

## Final Assignment Write-up

### 1. Introduction

In this assignment, multiple visualization plots including boxplots, scatterplots, bar plots, and line plots were illustrated based on the UN Migrant Stock datasets. The graphs visualized the international migrant stock, total population, the annual rate of change of the migrant stock, female migrant stock, estimated refugee stock, percentage of refugee stock and the annual rate of change of the refugee stock in global and individual continent aspects. Tufte's 6 data visualization principles would be followed to ensure the graphical integrity.

### 2. Method/Result

The UN Migrant Stock datasets contain three tables: (1) a Location table; (2) a table containing year, sex, the international migrant stock, total population, and annual rate of change of the migrant stock (refer as table 2 in the following passage); (3) a table containing year, female migrant stock, estimated refugee stock, percentage of refugee stock and the annual rate of change of the refugee stock (refer as table 3 in the following passage). The location table was merged into each table 2 and table 3 for visualization purposes.

Tufte's six principles were used to visualize the data in efficient and correct way. (1) The y-axis ticks for most of the plots were not modified, therefore the bin size and trend of the line and scatter points reflect the real proportions of the data. (2) All the graphs were clearly labelled based on the variable we used. (3) Each plot derives only the data variation preventing from containing too much information. For example, the bar plot only shows the trend of the associated variable across the years. (4) The unit in graphs are consistent when comparing different continents and global aspect. (5) Graphs were illustrated in two-dimensional to transfer the information in clear and efficient way. (6) All the data points were included in plots to avoid misleading information from quoting data out of context.<sup>1</sup>

#### 2.1 International Migrant Stock

A boxplot was illustrated in Figure 1 to visualize the central tendency of the global international migrant stock.

---

<sup>1</sup> R., R. (2021, February 7). *Edward Tufte's six principles of graphical integrity*. LinkedIn. Retrieved December 12, 2022, from <https://www.linkedin.com/pulse/edward-tuftes-six-principles-graphical-integrity-radhika-raghu>

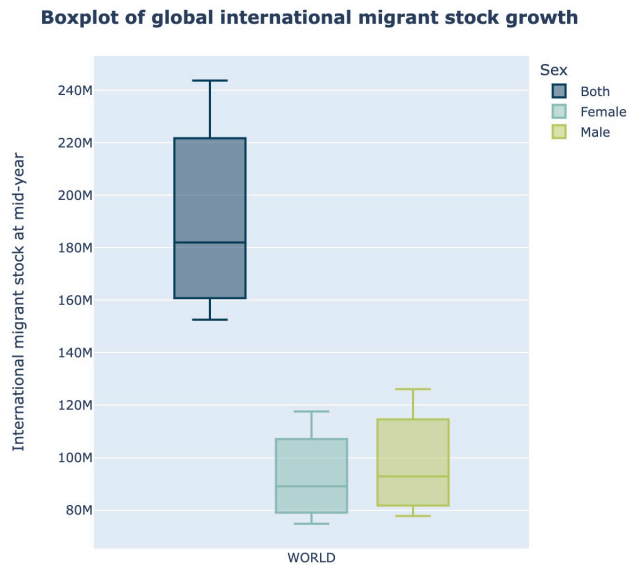


Figure 1: Boxplot shows the statistics summary of global international migrant stock in different sex categories.

The interquartile range of the international migrant stock for both sexes is much larger than female and male migrant stock, which is reasonable because it is the combination of male and female data. Box plots for female and male migrant stock are similar in size and level, which indicates the similar migrant stock in two sex categories within 25 years (from 1990 to 2015). The medians for male and female migrant stock are also similar, this also suggests the similar growth in the male and female migrant stock in the global aspect.

To zoom in on the global growth of the international migrant stock, a bar plot was illustrated in Figure 2.



Figure 2: Bar plot shows and increasing trend of global international migrant stock in three sex categories.

Based on Figure 2, we could see a gradually increasing trend in all three sex categories, which suggests that

the international migrant stock was increasing year by year from 1990 to 2015 in the global aspect. However, the growth of female migrant stock was slightly slower than male migrant stock, which increased the gap between two sex categories as time goes on.

However, if we visualized the international migrant stock in developing and developed regions in Figure 3, the distributions for male and female migrant stock are surprisingly different.

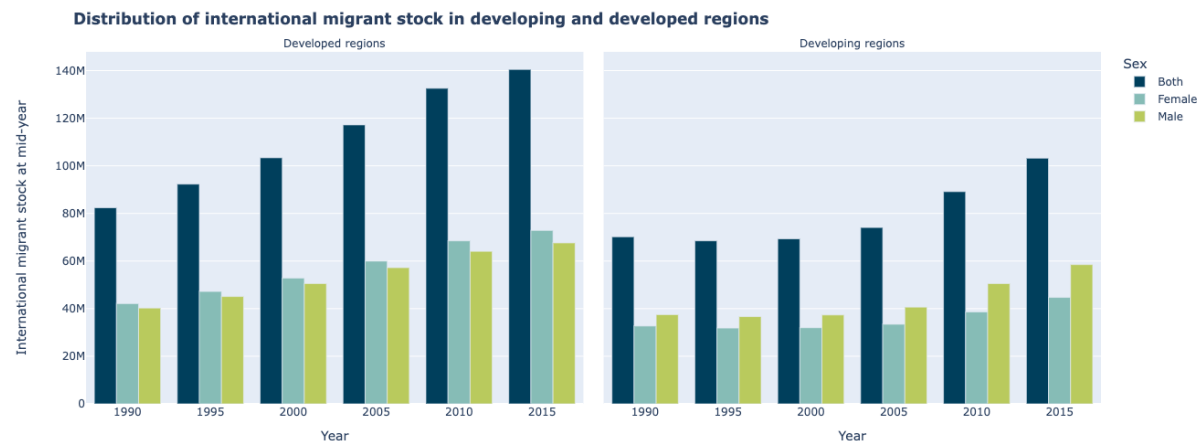


Figure 3: Bar plot shows the growth of international migrant stock in developing and developed regions in three sex categories. Developing regions have a slower growth than developed regions within 25 years. In Figure 3, we discovered that developed regions have more female international migrants, while developing regions have more male international migrants. Both regions have an increasing trend for the migrant stock, but the growth in developed regions is steeper than in the developing regions. The migrant stock in developed regions increases linearly within 25 years, in comparison, developing regions started to have noticeable growth from 2010.



Figure 4: This plot shows the increasing trend of international migrant stock in five continents. Europe, Asia, and Northern America have significant growth.

Figure 4 shows the growth of international migrant stock in six continents: Asia, Europe, Northern America, Africa, Latin America and the Caribbean, and Oceania. Both Europe and Northern America have a linear growth from 1990 to 2015. Moreover, both of these two continents have higher female migrant stock than males. The growth of international migrant stock in Asia became noticeable in 2010, but female migrants were continuously less than male migrants from 1990 to 2015. Africa, Latin America and the Caribbean and

Oceania did not have any significant growth in migrant stock compared to the trend in other three continents.

## 2.2 Total Population

A box plot and bar plot were illustrated in Figure 5 to visualize the descriptive statistics and trend of global population.

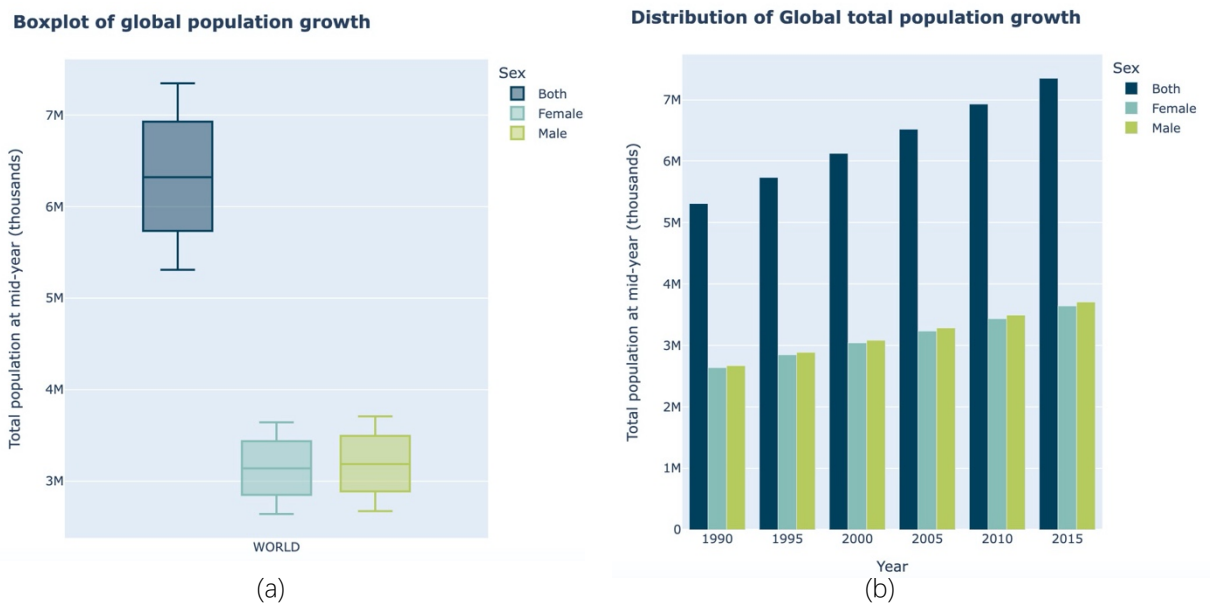


Figure 5: (a) Box plot shows the statistics summary of global population growth. The population of female and male are similar in global aspect. (b) Bar plot shows a linear growth in global population

Similar to the international stock migrant data, the global population did not have significant differences between males and females. The similarity in box size and median suggests that the population growth in males and females to identical to each other. Figure 5b shows a linear increasing trend of global population growth. There has no dramatic difference between the female and male populations.

Line plots were illustrated in Figure 6 to distinguish the population growth in each continent.

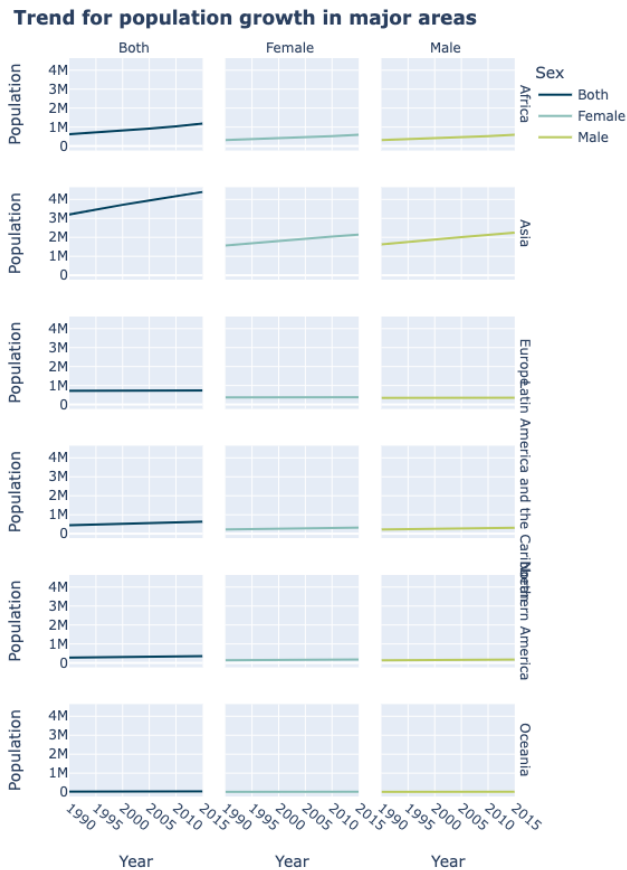


Figure 6: Line plot shows trend of population growth in each continent.

It can be concluded from Figure 6 that there was no markedly increase in population growth in Europe, Latin America and the Caribbean, Northern America, and Oceania. However, Asia and Africa have dramatically upward trends in population growth within 25 years. The sex categories shared the same growing trend.

### 2.3 International Migrant Stock as a Percentage of Population

A Box plot and bar plot were graphed in Figure 7 to visualize the descriptive statistics and trend of the global percentage of migrant stock based on the population.

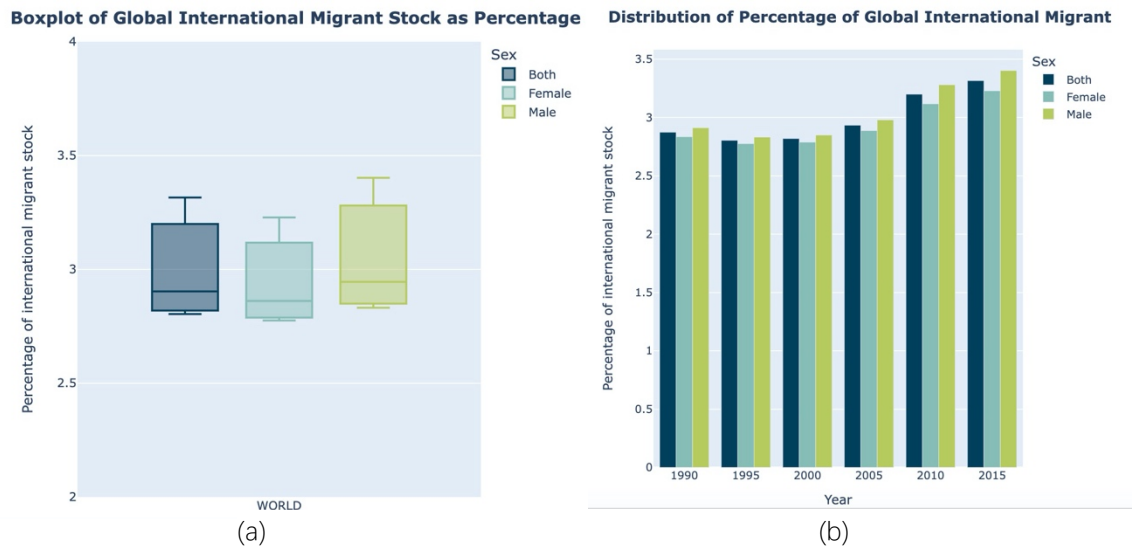


Figure 7: (a) Percentage of male migrant stock has wider IQR than other two sex categories. (b) There is slight downfall of percentage of migrant stock in 1995 and 2000, then pick up after 2005.

Figure 7a shows a different pattern from the previous sections. The percentage of migrant stock was calculated based on the total population of the associated location. The maximum and median of the percentage of male migrant stock were larger than the other two categories, which indicates that among the international migrants, there were more males than females. Figure 7b shows that the percentage of international migrant stock fell before 2010 and began to make recovery after 2005. The percentage of female migrant stock was always lower than the percentage of male migrant stock within 25 years.

If we zoom in on the continent's aspects, the following plot shows the fluctuation of the percentage of migrant stock and the international migrant stock over 25 years. The size of the circle represents the total population of the associated continent in each sex category.

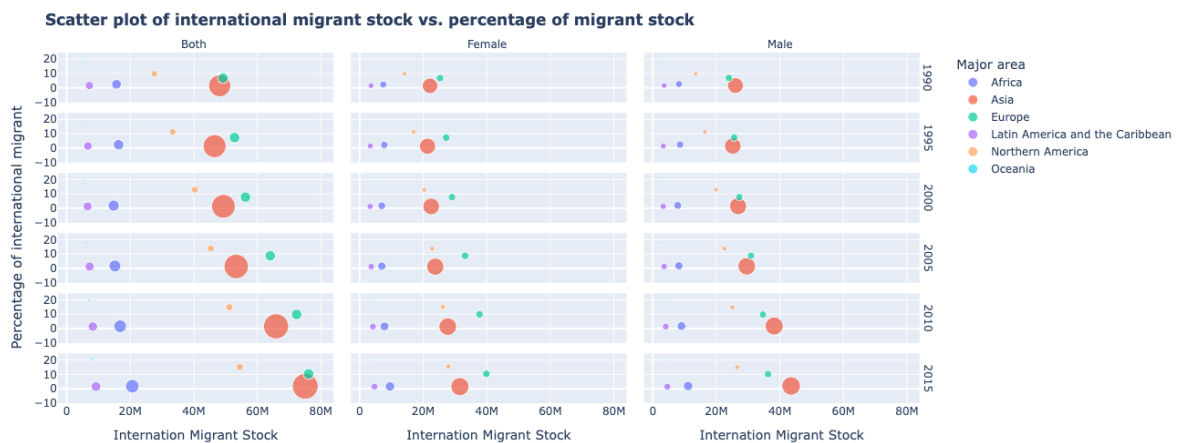


Figure 8: Scatter plot shows the change of international migrant stock and the percentage of migrant stock in three sex categories within 25 years.

It can be seen from Figure 8 that Asia has a larger population and its international migrant stock grow continuously over 25 years in all sex categories. There was no drastic change in the percentage of migrant stock since the number of international migrants increases as the population increases.

Oceania has no significant change in all international migrant stock, the percentage of the migrant stock and the total population which match the visualization results in the above sections.

## 2.4 Annual Rate of Change of the Migrant Stock

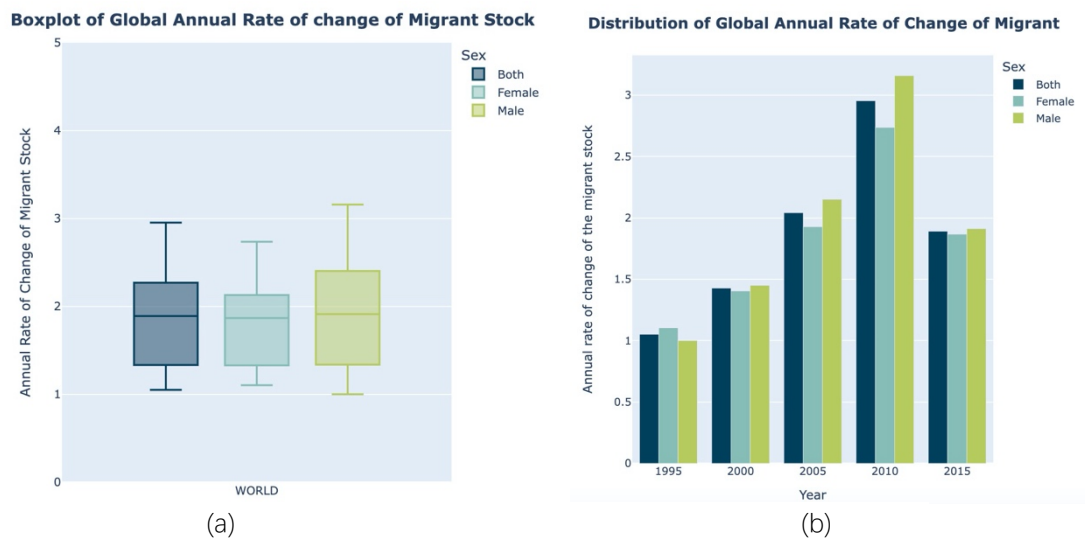


Figure 9: (a) The annual rate of change for male migrant stock varies significantly within 25 years across the sex categories. Compare to other two categories, the annual rate of change for female migrant stock seems more stable. (b) The annual rate of change increased steadily during 1995 and 2010 but fell sharply in 2015, especially for male migrant stock.

Based on Figure 9a, the annual rate of change for the male migrant stock has a wider range than the other two categories. However, the medians for all three categories are closer to each other. This suggests that the male dataset has more diversity in numbers. Figure 9b shows a fluctuating trend of the annual rate of change of the migrant stock. The annual rate of change increased exponentially between 1995 and 2010 but shapely dropped back to 2005's level in 2015.

A scatterplot was illustrated to determine the trend of the annual rate of change in each continent.

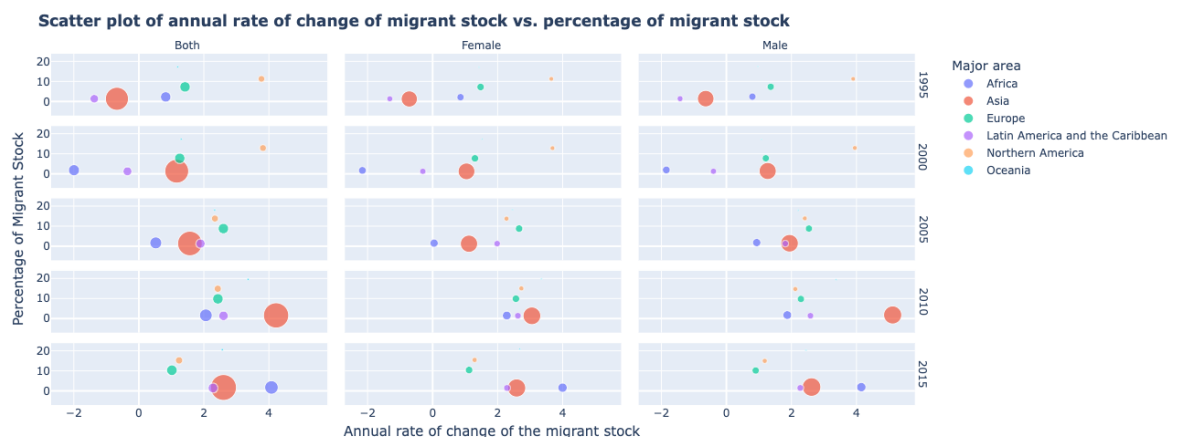


Figure 10: Scatter plot shows the trend for annual rate of change of the migrant stock from 1995 to 2015 in three difference sex categories and six continents. The size of the circle represents the population of the associated continent.

Based on Figure 10, we could see that Asia follows the same trend as the global case, which is a dramatic

drop in 2015. The annual rate of change of male migrant stock is much higher than the one for female migrant stock in 2010, but the annual rate of change of male migrant stock dropped to the same level as female migrant stock. Conversely, Africa did not experience any significant downfall in 2010. Its annual rate of change of the migrant stock rise starting from 2000 to 2015 in all sex categories. On the other hand, Europe had a periodic behaviour in its annual rate of change of migrant stock. The trend was nearly in wave shape –reached its highest point in 2005, then fell back to the level close to the starting point in 2015. Therefore, Europe had a similar annual rate of change of migrant stock in 1995 and 2015. Latin America and the Caribbean had continuous growth in the annual rate of change of the migrant stock from 1995 to 2015. In contrast, Northern America had a steady decline in the annual rate of the migrant stock during the same period.

## 2.5 Percentage of Female Migrant Stock

The percentage of female migrant stock value was calculated based on the international migrant stock. In other words, it is the proportion of female migrants in the total international migrant stock. The following graphs show the central tendency of the percentage of global female migrant stock.

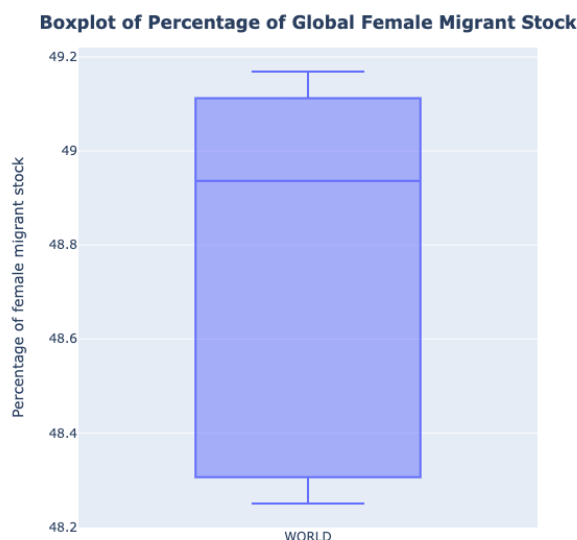


Figure 11: The median of the percentage of global female migrant stock is 48.93607, the minimum and maximum are 48.30566 and 49.16879, respectively.

The percentage for global female migrant stock did not have dramatic difference across the year. The difference between the minimum and maximum was only around 0.8%.

The growing trend of the percentage of female migrant stock in each continent was illustrated in Figure 12.



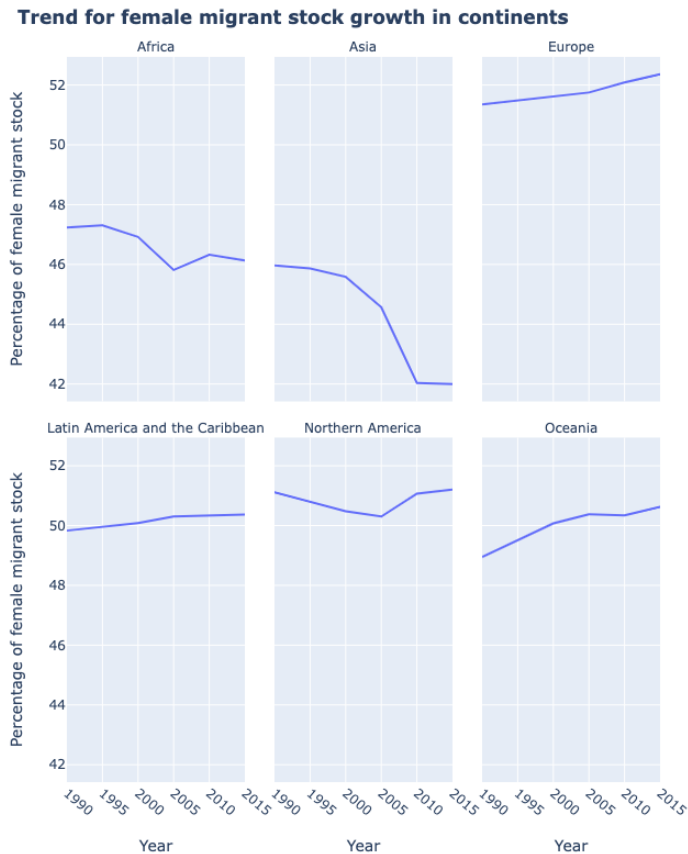


Figure 11: This plot shows a decreasing trend in Africa and Asia, and an increasing trend in Europe, Oceania and Latin America and the Caribbean. Northern America has fluctuated inclination across the years.

It could be seen from Figure 11 that both Africa and Asia had a great decline in the percentage of female migrant stock. Asia dropped almost 4% within 25 years. However, Europe and Oceania had 1 to 2% increase on the percentage of female migrant stock. Northern America had a fluctuating trend –the percentage dropped until 2005 and pull back to the original level in 2015.

## 2.6 Estimate Refugee Stock

The following boxplot visualized the descriptive statistics of the global refugee stock.

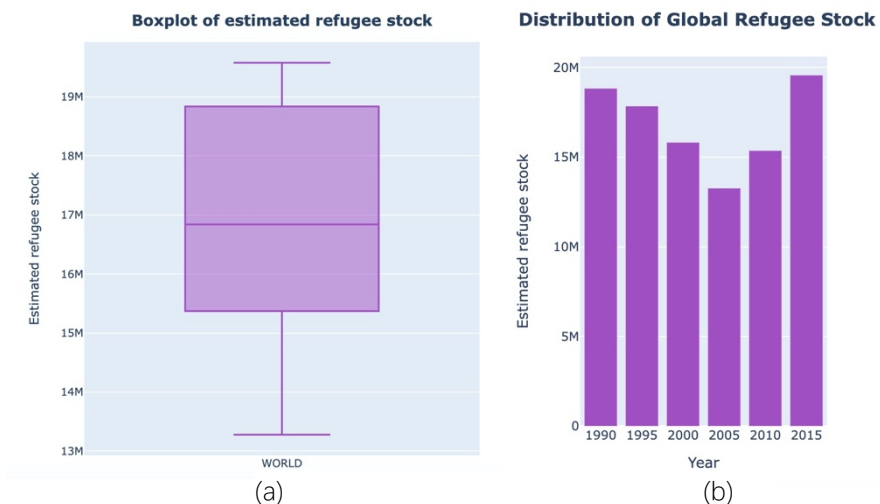


Figure 12: (a) The median of the estimated refugee stock is 16.84 million; the minimum and maximum of the refugee stock are 13.28 million and 19.28 million, respectively. (b) This bar plot shows a concave up trend in the global refugee stock.

By following the Tufte six principles, the unit of million was used to measure the estimated refugee stock. The bins in Figure 12b were made geometrically proportional to the actual numeric data of the estimated refugee stock. Each bin represented the refugee stock in the associated year. To avoid quoting data out of context, all data collected years (1990 to 2015) were included in the figure. Therefore, all six principles are fulfilled in Figure 12.

Figure 12a shows the statistics summary of the global refugee stock within 25 years. The maximum refugee stock is 19.28 million, and the minimum is 13.28 million. The details distribution of the global refugee stock can be seen in Figure 12b. The global refugee stock had a decreasing trend from 1990 to 2005, but increased again in 2010. The highest point was reached in 2015, and the lowest refugee stock happened in 2005 (13.28 million).

By applying Tufte's principles, we could visualize the data in a small-multiple way, which is by zooming in the continent aspect to see the performance of each continent.



Figure 13: The estimate refugee stock in six different continents.

Similar to Figure 12b, the continent plots included all year categories. The bins represent the count of the refugee stocks in each individual continent. Tufte's six principles are all satisfied.

It can be seen from Figure 13 that Africa, Asia and Europe have a larger number of refugee stock than

Latin America and the Caribbean, Northern America, and Oceania. Asia had the highest refugee stock among the six continents. Instead of having a clear increasing or decreasing trend, Asia's refugee stock had a fluctuating pattern. It rose gradually from 2005 to 2015, and the stock increased by 5 million. Both Africa and Europe had a decline in refugee stock, but Africa's refugee stock pulled back in 2015, while Europe did not have any change other than decreasing. Latin America and the Caribbean had a significant drop in their refugee stock in 2000 and 2005, but began to increase in 2010. Northern America followed a similar pattern as Europe but with much lower stock.

## 2.7 Refugees as a Percentage of International Migrant Stock

Since each continent has a different number of international migrant stock, only inspecting the estimated refugee stock would not represent the proportion of refugees in each continent. Therefore, Tufte's small-multiple technique was also applied to visualize the percentage of refugees in the international migrant stock so that we could determine the real proportion of refugees in the international migrant stock.

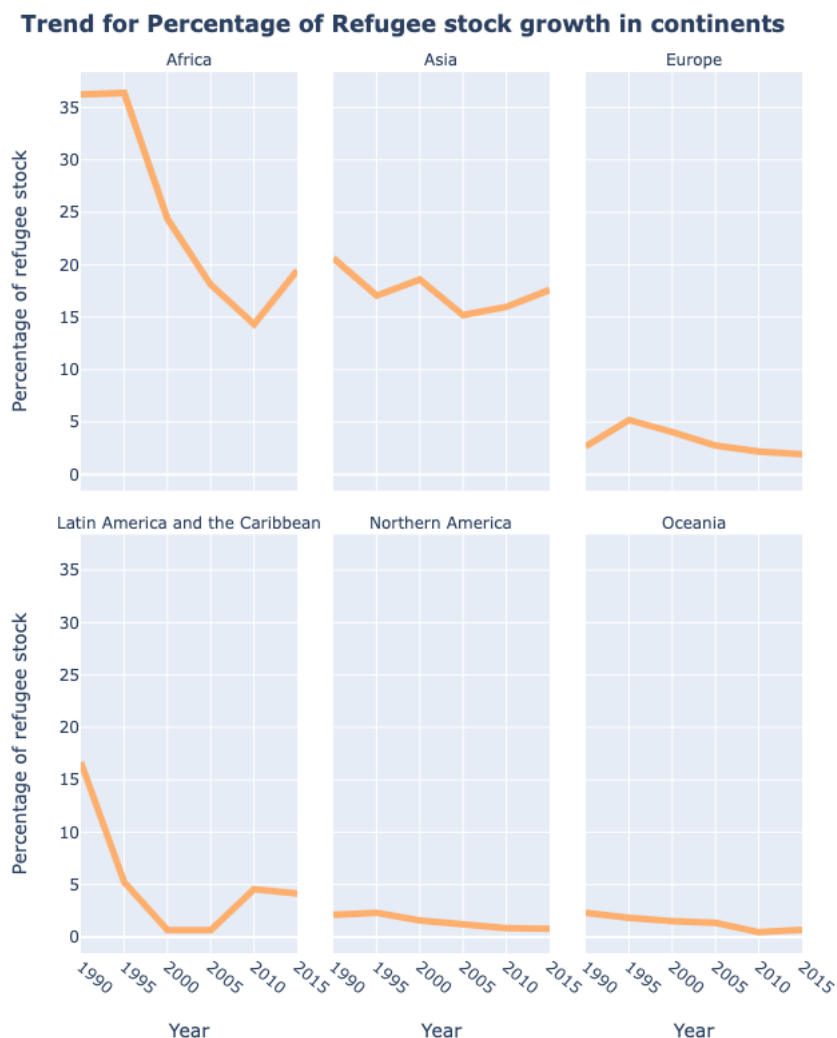


Figure 14: Trend for percentage of refugee stock in international migrant stock in six continents.

When considering the refugee stock as the proportion in international migrant stock, we got a slightly different result from Figure 13. In Figure 14, Asia no longer had the highest portion of refugee stock, but instead, Africa had the highest percentage of refugee stock among the six continents, because it had

relatively low international migrant stock compared to Asia and Europe which we can visualize in Figure 4. Unlike the performance in Figure 13, Asia had a fluctuating trend with no significant increase. Europe, Northern America and Oceania had a decreasing trend as time goes on which matched their performance in Figure 13. The latter two continents had a similar percentage of refugee stock. Latin America and the Caribbean surprisingly had a relatively high percentage of refugee stock in 1990 (16.70%), but dropped by 16% in 2000 and 2005.

## 2.8 Annual Rate of Change of the Refugee

We could also visualize the annual rate of change of the refugee stock to determine the fluctuation of the change of the refugee stock in each continent. Similar to the last section, Tufte's small multiple technique was also used to compare the annual rate of change of refugee stock in all continents.

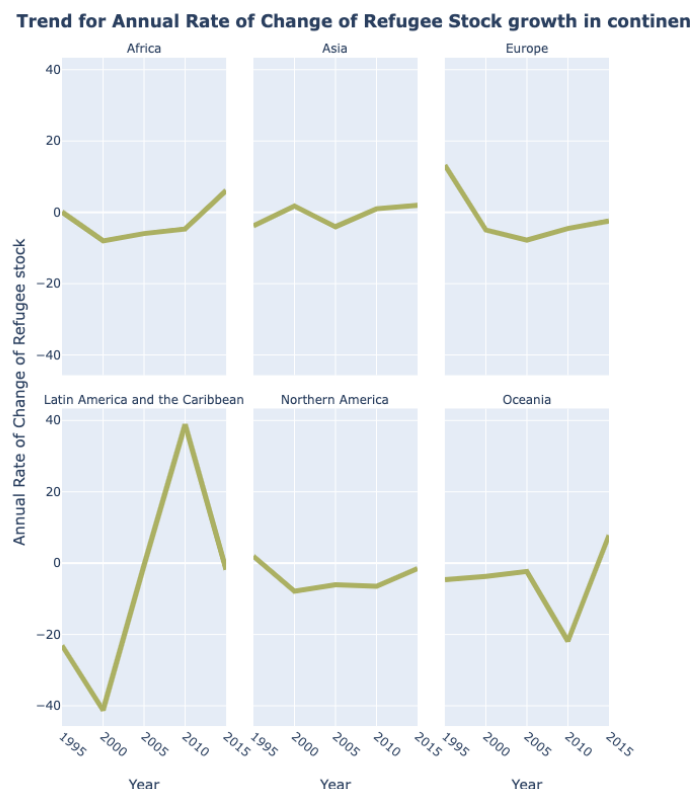


Figure 15: Trend of the annual rate of change of the refugee stock in six continents.

For keeping the data table consistent, the categories of the year were changed from interval to the upper limit of the associated interval. For example, “1990-1995” was changed to “1995”. All the interval ranges were included in the plot to avoid the violation of Tufte's sixth principle – not quote data out of context. There was no additional change made on the y-axis ticks, the patterns were also made geometrically based on the actual number, which satisfies Tufte's first principle –measurements on a graph are directly proportional to the actual numeric data. All the axes were labelled, which fulfilled the second principle. Therefore, Tufte's principles were all satisfied.

Based on Figure 15, we could see that the refugee stock in Africa, Asia and Northern America was relatively stable from 1990 to 2015. But they all experienced a decline in refugee stock compared to the previous year due to the negative number of the annual rate of change they had for some interval of years. Europe had a decline in refugee stock from 2000 to 2015 since its annual rate of change became negative from

2000. Latin America and the Caribbean had a significant increase in refugee stock between 2005 and 2010, its annual rate of change increased by 38.3. But it dropped by 37 again between 2010 and 2015. Similarly, Oceania had a drastic drop between 2005 and 2010, but pulled back in 2015.

### **3 Discussion**

Tufte's fourth principle emphasizes the importance of the unit we should use when representing money factors. Since the UN dataset does not contain any features related to money, this principle does not play an important role in this project. However, illustrating the population migrant and refugee stock, million was used as the unit due to the large count of the population.

Since the dataset does not contain any detailed information about the location, such as the region's happiness level and GDP, we were not able to explore more on the relationship between features.

### **4 Conclusion**

The UN datasets were visualized by following Tufte's six principles to determine the global performance and the contribution of each continent across the year and in different sex categories. We determine that even though in most of the features, there was an increasing trend in the global aspect, if we separate the case into continents, Latin America and the Caribbean and Oceania generally had less contribution to the number.