Meta-Document

Why Did You Choose This Topic?

Life expectancy is one of the most important metrics in global health, reflecting a combination of healthcare quality, socio-economic progress, and public health policies. It offers valuable insights into disparities between countries and over time. Additionally, examining gender differences in life expectancy reveals societal and biological factors affecting longevity. This topic was chosen because it is globally relevant and resonates with a wide audience interested in understanding health and development trends. The dataset from TidyTuesday provides a comprehensive and reliable source for exploring these patterns across decades.

Intended Audience

The intended audience for this project is a general global audience, including readers of health blogs, educators, policymakers, and individuals interested in data storytelling. The goal is to make the analysis accessible to non-specialists by using clear visuals and straightforward language. The blog post focuses on global insights and trends, emphasizing the human stories behind the data.

Data Source

The dataset was sourced from the TidyTuesday project (December 5, 2023). It contains:

- Historical life expectancy data (1950–2021) for countries and regions worldwide.

- Gender-specific life expectancy data, allowing for an analysis of disparities between female and male populations.

Dataset Links: Life Expectancy Dataset

The dataset is well-structured and includes numeric variables for life expectancy and categorical variables for countries and regions.

Data Cleaning and Wrangling

Tools Used:

- R Programming Language, specifically the dplyr package, for filtering, grouping, and summarizing data.

Steps Taken:

1. **Filtered Years:**

- Restricted the dataset to include only the years from 1950 to 2021.

- This ensured relevance by excluding earlier years with potentially incomplete or less accurate data.

2. **Removed Unnecessary Rows:**

- Excluded rows with missing values in the Code column.

- Removed rows where Code = "OWID_WRL" to avoid including aggregated world-level data.

3. **Grouped Data:**

- Grouped by Entity (country/region) and Code to calculate summary metrics:

    - Average Life Expectancy (1950–2021).

   - Average Female-Male Life Expectancy Difference (1950–2021).

4. **Exported Data:**

- Saved two cleaned datasets as CSV files for use in Tableau:

    - avg_life_expectancy.csv: Contains average life expectancy per country.

    - avg_life_expectancy_diff_female_male.csv: Contains average gender differences in life expectancy.

Visualizations

## 1. Average Life Expectancy Map (1950–2021)

Average Life Expectancy Map from 1950 to 2021
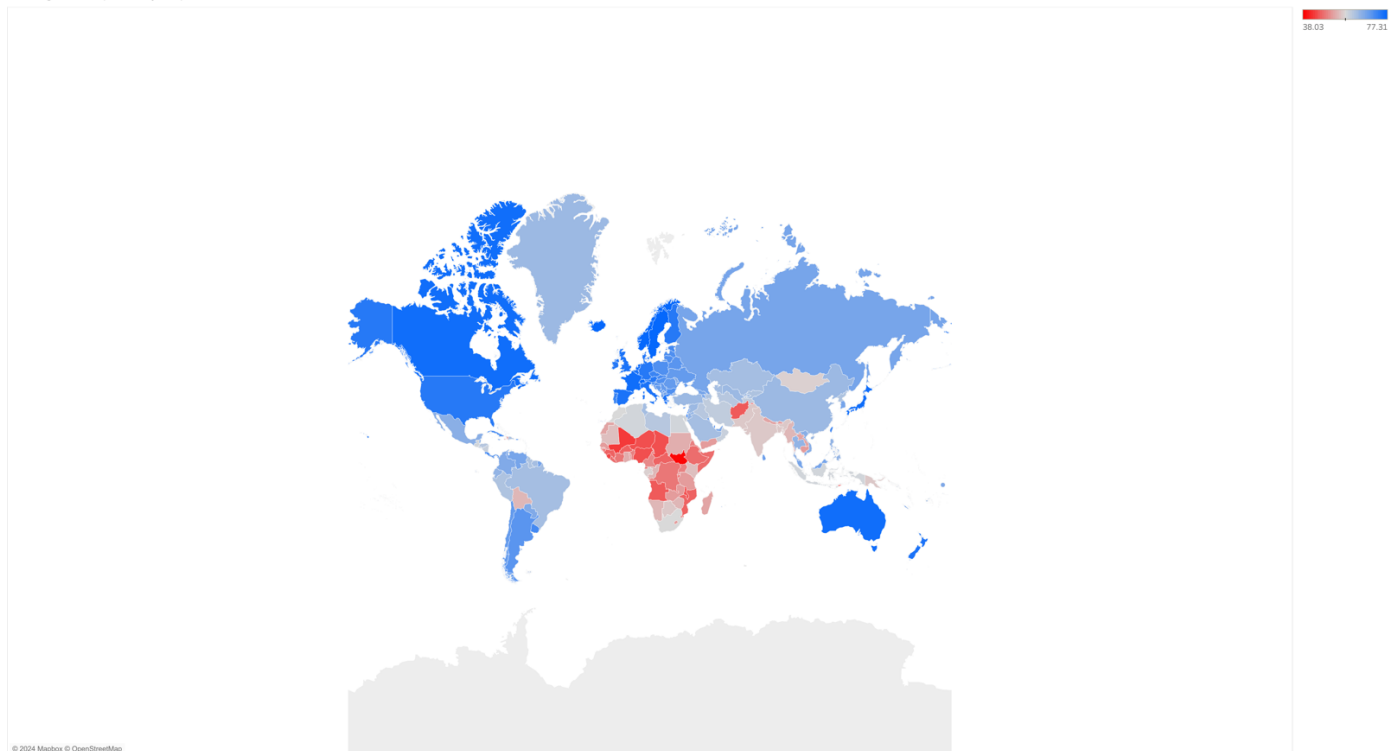
- **Purpose:** Show global disparities in life expectancy by country.

- **Platform:** Tableau.

- **Design Decisions:**

  - Used a color gradient to represent average life expectancy, with darker shades indicating higher values.

  - Simplified labels and tooltips for accessibility.

- **Insights:**
  - Countries in Europe, North America, and parts of Asia exhibit the highest life expectancies.
  - Sub-Saharan Africa displays lower average life expectancy, highlighting global health disparities.

2. Average Life Expectancy Difference Between Female & Male (1950–2021)
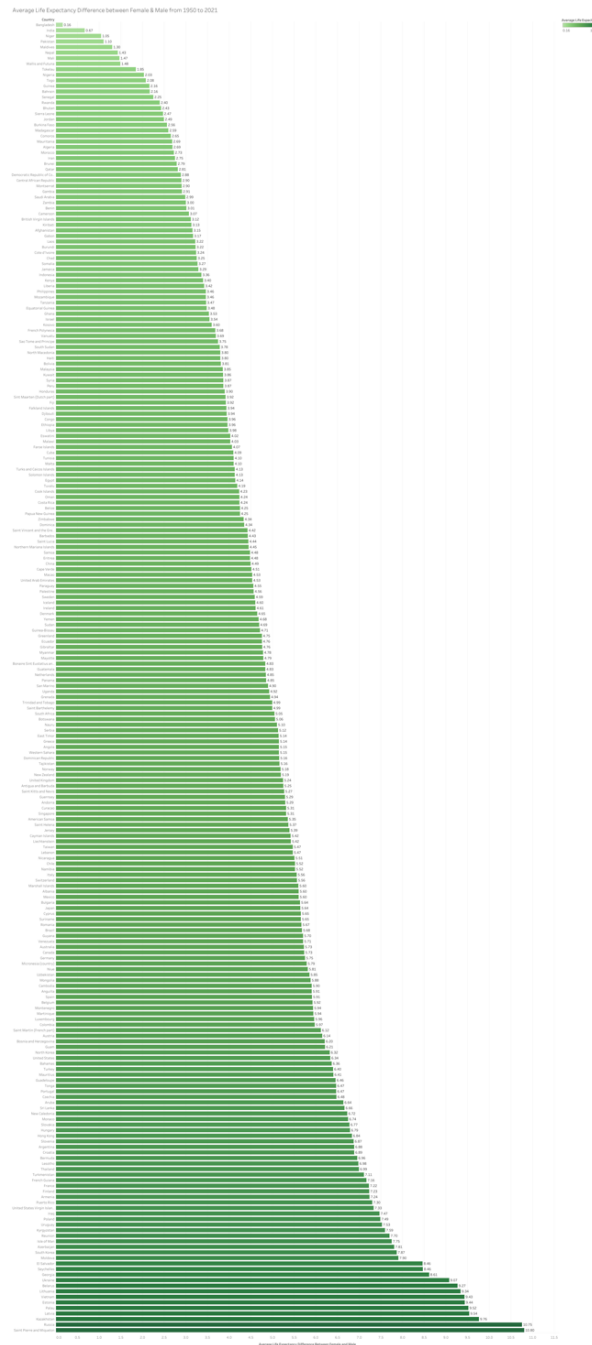
- **Purpose:** Highlight gender differences in life expectancy across countries.
- **Platform:** Tableau.

- **Design Decisions:**
  - Used a horizontal bar chart with countries sorted by the magnitude of gender differences.
  - Applied a consistent color scheme to enhance readability.
- **Insights:**
  - Countries with the largest gender gaps tend to have higher overall life expectancies.
  - Minimal gender differences are observed in some regions, possibly reflecting unique cultural or healthcare factors.

Justification for Design Choices

- **Visualization Types:**
  - A map provides an intuitive overview of geographical disparities.
  - A bar chart effectively compares numerical differences (e.g., gender gaps) across countries.
- **Color Schemes:**
  - Non-default color gradients were chosen for clarity and accessibility.
- **Labels and Layout:**
  - Titles and labels were simplified to ensure accessibility for a general audience.

Citations

- TidyTuesday. (2023). Life Expectancy Dataset.
- Tableau Public. (n.d.). Visualization best practices.

Reproducibility

R Code for Data Cleaning and Wrangling:

The following R code was used to clean and prepare the data:

```
# Load required libraries
library(tidyverse)

# Load datasets
life_expectancy <-
readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2023/2023-12-05/life_expectancy.csv')
life_expectancy_female_male <-
readr::read_csv('https://raw.githubusercontent.com/rfordatascience/tidytuesday/master/data/2023/2023-12-05/life_expectancy_female_male.csv')

# Filter and summarize data for average life expectancy
avg_life_expectancy <- life_expectancy |>
  filter(Year >= 1950 & Year <= 2021, !is.na(Code), Code != "OWID_WRL") |>
```

```
  group_by(Entity, Code) |>
  summarize(AvgLifeExpectancy = mean(LifeExpectancy, na.rm = TRUE)) |>
  ungroup()

# Filter and summarize data for gender differences
avg_life_expectancy_diff_female_male <- life_expectancy_female_male |>
  filter(Year >= 1950 & Year <= 2021, !is.na(Code), Code != "OWID_WRL") |>
  group_by(Entity, Code) |>
  summarize(AvgLifeExpectancyDiffFM = mean(LifeExpectancyDiffFM, na.rm = TRUE)) |>
  ungroup()

# Save cleaned data
write_csv(avg_life_expectancy, "avg_life_expectancy.csv")
write_csv(avg_life_expectancy_diff_female_male, "avg_life_expectancy_diff_female_male.csv")
```

Tableau Workflow:

1. Import CSV Files:

    • avg_life_expectancy.csv for the map.

    • avg_life_expectancy_diff_female_male.csv for the bar chart.

2. Map Creation:

    • Dragged Entity to the map.

    • Used AvgLifeExpectancy for color gradients.

    • Added tooltips for clarity.

3. Bar Chart Creation:

    • Plotted Entity on the Y-axis and AvgLifeExpectancyDiffFM on the X-axis.

    • Sorted bars by descending order of gender differences.