

# World Demographics Explorer: A Dashboard for UN Population Data

Shashwat SHARMA<sup>1</sup>

<sup>1</sup> Institut Polytechnique de Paris

## Abstract

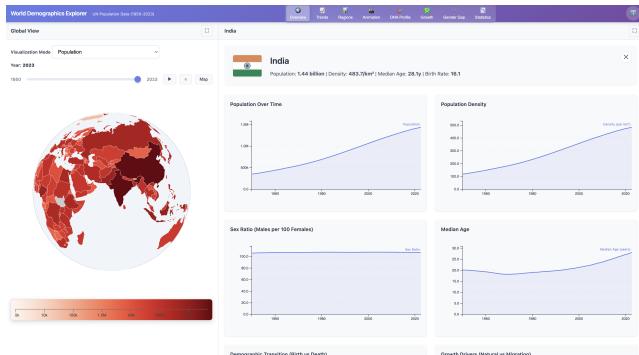
The *World Demographics Explorer* transforms the UN World Population Prospects 2024 dataset (237 countries, 1950-2023) into an interactive visualization through coordinated multiple views. The system implements eight visualization modes including a 3D globe with choropleth encoding, animated scatter plots, small multiples grids, and statistical correlation analysis. Using brushing and linking via `d3.dispatch`, the dashboard enables users to explore spatial, temporal, and multivariate demographic patterns simultaneously. Visual encodings are grounded in perceptual principles including logarithmic scales for skewed distributions and diverging color schemes for zero-centered variables, revealing insights about demographic transitions, migration patterns, and gender disparities in longevity.

You can try out World Demographics Explorer: <https://world-population.vercel.app/>

**Keywords:** Data Visualisation, Demographics.

## 1 Introduction and Analytical Context

Understanding global demographic shifts is fundamental to policy decisions affecting billions of people, yet the sheer volume of data across 237 countries and 74 years presents a cognitive challenge. The *World Demographics Explorer* addresses this by transforming the UN World Population Prospects 2024 dataset into an interactive visual narrative, enabling researchers and policymakers to rapidly identify patterns, validate demographic theories, and compare countries across multiple dimensions simultaneously.



**Figure 1.** The World Demographics Explorer: A multi-view coordinated interface displaying global population trends (1950-2023). The split-pane layout provides persistent spatial context (globe, left) alongside dynamic temporal analyses (right).

### 1.1 Dataset Description and Preprocessing Pipeline

The visualization leverages the United Nations World Population Prospects 2024 dataset, a comprehensive longitudinal repository comprising demographic indicators for 237 countries and territories spanning a 74-year observation period (1950-2023). The dataset exhibits a hierarchical structure organized at three distinct geographic granularities: global aggregates, regional classifications (six major regions), and country-level observations. Each observation contains over 50 demographic indicators, encompassing population metrics, vital rates (crude birth and death rates), life expectancy estimates, fertility trajectories, and migration statistics.

To facilitate high-performance, interactive visualization, the raw dataset underwent an extensive preprocessing phase using a custom Python-based pipeline (`prepare_dataviz.py`). A critical technical challenge involved the geographic reconciliation of disparate naming standards between the UN tabular data and TopoJSON geospatial layers. The pipeline implements a *country-NameMap* lookup table to resolve nomenclature inconsistencies, such as mapping "United States" to "United States of America", ensuring  $O(1)$  join performance during runtime. The original space-separated numerical format was transformed into 12 specialized JSON files, each year-nested to optimize for  $O(1)$  data retrieval during high-frequency temporal animations. This preprocessing architecture reduced typical page load times from over 30 seconds (raw CSV parsing) to under 3 seconds while maintaining full interactivity across 74 years of data.

### 1.2 Analytical Objectives

The dashboard was engineered to follow a "top-down" narrative structure, allowing users to transition from global spatial patterns to granular country-specific profiles. This architecture addresses three primary analytical questions fundamental to demographic research. First, the system examines temporal convergence by tracking longitudinal shifts in demographic indicators to determine whether regional trajectories have converged or diverged over the last seven decades. Second, it provides an empirical test for the classical four-stage demographic transition model through coordinated visualization of birth and death rate trajectories. Third, it investigates multivariate relationships between population growth drivers (natural change versus net migration), the persistence of gender disparities in longevity, and the evolution of reproductive health indicators.

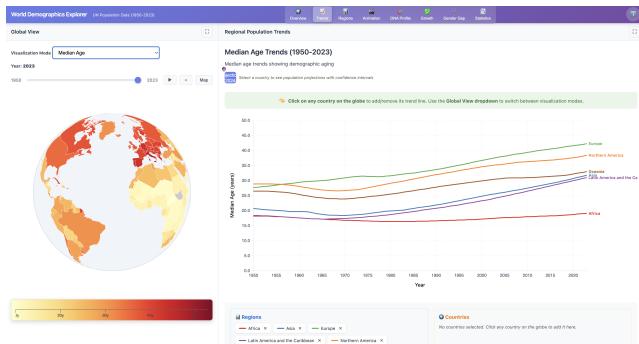
These objectives necessitated a Coordinated Multiple Views (CMV) interface, a design pattern extensively covered in the course that enables users to maintain analytical context while exploring different data dimensions. The implementation utilizes a custom event dispatcher (`d3.dispatch`) to enable seamless brushing and linking. This ensures that any interaction, such as selecting a country on the 3D globe, simultaneously reconfigures the temporal, statistical, and profile views. For instance, clicking Japan on the globe updates the time series chart to show Japan's trends,

adds Japan to the small multiples grid, highlights Japan in the animated scatter plot, and populates the radar chart with Japan's demographic profile, all while maintaining visual consistency across views.

## 2 Design Rationale and Visual Encoding Decisions

### 2.1 Architectural Framework: Coordinated Multiple Views

The dashboard implements a split-pane architecture consisting of a persistent globe visualization, which occupies 33% of the viewport width, and a dynamic details pane, which occupies the remaining 67%. This design adheres to the Overview + Detail pattern, ensuring that users maintain spatial context while exploring temporal or statistical dimensions.



**Figure 2.** Illustration of Coordinated Multiple Views interaction. Selecting a visualization mode from the globe view dropdown and clicking on countries on the globe dynamically updates the linked views, including the time series plots, small multiples, animated scatter plot, and radar chart. Countries can be interactively added to or removed from the line chart visualizations while preserving consistent object identity across views.

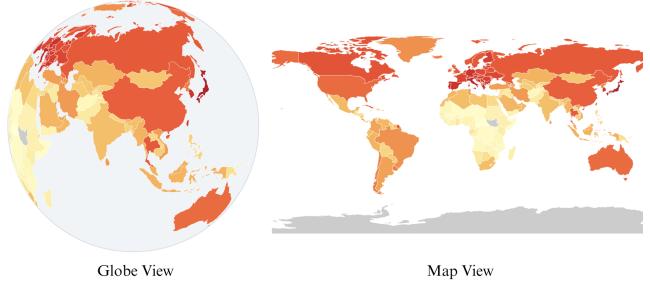
The system employs a custom event dispatcher (`d3.dispatch`) that enables brushing and linking across all views. When a user selects a country on the globe, this action simultaneously triggers four updates:

- The time series chart adjusts to display country-specific trends.
- A demographic profile is added to the small multiples grid.
- The selected country is highlighted in the animated scatter plot.
- The radar chart is populated with country-specific indicators.

This implementation directly applies the principle of Coordinated Multiple Views, which prevents cognitive overload by maintaining consistent object identity across various transformations.

### 2.2 Globe Visualization: Projection and Color Scale Design

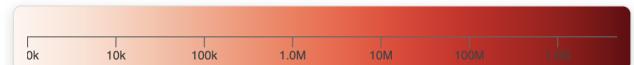
**2.2.1 Projection Selection** The globe supports two distinct projection modes tailored to different analytical needs. The *Orthographic Projection* (3D mode) is implemented using `d3.geoOrthographic` to simulate a three-dimensional sphere. This choice prioritizes spatial intuition over metric accuracy, serving as a navigational anchor for country selection. Conversely, the *Equirectangular Projection* (2D mode) is a cylindrical projection that preserves rectangular coordinates, which facilitates the rapid



**Figure 3.** Side-by-side comparison of the 3D Orthographic globe and the 2D Equirectangular map representations.

scanning of all countries without the occlusion inherent in a spherical view.

**2.2.2 Color Scale Strategy** The globe implements ten distinct color encoding modes selected based on the nature of the underlying data. For ratio variables such as Population and Density, the system uses a logarithmic scale (`d3.scaleLog`) with `d3.interpolateReds` and `d3.interpolateBlues`. This logarithmic transformation is critical because of the six-order-of-magnitude range between countries, as it prevents smaller nations from becoming imperceptible.



**Figure 4.** Close-up of the population legend on the globe view, illustrating logarithmic scaling and abbreviated numeric formatting (e.g., 1B).

For variables with meaningful midpoints, the system utilizes diverging scales. The *Gender Gap in Life Expectancy* uses `d3.interpolateRdBu` with a midpoint at zero, where red hues indicate a female advantage and blue hues indicate a male advantage. This encoding exploits pre-attentive processing to reveal disparities without requiring a numerical lookup. Similarly, the *Sex Ratio* is centered at the biological baseline of 105.

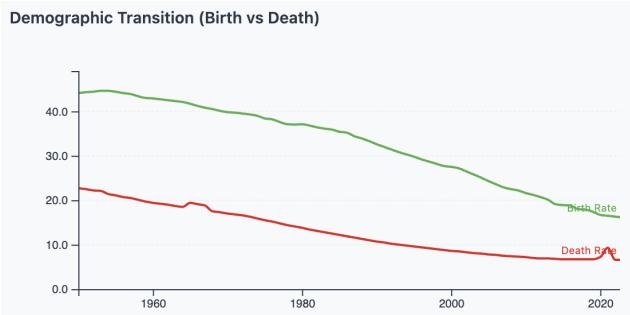
### 2.3 Time Series Visualization: Mode-Specific Encoding

The time series module supports nine visualization modes, each designed to convey specific demographic narratives. For univariate indicators, regional lines are encoded using position on a common scale, which Cleveland's empirical studies identified as the most accurate visual channel for quantitative comparison. A line thickness of 2.5px for regions and 2px for countries creates a clear figure-ground relationship.

#### 2.3.1 Multi-Metric Modes

Five modes visualize paired indicators to support complex analysis:

- **Demographic Transition:** Birth and death rates are plotted on a shared y-axis using green (#4daf4a) and red (#e41a1c) respectively, where the visual gap between lines illustrates the rate of natural change.
- **Growth Drivers:** This mode compares natural change against net migration (#984ea3), revealing whether growth is endogenous or exogenous.



**Figure 5.** The Demographic Transition chart showing birth and death rates over time (green/red lines), alongside the Growth Drivers chart illustrates population growth components (green/purple lines).

- Longevity and Gender Gap:** Male and female life expectancies are contrasted to show survival differences.
- Healthcare Quality:** This mode includes a logarithmic scale toggle to handle the 50-fold range in mortality rates across different development stages.

#### 2.4 Small Multiples: Demographic Transition Visualization

The birth and death rate grid utilizes Edward Tufte's small multiples concept, providing identical chart structures for each region to facilitate cross-comparison. This design was chosen over a single animated view because it enables simultaneous comparison across all regions, a cognitive task that animation cannot support effectively due to working memory limitations. By using area charts with translucent fills (0.7 opacity), the visualization emphasizes the magnitude of vital rates. The overlapping colored areas create a visual gestalt where the gap between birth and death areas directly encodes natural population change.



**Figure 6.** Small multiples grid displaying birth and death rates across all world regions, enabling comparative analysis of regional demographic patterns.

#### 2.5 Animation: Hans Rosling-Style Scatter Plot

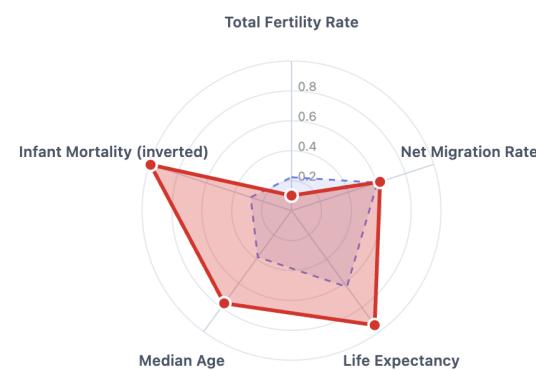
The animated scatter plot replicates the Gapminder methodology, encoding four dimensions: fertility rate (x-axis), life expectancy (y-axis), population (circle size via `d3.scaleSqrt`), and region (categorical hue). The animation maintains object constancy through a data join pattern with a key function, ensuring smooth transitions rather than disorienting jumps as the years progress.



**Figure 7.** Hans Rosling-style animated scatter plot illustrating demographic transitions over time. Each point represents a country and can be interactively hovered to reveal detailed demographic values for a given year.

#### 2.6 Radar Chart: Dimensional Profile Visualization

The radar chart transforms parallel coordinates into a polar form to visualize five normalized indicators: fertility, migration, life expectancy, median age, and infant mortality. This polar encoding was chosen over traditional parallel coordinates for its compactness and its ability to create a recognizable "shape signature" through gestalt perception. Each indicator is min-max scaled to a range of [0, 1] using global extrema. The resulting pentagonal shape represents a country's "demographic DNA," allowing users to immediately distinguish between different stages of national development.



**Figure 8.** Radar chart comparing a selected country (Japan) with its regional average (Asia) across key demographic indicators.

#### 2.7 Growth Drivers Scatter Plot

This visualization employs a quadrant analysis framework based on the Rate of Natural Change and the Net Migration Rate. The orthogonal reference lines at zero create four semantic quadrants that identify whether a country is growing via births, immigration, or shrinking through natural decline. This quadrant structure

leverages spatial grouping to enable rapid categorical assessment without requiring precise value reading. Circle size encodes population to maintain consistency with the animation module.



**Figure 9.** Drivers of population growth by country, contrasting natural change and net migration. This figure represents a snapshot of an animated visualization enabling exploration of temporal dynamics.

## 2.8 Gender Gap Line Chart

This module visualizes the life expectancy differential (female minus male) as a single longitudinal line. The color encoding uses a conditional diverging scheme where red indicates the global norm of female advantage and blue indicates a male advantage. The horizontal dashed line at zero serves as the perceptual anchor for this comparison.



**Figure 10.** Life expectancy gender gap showing the temporal evolution of differences between female and male life expectancy, illustrated for Nigeria.

## 2.9 Uncertainty Visualization: Graduated Confidence Intervals

To address the challenge of conveying data reliability and prevent deterministic construal errors when users interpret projections, we implemented graduated confidence intervals using nested translucent bands. This design directly responds to Session 4's emphasis on uncertainty communication in visualization.

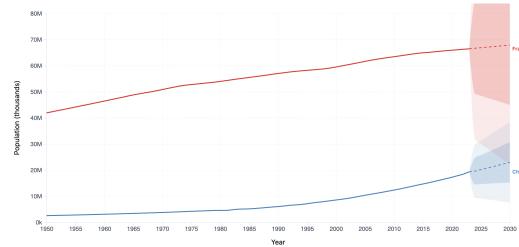
**2.9.1 Design Rationale** Following Cleveland's empirical hierarchy, position along a common scale provides the most accurate quantitative decoding. However, for uncertainty communication, we deliberately employ *nested semi-transparent areas* to leverage multiple perceptual channels simultaneously:

- **Opacity Gradient:** The 50% confidence interval is rendered at 30% opacity, while the 95% interval uses 10% opacity. This creates a natural luminance gradient where darker regions signal higher confidence.
- **Progressive Widening:** Confidence bands expand proportionally with projection distance ( $\text{width} \propto 1 + 0.05 \times \Delta t$ ), visually encoding increasing epistemic uncertainty as forecasts extend further into the future.
- **Layered Rendering:** Drawing the widest (95%) band first, followed by the 50% band, and finally the median line ensures proper visual hierarchy and prevents occlusion artifacts.

**2.9.2 Implementation Details** The preprocessing pipeline (`prepare_dataviz.py`) generates projections by fitting linear regression models to the terminal decade (2014–2023) of observed data, then applies time-dependent uncertainty inflation. On the front-end, D3's `d3.area()` generator constructs the confidence polygons, with a vertical dashed line at 2023 explicitly demarcating historical observations from forecasts.

Importantly, we maintain color consistency (using the primary interface hue #667eee) rather than introducing new colors for uncertainty bands. This reduces cognitive load and ensures accessibility for colorblind users, as opacity modulation remains perceptible across all forms of color vision deficiency.

**2.9.3 Cognitive Benefits** This approach addresses the well-documented tendency for users to treat probabilistic forecasts as deterministic outcomes. By making uncertainty visually salient through graduated bands rather than discrete error bars, we encourage probabilistic thinking about demographic futures—a critical capability for evidence-informed policy planning in domains such as infrastructure development and pension fund management.

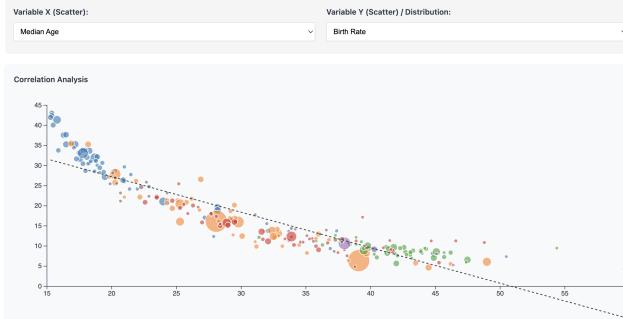


**Figure 11.** Population projection with graduated confidence intervals (2024–2030). Nested bands represent 95% and 50% confidence ranges, with increasing width indicating growing uncertainty over longer forecast horizons.

## 2.10 Statistical Analysis: Correlation and Distribution

The statistics module provides two linked views. The Correlation Scatter Plot allows users to select two variables for regression analysis, featuring a linear least-squares fit and  $R^2$  value. To ensure robust visualization, the system includes an automated outlier removal feature using the Interquartile Range (IQR) method, which was chosen over standard deviation approaches because it remains robust when the underlying data distribution is skewed or contains extreme values. This prevents extreme observations from compressing the scale and obscuring the relationship for the

majority of countries.



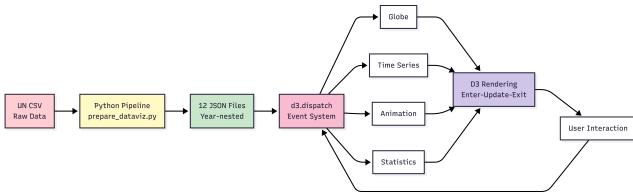
**Figure 12.** Statistics view showing the relationship between median age and birth rate using a scatter plot with regression line and accompanying box plots. Analysis includes 236 countries (1 outlier removed), yielding a strong negative correlation ( $R^2 = 0.847$ , slope =  $-0.883$ ).

Finally, the *Distribution Box Plots* implement Tukey's five-number summary (minimum, first quartile, median, third quartile, and maximum) for regional comparison. The box labels are rotated 45 degrees to avoid occlusion and maintain readability.

### 3 Technical Implementation

#### 3.1 Data Preprocessing Pipeline

The Python preprocessing script, `prepare_dataviz.py`, executes a multi-stage transformation to convert high-volume longitudinal data into visualization-ready structures. The pipeline begins with data cleaning, where space-separated values are stripped and converted to numeric types using `pd.to_numeric` with coercion to handle missing entries.



**Figure 13.** Technical architecture showing the data flow from preprocessing through rendering. The event-driven coordination layer (center) enables synchronized updates across eight visualization modules using a publish-subscribe pattern.

A significant technical feature is the hierarchical mapping logic. The script traverses parent-child relationships via the *Parent code* column to construct a comprehensive region-to-country mapping, which enables accurate regional assignment for all 237 territories. Additionally, the pipeline calculates derived metrics such as the rate of natural change and the life expectancy gender gap to support complex analytical views.

For radar chart profiling, the script performs global min-max normalization for five key indicators. The resulting `radar_chart_data.json` stores both raw and normalized values, enabling the dashboard to render the visual shape of a country's demographic DNA while providing precise raw data within tooltips. Finally, the output generation phase produces 12 specialized JSON files. Files such as `globe_data_all_years.json` are

nested by year to enable  $O(1)$  lookup performance during temporal animations, significantly reducing the computational overhead on the client side.

#### 3.2 Front-End Architecture

The JavaScript implementation follows a modular controller-view pattern to manage the complexity of the multi-view interface. The `main.js` module serves as the application controller, maintaining the global `AppState` object and coordinating the lifecycle of each visualization. Data management is handled by the `dataLoader.js` singleton, which abstracts the JSON structures and provides standardized accessor methods for all modules.

The system utilizes an event-driven coordination model powered by a namespaced event dispatcher. By using `d3.dispatch`, the dashboard implements a robust publish-subscribe pattern. This allows modules to subscribe to relevant events using unique identifiers, such as `countrySelected.radar` or `countrySelected.timeseries`. This namespaced architecture is essential for preventing callback conflicts and ensuring that a single user interaction can trigger synchronized updates across independent components without compromising code maintainability.

#### 3.3 Performance and Interactivity Optimizations

To ensure a fluid user experience across 75 years of data, the system implements several performance-critical features. The `DataLoader` fetches all JSON assets during initialization and stores them in a local cache, eliminating redundant network requests during view transitions. Visualization modules utilize the D3 enter-update-exit pattern, ensuring that only the affected DOM elements are modified during state changes, which is particularly vital for the high-density line charts in the time series view. Temporal sliders and input controls trigger updates on the change event rather than the input event, preventing excessive re-rendering during active dragging and maintaining a stable frame rate. SVG containers calculate their dimensions dynamically via `getBoundingClientRect`, allowing the interface to adapt to various viewport sizes without hardcoded pixel values.

#### 3.4 Temporal Navigation and Animation

The dashboard implements three distinct temporal interaction modes to support different analytical workflows. A global year slider enables synchronized cross-view updates, where moving the slider updates the globe's choropleth encoding, the animated scatter plot positions, and the growth drivers quadrant plot simultaneously. For the animated scatter plot module, the system employs D3's transition framework with a data join keyed by ISO country code, ensuring object constancy as circles move through demographic space over time. The animation progresses at 500ms per year with easing functions (`d3.easeCubic`) to create smooth, perceptually uniform motion rather than discrete jumps. This implementation mirrors the temporal storytelling approach pioneered by Gapminder, where animation reveals trajectories and convergence patterns that remain invisible in static snapshots.

#### 3.5 Layout Orchestration and User Agency

The dashboard includes an interface orchestration layer that provides users with significant control over the viewing environment. Users can toggle the globe and details panes between normal, min-

imized, and maximized states to suit their current analytical focus. Furthermore, the system supports mode syncing, where selecting a new indicator on the globe automatically reconfigures the primary time series chart. This reduces interaction friction and facilitates a more intuitive discovery process.

### 3.6 Accessibility and Universal Design

The project adheres to several accessibility principles to ensure the data remains interpretable for all users. Colorblind-safe palettes derived from qualitative schemes are used for regional grouping to avoid problematic red-green combinations. The time series charts utilize direct labeling, where text labels are placed at the terminus of each line to reduce the cognitive load associated with legend lookups. Additionally, the implementation uses semantic HTML and ARIA labels to provide better compatibility with screen readers and keyboard-based navigation.

## 4 Insights and Analytical Findings

### 4.1 Demographic Transition Convergence

The small multiples grid empirically validates the stages of the demographic transition model across the six major UN regions. Analysis of the 1950 regional profiles reveals that Africa exhibited a high-growth Stage 2 pattern, with crude birth rates near 48 per 1,000 and death rates at approximately 27 per 1,000. By 2023, the visualization indicates a significant global convergence toward Stage 4, characterized by low vital rates and natural change approaching zero. A notable finding is the **temporal compression** of this transition; while European nations historically required over a century to reach demographic stability, regions such as East Asia achieved similar transitions within a 50 to 70-year window which we can visualise.

### 4.2 Migration as a Population Growth Driver

The growth drivers scatter plot reveals a **spatial polarization** in how modern populations sustain themselves. Using the quadrant analysis framework, we can see that population growth is not determined solely by fertility, highlighting the critical compensatory role of migration in aging societies.

### 4.3 Gender Gap Persistence in Longevity

The gender gap analysis confirms a universal female survival advantage, with a global median life expectancy gap of 4.8 years as of 2023. The data displays significant geographic heterogeneity.

### 4.4 Statistical Correlation Findings

The statistics module allowed for rigorous hypothesis testing of demographic theories using the Interquartile Range (IQR) method for robust regression analysis. The following results were obtained:

- **Total Fertility Rate vs. Median Age:** A **strong negative correlation** was identified ( $R^2 = 0.726$ ,  $n = 230$ , 7 outliers removed), validating the theory of demographic momentum as populations with lower fertility rates experience significant structural aging.
- **Infant Mortality vs. Life Expectancy at Birth:** A **strong correlation** exists between these healthcare proxies ( $R^2 = 0.743$ ,  $n = 223$ , 14 outliers removed), with a slope of  $-0.429$

indicating that improvements in early-childhood survival remain a primary driver for increasing national longevity.

- **Net Migration Rate vs. Crude Birth Rate:** A **weak correlation** was observed ( $R^2 = 0.031$ ,  $n = 196$ , 41 outliers removed). This suggests that international migration patterns are largely decoupled from domestic fertility rates and are instead driven by external geopolitical and economic factors.

## 5 Conclusion

The *World Demographics Explorer* demonstrates the successful application of perceptual psychology principles and interaction design patterns to a high-dimensional spatiotemporal dataset. By carefully selecting visual encodings matched to specific data characteristics, such as logarithmic scales for skewed population distributions, diverging color schemes for zero-centered variables, and small multiples for regional comparison, the dashboard transforms a vast repository of 237 territories across a 74-year timeline into an interpretable analytical instrument.

The implementation of **Coordinated Multiple Views** (CMV) with brushing and linking directly addresses the challenge of enabling users to construct complex mental models of multivariate relationships. Through interactive exploration rather than the passive consumption of static charts, users can uncover the nuances of global demographic shifts. The technical architecture, which is grounded in the D3.js data-join paradigm and modular JavaScript design patterns, ensures both computational efficiency and extensibility for future demographic datasets.

Ultimately, this project validates the core thesis that well-designed visualizations significantly amplify human cognition. The dashboard reveals critical insights that remain invisible in traditional tabular summaries, including the acceleration of demographic transition convergence, the increasing compensatory role of migration in aging societies, and the persistent gender disparities in longevity. By bridging the gap between raw statistical aggregates and interactive visual narratives, the explorer provides a robust framework for understanding the evolving demographic DNA of the global population.

## Repositories

You can try out World Demographics Explorer:

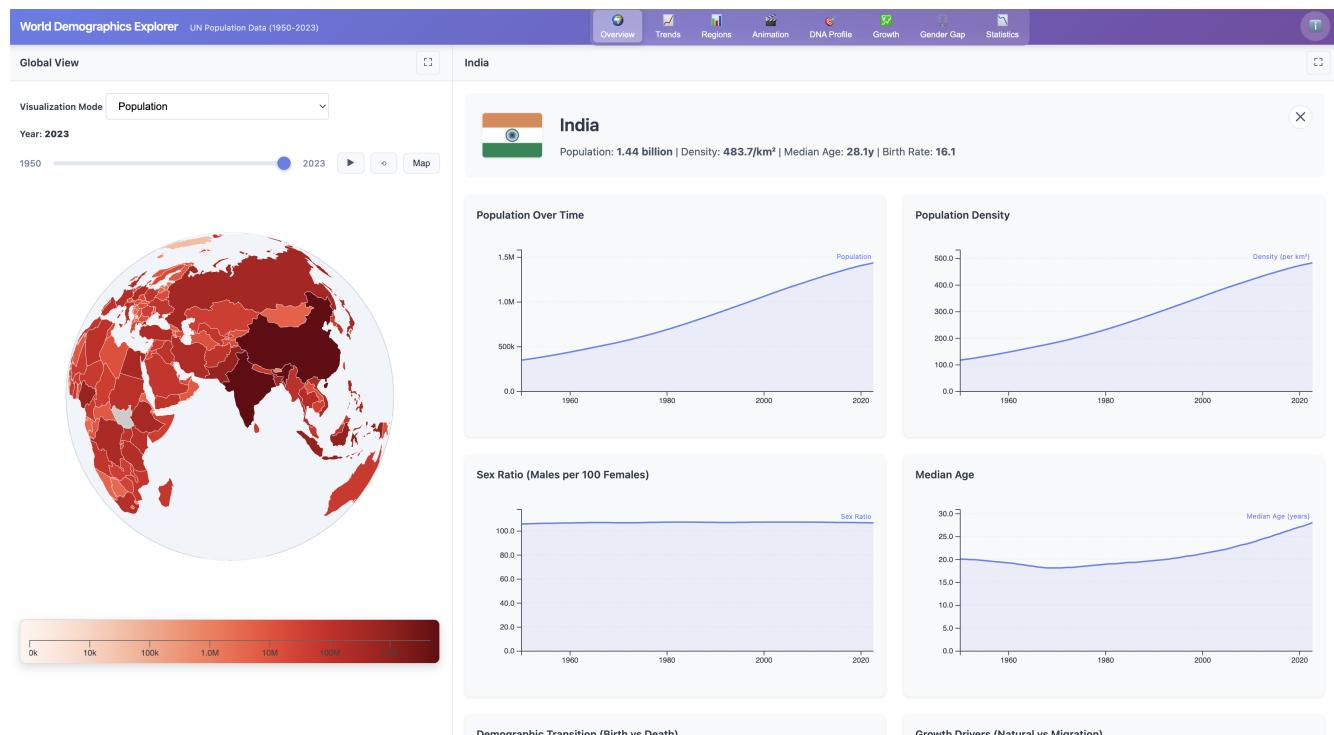
<https://world-population.vercel.app/>

The source code used for the experiments is on GitHub:

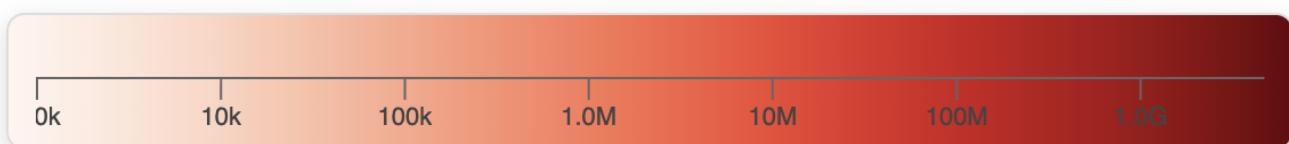
[https://github.com/shashuat/world\\_population](https://github.com/shashuat/world_population)

## Appendix: Image Galleries

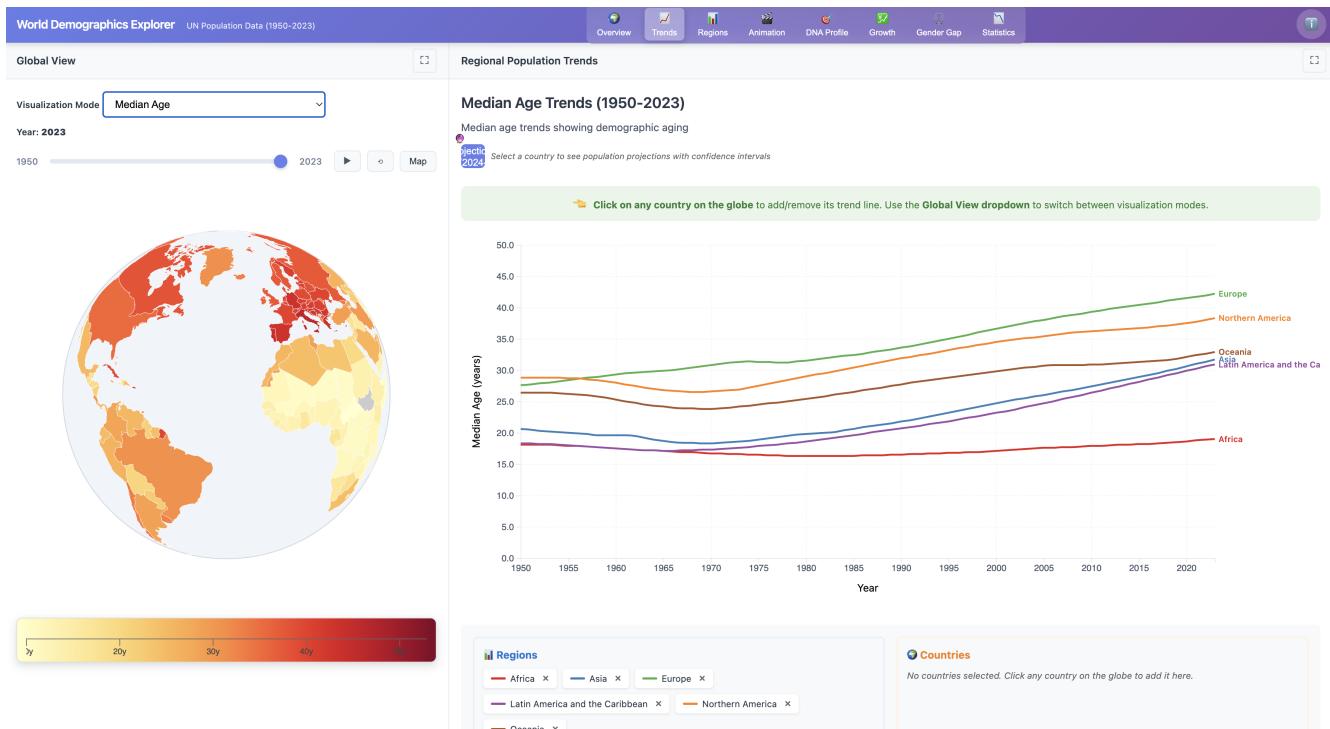
This appendix presents a full-resolution gallery of all visualizations used in the World Demographics Explorer. The figures are reproduced here to support detailed inspection of visual encodings, interaction techniques, and coordinated behaviors discussed in the main text.



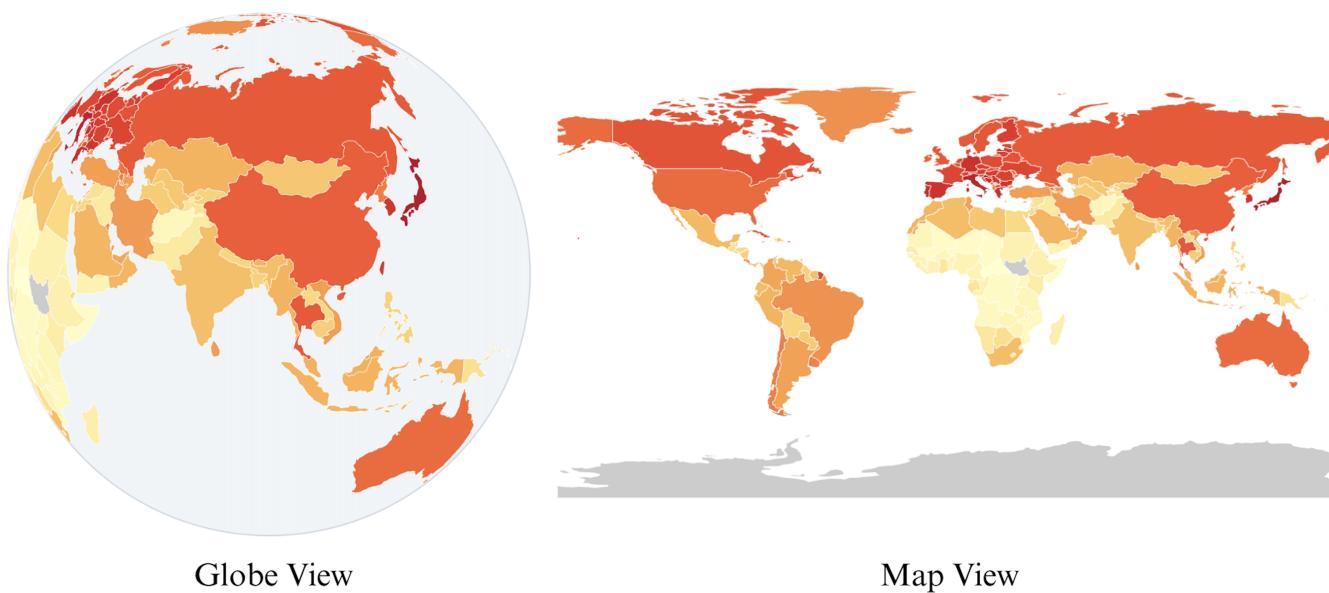
**Figure 14.** Dashboard overview showing the split-pane layout with a persistent globe view and dynamic analytical panels.



**Figure 15.** Population legend showing logarithmic scaling and abbreviated numeric formatting.



**Figure 16.** Coordinated Multiple Views interaction demonstrating linked updates across visualizations when selecting countries and visualization modes.



**Figure 17.** Comparison between the 3D Orthographic globe and the 2D Equirectangular map projection.

## Demographic Transition (Birth vs Death)

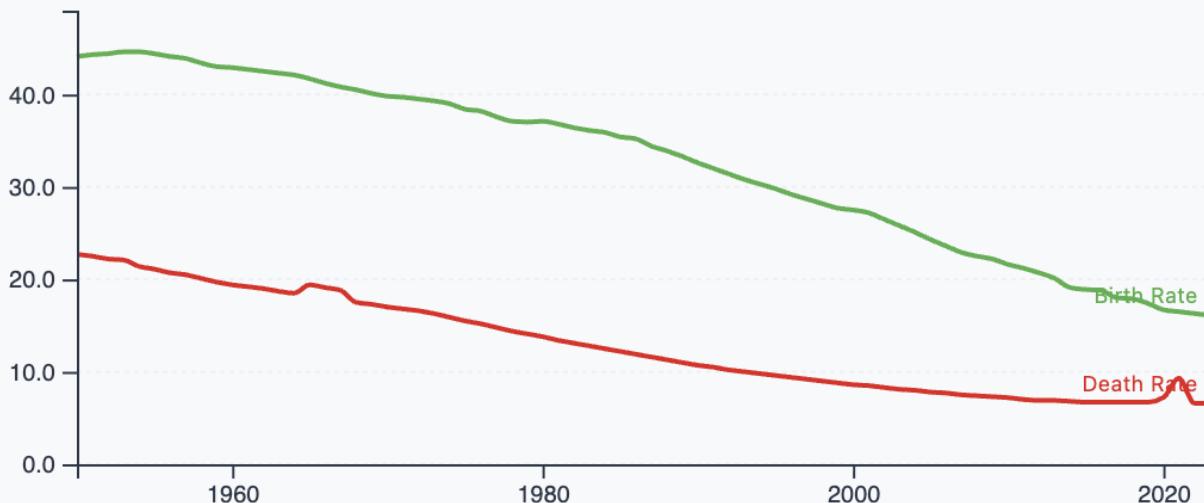


Figure 18. Demographic Transition and Growth Drivers charts illustrating population dynamics over time.

■ Birth Rate (per 1,000) ■ Death Rate (per 1,000)

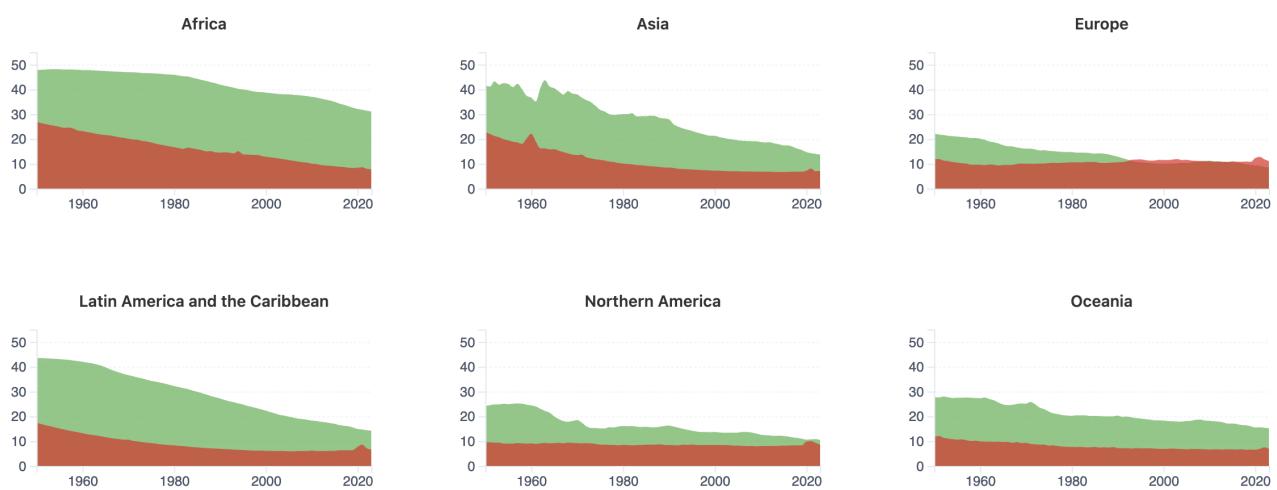


Figure 19. Small multiples grid comparing birth and death rates across world regions.

### Demographic Transition Animation

Fertility vs. Life Expectancy (1950–2023)



**Figure 20.** Hans Rosling-style animated scatter plot depicting demographic transitions over time.

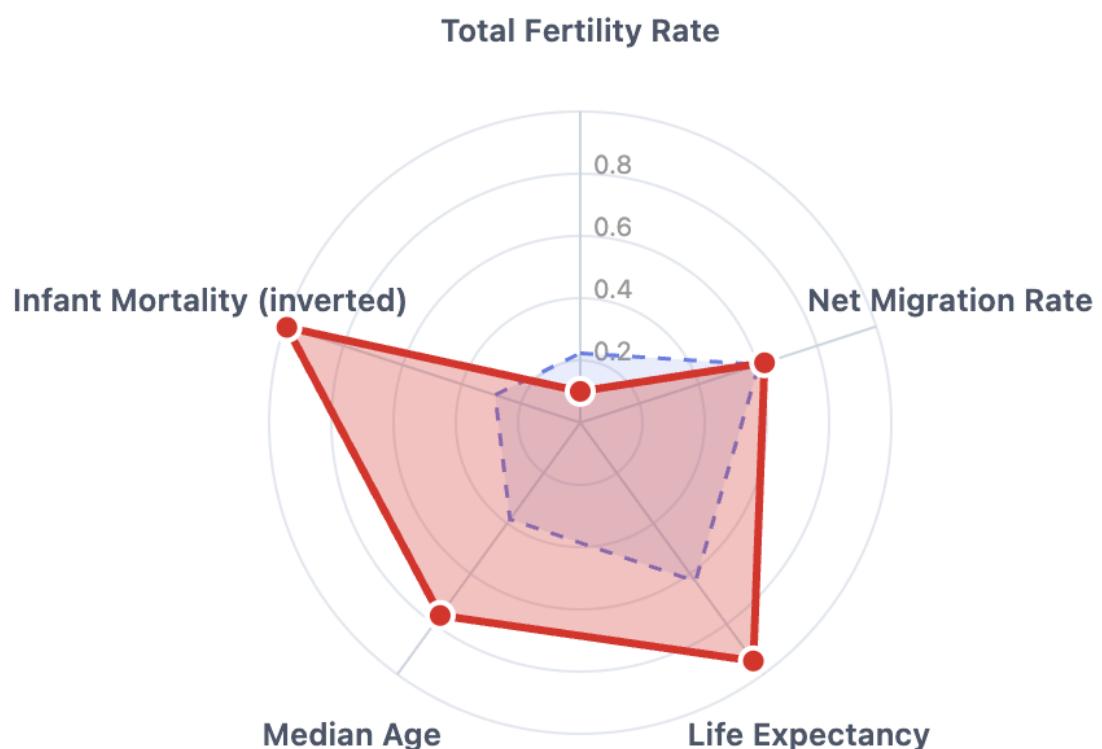
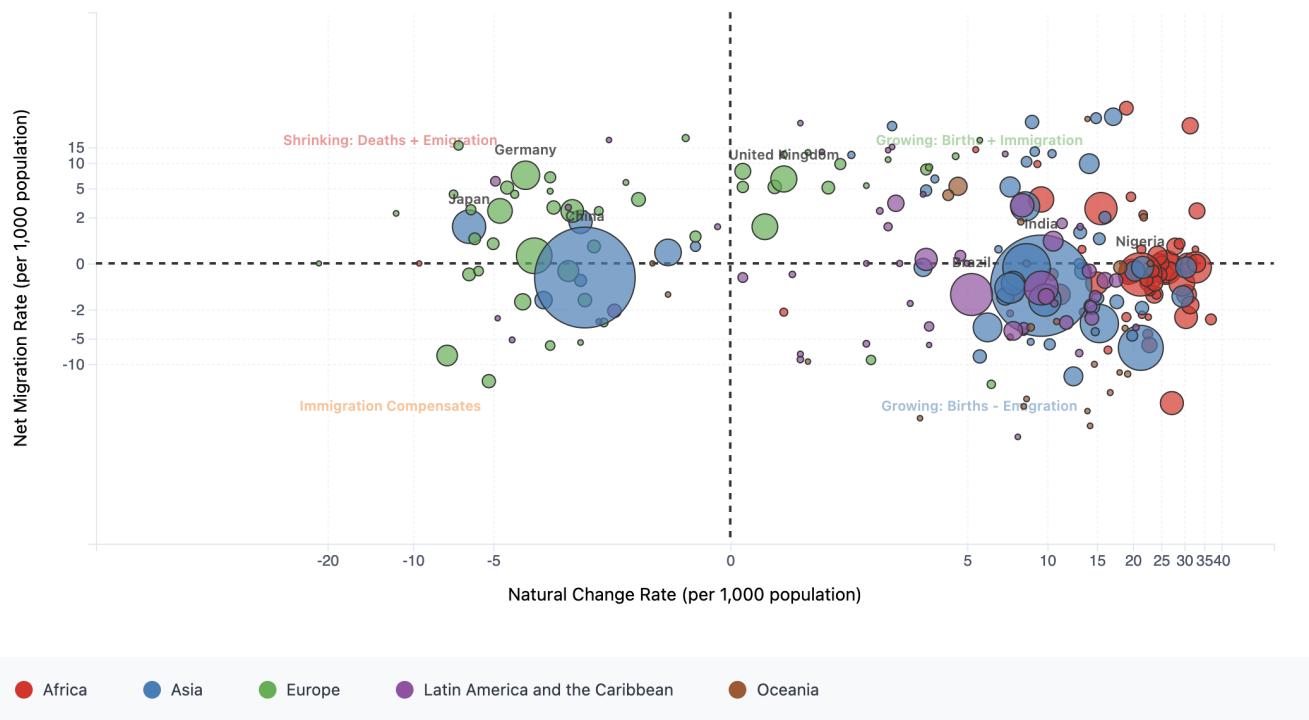
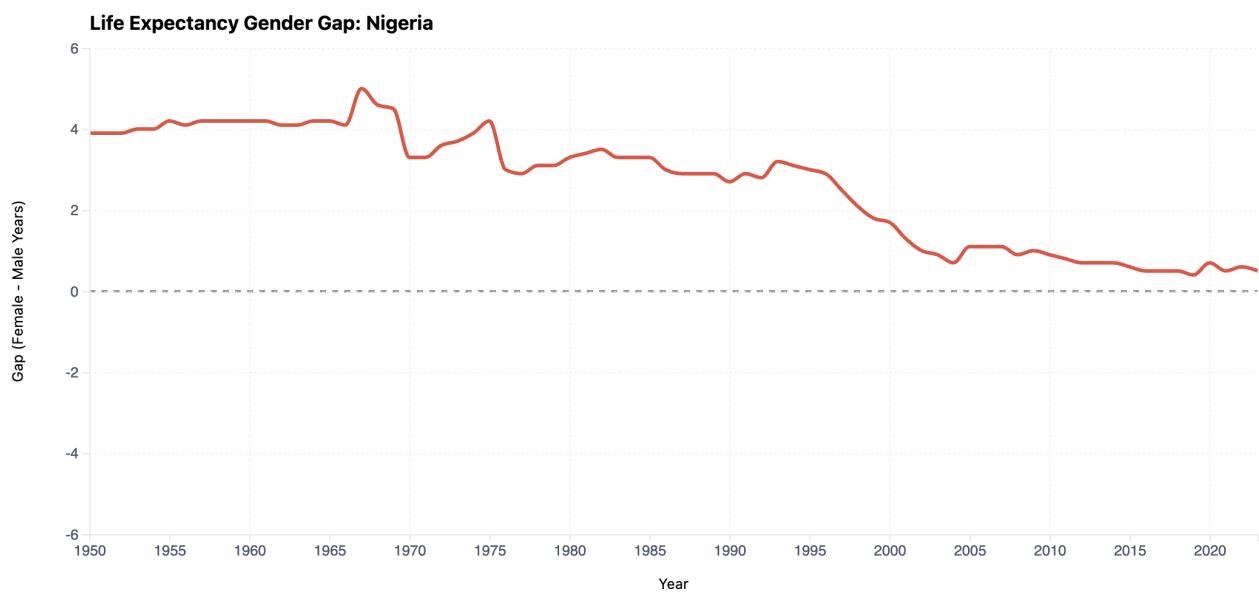


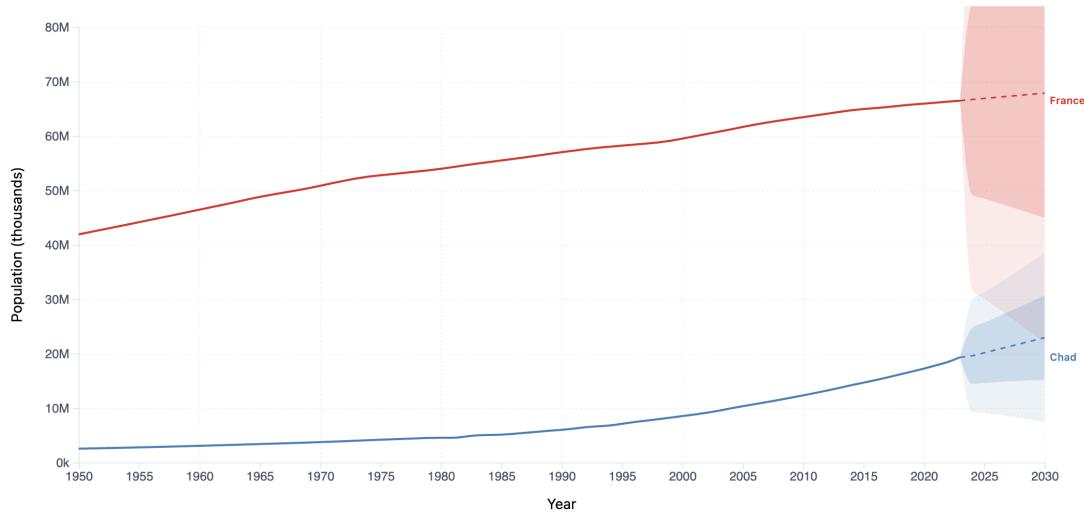
Figure 21. Radar chart comparing Japan with the Asian regional average across demographic indicators.



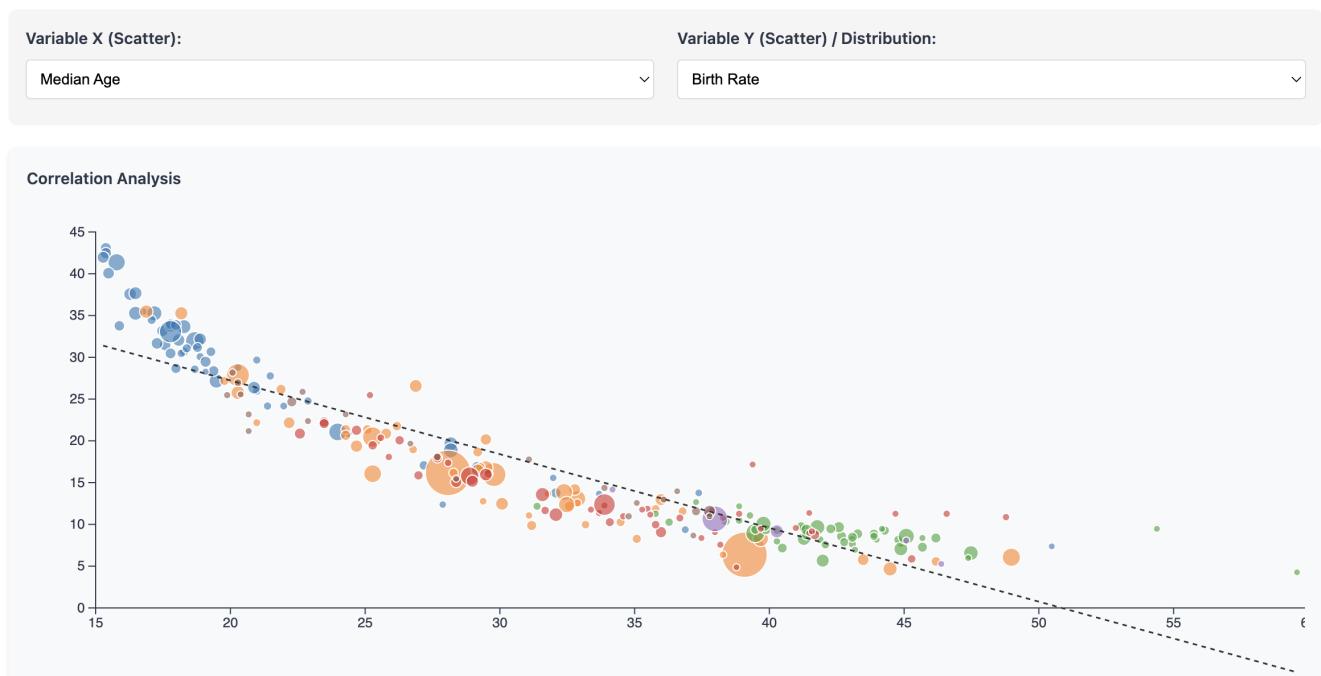
**Figure 22.** Drivers of population growth contrasting natural change and net migration (animation snapshot).



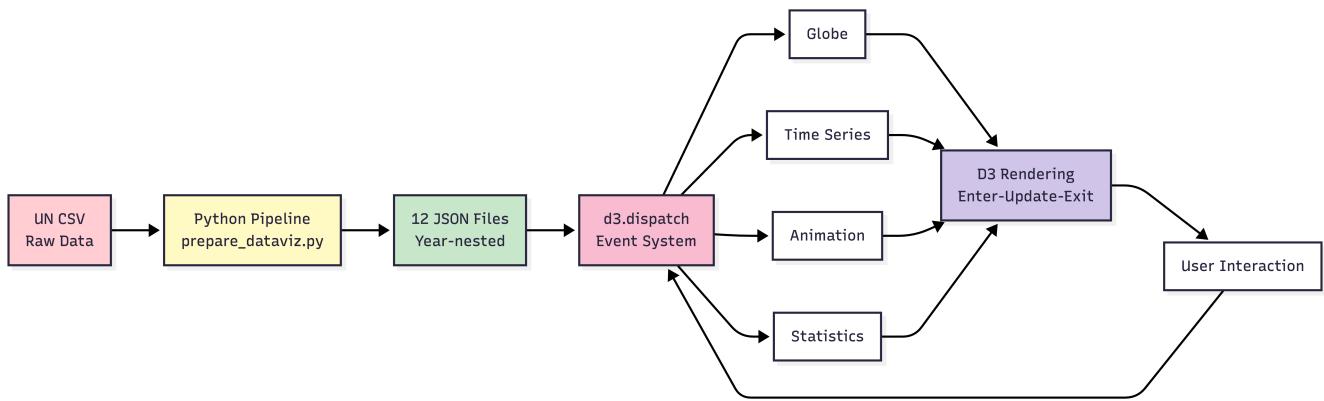
**Figure 23.** Life expectancy gender gap over time for Nigeria.



**Figure 24.** Population projections with graduated confidence intervals indicating forecast uncertainty.



**Figure 25.** Statistical analysis view showing the relationship between median age and birth rate.



**Figure 26.** System architecture illustrating data flow and event-driven coordination across visualization modules.