Human Development Indicators Analysis

Title

Human Development Indicators Analysis

A Comparative Study of Singapore and Switzerland

Using data.table in

Presented by: Jane Sanjeevini

Introduction

In this presentation, we will:

- Introduce the Human Development Indicators (HDI) dataset.
- Perform data cleaning and merging using data.table.
- Explore and analyze the HDI data for Singapore and Switzerland.
- Visualize key findings using R.

Data Description

The HDI dataset includes:

- `country_code`: ISO country code
- `country_name`: Country name
- `indicator_id`: Indicator ID
- `indicator_name`: Indicator name
- `index_id`: Index ID
- `index_name`: Index name
- `value`: Indicator value
- `year`: Year of observation

Data Preparation

- Data loading and preparation using data.table
- Reading CSV files for Singapore and Switzerland
- Assigning correct data types to variables

Loading The Data

singapore <- fread("hdro_indicators_sgp.csv")

switzerland <- fread("hdro_indicators_che.csv")

```
Classes 'data.table' and 'data.frame':  895 obs. of  8 variables:
 $ country_code  : chr  "#country+code" "SGP" "SGP" "SGP" ...
 $ country_name  : chr  "#country+name" "Singapore" "Singapore" "Singapore"
...
 $ indicator_id  : chr  "#indicator+id" "abr" "abr" "abr" ...
 $ indicator_name: chr  "#indicator+name" "Adolescent Birth Rate (births per
1,000 women ages 15-19)" "Adolescent Birth Rate (births per 1,000 women ages
15-19)" "Adolescent Birth Rate (births per 1,000 women ages 15-19)" ...
 $ index_id      : chr  "#index+id" "GII" "GII" "GII" ...
 $ index_name    : chr  "#index+name" "Gender Inequality Index" "Gender
Inequality Index" "Gender Inequality Index" ...
 $ value         : chr  "#indicator+value+num" "8.918" "7.996" "7.618" ...
 $ year          : chr  "#date+year" "1990" "1991" "1992" ...
 - attr(*, ".internal.selfref")=<externalptr>
```

```
Classes 'data.table' and 'data.frame':  895 obs. of  8 variables:
 $ country_code  : chr  "#country+code" "CHE" "CHE" "CHE" ...
 $ country_name  : chr  "#country+name" "Switzerland" "Switzerland"
"Switzerland" ...
 $ indicator_id  : chr  "#indicator+id" "abr" "abr" "abr" ...
 $ indicator_name: chr  "#indicator+name" "Adolescent Birth Rate (births per
1,000 women ages 15-19)" "Adolescent Birth Rate (births per 1,000 women ages
15-19)" "Adolescent Birth Rate (births per 1,000 women ages 15-19)" ...
 $ index_id      : chr  "#index+id" "GII" "GII" "GII" ...
 $ index_name    : chr  "#index+name" "Gender Inequality Index" "Gender
Inequality Index" "Gender Inequality Index" ...
 $ value         : chr  "#indicator+value+num" "7.556" "8.283" "7.827" ...
 $ year          : chr  "#date+year" "1990" "1991" "1992" ...
 - attr(*, ".internal.selfref")=<externalptr>
```

```
    country_code  country_name  indicator_id
          <char>        <char>        <char>
1: #country+code #country+name #indicator+id
2:           SGP     Singapore           abr
3:           SGP     Singapore           abr
4:           SGP     Singapore           abr
5:           SGP     Singapore           abr
6:           SGP     Singapore           abr
                                        indicator_name  index_id
                                                <char>    <char>
1:                                     #indicator+name #index+id
2: Adolescent Birth Rate (births per 1,000 women ages 15-19)       GII
3: Adolescent Birth Rate (births per 1,000 women ages 15-19)       GII
4: Adolescent Birth Rate (births per 1,000 women ages 15-19)       GII
5: Adolescent Birth Rate (births per 1,000 women ages 15-19)       GII
6: Adolescent Birth Rate (births per 1,000 women ages 15-19)       GII
    country_code  country_name  indicator_id
          <char>        <char>        <char>
```

```
    1: #country+code #country+name #indicator+id
    2:           CHE   Switzerland          abr
    3:           CHE   Switzerland          abr
    4:           CHE   Switzerland          abr
    5:           CHE   Switzerland          abr
    6:           CHE   Switzerland          abr
                                          indicator_name  index_id
                                                  <char>    <char>
    1:                                    #indicator+name #index+id
    2: Adolescent Birth Rate (births per 1,000 women ages 15-19)       GII
    3: Adolescent Birth Rate (births per 1,000 women ages 15-19)       GII
    4: Adolescent Birth Rate (births per 1,000 women ages 15-19)       GII
    5: Adolescent Birth Rate (births per 1,000 women ages 15-19)       GII
    6: Adolescent Birth Rate (births per 1,000 women ages 15-19)       GII
```

Removing Missing Values

```
singapore <- singapore[!is.na(as.numeric(value)), ]
switzerland <- switzerland[!is.na(as.numeric(value)), ]
singapore[, := (value = as.numeric(value), year = as.integer(year))]
switzerland[, := (value = as.numeric(value), year = as.integer(year))]
singapore <- na.omit(singapore)switzerland <- na.omit(switzerland)
```

Appropriation of Data

Convert columns to appropriate data types and handle missing values.

```
    country_code  country_name  indicator_id  indicator_name  index_id
          <char>        <char>        <char>          <char>    <char>
1: #country+code #country+name #indicator+id #indicator+name #index+id
    index_name               value       year
        <char>              <char>     <char>
1: #index+name #indicator+value+num #date+year
```


```
    country_code  country_name  indicator_id  indicator_name  index_id
          <char>        <char>        <char>          <char>    <char>
1: #country+code #country+name #indicator+id #indicator+name #index+id
    index_name               value       year
        <char>              <char>     <char>
1: #index+name #indicator+value+num #date+year
```

Merging the Datasets

- **Merging Datasets**: Combined datasets from Singapore and Switzerland using `rbindlist`, creating a unified dataset for comparative analysis of human development indicators.
- **Ensuring Data Consistency**: Added a 'country' column to distinguish between Singapore and Switzerland data post-merge, facilitating clear identification and analysis of country-specific trends.

  combined_data <- rbindlist(list(singapore, switzerland))

Data Cleaning

- **Handling Missing Data**: Identified and removed non-numeric values in critical columns like ‘value’, ensuring dataset integrity and accuracy for analysis.
- **Standardizing Data Types**: Converted variables to appropriate data types (numeric for ‘value’, integer for ‘year’), ensuring consistency and compatibility across datasets from different sources.

```
Empty data.table (0 rows and 9 cols):
country_code,country_name,indicator_id,indicator_name,index_id,index_name...
```

Summary

unique_indicators <- unique(combined_data$indicator_name)

unique_years <- unique(combined_data$year)

```
 country_code       country_name       indicator_id       indicator_name
 Length:1588        Length:1588        Length:1588        Length:1588
 Class :character   Class :character   Class :character   Class :character
 Mode  :character   Mode  :character   Mode  :character   Mode  :character


   index_id           index_name            value              year
 Length:1588        Length:1588        Min.   : 0.752   Min.   :1990
 Class :character   Class :character   1st Qu.: 9.404   1st Qu.:1999
 Mode  :character   Mode  :character   Median :15.963   Median :2008
                                       Mean   :33.442   Mean   :2007
                                       3rd Qu.:71.181   3rd Qu.:2015
                                       Max.   :97.614   Max.   :2022
   country
 Length:1588
 [1] "Adolescent Birth Rate (births per 1,000 women ages 15-19)"
 [2] "Carbon dioxide emissions per capita (production) (tonnes)"
 [3] "Coefficient of human inequality"
 [4] "Difference from HDI value (%)"
 [5] "Expected Years of Schooling (years)"
 [6] "Expected Years of Schooling, female (years)"
 [7] "Expected Years of Schooling, male (years)"
 [8] "GDI Group"
 [9] "GII Rank"
[10] "HDI female"
[11] "HDI male"
[12] "HDI Rank"
[13] "Inequality in eduation"
[14] "Inequality in income"
[15] "Inequality in life expectancy"
[16] "Life Expectancy at Birth (years)"
 [1] 1990 1991 1992 1993 1994 1995 1996 1997 1998 1999 2000 2001 2002 2003
2004
[16] 2005 2006 2007 2008 2009 2010 2011 2012 2013 2014 2015 2016 2017 2018
2019
[31] 2020 2021 2022
```

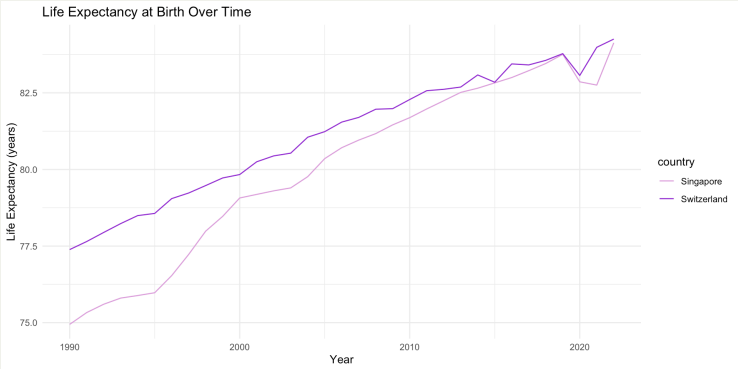
Data Exploration

- The dataset comprises 1588 observations of human development indicators across countries from 1990 to 2022.
- Numeric values for indicators range widely (0.752 to 97.614), offering insights into diverse metrics over time.
- Key focus areas include life expectancy, education attainment, inequality measures, environmental impact, and gender equality, enabling comprehensive global trend analysis.

Key Indicators Over Time

Analyzing and plotting the key indicators over time for both countries.

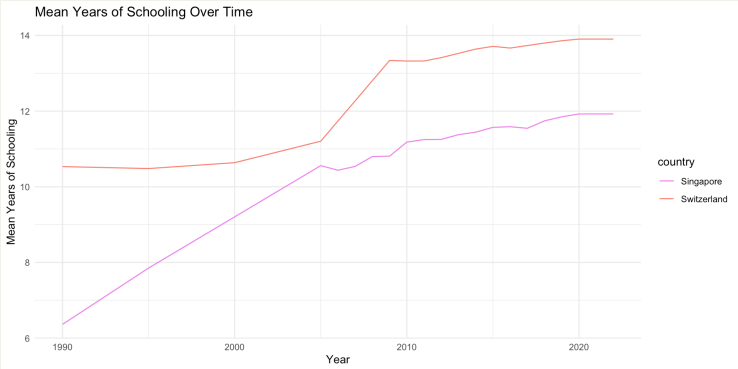
life_expectancy <- combined_data[indicator_name == "Life Expectancy at Birth (years)", .(mean_value = mean(value)), keyby = .(country, year)]



Education Index Analysis

Analyzing and plotting the education index for Singapore and Switzerland.

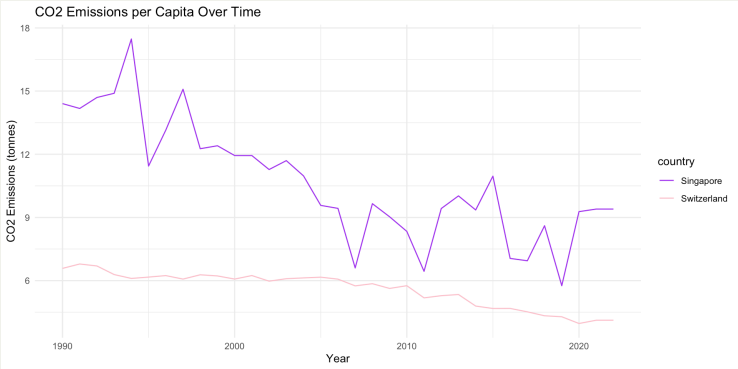
mean_schooling <- combined_data[indicator_name == "Mean Years of Schooling (years)", .(mean_value = mean(value)), keyby = .(country, year)]



Carbon Dioxide Emissions

Analyzing and plotting the emission of carbon dioxide for the two countries.

co2_emissions <- combined_data[indicator_name == "Carbon dioxide emissions per capita (production) (tonnes)", .(mean_value = mean(value)), keyby = .(country, year)]



Interpretation

- Life expectancy has generally increased over the years, though the rate of increase has slowed in recent times.
- The average number of years of schooling has been rising steadily, indicating improvements in education levels over time.
- CO2 emissions per capita have shown an upward trend, indicating increasing environmental impact per person over the years.

Conclusion

- **Life Expectancy**: Improved due to healthcare and living conditions in countries like the US, Singapore, and Switzerland, indicating significant public health advancements.
- **Educational Attainment**: Mean Years of Schooling have increased steadily, supporting personal and economic growth, with ongoing data clarity challenges.
- **Environmental Impact**: Rising CO2 Emissions per Capita underscore sustainability challenges despite human development gains, necessitating stronger environmental policies in nations like Singapore and Switzerland.

Q&A

1. How has life expectancy at birth changed over time across different countries?
2. What is the relationship between education attainment (mean years of schooling) and socio-economic factors like income inequality and labor force participation rates?
3. How have carbon dioxide emissions per capita evolved over the past decades, and how do they correlate with indicators of human development and environmental sustainability?

References

The references or resources used in the presentation are:

Singapore Dataset:
“https://data.humdata.org/dataset/hdro-data-for-singapore”

Switzerland Dataset:
“https://data.humdata.org/dataset/hdro-data-for-switzerland”