sma2

August 24, 2024

[1]:

```
import pandas as pd
```

[2]:

```
df = pd.read_csv(r"C:\Users\admin\Downloads\soci_econ_country_profiles.csv")
df.head()
```

[2]:

```
   Unnamed: 0    country         Region  Surface area (km2)  \
0           0  Argentina   SouthAmerica             2780400   
1           1  Australia        Oceania             7692060   
2           2    Austria  WesternEurope               83871   
3           3    Belarus  EasternEurope              207600   
4           4    Belgium  WesternEurope               30528   

   Population in thousands (2017)  Population density (per km2, 2017)  \
0                           44271                                16.2   
1                           24451                                 3.2   
2                            8736                               106.0   
3                            9468                                46.7   
4                           11429                               377.5   

   Sex ratio (m per 100 f, 2017)  \
0                           95.9   
1                           99.3   
2                           96.2   
3                           87.0   
4                           97.3   

   GDP: Gross domestic product (million current US$)  \
0                                             632343   
1                                            1230859   
2                                             376967   
3                                              54609   
4                                             455107   

   GDP growth rate (annual %, const. 2005 prices)  \
0                                             2.4   
1                                             2.4   
2                                             1.0   
```

```
3                                            -3.9
4                                             1.5

   GDP per capita (current US$)  … Inflation, consumer prices (annual %)  \
0                       14564.5  …                                   NaN
1                       51352.2  …                              1.948647
2                       44117.7  …                              2.081269
3                        5750.8  …                              6.031837
4                       40277.8  …                              2.125971

   Life expectancy at birth, female (years)  \
0                                    79.726
1                                    84.600
2                                    84.000
3                                    79.200
4                                    83.900

   Life expectancy at birth, male (years)  \
0                                  72.924
1                                  80.500
2                                  79.400
3                                  69.300
4                                  79.200

   Life expectancy at birth, total (years)  Military expenditure (% of GDP)  \
0                                76.372000                         0.856138
1                                82.500000                         2.007966
2                                81.643902                         0.756179
3                                74.129268                         1.162417
4                                81.492683                         0.910371

   Population, female  Population, male Tax revenue (% of GDP)  \
0          22572521.0        21472290.0              10.955501
1          12349632.0        12252228.0              21.915859
2           4478340.0         4319226.0              25.355237
3           5077542.0         4420722.0              13.019006
4           5766141.0         5609017.0              23.399721

   Taxes on income, profits and capital gains (% of revenue)  \
0                                          12.929913
1                                          64.110306
2                                          27.024073
3                                           2.933101
4                                          33.727746

   Urban population (% of total population)_y
0                                      91.749
```

```
1                                        85.904
2                                        58.094
3                                        78.134
4                                        97.961

[5 rows x 96 columns]
```

[3]:

```
pip install wordcloud
```

```
Requirement already satisfied: wordcloud in c:\users\admin\anaconda3\lib\site-
packages (1.9.3)
Requirement already satisfied: numpy>=1.6.1 in
c:\users\admin\anaconda3\lib\site-packages (from wordcloud) (1.26.4)
Requirement already satisfied: matplotlib in c:\users\admin\anaconda3\lib\site-
packages (from wordcloud) (3.5.2)
Requirement already satisfied: pillow in c:\users\admin\anaconda3\lib\site-
packages (from wordcloud) (9.2.0)
Requirement already satisfied: python-dateutil>=2.7 in
c:\users\admin\anaconda3\lib\site-packages (from matplotlib->wordcloud) (2.8.2)
Requirement already satisfied: pyparsing>=2.2.1 in
c:\users\admin\anaconda3\lib\site-packages (from matplotlib->wordcloud) (3.0.9)
Requirement already satisfied: fonttools>=4.22.0 in
c:\users\admin\anaconda3\lib\site-packages (from matplotlib->wordcloud) (4.25.0)
Requirement already satisfied: cycler>=0.10 in
c:\users\admin\anaconda3\lib\site-packages (from matplotlib->wordcloud) (0.11.0)
Requirement already satisfied: packaging>=20.0 in
c:\users\admin\anaconda3\lib\site-packages (from matplotlib->wordcloud) (21.3)
Requirement already satisfied: kiwisolver>=1.0.1 in
c:\users\admin\anaconda3\lib\site-packages (from matplotlib->wordcloud) (1.4.2)
Requirement already satisfied: six>=1.5 in c:\users\admin\anaconda3\lib\site-
packages (from python-dateutil>=2.7->matplotlib->wordcloud) (1.16.0)
Note: you may need to restart the kernel to use updated packages.
```

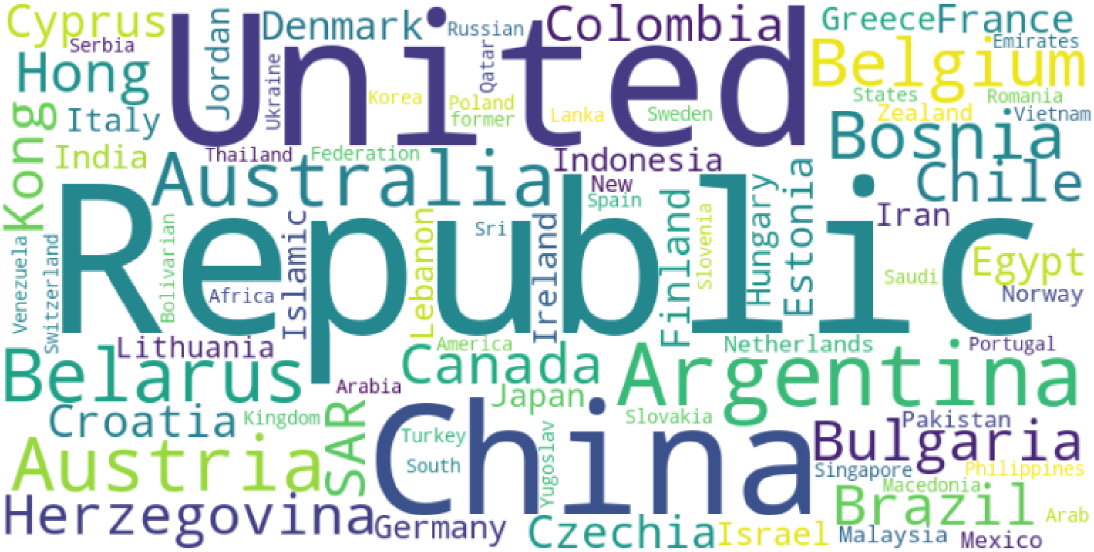
[4]:

```python
from wordcloud import WordCloud
```

1. Word Chart

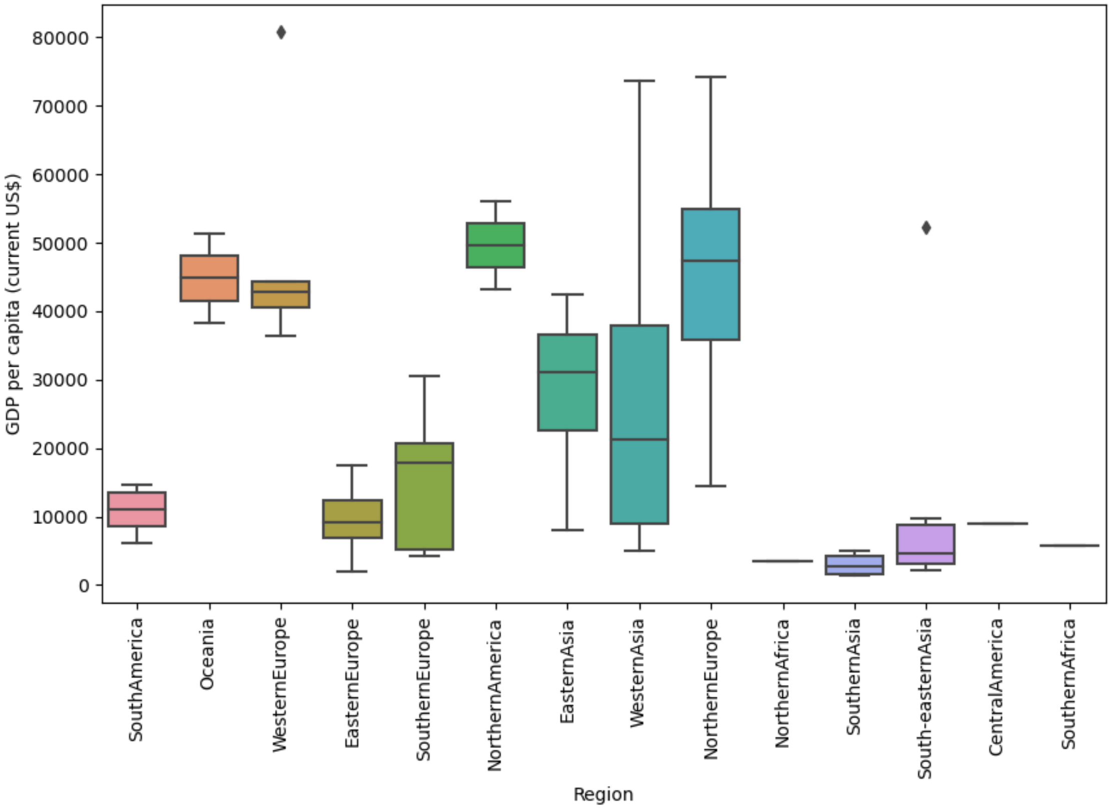
[5]:

```python
from wordcloud import WordCloud
import matplotlib.pyplot as plt

# Creating a Word Cloud from the country names or relevant socio-economic terms
text = ' '.join(df['country'])
wordcloud = WordCloud(width=800, height=400, background_color='white').
 ↪generate(text)

plt.figure(figsize=(10, 5))
plt.imshow(wordcloud, interpolation='bilinear')
plt.axis('off')
plt.show()
```



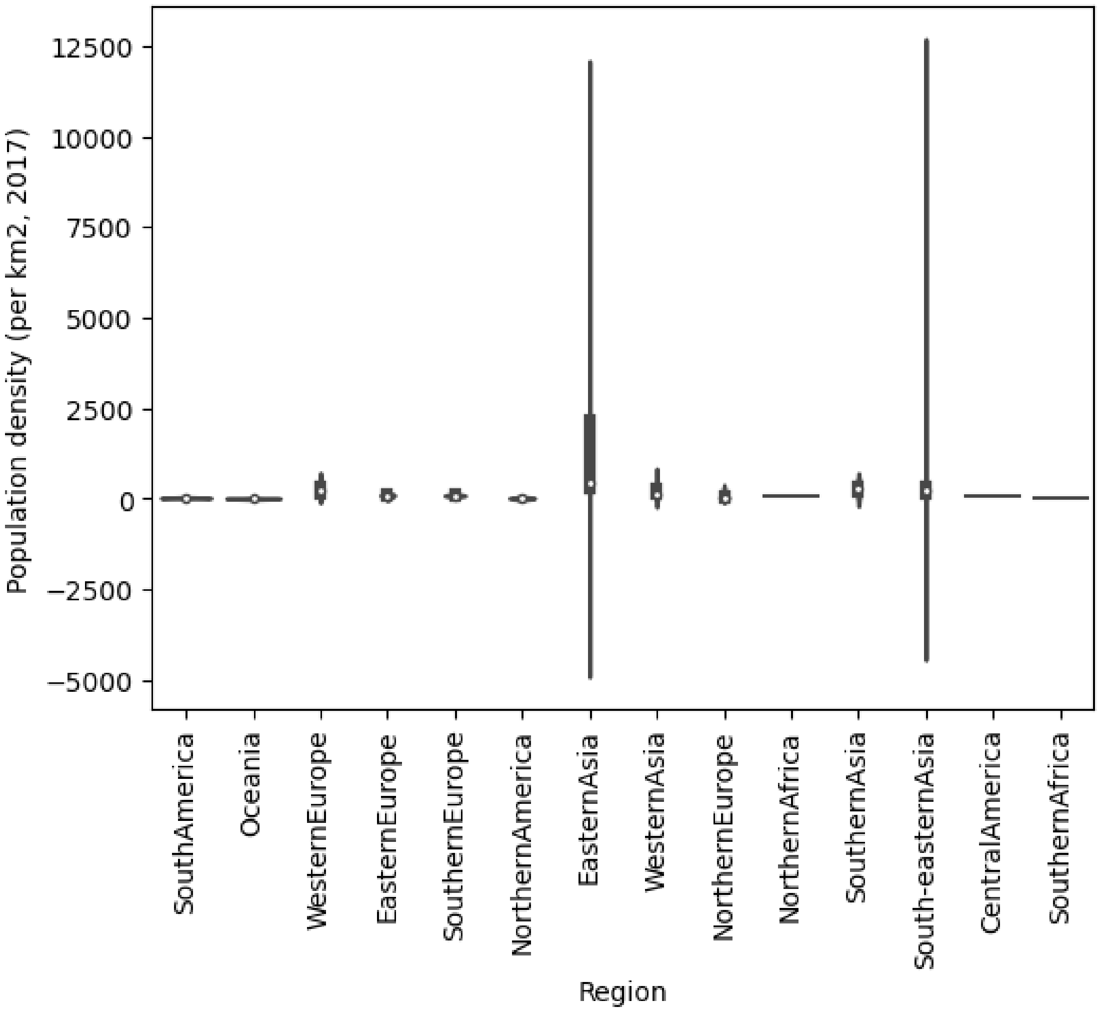
Observation: The Word Cloud highlights that terms like “Republic,” “China,” “United,” and “Australia” are prominently displayed, indicating that these countries or regions are mentioned frequently within the dataset. The larger size of these words suggests that they either have more data points or are more central to the analysis compared to other countries. For example, “China” and “United” (possibly referring to countries like the United States or the United Kingdom) appear frequently, indicating that these countries may have significant socio-economic indicators that are prominently featured in the dataset. The presence of “Republic” might suggest that several countries with “Republic” in their name (like “Republic of Korea,” “Czech Republic”) are included and have a substantial number of entries.

[6]:
```
import seaborn as sns

# Box plot for GDP per capita across regions
plt.figure(figsize=(10, 6))
sns.boxplot(x='Region', y='GDP per capita (current US$)', data=df)
plt.xticks(rotation=90)
plt.show()
```



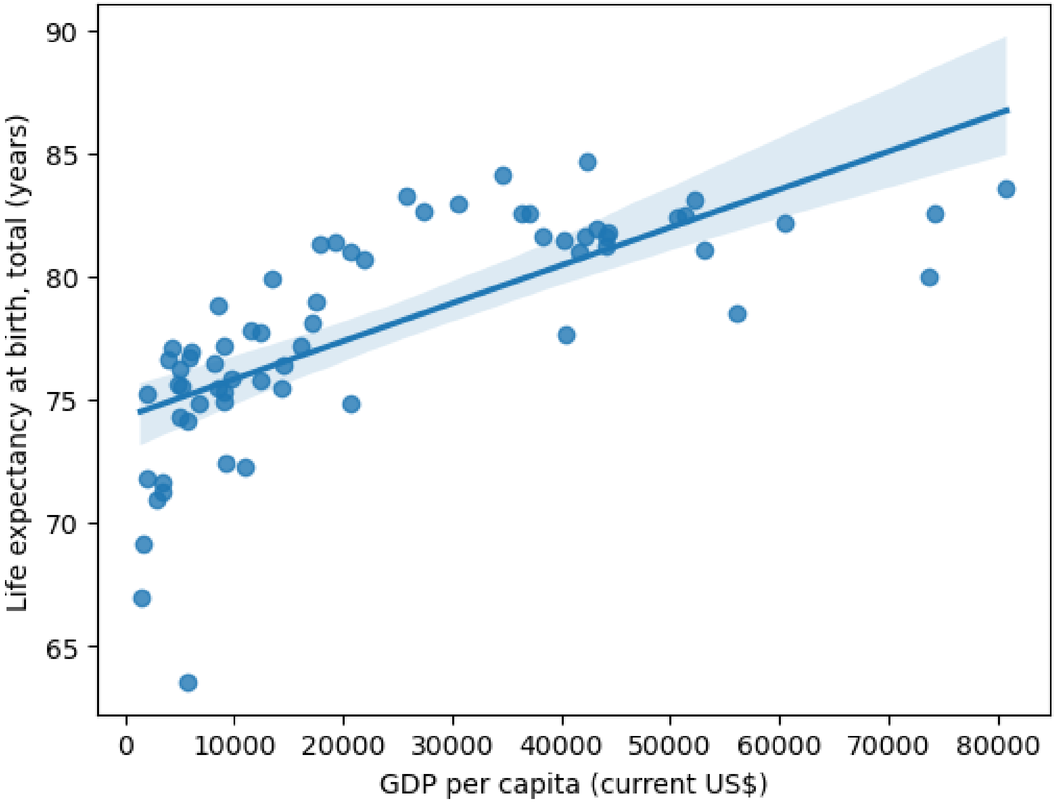
Observation: GDP Distribution Across Regions: The box plot shows significant variation in GDP per capita among different regions. Western Europe, Northern America, and Oceania have higher median GDP per capita compared to other regions, indicating higher economic prosperity. Variation and Outliers: Western Europe and Northern America display a wide range of GDP per capita, as indicated by the larger interquartile range (IQR). These regions also have outliers, suggesting that some countries within these regions have exceptionally high GDP per capita compared to others. Regions with Lower GDP: Regions like South-eastern Asia, Central America, and Southern Africa show much lower median GDP per capita with smaller IQRs, indicating less economic diversity within these regions. Comparison between Regions: Western Asia shows a moderate median GDP per capita but with significant variation, suggesting a mix of wealthy and less wealthy countries. In contrast, South America has a lower and more consistent GDP per capita across its countries.

[7]:
```
sns.violinplot(x='Region', y='Population density (per km2, 2017)', data=df)
plt.xticks(rotation=90)
plt.show()
```



Observation: Population Density Variations: The plot reveals that population density varies widely across different regions. Eastern Asia exhibits extreme values with very high population density, likely influenced by highly populous countries like China and Japan. Regions with Low Population Density: Regions like Oceania, Northern Europe, and South America show low and consistent population densities. This indicates that these regions have fewer people per square kilometer, possibly due to larger land areas relative to their population. Outliers and Data Spread: Some regions, like Eastern Asia and Northern Africa, display significant outliers and a wide range of population densities, suggesting both densely and sparsely populated countries within these regions. The presence of outliers (e.g., negative values) might indicate anomalies or errors in data entry that require further investigation. Central Tendency: Most regions, including Western Europe and Northern America, show a more centralized distribution of population density, suggesting that countries within these regions have more uniform population densities.

```
[8]: sns.regplot(x='GDP per capita (current US$)', y='Life expectancy at birth,␣
      ↪total (years)', data=df)
     plt.show()
```



observations:

There is a positive correlation between GDP per capita and life expectancy, meaning that as GDP per capita increases, life expectancy also tends to increase. The scatter plot shows a general upward trend, but with some variability among the data points, suggesting that while higher GDP per capita is generally associated with higher life expectancy, there are other factors at play.

```
[9]: import plotly.express as px

fig = px.scatter_3d(df, x='GDP per capita (current US$)', y='Population density␣
 ↪(per km2, 2017)',
                    z='Life expectancy at birth, total (years)', color='Region')
fig.show()
```

Observations: Correlation with GDP per Capita:

Similar to the previous plot, there appears to be a positive relationship between GDP per capita and life expectancy, where regions with higher GDP per capita generally show higher life expectancy. Impact of Population Density:

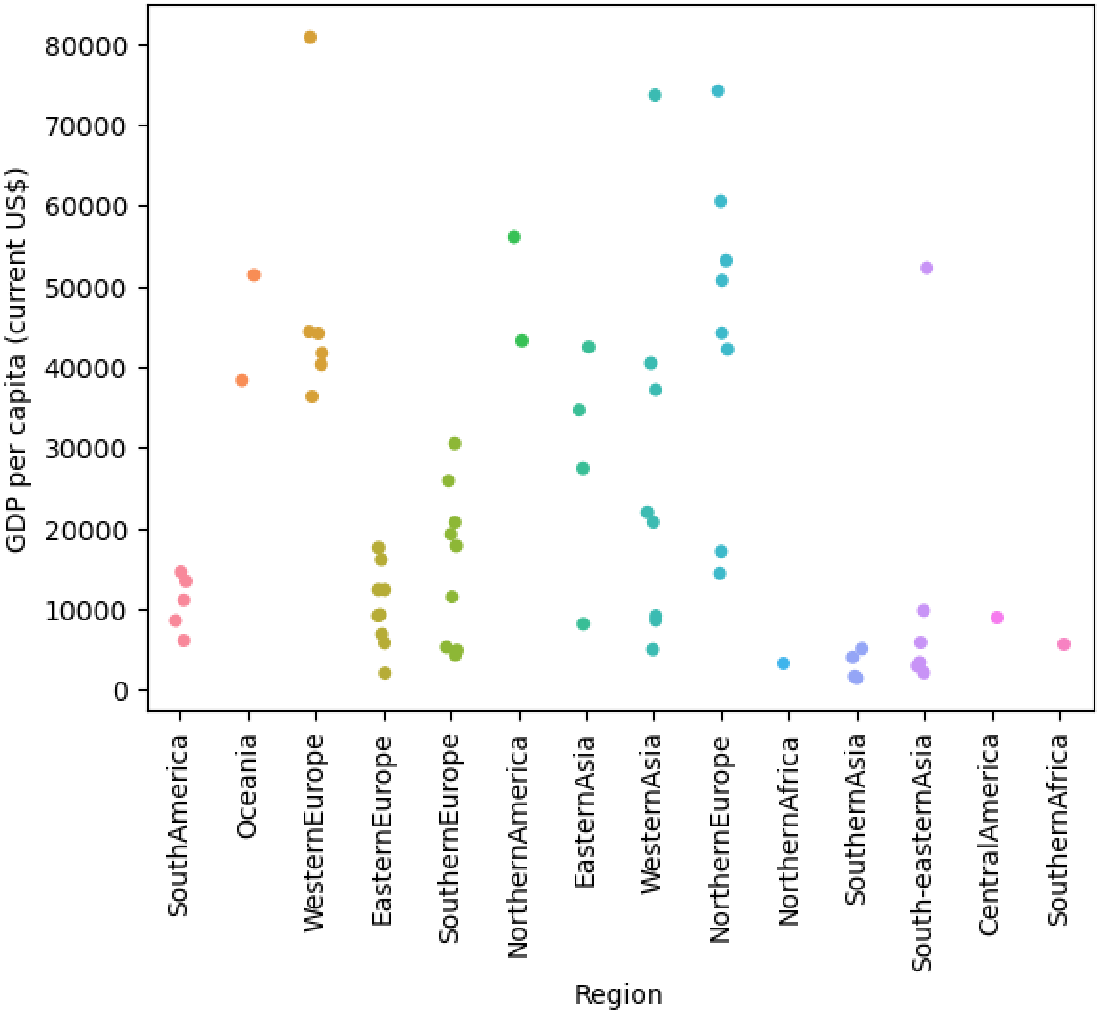
The plot also incorporates population density as a third variable. It seems that population density varies widely across regions with similar GDP per capita and life expectancy. This suggests that while population density might be relevant, it does not show a clear pattern in this 3D space.

Regional Clustering:

There is some clustering based on regions. For example, Western Europe and Northern Europe regions (orange and green, respectively) seem to cluster in areas of higher GDP per capita and life expectancy. Some regions, like Southern Asia (yellow), exhibit more spread, indicating a wider range of GDP per capita and population densities. Outliers:

A few data points, such as those from Southern Africa (purple) and Oceania (light green), seem to be outliers, possibly due to either exceptionally high or low values in GDP, population density, or life expectancy.

[10]:
```
sns.stripplot(x='Region', y='GDP per capita (current US$)', data=df,␣
  ↪jitter=True)
plt.xticks(rotation=90)
plt.show()
```



Observations: Variation in GDP per Capita Across Regions:

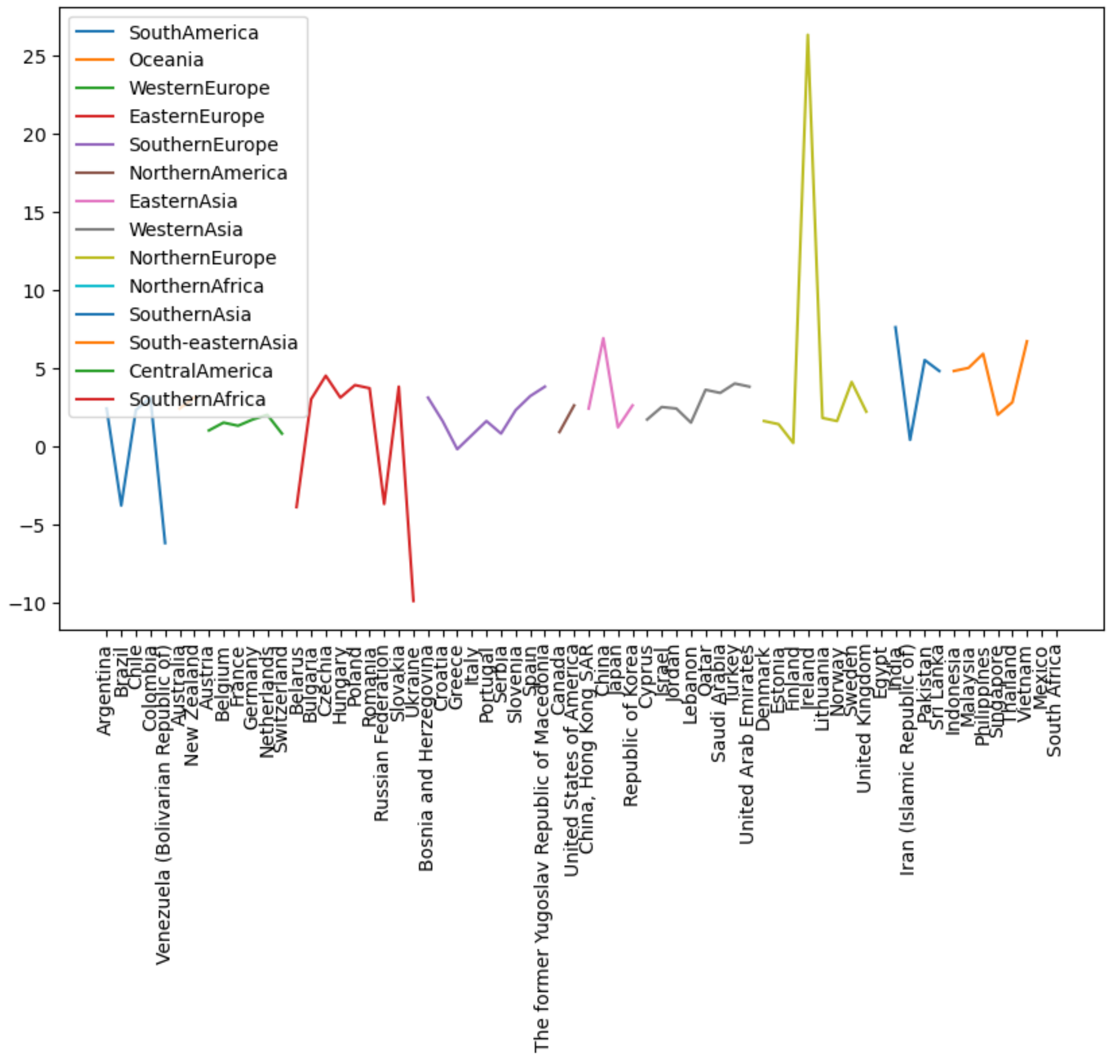
Western Europe and Northern Europe: These regions (orange and green dots) show the highest GDP per capita, with many countries in these regions clustered towards the top of the chart (above $40,000). Eastern Europe and Eastern Asia: These regions (light green and cyan dots) display a moderate range of GDP per capita, with some countries reaching up to $30,000-$40,000, but generally lower than Western and Northern Europe. Southern Africa, Central America, and South-eastern Asia: These regions (purple, light blue, and pink dots) have countries with lower GDP per capita, mostly below $20,000, with some even below $10,000. Regional Distribution:

South America, Oceania, and Southern Europe: These regions (red, blue, and light brown dots) exhibit a wide range of GDP per capita, with South America and Oceania showing some countries with low GDP per capita, while Southern Europe includes countries with higher GDP per capita. Northern Africa and Western Asia: These regions (light yellow and dark green dots) show a diverse distribution of GDP per capita, indicating significant economic disparities within the regions. Clusters and Outliers:

The plot shows clustering within certain regions, particularly in Western Europe and Northern Europe, where countries tend to have high GDP per capita. There are outliers in regions like Oceania and Southern Africa, where a few countries have significantly higher or lower GDP per capita compared to the regional average.

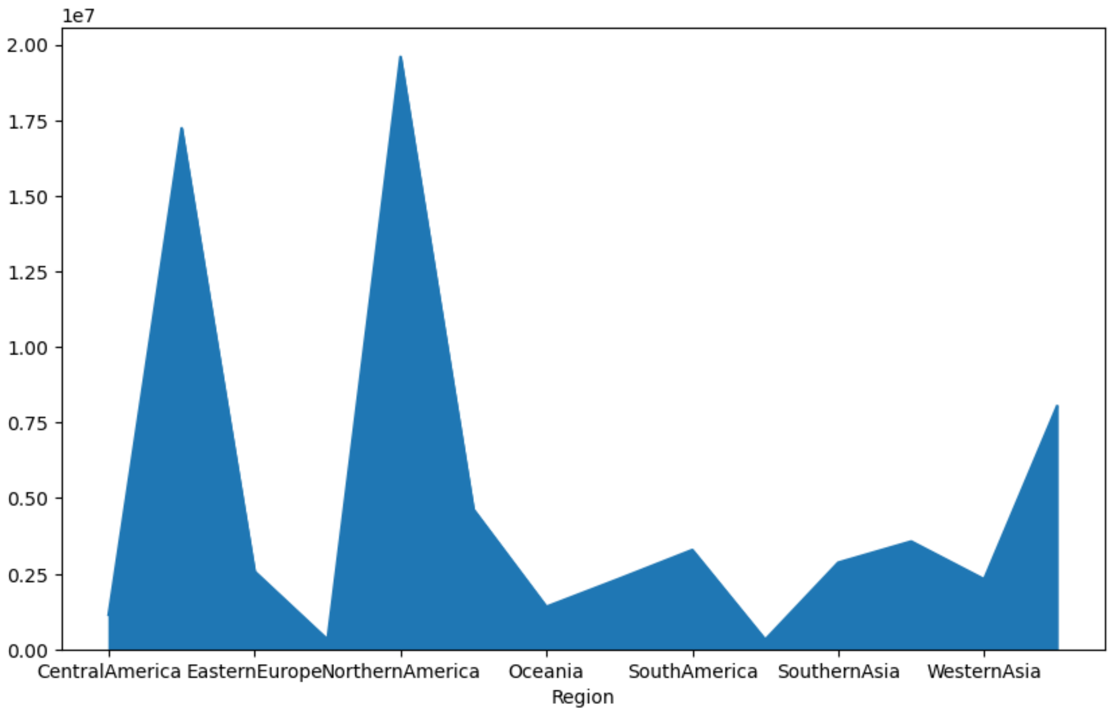
[11]:

```
plt.figure(figsize=(10, 6))
for region in df['Region'].unique():
    subset = df[df['Region'] == region]
    plt.plot(subset['country'], subset['GDP growth rate (annual %, const. 2005␣
 ↪prices)'], label=region)

plt.xticks(rotation=90)
plt.legend()
plt.show()
```

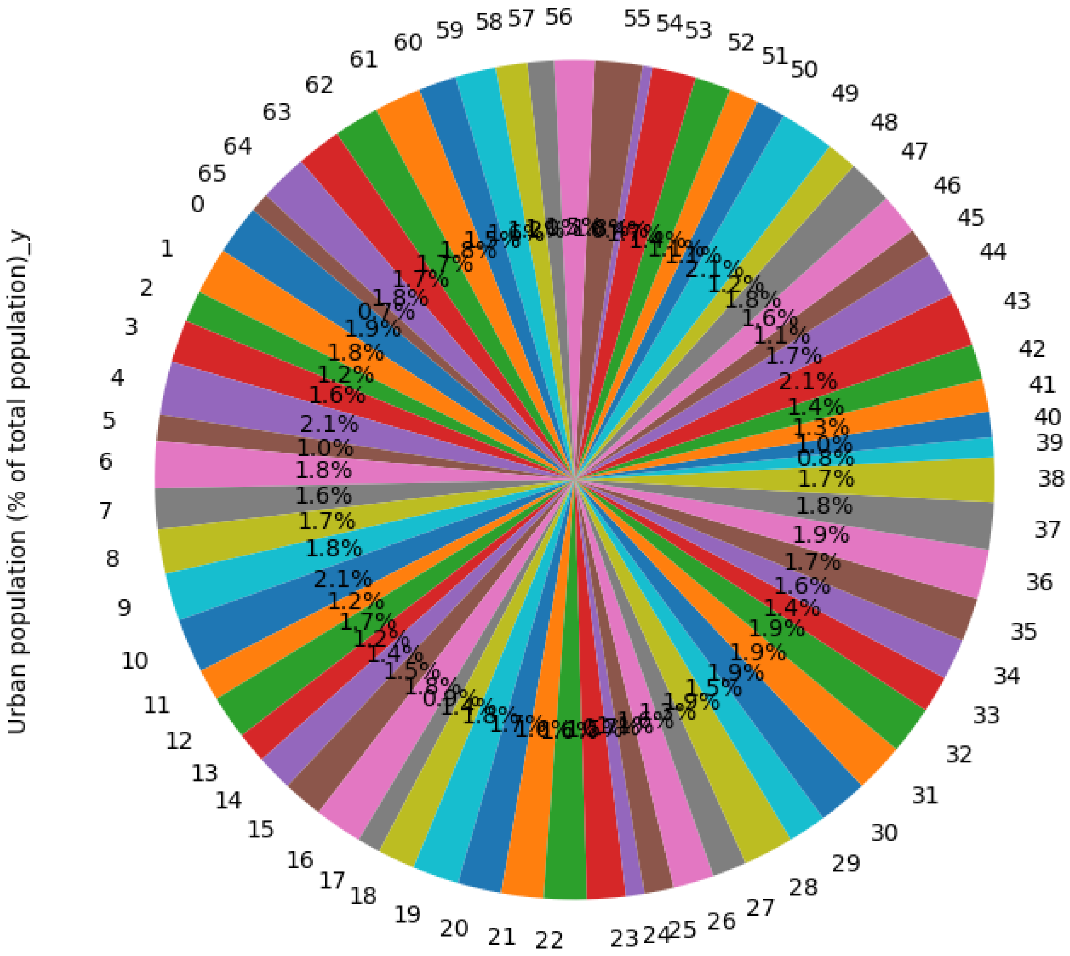


[12]:

```
df.groupby('Region')['GDP: Gross domestic product (million current US$)'].sum().
  ↪plot(kind='area', figsize=(10, 6))
plt.show()
```



```
[13]: df['Urban population (% of total population)_y'].plot(kind='pie', figsize=(8,␣
        ↪8), autopct='%1.1f%%', startangle=140)
      plt.gca().set_aspect('equal')
      plt.show()
```



[14]:

```
import plotly.express as px

fig = px.treemap(df, path=['Region', 'country'], values='GDP: Gross domestic␣
 ↪product (million current US$)')
fig.show()
```

United States: The largest block on the map, indicating it has the highest GDP compared to other countries.

China: Another significant block, representing the highest GDP in Eastern Asia.

Regions: The countries are grouped by their regions such as Northern America, Western Europe, Eastern Asia, etc.

Switzerland: The label in the image shows Switzerland with a GDP of 670,790 million current US$, categorized under Western Europe.

Relative Sizes: The size of each block corresponds to the GDP of that country. Larger blocks indicate higher GDPs, while smaller blocks indicate lower GDPs.

Color Coding: Each region appears to be color-coded differently to distinguish between them

[15]:

```
import plotly.express as px

stages = ['Planning', 'Investment', 'Implementation', 'Review']
values = [80, 60, 40, 20]  # Example values for each stage
fig = px.funnel(y=stages, x=values)
fig.show()
```

Planning: The widest segment at the top with a value of 80. Investment: The next stage down, with a value of 60. Implementation: The following stage, with a value of 40. Review: The final and narrowest stage with a value of 20. Conversion Rate: As you move down the funnel, the values decrease, indicating that some portion of whatever is being tracked (like leads, projects, or resources) does not continue to the next stage. This is typical of a funnel chart, where each stage represents a conversion from the previous stage.

Visual Representation: The decreasing width of the funnel segments represents the diminishing numbers as the process moves forward from Planning to Review.

Purpose: This chart is useful for understanding where drop-offs occur in a process and can be used in contexts like sales pipelines, project management, or process optimization to identify bottleneck