Yiran Zheng, 1003993725
INF1340, Prof. Guha
Dec 9th, 2022

The Exploratory Data Analysis of the Change of International Migrants

Introduction:

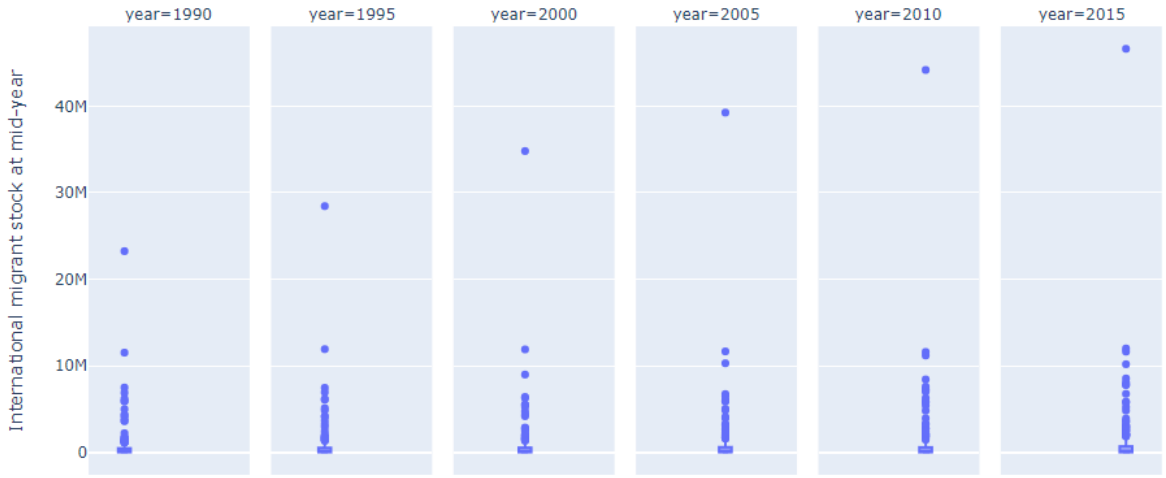
The international migration is important to economies and societies. It brings benefits to countries of origin and destination. Migrants promote the international trade flow for their countries of origin through the remittances, expand the workforce for their host countries, also contribute to taxes for host countries, even bring an increase in per-capita income. International migration is also a channel for individuals especially form less developed countries to improve their material lives. For countries of origin, more international migrants reduce unemployment in their home countries and reduce the demand for natural resources and services such as food, water, education and health care. However, more international migration can put pressure on the host country's resource allocation. Therefore, the trend of change in international migration is an important topic we need to focus on and research. Exploring this topic, I used the data set named “Trends in International Migrant Stock: The 2015 Revision” and Tufte’s principles for data visualization. Its data describes the change of international migrants over 200 countries or areas between 1990~2015 from 6 aspects: international migrants’ stock, the total population of each country or area, international migrants as a percentage, female international migrants as a percentage, annual rate of change of the migrant stock, estimated refugee stock at mid-year. The data is classified by country’s geography information and level of development information, and most of them are also classified by sex. The Exploratory Data Analysis of the change of international migrants and refugees is focused on two aspects: geography aspect (major areas and regions) and developing aspect (whether the country is a developed country, developing country or a less develop country).

Method:

Before the Exploratory Data Analysis, I tried to make graph that contains all countries and the graph turned out to be not that informative (as showed in Figure 1-0

below), and I discovered that display information based on geography information and level of development information shows the relationship between a country's immigrant information and its geography information/level of development (show causality). I could also compare immigrant information of different regions and level of development (show comparisons). In order to visualize the data set, I imported "pandas" to handle data frames and "plotly" to draw graphs.

Figure 1-0: International migrant stock at mid-year



Firstly, I read data from "UN_MigrantStockTotal_2015.xlsx" and set up some new data frames with the ANNEX table of the excel file. I made four data frames from the ANNEX table. One data frame contains only regions, one data frame contains only major areas, one data frame contains both regions and major areas, and one data frame contains the level of develop information so that I can deal with all kinds of situations:

1. when I only need region information.
2. when I only need major area information.
3. when I need the relationship information between the region and major area.
4. when I need the level of development information of a country.

Then, I cleaned each table so that my data frames are tidy. All tables (1~6) have some columns that are data values (e.g., Some of them contain year information and also contain gender information), therefore, I converted these column names(header) to column values and insert some new columns to store the information in them (e.g., create a new column called "year"). Next, I discovered that some converted columns

have more than 1 variable (e.g., “female” + “1990” is actually “gender” + “year”), so the next thing to do is to split those kinds of columns. Besides, in table 6, “Estimated refugee stock at mid-year (both sexes)”, the columns “Refugees as a percentage of the international migrant stock” and “Annual rate of change of the refugee stock” are different types of data which violates principle 4. Thus, I split them into three independent tables (UN_6_1, UN_6_2, UN_6_3). At last, I dropped all columns that are not directly related to the main topic, including “sort order”, “country code”, “notes”, etc.

Thirdly, I identify from which aspect I want to divide the data set and merge the data set with related table of ANNEX.

Finally, I draw graphs based on these data frames. For the table 1, 2, 3, 5, I used regions as a unit and display them by a bar chart. As the data is from 1990 to 2015, I separate every table into 6 graphs and displayed it 3 * 2 for table 1,2,3. I use a different color for different gender, and it is also easy to see the amount of both gender as the graph combines males and females for every region (Use appropriate encodings: color). For the third table and fourth table and UN_6_3, I used boxplot and divides the data based on its development information. Similarly, I separate every table into several graphs according to time information. For the last table (6th table), similarly to the fourth one, I made 2 independent line charts for UN_6_1 and UN_6_2.

Result:

Based on the methods described above, there are 9 figures, and interpretations are stated as follows.

From 1990-2015, the total number of international migrant stock has been on an upward trend as shown in Figure 1 below. The total number of international migrant stock in North America has been the largest and increasing trend during these 25 years (around 50 million by 2015). Western Asia, Eastern Europe, and Western Europe also have the larger amount of international migrant stock in the global atmosphere. In particular, Western Asia was second only to North America in terms of the total amount of international migrant stock as of 2015.

Figure 1: International migrant stock at mid-year

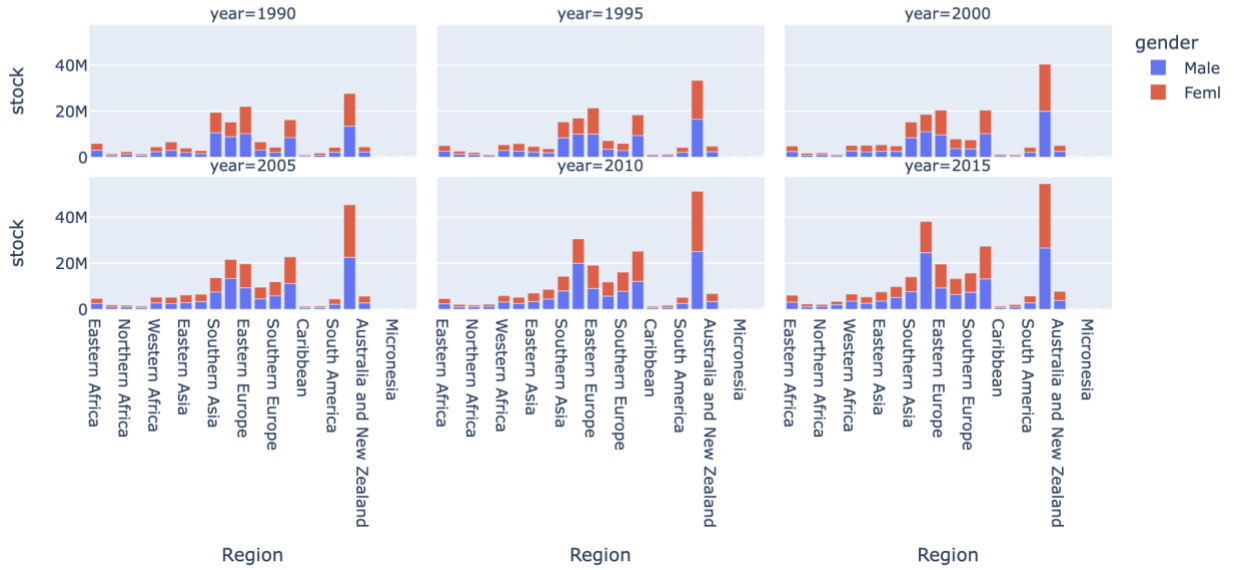
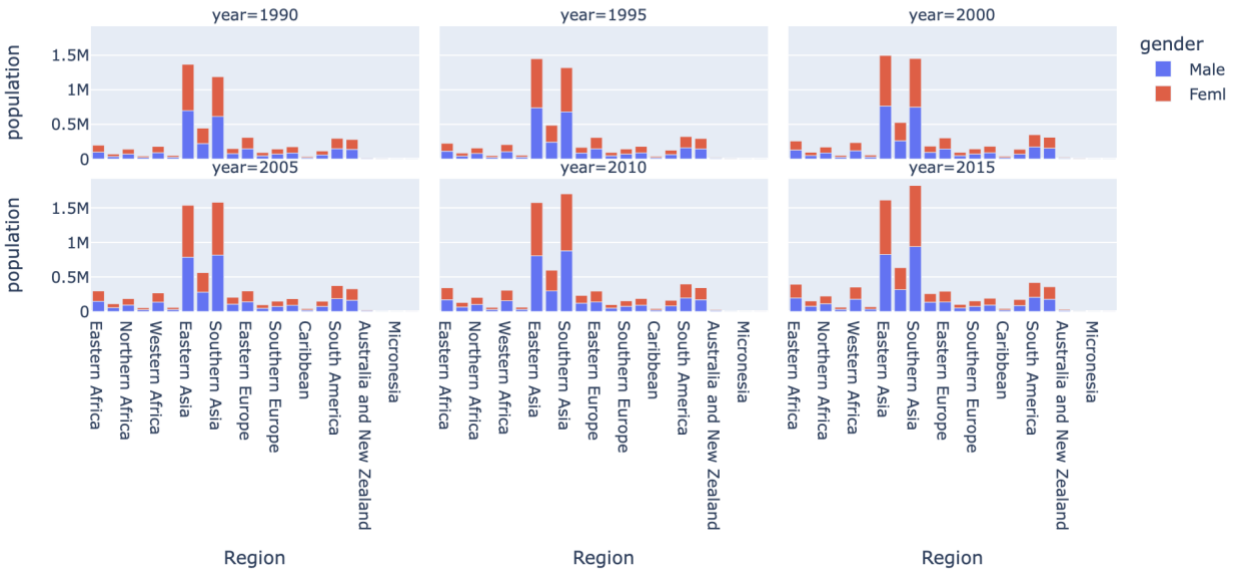


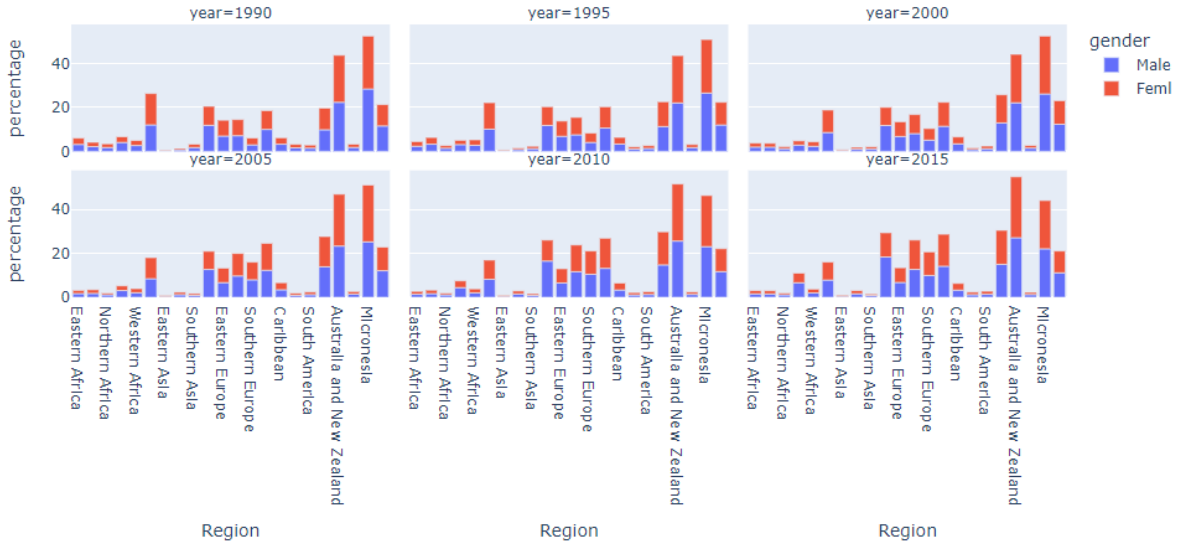
Figure 2: Total population at mid-year



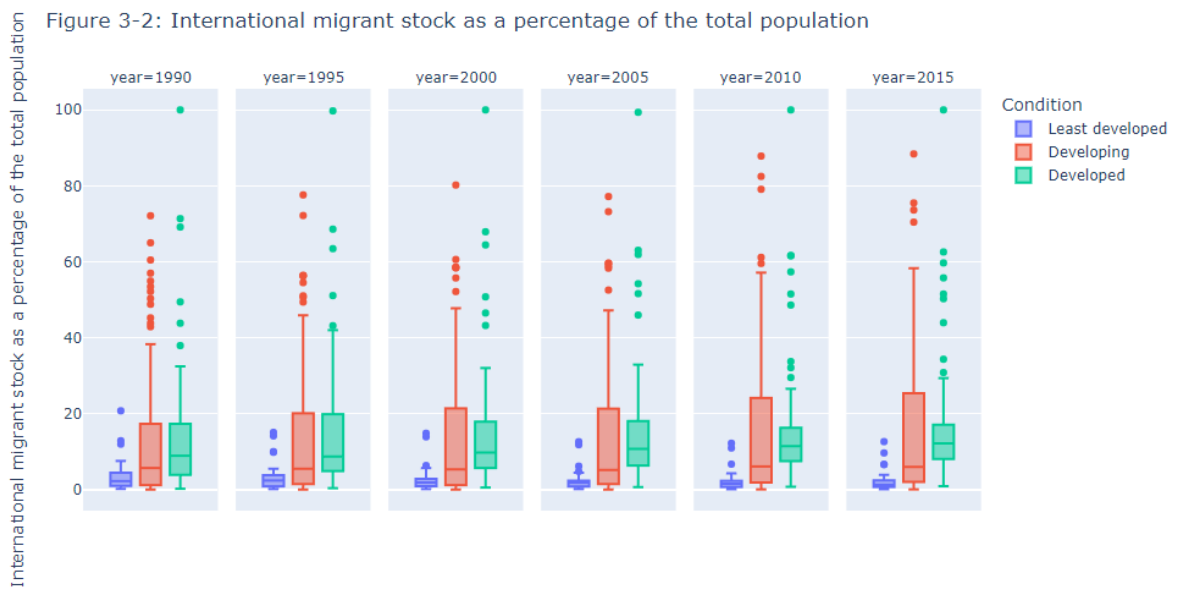
In Figure 2 above (in thousands), the population of each Eastern Asia, Southern Asia, and South-Eastern Asia has been the largest three in the world in every mid-year between 1990 and 2015. In 1990, South Asia had the largest total population in the world, followed by East Asia, both of them have a population larger than 1000 million. However, East Asia's population surpassed that of Southern Asia to become the world's

first in 2015. The change in the total population of North America and South has been stable over these 25 years.

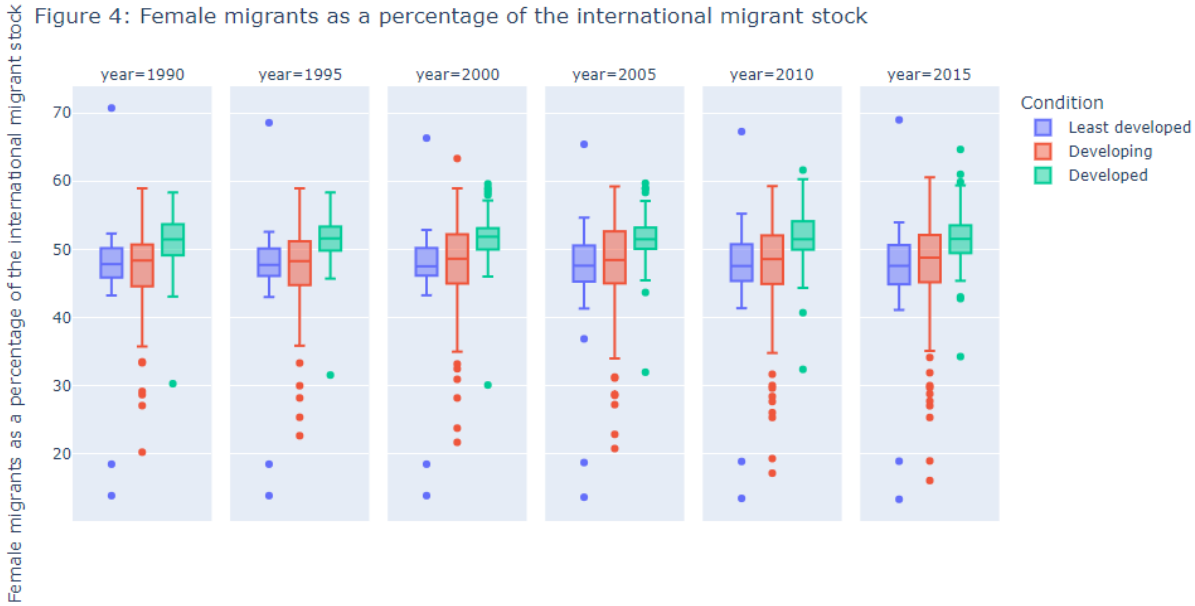
Figure 3-1: International migrant stock as a percentage of the total population



From Figure 3-1, the percentage of international migrant stock had only increased slightly from the macro perspective, but some regions had distinct changes in those 25 years. The percentage of migrants in Australia and New Zealand had been increasing steadily. In 2010, Australia and New Zealand exceeded Micronesia and became the highest in the world. Both of them have a percentage over 40%. International migrants are also a large component of Western Europe, Northern America, Western Asia, and some other regions, the percentage in these countries is around 20 or higher.



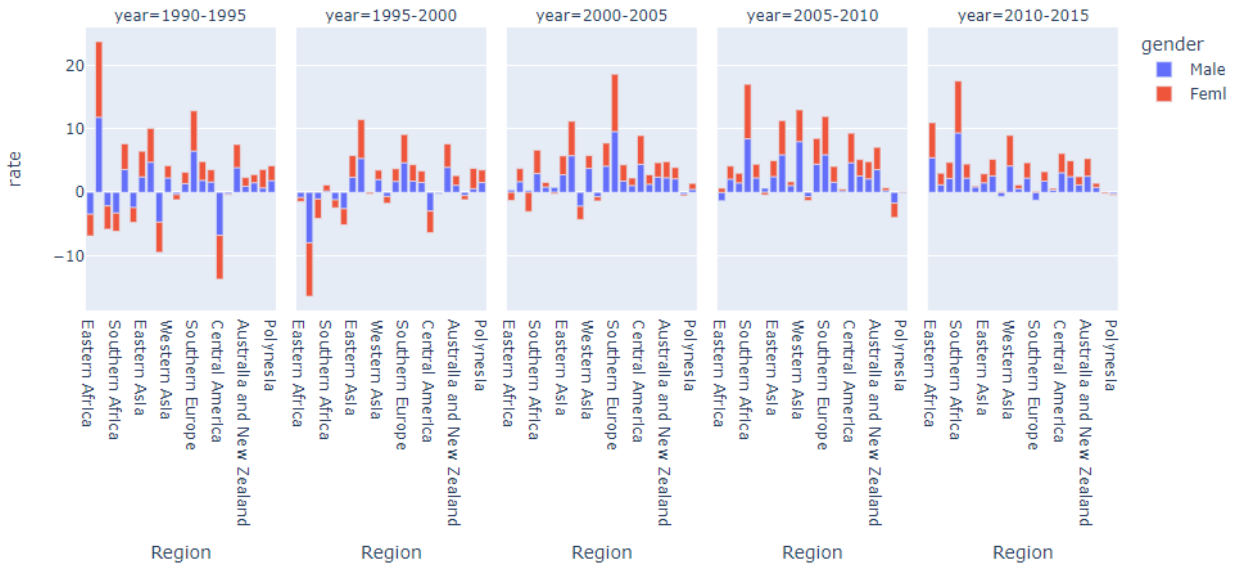
From Figure 3-2 above, the value of international migrant stock as a percentage of the total population is related to the level of development of the country. The more developed the country is, the higher the percentage is. The percentage of developing country is increasing in the recent 25 years while the other two remains stable.



From Figure 4 above, the value of female migrants as a percentage of the international migrant stock is related to the level of development of the country. The more developed the country is, the higher the percentage is. Only the median of

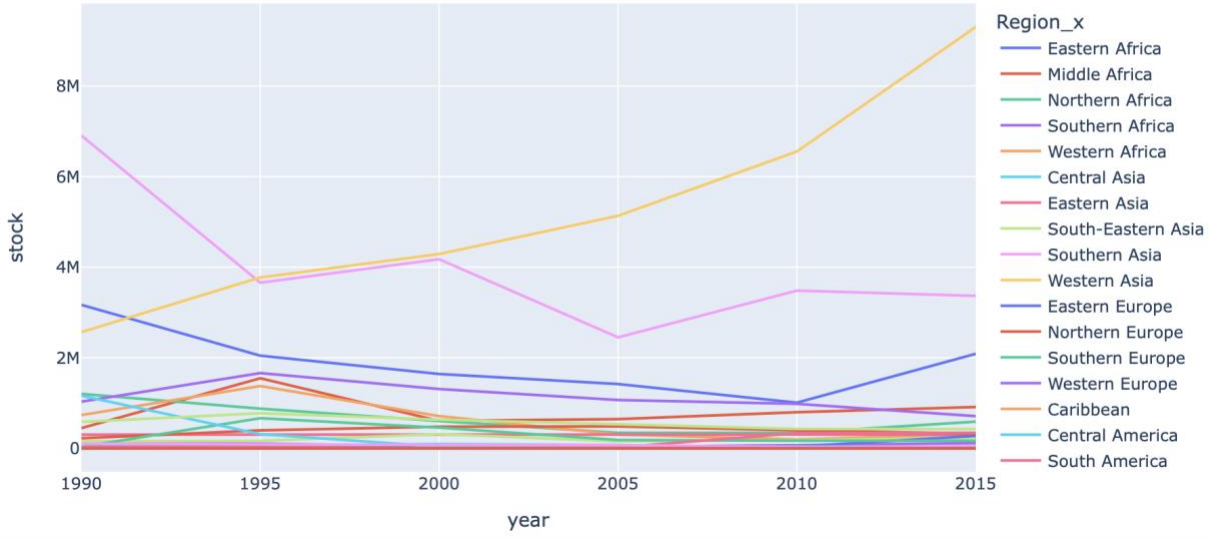
developed region is higher than 50% while the other two are lower than 50%. All of them are quite stable in 1990~2015.

Figure 5: Annual rate of change of the migrant stock



In Figure 5 above, we see that the annual rate of change of migrant stock is always the same for both gender in one region, and the general trend of migrants is increasing. There is no specific pattern for this graph.

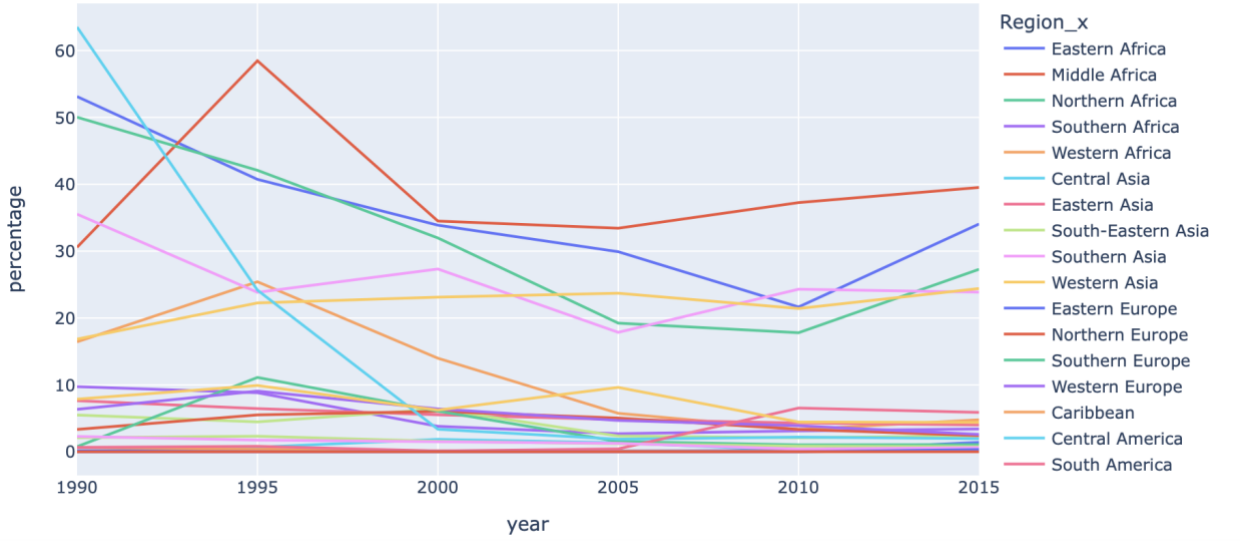
Figure 6.1: Estimated refugee stock at mid-year (both sexes)



From Figure 6.1 above, the majority of the world has refuges of less than 2 million, but Western Asia always has a refugee stock larger than 2 million, and the

number was increasing substantially from 1990 to 2015 and break 8 million by 2015. The refugee stock of Southern Asia was decreasing in those 25 years and kept stable at around 3 million. The refugee stock of Eastern Africa was also over 2 million in 1990 but it kept decreasing to around 1 million until 2010 and rise again to 2 million in 2015.

Figure 6.2: Refugees as a percentage of the international migrant stock



In Figure 6.2, the trend of the percentage of refugees of the international migrant stock in decreasing from 1990 to 2010 and had a rise in 2015. Most of them are below 2010 all the time. From 1990 to 1995, some regions have refugees as a big component of international migrants (over 30%), such as Central America, most regions of Africa, and some parts of Asia, and the majority of them kept decreasing in those 25 years. Central America even got below 10% in 2000 though the percentage was over 60% in 1990. The percentage kept stable since 2000 and only part of Asia and Africa are over 10%, most of them are around 20%, only middle Africa is around 40%.

Figure 6.3: Annual rate of change of the refugee stock

From Figure 6.3, most countries have a rate of change around 0, no obvious pattern for this rate.

Discussion & Conclusion:

According to the exploratory data analysis of the data set "Trends in International Migrant Stock: The 2015 Revision", Figure 1 shows that the number of international migrants living in countries other than their place of birth around the world has been increasing every year since 1990. This could be a sign of globalization. However, there are considerable differences in the number of international migrants between large regions of the world as shown in Figure 3-1. Asia has seen a rapid increase in the number of migrants, more than any other major region (see Figure 1). Migration occurs mainly between countries within the same geographic region. Based on Figure 6-1, the excess of refugees in West Asia over other regions and the possible is the outbreak of the war. For example, between 2010 and 2015, wars broke out in the Middle East, which led to an increase in the number of refugees in and around the Middle East, and also affected the number of international migrants. Combing Figure 1 and Figure 3-1, most international migrants are concentrated in only a few major regions, such as North America and Western Europe. From Figure 3-2, we acknowledged that more developed countries are more attractive to migrants, and the reason for this phenomenon may be that more developed countries have better environment and living facilities and offer more job opportunities. Similarly, developed countries are more attractive to female

migrants, probably because the more developed places the better the living conditions of women will be (see Figure 4). The rising trend of international migration is a phenomenon that we need to pay attention to, as it relates to economic and social development, as well as resource allocation. We need to further explore the reasons that have led to the increase in the number of international migrant and the reasons for the concentrated distribution of international migrants, which will help us understand social and economic development trends. We should also be concerned about the proportion of refugees in international migration, and the increase in the total number of refugees may trigger a humanitarian crisis.

The entire data set is very detailed and needs more professional, more integrated staff, and more advanced analysis data technology to better reflect its value. We also need more global political and financial knowledge to better understand these data and do more comprehensive exploratory data analysis.