

INF1340 Final Project
UN Dataset Visualization Write-Up

Mingli Wang

1003100716

INF1340
Professor Shion Guha
University of Toronto
December 13, 2022

Table of Contents

Introduction.....	3
Migrant Stock and Total Population Overview	3
Migrant Stock Sex Ratio.....	5
Relationship Between Population and Migrants	6
Migrant Stock's Distribution & Impact on Demographic Composition ..	9
Global Refugee Stock Distribution	10
Conclusion	11

1. Introduction

This project looks at the data on global migrant stock including its geographical allocation, sexual composition, and refugee distribution. The project uses data from Table 1 – 6 and the ANNEX table of the UN dataset. The script is programmed using Python on Google Colab.

TABLE	TITLE
Table 1	International migrant stock at mid-year by sex and by major area, region, country or area, 1990-2015
Table 2	Total population at mid-year by sex and by major area, region, country or area, 1990-2015 (thousands)
Table 3	International migrant stock as a percentage of the total population, 1990-2015
Table 4	Female migrants as a percentage of the international migrant stock by major area, region, country or area, 1990-2015
Table 5	Annual rate of change of the migrant stock by sex and by major area, region, country or area, 1990-2015 (percentage)
Table 6	Estimated refugee stock at mid-year by major area, region, country or area, 1990-2015
ANNEX	Classification of countries and areas by major area and region
NOTES	NOTES

2. Migrant Stock and Total Population Overview

To start off exploring the UN dataset, an overview on the size and composition of the global migrant stock and population will help to set the baseline for this explorative research. Figure 1.1 shows global migrant stock of major areas from 1990 to 2015. Europe, Asia, and North America lead other major areas in migrant stocks by considerable numbers. Figure 1.2 shows the global population by major areas, where you can see Asia holds the majority of the population in the world, and it is continuously growing in a stable rate. Whereas the population of other major areas remains relatively flat. Figure 1.1 uses data from Table 1, and figure 1.2 uses data from Table 2 of the UN dataset.

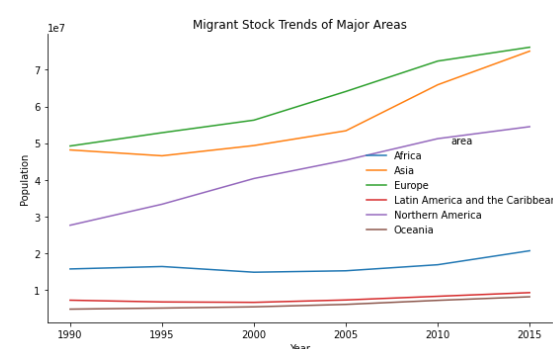


Figure 1.1 Migrant Stock by Major Areas

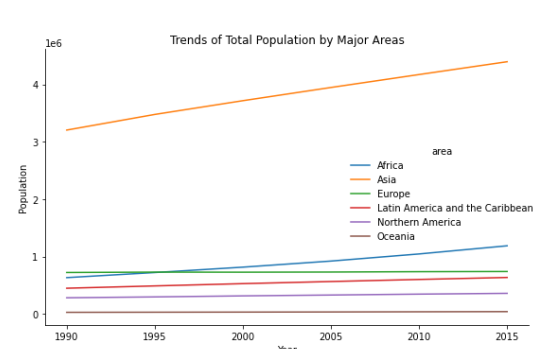


Figure 1.2 Global Population by Major Areas

● Methods

To generate these 2 figures, both figures have undergone a similar process to extract only data of major areas from the whole set. A set of 3 definitions were used for this process, as shown

in Figure 1.3. Lists of major areas, regions, and countries are generated based on the Annex Table given in the UN dataset, and then the lists are used as indexes to extract desired data. The data of major area of year 1990, 1995, 2000, 2005, 2010, and 2015 are melted to form a tidy data frame, as shown in figure 1.4. The unnecessary grid lines and marks are also removed to comply with Tufte’s principles on the data-ink ratio.

```
annex = pd.read_excel('UN_MigrantStockTotal_2015.xlsx', 'ANNEX')
annex.columns = [
    'Code', 'Country or area', 'Order', 'Major area',
    'Code', 'Order', 'Region', 'Code', 'Order', 'Developed region',
    'Least developed country', 'Sub-Saharan Africa']
annex = annex.drop(annex.index[0 : 14])

#Identify unique major areas present in the excel
major_area_list = annex['Major area'].unique()
#Identify unique regions present in the excel
region_list = annex['Region'].unique()

#Define function for extracting country and area from column "area"
def extract_country(table):
    table = table[~table['area'].isin(region_list)]
    table = table[~table['area'].isin(major_area_list)]
    return table

#Define function for extracting regions from column "area"
def extract_region(table):
    table = table[table['area'].isin(region_list)]
    return table

#Define function for extracting major area from column "area"
def extract_majorarea(table):
    table = table[table['area'].isin(major_area_list)]
    return table
```

Figure 1.3 3 Definitions used to separate data of major area, regions, and country or area

	area	year	both sexes
0	Africa	1990	15690623.0
64	Asia	1990	48142261.0
120	Europe	1990	49219200.0
173	Latin America and the Caribbean	1990	7169728.0
225	Northern America	1990	27610542.0
231	Oceania	1990	4730858.0
259	Africa	1995	16352814.0
323	Asia	1995	46548225.0
379	Europe	1995	52842663.0
432	Latin America and the Caribbean	1995	6694640.0
484	Northern America	1995	33341147.0
490	Oceania	1995	5022263.0
518	Africa	2000	14800306.0

Figure 1.4 Migrant stock by major areas

3. Migrant Stock Sex Ratio

In terms of migrant stock’s sex ratio, as shown by Figure 3.1, from 1990 to 2015, males are the dominant sex in almost every major area, suggesting that females are often placed in a disadvantageous position.

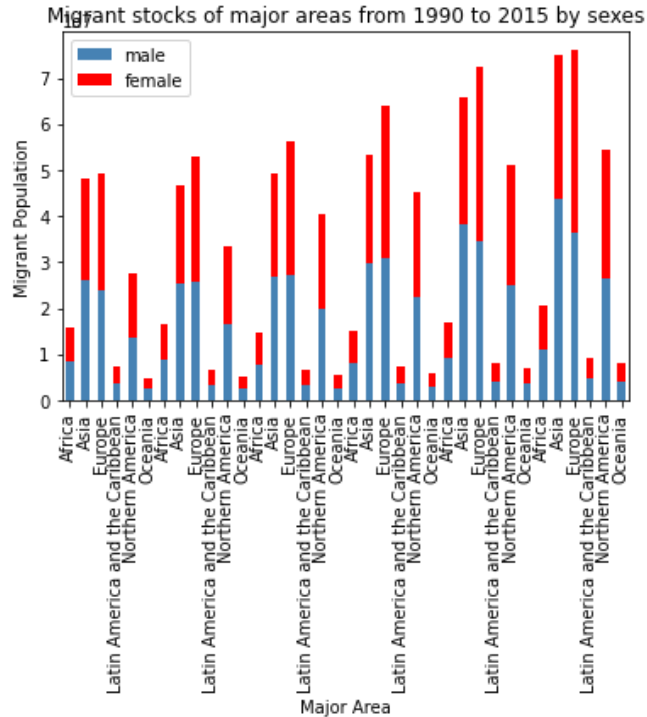


Figure 3.1 Migrant Stock of Major Areas by Sexes, 1990 -2015

According to Tufte’s principles, to better determine the difference between male and female migrant stock in each country, and to enable a more direct comparison, figure 2.3 utilized data from *Table 4 Female migrants as a percentage of the international migrant stock by major area, region, country or area, 1990-2015* of the UN dataset, it counts the number of countries with more female than male migrant stock between 1990 to 2015. As you can see, male immigrants have been the majority over the years, but the number of female immigrants has been slowly increasing year by year.

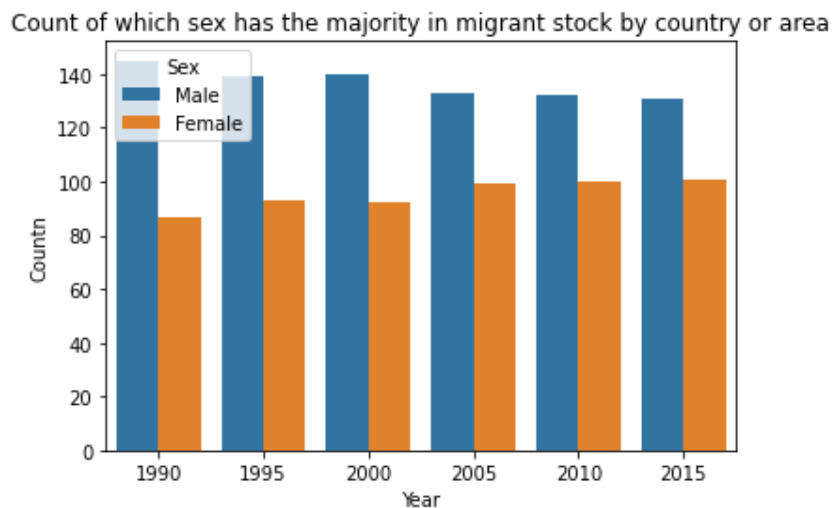


Figure 3.2 Count of countries with more female than male migrant stock between 1990 to 2015

● Method

Figure 3.1 is generated based on the data from Table 1 of the UN dataset, the data for males and females in different years are melted into two columns respectively using `.melt`. The stacked

bar chart then takes input from these 2 columns to generate the figure.

As for figure 3.2, a slightly more interesting approach was taken. The count of countries with more female migrants was calculated using the definition shown in figure 3.3. The definition returns true (1) if more than 50% of a country's migrant stock is female. If the share of females in the overall immigrant stock is less than 50%, then it means that males are in the majority of the immigrant stock and the definition will return an error (0). This definition is applied to the years 1990, 1995, 2000, 2005, 2010, and 2015 and generates six new columns to store the T/F values. Sns.countplot then counts the 1s and 0s in these new columns and visualizes the results in figure 3.1.

```
def is_half (row):
    if row["Female migrants as a percentage of the international migrant stock"] > 50:
        return 1
    else:
        return 0
```

Figure 3.3 Definition to determine if a country have more female migrants.

4. Relationship Between Population and Migrants

Figure 4.1 shows the relationship between a country's population and its migrant stock from 1990 to 2015. As you can see, there are 3 major outliers – United States, India, and China. Compared to India and China, the US has a smaller population but bigger migrant stock, and it is increasing dramatically every year. On the other hand, China leads the world in population, but has very low migrant stock. India has the second largest population in the world, and it is rapidly increasing, although India's migrant stock are decreasing over the years. This potentially suggests that India's population increase is the result of a high birth rate.

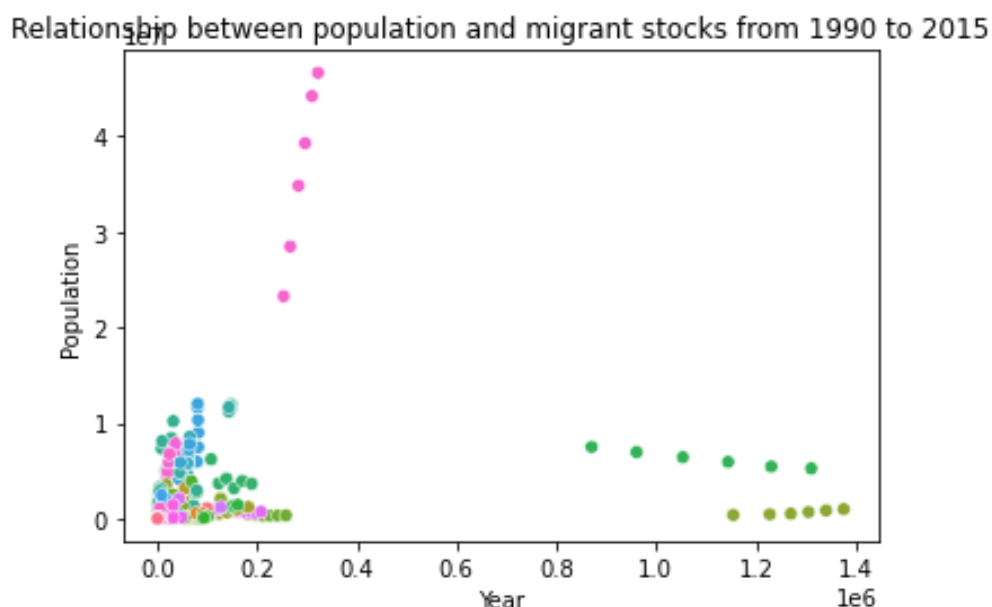


Figure 4.1 Relationship between population and migrant stock, 1990-2015¹

¹ Legends of this figure is omitted

● Method

Figure 4.1 uses data from both Table 1 and Table 2 to show the relationship, only data of country or area is extracted to form a new dataframe using the `extract_country` function as previously mentioned in figure 1.3. The new data frame, as shown in figure 4.2, contains migrants and population data for each country from 1990 to 2015. `Sns.scatterplot` then takes these as inputs to plot figure 4.1. According to Tufte's principles, to emphasize important data and avoid distraction, since the legend is too long to show because it contains all country names, the legend is turned off for better visibility.

	area	year	migrants	population
2	Burundi	1990	333110.0	5613.141
3	Comoros	1990	14079.0	415.144
4	Djibouti	1990	122221.0	588.356
5	Eritrea	1990	11848.0	3139.083
6	Ethiopia	1990	1155390.0	48057.094
...
1549	Samoa	2015	4929.0	193.228
1550	Tokelau	2015	487.0	1.250
1551	Tonga	2015	5731.0	106.170
1552	Tuvalu	2015	141.0	9.916
1553	Wallis and Futuna Islands	2015	2849.0	13.151

1392 rows × 4 columns

Figure 4.2 Dataframe used by the scatter plot

5. Migrant Stock's Distribution and Impact on Demographic Composition

After exploring the distribution and trends of migrant stock across major areas of the world from 1990 to 2015, how does these trends impact the demographic composition of individual countries or areas?

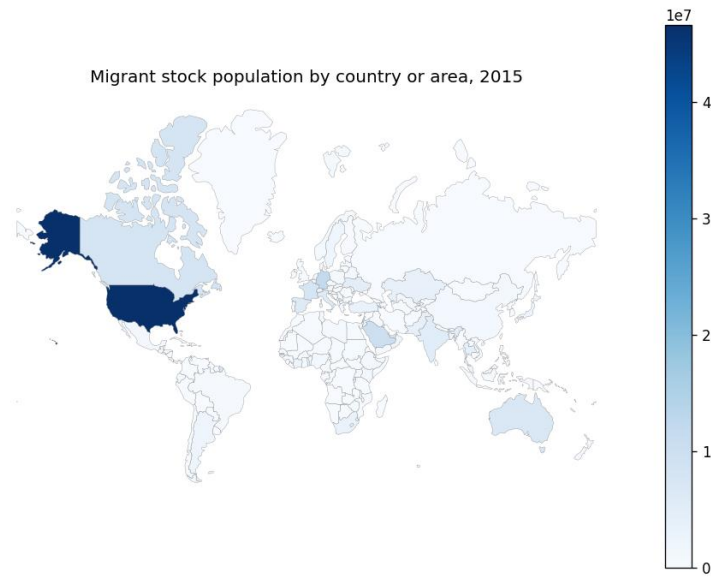


Figure 5.1 *Migrant stock by country or area, 2015*

Figure 5.1 illustrates the results of mapping the migrant stock population for each country or area onto a world map, where blue indicates a high migrant population and white indicates a low migrant population. As you can see, well-known popular countries to immigrate such as North American countries, European countries, Australia, and New Zealand all appear more blue than the rest of the world, with the US being the country with the most migrant stock. However, the size and population of each country or area greatly differs. This illustration only reflects the population of migrant stock of each country, but it is not enough to tell the impact of migrant stock on the country's demographic composition. For example, Canada has a population of about 36 million, and 3 million migrant stock would have huge impact on Canada's demographic. Whereas the US has 322 million population, 3 million migrant stock would have a much smaller impact on the US's demographic composition. Therefore, the impact of the migrant stock on the demographic composition must be discussed in proportion to the size of country or area, and the data in Table 3 will illustrate the other side of the story.

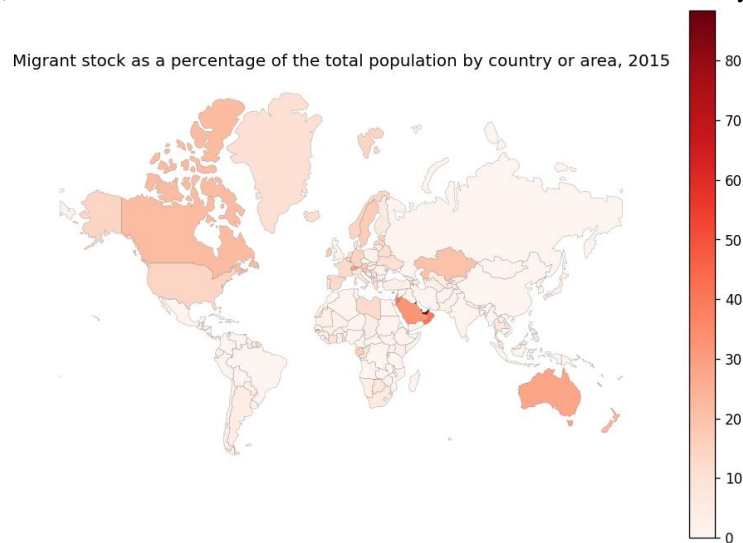


Figure 5.2 *Migrant stock as percentage of the total population by country or area, 2015*

If we instead look at the percentage of migrant stock to the country's total population, and use it to map the world map (Figure 5.2), the US would not be the top in terms of migrant percentage.

Countries including Canada, Australia, and Saudi Arabia have much higher percentage of migrants in their population. In particular, western Asia countries appear to have some of the highest percentages of migrant population, potentially because people are drawn by their rich natural resources.

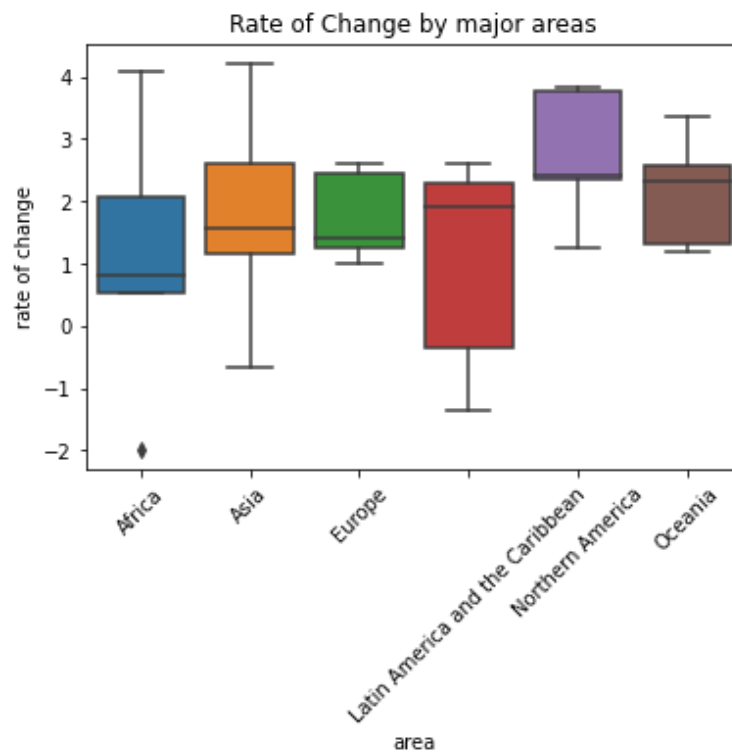


Figure 5.3 Migrant stock rate of change from 1990 to 2015 by major areas

Figure 5.3 supplements the findings we've made above and shows the migrant stock's rate of change from 1990 to 2015 by major areas across the world. Among the six major areas, North America has a consistent high rate of increase of migrants, whereas Latin America and the Caribbean suffers from obvious fluctuations through the years.

● Method

From Table 1 and 3 of the UN dataset, 2 sets of data were extracted to explore: *1) Total international migrant stock at mid-year by country or area of 2015, 2) Total international migrant stock as a percentage of the total population by country or area of 2015*. Which means only data of both sexes from 2015 is included in this case (as shown by figure 5.4). The two sets of data are then projected to a world map using GeoPandas.² Then, the unnecessary grid lines and marks are also removed to comply with Tufte's principles on the data-ink ratio.

² This world map visualization is inspired by: <https://zhuyuhao.com/2020/04/05/data-map-1/>

	area	2015 all
23	Burundi	286810.0
24	Comoros	12555.0
25	Djibouti	112351.0
26	Eritrea	15941.0
27	Ethiopia	1072949.0
...
275	Samoa	4929.0
276	Tokelau	487.0
277	Tonga	5731.0
278	Tuvalu	141.0
279	Wallis and Futuna Islands	2849.0

232 rows × 2 columns

Figure 5.4 Dataframe for migrant stock in by country and area

For figure 5.3, data is drawn from Table 5. First, data for major areas is separated from the rest using the `extract_majorarea` function, and then data of different years stored in separate columns are melted into a single column called “period” (as shown in figure 5.5). Finally the `sns.boxplot` function is called to plot the chart.

	area	period	rate of change
0	Africa	1990-1995	0.826734
64	Asia	1990-1995	-0.673431
120	Europe	1990-1995	1.420702
173	Latin America and the Caribbean	1990-1995	-1.371210
225	Northern America	1990-1995	3.771892
231	Oceania	1990-1995	1.195481
259	Africa	1995-2000	-1.995043
323	Asia	1995-2000	1.165255
379	Europe	1995-2000	1.257523
432	Latin America and the Caribbean	1995-2000	-0.350228
484	Northern America	1995-2000	3.816898
490	Oceania	1995-2000	1.301768
518	Africa	2000-2005	0.521298

Figure 5.5 Dataframe for Migrant stock rate of change from 1990 to 2015 by major areas

6. Global Refugee Stock Distribution

Table 6 of the UN dataset has offered us insight on the status of global refugee stock. The number of refugees has been fluctuating between 1990 to 2015, and Asia constantly holds the most refugees among other major areas, as shown in figure 6.1. Between the year 1990 and 2015, the overall refugee stock of most major areas has decreased except for Asia, whose refugee stock has increased.

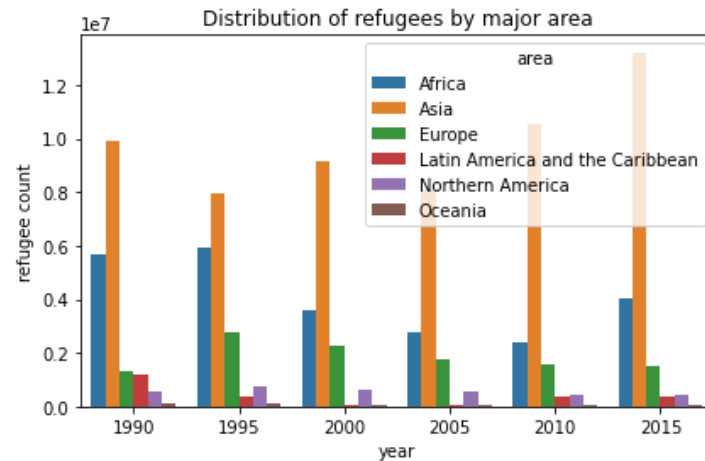


Figure 6.1 Distribution of refugees by major area

If we view the refugee stock distribution by country or area, Africa, Western and Southern Asia have the highest percentage of refugees in their migrant stock, as shown in figure 6.2. Most of the refugee stock is located in developing countries, and developed countries hold a much smaller refugee population.

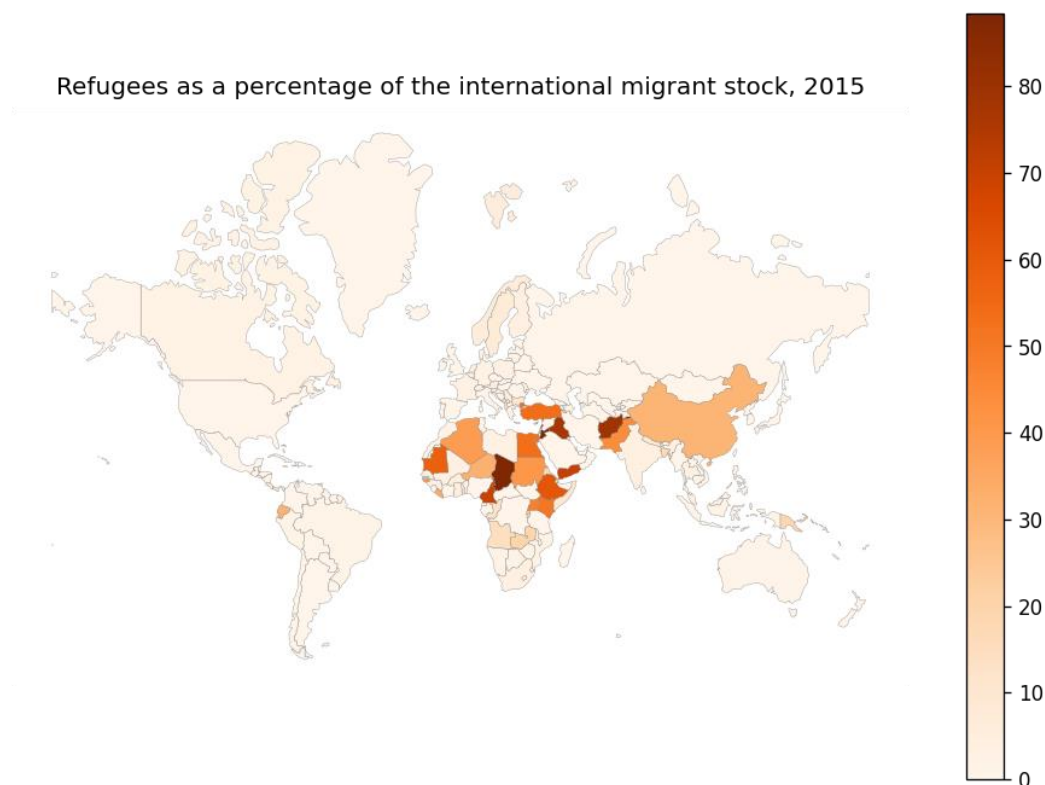


Figure 6.2 Refugees as percentage of the migrant stock by country or area.

● Method

Data from Table 6 of the UN dataset is used to generate figure 6.1 and 6.2. For figure 6.1, data for refugee stock from 1990 to 2015 are firstly melted to have the year variable in one column, as shown in figure 6.3. And then `sns.barplot` function is called to generate the bar chart.

	area	year	refugee count
0	Africa	1990	5687352.0
64	Asia	1990	9937007.0
120	Europe	1990	1321884.0
173	Latin America and the Caribbean	1990	1197198.0
225	Northern America	1990	583450.0
231	Oceania	1990	109680.0
259	Africa	1995	5949953.0
323	Asia	1995	7937682.0
379	Europe	1995	2746090.0
432	Latin America and the Caribbean	1995	352256.0
484	Northern America	1995	775419.0
490	Oceania	1995	92440.0
518	Africa	2000	3609138.0
582	Asia	2000	9175210.0
638	Europe	2000	2283959.0
691	Latin America and the Caribbean	2000	44088.0
743	Northern America	2000	633376.0
749	Oceania	2000	82032.0
777	Africa	2005	2750644.0
841	Asia	2005	8109615.0

Figure 6.3 Dataframe for Distribution of refugees by major area

Figure 6.2 takes a similar approach to figure 5.1 and 5.2, where only the percentage data for 2015 was extracted directly from the raw data set using the `extract_country` function. The list is then fed to GeoPandas to generate the world map visualization. Again, as a finishing touch, unnecessary grid lines and marks are also removed to comply with Tufte's principles on the data-ink ratio.

7. Conclusion

The migrant stock has been increasing over the years 1990 to 2015, and it will likely continue to increase as suggested by the trends. Refugee stock as a part of the global migrant stock is also increasing overall, despite it was once decreasing between the year 1995 and 2005. Among the migrants, females have been the minority most of the time, suggesting there could be more societal, ethical, cultural, or political restrictions on females that prevented them from migration, and have placed females in a disadvantageous position. Overall, developing countries or areas often have more migrant stock than developed countries, and it is heavily influenced by factors such as economic development, politics, and conflicts.