University of Toronto

Trends in International Migrants Stock from 1990 to 2015

By:

Zhen Wu

Faculty of Information

INF1340 Programming for Data Science

Instructor: Shion Guha

December 15, 2022

Introduction

The United Nation conduct regular research to track and collect empirical data on the number of international migrants every year by major area, region, country or area, and sex. The dataset that is used in this report aims at finding the trends in international migrant stock, and it was last modified in 2015. The original dataset contains complex data of six tables which were hard to read and visualize. Therefore, the data cleaning process, followed by tidydata principles, was performed prior to this report. After the cleaning process, the original Table 6 was divided into three tables and rest of the five tables were reorganized with same title and contents of information. The new dataset contains information related to the change in migrant stock in more than 200 countries from 1990 to 2015. For instance, Table 1 and Table 2 tracks the total international migrant stock and total population by sex from 1990 to 2015. Table 3, 4, and 5 shows the relationship between migrant stock, total population, and sex. Table 6 goes one step further and seperate refugee from international migrant stock, and collects the refugee stock data by country and area, as a percentage of the migrants stock, and annual rate changes.

Method

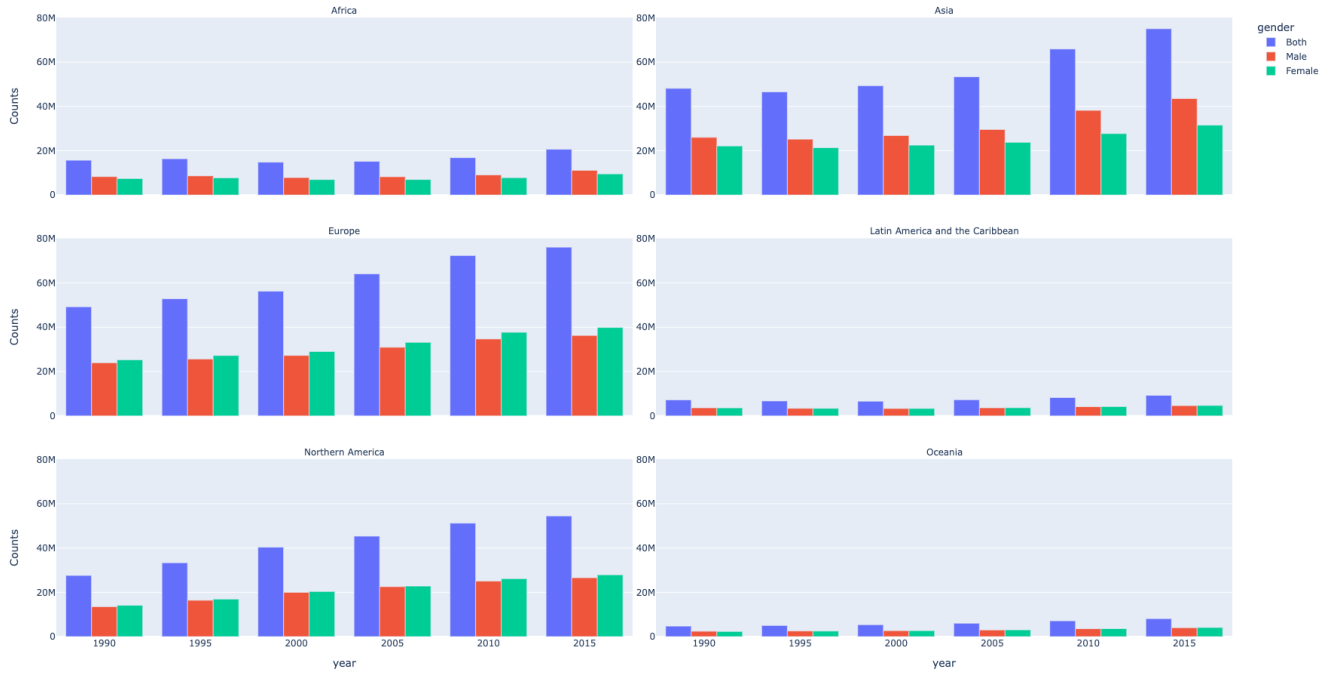
With rapid increase in the amount of generated data, data visualization gives the audience an opportunity to better understand and interpret the relationship between the raw data. This report aims at finding out the the trends of international migrants among five continent including Africa, Asia, Europe, Latin America and the Caribbean, Northern America, and Oceania. The visulizations come with total of seven graphs in the form of four different types, including bar charts, line charts, box plots, and violin plot. All the data visualizations within this report follow the Tufle's principles including but not limited to showing the accurate data, avoid distorting the data, present various data in a small space, make large datasets coherent, reveal data at several levels of detail, serve a reasonably clear purpose, and more. Due to the convenience in labeling and customization, Plotly is the main graphic tool used in the report.

Results

Results for Table 1: Migrant Stock by Continents from 1990 to 2015

As shown in Figure 1, the color of blue represents the overall international migrant stock, red represent male migrants, and green represents femail migrants. Ths graph uses suitable and distinguished colors to help audience to easily track the three types of migrnats information which follows Tufle's principle of encourage eyes to compare data. Overall, the migrant stock raises from 1990 to 2015 with Asia and Europe as two of the most popular continents to attract international migrants. Out of the five continents, Asia has the most conspicuous differences between male and female international migrant stock. The number of male international migrant stock was greater than female in 1990, and the increase in male migrant stock was faster than female migrants in Asia. On the opposite, Europe has always been having more female international migrants than male international migrants. Latin America and the Caribbean and Oceania turned out to have the lowest number of international migrants stock among the five continents. These two continents did not show a significant increase in the past years.

Figure 1: Migrant Stock by Continents from 1990 to 2015



Results for Table 2: Total Population by Continents from 1990 to 2015

Under Tufel's principle, data visualization need to serve a reasonably clear purpose. Table 2 contains information about the total population by sex and countries. Since the purpose for this report is to analyze the trend of international migrants stock in the five continents. The Figure 2, as shown below, was designed to show the total population data in the same five continents. For easier compare, the color setting stays the same as Figure 1. From the graph, the total population in Asia and Europe show some interesting results. As indicated from Figure 1, Aisa and Europe are two of the most attractive area for international migrants. The increase in migrants stock supposed to have a significant impact on the total population. However, Europe did not have a significant growth in total population which indicates that there is a potentially high probability of internal flows in European countries. In Asia, we do see an increase in the total population, but the increase in male population is relatively slow compare with the increase in male imigrants stock.

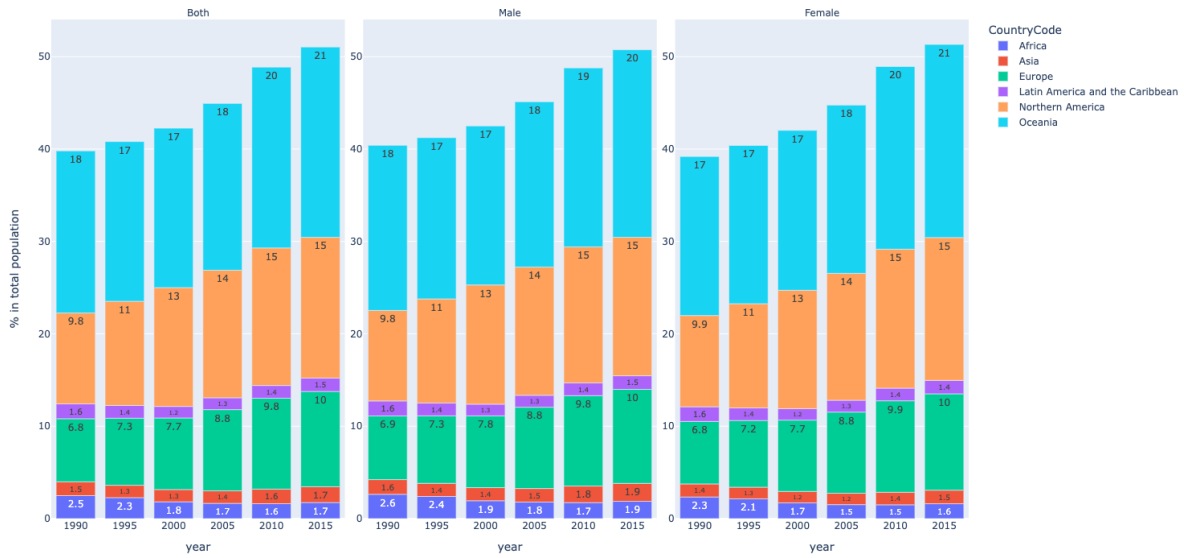
Figure 2: Total Population by Continents from 1990 to 2015

Results for Table 3: Migrants percentage in Total Population by Sex and Continents from 1990 to 2015

Table 1 and 2 shows a detailed information about the international migrants stock and total population in the five continents. Table 3 provided us with the migrants percentage in total population which is essential for us to understand the

relationship. Therefore, another bar chart shown in Fugure 3 is designed based on Tufel's principal of avoid distorting the data. In Fugure 3, the x still referring year, and y is not referring the imigrants percentage in total population. Differs from the previous two figures, continents are now separated by colors instead of gender. On each of the bar, there is a clear lable of the actual percentage. This way, audience can easily see if the percentage has been increasing or not. From this setting, we can easily see that Oceania has the largest percentate of international migrants followed by Northern America and Europe as the second and third of the largest. Asia instead is one of the continents which has the lowest percentage of international migrants stock.

Figure 3: Migrants % in Total Population by Sex and Continents from 1990 to 2015



Results for Table 4: Female Percentage in Migrants by Continents from 1990 to 2015

One of Tufel's principles is to reveal data at several levels of detail. Now we have figured out the international migrants percentage among the five continents, we take a step further to try and see if there is any trends in the female migrants. From Figure 4, two of the continents have been facing a decreasing in femal migrants which including Africa and Asia. Three out of the five continents have attracted more femal migrants including Europe, Latin America and the Carbbean, and Oceania. Northern America faced a decreasing at the beginning, and got a sharpe turn around after 2005. Another fact is that not just Africa and Asia have been loosing female

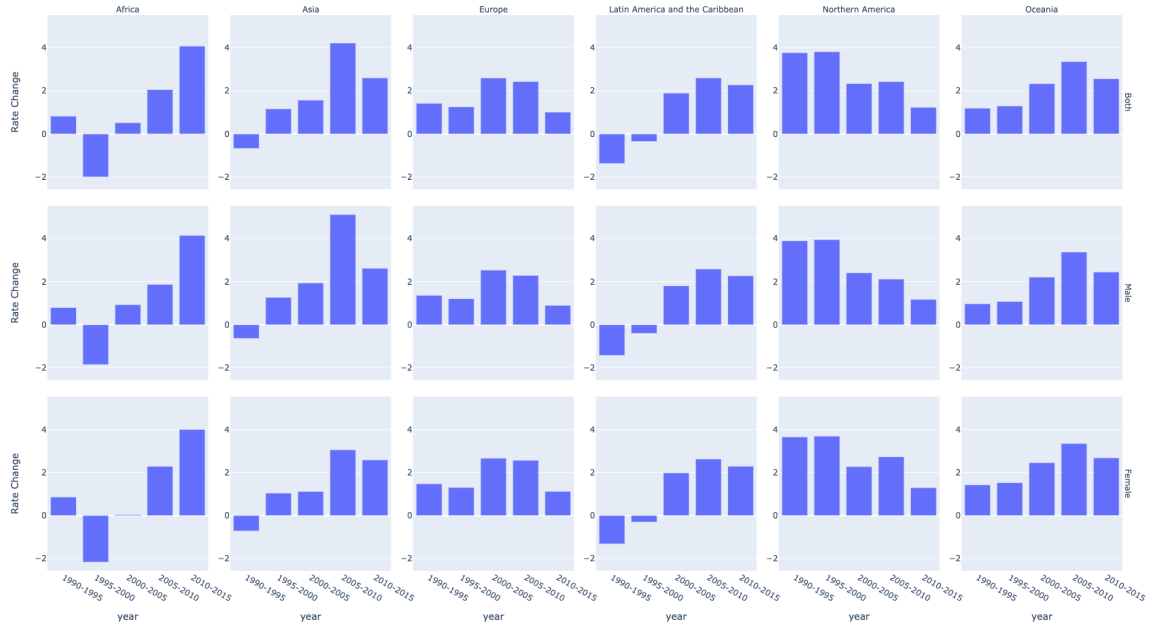
migrants, these two continents had a comparatively lower percentage in femail migrants from the beginning.

Figure 4: Female % in Migrants by Continents from 1990 to 2015

Results for Table 5: Annual Rate Change in Migrant Stock in Sex by Continents from 1990 to 2015

From Table 3, we visualized the percentage of migrants among continents. Table 5 provides us with the calculated annual rate change among the same group of data. Therefore, Figure 5 uses this information to make a bar chart with positive annual rate above 0, and negative annual rage change below 0. This visualization also reflect back Tufel's principle of revealing data at several levels of detail. As shown in Figure 5, Africa had a decrease in annual rage change from 1995 to 2000, and Asia had a this situation happened from 1990 to 1995. Lartin America and the Caribbean had the negative annual rage change in the first decate of this dataset.

Figure 5: Annual Rate Change in Migrant Stock in Sex by Continents from 1990 to 2015



Results for Table 6: Estimated refugee stock by Continents from 1990 to 2015

Table 6 contains a complex dataof the estimated refugee stock. This table is aims at helping the audience to better understand the international migrants group. And for the visualization purpose of this report, we care about the estimated refugee stock more than its percentatge in migrants and the annual rate change. With Tufel's principle of present various data in a small space, we used a boxplot (Figure 6) and a violin plot (Figure 7) to visualize the data. From Figure 6. we have x as continents and y as estimated refugee. The continents are separated by different colors for easier comparison. The top line of each box plot refers to the highest number of refugee a continent has, and the bottom line refers to the lowest number of refugee. The line inside of the box refers to the mean of the estimated refugee in that incontent. With that in mind, we can see that Asia has the highest mean followed by Africa.

Figure 6: Estimated refugee stock by Continents from 1990 to 2015

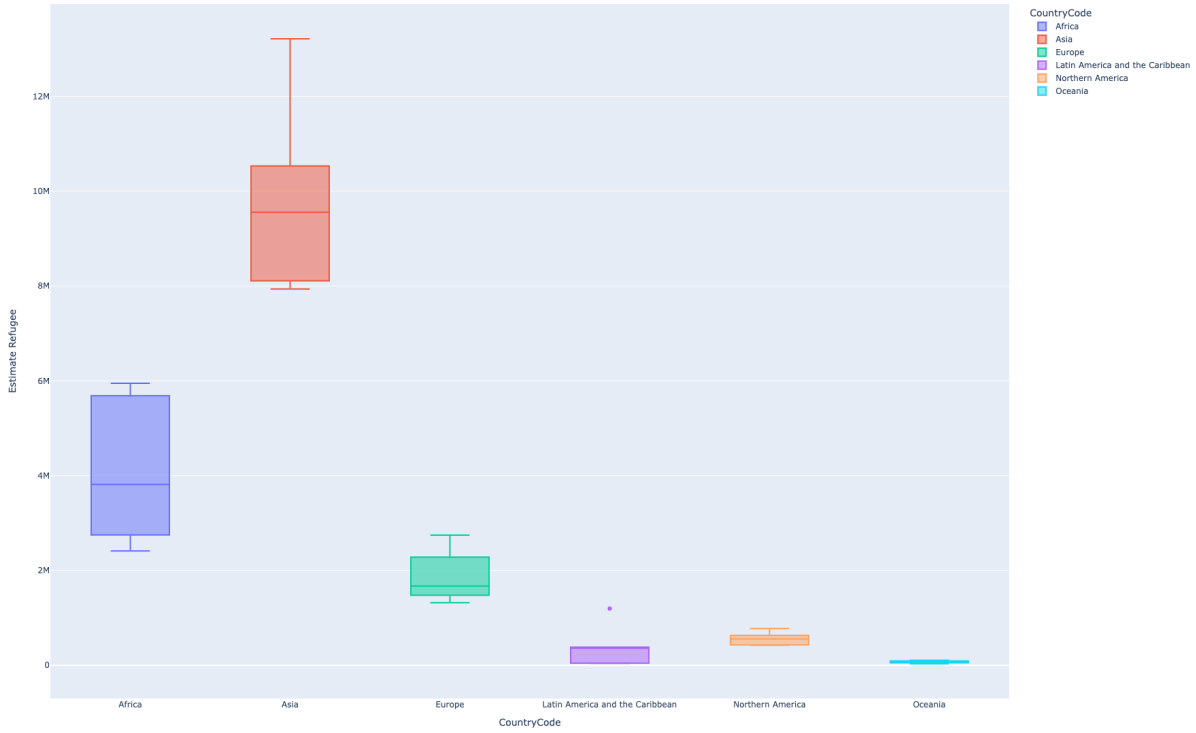
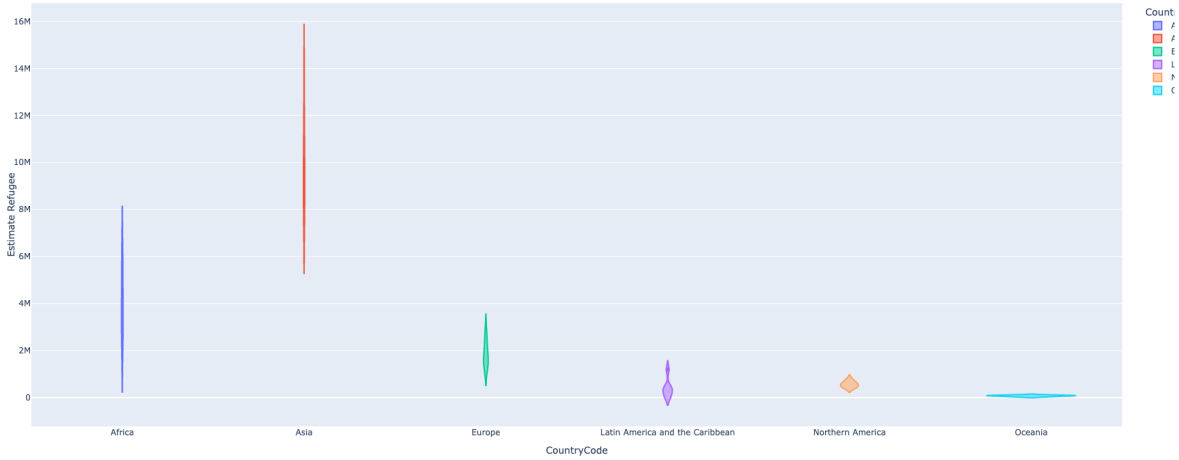


Figure 7 shows that how close the actually data is to the mean. The thicker graph mens a more different between mean and actual data. For example, Asia and Africa almost look like a line, which indicates that the estimated refugee number increase constantly over the years. On the other hand, Oceania seems to have been having up and down in their number of refugees.

Annual Rate Change in Migrant Stock in Sex by Major Area from 1990 to 2015



Discussion

From the result, we see that Asia and Europe has attracted many migrants, but the migrants they have are very different. Asia seems to be more attractive to male migrants. And Europe seems to have a heavy internal flow between the countries located in the area.

There are some limitations towards the data visualization as well. One of the limitation is that we set the purpose to discuss the trende between the five contents instead of the actual country. Therefore, the data we are showing can only tell the audience a general idea of what happened in the past years. Another limitation is that we did not include reference of immigrants policies between these incontinents, and it limits our ability to analyse the reason behind all the increase or decrease in annual rate change or female percentage.

Conclusion

Data visualization is essential in data science since it is where the audience can finally see what the team has been working on. And it is the only way for business partners and colleagues to find useful information effectively. Visualization by itself won't provide enough information, it has to combine with human and with the real world. For further analysis of this report, two action should be conducted. First, look into more details of the countries with large number of immigrants. Second, have more research on the global events happed from 1990 to 2015.