EXPLORATORY DATA ANALYSIS AND VISUALIZATIONS ON THE UNITED NATIONS TRENDS IN INTERNATIONAL MIGRANT STOCK: THE 2015 REVISION

Karishma Savani

INF1340- FINAL ASSIGNMENT

1002147731

Introduction

The final project is a continuation of the midterm project whereby I cleaned the '*United Nations Trends in International Migrant Stock: The 2015 Revision*' dataset using 'Tidy Data Principles'. For the final project, I will conduct Exploratory Data Analysis (EDA) and Visualizations using Tukey and Tufte's principles. Tukey's principles include sorting, grouping, comparing and subseting the data to easily and explore and analyze the data. Tufte's principles include reducing chart junk and separating the data in small multiples for better and cleaner visualizations. I will be fulfilling these principles using Python and various libraries available with Python such as numpy, pandas, seaborn, matplotlib, and plotly. The current data set is divided into 6 different tables with some information that is overlapping and some that can be compared between the tables, which will assist in providing a wholesome analysis of the visualizations and data and contribute to future predictions and strategies without conducting lots of detailed statistical analyses.

Methods and Results

In order to do the visualizations, I had to sort, compare, group and subset the data since the data was too crowded with a lot of information that I did not think was relevant to analyze the data. I grouped the data by major areas which include Asia, Africa, Europe, Latin America and the Caribbean, Northern America, and Oceania. This ensured that the countries in these major areas were also accounted for and gave an overarching view of change of data by the location. I sorted and compared the data to only have male and females when comparing for the two, or only both sexes when looking at the population data and for all three to compare which had more or less. Additionally, I subsetted the data by removing columns such as notes, type of data, sort and in cases when looking at females and males I removed both sexes and vie versa. I also combined data across tables to analyze data and gather predictions for the International Migrant stock.

In order to reduce the chartjunk, I ensured the visualizations are minimal and not overcrowded with too much information such as only using major areas (as mentioned above) for the location. Where I felt it necessary, I had included more information but overall was either comparing genders or variables. The colors used are simple and easy to discern differences in the comparisons. I also used small multiples where I took subparts and broke them into individual graphs to show all the facets. These will be shown below.

I started the visualizations by showing the international migrant stock around the world to get a rough idea of the difference in gender from 1990-2015.

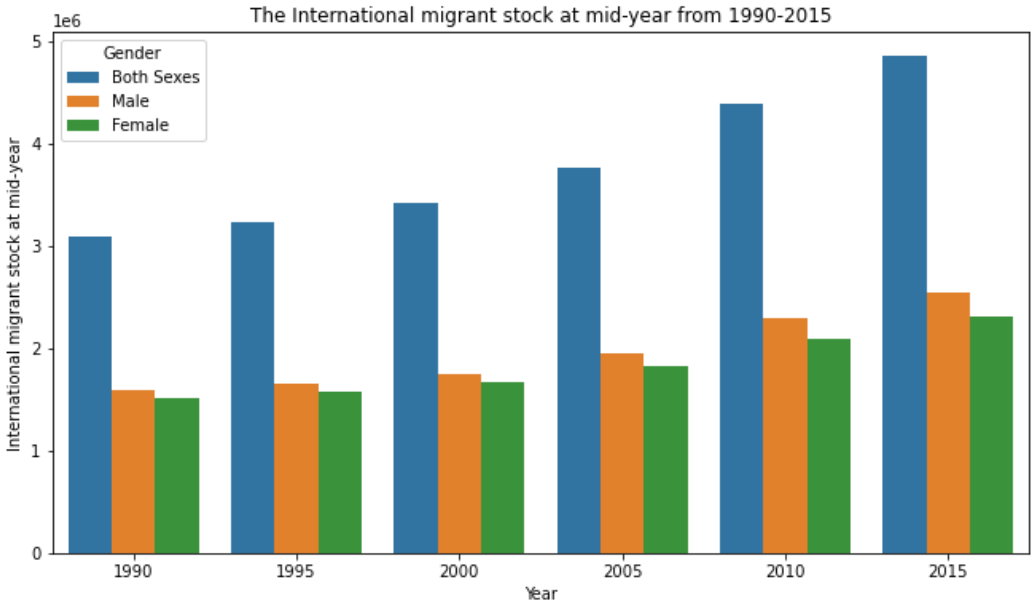

Figure 1: Comparison of Both sexes, Male and Female international migrant stock at mid-year from 1990-2015 around the World

This bar graph shows us that over the year there has been an increase in the international migrant stock around the world, in which men are increasing more that women (despite an increase in both genders). This shows us and the UN that it can be predicted that there will be more international migrant men in the coming years than women. This graph is also clean with clean bars and clean colors eliminating any chart junk.

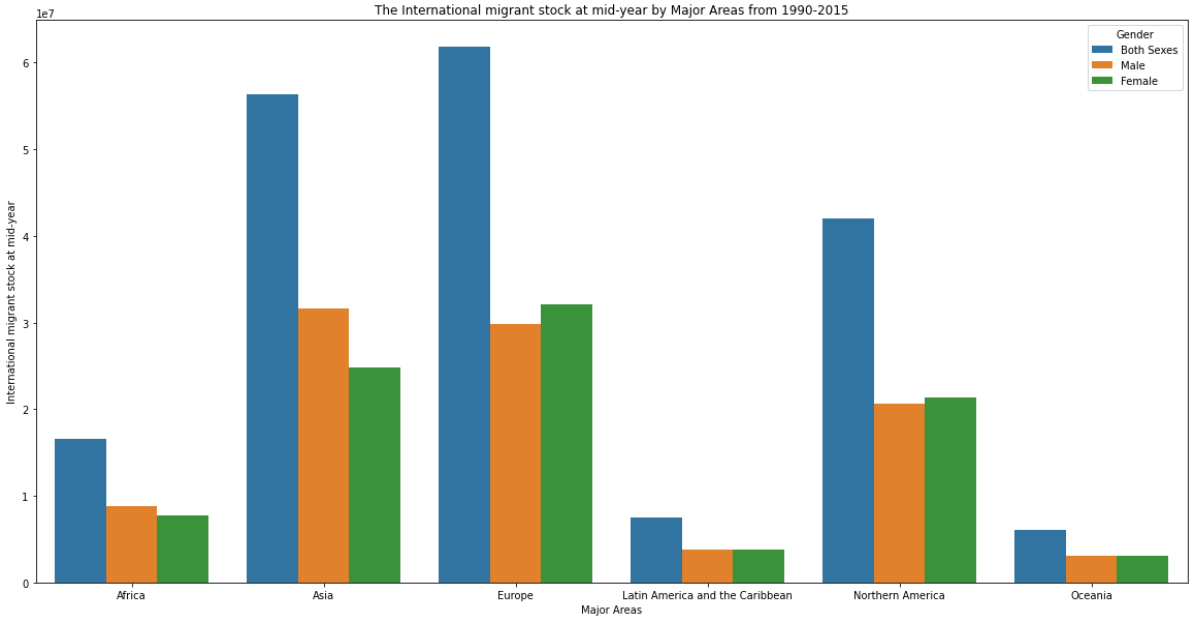

Figure 2: *Comparison of Both sexes, Male and Female international migrant stock at mid-year from 1990-2015 in Major Areas*

This second bar graph shows us in which major areas around the world the population has gone up. We can see that Asia and Europe have the highest increase in international migrant stock and third being Nothern America. This gives perspective on the previous graph and in this graph, we can see that the males not consistently more than women in all major areas except Africa and Asia which could also be due to the culture in those areas. This graph is also clean with clean bars and clean colors eliminating any chart junk.

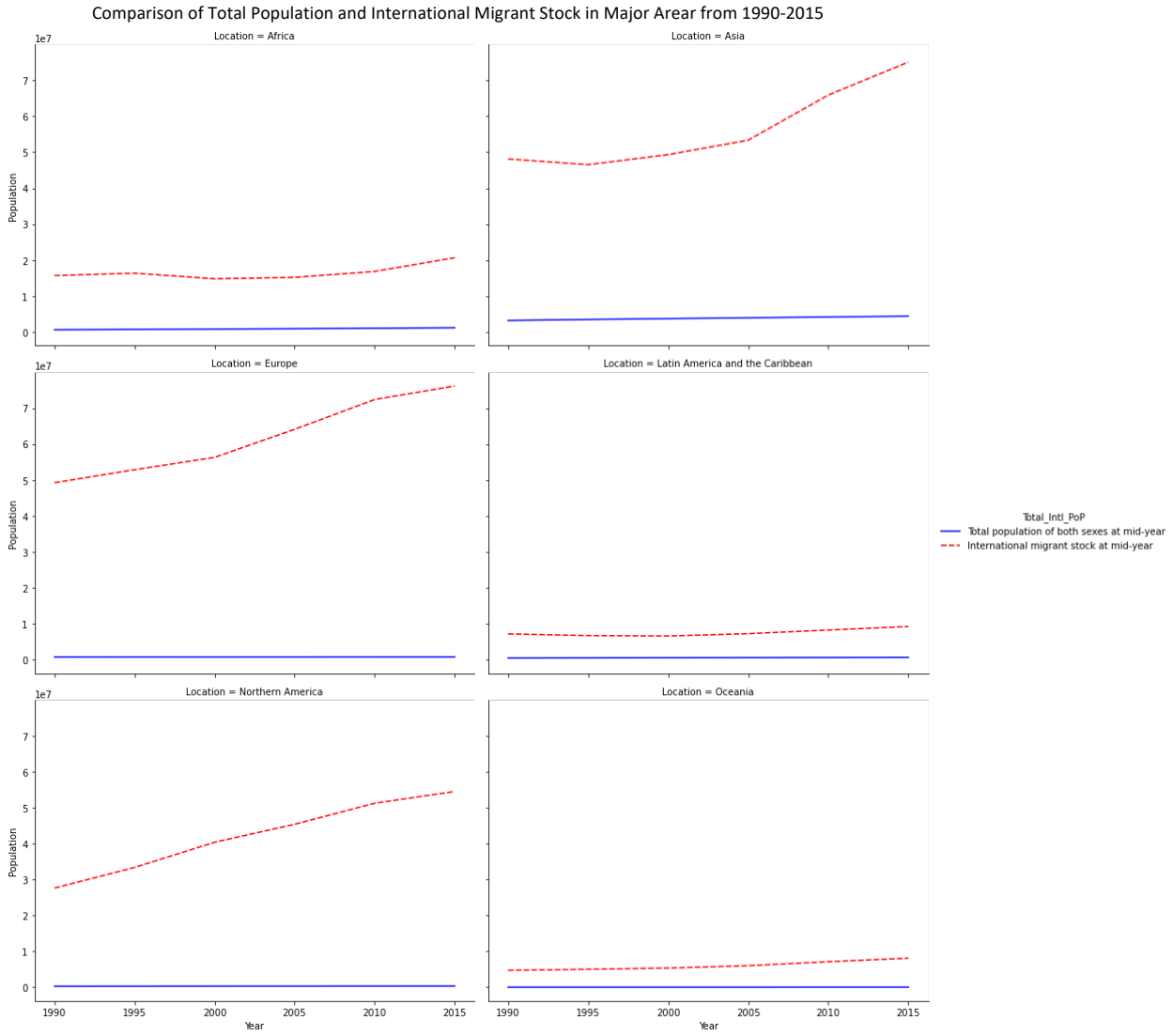

Figure 3: Comparison of Total Population and the International Migrant Stock in Major Areas from 1990-2015

This is a relplot from the seaborn library which shows the relationships between 2 variables with subsets. In this graph we can see the 6 major areas and can compare the total population from table 2 vs the international migrant stock from table 1 to get an idea of how many are international migrants from the whole population. From the graphs, which also show small multiples, we can see that in Asia, Europe and Northern America the international population is

much higher across the years and the total population is almost the same across the years, with Asia's total population and international migrant stock being the highest. This can help the UN understand and prepare for an increased number of international migrants in Asia, Europe, and Northern America.

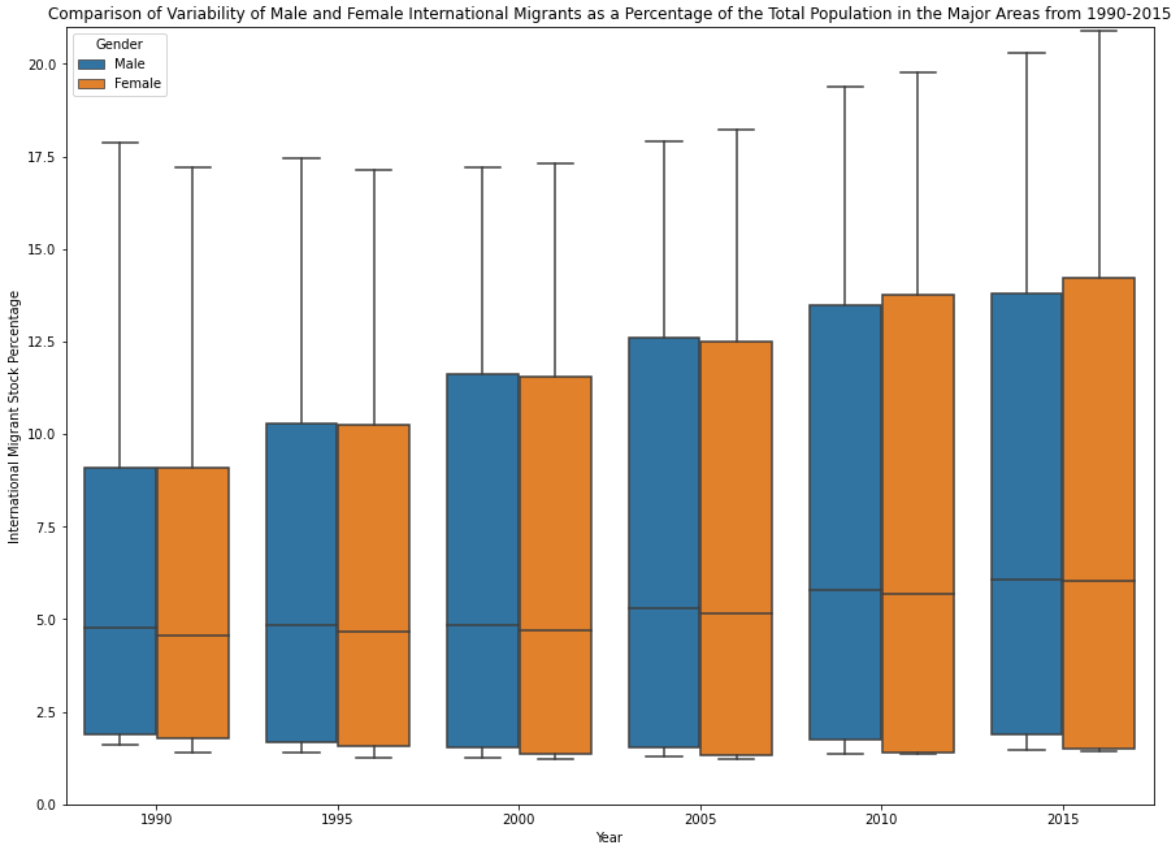

Figure 4: Boxplot showing the Comparison of Variability of Male and Female International Migrants as a Percentage of the Total Population in the Major Areas from 1990-2015

This boxplot is from table 3 comparing male and females in major areas from 1990-2015. We can ascertain from this boxplot that the variability in the male and female international migrants increases as the years go by and the data is mostly positively skewed with the median (middle line) leaning towards the minimum- lower end of the quartile. This suggests that there has been a lot of variability every 5 years in the spread of the international migrants in the different major areas. Majority are towards the lower end but the long Q3 whiskers suggest some countries may proportions that are disabling a normal distribution from occurring. This shows there are more and more countries where the international migrant population is expanding.

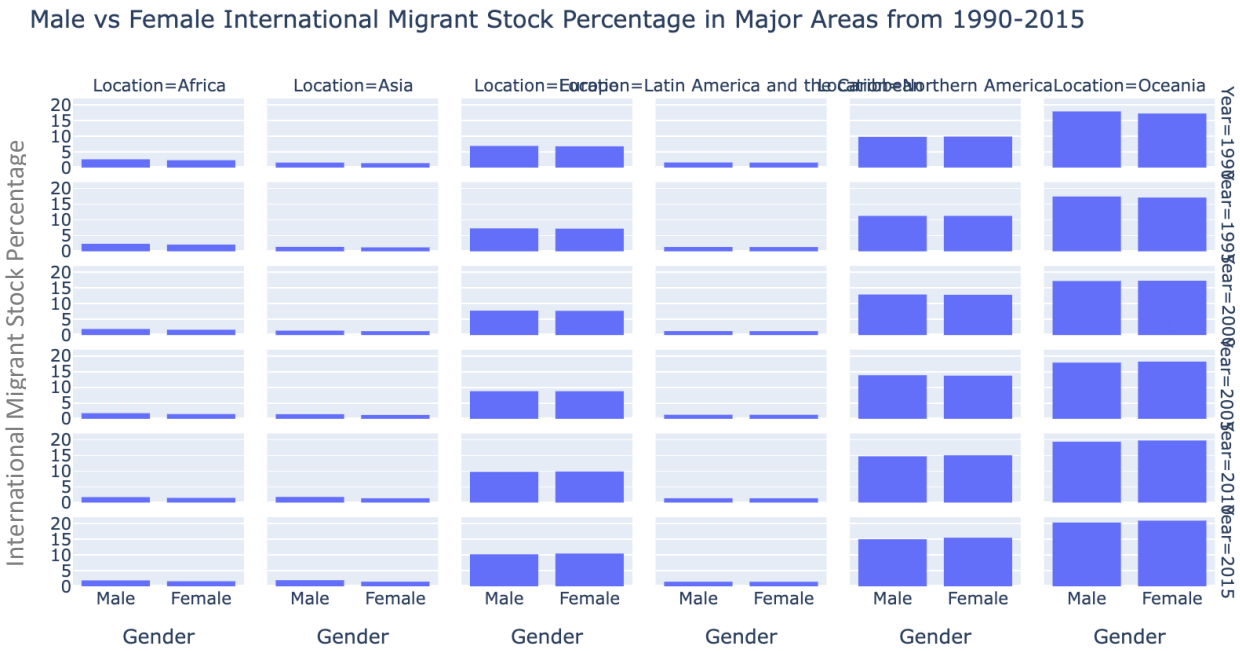

Figure 5: Histogram Showing the Comparison Between Male and Females International Migrant Stock Percentage in Major Areas from 1990-2015

This histogram is different from figure 2 because this is a percentage of the male and female international migrants as a percentage of the total population. Therefore, this tells us how many international migrants were there divided by the total population and gives us a better idea overall of the difference. From this graph we can see that Oceania male and female international migrant stock is the highest year over year and Africa, Asia and Latin America and Caribbean is the lowest. This shows that the international migrant population is increasing compared to the total population and in the last year females more than male. This histogram is split into small multiples, where we can assess by major area, year, and gender.

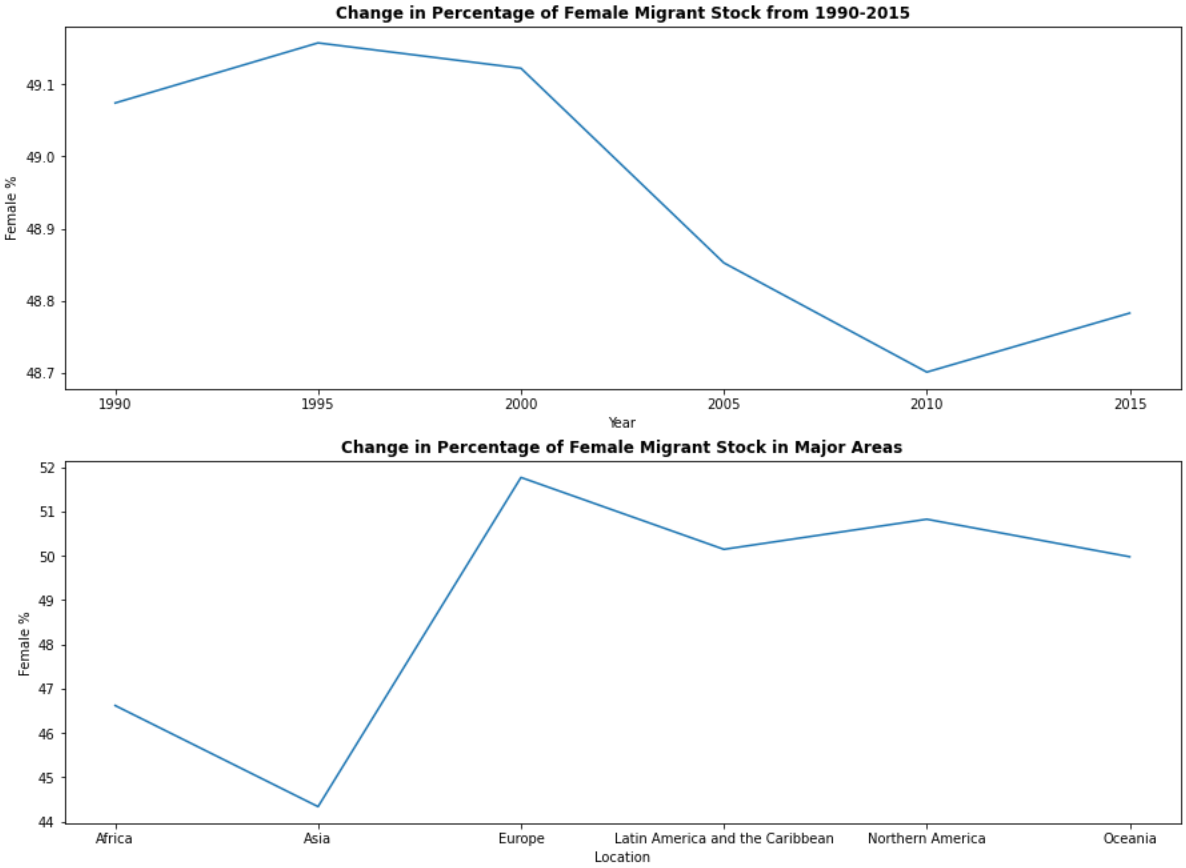

Figure 6: Change in the International Female Migrant Stock Percentage in Major Areas from 1990-2015

These two graphs are supposed to be looked at to compare, and they show us that the female international migrant population decreased in 2010 and then is increasing again and this is validated by the previous graphs where males were dominated except in some areas. Secondly, we can see that the females are mostly coming in from all major areas except Asia and Africa. This gives us and the UN and idea that they can expect more females in the other major areas and plan accordingly and strengthen their strategy. These graphs are also clean with clean lines and clean colors eliminating any chart junk.

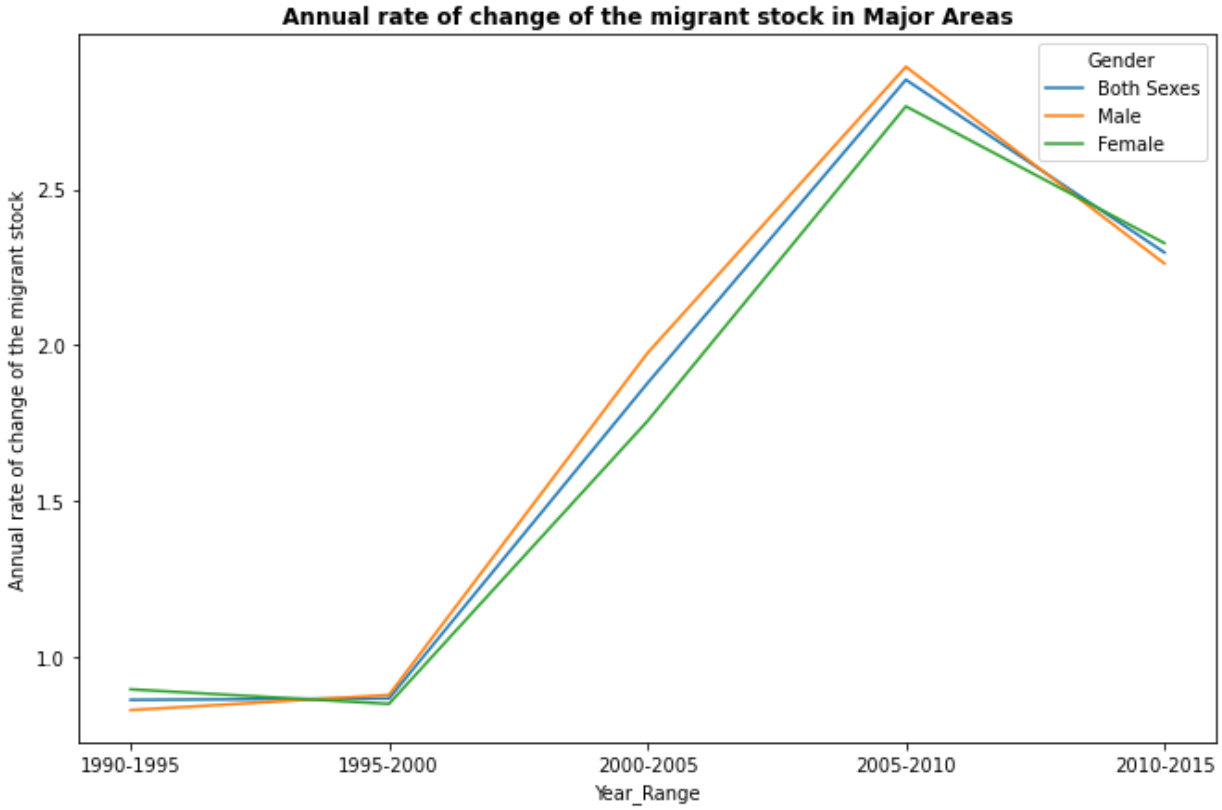

Figure 7: The Annual Rate of Change in the Migrant Stock in Major Areas from 1995-2015

This line graph shows us the year ranges and compares the different genders to show annual rate of change if the migrants. There is an upward trend that stops from 2005-2010 and is decreasing with males decreasing the most and females increasing. This shows there might be a decrease in the annual rate of change, but more females than males are migrating as we saw in figure 4 and 5. However, I wonder the reason for the decrease in migrant stock since previous figures have showed an increase in some of the major areas. Since this is combining all the major areas it could be that the ones not increasing, rather decreasing may be affecting the overall rate of change. This graph is also clean with clean lines and clean colors eliminating any chart junk.

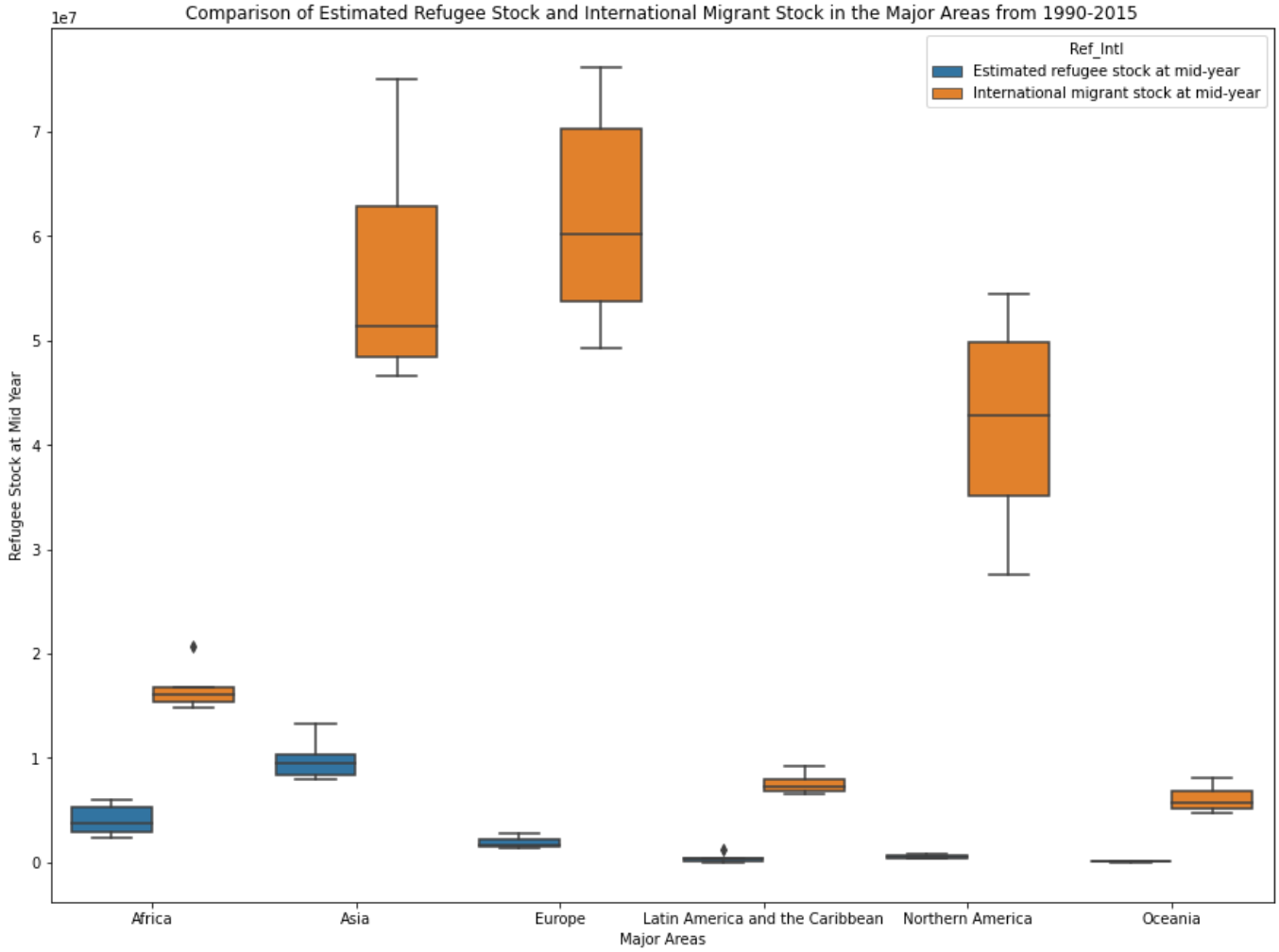

Figure 8: Comparing the Estimated Refugee Stock from the International Migrant Stock at Mid-Year in Major Areas since 1990-2015

This boxplot is comparing refugees from the total international migrant stock (data from tables 1 and 6). Here we can see that there is a great difference in the variability of the international migrant stock as they are much higher especially in Asia, Europe, and Northern America. From them the refugee stock is much lower in numbers with Latin America and the Caribbean, Northern America, Oceania, and Europe being the lowest. This shows that refugees are much lower in number, and they do not have much variability over the years showing their consistency in the different major areas.

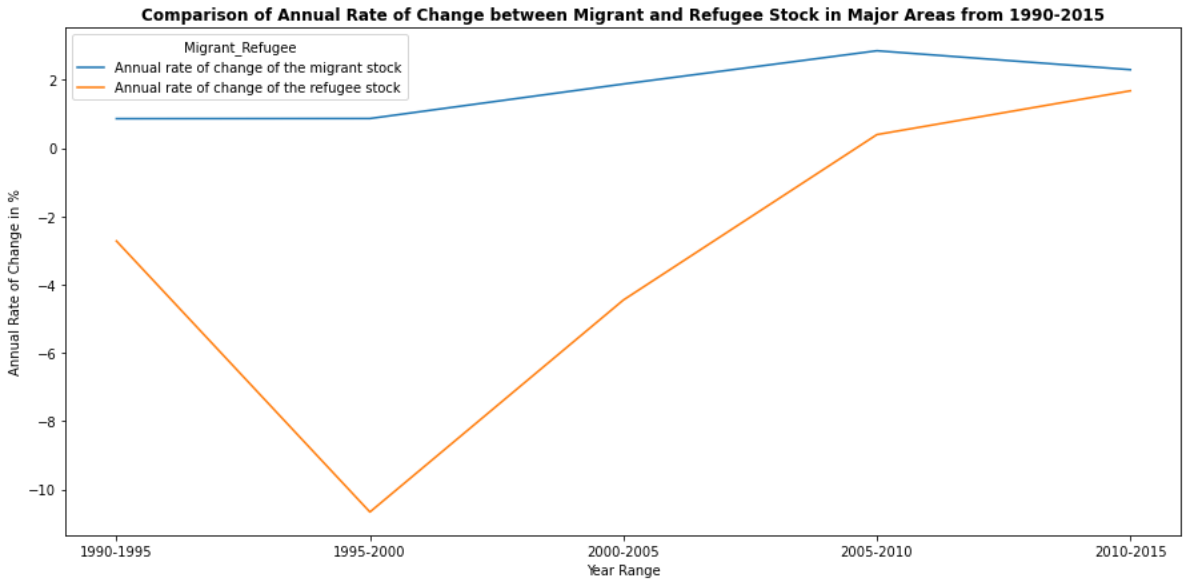

Figure 9: Annual Rate of Change of Refugees vs Migrant Stock in Major Areas from 1990-2015

This graph shows us that migrants are remaining at a constant compare the refugees who reduced drastically and then increased again. This helps us recognize that migrants will continue to be there, and the UN can prepare for that accordingly. This graph is also clean with clean lines and clean colors eliminating any chart junk.

Discussion

From this dataset, I have learnt a lot about the different trends for international migrants, refugees, and total populations. I have seen that there is a huge growth in the international migrants over the years and I can assume that they will continue to grow. This is mostly in Asia, Europe, and Northern America. This can allude to the westernization and modern lifestyles of these areas that cause the migration here and further settlements too. I also noticed that females tend to have a larger spread than men and females and after the big drop in 2010 they are migrating more since, but mostly to the major areas other than Africa and Asia perhaps due to the culture or future opportunities. I also noticed that there seems to be an increase in the refugee stock after 2005-2010 suggesting the political issues in some of the major areas getting worse and them going to Africa, Asia, and Europe. These trends show us that there is a lot of work that needs to be done to protect and accommodate for the international migrants and refugees. Working on this dataset, has encouraged me to be more mindful of the trends around the world and work towards helping more migrants. Tukey and Tufte's principles really helped to draw these conclusions which will help the UN in their future planning and in me being more mindful

Conclusion

Overall, this was a very fun data set to work with and gave me deep insights into the different ways we can explore and analyze data. I also learnt a lot about the international migrants,

refugees and total populations and the amount of work the United Nations must plan for and carry out to help protect international individuals. I also learnt a lot about data science and the touched on the extent of visualizations and forecasting that can be conducted. These visuals are an easier way to see the trends and trajectories and not most importantly having to do statistics on these. Tukey and Tufte's principles really helped to guide my data analysis and visualizations. Overall, I can safely say that I have only touched the surface of what can be done with this data set and while making the graphs I was thinking about different ways something can be shown using different libraries. There is a lot more to learn and surely these trends are what inform the forward planning on many companies as well.