World Development Indicators Analysis - 2022

Daisy Cossio Garcia

2025-02-26

```python
# Install the necessary libraries
# pip install pandas
# pip install wbgapi

# Import the libraries
import pandas as pd
import wbgapi as wb
# Define the indicators to download
indicators = {
    'gdp_per_capita': 'NY.GDP.PCAP.CD',
    'gdp_growth_rate': 'NY.GDP.MKTP.KD.ZG',
    'inflation_rate': 'FP.CPI.TOTL.ZG',
    'unemployment_rate': 'SL.UEM.TOTL.ZS',
    'total_population': 'SP.POP.TOTL',
    'life_expectancy': 'SP.DYN.LE00.IN',
    'adult_literacy_rate': 'SE.ADT.LITR.ZS',
    'income_inequality': 'SI.POV.GINI',
    'health_expenditure_gdp_share': 'SH.XPD.CHEX.GD.ZS',
    'measles_immunisation_rate': 'SH.IMM.MEAS',
    'education_expenditure_gdp_share': 'SE.XPD.TOTL.GD.ZS',
    'primary_school_enrolment_rate': 'SE.PRM.ENRR',
    'exports_gdp_share': 'NE.EXP.GNFS.ZS'
}

# Get the list of country codes for the "World" region
country_codes = wb.region.members('WLD')

# Download data for countries only in 2022
df = wb.data.DataFrame(indicators.values(), economy=country_codes, time=2022, skipBlanks=True

# Delete the 'economy' column
```

```
df = df.drop(columns=['economy'], errors='ignore')

# Create a reversed dictionary mapping indicator codes to names
# Rename the columns and convert all names to lowercase
df.rename(columns=lambda x: {v: k for k, v in indicators.items()}.get(x, x).lower(), inplace=

# Sort 'country' in ascending order
df = df.sort_values('country', ascending=True)

# Reset the index after sorting
df = df.reset_index(drop=True)

# Display the number of rows and columns
print(df.shape)

# Display the first few rows of the data
print(df.head(3))

# Save the data to a CSV file
df.to_csv('wdi.csv', index=False)
```

```
(217, 14)
      country  inflation_rate  exports_gdp_share  gdp_growth_rate  \
0  Afghanistan             NaN          18.380042        -6.240172
1      Albania        6.725203          37.197085         4.826688
2      Algeria        9.265516          30.808979         3.600000

   gdp_per_capita  adult_literacy_rate  primary_school_enrolment_rate  \
0      357.261153                  NaN                            NaN
1     6846.426143                 98.5                      96.371231
2     4961.552577                  NaN                     108.343933

   education_expenditure_gdp_share  measles_immunisation_rate  \
0                              NaN                       56.0
1                         2.744330                       86.0
2                         4.749247                       79.0

   health_expenditure_gdp_share  income_inequality  unemployment_rate  \
0                           NaN                NaN             14.100
1                           NaN                NaN             10.137
2                           NaN                NaN             12.346
```

```
    life_expectancy  total_population
0           62.879         40578842.0
1           76.833          2777689.0
2           77.129         45477389.0
```

```
df.info()
df.head()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 217 entries, 0 to 216
Data columns (total 14 columns):
 #   Column                         Non-Null Count  Dtype
---  ------                         --------------  -----
 0   country                        217 non-null    object
 1   inflation_rate                 173 non-null    float64
 2   exports_gdp_share              179 non-null    float64
 3   gdp_growth_rate                206 non-null    float64
 4   gdp_per_capita                 207 non-null    float64
 5   adult_literacy_rate            54 non-null     float64
 6   primary_school_enrolment_rate  156 non-null    float64
 7   education_expenditure_gdp_share 137 non-null   float64
 8   measles_immunisation_rate      193 non-null    float64
 9   health_expenditure_gdp_share   20 non-null     float64
 10  income_inequality              28 non-null     float64
 11  unemployment_rate              186 non-null    float64
 12  life_expectancy                209 non-null    float64
 13  total_population               217 non-null    float64
dtypes: float64(13), object(1)
memory usage: 23.9+ KB
```

	country	inflation_rate	exports_gdp_share	gdp_growth_rate	gdp_per_capita	adult_lite
0	Afghanistan	NaN	18.380042	-6.240172	357.261153	NaN
1	Albania	6.725203	37.197085	4.826688	6846.426143	98.5
2	Algeria	9.265516	30.808979	3.600000	4961.552577	NaN
3	American Samoa	NaN	46.957520	1.735016	18017.458938	NaN
4	Andorra	NaN	NaN	9.564612	42414.059009	NaN

Exploratory Data Analysis

I will be exploring the following indicators: "GDP Per Capita", "Life Expectancy", and "Un-employment Rate" for the year 2022. The analysis includes summary statistics, handling of

missing values, and visualizations to understand the relationships between these variables.

GDP Per Capita

GDP per capita measures a country's economic output per person and is often used as an indicator of living standards. Higher GDP per capita typically correlates with better access to goods, services, and overall quality of life.

Summary Statistics

```
gdp_summary = df['gdp_per_capita'].describe()
gdp_summary
```

```
count       207.000000
mean      20520.336828
std       30640.741594
min         250.634225
25%        2599.752468
50%        7606.237525
75%       27542.145523
max      226052.001905
Name: gdp_per_capita, dtype: float64
```

Insights:

- The average GDP per capita is 20520.336828 USD.
- The minimum GDP per capita is 250.634225 USD, while the maximum is 226052.001905 USD.
- There is a wide range, indicating significant economic disparity across countries.

Distribution of GDP Per Capita

```
import matplotlib.pyplot as plt

plt.figure(figsize=(8, 6))
plt.hist(df['gdp_per_capita'].dropna(), bins=30, edgecolor='black')
plt.title('Distribution of GDP Per Capita (2022)')
plt.xlabel('GDP Per Capita (USD)')
plt.ylabel('Number of Countries')
```

```
plt.grid(axis='y')
plt.show()
```

Distribution of GDP Per Capita (2022)
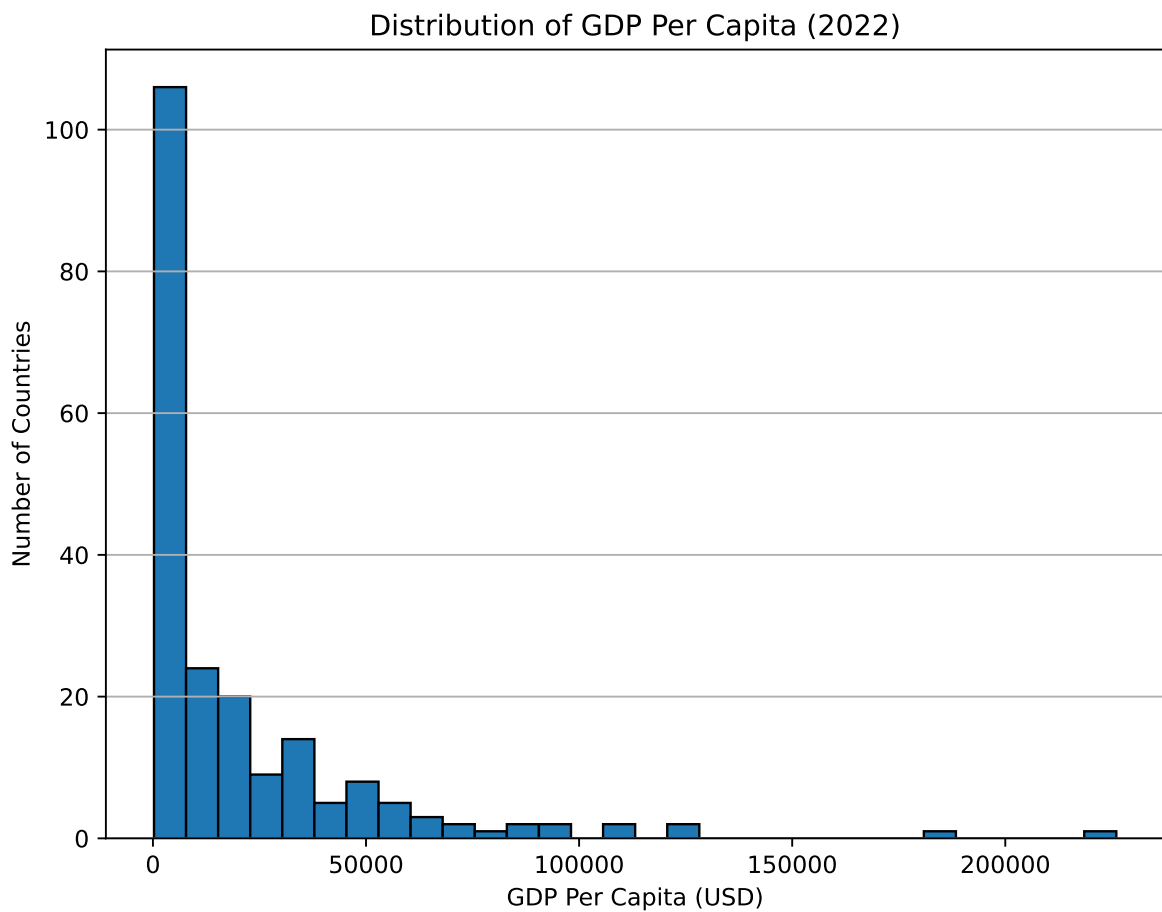
The distribution is right-skewed, with most countries having a GDP per capita below $10,000. A few high-income countries significantly increase the range.

Life Expectancy

Life expectancy at birth reflects the average number of years a newborn is expected to live under current mortality levels. It serves as an indicator of a country's overall health and development.

Summary Statistics

```
life_exp_summary = df['life_expectancy'].describe()
life_exp_summary
```

```
count    209.000000
mean      72.416519
std        7.713322
min       52.997000
25%       66.782000
50%       73.514634
75%       78.475000
max       85.377000
Name: life_expectancy, dtype: float64
```

Insights:

- The mean life expectancy is 72.416519 years.
- The lowest life expectancy observed is 52.997000 years, while the highest is 85.377000 years.
- The range suggests disparities in healthcare access and living conditions.

Relationship Between GDP Per Capita and Life Expectancy

```
import matplotlib.pyplot as plt

plt.figure(figsize=(8, 6))
plt.scatter(df['gdp_per_capita'], df['life_expectancy'], alpha=0.7, color='blue')
plt.xlabel('GDP Per Capita (USD)')
plt.ylabel('Life Expectancy (Years)')
plt.title('GDP Per Capita vs. Life Expectancy (2022)')
plt.grid(True)
plt.show()
```

GDP Per Capita vs. Life Expectancy (2022)
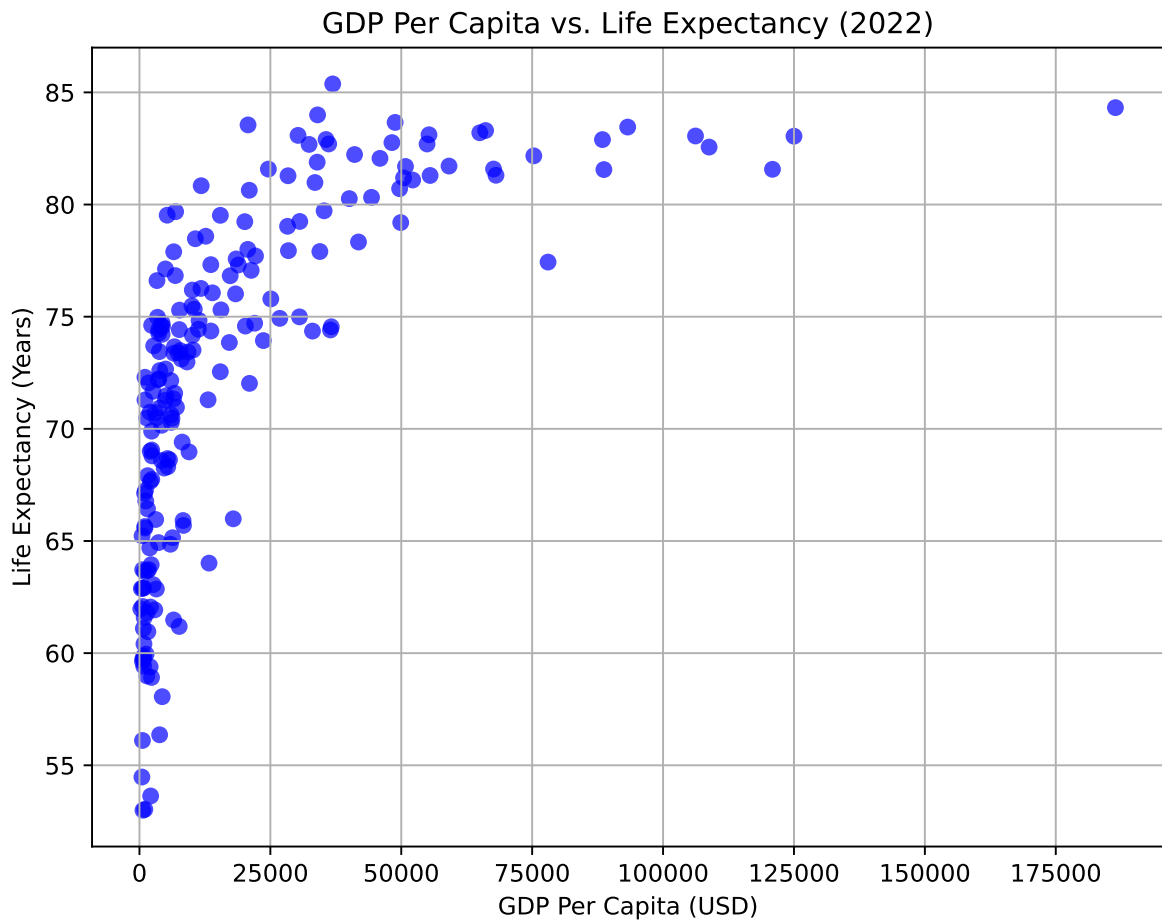
As shown in @fig-gdp-life, there is a positive correlation between GDP per capita and life e
There appears to be a positive correlation between GDP per capita and life expectancy, especi

Unemployment Rate

The unemployment rate represents the percentage of the labor force that is unemployed but act

Summary Statistics

::: {.cell execution_count=7}
``` {.python .cell-code}
unemployment_summary = df['unemployment_rate'].describe()
unemployment_summary
```

count   186.000000

```
mean        7.227344
std         5.844462
min         0.130000
25%         3.478000
50%         5.334000
75%         9.261750
max        35.359000
Name: unemployment_rate, dtype: float64
```

:::

Insights:

- The average unemployment rate is 7.227344%.
- Some countries experience unemployment rates as high as 35.359000%, reflecting potential economic challenges.

Top 10 Countries by Unemployment Rate

```
top_unemployment = df[['country', 'unemployment_rate']].nlargest(10, 'unemployment_rate')
top_unemployment
```
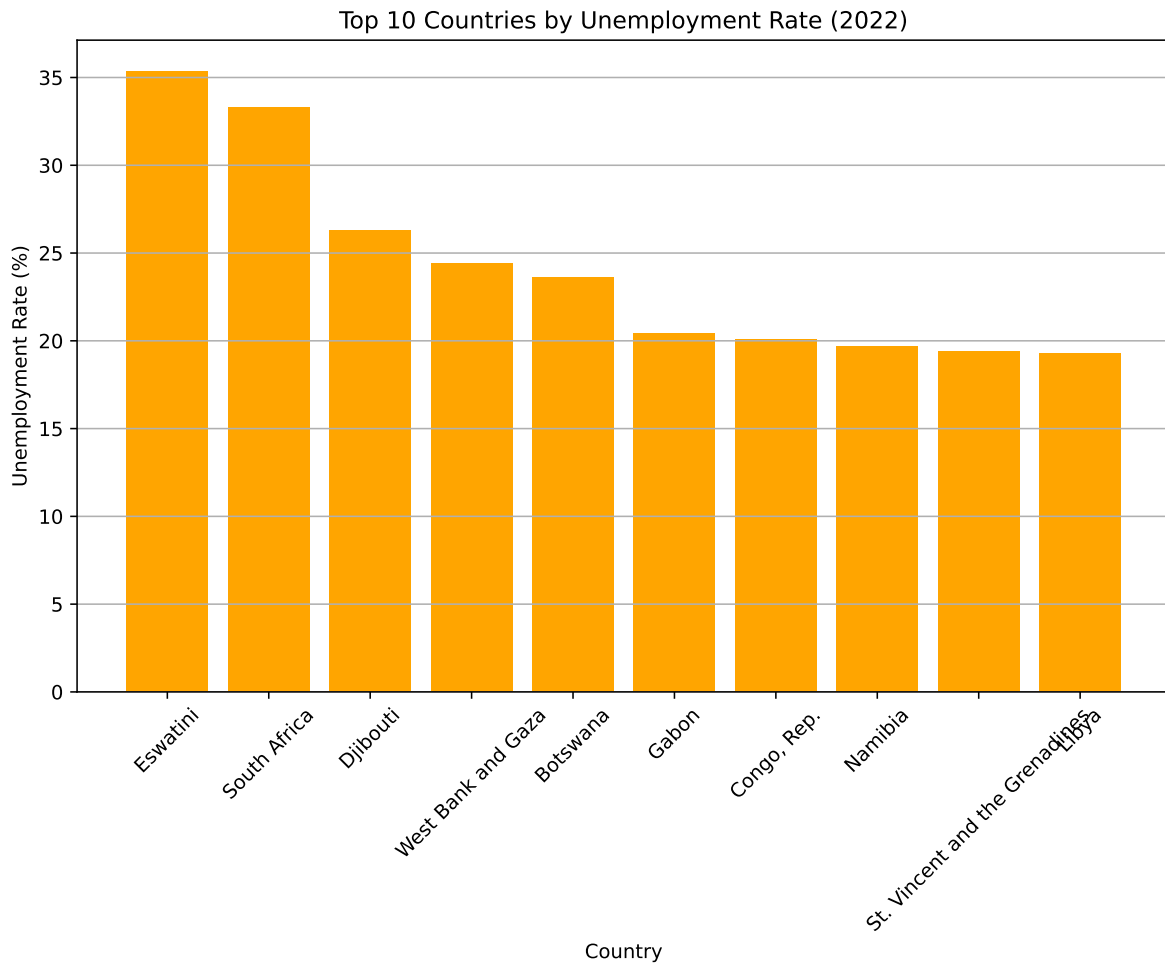
	country	unemployment_rate
63	Eswatini	35.359
177	South Africa	33.268
54	Djibouti	26.307
213	West Bank and Gaza	24.420
25	Botswana	23.615
70	Gabon	20.399
45	Congo, Rep.	20.108
136	Namibia	19.695
184	St. Vincent and the Grenadines	19.417
113	Libya	19.282

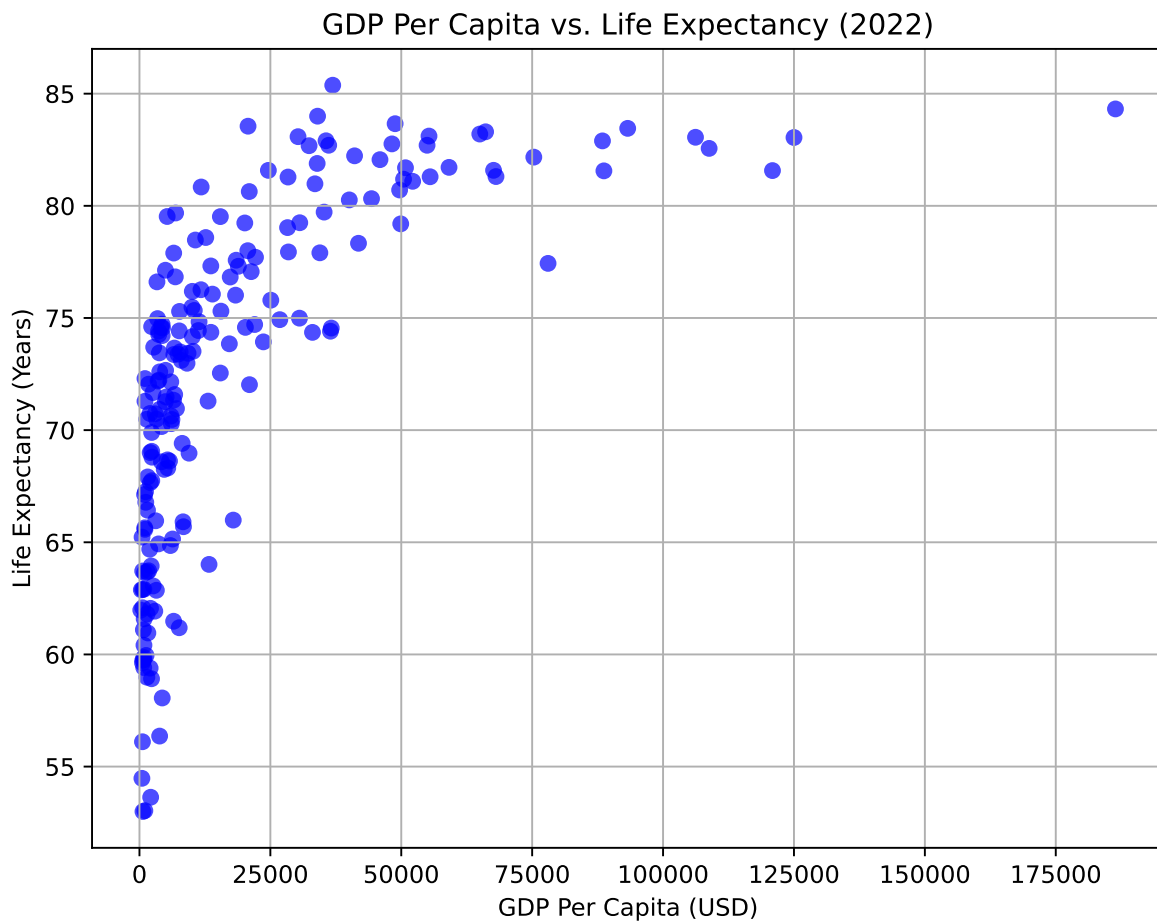**Visualization: Top 10 Unemployment Rates**

```
plt.figure(figsize=(10, 6))
plt.bar(top_unemployment['country'], top_unemployment['unemployment_rate'], color='orange')
plt.title('Top 10 Countries by Unemployment Rate (2022)')
plt.xlabel('Country')
plt.ylabel('Unemployment Rate (%)')
plt.xticks(rotation=45)
plt.grid(axis='y')
plt.show()
```


The countries with the highest unemployment rates are primarily developing economies. This may reflect structural economic issues, political instability, or challenges in labor market policies.

As shown in **?@fig-gdp-life**, GDP per capita has a clear relationship with life expectancy.

```
plt.figure(figsize=(8, 6))
plt.scatter(df['gdp_per_capita'], df['life_expectancy'], alpha=0.7, color='blue')
plt.xlabel('GDP Per Capita (USD)')
plt.ylabel('Life Expectancy (Years)')
plt.title('GDP Per Capita vs. Life Expectancy (2022)')
plt.grid(True)
plt.show()
```


GDP Per Capita vs. Life Expectancy (2022)

```
plt.figure(figsize=(10, 6))
plt.bar(top_unemployment['country'], top_unemployment['unemployment_rate'], color='orange')
plt.xlabel('Country')
plt.ylabel('Unemployment Rate (%)')
plt.title('Top 10 Countries by Unemployment Rate (2022)')
plt.xticks(rotation=45)
```

```
plt.grid(axis='y')
plt.show()
```


Top 10 Countries by Unemployment Rate (2022)
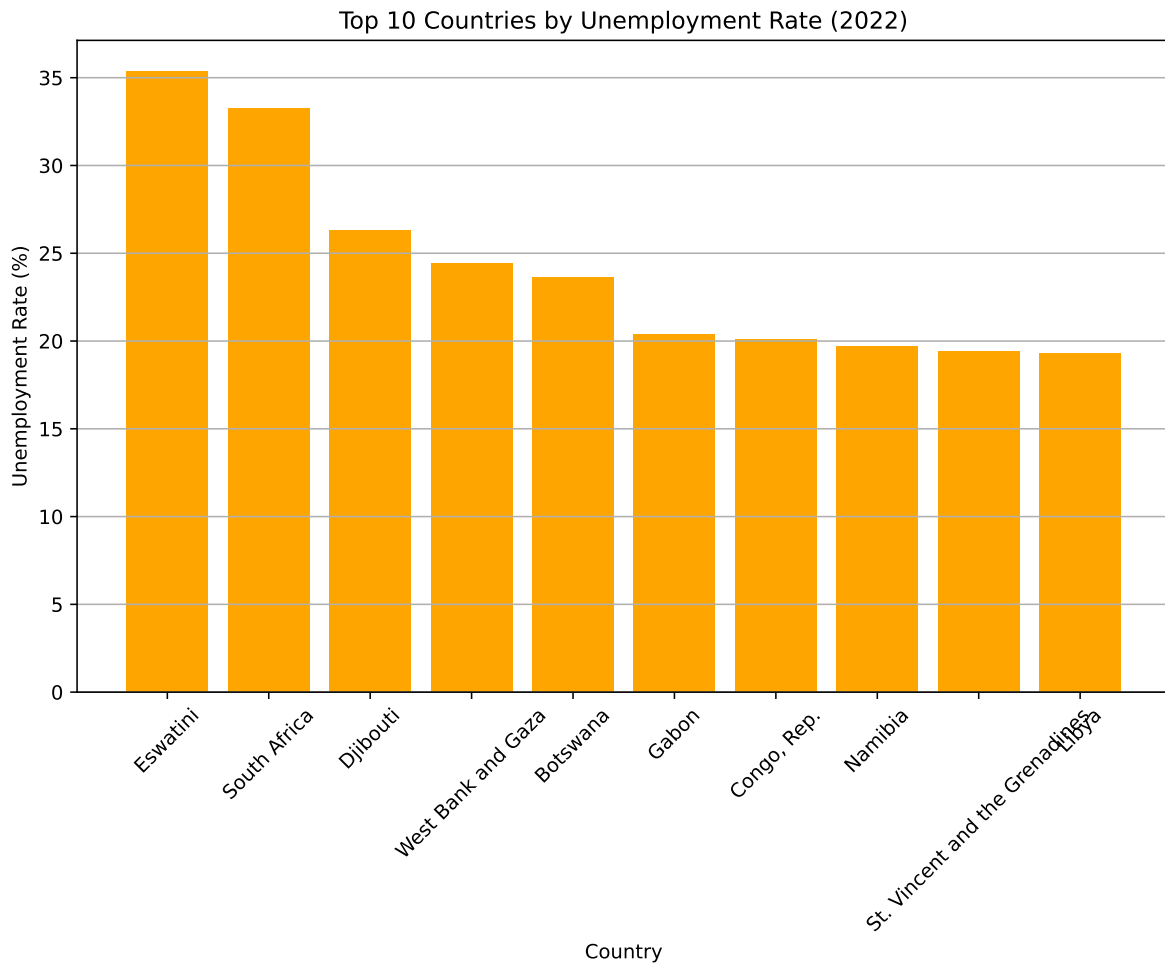
Table 1: Summary Statistics for Selected Indicators (2022)

Summarize the following indicators: -GDP Per Capita (gdp_per_capita) -Life Expectancy (life_expectancy) -Unemployment Rate (unemployment_rate)

The table will include: -Mean (Average value) -Median (Middle value) -Minimum (Lowest recorded value) -Maximum (Highest recorded value) -Standard Deviation (Measure of data spread)

This table provides key summary statistics for GDP per capita, life expectancy, and unemployment rate across all available countries.

```python
import pandas as pd

# Select relevant indicators
summary_stats = df[['gdp_per_capita', 'life_expectancy', 'unemployment_rate']].describe().T

# Rename the columns for readability
summary_stats = summary_stats[['mean', '50%', 'min', 'max', 'std']]
summary_stats.columns = ['Mean', 'Median', 'Min', 'Max', 'Std Dev']

# Round values for clarity
summary_stats = summary_stats.round(2)

# Display the table using Pandas
summary_stats
```

	Mean	Median	Min	Max	Std Dev
gdp_per_capita	20520.34	7606.24	250.63	226052.00	30640.74
life_expectancy	72.42	73.51	53.00	85.38	7.71
unemployment_rate	7.23	5.33	0.13	35.36	5.84

Key Observations: - The average GDP per capita is around 20520.34 USD, but there is a large variation, indicating economic inequality. - Life expectancy varies from 53.00 to 85.38 years, with a median of 73.51. - Unemployment rates differ significantly, ranging from 0.13% to 35.36%, highlighting labor market disparities.

Cite the Sources

The **World Development Indicators (WDI) dataset** provides global economic and social metrics, essential for policy analysis (Bank 2022).

GDP per capita has long been linked to **health and economic development**, a relationship observed by Preston (1975) in the **Preston Curve** (Preston 1975).

Additionally, effective **data visualization** is critical for communicating trends in economic indicators (Knaflic 2015). ## References ::: {#refs}

Quarto

Quarto enables you to weave together content and executable code into a finished document. To learn more about Quarto see https://quarto.org.

Bank, World. 2022. "World Development Indicators 2022." https://databank.worldbank.org/source/world-development-indicators.

Knaflic, Cole Nussbaumer. 2015. *Storytelling with Data: A Data Visualization Guide for Business Professionals.* Wiley.

Preston, Samuel H. 1975. "The Changing Relation Between Mortality and Level of Economic Development." *Population Studies* 29 (2): 231–48.