Plots for data with a single variable

```
# Importing pandas for data & seaborn, pyplot for visualizations
import pandas as pd
import seaborn as sns
import plotly.express as plx
import matplotlib.pyplot as plt
```

Importing the Dataset

Socio-Economic Country Profiles

This dataset contains about 95 statistical indicators of the 66 countries. It covers a broad spectrum of areas including

General Information; Broader Economic Indicators ; Social Indicators ; Environmental & Infrastructure Indicators Military Spending ; Healthcare Indicators ; Trade Related Indicators e.t.c.

```
df = pd.read_csv('/content/soci_econ_country_profiles.csv')
df.head(10)
```

	Unnamed: 0	country	Region	Surface area (km2)	Population in thousands (2017)	Population density (per km2, 2017)	Sex ratio (m per 100 f, 2017)	GDP: Gross domestic product (million current US$)	GDP growth rate (annual %, const. 2005 prices)	GDP per capita (current US$)	...	Inflation, consumer prices (annual %)	expect at bi fe (ye
0	0	Argentina	SouthAmerica	2780400	44271	16.2	95.9	632343	2.4	14564.5	...	NaN	7
1	1	Australia	Oceania	7692060	24451	3.2	99.3	1230859	2.4	51352.2	...	1.948647	8
2	2	Austria	WesternEurope	83871	8736	106.0	96.2	376967	1.0	44117.7	...	2.081269	8
3	3	Belarus	EasternEurope	207600	9468	46.7	87.0	54609	-3.9	5750.8	...	6.031837	7
4	4	Belgium	WesternEurope	30528	11429	377.5	97.3	455107	1.5	40277.8	...	2.125971	8
5	5	Bosnia and Herzegovina	SouthernEurope	51209	3507	68.8	96.4	16251	3.1	4265.0	...	0.810133	7
6	6	Brazil	SouthAmerica	8515767	209288	25.0	96.6	1772591	-3.8	8528.3	...	3.446373	7
7	7	Bulgaria	EasternEurope	111002	7085	65.3	94.6	48953	3.0	6846.8	...	2.064355	7
8	8	Canada	NorthernAmerica	9984670	36624	4.0	98.5	1552808	0.9	43205.6	...	1.596884	8
9	9	Chile	SouthAmerica	756102	18055	24.3	98.2	240796	2.3	13416.2	...	2.182718	8

10 rows × 96 columns

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 66 entries, 0 to 65
Data columns (total 96 columns):
 #   Column                                                Non-Null Count  Dtype
---  ------                                                --------------  -----
 0   Unnamed: 0                                            66 non-null     int64
 1   country                                               66 non-null     object
 2   Region                                                66 non-null     object
 3   Surface area (km2)                                    66 non-null     int64
 4   Population in thousands (2017)                        66 non-null     int64
 5   Population density (per km2, 2017)                    66 non-null     float64
 6   Sex ratio (m per 100 f, 2017)                         66 non-null     float64
 7   GDP: Gross domestic product (million current US$)     66 non-null     int64
 8   GDP growth rate (annual %, const. 2005 prices)        66 non-null     float64
 9   GDP per capita (current US$)                          66 non-null     float64
 10  Economy: Agriculture (% of GVA)                       66 non-null     object
 11  Economy: Industry (% of GVA)                          66 non-null     float64
 12  Economy: Services and other activity (% of GVA)       66 non-null     float64
 13  Employment: Agriculture (% of employed)               66 non-null     float64
 14  Employment: Industry (% of employed)                  66 non-null     float64
 15  Employment: Services (% of employed)                  66 non-null     float64
 16  Unemployment (% of labour force)                      66 non-null     float64
 17  Labour force participation (female/male pop. %)       66 non-null     object
 18  Agricultural production index (2004-2006=100)         66 non-null     int64
 19  Food production index (2004-2006=100)                 66 non-null     int64
 20  International trade: Exports (million US$)            66 non-null     int64
 21  International trade: Imports (million US$)            66 non-null     int64
 22  International trade: Balance (million US$)            66 non-null     int64
```

```
 23  Balance of payments, current account (million US$)          66 non-null     int64
 24  Population growth rate (average annual %)                  66 non-null     object
 25  Urban population (% of total population)_x                  66 non-null     float64
 26  Urban population growth rate (average annual %)             66 non-null     object
 27  Fertility rate, total (live births per woman)               66 non-null     float64
 28  Population age distribution (0-14 / 60+ years, %)           66 non-null     object
 29  International migrant stock (000/% of total pop.)           66 non-null     object
 30  Refugees and others of concern to UNHCR (in thousands)      66 non-null     object
 31  Infant mortality rate (per 1000 live births                 66 non-null     float64
 32  Health: Total expenditure (% of GDP)                        66 non-null     float64
 33  Health: Physicians (per 1000 pop.)                          66 non-null     object
 34  Education: Government expenditure (% of GDP)                66 non-null     object
 35  Education: Primary gross enrol. ratio (f/m per 100 pop.)    66 non-null     object
 36  Education: Secondary gross enrol. ratio (f/m per 100 pop.)  66 non-null     object
 37  Education: Tertiary gross enrol. ratio (f/m per 100 pop.)   66 non-null     object
 38  Seats held by women in national parliaments %               66 non-null     float64
 39  Mobile-cellular subscriptions (per 100 inhabitants)         66 non-null     float64
 40  Mobile-cellular subscriptions (per 100 inhabitants).1       66 non-null     float64
 41  Individuals using the Internet (per 100 inhabitants)        66 non-null     int64
 42  Threatened species (number)                                 66 non-null     float64
 43  Forested area (% of land area)                              66 non-null     object
 44  CO2 emission estimates (million tons/tons per capita)       66 non-null     int64
 45  Energy production, primary (Petajoules)                     66 non-null     int64
 46  Energy supply per capita (Gigajoules)                       66 non-null     object
 47  Pop. using improved drinking water (urban/rural, %)         66 non-null     object
 48  Pop. using improved sanitation facilities (urban/rural, %)  66 non-null     object
 49  Net Official Development Assist. received (% of GNI)        66 non-null     int64
 50  Quality Of Life Index                                       66 non-null     float64
 51  Purchasing Power Index                                      66 non-null     float64
 52  Safety Index                                                66 non-null     float64
```

```
df['Education: Primary gross enrol. ratio (f/m per 100 pop.)']
```

	Education: Primary gross enrol. ratio (f/m per 100 pop.)
0	109.8/110.2
1	102.1/102.3
2	102.2/103.7
3	101.3/101.4
4	104.2/104.2
...	...
61	116.0/116.7
62	108.1/108.4
63	100.0/100.3
64	98.6/101.3
65	108.4/109.3

66 rows × 1 columns

dtype: object

Data Cleaning and pre=processing

```
df.isnull().sum()
```

	0
Unnamed: 0	0
country	0
Region	0
Surface area (km2)	0
Population in thousands (2017)	0
...	...
Population, female	0
Population, male	0
Tax revenue (% of GDP)	8
Taxes on income, profits and capital gains (% of revenue)	9
Urban population (% of total population)_y	0

96 rows × 1 columns

dtype: int64

```
#Drop row with null values
df.dropna(inplace = True)
```

```
# Select only numeric columns and keep 'Region' and 'country'
df = pd.concat([df[['Region', 'country']], df.select_dtypes(include=[float, int])], axis=1)

# Check the updated DataFrame types
df.info()
```

```
 19  International trade: Imports (million US$)                66 non-null     int64
 20  International trade: Balance (million US$)                66 non-null     int64
 21  Balance of payments, current account (million US$)        66 non-null     int64
 22  Urban population (% of total population)_x                66 non-null     float64
 23  Fertility rate, total (live births per woman)             66 non-null     float64
 24  Infant mortality rate (per 1000 live births               66 non-null     float64
 25  Health: Total expenditure (% of GDP)                      66 non-null     float64
 26  Seats held by women in national parliaments %             66 non-null     float64
 27  Mobile-cellular subscriptions (per 100 inhabitants)       66 non-null     float64
 28  Mobile-cellular subscriptions (per 100 inhabitants).1     66 non-null     float64
 29  Individuals using the Internet (per 100 inhabitants)      66 non-null     int64
 30  Threatened species (number)                               66 non-null     float64
 31  CO2 emission estimates (million tons/tons per capita)     66 non-null     int64
 32  Energy production, primary (Petajoules)                   66 non-null     int64
 33  Net Official Development Assist. received (% of GNI)      66 non-null     int64
 34  Quality Of Life Index                                     66 non-null     float64
 35  Purchasing Power Index                                    66 non-null     float64
 36  Safety Index                                              66 non-null     float64
 37  Health Care Index                                         66 non-null     float64
 38  Cost of Living                                            66 non-null     float64
 39  Property price to income ratio                            66 non-null     float64
 40  Traffic commute time index                                66 non-null     float64
 41  Pollution index                                           66 non-null     float64
 42  Climate index                                             66 non-null     float64
 43  Gross Rental Yield City Center                            66 non-null     float64
 44  Gross Rental Yield Outside Center                         66 non-null     float64
 45  Property Price to Rent Ratio City Center                  66 non-null     float64
```

```
70  Inflation, consumer prices (annual %)          64 non-null     float64
71  Life expectancy at birth, female (years)       66 non-null     float64
72  Life expectancy at birth, male (years)         66 non-null     float64
73  Life expectancy at birth, total (years)        66 non-null     float64
74  Military expenditure (% of GDP)                63 non-null     float64
75  Population, female                             66 non-null     float64
76  Population, male                               66 non-null     float64
77  Tax revenue (% of GDP)                         58 non-null     float64
```

Data Visualisations

```
!pip install pillow
!pip install wordcloud
```

```
Requirement already satisfied: pillow in /usr/local/lib/python3.10/dist-packages (9.4.0)
Requirement already satisfied: wordcloud in /usr/local/lib/python3.10/dist-packages (1.9.3)
Requirement already satisfied: numpy>=1.6.1 in /usr/local/lib/python3.10/dist-packages (from wordcloud) (1.26.4)
Requirement already satisfied: pillow in /usr/local/lib/python3.10/dist-packages (from wordcloud) (9.4.0)
Requirement already satisfied: matplotlib in /usr/local/lib/python3.10/dist-packages (from wordcloud) (3.7.1)
Requirement already satisfied: contourpy>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (1.3.0)
Requirement already satisfied: cycler>=0.10 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (0.12.1)
Requirement already satisfied: fonttools>=4.22.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (4.53.1)
Requirement already satisfied: kiwisolver>=1.0.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (1.4.7)
Requirement already satisfied: packaging>=20.0 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (24.1)
Requirement already satisfied: pyparsing>=2.3.1 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (3.1.4)
Requirement already satisfied: python-dateutil>=2.7 in /usr/local/lib/python3.10/dist-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.10/dist-packages (from python-dateutil>=2.7->matplotlib->wordcloud
```

Word Cloud

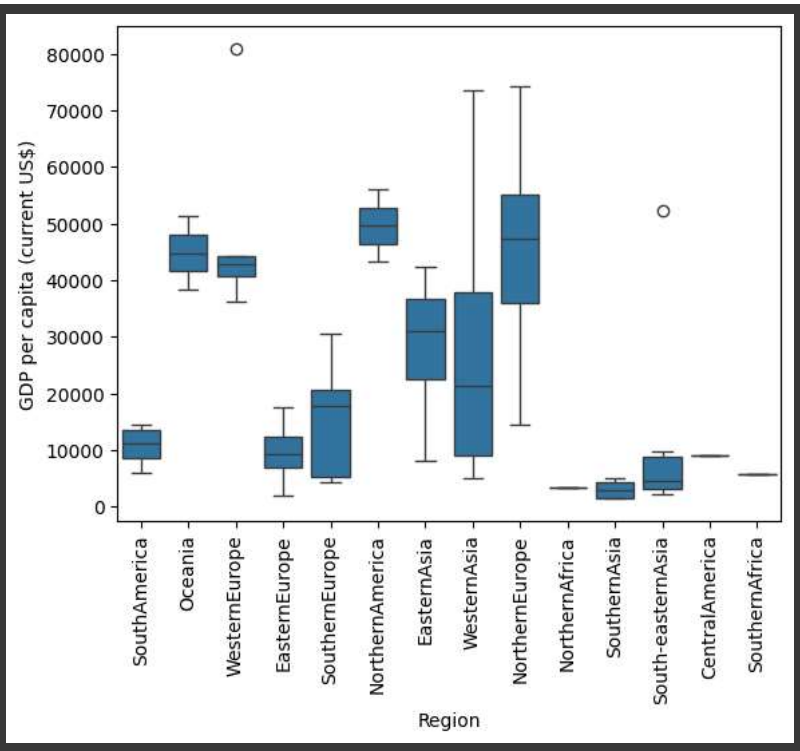
```
from PIL import Image
from wordcloud import WordCloud
Regions = " ".join(df['Region'].astype(str))

wordcloud = WordCloud(width=800, height=400, background_color='white').generate(Regions)

plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')  # No axis to show
plt.show()
```

Question answered : Which regions in the world are featured most in this dataset?

The word cloud has the most frequent values in large size and the less frequent ones with tiny font

Box and Whisker Plot

```
sns.boxplot(x='Region', y='GDP per capita (current US$)', data=df)
plt.xticks(rotation=90)
plt.show()
```

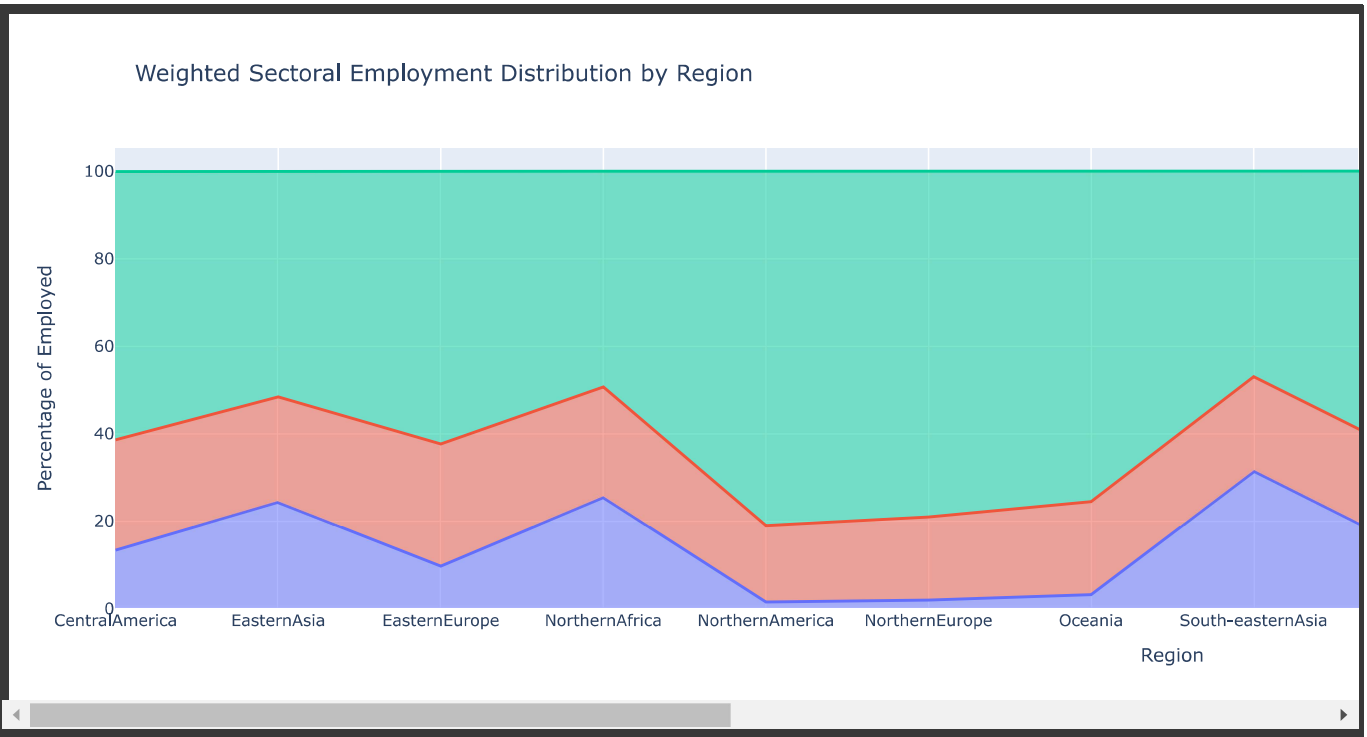


Question answered : How does the GDP epr capita vary across different regions?

A box and whisker plot shows us the spread of values for a numeric data. The box showcases the middle 50% of the data. The whiskers showcase the

Area Chart

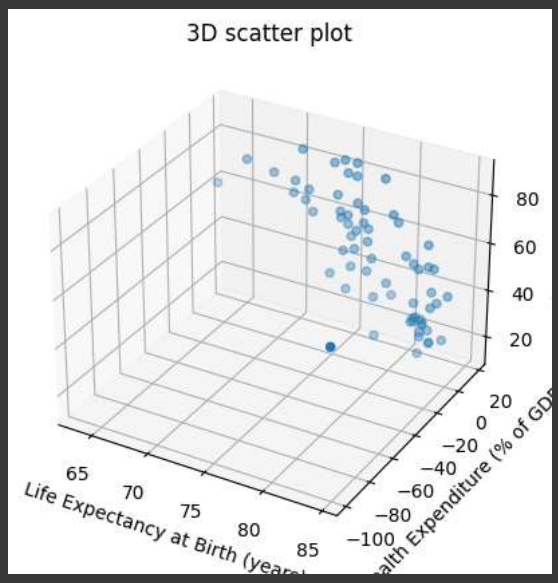
```
import pandas as pd
import plotly.express as px

df['Population (2017)'] = df['Population in thousands (2017)'] * 1000

# Aggregate total population by region
region_population = df.groupby('Region')['Population (2017)'].sum().reset_index()

# Merge with the original DataFrame to get region-wise data
df = df.merge(region_population, on='Region', suffixes=('', '_total'))

# Calculate weighted averages for employment data
df['Weighted_Agriculture'] = df['Employment: Agriculture (% of employed)'] * df['Population (2017)'] / df['Population (2017)_total']
df['Weighted_Industry'] = df['Employment: Industry (% of employed)'] * df['Population (2017)'] / df['Population (2017)_total']
df['Weighted_Services'] = df['Employment: Services (% of employed)'] * df['Population (2017)'] / df['Population (2017)_total']

# Aggregate the weighted data by region
region_weighted_employment = df.groupby('Region')[['Weighted_Agriculture', 'Weighted_Industry', 'Weighted_Services']].sum().reset_index(

# Step 2: Plot the aggregated data
fig = px.area(region_weighted_employment, x='Region',
              y=['Weighted_Agriculture', 'Weighted_Industry', 'Weighted_Services'],
              labels={'value': 'Percentage of Employed', 'variable': 'Sector'},
              title='Weighted Sectoral Employment Distribution by Region')

# Show the chart
fig.show()
```



Question answered : Is there a regional variation in the percentage employyes by the three sectors ?

3D Chart

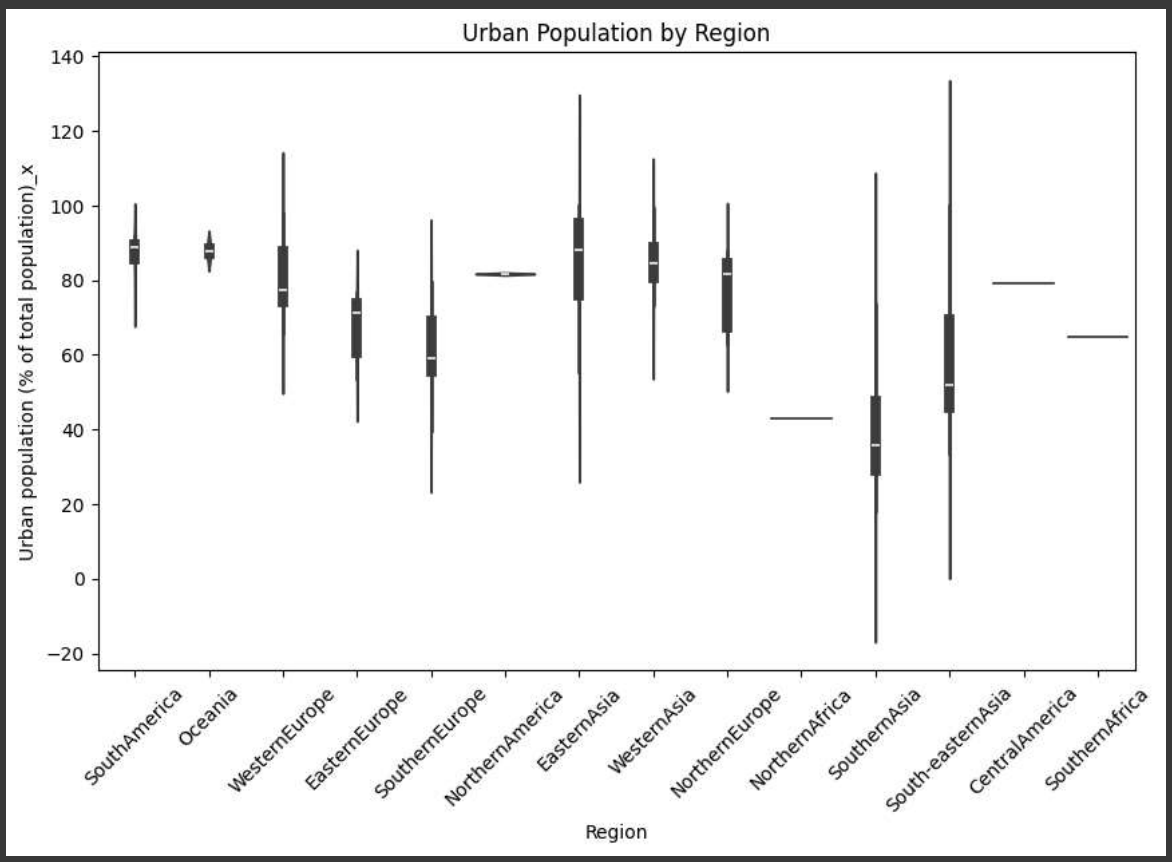
```
ax = plt.axes(projection ="3d")

ax.scatter(df['Life expectancy at birth, total (years)'],
               df['Health: Total expenditure (% of GDP)'],
               df['Pollution index'])

plt.title("3D scatter plot")

ax.set_xlabel('Life Expectancy at Birth (years)')
ax.set_ylabel('Health Expenditure (% of GDP)')
ax.set_zlabel('Pollution Index')
# show plot
plt.show()
```



Violin Plot

```
plt.figure(figsize=(10, 6))
sns.violinplot(x='Region', y='Urban population (% of total population)_x', data=df)
plt.xticks(rotation=45)
```

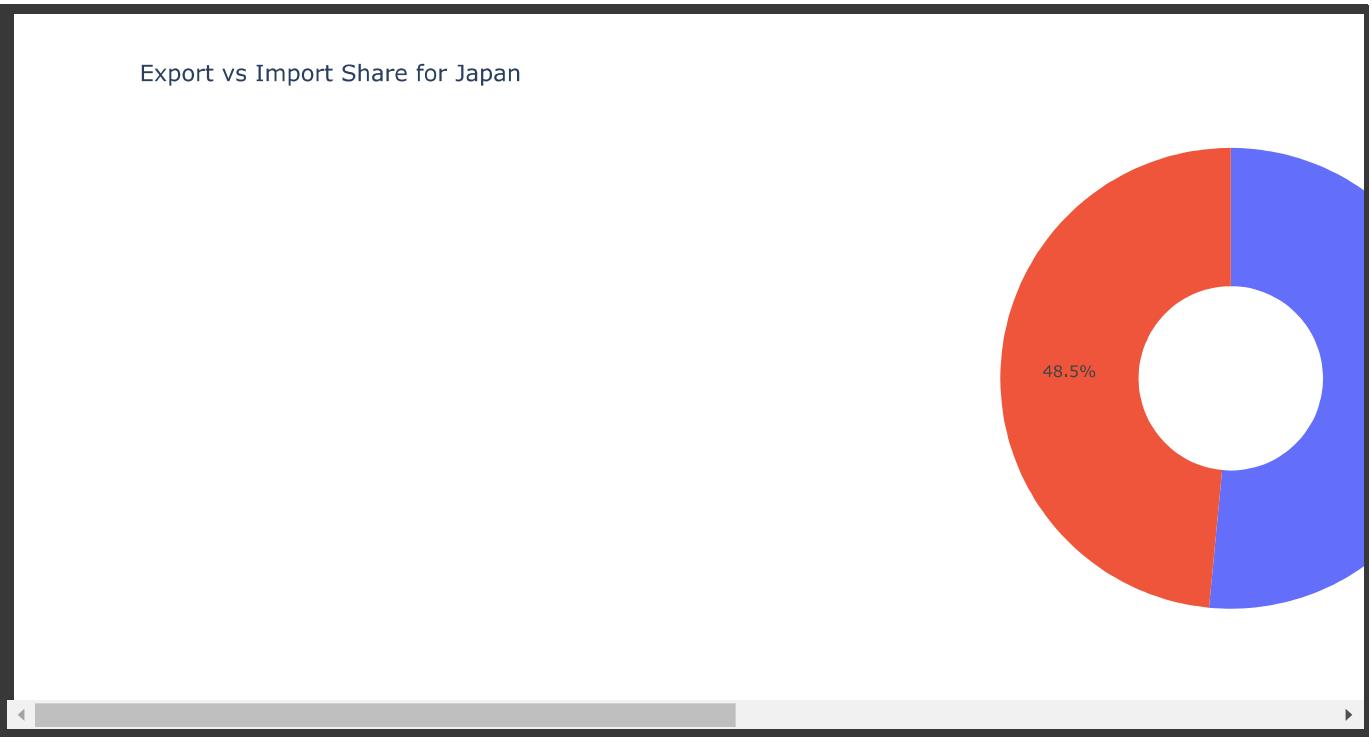
```
plt.title('Urban Population by Region')
plt.show()
```



Double-click (or enter) to edit

Donut Chart

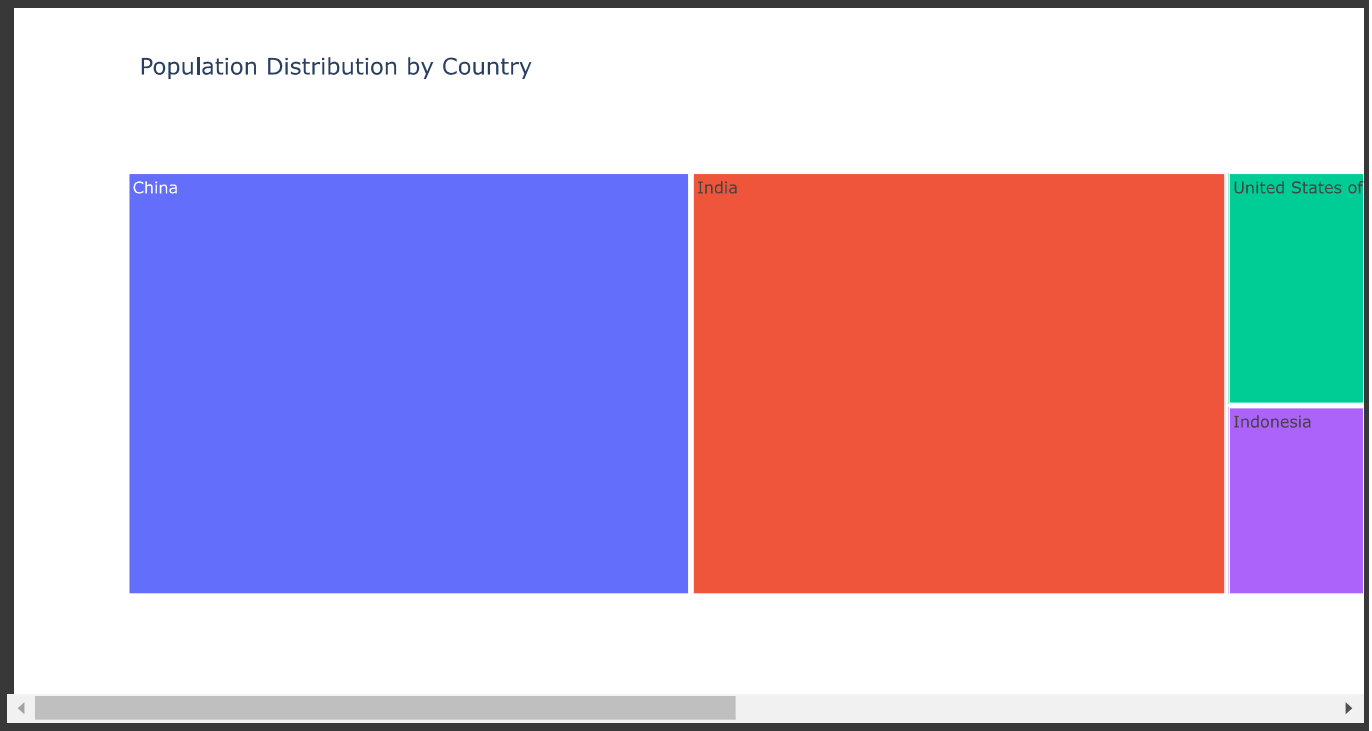
```
import plotly.express as px

country_data = df[df['country'] == 'Japan']

# Aggregate the total exports and imports
total_exports = country_data['International trade: Exports (million US$)'].sum()
total_imports = country_data['International trade: Imports (million US$)'].sum()

# Prepare data for the pie chart
values = [total_exports, total_imports]
labels = ['Exports', 'Imports']

# Create the pie chart
fig = px.pie(values=values, names=labels, title="Export vs Import Share for Japan", hole=0.4)

# Show the chart
fig.show()
```



Question Answered : How does the trade baalnce for (any particular country) look?

TreeMap

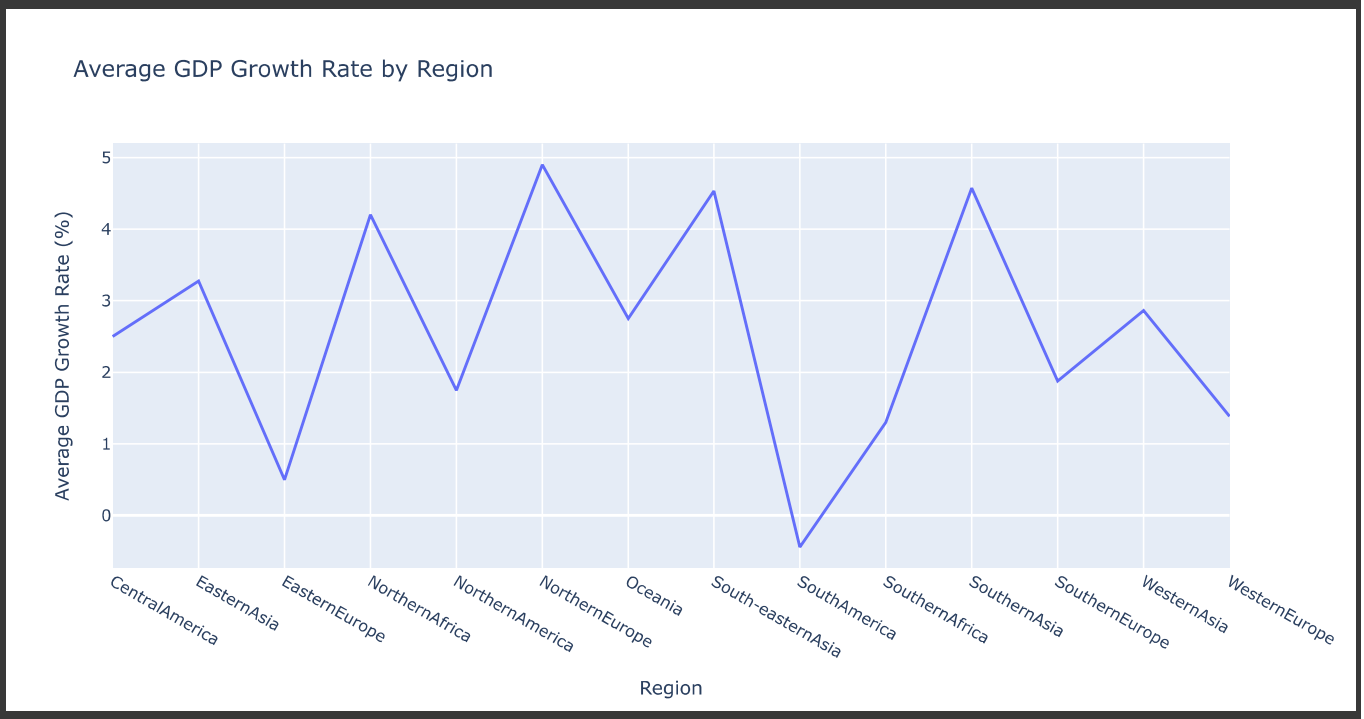
```
import plotly.express as px

# Create a treemap of population by country
fig = px.treemap(df,
                 path=['country'],
                 values='Population in thousands (2017)',
                 title="Population Distribution by Country")

# Show the chart
fig.show()
```



How do the populations of various countries compare with each other?

Line Chart

```
df['GDP growth rate (annual %, const. 2005 prices)'] = pd.to_numeric(df['GDP growth rate (annual %, const. 2005 prices)'], errors='coerc

df_grouped = df.groupby('Region', as_index=False)['GDP growth rate (annual %, const. 2005 prices)'].mean()

fig = px.line(df_grouped,
              x='Region',
              y='GDP growth rate (annual %, const. 2005 prices)',
              title="Average GDP Growth Rate by Region",
              labels={'GDP growth rate (annual %, const. 2005 prices)': 'Average GDP Growth Rate (%)',
                      'Region': 'Region'})

fig.show()
```



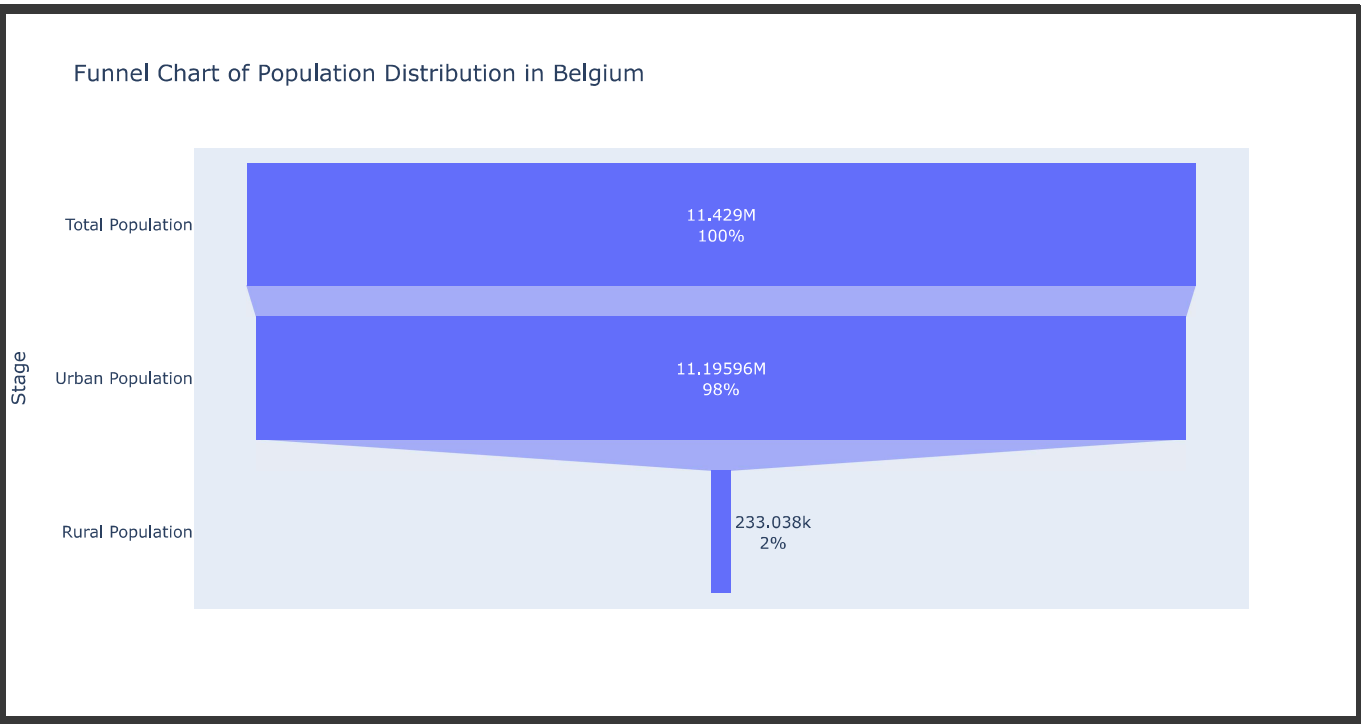
Questions answered :

Which regions show the highest GDP growth rate?

Which regions show the least GDP growth rate?

Funnel Chart

```
import plotly.graph_objects as go
country_data = df[df['country'] == 'Belgium']

population_in_thousands = country_data['Population in thousands (2017)'].iloc[0]
total_pop = int(population_in_thousands) * 1000

urban_population = country_data['Urban population (% of total population)_y'].iloc[0]
urban_population_number = int(total_pop * (urban_population / 100))

rural_population_number = total_pop - urban_population_number

# Create a funnel chart using Plotly
fig = go.Figure()

fig.add_trace(go.Funnel(
    y=["Total Population", "Urban Population", "Rural Population"],
    x=[total_pop, urban_population_number, rural_population_number],
    textinfo="value+percent initial",
))

fig.update_layout(
    title="Funnel Chart of Population Distribution in Belgium",
    xaxis_title="Population",
    yaxis_title="Stage"
)

fig.show()
```



How urbanised is a particular country?

Jitter Plot

```
plt.figure(figsize=(12, 6))
sns.stripplot(x='Region',
              y='Pollution index',
              data=df,
              jitter=True,
              palette='viridis')

plt.xlabel('Region')
plt.ylabel('Pollution Index')
plt.title('Jitter Plot of Pollution Index by Region')
plt.xticks(rotation=45)

plt.show()
```

```
<ipython-input-17-dde03821a72f>:2: FutureWarning:
```