Exploratory Data Analysis of the Trends in International Migrant Stock: The 2015 Revision

by

Keying Zhu #1003790953

INF1340H F: Programming for Data Science
Professor: Shion Guha
December 15, 2022

Table of Content

List of Figures

Introduction

This data set is about the trends in international migrant stock provided by the United Nations, Department of Economic and Social Affairs, Population Division. By analysis this data set that about the trends in international migrant stock provided by the United Nations, Department of Economic and Social Affairs, Population Division (2015). This report will conduct analysis on this data set around the questions raised.

Method

By following up the Tufte's visualization principles:

1. Make Representation of Numbers Proportional to Quantities
2. Use Clear, Detailed, Thorough Labeling
3. Show Data Variation, Not Design Variation
4. Use Standardized Units, Not Nominal
5. Depict N Data Dimensions with N Variable Dimensions
6. Quote Data in Full Context

By following up the classic designs:

1. Small multiples
2. Time series
3. Micro/Macro Composition

Questions will explore and analysis:

1. How did the stock of international migrants change from 1990-2015? Did the top five countries with the largest immigrant stocks change between 1990-2015?
2. For the countries mentioned in question 1, is the gender distribution of the international migrant stock equal?
3. For the countries mentioned in question 1, Which country has the most international migrant stock and which country has the most estimated refugee stock?
4. Is the gender distribution same across continents? If not equal, which continents will have the highest female's migrant's percentage? Which one has the lowest female's migrant's percentage?
5. How about the Estimated Refugee Stock across continents? What about the Annual Rate of Change of the Refugee Stock of Continents?

Data Selection：

Since questions 1-3 will focus on the top five countries in 1990 and the top five countries in 2015, these countries (the United States of America, the Russian Federation, India, Ukraine, Pakistan, Germany, Saudi Arabia, and the United Kingdom of Great Britain and Northern Ireland) are referred to as 'top countries' in the analysis. Question 4 and 5 will focus on the continents (Africa, Asia, Europe, Latin America and the Caribbean, Northern America and Oceania) are referred to as 'Continents' in the analysis.

Results

In response to the questions mentioned in method, they will be given results in this section.

3.1 Results for Question 1

How did the stock of international migrants change from 1990-2015? Did the top five countries with the largest migrant stocks change between 1990-2015?

First, select the world's data from the table_1 by Country code (900) in 1990-2015, to show the trend, plotting the line chart of world international migrant stock from 1990 to 2015.

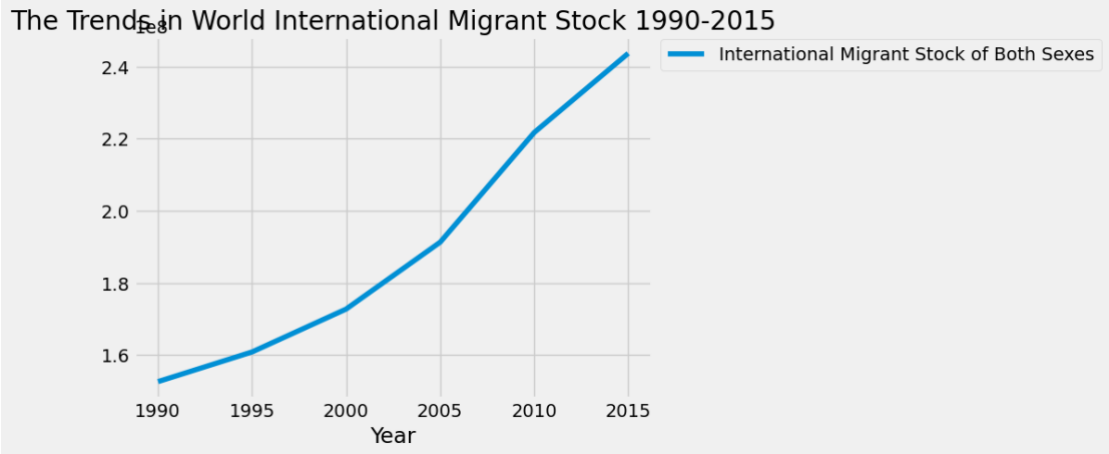

Figure 1 The Trends in World International Migrant Stock 1990-2015

As the Figure 1 shown, the trend of the world international migrant stock is showing an upward trend. Corresponding to world international migrant stock above, the following will show the change of world annual migrant stock percentage in Figure 2.

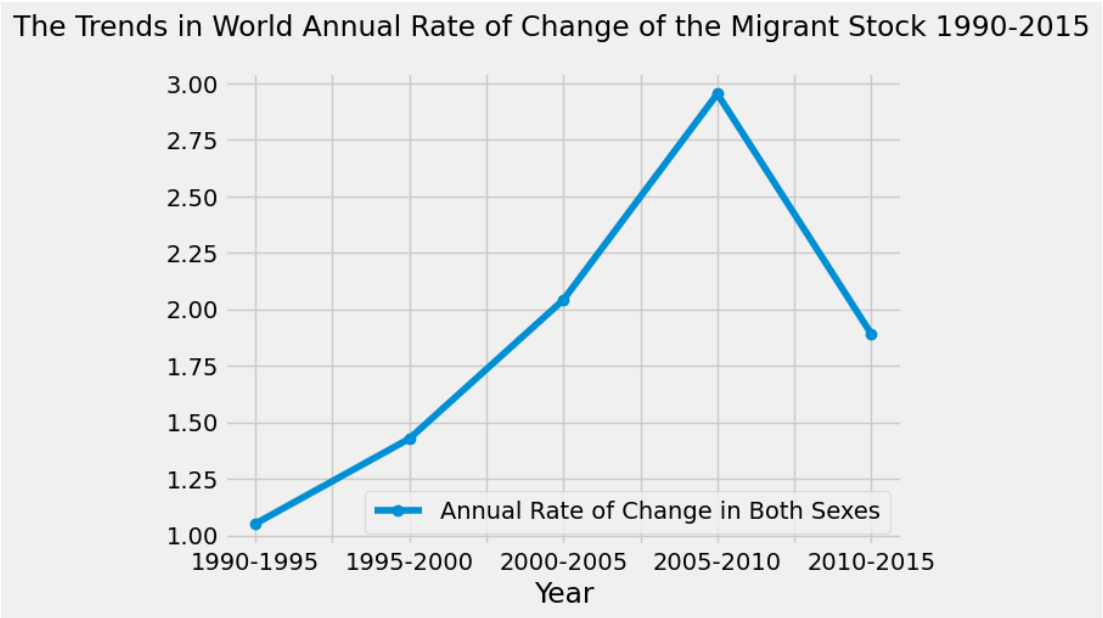

Figure 2 The Trends in World Annual Rate of Change of the Migrant Stock 1990-2015

As Figure 1 and Figure 2 shown, even if the population of world international migrants stock continued to increase in 1990-2015. The rate of change of the world international migrants stock only continued to increase in 1990-2010, and it had peak in 2005-2010, but was decline in 2010-2015.

Select the top five countries with largest international migrant stock in 1990 and plot the bar chart.

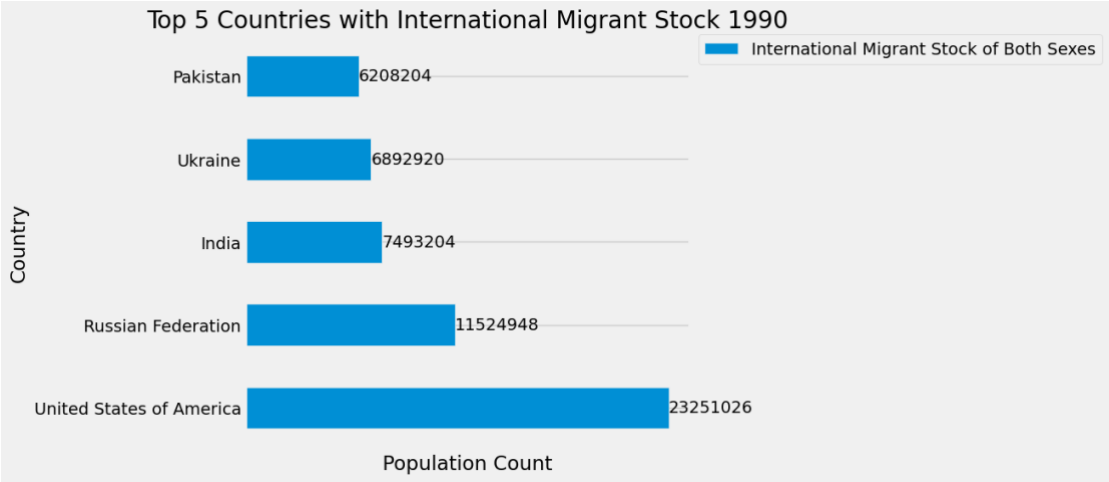

Figure 3 Top 5 Countries of International Migrant Stock in 1990

Select the top five countries with largest international migrant stock in 2015 and plot the bar chart.

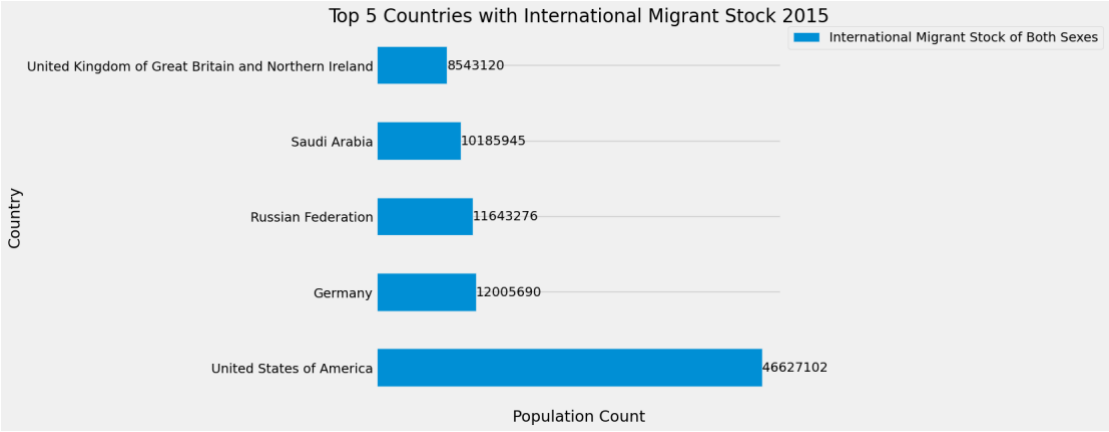

Figure 4 Top 5 Countries with International Migrant Stock 2015

As shown in the figure 4 and figure 5, United of States of America, Russian Federation, India, Ukraine and Pakistan had the largest international migrant stock in 1990; United of States of America, Germany, Russian Federation, Saudi Arabia and United Kingdom of Great Britain and Northern Ireland had the largest international migrant stock in 2015. United of States of America had the highest of international migrant stock in both 1990 and 2015. In both 1990 and 2015 the only countries in the top five international migrant stock countries are the United States of America and Russian Federation.

3.2 Results for Question 2

For the countries mentioned in question 1, is the gender distribution of the international migrant stock equal?

The eight countries mentioned in question 1 are the United States of America, the Russian Federation, India, Ukraine, Pakistan, Germany, Saudi Arabia, and the United Kingdom of Great Britain and Northern Ireland. These eight countries are collectively referred to as "Top Country" in the following analysis. In order to explore the gender distribution of each country, we need to use the small multiple principle to draw 8 separate line charts.

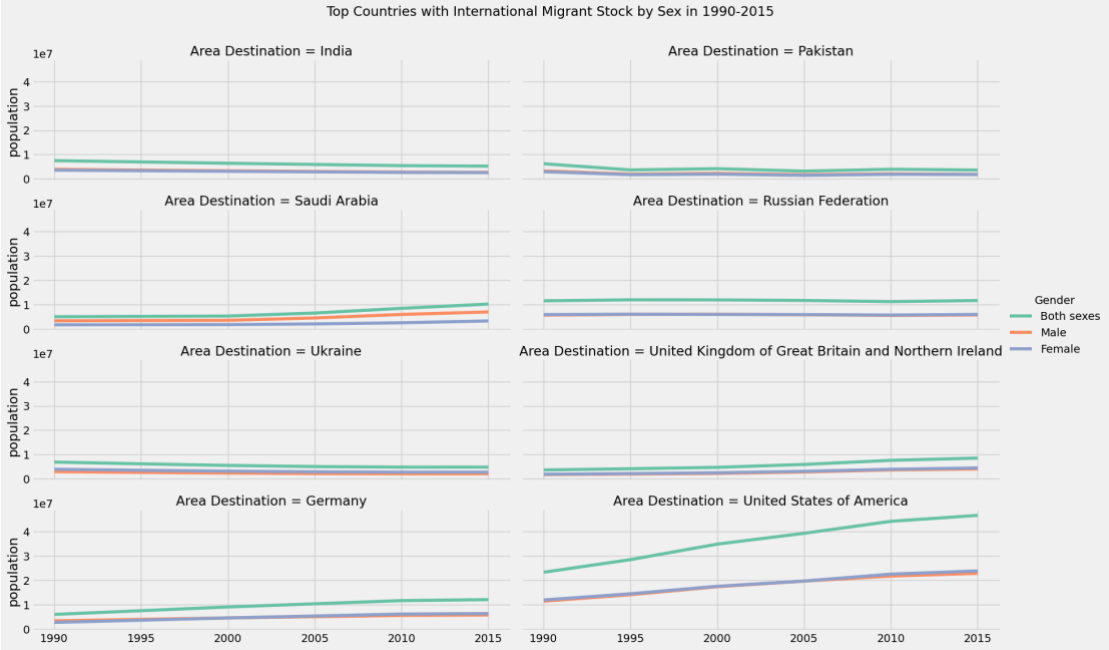

Figure 5 Top Countries with International Migrant Stock by Sex in 1990-2015 (Separate)

As Figure 5 shown, India, Pakistan and Russian Federation have a fairly equal gender distribution; Ukraine, Germany and United States of America have a relatively equal gender distribution; Saudi Arabia males are much higher than females, there is an un equal gender distribution. It can also reflect that the stock of international migrants in the United States far exceeds that of other countries.

3.3 Results for Question 3

For the countries mentioned in question 1, which country has the most international migrant stock and which country has the most estimated refugee stock?

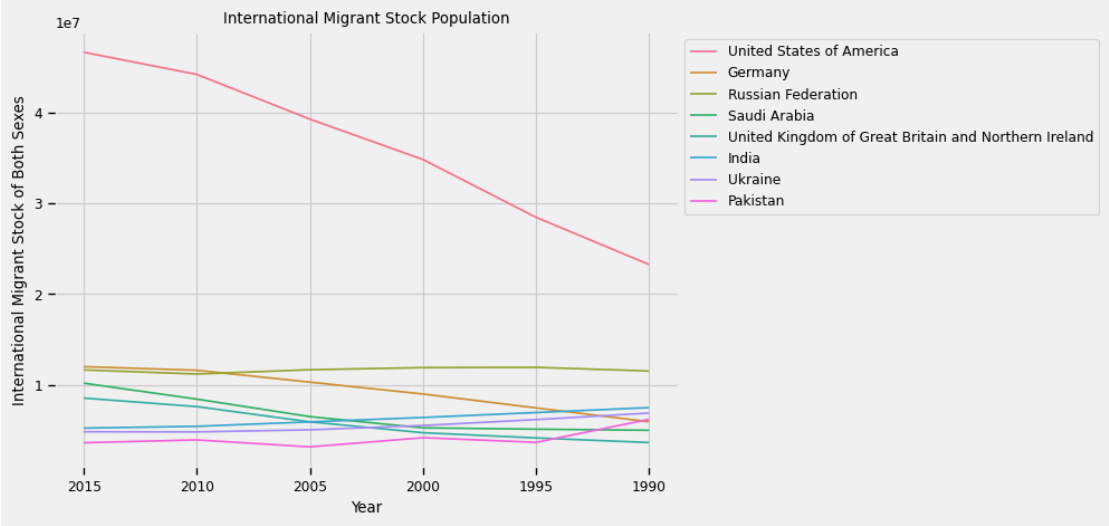

Figure 6 Top Countries with International Migrant Stock 1990-2015

Figure 6 shows the United States of America is the country with the most international migrant stock. In order to more intuitively show how much international migrant stock is in the United States, a histogram comparing the number of international migrant stock in the United States of America with international migrant stock in the world is shown in Figure 7.

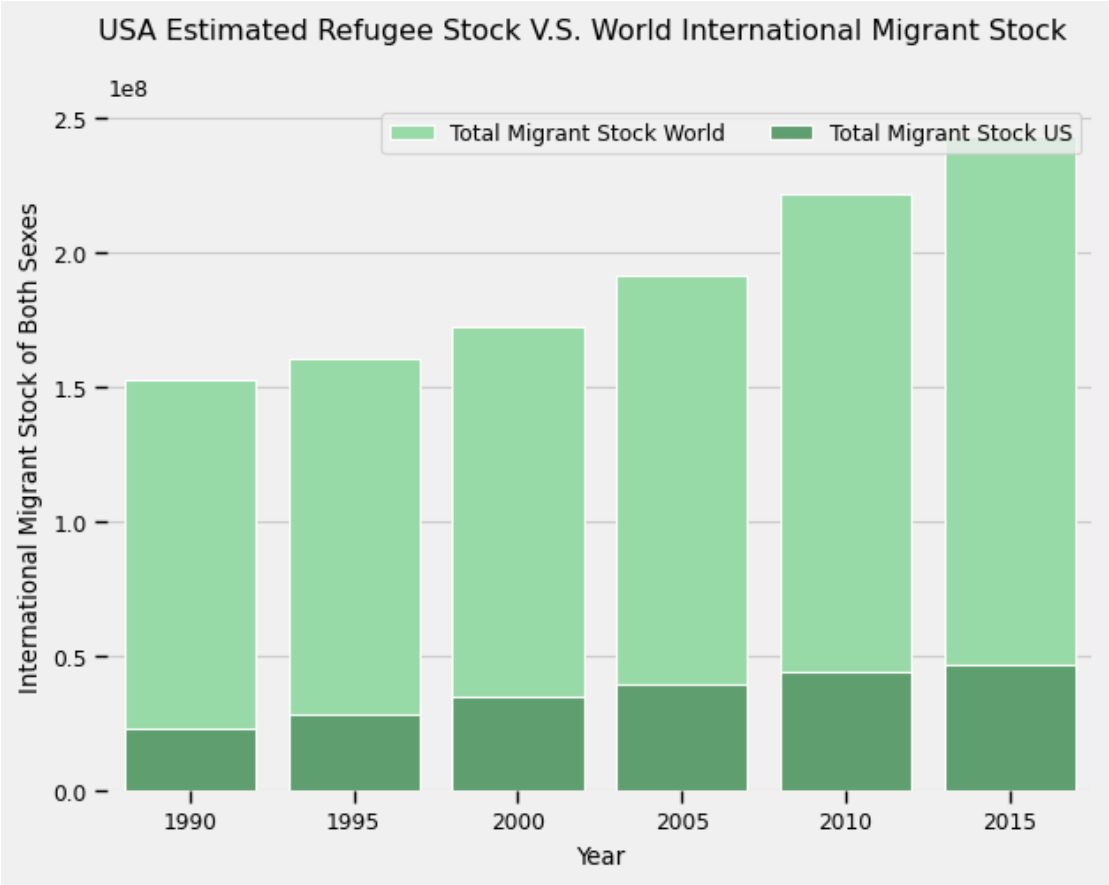

Figure 7 USA International Migrant Stock V.S. World International Migrant Stock

As Figure 7 implies, from 1990 to 2015, the United States of America contributed a significant portion of the world's international migrant stock.

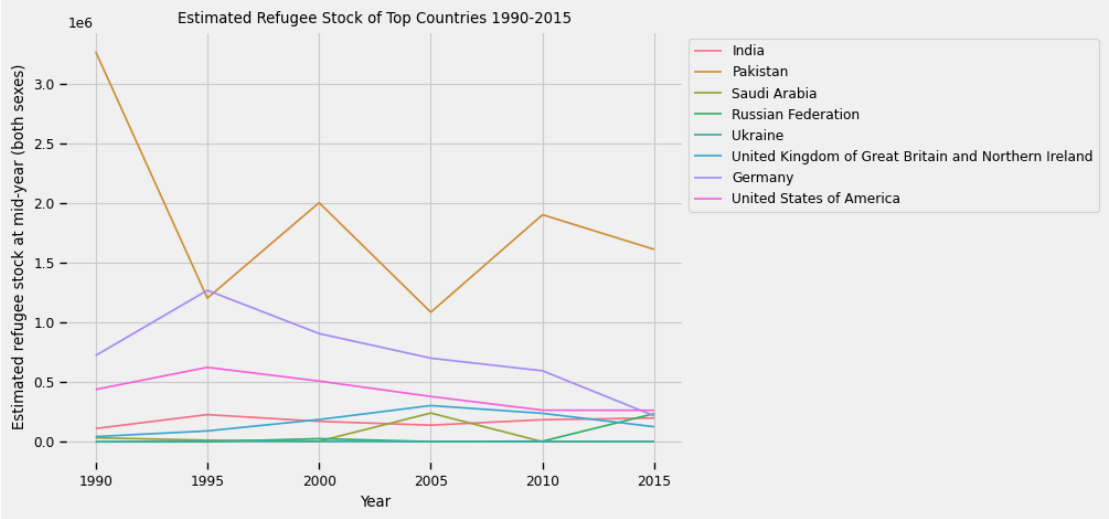

Figure 8 Estimated Refugee Stock of Top Countries 1990-2015

Figure 8 shows, Pakistan is the country with the most estimated refugee stock, except for the 1995 slightly lower than Germany. In order to more intuitively show how much estimated refugee stock is in Pakistan, a histogram comparing the number of estimated refugee stock of Pakistan with its international migrant stock is shown in Figure 9.

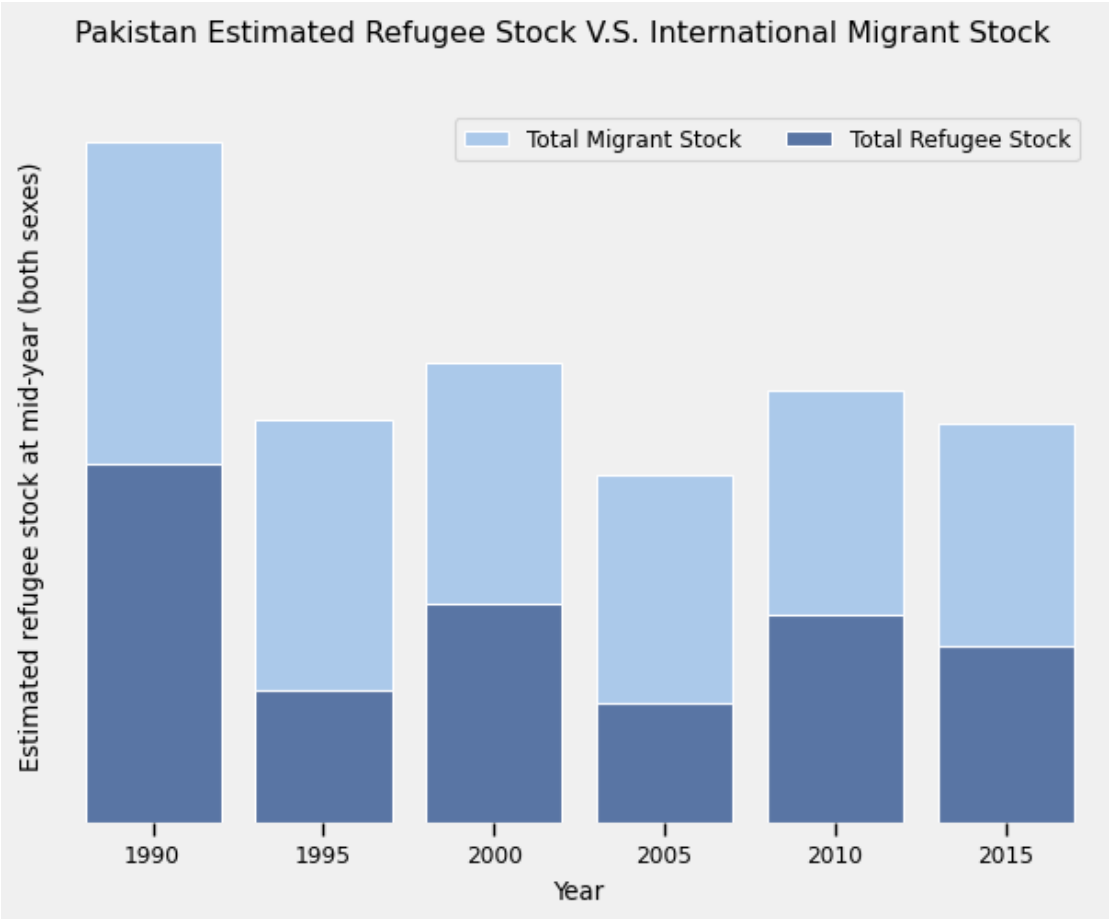

Figure 9 Pakistan Estimated Refugee Stock V.S. International Migrant Stock 1990-2015

As Figure 9 implies, from 1990 to 2015, the estimated refugee stock of Pakistan contributed almost half portion of its international migrant stock. This proportion was largest in 1990 and fluctuated in the 1995-2015.

3.4 Results for Question 4

Is the gender distribution same across continents? If not equal, which continents will have the highest female's migrant's percentage? Which one has the lowest female's migrant's percentage?

In this question, the data in table_4: Female migrants as a percentage of the international migrant stock by major area, region, country or area, 1990-2015 will be used, first filter out the data of each continent (Africa, Asia, Europe, Latin America and the Caribbean, Northern America and Oceania), and then generate plots.

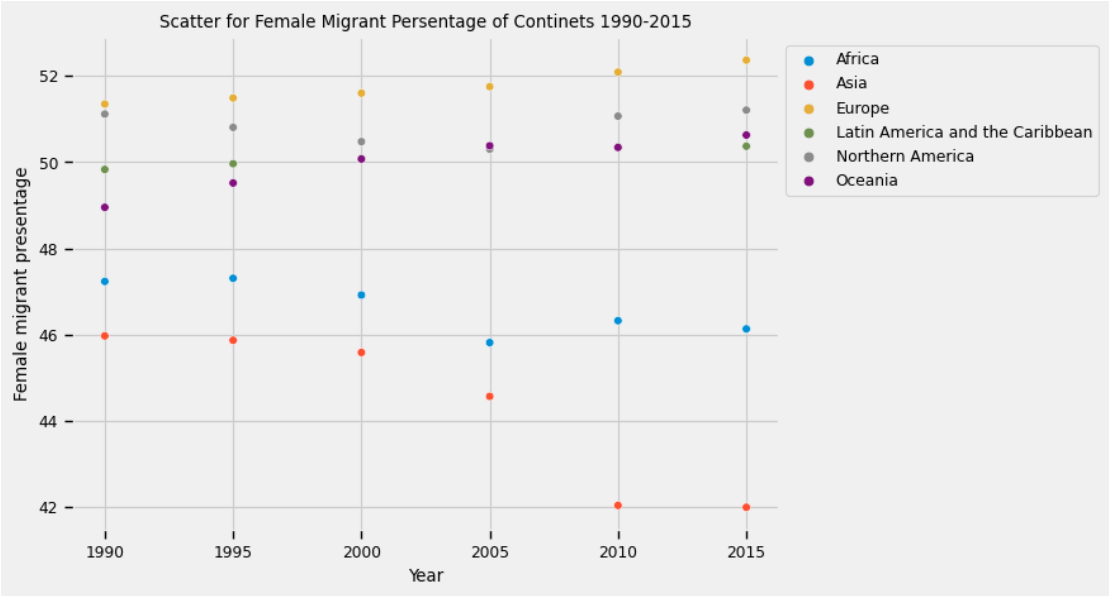

Figure 10 Scatter for Female Migrant Percentage of Continents 1990-2015

Figure 10 is a scatter plot for female migrant percentage of continents 1990-2015. As the figure shows, there is an unequal distribution across continents. Significant differences exist between continents, female migrants' percentage of Europe and Northern America have remained above 50% from 1990 to 2015. The percentage of female migrant in Oceania continued to increase between 1990 and 2015, reaching more than 50% in 2000. The percentage of female migrant in Africa and Asia continued to decrease between 1990 and 2015 and never reach 50%.

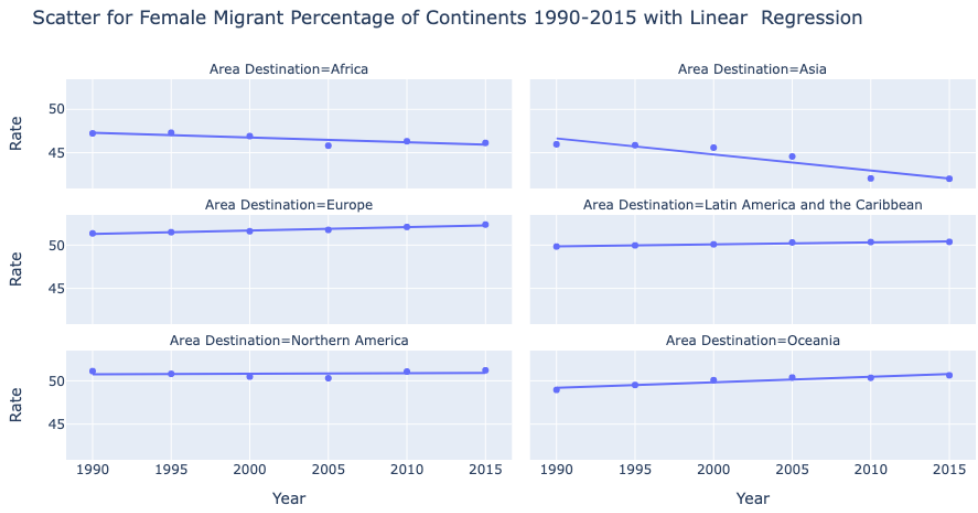

Figure 11 Scatter for Female Migrant Percentage of Continents 1990-2015 with Linear Regression (Separate)

Following the principle of small multiples, we use the assistance of linear regression to look at the percentage of female migrant change of each continent separately. As the trend line shows, the percentage of female migrant in Europe, Latin America and the Caribbean, and Oceania would increase; the percentage of female migrant in Asia and Africa would decrease; and the percentage of female migrant in Northern America would slightly increase.

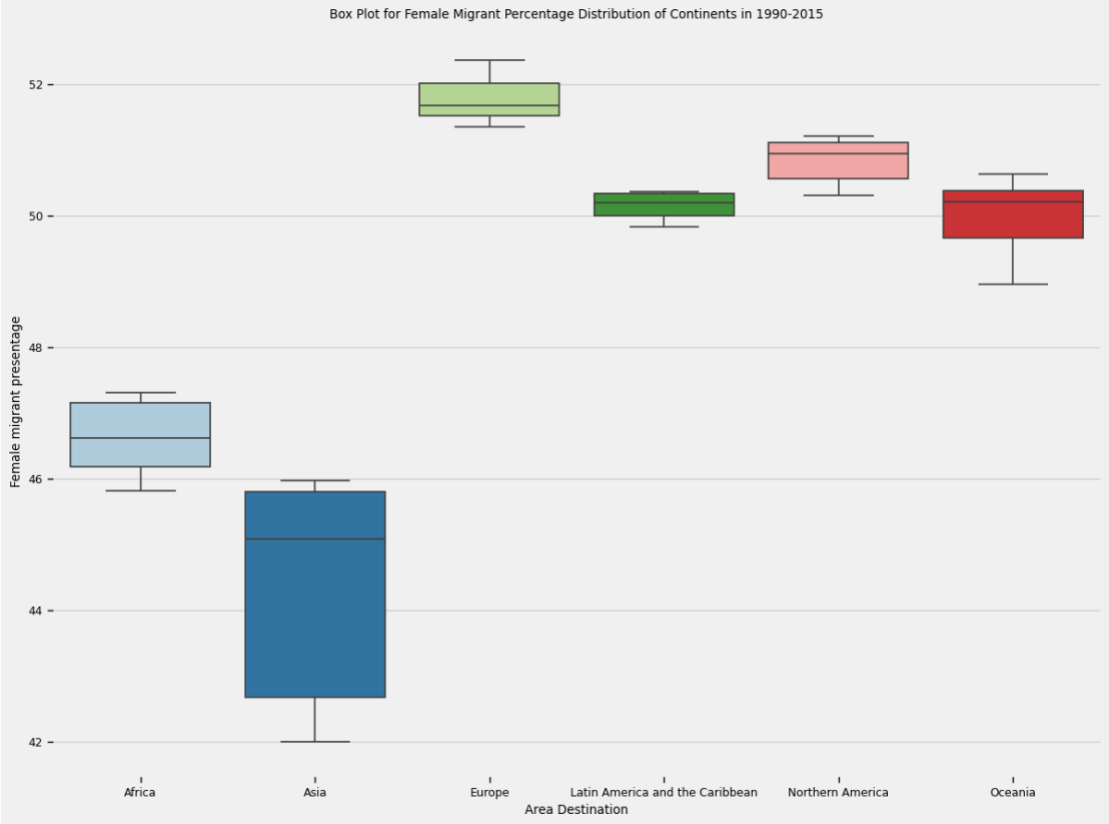

Figure 12 Box Plot for Female Migrant Percentage Distribution of Continents in 1990-2015

As figure 12 shows the box plot for female migrant percentage of continents in 1990-2015.X-axis is continents, y-axis is the percentage of female migrant. Draw a box plot

for the 1990-2015 data of each continent, which can reflect the data distribution of each continent. The female migrant percentage of Europe, Latin America and the Caribbean, Northern America and Oceania are stable; the female migrant percentage of Africa is relatively stable, and the female migrant percentage of Asia is extremely volatile.

3.5 Results for Question 5

How about the Estimated Refugee Stock across continents? What about the Annual Rate of Change of the Refugee Stock of Continents?

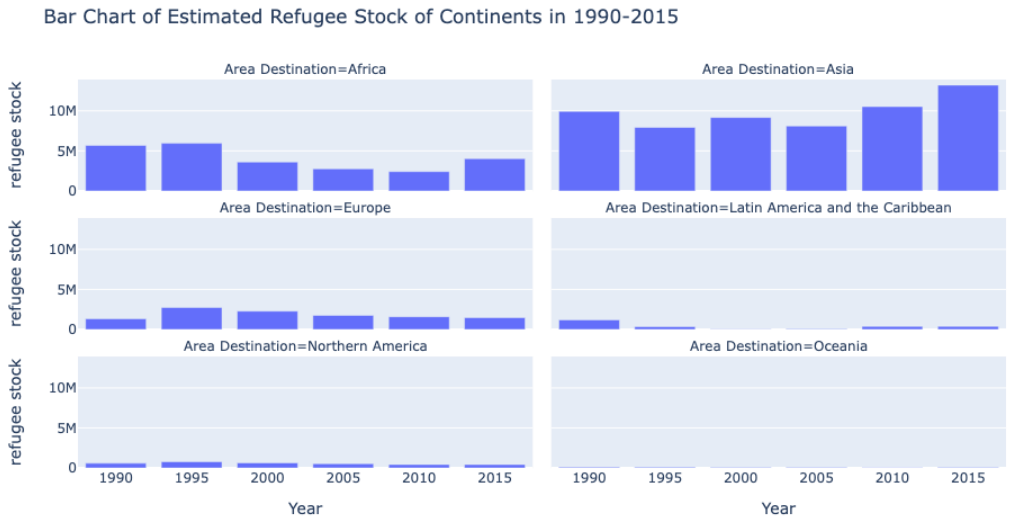

Figure 13 Bar Chart of Estimated Refugee Stock of Continents in 1990-2015

Figure 13 shows the estimate refugee stock for each continent. Asia has the most of estimate refugee stock overtime and Oceania has the lowest estimate refugee stock. Most of the estimate refugee stock is in Africa and Asia.

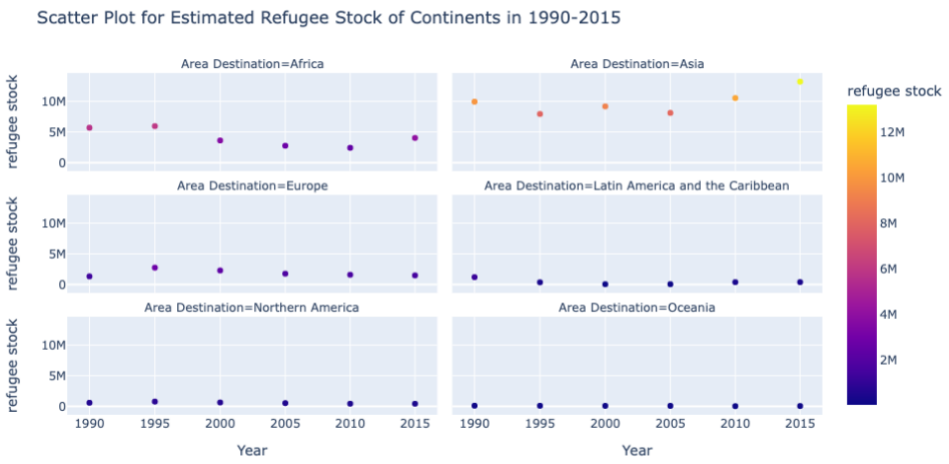

Figure 14 Scatter Plot for Estimated Refugee Stock of Continents in 1990-2015

Figure 14 is the scatter plot for estimated refugee stock of continents in 1990-2015, shows the highest estimate refugee stock occurs in Asia 2015. The estimated refugee stock in Asia is far higher than that of other continents every 5 years.

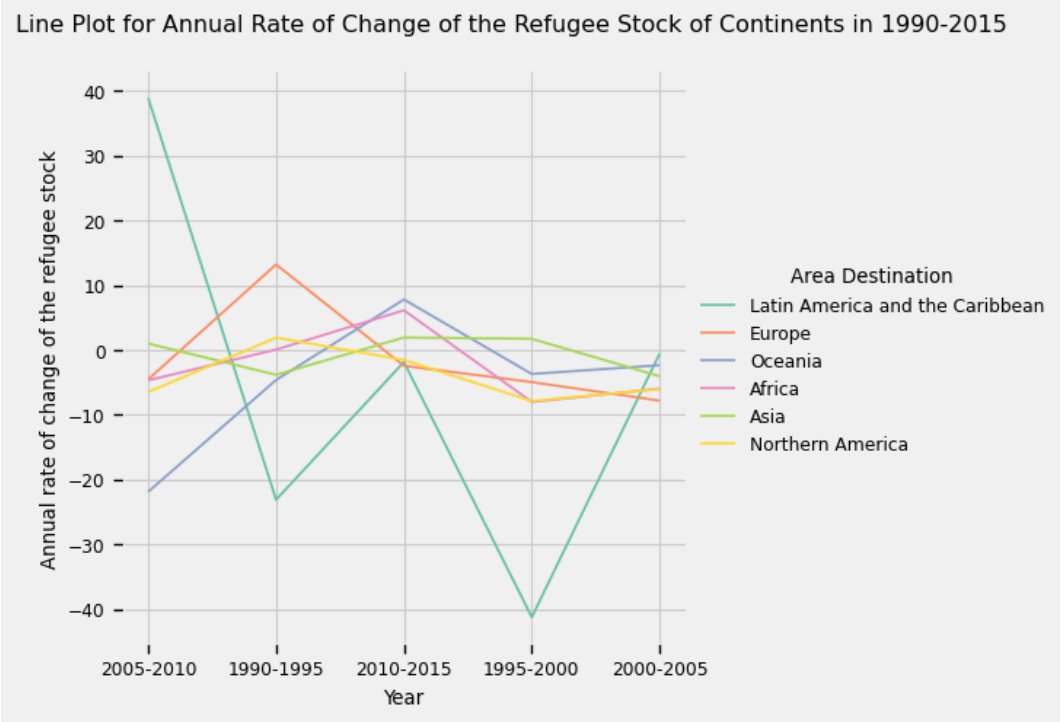

Figure 15 Line Plot for Annual Rate of Change of the Refugee Stock of Continents in 1990-2015

For the annual rate of change of the refugee stock across continents, figure 15 shows Latin America and the Caribbean fluctuated violently between 1990 and 2015, but the overall trend is downward. This is a problem that cannot be reflected in Figure 14, because compared with Asia, Latin America and the Caribbean has a smaller population base.

Figure 16 Violin Plot for Annual Rate of Change of the Refugee Stock of Continents in 1990-2015

To better reflect the fluctuation of the annual rate of change of the refugee stock across continents, based on the years for each continent, a violin plot (Figure 16) is drawn. Violin plots are used to show data distributions and their probability densities. Largest difference in distribution shape between continents occurs in Africa. Asia, Europe, Latin America and the Caribbean, Northern America and Oceania are having relative. The wider parts of the violin plot represent observations with higher probabilities, and the narrower parts correspond to lower probabilities. Therefore, Europe, Latin America and the Caribbean, Northern America and Oceania have a higher probability of decreasing annual rate of change of the refugee stock, while Asia has a higher probability of increasing annual rate of change of the refugee stock.

Discussion

In this exploratory data analysis, by following the Tufte's visualization principles and classic designs, a total of 16 valid figures were plotting form around 6 tables and 5 questions raised. According to the type and context of the data, the data should be selected for drawing, and it is necessary to think about how to show the characteristics of the data more intuitively. Thousands of figures can be drawn from this data set, generate the plots for the questions raised can make the report look more consistent.

Conclusion

To communicate information clearly and effectively, data visualization uses statistical graphics, charts, infographics, and other tools. Numerical data can be encoded using dots, lines, or bars to visually convey quantitative information. Visualization is an essential part of data science results. It seeking to convey the story behind the data, one will often find visualization or what we describe in more detail as the visual process storytelling is one of the most effective ways.

Reference

Aragon, C. R., Guha, S., Kogan, M., Muller, M., & Neff, G. (2022). *Human-centered data science: An introduction*. The MIT Press.

Edward Tufte *The Visual Display of Quantitative Information and Envisioning Information*, Graphics Press, PO Box 430, Cheshire, CT 06410.