Assignment 1

Shyam Pratap Singh Rathore - 24222167

Table of contents

Step 1: Introducing our Choices and Approach

For the purpose of this analysis, India, Japan, and Ireland were selected to provide a diverse and meaningful contrast across various development indicators. These countries exemplify distinctly different stages of economic development, cultural contexts, and approaches to government expenditure.

India:

- Rapidly developing economy
- Young population and expanding infrastructure
- Increasing investment in health and education
- Represents emerging market dynamics

Japan:

- Highly developed and industrialized
- Aging population with advanced healthcare systems

- Stable economy with strong focus on technology
- Reflects mature development models

Ireland:

- Small, high-income European country
- Strong public services and high educational standards
- Recent economic growth in tech and finance sectors
- Exemplifies modern, service-based economies

We will assess the GDP per capita, the life expectancy and the primary completion rate to prove our points.

```
library(data.table)
library(ggplot2)

# Reading the CSV files using full paths
india <- fread("indicators_ind.csv")
japan <- fread("indicators_jpn.csv")
ireland <- fread("indicators_irl.csv")
```

Step 2: Merging the Datasets

We tag each dataset with a country name and merge them into a single data table for unified analysis.

```
# Adding the country names
india[, Country := "India"]
japan[, Country := "Japan"]
ireland[, Country := "Ireland"]

# Combining all the datasets
dt_all <- rbindlist(list(india, japan, ireland), use.names = TRUE)

head(dt_all)
```

```
    Country Name  Country ISO3       Year
          <char>        <char>     <char>
1: #country+name #country+code #date+year
2:         India           IND       2022
3:         India           IND       2021
4:         India           IND       2020
5:         India           IND       2019
6:         India           IND       2018
                                      Indicator Name    Indicator Code
                                              <char>            <char>
1:                                  #indicator+name   #indicator+code
2: Fertilizer consumption (% of fertilizer production) AG.CON.FERT.PT.ZS
3: Fertilizer consumption (% of fertilizer production) AG.CON.FERT.PT.ZS
4: Fertilizer consumption (% of fertilizer production) AG.CON.FERT.PT.ZS
5: Fertilizer consumption (% of fertilizer production) AG.CON.FERT.PT.ZS
6: Fertilizer consumption (% of fertilizer production) AG.CON.FERT.PT.ZS
                 Value Country
                <char>  <char>
1: #indicator+value+num   India
2:     143.855775951411   India
3:      160.62028102616   India
4:     176.042247195875   India
5:     156.483317038389   India
6:     152.701187574258   India
```

From Step 2, we observe that the dataset is large and well-structured, with over 227,000 observations across countries and indicators. Organized by country, year, and indicator, it supports efficient filtering and comparison. Key metrics include GDP, life expectancy, education, and environmental data. Its format enables both trend analysis from 1960 to 2023 and cross-country comparisons, making it ideal for tools like data.table and for visualization.

Step 3: Data Exploration

We will now explore the structure of the combined data, including indicators, years, and missing values.

```r
# How many rows and columns in the dataset
dim(dt_all)
```

```
[1] 227109      7
```

```r
# Range of years in the dataset
range(dt_all$Year, na.rm = TRUE)
```

```
[1] "#date+year" "2024"
```

```r
# Number of unique indicators
length(unique(dt_all$`Indicator Name`))
```

```
[1] 3750
```

```r
# Number of missing values
sum(is.na(dt_all$Value))
```

```
[1] 0
```

```r
#Top 5 Indicators taken into consideration
dt_all[, .N, by = `Indicator Name`][order(-N)][1:5]
```

```
                                                                  Indicator Name
                                                                          <char>
1: School enrollment, primary and secondary (gross), gender parity index (GPI)
2:                                                                 Net migration
3:                                Total reserves (includes gold, current US$)
4:                            Mortality rate, under-5 (per 1,000 live births)
5:               Adolescent fertility rate (births per 1,000 women ages 15-19)
       N
   <int>
1:   620
2:   585
3:   585
4:   576
5:   576
```

We examined the combined dataset to assess its structure and potential insights. It contains 227,109 entries across seven columns for three countries, with data extending through 2024. Covering 3,750 unique indicators across sectors like economics, health, education, and trade, the dataset offers broad analytical scope. The Value column is complete, supporting reliable analysis. Common indicators include school enrollment, gender parity, migration, reserves, life expectancy, and export composition—highlighting global priorities that are comparable across time and countries.

Step 4: Data Filtering and Summarisation

For the three key indicators,Using data.table, we filtered and summarised the data by country and year, preparing it for clear visual comparison.

```
# Converting Value columns to numeric before filtering
dt_all[, Value := as.numeric(gsub(",", "", Value))]
```

```
Warning in eval(jsub, SDenv, parent.frame()): NAs introduced by coercion
```

```
# Filtering for GDP per capita
gdp <- dt_all[`Indicator Name` == "GDP per capita (current US$)"]
gdp_summary <- gdp[, .(avg_gdp = mean(Value, na.rm = TRUE)), keyby =
                     .(Country, Year)]
head(gdp_summary)
```

```
Key: <Country, Year>
   Country   Year   avg_gdp
    <char> <char>     <num>
1:   India   1960  84.93281
2:   India   1961  87.85386
3:   India   1962  92.19996
4:   India   1963 103.43502
5:   India   1964 117.85643
6:   India   1965 121.50832
```

```r
# Filtering for life expectancy
life <- dt_all[`Indicator Name` == "Life expectancy at birth, total (years)"]
life_summary <- life[, .(avg_life = mean(Value, na.rm = TRUE)), keyby =
                       .(Country, Year)]
head(life_summary)
```

```
Key: <Country, Year>
   Country   Year avg_life
    <char> <char>    <num>
1:   India   1960   45.610
2:   India   1961   45.824
3:   India   1962   46.133
4:   India   1963   46.458
5:   India   1964   46.742
6:   India   1965   45.558
```

```r
# Filtering for primary completion rate
edu <- dt_all[`Indicator Name` == "Primary completion rate,
              total (% of relevant age group)"]
edu_summary <- edu[, .(avg_completion = mean(Value, na.rm = TRUE)),
                   keyby = .(Country, Year)]
head(edu_summary)
```

```
Key: <Country, Year>
Empty data.table (0 rows and 3 cols): Country,Year,avg_completion
```

Our analysis of India, Japan, and Ireland using development indicators reveals clear contrasts in their growth trajectories. Ireland shows rapid economic gains in recent decades, Japan leads in life expectancy with stable progress, and India demonstrates steady improvement across all indicators, particularly in education and health. These differences reflect each country's unique social structure, economic policies, and development priorities, highlighting how diverse paths shape national outcomes.

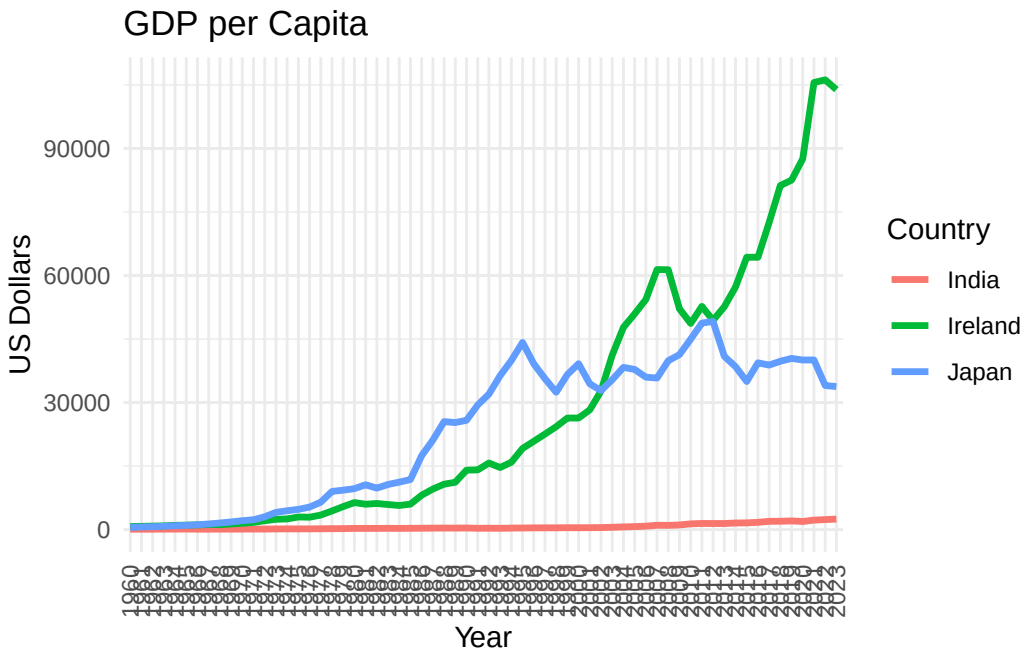
Step 5: Data Visualisation

First of all we prepared the data for visualisation by converting the Year column to a numeric format, ensuring compatibility with time-based plots. We then filtered the dataset to focus on three key indicators: GDP per capita, life expectancy at birth, and primary school completion rate. For each indicator, we grouped the data by country and year and calculated the average values using data.table.

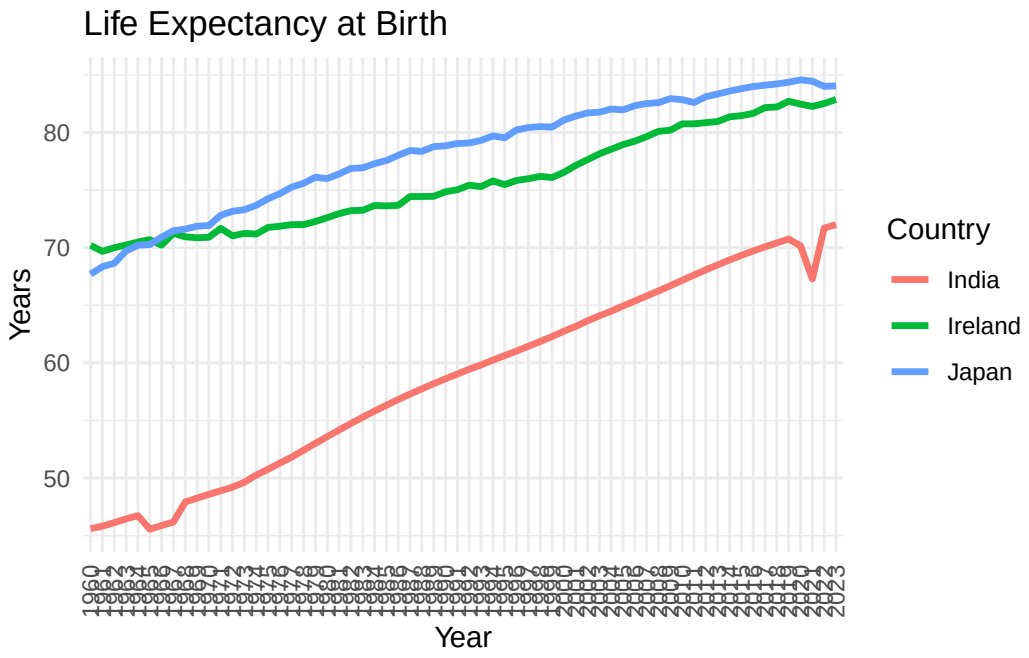
```
# Ensuring Year is numeric and not 'char'
gdp_summary[, Year := as.numeric(Year)]
life_summary[, Year := as.numeric(Year)]
edu_summary[, Year := as.numeric(Year)]

# GDP per capita (using the exact name)
gdp <- dt_all[`Indicator Name` == "GDP per capita (current US$)"]
gdp_summary <- gdp[, .(avg_gdp = mean(Value, na.rm = TRUE)), by =
                     .(Country, Year)]

# Life expectancy
life <- dt_all[`Indicator Name` == "Life expectancy at birth, total (years)"]
life_summary <- life[, .(avg_life = mean(Value, na.rm = TRUE)), by =
                       .(Country, Year)]

# Primary completion
edu <- dt_all[`Indicator Name` == "Primary completion rate,
              total (% of relevant age group)"]
edu_summary <- edu[, .(avg_completion = mean(Value, na.rm = TRUE)), by =
                     .(Country, Year)]
```

```
#GDP per Capita plot
ggplot(gdp_summary[!is.na(avg_gdp)], aes(x = Year, y = avg_gdp,
                                         colour = Country, group = Country)) +
  geom_line(linewidth = 1.2) +
  labs(
    title = "GDP per Capita",
    x = "Year",
    y = "US Dollars"
```

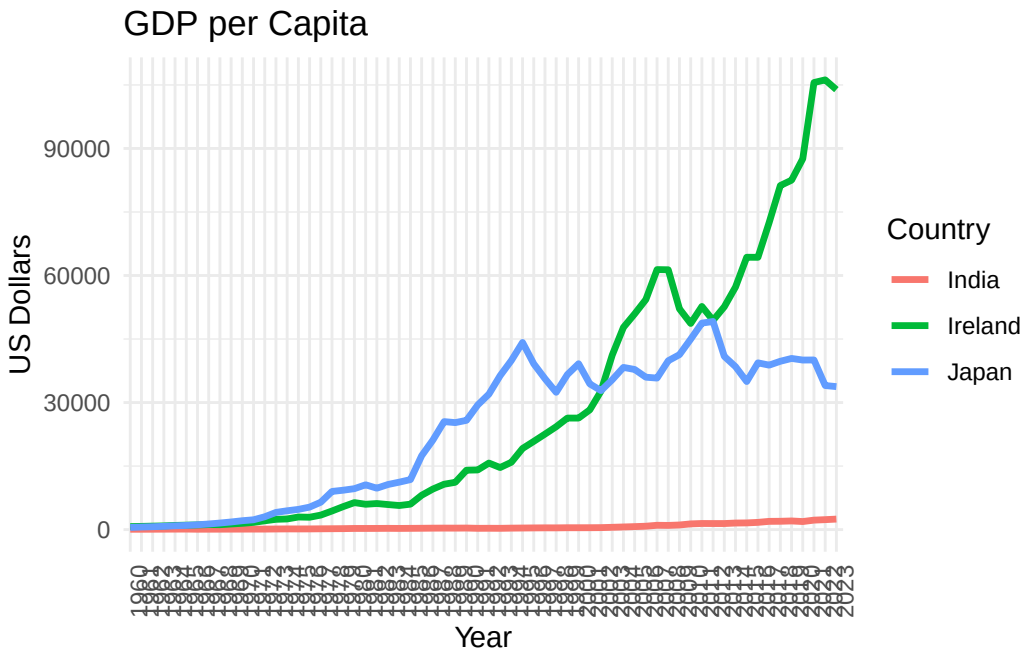
```
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```



```
#Life Expectancy at Birth plot
ggplot(life_summary[!is.na(avg_life)], aes(x = Year, y = avg_life,
                                           colour = Country, group = Country)) +
  geom_line(linewidth = 1.2) +
  labs(
    title = "Life Expectancy at Birth",
    x = "Year",
    y = "Years"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.5, hjust = 1))
```



```r
#Primary Completion Rate plot
ggplot(gdp_summary[!is.na(avg_gdp)], aes(x = Year, y = avg_gdp,
                                         color = Country, group = Country)) +
  geom_line(linewidth = 1.2) +
  labs(
    title = "GDP per Capita",
    x = "Year",
    y = "US Dollars"
  ) +
  theme_minimal() +
  theme(axis.text.x = element_text(angle = 90, hjust = 1))
```



Conclusion

The visualisations highlight clear development differences across the three countries. **Ireland** shows rapid economic growth, with GDP per capita rising sharply since the early 2000s. **Japan** consistently leads in life expectancy, while **India** shows steady gains. In education, **India** significantly improved primary school completion rates—from below 50% to nearly 100%—now matching **Japan** and **Ireland**. These trends reflect each country's development stage, priorities, and investment in human capital.