INF1340: Final Project

Wenxuan Li

Introduction

To investigate the population changes of a country or the whole world, the international migrant stock is an unavoidable topic. In the midterm project, we focused on the UN data file, which provides the information on the international migrant stock, to do the initial data cleaning and make the data set tidy. In this final project, we will continue the analysis on the UN data file – based on the tidy data, the exploratory data analysis is conducted to find the future appropriate data analysis directions.

The tidy data set, from the midterm project, contains 4770 rows and 19 variables. The variables include the country-level information, such as the country name, the major area, the region, whether it is a developed country, etc.; the time information – the year number; the gender information – female, male, or all gender; and the migrant stock related variables, such as the migrant stock, the total population, the gender proportion, the refugee stock, etc. Based on these variables, in the following report, we will summarize some basic features of the UN data set on migrant stock.

Methods and Results

In this part, we will focus on three major variables to explore their features, population, migrant stock, and refugees, due to their importance. In the meantime, we also include the year information and the gender information (for the population and the migrant stock) to better summarize the data pattern. However, instead of directly visualizing different values on the country level, we choose to visualize on the major area level because the number of countries or regions is too large to be visualized within in one plot – it will not only cause a problem to the coding language, but also cause a problem when interpreting the visualization results. In the following context, we will introduce how we generate these plots one after the other.

Figure 1 gives the total population of six major areas – for all genders, female, and male, from 1990 to 2015. From the figure, we can observe that Asia has the largest population, followed by Africa, Europe, Latin America and Caribbean, Northern America, and Oceania. Except Asia, female occupies the larger proportion of population compared to male among six major areas. However, Asia has a totally different story. This general distribution does not change much over years. On the other hand, there is general decreasing trend for the population in Europe, but an obvious increasing trend for the population in Asia and Africa.

Figure 2 gives the migrant stock of six major areas – for all genders, female, and male, from 1990 to 2015. Unlike what we observed in Figure 1, the differences of migrant stocks do not vary much across different major areas. Meanwhile, Asia and Europe occupy the leading position, followed by Northern America, Africa, Latin American and the Caribbean, and Oceania. Overall, there is an increasing trend for the migrant stock across different major areas. However, since the

maximum keeps changing overtime, we can only directly capture the migrant stock changes on Europe and Asia, all other major areas are not clear enough.

Figure 3 gives the refugee of six areas from 1990 to 2010. Since the original data file does not provide the gender information, we cannot visualize the gender, either. From the figure, we observe that Africa has the highest number of refugees in 2010, followed by Asia, Europe, Latin American and the Caribbean, Oceania, and Northern American. However, in 1990, things are totally different – Asia has the highest number of refugees at that time, followed by Europe, Latin American and Caribbean, Africa, Oceania, and Northern American. There is no general trend for the number of refugees, each major area has totally different patterns. However, Northern America is quite stable – the number of refugees keeps in the lowest state.

Discussion

From the results, we can have a rough idea on the whole data set. Based on the observations made previously, the potential interesting directions to be explored may include: 1) why Europe does not have a large population but has a large migrant stock? 2)why the number of refugees sharply decreased between 1995 and 2005, then bump back? 3)why Africa has an increasing trend on the refugees recently?

Conclusion

In the final project, based on the tidy data set, we conducted the explorative data analysis. Three bar plots are plotted for the population, migrant stock, and the number of refugees, which are the most representative variables from the data file. We have made several observations, like the general patterns or some detailed observations for each plot and several potential research directions are identified correspondingly.

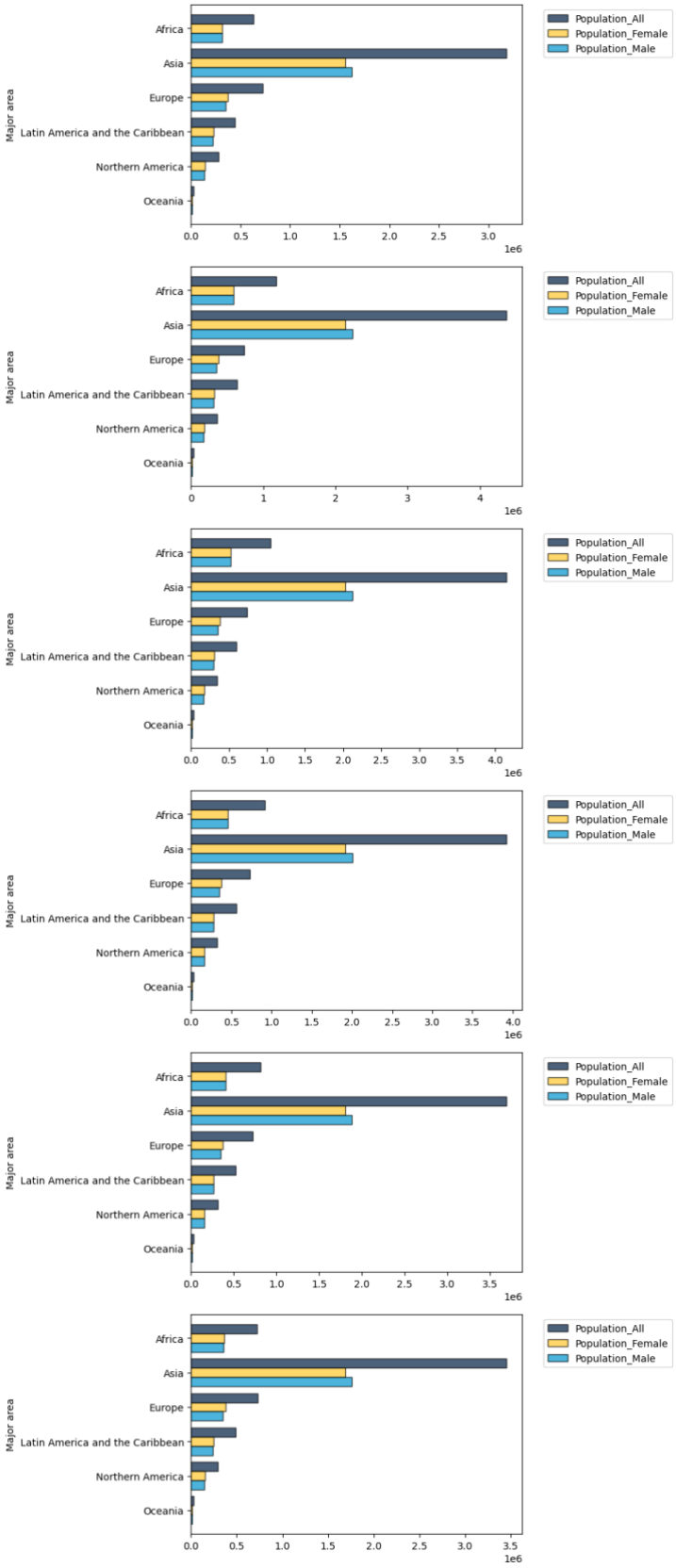

Figure 1. Barplots for population over years

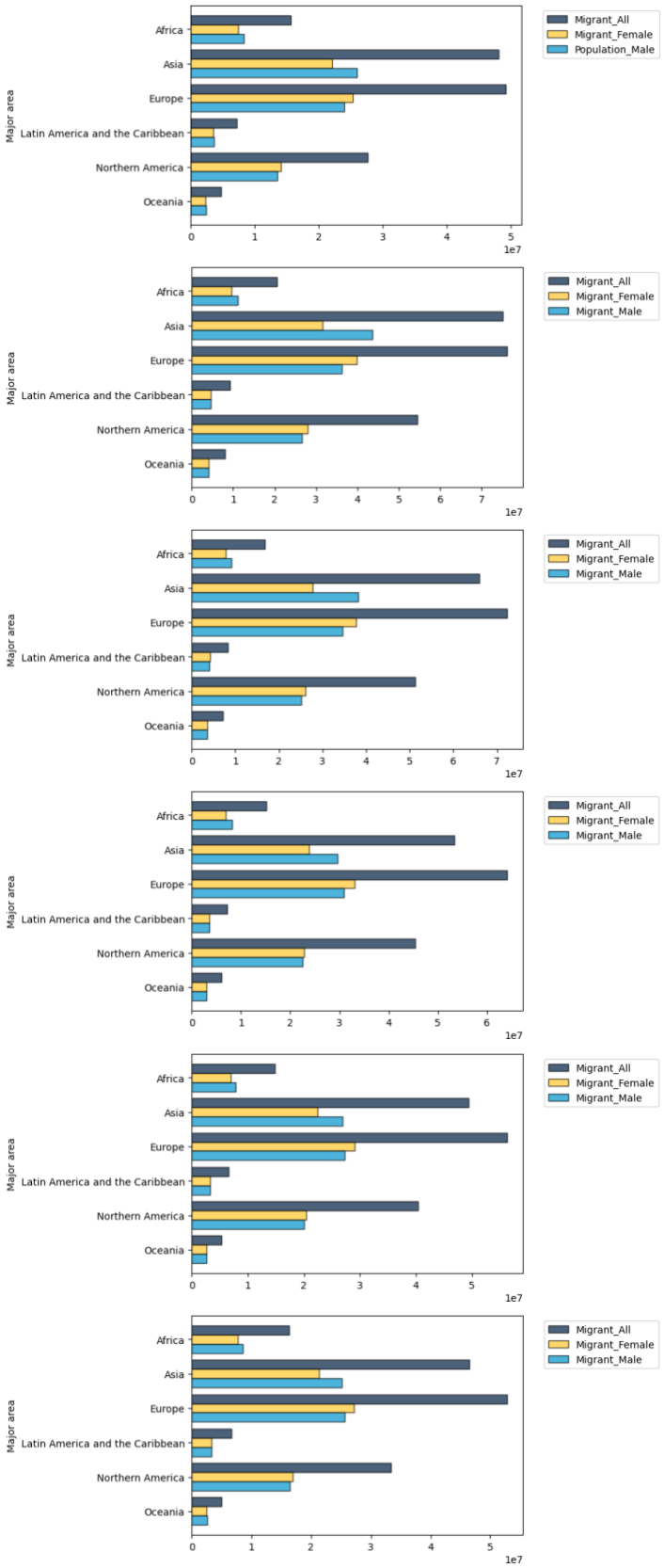

Figure 2. Migrant stock distribution over years

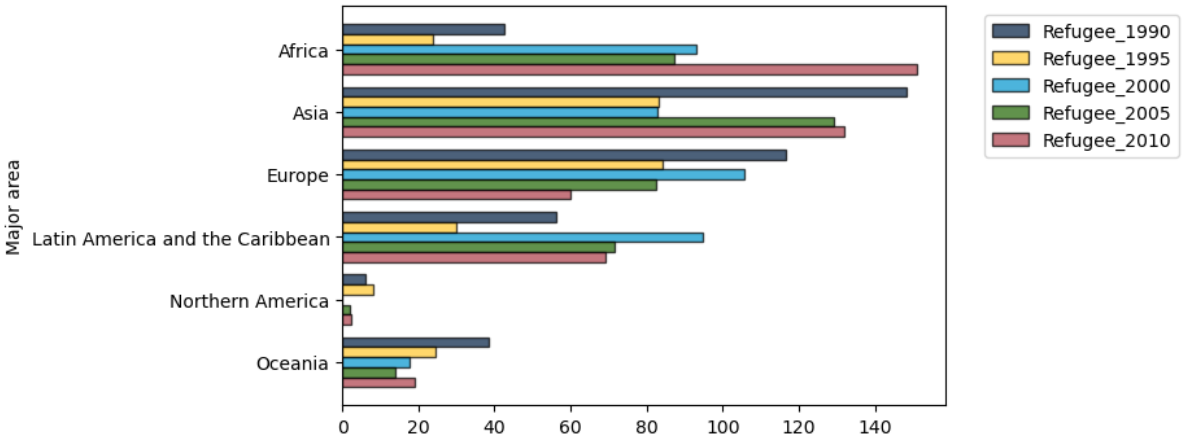

Figure 3. Refugee data over years