Yiwen Ma 1009691576

Introduction

Data visualization refers to the way that presents information and data in a graphical way that provides an easier accessible way to see and understand trends, patterns and findings in a dataset. Especially, it can make audiences without pre-contexts understand the messages in the graphs without confusion.

The aim of this write up is to provide graphic representations of International Migrant Trends in different continents from 1995 to 2015 with a focus on female migrants. There are some interesting macro level findings that attract my attention from International Migration 2020 Highlights by UN. This reported stated that Europe and North America are the top two destination for international migrants and nearly half of all international migrants worldwide were women or girls (United Nations, 2021). With this result as of 2020, I am curious about what the trends and patterns will be like in the past as for migrant's destinations and gender distribution. Thus, this writeup will focus on analyzing international migrant stock in continent and gender level, with focus on female.

All datasets used for this assignment are from previous data cleaning projects. Cleaned Table 1, Table 2, Table 3, Table 4 and Table 5 are selected to use which provide comprehensive data on international migrant stock and stock trends from 1995 – 2015. All previous data cleaning process can be found in previous assignment.

Method

First, descriptive analysis will be used for each table to describe and show some basic features and measurements of a table. Measurements of descriptive analysis will be min, max, mean, median and standard deviation. After that, each table will be analyzed using different graphs. Six types of plots are used in this assignment namely: bar plot, line graph, box plot, histogram, scatter plot and choropleth Map. Seaborn and Polty are the main tools to create these graphs.

As mentioned in previous section, this writeup will focus on analyzing international migrant stocks in major continents, for this UN dataset, it included 6 major continents, namely Africa, Asia, Europe, Latin America and the Caribbean, Northern America and Oceania. All plots will focus on data from these continents level with a comparison between gender.

Tufte's visualization principles are used to guide plots choices on different tables. It is worth noting that Tufte’s principles may have different names depending on which perspective people interpreted with, but the main ideas will be the same. Three of his major principles used in guiding visualization, including content focus, comparison and small multiples in Time series. Detailed explanations on each plot will be elaborated in sections below.

Result & Discussion

Table 1 - International migrant stock at mid-year

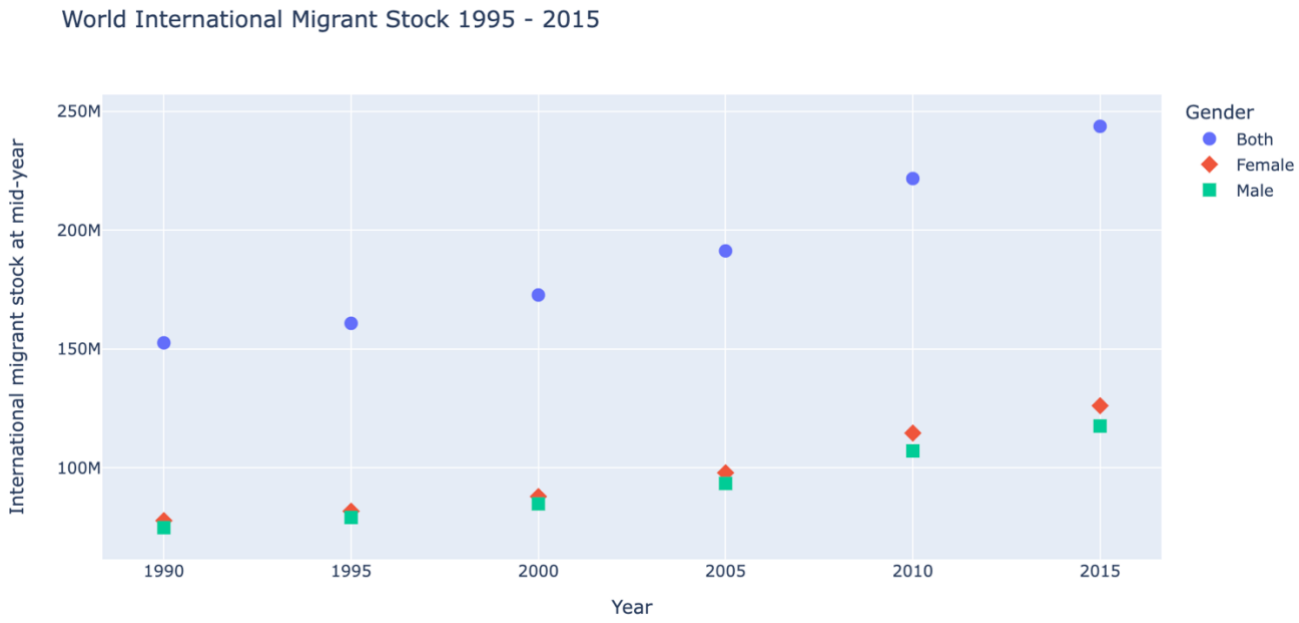
First, I used descriptive measurements to get some basic information about overall international migrant stock from 1995 to 2015. Below is the descriptive analysis result.

Figure 1.1 - Descriptive Analysis of Table 1.

```
max       2.437002e+08
min       6.300000e+01
mean      2.551538e+06
median    1.123510e+05
std       1.172720e+07
Name: International migrant stock at mid-year
```

The first two principles of Tufte are content focus and comparison rather than merely description, in order to have a more straightforward overview and comparison of the world international migrant stock over 1995 ~ 2015, a scatter plot used here to better illustrate the information.

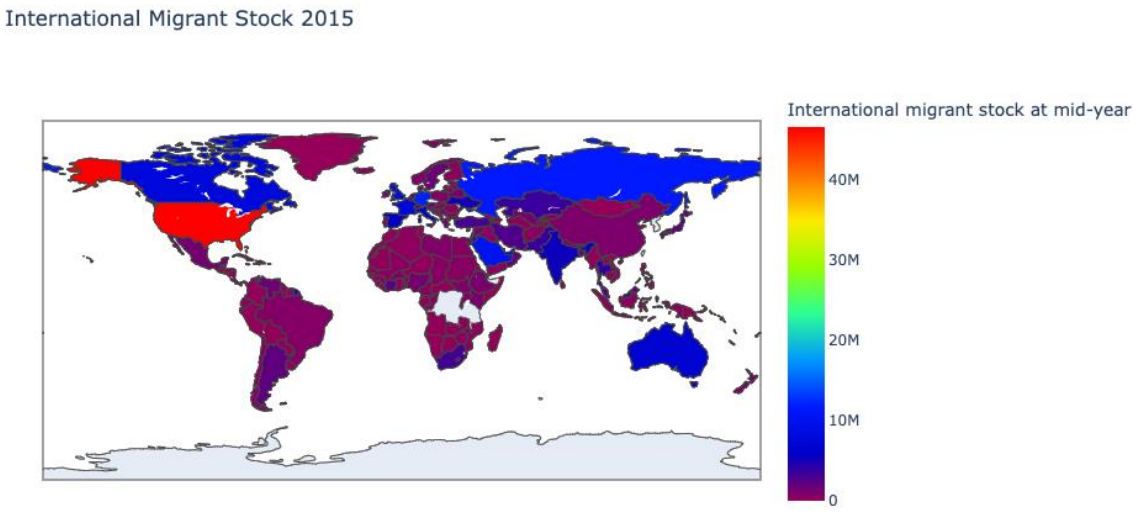
Figure 1.2 - World International Migrant Stock between Year 1995 ~ 2015 (Table 1)



It is quite clear that from the world overall level, male and female migrants are increasing overtime and total female migrant stocks increased slightly more than male groups, and the differences between the amount of increase is getting bigger in recent time, year 2010 and 2015. Then, I used an interactive choropleth map to illustrate the world international migrant stock trends layered with continents (Figure

1.3). Since this figure is an interactive figure, it will show more information in coding documents where data information is shown when cursor moves to a specific area.

Figure 1.3 - World International Migrant Stock 2015 by Continents (Table 1)



The above figure clearly showed that Northern America and Europe continents attracted the most migrants in 2015. As far as for now, the trends of continents attractiveness and amount of female migrant stocks are adhered with the recent trends reported by UN. The second Tufte's principle is comparison rather than mere description. In order to get a clear comparison and in-depth understandings between gender within different continents on a macro level, I used bar plot with gender layer, see Figures below.

Figure 1.4 - 1990 ~2015 Africa International Migrant Stock between Genders

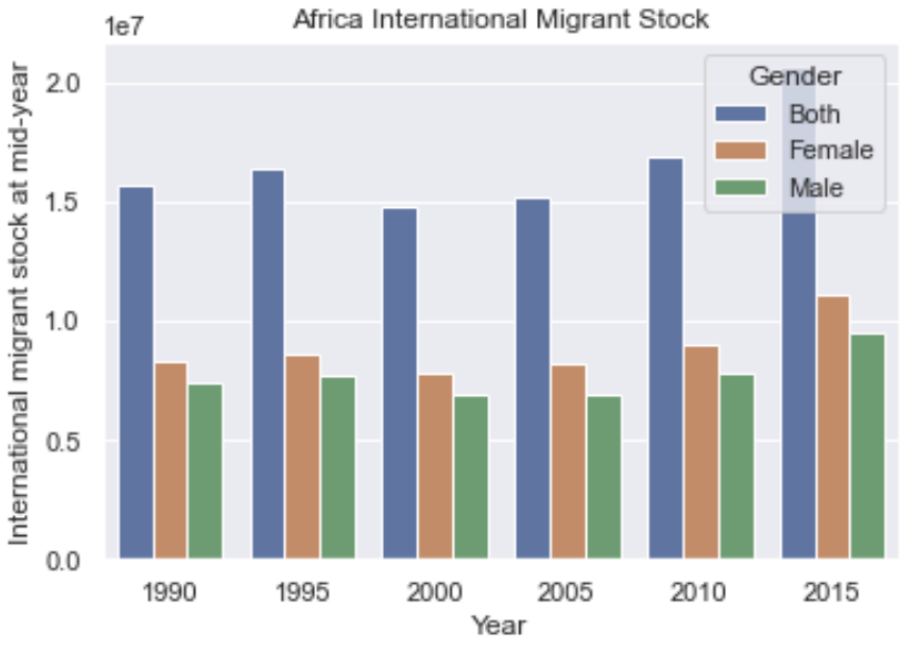


Figure 1.5 - 1990 ~2015 Asia International Migrant Stock between Genders

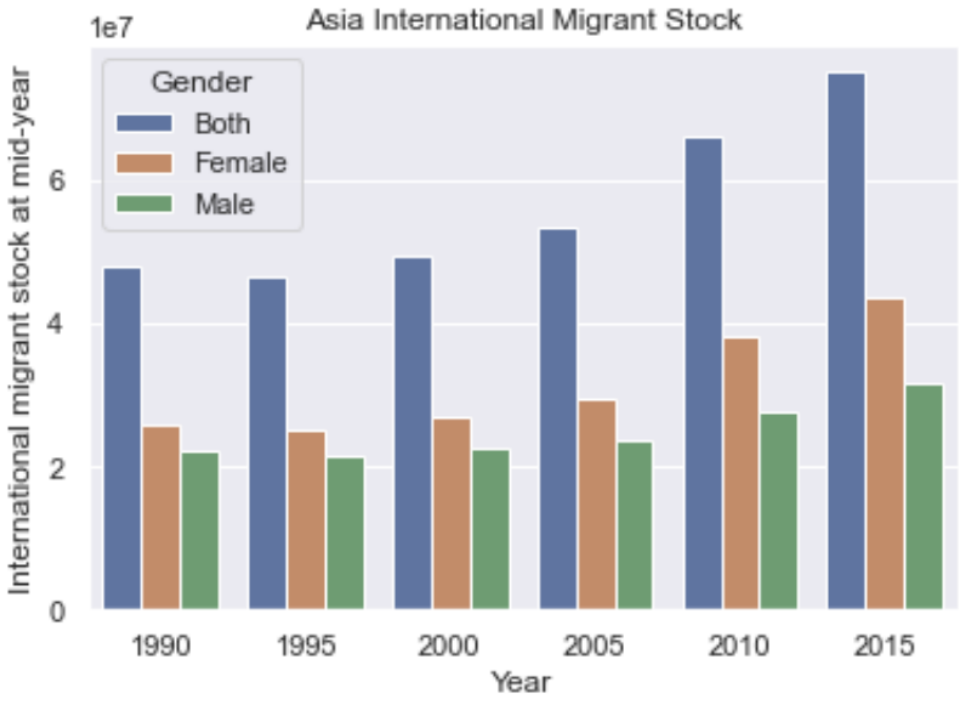


Figure 1.6 - 1990 ~2015 Europe International Migrant Stock between Genders

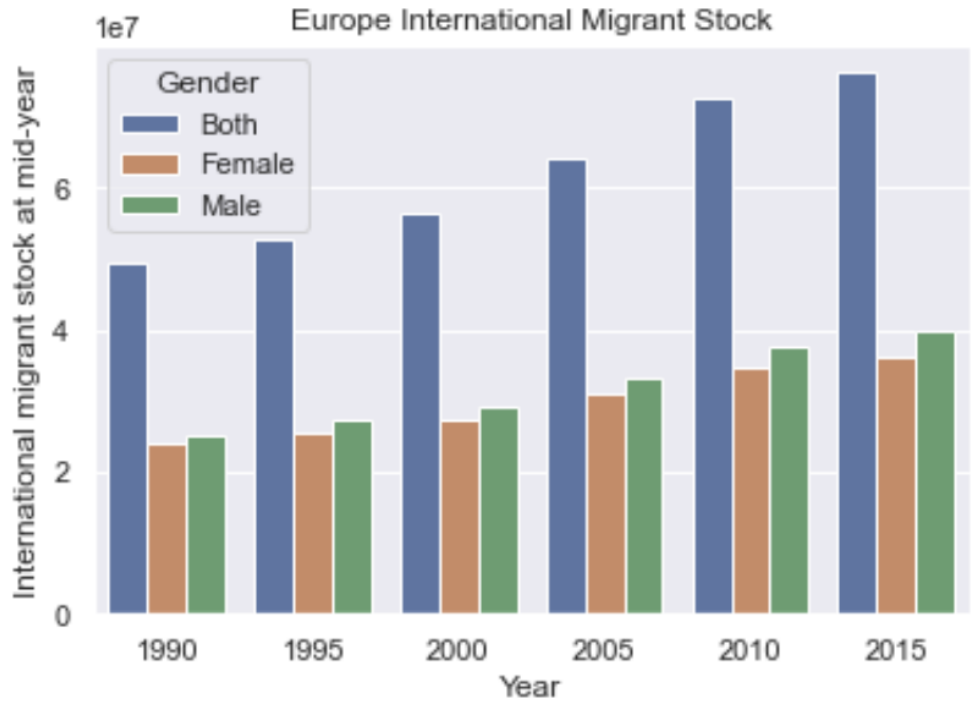


Figure 1.7 - 1990 ~2015 Latin America and the Caribbean International Migrant Stock between Genders

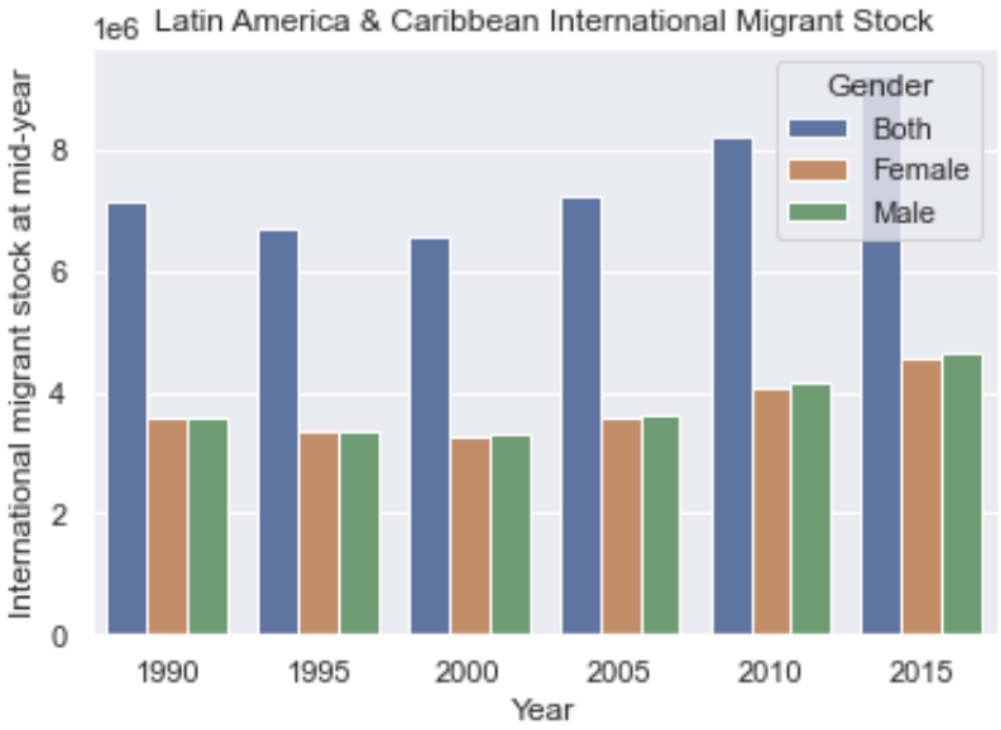


Figure 1.8 - 1990 ~2015 Oceania International Migrant Stock between Genders

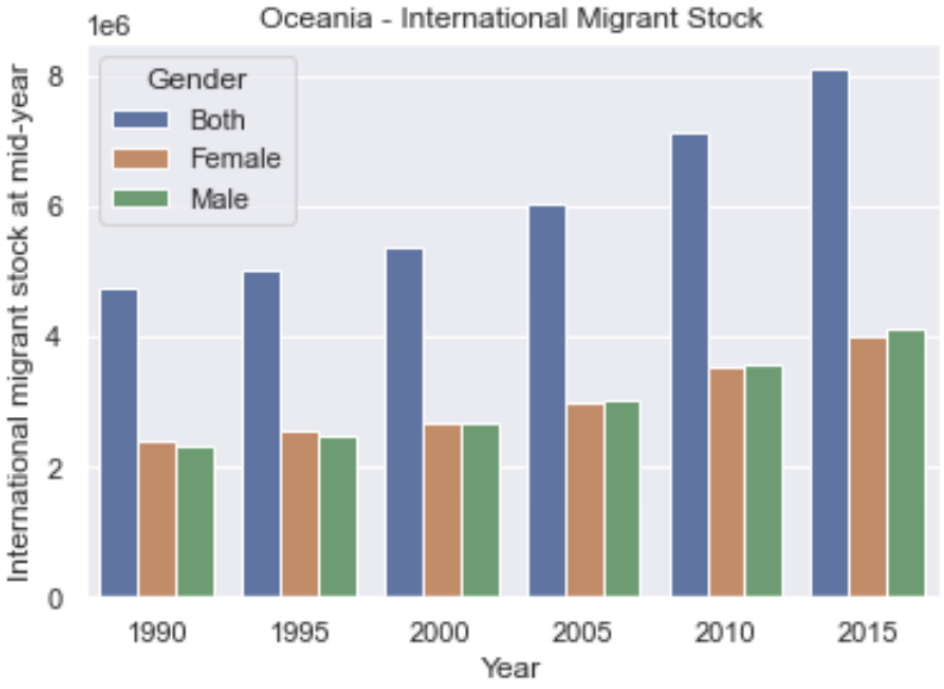
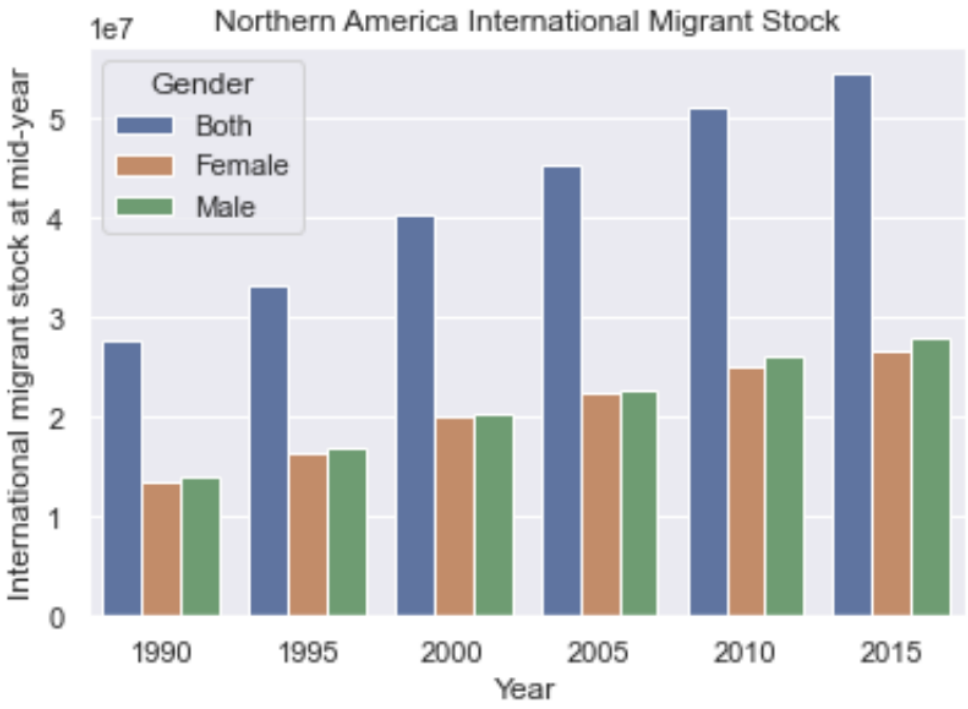


Figure 1.9 - 1990 ~2015 Northern America International Migrant Stock between Genders



The above bar plots showed clear value differences in male and female stocks in different continents over a time period. It is obvious that female in Asia and Africa regions are more likely to migrate to other country and regions. Moreover, for all region, female migrants are increasing over the time, however, in some regions like Latin America & Caribbean, Asia and Africa, male migrants trend experienced a slightly decrease between 2000 to 2005. This may be because women's rights can be more protected in other regions and female seek to going abord to improve their living standard.

Table 2 - Total Population at Mid-year

First, descriptive analysis will be used to get a basic idea of the total amount of total population stock from 1990 to 2015.

```
max       7.349472e+09
min       7.700000e+02
mean      1.015966e+08
median    4.713626e+06
std       4.743601e+08
Name: Total population at mid-year,
```

Followed by Tufte's first two principles of content focusing and comparison than mere description, boxplots are used for this table, comparing each year's total population by gender, see figures below.

Figure 2.1 - Total Population at Year 1990 by Gender

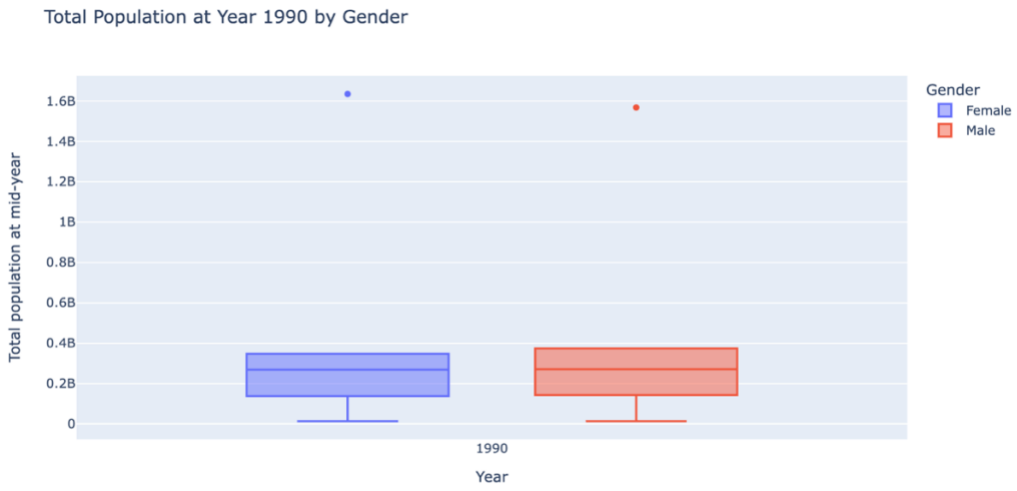


Figure 2.2 - Total Population at Year 1995 by Gender

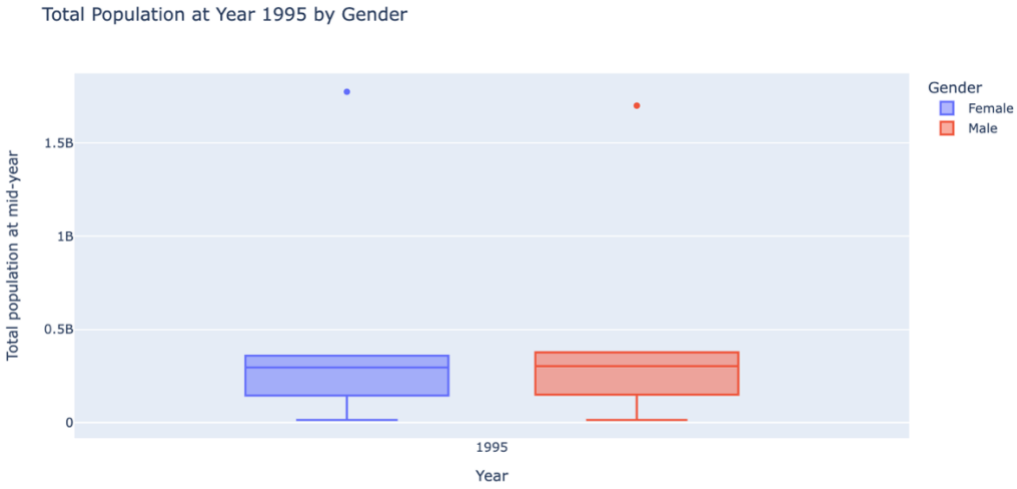


Figure 2.3 - Total Population at Year 2000 by Gender

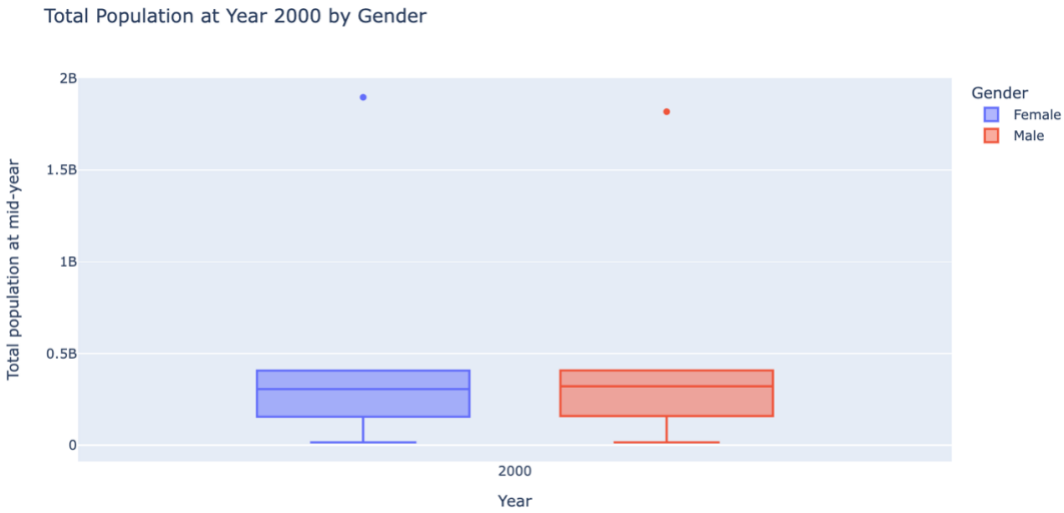


Figure 2.4 - Total Population at Year 2005 by Gender

Figure 2.5 - Total Population at Year 2010 by Gender

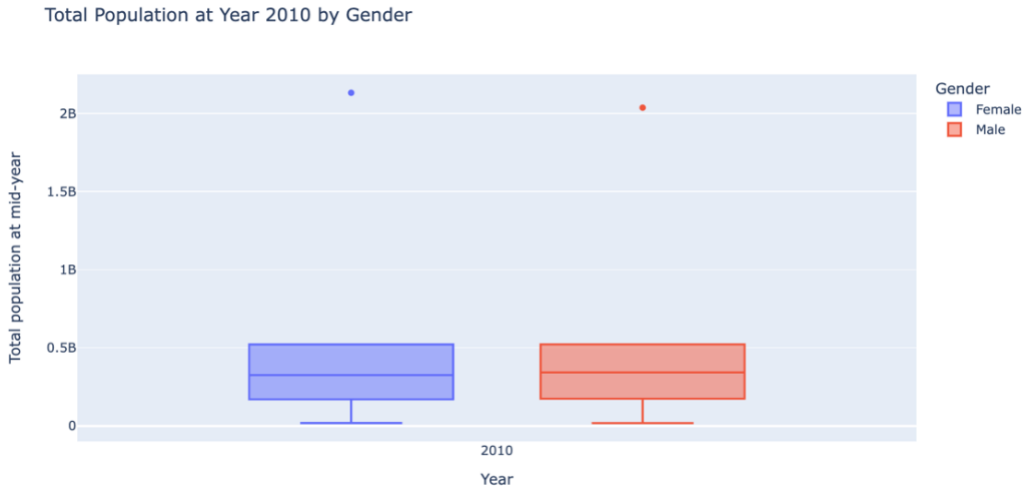
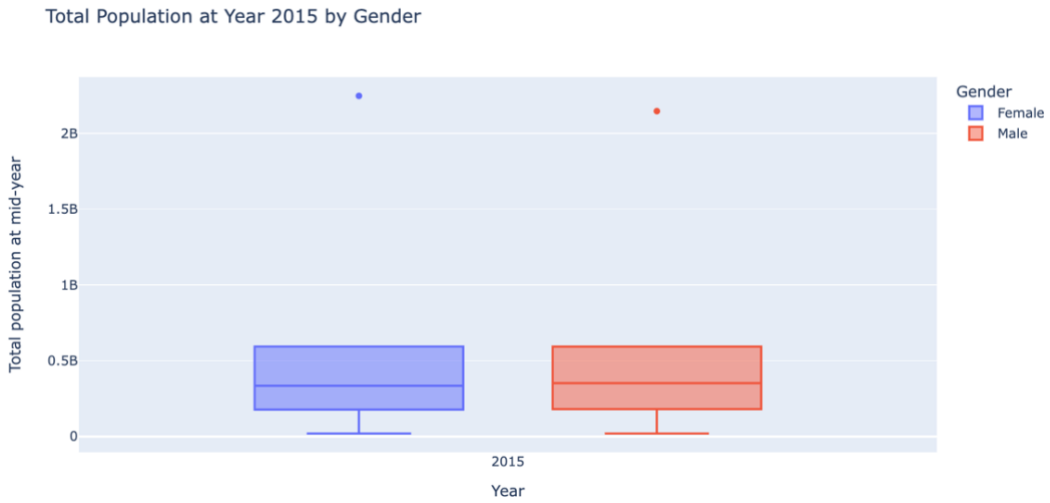


Figure 2.6 - Total Population at Year 2015 by Gender



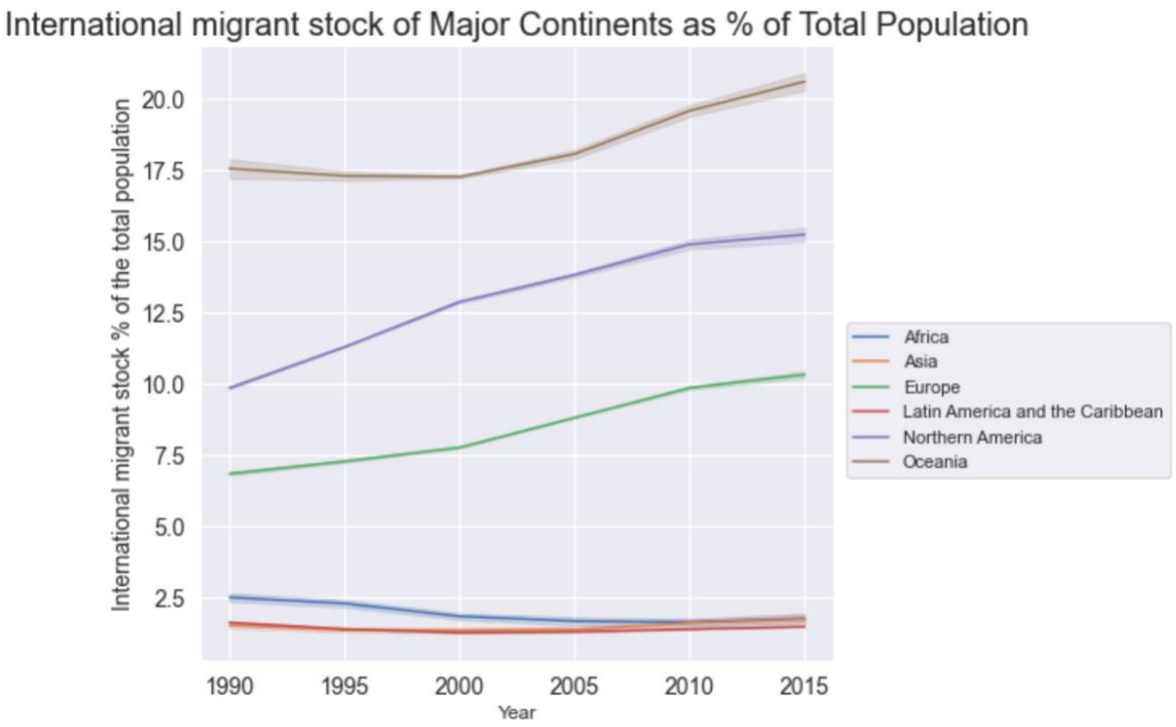
The above box plots illustrate clear outliers of total population across all year periods and maximum lines cannot be able to observe. The minimum and maximum numbers are also having a big difference in both male and female populations. This may be due to the different development stages and processes between countries and regions, that will lead to a distinct population stock. All the boxes do not show a significant difference, most of the figures stayed within the middle 50% quartile with similar medians between genders.

There are not huge differences in total population between male and female in majority of the numbers, still international migrant stock is show differences between genders. The next table will examine the trends and patterns of international migrants as percentages of total population.

Table 3 - International Migrant Stock as a Percentage of Total Population

Since this write up will focus on macro continental level, I created a new data frame based on the previously cleaned Table 3 used for the following visualization (see code file for process of creating new data frame).

Figure 3.1 - International Migrant Stock of Major Continents as Percentage of Total Population

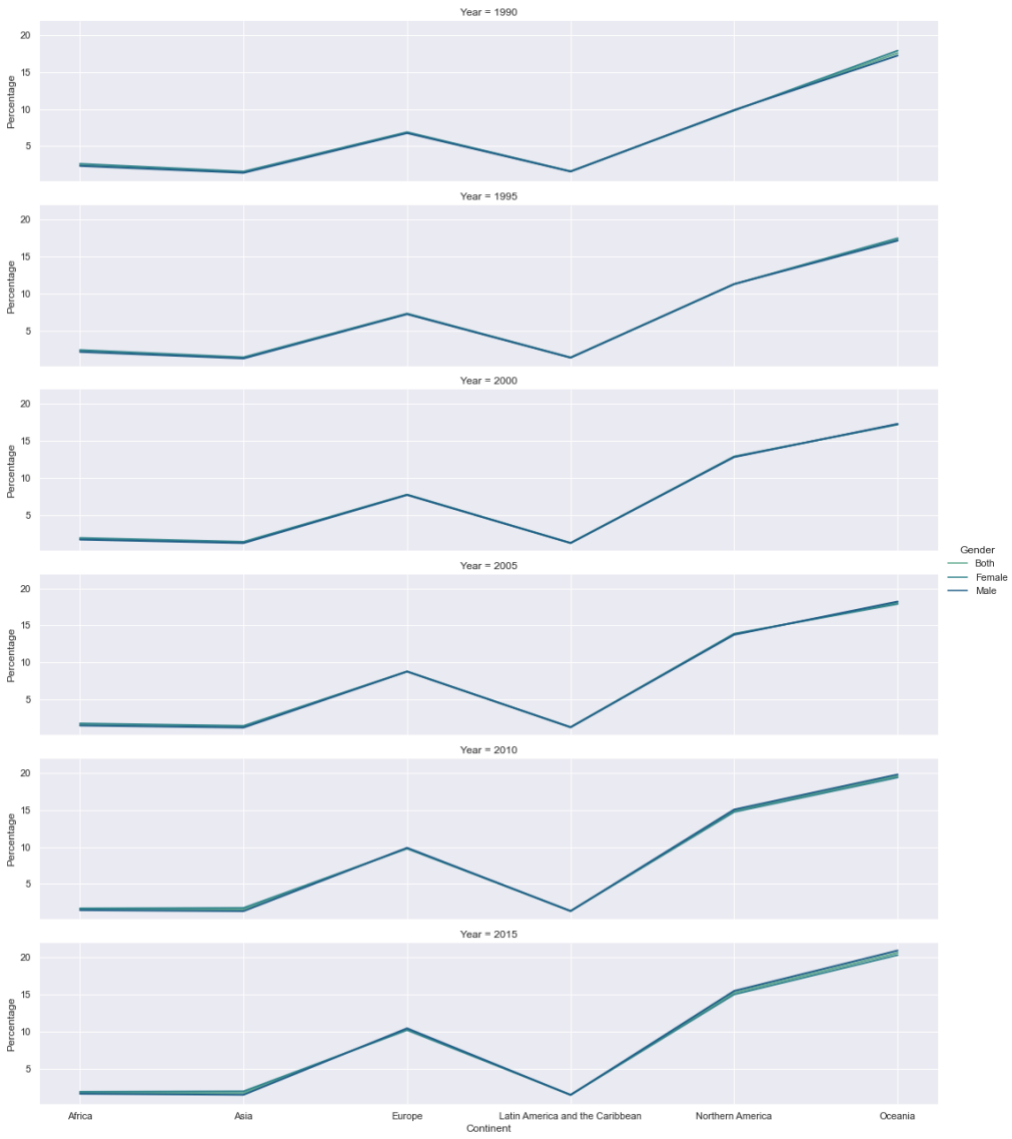


The above line graph shows the changing trends of international migrants as percentages over 1990 to 2015 time period of 6 different continents. As indicated in the above line graph, the content comparison is obvious and straightforward with enrich content that adheres with Tufte's principle. From Figure 3.1,

we can see ghat Oceania, Northern America and Europe attacks more international migrants than other continents. This result is very similar to the results of the previously mentioned UN 2020 report where Europe and Northern America are the top 2 migrants' destination. Moreover, as for 2015, from the direction of the lines shows that these three continents will continue to welcome more new immigrants subsequently.

Small multiples is another principle that Tufte addressed which comparisons can be enforced within the scope of eye span. Figure 3.2 illustrated this table with small multiples in time series. Tufte also emphasized the resolution of data visualizations, thus the below figure is saved as a high-resolution picture not just screenshotted.

Figure 3.2 - International Migrant Stock of Major Continents as Percentage of Total Population (Small Multiples)



The small multiples figure provides the opportunity to see the trends overtime all at once with all independent values (6 continents) showing different percentages from 1990 to 2015 time span. From the above figure we can tell that all gender types show similar patterns for each 5 year time in all 6 continents, which is not very similar to the trends shown in Figure 3.1. However, Europe, Norther America and Oceania is the top 3 continents for migrants' destination with similar amount of attractiveness for both male and female migrants.

Table 4 - Female Migrants as a percentage of the International Migrant stock

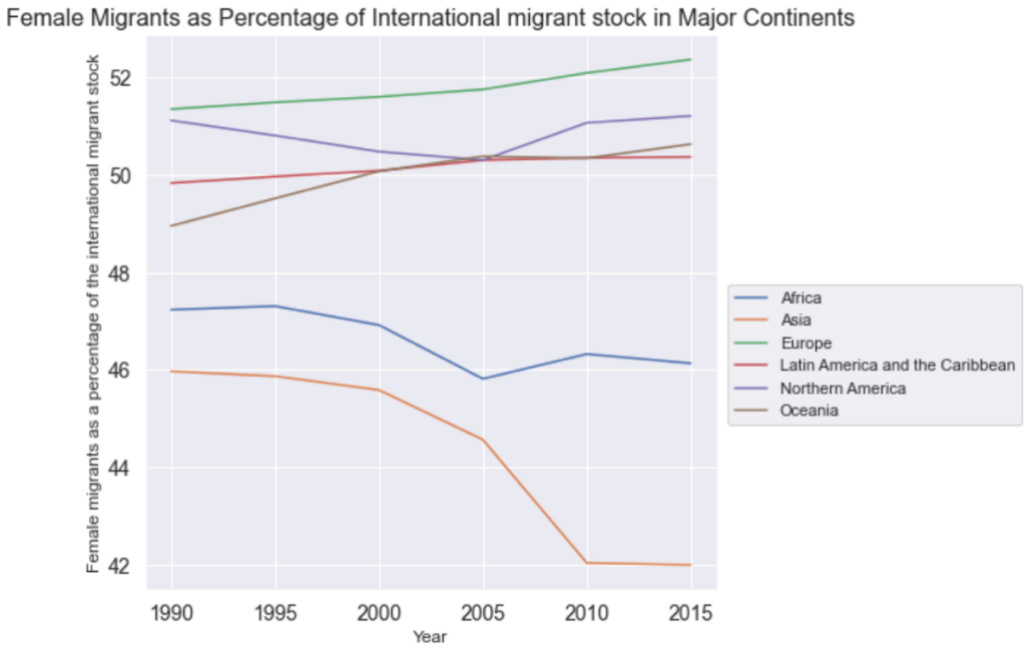
First, a descriptive measurement is used here to get some basic information about female migrants' proportion of International migrant stock.

Figure 4.1 - Descriptive Analysis of Table 4.

```
max       70.703810
min       13.325719
mean      48.316114
median    49.101998
std        6.532569
Name: Female migrants as a percentage of the international migrant stock(%)
```

There are big differences observed between max and min number, this might because of the differences between continents, which female may have preferences of choosing their migration destination. Since this write up will focus on macro continental level, I created a new data frame based on the previously cleaned Table 4 used for the following visualization (see code file for process of creating new data frame).

Figure 4.2 - Female Migrants as Percentage of International Migrant Stock in Major Continents



According to Tufte's content focusing and comparison principle, this line graph shows a clear comparison and trends of female migrants in 6 major continents overtimes. As elaborated in the graph, Europe is the top 1 destination of female's migration choices, followed by Northern America and Oceania. To have a more in-depth understanding and comparison, the small multiples principle is used here to create a time series graph as shown in Figure 4.3. Tufte also emphasized the resolution of data visualizations, thus the below figure is saved as a high-resolution picture not just screenshotted.

Figure 4.2 - Female Migrants as Percentage of International Migrant Stock in Major Continents (small multiples)

The figure above shows the changing trends of all major continents overtime all at once from 1990 to 2015. It is quite clear that female migrants are prefer European countries and the percentage of immigrants has been high during these years, the line of indication almost turn into straight line. Females also showed a significant fever towards Northern America, Oceania and Latin America and the Caribbean. Whereas, Africa and Asian does not interests female migrants as the rest continents do. Especially, Asia experienced a significant drop in female migrants in recent decades. It would be interesting to further investigate the female migration destinations between 2005 and 2015 for Asian countries. The small multiples do help a lot in visually enforcing comparisons of changes.

Table 5 - Annual rate of change of the Migrant Stock

Histogram will be used to easy understanding comparisons between male and female migrant stock annual rate of change, adhere to Tufte content focus and comparisons not just merely description principles.

Figure 5.1 - Annual Rate of Change in Percentage of Migrant Stock (1995~2000)

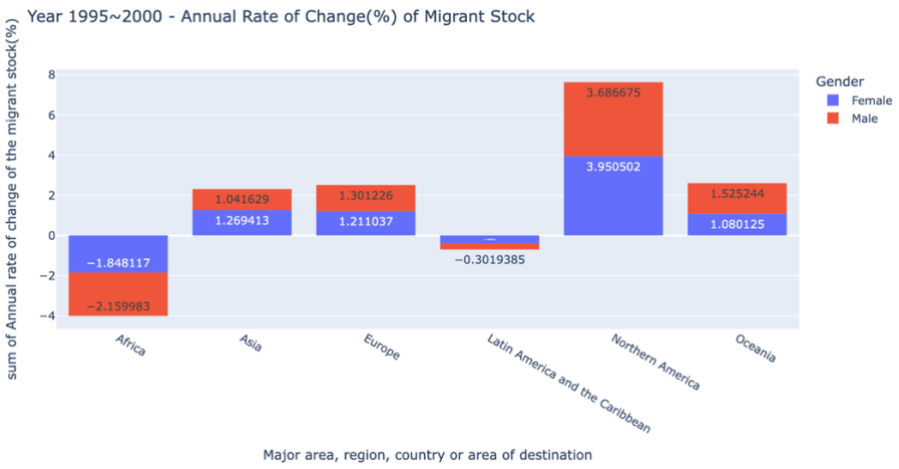


Figure 5.2 - Annual Rate of Change in Percentage of Migrant Stock (2000~2005)

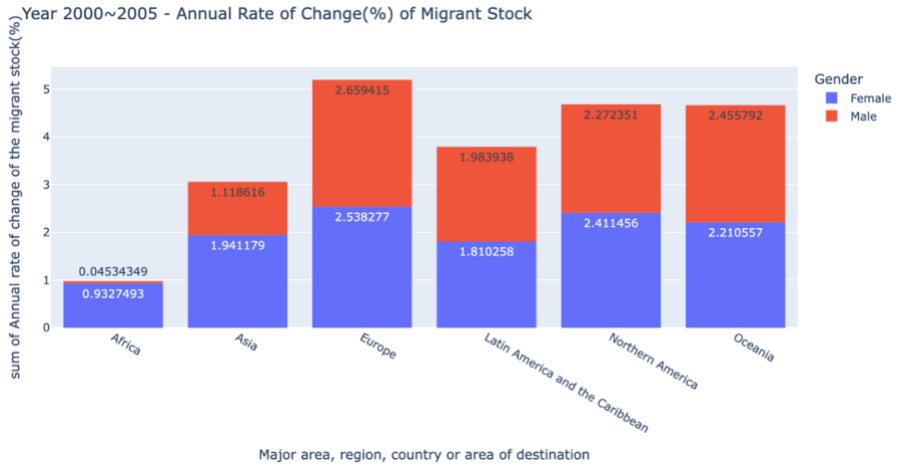


Figure 5.3 - Annual Rate of Change in Percentage of Migrant Stock (2005~2010)

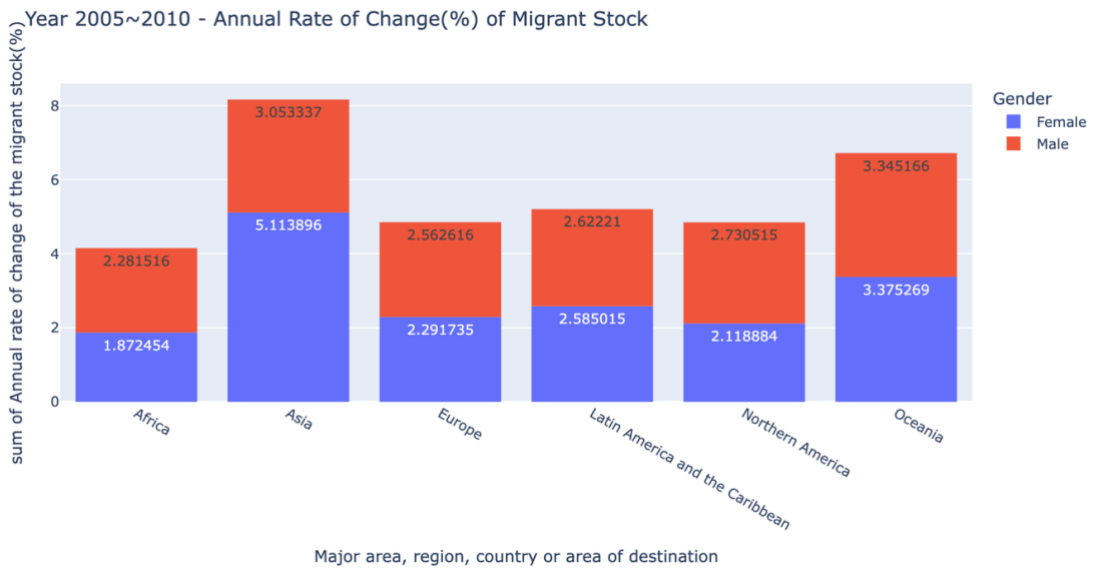
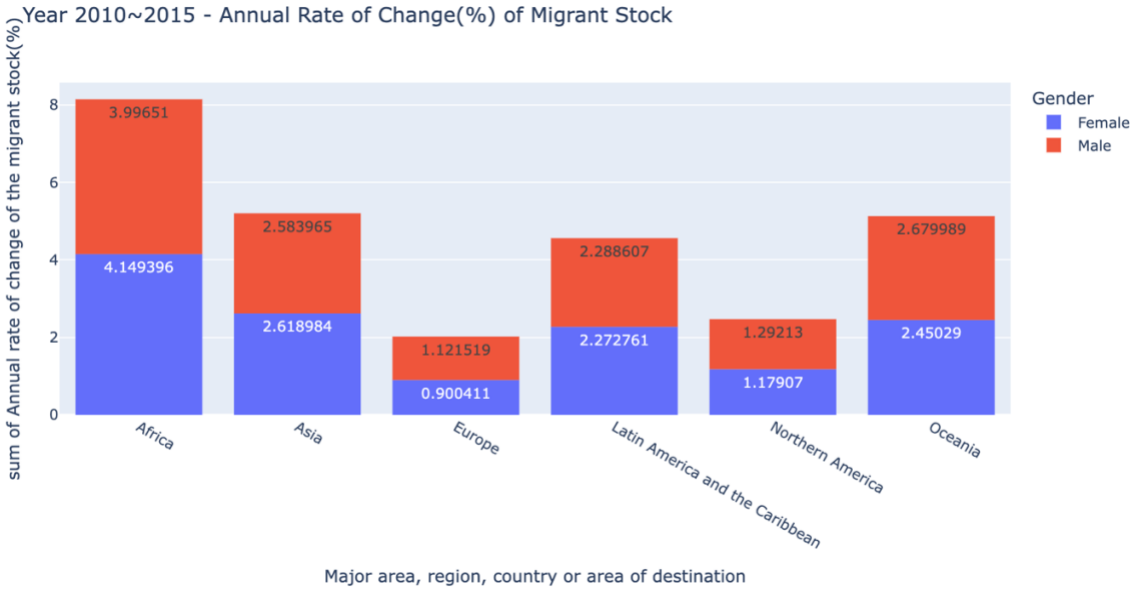


Figure 5.4 - Annual Rate of Change in Percentage of Migrant Stock (2010~2015)

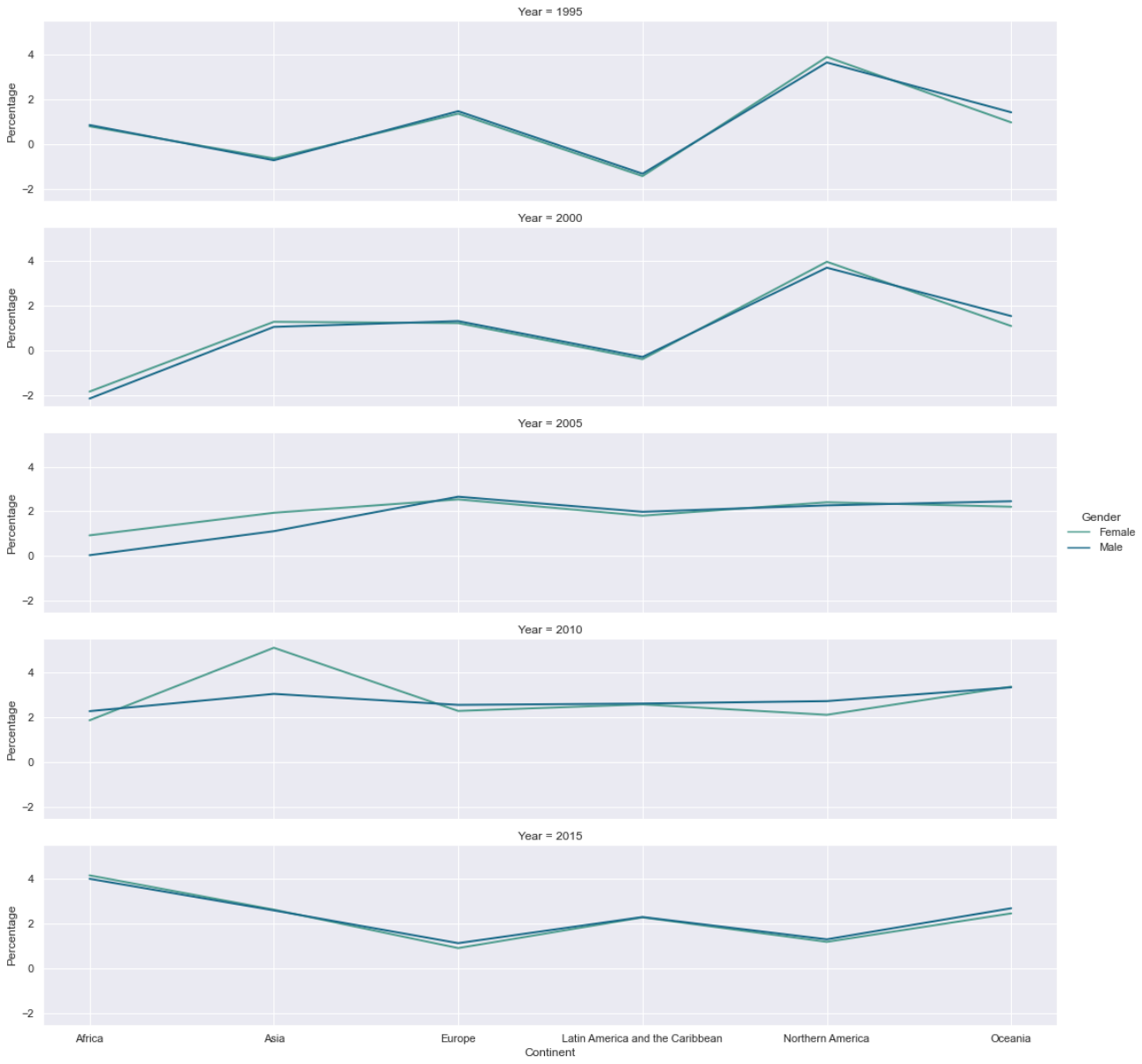


The histogram above shows a steady increase in male and female migration in most regions, except Africa and Latin America and the Caribbean. Between 1995 to 2000, these two regions show a significant negative growth in both male and female migrants, that is consistent with the previous analysis that most immigrants would not choose these two regions. This may be due to the under development of social, economic and educational aspects. Over time, both regions have seen significant growth in overall development, thus a steady increase in the total number of new immigrants have been observed. Surprisingly, Europe and Northern America have experienced falls in annual rate of change in year 2010

to 2015, which is showing different trends with 2020 UN developments report, and this could be further investigated.

As mentioned before, I created a new data frame based on the previously cleaned Table 5 used for the following visualization (see code file for process of creating new data frame).

Figure 5.5 - Annual Rate of Change in Percentage of Migrant Stock (2010~2015)

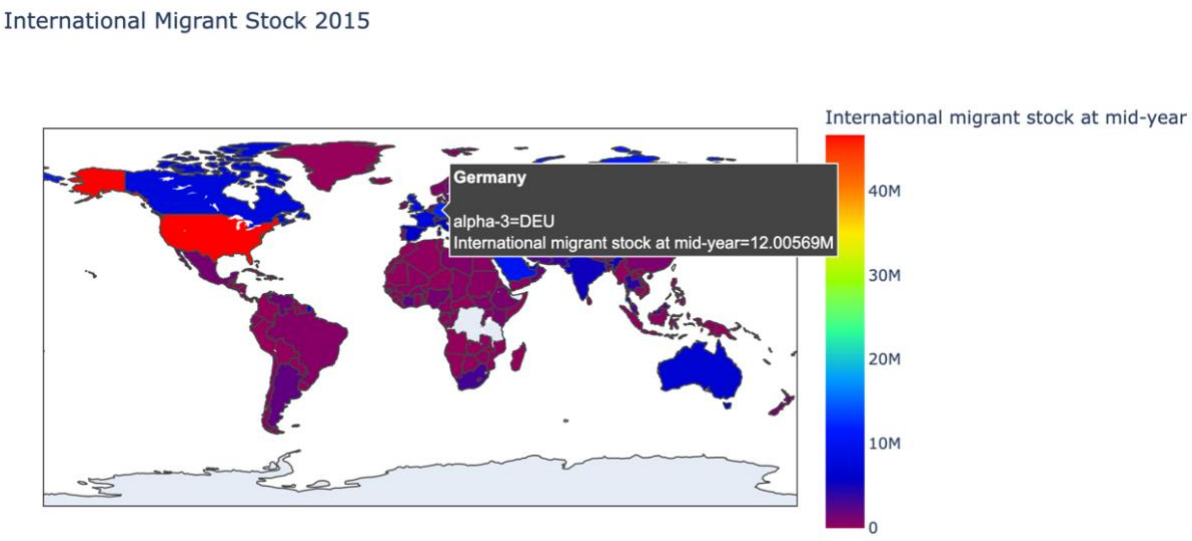


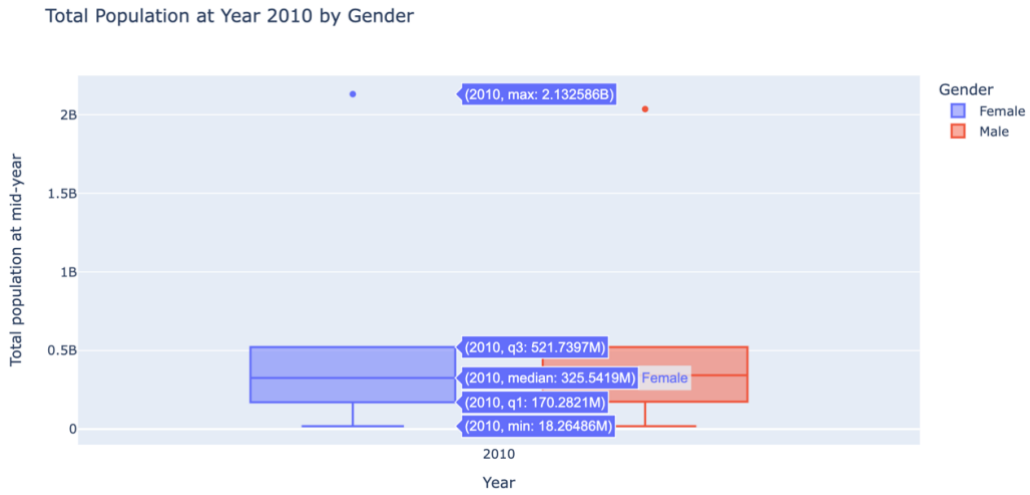
Small multiples principle is also used for Table 5 to have a better comparison by time span of each time period. The figure above shows the changing trends of annual rate of change of all major continents overtime 1995~2015 period. It is quite clear that all six regions have experienced some ups and downs through years in increases of migration of all genders. Surprisingly, Asian has experienced a major

increase in female migrants in 2005~2010 time period and then dropped sharply in 2010 ~ 2015 period. Moreover, both Europe and Northern America continents has experienced falls in migrants of both genders since 2005. However, the annual rate of change is calculated based on previous year, a decrease in annual rate of change does not necessarily lead to a decrease in its overall migration performance. Migration will normally be related to the economic and social environment of a country and region, and it will vary from time to time according to numerous social and political factors.

Limitation

There are multiple limitations were discovered during the coding process about figure presentations. First, the presentation of color and font size of figures created by seaborn will differ each time running the code even on the same computer. Initially, I set the figure size as a certain proportion to make the figure in higher resolution with clear indicates of the content. However, the figures changed each time run the code based on previous figure proportion. Thus, I decided to just keep the simple code using default color and font size setting. Secondly, plotly will create high quality interactive figures and it works well in coding documents. However, the interactive motions are hard to capture and frame in the writeups which will lead to a loss in information display, like the scatter plot, choropleth map and boxplot mentioned in above sections. The actual interactive display will show as the figure below.





Conclusion

To conclude, through the visual analysis of the above tables, migrants showed a clear favor to Europe and Northern America continents for their migrants' destinations. Moreover, female also is more likely to migrant than male with slightly differences over time. These two findings are in accordance with the recent trends reported by the UN as of 2020. However, there are still some interesting findings that can be explored later. For instance, Oceania also attracted lots of migrants in the past and its ranking sometimes can higher than Europe and Northern America. The patterns and reasons behind this can be further studied for interest.