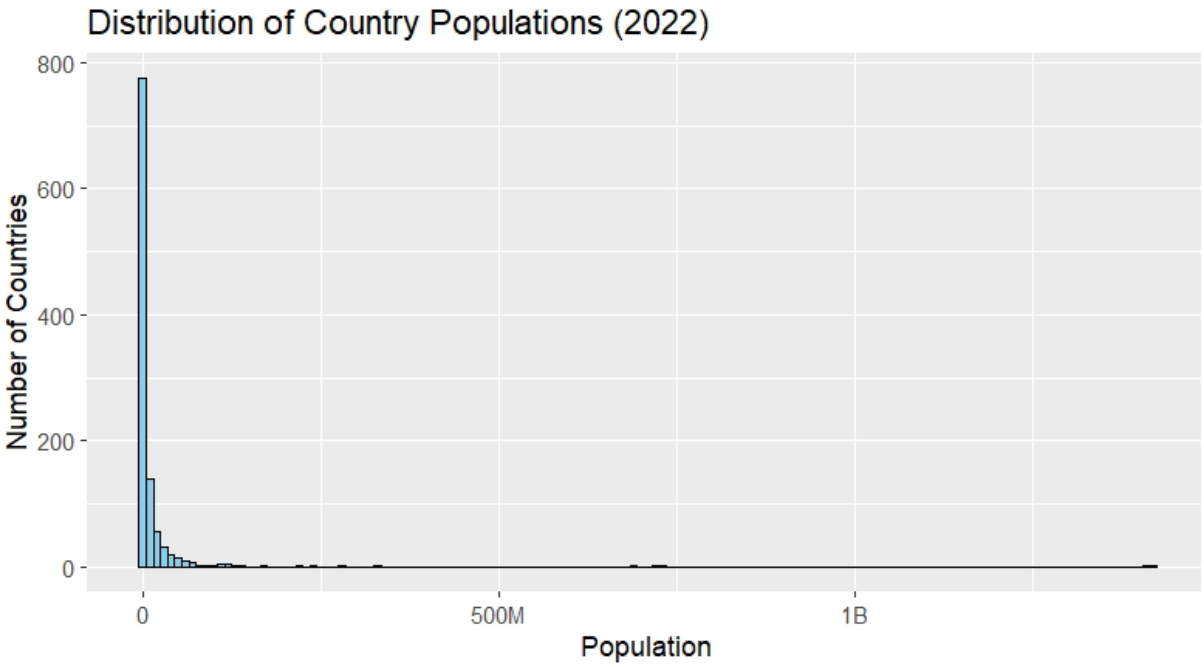
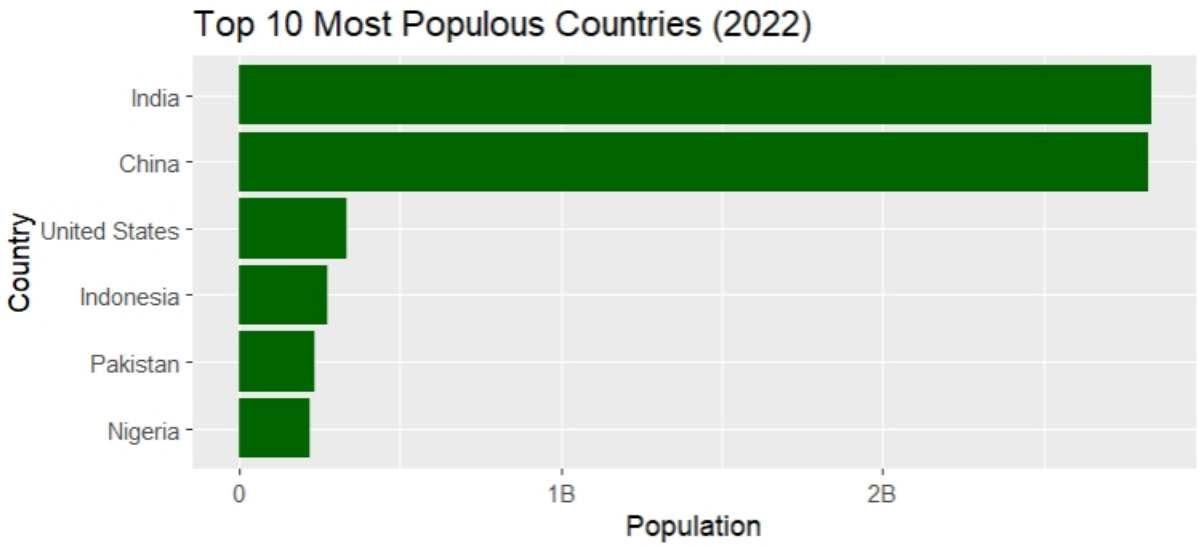
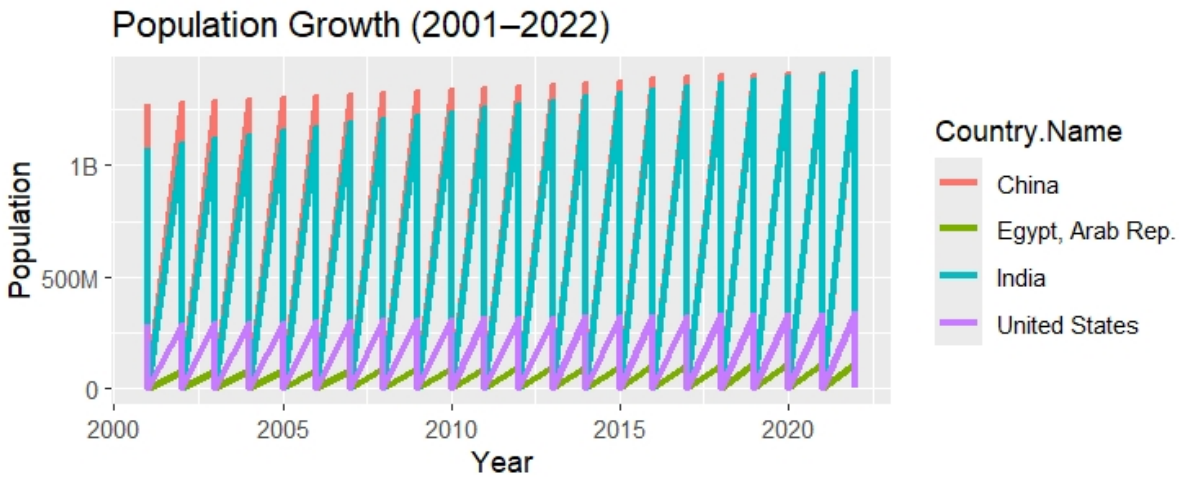
Task 01 Understand population spread



This histogram shows the distribution of populations across countries in 2022. The majority of countries have relatively small populations (below 50 million), while only a few countries exceed several hundred million. This highlights the skewed distribution of population worldwide, where a handful of countries account for a large share of the global population



This bar chart presents the ten most populous countries in 2022. India and China dominate with populations exceeding 1.4 billion each, followed by the United States, Indonesia, Pakistan, and Nigeria. These top countries alone account for a significant portion of the global population, underscoring regional demographic concentration.



This line chart illustrates population growth trends between 2001 and 2022 for selected countries (Egypt, India, China, and the United States). India shows rapid growth, overtaking China in recent years. Egypt demonstrates steady population increase, while the United States grows at a slower but consistent rate. China's growth appears to be stabilizing. These trends reflect different demographic dynamics such as fertility rates, urbanization, and policy impacts.

Appendix:

```
##Task 01
# Task 1 using the sample dataset from Prodigy

library(ggplot2)
library(dplyr)

# Load population data
pop <- read.csv("worldpopulationdata.csv")

# Inspect
str(pop)
summary(pop)

colnames(pop)

##reshape wide format into long format

library(dplyr)
library(tidyr)

pop_long <- pop %>%
  pivot_longer(
    cols = starts_with("X"),    # all columns that start with "X"
    names_to = "Year",
    values_to = "Population"
  )

# Remove "X" from the year names
pop_long$Year <- as.numeric(gsub("X", "", pop_long$Year))

pop_2022 <- pop_long %>% filter(Year == 2022)

library(scales)

##Histogram to visualize the distribution of Country Populations (2022)
```

```r
##Histogram to visualize the distribution of Country Populations (2022)

ggplot(pop_2022, aes(x = Population)) +
  geom_histogram(binwidth = 1e7, fill="skyblue", color="black") +
  scale_x_continuous(labels = label_number(scale_cut = cut_short_scale())) +
  labs(title="Distribution of Country Populations (2022)",
       x="Population", y="Number of Countries")


##Bar chart to visualize the top 10 Most Populous Countries (2022)

pop_top10 <- pop_2022 %>%
  arrange(desc(Population)) %>%
  head(10)

ggplot(pop_top10, aes(x = reorder(Country.Name, Population), y = Population)) +
  geom_col(fill = "darkgreen") +
  coord_flip() +
  scale_y_continuous(labels = label_number(scale_cut = cut_short_scale())) +
  labs(title = "Top 10 Most Populous Countries (2022)",
       x = "Country",
       y = "Population")

# Line plot for some selected countries to analyze the Population Growth (2001-2022) for them
countries_selected <- c("Egypt, Arab Rep.", "India", "China", "United States")

pop_selected <- pop_long %>%
  filter(Country.Name %in% countries_selected, Year >= 2001, Year <= 2022)

ggplot(pop_selected, aes(x=Year, y=Population, color=Country.Name)) +
  geom_line(size=1.2) +
  labs(title="Population Growth (2001-2022)",
       x="Year", y="Population") +
  scale_y_continuous(labels = label_number(scale_cut = cut_short_scale()))
```