Key Demographic and Development Indicators from 1960 to 2016

Cherilyn Christen
108780
IST – Alameda
cherilyn.christen@gmail.com

Rassene M'sadaa
108795
IST – Alameda
rassene.msadaa@gmail.com

Zeineb Mellouli
1108825
IST – Alameda
zaynebmellouli25@gmail.com

ABSTRACT

UPDATED—27 October 2023. This sample paper describes the formatting requirements for key demographic and development indicators, and offers recommendations on writing for the worldwide demographic trends. Please review this document even if you have submitted to the world bank data conferences before, as some format details have changed relative to previous years. Abstracts should be about 150 words and are required.

INTRODUCTION

This project offers dynamic visualizations of key demographic and development indicators from 1960 to 2016. The dataset encompasses Fertility rates, Natural rates, Human Development Index (HDI), Replacement rates, and Life expectancies for countries globally. The initiative aims to provide insights into global socio-economic dynamics and address questions using animated plots and visual representations.

We aim to investigate the relationship between increasing life expectancy and decreasing fertility rates, focusing on tracking these trends over time using choropleth maps and animated plots. We identify countries with consistently higher fertility rates than replacement rates, visualizing these demographic dynamics through choropleth maps and Sankey diagrams. Additionally, our visualizations explore the implications of declining fertility rates, particularly in countries experiencing population aging. We differentiate trends in fertility rates and life expectancy between developed and developing countries by incorporating the Human Development Index (HDI). Color-coded visualizations facilitate meaningful comparisons.

Possible Tasks

This visualization will allow to:

Search

- Search a country in the map and visualize their current information about different demographic indicators in different plots.

Explore

- Visualize the evolution of the demographic indicators in a specific country through the years.
- Visualize which are the countries with different values of fertility rate, life expectancy, HDI, natural rate and replacement rate (low, intermediate and high).
- Visualize which are the countries with lower and higher value of life expectancy.

Identify

- Observe a country's demographic indicators in a certain year.

Compare

- Compare the demographic indicators between continents and countries.
- Compare the countries with a higher human development index and the countries with more positive demographic indicators.

Example Questions

This visualization can help us answer the following example questions:

1. *Is there a discernible relationship between the upward trend in life expectancy at birth and the concurrent decline in fertility rates across countries from 1960 to 2016?*
2. *Are there countries where the fertility rate has consistently exceeded the replacement rate throughout this timeframe?*
3. *What are the demographic implications of countries with declining fertility rates, particularly in terms of population aging?*
4. *How do fertility rates and life expectancy trends differ between developed and developing countries?*
5. *Can you detect any significant anomalies or outliers in the dataset that require further investigation, such as countries with exceptionally low fertility rates or life expectancies?*

RELATED WORK

Regarding the inspiration and motivation for this work, it all began with our desire to present a work that could corelate important subjects. However, theoretical lectures were helpful to gather useful information that guided us to

structure the idea of how to encode the data, and what types of idioms to use in order to get the most out of available tools.

We used different sources such as websites, and scientific articles, to gather the highest amount of information, and data to support our subject. This led us to try to find correlation between different demographic indications and to really emphasize about this correlation we tried to combine 2 different indicators each time in a bivariate choropleth map. This website helped us a lot to encode our first idiom:

https://waterprogramming.wordpress.com/2022/09/08/bivariate-choropleth-maps/

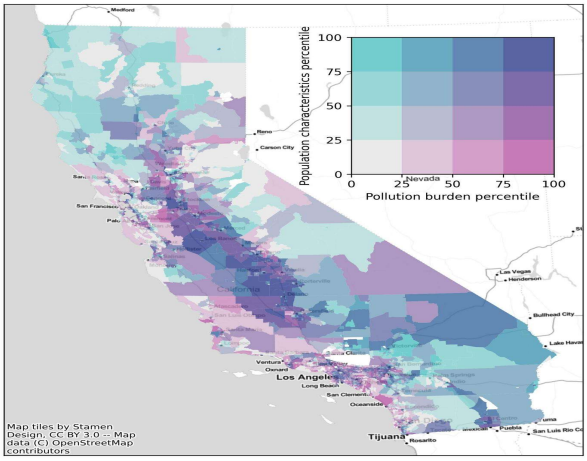


Figure 1. Bivariate choropleth showing pollution burden and population characteristic score percentiles simultaneously.

The idea of including the human development index was inspired from the website our world in data.

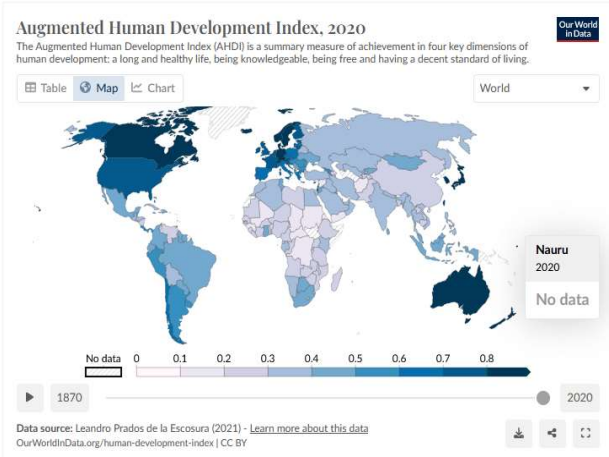


Figure 2. Example of visualization of HDI

This visualization was made using a single variate choropleth chart, so it did not provide an easy assessment of the comparison with the other demographic indicators.

For both Sankey plot and the slider, it was very hard to find some inspirations. We wanted a 3 slider in one (3 cursors) and the single library that worked was no-UI slider.

During the laboratories throughout the semester, we were able to learn from other groups approach to their subjects and obtain good feedback on how we could improve our work.

THE DATA

Our main source of data was the World Bank Open Data website that made the data available for everyone with no restrictions. You can sort through the Data Catalog to find the database of your interest. Its purpose is to provide high quality statistics regarding the whole globe, which is why we used so much information from it.

Raw Data

Originally our dataset was composed by 8 different .csv files (approximately 1351 Ko of crude data), related to multiple countries from a time period of, roughly, 1990 to 2016 – however, the most complete period of time in terms of data was approximately from the 1960 to 2018.

We faced some challenges when acquiring and processing the data since at the beginning we started to gather all our questions but the lack of variables that would relate these metrics and provide valuable information for the project on the available datasets led us to add some data and calculate derived attributes in order to have a complete data that will answer all our questions.

Missing Values, and Data Cleaning

We had multiple data files with heterogenous formats gathered from the same source but the .csv we added were from a different source. One of the initial challenges was to standardize these formats so that the dataset would become homogeneous.

To ensure the quality of the data needed to answer the raised questions, it was necessary to attend a process of cleaning, that involved eliminating unwanted columns with unnecessary attributes like the country code. To solve the issue of having missing values, we had to decide the best strategy to address this problem: if more than ¼ of the information for the item is missing we erase it otherwise we repeat the closest value.

Another problem we faced was that two of our data was initially represented by country, year and value. So, there are multiple lines where there is the name of the country and we wanted it to be represented by country, 1950, …, 2016. We managed to do it with the method ‘pivot’ (using pandas).

The last part of this standardization process was to preserve only the name of the country and erase its ISO code because it was irrelevant for our visualization.

VISUALIZATION

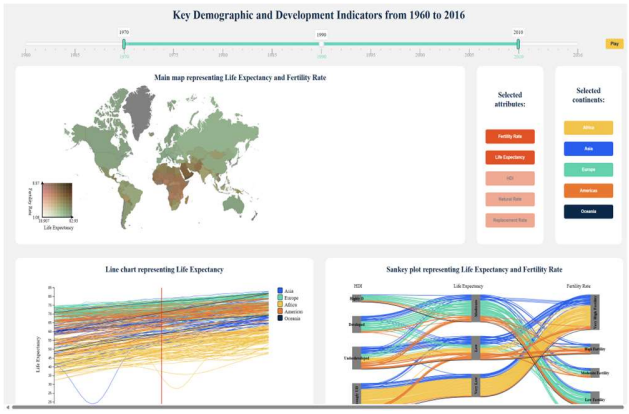


Figure 2. Overview of the Visualization Layout

Overall Description

Our final visualization layout consists in three different idioms, a choropleth map on the top right, a line chart on the bottom right corner, filter buttons for attributes and continents on the top left and a Sankey plot on bottom left corner. The full layout of the visualization is shown above in figure 2.

Choropleth Map

The centerpiece of our visualization is the choropleth map. We opted for this data visualization technique due to its ability to intuitively convey the data's geographical distribution. The inclusion of a two-dimensional color scale within our choropleth map was driven by the inherent complexity of our dataset, which comprises multiple attributes across various countries and years. It allows us to encode two attributes in a single visual channel—color. If you want to show only one single attribute on our choropleth map you can do it too.

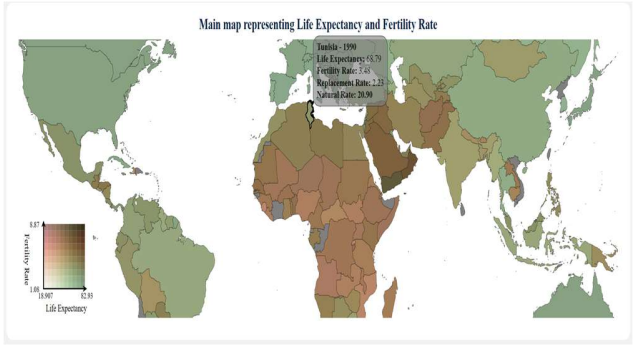


Figure 3. Choropleth Map for Life expectancy and fertility rate, with mouse pointer hovering over Tunisia

From the color scale it is possible to understand the differentiation of the countries life expectancy and fertility rate, the countries that have no data available have a filling in grey.

When the mouse is over a country, it gets a dark border and a pop-up tooltip shows the country's name and all the corresponding demographic indicators, as seen in figure 3.

Line Plot

Line plots excel at representing changes over time, making them an obvious choice for displaying the temporal evolution of demographic and development indicators. It allows the user to compare the selected attribute between different countries and continents.

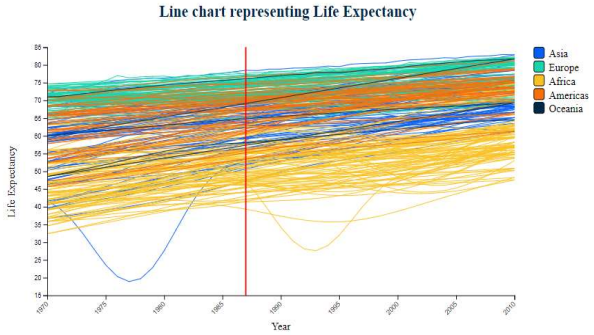


Figure 4. Line Chart for Life Expectancy

By encoding the temporal dimension on the x-axis and the magnitude of the selected attributes on the y-axis, we create a clear and intuitive representation of how these indicators have evolved for different countries across the selected time span.

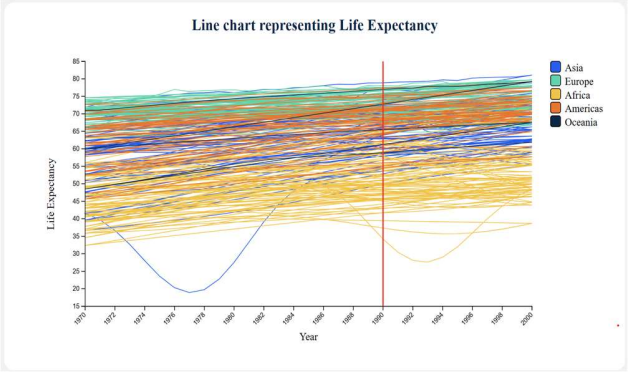


Figure 5. Line Chart for Life Expectancy rate from selected period (1970-2010)

The line chart is filtered through the selected period time (from the slider), the user can select the period s/he wants to observe by sliding the cursor at the top of the layout, just as shown in figure 5.

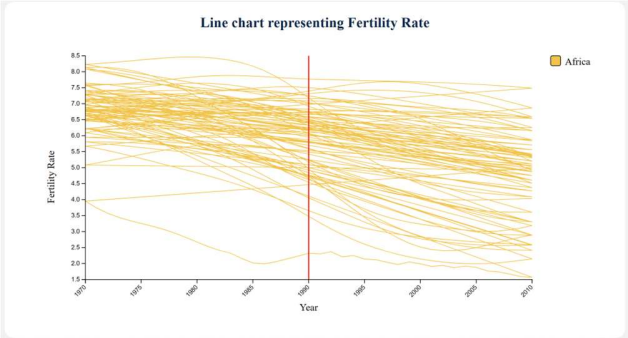


Figure 6. Line Chart for Fertility rate for Africa only

Furthermore, the utilization of colored lines corresponding to continents adds an essential layer of context. Users can readily distinguish trends and outliers associated with specific continents. The user could filter between continents from the continent's buttons top for a good observation, as you can see from figure 6.

Besides, we created a red bar positioned exactly on the year selected in the slider. The bar scrolls over the line chart when you're sliding the cursor on the slider on top it will guide you well through the years displayed in the line chart.

While our choice of line plot aligns well with our project's objectives, one prominent limitation of line plots is their inherent restriction to higher-dimensional representations. In our initial design considerations, we explored the possibility of utilizing two y-axes on a single line plot to represent two distinct attributes concurrently, effectively creating a three-dimensional plot. However, this approach presented practical challenges, primarily related to the management of an overwhelming amount of data within the confines of a single plot.

Sankey Plot

We chose to introduce the Sankey plot as an innovative visualization idiom, as it illustrates the intricate flow of data across categories and assist users in grasping the complex interplay between these variables.

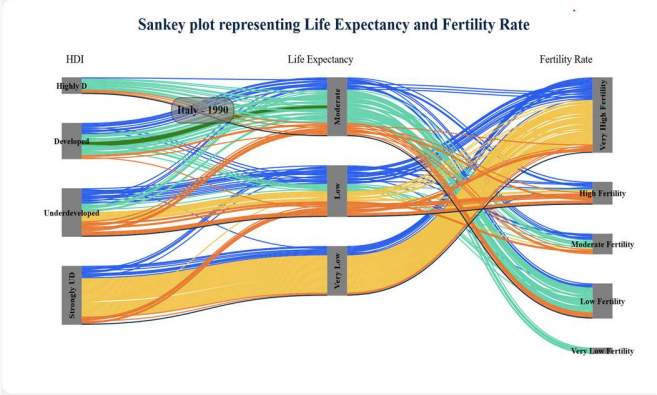


Figure 7. Sankey Plot representing Life expectancy and Fertility rate with mouse pointer hovering over Italy

In our implementation, the Sankey plot consists of two to three layers, each serving distinct roles. The first layer of nodes remains fixed, representing the developmental status of each country. The second and third layers, however, adapt dynamically based on the user's attribute selection. In the case of a single attribute selection, the plot transforms into a simplified two-layer structure, with the Human Development Index (HDI) value in the initial layer and the chosen attribute in the second. The inclusion or exclusion of the HDI attribute doesn't count as an additional attribute, as it serves as a fundamental baseline layer in the visualization.

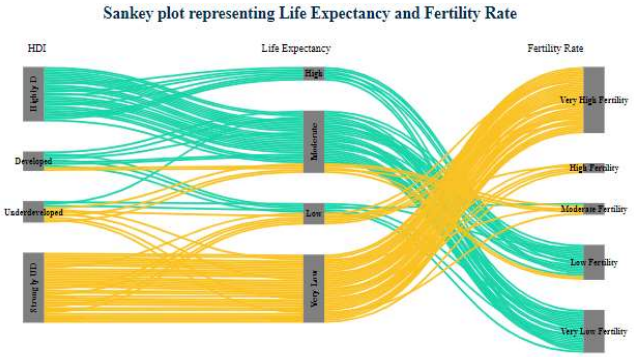


Figure 8. Sankey Plot representing Life expectancy and Fertility rate with selected countries: Africa and Europe.

You can also select only the continents you want to display on your Sankey plot for a good visualization and comparison, as you can see on figure 8. When the mouse pointer hovers over a particular country it highlights it on the choropleth map.

Furthermore, all attributes displayed in the Sankey plot are systematically categorized, derived from authoritative sources and international standards. This categorization provides users with a structured and systematic representation of the data adds a layer of coherence to the visualization.

Interactive Attribute Selection

An essential feature of our project is the incorporation of interactivity. Users can dynamically select attributes, such as fertility rate, life expectancy, and others, to observe their correlations and evolutions. This dynamic functionality promotes user-driven exploration and enables them to tailor the visual representation to address specific inquiries.

Filtering for Continents: To facilitate a more refined exploration of the data, we implemented the option to filter data by continents. This intuitive tool provides users with the capability to isolate specific geographical regions, thus offering a deeper understanding of the dataset's patterns and enabling more precise investigation into continental trends. This filtering mechanism facilitates the identification of outliers and anomalies in the data. By isolating data from individual continents, users can scrutinize the unique demographic and development characteristics of each region. It becomes easier to distinguish whether certain

countries exhibit exceptional trends or if there are notable variations within specific continents.

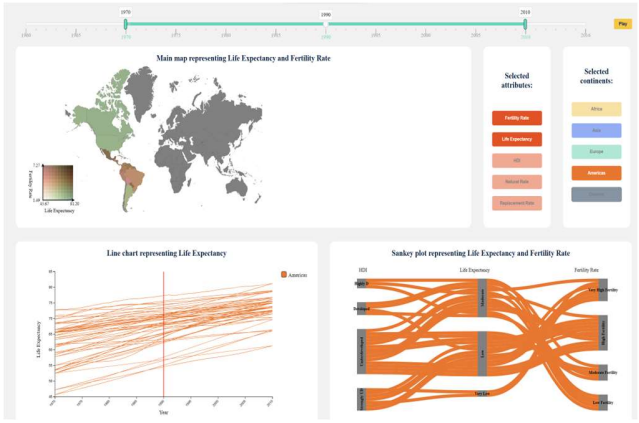


Figure 9. The overall data for the North America and South America only

Animation Over Years

Our project offers a dynamic experience that showcases temporal trends effectively, allowing users to explore demographic and development indicators spanning from 1960 to 2016. This animated approach serves the dual purpose of highlighting historical shifts and engaging users in an immersive exploration of the dataset. Users can choose to focus on a specific year, resulting in a static representation of the dataset for that single year. Alternatively, users can select any desired timeframe within the 1960 to 2016 range, initiating an animated presentation of the data. This dynamic visualization permits users to witness the evolution of their selected attributes over their chosen time frame, offering insights into long-term trends and transformations. The animation can be pause at any given moment by simply pressing the play button the second time.

Figure 10. Slider from 1970 to 2016 with the selected year 1990

Hover Text for Additional Information

To pinpoint outliers, we incorporated hover text functionality. This feature equips users with supplementary context when they interact with visual elements. For the Choropleth map, the hover text feature provides a complete view of the data for the selected year. Users can access information related to all attributes at once, allowing them to explore a broad spectrum of data with a simple mouse hover.
In contrast, when interacting with the Sankey Plot, users can hover over the links to access the name of the country. This selective approach ensures that the Sankey plot remains a concise and effective tool for exploring the data's interconnections.

Rationale

When we started to work on our layout sketch, one of the first steps was to analyze and select the visualizations that would better suit our data. As a group we had several exchanges of ideas and opinions, so that in the end we would come up with the best overall visualization layout.

Looking at our initial visualization prototype (figure 12) and the final layout version (figure 2), the two are different, the idioms remained the same but the transposition of the items changed completely and some modifications were added (mostly about interaction).

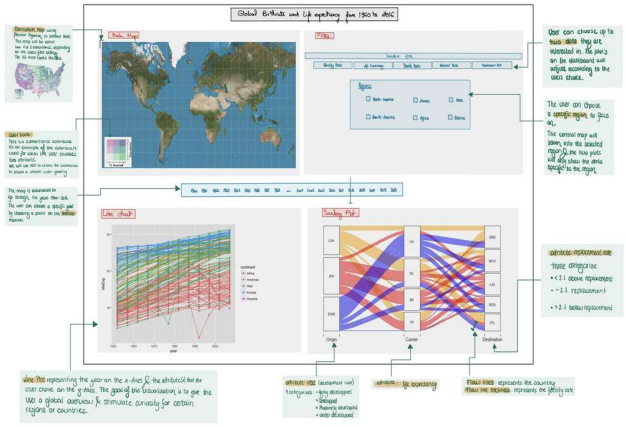


Figure 11. Initial Layout low fidelity prototype

When building the low fidelity prototype, we came out with different ideas but we were concentrated about doing three idioms the customized one is the choropleth map with its unique functionality to be a bivariate plot and represent two attributes at once. All of our questions were oriented to find the relationships between different demographic indicators and to find if there's a sort of correlation between them.

For our final visualization, we chose to stick with our initial prototype and add some modifications like the tooltip for choropleth map and Sankey plot, the red bar on the line chart. Also, at fist our Sankey plot in the prototype was designed to be fixed and not interactive at all but for our visualization it seemed more relevant to transform our Sankey plot into an interactive idiom and it could also show only one attribute at a time or two depending on the number of nodes' level.

Demonstrate the potential

For this part, we chose to answer the first question of our list.

Is there a discernible relationship between the upward trend in life expectancy at birth and the concurrent decline in fertility rates across countries from 1960 to 2016?

In order to provide an answer to this question, let's take a look at first to our choropleth map from 1960 to 2016. You can hit the button play to see the whole animation year by year and it's very obvious from the different degradation of colors you could tell there's a big correlation between a high

life expectancy and a low fertility rate. Let's take a country as an example: Tunisia.

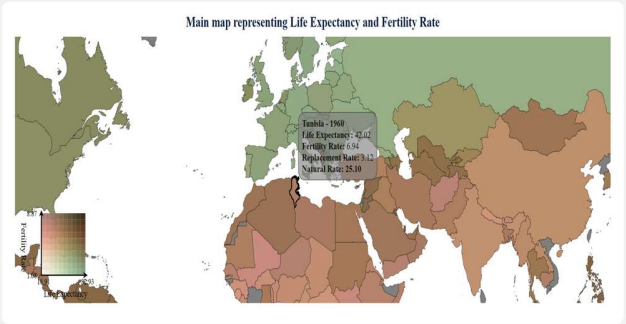


Figure 12. High fertility rate and low life expectancy in 1960 for Tunisia in the choropleth map

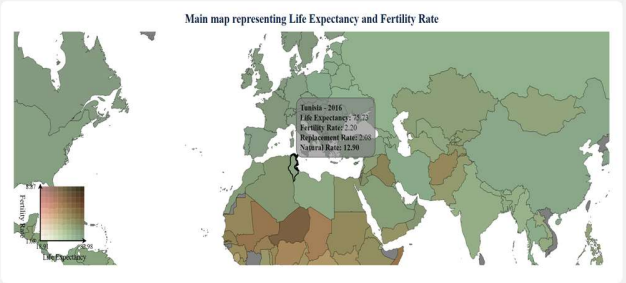


Figure 12. High life expectancy and low fertility rate in 2016 for Tunisia in the choropleth map

Secondly, we can move to our line chart to a best visualization of the evaluation through the years. We can deselect all the continents and keep only Africa. Life expectancy chart: When you hit the button play, the red bar will guide you through the years. It's easily noticeable that almost all the lines are increasing from low life expectancy (around 50 years for average in 1970-1980) to a high life expectancy reaching 75 years in 2016. Fertility chart: exactly the inverse behavior of life expectancy line chart.

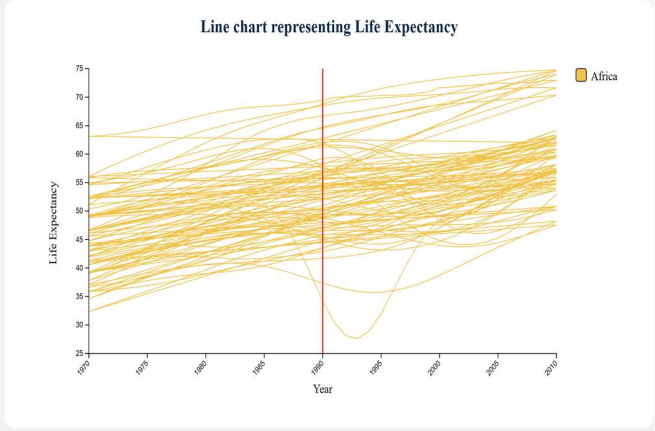
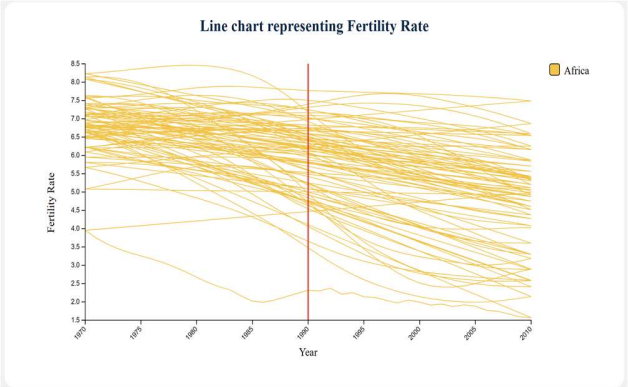


Figure 13. Increasing life expectancy in Africa



Finally, let's explore our Sankey plot. The difference between the two plots in 1960 and 2016 for Tunisia speaks for itself. In 1960, Tunisia is connected to a very low life expectancy and a very high fertility rate but for 2016 it's connected to a moderate life expectancy and moderate fertility rate.

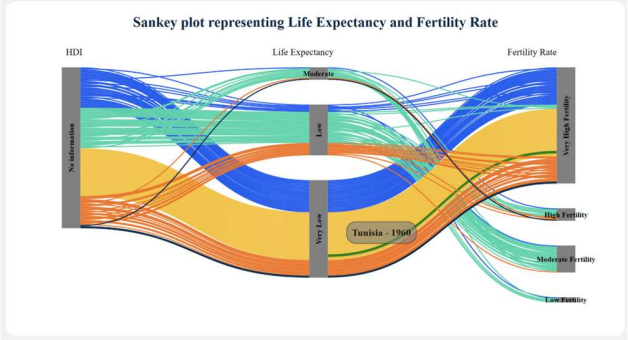


Figure 12. High fertility rate and low life expectancy in 1960 for Tunisia in the Sankey Plot

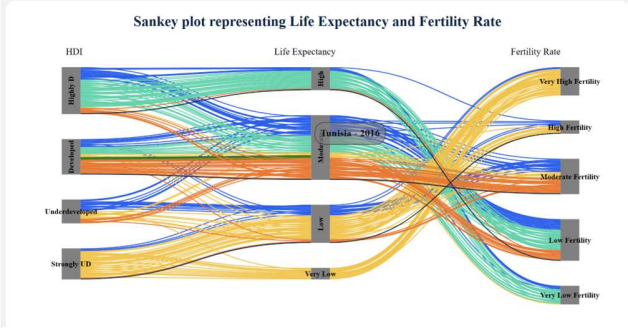


Figure 12. Moderate life expectancy and moderate fertility rate in 2016 for Tunisia in the choropleth map

Through the process of designing our prototype, we faced a lot of weird values and outliers. For example, for Camboge in 1975 the replacement rate was about 20, knowing that a normal value would be around 3.

IMPLEMENTATION DETAILS

One of the requirements of implementation for this project was to use d3.js, which is a *JavaScript* library that facilitates

the production of dynamic and interactive data visualizations.

Not all the idioms were adapted from D3 examples, for example, the slider was implemented from scratch, since none of the examples we found wasn't even near of what we wanted for our visualization. The Sankey plot was almost implemented from scratch because the library mostly used was the Sankey.

The idioms that were modified and adapted to fit the desired functionality of our visualization were the line chart and the choropleth map.

CONCLUSION

In this project, we set out to explore and visualize a wide array of global data pertaining to demographic and development indicators from 1960 to 2016. Our primary objective was to provide insights and answers to key questions through the use of various data visualization techniques. Let's reflect on what we've learned and accomplished.

Demographic and Development Trends: We gained a comprehensive understanding of the global demographic and development trends over the past five decades.

Bivariate Analysis: Our choropleth map with bivariate filter buttons allowed us to observe the correlations between different indicators. This feature provided a nuanced view of how these indicators interrelate.

Temporal Analysis: The line chart enabled us to analyze how these indicators evolved over time. We could track changes and assess whether progress was consistent or varied across continents.

Flow of Resources: The Sankey plot helped us visualize the flow of resources, further deepening our understanding of global development. It highlighted the distribution of resources and their impact on socio-economic factors.

We successfully addressed the key questions posed in the project. Our visualizations provided insights into the data, helping to answer questions about demographic shifts, development disparities, and the relationships between various indicators.

If we were to start over, there are a few aspects we might approach differently:

Enhanced User Interaction: We would focus on improving user interaction and customization. Allowing users to select specific countries in the choropleth map and display all the other charts with this specific country only.

Deeper Analysis: We could explore more advanced analytical techniques, such as predictive modeling or machine learning, to uncover hidden patterns and trends within the data.

With an additional month and a budget of €3000, we could further enrich our solution:

Geospatial Analysis: We could incorporate geospatial analysis to dive deeper into regional disparities. This might involve adding geographic heatmaps or more complex spatial analyses.

Mobile-Friendly Interface: We would invest in making the solution mobile-friendly, ensuring accessibility to a broader audience.

Real-Time Data: Implementing real-time data updates would provide the latest information for decision-makers. From 2016-2023.

Machine Learning: We might develop machine learning models to forecast future demographic and development trends, making the solution more predictive.

Community Engagement: Allocate resources for community engagement and feedback collection to understand the needs of our users better.

In conclusion, our project has offered valuable insights into global demographic and development indicators through the use of data visualization. While we've successfully addressed the initial questions, there is always room for improvement and expansion. With more time and resources, we could take this solution to new heights and provide even more comprehensive and insightful analyses for our users. The journey doesn't end here; it's a stepping stone to more in-depth exploration and understanding.

REFERENCES

1. World Bank Data

https://data.worldbank.org/

2. Choropleth Map D3 Example

https://bl.ocks.org/LemoNode/5a64865728c6059ed89388b5f83d6b67

3. Line Chart D3 Example

https://d3-graph-gallery.com/graph/line_basic.htmlSankey Plot library

4. Sankey Plot Library

https://ec.europa.eu/eurostat/web/products-eurostat-news/-/EDN-20180423-1

5. noUiSlider Javascript

https://refreshless.com/nouislider/

6. Sankey plot repository
https://github.com/topics/sankey-diagram?l=javascript