Programming for Data Analytics

Assignment 3 - Group 1(Fahid Ansar - 960465, Gurmeet Deshwal - 124670, YaswanthKumar - 809732, Raafay- 927876)

```
# Load necessary libraries
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(readr)
```

```
#Load the dataset and check basic information

df = read.csv("gapminder.csv")

# Display first few rows
head(df)
```

```
##   X     country continent year lifeExp      pop gdpPercap
## 1 1 Afghanistan      Asia 1952  28.801  8425333  779.4453
## 2 2 Afghanistan      Asia 1957  30.332  9240934  820.8530
## 3 3 Afghanistan      Asia 1962  31.997 10267083  853.1007
## 4 4 Afghanistan      Asia 1967  34.020 11537966  836.1971
## 5 5 Afghanistan      Asia 1972  36.088 13079460  739.9811
## 6 6 Afghanistan      Asia 1977  38.438 14880372  786.1134
```

```
# Checking for unique years in the dataset to understand time span
unique_years = unique(df$year)

# Checking for unique continents
unique_continents = unique(df$continent)

# General statistics of population distribution
pop_stats = summary(df$pop)

# Display results
list(unique_years, unique_continents, pop_stats)
```

```
## [[1]]
##  [1] 1952 1957 1962 1967 1972 1977 1982 1987 1992 1997 2002 2007
##
## [[2]]
## [1] "Asia"     "Europe"   "Africa"   "Americas" "Oceania"
##
## [[3]]
##      Min.   1st Qu.    Median      Mean   3rd Qu.      Max.
## 6.001e+04 2.794e+06 7.024e+06 2.960e+07 1.959e+07 1.319e+09
```

The dataset covers the years 1952 to 2007 and includes data for five continents: Asia, Europe, Africa, Americas, and Oceania.

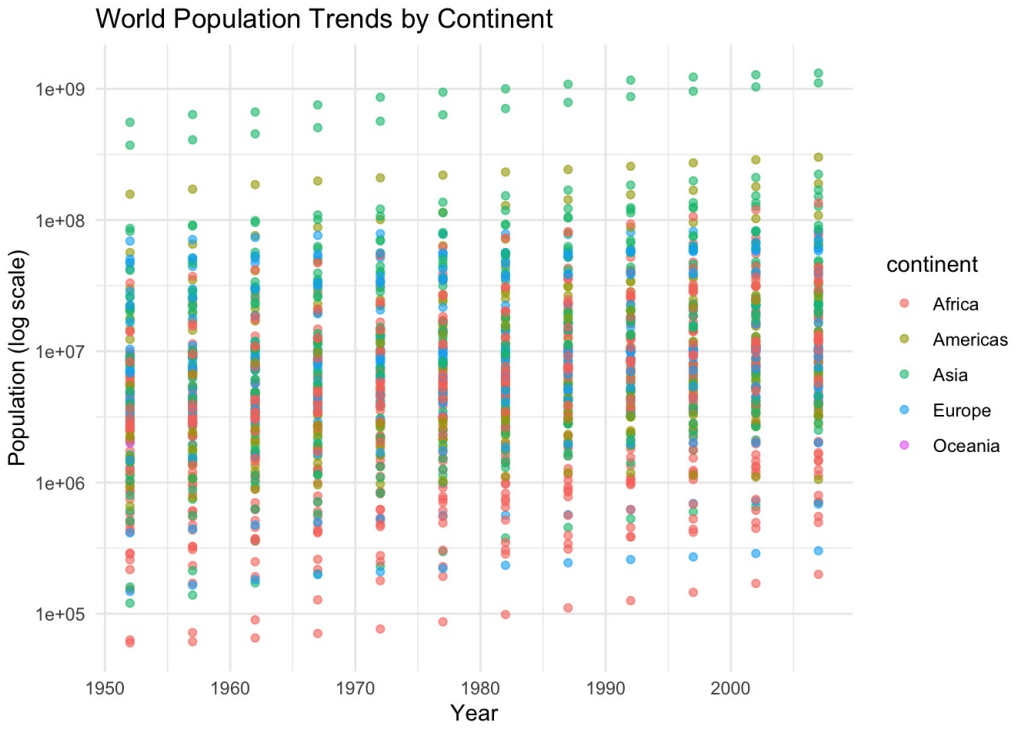
Population Distribution Overview:

- The average population across all countries and years is approximately 29.6 million.
- The smallest recorded population is 60,011.
- The largest recorded population is 1.318 billion, which is likely from China or India.
- The median population is around 7 million, indicating that most countries have significantly lower populations compared to the most populous ones.

1.Investigate visually and comment on the distribution of world population as provided in this data – use two different plotting tools.

Visualizing the Distribution of World Population

```
ggplot(df, aes(x=year, y=pop, color=continent)) +
  geom_point(alpha=0.6) +
  scale_y_log10() +
  labs(title="World Population Trends by Continent",
       x="Year", y="Population (log scale)") +
  theme_minimal()
```



This scatter plot illustrates population trends across continents over time.

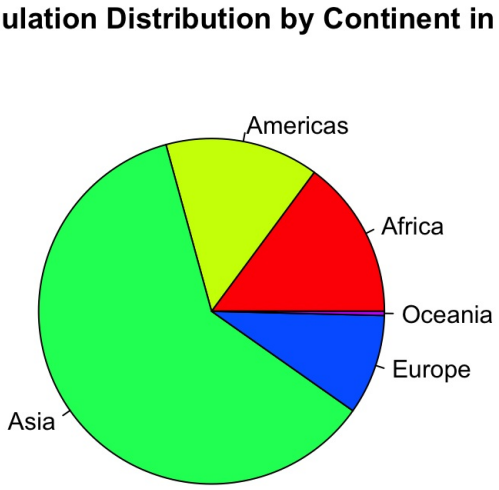
The key insights from this visualization:

- Asia (marked in one color) has the largest population values, showing significant growth over the years.
- Africa also exhibits a strong growth trend, suggesting rapid population increase in more recent years.
- Europe, the Americas, and Oceania have relatively lower populations compared to Asia and Africa.
- The log scale helps in distinguishing trends more clearly, especially for countries with smaller populations.

```
# Selecting the most recent year in the dataset
latest_year = max(df$year)
df_latest = df[df$year == latest_year,]

# Summing population by continent
pop_by_continent = aggregate(pop ~ continent, data=df_latest, sum)

# Creating a pie chart
pie(pop_by_continent$pop, labels=pop_by_continent$continent,
    main=paste("World Population Distribution by Continent in", latest_year),
    col=rainbow(length(pop_by_continent$continent)))
```



This pie chart represents the world population distribution by continent for the most recent year in the dataset.

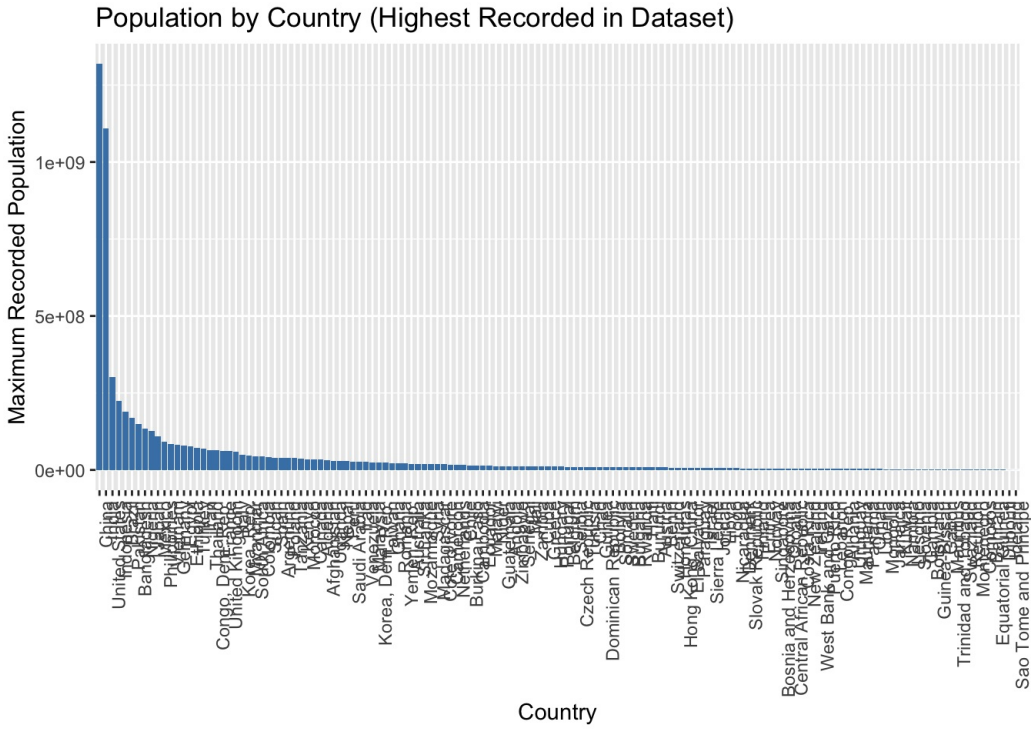
Key Observations:

- Asia holds the largest share of the global population.
- Africa follows as the second-largest, with a growing share.
- Europe, the Americas, and Oceania have significantly smaller population proportions.
- This visualization highlights the concentration of population in certain regions, particularly Asia and Africa.

Aggregating Population Data for All Countries

```
# Aggregating population data for all countries in the dataset
country_population = df %>% group_by(country) %>% summarise(max_pop = max(pop, na.rm=TRUE)) %>% arrange(desc(max_
pop))

# Creating a bar chart for population by country
ggplot(country_population, aes(x=reorder(country, -max_pop), y=max_pop)) +
  geom_bar(stat="identity", fill="steelblue") +
  labs(title="Population by Country (Highest Recorded in Dataset)", x="Country", y="Maximum Recorded Population")
+
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



This bar chart represents the maximum recorded population for each country in the dataset.

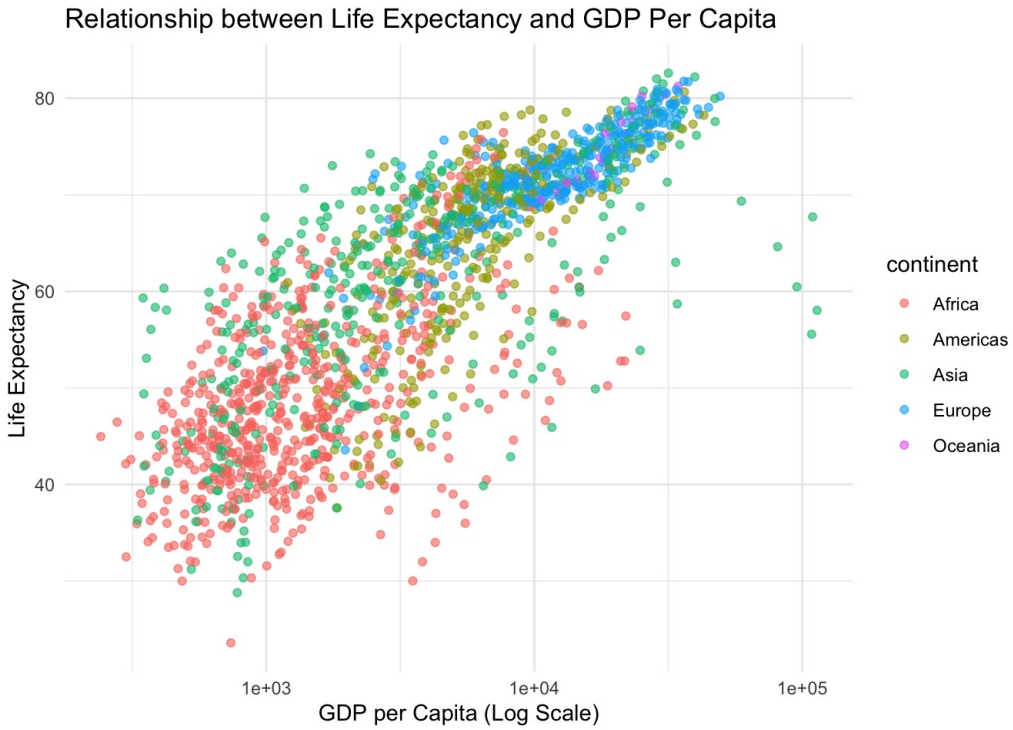
Key Observations:

- The most populated countries include China, India, and the United States, reflecting global population trends.
- Many countries have significantly smaller populations, highlighting disparities in population sizes.
- The chart provides a clear comparison across all countries, with population numbers varying widely.

2.Establish the relationship between life expectancy and GDP per capita for all world countries, then for different continents using a single suitable plot for each. Comment on the visual insights.

Relationship between Life Expectancy and GDP Per Capita

```
ggplot(df, aes(x=gdpPercap, y=lifeExp, color=continent)) +
  geom_point(alpha=0.6) +
  scale_x_log10() +
  labs(title="Relationship between Life Expectancy and GDP Per Capita",
       x="GDP per Capita (Log Scale)", y="Life Expectancy") +
  theme_minimal()
```

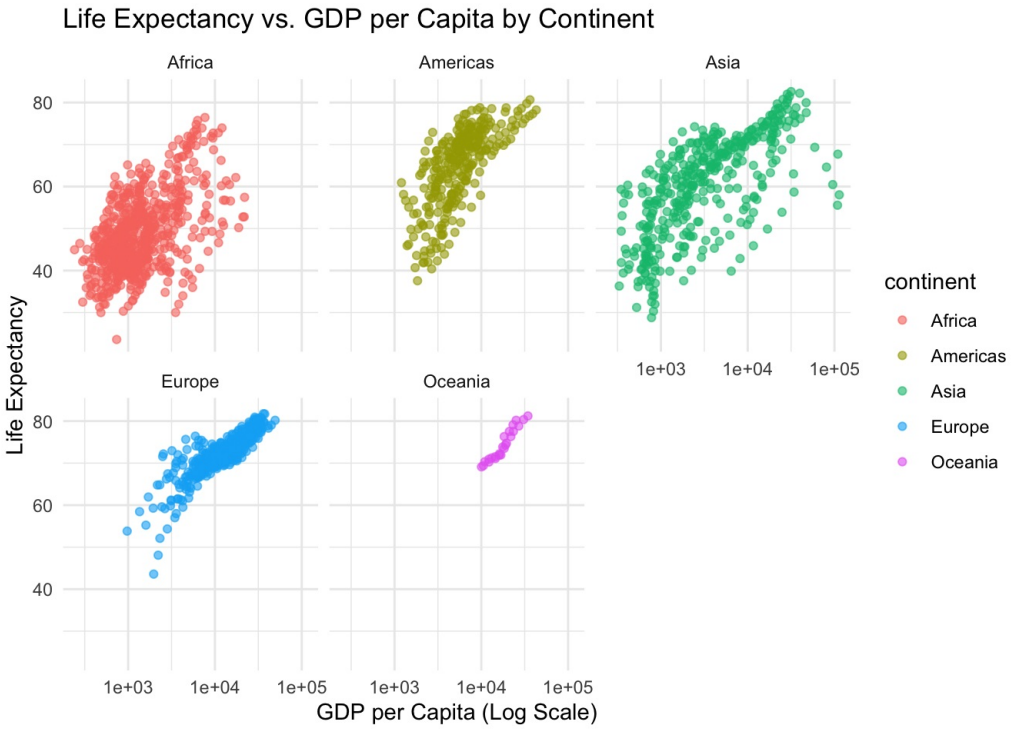


This scatter plot visualizes the relationship between Life Expectancy and GDP per Capita for all world countries, with different colors representing different continents.

Key Insights:

- Positive Correlation: Countries with higher GDP per capita tend to have higher life expectancy, indicating that economic prosperity often leads to better healthcare, living conditions, and overall well-being.
- Strong Regional Differences: European and North American countries (higher GDP per capita) generally have higher life expectancy.
- African countries tend to have lower GDP per capita and lower life expectancy, highlighting economic and healthcare disparities. Asian and South American countries fall in between, with a mix of lower and higher GDP per capita and corresponding variations in life expectancy.
- Log Scale for GDP per Capita: This helps visualize data better, as GDP values vary widely across countries.

```
# Creating scatter plots for each continent separately
ggplot(df, aes(x=gdpPercap, y=lifeExp, color=continent)) +
  geom_point(alpha=0.6) +
  scale_x_log10() +
  facet_wrap(~continent) +
  labs(title="Life Expectancy vs. GDP per Capita by Continent",
       x="GDP per Capita (Log Scale)",
       y="Life Expectancy") +
  theme_minimal()
```



This scatter plot presents the relationship between Life Expectancy and GDP per Capita for different continents using distinct colors.

Key Insights:

- The general positive correlation between GDP per capita and life expectancy remains evident across all continents.
- Europe and North America show high GDP per capita and high life expectancy, with tight clustering.
- Africa has the widest spread, with many countries having low GDP per capita and life expectancy.
- Asia and South America show varied distributions, with some countries approaching the higher GDP and life expectancy range.
- Oceania (mostly represented by Australia and New Zealand) falls into the high GDP, high life expectancy category.

3. Extract the data corresponding to G7 (or G8) countries, then compare the life expectancy of the peoples' in these countries on one chart. Provide commentary interpretation on the obtained chart.

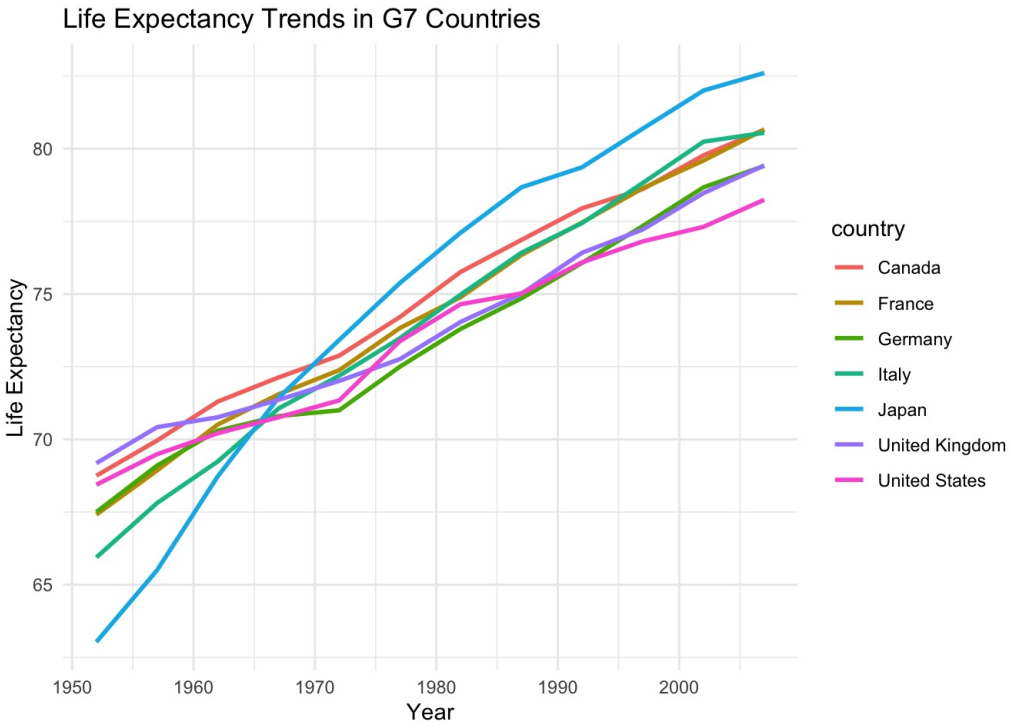
Life Expectancy in G7 Countries

```
# Defining the G7 countries
g7_countries = c("Canada", "France", "Germany", "Italy", "Japan", "United Kingdom", "United States")

# Extracting data for G7 countries
df_g7 = df[df$country %in% g7_countries, ]

ggplot(df_g7, aes(x=year, y=lifeExp, color=country)) +
  geom_line(size=1) +
  labs(title="Life Expectancy Trends in G7 Countries", x="Year", y="Life Expectancy") +
  theme_minimal()
```

```
## 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.
```



This line chart tracks the life expectancy trends in the G7 countries over time.

Key Insights:

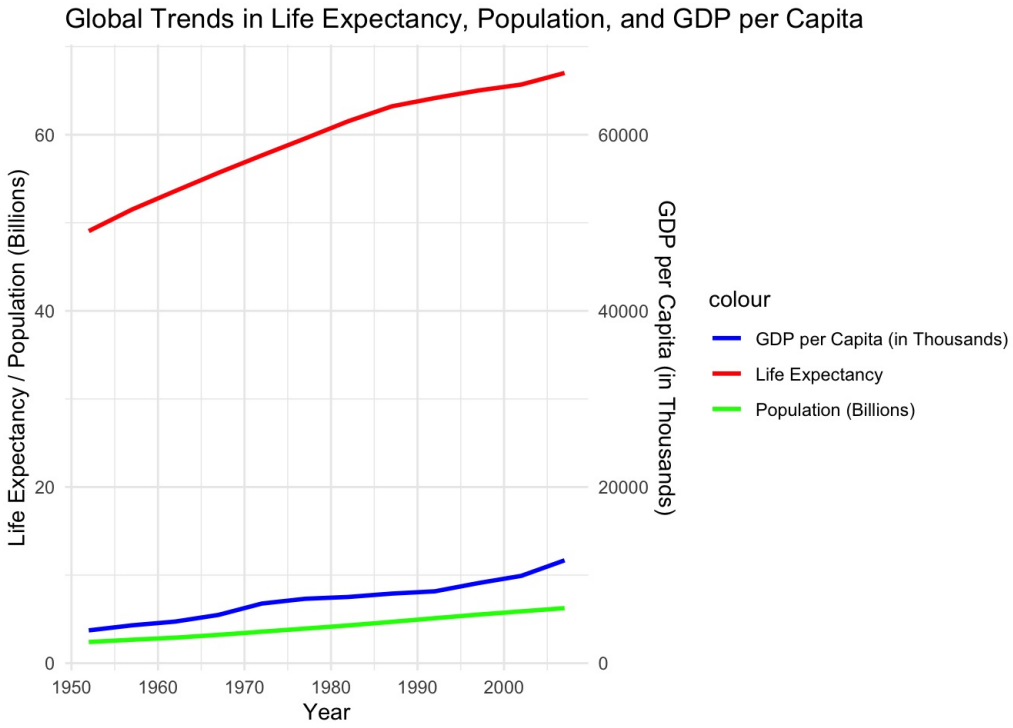
- Consistently High Life Expectancy: All G7 countries show high life expectancy, generally above 70 years.
- Gradual Increase Over Time: Life expectancy has improved in all these countries from 1952 to recent years.
- Japan Leads in Life Expectancy: Japan consistently has the highest life expectancy among the G7 nations.
- Steady Growth with Convergence: While there were some differences in earlier years, the G7 countries seem to converge towards a similar high life expectancy in recent years.
- This highlights how developed economies with strong healthcare and social systems ensure longer life expectancy for their populations.

4. Using an appropriate chart, visualize life expectancy,population, and GDP patterns over the years, then briefly comment the obtained insights.

Global Trends in Life Expectancy, Population, and GDP per Capita

```
# Aggregating data for global trends
global_trends = df %>% group_by(year) %>%
  summarize(lifeExp_avg = mean(lifeExp, na.rm=TRUE),
            pop_sum = sum(pop, na.rm=TRUE),
            gdp_avg = mean(gdpPercap, na.rm=TRUE))

ggplot(global_trends, aes(x=year)) +
  geom_line(aes(y=lifeExp_avg, color="Life Expectancy"), size=1) +
  geom_line(aes(y=pop_sum / 1e9, color="Population (Billions)"), size=1) +
  geom_line(aes(y=gdp_avg / 1000, color="GDP per Capita (in Thousands)"), size=1) +
  scale_y_continuous(sec.axis = sec_axis(~.*1000, name="GDP per Capita (in Thousands)")) +
  labs(title="Global Trends in Life Expectancy, Population, and GDP per Capita",
       x="Year", y="Life Expectancy / Population (Billions)") +
  scale_color_manual(values=c("blue", "red", "green")) +
  theme_minimal()
```



This visualization presents global trends in life expectancy, population, and GDP per capita over the years.

Key Insights:

1. Life Expectancy There is a steady increase in global life expectancy over time. Improvements in healthcare, sanitation, and economic conditions have contributed to this rise.
2. Population Growth The global population has been increasing rapidly, particularly in the latter half of the 20th century. This reflects demographic changes,longer life expectancy, and declining mortality rates.
3. GDP per Capita The average GDP per capita has also risen over time, indicating economic growth across many countries. However, there may still be income disparities that are not captured in an average measure.

Conclusion:

This assignment provided a detailed exploration of global demographic and economic trends using visualizations. The key takeaways include:

1. Population Distribution:

- Asia holds the largest population share, followed by Africa, which is rapidly growing.
- A significant disparity exists between continents in terms of population sizes.

2. Relationship between GDP and Life Expectancy:

- Countries with higher GDP per capita tend to have higher life expectancy.
- The relationship is more prominent in developed regions compared to lower-income regions.

3. G7 Countries' Life Expectancy Trends:

- Life expectancy has steadily increased in G7 nations.
- Japan consistently leads in life expectancy among G7 nations.

4. Global Trends in Life Expectancy, Population, and GDP:

- Overall improvements in living conditions, economic growth, and population increase are evident.
- A sharp rise in GDP per capita over time suggests global economic progress.