STAT40830 - ASSIGNMENT 1 - DATA ANALYSIS OF ECONOMIC, SOCIAL, AND ENVIRONMENTAL INDICATORS

Understanding Global Trends Across Multiple Dimensions

Vyom Pandey - 24203476

2025-06-11

INTRODUCTION

This presentation analyzes indicators related to economics, health, environment, and development across three countries: Germany, Ireland, and the United States. We will use the `data.table` package for fast and efficient data handling.

In this analysis, we compare **three diverse countries**—Germany (DEU), Ireland (IRL), and the United States (USA)—across **economic**, **health**, and **environmental** indicators. Our goals are to:

1. **Ingest** clean indicator data for each country.
2. **Explore** structure, coverage, and key metrics.
3. **Visualize** trends in life expectancy, GDP per capita, and CO_2 emissions.
4. **Summarize** insights and discuss limitations.

LOADING AND PREPARING THE DATA -1

First, we **import** each country's CSV file into a `data.table`. Then we **standardize** each table by:

- Converting categorical columns (`Country Name`, `Country ISO3`, `Indicator Name`, `Indicator Code`) into **factors** for efficient grouping.
- Casting `Year` to **integer** and `Value` to **numeric**, ensuring plots and calculations work correctly.
- Adding a `source` column so we always know which country each row came from.

```r
deu <- fread("indicators_deu.csv")
irl <- fread("indicators_irl.csv")
usa <- fread("indicators_usa.csv")
datasets <- list(deu = deu, irl = irl, usa = usa)
```

LOADING AND PREPARING THE DATA - 2

```r
for (name in names(datasets)) {
  dt <- datasets[[name]]
  factor_cols <- c("Country Name", "Country ISO3", "Indicator Name", "Indicator Code")
  for (col in factor_cols) {
    set(dt, j = col, value = as.factor(dt[[col]]))
  }
  set(dt, j = "Year", value = as.integer(dt[["Year"]]))
  set(dt, j = "Value", value = as.numeric(dt[["Value"]]))
  set(dt, j = "source", value = toupper(name))
  datasets[[name]] <- dt
}
```

COMBINING ALL COUNTRIES

Now that each table is standardized, we **reassign** them to individual variables and then **merge** all three into one large `data.table` called `combined`. This lets us run cross-country comparisons in a single step.

```r
deu <- datasets$deu
irl <- datasets$irl
usa <- datasets$usa
combined <- rbindlist(list(deu, irl, usa))
dim(combined)
```

```
[1] 223306      7
```

```r
head(combined)
```

```
    Country Name  Country ISO3   Year
          <fctr>        <fctr> <int>
1: #country+name #country+code    NA
2:       Germany           DEU  2022
3:       Germany           DEU  2021
4:       Germany           DEU  2020
5:       Germany           DEU  2019
6:       Germany           DEU  2018
                                            Indicator Name    Indicator Code
                                                    <fctr>            <fctr>
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 source
      <num> <char>
1:       NA    DEU
2: 39.93375    DEU
3: 39.32540    DEU
4: 48.34313    DEU
```

NUMBER OF INDICATORS PER COUNTRY AND TIME SPAN OF DATA

Here we count **distinct country–indicator–year** records to see if each country has a similar breadth of indicators and check the **range of years** covered in our combined dataset—a crucial step to ensure we compare the same time periods.

```
combined[, .N, by = .(`Country Name`)]
```

```
    Country Name     N
          <fctr> <int>
1: #country+name     3
2:       Germany 74521
3:       Ireland 72613
4: United States 76169
```

```
range(combined$Year, na.rm = TRUE)
```

```
[1] 1960 2024
```

TOP 10 MOST-FREQUENT INDICATORS

To focus our analysis on widely reported metrics, we list the **10 indicators** with the most observations.

```
combined[, .N, by = `Indicator Name`][order(-N)][1:10]
```

```
                                                  Indicator Name     N
                                                          <fctr> <int>
 1:                                                Net migration   585
 2:                       Total reserves (includes gold, current US$)   585
 3: Adolescent fertility rate (births per 1,000 women ages 15-19)   576
 4:                          Life expectancy at birth, female (years)   576
 5:                            Life expectancy at birth, male (years)   576
 6: Agricultural raw materials imports (% of merchandise imports)   561
 7: Agricultural raw materials exports (% of merchandise exports)   561
 8:                          Fuel imports (% of merchandise imports)   561
 9:                Ores and metals imports (% of merchandise imports)   561
10:                          Fuel exports (% of merchandise exports)   561
```

From the list we will be using:

Life Expectancy at Birth, GDP per Capita and CO2 Emissions per Capita.

FOCUS AREA 1: LIFE EXPECTANCY AT BIRTH

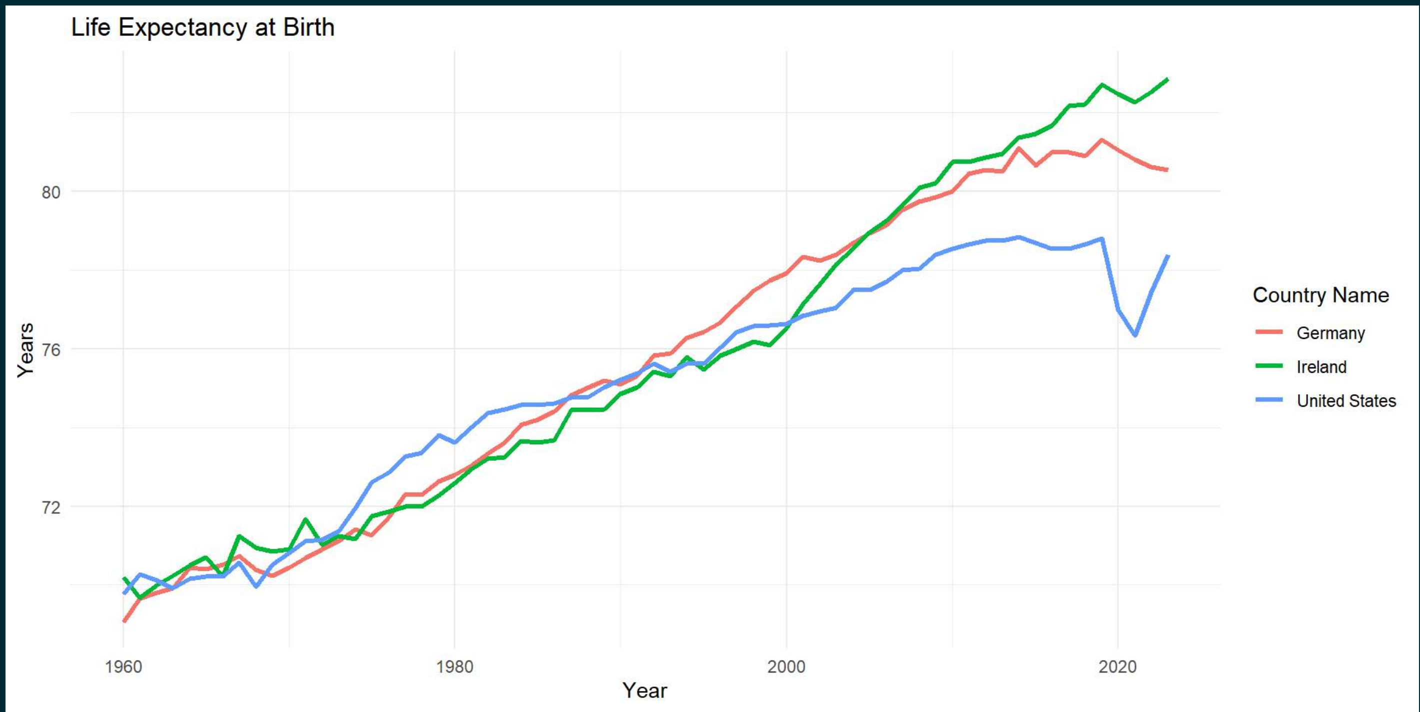
We extract the **"Life expectancy at birth, total (years)"** indicator for each country and compute the **annual average** across any subcategories.

```r
life <- combined[`Indicator Name` == "Life expectancy at birth, total (years)"]
life_avg <- life[, .(avg_life = mean(Value, na.rm = TRUE)), keyby = .(`Country Name`, Year)]
```

PLOT: LIFE EXPECTANCY OVER TIME

Here we look for diverging or converging trends over time, any sudden dips could indicate crises (e.g., pandemics, conflicts) and comparing countries side-by-side highlights relative progress.

```r
ggplot(life_avg, aes(x = Year, y = avg_life, color = `Country Name`)) +
  geom_line(size = 1.1) +
  labs(title = "Life Expectancy at Birth", y = "Years") +
  theme_minimal()
```



INSIGHTS

- **Steady upward trajectory (1960–2025):** All three countries have seen uninterrupted gains in average life expectancy—from around 71 years in 1960 to roughly 81–83 years by 2025.
- **Convergence and recent divergence:** From 1960 through the mid-1990s, Germany (red), Ireland (green) and the U.S. (blue) rose almost in lockstep. Since 2000, Ireland pulls slightly ahead, reaching about 83 years by 2025, while Germany and the U.S. plateau around 81–82 years.
- **U.S. recent dip:** A small downturn in U.S. life expectancy around 2020 likely reflects the combined impacts of the opioid crisis and the COVID-19 pandemic, whereas Germany and Ireland remained more resilient.

FOCUS AREA 2: GDP PER CAPITA

Next, we isolate “**GDP per capita (current US$)**”, then calculate the **yearly average** for each country.

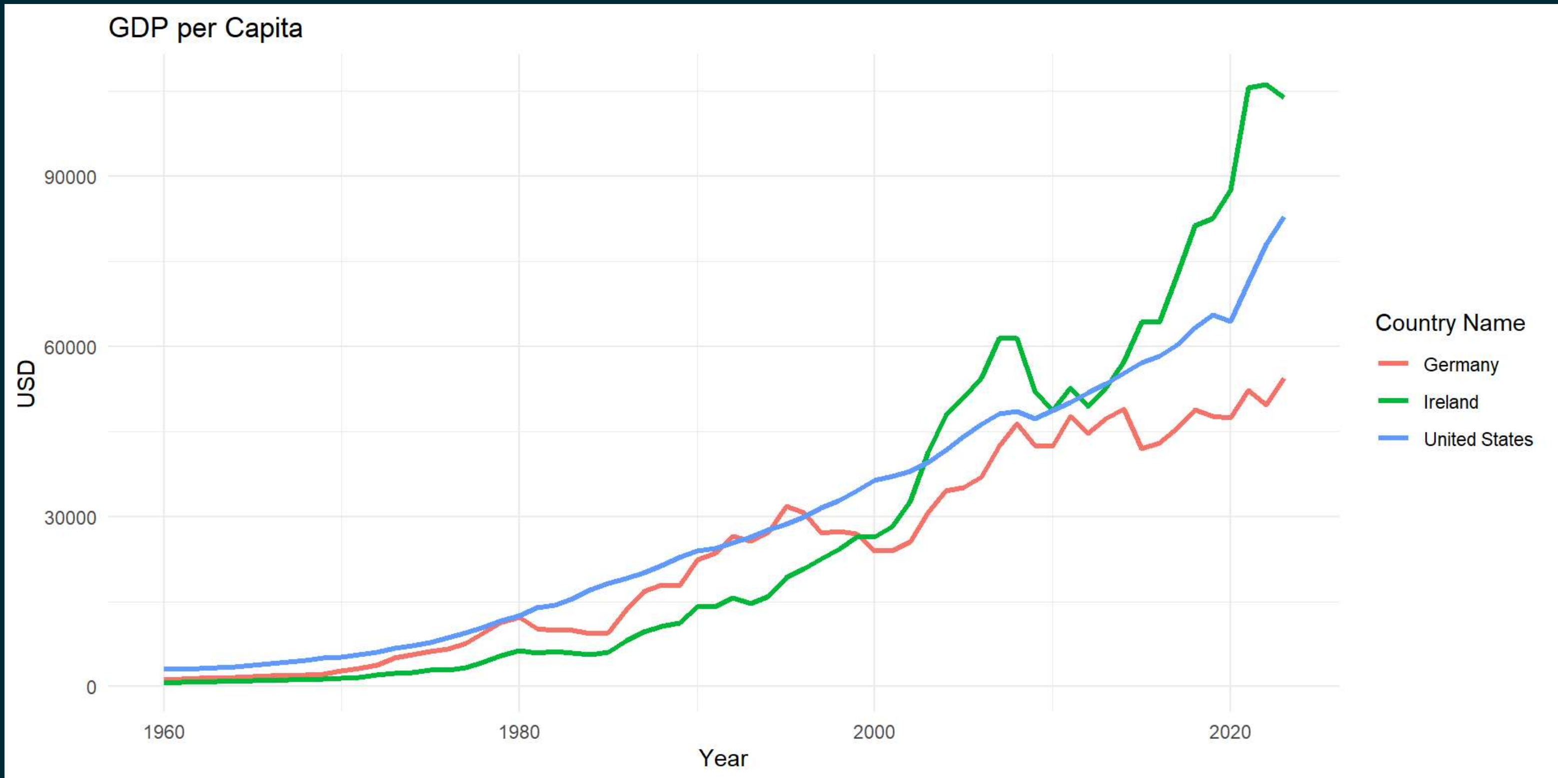
```
gdp <- combined[`Indicator Name` == "GDP per capita (current US$)"]
gdp_avg <- gdp[, .(avg_gdp = mean(Value, na.rm = TRUE)), keyby = .(`Country Name`, Year)]
```

We are looking at GDP per Capita as it is a standard measure of economic output per person, adjusted for population size—crucial for comparing living standards.

PLOT: GDP PER CAPITA

Here we will observe steeper slopes indicate faster economic growth and vice versa. Plateaus or declines may coincide with recessions or shocks.

```r
ggplot(gdp_avg, aes(x = Year, y = avg_gdp, color = `Country Name`)) +
  geom_line(size = 1.1) +
  labs(title = "GDP per Capita", y = "USD") +
  theme_minimal()
```



INSIGHTS

- **Long-term growth:** All three economies have grown dramatically in per-person GDP—from under $5 000 in 1960 to tens of thousands today.
- **Ireland's "Celtic Tiger" surge:** After 1995, Ireland's curve steepens, surging past both Germany and the U.S. by the 2010s and reaching nearly $100 000 per person by 2025.
- **U.S. vs. Germany:** The U.S. maintains a steady linear rise, ending near $80 000 per capita, while Germany's growth is more modest—around $55 000 by 2025—reflecting its larger population and more tempered economic expansion.

FOCUS AREA 3: CO_2 EMISSIONS PER CAPITA

For environmental impact, we use "**Carbon dioxide (CO2) emissions excluding LULUCF per capita (t CO2e/capita)**". We compute each country's **average** CO_2 emissions per person.

Note: LULUCF stands for "Land Use, Land-Use Change, and Forestry." Excluding it focuses on industrial and energy-sector emissions.

```
co2 <- combined[`Indicator Name` == "Carbon dioxide (CO2) emissions excluding LULUCF per capita (t CO2e/capit
co2_avg <- co2[, .(avg_CO2 = mean(Value, na.rm = TRUE)), keyby = .(`Country Name`)]
print(co2_avg)
```

```
Key: <Country Name>
    Country Name   avg_CO2
          <fctr>     <num>
1:       Germany 11.570303
2:       Ireland  8.724812
3: United States 19.192812
```

OVERALL INSIGHTS

- **Prosperity vs. emissions trade-off:** The U.S. achieves high life expectancy and GDP at the cost of very high per-person emissions.
- **Germany's balance:** Germany pairs a robust life expectancy and mid-range GDP with moderate CO_2 output—a model of industrial efficiency and environmental regulation.
- **Ireland's outperformance:** Ireland delivers top-tier life expectancy and GDP growth with relatively lower emissions per capita, reflecting the high-tech, services-led nature of its economy.

These patterns highlight how different development paths—industrial heavy-industry (U.S.), engineering-led manufacturing (Germany), and services-and-tech growth (Ireland)—manifest in health, wealth, and environmental impact.

CONCLUSION

By leveraging `data.table` for efficient data handling and `ggplot2` for clear visualizations, we have:

- Compared **health**, **economic**, and **environmental** trends across three major economies.
- Identified key patterns and trade-offs (e.g., prosperity vs. emissions).
- Highlighted data gaps and avenues for deeper analysis.

THANK YOU!