Course: INF1340

Term: 2021 Fall

Name: Kuo-Lun Chang

Student number: 1007618641

Exploratory Data Analysis- UN Migrant and Refugee Stock

1. Introduction

1.1. Problem Statement and Importance

When a person decides to leave his/ her born country, he/she can be categorized as an international migrant or refugee depending on the situation. In the typical migrant analysis, researchers prefer to use the "push and pull" model [1] to explain migrants' motivations. For example, human ancestors migrate to look for abundant food supplies, a safe living environment, and protection from physical dangers. Nowadays, in the modern and globalized society, people migrate to seek better opportunities related to safety, education, employment, living quality, and human rights, among other factors. As economic globalization continues to grow and international migrants keep muddling the concept of country boundaries, the migrant phenomenon and the influence of countries' original citizens become an important issue that deserves a profound discussion.

1.2. Background

The purpose of this project was inspired by international migrants and refugees' data from the United Nations. International migrations and refugees are broadly understood as permanent or semi-permanent changes of residence [2]. In today's political climate, where migrants are counted as one of the primary sources to stabilize the total population in developed countries because fewer young citizens in developed countries are willing to give the same level of births as their parents did in the past. It is essential to investigate how the numbers of migrant and refugee stocks changed across centuries (1990-2015) and compare countries to understand the various factors influencing the migrant process.

In addition, the number of migrants and refugees will force governments to enact different levels of laws on border control. It reflects countries' attitudes and rationalities toward non-nationals into the territory of a nation-state and consequently influences their citizens to treat migrants. Also, border control is a method to prevent illegal migrants from entering the country and unlawfully consuming countries' resources [3]. In the early 1990s, the U.S government started enforcing some border policies to stop illegal migrants at its border. After September 11, 2001, tourists attacks, the U.S. government executed many rigorous regulations to prevent illegal migrants. As a result, these regulations could potentially shape how U.S citizens view foreign migrants [3].

1.3 Research Question and Objectives

This paper will provide multiple comparisons and time series trends to find the trends and underlying border control implications among countries, genders, migrant stocks, refugee stocks, and annual change rates across centuries (1990-2015). The proposed project aims to extract hidden information from the data and further make assumptions that support countries' border control regulations set up rationalities.

The following hypotheses were derived from our research question: (1) developed countries attract more migrants across centuries, (2) the number of male migrants is more significant than female migrants across all seven major areas, (3) refugees account for the most migrant population in less developed countries.

2. Methodology

2.1 Data Collection

The final dataset, which contains nine excel sheets, was collected from United Nations-Population Division- International Migrant Stock data. The data consists of the numbers of international migrants, including refugees, with their genders and 232 countries or area of origin, based on national statistics, in most cases obtained from population censuses, across centuries (1990-2015). In addition, the definition of the numbers of migrants in this research are measured by foreign-born population, foreign citizens, and refugees.

2.2 Data Analysis Method

In the research, I summarized the UN data using descriptive statistics. The research includes creating multiple bar charts to compare gender, countries, regions, migrant stocks, populations, refugee stocks, and annual change rate across centuries (1990-2015). I start to make the comparison by looking into the individual table, trying to figure out the trend within a specific topic. After retrieving valuable information from an individual table, I changed my views by comparing different factors across tables to gain more insights. Also, for some outliers that were extracted from visualization, I will use the news as supportive evidence to illustrate or explain the phenomenon.

2.3 Data Cleaning Method

In this research, I follow Wickham‘s “Tidy data” concepts to perform the data cleaning process [4]. All five principles are well implemented to the dataset. Although there is a total of nine excel sheets in the dataset, three of them are content explanations, annotations, and notes that do not require performing the data cleaning process. Therefore, only the remaining 6 excel sheets were transformed into six tables in Jupyter notebooks to complete the data cleaning process. The below descriptions show the detailed cleaning process on the selected table.

1) Principle 1- informative column names, cannot be values: Across six tables, instead of using variables as column names, ‘Year’ and ‘Gender’ need to be used as separate column names.
2) Principle 2- one and only one variable in each column: Across six tables, the original dataset includes ‘Migrant/refugee number’ and ‘Gender’ together in the same column. Separate them into different columns.
3) Principle 3- variables in cells: Across six tables, ‘Year’ and ‘Gender’ were illustrated as numbers and m/f in columns, where these two variables should appear in where

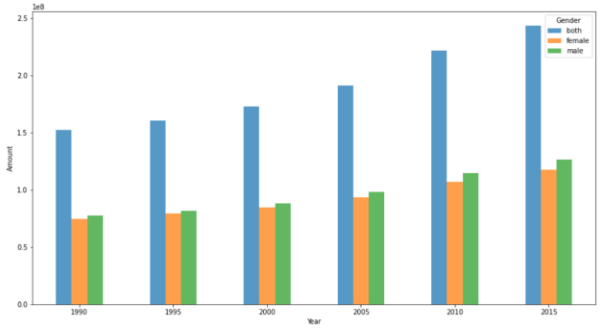
these two variables should appear in cells. Thus, create 'Year' and 'Gender' columns to store data.

4) Principle 4- singular data type in each table: Table six includes refugee population data, refugee percentage data, and the annual rate of change data. Since refugee percentage data and the annual rate of change data use percentage as calculating units, they need to be differentiated from population data (actual number) by separating them into a new table, which is table seven.
5) Principle 5- same observational unit in one table: Across six tables, 'Major', 'Region', 'Notes', and 'Type of data(a)' repeatedly show up. Instead of repeatedly showing in across the table, drop them from tables and store in annex table only.
6) Others:
    a) Use 'Country code' as the primary key across all tables.
    b) Drop non-country rows, such as "world", "developed region," etc. because this information can be referenced to the annex table and calculated by adding up values from the individual country. After dropping non-country rows, re-index the remaining data.
    c) Re-organize table1 migrant stocks, table 2 total population, table 3 Estimated refugee stock at mid-year (both sexes) to integer rather than a string for future calculating.
    d) Inevitably, missing values appeared in the original data source. In order to make the analysis understandable and straightforward, missing values were tagged as 'NaN' and were not counted into the data set. Thus, no statistical process had performed on missing values.

3. Result and Review

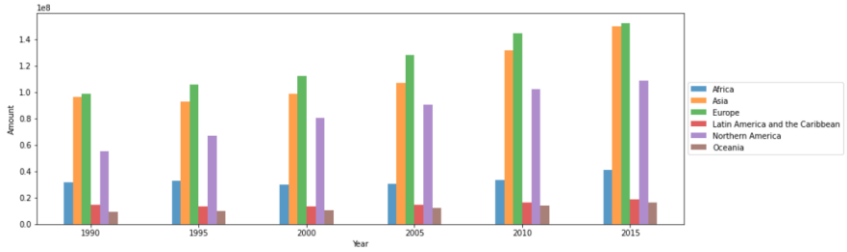
3.1 The number of migrant stocks changes- total number, gender, and region

a) According to the dataset, the total number of migrant stocks was 152 million in 1990 and 243 million in 2015, which is around a 60% increase in the past 25 years. Also, in 1990 the number of male and female migrants was mainly the same and remained at the same level for ten years. After that, we can see that male migrant stocks account for more proportion than female migrant stocks as the time comes to near recent.

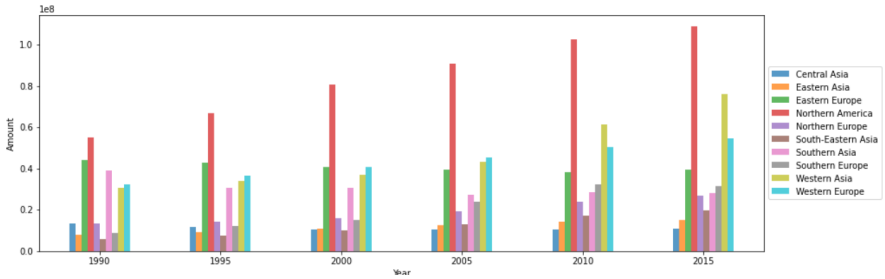


b) From the spatial perspective, Europe, Asia, and North America are the top three major areas in the world that have the most migrant stocks during 1990-2015. Interestingly, even though the total number of migrant stocks kept increasing, these top three major areas' migrant stocks grew in a proportional way, which means migrant

stocks in Asia was roughly 85%-90% of Europe's and North America's is roughly 70%-75% of Asia's.

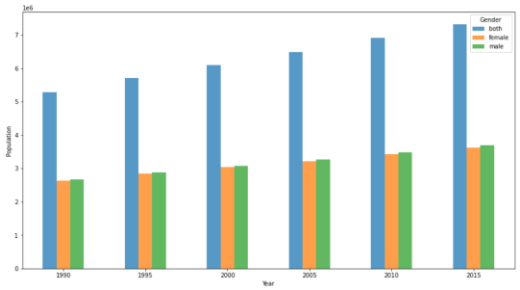


c) If we take a detailed look into Europe and Asia's region, we can see that Eastern Europe remained a steady number of total migrants through the years, while Southern Asia had the most migrant stocks over other Asia regions in 1990. However, as time evolved, Western Asia started to have more migrant stocks than Southern Asia since 1995. At the same time, in 2000, Western Europe started to accept more migrants than Eastern Europe.

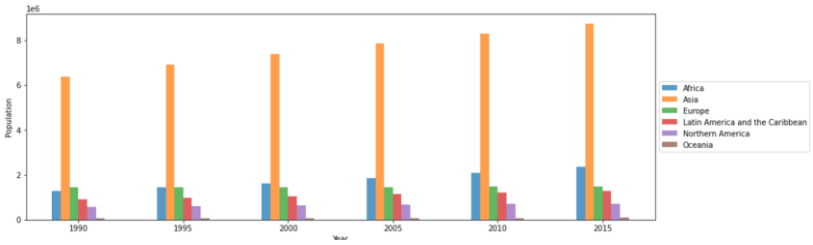


3.2 The number of population changes- total number, gender, and region

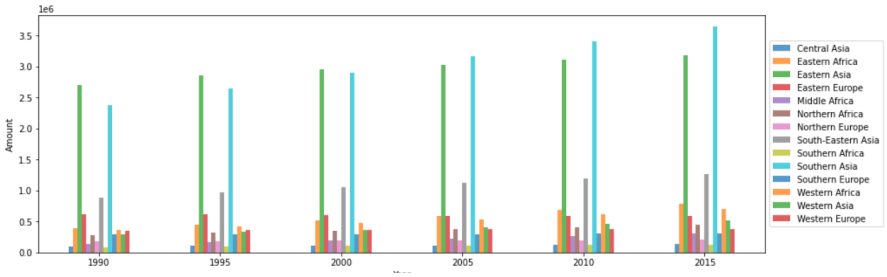
a) From a high level of view, the total world population has grown steadily through the years, and the number of differences between male and female populations remains the same.



b) By looking into the spatial data, Asia provided the most significant number of populations, roughly three times more than the second greatest area, which is Africa. Interestingly, before 2000, Europe had the second greatest population area among the worlds. However, since 2000, Africa started to have not only more population than Europe, but the population also grew much faster than Europe.

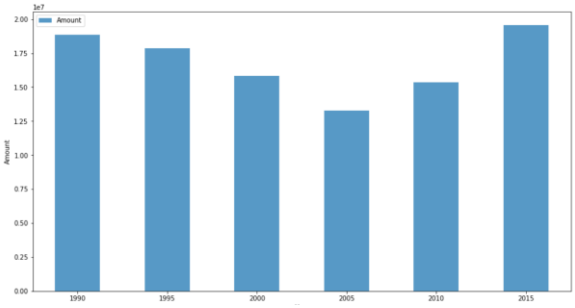


c) Diving into the regional data, Eastern Asia, Southern Asia, and South-Eastern Asia are the places that have the greatest number of populations, where since 2005, South-Eastern Asia replaced Eastern Asia as the largest population provided places. For Africa and Europe, Eastern Africa and Eastern Europe are the top population providers, respectively.



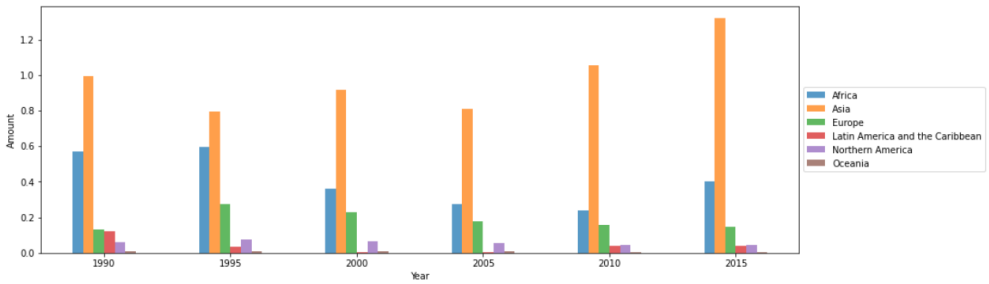
3.3 The number of refugee changes- total number and region

a) The chart shows that the total number of refugees declined between 1990 and 2005 but started to increase significantly since then. In 2015, the number of refugees was the highest among all-time record.

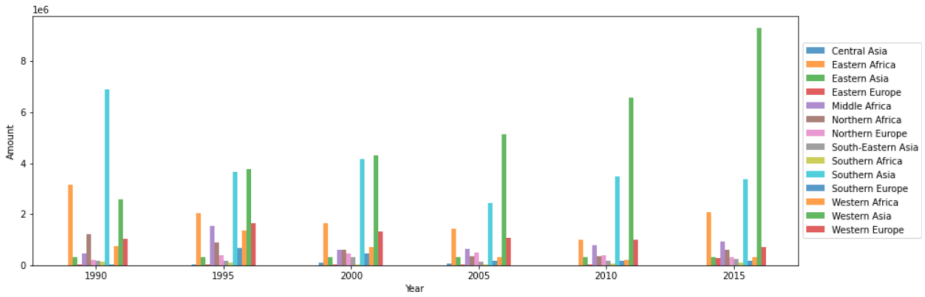


b) From a spatial perspective, Asia is the region that has the greatest number of refugees. Although Asia had the most refugees, it did not strictly follow the global refugee pattern. The number of refugees in Asia countries fluctuated from 1990 through 2005 and steadily increased after 2005.

At the same time, Africa and Europe took second and third place of having the greatest number of refugees. In contrast to other major areas, Northern America remained at the same level as the number of refugees across centuries.

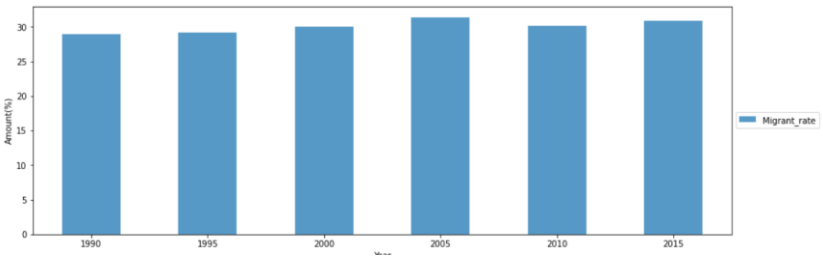


c) Looking into the detailed regional data, the number of refugees grew much faster than other regions in the world, while refugee stocks within other regions fluctuated through the years.

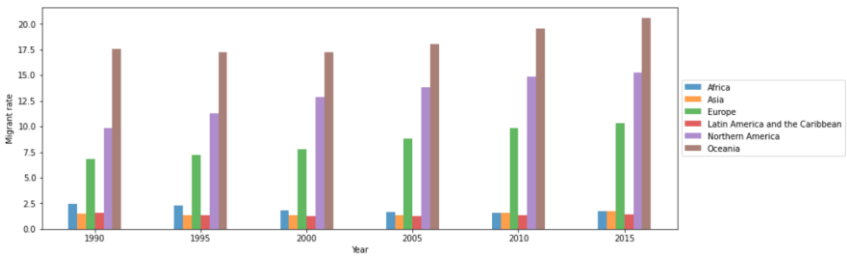


3.4 Compare the number of migrant changes and total population

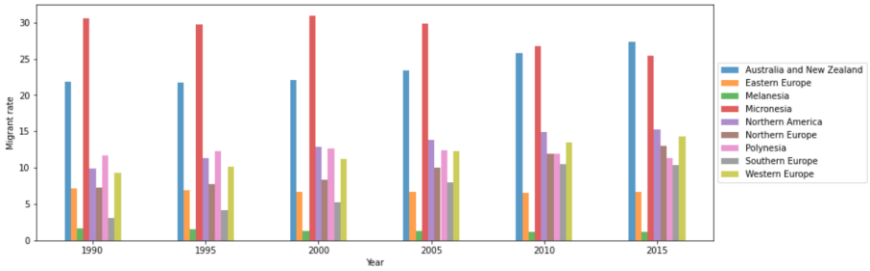
a) Migrant rate (%) = total number of migrant stocks / total population. We can see from the chart that the world migrant rate during the period of 1990-2015 was around 30%. Interestingly, even though the total number of migrant stocks and population grew and fluctuated during this time, the migrant rate remained relatively steady.



b) When looking into a spatial dimension, Oceania, Northern America, and Europe had the highest migrant ratio compared to other areas. Although I mentioned that Europe, Asia, and North America had the greatest number of migrants (3.1), migrant ratios show the different points of view by considering the total population.

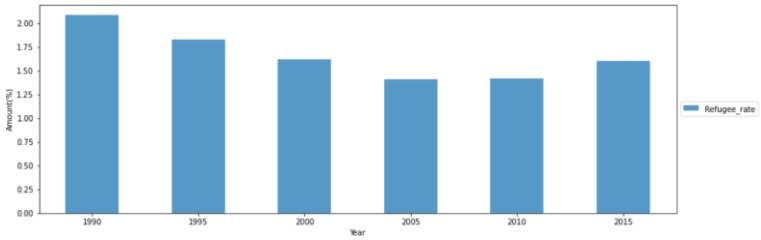


c) If we dive into the top three areas' regional data on migrant rate, Micronesia, Australia, and New Zealand had the highest migrant rate. Before 2010, Micronesia had a higher migrant rate than Australia and New Zealand. However, since 2010, Australia and New Zealand started to have a higher migrant rate. We can also see that Northern America's migrant rate grew faster than other regions. Thus, Northern America had the third highest migrant rate in 2015. Also, interesting to see that Polynesia had the third highest migrant rate before 2000. After that time, other regions migrant rate surpassed it.

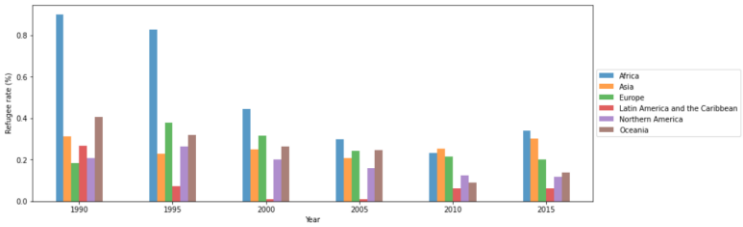


3.5 Compare the number of refugees changes and total population

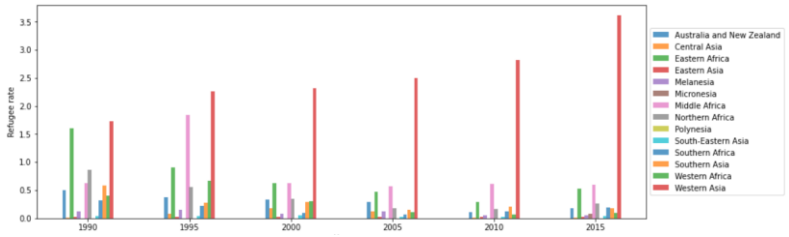
a) Refugee rate (%) = total number of refugee stocks / total population. We can see the graphic that refugee accounts for around 1.75% of the total population around the world. Although there were some fluctuations during the period of 1990-2015, it did not change a lot from the world's total population's perspective.



b) At the spatial level, Africa had the highest refugee rate almost across the records, excepts for 2010, when Asia took the lead in that year. However, the graphic shows that Africa's refugees' rate dropped significantly, almost 50%, since 1990. At the same time, other areas followed the same pattern that refugee rate fluctuated before 2000 and grew up in the following years.



c) From the regional perspective of the highest refugee rate areas, Eastern Africa had the highest refugee rate in 1990 in Africa but dropped in the following years. Interestingly, middle Africa's refugee rate soared in 1995 and returned to its old level. At the same time, Easter Asia had the highest refugee rate among all other regions since 1990 and kept the records till 2015.



4. Implications, Considerations, and Potential Solutions

4.1 Migrant statistics

There are plenty of reasons why people migrate to other countries, one of the main reasons is that people seek better work and education opportunities and living environments [5]. As 3.1.B chart presents, most major areas' migrant stocks grew in a steady pattern, except for Asia. Asia's total migrant stocks grew much faster than other regions and almost caught up the same migrant stocks with European countries in 2015. According to the Asian Development Blog, the People's Republic of China (PRC), the Russian Federation, and Bangladesh are the leading source countries of migrants. At the same time, Japan, Singapore, and Thailand are the leading destination for Asia migrants because of their better educational and economic opportunities. However, since Asia's populations are enormous, the migrant ratio was not as high as Oceania countries. It is worth noting that even though Australia and New Zealand, categorized under the Oceania region, did not have that many migrant stocks in total, they had the highest migrant ratio among all regions in the world because migrants were attracted by their better living environment [7].

4.2 Refugee statistics

According to the Global Citizen's article, five main reasons cause people to become refugees. They are religious/national/social/racial/political persecution, war, gender/ sexual orientation, hunger, and climate change [8]. We can see from chart 3.5.C that Western Asia, Eastern Africa, and Middle Africa had occupied the top three highest refugee rates across centuries. Countries in these regions were the most war-torn countries around the world as well [9]. In addition to the war situation, religious and political persecution, low women's rights, and hungry also played essential roles that worsened the refugee's scenario.

5. Conclusions

In this paper, I have introduced the background of having the research. I then introduced the methodology used in the research to collect data, deal with missing values, and conduct data cleaning and data analysis processes. After extracting results from the data, I have illustrated migrant stocks and refugee stocks' variations between 1990 and 2015 through bar charts visualizations. Finally, some implications on migrant and refugee stocks' issues that drowned from the original dataset and further actions are needed to make the research more comprehensive.

From the migrant populations' perspective, they sought more work and education opportunities and a better living environment. Also, we need to compare both total migrant stocks and migrant rates to retrieve a more accurate picture of migrant statistics because the total population among countries could have huge differences. In short, people migrated because they were pulled by positive motivations most of the time. However, from the

refugees' perspective, most of them became refugees not by their choices but the hostile environment that pushed them to leave their original countries.

6. Limitation and future analysis

In this research, I only focus on the descriptive statistics of migrant and refugee stocks that the United Nations provided from 1990-2015, which lack the most recent data that indeed need to be included and investigated. Also, in the scope of the research, I only consider the time series' impact on migrant and refugee stocks. Thus, only bar plots were used in the research. Further, the correlations among migrant/ refugee stocks, gender, and populations are indeed worthy of detailed investigation and the construction of a migration model for future prediction.

Migrant statistic is not plain numbers. It consists of a wide range of reasons, stories, and impacts on the societies behind the scenes. To complement the results in this report, considering other factors such as economic impacts on border controls, social impact on crime rate, and policy impact on immigration policies could potentially tell different stories. It would also be worth looking into factors regarding cultural influence between original citizens and immigrations within countries. Besides, because the project focuses heavily on the general facts and trends, it would be essential to investigate more hidden implications and potential interest conflicts among countries on the issue of border controls and human rights to reduce the number of illegal immigrants and refugees.

Reference:

[1] Tasleem, Z., Muhammad, S.A., Na'eim Ajis, M. et al. Pakistan to Malaysia: What Expectations Behind Migration?. Int. Migration & Integration (2021). https://doi.org/10.1007/s12134-021-00860-y

[2] National Geographic Society. (2015, August 7). Introduction to Human Migration. https://www.nationalgeographic.org/activity/introduction-human-migration/

[3] Koslowski, R. (2011, February). The Evolution of Border Controls as a Mechanism to Prevent Illegal Immigration. https://www.migrationpolicy.org/pubs/bordercontrols-koslowski.pdf

[4] Wickham, H. (2014). Tidy Data. Journal of Statistical Software, 59(10). https://doi.org/10.18637/jss.v059.i10

[5] Cerdin, JL., Diné, M. & Brewster, C. Qualified immigrants' success: Exploring the motivation to migrate and to integrate. J Int Bus Stud 45, 151–168 (2014). https://doi.org/10.1057/jibs.2013.45

[6] ÉDes, B. (2019, February 4). Migration in Asia: Where is Everybody Going? Asian Development Blog. https://blogs.adb.org/blog/migration-asia-where-everybody-going

[7] Labour migration in Asia and the Pacific. (2020, October 23). International Labour Migration. https://www.ilo.org/asia/areas/labour-migration/WCMS_634559/lang--en/index.htm

[8] Rueckert, P. (2017, August 14). 5 Of The Biggest Reasons Why People Become Refugees. Global Citizen. https://www.globalcitizen.org/en/content/reasons-why-people-become-refugees/

[9] Nag, O. S. (2019, June 6). The World's Most War-Torn Countries. WorldAtlas. https://www.worldatlas.com/articles/the-world-s-most-war-torn-countries.html