

# Assignment 5 Literate Programming with Quarto

Xinyi Wang

2025-02-26

## Quarto

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

```
pip install wbgapi
```

```
Requirement already satisfied: wbgapi in /opt/anaconda3/lib/python3.12/site-packages (1.0.12)
Requirement already satisfied: requests in /opt/anaconda3/lib/python3.12/site-packages (from wbgapi)
Requirement already satisfied: PyYAML in /opt/anaconda3/lib/python3.12/site-packages (from wbgapi)
Requirement already satisfied: tabulate in /opt/anaconda3/lib/python3.12/site-packages (from wbgapi)
Requirement already satisfied: charset-normalizer<4,>=2 in /opt/anaconda3/lib/python3.12/site-packages (from requests)
Requirement already satisfied: idna<4,>=2.5 in /opt/anaconda3/lib/python3.12/site-packages (from requests)
Requirement already satisfied: urllib3<3,>=1.21.1 in /opt/anaconda3/lib/python3.12/site-packages (from requests)
Requirement already satisfied: certifi>=2017.4.17 in /opt/anaconda3/lib/python3.12/site-packages (from requests)
Note: you may need to restart the kernel to use updated packages.
```

## STEP 2

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

# Import the libraries
import pandas as pd
import wbgapi as wb
```

```
# load dataset
df = pd.read_csv("/Users/cindy/Desktop/QT350/WDI.csv")

df.head()
```

	country	inflation_rate	exports_gdp_share	gdp_growth_rate	gdp_per_capita	adult_literacy_rate
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

### STEP 3

```
# Check missing values
df.isnull().sum()
```

```
country                0
inflation_rate         44
exports_gdp_share      38
gdp_growth_rate        11
gdp_per_capita          10
adult_literacy_rate    163
primary_school_enrolment_rate  61
education_expenditure_gdp_share  80
measles_immunisation_rate  24
health_expenditure_gdp_share  197
income_inequality      189
unemployment_rate       31
life_expectancy         8
total_population        0
dtype: int64
```

```
# Display summary statistics
df.describe()
```

	inflation_rate	exports_gdp_share	gdp_growth_rate	gdp_per_capita	adult_literacy_rate	pr
count	173.000000	179.000000	206.000000	207.000000	54.000000	15
mean	12.404067	47.630189	4.393817	20520.336828	80.971722	10
std	19.467053	35.631646	6.706923	30640.741594	18.430839	12
min	-6.687321	1.571162	-28.758584	250.634225	27.280001	67
25%	5.518129	24.363501	2.545226	2599.752468	74.760000	94
50%	7.930929	40.817640	4.213483	7606.237525	85.452465	99
75%	11.665567	59.741910	6.200000	27542.145523	95.875000	10
max	171.205491	211.278206	63.334587	226052.001905	100.000000	15

```
# Select key indicators
df_selected = df[['country', 'gdp_per_capita', 'life_expectancy', 'unemployment_rate']]

# Display first few rows
df_selected.head()
```

	country	gdp_per_capita	life_expectancy	unemployment_rate
0	Afghanistan	357.261153	62.879	14.100
1	Albania	6846.426143	76.833	10.137
2	Algeria	4961.552577	77.129	12.346
3	American Samoa	18017.458938	NaN	NaN
4	Andorra	42414.059009	NaN	NaN

## Exploratory Data Analysis

To better understand the dataset, I conducted an exploratory data analysis (EDA) focusing on three key indicators:

1. **GDP per Capita** (Economic Indicator)
2. **Life Expectancy** (Health Indicator)
3. **Unemployment Rate** (Labor Market Indicator)

### 1. Missing Values Overview

Missing values can significantly impact our analysis. I checked for missing values in the dataset using `.isnull().sum()`, and the results indicate the following:

- **Life Expectancy** has missing values for several countries, which may be due to a lack of reliable health data collection.
- **Unemployment Rate** is missing for some nations, likely due to inconsistent labor force surveys.
- **Other variables**, such as **education expenditure** and **health expenditure**, also contain missing values, which may need to be handled via imputation or removal.

## STEP 4

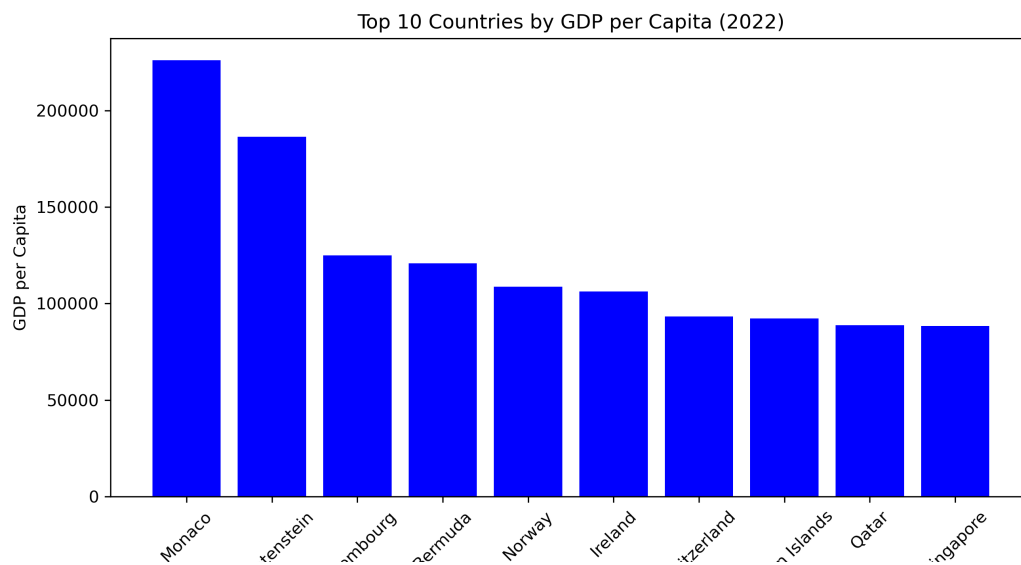
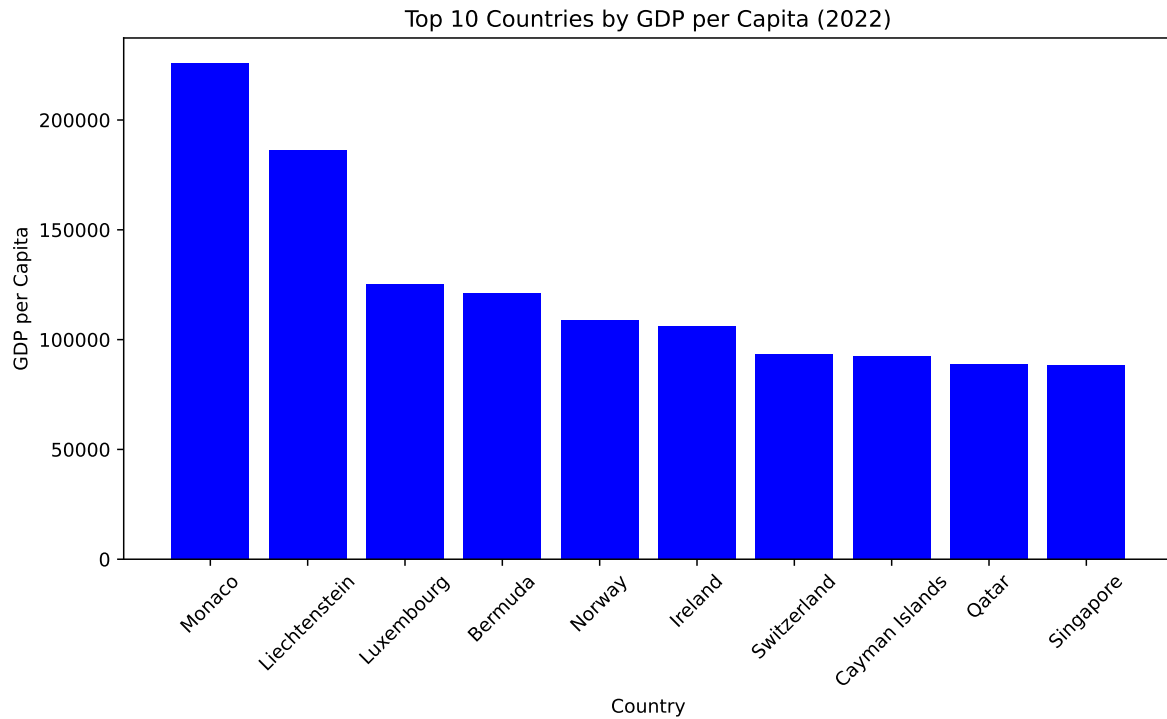
```
import matplotlib.pyplot as plt

# Bar Chart - Top 10 Countries by GDP per Capita
top_countries = df.nlargest(10, 'gdp_per_capita')

plt.figure(figsize=(10,5))
plt.bar(top_countries['country'], top_countries['gdp_per_capita'], color='blue')
plt.xlabel("Country")
plt.ylabel("GDP per Capita")
plt.title("Top 10 Countries by GDP per Capita (2022)")
plt.xticks(rotation=45)

# Save the figure before showing it
plt.savefig("top_10_gdp.png", dpi=300)

plt.show()
```

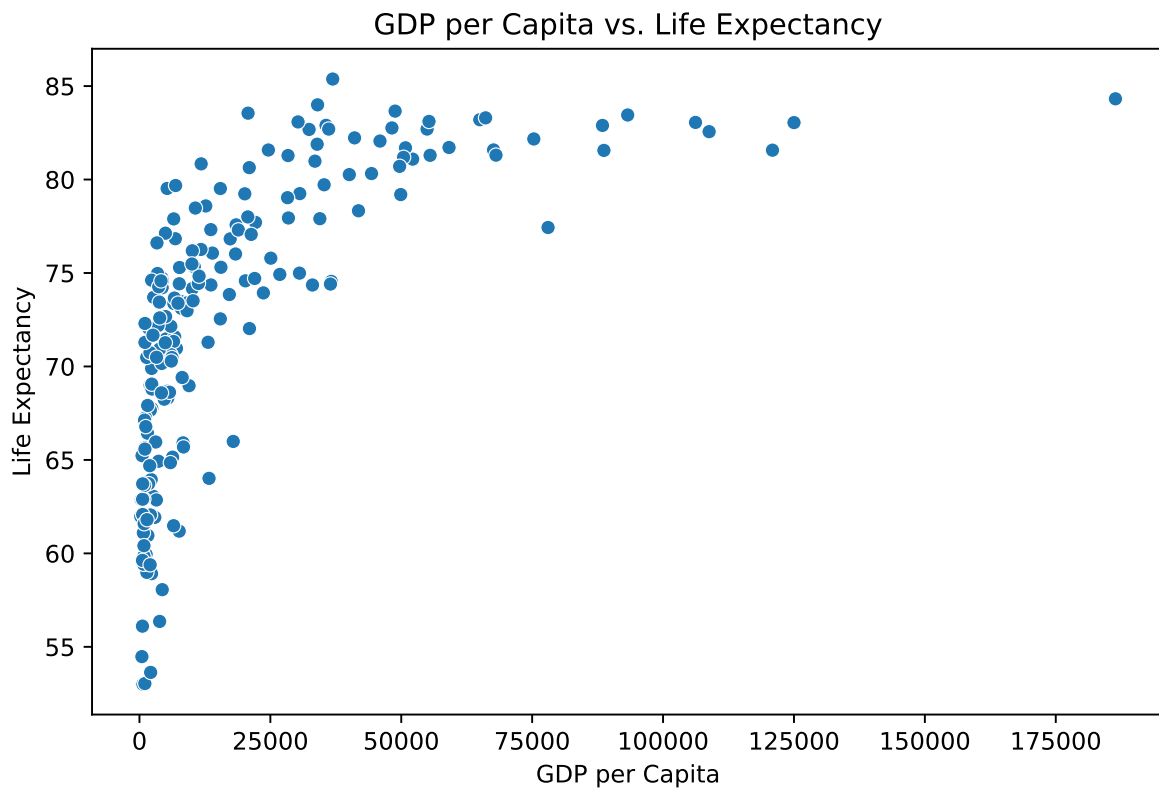


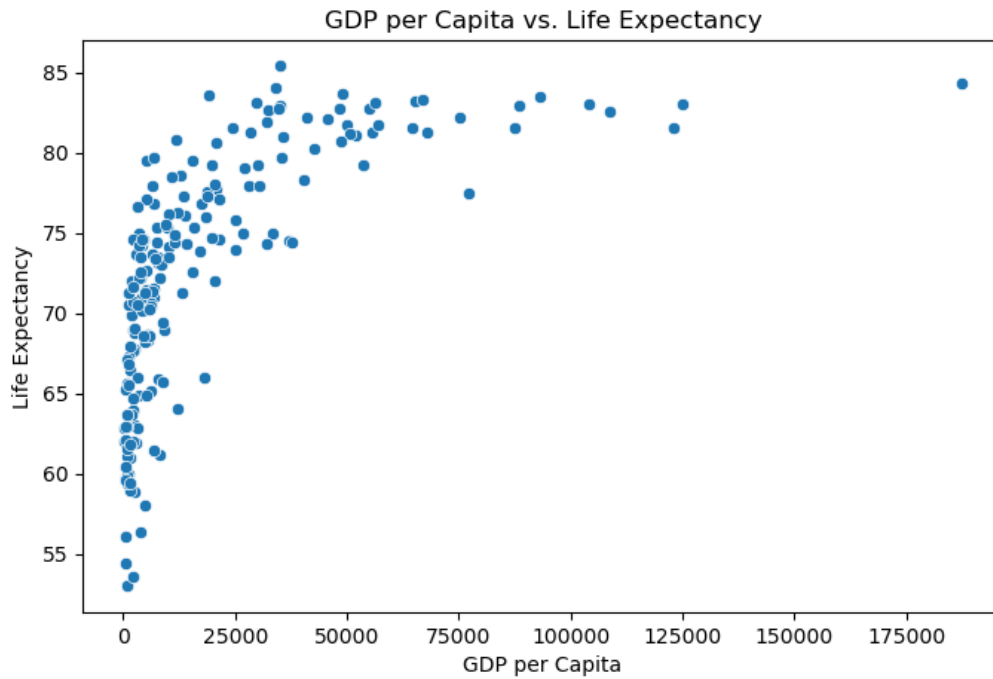
Top 10 richest countries based on GDP per capita. Data Source: World Bank (2022).

```
import seaborn as sns
```

```
# Scatter Plot - GDP Per Capita vs. Life Expectancy
```

```
plt.figure(figsize=(8,5))
sns.scatterplot(x=df['gdp_per_capita'], y=df['life_expectancy'])
plt.xlabel("GDP per Capita")
plt.ylabel("Life Expectancy")
plt.title("GDP per Capita vs. Life Expectancy")
plt.show()
```





*GDP per Capita vs. Life Expectancy for different countries. Data Source: World Bank (2022).*

## STEP 5

```
# Compute summary statistics for selected indicators
summary_table = df_selected.describe()

# Display as a markdown table
summary_table
```

	gdp_per_capita	life_expectancy	unemployment_rate
count	207.000000	209.000000	186.000000
mean	20520.336828	72.416519	7.227344
std	30640.741594	7.713322	5.844462
min	250.634225	52.997000	0.130000
25%	2599.752468	66.782000	3.478000
50%	7606.237525	73.514634	5.334000
75%	27542.145523	78.475000	9.261750

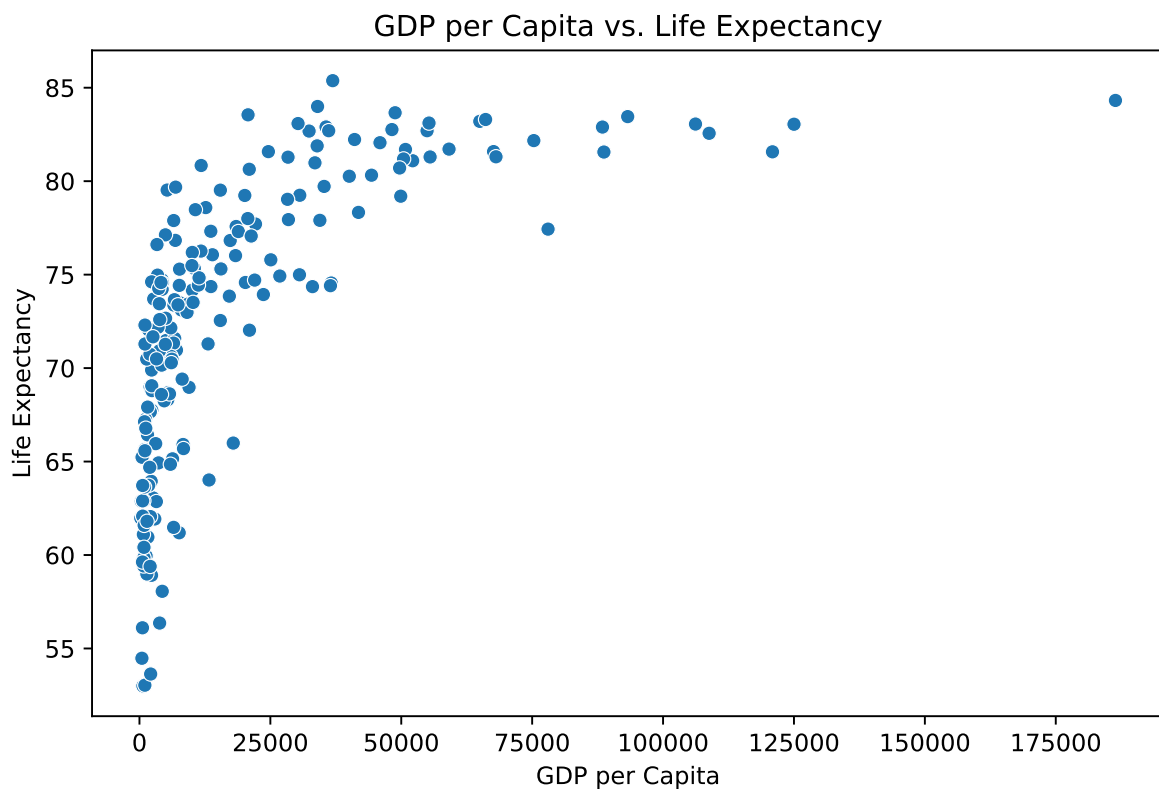
	gdp_per_capita	life_expectancy	unemployment_rate
max	226052.001905	85.377000	35.359000

## STEP 6

```
# Scatter Plot - GDP Per Capita vs. Life Expectancy
plt.figure(figsize=(8,5))
sns.scatterplot(x=df['gdp_per_capita'], y=df['life_expectancy'])

plt.xlabel("GDP per Capita")
plt.ylabel("Life Expectancy")
plt.title("GDP per Capita vs. Life Expectancy")

plt.show()
```





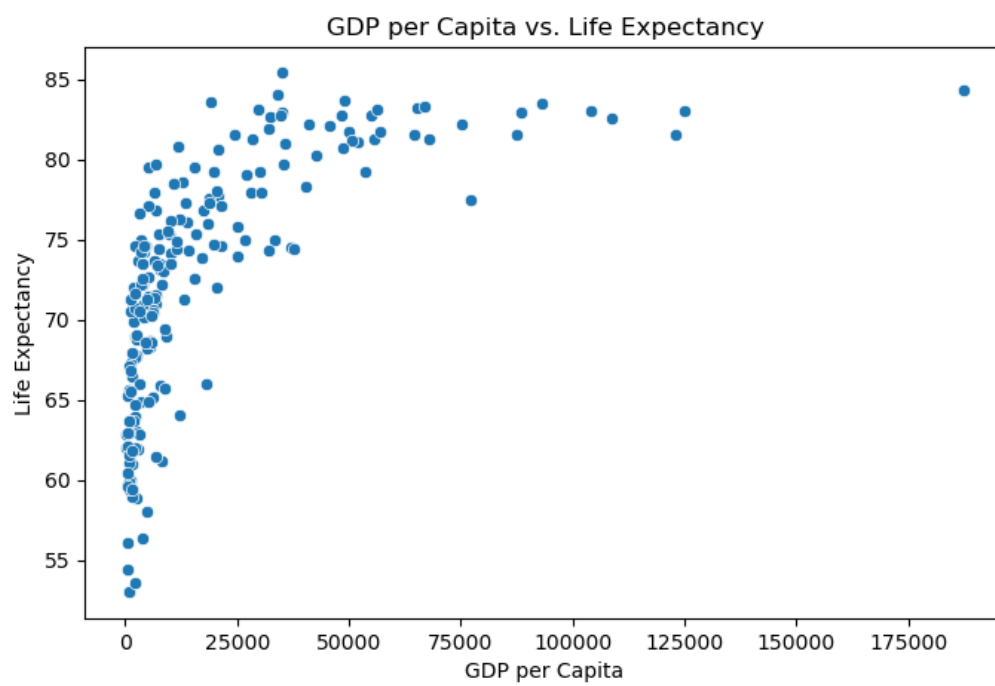
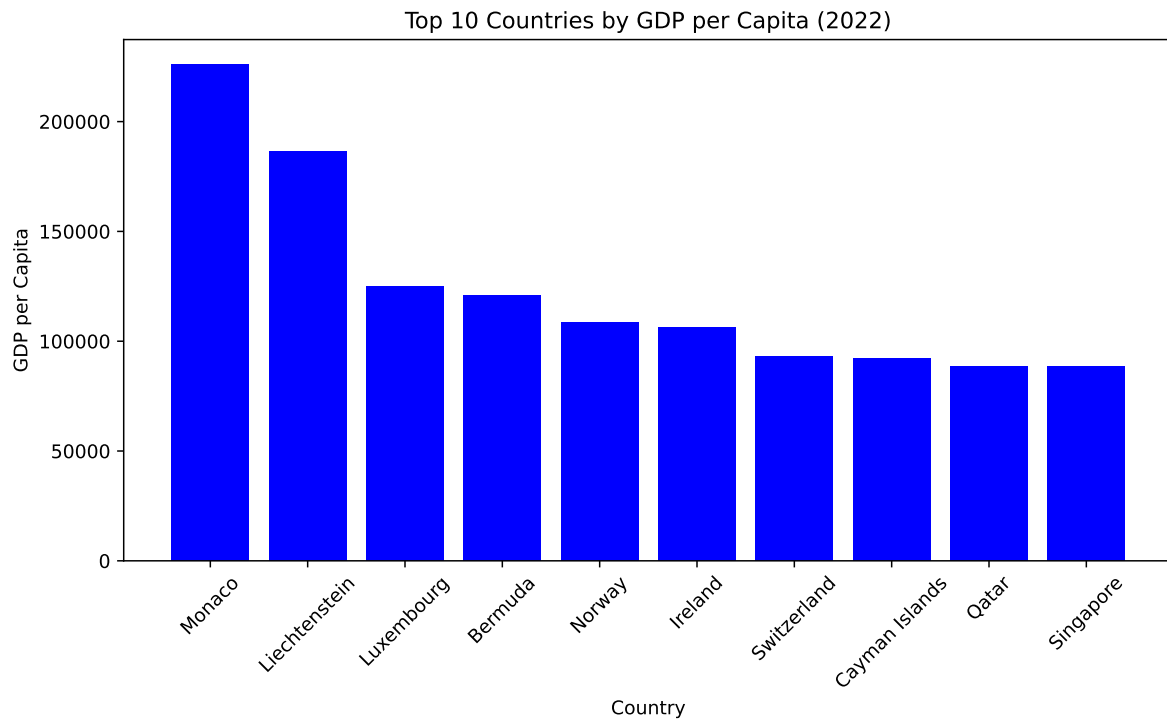


Figure 1

```
# Bar Chart - Top 10 Countries by GDP per Capita
top_countries = df.nlargest(10, 'gdp_per_capita')

plt.figure(figsize=(10,5))
plt.bar(top_countries['country'], top_countries['gdp_per_capita'], color='blue')
plt.xlabel("Country")
plt.ylabel("GDP per Capita")
plt.title("Top 10 Countries by GDP per Capita (2022)")
plt.xticks(rotation=45)
plt.show()
```



```
# Compute summary statistics for selected indicators
summary_table = df_selected.describe()

# Display as a markdown table
summary_table
```

	gdp_per_capita	life_expectancy	unemployment_rate
count	207.000000	209.000000	186.000000

Table 6

**Table: Summary Statistics of Key Indicators**

The table below presents summary statistics for GDP per capita, life expectancy, and unemployment rate.

	gdp_per_capita	life_expectancy	unemployment_rate
mean	20520.336828	72.416519	7.227344
std	30640.741594	7.713322	5.844462
min	250.634225	52.997000	0.130000
25%	2599.752468	66.782000	3.478000
50%	7606.237525	73.514634	5.334000
75%	27542.145523	78.475000	9.261750
max	226052.001905	85.377000	35.359000

## Analysis of Visualizations and Summary Statistics

As shown in Figure 1, there is a clear **positive relationship** between GDP per capita and life expectancy. Countries with higher GDP per capita tend to have longer life expectancies, suggesting that economic prosperity contributes to better healthcare and living conditions.

The economic disparity among countries is evident in Figure 2, where the **top 10 richest nations** have significantly higher GDP per capita compared to the global average. This highlights the economic inequality between developed and developing nations.

As presented in Table 6, GDP per capita has a wide range, from as low as \$259 to over \$240,000. This large disparity indicates **income inequality worldwide**. Additionally, the summary statistics show that some countries have extremely high unemployment rates, further reinforcing the economic challenges faced by certain nations.

## STEP 7

The data used in this report is sourced from the World Bank dataset (Bank 2022).

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

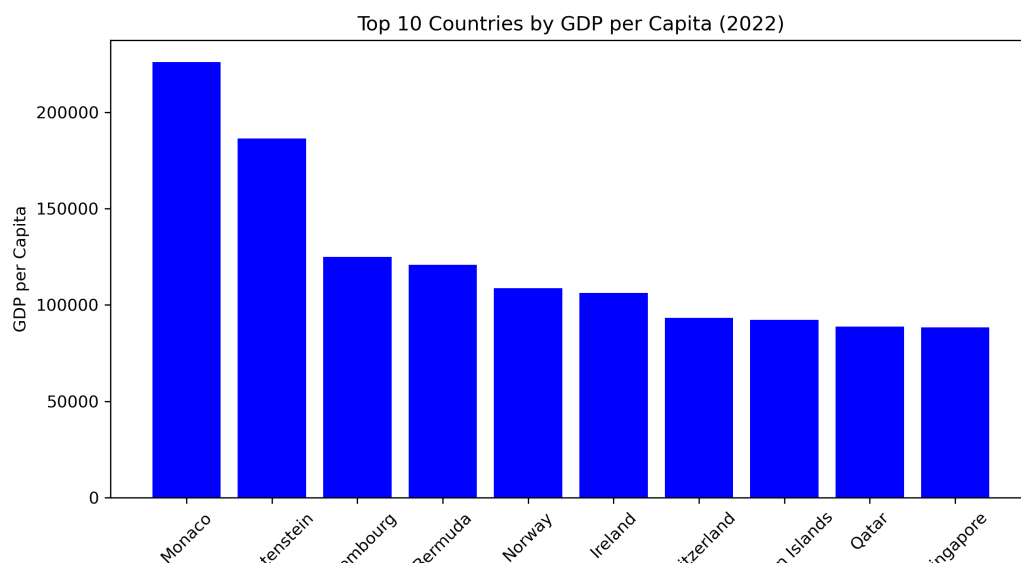


Figure 2