

**Name: Jiahao Li**

### **MyColab Link**

<https://colab.research.google.com/drive/1r3Ejsd2AzeAyCaXNhfQmbZw3-zki6RiY?usp=sharing>

### **Introduction**

In this data visualization project, I will visualize the table “Trends in international migrant stock: the 2015 Revision” using several python visualization libraries. In the previous project, I had already cleaned and wrangled the data according to the tidy data principles. In this assignment, I used tables cleaned in the last project. The main goal of this project is to visualize the cleaned/wrangled tables in multiple charts, using different visualization techniques to present the data from various perspectives.

The overall data set includes data for 265 regional countries (Major area, region, country, or area of destination), and the data are available every five years (mid of each year) from 1990 to 2015. This data set consists a total of six main tables (from table1 to table6), which are “International migrant stock at mid-year by sex and by major area, region, country or area, 1990-2015”, “Total population at mid-year by sex and by major area, region, country or area, 1990-2015 (thousands)”, “International migrant stock as a percentage of the total population, 1990-2015”, “Female migrants as a percentage of the international migrant stock by major area, region, country or area, 1990-2015”, “Annual rate of change of the migrant stock by sex and by major area, region, country or area, 1990-2015 (percentage)”, “Estimated refugee stock at mid-year by major area, region, country or area, 1990-2015”.

In the process of visualizing data, I use a variety of charts to visualize the content and characteristics of the data, including distribution plots, scatter plots, line plots, boxplot, bar charts, and violin plots. In the next part of this report, I will discuss in detail the name of each graph, its content, and the methods and techniques used in visualization.

### **Methods**

In this section, I will elaborate on the methodology and techniques I was involved in this visualization project and my logic and process. In the visualization process, I first learned about the roles and characteristics of the various charts, as different charts serve different purposes and convey other points. Secondly, I went through all the data, whether it was the

data from the original Excel file or data that I had cleaned up/sorted, because I believe that only by understanding the content of the data can I choose the appropriate chart. Third, I followed Tufte's Principles in creating each chart, using different charts to present and explain the data as comprehensively as possible from different perspectives. In the remainder of this report, I will introduce each chart, describing in general terms the content of the chart, my reasons and motivations for designing such a chart, and explaining in general terms the results and patterns shown on the chart.

## **Results**

### List of Figures

Figure 1.1 Compare the international migrant of major area (both sexes)

Figure 1.2 Population of International migrant in North America in 2005 - Boxplot

Figure 2.1 Total Population in East Asia Countries from 1990-2015

Figure 2.2 Violin Plot for total population of Eastern Africa Countries from 1990-2015

Figure 3.1 Compare the migrant stock percentage of middle Africa countries (both sexes)

Figure 3.2 Migrant percentage of the total population in Middle Africa Countries

Figure 4.1 Compare the migrant stock percentage of Middle Africa Countries (both sexes)

Figure 5.1 Annual rate of change the migrant stock Asian countries from 2010 to 2015 (both sexes)

Figure 6.1 Estimated refugee stock at midyear in Southern Africa

Figure 6.2 Refugees as a percentage of the international migrant stock in Eastern Europe Countries

Figure 6.3 Annual rate of change the refugee in Major Areas

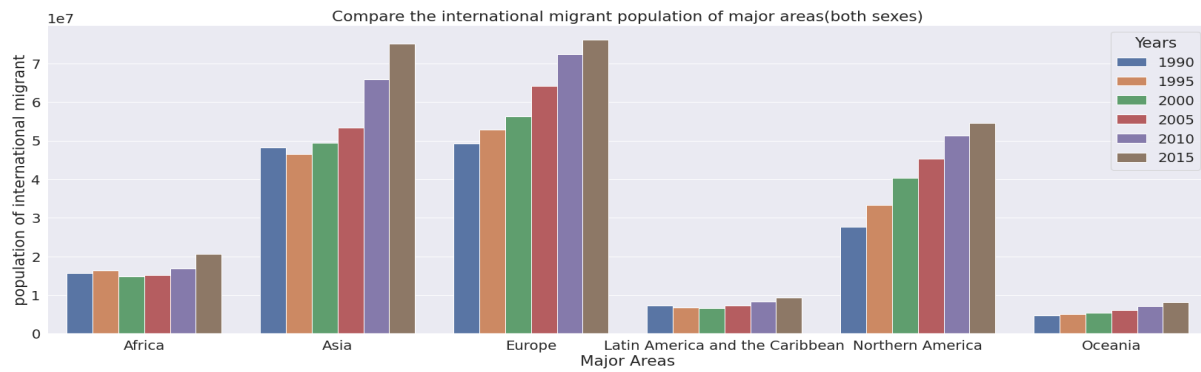


Figure 1.1 Compare the international migrant of major area (both sexes)

In Figure 1.1, I compare international migrant populations in six major regions of the world. To compare changes in trends in the total population I chose the total population as the data (i.e., the data that count the sum of the two genders). Barcharts are used to compare and contrast different types of data, frequencies, or other measures of distinct categories of data. Because I want to compare overall migration trends from 1990 to 2015 for each major area, I used a bar chart as the graph, with the x-axis representing the six continents. The y-axis represents the total number of international migrants, and the different colors distinguish the different years. It is clear to see that the international migrant population is on the rise in all major regions, with Europe and Asia having the highest number of international migrants.

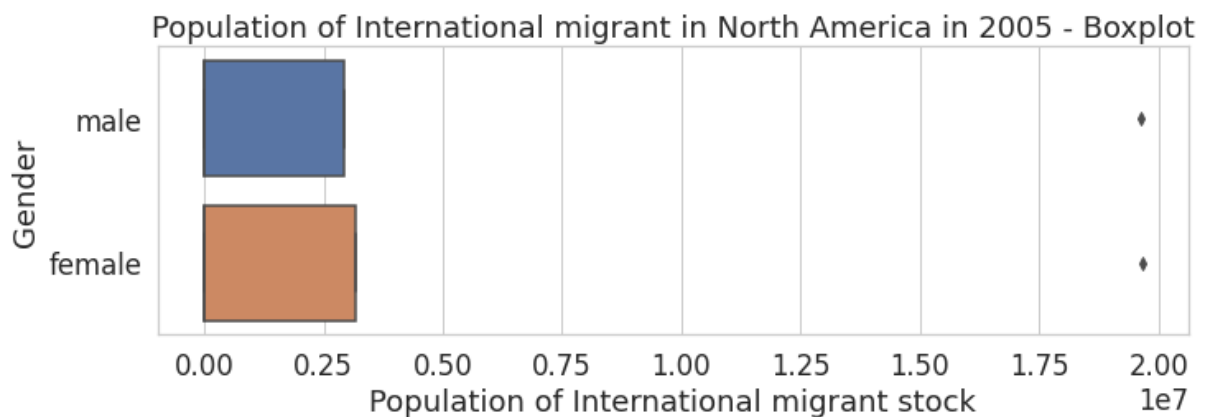


Figure 1.2 Population of International migrant in North America in 2005 - Boxplot

In Figure 1.2, I use box plots to visualize the total number of international migrant populations in four countries/regions of North America (Bermuda, Canada, Greenland, Saint

Pierre and Miquelon, United States of America). Box plots visually show the distribution of numerical data and skewness by displaying the data quartiles (or percentiles) and averages. The plot's x-axis represents the total international immigrant population, and the y-axis represents the genders (male and female). The data I used was the number of immigrants in the four countries/regions of North America in 2015.

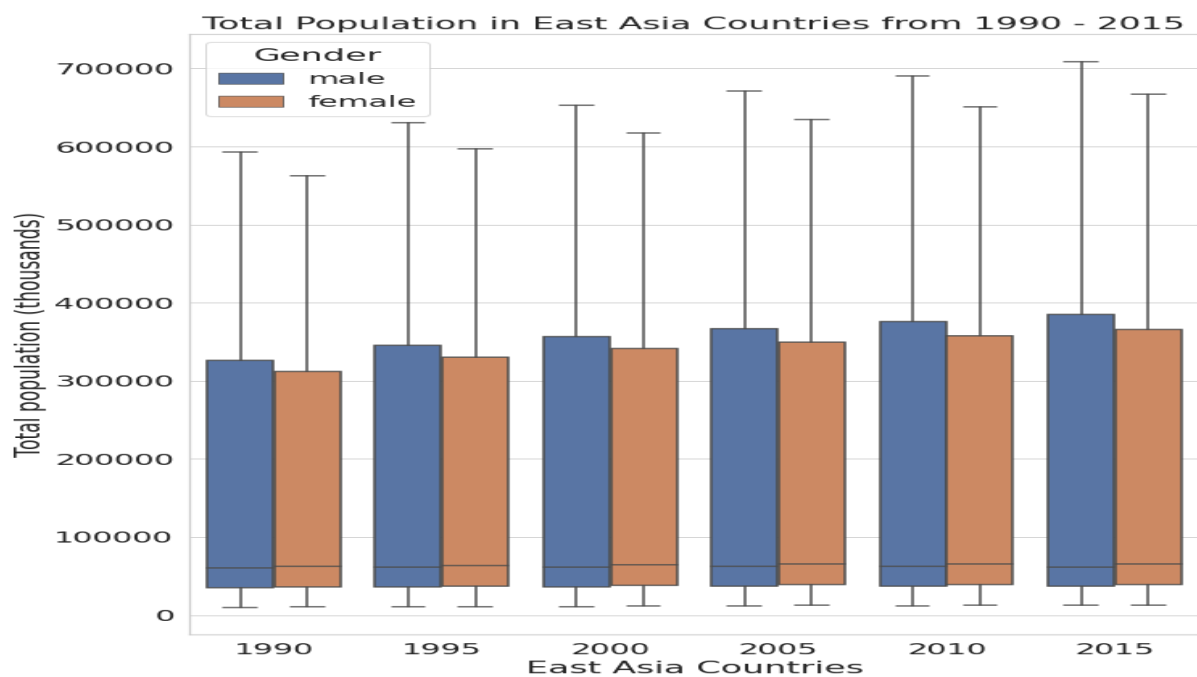


Figure 2.1 Total Population in East Asia Countries from 1990-2015

In Figure 2.1, I use box plots to visualize the total number of international migrant populations in East Asian countries/regions (China, Hong Kong Special Administrative Region China, Macao Special Administrative Region, Democratic People's Republic of Korea, Japan, and Mongolia Republic of Korea). Box plots visually show the distribution of numerical data and skewness by displaying the data quartiles (or percentiles) and averages. The x-axis in the plot represents the year, the y-axis represents the total number of the international migrant population, and the different colours indicate the different genders.

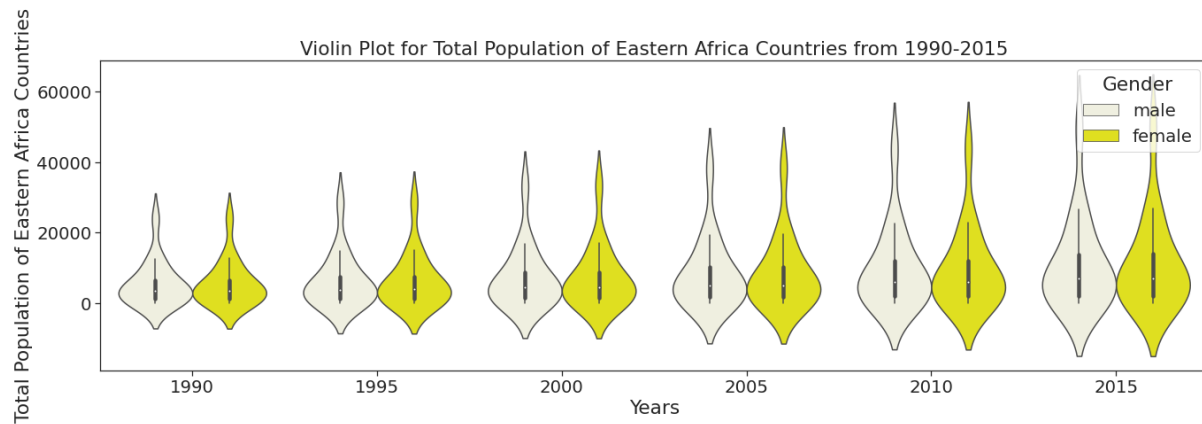


Figure 2.2 Violin Plot for total population of Eastern Africa Countries from 1990-2015

Figure 2.2 , in which I use a violin plot to show the total population of East African countries because a violin plot is a method of plotting numeric data. It is similar to a box plot, with the addition of a rotated kernel density plot on each side. In this plot, I set the year as the horizontal axis (x-axis) and the vertical axis (y-axis) to show the total population, with different colours to distinguish the gender. Similar to the box plot, the violin plot shows the normal distribution of a data set and the mean/median, which are important indicators used in data analysis. On a macro level, this graph illustrates the total population of East African countries increasing from 1990 to 2015.

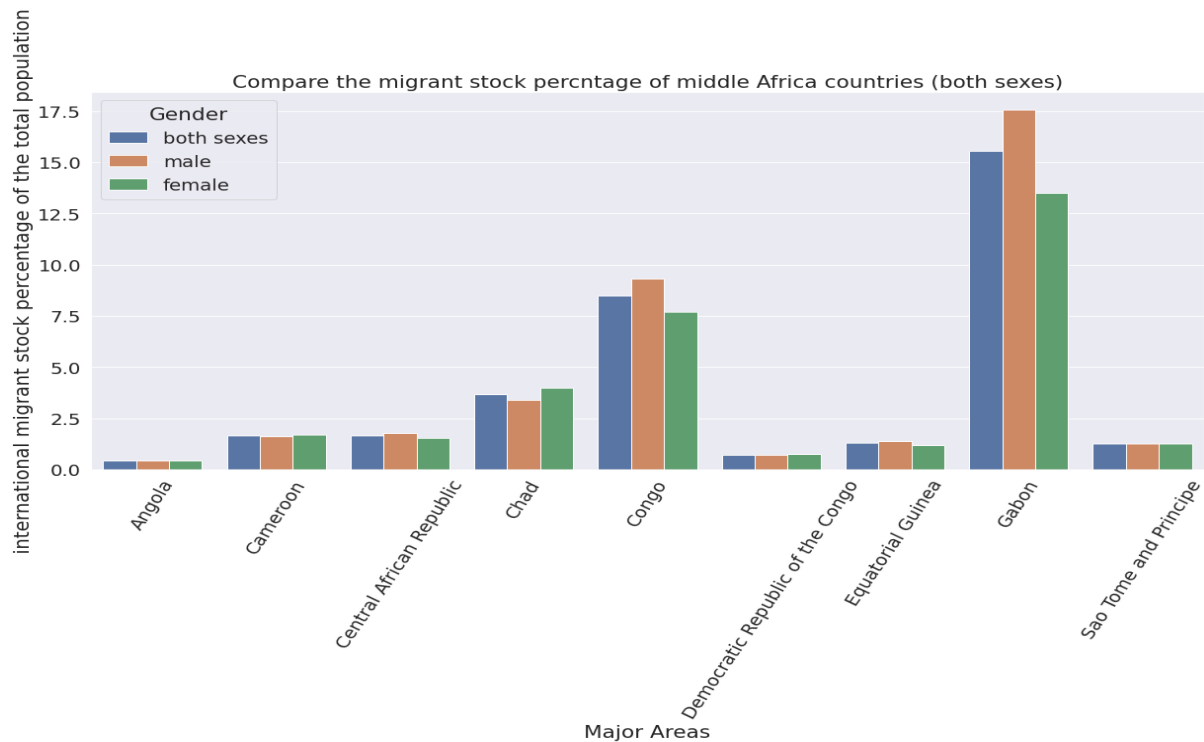


Figure 3.1 Compare the migrant stock percentage of middle Africa countries in 2015 (both sexes)

In Figure 3.1, I show the percentage of the international migrant population in Central African countries as a percentage of the total population as of 2015. I think the bar chart can better illustrate the percentage of the total population so that it can more visually show the audience which gender or region has the most significant or most minor percentage of the international migrant population. For example, this chart shows that Gabon has the highest percentage of the total global migrant population. In this graph, I set the x-axis to nine Central African countries to compare the differences and relationships between the countries; the y-axis I set is the proportion, which is the percentage of the total migrant population to the total population.

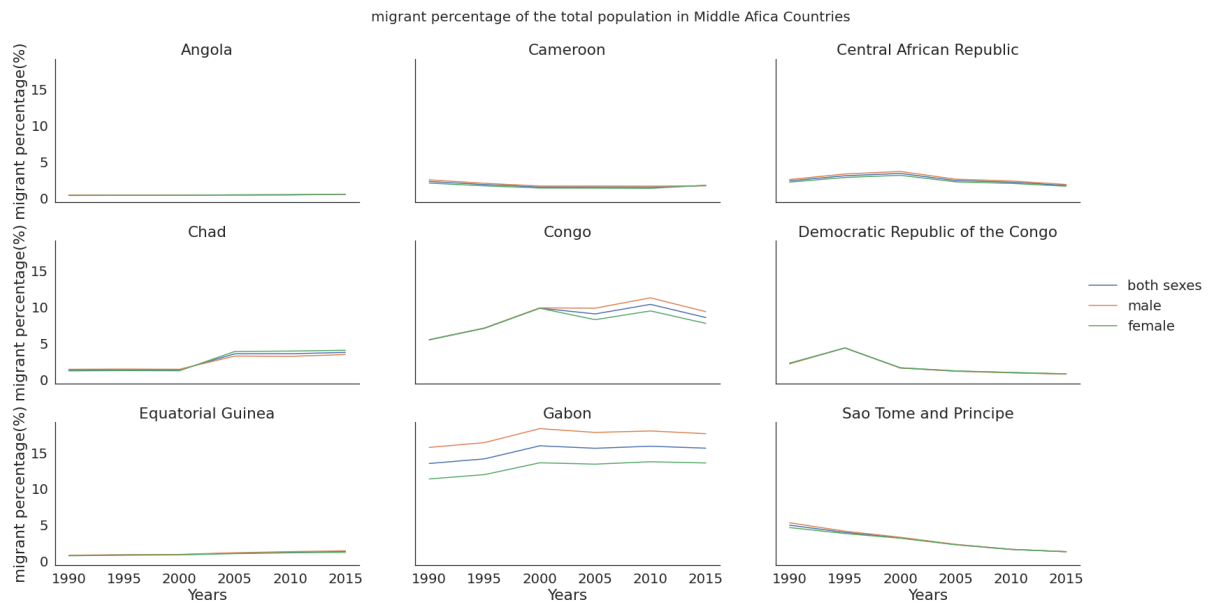


Figure 3.2 Migrant percentage of the total population in Middle Africa Countries

Figure 3.2 and Figure 3.1 use the same set of data, i.e., the proportion of the international migrant population to the total population in nine countries in Central Africa as of 2015. I use the Seaborn Facetgrid nine subplots, i.e., each subplot represents a country. The x-axis represents the year, and we can see the trend of the migrant population to the total population from 1990 to 2015. The y-axis represents the percentage of the migrant population. The reason I am using line graphs is to see more clearly the change in trend over the years.

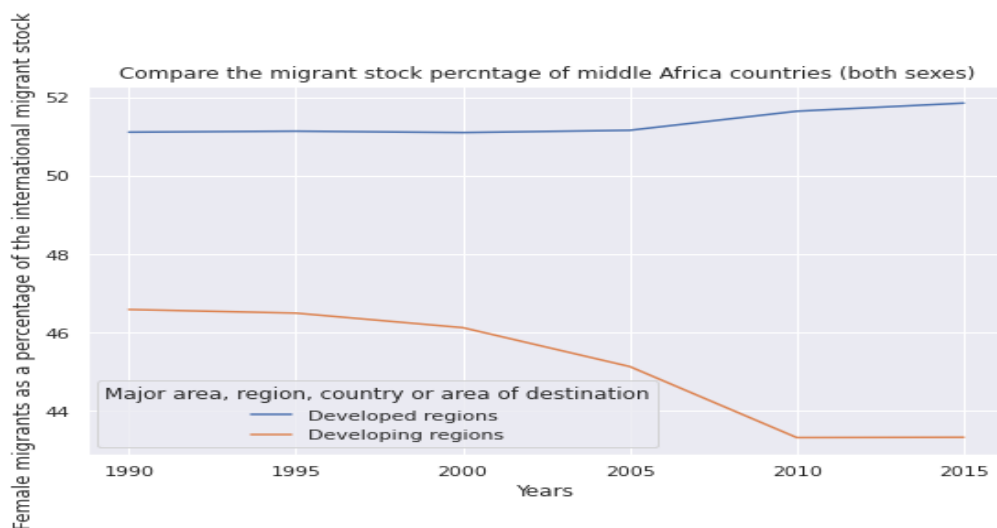


Figure 4.1 Compare the migrant stock percentage of Middle Africa Countries (both sexes)

In Figure 4.1, I used line graphs to represent the proportion of female international migrants to the total population from 1990 to 2015. I think the line graphs show the trend because I set the horizontal axis (x-axis) to the year so that I can clearly see the trend since those years. I set the y-axis to be the value of the rate of change, using two colours to indicate developed and developing regions respectively.

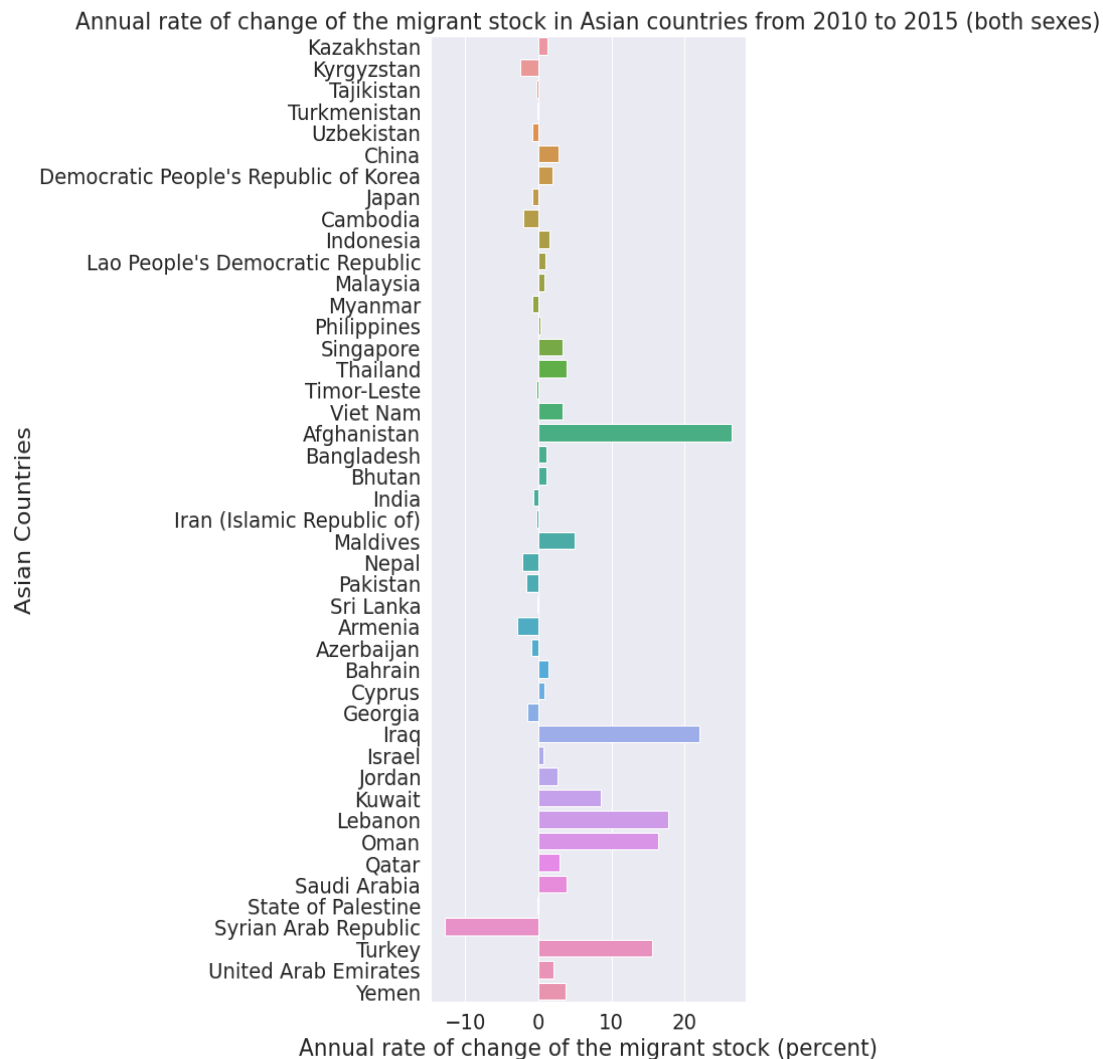


Figure 5.1 Annual rate of change the migrant stock Asian countries from 2010 to 2015 (both sexes)

In Figure 5.1, I used a bar chart to show the rate of change in the international migrant population in Asian countries between 2010 and 2015. Because it is a rate of change, it has the potential to increase or decrease. I set the change rate on the x-axis to show the increase and decrease of the rate of change. With shading to the left representing a decrease and shading to the right representing an increase, it is very straightforward to see the trend of



change and visually read the value of the difference. I set the vertical coordinates to countries, and the rate of change for all Asian countries is counted in this figure. From the graph, we can see that from 2010 to 2015, the number of international migrants in Afghanistan changed significantly, with a growth rate of nearly thirty percent in just five years.

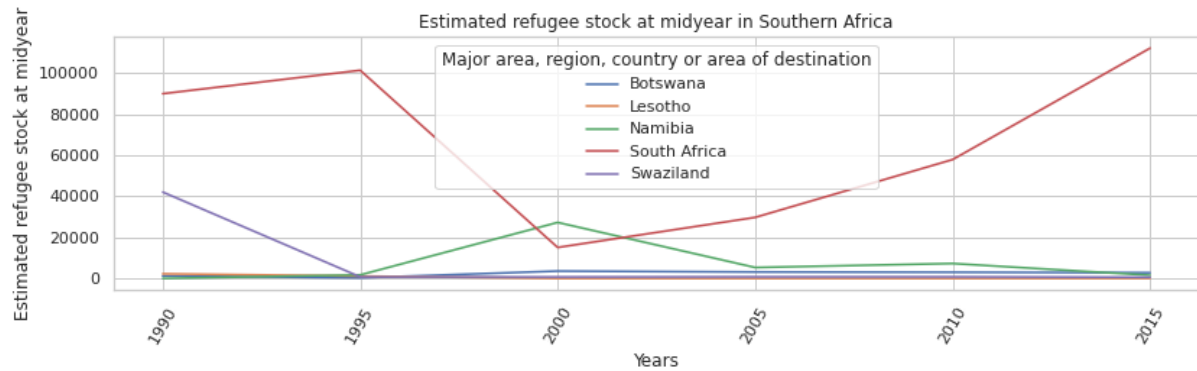


Figure 6.1 Estimated refugee stock at midyear in Southern Africa

The data in Figure 6.1 are for the international refugee population. Similarly, I set the x-axis to time from 1990 to 2015. For the y-axis, I put the estimated population of refugees. For this graph, I wanted to represent the trend of the refugee population in the countries of southern Africa over these years, so I chose to use a line graph. A line chart is a type of chart used to show information over time. As can be seen in this line chart, the refugee population in South Africa has the most extensive ups and downs compared to other countries, and the trend in other countries is relatively flat.

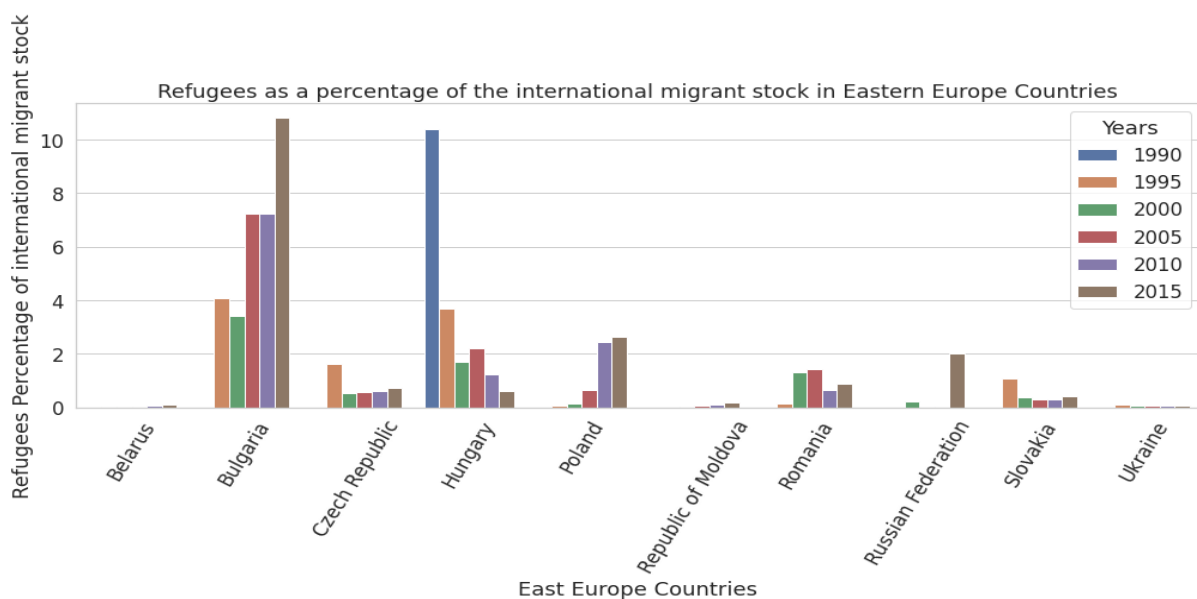


Figure 6.2 Refugees as a percentage of the international migrant stock in Eastern Europe Countries

In Figure 6.2, I compared refugees as a percentage of the total population in ten Eastern European countries. To reach the differences between countries, I chose the entire population as the sample (i.e., the data for which the sum of the two genders). Bar charts are used to compare and contrast different data types, frequencies, or other measures of various categories of data. Because I wanted to compare trends in the overall refugee population in Eastern European countries from 1990 to 2015, I used bar charts and set the x-axis to ten Eastern European countries. Y-axis represents the proportion of refugees to the country's total population, with different colors distinguishing the years. Overall, the percentage of refugees in Bulgaria has been higher than in other Eastern European countries and has been upward from 1990 to 2015.

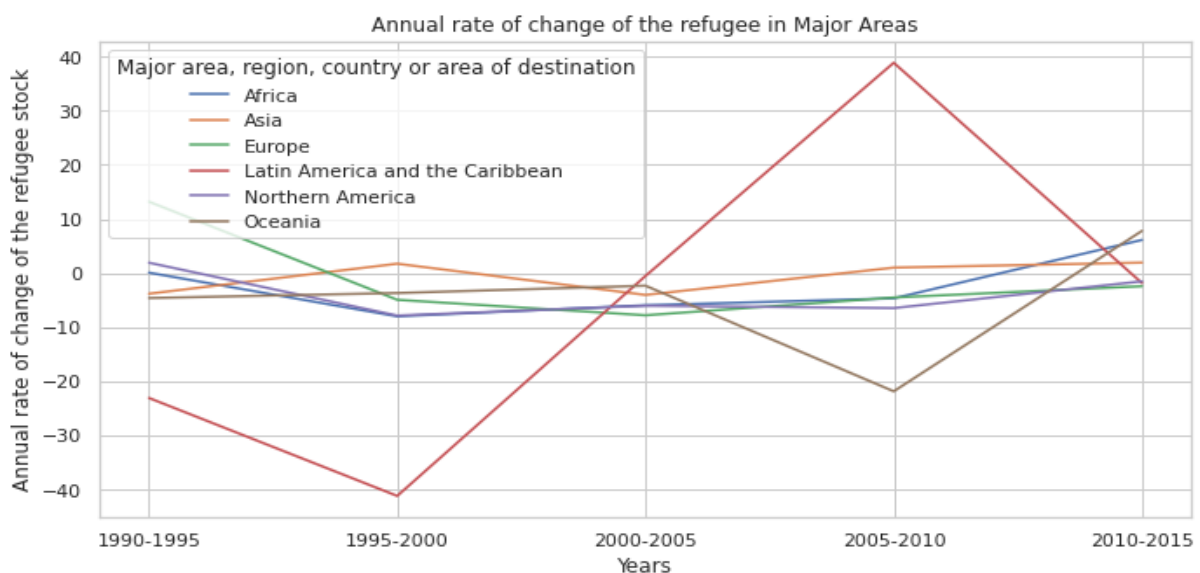


Figure 6.3 Annual rate of change the refugee in Major Areas

In Figure 6.3, I applied the line chart to represent the annual rate of change of the refugee population from 1990 to 2015; the line chart can show the trend of the rate of change. In this graph, I set the horizontal axis (x-axis) to the year and the vertical axis (y-axis) to the rate of change, with different colors, to distinguish the data from other continents.

## Discussion

In this data visualization project, our data focuses on migration, refugees, and total population counts. The data is also broken down by region and gender. Trends in the migration of

refugees are also counted (i.e., annual rates of change). The data provided me with three available variables, country/region, year, and gender. Based on these three general factors, I can visualize and analyze the same data from multiple perspectives. For example, we can compare male and female immigrants in a particular country or even explore the trend of changes in a specific year. For this project, I used various charts based on analysis objectives, such as line charts for trends and bar graphs for comparing population changes and numbers between time (years). Based on this population-based data set, I have used line charts, bar charts, box plots, and violin plots to visualize these data. In this project, since the data set involved more than 250 countries and regions worldwide, there was no way to visualize each area separately, so I had to affect as many different areas and countries as possible. For the use and design of charts, I also tried to experiment with different types and forms of charts to analyze the data from different angles and dimensions.