**World Fertility Data Visualization Project**

**Table of Contents**

1. [**Introduction**](#introduction)
2. [**Data Overview**](#data-overview)
3. [**Key Fertility Metrics**](#key-fertility-metrics)
   * [**Age-Specific Fertility Rate (ASFR)**](#age-specific-fertility-rate-asfr)
   * [**Total Fertility Rate (TFR)**](#total-fertility-rate-tfr)
   * [**Mean Age at Childbearing (MAC)**](#mean-age-at-childbearing-mac)
4. [**Data Preparation**](#data-preparation)
   * [**Data Sources**](#data-sources)
   * [**Data Loading and Cleaning**](#data-loading-and-cleaning)
   * [**Date Conversion**](#date-conversion)
   * [**Combining World Map Data**](#combining-world-map-data)
5. [**Visualization Design**](#visualization-design)
6. [**Interactive Map**](#interactive-map)
7. [**Evaluation and Feedback**](#evaluation-and-feedback)
   * [**Evaluation Process**](#evaluation-process)
   * [**Synthesis of Findings**](#synthesis-of-findings)
8. [**Conclusion**](#conclusion)
9. [**References**](#references)

**Introduction**

**The global fertility landscape has undergone significant changes over the past decades, influencing demographic structures, economic development, and social policies worldwide. Understanding these trends is crucial for policymakers, researchers, and the public. This project aims to visualize global fertility trends using the World Fertility Data 2019 provided by the United Nations.**

**Data Overview**

**World Fertility Data 2019 presents age-specific fertility rates, total fertility rates, and mean age at childbearing for 201 countries or areas globally. It includes empirical data from civil registration systems, population censuses, and sample surveys available as of August 2019, covering the period from 1950 to the present.**

**Data Sources Include:**

* **National Statistical Offices**
* **Demographic and Health Surveys (DHS)**
* **Multiple Indicator Cluster Surveys (MICS)**
* **Reproductive Health Surveys (RHS)**
* **Eurostat**
* **Human Fertility Database (HFD)**
* **Human Fertility Collection (HFC)**
* **Pan-Arab Project for Child Development Surveys (PAPCHILD)**
* **Pan-Arab Project for Family Health Survey (PAPFAM)**
* **National surveys and fertility estimates produced by the UN DESA Population Division**

**Note: The empirical data used for deriving estimates in the *World Population Prospects 2019* may differ due to publication time lags and additional adjustments for cohort sizes in consecutive censuses.**

**Key Fertility Metrics**

**Understanding the following key fertility metrics is essential for interpreting the data visualizations:**

**Age-Specific Fertility Rate (ASFR)**

**The Age-Specific Fertility Rate (ASFR) represents the number of live births per 1,000 women in a specific age group during a defined period, typically one year.**

**Formula:**

**ASFR=(Number of live births to women in a specific age groupNumber of women in that age group)×1000ASFR = \left( \frac{\text{Number of live births to women in a specific age group}}{\text{Number of women in that age group}} \right) \times 1000ASFR=(Number of women in that age groupNumber of live births to women in a specific age group​)×1000**

**Example:**

**If there are 50 live births to women aged 20-24 in a year, and the population of women in that age group is 10,000:**

**ASFR = (50/10,000)×1000=5 live births per 1,000 women**

**Total Fertility Rate (TFR)**

**The Total Fertility Rate (TFR) estimates the average number of children a woman would have if she experienced the current ASFRs throughout her reproductive years (usually ages 15-49).**

**Formula:**

**TFR=∑I =1n(ASFRi×5/1000)**

**Where:**

* **ASFR\_i= Age-Specific Fertility Rate for age group (i)-th age group.**
* **n = Number of age groups**

**Mean Age at Childbearing (MAC)**

**The Mean Age at Childbearing (MAC) represents the average age at which women give birth, calculated by weighting each age group's mid-point by its ASFR.**

**Formula:**

**MAC = ∑I =1n(ai × ASFRi ) / ∑i=1 nASFRi ​**

**Where:**

* **a\_i​ = Mid-point of age group (e.g., 17.5 for ages 15-19)**
* **ASFR\_i​ = Age-Specific Fertility Rate for age group (i)-th age group.**

**Data Preparation**

**Data Sources**

* **Fertility Data:** [**World Fertility Data 2019**](https://www.un.org/development/desa/pd/data/world-fertility-data)
* **World Map Data:** [**Natural Earth Data**](https://www.naturalearthdata.com/downloads/110m-physical-vectors/)

**Data Loading and Cleaning**

**Loading Fertility Indicators:**

**python**

**Copy code**

**import pandas as pd**

**df\_fertility\_indicators = pd.read\_excel(**

**'./data/undesa\_pd\_2019\_world\_fertility\_dataset.xlsx',**

**sheet\_name='FERTILITY INDICATORS',**

**skiprows=6**

**)**

**df\_fertility\_indicators.rename(**

**columns={'Country or Area': 'country', 'Country or Area Code': 'country\_code'},**

**inplace=True**

**)**

**Loading World Map Data:**

**python**

**Copy code**

**import geopandas as gpd**

**world\_map\_data = gpd.read\_file('./Data/Map\_data/ne\_110m\_admin\_0\_countries.shp')**

**world\_map\_data['country\_code'] = world\_map\_data['ISO\_N3'].astype(int)**

**Date Conversion**

**The 'Date' field in the dataset is in a fractional year format. It needs to be converted to a standard date format:**

**python**

**Copy code**

**import datetime**

**def convert\_fractional\_year(fractional\_year):**

**if pd.isnull(fractional\_year):**

**return None**

**year = int(fractional\_year)**

**fraction = fractional\_year - year**

**days\_in\_year = 366 if (year % 4 == 0 and (year % 100 != 0 or year % 400 == 0)) else 365**

**days\_to\_add = int(fraction \* days\_in\_year)**

**start\_date = datetime.date(year, 1, 1)**

**final\_date = start\_date + datetime.timedelta(days=days\_to\_add)**

**return final\_date**

**df\_fertility\_indicators['Date'] = df\_fertility\_indicators['Date'].apply(convert\_fractional\_year)**

**Combining World Map Data**

**Merging Fertility Data with World Map:**

**python**

**Copy code**

**df\_world\_map = world\_map\_data.merge(**

**df\_fertility\_indicators, on='country\_code', how='left'**

**)**

**df\_world\_map['Date'] = pd.to\_datetime(df\_world\_map['Date'])**

**df\_world\_map = df\_world\_map[['country', 'geometry', 'Date', 'Indicator', 'Value']]**

**df\_world\_map.dropna(subset=['geometry', 'Value'], inplace=True)**

**Extracting and Preparing Indicators:**

**python**

**Copy code**

**# Extract unique indicators**

**indicators = df\_world\_map['Indicator'].unique()**

**# Filter and prepare ASFR indicators**

**asfr\_indicators = [i for i in indicators if str(i).startswith('ASFR')]**

**df\_asfr\_indicators = df\_world\_map[df\_world\_map['Indicator'].isin(asfr\_indicators)].copy()**

**# Similarly, prepare TFR and MAC indicators**

**tfr\_indicators = [i for i in indicators if str(i).startswith('TFR')]**

**df\_tfr\_indicators = df\_world\_map[df\_world\_map['Indicator'].isin(tfr\_indicators)].copy()**

**mac\_indicators = [i for i in indicators if str(i).startswith('MAC')]**

**df\_mac\_indicators = df\_world\_map[df\_world\_map['Indicator'].isin(mac\_indicators)].copy()**

**Preparing ASFR Dataframes by Age Group:**

**python**

**Copy code**

**def prepare\_asfr\_dataframes(asfr\_indicators, df\_asfr\_indicators):**

**asfr\_dfs = {}**

**for indicator in asfr\_indicators:**

**df\_indicator = df\_asfr\_indicators[df\_asfr\_indicators['Indicator'] == indicator].copy()**

**asfr\_dfs[indicator] = df\_indicator**

**age\_group\_dfs = {**

**'15-19': asfr\_dfs.get('ASFR1519', pd.DataFrame()),**

**'20-24': asfr\_dfs.get('ASFR2024', pd.DataFrame()),**

**# ... other age groups**

**}**

**# Ensure 'geometry' column is of correct type**

**for df in age\_group\_dfs.values():**

**if not df.empty and df['geometry'].dtype == object:**

**df['geometry'] = df['geometry'].apply(**

**lambda x: x if isinstance(x, shapely.geometry.base.BaseGeometry) else shapely.wkt.loads(x)**

**)**

**return age\_group\_dfs**

**age\_group\_dfs = prepare\_asfr\_dataframes(asfr\_indicators, df\_asfr\_indicators)**

**Visualization Design**

**The primary design choice was to develop an interactive world map that allows users to explore fertility data across different countries and time periods. This approach provides:**

* **Interactivity: Users can select specific indicators, age groups, and years.**
* **Visual Clarity: A color-coded map facilitates immediate comparison between regions.**
* **Temporal Analysis: A slider enables the exploration of trends over time.**

**Design Justifications:**

* **User Engagement: Interactivity enhances user engagement and understanding.**
* **Data Accessibility: Visual representation makes complex data more accessible.**
* **Insightful Analysis: Users can identify patterns and anomalies across regions and time periods.**

**Interactive Map**

***[Note: In a static document, screenshots of the interactive map at different time points and with various indicators selected should be included here.]***

**The interactive map was created using the following technologies:**

* **Python Libraries: ipyleaflet, ipywidgets, branca, geopandas, pandas**
* **Features:**
  + **Dropdown menus for selecting fertility indicators and age groups.**
  + **A slider for selecting the year.**
  + **Color-coded choropleth map representing fertility rates.**
  + **Pop-up information on clicking a country.**

**Code Snippet for Map Creation:**

**python**

**Copy code**

**from ipyleaflet import Map, GeoJSON**

**import ipywidgets as widgets**

**# Create the map**

**m = Map(center=(20, 0), zoom=2)**

**# Define the style function**

**def style\_function(feature):**

**value = feature['properties']['Value']**

**color = cmap(value) if value is not None else '#8c8c8c'**

**return {**

**'fillColor': color,**

**'color': 'black',**

**'weight': 1,**

**'fillOpacity': 0.6**

**}**

**# Add GeoJSON layer**

**geo\_json = GeoJSON(**

**data=geojson\_data,**

**style\_callback=style\_function,**

**hover\_style={'fillColor': 'white', 'fillOpacity': 0.2}**

**)**

**m.add\_layer(geo\_json)**

**# Display the map with controls**

**display(widgets.VBox([title\_widget, indicator\_dropdown, age\_group\_dropdown, date\_slider, m]))**

**Evaluation and Feedback**

**Evaluation Process**

* **Participants: One English tutor and two mathematics professors.**
* **Procedure: Participants interacted with the map, exploring different regions, indicators, and time periods. They were asked to provide feedback on usability, clarity, and overall effectiveness.**

**Synthesis of Findings**

**What Worked Well:**

* **Ease of Navigation: The map's interactivity made it intuitive to explore the data.**
* **Visual Appeal: Color-coding effectively highlighted differences between regions.**
* **Engagement: Users found the map engaging and informative.**

**Areas for Improvement:**

1. **Date Slider Granularity: The slider cycled through days, which was too granular for annual data.**
   * **Solution: Adjust the slider to increment by years instead of days.**
2. **Consistency Between Datasets: The selected date did not remain consistent when switching indicators.**
   * **Solution: Implement code to retain the selected date across different datasets, defaulting to the closest available date if necessary.**
3. **Color Mapping Clarity: The color scale was challenging to interpret due to a concentration of data in lighter shades.**
   * **Solution: Adjust the color palette or use data normalization techniques to enhance visual differentiation.**

**Conclusion**

**This project successfully visualized global fertility trends using an interactive map, providing valuable insights into how fertility patterns have evolved over time and across regions. Visualization aids in understanding complex demographic data, making it accessible to a broader audience.**

**Future Improvements:**

* **Enhanced Legends and Tooltips: Provide clearer legends and more detailed tooltips with additional country-specific information.**
* **Performance Optimization: Optimize data loading and processing for smoother user experience.**
* **Extended Data Integration: Incorporate additional demographic indicators for a more comprehensive analysis.**

**References**

* **United Nations, Department of Economic and Social Affairs, Population Division (2019). *World Fertility Data 2019*. POP/DB/Fert/Rev2019.**
* **Natural Earth Data. *1:110m Cultural Vectors*. Retrieved from Natural Earth Data Downloads.**
* [**The final project is published on GitHub**](https://github.com/DJ-Greenwood/Data_visualization.git)**.**