World Demographics

Group - 8

Submitted To-

Dr. Dootika Vats

Submitted by-

- Sameer Verma
- Suresh Rana
- Vishnu
- Jiyanshu Dhaka

Introduction

While studying World Demographic,we may assume that population growth is uniformly high worldwide, while in reality, growth rates vary significantly among regions and countries.

We all have different views on how the world should work. For example, some may think that if people move out of a country(Migra), its population density will decrease. However, this may not be true, as the country's high fertility rate could actually increase population density.

There might be a misconception that aging populations are a concern only in developed countries. However, many developing nations are also experiencing demographic transitions. There might be an assumption that urbanization is a universal trend, but the pace and nature of urbanization vary across regions and countries.

It's important to know that not every country follows the same patterns or beliefs. Also, these factors are not influenced by just one thing. They are connected in complex ways with many different factors playing a role.

It is important to realise that not all countries follow any certain trend or our conventional beliefs. Further these factors aren't solely influenced by a single variable; Instead, they are subject to complex interdependencies with numerous contributing factors.

Motivation

Studying world demographics helps in understanding global population trends, such as population growth, aging populations,Mortality rates,urbanization etc.

Policy Planning: Governments and international organizations use demographic data to plan policies related to healthcare, education, immigration, and social services.

Economic Development: Demographic factors impact economic development. Population size, age structure like Median Age, and migration patterns influence labor markets, consumer behavior, and economic growth.

Healthcare Planning: Median Age and age distributions are crucial for healthcare planning. They help in determining healthcare needs, allocating resources, and addressing public health challenges.

Urbanization Studies: Urban population and population density data are essential for studying urbanization trends, planning infrastructure, and addressing challenges associated with rapid urban growth.

Migration Patterns: Understanding migration patterns is important for policymakers and researchers. It helps address issues related to immigration, emigration, refugee movements, and their social and economic impacts.

Aim

We started this project to understand how different indicators are connected. We believe these factors influence each other in some way, and that's what we wanted to figure out. While working on it, we found some interesting results that we'll talk about later in the report.

Data

We have obtained the data by data scraping and downloading csv files and then arranging the data in a compact combined data table.

Data Scraping :

Data obtained by data scraping: This Data was scraped from Worldometer and Numbeo websites.

<**https://www.worldometers.info/world-population/population-by-country/ (https://www.worldometers.info/world-population/population-by-country/)**> <**https://www.numbeo.com/cost-of-living/ (https://www.numbeo.com/cost-of-living/)**>

Data contained information of more than 200 countries from 1955 to 2023, with a gap of 5 years between data of 1955 to 2020 years and then data of 2022 and 2023. We cleaned the data and selected 115 countries for our analysis.When scraping data from worldometer and Numbeo, we frequently encountered connection time-out errors, requiring us to restart the entire process.

####The variables included in this dataset are :

- Population
- Median age
- Yearly change in population
- Urban Population Percentage
- Migration
- Population Density
- Fertility rate
- Urban population
- Country share in world’s total population

From .csv files:

.CSV file: We needed to add Morality Rate in the final dataset as it is one of the important demographic factor. we downloaded a .csv file from ourworldindata.org the .csv file had data of Number of deaths from 1950 to 2100 for all countries . this data was half historical and half predicted . We took only the historical data and the years were different for different countries. The countries names were not matching as well, it was needed to be cleaned and organised according to the previous data. As the data was very unorganised we had to manually make a new .csv file which had data of number of deaths of required 115 countries with their correct names and year range.Then we had to divide the number of deaths by country’s population to get the mortality rate.

Normalising data

We Normalised the data to do cross sectional analysis of different factors.

Arranging, Normalising and Scaling Discrepancies: We faced challenges with the scaling of data, as many variables were in different scales and units. For instance,population density was in people/square km , while median Age was in years.

Biases in the data

Certain countries had incomplete data, it can introduce bias. For instance, if a country has a lack of infrastructure for data collection, its demographic information may be underrepresented. Data on migration seemed biased as certain countries were more likely to be undercounted or as there might be discrepancies in reporting immigration and emigration.In our heatmap analysis, we focused on data for discrete timeframe due to the presence of data values in between the years. This decision could introduce a selection bias and potentially impact the correlation values.

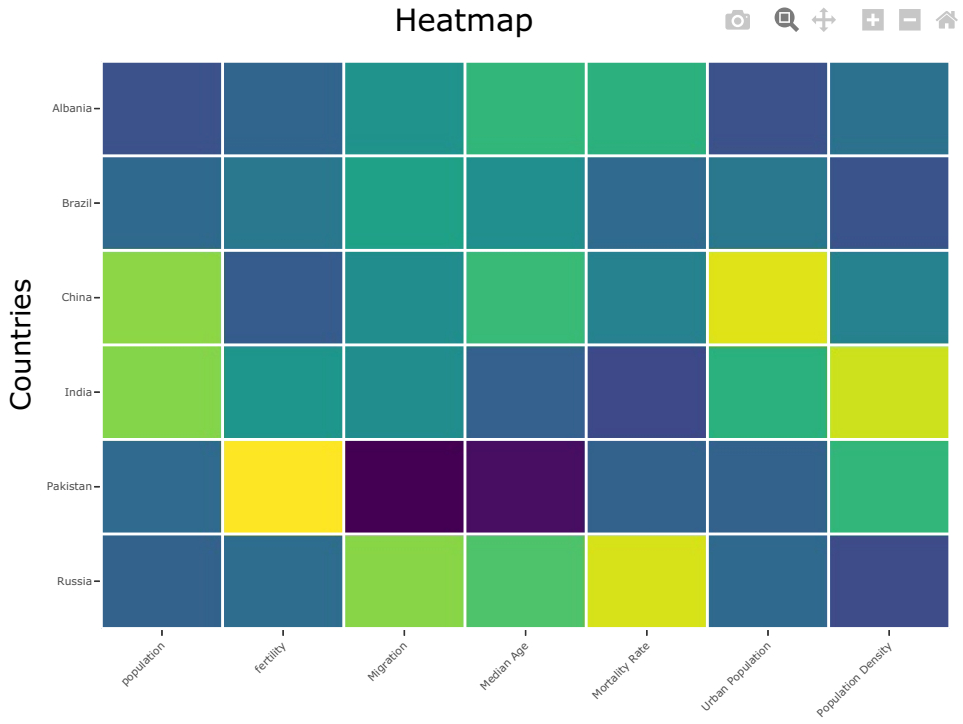
Key Questions

How are different development indicators dependent on each other and how they affect a country?

Are there any countries that show an abnormal pattern or are deviating from the expected pattern?

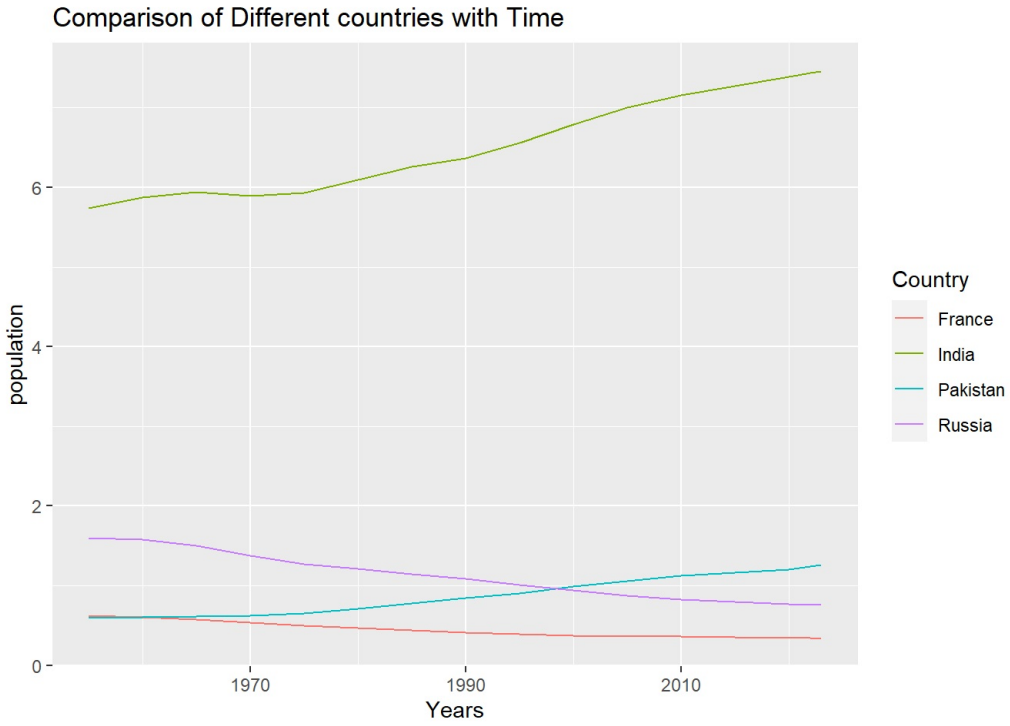
Important Visualizations

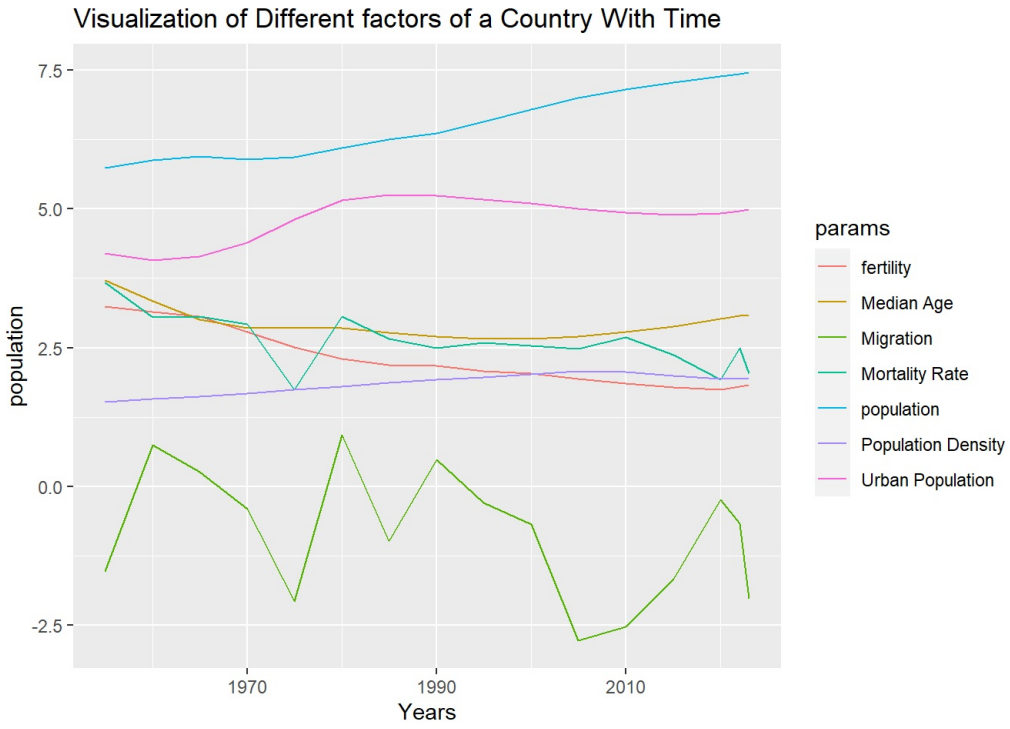
Firstly, we've tried to give a visual representation of our dataset in the form of a heatmap.



####The heatmap above provides a visual representation of a subset of our dataset. This specific heatmap is for the year 2020, featuring countries like India, China, Russia, etc. Different colors indicate the individual values of factors on the x-axis for each country on the y-axis. #### the data for various factors are in different scales, leading to significant variations among them. To address this, we've normalized the data, allowing for comprehensive visualization of all factors in a single graph. The colorbar on the side illustrates that yellow represents high values, dark blue represents the lowest values.

```
## [1] "character"
```


Comparison of Different countries with Time
population
6
4
2
0
1970
1990
2010
Years
Country
France
India
Pakistan
Russia


Visualization of Different factors of a Country With Time
population
7.5
5.0
2.5
0.0
-2.5
1970
1990
2010
Years
params
fertility
Median Age
Migration
Mortality Rate
population
Population Density
Urban Population

Conclusions

From the observed graphs, it becomes evident that the factors are interconnected. However, it's crucial to avoid generalizing assumptions to all countries, as we've already witnessed instances of contrasting behaviors.

References

- **<https://www.worldometers.info/world-population/population-by-country/ (https://www.worldometers.info/world-population/population-by-country/)>**
- **<https://www.numbeo.com/ (https://www.numbeo.com/)>**
- **<https://ourworldindata.org/ (https://ourworldindata.org/)>**
- **<https://r-graph-gallery.com/ (https://r-graph-gallery.com/)>**