

Name : **Omkar Shinde**

Div : **E**

Roll No : **582**

Assignment "5" : **Operations on Matplotlib**

PRN : **202201040190**

## DataSet :



KUMARRAJARSHI · UPDATED 5 YEARS AGO

1006

New Notebook

Download (121 kB)



## Life Expectancy (WHO)

Statistical Analysis on factors influencing Life Expectancy



Data Card

Code (198)

Discussion (24)

### About Dataset

#### Context

Although there have been lot of studies undertaken in the past on factors affecting life expectancy considering demographic variables, income composition and mortality rates. It was found that affect of immunization and human development index was not taken into account in the past. Also, some of the past research was done considering multiple linear regression based on data set of one year for all the countries. Hence, this gives motivation to resolve both the factors stated previously by formulating a regression model based on mixed effects model and multiple linear regression while considering data from a period of 2000 to 2015 for all the countries. Important immunization like Hepatitis B, Polio and Diphtheria will also be considered. In a nutshell, this study will focus on immunization factors, mortality factors, economic factors, social factors and other health related factors as well. Since the observations this dataset are based on different countries, it will be easier for a country to determine the predicting factor which is contributing to lower value of life expectancy. This will help in suggesting a country which area should be given importance in order to efficiently improve the life expectancy of its population.

#### Usability

8.24

#### License

Other (specified in description)

#### Expected update frequency

Not specified

#### Tags

Health

Social Science

Data Cleaning

Global

Dataset Source : <https://www.kaggle.com/datasets/kumararajarshi/life-expectancy-who>

```
import matplotlib.pyplot as plt
import seaborn as sns
```

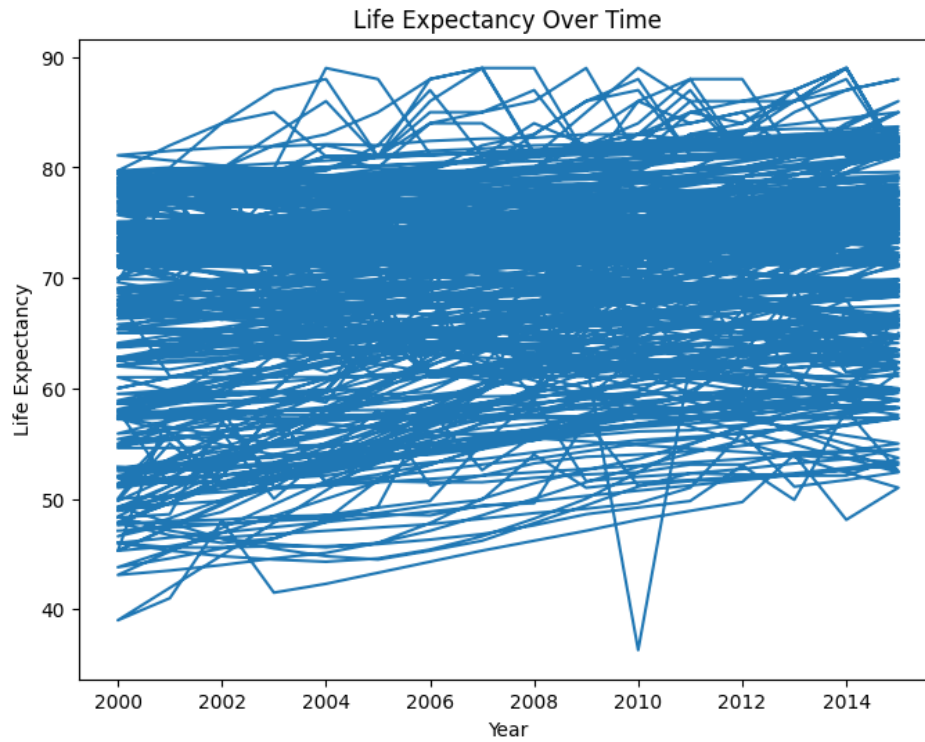
```
import pandas as pd
df = pd.read_csv("Life Expectancy Data.csv")
```

```
print(df.columns)
```

```
Index(['Country', 'Year', 'Status', 'Life expectancy', 'Adult Mortality',
       'infant deaths', 'Alcohol', 'percentage expenditure', 'Hepatitis B',
       'Measles ', ' BMI ', 'under-five deaths ', 'Polio', 'Total expenditure',
       'Diphtheria ', ' HIV/AIDS', 'GDP', 'Population',
       ' thinness 1-19 years', ' thinness 5-9 years',
       'Income composition of resources', 'Schooling'],
      dtype='object')
```

```
plt.figure(figsize=(8, 6))
plt.plot(df['Year'], df['Life expectancy'])
```

```
plt.title('Life Expectancy Over Time')
plt.xlabel('Year')
plt.ylabel('Life Expectancy')
plt.show()
```



```
# Scatter Plot
plt.figure(figsize=(8, 6))
plt.scatter(df['GDP'], df['Life expectancy'])
plt.title('Life Expectancy vs. GDP')
plt.xlabel('GDP')
plt.ylabel('Life Expectancy')
plt.show()
```

```
plt.figure(figsize=(8, 6))
plt.bar(df['Adult Mortality'])
plt.title('Adult Mortality by Country')
plt.xlabel('Country')
plt.ylabel('Adult Mortality')
plt.xticks(rotation=90)
plt.show()
```

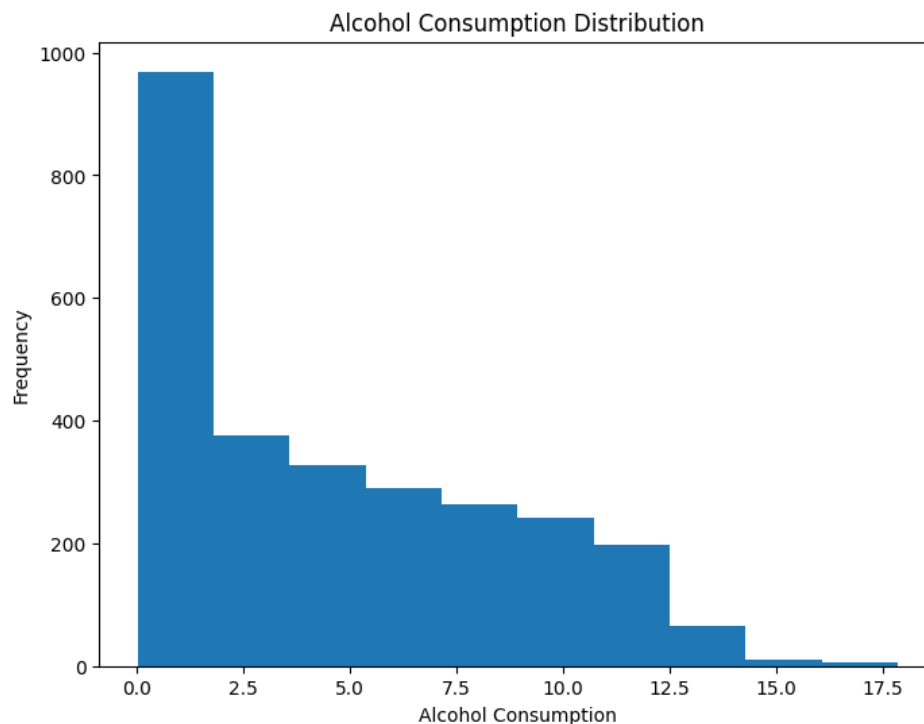
```
-----
TypeError                                Traceback (most recent call last)
<ipython-input-20-fb2bad83a13e> in <cell line: 2>()
      1 plt.figure(figsize=(8, 6))
----> 2 plt.bar(df['Adult Mortality'])
      3 plt.title('Adult Mortality by Country')
      4 plt.xlabel('Country')
      5 plt.ylabel('Adult Mortality')
```

**TypeError:** bar() missing 1 required positional argument: 'height'

[SEARCH STACK OVERFLOW](#)

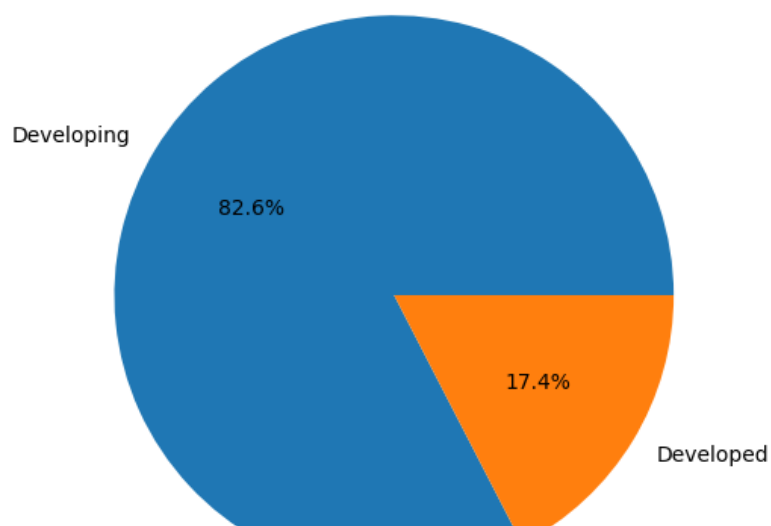
<Figure size 800x600 with 0 Axes>

```
plt.figure(figsize=(8, 6))
plt.hist(df['Alcohol'], bins=10)
plt.title('Alcohol Consumption Distribution')
plt.xlabel('Alcohol Consumption')
plt.ylabel('Frequency')
plt.show()
```

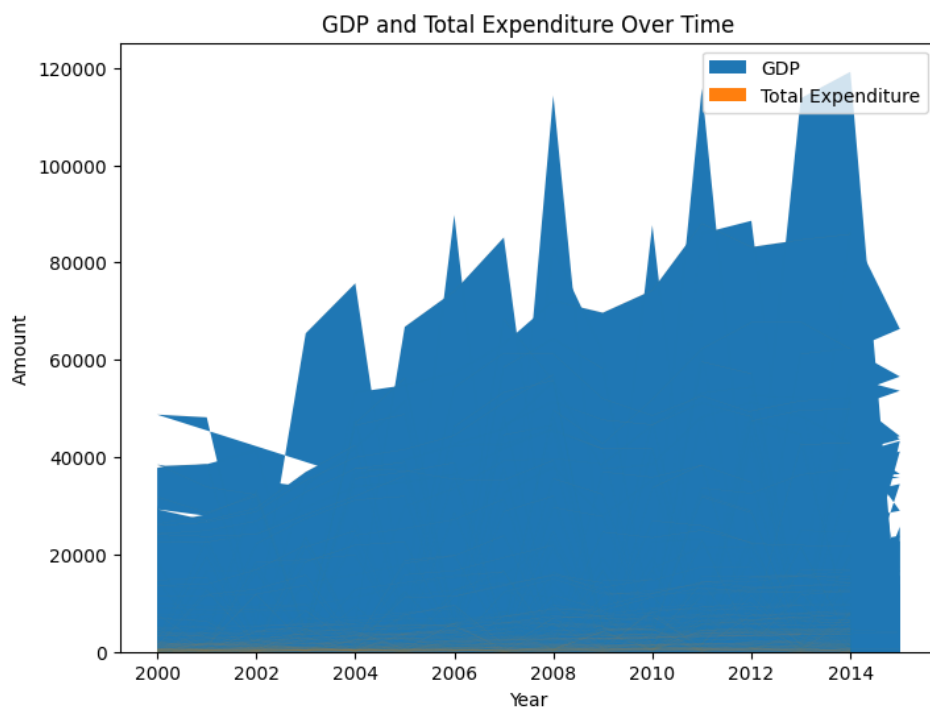


```
# Pie Chart
plt.figure(figsize=(8, 6))
plt.pie(df['Status'].value_counts(), labels=df['Status'].unique(), autopct='%1.1f%%')
plt.title('Distribution of Country Status')
plt.show()
```

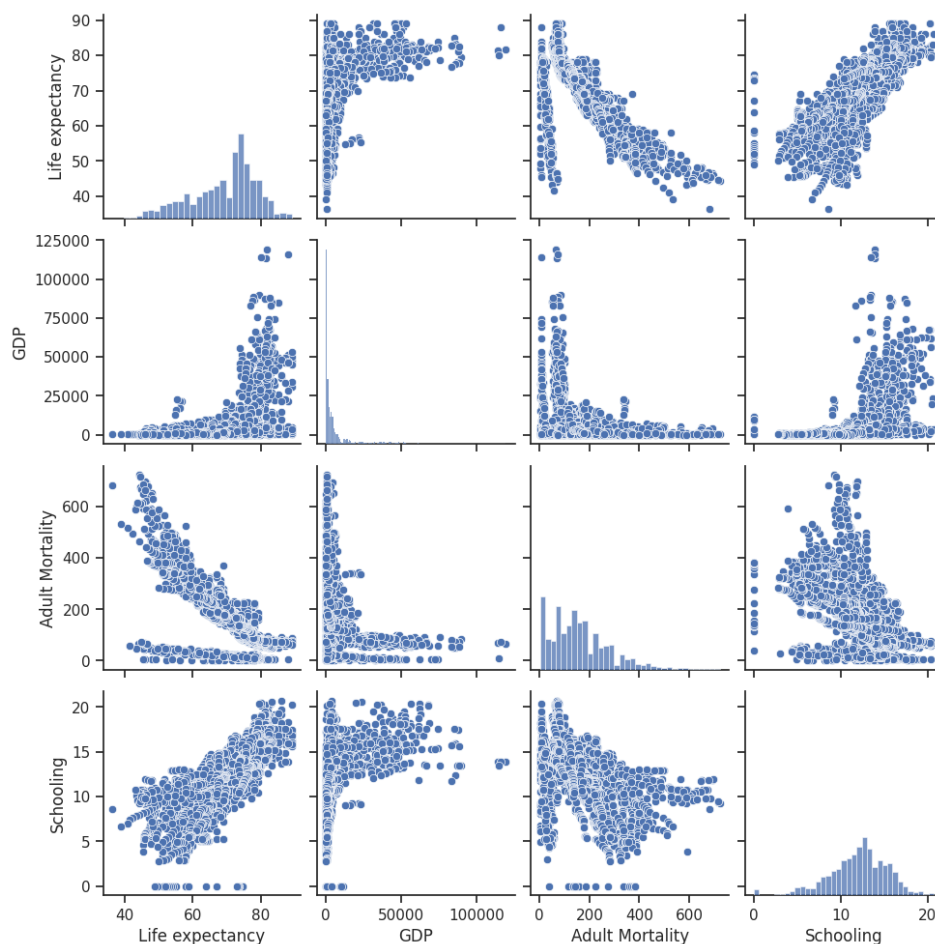
Distribution of Country Status



```
# Area Plot
plt.figure(figsize=(8, 6))
plt.stackplot(df['Year'], df['GDP'], df['Total expenditure'], labels=['GDP', 'Total Expenditure'])
plt.title('GDP and Total Expenditure Over Time')
plt.xlabel('Year')
plt.ylabel('Amount')
plt.legend()
plt.show()
```



```
# Scatter Matrix
sns.set(style='ticks')
sns.pairplot(df[['Life expectancy', 'GDP', 'Adult Mortality', 'Schooling']])
plt.show()
```

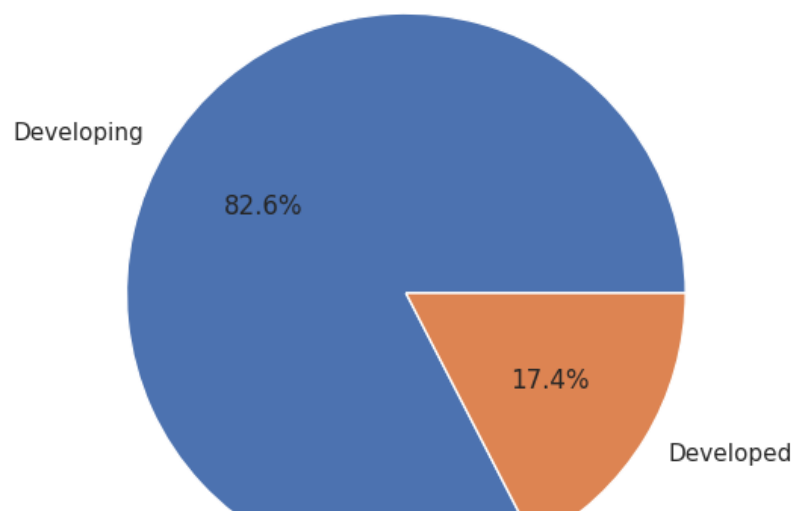


```
country_counts = df['Country'].value_counts()
print(country_counts)
```

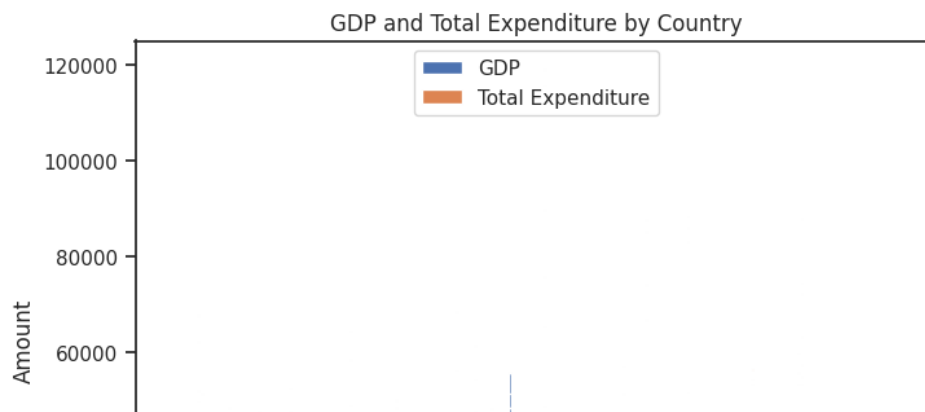
```
Afghanistan      16
Peru              16
Nicaragua        16
Niger            16
Nigeria          16
..
Niue              1
San Marino        1
Nauru             1
Saint Kitts and Nevis 1
Dominica          1
Name: Country, Length: 193, dtype: int64
```

```
plt.figure(figsize=(8, 6))
plt.pie(df['Status'].value_counts(), labels=df['Status'].unique(), autopct='%1.1f%%')
plt.title('Distribution of Country Status')
plt.show()
```

Distribution of Country Status

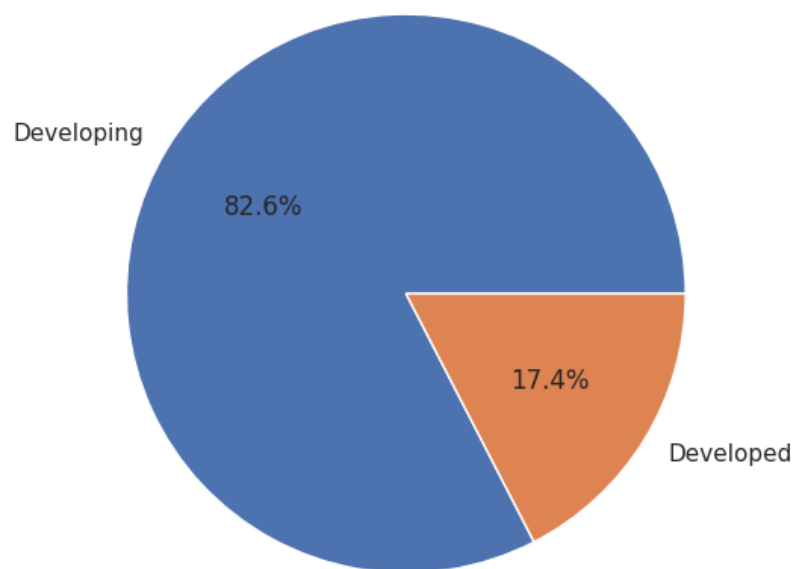


```
plt.figure(figsize=(8, 6))
x = df['Country']
y1 = df['GDP']
y2 = df['Total expenditure']
width = 0.35
plt.bar(x, y1, width, label='GDP')
plt.bar(x, y2, width, label='Total Expenditure', bottom=y1)
plt.title('GDP and Total Expenditure by Country')
plt.xlabel('Country')
plt.ylabel('Amount')
plt.xticks(rotation=90)
plt.legend()
plt.show()
```

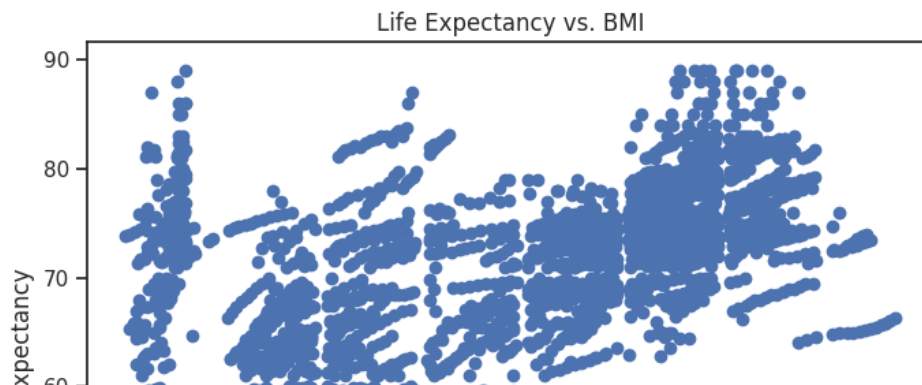


```
plt.figure(figsize=(8, 6))
plt.pie(df['Status'].value_counts(), labels=df['Status'].unique(), autopct='%1.1f%%')
plt.title('Country Status Distribution')
plt.show()
```

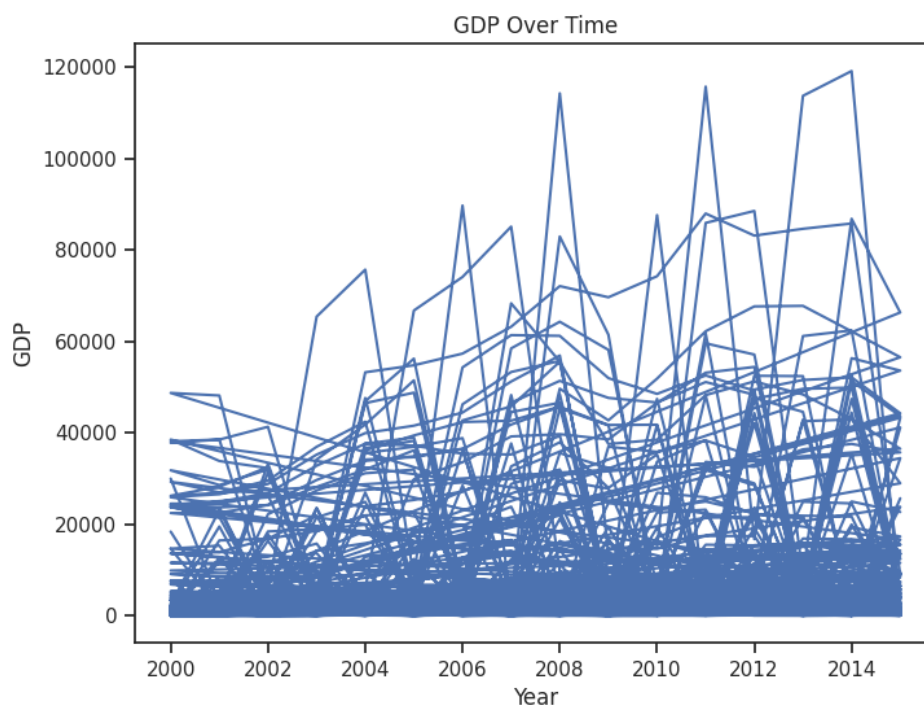
Country Status Distribution



```
# Scatter Plot
plt.figure(figsize=(8, 6))
plt.scatter(df[' BMI '], df['Life expectancy'])
plt.title('Life Expectancy vs. BMI')
plt.xlabel('BMI')
plt.ylabel('Life Expectancy')
plt.show()
```

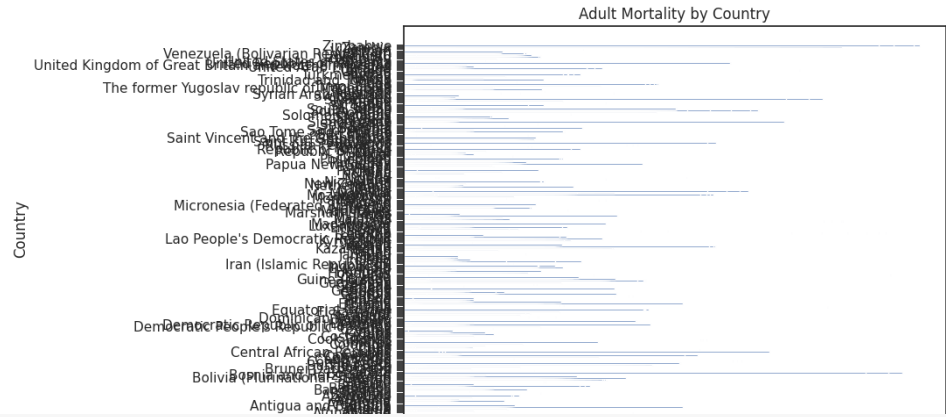


```
# Line Plot
plt.figure(figsize=(8, 6))
plt.plot(df['Year'], df['GDP'])
plt.title('GDP Over Time')
plt.xlabel('Year')
plt.ylabel('GDP')
plt.show()
```



```
# Horizontal Bar Chart
plt.figure(figsize=(8, 6))
plt.barh(df['Country'], df['Adult Mortality'])
plt.title('Adult Mortality by Country')
plt.xlabel('Adult Mortality')
plt.ylabel('Country')
plt.show()
```





```
plt.figure(figsize=(8, 6))
plt.bar(df['Country'], df['GDP'], label='GDP')
plt.bar(df['Country'], df['Total expenditure'], bottom=df['GDP'], label='Total Expenditure')
plt.title('GDP and Total Expenditure by Country')
plt.xlabel('Country')
plt.ylabel('Values')
plt.xticks(rotation=90)
plt.legend()
plt.show()
```

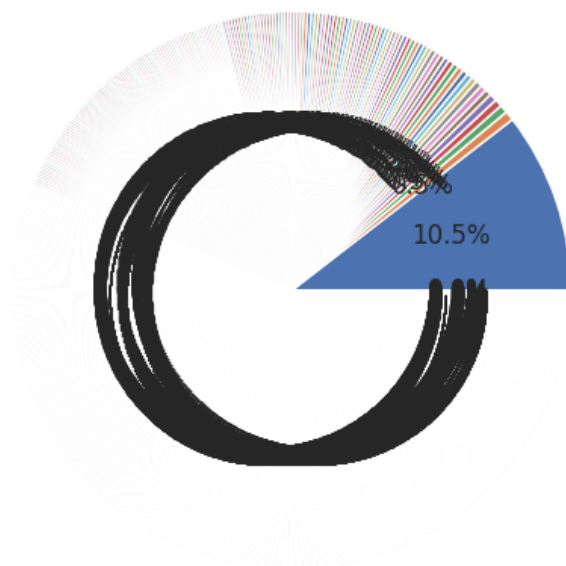
GDP and Total Expenditure by Country

120000



```
plt.figure(figsize=(8, 6))
plt.pie(df['Alcohol'].value_counts(), autopct='%1.1f%%')
plt.title('Alcohol Consumption Categories')
plt.show()
```

Alcohol Consumption Categories

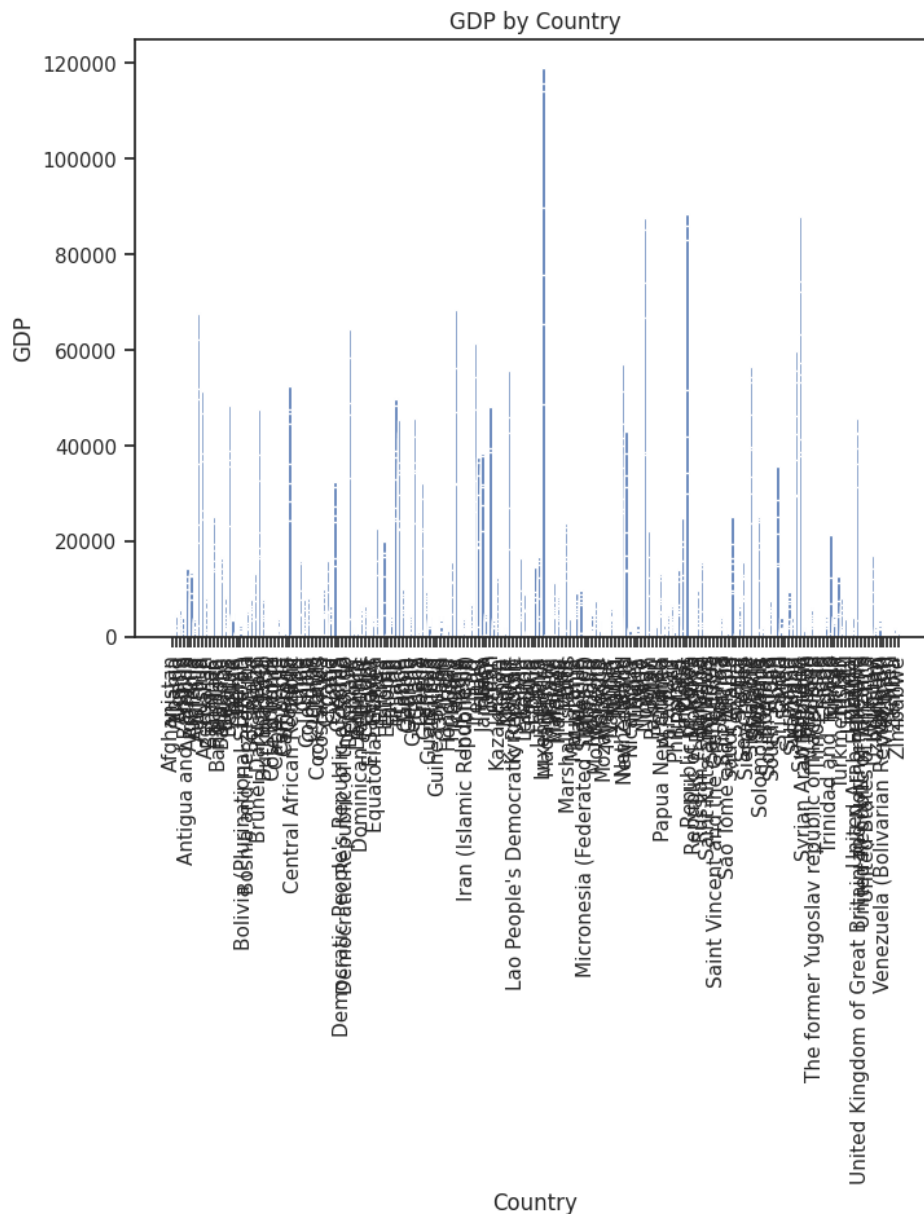


he dc

```
plt.figure(figsize=(8, 6))
plt.pie(df['Income composition of resources'], autopct='%1.1f%%')
plt.title('Income Composition of Resources')
plt.show()
```

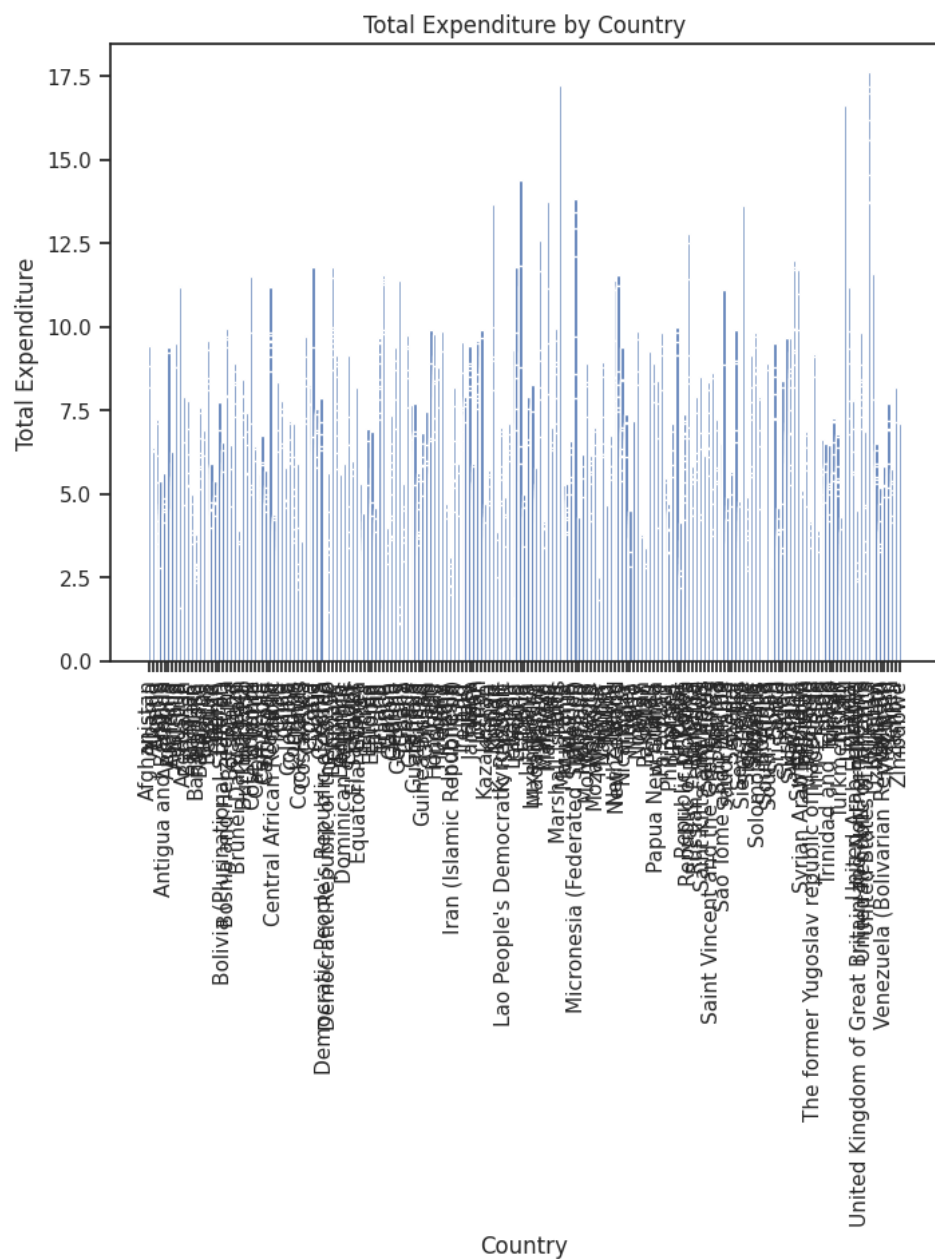
```
ValueError                                Traceback (most recent call last)
<ipython-input-38-35150302c3bb> in <cell line: 2>()
    1 plt.figure(figsize=(8, 6))
----> 2 plt.pie(df['Income composition of resources'], autopct='%1.1f%%')
    3 plt.title('Income Composition of Resources')
    4 plt.show()
```

```
plt.figure(figsize=(8, 6))
plt.bar(df['Country'], df['GDP'])
plt.title('GDP by Country')
plt.xlabel('Country')
plt.ylabel('GDP')
plt.xticks(rotation=90)
plt.show()
```



```
plt.figure(figsize=(8, 6))
plt.bar(df['Country'], df['Total expenditure'])
plt.title('Total Expenditure by Country')
plt.xlabel('Country')
plt.ylabel('Total Expenditure')
plt.xticks(rotation=90)
```

```
plt.show()
```



✓ 7s completed at 10:16 AM

