

Table of Contents

1. Introduction
2. Methods
3. Results
4. Discussion
5. Conclusion
6. Work Cited
7. Appendix

1. Introduction

After cleaning the UN Migrant Dataset in the mid-term with Tidy Data principles, we now revisit the data set with the aim of executing exploratory data analysis (EDA). We will follow the principles of EDA as elaborated by John Tukey in his book “Exploratory Data Analysis” and use it to tell a story. We will also employ Edward Tufte’s “Envisioning Information”.

In the following sections, we will dive into the procedure I used for each table (Chapter 2), followed up by the results containing the story the data tells us. Lastly, we will wrap up the final with potential next steps and interesting insights from the data (Chapter 4 & 5). The Appendix contains any interesting graphs that don’t fit with the narrative.

Data on migrants and refugees is a great way to peer into how humans move across borders. Now this movement may be voluntary or involuntary, regardless, you can seek patterns that may force people to move such as wars. During my exploration, I found trends in the Middle East that aligned with the Syrian Civil War and how it shaped movements to neighboring countries like Turkey and Lebanon as well as non-neighbouring countries like Germany.

2. Methods

There are four components in the Google Colab file:

1. **Imports** – a cell that imports relevant graphing libraries. Predominantly seaborn and matplotlib were the libraries of choice
2. **Data Cleaning** – cells that convert data type of columns where necessary and replace some values with NaNs.
3. **Exploratory Data Analysis** – these cells play around with the data and create different plots

4. **The Story** – this set of cells tells a story with the data. This section feeds into the Results section of this paper.

In terms of the process, I followed key takeaways – which we discussed in class – from Tukey’s book on EDA which recommends essentially to do the following,

1. Sort
2. Group
3. Subset
4. Compare

The steps in this process can be interchanged in terms of order as well. Along with this, I tried to minimize *chartjunk* as mentioned in Tufte’s book. Any unnecessary information was removed while critical information – like labels – which help readability were kept in. The principle of *small multiples* was used whenever I tried to plot data with categories like region and year. This is exemplified with the graphs with facet wraps. Additionally, to be as human-centered as possible, I used a colorblind color palette from Seaborn in order to make visuals as inclusive as possible.

3. Results

The exploratory data analysis began with looking at the refugee stock at a zoomed out level. I looked at firstly, a box-plot of refugee stock across continents.

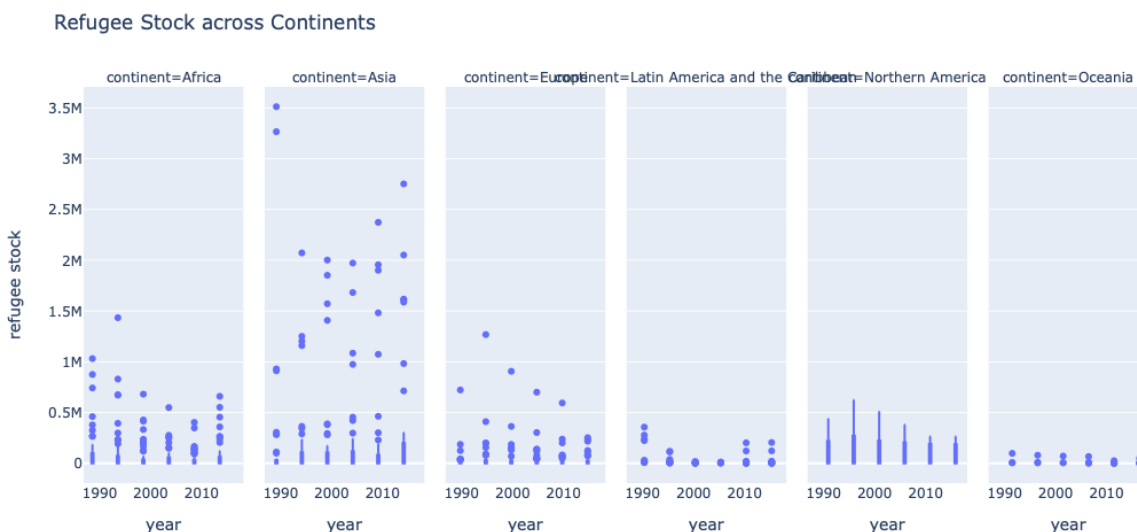


Figure 1: Boxplots of Refugee Stock Across Continents and Years

A clear pattern that can be detected from Figure 1 is that there is much higher variability in the number of refugees in Asia compared to all other continents. It has a median refugee stock much greater than 0 – as seen in Latin America as an example – with plenty of outliers.

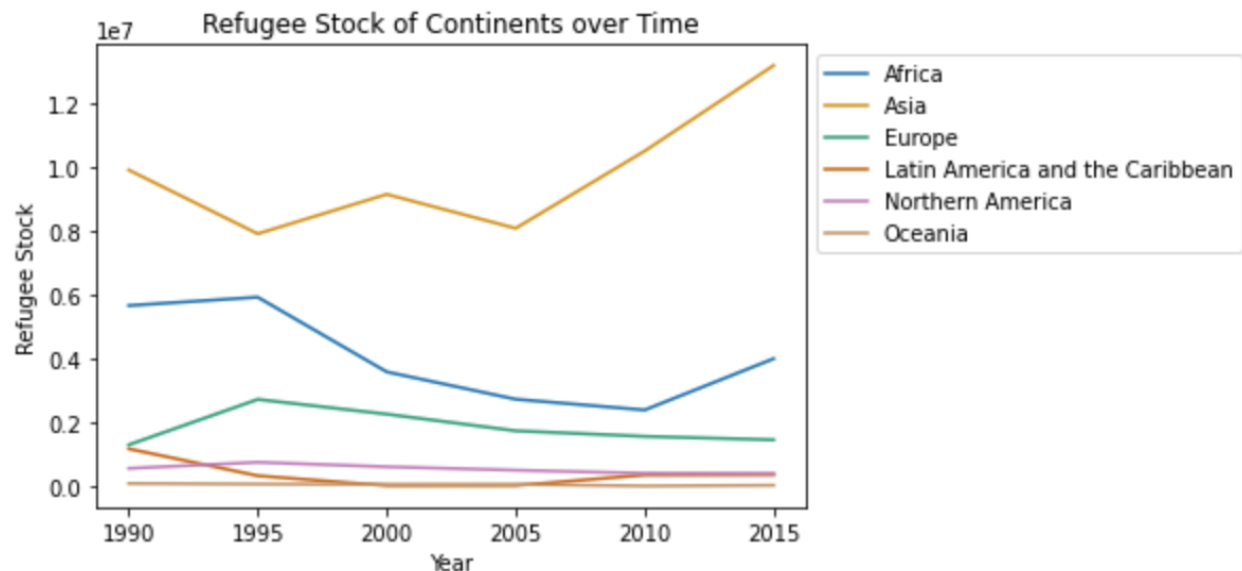


Figure 2: Refugee Stock of Continents over Time

We can see from the graph above that Asia is a clear, yet sad, leader in refugee stock. The pattern we see in Figure 2 can be attributed to two things. Firstly, Asia is the largest continent and as such has much more variability than other continents. Secondly, Asia has much more wars and conflicts compared to say North America or Oceania. This higher amount of conflict leads to higher migratory and refugee movements.

Before I proceeded, I would like to dive deeper into the Asian continent. I filtered the data for the rate of change of refugee data and kept only the data for Asia. Then I grouped it by region and created a facet wrapped box plot below.



Figure 3: Boxplots of Rate of Change of Refugee Stock across Time with regions in Asia facet wrapped

No conclusive results can be drawn from these boxplots in Figure 3. Some regions have medians fluctuating around -10 while others hover just above 0. I had to go back to refugee stock and look at it again from a different angle. To explore if there are any leaders in refugee stocks, I sorted the refugee stock in descending order and filtered for the countries with highest refugee stock for the year 2015.

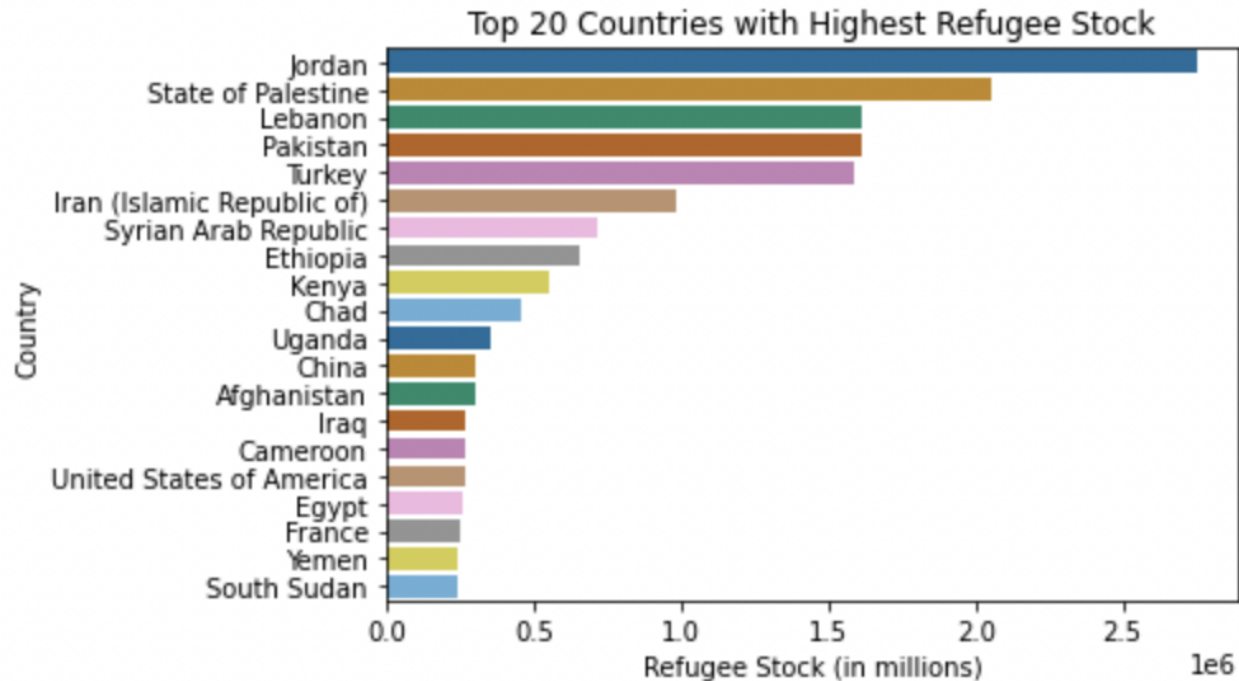


Figure 4: Top 20 Countries with Highest Refugee Stock in 2015

A glance over the first few rows of Figure 4 and another interesting pattern emerges. Four out of the top five countries with the most refugee stock are in or are in the “Western Asia” region of Asia – more commonly known as the Middle East. Seeing this pattern I was quickly able to gauge that this pattern that we are seeing here is being caused by a war: the Syrian Civil War. Jordan had the highest number of refugees with a behemoth 2.7 million refugees. The refugee crisis that emerged from this conflict, which started out in 2011 (BBC News), led to refugee movements to neighboring countries like Turkey, Iraq, Lebanon, Jordan, and Egypt. All these countries make it to the countries in the world with the highest refugees. While the data does not indicate where refugees originated from, a small proportion may be in countries like the United States of America in the form of asylum seekers.

I wanted to gauge whether or not my hypothesis is correct. The next step was to dive back into the rate of change of refugee stock – to create a violin plot of rate of change of refugee stock.

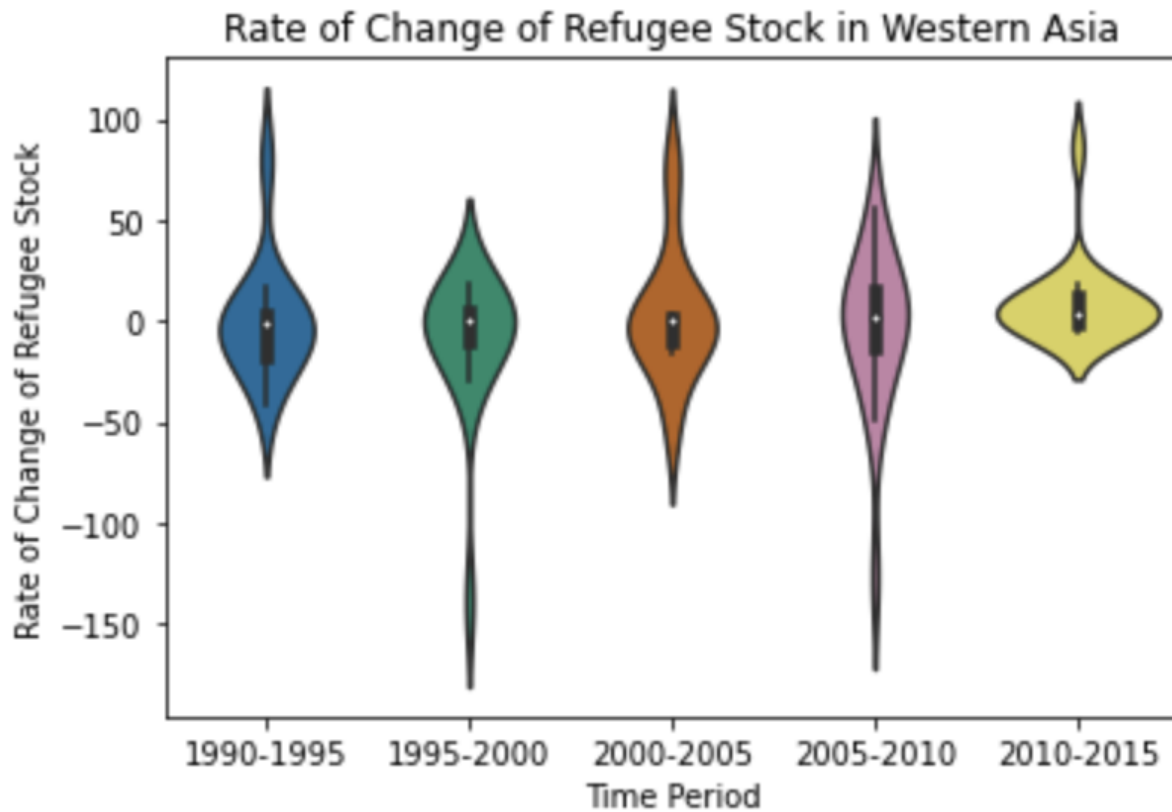


Figure 5: Violin Plot of Rate of Change of Refugee Stock in Western Asia

While the median (indicated by the white dot in the violin plots) is close to zero, you can note that in the yellow violin plot for 2010-2015 of Figure 5, we can see that there is very little probability of the Rate of Change being in the negative. Additionally, there seems to be some outliers in the positive side of the y-axis, indicating an increase in the number of refugees. This was likely caused by the Syrian Civil War.

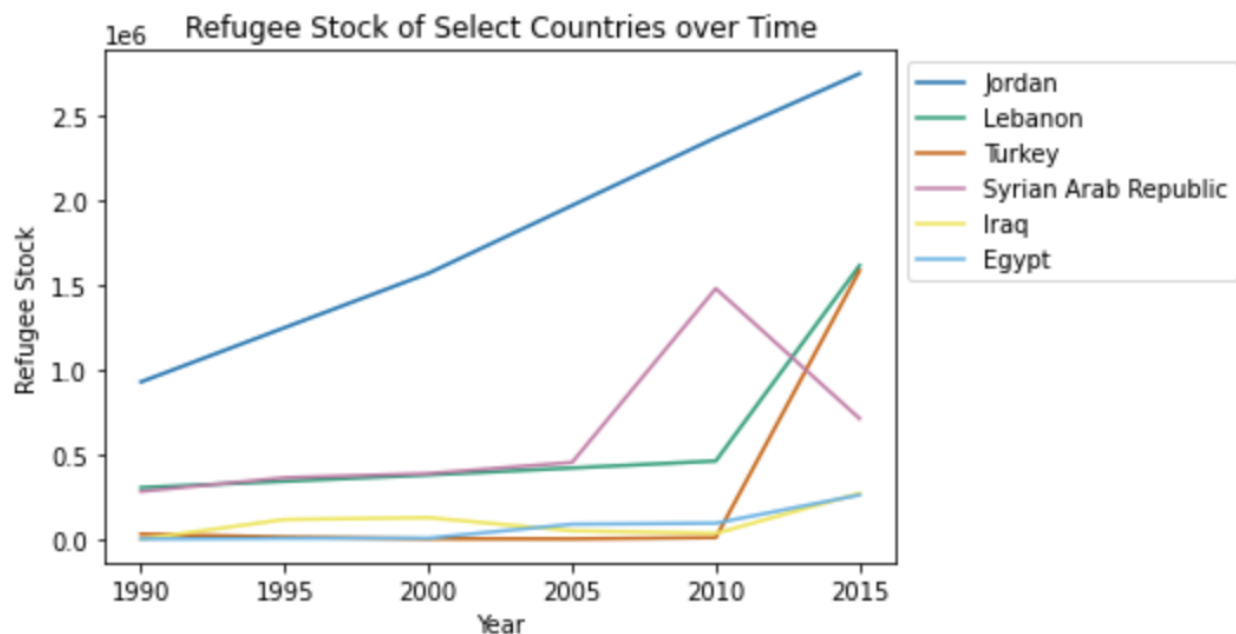


Figure 6: Refugee Stock across time of Countries surrounding Syria

This war has had serious consequences for the stability of the region. As always, in a war, the people are the true sufferers. They lose their homes and are forced to migrate. This is what the data is telling us here too. We can see the same stark story in Figure 6. There is a sharp increase in refugee stock in countries neighboring Syria. Turkey's refugee population increased a whopping 158 times from 2010 to 2015. Lebanon's refugees increased more than three-fold as well. Additionally, there is a definitive decrease in refugees in Syria. This might be categorized by the fact that refugees in Syria were probably forced to other countries when the civil war broke out.

After seeing these patterns in refugee stock, I switched my variable of interest to migrant stock and the various ways it was presented in my cleaned dataset. Firstly we look at the total population of Syria. In Figure 7 below, there is a sharp decline in total population for both male and female Syrians – approximately 2 million people moved out of the country. Yet another graph confirming the consequences of the war. A substantial portion of these people will have moved to the neighboring countries. If my inference is correct, we should see an increase in migratory stock as in the neighboring countries.

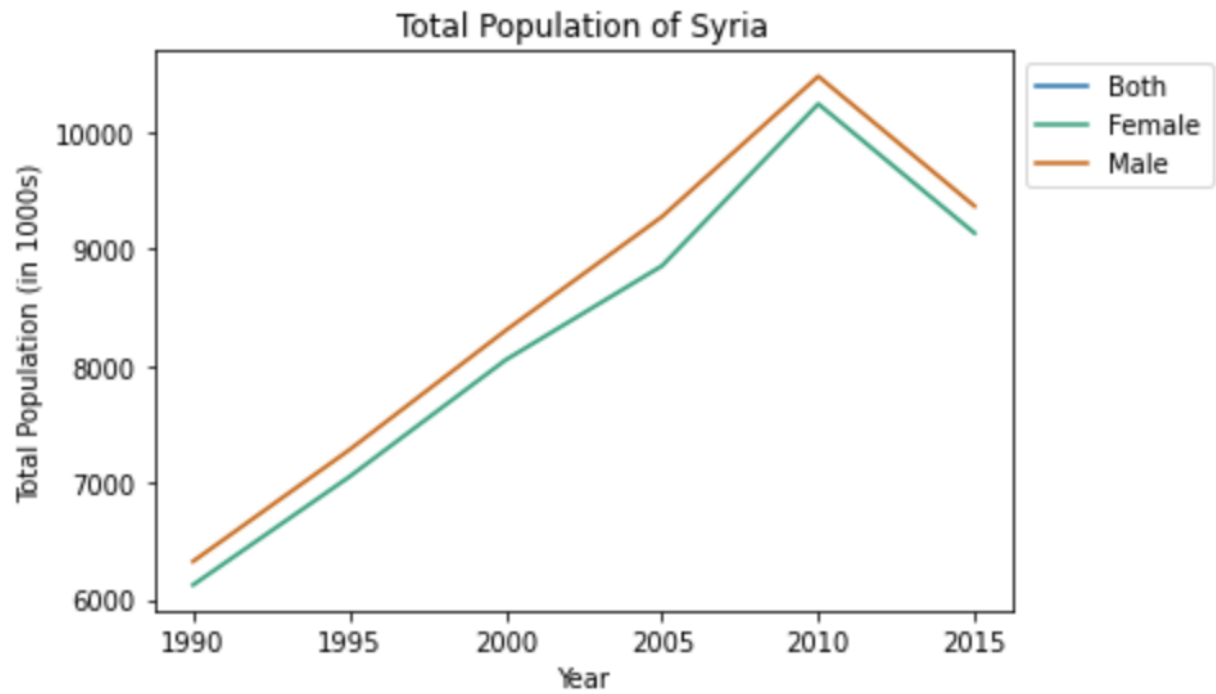


Figure 7: Total Population of Syria broken down by Gender

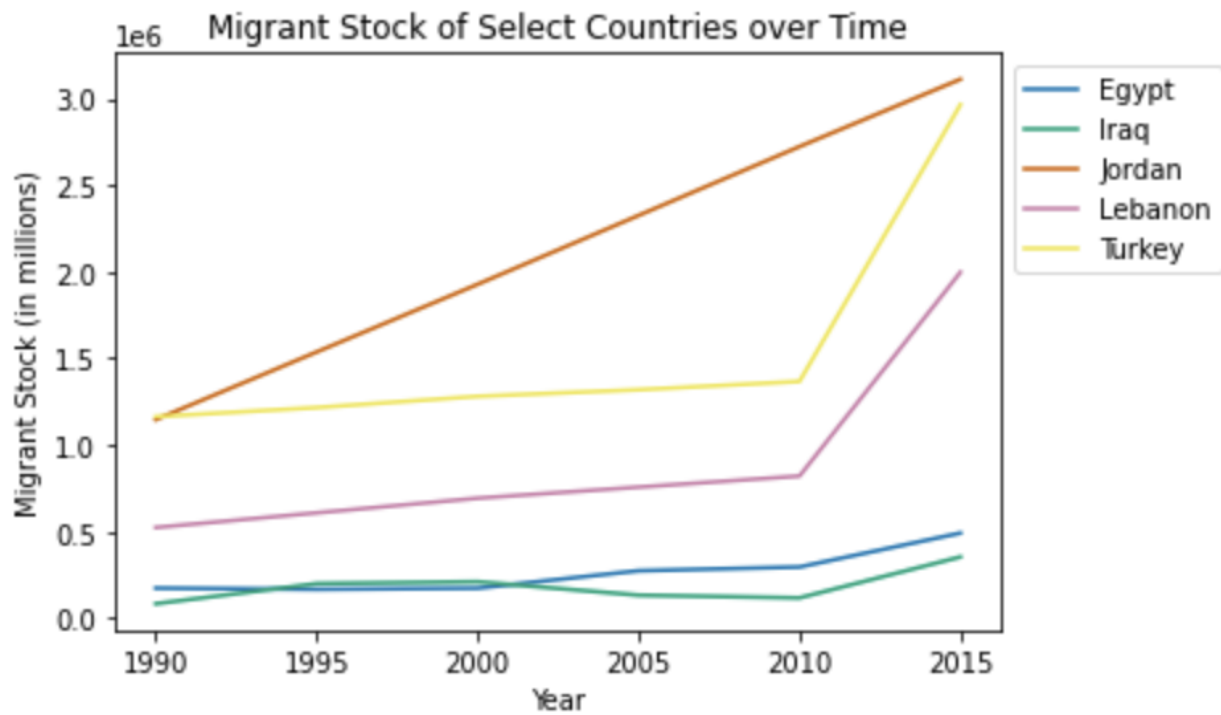


Figure 8 : Migrant Stock of Countries Neighboring Syria

The results we see in Figure 8 is almost identical to the refugee stock plot we see in Figure. This confirms my inference that the war led to a large displacement of Syrian people.

According to the UNHCR, Germany is the largest non-neighbouring host country for refugees from the war. Sweden and Netherlands were also among these non-neighbouring host countries (BBC News). That is why I looked at the migrant stock of these aforementioned countries to see if any increase is present.

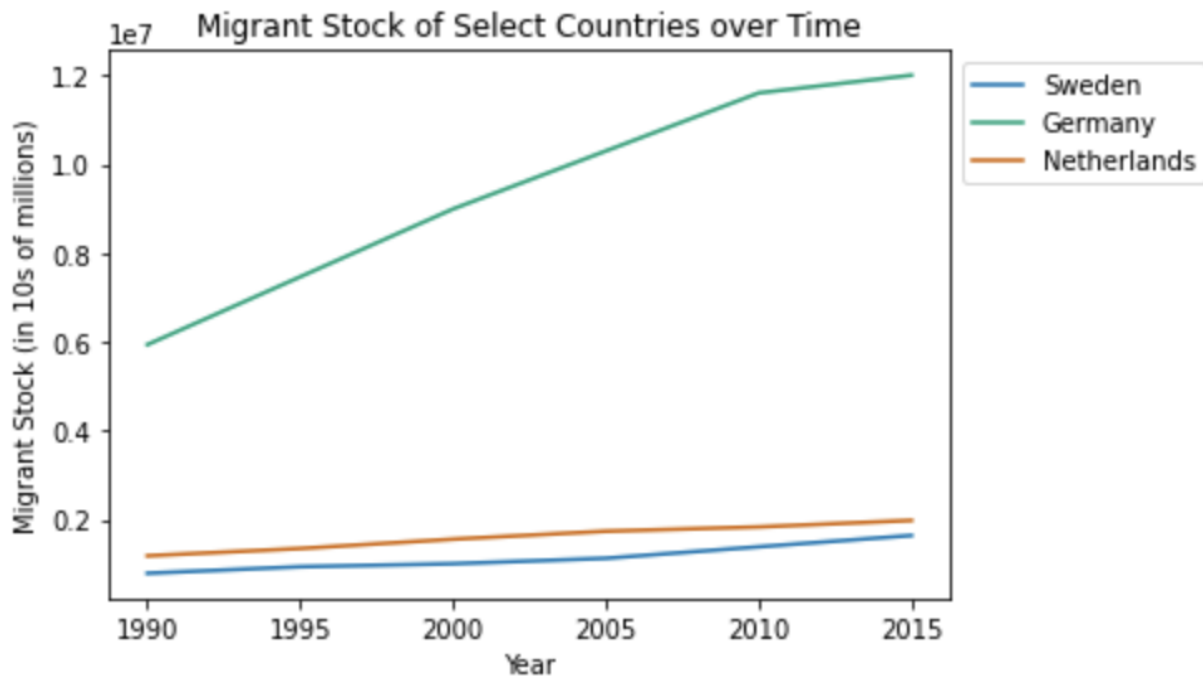


Figure 9: Migrant Stock of non-neighbouring Countries

There is a slight increase for Germany but no significant increase for the Swedes or the Dutch (See Figure 9). The size of the Syrian refugees is likely being drowned out by the already present migrant population in these countries.

4. Discussion

I investigated migratory and refugee patterns in the context of the Syrian Civil War in the Middle East. The results we saw were in-line with reports from the United Nations High Commissioner for Refugees. This violent civil war has led to a displacement of over 5.2 million Syrians since 2012 (UNHCR). For a vast majority of these refugees it's easier to go to host countries surrounding it. In time with so much turmoil, access to flights is greatly reduced. But some find opportunities of seeking asylum elsewhere

such as Germany. A large number of Syrians will have moved to Europe by getting flights after fleeing to more stable neighbors.

One question that comes to mind from this data is that with time when refugees are given citizenship of countries where they migrate to, does that lead to a decrease in refugee stock for that country or not? This ambiguity was not answered anywhere in the dataset. Additionally, there is no clear definition of how this dataset defines migrants and refugees. At multiple occasions this created confusion in my analysis.

With more time and resources, one could use this data to explore other conflicts in various regions. In Figure 2, you can see Africa has the second-highest refugee stock and as such a potential next step would be to explore conflicts in that part of the world.

Moreover, the graphs that I created as part of this paper didn't incorporate maps. Maps are visually powerful and can tell a much more visceral story. That is another interesting potential next step.

5. Conclusion

This project was an engaging experience and an effective way to practice storytelling with data. I have always been a fan of geopolitics. After a long time, I got a chance to practice what I read up about in my free time. Looking at the grim effects of the Syrian Civil War only reaffirmed how devastating wars can be. Hopefully in the future, I can do a deep dive of this again with another dataset.

6. Work Cited

BBC News. "Syria: The Story of the Conflict." *BBC News*, 11 Mar. 2016,

www.bbc.com/news/world-middle-east-26116868.

Syria Refugee Crisis Explained. www.unrefugees.org/news/syria-refugee-crisis-explained.

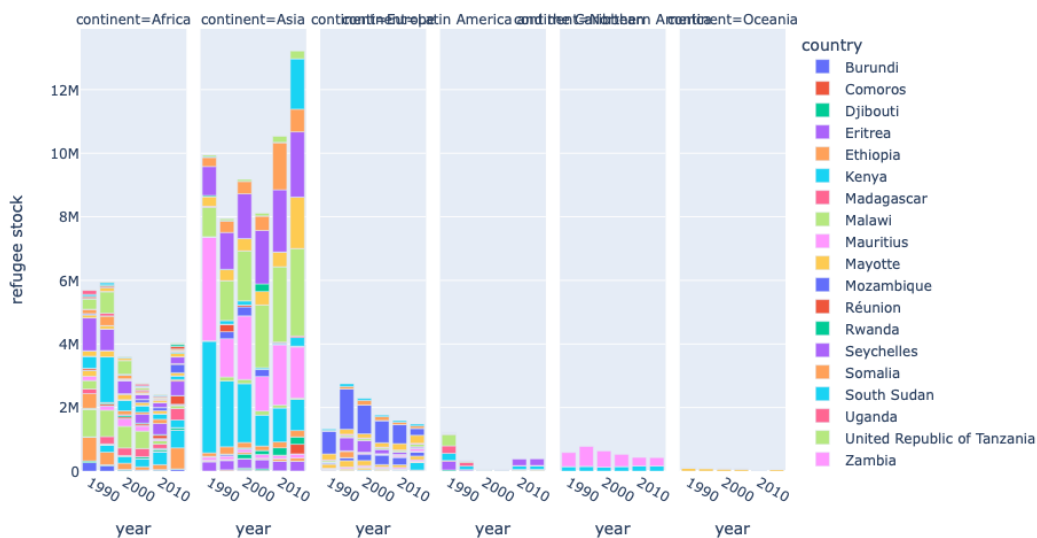
Situation Syria Regional Refugee Response. data.unhcr.org/en/situations/syria.

Tukey, John. *Exploratory Data Analysis (Classic Version) (Pearson Modern Classic)*. 1st ed.,
Pearson, 2019.

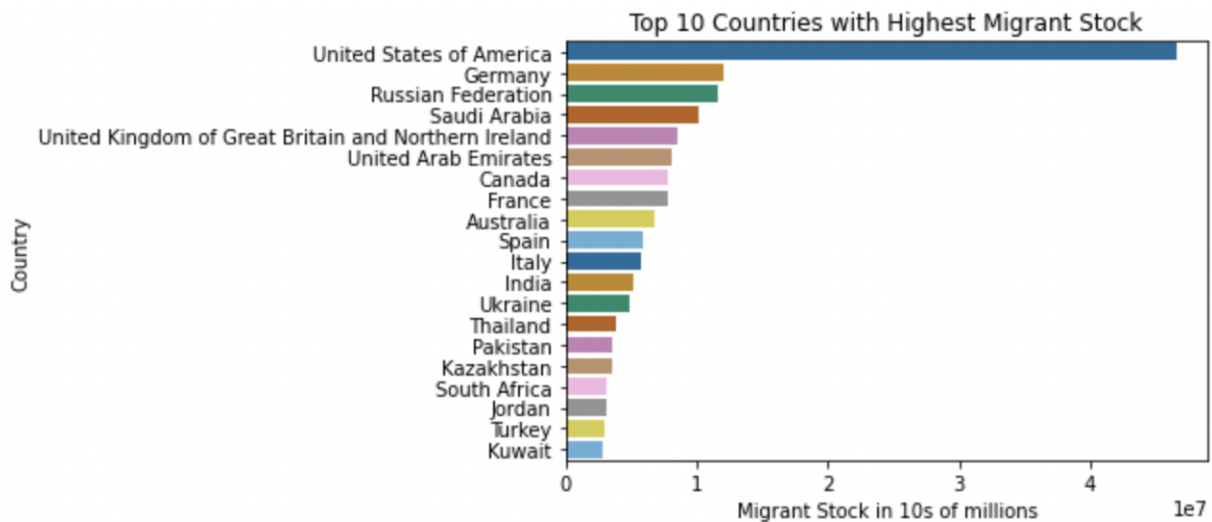
Tufte, Edward. *Envisioning Information*. Graphics Pr, 1990.

7. Appendix

Below is a plot that I really like since it portrays a lot of information in a compact, dense graph. It breaks the chartjunk principle and as such I did not add it.



The graph is a stacked bar plot that is facet wrapped by continent. Each color corresponds to a country. This bird's eye view can tell a lot about the stability of regions and which countries tend to receive the most refugees because of conflicts. You can interact with this plotly plot on my Google Colab.



This graph itself can tell a thought-provoking story. If you add on top of this a graph on the top 10 countries with the best universities, then more than half of them align with the top 10 countries in the figure above. That could be one source of migratory patterns in these developed countries. You can also see oil-rich Middle Eastern nations like Saudi Arabia, Kuwait, and UAE which have populations mostly consisting of expatriates.