Development Analysis of Austria, Latvia, and Poland

Economic, Social, Environmental and Health Indicators

Vimarish K M | 24229318

2025-06-10

Objective

- Compare economic, social, and development indicators across three countries.
- Focus: **data.table** speed and efficiency.
- Countries: Austria, Latvia, Poland.
- Data Source: World Bank (HDX).

Load Packages

```
library(data.table)
library(ggplot2)
library(stringr)
library(dplyr)
library(knitr)
library(latex2exp)
library(readr)
library(tidyr)
library(plotly)
```

Read and Clean Data

```
# Read the data (skipping metadata row)
correct_names = c("Country Name", "Country ISO3",
                  "Year","Indicator Name",
                  "Indicator Code", "Value")
aut = fread("indicators_aut.csv", skip = 2,
            col.names = correct_names)
lva = fread("indicators_lva.csv", skip = 2,
            col.names = correct_names)
pol = fread("indicators_pol.csv", skip = 2,
            col.names = correct_names)
# Merge all datasets into one
data_all = rbindlist(list(aut, lva, pol))
# Assign correct classes
data_all[, Year := as.integer(Year)] # Year to integer
data_all[, Value := as.numeric(Value)] # Value to num
```

Quick Data Exploration

```r
# 1. Dimensions and structure
str(data_all, vec.len = 1.4)
```

```
Classes 'data.table' and 'data.frame':  224813 obs. of  6 variab
 $ Country Name  : chr  "Austria" ...
 $ Country ISO3  : chr  "AUT" ...
 $ Year          : int  2001 2000 1999 1998 ...
 $ Indicator Name: chr  "Fertilizer consumption (% of fertilizer
 $ Indicator Code: chr  "AG.CON.FERT.PT.ZS" ...
 $ Value         : num  75.2 95.1 ...
 - attr(*, ".internal.selfref")=<externalptr>
```

Summary:

The combined data set comprises 224,813 observations across 6 variables. The key fields include:

- **Country Name**, **Country ISO3**, **Indicator Name**, and **Indicator Code**: stored as character variables.
- **Year**: correctly recognized as an integer.
- **Value**: stored as a numeric variable representing the actual measurement for each indicator.

```
kable(data_all[, .(Indicators = uniqueN(`Indicator Name`),
                   Years = paste0(min(Year), "-", max(Year))),
               keyby = `Country Name`])
```

Country Name	Indicators	Years
Austria	1998	1960-2024
Latvia	3161	1960-2024
Poland	3584	1960-2024

Interpretation

The dataset includes indicator data for the following countries:

- **Austria** has data for **1,998 indicators** spanning the years **1960 to 2024**.
- **Latvia** has data for **3,161 indicators** over the same period (**1960 to 2024**).
- **Poland** has the most extensive coverage, with **3,584 indicators** from **1960 to 2024**.
- For the Indicator in terms of range of years, we group by country using "keyby".

Country Count

```
kable(data_all[, .(Total_Observations = .N),
         keyby = `Country Name`])
```

Country Name	Total_Observations
Austria	73499
Latvia	64771
Poland	86543

- Austria has 73,499 records, Latvia has 64,771 records, and Poland has 86,543 records in the data set.
- All three countries are well-represented, with no major imbalance in data volume. This balanced representation supports a fair and reliable comparative analysis across countries.

Indicator Frequency Table

```
kable(data_all[, .N,
               keyby = `Indicator Name`][order(-N)][1:10])
```

Indicator Name	N
Net migration	585
Adolescent fertility rate (births per 1,000 women ages 15-19)	576
Life expectancy at birth, female (years)	576
Life expectancy at birth, male (years)	576
School enrollment, primary and secondary (gross), gender parity index (GPI)	525
Mortality rate, under-5 (per 1,000 live births)	519
Adjusted savings: mineral depletion (current US$)	468
Agricultural land (% of land area)	465
Arable land (% of land area)	465
Average precipitation in depth (mm per year)	456

Top Indicators (by Mean Value)

```
# Summary of combined data by mean value
ind_summary = data_all %>%
  group_by(`Indicator Name`) %>%
  summarise(
    #min = min(Value, na.rm = TRUE),
    mean_value = mean(Value, na.rm = TRUE),
    #max = max(Value, na.rm = TRUE),
    Total = n()
  ) %>%
  arrange(desc(mean_value)) %>%
  select(-mean_value)
```

```r
print(ind_summary,n = 10)
```

```
# A tibble: 3,632 x 2
   `Indicator Name`                                  Total
   <chr>                                             <int>
 1 GNI (constant LCU)                                   85
 2 Broad money (current LCU)                            46
 3 Gross domestic income (constant LCU)                112
 4 Gross national expenditure (constant LCU)           112
 5 GNI, PPP (constant 2021 international $)             85
 6 GDP: linked series (current LCU)                    102
 7 GNI: linked series (current LCU)                    102
 8 GDP, PPP (constant 2021 international $)            102
 9 GDP (constant LCU)                                  132
10 Gross national expenditure (current LCU)            112
# i 3,622 more rows
```

Interpretation

The top indicators, based on average values across years and countries, include:

1. GNI (constant LCU)
2. Broad money (current LCU)
3. Gross Domestic Income (constant LCU)
4. Gross national expenditure (constant LCU)
5. GNI, PPP (constant 2021 international $)

These are primarily **economic indicators** with high monetary values, mostly expressed in local currency units (LCU). This suggests that the data set offers rich coverage of **macroeconomic metrics**, making it well-suited for analyzing national income, expenditure, and monetary trends over time.

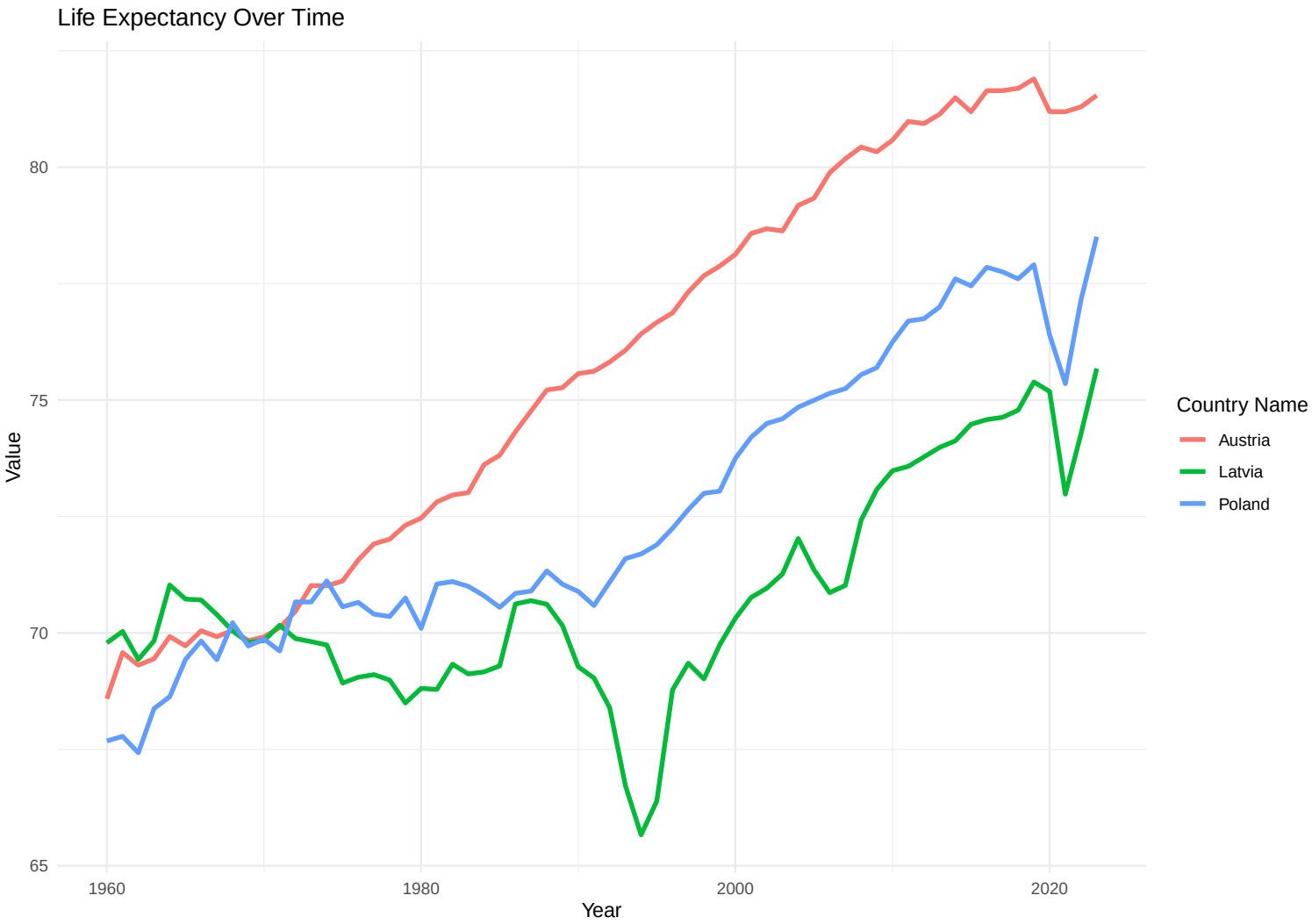
Electricity Access (%)

```
# Electricity access by country
elec = data_all[`Indicator Name` ==
    "Access to electricity (% of population)"]

elec[, .(MeanAccess = mean(Value, na.rm = TRUE)),
     keyby = `Country Name`]
```

```
Key: <Country Name>
   Country Name MeanAccess
          <char>      <num>
1:      Austria  100.00000
2:       Latvia   99.92353
3:       Poland   99.97059
```

Life Expectancy Over Years by Country

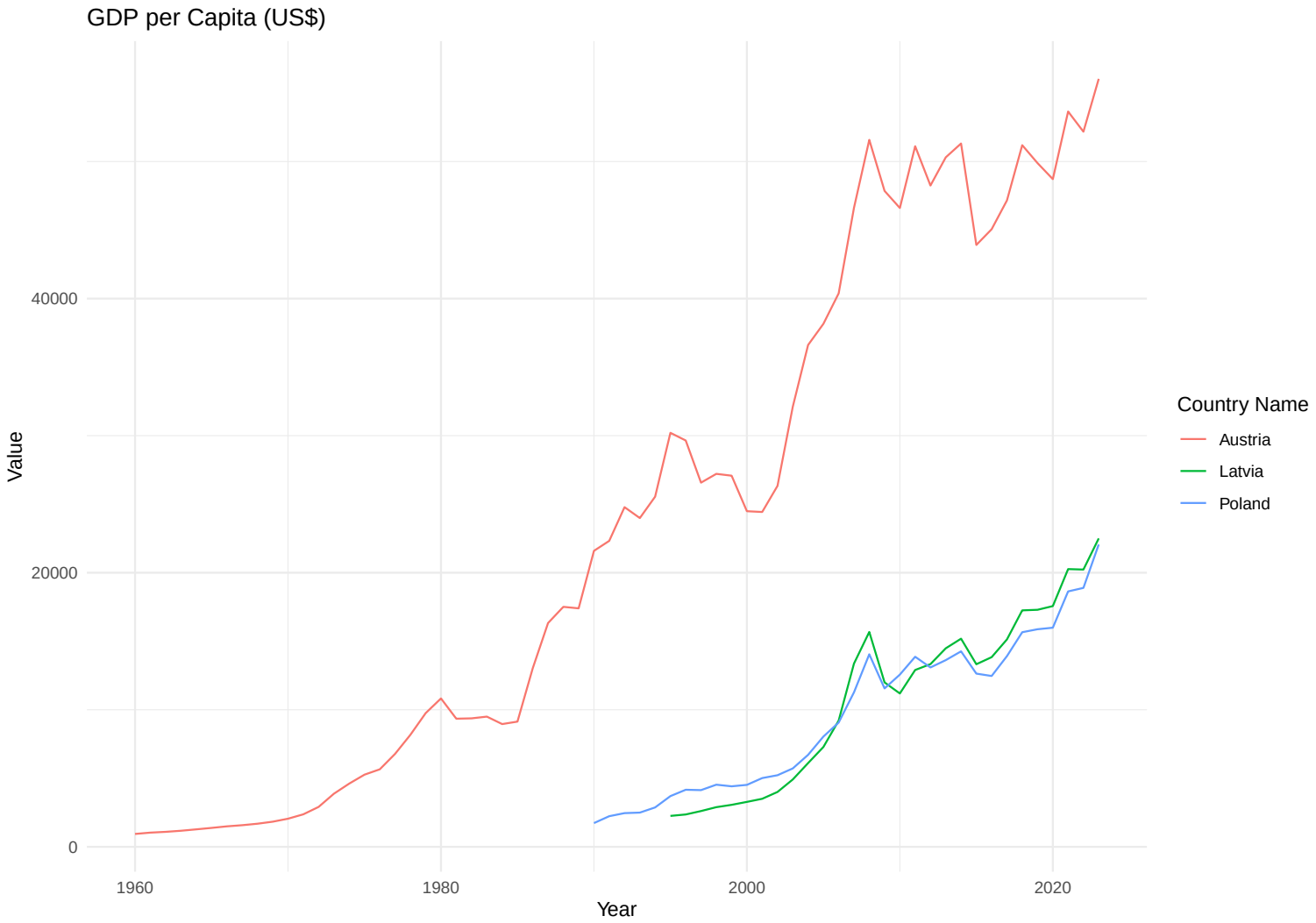


Interpretation

The chart shows a clear upward trajectory in life expectancy across all three countries (Austria, Latvia, and Poland) from 1960 to 2020, reflecting global improvements in healthcare and living standards.

- **Austria** maintains the highest life expectancy throughout the period (reaching ~80 years by 2020)
- **Latvia** and **Poland** show parallel growth patterns, with Latvia slightly ahead in recent years
- A noticeable dip occurs around 1990-2000 for Latvia/Poland, likely reflecting post-Soviet transition challenges
- The most rapid improvements occurred between 1960-1980
- Convergence trend appears post-2000 as Latvia/Poland narrow the gap with Austria

GDP per Capita



Interpretation

The chart shows **steady growth in GDP per capita** from 1980 to 2000, indicating economic expansion over two decades.

- GDP per capita **doubled** from ~$20,000 (1980) to ~$40,000 (2000), reflecting significant improvements in productivity or income distribution.
- A sharp rise occurs after 2000, driven largely by Austria high GDP values.
- Fluctuations post-2008 reflect the global financial crisis, and the dips around 2020 likely capture the COVID-19 economic impact for Austria where it doesn't affect much for Latvia and Poland.

Thank You