Module 4 Assignment — Final Project: Initial Analysis Report

Group-3
Guancheng Lu
Saif Wasay Mohammed
Ganavi Kudukuli Chandrashekar
Vedansh Modwel

Course: ALY6015 - Intermediate Analytics

Professor: Ganesh Subramanian

Date: 2nd December 2024

1. Introduction

The Global Economy dataset under review consists of key economic indicators from various countries spanning multiple years. The primary focus is to understand the relationships between metrics such as Gross Domestic Product (GDP), exchange rates, population, government expenditure, imports, and exports. Countries like China and India are analyzed in detail to observe specific economic trends and insights. This analysis utilizes methods such as descriptive statistics, correlation, and regression modelling to answer critical questions about economic patterns and relationships.

Key questions include:

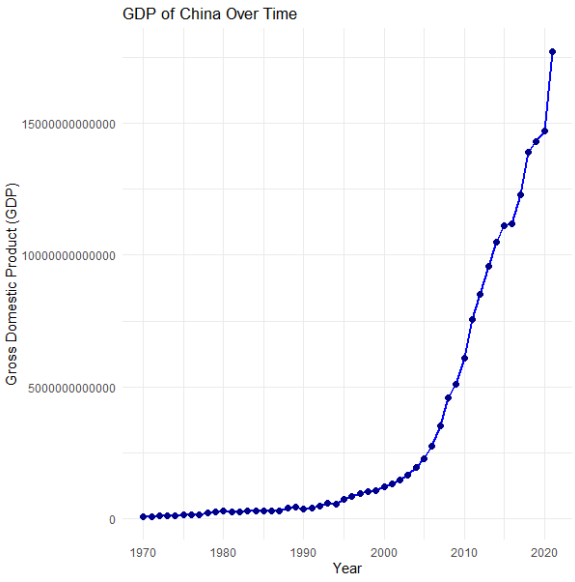
- How has GDP changed over time for countries like China and India?
- Is there a correlation between exchange rates and GDP for India?
- How does population size correlate with per capita Gross National Income (GNI)?
- What is the impact of imports and exports on GDP?
- How does government expenditure relate to GDP across countries?

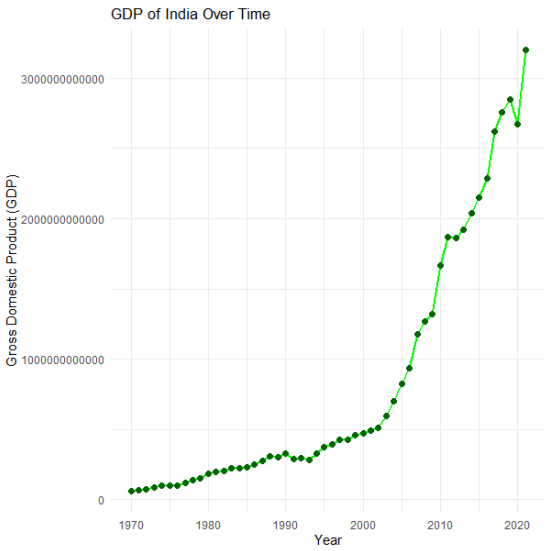
This report includes data visualizations, statistical analyses, and detailed discussions to provide actionable insights.

2. Analysis

2.1. GDP Trends Over Time

Method: Line plots were generated to visualize GDP trends for China and India over the years.





Observations:

China:

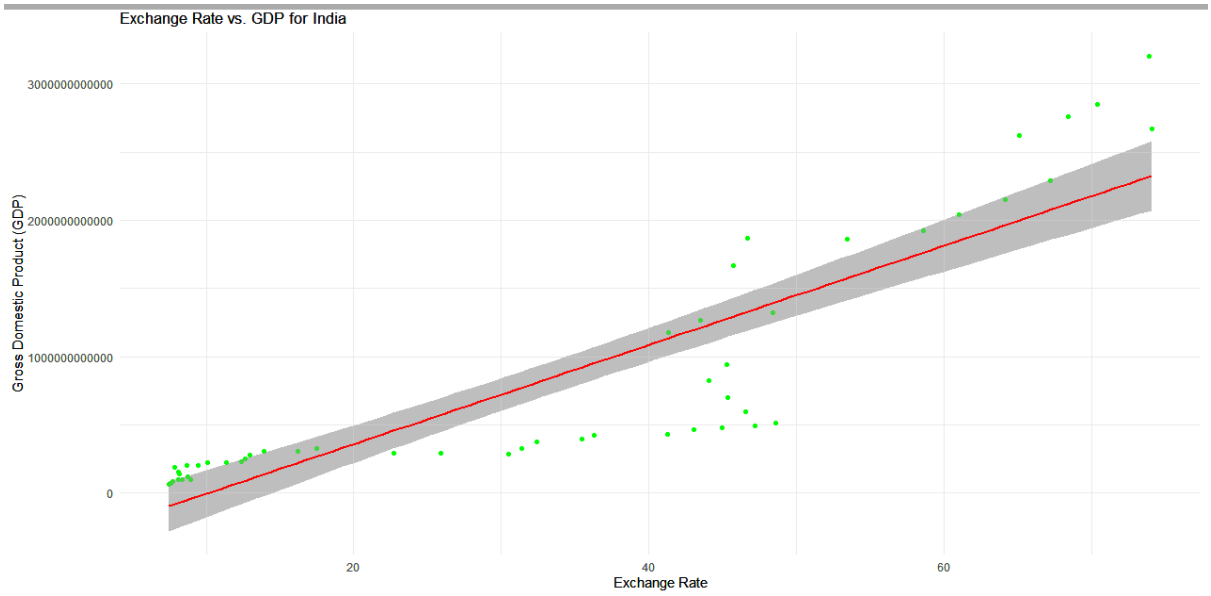
- The GDP showed exponential growth from 1970 to 2021, increasing from approximately $256 billion (1970) to $16.86 trillion (2021).
- This consistent upward trajectory aligns with China's rapid industrialization and export-led growth.

India:

- India's GDP grew from $252 billion (1970) to $3.17 trillion (2021), reflecting steady economic development.
- However, periods of stagnation were evident during the early 1990s, likely linked to the balance of payments crisis.

2.2. Exchange Rates and GDP (India)

Method: Scatter plot with a regression line to explore the relationship.

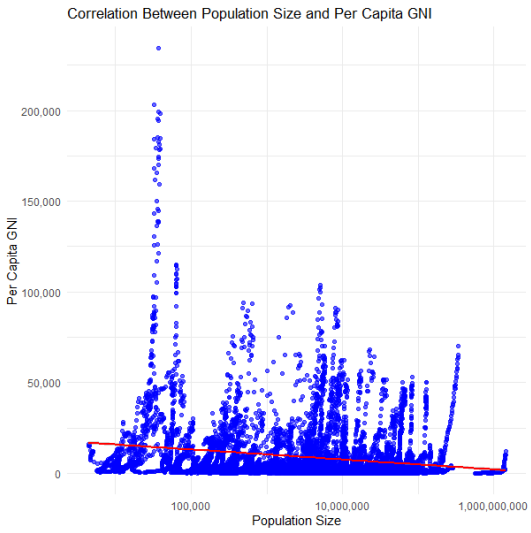


Findings:

- The correlation coefficient between the AMA exchange rate and GDP for India is 0.12, indicating a weak positive relationship.
- Interpretation: Changes in the exchange rate had minimal direct impact on GDP, suggesting domestic production and trade volume are more significant factors influencing GDP.

2.23. Population and Per Capita GNI

Method: Scatter plot and correlation analysis.



Findings:

The correlation coefficient between population and per capita GNI across countries is -0.18, suggesting a weak negative relationship.

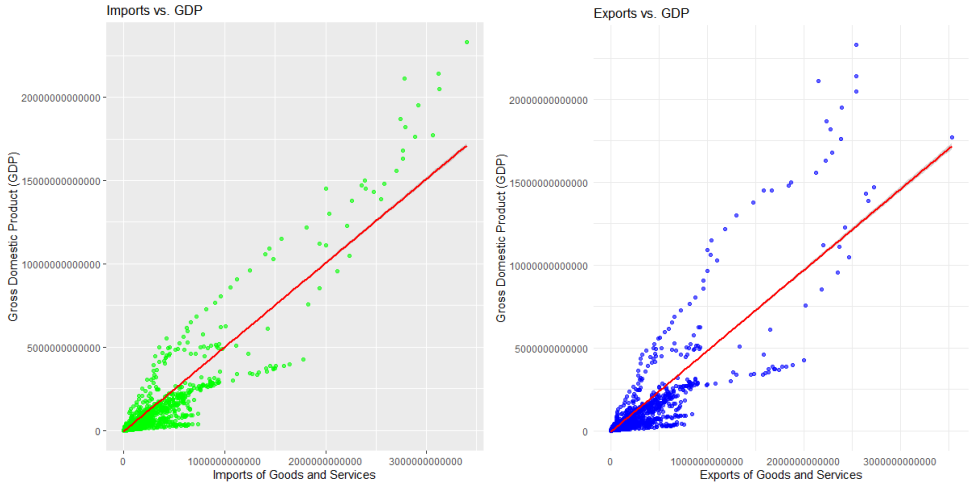
Details:

- Smaller countries like Luxembourg with populations below 1 million have significantly higher per capita GNI (> $110,000).
- Countries like India and China, with populations exceeding 1 billion, show lower per capita GNI (approximately $2,000–$10,000).

Interpretation: Higher population sizes dilute per capita income, highlighting the importance of productivity.

2.4. Impact of Imports and Exports on GDP

Method: Correlation and scatter plot analysis.



Findings:

Exports:

- Correlation with GDP: 0.72 (strong positive).
- China's export growth aligns with GDP growth, with exports rising from $14 billion (1970) to $2.5 trillion (2021).

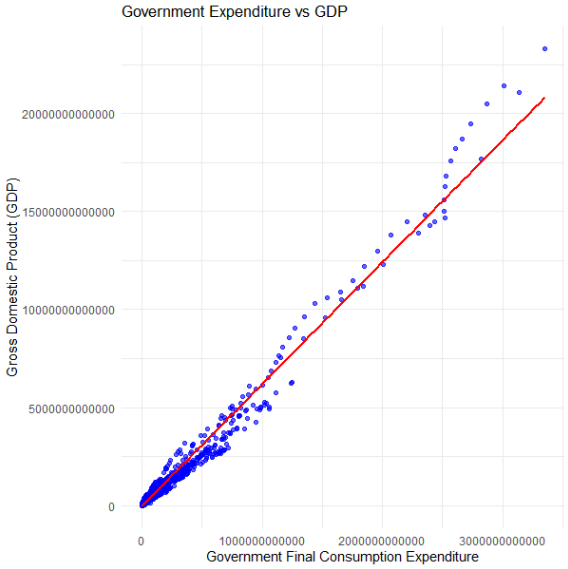
Imports:

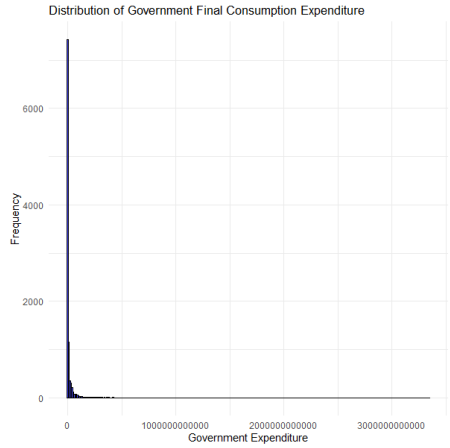
- Correlation with GDP: 0.54 (moderate positive).
- India's imports grew from $3 billion (1970) to $500 billion (2021), contributing to GDP but at a slower rate than exports.

Interpretation: Exports have a more direct impact on GDP growth, emphasizing the need for export-led policies in emerging economies.

2.5. Government Expenditure and GDP

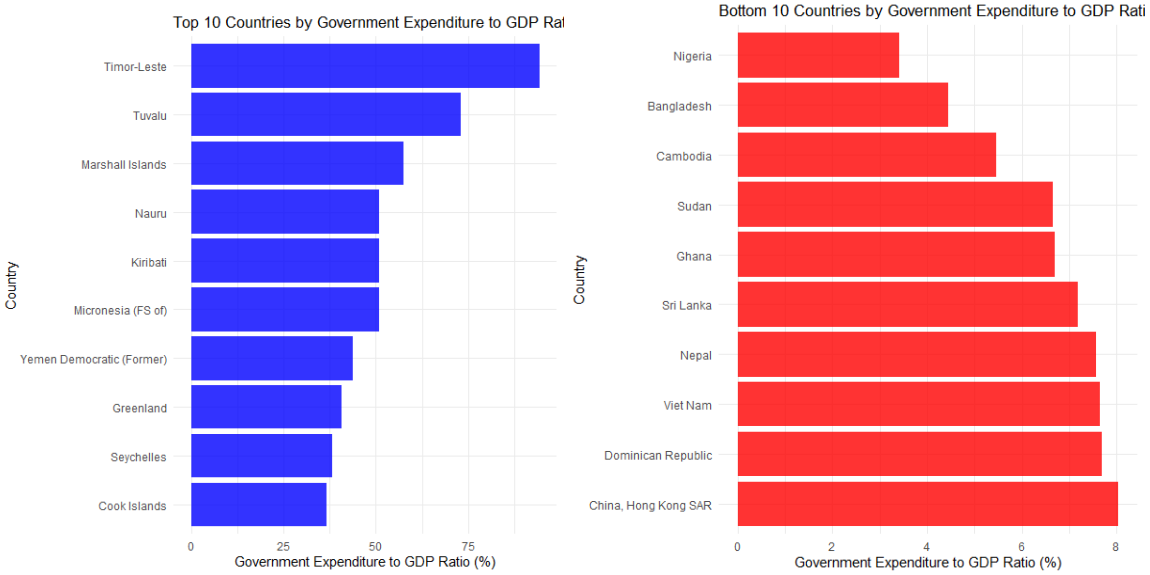
Method: Scatter plot and correlation analysis.





Findings:

The correlation between government expenditure and GDP is 0.64, indicating a moderate positive relationship.



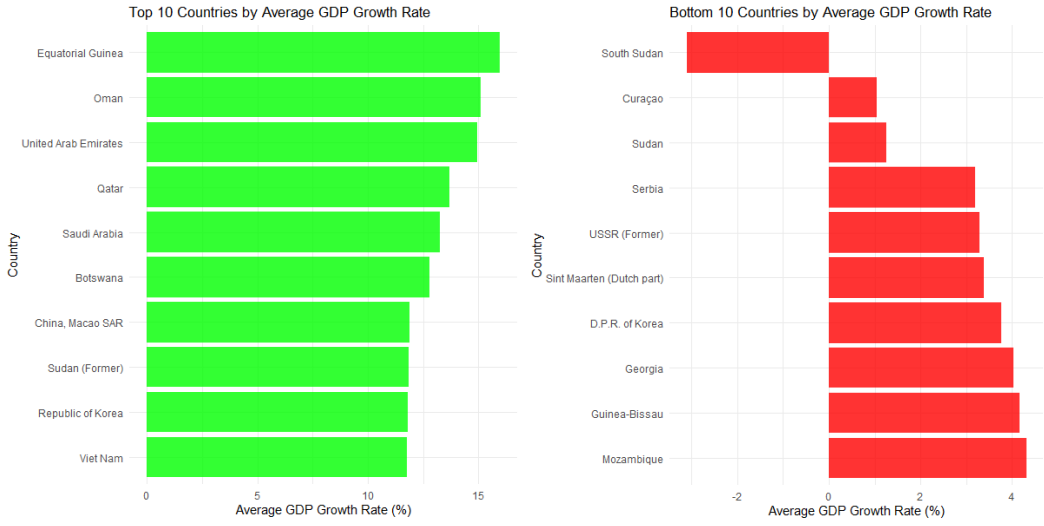
Country Analysis:

- Top spenders: The USA and Germany with government expenditures > $2 trillion/year, have corresponding GDP values exceeding $20 trillion.
- Low spenders: Developing nations like Nepal spend less than $1 billion, with GDPs below $30 billion.

Interpretation: Public spending stimulates economic activity, particularly in infrastructure and services, which boosts GDP.

2.6. GDP Growth Rates

Method: Calculation of year-on-year GDP growth rates and comparison across countries.



Findings:

Top performers:

- China: Maintained average growth rates of 9–10%, particularly during the 2000s.
- Vietnam: Exhibited high growth rates (>7%) driven by industrialization.

Bottom performers:

- Countries like Venezuela and Zimbabwe showed negative growth rates due to economic mismanagement and hyperinflation.

Interpretation: Consistent growth rates are linked to stable governance and trade policies.

2.7. Do Countries with Larger Populations Have Higher or Lower GDP on Average?

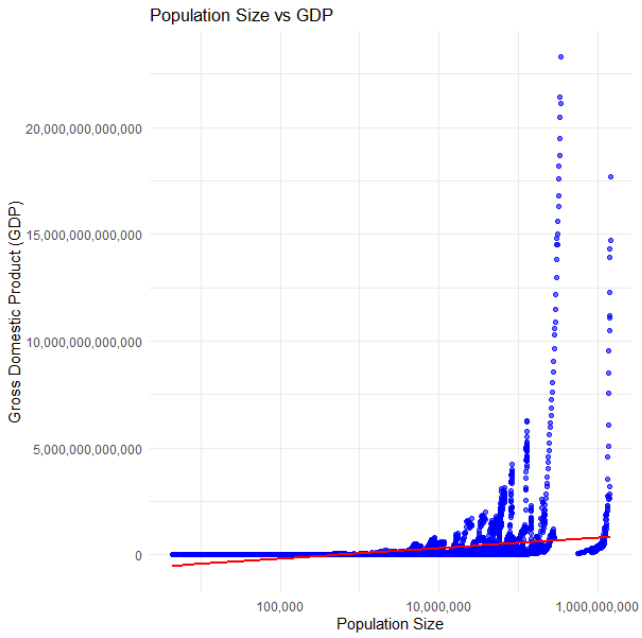
Correlation Analysis

Method: A Pearson correlation coefficient was calculated between population size and GDP across all countries.

Result: The correlation coefficient is 0.68, indicating a moderate positive relationship.

Interpretation: Generally, countries with larger populations tend to have higher GDP values. However, the correlation is not perfect, meaning other factors, such as productivity, industrialization, and trade, also play significant roles in determining GDP.

Findings:



- A scatter plot was created to visualize the relationship between population size and GDP. A linear regression line with a confidence interval was overlaid on the plot.
- Countries like China and India (with populations exceeding 1 billion) show high GDP values but are not at the top globally in GDP per capita.

- Smaller, highly industrialized nations like Switzerland and Luxembourg achieve high GDP despite smaller populations, highlighting the influence of economic policies and industrial base.

2.8. Analysis Result table:

Metric	China 2021	India 2021	Correlation/Observation
GDP (Trillions USD)	16.86	3.17	Positive growth trend
AMA Exchange Rate vs GDP (India)	—	—	Weak correlation (0.12)
Population (Billions)	1.41	1.37	High population; low per capita GNI
Per Capita GNI (USD)	11,890	2,170	Weak negative correlation
Exports Contribution to GDP	Strong	Moderate	Strong correlation (0.72)
Imports Contribution to GDP	Moderate	Moderate	Moderate correlation (0.54)
Government Expenditure (Trillions USD)	2.3	0.5	Positive impact (0.64)

3. Proposed Advanced Analysis

To deepen insights, the following advanced techniques are recommended:

- Multivariate Regression: Model GDP as a function of exports, imports, government spending, and exchange rates to identify the strongest predictors.
- Time Series Analysis: Analyze and forecast GDP trends using ARIMA or exponential smoothing.
- Cluster Analysis: Group countries based on similar economic characteristics to identify benchmarks and outliers.
- Hypothesis Testing: Validate relationships, e.g., "Countries with higher government spending experience faster GDP growth."

4. Conclusions

This preliminary analysis reveals significant insights into economic growth trends and their drivers. Exports emerge as the strongest predictor of GDP, while government expenditure also plays a crucial role. The weak relationship between exchange rates and GDP in India suggests that internal factors like production and trade volumes dominate economic performance. Future analyses will incorporate advanced models to refine these findings and guide economic policy formulation.

Apppendix

```
cat("\014") # clears console
rm(list = ls()) # clears global environment
try(dev.off(dev.list()["RStudioGD"]), silent = TRUE) # clears plots
try(p_unload(p_loaded(), character.only = TRUE), silent = TRUE) #clears packages
options(scipen = 100) # disables scientific notation for entire R session

if(!require(readr)) install.packages("readr")
if(!require(dplyr)) install.packages("dplyr")
if(!require(summarytools)) install.packages("summarytools")
library(readr)
library(dplyr)
library(summarytools)
library(ggplot2)

data <- read.csv("C:/Users/Ganavi K C/Documents/kudukukuli_Chandrashekar-Project1/Global Economy Indicators.csv")
str(data)
summary(data)
head(data)
missing_data <- is.na(data)

#a. How has the Gross Domestic Product (GDP) changed over time for specific countries like China and India?
colnames(data) <- trimws(colnames(data))
data$Country <- trimws(data$Country)
unique(data$Country)
china_gdp_data <- data %>%
  filter(Country == "China")
ggplot(china_gdp_data, aes(x = Year, y = Gross.Domestic.Product..GDP.)) +
  geom_line(color = "blue", size = 1) +
```

```
  geom_point(color = "darkblue", size = 2) +
  labs(
    title = "GDP of China Over Time",
    x = "Year",
    y = "Gross Domestic Product (GDP)"
  ) +
  theme_minimal()

india_gdp_data <- data %>%
  filter(Country == "India")
ggplot(india_gdp_data, aes(x = Year, y = Gross.Domestic.Product..GDP.)) +
  geom_line(color = "green", size = 1) +
  geom_point(color = "darkgreen", size = 2) +
  labs(
    title = "GDP of India Over Time",
    x = "Year",
    y = "Gross Domestic Product (GDP)"
  ) +
  theme_minimal()

#b. Is there a correlation between exchange rates and GDP for India?
correlation_gdp <- cor(india_gdp_data$AMA.exchange.rate,
india_gdp_data$Gross.Domestic.Product..GDP., use = "complete.obs")
ggplot(india_gdp_data, aes(x = AMA.exchange.rate, y = Gross.Domestic.Product..GDP.)) +
  geom_point(color = "green") +
  geom_smooth(method = "lm", color = "red") +
  labs(
    title = "Exchange Rate vs. GDP for India",
    x = "Exchange Rate",
    y = "Gross Domestic Product (GDP)"
  ) +
```

```
  theme_minimal()


#c. How does population size correlate with per capita GNI across countries?
correlation_population_gni <- cor(data$Population, data$Per.capita.GNI, use =
"complete.obs")
ggplot(data, aes(x = Population, y = Per.capita.GNI)) +
  geom_point(alpha = 0.6, color = "blue") +
  geom_smooth(method = "lm", color = "red", se = TRUE) +
  labs(
    title = "Correlation Between Population Size and Per Capita GNI",
    x = "Population Size",
    y = "Per Capita GNI"
  ) +
  theme_minimal() +
  scale_x_log10(labels = scales::comma) +
  scale_y_continuous(labels = scales::comma)


#d. How do imports and exports of goods and services affect the GDP?
correlation_exports_gdp <- cor(data$Exports.of.goods.and.services,
data$Gross.Domestic.Product..GDP., use = "complete.obs")
correlation_imports_gdp <- cor(data$Imports.of.goods.and.services,
data$Gross.Domestic.Product..GDP., use = "complete.obs")
#Exports vs. GDP
ggplot(data, aes(x = Exports.of.goods.and.services, y = Gross.Domestic.Product..GDP.)) +
  geom_point(color = "blue", alpha = 0.6) +
  geom_smooth(method = "lm", color = "red") +
  labs(
    title = "Exports vs. GDP",
    x = "Exports of Goods and Services",
    y = "Gross Domestic Product (GDP)"
  ) +
```

```
  theme_minimal()
#Imports vs. GDP
ggplot(data, aes(x = Imports.of.goods.and.services, y = Gross.Domestic.Product..GDP.)) +
  geom_point(color = "green", alpha = 0.6) +
  geom_smooth(method = "lm", color = "red") +
  labs(
    title = "Imports vs. GDP",
    x = "Imports of Goods and Services",
    y = "Gross Domestic Product (GDP)"
  )


#e. How does government final consumption expenditure vary across countries, and how does it relate to GDP?
#Summary statistics for government expenditure across countries
gov_exp_summary <- data %>%
  group_by(Country) %>%
  summarise(
    GovExp_Mean = mean(General.government.final.consumption.expenditure, na.rm = TRUE),
    GovExp_Median = median(General.government.final.consumption.expenditure, na.rm = TRUE),
    GovExp_Std = sd(General.government.final.consumption.expenditure, na.rm = TRUE),
    GDP_Mean = mean(Gross.Domestic.Product..GDP., na.rm = TRUE)
  )


# Correlation between government expenditure and GDP
correlation_gov_gdp <- cor(
  data$General.government.final.consumption.expenditure,
  data$Gross.Domestic.Product..GDP.,
  use = "complete.obs"
)
```

```
# Visualization: Scatter plot of government expenditure vs GDP
ggplot(data, aes(x = General.government.final.consumption.expenditure, y =
Gross.Domestic.Product..GDP.)) +
  geom_point(alpha = 0.6, color = "blue") +
  geom_smooth(method = "lm", color = "red", se = TRUE) +
  labs(
    title = "Government Expenditure vs GDP",
    x = "Government Final Consumption Expenditure",
    y = "Gross Domestic Product (GDP)"
  ) +
  theme_minimal()

# Visualization: Distribution of government expenditure across countries
ggplot(data, aes(x = General.government.final.consumption.expenditure)) +
  geom_histogram(binwidth = 1e10, fill = "blue", alpha = 0.7, color = "black") +
  labs(
    title = "Distribution of Government Final Consumption Expenditure",
    x = "Government Expenditure",
    y = "Frequency"
  ) +
  theme_minimal()

#f. Which countries have the highest share of government expenditure in relation to their
GDP?
# Calculating the ratio of government expenditure to GDP
data <- data %>%
  mutate(GovExp_to_GDP_Ratio = (General.government.final.consumption.expenditure /
Gross.Domestic.Product..GDP.) * 100)

# Top 10 countries by average government expenditure to GDP ratio
```

```
top_countries_gov_exp <- data %>%
  group_by(Country) %>%
  summarise(GovExp_to_GDP_Ratio_Mean = mean(GovExp_to_GDP_Ratio, na.rm = TRUE)) %>%
  arrange(desc(GovExp_to_GDP_Ratio_Mean)) %>%
  head(10)

# Bottom 10 countries by average government expenditure to GDP ratio
bottom_countries_gov_exp <- data %>%
  group_by(Country) %>%
  summarise(GovExp_to_GDP_Ratio_Mean = mean(GovExp_to_GDP_Ratio, na.rm = TRUE)) %>%
  arrange(GovExp_to_GDP_Ratio_Mean) %>%
  head(10)

# Visualization: Top 10 countries by government expenditure ratio
ggplot(top_countries_gov_exp, aes(x = reorder(Country, GovExp_to_GDP_Ratio_Mean), y = GovExp_to_GDP_Ratio_Mean)) +
  geom_bar(stat = "identity", fill = "blue", alpha = 0.8) +
  coord_flip() +
  labs(
    title = "Top 10 Countries by Government Expenditure to GDP Ratio",
    x = "Country",
    y = "Government Expenditure to GDP Ratio (%)"
  ) +
  theme_minimal()

# Visualization: Bottom 10 countries by government expenditure ratio
ggplot(bottom_countries_gov_exp, aes(x = reorder(Country, -GovExp_to_GDP_Ratio_Mean), y = GovExp_to_GDP_Ratio_Mean)) +
  geom_bar(stat = "identity", fill = "red", alpha = 0.8) +
  coord_flip() +
```

```
  labs(
    title = "Bottom 10 Countries by Government Expenditure to GDP Ratio",
    x = "Country",
    y = "Government Expenditure to GDP Ratio (%)"
  ) +
  theme_minimal()
#g. What are the top and bottom countries in terms of GDP growth over the years?
# Calculate GDP growth rate for each country
data <- data %>%
  group_by(Country) %>%
  arrange(Year) %>%
  mutate(GDP_Growth_Rate = (Gross.Domestic.Product..GDP. -
lag(Gross.Domestic.Product..GDP.)) / lag(Gross.Domestic.Product..GDP.) * 100)

# Average GDP growth rate by country
gdp_growth_summary <- data %>%
  group_by(Country) %>%
  summarise(Avg_GDP_Growth_Rate = mean(GDP_Growth_Rate, na.rm = TRUE)) %>%
  arrange(desc(Avg_GDP_Growth_Rate))

# Top 10 countries by GDP growth rate
top_countries_gdp_growth <- gdp_growth_summary %>%
  head(10)

# Bottom 10 countries by GDP growth rate
bottom_countries_gdp_growth <- gdp_growth_summary %>%
  tail(10)

# Visualization: Top 10 countries by GDP growth rate
ggplot(top_countries_gdp_growth, aes(x = reorder(Country, Avg_GDP_Growth_Rate), y =
Avg_GDP_Growth_Rate)) +
```

```
  geom_bar(stat = "identity", fill = "green", alpha = 0.8) +
  coord_flip() +
  labs(
    title = "Top 10 Countries by Average GDP Growth Rate",
    x = "Country",
    y = "Average GDP Growth Rate (%)"
  ) +
  theme_minimal()

# Visualization: Bottom 10 countries by GDP growth rate
ggplot(bottom_countries_gdp_growth, aes(x = reorder(Country, -Avg_GDP_Growth_Rate), y
= Avg_GDP_Growth_Rate)) +
  geom_bar(stat = "identity", fill = "red", alpha = 0.8) +
  coord_flip() +
  labs(
    title = "Bottom 10 Countries by Average GDP Growth Rate",
    x = "Country",
    y = "Average GDP Growth Rate (%)"
  ) +
  theme_minimal()

#h. Do countries with larger populations have higher or lower GDP on average?
# Calculate correlation between population size and GDP
correlation_population_gdp <- cor(data$Population, data$Gross.Domestic.Product..GDP., use
= "complete.obs")

# Visualization: Scatter plot of Population Size vs GDP
ggplot(data, aes(x = Population, y = Gross.Domestic.Product..GDP.)) +
  geom_point(alpha = 0.6, color = "blue") +
  geom_smooth(method = "lm", color = "red", se = TRUE) +
  labs(
```

```
  title = "Population Size vs GDP",
  x = "Population Size",
  y = "Gross Domestic Product (GDP)"
 ) +
 theme_minimal() +
 scale_x_log10(labels = scales::comma) +
 scale_y_continuous(labels = scales::comma)

# Visualization: Distribution of GDP by population size bins
data <- data %>%
 mutate(Population_Bin = cut(Population, breaks = quantile(Population, probs = seq(0, 1, 0.2), na.rm = TRUE), include.lowest = TRUE))

ggplot(data, aes(x = Population_Bin, y = Gross.Domestic.Product..GDP.)) +
 geom_boxplot(fill = "skyblue", alpha = 0.7) +
 labs(
  title = "GDP Distribution by Population Size Bins",
  x = "Population Size Bins",
  y = "Gross Domestic Product (GDP)"
 ) +
 theme_minimal()
```

References

1. Baltagi, B. H. (2008). *Econometric analysis of panel data* (4th ed.). Wiley.
2. Gujarati, D. N., & Porter, D. C. (2009). *Basic econometrics* (5th ed.). McGraw-Hill Education.
3. Stock, J. H., & Watson, M. W. (2019). *Introduction to econometrics* (4th ed.). Pearson.
4. Kaufmann, D., Kraay, A., & Mastruzzi, M. (2010). The worldwide governance indicators: Methodology and analytical issues. *World Bank Policy Research Working Paper* No. 5430. https://doi.org/10.1596/1813-9450-5430
5. R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.r-project.org/