<h1>Introduction</h1>

Following the midterm project on cleaning the data set from United Nations (2015), the task for this project is to practice exploratory data analysis on the cleaned dataset. For the purpose of this assignment, Tufte's 6 principles on data visualization serve as the guideline for completing the task. Multiple libraries such as seaborn, numpy, and matplotlib.pyplot were used to plot the graphs. The following texts will explain how EDA was performed on the first five tables in the UN dataset and draw attention to the changes of international migrant stock and the interpretation behind such changes. To note, Table 6 will not be used for the purpose of this project. This decision will be explained in the discussion section for future study purposes.

<h1>Methods & Results</h1>

<h2>Table 1: International migrant stock at mid-year</h2>

To start this analysis, I initially generated a multi-bar chart showcasing the total international migration stock from the year 1990-2015. Two insights can be interrupted from this chart. For one, the generated chart demonstrates an increasing trend of international migration on a macro level. On the other hand, one notices a slightly higher population of male international migration stock compares to female migration stock.
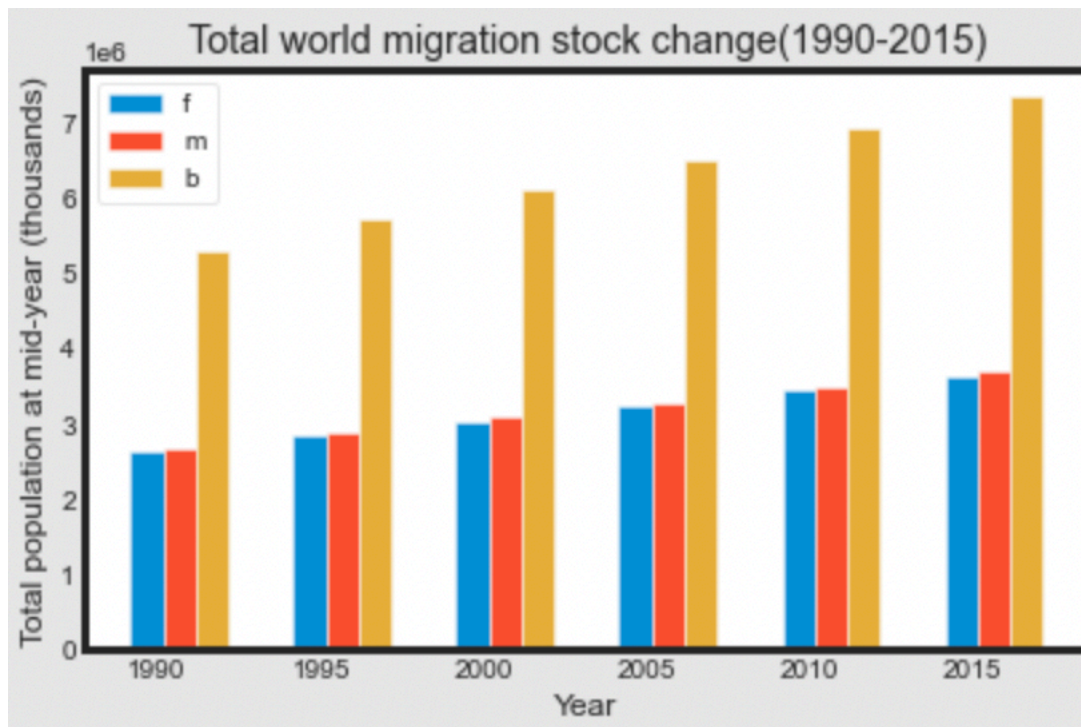
Continuing the analysis, two line plots were generated because I wanted to look at the trend of growing at a relatively micro level. Firstly, the left side graph shows the trend of international migration stock for four different regions. From the graph, one notice the statistics noticeably increases for the developed regions (blue line), whereas the least developed regions demonstrate little change in international migration stock throughout the years. Which further suggests the developed regions might have a stronger economy and higher living standards which may result in such increase. On the other hand, the right graph demonstrates the trend of growing for areas of destination, which takes an even closer look at the areas that increases drastically throughout the years and the ones are relatively stable.
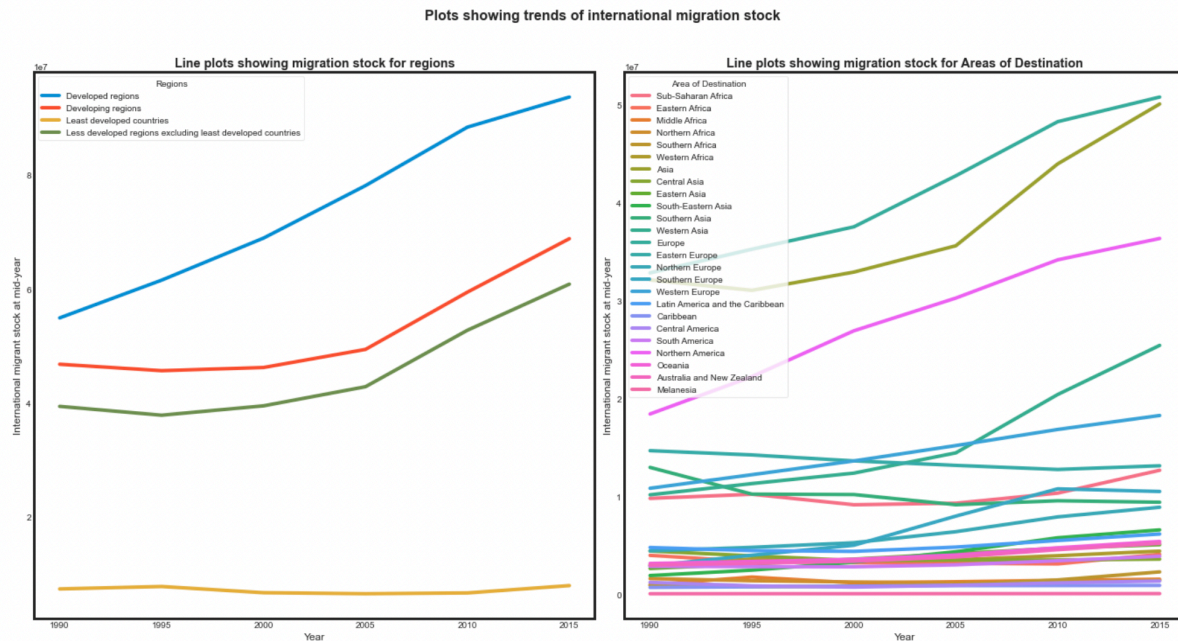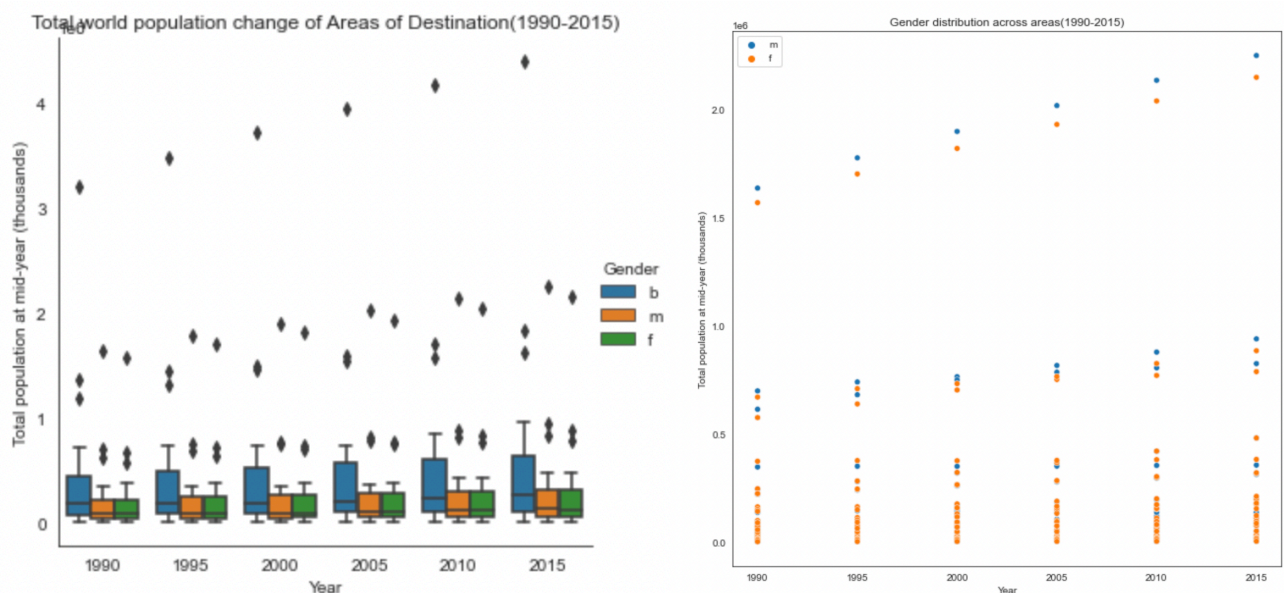

All graphs generated for table 1 closely follows **Tufte's principle 1, 2, 3, 4 & 6.** All graphs are labeled in detail and made sure to include variation of data and explained according to contexts.

Table 2: Total population at mid year

After knowing the international migration stock change, the second table provide us with a general idea on how total population changes throughout the year. For this table, I generated a

graph (left) showcasing the total population change that compares different genders across different areas of destination. Using a box plot, we can see that the total population across the areas (gender:b) is gradually increasing throughout the years because the median (middle line inside the blue box) increases. Furthermore, the population distribution among areas of destination are uneven as seen on the chart. Multiple outliers and the long upper whisker suggests many countries consists population that are much higher than the median population.

The result from the box plot can further be interpreted by a scattered plot. For this plot, I removed the both sex data and trying to compare the distribution of male and female across the areas. From the result, one can suggest that the median population of male is relatively higher than female across the areas, and the scattered plot correspond well to the box plot.


All graphs generated for table 1 closely follows **Tufte's principle 1, 2, 3, 4 & 6.** All graphs are labeled in detail and made sure to include variation of data and explained according to contexts.

Table 3: International migrant stock as a percentage of the total population

For table 3, I first generated a line graph to compare the worldwide growth of international migrant stock (percentage) by gender. From this graph, I understand the male and female growth of migrant stock appears to be in a similar pattern and speed.

Line plots showing change in migrant stock (percentage) by gender
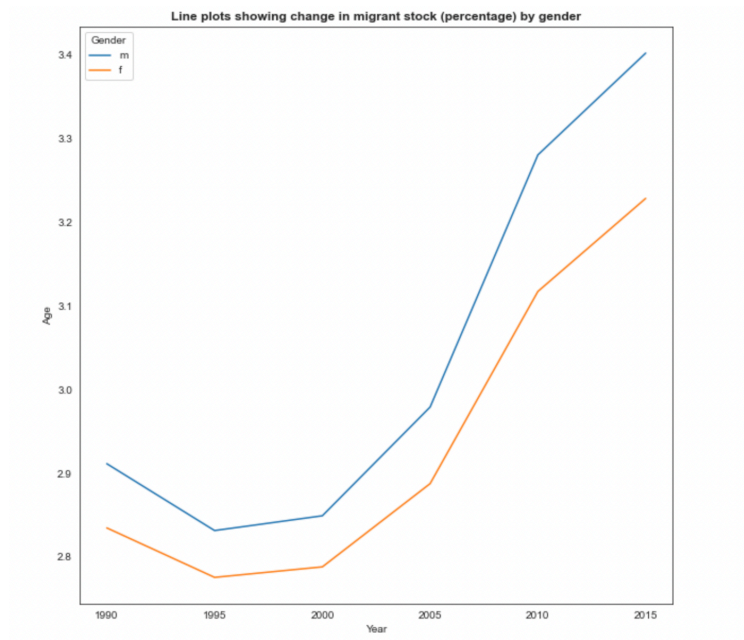
Next, I want to compare the differences between male and female when it comes to the countries with most migrant stock percentage between 1990-2015. To do so, I first calculated the

Top 10 contries with most female migrant stock percentage (1990-2015)
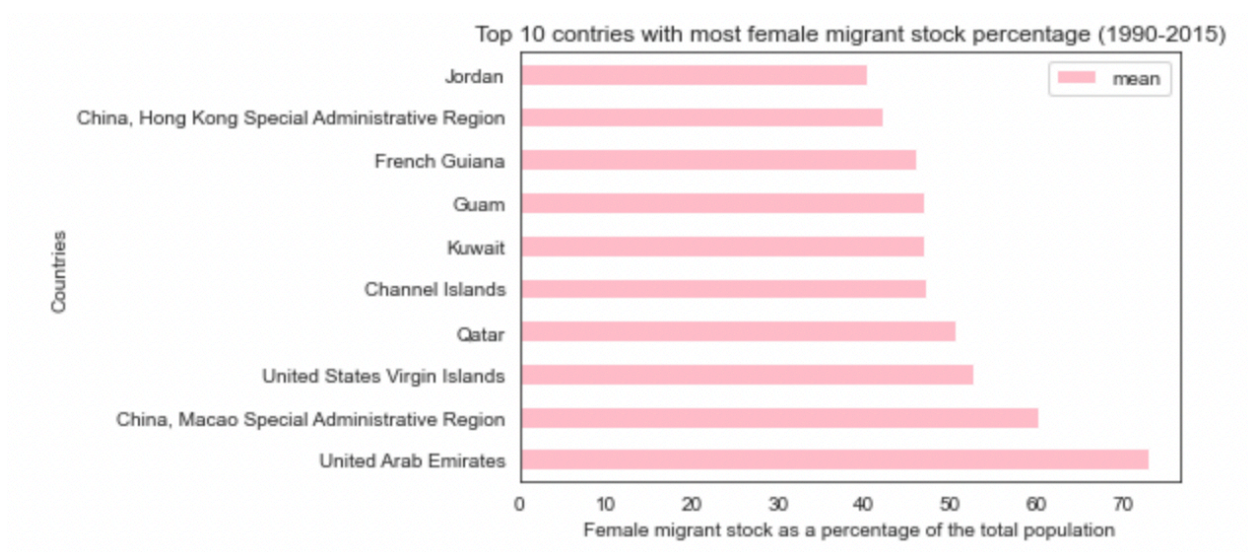
female migrant stock percentage mean value for each country from 1990-2015 and sorted them from descending order. Then, I selected the 10 countries with most female migrant stock percentage and created a horizontal bar chart.

I repeated the above step for the top 10 countries with male migrant stock percentage (of total population) and got another bar chart.
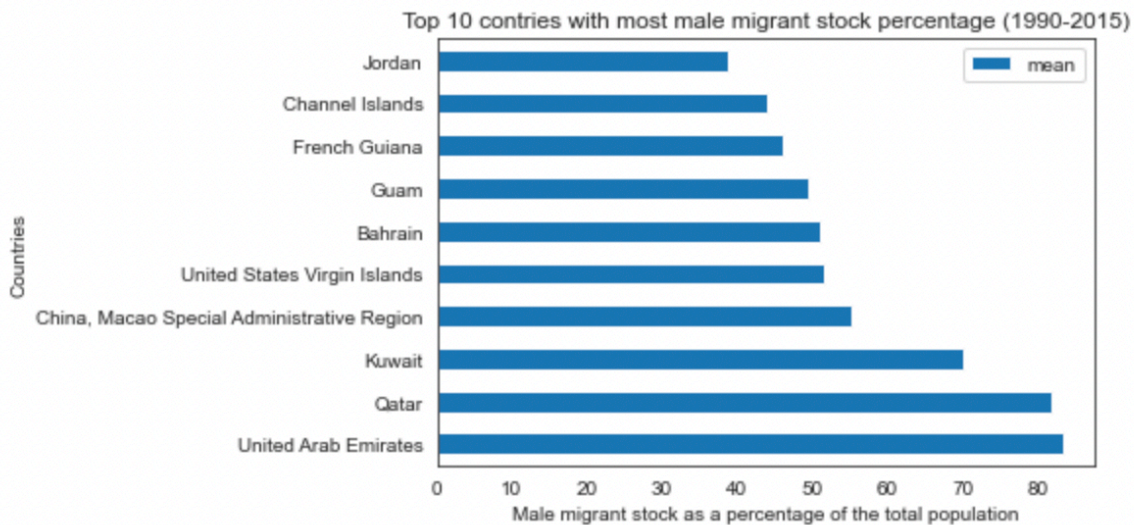

Compare the two graphs together, I can notice some difference between the countries and come to the conclusion that some countries have lesser male migrant throughout the years and some have more. However, multiple countries appears twice for each chart, which indicates they may have a significantly high migrant percentage of total population throughout the years for both male and female. This result can be further interpreted as these countries may have the history of immigration the most between the period of 1990-2015.

These graphs are also following the **Tufte's principles 1,2,3,&6** and are labeled in detail to compare gender difference data in the context of change in the length of 1990-2015.

Table 4: Female migrants as a percentage of the international migrant stock

For table 4, I want to see how female migrant stock percentage of the International migrant stock changes throughout the years. A scatter plot using the data from female migrant stock percentage across deferent areas of destination was generated for this purpose. From this graph, we can interpret that the female migrant stock percentage for majority of the areas range

around 50% of international migrant stock throughout the years. However, there is a decreasing tendency starting from 2005 in the graph which can suggest the female proportion of a population that consists of individuals who have migrated to a particular area from another location is relatively lower. To note, female migrant stock percentage in all areas are less than 55% of the international migrant stock in the year of 2015.


Female Migrant stock percentage (of international migrant stock) distribution across areas(1990-2015)
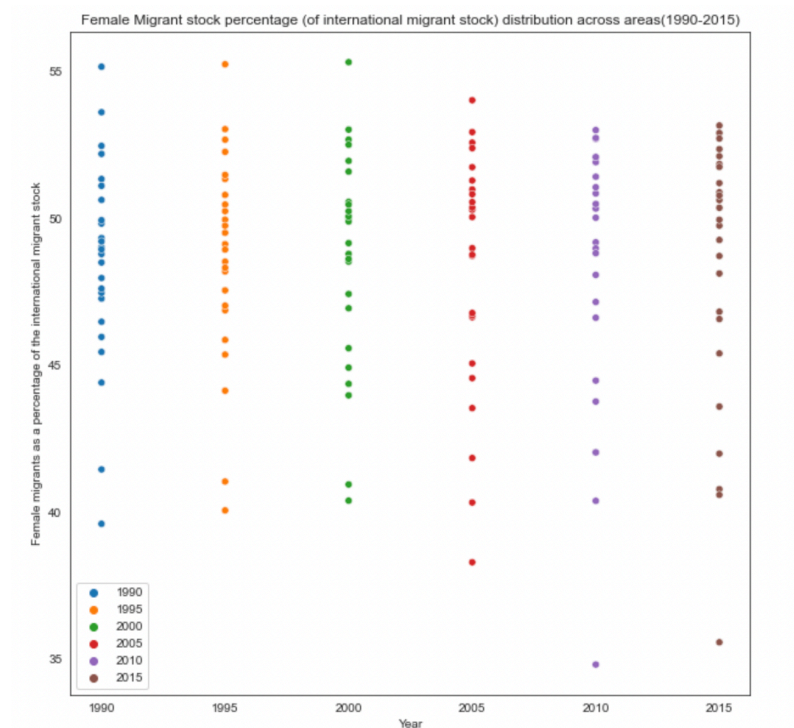
The graph generated for table 4 follows closely to **Tufte's principle 2 &6** with detailed labels and title. Moreover, the result was analyzed in respect to the context of female percentage of international migrant stock.

Table 5: Annual rate of change of the migrant stock

For the last table I will be analyzing for this project, a line plot was generated to interpret the pattern of annual rate in migrant stock for different regions (1990-2015). One can analyze this chart in comparison to the line plot generated for table 1. From this chart, one can notice a pattern of decrease in annual rate in migrant stock for developed regions starting around 2005-2010. This is also reflected in the line plot in table 1 in which the growth of international migrant stock has slowed down from the year of 2010. This information may indicate various

changes that are happening in the developed regions such as economic opportunities, political stability, or other attractive features of the area that may lead to lesser people migrant to the region. Similar trend of decrease in annual rate in migrant stock applies to the developing regions and less developed regions (excluding least developed countries). On the other hand, annual rate in migrant stock drastically increases from negative to positive in the least developed countries starting around 2000-2005. Which may indicate a positive change in its economic opportunities, political stability, or immigration policies.
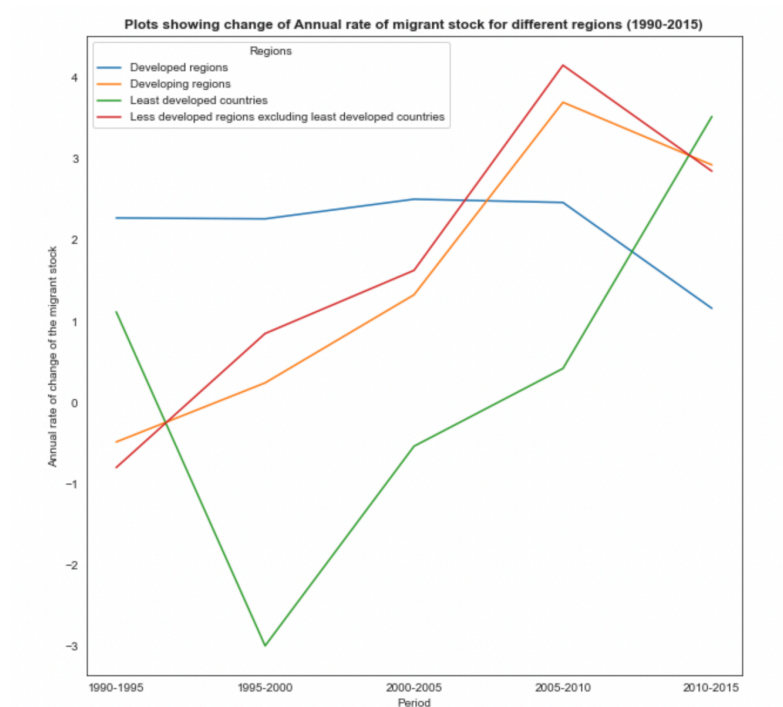
Discussion & Conclusion

In the final analysis, the five tables gives one a general idea on how different regions/ countries/areas of destination change in migrant stock throughout the years of 1990-2015. The above analysis also compares how male and female distribution across areas and the top ten countries that has the most migrant stock percentage throughout the period, which can lead to further sociological studies. We have also investigate on how male and female population distribute across the areas (1990-2015). On the other hand, we have learned the tendency of either decrease or increase in immigrant stock/ annual rate in migrant stock for different regions.

We have come to the conclusion that some regions are having less and less annual rate in migrant stock overtime whereas some regions change from negative annual rate to positive annual rate over time. From this data visualization, one can further study the various social/economical/ political factors that contribute to such result. For the propose of this study, table 6 is not included because it contains data of refugee stock in relation to migrant stock which is not the main focus of this project. However, this is important data that can be studied in the future to analyze how refugee stock contributes to the change of pattern in migrant stock on an international level, and can be further investigated with social and political disciplines.

To conclude, all charts generated for this project follow closely to Tufte's six principles on data visualization. Instead of creating various forms of visualization, I tried to include data variety instead of visualization variety. On the other hand, all numbers are made proportional to quantities and I ensured to use clear and detailed labeling for every graph. Furthermore, all nominal units are in standardized units and data are quoted in full contexts.