

Assignment 1

Shikan Chen & 23202508

Download the Data

Countries selected: Singapore and Ireland

Read the Data

```
1 # Load the required package
2 library(data.table)
3 library(ggplot2)
4
5 # Read the data for Singapore and Ireland, removing the first row as the metadata
6 spg_data <- fread("hdro_indicators_sgp.csv")[-1,]
7 irl_data <- fread("hdro_indicators_irl.csv")[-1,]
```

Peak of spg_data

```
1 head(spg_data)
  country_code country_name indicator_id
1:          SGP      Singapore          abr
2:          SGP      Singapore          abr
3:          SGP      Singapore          abr
4:          SGP      Singapore          abr
5:          SGP      Singapore          abr
6:          SGP      Singapore          abr

                                indicator_name index_id
1: Adolescent Birth Rate (births per 1,000 women ages 15-19)      GII
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

                                index_name value year
1: Gender Inequality Index 8.918 1990
2: Gender Inequality Index 7.996 1991
3: Gender Inequality Index 7.618 1992
4: Gender Inequality Index 7.418 1993
- - - - -
```

Peak of irl_data

```
1 head(irl_data)

  country_code country_name indicator_id
1:          IRL      Ireland          abr
2:          IRL      Ireland          abr
3:          IRL      Ireland          abr
4:          IRL      Ireland          abr
5:          IRL      Ireland          abr
6:          IRL      Ireland          abr

                                indicator_name index_id
1: Adolescent Birth Rate (births per 1,000 women ages 15-19)      GII
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

      index_name  value year
1: Gender Inequality Index 15.814 1990
2: Gender Inequality Index 16.584 1991
3: Gender Inequality Index 16.457 1992
4: Gender Inequality Index 15.454 1993
- - - - -
```

Merge the data

We need to assign the Correct Class to Variables for merging

```
1 # Merge the datasets
2 merged_data <- unique(rbind(spg_data, irl_data, use.names=TRUE))
```

Quick Data Exploration

Summary statistics

Unique indicator names

Range of years in the dataset

Number of records per country

Summary statistics

```
1 summary(merged_data)
```

country_code	country_name	indicator_id	indicator_name
Length:1788	Length:1788	Length:1788	Length:1788
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character
index_id	index_name	value	year
Length:1788	Length:1788	Length:1788	Length:1788
Class :character	Class :character	Class :character	Class :character
Mode :character	Mode :character	Mode :character	Mode :character

Unique indicator names

```
1 # Unique indicator names
2 unique_indicators <- unique(merged_data$indicator_name)
3 print(unique_indicators)
```

```
[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] "Gross National Income Per Capita, female (2017 PPP$)"
[11] "Gross National Income Per Capita, male (2017 PPP$)"
[12] "Gross National Income Per Capita (2017 PPP$)"
[13] "HDI female"
[14] "HDI male"
[15] "HDI Rank"
[16] "Inequality in education"
[17] "Inequality in income"
[18] "Inequality in life expectancy"
[19] "Life Expectancy at Birth (years)"
```

Range of years in the dataset

```
1 # Range of years in the dataset
2 year_range <- range(merged_data$year, na.rm = TRUE)
3 print(year_range)
```

```
[1] "1990" "2022"
```

Number of records per country

```
1 # Number of records per country
2 records_per_country <- merged_data[, .N, by = country_name]
3 print(records_per_country)
```

```
country_name    N
1:   Singapore 894
2:    Ireland 894
```

From Quick Data Exploration Results

We note that

- The dataset contains 1,788 rows and 8 columns.
- All columns are character types. We'll need to convert some columns to numeric or integer types for further analysis.
- Unique indicator names and years range from 1990 to 2022.
- The number of records per country:
 - Singapore: 894 records
 - Ireland: 894 records

Data Analysis

Aggregate data for annual average Adolescent Birth Rate in Singapore

```
1 # Ensure 'value' is numeric
2 merged_data[, value := as.numeric(value)]
3
4 # Calculate the annual average Adolescent Birth Rate in Singapore
5 annual_avg_adol_birth_rate <- merged_data[indicator_name == "Adolescent Birth Rate (births per
6
7 # Print the annual average data
8 head(annual_avg_adol_birth_rate)
```

	year	Average_Birth_Rate
1:	1990	8.918
2:	1991	7.996
3:	1992	7.618
4:	1993	7.418
5:	1994	6.755
6:	1995	7.148

Analysis

Overall Trend:

- The adolescent birth rate has shown a consistent decline over the 32-year period.
- Starting from a rate of 8.918 births per 1000 women aged 15-19 in 1990, it reduced to 2.462 by 2022.

Rate of Decline:

- The rate of decline was not uniform but was more pronounced in the later years. For example, from 2016 to 2017, the drop from 2.393 to 2.275 represents one of the sharper year-over-year declines in the dataset.
- The largest year-over-year percentage decrease appears to be after 2015, highlighting potentially effective interventions or changes in societal behavior around that time.

Aggregate data for average of each indicator by gender across years

```
1 # Calculate the average of each indicator by gender across years
2 gender_analysis <- merged_data[indicator_name %in% c("Labour force participation rate, female (",
3
4 # Output the result
5 print(gender_analysis)
```

	indicator_name	country_name
1:	Labour force participation rate, female (% ages 15 and older)	Ireland
2:	Labour force participation rate, female (% ages 15 and older)	Singapore
3:	Labour force participation rate, male (% ages 15 and older)	Ireland
4:	Labour force participation rate, male (% ages 15 and older)	Singapore
5:	Mean Years of Schooling, female (years)	Ireland
6:	Mean Years of Schooling, female (years)	Singapore
7:	Mean Years of Schooling, male (years)	Ireland
8:	Mean Years of Schooling, male (years)	Singapore

	Average_Value
1:	49.735758
2:	51.984848
3:	70.069394
4:	73.623939
5:	10.346121
6:	9.512333
7:	10.017273
8:	10.469758

Analysis

Labor Force Participation Rate:

- Singapore: Males (73.62%) have a higher participation rate than females (51.98%).
- Ireland: Similar to Singapore, males (70.07%) have a higher participation rate than females (49.74%).
- Comparison: The gender gap in labor force participation is slightly wider in Singapore than in Ireland.

Mean Years of Schooling:

- Singapore: Males (10.47 years) have slightly more schooling on average compared to females (9.51 years).
- Ireland: Females (10.35 years) have nearly the same years of schooling as males (10.02 years).
- Comparison: Ireland shows less disparity in education between genders

**Plot using some output from
the analysis**

Plot 1

for the differences in labor force participation and educational attainment between males and females in both countries

```
1 library(dplyr)
2
3 # Transform data for plotting
4 plot_data <- gender_analysis %>%
5   mutate(Gender = ifelse(grepl("female", indicator_name), "Female", "Male"),
6          Indicator = gsub(".$", "", indicator_name)) %>%
7   select(country_name, Gender, Indicator, Average_Value)
8
9 # Create the plot
10 ggplot(plot_data, aes(x = Indicator, y = Average_Value, fill = Gender)) +
11   geom_bar(stat = "identity", position = position_dodge()) +
12   facet_wrap(~country_name) +
13   scale_fill_brewer(palette = "Pastell1") +
14   labs(title = "Gender Comparison in Labor Force Participation and Education",
15        x = "Indicator",
16        y = "Average Value (%)") +
17   theme_minimal() +
18   theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



Analysis

Labor Force Participation:

- Ireland: Males show a significantly higher labor force participation rate compared to females.
- Singapore: A similar trend is observed; however, the gender gap is somewhat less pronounced than in Ireland, yet still substantial.

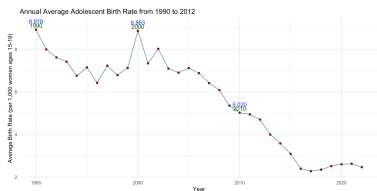
Mean Years of Schooling:

- Ireland: The mean years of schooling are slightly higher for males than for females, indicating a marginal gender disparity in educational attainment.
- Singapore: Contrary to Ireland, males in Singapore also have higher mean years of schooling than females, with a noticeable difference suggesting a persistent educational gap.

Plot 2

for point comparison of gender indicators between ireland and Singapore

```
1 # Ensure to numeric
2 annual_avg_adol_birth_rate$year <- as.numeric(annual_avg_adol_birth_rate$year)
3
4 # Create the line plot
5 ggplot(annual_avg_adol_birth_rate, aes(x = year, y = Average_Birth_Rate)) +
6   geom_line(color = "steelblue") +
7   geom_point(color = "darkred") +
8   labs(title = "Annual Average Adolescent Birth Rate from 1990 to 2012",
9        x = "Year",
10       y = "Average Birth Rate (per 1,000 women ages 15-19)") +
11   theme_minimal() +
12   # Annotations to highlight specific years or changes
13   geom_text(aes(label = ifelse(year %in% c(1990, 2000, 2010), as.character(year), "")),
14            nudge_y = 0.2, check_overlap = TRUE, color = "darkgreen") +
15   geom_text(aes(label = ifelse(year %in% c(1990, 2000, 2010),
16                               sprintf("%.3f", Average_Birth_Rate), ""),
17            year, Average_Birth_Rate),
18            vjust = -1.5, hjust = 0.5, check_overlap = TRUE, color = "blue")
```



Analysis

- The graph shows a steady decrease in the adolescent birth rate over the 22-year period, from 8.918 births per 1,000 women in 1990 to 4.700 births per 1,000 women in 2012.
- There are noticeable fluctuations in the early years, particularly a sharp increase in 2000 when the rate peaked at 8.863 before beginning a more pronounced decline.
- The year 2010 marks the onset of the most rapid declines in the series, reaching rates below 5 births per 1,000 by the end of the period observed.