INF 1340
FINAL PROJECT WRITE UP

KUAN YI CHOU
1003850674

Introduction

International migrant stock is an important indicator to track in order to see the movement of human capital. The dataset “UN_Migrant Stock Total”, tracks the international migrants count between the years 1990 to 2015. There are originally 6 tables, after a careful data cleaning process, we have 8 tables with information about migrants and 1 table with information about the background of the dataset. In this paper, we will be focused on creating data visualization on the 8 tables. Table 1 tracks international migrant counts. Table 2 tracks the total population. Table 3 tracks international migrant stock as a percentage of the total population. Table 4 tracks female international migrant stock as a percentage of total international migrant stock. Table 5 tracks the annual rate of change of migrant stock. The original table 6 is separated into three different tables, table 6-1 tracks the estimated number of refugees. Table 6-2 tracks refugees as a percentage of international migrant stocks. Table 6-3 tracks the annual rate of change in refugees. All tables are separated by year (1990, 1995, 2000, 2005, 2010, and 2015), while applicable, tables are separated by gender as well (both sexes, female and male). All tables are sorted by sort order, gender if applicable, year, and finally data input. Each data entry refers to a region of interest, starting from individual countries, sub continents, continents, and world. There are economically divided regions as well such as least developed regions, developing regions and developed regions.

Method

Our main goal for the exploratory data analysis is to analyze and create visualization for the 8 dataframe mentioned above. Since the dataset has been cleaned previously, the majority of the work will be focused on creating the visualization. For the guideline of data visualization, we want to follow the work of Edward Tufte. Tufte has several principles regarding data visualization. Firstly, it is to show the comparison between groups in data. Secondly, it is to show casualty and how a variable affects another variable. Thirdly, it is to show as much data as possible. Tufte also suggested maximizing the data-ink ratio, using appropriate scales, labeling the graphs, and making the outputs clear.

Following the principles mentioned above, we import the seaborn package in python. Due to the different variety in data, we want to produce different graphs that include and represent the data mentioned. Starting with Table 1, we created a “region” list that includes the data of developed and developing regions. Using the region list, we created a boxplot that visualizes the difference between developed and developing regions. However, the table does not portray sufficient data, which does not satisfy Tufte’s principles. In order to enhance the visualization, we pivot the table so that we separate the data by different genders and year and create a barchart. Within each combination

of gender and year, there is a comparison between the international migrant stock of developed and developing regions. Lastly, we need to give the plot labels, titles and legends. The plot created satisfies multiple qualifications for a good visualization, it shows a comparison, includes as much data as possible, shows causality, and has appropriate scales, labels and titles.

Similarly, we applied the same logic whilst changing the regions to continents. Firstly, we picked out the different six continents as regions and created a bar plot. We then pivoted the data and created a plot for the variable gender and a plot for the variable year. We chose the continents over the year plot for two reasons, one is that it shows more causality through time and the second is that we already have a general idea of gender trends through time from Figure 1. After pivoting the data and finalizing the labels, title, and legend we have completed our second plot.

Moving on to the second table, we want to visualize the trend in the growth of total population in a region. We decided to first create a “sum_list” which is a total list that includes all regions that are not a country. We then derive the list of countries with the total list minus “sum_list”. Firstly, we decided to make a boxplot with the given data separated by gender, however, the boxplot only gives limited information and does not explain and change in trend since it does not include any time variable. Therefore, we decided to separate the data by the six years and created a boxplot of each year.

These are the three major ways we derive figures from the tables, further information is included within the code. Exploratory data analysis and interpretation of the tables are given below.

Results

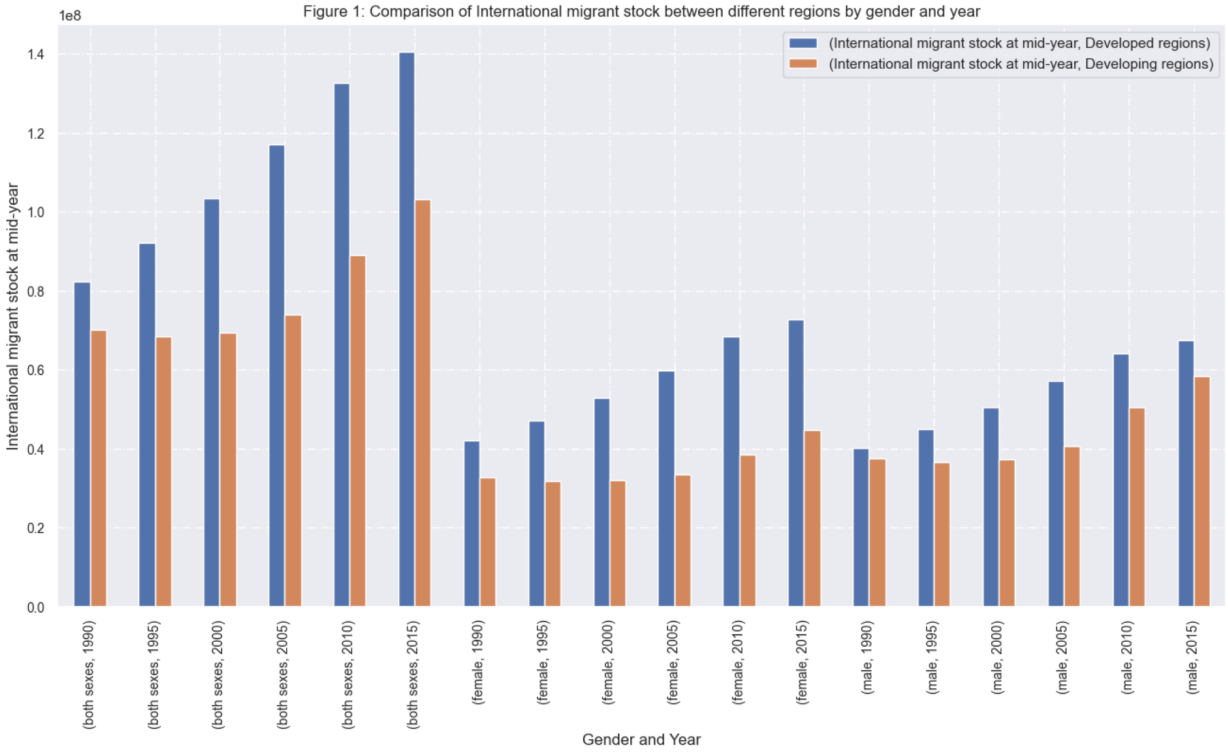


Figure 1 is a bar chart that exhibited the growth of international migrant stocks in developing and developed regions separated by gender. Our initial discovery is that there is clear growth of international migrant stock in developed regions since data is recorded. While for developing regions, growth started after 2005. Moving onto individual genders, we see a clear discrepancy between female & male. For female data, we see a bigger gap between international migrants in developed & developing regions. While for male data, the gap between developed & developing regions has been shrinking since 2005, which helps explain the increase in developing region migrant count since that year. We see a preference that females have for developed regions while a shrinking discrepancy for male data.

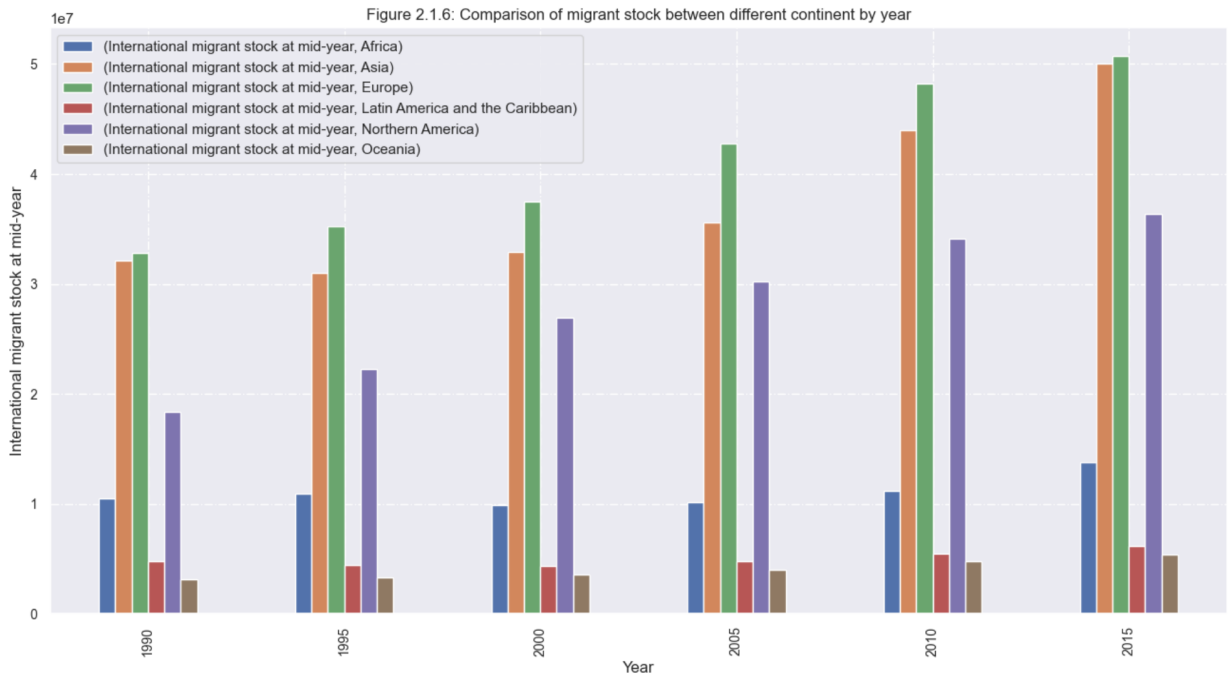


Figure 2 is a bar chart that shows the growth trend in international migrant stock between different continents through the 25 years of data collection. We see that the continents Asia, Europe and Northern America exhibit growth between the years 1990-2015, with one exception for Asia who declined only between 1990-1995. We also see that there is limited growth in the other three continents Africa, Latin America and Caribbean, and Oceania, though there is visible growth between 2010-2015 for Africa after 20 years of stagnation. We decided to include this plot because it is important to monitor the trend of movement between continents to see where people are moving to and where people are not moving to.

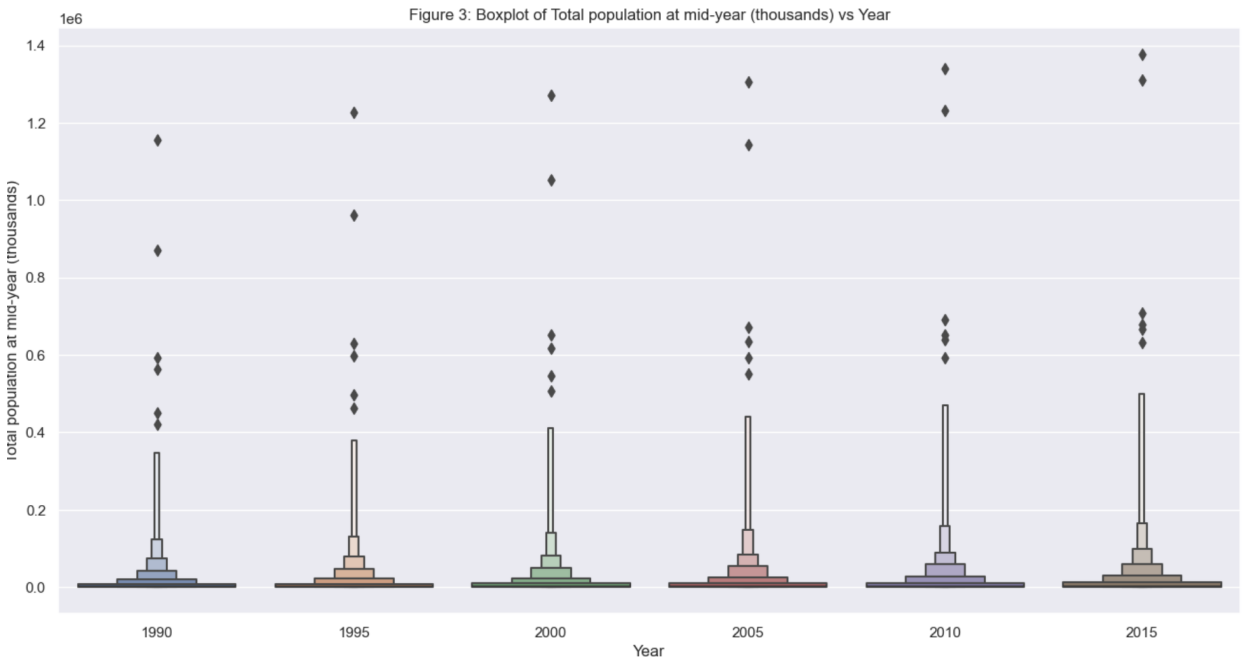


Figure 3 is a box plot of the total population of every country in the dataset. The top few data points in 2015 data are China, India, USA, Indonesia, Pakistan and Brazil. We can see an increase in both the first two quantiles which suggests that the total population increased overall. A portion of the increase may also be credited to the huge increase in population of the top 6 countries listed, especially India which almost caught up to China in terms of total population. To further investigate where the population increase should be accredited to, we move towards plot 4.

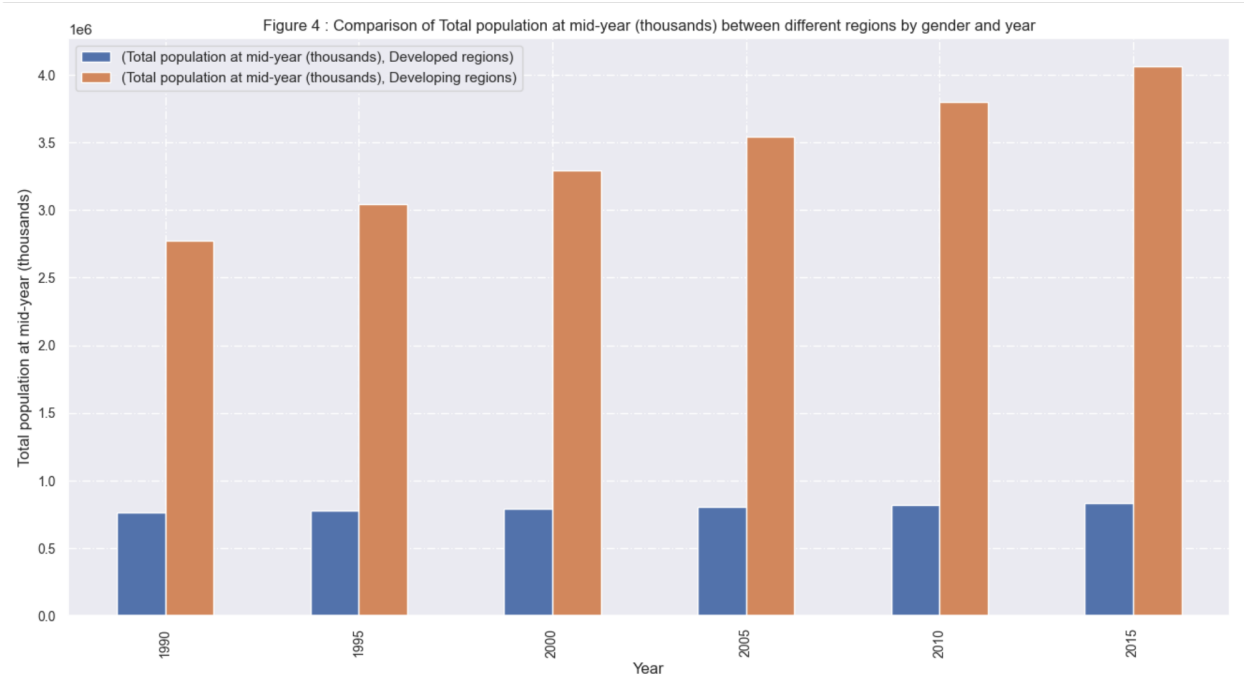


Figure 4 is a bar plot that shows the growth in total population in developing and developed regions over time. The plot describes the data by both genders only. We observe that the total population for developing regions grew significantly through the 25 years of record. While the total population of developed regions remains rather stagnant through the years in comparison to the developing regions. From previous plots, we observe a great migrant influx in developed regions, which suggests that the developed country has a low birth rate which leads to less population even though it is being injected with international migrants.

Please view Appendix Plot 1 for data on all three genders and observation on constant growth between genders.

Figure 5: Comparison of International migrant stock as a percentage of the total population between different regions by gender and year

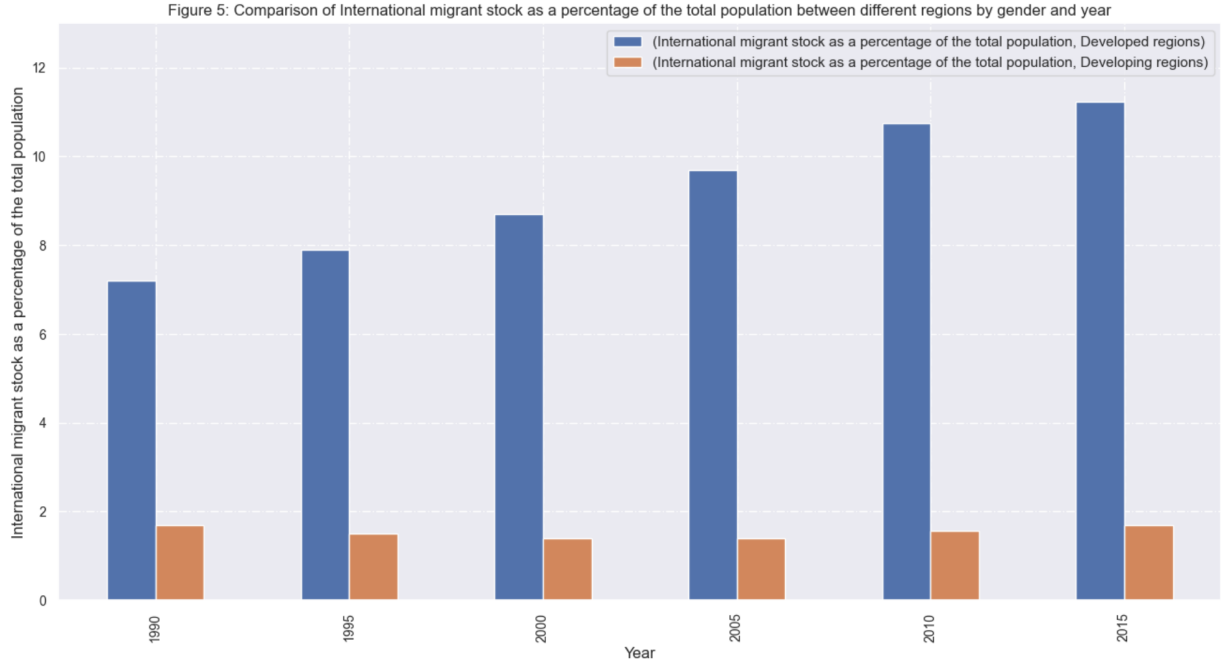


Figure 5 is a bar plot that shows the trend of international migrant stock as a percentage of total population through the 25 years of data recording. We observe that there is a significant increase in the import of international migrant stock in developed regions. Starting at around 7% in 1990, the number quickly rose to 11% in 2015. This suggests around 1 in 9 individuals in a developed country is an international migrant. We also observe a decrease in migrants as a percentage of population in developing regions between the years 1990-2005. This may be due to a local population boom or decreased foreign interest.

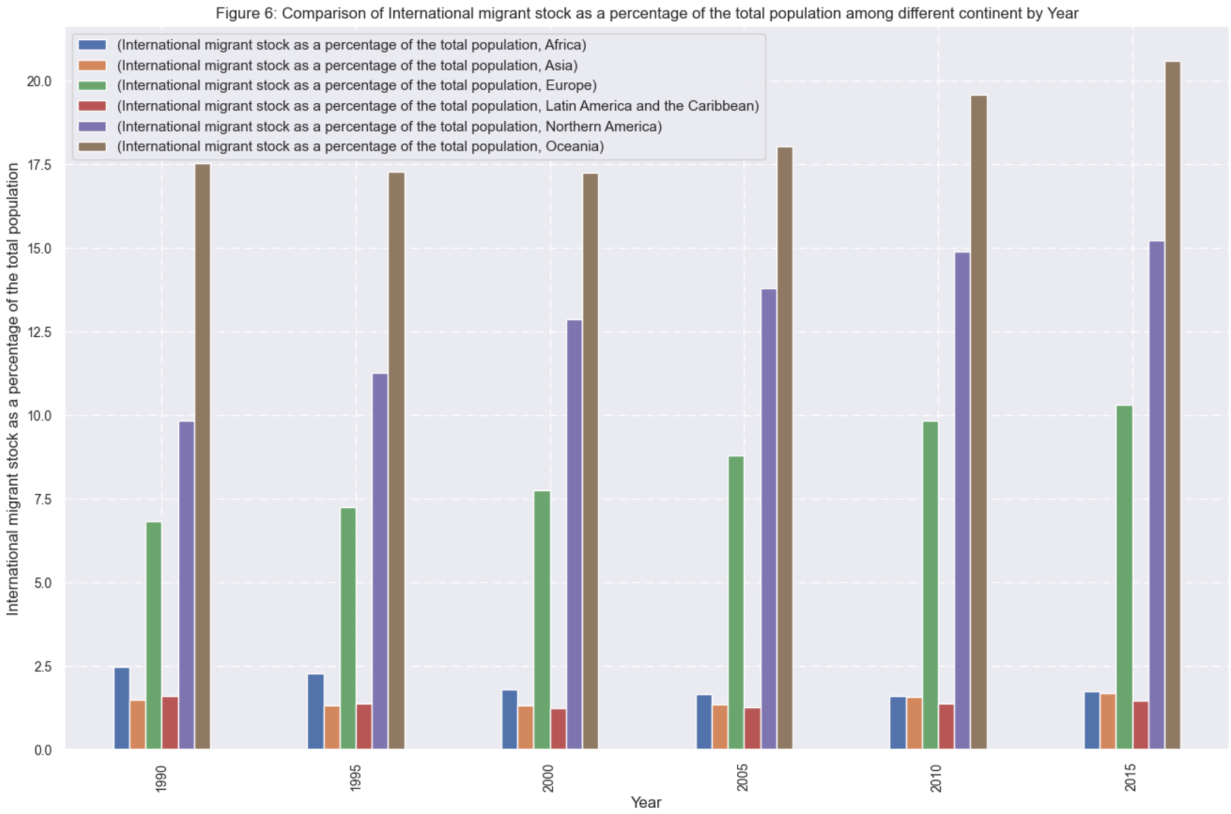


Figure 6 is a bar plot of international migrant stock as a percentage of total population. We observe a growth in the three continents Europe, North America and Oceania. Latin America and the Caribbean along with Asia appears to be neutral with slight growth while Africa seems to have declined from 2.5% to 2%. There are several factors to consider while viewing the graph, first off it is that the three regions with big growth are developed regions that have a lower amount of population growth. While the remaining regions are developing regions with high population growth. Corresponding with previous insights, although the total number of international migrants might have increased, the percentage decreased due to population boom.

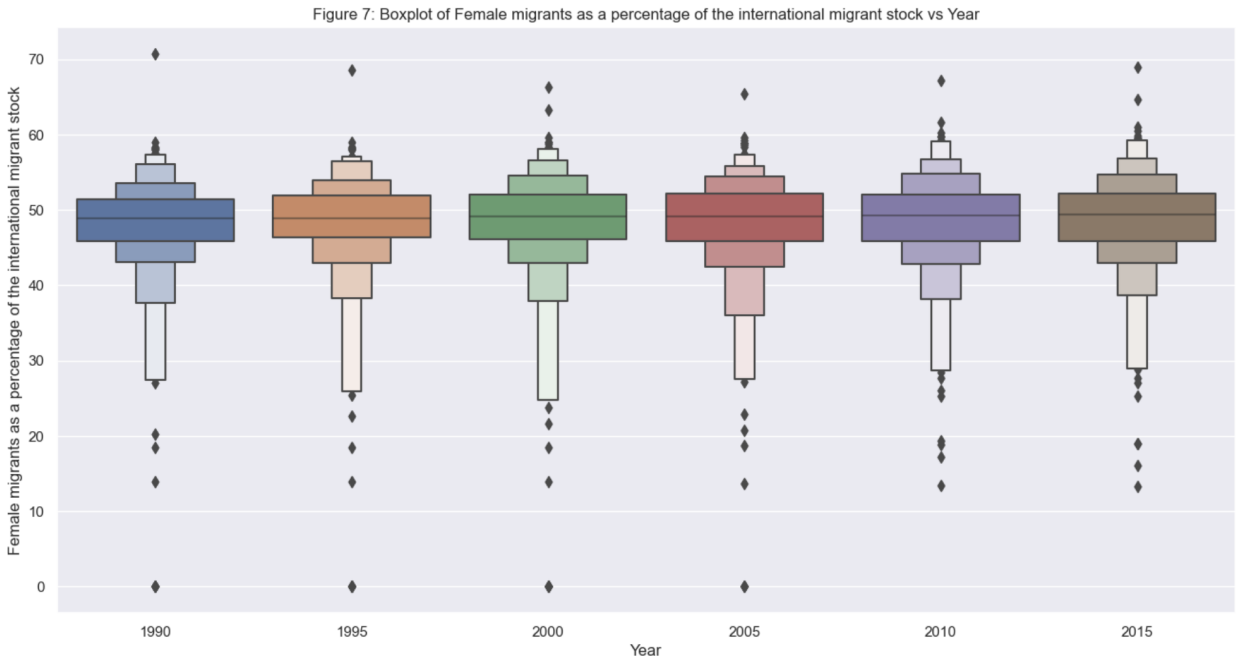


Figure 7 is a box plot of female migrants as a percentage of international migrants. We observe that the overall trend does not change drastically. The mean percentage did increase slightly over the 25 years of data recording. However, this may be due to the outlier of Sudan which has no recording of any data through the first 4 entries.

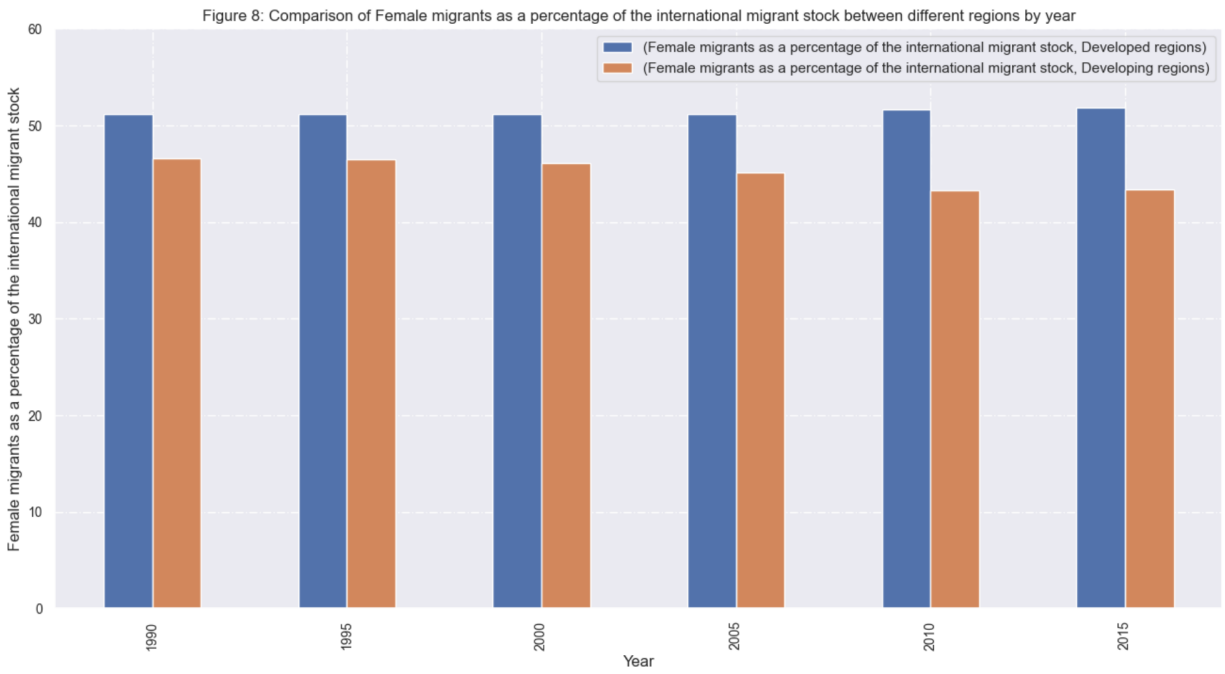


Figure 8 is a bar plot that describes the discrepancy between female migrants' choice of destination. We see that in the 1990s, there were already more female migrants than male migrants in developed regions and vice versa in developing regions. However, the percentage of female migrants in developing regions continued to decrease and the difference was 8.6% by 2015. This suggests that female migrants make up less of the migrant population in developing regions while gender is closer to equal in developed regions where females have a slight percentage advantage.

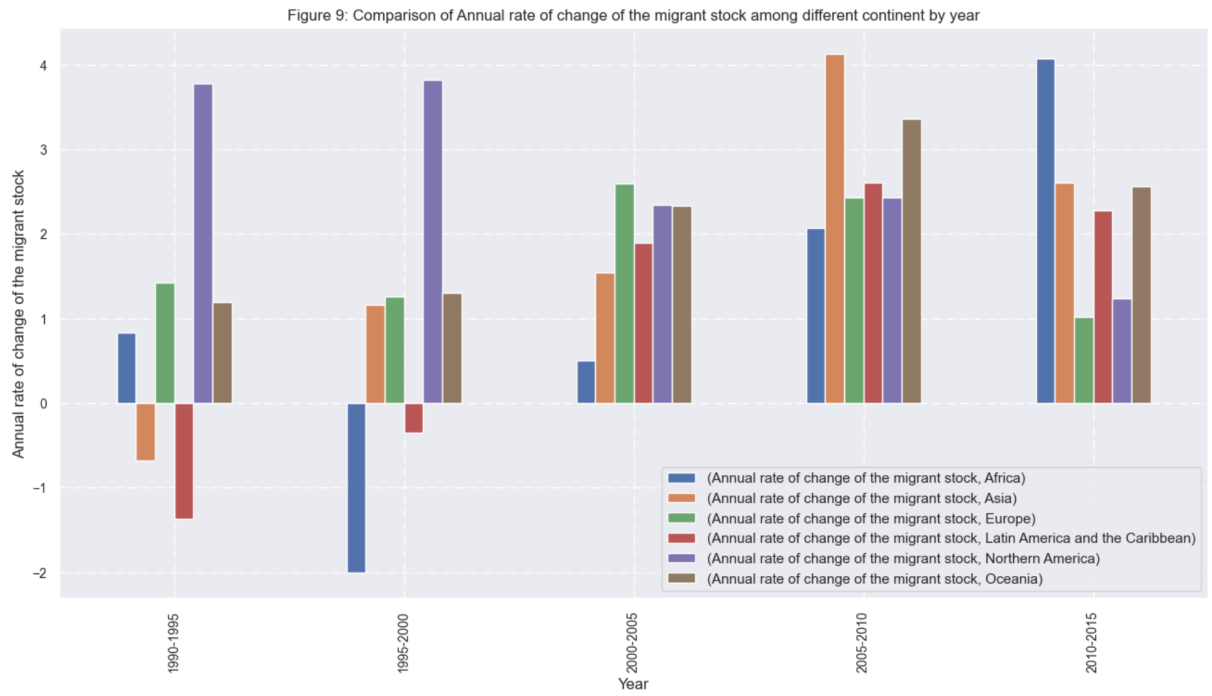


Figure 9 is a bar plot that records the annual rate of change of migrant stock among different continents. Different from previous plots, this plot shows bigger swings for continents such as Africa, Asia, and Latin America and the Caribbean. We hypothesize factors such as war, economy or political instability of major countries in the continent may alter the rate of change significantly. Overall Africa, Asia, Latin America and the Caribbean and Oceania have positive increases in annual rate of change, while Europe remains neutral and North America have a slowdown in growth.

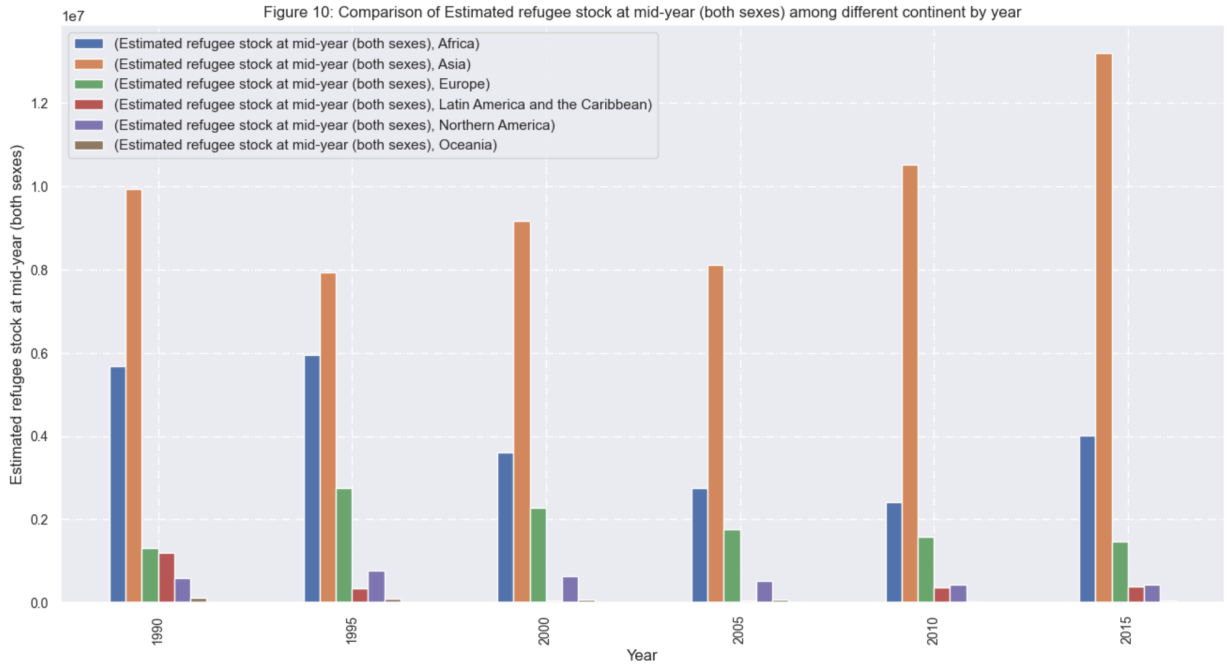


Figure 10 is a bar plot that tracks the number of refugees in the six continents between the years 1990-2015. We observe that most continents have a decreased amount of refugee intake including Africa, Latin America and the Caribbean, Northern America and Oceania. Europe experienced a small increase while Asia experienced a significant increase. The change in refugee counts may be due to increased political stability and food production or due to the reluctance of developed countries to intake refugees as shown in Figure 11 & 12.

Figure 11: Comparison of Refugees as a percentage of the international migrant stock among different continent by year

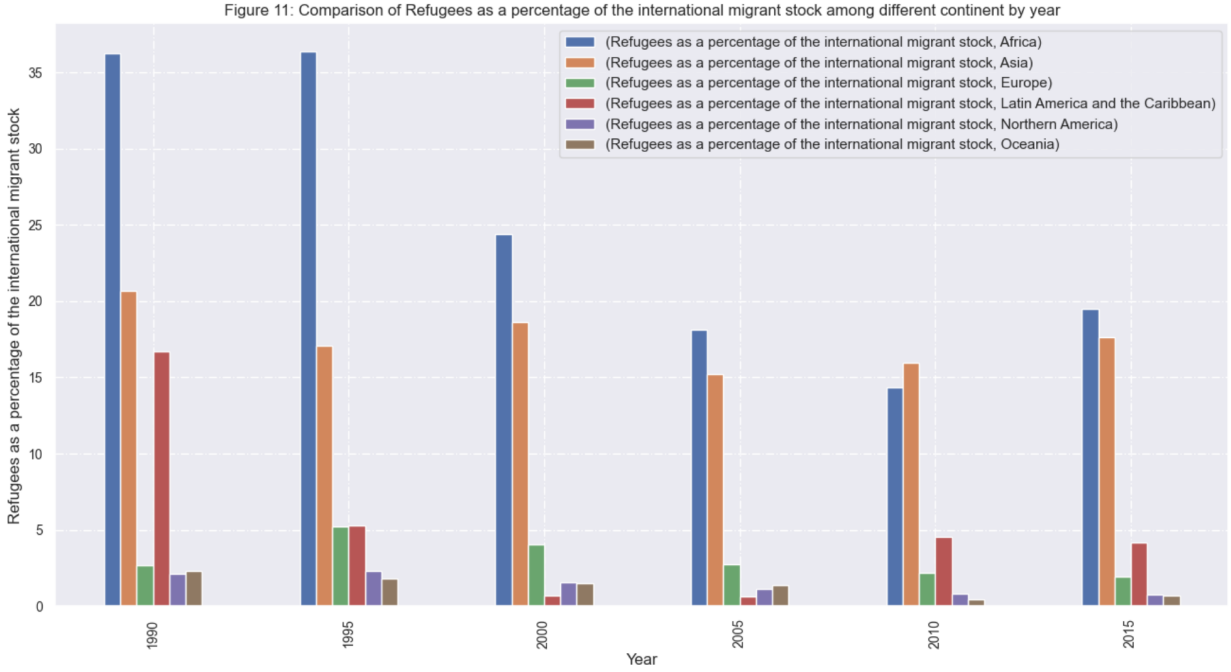


Figure 11 is a barplot that describes the number of refugees as a percentage of total migrants. As we can see, between 1990-2005, the general trend is decreasing. However, between 2005-2015, the trend is increasing for developing countries and decreasing for developed countries. From Figure 10, we observe that this is due to developed countries taking in fewer and fewer refugees.

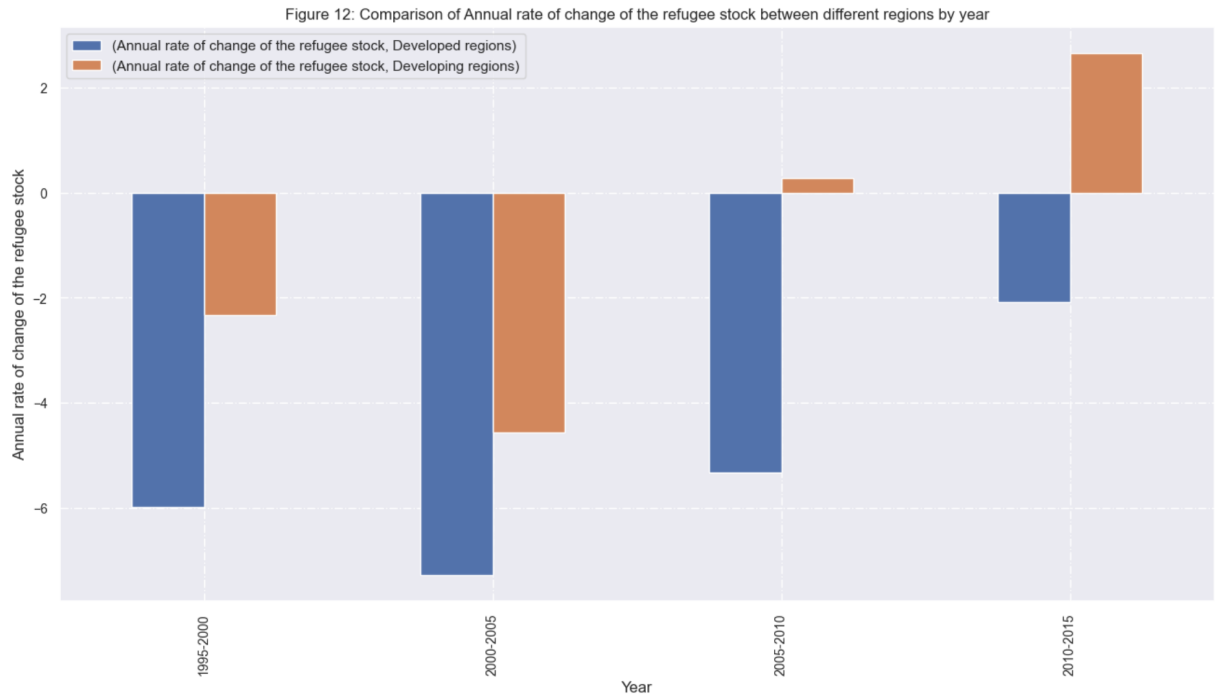


Figure 12 is a bar plot that describes the annual rate of change of refugee stocks between the developing and developed regions by year interval. We observe that developed regions have a decreasing rate of change in all intervals while developing regions started to accept a higher amount of refugees since the 2005-2010 interval.

Discussion

The 12 plots that we have prepared for the exploratory data analysis & visualization have a goal in order to visualize, describe and analyze the different trends of international migrants through the years 1990-2015. International migrant data is very important for any government organization to keep track of because it represents the flow of human capital into or out of their respective country. Developed countries often face an issue of low birth rate due to the increase of economic pressure and other social factors. Thus in result, we observe an increase in international migrants from developed countries to help mitigate the declining birth rate. Vice versa, developing regions often have higher population growth, less human capital import and more human capital export.

In our exploratory data analysis, we created 12 plots that tracked migrant stock trends by gender, region or continent. Our main observations are, continuous increase in international migrant counts, increases are greater in developed regions. Developing regions have slower or negative migrant/total population change due to high birth rate and the reluctance of moving towards a developing country. Developed countries are using migrants to replenish the slowed down birth rate and to help support the age pyramid. Total migrant counts are gender neutral, however, female migrants have a higher preference towards developed regions. As shown in Figure 9, developed regions may have reached a soft cap of international migrants intake, since we see a smaller increase in annual change of migrant stock in regions with more developed countries. We also observe that there is a decrease in the willingness of developed regions to take in refugees. Although studies show there is still a high amount of refugees, developed nations still exhibit a decline willingness in refugee intake.

We anticipate the brain drain from developing countries to developed countries to continue, and it would harm developing nations to a certain degree. Countries such as the U.S.A. are relying on the migrant “brains” from China and India to help boost their global dominant position. They are also relying on the low and medium end human capital to do day-to-day manual labor work as well. Currently, the international migrant intake is closing in on an equilibrium, with small growth in forecast. However, global political actions may alter the scenario quickly. The COVID-19 pandemic for example will have slowed down the immigration process for millions of people if not more. Thus in the near future, we may see a rapid increase in annual rate of change in migrants for 2023 in comparison to 2019-2022.

Although international migrants are a left wing ideology, we see a decrease in change of annual refugee stock in developed countries, which is rather right wing thinking. With this information, we come to a hypothesis that the main driving force behind the rise of

migrants is not internationalism but actually nationalism under its cover. Brain drain and human capital drain from developing nations will keep developing countries suppressed while adding stock and capital to developed countries' arsenal. Since migrants are all interviewed, filtered and chosen, developed nations have an idea of the social-demographic background of families that will be reporting into the country. On the contrary, it will be difficult to monitor and check the background of refugees as documentation might have been destroyed. This is not an open interrogation towards every country, as some developed nations such as Belgium, Italy and Sweden have increased refugee acceptance. The decrease in refugee intake may also be an economic burden issue as well, since refugees are usually an economic net loss while migrants are a net gain in the short term.

Conclusion & Limitation

There are several limitations in the exploratory data analysis. First of all, it is due to Figure 1. The limitation is that they violate the tidy data principle by having (gender,year) together on the x-axis. The issue could be solved easily if we separated the plot into 3 different plots. However, that will be redundant and is why we opted to use a single plot instead. Furthermore, the plot with all genders combined can show the trend and growth more clearly than if you separate them into different plots. Thus I think that the limitation is not that detrimental.

Another limitation on the report is that we focused more heavily on the general trend worldwide and did not focus on each country. The reasoning behind our decision is that we do not have sufficient data to determine WHY a country is gaining or losing migrants. Without the data to analyze why each country is undergoing change, the value of exploring trends of an individual country is weakened. Albeit it will be an interesting topic and would not have been a bad idea to explore the trend in individual countries. We reckon that having a big data overview is better since we will be able to utilize the “Notes” section with developed & developing regions and continents.

Critically and from the data given from the dataset, we observe that the brain drain on developing nations is true and developed nations are being more and more selective on the migrants import with slower yet positive growth. We will be able to generate multiple different and interesting topics from the given findings above. We can stand in a point of view of a developed country's government and determine the immigration policy, how immigration positively impacts economic growth and how immigrants may increase or decrease job opportunities for locals. We can also stand in the point of view of a developing nation, while experiencing big growth in population, they are losing top talents to developed nations. Do international migrants eventually move back to their

birth nation or do they tend to settle in their new home. What benefit do they bring to the developing country they are from after they leave the country? Considering findings, should we try to limit the loss of human capital or are they bringing home more wealth than they can create at home? From a macro and humanitarian standpoint, we are concerned about the lowered willingness that developed countries such as Europe, North America, and Japan have on refugee intake.

REFERENCE

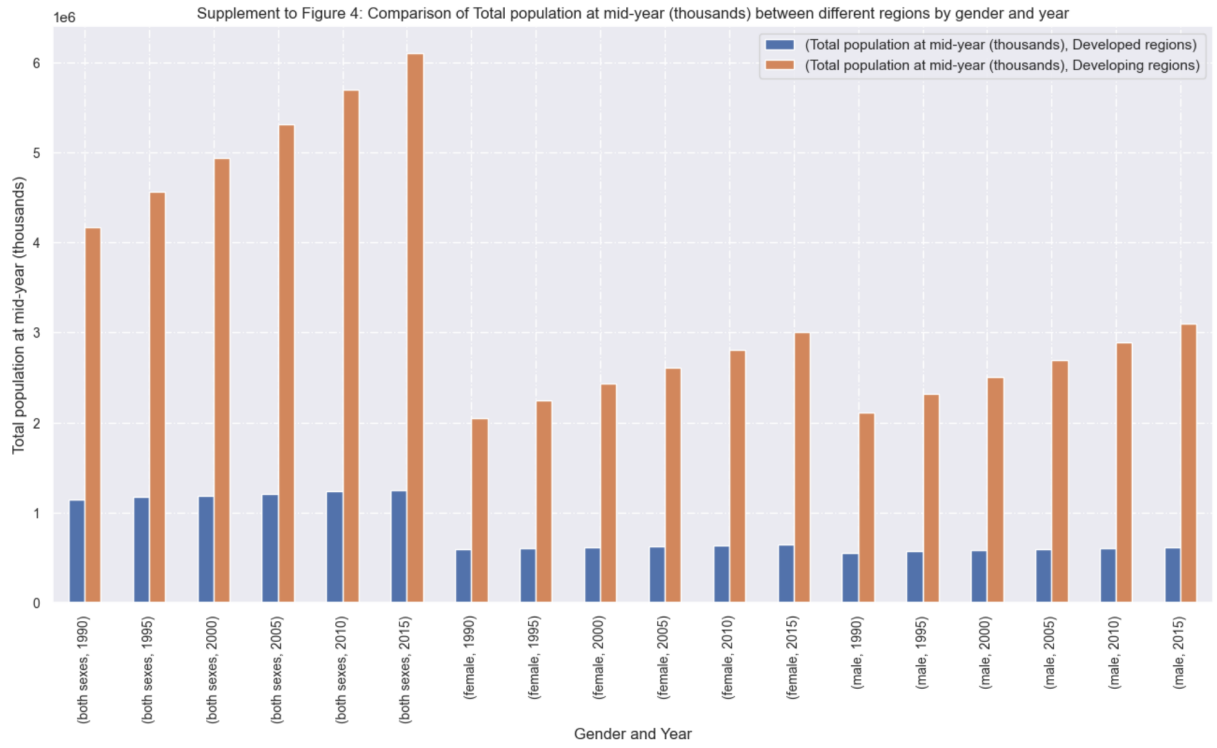
Docquier, F. (2014, May 1). *The brain drain from developing countries*. IZA World of Labor. Retrieved December 15, 2022, from https://wol.iza.org/articles/brain-drain-from-developing-countries/long#:~:text=Motivation

Geoba.se. (n.d.). *Gazetteer - the world - population - top 100+ by country (2022)*. Geoba.se: Gazetteer - The World at Your Fingertips. Retrieved December 15, 2022, from http://www.geoba.se/population.php?pc=world&type=28

World Refugee Statistics 1990-2022. MacroTrends. (n.d.). Retrieved December 15, 2022, from https://www.macrotrends.net/countries/WLD/world/refugee-statistics

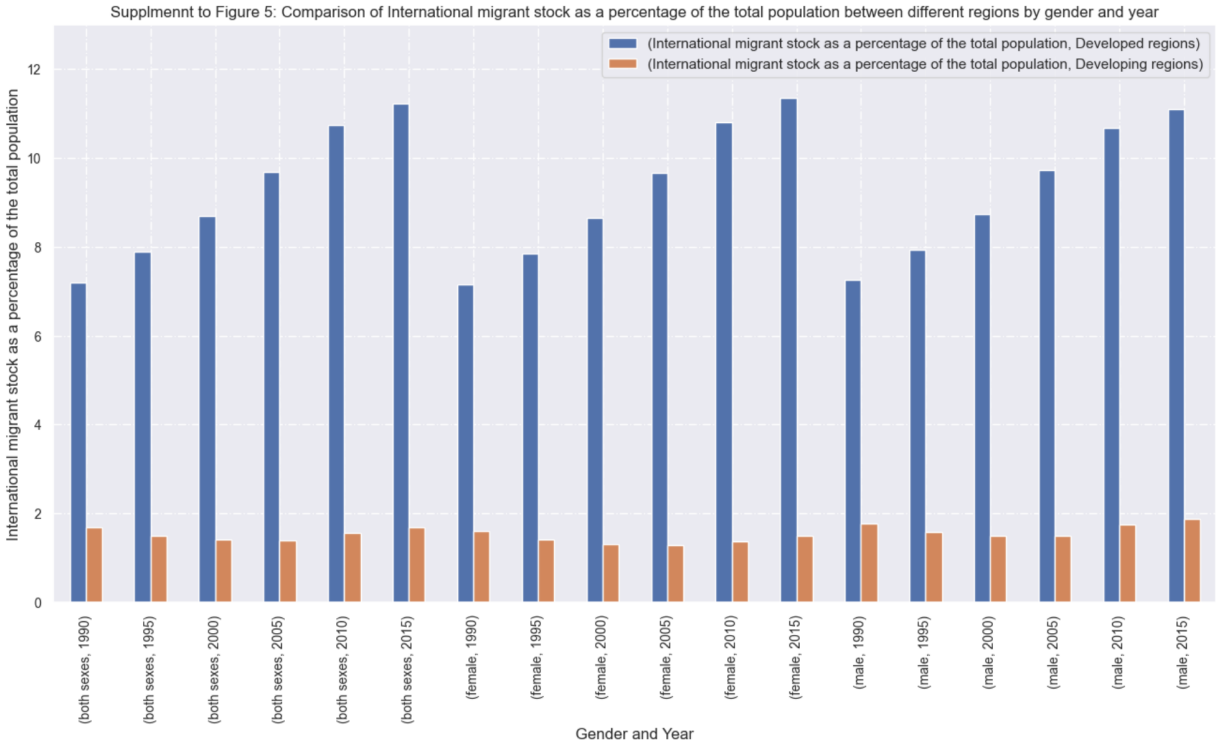
APPENDIX

Appendix plot 1



Supplement material to figure 4. We see constant growth in both genders thus to uphold tidy data principle we only included the both gender section.

Appendix Plot 2



Supplement for Figure 5. We observe that there is a similar trend between genders thus to uphold tidy data principle we only included the both gender section.