

# ASSIGNMENT 1

Shyam Pratap Singh Rathore - 24222167

## **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.

```
1 library(data.table)
2 library(ggplot2)
3
4 # Reading the CSV files using full paths
5 india <- fread("indicators_ind.csv")
6 japan <- fread("indicators_jpn.csv")
7 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.

```
1 # Adding the country names
2 india[, Country := "India"]
3 japan[, Country := "Japan"]
4 ireland[, Country := "Ireland"]
5
6 # Combining all the datasets
7 dt_all <- rbindlist(list(india, japan, ireland), use.names = TRUE)
8
9 head(dt_all)
```

	Country Name <char>	Country ISO3 <char>	Year <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 <char>	Indicator Code <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

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.

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

```
[1] 227109      7
```

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

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

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

```
[1] 3750
```

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

```
[1] 0
```



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

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

```
1 # Converting Value columns to numeric before filtering
2 dt_all[, Value := as.numeric(gsub(",", "", Value))]
3
4 # Filtering for GDP per capita
5 gdp <- dt_all[`Indicator Name` == "GDP per capita (current US$)"]
6 gdp_summary <- gdp[, .(avg_gdp = mean(Value, na.rm = TRUE)), keyby = .(Country, Year)]
7 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

```
1 # Filtering for life expectancy
2 life <- dt_all[`Indicator Name` == "Life expectancy at birth, total (years)"]
3 life_summary <- life[, .(avg_life = mean(Value, na.rm = TRUE)), keyby = .(Country, Year)]
4 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

```
1 # Filtering for primary completion rate
2 edu <- dt_all[`Indicator Name` == "Primary completion rate, total (% of rel
3 edu_summary <- edu[, .(avg_completion = mean(Value, na.rm = TRUE)), keyby =
4 head(edu_summary)
```

Key: <Country, Year>

	Country	Year	avg_completion
	<char>	<char>	<num>
1:	India	1971	39.65075
2:	India	1974	41.70239
3:	India	1987	63.32994
4:	India	1988	64.54919
5:	India	1995	73.25361
6:	India	1999	68.89942

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`.

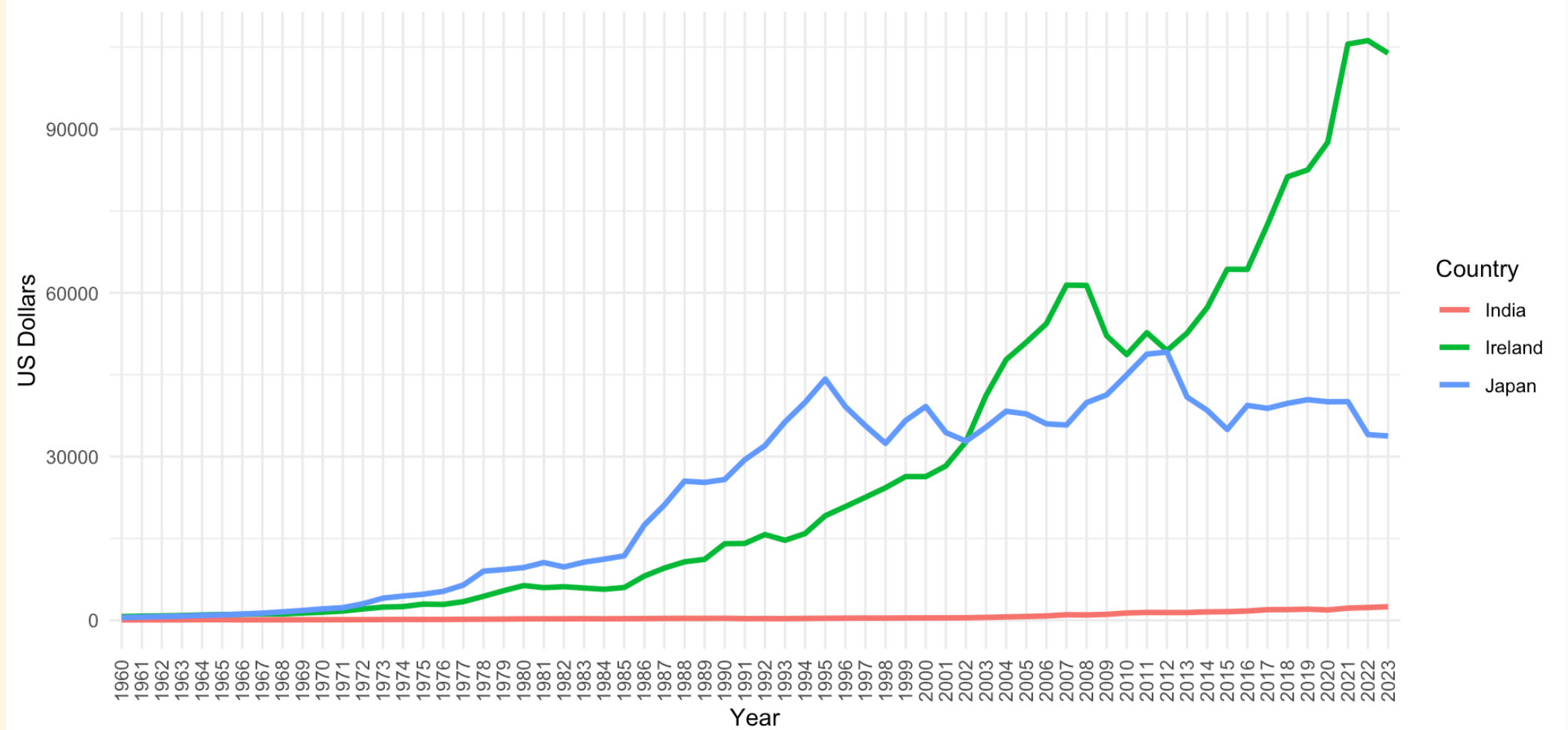
```

1 # Ensuring Year is numeric and not 'char'
2 gdp_summary[, Year := as.numeric(Year)]
3 life_summary[, Year := as.numeric(Year)]
4 edu_summary[, Year := as.numeric(Year)]
5
6 # GDP per capita (using the exact name)
7 gdp <- dt_all[`Indicator Name` == "GDP per capita (current US$)"]
8 gdp_summary <- gdp[, .(avg_gdp = mean(Value, na.rm = TRUE)), by = .(Country
9
10 # Life expectancy
11 life <- dt_all[`Indicator Name` == "Life expectancy at birth, total (years)
12 life_summary <- life[, .(avg_life = mean(Value, na.rm = TRUE)), by = .(Coun
13
14 # Primary completion
15 edu <- dt_all[`Indicator Name` == "Primary completion rate, total (% of rel
16 edu_summary <- edu[, .(avg_completion = mean(Value, na.rm = TRUE)), by = .(

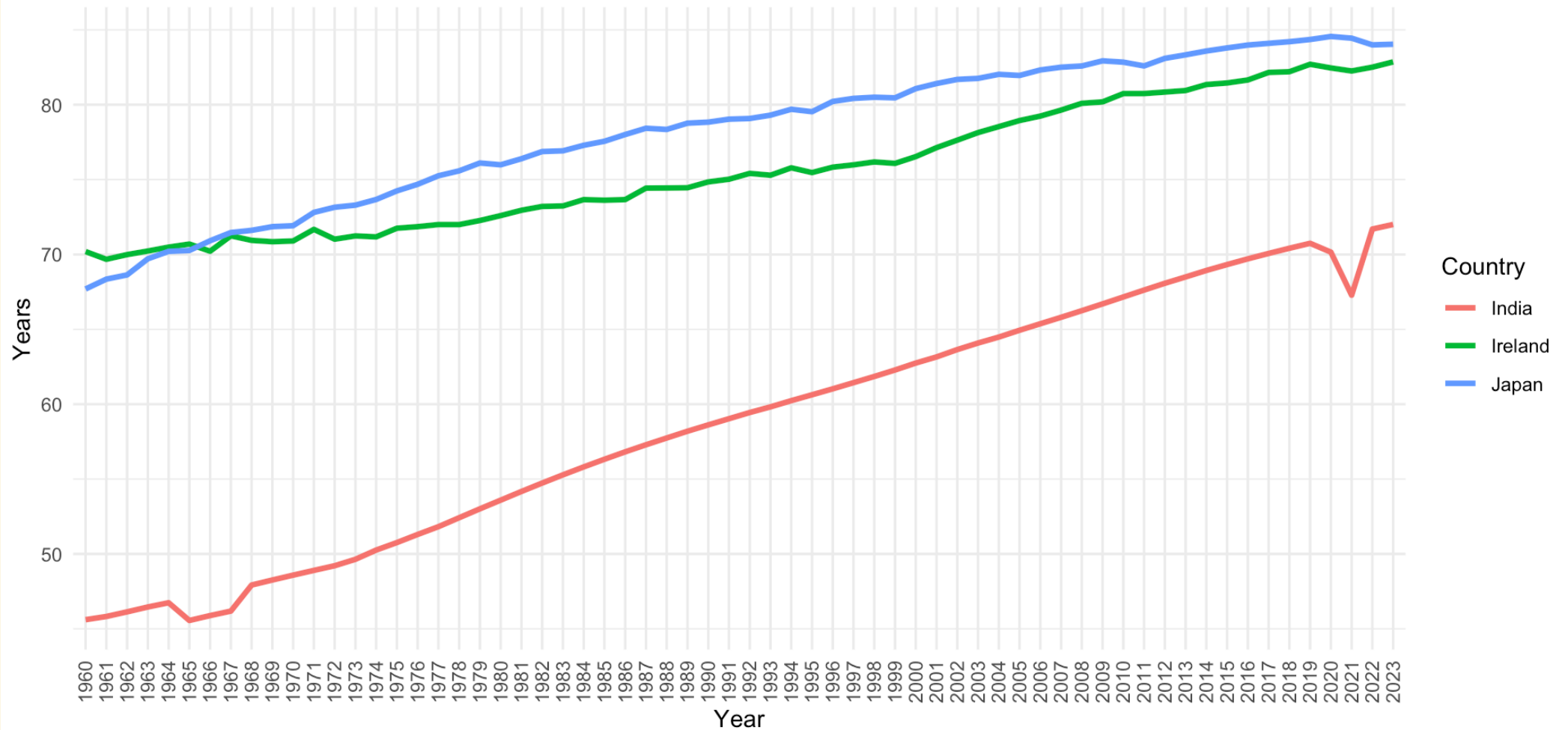
```

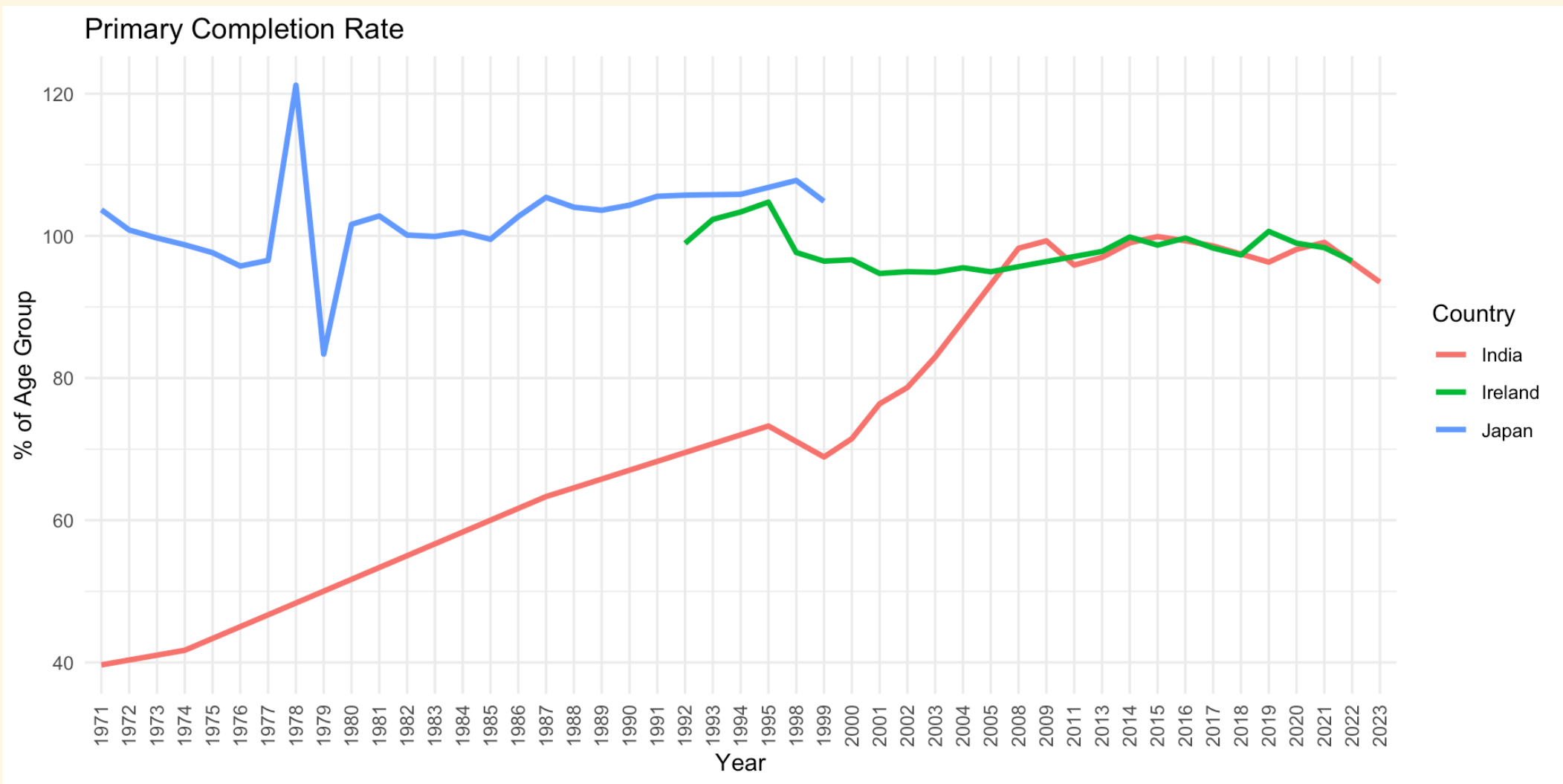


GDP per Capita



## Life Expectancy at Birth





# 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.