
# INF1340 Final Project: UN Data Visualization

Yanke Mo (1005939656)

A final paper submitted in conformity with the requirement

for the course of INF1344H

Faculty of Information

University of Toronto

## Table of Content

# 1. Introduction:

The aim of this project is to study the performance of international migration. The project is interested in finding out whether the choice of migration will be impacted by regions or gender. The dataset that will be analyzed is the "United Nation: Trends in International Migrant Stock", it captures the migrant stock for all regions based on gender from year 1990 to year 2015.

In the following content, the project will mainly focus on applying data visualization. By using descriptive statistics and small multiple figures, the project will analyze the change in migration in terms of regions and genders. In addition, the written project will cover tables that are significant and meaningful to the migration. These tables are table 1, table 3, table 4, and table 5 in the UN dataset; the rest of the tables will not be considered in this analysis.

## 2. Data Cleaning Process:

➢ **Step One: Remove Cover Page Error**

Importing tables and labeling tables are the first step of data cleaning. All of the 6 tables in UN excel file are import as "Table_#"(# means the order of the table). The first problem appears in the first 12 rows of every table.It is supposed to be the cover page of the excel file, but it is converted into NaN in Jupyter.

Meanwhile, the column header of the tables has been imported as values (row 13 and row 14). This issue has violated tidy data principles 1, given that the column names should be stored in the header, rather than the cell. Missing column headers will also leave the value undefined. If the data is going to be visualized, it is impossible to locate the values by column header. (pic1)

| index | Unnamed: 0 | Unnamed: 1 | Unnamed: 2 | Unnamed: 3 | Unnamed: 4 | Unnamed: 5 | Unnamed: 6 | Unnamed: 7 | Unnamed: 8 | Unnamed: 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 1 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 2 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 3 | NaN | NaN | NaN | NaN | United Nations | NaN | NaN | NaN | NaN | NaN |
| 4 | NaN | NaN | NaN | NaN | Population Division | NaN | NaN | NaN | NaN | NaN |
| 5 | NaN | NaN | NaN | NaN | Department of Economic and Social Affairs | NaN | NaN | NaN | NaN | NaN |
| 6 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 7 | NaN | NaN | NaN | NaN | Trends in International Migrant Stock: The 2015 Revision | NaN | NaN | NaN | NaN | NaN |
| 8 | NaN | NaN | NaN | NaN | Table 1 - International migrant stock at mid-year by sex and by major area, region, country or area, 1990-2015 | NaN | NaN | NaN | NaN | NaN |
| 9 | NaN | NaN | NaN | NaN | POP/DB/MIG/Stock/Rev.2015 | NaN | NaN | NaN | NaN | NaN |
| 10 | NaN | NaN | NaN | NaN | December 2015 - Copyright © 2015 by United Nations. All rights reserved | NaN | NaN | NaN | NaN | NaN |
| 11 | NaN | NaN | NaN | NaN | Suggested citation: United Nations, Department of Economic and Social Affairs (2015). Trends in International Migrant Stock: The 2015 revision (United Nations database, POP/DB/MIG/Stock/Rev.2015). | NaN | NaN | NaN | NaN | NaN |
| 12 | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN | NaN |
| 13 | Sort order | Major area, region, country or area of destination | Notes | Country code | Type of data (a) | International migrant stock at mid-year (both sexes) | NaN | NaN | NaN | NaN |
| 14 | NaN | NaN | NaN | NaN | NaN | 1990 | 1995 | 2000 | 2005 | 2010.0 |
| 15 | 1 | WORLD | NaN | 900 | NaN | 152563212 | 160801752 | 172703309 | 191269100 | 221714243.0 |
| 16 | 2 | Developed regions | (b) | 901 | NaN | 82378628 | 92306854 | 103375363 | 117181109 | 132560325.0 |
| 17 | 3 | Developing regions | (c) | 902 | NaN | 70184584 | 68494898 | 69327946 | 74087991 | 89153918.0 |

One of the solutions is that the first 12 rows of the cover page should be removed. Yet, if the previous 12 rows are removed, row 13 will automatically become the column names, leaving the attribute of year (1990-2015) in value. This problem will still leave the values impossible to define, and it is necessary to figure out a method that can store the attribute of years back to column headers with the other attributes.

The decision I make is to remove the entire rows from 0-15, the only data that is kept in the table are values, and all column headers are removed. The next step is to rename the entire column names. Removing headers and rename columns helps me to remove the cover page error, and I am able to define the values I need for visualization. (pic_2)

```
table_1_1 = table_01.drop(table_01.index[0:15])
table_1_1 #Drop 0-15 rows
```

| | Unnamed: 0 | Unnamed: 1 | Unnamed: 2 | Unnamed: 3 | Unnamed: 4 | Unnamed: 5 | Unnamed: 6 | Unnamed: 7 | Unnamed: 8 | Unnamed: 9 | ... | Unnamed: 13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **15** | 1 | WORLD | NaN | 900 | NaN | 152563212 | 160801752 | 172703309 | 191269100 | 221714243.0 | ... | 87884839 |
| **16** | 2 | Developed regions | (b) | 901 | NaN | 82378628 | 92306854 | 103375363 | 117181109 | 132560325.0 | ... | 50536796 |
| **17** | 3 | Developing regions | (c) | 902 | NaN | 70184584 | 68494898 | 69327946 | 74087991 | 89153918.0 | ... | 37348043 |

## ➢ Step Two: Rename Columns:

Based on the content of "United Nation: Trends in International Migrant Stock" , the values are stored by the names of areas and the migrant stock of the area from different genders and years. To ensure all attributes are kept in the column headers, I rename the years, migrant stock, and gender in the column headers as below.

```
[5] table_1_1.columns
```

```
[6] table_1_1.columns=['Order','Name of Country','Notes','Country Code','Type of Data(a)','1990 Migrant Stock Both Sex',
                       '1995 Migrant Stock Both Sex','2000 Migrant Stock Both Sex','2005 Migrant Stock Both Sex',
                       '2010 Migrant Stock Both Sex','2015 Migrant Stock Both Sex','1990 Migrant Stock Male',
                       '1995 Migrant Stock Male','2000 Migrant Stock Male','2005 Migrant Stock Male','2010 Migrant Stock Ma
                       '2015 Migrant Stock Male','1990 Migrant Stock Female','1995 Migrant Stock Female','2000 Migrant Stock
                       '2005 Migrant Stock Female','2010 Migrant Stock Female',
                       '2015 Migrant Stock Female']
```

```
[7] table_1_1 #rename the columns
```

| | Order | Name of Country | Notes | Country Code | Type of Data(a) | 1990 Migrant Stock Both Sex | 1995 Migrant Stock Both Sex | 2000 Migrant Stock Both Sex | 2005 Migrant Stock Both Sex | 2010 Migrant Stock Both Sex | ... | 2000 Migrant Stock Male | 2005 Migrant Stock Male |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **15** | 1 | WORLD | NaN | 900 | NaN | 152563212 | 160801752 | 172703309 | 191269100 | 221714243.0 | ... | 87884839 | 97866674 |
| **16** | 2 | Developed regions | (b) | 901 | NaN | 82378628 | 92306854 | 103375363 | 117181109 | 132560325.0 | ... | 50536796 | 57217777 |
| **17** | 3 | Developing regions | (c) | 902 | NaN | 70184584 | 68494898 | 69327946 | 74087991 | 89153918.0 | ... | 37348043 | 40648897 |

The data of migrant stock in table_01 are now categorized from every five years for each gender. (pic_3)

## ➢ Step Three：Separate Informations in Column Headers

The column headers are now able to explain the variable below them, but another problem appears. Too much information is stored in one column header.

The best method I could first consider is using the functions of assign and lambda to separate the column headers based on string. But before the function separates strings, I decided to store attributes of year and gender as unique columns. I first used the melt function to pull the column headers which contain years, migrant stock, and gender into cells, (pic_4) and then I used the assign and lambda functions to split years, migrant stocks, and gender into different cells. (pic_5)

```
[8] tidy_table_01 = table_1_1.melt(id_vars=["Order","Name of Country","Notes","Country Code","Type of Data(a)"],var_name="Year",value_name="Migrant Stock")
```

```
tidy_table_01
```

| | Order | Name of Country | Notes | Country Code | Type of Data(a) | Year | Migrant Stock |
|---|---|---|---|---|---|---|---|
| 0 | 1 | WORLD | NaN | 900 | NaN | 1990 Migrant Stock Both Sex | 152563212 |
| 1 | 2 | Developed regions | (b) | 901 | NaN | 1990 Migrant Stock Both Sex | 82378628 |

```
tidy_table_01=(tidy_table_01.assign(year = lambda x: x.Year.str[:4].astype(str), Gender = lambda x: x.Year.str[18:].astype(str)).drop("Year",axis=1))
tidy_table_01.sort_values(by ='Order').reset_index().drop(columns = 'index') #Separate the Year cloumns into year, gender, migrant stock
```

| | Order | Name of Country | Notes | Country Code | Type of Data(a) | Migrant Stock | year | Gender |
|---|---|---|---|---|---|---|---|---|
| 0 | 1 | WORLD | NaN | 900 | NaN | 152563212 | 1990 | Both Sex |
| 1 | 1 | WORLD | NaN | 900 | NaN | 93402426 | 2005 | Female |

Applying the melt function and the assign & lambda functions successfully separate the attributes of years, migrant stock, and gender.

## ➢ Step Four: Replace Missing Value

After reorganizing the tables, I discovered that there are spaces which are filled out by ".." rather than observational values. These spaces are the loss of data in the UN dataset. (Pic_7)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 416 | 24 | South Sudan | NaN | 728 | B R | .. | 1995 | Both Sex |

Based on the fact that all missing values are labeled by "..", I decided to replace the spaces that contain ".." with the numeric data of NaN. (Pic_8)

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 23 | 24 | South Sudan | NaN | 728 | B R | NaN | 1990 | Both Sex |

Replacing the missing values by NaN is the optimal solution for the issue. I believe that NaN is able to express the content of the missing space accurately, and it can also be treated as numeric data which enables data visualization. Compared with NaN, replacing missing values with 0 or string performs less desirably. Even though replacing missing value with 0 could also enable data visualization, it does not represent that the object has 0 observational values. 0 could alter the data by giving false information. String such as "Missing_Value" could be a desirable replacement in data cleaning, because it represents the state of the observation accurately. However, string is inconvertible to numeric value. Using string to replace missing values will make data visualization impossible to achieve. Thus, I decided to use NaN to replace all the missing values in the data.

## ➢ Step Five: Categorized Observations

To analyze gender performance in different areas, I decided to categorize observations into country and region & continent. Since there are a considerable number of countries inside the UN dataset, studying the performances of all countries is unrealistic. Instead of studying gender performance by countries, I believe that studying gender performance by continent is the optimal solution. This solution can reduce a significant amount of workload, more importantly, the continent can demonstrate an average performance of the countries. If a

continent has an outstanding performance, the project can conduct a visualization for all the countries within the continent. Therefore, I believe that starting the visualization by analyzing continents is the optimal approach, and I decided to split the tables into countries and continents. (pic_9, pic_10)

```
[12] #Selected Region and Continue to form a new dataframe
     table1_Region_Conti = table1_RM.loc[(table1_RM['Name of Country']=='WORLD')|(table1_RM['Name of Country']=='Developed regions')|(table1_RM['Name of Country']=='Developing regions')|
                                         (table1_RM['Name of Country']=='Sub-Saharan Africa')|
                                         (table1_RM['Name of Country']=='Africa')|
                                         (table1_RM['Name of Country']=='Eastern Africa')|
                                         (table1_RM['Name of Country']=='Middle Africa')|
                                         (table1_RM['Name of Country']=='Northern Africa')|
                                         (table1_RM['Name of Country']=='Southern Africa')|
                                         (table1_RM['Name of Country']=='Western Africa')|
                                         (table1_RM['Name of Country']=='Asia')|
                                         (table1_RM['Name of Country']=='Central Asia')|
                                         (table1_RM['Name of Country']=='Eastern Asia')|
                                         (table1_RM['Name of Country']=='South-Eastern Asia')|
```

| index | Order | Name of Country | Notes | Country Code | Type of Data(a) | Migrant Stock | year | Gender |
|---|---|---|---|---|---|---|---|---|
| 0 | 1 | WORLD | NaN | 900 | NaN | 152563212 | 1990 | Both Sex |
| 1 | 2 | Developed regions | (b) | 901 | NaN | 82378628 | 1990 | Both Sex |
| 2 | 3 | Developing regions | (c) | 902 | NaN | 70184584 | 1990 | Both Sex |
| 3 | 4 | Least developed countries | (d) | 941 | NaN | 11075966 | 1990 | Both Sex |
| 4 | 5 | Less developed regions excluding least developed countries | NaN | 934 | NaN | 59105261 | 1990 | Both Sex |
| 5 | 6 | Sub-Saharan Africa | (e) | 947 | NaN | 14690319 | 1990 | Both Sex |
| 6 | 7 | Africa | NaN | 903 | NaN | 15690623 | 1990 | Both Sex |
| 7 | 8 | Eastern Africa | NaN | 910 | NaN | 5964031 | 1990 | Both Sex |
| 28 | 29 | Middle Africa | NaN | 911 | NaN | 1460530 | 1990 | Both Sex |
| 38 | 39 | Northern Africa | NaN | 912 | NaN | 2403200 | 1990 | Both Sex |
| 46 | 47 | Southern Africa | NaN | 913 | NaN | 1392359 | 1990 | Both Sex |
| 52 | 53 | Western Africa | NaN | 914 | NaN | 4470503 | 1990 | Both Sex |
| 70 | 71 | Asia | NaN | 935 | NaN | 48142261 | 1990 | Both Sex |
| 71 | 72 | Central Asia | NaN | 5500 | NaN | 6630683 | 1990 | Both Sex |

## ➢ Step Six: Visualization Methods

After data cleaning, Table 1, table 3, table 4, and table 5 were ready for visualization. The first figure I plan to conduct is a basic line graph, which includes year as an independent variable, migration value for all the observations as dependent variables, and genders in different colors of lines. This figure will provide general information about the performance of migration of all genders, and it also indicates a change in migration value from 1990 to 2015. See the screenshot below: (pic_11)

```
#Figure 1
#Shows gender development throught out 1990-2015 by country
fig1 = table1_RMC.to_dict()
fig11 = sns.relplot(data=fig1, x="year", y="Migrant Stock", hue="Gender", kind="line")
fig11.figure.suptitle("International Migrant Stock for All Gender From 1995 to 2015(Table 1.1)")
```

However, the line figure is not able to capture the descriptive data of the tables. In order to study values such as maximum, minimum, and average, I will conduct a small multiple figure in violin plot based on regions and continents. The violin plot not only interprets the maximum, minimum, average value of the data, but it also shows the area where the majority of the values are located. See screenshot below for pci_12.

```
#Convert year and migrant stock into numeric
table1_Region_Conti['year'] = pd.to_numeric(table1_Region_Conti['year'], errors = 'raise')
table1_Region_Conti['Migrant Stock'] = pd.to_numeric(table1_Region_Conti['Migrant Stock'], errors = 'raise')
```

```
#Figure 2: Violin graph
sns.set(style="darkgrid")
fig1_2 = sns.violinplot(x="year", y="Migrant Stock", hue="Gender", data=table1_Region_Conti, palette="Pastel1", width=0.9)
fig1_2.figure.suptitle("International Migrant Stock of Region or Continents from 1995 to 2015(Table 1.2)")
```

The violin plot provides the descriptive data, but the limitation of the violin plot is that it does not demonstrate the performance of individual continents. To learn the performances of regions and continents respectively, I plan to conduct another small multiple

figure with histogram. The histograms are able to indicate the performance of every region and continents. Based on the histograms, this analysis can observe which continent performed outstandingly from 1990 to 2015, and another visualization can also be conducted to the countries within that continent. See screenshot below pic_13:

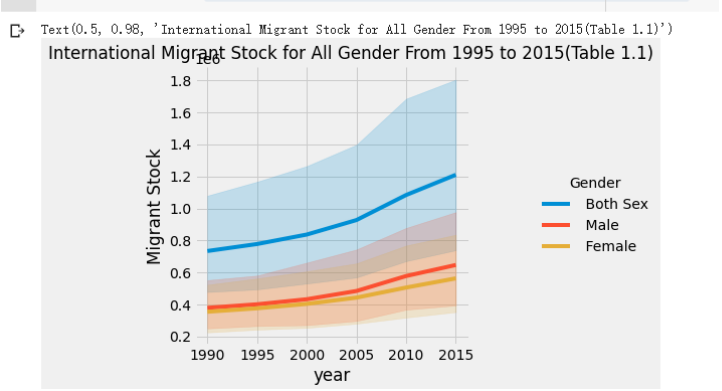
```
#Figure 3 shows the trend of migrant stock in both gender and year based on continents
fig1_3 = px.histogram(data_frame = table1_Region_Conti, x = 'Name of Country', y = 'Migrant Stock', facet_col = 'year', facet_row= 'Gender')
#fig1_3.suptitle("International Migrant Stock of Region or Continents from 1995 to 2015(Table 1.3)")
fig1_3
```

# 3. Figure Interpretation

## ➢ 3. 1 Analysis of International Migrant Stock (Table 1)

Table 1 in the UN dataset contains detailed information of migrant stocks. To study the performance of migration, I conducted a basic line graph for table 1. The line graph (table 1.1) contains year as x value, and international migrant stock as y value. Three gender types are defined by lines with different colors. The basic line graph enables the project to study the gender performance of the world as a whole.

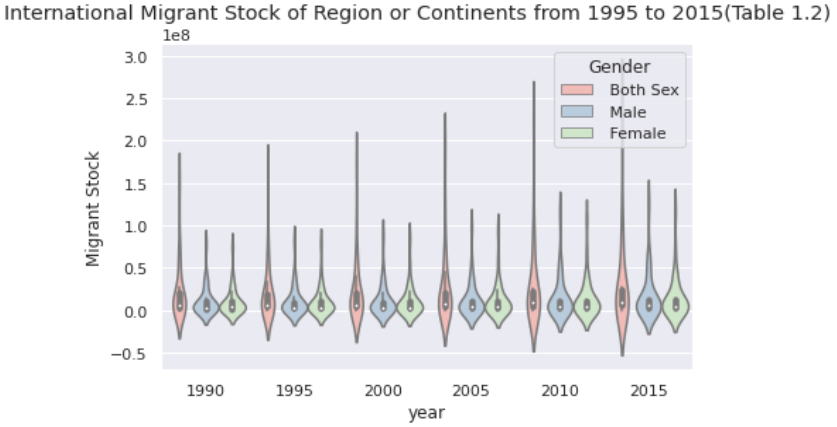
*International Migrant Stock for All Gender and Countries From 1995 to 2015(Table 1.1)*




In the graph, all genders show an increase in willingness to migrate. The overall migrant stock increased steadily before 2005, but the number of migrant stock accelerated after that. In terms of gender performance, male and females had shown similar performance in migration before 2005. Yet, male increased in migrant stock slightly more than females after 2000, the gap between the number of male migrants and number of female migrants grew wider in the following years. It suggests gender difference can impact the choice of migration, but the size of the impact is relatively small.

The table of international migrants performance(Table 1.1) provides an overall analysis for all countries, but it does not indicate descriptive data such as mean, maximum and minimum. In order to capture the descriptive data of migration, I visualized the gender performance based on regions and continents.

By applying a violin plot, I am able to study the descriptive data of the international migrant stock of all genders based on region and continent. The table 1.2 expresses male in blue, female as green, all gender as pink, and the migrant stocks are stored as 1×10^8 for every unit of migration. The table illustrates the fact that the maximum number of people

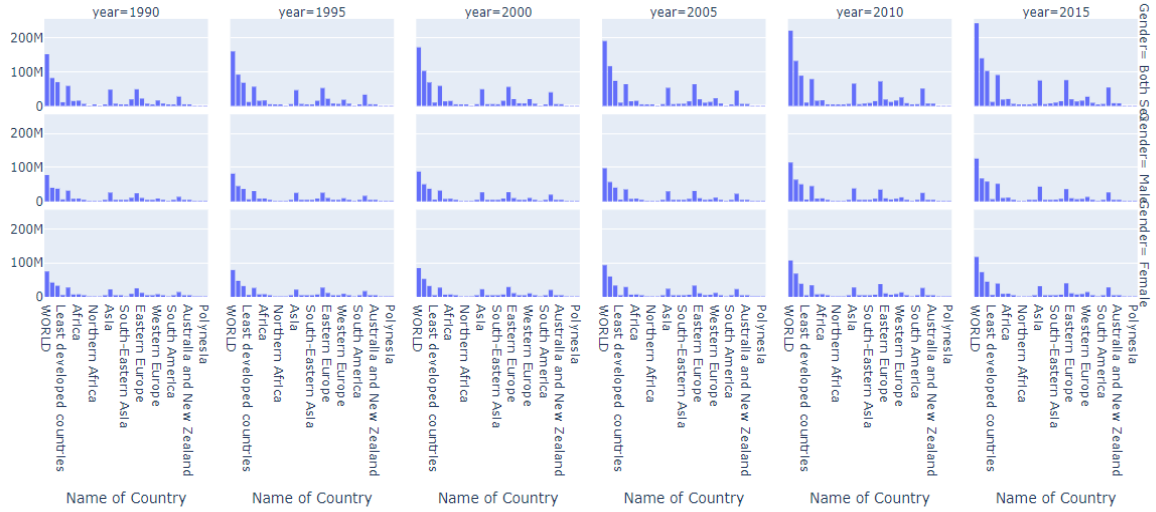


who are willing to migrate increases significantly each year, and it has reached almost 3×10^8 in 2015. After 2000, male showed larger desire in migration compared with females.

However, the mean of migrant stocks remains unchanged. The mean, median, and mode stay in the range between 0 and 0.5×10^8 for all years, while the maximum migrant stock was increasing. It suggests that certain regions had a growing interest in migration, while the majority of the regions remain unchanged. Thus, the violin plot indicates that the maximum migrant stock increased, but the average of migrant stock stays constant.

The violin plot captures the descriptive data of the regions and continents, but it fails to observe the performance of individual regions or continents. To study the performance by regions and continents respectively, I conducted small multiple figures by using histograms.

International Migrant Stock of Region or Continents from 1995 to 2015(Table 1.3)



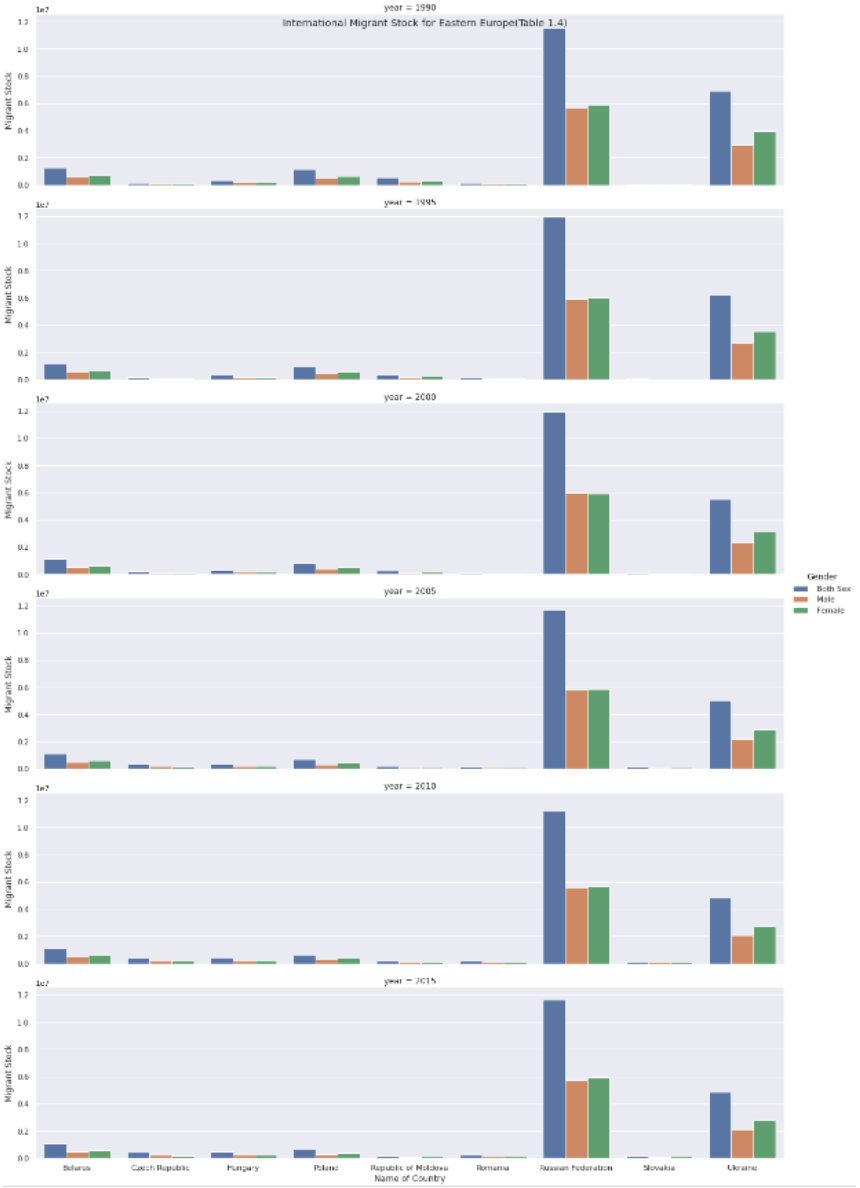

According to figure 1.3, the overall number of migrant stock (given by observation WORLD) increases every five years, and several continents show significant interest in migration. The continents that are particularly interested in migration are Asia, Eastern Europe, and Northern America. They have an outstanding migrant stock compared with others. Asia expresses a strong and increasing desire in migration, their migrant stock grew from 48 million to 75 million from 1990 to 2015. Europe was also interested in migration, their migrant stock grew from 49 million to 76 million from 1990 to 2015. Based on this observation, I believe that willingness to migrant is impacted by regions or continents.

Interestingly, Asia and Europe were very similar in their choice of migration. Asia and Europe have a stable migration stock around 50 million before 2005, and their migrant stock both accelerated after that. Asia has an increase in migrant stock of 12 million in from 2005 to 2010, whereas their previous migrant stock only increased by 4 millions. The growth of migrants in Asia triplel. Similar to Asia, Europe experienced a jump in migrant stock around 2005. Europe's migrant stock increased significantly by 8 million from 2000 to 2005, which is remarkably higher than previous years. The jump in migrant stocks from Asia and Europe is consistent with the jump in migrants stocks for all countries from table 1.1. The consistency of table 1.1 and table 1.2 delivers two possible messages. First of all, all observed countries experienced an increase in migrant stocks in 2005, thus, the migrant stock jumped

around 2005. Second of all, Asia and Europe experienced a significant increase in migrant stock. The increase of their migrant stock contributes to the overall numbers of migrant stocks, therefore, migrant stocks for all countries increase. Both messages can contribute to the jump in migrant stocks.

Meanwhile, table 1.3 discovered a diverse performance of developed countries, developing countries, and least developed countries. The developed countries had higher interest in migration compared to the developing countries and least developed countries. Developed countries always have a higher number of migrant stocks, whereas the least developed countries have one of the lowest migrant stock of all observational regions. Least developed countries have only 5 million migrant stocks, and it remained 5 million for the rest of the years. This analysis suggests that regions and continents have significant influence on migration, and one of the driving reasons behind it are the economic strength and standard of living. Regions which have better standards of living and stronger economic power provide more opportunities for their people. People who are born in these areas are able to make choices on where to live. Compared with the developed world, people in the least developed world have significantly less opportunity. The environment provides very limited resources for them to choose. Thus, Developed countries have higher migrant stock, and the least developed countries have the least migrant stock.

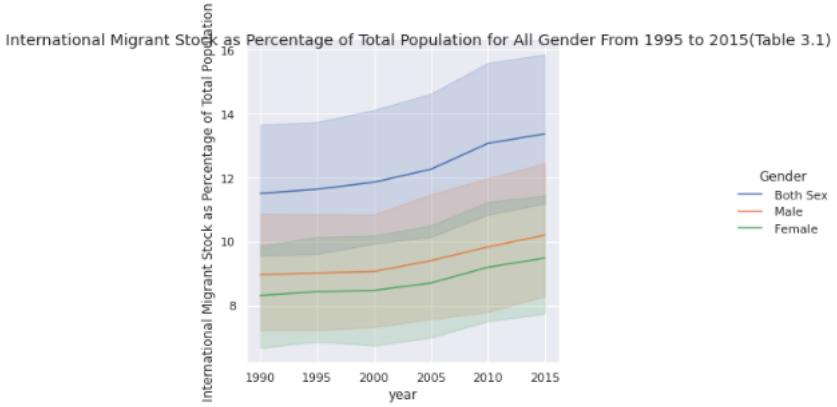
In table 1.3, gender does not perform a remarkable difference in choice of migration. The number of migrant females are only slightly lower than male, and male and females grow in very similar patterns. Yet, Eastren Europe was different from the majority of the regions. According to table 1.3, only Eastern Europe has a high migrant stock in females compared with migrant stock in male. To study the gender performance of male and female in Eastern Europe, I conducted a small multiple figure in histogram



for all the countries in Eastern Europe.

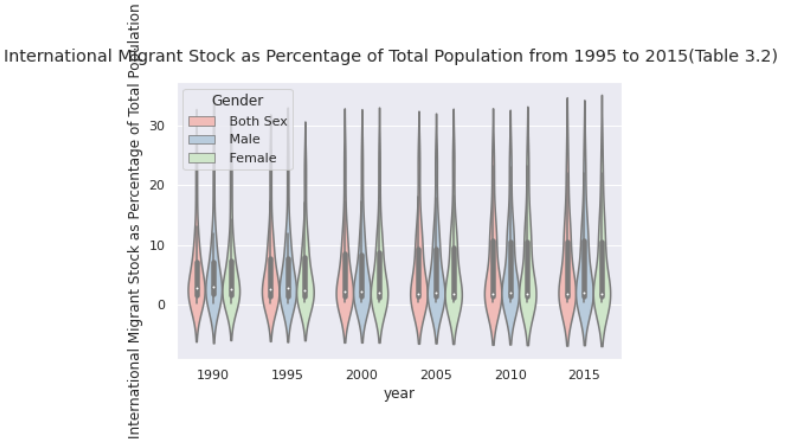
The horizontal small multiple figure is table 1.4, it indicates the gender performance of all the countries in Eastern Europe. In table 1.4, two countries are worth discussing. First of all, the Russian Federation has the highest number of migrants of all the other countries. Male and female do not perform very differently on the choice of migration, they have an equal number of migrant stock in all gender. On the right side of the table, Ukraine also contributes a large number of migrant stock. However, male and females have very diverse opinions on migration in Ukraine. Female in Ukraine preferred migration significantly more than male, and it remained the same diversity from 1990 to 2015. It suggests that male and females have different opinions about migration Ukraines. Thus, gender is one of the factors that shapes the performance of migrant stocks.

## ➢ 3. 2 Analysis of International Migrant in Percentage of Total Population(Table 3)




In the UN dataset, table 3 also provides comprehensive information for migration in percentage of all population. I first conduct a visualization for all of the observed countries. In the line figure of 3.1, genders are described by different colors of lines. The observed years are expressed as independent value, and percentage of migrant stock is represented by dependent value. According to figure 3.1, male and females have similar patterns of growth in percentage of migrant stock. The percentages of migrant stock remained constant before 2000 for male and females, and the percentages boosted around 2005. However, female migrant stock in percentage is always 0.5 percentage lower than male. It proves that the migration percentage is differentiated by genders.

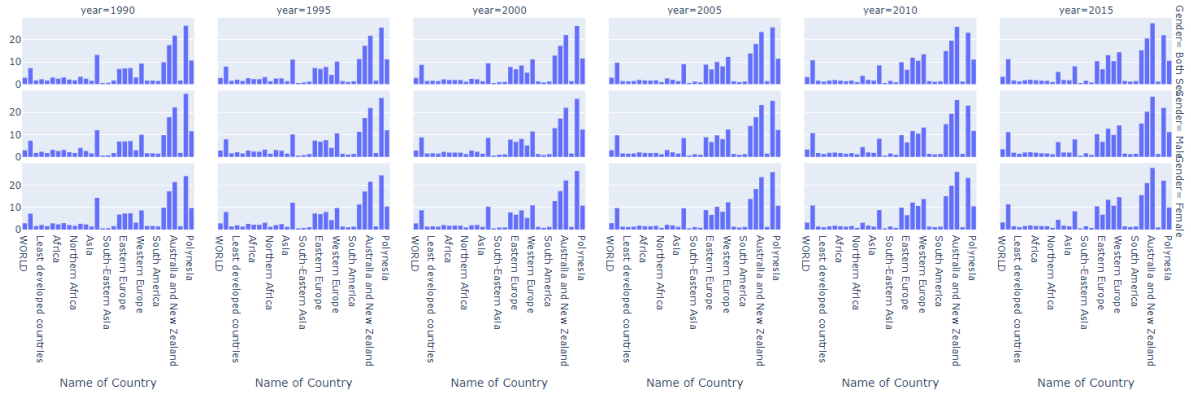
To study the percentage performance of regions and continents, I conducted figure 3.2 and figure 3.3. Figure 3.2 captures the descriptive data of the percentage migrant stock for all genders from 1990 to 2015. Based on the information in the figure, the majority of the regions and continents have their migrant percentages between 0% to 10%. The maximum number of migration percentage reached almost 35% of their total population by



2015, and the minimum number of migration percentage stayed below 0. The majority of the regions and continents had a slow and stable performance in their migrant percentage. They had a very small increase in migrant percentage throughout 1990 to 2015.

Figure 3.3 indicates the performance of regions and continents individually. From figure 3.3, the project is able to learn the continents that have outstanding performances.

International Migrant Stock of Region or Continents from 1995 to 2015(Table 3.3)



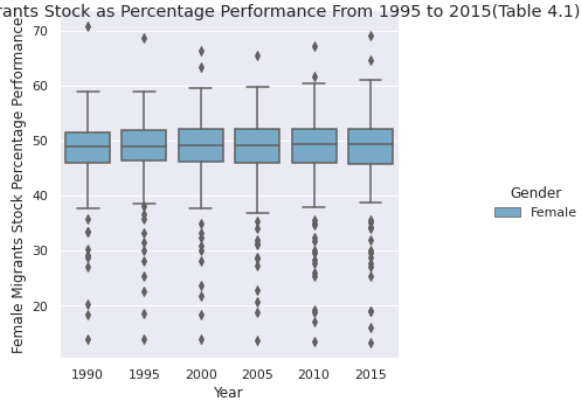

Based on the figures, several regions were particularly interested in migration. Central Asia, Western Europe, North America, Oceania, Australia and Newzeland, Micronesia, and Polynesia express strong interest in migration. All of them had a migration percentage above 10%, which is significantly higher than other areas. Australia & New Zealand and Micronesia, in particular, had migrant percentages above 20% since 1990. They used to have a migrant percentage lower than Micronesia, but their migrant number increased significantly. Its migrant percentage exceeded Micronesia in 2005, and it had approached almost 30% by the end of 2015.

Europe in general also shows an increasing interest in migration. Western Europe, Eastern Europe, North Europe, and South Europe had an uniform increase in migrant percentage. Their migrant percentage was below 10% by 1990, but their migrant percentage increased uniformly, and they had all reached 10% of migrant percentages by 2015. The reason behind this uniform increase in migrant percentage can be explained by cultural background. Since all regions in Europe share similar cultural backgrounds, they performed a similar pattern in their choice of migration. Thus, the choice of migration is impacted by regions and continents, and regions that share the same cultural background will have similar performance in choice of migration.

## ➢ 3.3 Female Migration as Percentage in International Migrant Stock (Table 4)

In the UN datasheet, table 4 studies the performance of female migrants individually, it observed the percentage of female migrants in terms of total number of migrants. Based on figure 4.1, the percentage of female migrant stock

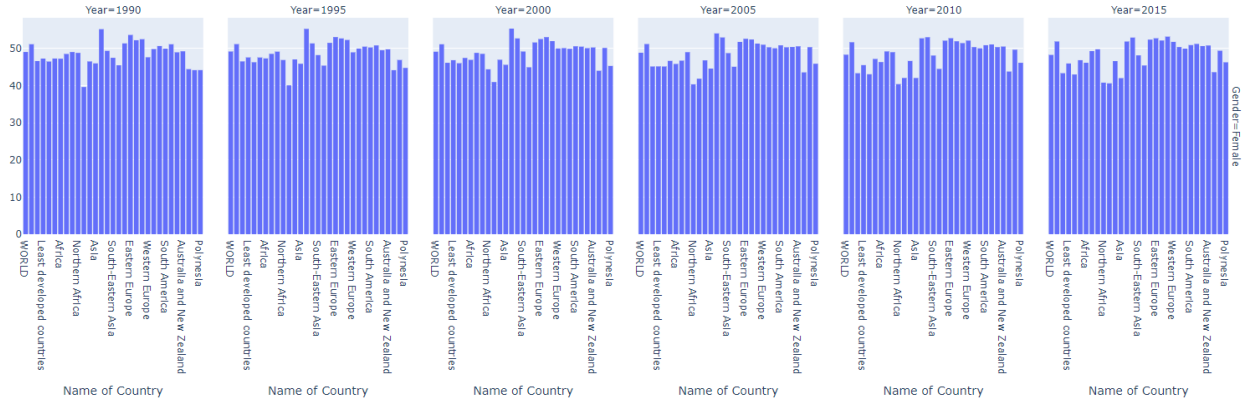
Female Migrants Stock as Percentage Performance From 1995 to 2015(Table 4.1)



performed steadily. The average percentages of female migrant stocks were around 50% from 1990 to 2015, and the maximum percentage of female migrant stock increased slightly. The minimum percentage of female migrants fluctuated around 40%, most of the female migrant percentage were located between 60% to 40% for all the countries. One of the issues that is reflected by this figure is that there are significantly more minimum outliers than maximum outliers. In other words, the percentage of female migrants could be significantly less than the percentage of male migrants, in a small number of countries. In the analysis of figure 4.1, the project discovers that even though the majority of the countries had an equal distribution in male and female migration population, females are more likely to have low migration percentage in a small number of countries. This phenomenon also suggests that choice of migration is influenced by gender differences, but the impact is relatively small, despite the fact that only a small number of countries had an imbalanced migration rate between male and female.

The figure 4.1 reports the descriptive data of the female migration in percentage, but it did not describe the regional performance of it. To study the regional difference in female migration, I conducted a small multiple histogram to examine it.

International Migrant Stock of Female from 1995 to 2015(Table 4.3)



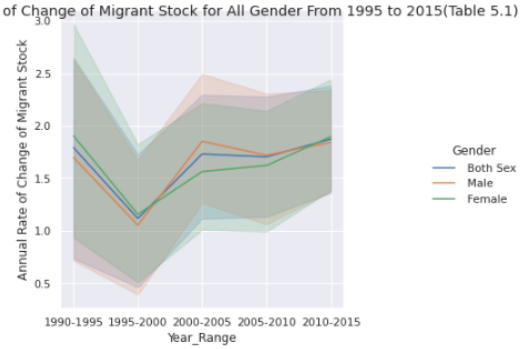

Based on figure 4.3, Central Asia had the highest female migration percentage in 1990, but the percentage had fallen in the following year. Female Europeans were also more interested in migration compared with male Europeans, the majority of European regions had a female migration percentage above 50%, and the percentage remained slightly above 50% until 2015. In general, the percentage of female migration improved overall. Comparing the histogram from 1990 to the histogram of 2015, more female migration percentage exceeded 50% in 2015, and there were only a few regions that had female migration percentage exceeding 50% back in 1990. Based on the analysis on female migration percentage, the project finds that the majority of the countries have an equal distribution in migration between male and female, only a small number of countries have a particularly low migration percentage in females. The female migration percentage was also increasing, by the year 2015, many regions have a higher female migration percentage than male migration percentage.

## ➢ 3.4 Analysis in Annual Rate of Change in Migrant Stock (Table 5)

In the UN dataset, table five reports the annual rate of change for all genders, and the change is measured every five years. To learn the change in migrant stock by visualization, I first conducted a basic line graph for the annual rate of change in all countries. According to figure 5.1, the annual rate of change fluctuates frequently. The rate of change for all countries decreased in the beginning, and it

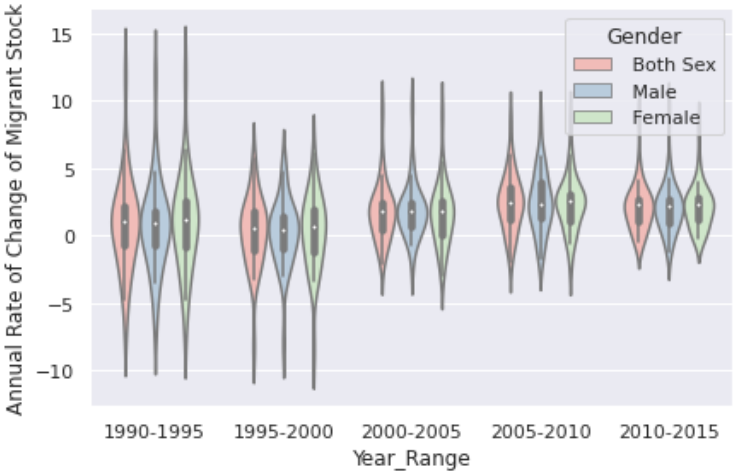
reached the lowest point of 1 at the end of 2000. The rate of change started to grow after 2000, and it approached a peak in 2005. The peak in 2005 reached almost 2, and it has the highest annual rate of change among all observational years. The rest of the years showed a slight fluctuation, but the average annual rate of change had remained stable at 1.8. In general, the annual rate of change in migration is increasing overall.

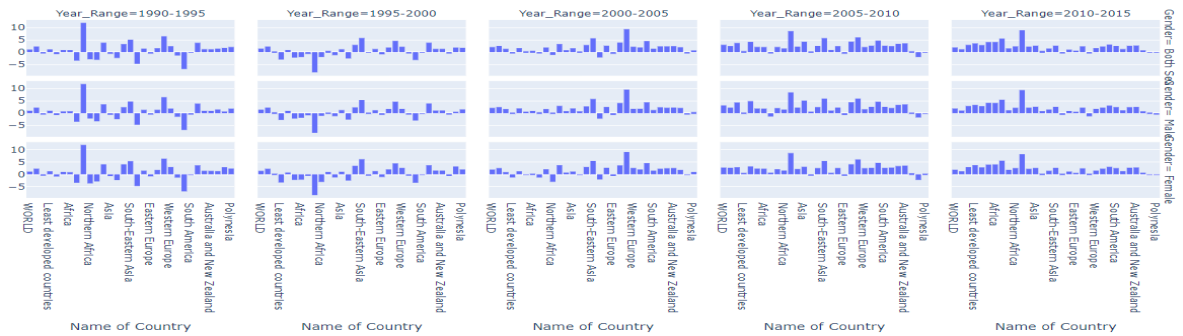
Annual Rate of Change of Migrant Stock for All Gender From 1995 to 2015(Table 5.1)




After conducting a line graph. I also visualized the annual rate of change based on regions and continents. In the figure of 5.2, the data is categorized into regions and continents, and a violin plot is applied to study the descriptive data of the annual rate of change. According to figure 5.2, the majority of the regions had an annual rate of change between -5 and 5. In the year of 1990, the rate of change for regions and continents were very diverse in annual rate of change in migrants stock. The lowest annual rate of change reached -10, and the highest annual rate of change approaches 15. However, such diversity slowly disappeared after 2000. Between the year of 2000 and 2005, the lowest annual rate of change increased, the minimum annual rate of change shrunk to -5. In the following year, the minimum annual rate shrunk continually, until it reached -2.5 by 2010 to 2015. At the same time, the maximum annual rate of change was shrinking as well. The highest point of annual rate of change used to be 15, but until 2015, the maximum annual rate of change lowered to 10. Even though the maximum and minimum of the annual rate of change shrunk, most of the regions and continents experienced a slow growth in their annual rate of change in migration. The average annual rate of change was almost 0 by the year 1990 to 1995, and it gradually climbed to 2.5 at the end of 2015.

Annual Rate of Change of Migrant Stock 1995 to 2015(Table 5.2)




International Migrant Stock of Region or Continents from 1995 to 2015(Table 5.3)

To study the annual rate of change of regions and continents respectively, I conducted a small multiple histogram (Figure 5.3). According to figure 5.3, the annual rate of change for migration varied significantly between the years 1990 to 1995. Middle Africa, in particular, had an outstanding performance in 1990. They had an annual rate of change of more than 10, whereas the annual rate of change for other regions was mostly below 5. Between 1995 and 2000, Middle Africa still acted significantly different from all the other areas. The annual rate of Change for Middle Africa reached the minimum number of annual rate of change, it had an annual rate of change of -7. Regions such as Europe also performed outstandingly, the annual rate of change for Europe rose to 10 between 2000 and 2005, and it soon fell to almost 0 by the end of 2015. Interestingly, the annual rate of changes for all regions and continents were extremely diverse initially. Some regions perform an remarkably low annual rate of change in migration, while other regions perform an remarkably high annual rate of change in migration. The diversity disappears gradually. By the year 2000-2005, half of the regions shift from negative annual rate of change to positive annual rate of change, and by the year of 2010 to 2015, none of the regions shows an negative annual rate of change for migration. In general, the annual rate of change for all regions and continents was rising.

## 4. Discussion and Limitation:

Based on the study from table 1, table 3, table 4, and table 5, the analysis discovers that the world experienced an increase in migration stock from 1990 to 2015 in general. The majority of the regions and continents have a stable rise in the scale of migration, and several continents had extremely outstanding performances. Yet, in the process of data visualization, there are few limitations. First of all, the migration is categorized by regions and continents, instead of countries. Even though analyzing continents reduces workload, it only reflects the average performance of all the countries. If a country performs outstandingly while the other countries perform poorly, the analysis will fail to capture it because the average performance in the continent is normal. Second of all, the UN dataset only reports the observation between 1990 and 2015. In the years after 2015, the world experienced several major changes. Changes such as the American presidential election in 2016, Covid pandemic from 2019 until now, and the War between Ukraine and Russia Federation in 2022 can each produce a significant impact on migration. The UN dataset fails to observe the change in migration for all these events. This limitation might constrain the analysis to make a less conclusive result.

## 5. Conclusion

The analysis on the UN dataset studies the International migration performances from 1990 to 2015. It analyzes the change in international migration value based on different genders and continents. According to the figures from the data visualization, gender difference and regional difference are influential to the choice of migration. Female migrants are slightly less than male migrants in terms of percentage in migration and migrant stock. In a small number of areas, the percentage of female migrants is significantly less than male. In different regions, the migration value also performs differently. Developed countries have the highest migrant stocks over developing countries and least developed countries. Regions such as Asia, Europe, and North America in particular, show outstanding interest in migration. Their migrant stocks are significantly higher among all the other regions in the world. Based on these findings, I argue that the choice of migration is varied based on gender and region, females tend to have slightly lower migration opportunities, and developed countries are more interested in migration.