Project 1

Anish Goli

2024-11-05

<style type="text/css"> body p, div, h1, h2, h3, h4, h5 { color: black; font-family: "Modern Computer Roman"; } h1, h2 { color: #8C1D40; } .smaller-text { font-size:0.8em; } </style>

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(plotly)
```

```
##
## Attaching package: 'plotly'
##
## The following object is masked from 'package:ggplot2':
##
##     last_plot
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following object is masked from 'package:graphics':
##
##     layout
```

```
library(ggplot2)
library(dplyr)
library(reshape2)
```

```
##
## Attaching package: 'reshape2'
##
## The following object is masked from 'package:tidyr':
##
##     smiths
```

```r
library(ggthemes)
library(rgl)
library(maps)
```

```
##
## Attaching package: 'maps'
##
## The following object is masked from 'package:purrr':
##
##     map
```

```r
library(viridis)
```

```
## Loading required package: viridisLite
##
## Attaching package: 'viridis'
##
## The following object is masked from 'package:maps':
##
##     unemp
```

Introduction

In this project, we analyze a dataset containing socio-economic indicators from various countries. Using R, we explore the data to identify trends and relationships across different metrics. This study focuses on variables such as population, growth rates, density, and other country-level indicators.

Problem Statement

The objective of this analysis is to answer the following questions:

1. Which countries have the highest and lowest population?
2. How has the population changed over the years?
3. What are the patterns in population density across countries?
4. How do growth rate and net population change correlate?
5. What regional patterns can be identified in socio-economic indicators?

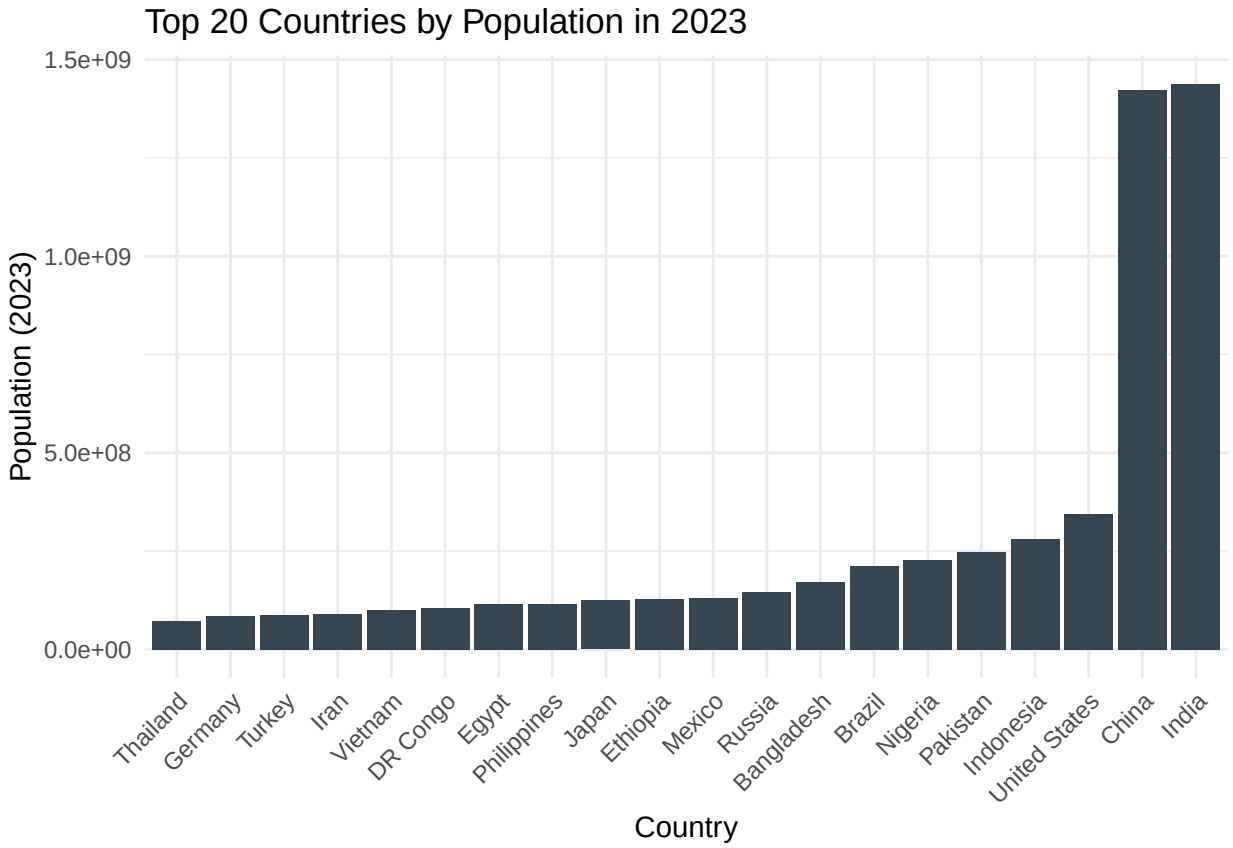
To address these questions, we perform exploratory data analysis and visualize the data using various graph types.

```r
countries_data <- read.csv("countries-table.csv")
countries_data$country <- as.factor(countries_data$country)
```

1. Population in 2023 by Country

```r
top_countries <- countries_data %>%
  arrange(desc(pop2023)) %>%
  head(20)

ggplot(top_countries, aes(x = reorder(country, pop2023), y = pop2023)) +
  geom_bar(stat = "identity", fill = "#36454F") +
```

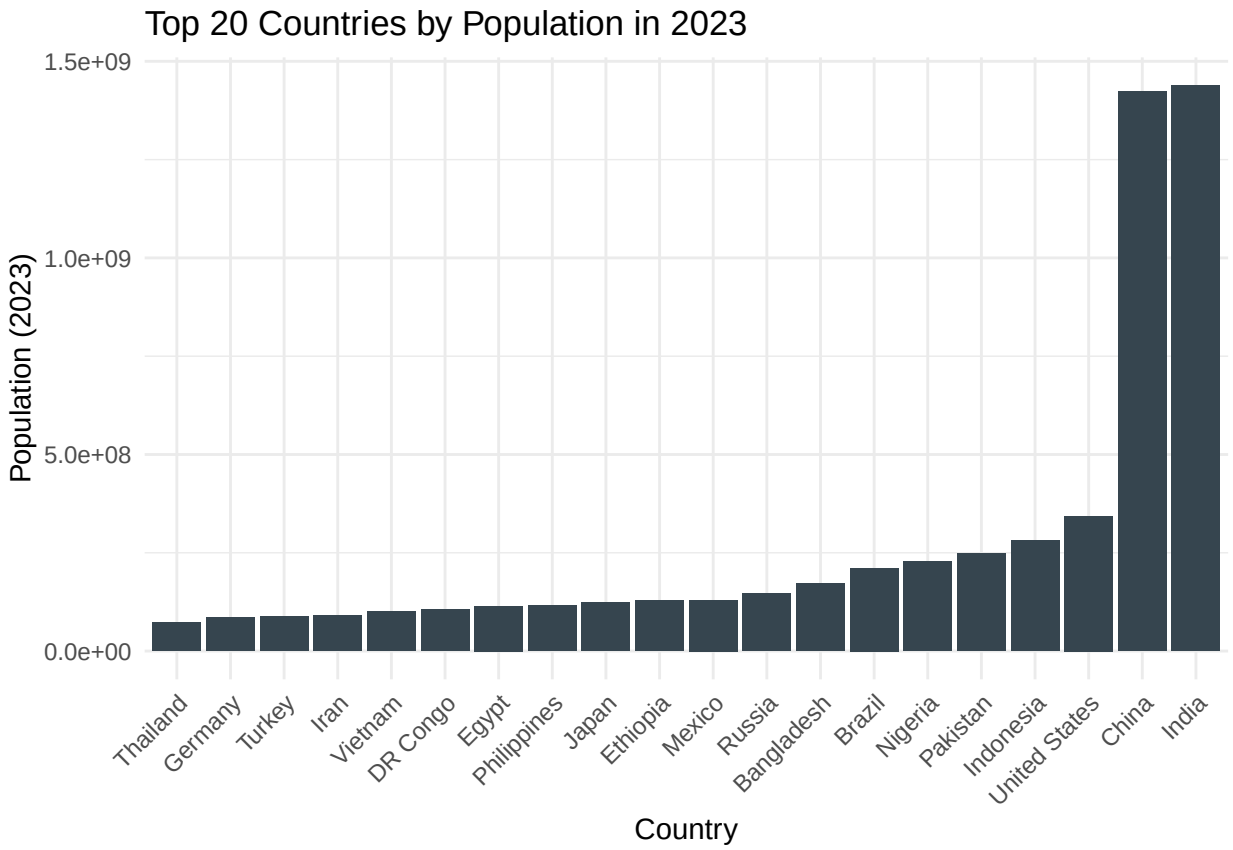
```
  labs(title = "Top 20 Countries by Population in 2023", x = "Country", y = "Population (2023)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



This bar plot shows the top 20 countries by population in 2023. India and China lead with the highest populations, reflecting their significant demographic impact.

2. Population Growth from 1980 to 2023

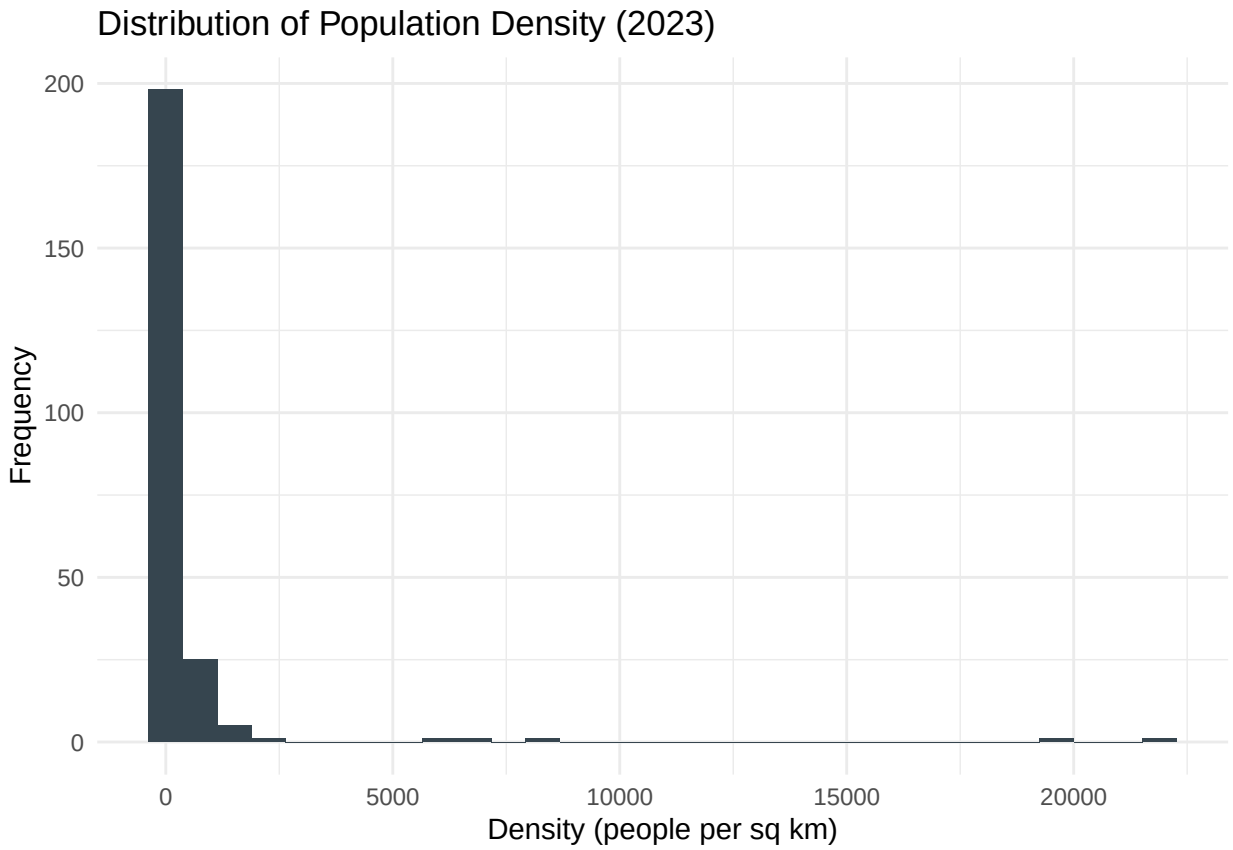
```
top_countries <- countries_data %>%
  arrange(desc(pop2023)) %>%
  head(20)

ggplot(top_countries, aes(x = reorder(country, pop2023), y = pop2023)) +
  geom_bar(stat = "identity", fill = "#36454F") +
  labs(title = "Top 20 Countries by Population in 2023", x = "Country", y = "Population (2023)") +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



The line plot illustrates population trends over time for the top 10 countries. Most countries show a steady increase, highlighting global population growth.

3. Distribution of Population Density

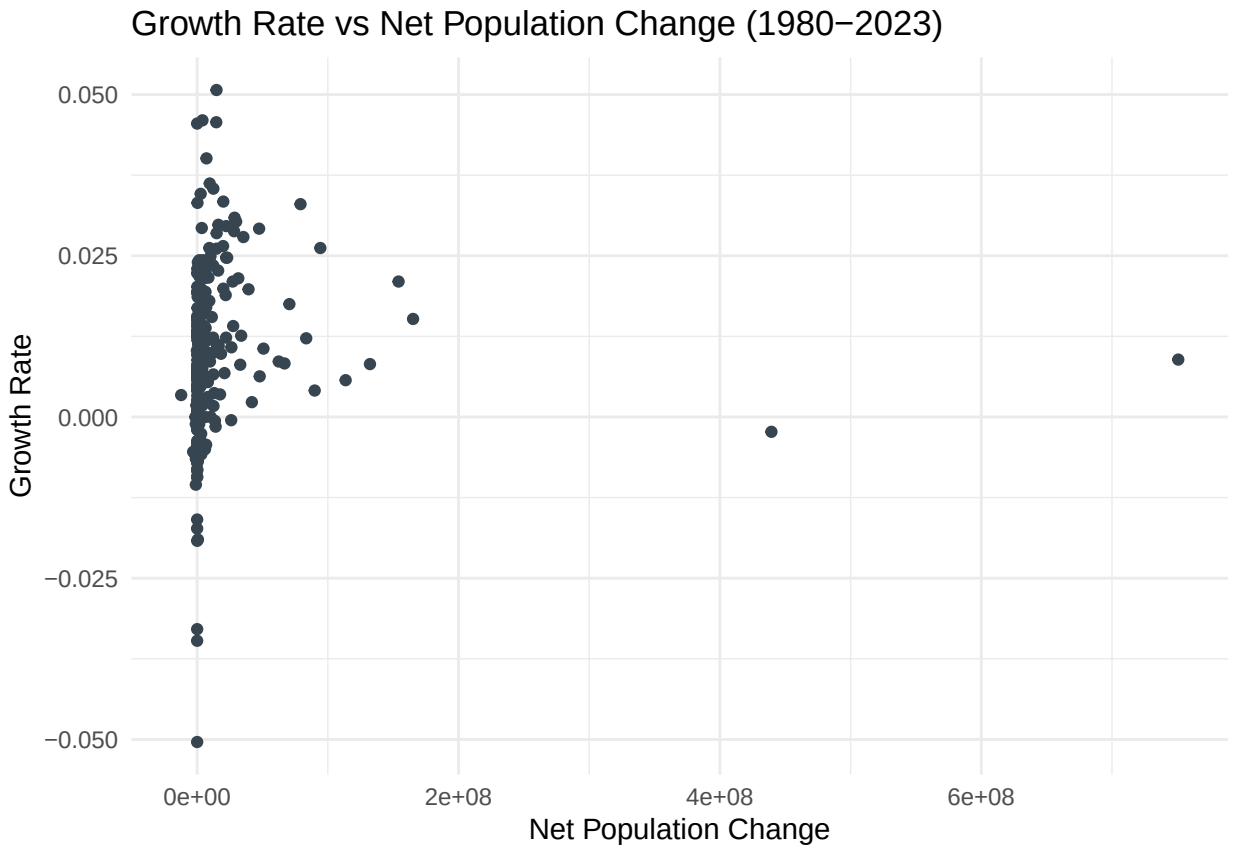
```
ggplot(countries_data, aes(x = density)) +
  geom_histogram(fill = "#36454F", bins = 30) +
  labs(title = "Distribution of Population Density (2023)", x = "Density (people per sq km)", y = "Frequ
  theme_minimal()
```



The histogram indicates that most countries have low to moderate population density. A few countries have exceptionally high densities, suggesting urban concentration.

4. Population Change from 1980 to 2023

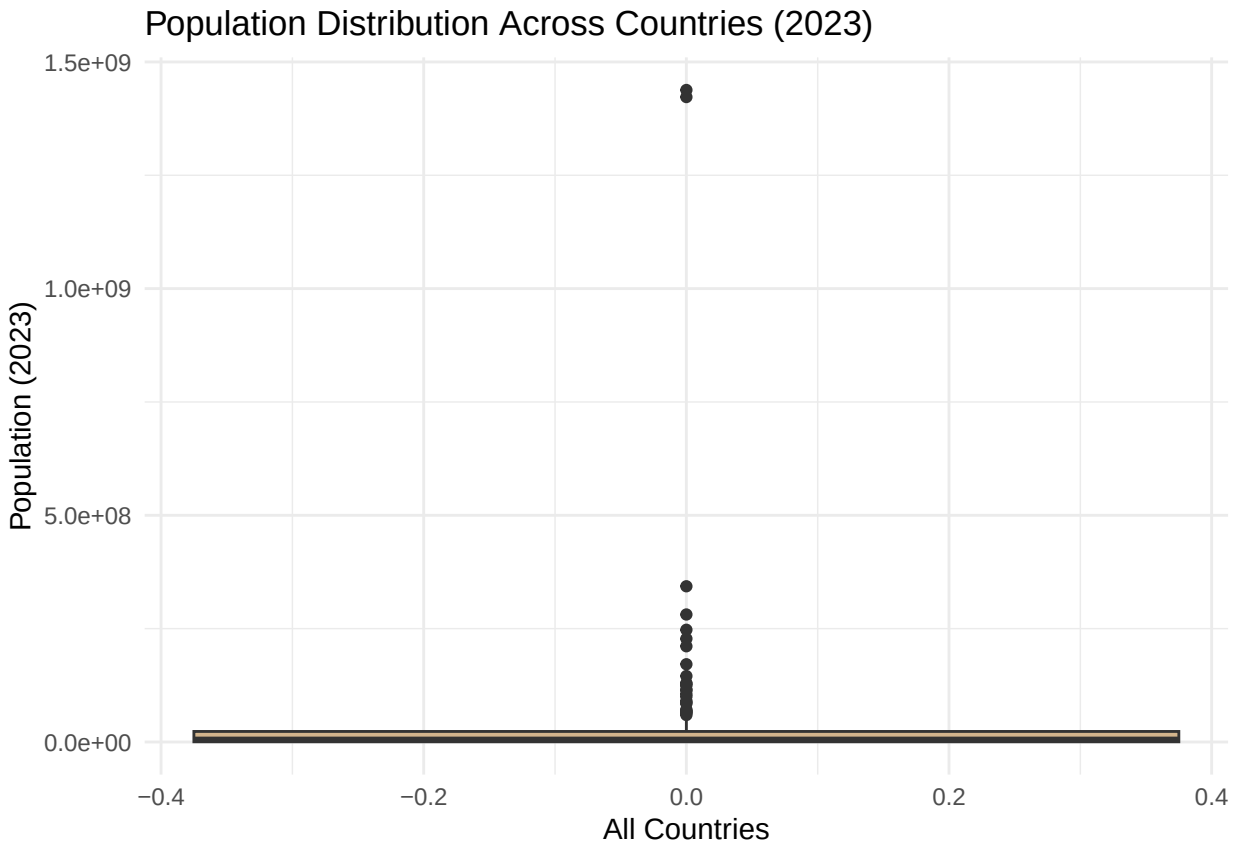
```
countries_data <- countries_data %>%
    mutate(net_change = pop2023 - pop1980)

ggplot(countries_data, aes(x = net_change, y = growthRate)) +
    geom_point(color = "#36454F") +
    labs(title = "Growth Rate vs Net Population Change (1980-2023)", x = "Net Population Change", y = "
    theme_minimal()
```



This scatter plot shows a positive correlation between net population change and growth rate. Countries with higher net changes tend to have higher growth rates.

5. Population Distribution by Country

```
ggplot(countries_data, aes(y = pop2023)) +
    geom_boxplot(fill = "tan") +
    labs(title = "Population Distribution Across Countries (2023)", x = "All Countries", y = "Populatio
    theme_minimal()
```

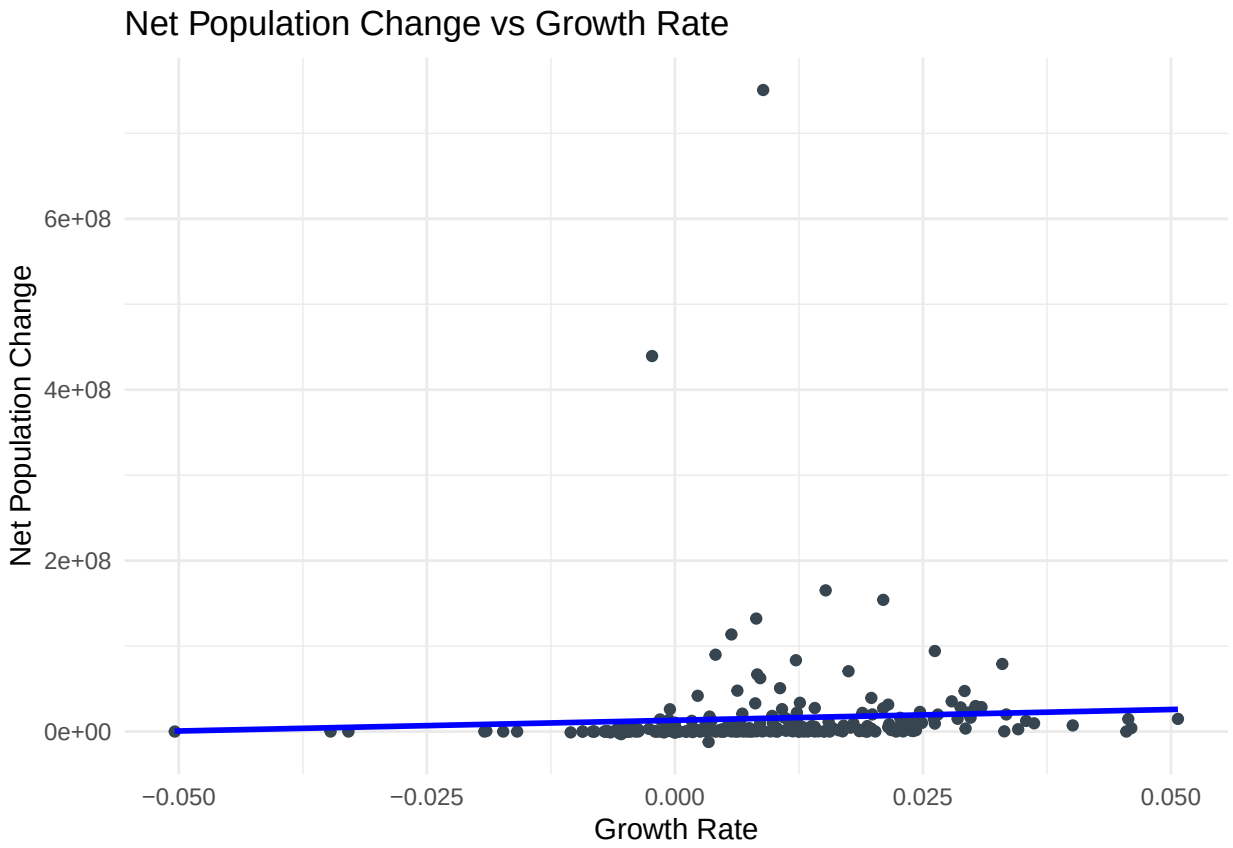


The box plot demonstrates the wide variation in country populations. The presence of outliers indicates countries with significantly higher populations.

6. Growth Rate vs Net Change

```
ggplot(countries_data, aes(x = growthRate, y = net_change)) +
    geom_point(color = "#36454F") +
    geom_smooth(method = "lm", se = FALSE, color = "blue") +
    labs(title = "Net Population Change vs Growth Rate", x = "Growth Rate", y = "Net Population Change")
    theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

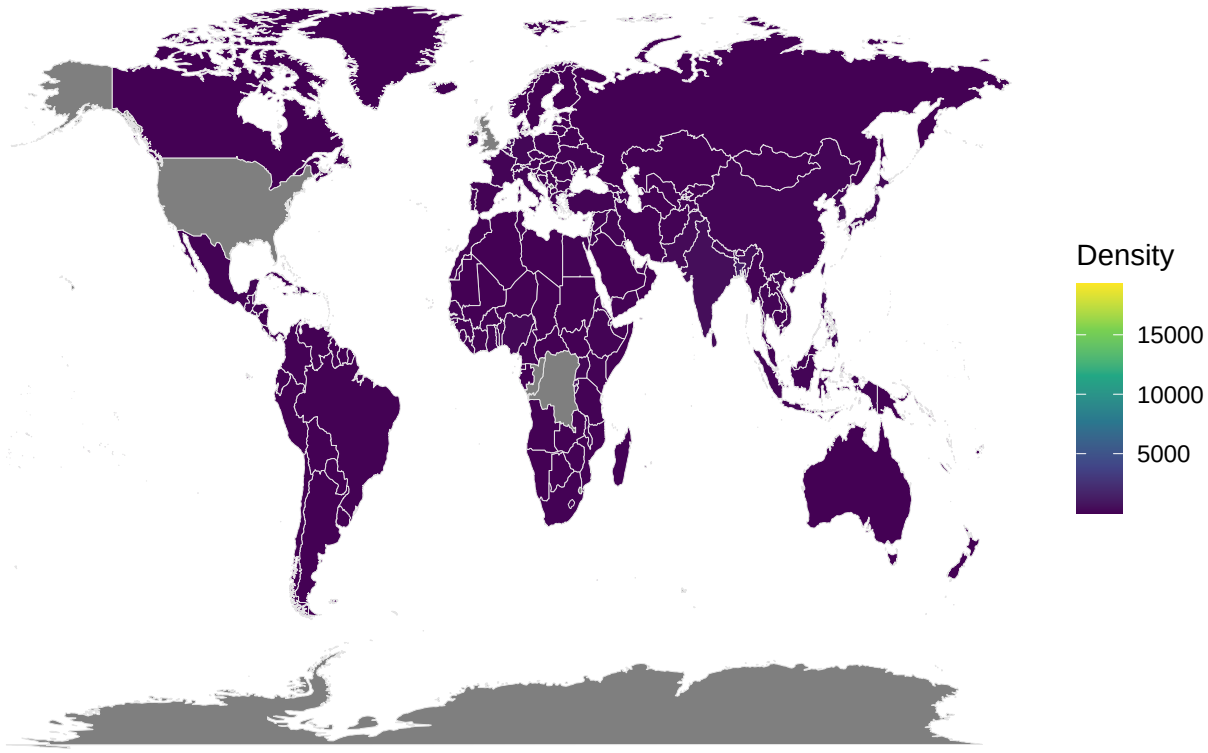


This plot adds a regression line to emphasize the relationship between growth rate and net change. The strong positive slope confirms the correlation.

7. Population Density Map

```
world_map <- map_data("world")
world_map$region <- tolower(world_map$region)
countries_data$country <- tolower(countries_data$country)

map_data_joined <- left_join(world_map, countries_data, by = c("region" = "country"))

ggplot(map_data_joined, aes(x = long, y = lat, group = group, fill = density)) +
  geom_polygon(color = "gray90", size = 0.1) +
  scale_fill_viridis_c(option = "viridis", na.value = "grey50") +
  labs(title = "Global Population Density Map", fill = "Density") +
  theme_void()
```

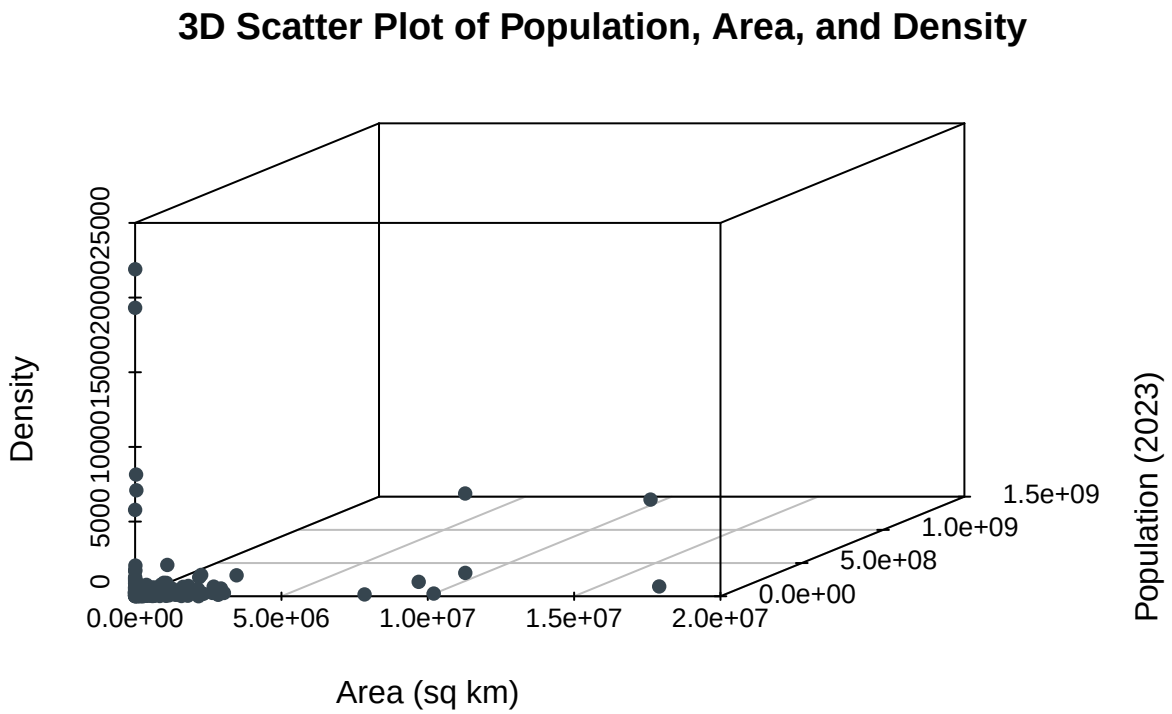
```
## Warning: Using `size` aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use `linewidth` instead.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.
```



The choropleth map visualizes population density worldwide. Regions with higher density are highlighted, showing patterns of population concentration.

8. 3D Scatter Plot of Population, Area, and Density

```
library(scatterplot3d)

plot_data <- countries_data %>%
  select(area, pop2023, density) %>%
  na.omit()

scatterplot3d(plot_data$area, plot_data$pop2023, plot_data$density,
              pch = 16, color = "#36454F",
              xlab = "Area (sq km)",
              ylab = "Population (2023)",
              zlab = "Density",
              main = "3D Scatter Plot of Population, Area, and Density")
```



The 3D scatter plot illustrates the interaction between area, population, and density. It provides a multidimensional perspective on how these variables relate across countries.

Conclusion

Through this analysis, we uncovered several patterns in socio-economic indicators:

1. **Population Trends**: The global population has been increasing, with certain countries experiencing rapid growth.
2. **Population Density**: There's significant variation in population density, affecting resource allocation and urban planning.
3. **Growth Rate Correlation**: A strong positive correlation exists between growth rate and net population change.
4. **Geographical Patterns**: The choropleth map highlights regions with high population densities, often associated with urban centers.

These insights can inform policymakers and researchers about demographic trends and potential areas for further study.

References

- **Countries Data**: https://www.kaggle.com/datasets/arpitsinghaiml/world-population