

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
import seaborn as sns
import warnings
warnings.filterwarnings('ignore')
```

```
import plotly.express as px
import plotly.graph_objects as go
```

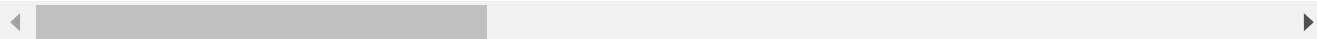
```
df = pd.read_csv('/content/soci_econ_country_profiles.csv')
```

```
df.head()
```



	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\$)
0	0	Argentina	SouthAmerica	2780400	44271	16.2	95.9	632343	2.4	14564.5
1	1	Australia	Oceania	7692060	24451	3.2	99.3	1230859	2.4	51352.2
2	2	Austria	WesternEurope	83871	8736	106.0	96.2	376967	1.0	44117.7
3	3	Belarus	EasternEurope	207600	9468	46.7	87.0	54609	-3.9	5750.8
4	4	Belgium	WesternEurope	30528	11429	377.5	97.3	455107	1.5	40277.8

5 rows × 96 columns



```
df.columns.tolist()
```



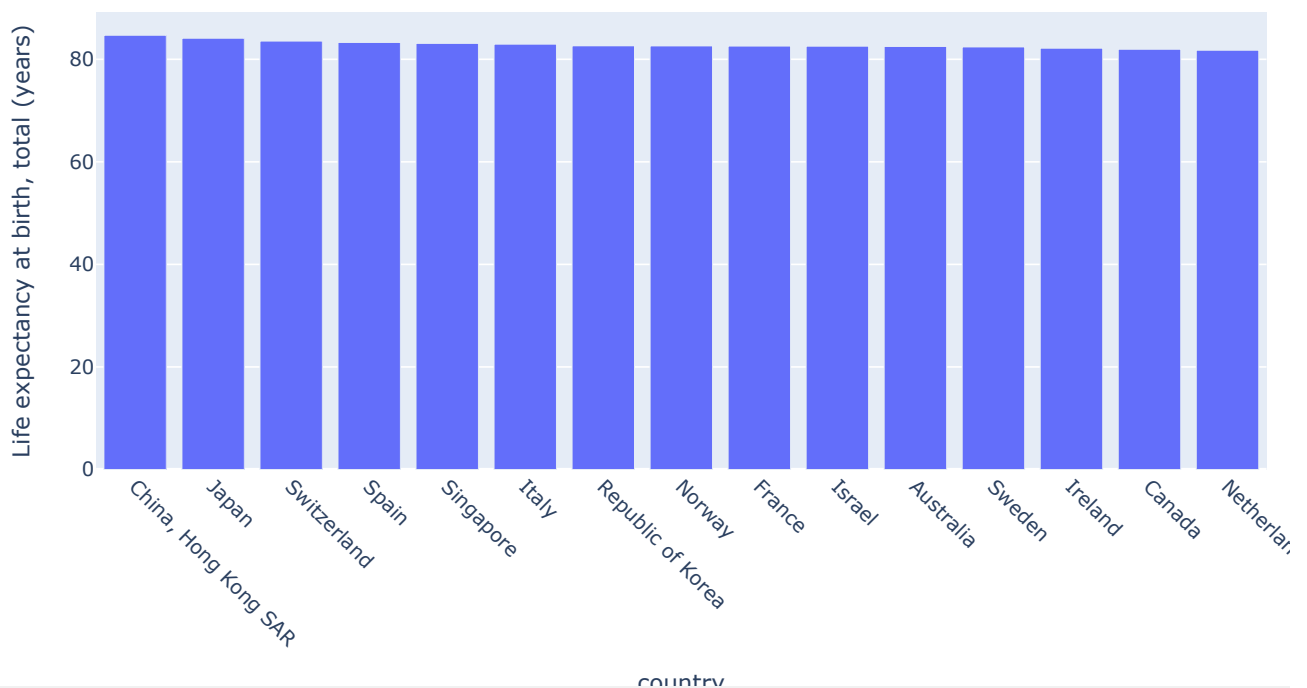
```
['Unnamed: 0',
'country',
'Region',
'Surface area (km2)',
'Population in thousands (2017)',
'Population density (per km2, 2017)',
'Sex ratio (m per 100 f, 2017)',
'GDP: Gross domestic product (million current US$)',
'GDP growth rate (annual %, const. 2005 prices)',
'GDP per capita (current US$)',
'Economy: Agriculture (% of GVA)',
'Economy: Industry (% of GVA)',
'Economy: Services and other activity (% of GVA)',
'Employment: Agriculture (% of employed)',
'Employment: Industry (% of employed)',
'Employment: Services (% of employed)',
'Unemployment (% of labour force)',
'Labour force participation (female/male pop. %)',
'Agricultural production index (2004-2006=100)',
'Food production index (2004-2006=100)',
'International trade: Exports (million US$)',
'International trade: Imports (million US$)',
'International trade: Balance (million US$)',
'Balance of payments, current account (million US$)',
'Population growth rate (average annual %)',
'Urban population (% of total population)_x',
'Urban population growth rate (average annual %)',
'Fertility rate, total (live births per woman)']
```

```
'Population age distribution (0-14 / 60+ years, %)',
'International migrant stock (000/% of total pop.)',
'Refugees and others of concern to UNHCR (in thousands)',
'Infant mortality rate (per 1000 live births',
'Health: Total expenditure (% of GDP)',
'Health: Physicians (per 1000 pop.)',
'Education: Government expenditure (% of GDP)',
'Education: Primary gross enrol. ratio (f/m per 100 pop.)',
'Education: Secondary gross enrol. ratio (f/m per 100 pop.)',
'Education: Tertiary gross enrol. ratio (f/m per 100 pop.)',
'Seats held by women in national parliaments %',
'Mobile-cellular subscriptions (per 100 inhabitants)',
'Mobile-cellular subscriptions (per 100 inhabitants).1',
'Individuals using the Internet (per 100 inhabitants)',
'Threatened species (number)',
'Forested area (% of land area)',
'CO2 emission estimates (million tons/tons per capita)',
'Energy production, primary (Petajoules)',
'Energy supply per capita (Gigajoules)',
'Pop. using improved drinking water (urban/rural, %)',
'Pop. using improved sanitation facilities (urban/rural, %)',
'Net Official Development Assist. received (% of GNI)',
'Quality Of Life Index',
'Purchasing Power Index',
'Safety Index',
'Health Care Index',
'Cost of Living',
'Property price to income ratio',
'Traffic commute time index',
'Pollution index'.
```

```
top_life_expectancy = df.sort_values('Life expectancy at birth, total (years)', ascending=False).head(15)
fig = px.bar(top_life_expectancy, x='country', y='Life expectancy at birth, total (years)',
             title='Top 15 Countries with Highest Life Expectancy')
fig.update_xaxes(tickangle=45)
fig.show()
```



Top 15 Countries with Highest Life Expectancy



The bar chart illustrates the top 15 countries with the highest life expectancy at birth, with all countries exhibiting a life expectancy slightly above 80 years. The countries represented span across different regions, including Asia, Europe,

and North America, indicating that high life expectancy is a global phenomenon rather than region-specific.

Countries like China (Hong Kong SAR) and Japan, both in Asia, are known for their advanced healthcare systems and healthy lifestyles, contributing to their high life expectancy. Similarly, European countries like Switzerland, Spain, and Italy, along with others such as France and Norway, reflect strong healthcare systems, social welfare, and high standards of living.

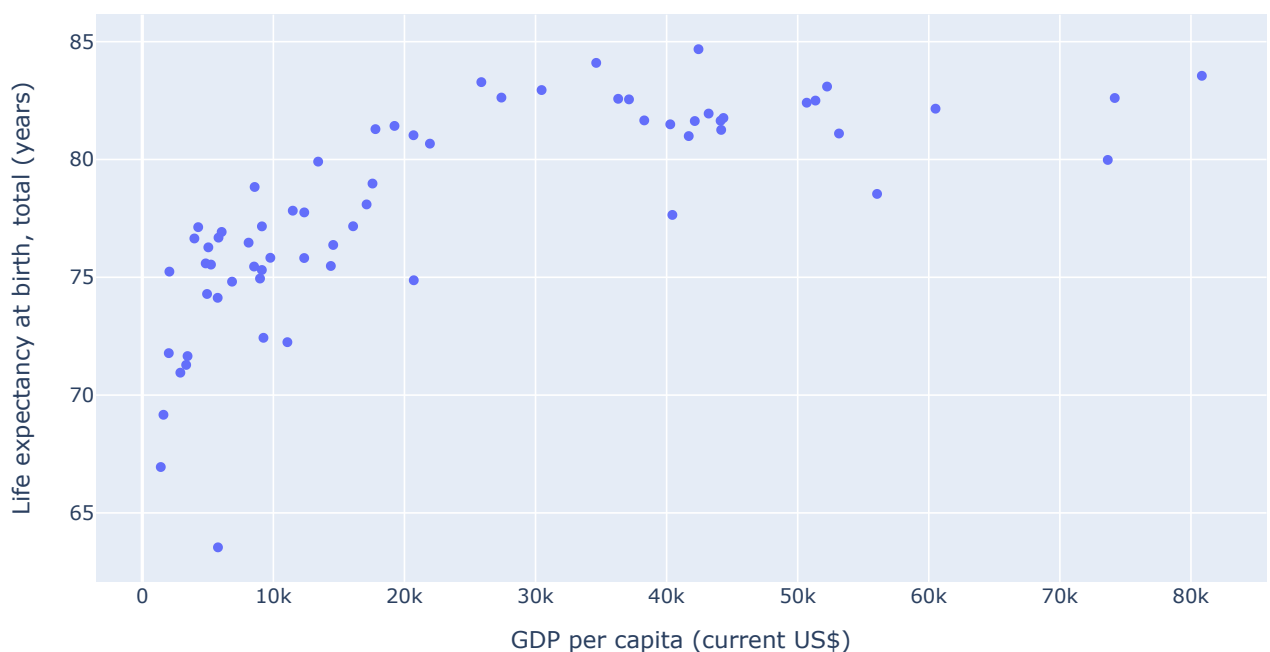
Singapore and the Republic of Korea represent Southeast Asia and East Asia, respectively, showing the impact of rapid economic development and improved healthcare. Israel and Australia highlight the importance of robust healthcare infrastructure in maintaining high life expectancy. Canada, a representative of North America, also ranks high, reflecting its well-developed healthcare system.

Overall, the chart underscores the close clustering of life expectancy figures among these top 15 nations, implying that once a certain threshold of healthcare, nutrition, and social services is reached, life expectancy levels off, with minimal variation among the leading countries.

```
fig = px.scatter(df, x='GDP per capita (current US$)', y='Life expectancy at birth, total (years)',
                hover_name='country', title='GDP per capita vs Life Expectancy')
fig.show()
```



GDP per capita vs Life Expectancy



The scatter plot illustrates the relationship between GDP per capita (in current US dollars) and life expectancy at birth (in years). The data points suggest a positive correlation between GDP per capita and life expectancy, meaning that, generally, as GDP per capita increases, life expectancy also tends to rise.

Key observations include:

- Initial Increase:** For countries with lower GDP per capita (below \$20,000), there is a noticeable increase in life expectancy as GDP per capita increases. Most countries in this range have life expectancies between 70 to 80 years.
- Plateau Effect:** Beyond a certain threshold of GDP per capita (around 30,000 to 40,000), the increase in life expectancy starts to plateau. Countries with very high GDP per capita (above \$40,000) tend to have life expectancies clustered around 80 to 85 years, with less variation.

3. **Outliers:** There are a few outliers where countries with lower GDP per capita have a significantly lower life expectancy (below 70 years). Conversely, there are also countries with high GDP per capita but with a slightly lower than expected life expectancy (around 75 years).
4. **General Trend:** The scatter plot indicates that while higher GDP per capita is generally associated with higher life expectancy, other factors beyond GDP also influence life expectancy, particularly at higher income levels. This suggests that after reaching a certain economic level, other determinants such as healthcare quality, lifestyle, and social factors play a more significant role in determining life expectancy.

```
economic_indicators = ['GDP per capita (current US$)', 'Inflation, consumer prices (annual %)',
                       'Unemployment (% of labour force)', 'Exports of goods and services (% of GDP)']
correlation = df[economic_indicators].corr()
```

```
fig = go.Figure(data=go.Heatmap(
    z=correlation.values,
    x=correlation.columns,
    y=correlation.index,
    colorscale='RdBu'))
fig.update_layout(title='Correlation between Economic Indicators')
fig.show()
```



Correlation between Economic Indicators



1. GDP per capita (current US\$) shows:
 - positive correlation with exports of goods and services
 - negative correlation with inflation
 - Moderate negative correlation with unemployment
2. Inflation (consumer prices, annual %) shows:
 - positive correlation with unemployment
 - Weak negative correlation with exports
3. Unemployment (% of labor force) shows:

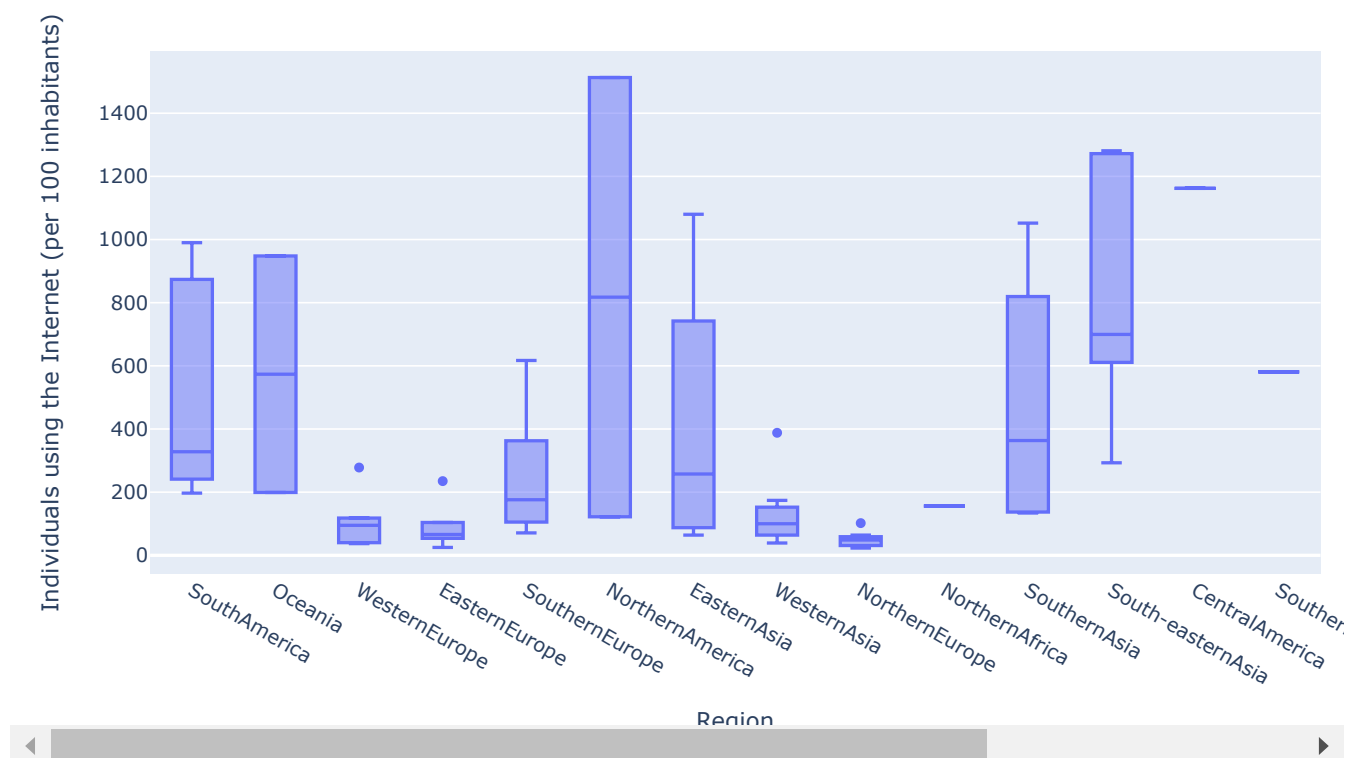
- negative correlation with exports

- The strongest positive correlation appears to be between GDP per capita and exports.
- The strongest negative correlation is between inflation and GDP.
- Inflation and unemployment show a positive correlation, which aligns with economic theories about the relationship between these factors.

```
fig = px.box(df, x='Region', y='Individuals using the Internet (per 100 inhabitants)',
            title='Distribution of Internet Usage across Regions')
fig.show()
```

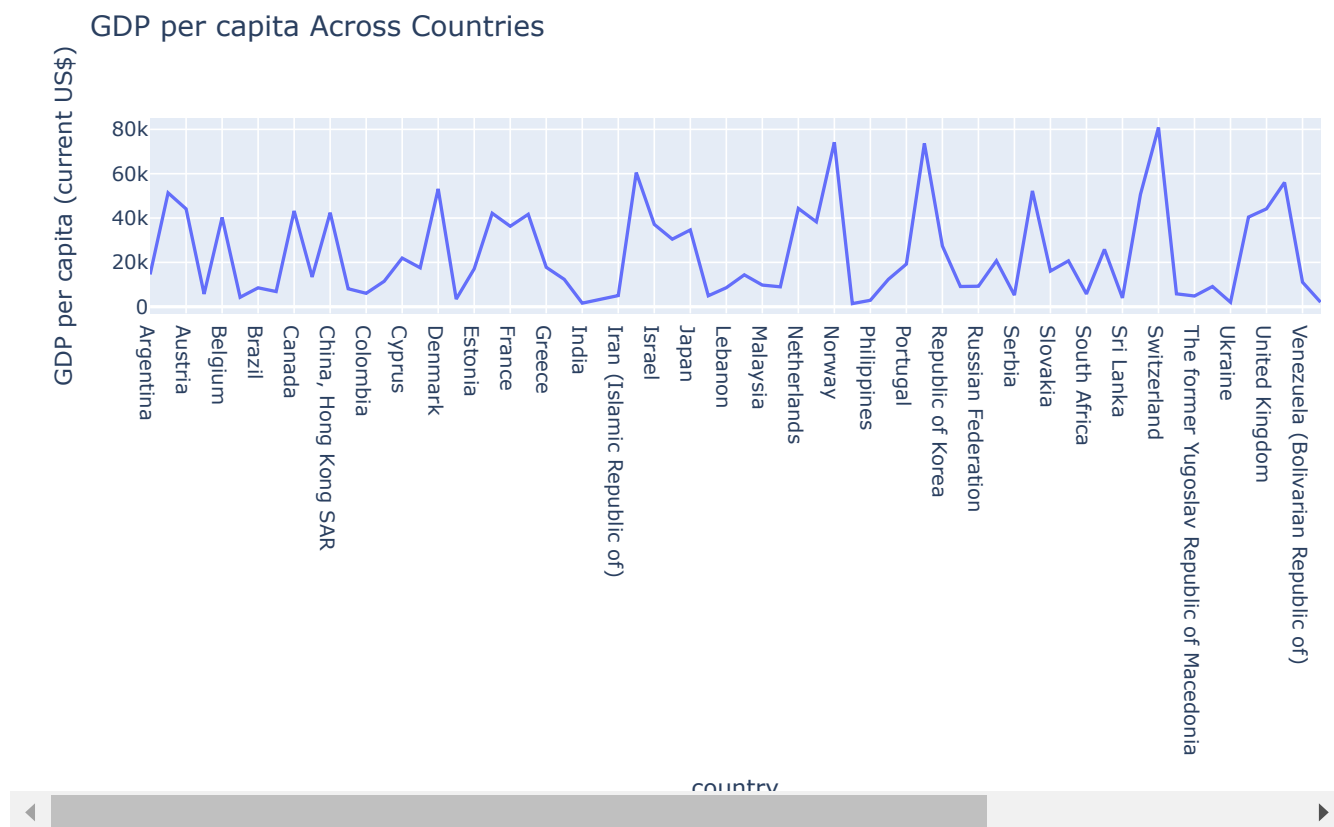


Distribution of Internet Usage across Regions



- Northern America has the highest median Internet usage and the largest range, indicating high but varied usage across the region.
- South-eastern Asia and Eastern Asia have wide ranges, indicating significant disparities in Internet access within these regions.
- Southern Asia has a low median but a wide range, suggesting uneven Internet penetration across the region.
- Africa (Northern and Southern) generally shows lower Internet usage compared to other regions.
- Central America and South America have moderate usage levels with some variability.
- There are some outliers, particularly in Western Asia and Northern Europe, represented by dots above the whiskers.
- The overall trend suggests that more economically developed regions tend to have higher Internet usage rates.
- There's significant variation both between and within regions, highlighting global digital divides.
- Oceania shows a particularly wide box, indicating a large spread in the middle 50% of its data.

```
# Line Chart of GDP per capita across different countries
fig = px.line(df, x='country', y='GDP per capita (current US$)', title='GDP per capita Across Countries')
fig.show()
```

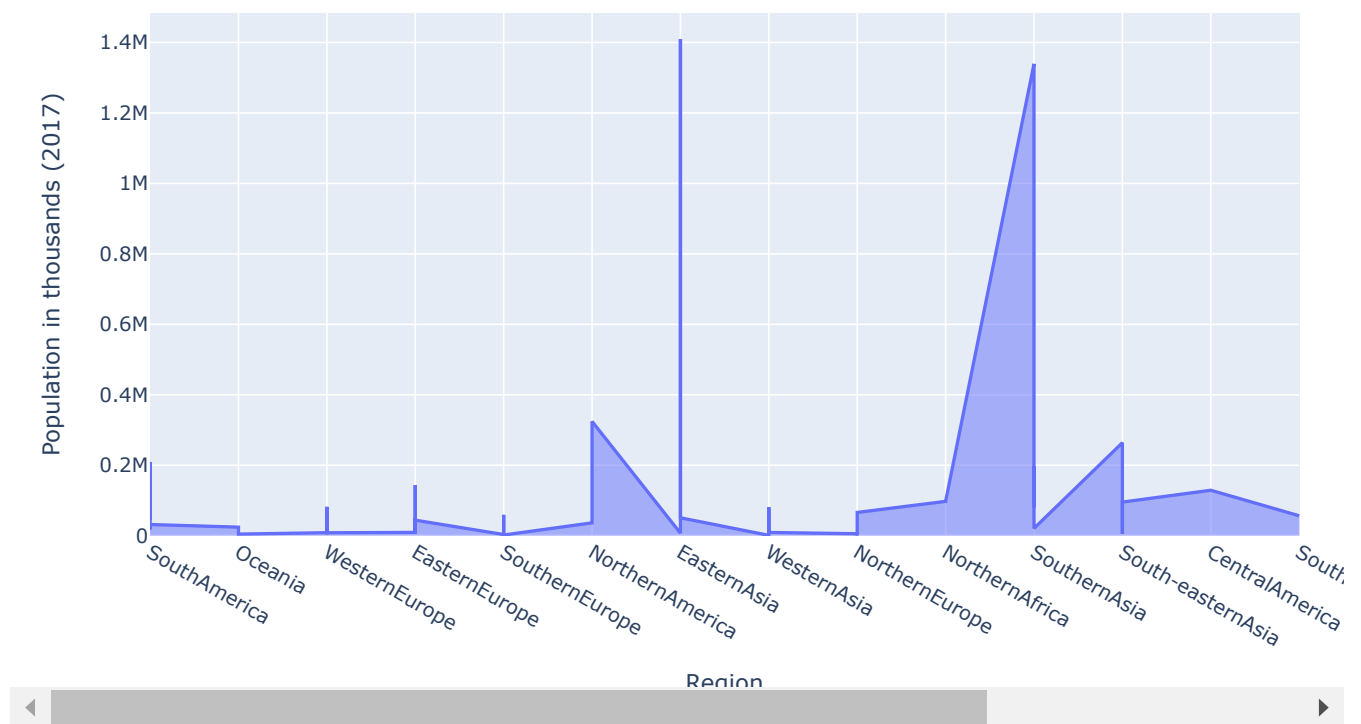


1. Switzerland has highest GDP per capita of about 80,000 US dollars.
2. Pakistan has lowest GDP of about 1000 US dollars per capita.
3. Many countries cluster in the lower range, below 20,000 US dollars per capita.
4. There are sharp spikes in the graph, indicating some countries have substantially higher GDP per capita than their neighbors in the graph's ordering.
5. The overall pattern is irregular, with frequent ups and downs, suggesting large economic disparities between countries.
6. Some regions of the graph show clusters of countries with similar GDP levels, possibly indicating regional economic similarities.
7. There appears to be a greater concentration of countries in the lower GDP per capita range, with fewer countries reaching the highest levels.
8. The graph effectively illustrates the economic divide between developed and developing/underdeveloped countries.

```
# Area Chart of Population in thousands (2017) across different regions
fig = px.area(df, x='Region', y='Population in thousands (2017)', title='Population Distribution Across Regions')
fig.show()
```



Population Distribution Across Regions

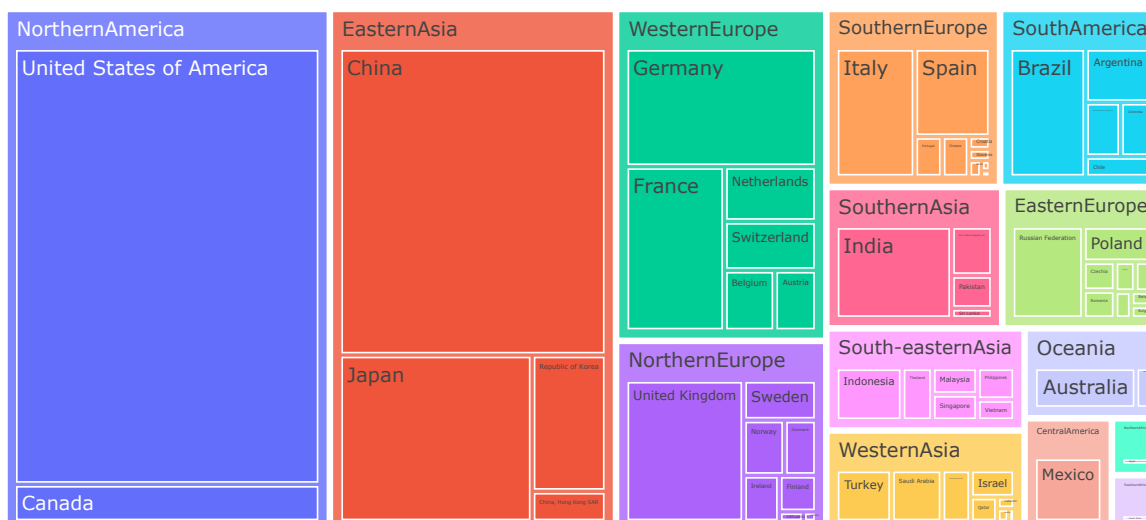


1. Eastern Asia has the highest population, with a sharp peak reaching about 1.4 million.
2. Southern Asia is the second most populous region, with over 1.2 million people.
3. South-eastern Asia is the third most populous region.
4. Northern America has a moderate population, forming a notable peak but much lower than Asia's regions.
5. Europe is divided into several regions (Western, Eastern, Southern, Northern), each with relatively low populations compared to Asia.
6. Africa is also divided (Northern, Southern), with both regions showing low populations relative to Asia.
7. Oceania has the lowest population among all regions shown.
8. There's a general trend of Asian regions having higher populations compared to other continents.
9. The Americas (excluding Northern America) have relatively low populations.
10. There's significant variation in population sizes across regions, from very low (Oceania) to very high (Southern Asia).

```
# Treemap of GDP across different regions and countries
fig = px.treemap(df, path=['Region', 'country'], values='GDP: Gross domestic product (million current US$)',
                 title='GDP Distribution Across Regions and Countries')
fig.show()
```



GDP Distribution Across Regions and Countries



- United States Dominance:** The United States has the largest block in the treemap, indicating that it has the highest GDP among all the regions and countries shown. The entire region of Northern America is predominantly represented by the U.S.
- China's GDP:** In Eastern Asia, China has the largest share, highlighting its significant contribution to the global GDP. China's block is the second largest after the United States.
- Western Europe:** Germany, France, and the United Kingdom are major contributors to GDP in Europe. Germany leads Western Europe, while the United Kingdom is the largest contributor in Northern Europe.
- Japan and South Korea:** In Eastern Asia, Japan has a significant GDP, followed by South Korea, though both are smaller compared to China.
- Brazil in South America:** Brazil is the leading country in South America in terms of GDP, with a large block relative to other countries on the continent.
- Southern Europe:** Italy and Spain are the major contributors in Southern Europe, with Italy having a slightly larger block than Spain.
- Other Notable Countries:**
 - Russia in Eastern Europe has a considerable GDP block.
 - India in Southern Asia also has a significant share, dominating its region.
 - Australia in Oceania has the largest GDP in that region.
 - Mexico stands out in Central America.
- Smaller Economies:** Several smaller blocks represent the GDPs of countries like Qatar, Israel, Norway, and Finland, among others, indicating their relatively smaller contributions to the global GDP.

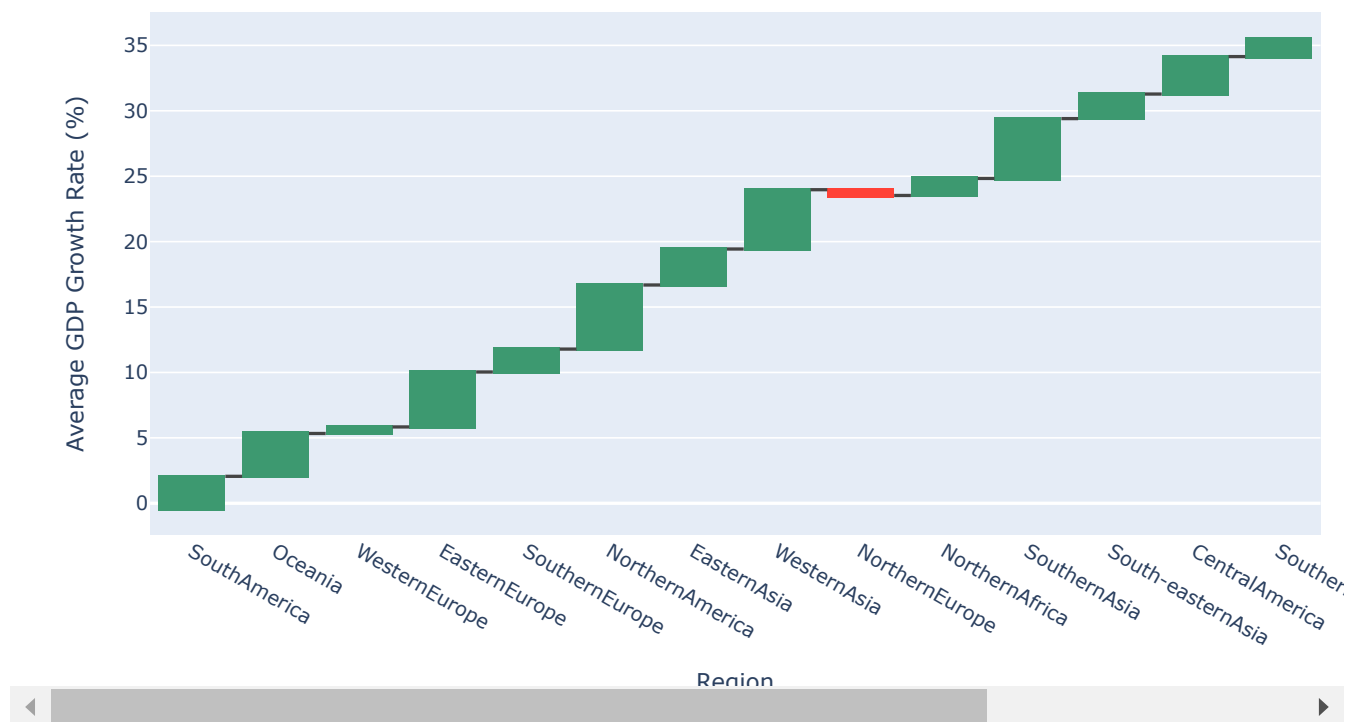

```
# Waterfall Chart showing GDP growth rate (annual %) for various regions
regions = df['Region'].unique()
gdp_growth = df.groupby('Region')['GDP growth rate (annual %, const. 2005 prices)'].mean().fillna(0)

fig = go.Figure(go.Waterfall(
    x=regions,
    y=gdp_growth,
    base=gdp_growth.min(),
    orientation="v"
))

fig.update_layout(title='Average GDP Growth Rate by Region', xaxis_title='Region', yaxis_title='Average GDP G
fig.show()
```



Average GDP Growth Rate by Region



1. **Southern Africa:** The region with the highest average GDP growth rate, exceeding 35%. This suggests that Southern Africa has experienced significant economic growth compared to other regions.
2. **Central America and South-eastern Asia:** These regions also show relatively high GDP growth rates, around 25-30%, indicating robust economic performance.
3. **Southern Asia:** The average GDP growth rate for Southern Asia is slightly above 20%, placing it among the higher-growth regions.
4. **Northern Africa and Northern Europe:** Both regions have an average GDP growth rate around 15-20%, showing moderate economic growth.
5. **Western Asia and Eastern Asia:** These regions have GDP growth rates in the range of 10-15%. The growth rate for Eastern Asia is notable given the size and economic significance of the region.
6. **Northern America:** The average GDP growth rate for Northern America is close to 10%, reflecting stable but lower growth compared to the rapidly growing regions.
7. **Southern Europe, Eastern Europe, and Western Europe:** These European regions show lower average GDP growth rates, generally under 10%, indicating slower economic growth.

8. **Oceania:** The GDP growth rate for Oceania is slightly above 5%, which is lower than many other regions.
9. **South America:** This region has the lowest average GDP growth rate, just above 0%, indicating minimal economic growth.
10. **Red Marker (Potential Outlier):** There is a red marker at around 15% in Northern Europe, which could indicate a specific data point that is an outlier or has been highlighted for some reason.

Funnel Chart of Sex Ratio across countries

```
fig = px.funnel(df, x='Sex ratio (m per 100 f, 2017)', y='country', title='Sex Ratio Across Countries')
fig.show()
```



Sex Ratio Across Countries



Key Observations:

1. **Symmetry Around the Center:** Most of the bars are centered around a specific value, suggesting that the sex ratios in these countries are close to an equal number of males and females. This would typically be around a ratio of 1 (or 100 if scaled per 100 females).
2. **Countries with Higher Ratios:** Some bars extend significantly further to the right. This indicates a higher number of males compared to females in those countries.
3. **Countries with Lower Ratios:** If any bars extend further to the left, it would indicate a higher number of females compared to males, but in this graph, it appears that most countries have a sex ratio around the central point.
4. **Notable Outliers:** A few countries have bars that are significantly longer, indicating that their sex ratios are far from the average. These countries might have a societal or cultural context leading to a skewed sex ratio.

Specific Anomalies:

- **Pakistan and India:** Both countries show a significantly higher ratio (with a longer bar to the right), indicating more males than females.

- **Portugal and the United States:** The bars for these countries are quite short, indicating sex ratios close to the average.

General Trend:

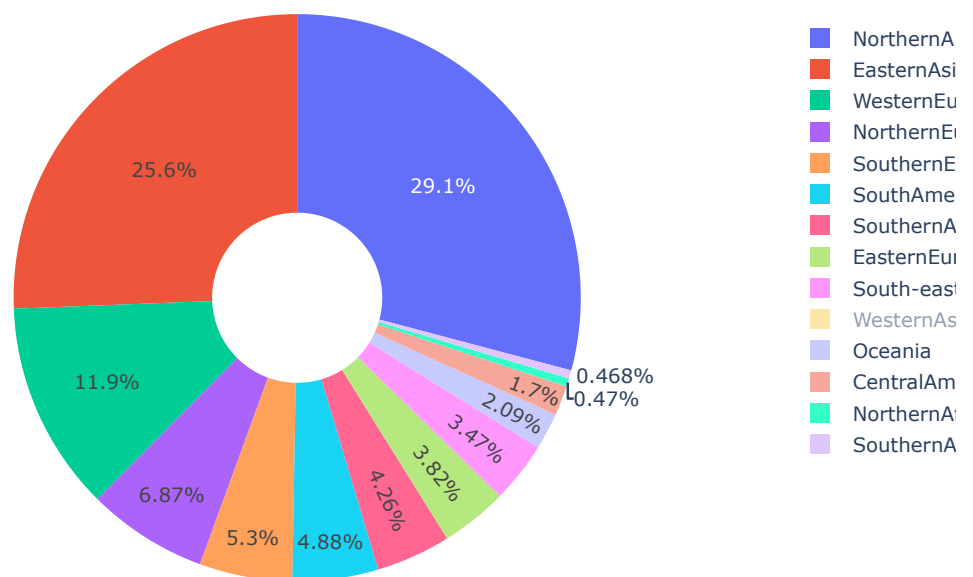
- The general trend across most countries is a sex ratio near the central value, suggesting a balanced population in terms of sex.

Donut Chart of GDP by Region

```
fig = px.pie(df, names='Region', values='GDP: Gross domestic product (million current US$)', hole=0.3, title=fig.show())
```



GDP Distribution by Region



Key Observations:

1. Dominant Regions:

- **Northern America (29.1%):** The largest segment in the pie chart, Northern America (likely including the United States and Canada), contributes the most significant portion to the global GDP.
- **Eastern Asia (25.6%):** This region, likely including economic powerhouses like China, Japan, and South Korea, also contributes a substantial share to the global GDP.

2. Moderately Contributing Regions:

- **Western Europe (11.9%):** Western Europe is another significant contributor, representing nearly 12% of the global GDP.
- **Northern Europe (6.87%):** This region contributes a smaller but still notable share to the global GDP.
- **Southern Europe (5.3%) and South America (4.88%):** Both regions make moderate contributions to the global GDP, with percentages around 5%.

3. Smaller Contributors:

- Regions like **Southern Asia (4.88%)**, **Eastern Europe (3.82%)**, **South-eastern Asia (3.47%)**, and others have smaller slices of the GDP pie, indicating a lower contribution to the global economy relative to the dominant regions.

4. Minor Contributors:

- **Western Asia (1.7%)**, **Oceania (1.09%)**, **Central America (0.468%)**, **Northern Africa (0.47%)**, and **Southern Africa (0.47%)** contribute the smallest shares to the global GDP, each representing less than 2% of the total.

General Trend:

- The chart highlights the economic disparity between regions, with Northern America and Eastern Asia dominating the global GDP distribution. Western Europe also plays a significant role, while other regions