Global Country Cross-Analysis

Visulization Project

Siddhesh Maheshwari

2023-09-29

Table of Contents

Abstract :

This report offers an extensive analysis of global country data,The main goal of this study is to find important patterns and connections between different countries, so we can compare them and understand what's happening on a global scale. It may serves as a valuable tool for researchers and analysts seeking deeper insights into global trends and international dynamics.

1 Introduction:

We have a comprehensive dataset provides a wealth of information about all countries worldwide. It can facilitate in-depth analysis and foster cross-country comparisons. The process involves a comprehensive examination of known global characteristics while simultaneously exploring new data insights to better understand the interplay of factors contributing to a country's economic growth. The primary objective of this study is to glean profound insights into worldwide trends and the dynamics that underlie international affairs, facilitating cross-country juxtapositions and comprehensive investigations.

2 Data Description:

This extensive dataset offers an abundance of data on countries across the globe. It encompasses demographic statistics, economic metrics, environmental factors, healthcare indicators, educational statistics, and a host of other information. With representation from every country, this dataset provides a comprehensive global view of various facets of nations, facilitating in-depth examinations and cross-national assessments. From the Global Country Information Dataset 2023, we'll focus on the most valuable data for our research needs,even though it offers a wide range of information.

While our dataset encompasses 35 different features, our primary attention will be directed towards specific key indicators. But we will try to study about GDP, Birth Rate, Agricultural Land (%), CO2 Emissions, CPI (Consumer Price Index), Life Expectancy, Unemployment Rate, Population, Tax Revenue, and more.

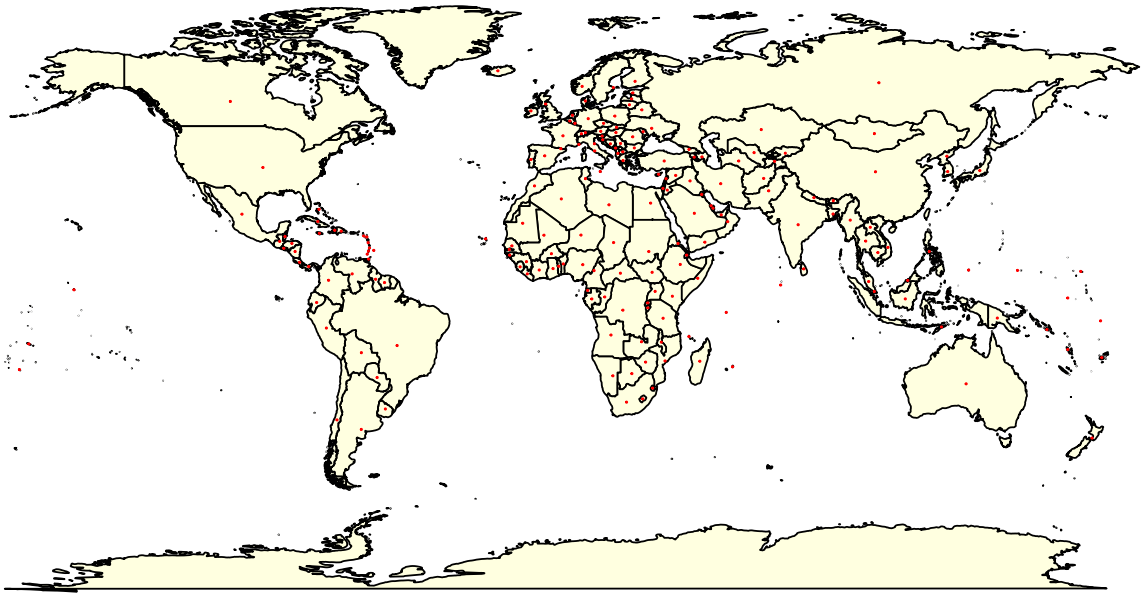
```
data=read.csv("/home/sidd/sidd/Visualization_/db/World-data-2023.csv")
library(ggplot2)
```

Table 1: Name of Features of Data Set

Density_P_Km2_	Abbreviation
Agricultural_Land_	Land_Area_Km2_
Armed_Forces_size	Birth_Rate
Calling_Code	Capital_Major_City
Co2_Emissions	CPI
CPI_Change_	Currency_Code
Fertility_Rate	Forested_Area_
Gasoline_Price	GDP
Gross_primary_education_enrollment_	Gross_tertiary_education_enrollment_
Infant_mortality	Largest_city
Life_expectancy	Maternal_mortality_ratio
Minimum_wage	Official_language
Out_of_pocket_health_expenditure	Physicians_per_thousand
Population	Population_Labor_force_participation_
Tax_revenue_	Total_tax_rate
Unemployment_rate	Urban_population
Latitude	Longitude

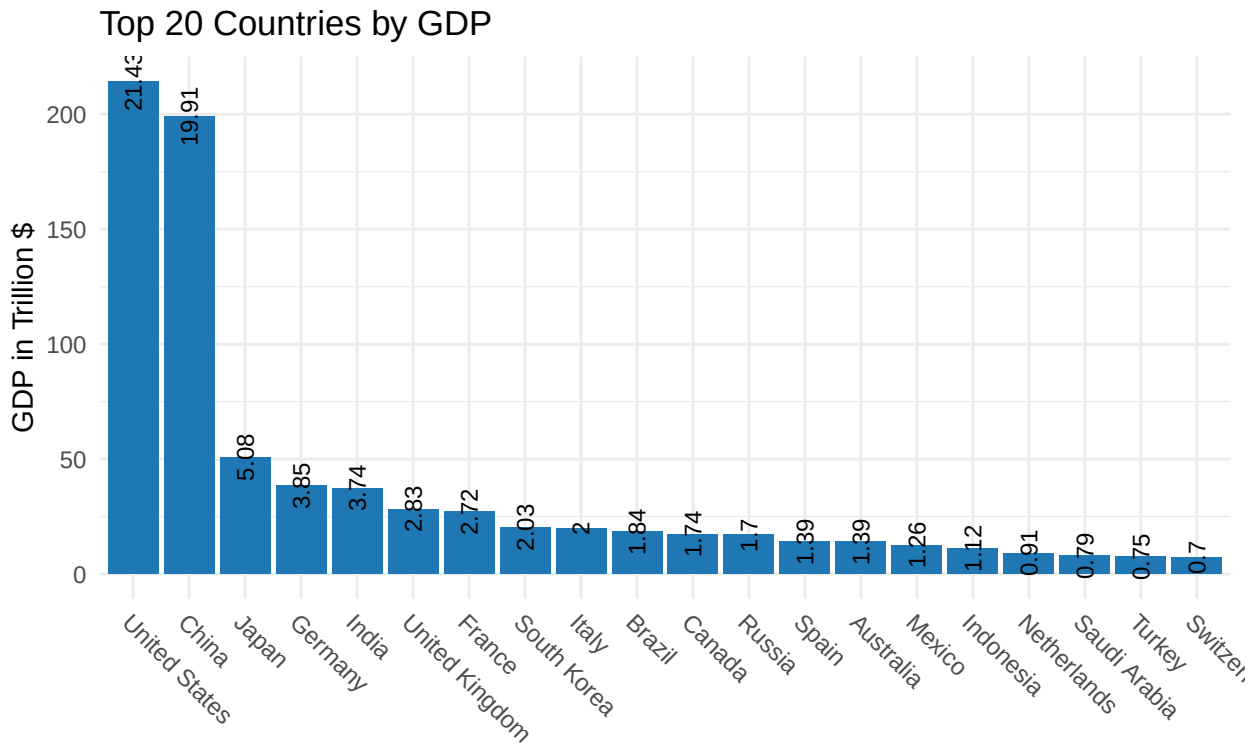
Some Important Terms :-
GDP: Gross Domestic Product, the total value of goods and services produced in the country.
CPI: Consumer Price Index, a measure of inflation and purchasing power and changes in the cost of living.
Infant Mortality: Number of deaths per 1,000 live births before reaching one year of age.
Tax Revenue (%): Tax revenue as a percentage of GDP measures the amount of government revenue collected from taxes relative to the size of the economy (GDP).
Unemployment Rate: Percentage of the labor force that is unemployed.
CO2 Emissions: Carbon dioxide (CO2) emissions refer to the release of carbon dioxide gas into the atmospher,.
Fertility Rate: The fertility rate represents the average number of children born to a woman over her lifetime.
Life Expectancy : It refers to the average number of years a person can expect to live, based on current mortality rates and other demographic factors.

Countries for which we have data

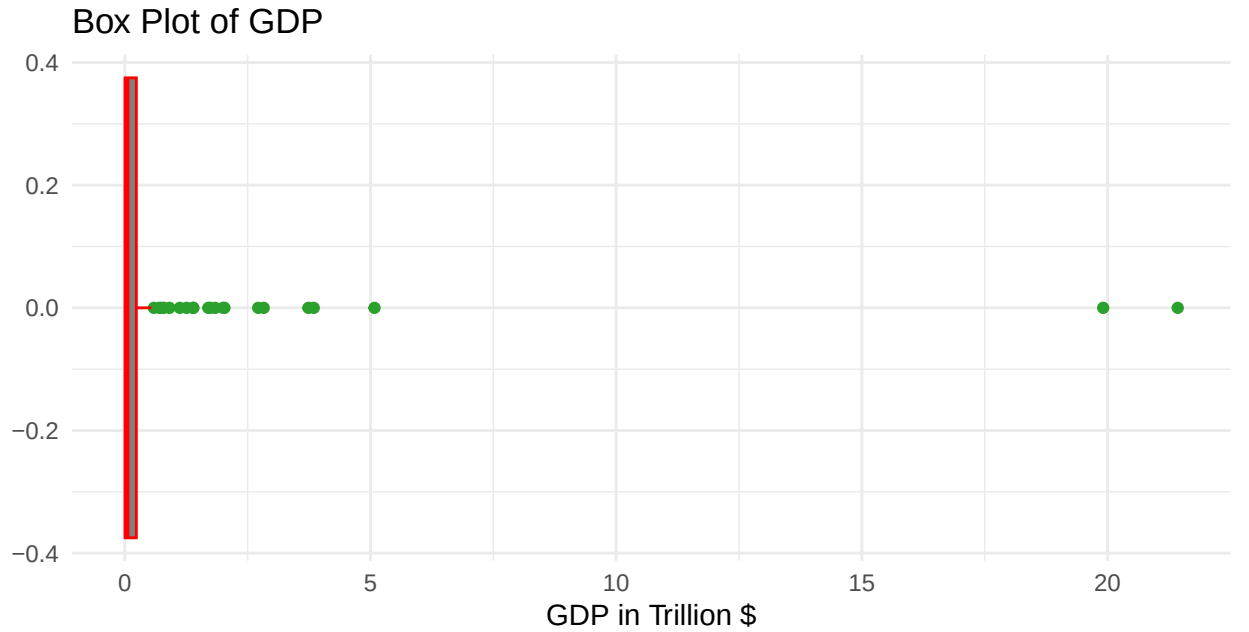


3 Exploratory Data Aalysis:

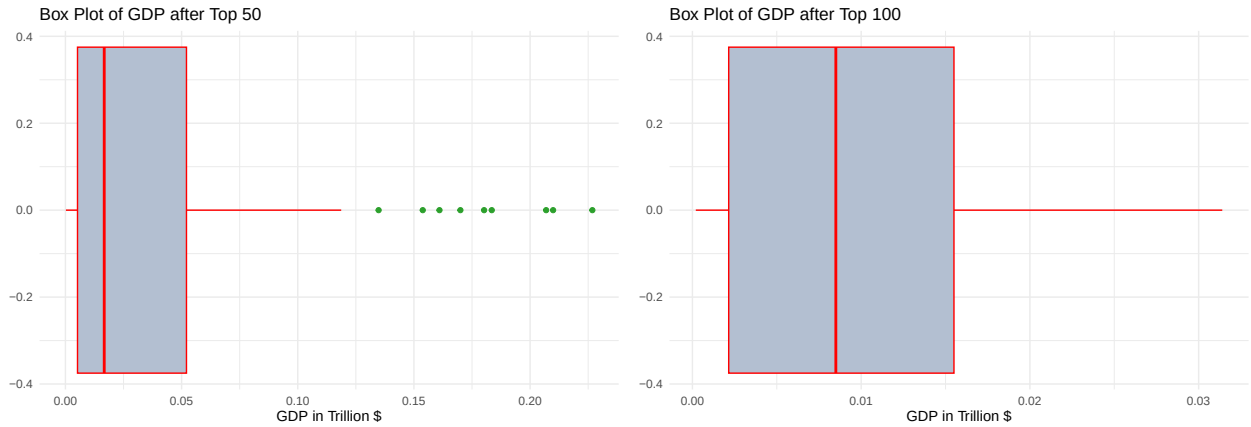
GDP (Gross domestic product):



It's evident that the **United States and China** make substantial contributions to the global GDP, while **INDIA** holds the fifth position.



Here We can Clearly See that GDP is highly affected from outliers like United States and China. If we remove some data then make a plot lets see what happens.



Now clearly we can see that If we remove Chine and US then our data is not abnormal.

Population and BirthRate :

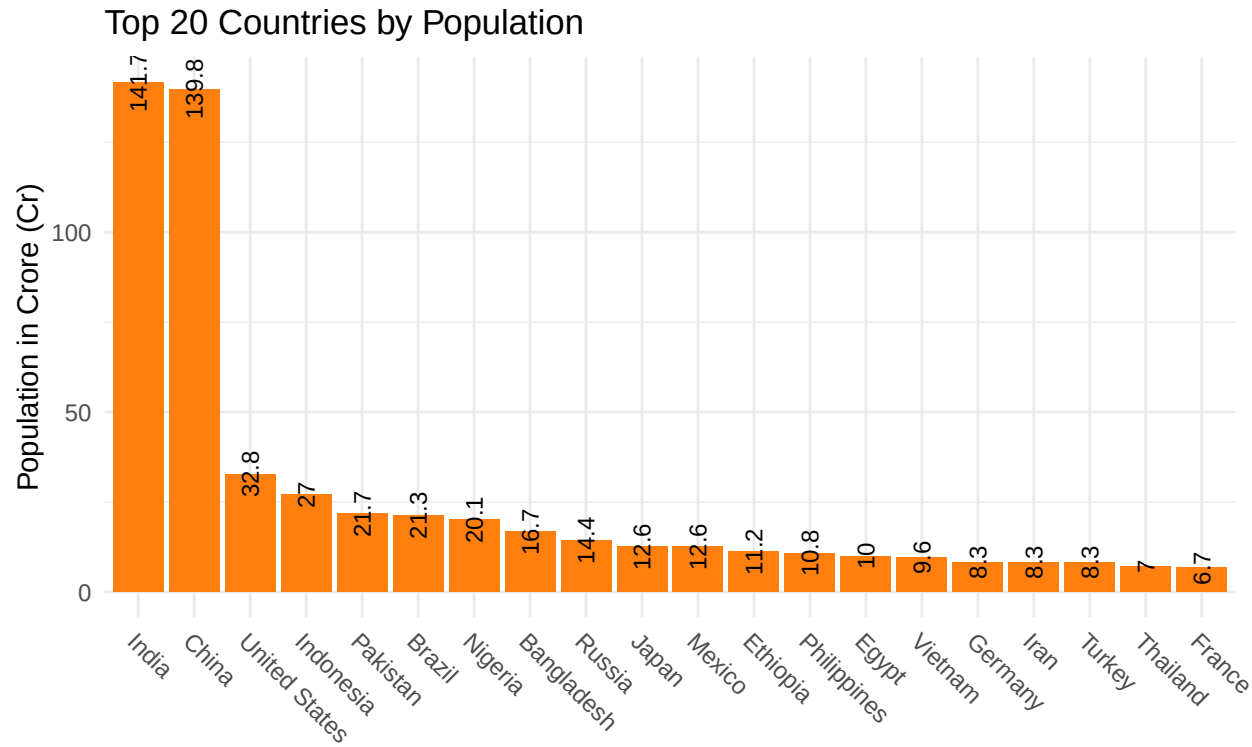
BirthRate : is a demographic measure that quantifies the number of live births in a given population during a specific period (typically per 1,000 people) within that population.

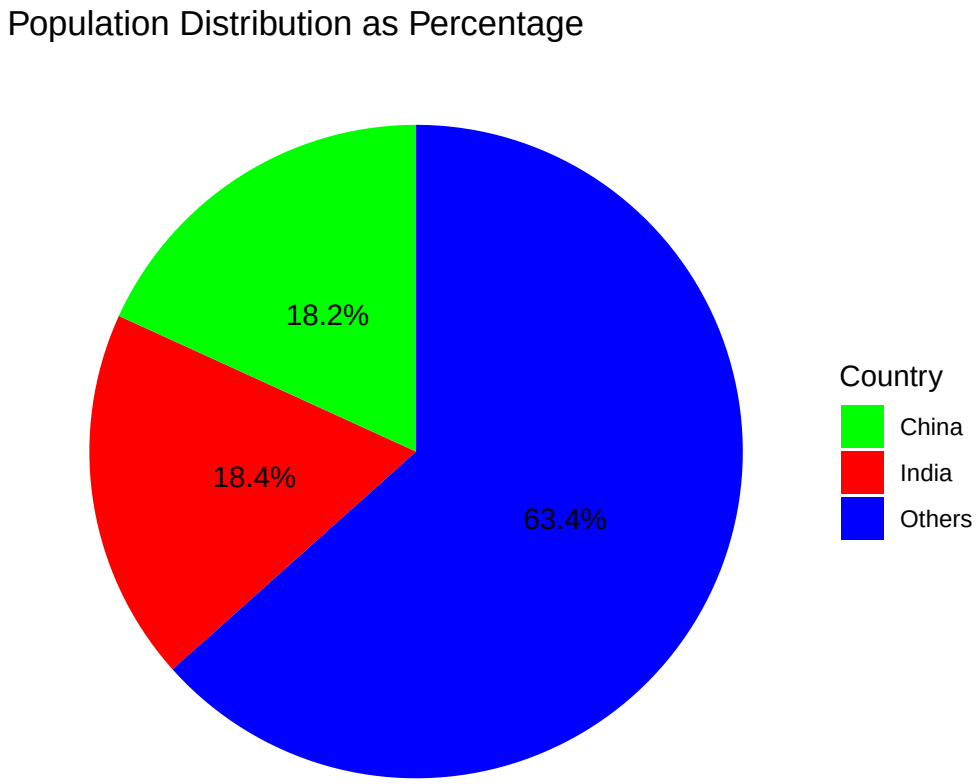
```
[1] "Total Population :  7690485345  # May be some missing Countries"

[1] "Avarage Birth Rate is :  20.1547340425532"

[1] "So we can predict That on the given population  154999687  will increase in a year "

[1] "It will be increased more that 0.15 biilon next year!"
```



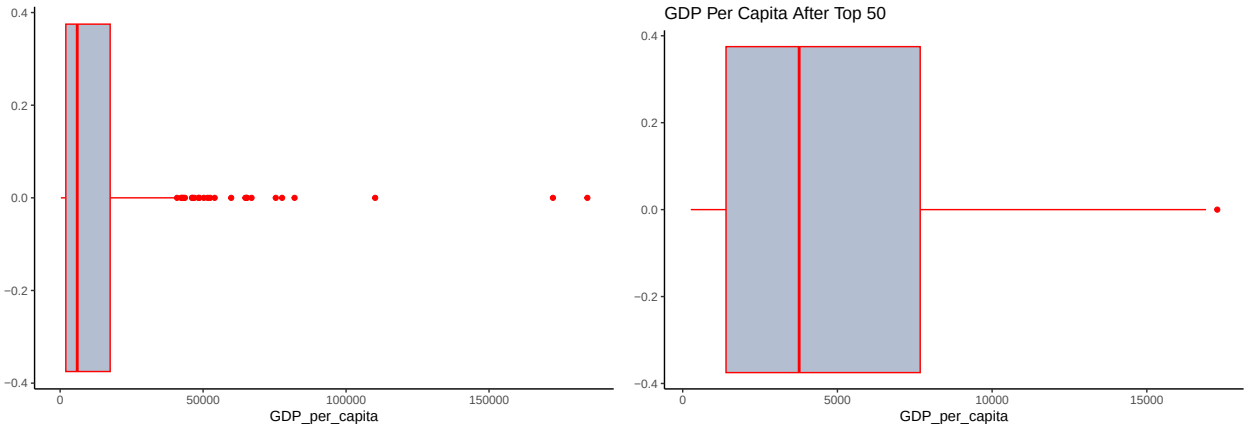


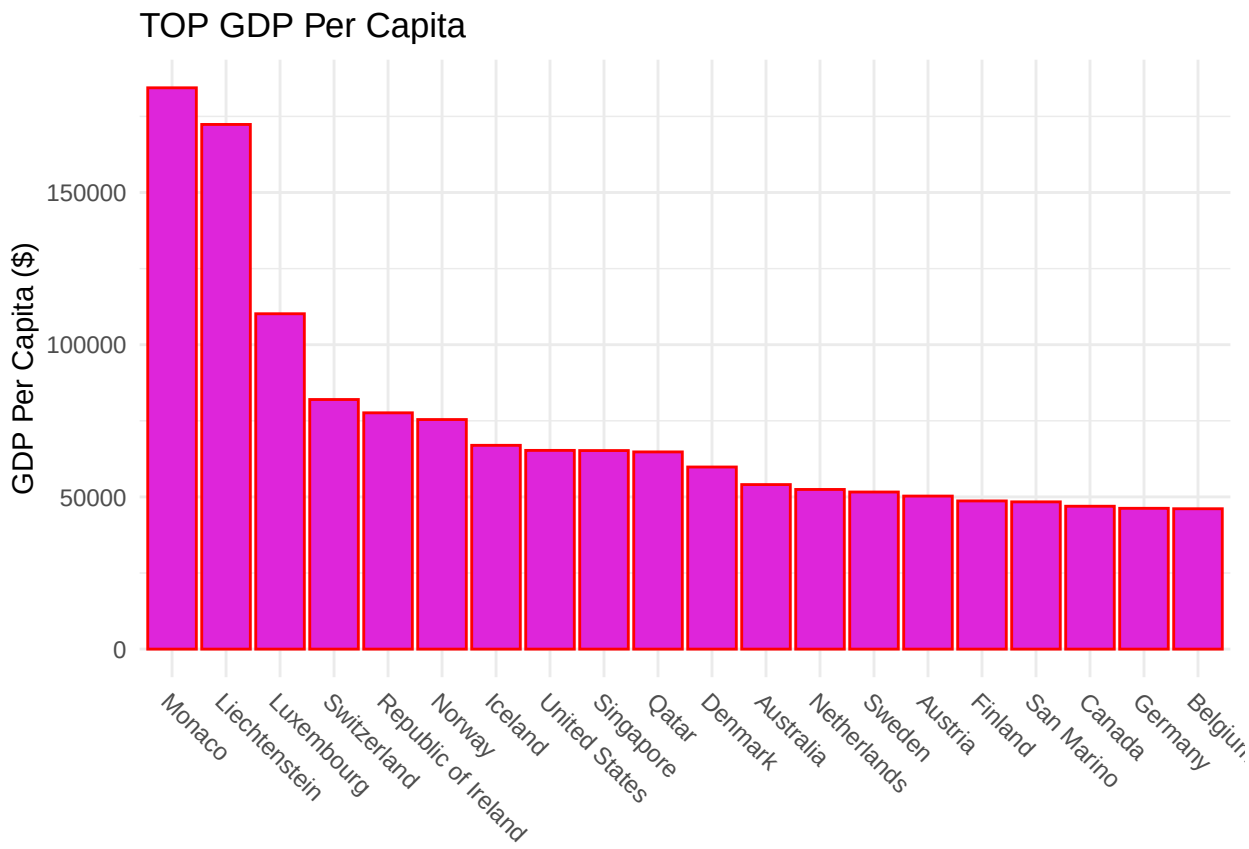
Upon analyzing the plot, it becomes apparent that India faces a challenge in achieving a harmonious equilibrium between GDP and its large population, whereas the United States seems to have effectively managed this balance.

India's population can be a valuable asset, but it also requires careful management and planning to ensure a high quality of life for its citizens and sustainable development for the country as a whole.

GDP Per Capita:

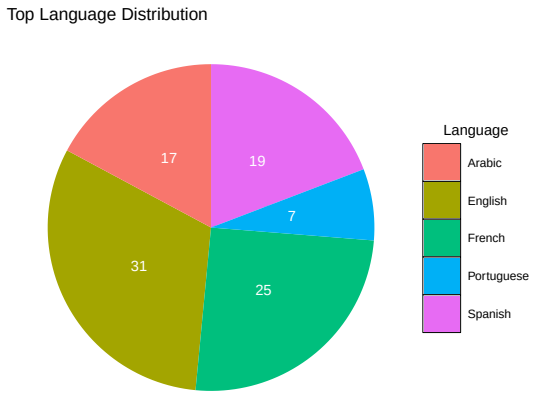
GDP per capita: Gross Domestic Product (GDP) per person, a measure of the average income or economic output per individual in a population.

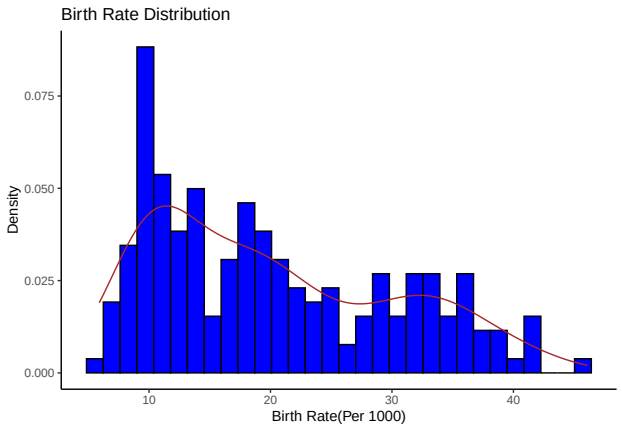




It appears that the GDP per capita is influenced by outliers, but beyond a certain point (among the top countries), there is a more even distribution.

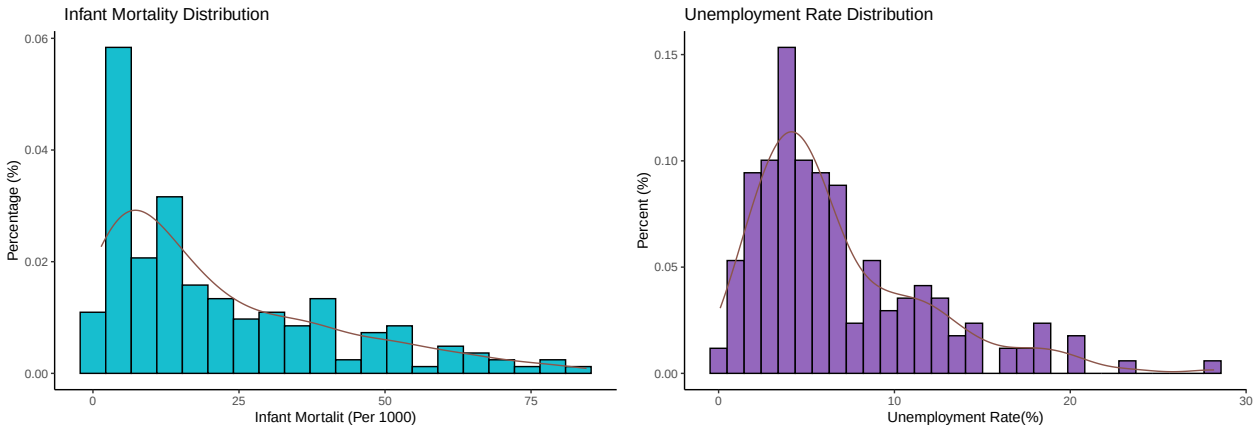
Some other Features distributions :





The pie chart strongly indicates that English and French are prevalent languages on a global scale.

The histogram illustrates that the average birth rate falls within the range of 10 to 20 persons per 1000.

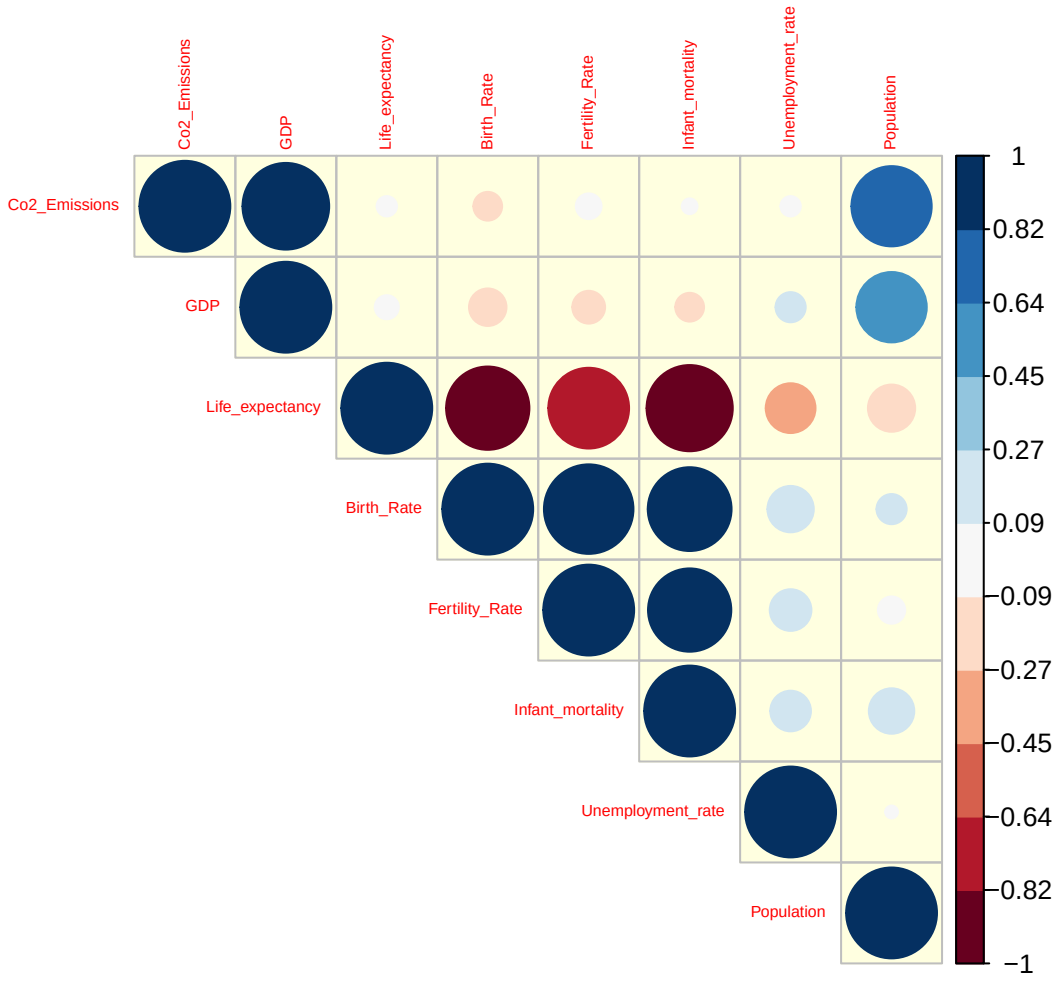


~The plot suggests that, on average, there are approximately 10 infant deaths per 1000 births.
~By calculation the mean of the unemployment rate and multiply it by the population data, it appears that approximately 24% of the population is unemployed.

Correlation Among Key Features of the Top 50 Countries:

Table 2: Correlation Matrix

	Po.	GDP	Un_Rt	Bir_Rt	Co2_Emi	Fer_Rt	Life_Ex	Infnt_Mor
Population	1.00	0.60	-0.02	0.11	0.78	0.09	-0.27	0.25
GDP	0.60	1.00	0.11	-0.17	0.91	-0.13	0.07	-0.10
Unemployment_rate	-0.02	0.11	1.00	0.26	0.05	0.21	-0.30	0.20
Birth_Rate	0.11	-0.17	0.26	1.00	-0.10	0.97	-0.84	0.85
Co2_Emissions	0.78	0.91	0.05	-0.10	1.00	-0.08	-0.05	-0.03
Fertility_Rate	0.09	-0.13	0.21	0.97	-0.08	1.00	-0.79	0.84
Life_expectancy	-0.27	0.07	-0.30	-0.84	-0.05	-0.79	1.00	-0.90
Infant_mortality	0.25	-0.10	0.20	0.85	-0.03	0.84	-0.90	1.00

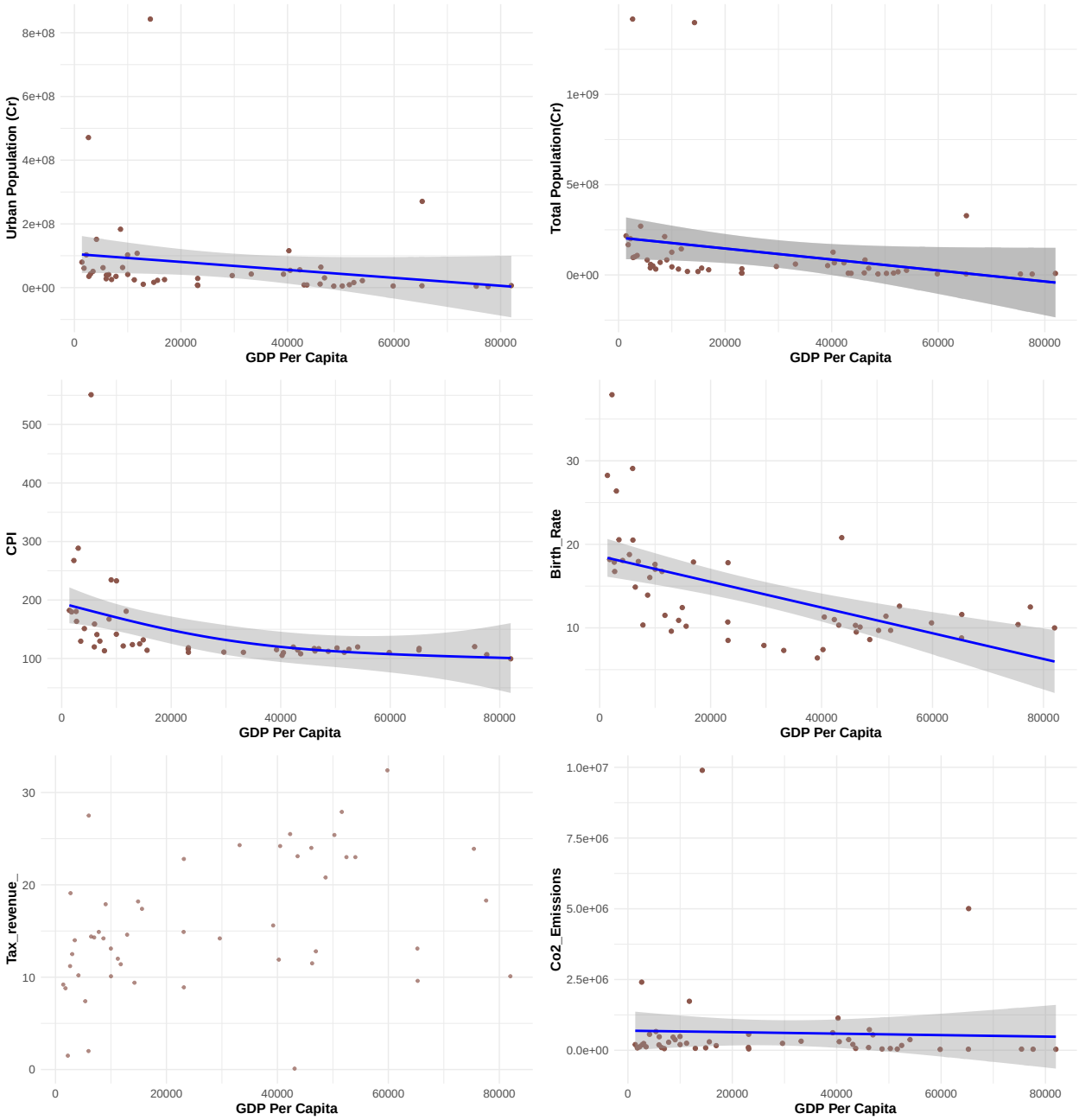


Observations :

- Based on the heat map and our observations, it appears that GDP is positively linked to both population size and is notably associated with CO2 emissions .
- Life expectancy is significantly linked to birth rate, fertility rate, and infant mortality rate, showing a strong positive correlation with these demographic indicators.
- The increase in population is a major factor contributing to higher CO2 emissions. More people mean more carbon dioxide being released into the atmosphere.
- Unemployment rates does not exhibit a strong correlation with any specific element or factor.
- The unemployment rate is not directly tied to the size of the population.
- Life expectancy is a key factor that demonstrates a strong correlation with various other features or indicators.

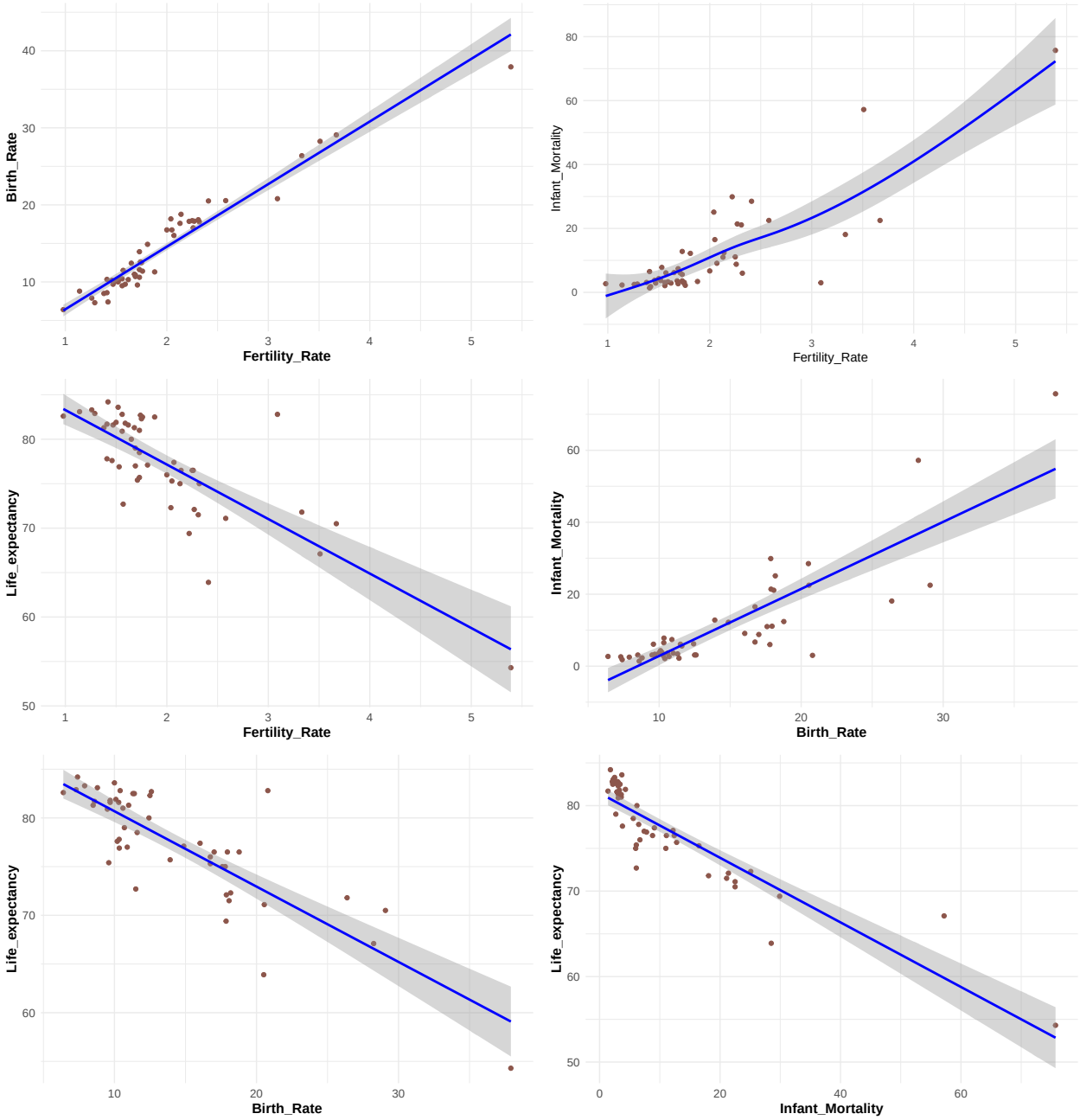
Trends Analysis :

The Relationships between various features and GDP



- GDP per capita is positively associated with CO2 emissions, indicating that wealthier countries tend to have higher carbon emissions due to increased industrialization and energy consumption.
- GDP per capita does not consistently correlate with tax revenue.

Highly Correlated Features :



• The scatter plots with linear fitting clearly demonstrate a positive correlation between Fertility Rate, Infant Mortality, and Birth Rate. This means that as one of these variables increases, the others also tend to increase, indicating a strong association among these key demographic factors

• Whereas Life expectancy is inversely proportional to Fertility Rate, Infant Mortality, and Birth Rate, meaning that as any of these factors increase, life expectancy tends to decreas.

4 Conclusions :

Based on the analysis and findings presented in the report, several conclusions can be drawn regarding the key features and trends in global country statistics:

1. **GDP Significance**: The global economic landscape is heavily influenced by a handful of countries, with the United States and China standing out as major contributors to the world's Gross Domestic Product (GDP). This concentration of economic power among a few nations has a substantial impact on the global economy.

2. **Population Challenges**: India's large population presents a unique challenge in balancing economic growth (GDP) with the need for effective population management. The significant increase in the average birth rate underscores the importance of policies aimed at addressing population growth

3. **Language and Birth Rates**: English and French emerge as prominent languages globally. Birth rate distribution reveals that many countries fall within a range of 10 to 20 births per 1,000 people, indicating a commonality in demographic trends.

4. **Unemployment and Infant Mortality**:The relatively high unemployment rate, at approximately 24%, underscores the global challenge of providing employment opportunities for populations. The unemployment rate is not necessarily influenced by the overall population size.Infant mortality rates, with an average of around 10 deaths per 1,000 births, reflect varying healthcare qualities across countries.

5. **Correlation Insights**: The correlation matrix analysis reveals several important relationships among key features. Notably, GDP is positively correlated with population size and CO2 emissions, suggesting that larger economies tend to produce more emissions. Life expectancy is strongly linked to birth rate, fertility rate, and infant mortality, highlighting the importance of healthcare and family planning in improving life expectancy.

6. **Highly Correlated Features**: Relationships between highly correlated features, such as fertility rate and birth rate, emphasize the importance of family planning and healthcare in managing population growth and reducing infant mortality.

In conclusion, this report offers valuable insights into the complex interplay of economic, demographic, and social factors on a global scale. It highlights the dominance of a few countries in shaping the world economy, the challenges posed by large populations, and the significance of healthcare and family planning in improving overall well-being. Researchers and analysts can leverage these findings to better understand and address global trends and dynamics.