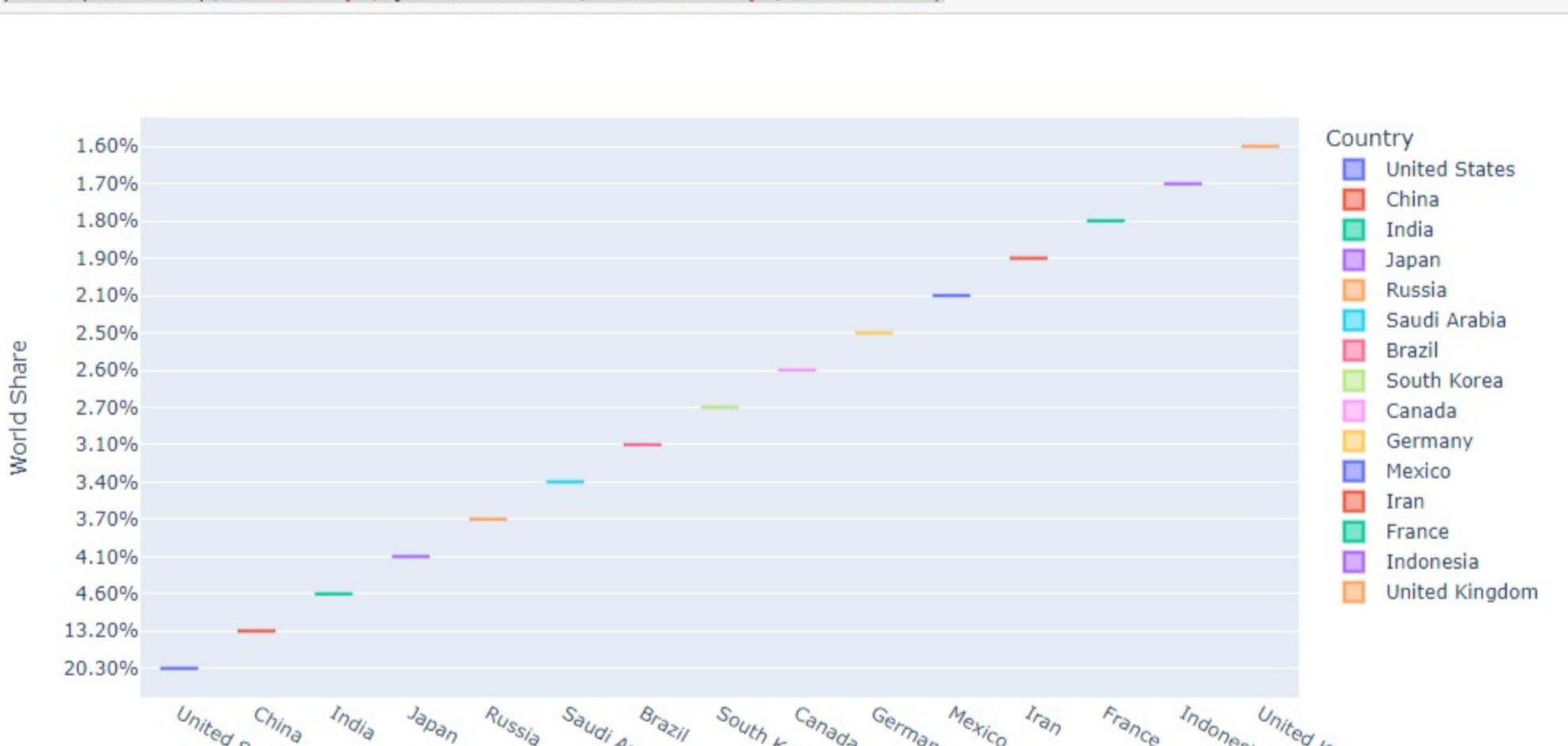
```
In [16]: DatasetTop = Dataset.sort_values(by=['Daily Oil Consumption (Barrels)'], ascending=False)
DatasetTop = DatasetTop.head(15)
DatasetTop
```



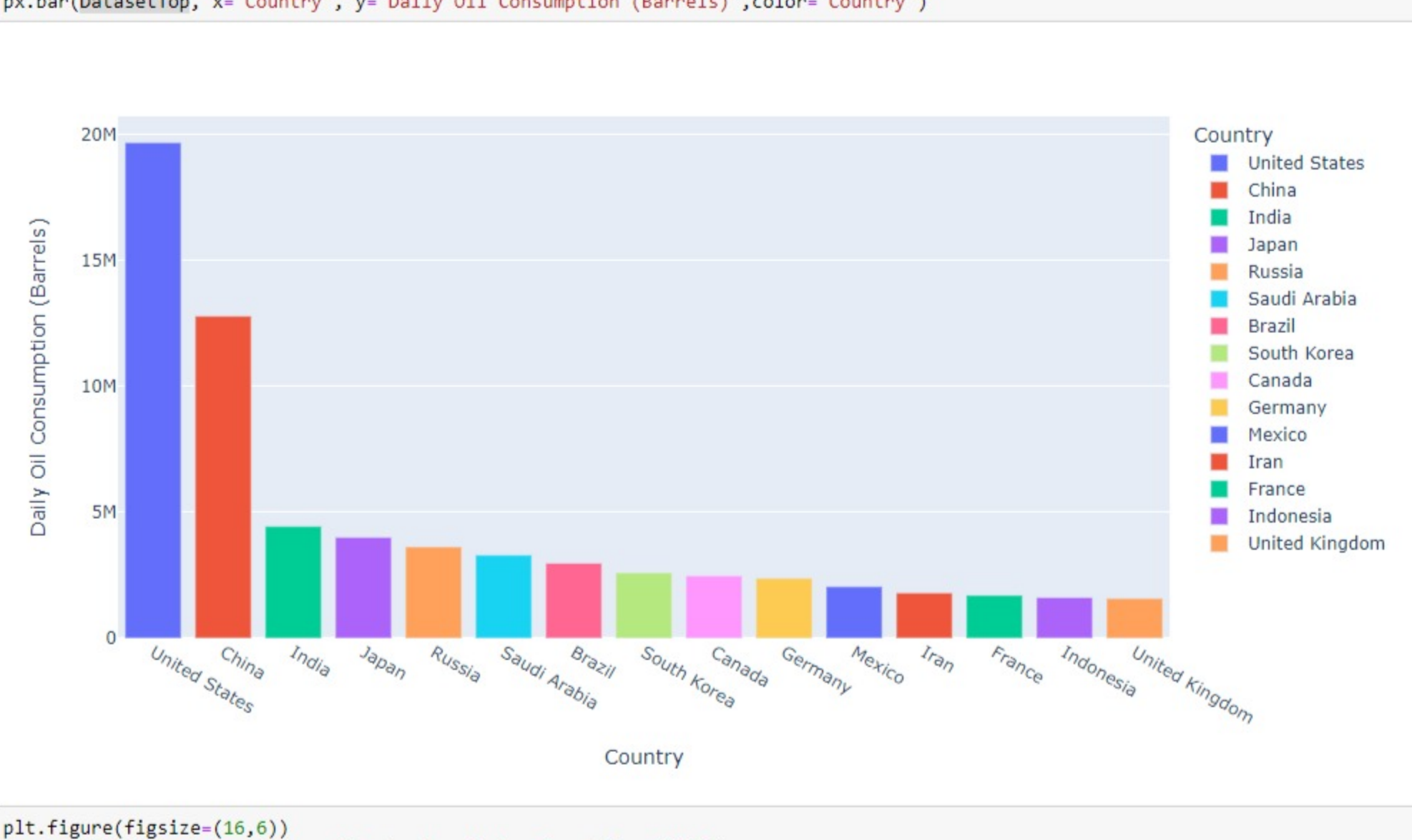
```
In [17]: sns.pairplot(DatasetTop,hue='Country',palette='Blues')
Out[17]: <seaborn.axisgrid.PairGrid at 8a21a27e7e48>
```



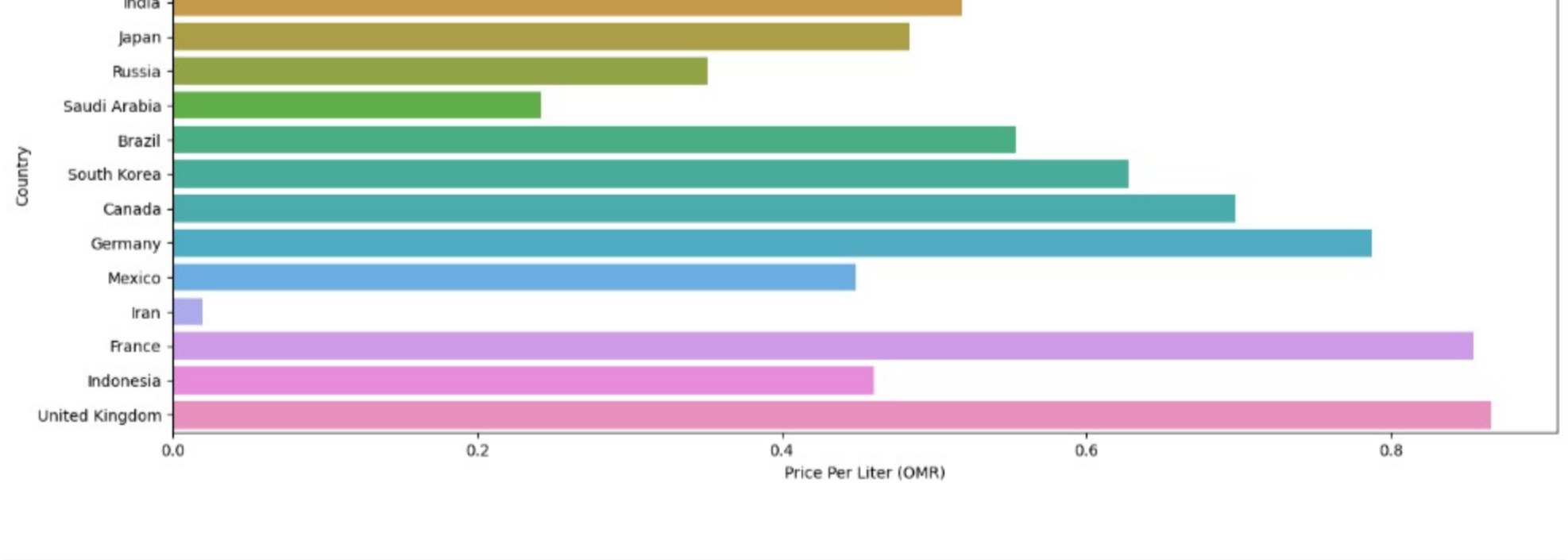
```
In [18]: px.box(DatasetTop, x='Country', y='World Share',color='Country',notched=True)
Out[18]:
```



```
In [19]: # Lets see the top 10 countries where have the Highest Daily Oil Consumption
px.bar(DatasetTop, x='Country', y='Daily Oil Consumption (Barrels)',color='Country')
```



```
In [20]: plt.figure(figsize=(16,6))
sns.barplot(data=DatasetTop,y='Country',x='Price Per Liter (QMR)')
Out[20]: <AxesSubplot:Label='Price Per Liter (QMR)', ylabel='Country'>
```



Analyzing the Lowest 15 Country

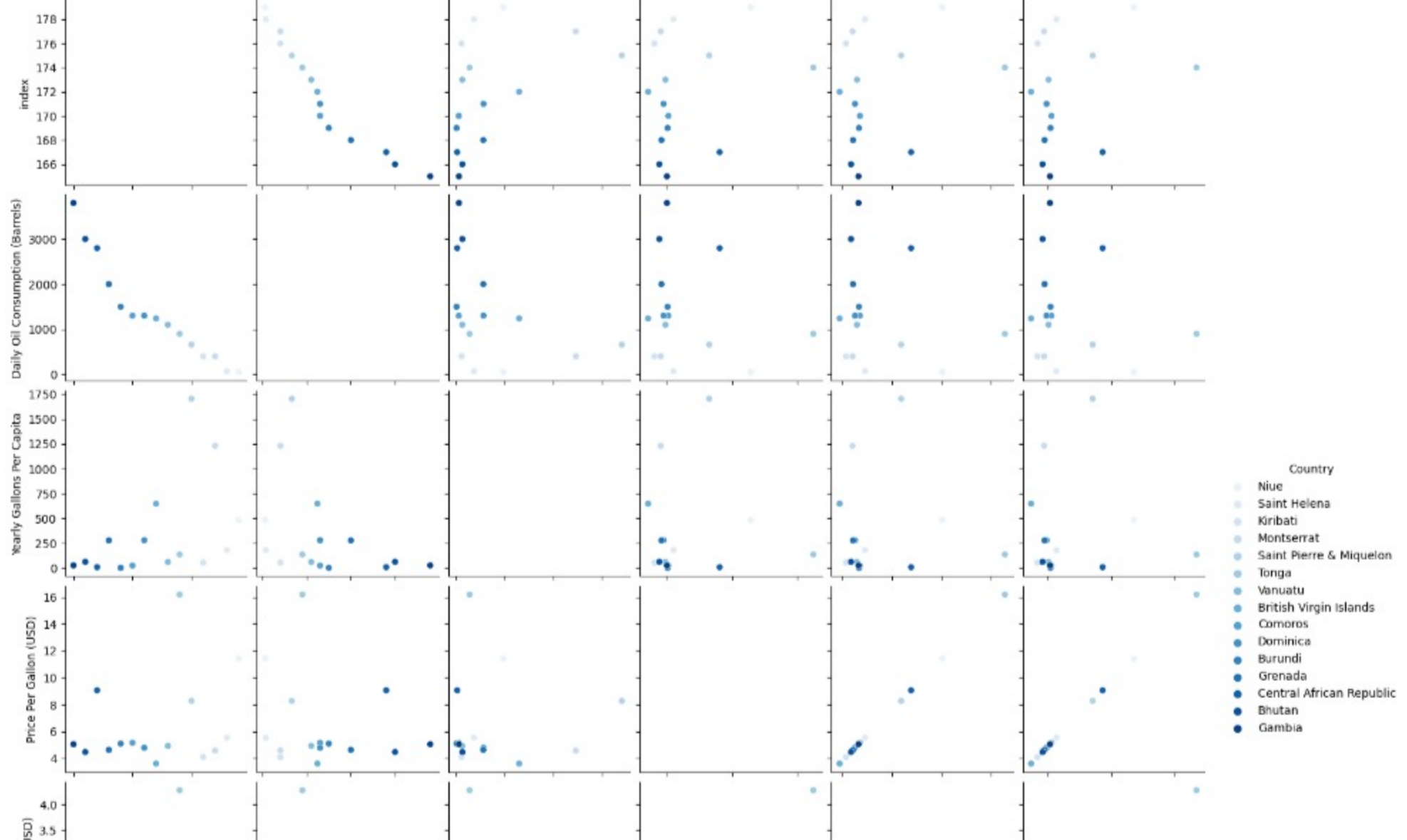
According to Daily Oil Consumption

```
In [21]: lowest_per_capita=Dataset.reset_index(15,'Daily Oil Consumption (Barrels)').reset_index()
lowest_per_capita
```

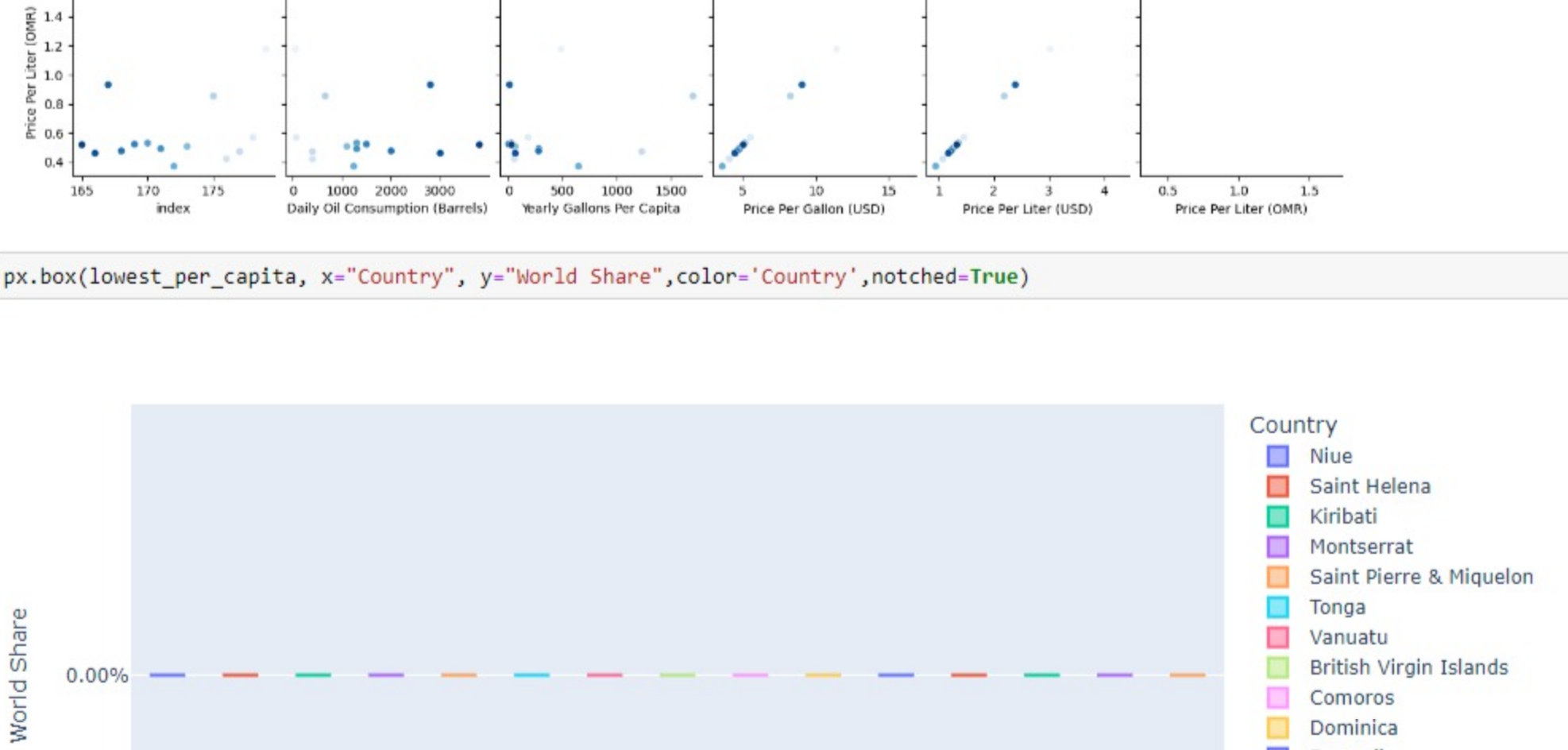
```
Out[21]:
```

Index	Country	Daily Oil Consumption (Barrels)	World Share	Yearly Gallons Per Capita	Price Per Gallon (USD)	Price Per Liter (USD)	Price Per Liter (QMR)	Code
0	179 Niue	51	0.00%	484.4	11.43	3.02	1.1778	NIU
1	178 Saint Helena	70	0.00%	180.2	5.53	1.46	0.5694	Unknown code
2	176 Kiribati	400	0.00%	54.5	4.09	1.08	0.412	KIR
3	177 Montserrat	400	0.00%	1231.1	4.57	1.21	0.4719	MSR
4	175 Saint Pierre & Miquelon	660	0.00%	1705.1	8.27	2.19	0.8541	Unknown code
5	174 Tonga	889	0.00%	136.3	16.20	4.28	1.6692	TON
6	173 Vanuatu	1100	0.00%	60.6	4.92	1.30	0.5070	VUT
7	172 British Virgin Islands	1240	0.00%	647.6	3.60	0.95	0.3705	Unknown code
8	170 Comoros	1300	0.00%	25.0	5.15	1.36	0.5304	COM
9	171 Dominica	1301	0.00%	279.7	4.78	1.26	0.4914	DMA
10	169 Burundi	1499	0.00%	2.2	5.99	1.54	0.5926	BUR
11	168 Grenada	2000	0.00%	278.1	4.62	1.22	0.4758	GRD
12	167 Central African Republic	2600	0.00%	9.5	9.06	2.39	0.9321	CAF
13	166 Bhutan	3001	0.00%	62.4	4.46	1.18	0.4602	BTN
14	165 Gambia	3800	0.00%	27.1	5.94	1.53	0.5187	GMB

```
In [22]: sns.pairplot(lowest_per_capita,hue='Country',palette='Blues')
Out[22]: <seaborn.axisgrid.PairGrid at 8a21a26b6b78>
```



```
In [23]: px.box(lowest_per_capita, x='Country', y='World Share',color='Country',notched=True)
Out[23]:
```



```
In [24]: # Lets see the top 10 countries where have the Highest Daily Oil Consumption
px.bar(lowest_per_capita, x='Country', y='Daily Oil Consumption (Barrels)',color='Country')
```



```
In [25]: plt.figure(figsize=(16,6))
sns.barplot(data=lowest_per_capita,y='Country',x='Price Per Liter (QMR)')
Out[25]: <AxesSubplot:Label='Price Per Liter (QMR)', ylabel='Country'>
```

