DSCI 5360\_03 Data Visualization for Analytics

Project Writeup Final

1. Introduction and Background

a. Title:

Global Development

b. Motivation:

We were motivated to choose the topic of global development because understanding the development of countries and regions of the globe can lead to a better understanding of why or how policies are chosen. Development level of a region is also important to understand for companies or organizations when considering investment potential outreach needs.(“World Development Indicators”, 2024)

c. Context:

In this project, we will be exploring the topic of development of different countries and regions around the world and try to categorize them into developed, developing, or conflict status.

d. Objectives and Queries

- We aim to learn how to use Tableau to answer questions about data.

- We aim to learn how to tell a story with data

- We aim to learn when to use certain types of visualizations to express information in the clearest way possible.

- We aim to learn how to create clear and thought-provoking visualizations that help us to learn more about global development and factors influencing it.

Research Questions:

Population and Urbanization Patterns:

-How do urbanization rates vary globally during 2021-2023?

-Which regions show significant urban population changes and what might be driving these changes?

-What is the relationship between urbanization and population growth/decline in different countries?

-Which countries show the highest and lowest population growth rates (2021-2023)?

-What factors (like conflict, economic conditions) might explain population decline in certain regions?

The insights from these visualizations will help us better understand the complex relationships between population dynamics, education, and development across different regions of the world.

e. Benefit

Understanding global development is key because it shows the health of a country and gives insight into its population, living standards, ect. It also helps both citizens and government officials make informed decisions on policies based on the country's level of development or regional issues they may be facing.

2. Objectives and Goals

a. Project Goals:

- We aim to express information in the clearest way possible.

- We aim to use visualizations to express the population growth of different countries throughout recent years.

- We aim to use visualizations to express the fertility rate of different countries throughout the years.

- We aim to use visualizations to express which countries have the highest and lowest fertility rates.

- We aim to use visualizations to express how education enrollment may play a role in fertility rate.

b. Visualization Objectives:

-Create interactive and informative maps to display global patterns of urbanization rates, fertility rates, and education enrollment

-Design clear and effective dashboards that combine multiple visualizations to show relationships between different development indicators

-Utilize appropriate chart types (bar charts, maps, line charts) to best represent different types of data

-Implement proper color schemes and legends to ensure data is easily interpretable

-Build visualizations that allow users to explore data through filtering and interactive features

-Create visualizations that tell a coherent story about global development patterns and relationships between different indicators

-Ensure visualizations effectively support our research questions and help reveal patterns in development over time

c. Impact:

This project benefits others by helping them gain a better understanding about different countries. Understanding development is a great indication of how a country is doing economically, in policy, in terms of living standards, and technologically, while country development is not a perfect measure of these things it is a great way to understand more about them. This project could be very helpful during a political climate so people can better understand which policies the government should or should not be enacting.

3. Datasets

a. Origins and Support:

World Bank Group collected the data and was funded by trust funds, a financing arrangement set up with contributions from one or more development partners, to complement core funding from the International Bank for Reconstruction and Development (IBRD), and the International Development Association (IDA). World Bank Data is used by many entities for all kinds of economic and social research projects.

b. Variable Insights:

This data set covers a multitude of economic factors across 2000-2023 which was obtained from World Bank Data. The variables we are interested in analyzing are urbanization rates, population change, fertility rates, and education enrollment levels. With this data we want to try and correlate which variables contribute to GDP growth or shrinkage.

c. Dataset Scope and Dimension:

- 14000 lines of countries in the world spanning 2000-2023

- Global

- Features include:

· Population

· Exports and Imports

· GDP

· Gender and age demographics

· Urbanization and rural population

· Education enrollment

4. Visualization and Data Preprocessing Plan

a. Visualization Plan:

For this project, the goal is to create at least ten visualizations, including at minimum one map, one table, and two dashboards. Because there are so many factors that impact and play a part in GDP and so many countries included in the dataset, there is a plethora of information that can be used to create even more visualizations than the multiple maps and two dashboards and the many other visualizations that have already been created. We do plan to elaborate further on the information about school enrollment and its impact on fertility rate and the relationship between fertility rate and GDP. We would also like to see the relationship between school enrollment and GDP, as the relationship between school enrollment and fertility rate we have already lightly explored was also interesting.

b. Chart Proposals:

Propose at least four chart types, listing the specific variables you intend to use for each. Provide detailed descriptions of these variables, including their source dataset, units (or categories), and any other relevant information.

- A line chart showing the change of fertility rate throughout the years of the countries with the highest and lowest fertility rates. We plan to use variable like fertility rate, country names, and specific years ranging from 2000 to 2022

- A histogram showing the GDP of the countries with the highest and lowest GDPs throughout the most recent years. We plan to filter countries with the top 5 and bottom 5 GDPs and use the years 2021, 2022, and 2023 to create this visualization.

- A map outlining primary and secondary school enrollment. We would use longitude and latitude and filter by school enrollment and countries. We would include marks like coloring by school enrollment and detailing by country name.

- A table expressing the fertility rate, school enrollment, and urbanization growth of the counties with the highest and lowest GDPs. We would use elements like country (filtered by the countries with the top 5 and bottom 5 GDPs), fertility rate, school enrollment, and urbanization growth. We would mark GDP, fertility rate, school enrollment, and urbanization growth by label.

c. Alignment with Goals:

Our proposed charts have helped us to learn how to use Tableau to answer questions. We learned that we can see which countries have the highest and lowest fertility rates and what they are just by using the filtering function. We learned about the school enrollment of the countries with the highest and lowest fertility rates by creating a histogram and making use of the filters. Our chart proposals will help us to learn even more about gross domestic population of countries by diving deeper into the topic and answering questions that have been brought up by the visualizations that we have already created. We have found that just by attempting to answer the questions brought up by our data, we learn how to make better use of Tableau in order to answer these questions.

5. **Visualization Design and Discussion:**

**Population Makeup Dash**:   
Urbanization growth is based on the percent increase of a country's urban areas. Rural population growth is based on percent increase of a country's rural areas.

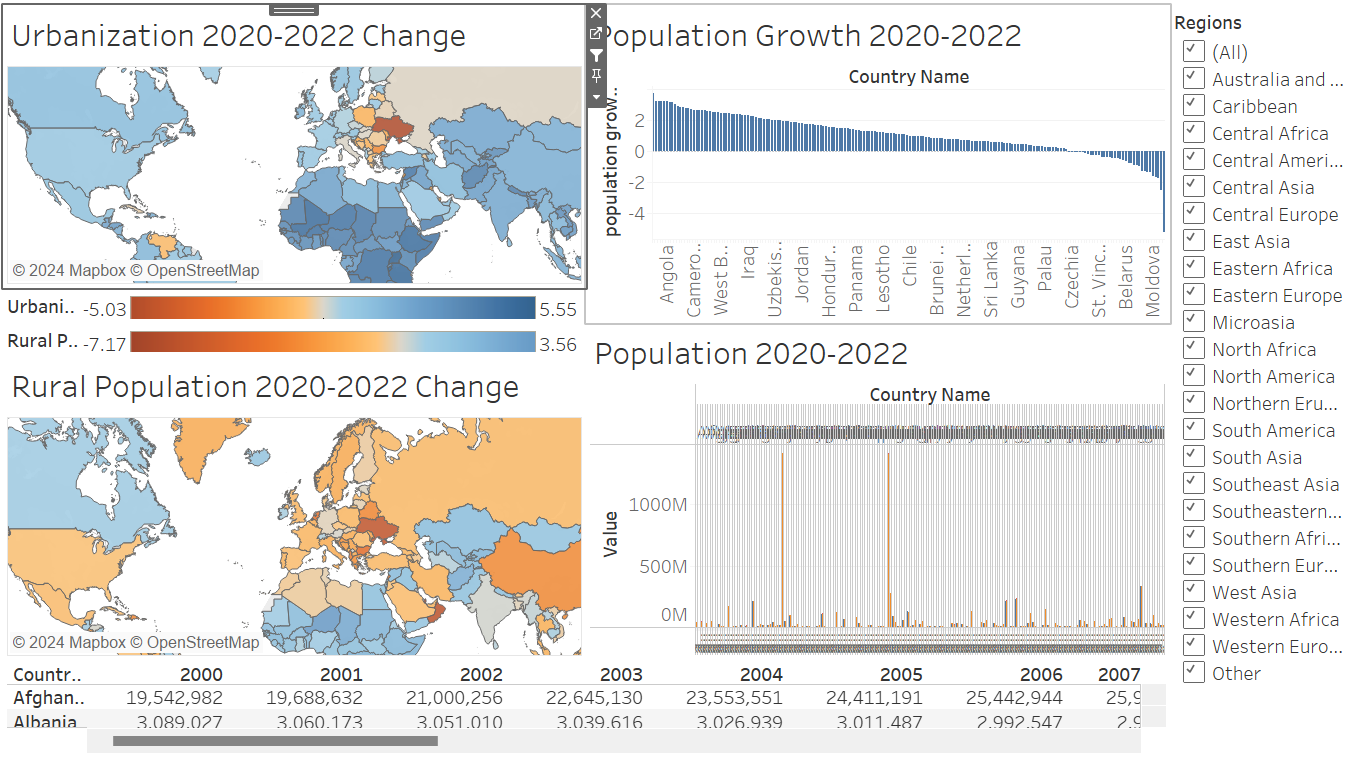
We grouped each country into their respective regions based on the United Nations Geoscheme.

From this information we can get an understanding of classifying countries into several groupings.

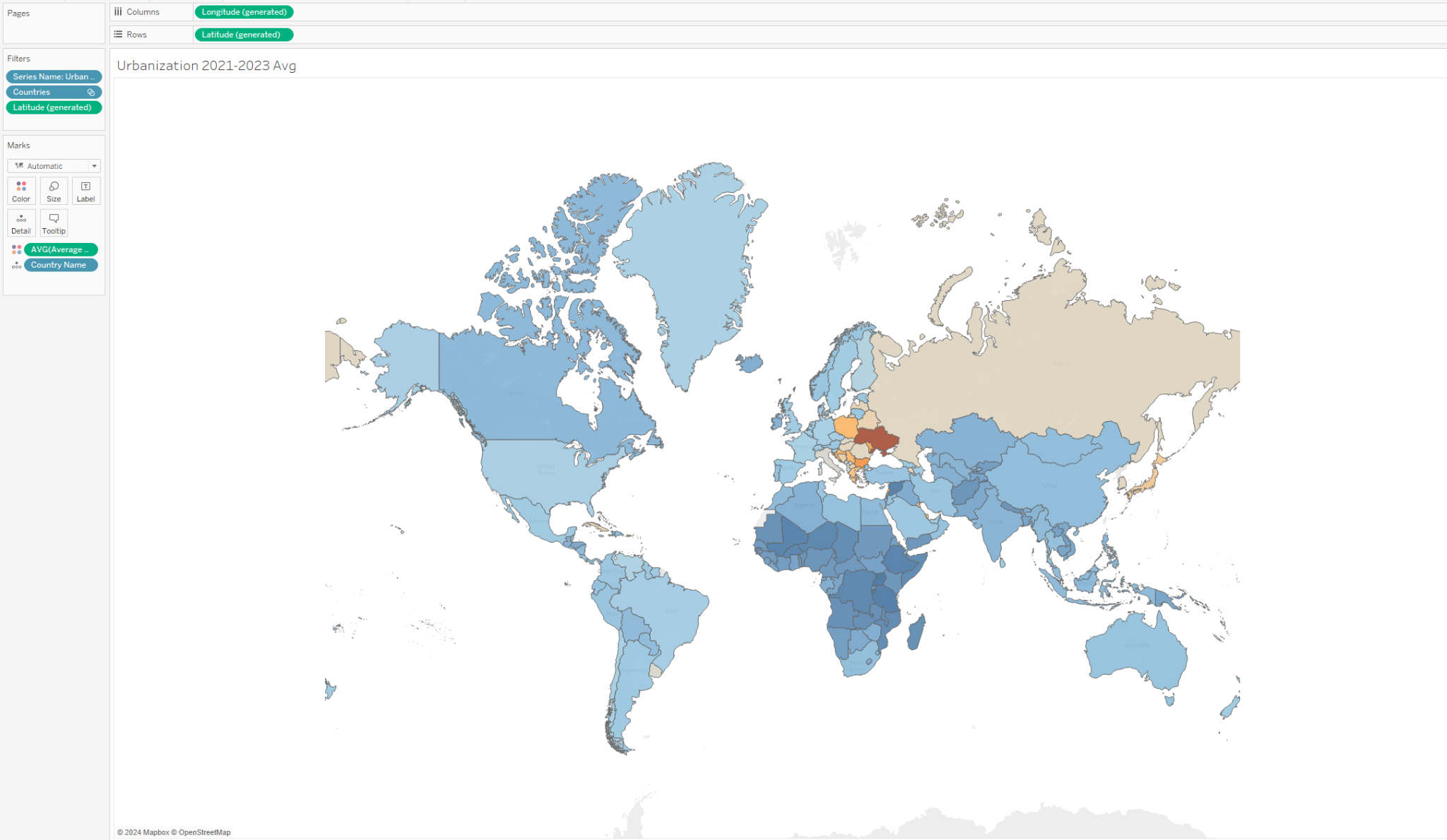
1. Countries which have seen both a decrease in urban and rural population. Meaning they are experiencing a shrinkage in their population. Ukraine is a top example of this due to the ongoing conflict with Russia. But conflict zones are not the only factors contributing to shrinkage. Balken states are also seeing a decrease in population.

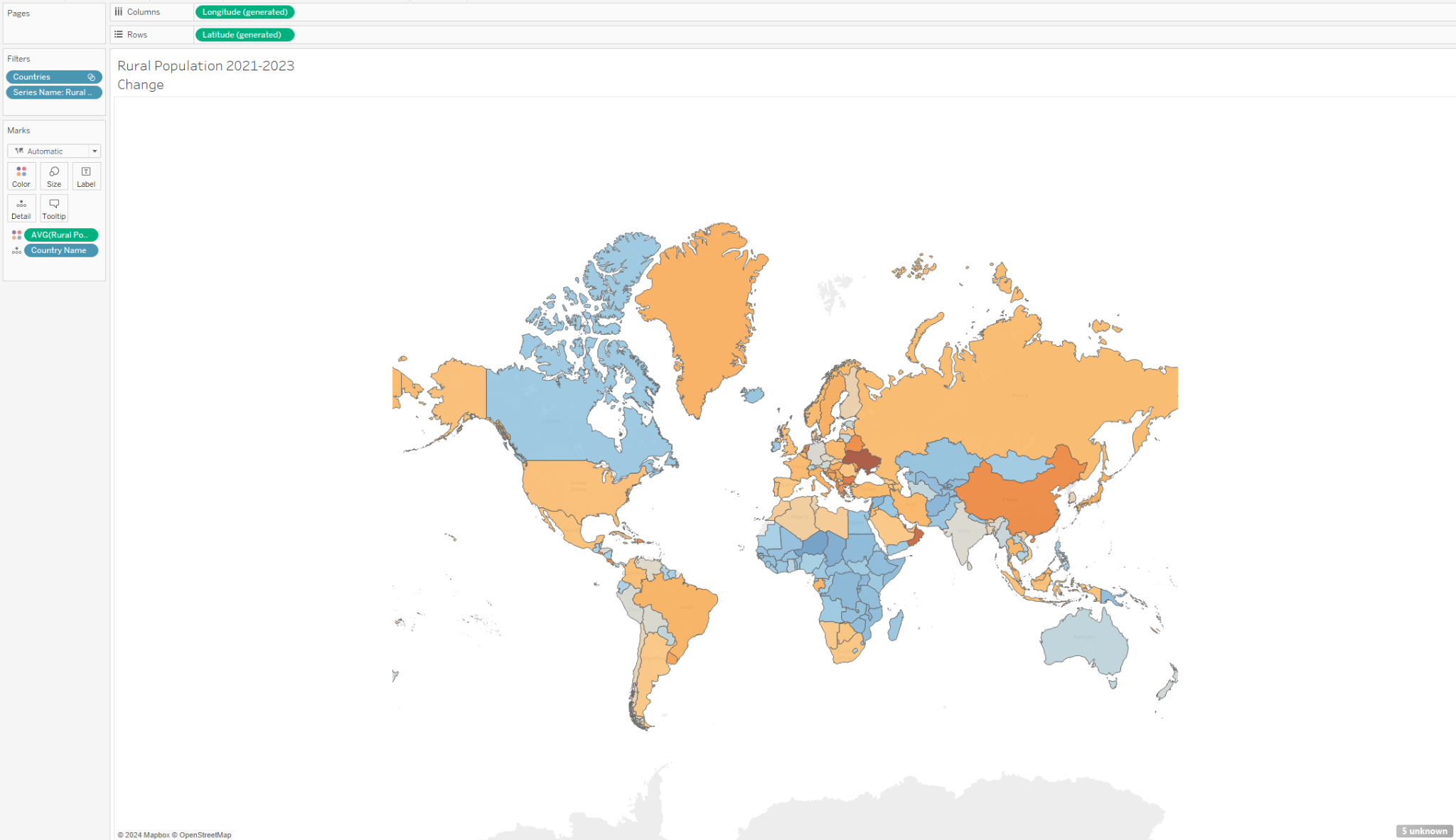
2. Countries seeing both growth in urban and rural population can be classified as an overall growing region. Mali has both an expanding urban and rural population and can be classified as overall expanding in population. An interesting country seeing overall growth but maybe for different reasons is Syria. The reasons being they are next to a conflict zone and could be receiving an influx of refugees/immigrants from Lebanon.

3. Countries experiencing a drop in rural population but an increase in urban areas can be classified as a more developed or industrializing country. Big players within this group are the U.S, China, Japan. Most already modernized/industrialized countries would belong in this classification.



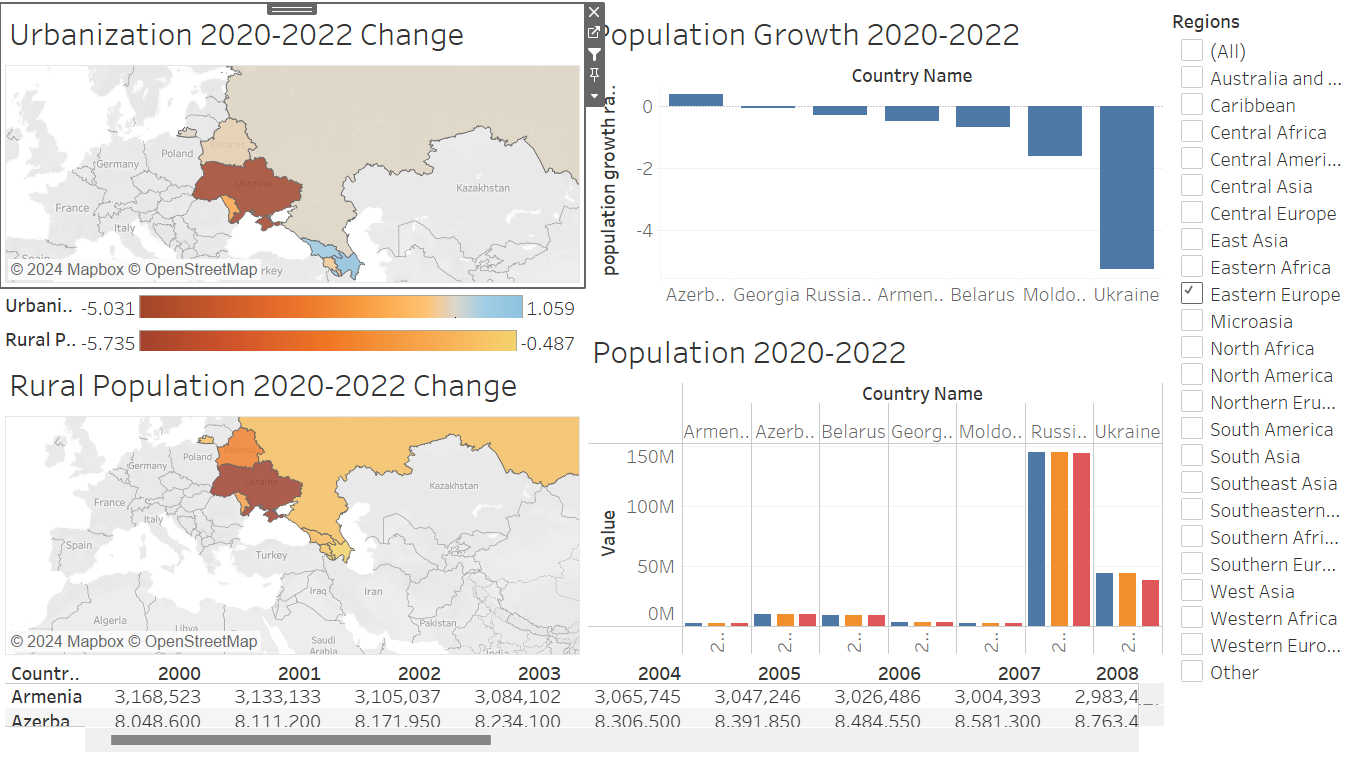
For mapping Urban Population and Rural population change maps we used the following configuration of features.   
Filtering for urban/rural feature set.  
Selecting target country values and removing regional aggregates.   
Calculation for average between the selected range: ([2023]+[2021]+[2022])/3

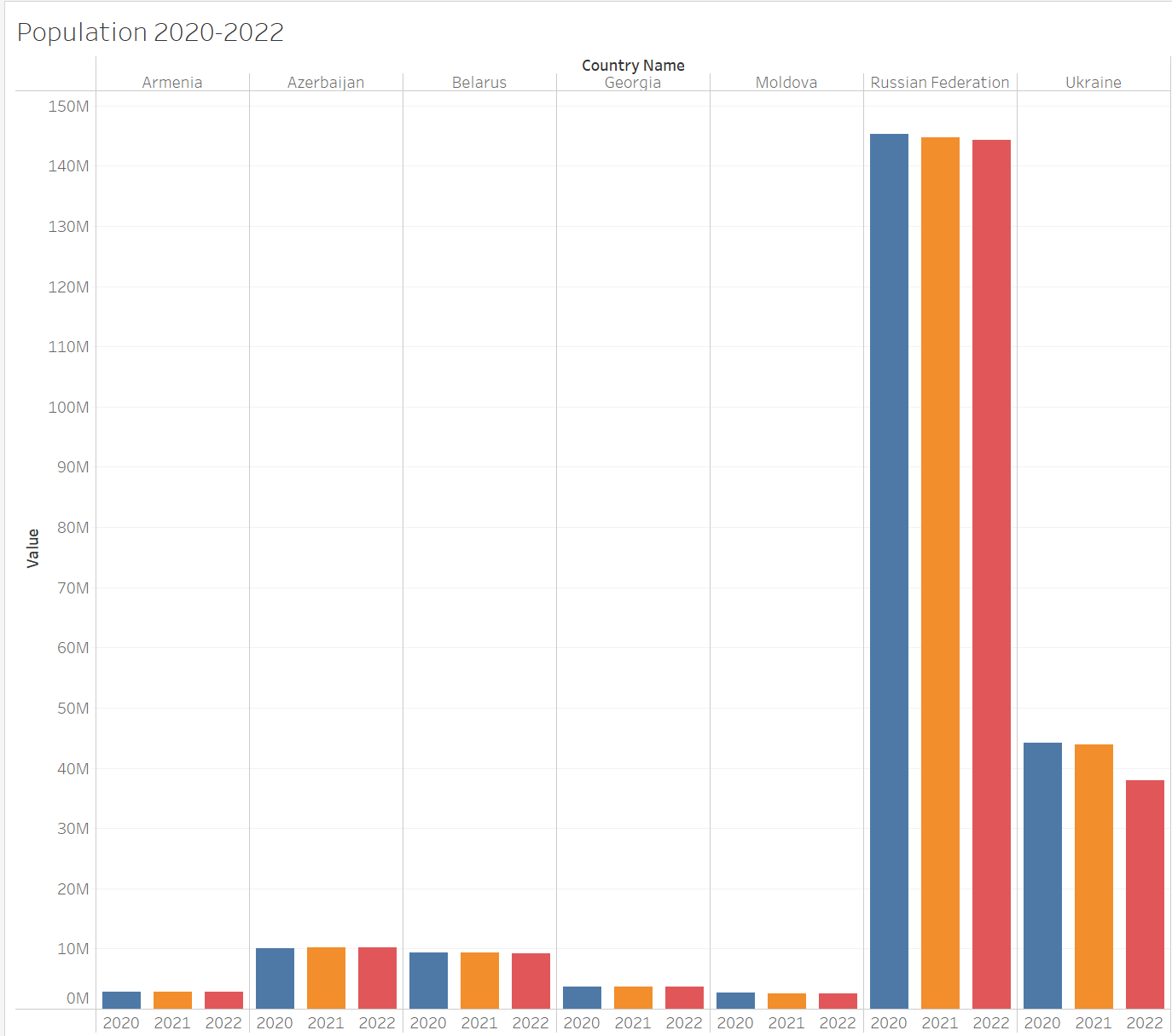
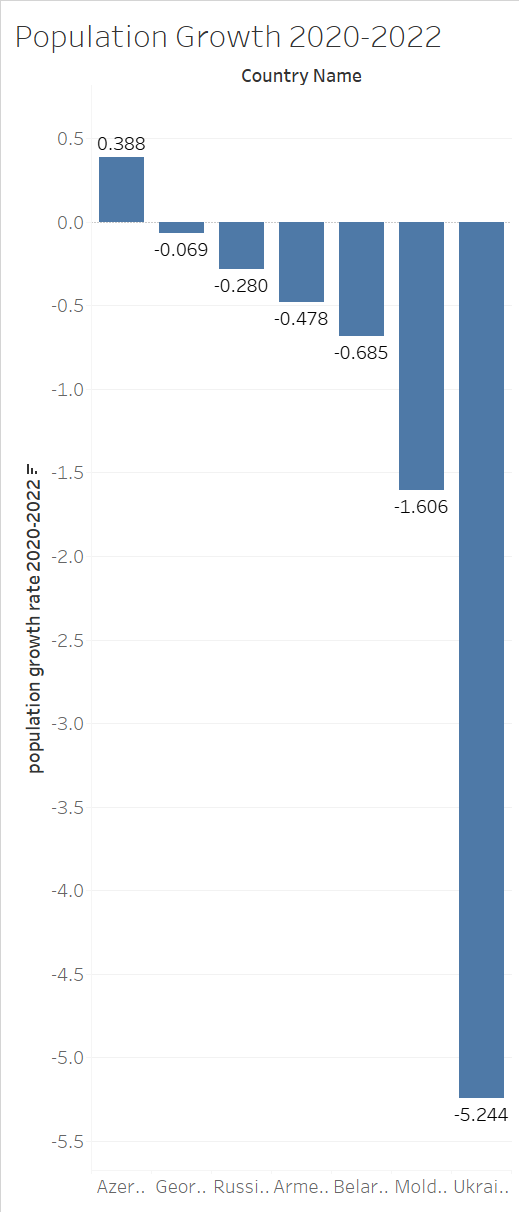




The mapping provided a good visual for quick identification of countries experiencing urban/rural change.  
A quick analysis of that change through this visualization shows the world is experiencing a greater increase in urban population then rural population.   
Globally:   
Urbanization rate for the 2021-2023 span is 1.58  
Ruralization rate for 2021-2023 was .73

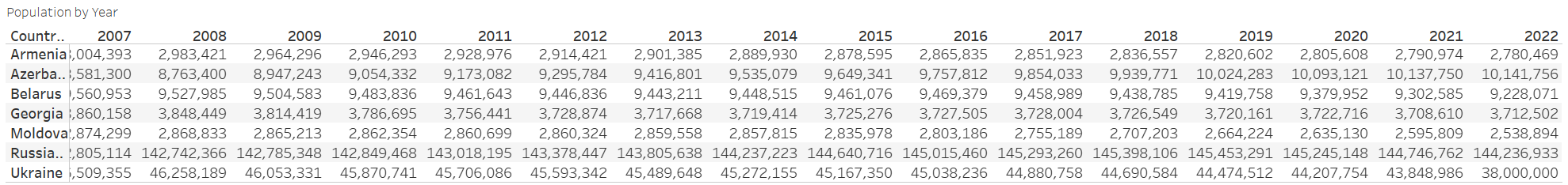
Now we will target a region of interest for analysis. For this example we will review the Eastern European region.

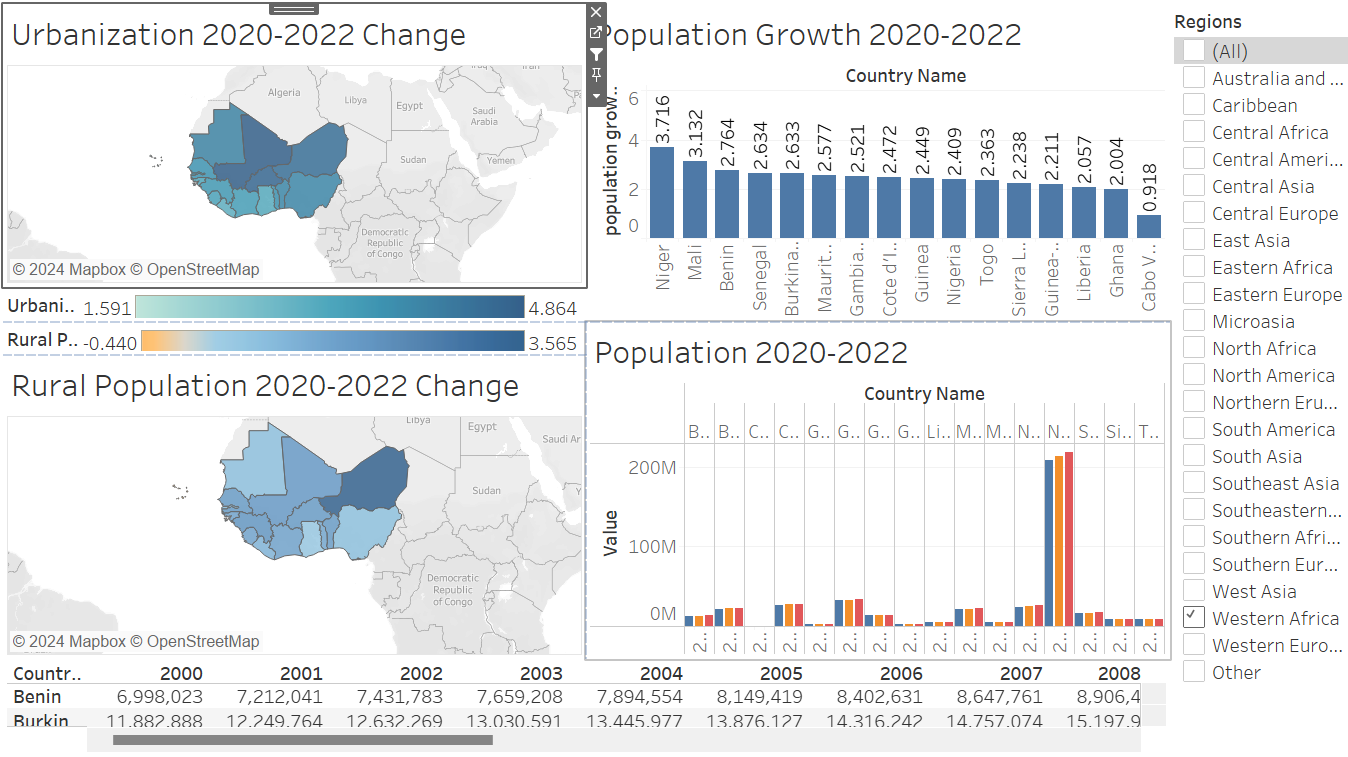




The above right chart displays total population change over the selected years (2021-2023).  
The left chart displays population growth averages for those select countries.

Overall the eastern european region is having a decline, Ukraine being the most obvious example of a decline. The reason behind this is the ongoing conflict with Russia who is also experiencing a decline in population growth for the period of time. The only country within this region group which has a positive population growth is Azerbaijan at .368.



While eastern europe is an interesting region to analyze with decreasing population West Africa on the other hand is showing to have the strongest population growth in the world. 

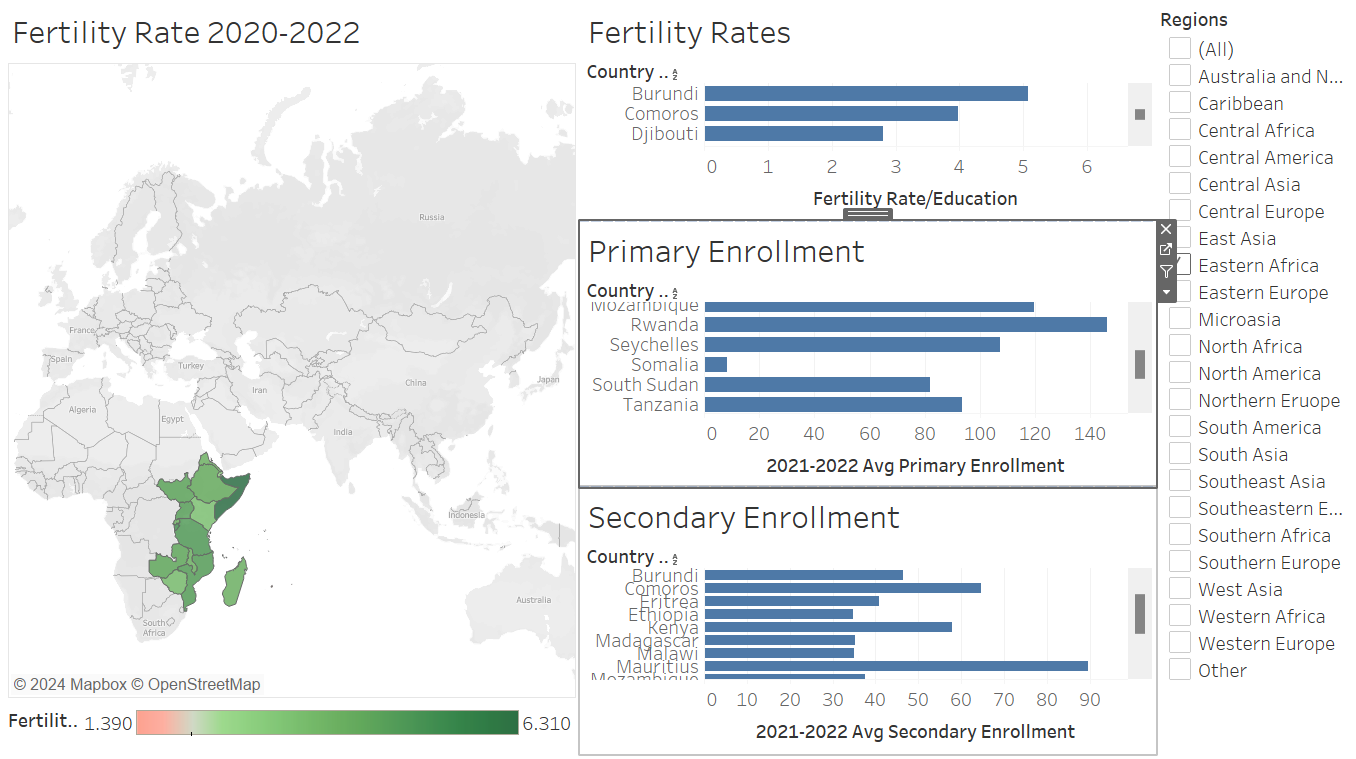
This regions has the highest population growth rate country being Niger at 3.716 and it’s lowest being .918.

It has both a strong increase in urban and rural population growth as well indicating it is more developing then developed of a nation.

**Education and Fertility Dash:**

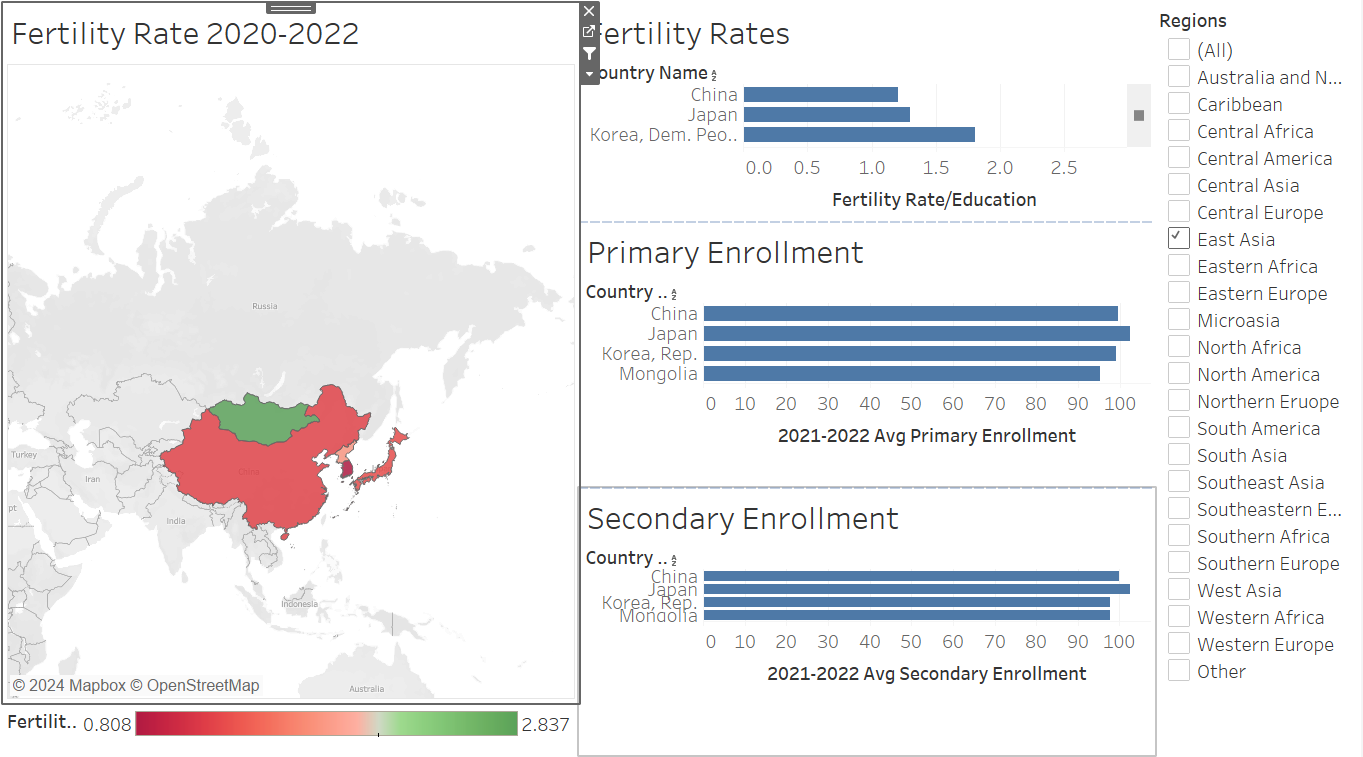
Fertility rates of a country are defined as children per woman of child bearing age.

The same grouping has been applied to the map as the grouping done in the urban and rural analysis. .

Comparing both fertility rate and education rate of a county we can see as countries have higher secondary school enrollment rates fertility rate drops. 

The most fertile region i the globe is east Africa. Specifically Somalia which has the highest fertility rate of 6.31. Somalia also has the lowest primary enrollment rate of 8.4 and an even lower secondary enrollment rate of 5.4.

Now compare that with the lowest fertility rate and highest education rate region East Asia.



The lowest fertility rate country is South Korea with a rate of .808. The education rate in that country though is near 100 with it being 98% for primary education and 97% for secondary.

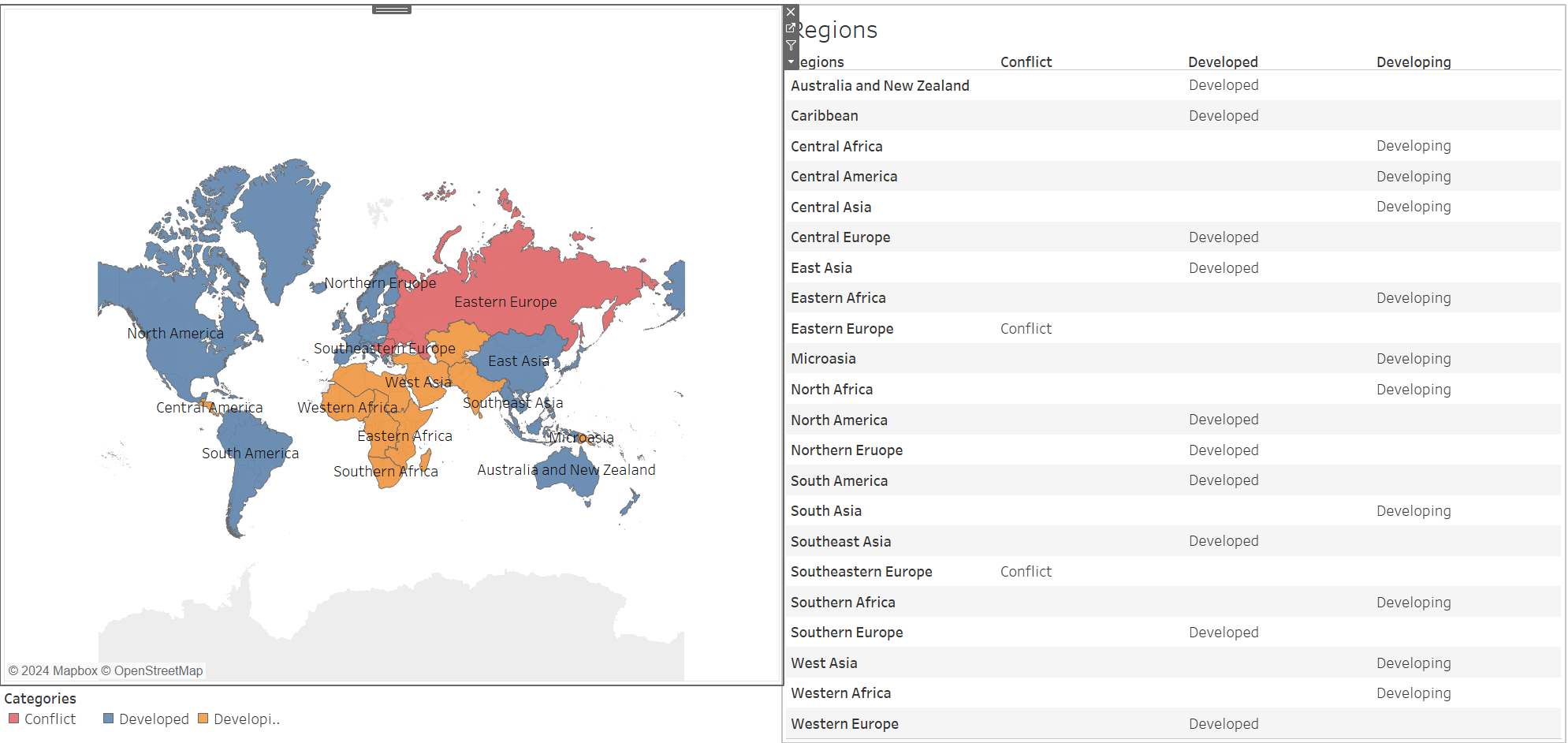
Overall the east asian region has a low fertility rate and high education rate with Mongolia being an outlier of the region possessing an fertility rate of 2.837 and an education rate of 95 primary and 97 secondary enrollment.

**Ranking:**

What we have learned about the regions in the world in the above dashboards is a general categorization they can be placed in. The whole of Eastern Europe is in a population decline, the whole of West Africa is seeing a population boom. This gives opportunity to classify these regions into 3 distinct categories:

1. Developed:
   * Countries which have population growth
   * Countries with a greater increase in urban centers over rural
   * Countries with >=97 gross enrollment for both primary and secondary schools averaged
   * Fertility rate is <=2.1
2. Developing:
   * Countries which have population growth
   * Countries with a increase in both urban and rural population
   * Countries with < 97 gross enrollment for both both primary and secondary schools averaged
   * Fertility rate is above 2.1
3. Conflict:
   * Countries which have population shrinkage
   * Fertility rate is below 2.1

We can visualize this to provide a better overview of global regions and how they can be classified.



As you can see above, by applying a categorization function to the urbanization, rural, population growth, fertility, and education enrollment factors we can get a better idea as to which regions can be better targets for market expansions or possibility outreach programs to send aid.

To summarize the findings there are two main conflict regions: Eastern and Southeastern Europe.

Ten developing regions.

And ten developed regions.

Contributions

Niki Bakshi:

* Proposed a dataset and provided questions
* Participated in vote on current dataset
* Wrote the Project Proposal Part 1
* Created the map expressing average fertility rates
* Created the visualization showing the fertility rates of the countries with the highest and lowest fertility rates
* Attended all group meetings (in person and Zoom)
* Wrote parts 1, 2, and 3 of Project Proposal Part 2

Chuyu Hu:

* Proposed a dataset and provided questions
* Participated in vote on current dataset
* Created visualization showing countries with the highest and lowest populations
* Created visualization of the map showing rural population change
* Attended all group meetings (in person and Zoom)
* Edited parts 1, 2, and 3 of Project Proposal Part 2

Preston Simpson:

* Proposed a dataset and provided questions
* Participated in vote on current dataset
* Cleaned current data set
* Created the visualization on Urbanization map
* Created visualization of populations of different countries from 2021 to 2023
* Complied the population dashboard
* Created visualization of primary education enrollment
* Created visualization of secondary education enrollment
* Attended all group meetings (in person and Zoom)
* Wrote part 4 of Project Proposal Part 2

References

Ku, Chih-Hao. "Basic Charts." Data Visualizations for Analytics, August 26, 2024,University of North Texas. Microsoft PowerPoint presentation

Ku, Chih-Hao. "Basic Visualizations." Data Visualizations for Analytics, September 16, 2024, University of North Texas. Microsoft PowerPoint presentation

Ku, Chih-Hao. "Connecting to Data." Data Visualizations for Analytics, August 26, 2024, University of North Texas. Microsoft PowerPoint presentation

Munzner, Tamara. Visualization Analysis and Design. A.K. Peters Visualization Series, CRC Press, 2015.

“World Development Indicators.” DataBank, World Bank Group, 19 Sept. 2024, databank.worldbank.org/source/world-development-indicators. Accessed 20 Sept. 2024

"United Nations Geoscheme." *Wikipedia*, Wikimedia Foundation, 6 Dec. 2024, https://en.wikipedia.org/wiki/United\_Nations\_geoscheme.