Visualisation of Singapore's Fertility Crisis

Akram, M., Guo, Z., Tan, G., Cheong, W. and Chew, T.

Keywords — AAA, BBB, CCC

I. Introduction

Singapore faces a demographic crisis with one of the world's lowest fertility rates. Understanding the underlying socioeconomic factors is crucial for policy formulation and national planning. This project analyses **three decades of fertility and labour force data** to identify patterns and relationships that visualisations from the source neglects. Using various packages in R, we will create a poster that thoughtfully displays the socioeconomic factors that influence fertility/birth rates in Singapore by using fertility rate data sourced from different websites.

> ⓘ Research Focus
>
> Understanding the underlying **socioeconomic factors** is crucial for policy formulation and national demographic sustainability.

II. Original Visualisation
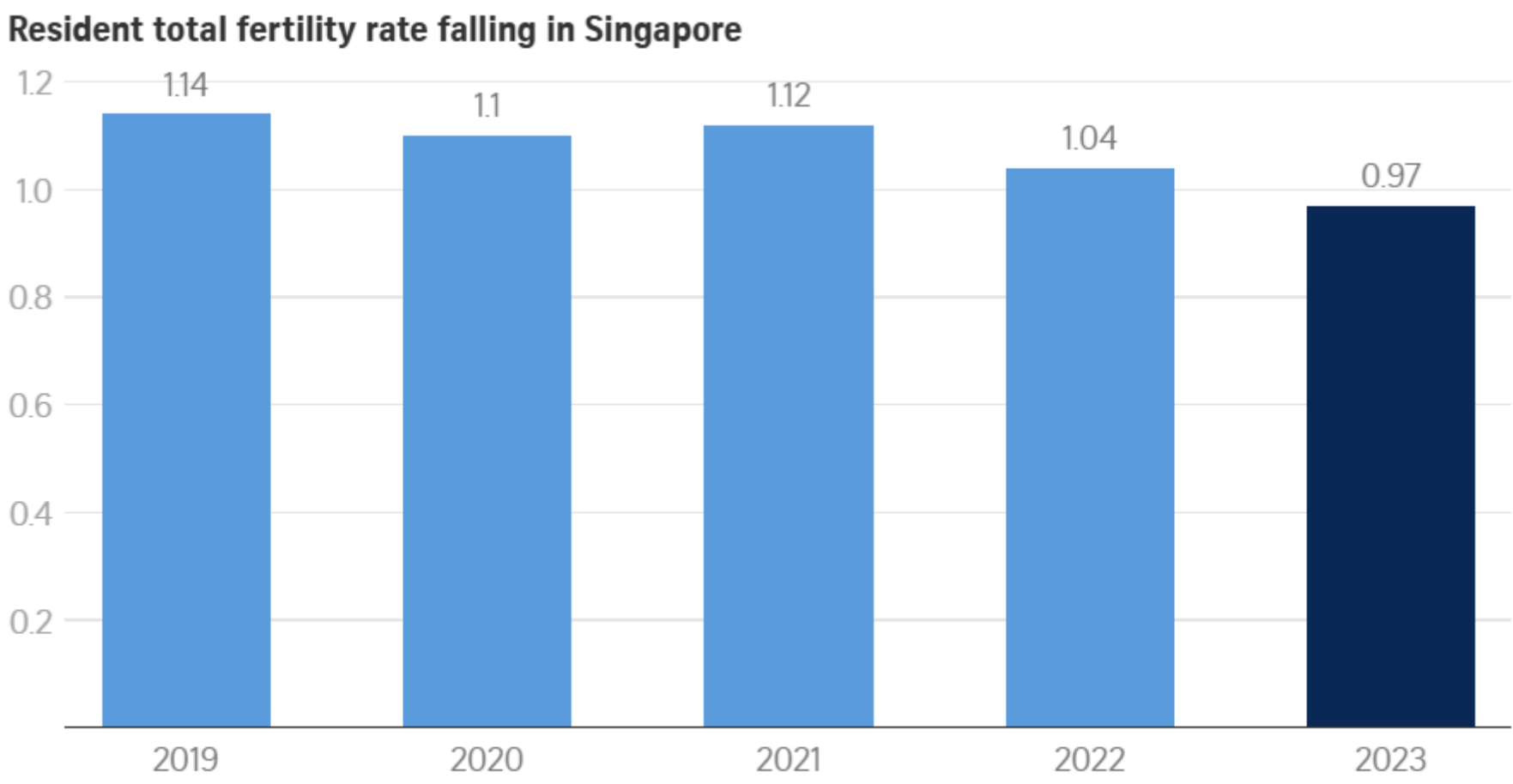
Figure 1: Total fertility rate from 2019 to 2023

III. Original Visualisation Analysis

1. **No data validation**: There is no data validation provided in the graph.
2. **Limited range:** The graph only displays data from 2019-2023. This is not a wide enough of a range to make conclusive statements.
3. **Missing socioeconomic factors**: There are no socioeconomic factors listed to support the decline in the fertility rate. They are crucial information that can reinforce the nature of the graph.

---

4. **Static visualisation**: The original visualisation lacks depth and interactive features that would otherwise allow the user to better understand the graphs. The static natures does not allow the user to hover or click on data points to gain more information.

IV. Suggested Improvements

1. **Inlcude data validation**: Add comprehensive data validation along with outlier analysis. This aims to ensure the accuracy and reliability of the data presented.
2. **Extended analysis**: Increase the range of the data to 1990-2022. This will provide a more comprehensive view of the fertility trends in Singapore which allows for better analysis and understanding of long-term patterns.
3. **Integrate socioeconomic factors**: Integrate socioeconomic factors such as labour force participation and marital status. This will provide a more holistic view of the factors influencing fertility rates. This allows for better policy formulation and understanding of the demographic trends.
4. **Dynamic visualisation**: Ensure that the visualisation consists of a fully interactive dashboard. This will allow users to interact with the data, such as hovering over data points to see more information. They can also filter by different socioeconomic factors and viewing age-specific fertility rates. This enhances user engagement and understanding of the data.

V. Implementation

i. Data Sources

Dataset	Source	Time Period	Variables	Records
Fertility Rates	SingStat	1960-2024	Age-specific fertility rates, Total fertility rate	17 variables wide format
Labour Force (Working)	data.gov.sg	1991-2022	Female labour force by age & marital status	5 columns long format
Labour Force (Not working)	data.gov.sg	1991-2022	Females outside labour force by age & marital status	5 columns long format

ii. Software
- `crosstalk` – Enables interactivity between HTML widgets
- `tidyverse` – Loads core tidy data science packages like `ggplot2`, `dplyr` and `tidyr`
- `viridis` – Provides colourblind-friendly color palettes for plots
- `ggpp` – Adds plot annotations like equations and labels in `ggplot2`
- `ggrepel` – Prevents overlapping text labels in `ggplot2` plots
- `RColorBrewer` – Offers pre-made color palettes for maps and plots
- `htmltools` – Tools for creating and customizing HTML content in R
- `dplyr` – Grammar of data manipulation
- `knitr` – For dynamic report generation
- `tools` – Base R utilities for package and file management
- `ggiraph` – Adds tooltips and interactivity to `ggplot2` plots
- `ggplot2` – Core package for creating elegant plots
- `plotly` – Converts static plots to interactive plots
- `janitor` – Cleans messy data
- `gt` – Creates beautiful tables for reporting
- `stringr` – Simplifies string operations
- `scales` – Formatting scales and labels in visualisations
- `forcats` – Handles categorical variables more easily

- `DT` – R interface to interactive DataTables (tables with filters/sorting).
- `glue` – Embeds R expressions in strings using `{}`
ii. Workflow
1) Exploratory Data Analysis:
2) Feature Engineering:

i. Data Cleaning and Reshaping Workflow
- Handle all missing values by converting `"na"` and `"-"` strings to `NA`.
- Standardise age bands by renaming columns to consistent labels like `"15-19"` and align across datasets.
- Filter data by keeping only years from **1990 to 2022** to match across fertility and labour datasets.
- Pivot fertility data from wide to long format for year-wise plotting.
- Rename age to `age_band`, clean column names using consistent formatting.
- Introduce `uom` (unit of measurement) column to specify rate scaling.
- Divide labour force counts by 1,000 to match y-axis scale.
- Keep age-specific fertility rates and Total Fertility Rate (TFR).
- Add `"All"` age group to show total counts by year and marital status for plots and filters.
- Use the same cleaning logic for `fertility`, `not_working`, and `work` tibbles for consistency.
- Check for missing data and outliers.
- Join the datasets to create a single tibble with all the necessary information for analysis.
3) Data Visualisation:
- Define Colors: Create a color palette representing the socioeconomic factors influencing fertility rates such as labour force participation and marital status.
- Graph Properties: Configure interactivity by allowing user to click and zoom on the data points.
- Layout: Set the title and overall layout properties for an informative and visually appealinggraph.

VI. Improved Visualisation
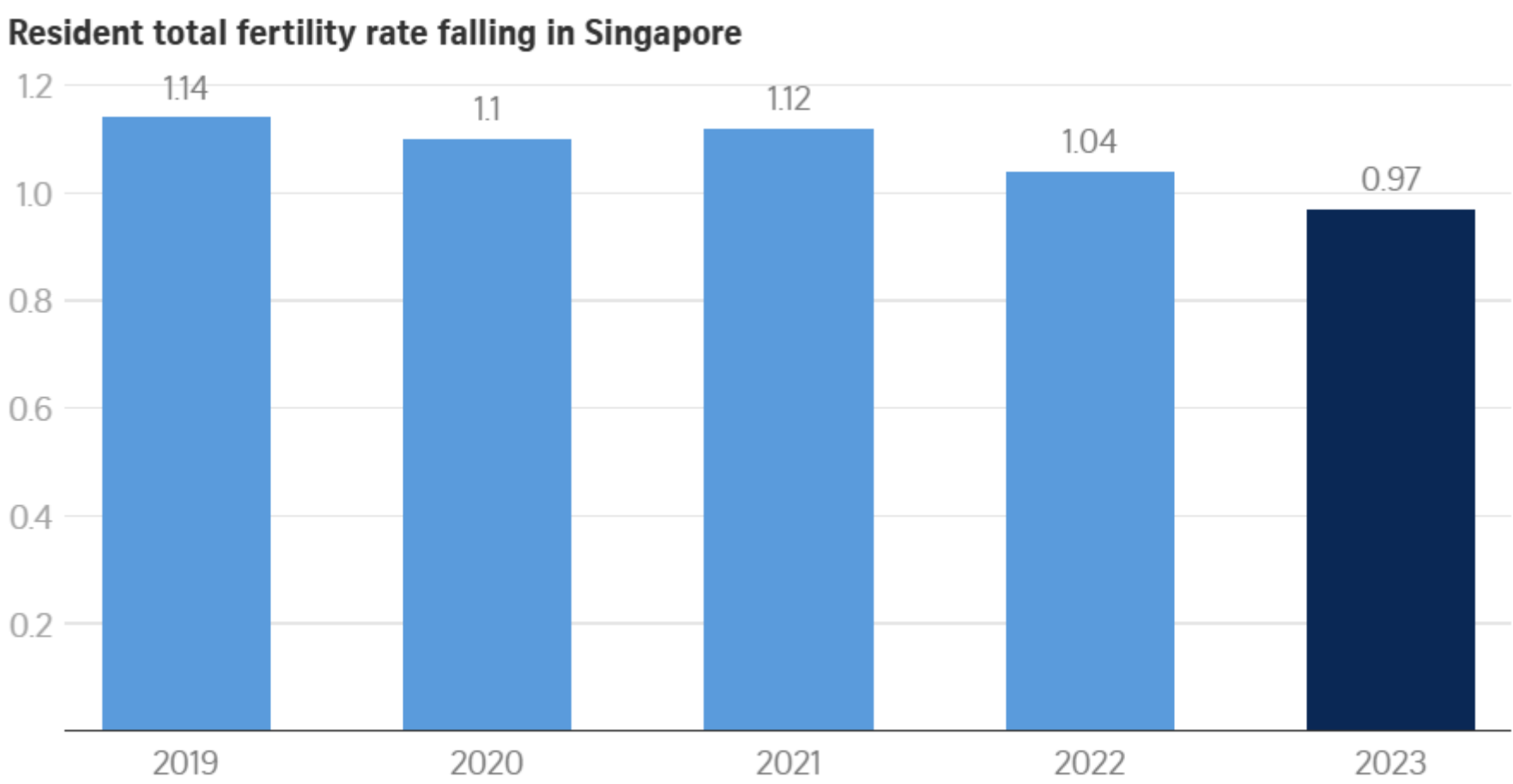
Figure 2: Total fertility rate from 2019 to 2023

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magnam aliquam quaerat voluptatem. Ut enim

:::