

INF1340 Final Assignment

DATA VISUALIZATIONS: PART 2 THE UN_DATASET
STEFAN NAVARRETE

Project Outline

This project is a continuation of the INF1340 Midterm assignment, which involved cleaning a data set provided by the United Nations. For the second portion of this assignment, I will be using the data set I cleaned previously to create data visualization that explore questions and explain outcomes. In doing so, I will be grounding the visuals in the context it is related to, while using story telling techniques to communicate the significance of the information graphed. As mentioned in our class textbook, *Human Centered Data Science*, contextualizing data is crucial to making more ethical decisions with data and communicating through story telling allows us to keep in mind the various audiences we are speaking to.

In this document, I am going to introduce the question that guided my exploration of the data and provide an explanation why I believe my question is important to explore.

Introduction

Using the data set provided by the United Nations, titled *Trends in International Migrant Stock: The 2015 Revision*, I explored the following questions:

1. How has migrant and refugee stock influenced Canadas population?
2. Has Canada been allowing more or less migrants and refugees into the country over the years?

Questions around the appropriate number of immigrants and refugees a country should accept is a contentious topic around the world because it raises complex and sensitive issues, including economic, social, and political concerns. Exploring the questions outlined above will provide insight on how many immigrants and refugees Canada believes appropriate to accept and can lead to other exploratory questions which may help inform future immigration policies.

In exploring these questions, I will be creating several data visualizations that utilize the design principles outlined by Dr. Edward Tufte. Tufte is famous for outlining principles for creating data visualizations as his principles have been used for over 50 years. Some of the design principles I will be following in my visuals will include¹:

1. Content focus
2. Comparison rather than mere description
3. Render graphs in high resolution
4. Have a properly chosen format and design
5. Use words, numbers, and drawings together
6. Reflect a balance and a sense of relevant scale
7. Display an accessible complexity of detail
8. Have a narrative quality, a story to tell about the data
9. Avoid content-free decoration, including chart junk

¹ Summary of Tufte's data visualization principles were gathered from "*Principles of Information Display for Visualization Practitioners*. By A1 Globus, NASA Ames Research Center. 1994

Methods

The following section describes the methods used to create the data visualizations for exploring the research questions previously outlined. All visualizations were created by using a previously cleaned data set and by using the Python programming language.

To create the visualizations, I first imported the appropriate libraries and packages that allow my code to be processed, as well as libraries that are specifically used for data visualization. These included numpy, pandas, matplotlib, seaborn, and plotly.

The first data set I explored was the number of migrants in Canada. To gather this data, I imported the code I had from my previous assignment and adjusted Step-6 in my lines of code to isolate the data associated to Canada. The table it produced had migrant data from 1990 to 2015 (collected every 5 years) for males and females, and for both sexes. Using the information in this table, I pulled from the matplotlib library to graph the total number of migrants (see code below, Figure -1). Using information from the same table, I also created a comparable bar graph to show the number of female and male migrants between 1990 and 2015 (see code below, Figure-2).

Figure-1

```
1 #Graph to display number of Migrants In Canada
2 year = ['1990', '1995', '2000', '2005', '2010', '2015']
3 migrants = [4333318, 4864778, 5511914, 6078985, 7011226.0, 7835502.0]
4
5 plt.bar(year, migrants)
6 plt.title("Number of Migrants in Canada")
7 plt.xlabel('Year')
8 plt.ylabel('Migrants')
9 plt.show()
```

Figure -2

```
1 #graph to display female and male migrants in Canada
2
3 year = ['1990', '1995', '2000', '2005', '2010', '2015']
4 male_migrants = [2109652, 2356105, 2659135, 2915838, 3356211.0, 3748267.0]
5 female_migrants = [2223666, 2508673, 2852779, 3163147, 3655015.0, 4087235.0]
6
7 X_axis = np.arange(len(year))
8
9 plt.bar(X_axis - 0.2, male_migrants, 0.4, label = 'Male')
10 plt.bar(X_axis + 0.2, female_migrants, 0.4, label = 'Female')
11
12 plt.xticks(X_axis, year)
13 plt.title("Number of Female and Male Migrants in Canada")
14 plt.xlabel('Year')
15 plt.ylabel(' Number of Migrants')
16 plt.legend()
17 plt.show()
```

Next, I wanted to get a sense of the rate of change over the years for the migrant data. For this, I imported the code I previously cleaned to generate a table containing the rate of change information. With this information, I used a box plot to create a visual of the average rate of change (Figure -3).

Figure -3

```
#Average rate of change for Migrant Data

plt.boxplot(table_5_region['rate of change'])
plt.title("Average Rate of Change in Migrant Stock Over 25 year period")
plt.ylabel('Rate of Change')

plt.show()
```

The next data set I explored was the population data set. Using the same method described above, I imported and adjusted the code to produce an isolated data frame of the population in Canada between 1990 and 2015 (recorded every 5 years). Because I am interested to know how much of the Canadian population is made up of migrant stock, I created a comparable bar chart with information from the migrant data set (both sexes) and the population from Canada data set I just imported (both sexes) (Figure-4). I also attempted to calculate and graph the rate of change for total population but was not successful in making the graph look clean and comparable. As such it did not make it into the results section.

Figure-4

```
1 #Graph to display migrants compared to total population in Canada
2
3 year = ['1990', '1995', '2000', '2005', '2010', '2015']
4 total_population = [27662440, 29299478, 30701903, 32256333, 34126173, 35939927]
5 Migrants = [4333318, 4864778, 5511914, 6078985, 7011226.0, 7835502.0]
6
7 X_axis = np.arange(len(year))
8
9 plt.bar(X_axis - 0.2, total_population, 0.4, label = 'Total Population')
10 plt.bar(X_axis + 0.2, Migrants, 0.4, label = 'Migrants')
11
12 plt.xticks(X_axis, year)
13 plt.title("Number Migrants compared to total population in Canada")
14 plt.xlabel('Year')
15 plt.ylabel('Population')
16 plt.legend()
17 plt.show()
```

Migrant stock as a percentage of the total population in Canada is the next data frame I printed. With this data I again created a box plot to generate a visual of the average percentage of total population (Figure -5).

Figure-5

```
1 #Plot Average Percent of Migrants in Canada
2
3 total= table_3_region.loc[233:1558]
4 plt.title("Average percentage of Migrants in Canada over a 25 year period")
5 plt.ylabel('Percentage of total population')
6
7 total
8 plt.boxplot(total['Percentage'])
```

Next, I imported the data frame containing information on refugee stock in Canada. To do this, I reused the code I used in my previous assignment for table 6A and again, adjusted Step-6 to isolate the data on

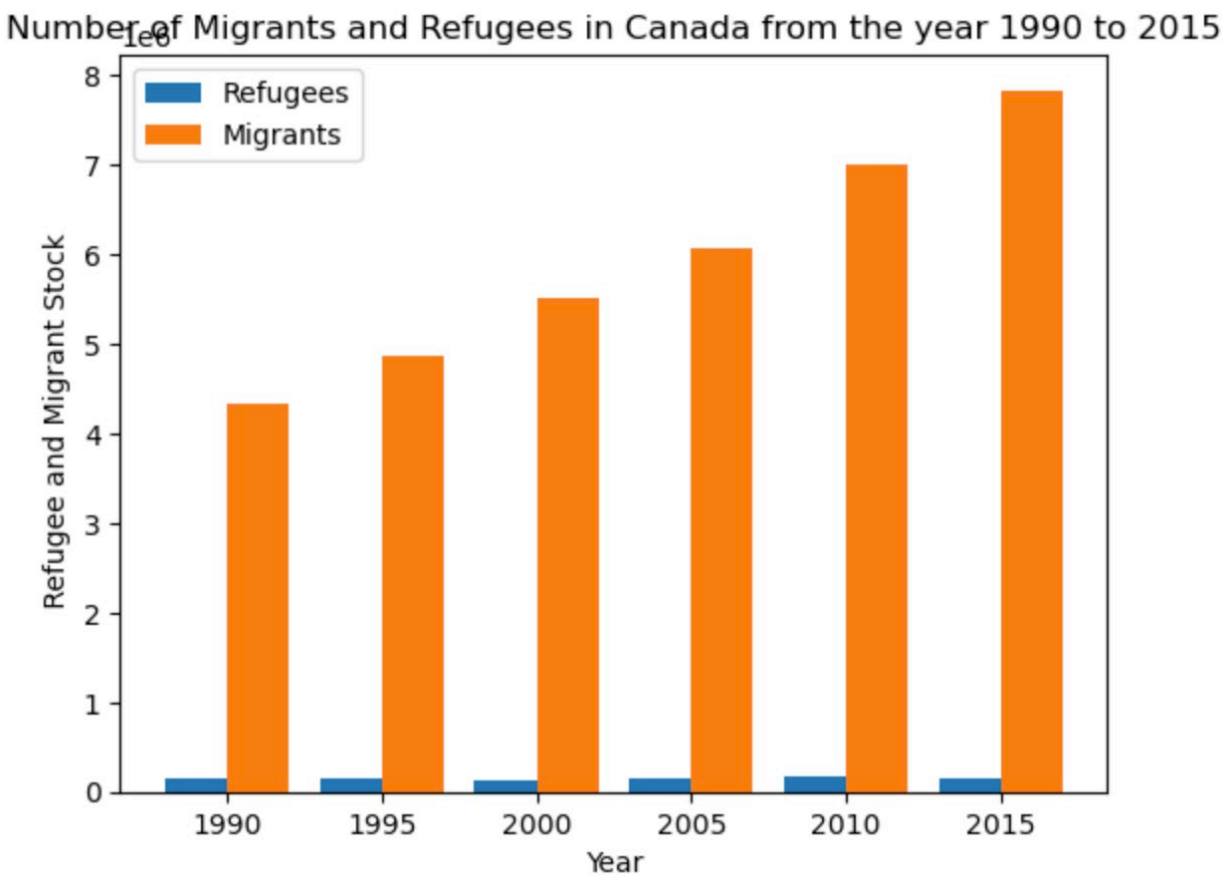
Canada. Using the information from the chart produced, I created another bar graph to display the information from the chart (Figure -6 shows this code).

Figure-6

```
1 #Graph to display refugees
2
3
4 year = ['1990', '1995', '2000', '2005', '2010', '2015']
5 refugees = [145257, 152125, 125154, 147171, 165549.0, 160279.0]
6
7 plt.bar(year, refugees)
8 plt.title("Refugee stock in Canada between 1990 and 2015")
9 plt.xlabel('Year')
10 plt.ylabel('Refugee stock')
11 plt.show()
```

Next, I wanted to graph a visual to show the number of refugees compared to the total number of migrants in Canada. To do this I created another comparable bar graph with data from both sexes of refugees and data from both sexes' migrants. However, because the values of each data set differed greatly, it was impossible to get an understanding of the number of refugees because they were so much less than the migrant numbers (See figure -7). Because this graph did not meet the design principles covered by Tufte, it was not used in the results section.

Figure-7



As a result of the poor visualization, I decided to use a circular packing graph to show the number of migrants, and the number of refugees as part of Canada's entire population between 1990 and 2015. Creating this visual required me to pip install circlify and then import the circlify library. After the installation, I copied a piece of code I found on the internet and adjusted the numbers and names in the data frame to represent Canada's total population, migrant stock, and refugee stock from 1990 to 2015 (Please see Jupyter notebook for code). However, the output of this graph also proved to be insufficient as the size of the circles for each data set are misleading, making it seem like Canada's population is mostly made up of migrants (see Figure -6). Overall, this visual turned out to be a great example of a bad visualization as it breaks several of Tufte's principles such as, not representing a relevant scale and failing to be non-ambiguous. Because of this, this graph did not make it into the results section.

Figure - 6

Refugees and Migrant stock compared to total population in Canada



Failing to successfully graph refugee data in a comparison chart, I settled with graphing the data in a regular bar graph. See code below (Figure-7)

Figure-7

```
: #Graph to display refugees

year = ['1990', '1995', '2000', '2005', '2010', '2015']
refugees = [145257, 152125, 125154, 147171, 165549.0, 160279.0]

plt.bar(year, refugees, color = '#EF4760')
plt.title("Refugee stock in Canada between 1990 and 2015")
plt.xlabel('Year')
plt.ylabel('Refugee stock')
plt.show()
```

Results

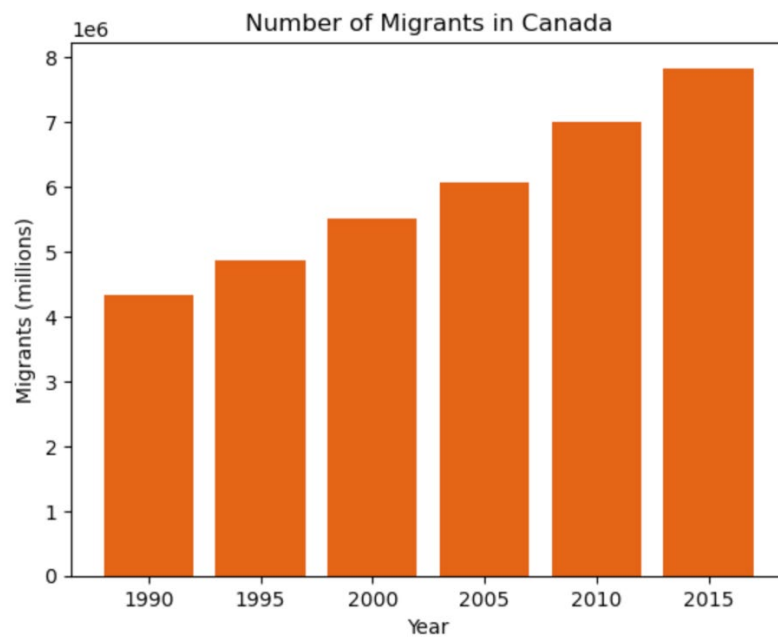


Figure-8, Number of Migrants in Canada

This graph shows the number of Migrants in Canada over a 5-year period from 1990 to 2015. This graph shows a steady incline in migrants.

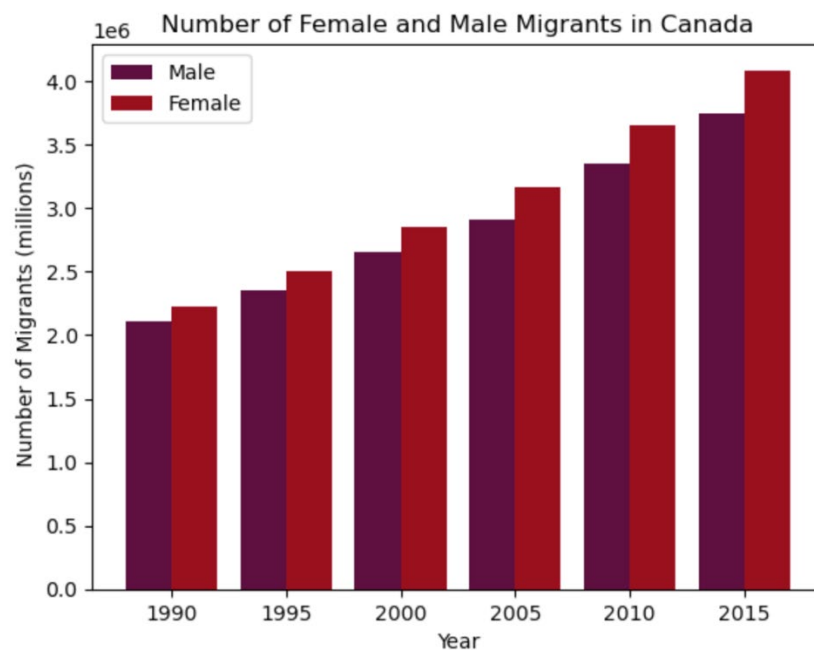


Figure – 9 Number of Female and Male Migrants in Canada

This graph shows the number of Female and Male Migrants in Canada over a 5-year period from 1990 to 2015. There is a steady upward trend in migrant population for both female and males, but each recorded year shows that there are more females than males.

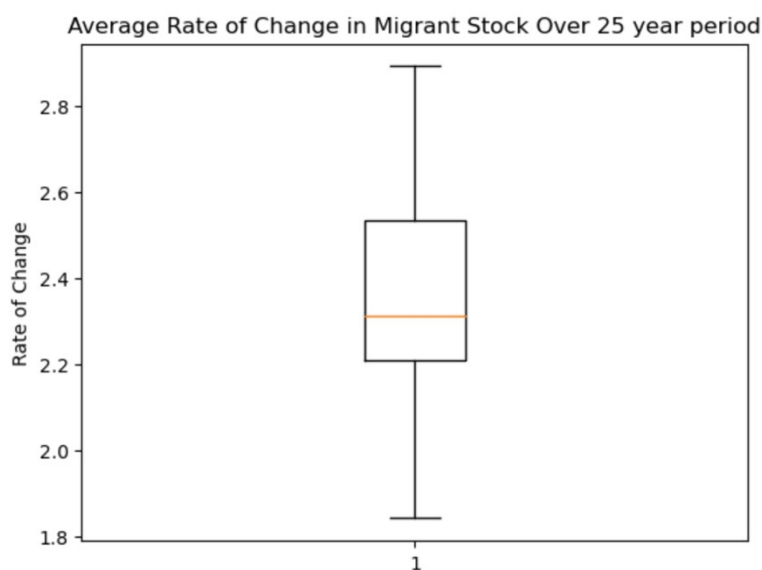


Figure- 10 – Average Rate of Change in Migrant Stock Over a 25 Year Period

This figure shows an average of the rate of change in total migrant stock over a 25-year period with the mean sitting at 2.3, the upper quartile sitting at 2.5, and the lower quartile sitting at 2.2.

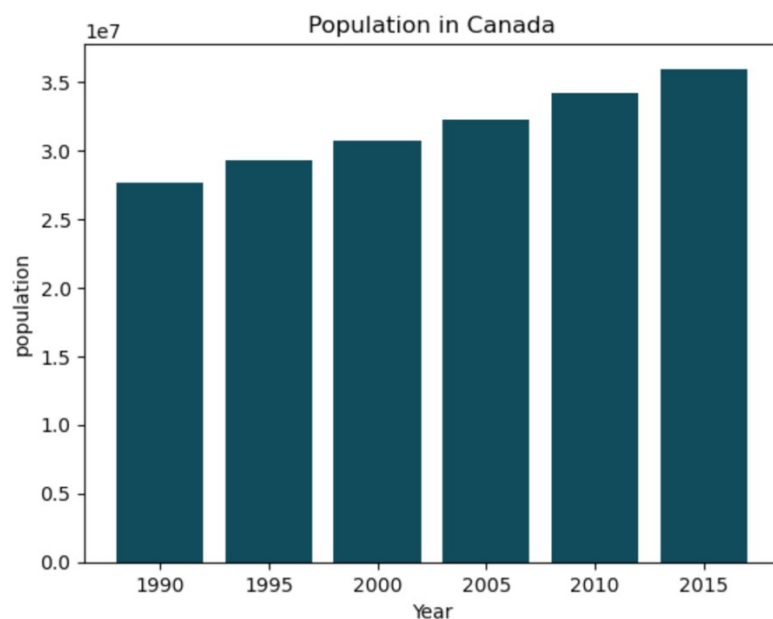


Figure – 11,

This graph shows the population in Canada from 1990 to 2015, recorded every 5 years.

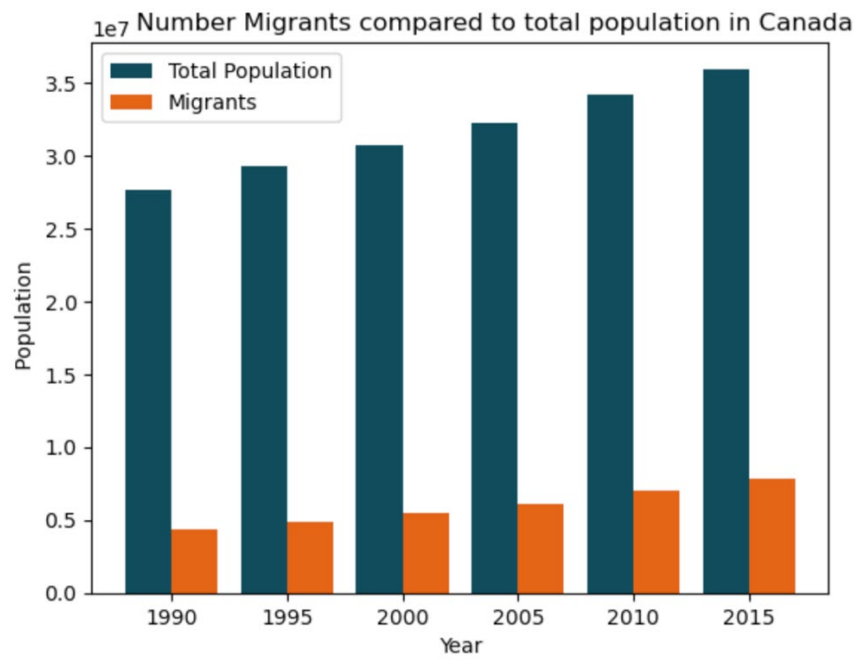


Figure – 12,

This graph shows the number of Migrants compared to total population in Canada every 5 years from 1990 to 2015. Both populations have been steadily increasing every 5 years.

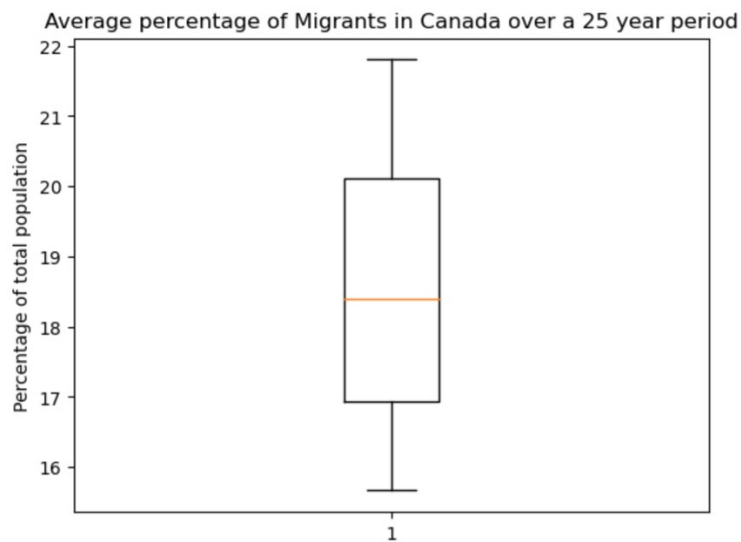


Figure- 13

This box plot shows an average percentage of migrants in Canada over a 25-year period with the mean sitting at 18.5%, the upper quartile sitting at 20%, and the lower quartile sitting at 17.

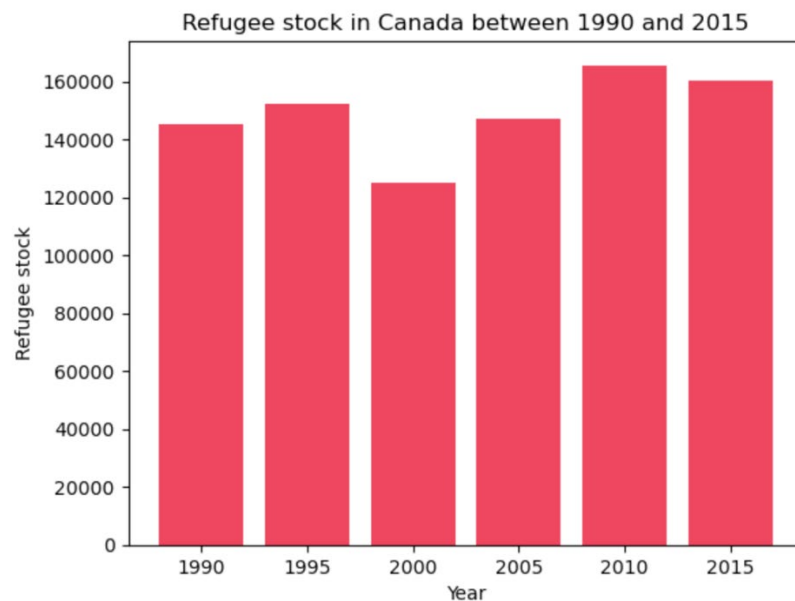


Figure – 14,

This graph shows the number of refugees in Canada between 1990 and 2015, recorded every 5 years. We see that refugee population has fluctuated over the 25 years with the lowest population sitting around 130000 and the highest point sitting around 160000

Discussion

The figures presented above demonstrate the trends and changes in the migrant population over a 25-year period from 1990 to 2015. Figure-7 shows the overall number of migrants in Canada over this time period, and displays a steading increase in the migrant population. Figure-8 provides a breakdown of the number of male and females migrants in Canada, showing that there are more females than males in the migrant population. This is an interesting observation and leads to the question of why are there consistently more female migrants than male migrants? This is likely due to a combination of factors, including differences in employment opportunities and family dynamics. Future research could Investigate the reasons behind this gender disparity among migrants in Canada and its potential impact on Canadian society.

As we can see in Figure 12, as Canada's population rises, so does the number of migrants. Figure 13 shows that over the last 25 years, migrants have on average made up 18-percent of the total population. When we look at the average rate of change in migrant stock, we see that the number of migrants in Canada has been growing at an average rate of 2.3 every 5 years from 1990 to 2015 (Figure-9). This trend indicates that Canada continues to be a popular destination for migrants from around the world.

Examining the refugee data (figure 14), we see fluctuations in the refugee population with the lowest population at around 130,000 and the highest at around 160,000.

Overall, the figures outlines in the results paint a picture of a growing and diverse migrant population in Canada. The steady increase in the number of migrants, particularly females, suggests that Canada is a desirable destination for individuals seeking a new home. The fluctuations in the number of refugees indicates that Canada is also a country that actively engages with global humanitarian issues. It is important for policymakers to continue monitoring and addressing the needs of this growing population to ensure their successful integration and well-being in Canada.

Conclusion

This report examined a subset of data from the United Nations, *Trends in International Migrant Stock: The 2015 Revision*, to explore how the migrant and refugee population has influenced Canada's total population. Overall, the findings show that Canada has been steadily increasing its migrant population from 1990 to 2015 and has remained open and accepting of refugees.

Immigration is a contentious topic because it can have significant economic implications, cultural implications, and political implications. It is therefore important to continue to monitor trends in the migrant and refugee data as it can help inform future policies and decisions around immigration. Future studies could examine the gender disparity revealed here and its potential impact on Canadian society.