DIGITAL HEALTH

Name (student): PHAN NHAT HOANG LINH

Matriculation Number: 22508311

Name of study: Master of Global Public Health

Project: Public Health Data Visualization Dashboard

Topic: Total fertility rate (TFR): estimates for 1950-2021 and projections for 2022-2100

1. USE CASE

At the global level, from 1950 to 2021, the total fertility rate (TFR) declined by more than 50%, decreasing from 4.84 (4.63-5.06) to 2.23 (2.09-2.38). This downward trend is expected to continue, with the global TFR likely to fall below the replacement level of 2.1 by 2030 (a TFR of 2.1 represents the level at which a population replaces itself from one generation to the next). This decline may lead to negative population growth in the long term.

At the national level, some countries have experienced particularly sharp declines. For example, in the Republic of Korea, the TFR fell nearly sevenfold, from 5.72 (5.37-6.08) in 1950 to just 0.82 (0.75-0.89) in 2021. Without effective government intervention, this decline is projected to continue through 2100.

The TFR is a crucial demographic and public health indicator. It represents the average number of children a hypothetical cohort of women would have by the end of their reproductive period if they were subject throughout their lives to the age-specific fertility rates of a given period and were not subject to mortality.

Declining TFRs will directly alter the future age structure of a country's population, with far-reaching implications for healthcare demand and supply, education and labor systems, gender equality, family planning services, and social support structures.

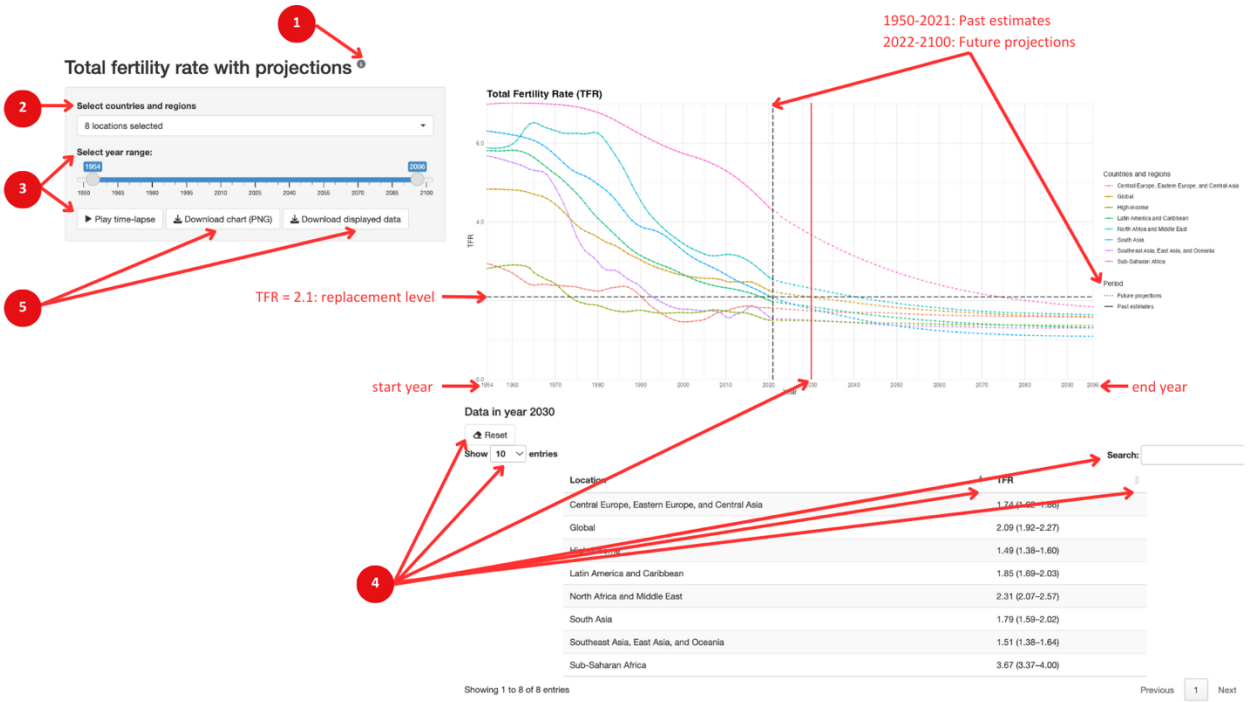
Therefore, accurate assessment of current and projected TFRs is essential for policymakers and healthcare professionals at both global and national levels to inform effective social, economic, and environmental policies.

In 2024, the Global Burden of Disease Study 2021 (GBD 2021), coordinated by the Institute for Health Metrics and Evaluation (IHME), published datasets estimating the burden of diseases, injuries, and risk factors for 204 countries, territories, and selected subnational locations.

These datasets are highly comprehensive and have been used in some scientific papers and reports. However, they are primarily presented as static images, which limits user interaction and customization. For example, a policymaker in a specific country may wish to view data only for their country, compare it with regional and global trends, and focus on a particular time period. These tasks cannot be easily supported by static images. An interactive data visualization dashboard is therefore needed to meet the diverse needs of users.

This dashboard is based on estimates for the period 1950-2021, with projections for 2022-2100 provided by IHME. Its purpose is to visualize TFR trends using interactive line charts and includes five main features: (1) a TFR definition popover, (2) country and region filters, (3) a selectable year range and time-lapse function, (4) hover-enabled data point displays, and (5) chart and dataset download options. These features enable policymakers and healthcare professionals to examine historical TFR trends and anticipate future patterns over the coming decades, both within individual countries and in comparison with other regions worldwide.

2. SOLLUTION



Strategies for building the solution

- A line chart was chosen because TFR data changes continuously over time, and users are interested in understanding long-term trends.

- Two different line styles were selected to represent two distinct periods: 1950-2021 for estimated data and 2022-2100 for projected data.

- The TFR value of 2.1 was highlighted, as it represents the replacement level of total fertility rate.

- A four-level location filter was created to handle the large number of locations (231 in total):

  - 8 priority selections: based on references from published scientific papers on this topic.
  - 1-50 selections: location names remain unchanged.
  - More than 50 selections: location names are abbreviated to three letters. Although this does not reduce chart complexity, it is useful when hovering over the chart to view data for a specific year, as I will mention below.
  - 0 selections: a warning message is displayed to the user.

- A year-range selector from 1950 to 2100 was added to allow users to freely choose their preferred time period.

- Inspired by several existing dashboards, a time-lapse feature was introduced to highlight changes over time and help users visually focus on year-by-year transitions. During implementation, several issues were identified:

  - When the year range reached 2100 (the maximum year), the time-lapse stopped functioning. To resolve this, an additional reset function (resetting the range to 1950-1951) was implemented.
  - The chart did not display the minimum and maximum years of the selected range during the time-lapse. This functionality was added.

- The year format included unwanted spaces "1 950 ", which were corrected to "1950".
- A download button was added to allow users to download charts for use in their own products or to download datasets for further research.

- A hover feature was implemented to display data for a specific year directly on the chart, providing users with more detailed information. However, several issues occurred:

- The hover information was expected to disappear when the mouse cursor left the chart area, but this did not occur. As a result, a manual reset button was added.
- The data table below the chart is currently quite basic. It is recommended to add search, sorting, and an option to control the number of displayed entries.

- Finally, a function was added to define the Total Fertility Rate (TFR) for beginners, and the data source was clearly stated.

3. IMPLEMENTATION

3.1. Data source

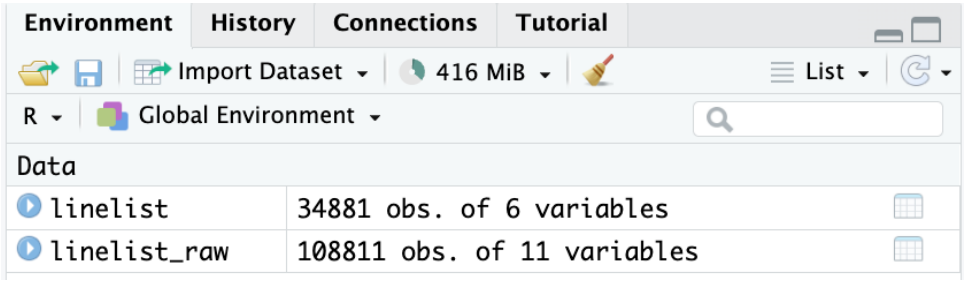
https://ghdx.healthdata.org/record/ihme-data/global-burden-disease-study-2021-gbd-2021-fertility-1950-2100

3.2. Data cleaning

The source code is located in ***cleaning_data.R***

The data cleaning process includes the following steps:

- Standardize column name syntax.
- Select relevant columns.
- Remove duplicate entries.
- Verify column classes (no need because they are already correct).
- Validate value coding (no need because they are already correct).
- Filter the dataset to "Reference" scenarios.



After running ***cleaning_data.R***, the Environment pane will appear on the right side of the screen, as shown above. The purpose of separating the data-cleaning code with the Shiny code is to store the ***linelist_raw*** and ***linelist*** and easy to determine what data ***linelist_raw*** and ***linelist*** contain.

3.3. Shiny dashboard

The source code is located in ***shiny.R***

3.4. Pushing code to GitHub

https://github.com/hoanglinh251205/DigitalHealth2025