Michaela Drouillard
December 15 2022
Final Project Submission
INF1340

Visualization and Analysis of Migration Patterns in Europe during the Bosnian War

Introduction

In this paper, I use John Tukey's principles of exploratory data analysis and Edward Tufte's principles of data visualization to guide my exploration of the 2015 UN Trends in International Migrant Stock dataset. By applying these principles, I was able to analyze migration patterns within Europe during the 1990s, using the Bosnian war as a case study.

Using small multiples, I identified patterns of interest across global rates of change in refugee and migrant stock, and noticed interesting trends in Europe during the 1990s. When I sorted my data, I found that Serbia had the highest rate of change in Migrant Stock from 1990-1995. I decided to focus on the Balkans, and proceeded to visualize Serbia's surrounding countries' migration data to look for any other patterns in migration. I then used secondary sources to provide context and understand the factors that influenced migration patterns in Europe during this time. Germany's high migrant stock correlated both with population trends in the Balkans and the secondary literature's analysis of German immigration policy during the 90s. The visualizations I use in this paper tell a story about how both policy and geographic proximity affected migration patterns in Europe in the 1990s.

Method

In building my data visualizations, I drew from the principles of data visualization outlined by Edward Tufte in his book Envisioning Information. Tufte's aesthetic principles for data visualization design focus on creating clear and concise visualizations that effectively tell a story. By applying these principles, I was able to better understand, clean, and communicate the data from the 2015 UN Trends in International Migrant Stock dataset.

I would roughly organize how I used these principles under three categories: 1) Understanding, 2) Cleaning, and 3) Communicating.

1) Understanding

To better understand the data, I used small multiples to compare and contrast the rates of change in refugee and migrant stock across different continents (Figure 1 and Figure 2). This allowed me to easily see patterns and identify significant differences between the rates of change in migrant stock across different continents.

I also used multivariate data visualization in Figure 3 to show trends in both migrant and refugee stock in the Balkans, which provided insight into the impact of the Bosnian war on migration patterns in the region.

I had initially visualized these variables across 6 different charts in a small multiples format. However, this somewhat distorted the data, since the Y axes in each chart were non-uniform. Visualizing the data in a single chart allowed me to more accurately and more rapidly interpret the trends and patterns in the data.

2) Cleaning

Many of Tufte's principles surrounding the aesthetics and design of data are encompassed by the idea of maximizing the data:ink ratio. In order to reduce clutter and present information clearly as possible, one should use as much "ink" as is needed to present the data, and no more (2018). This can mean using thin lines (which I did in Figures 1, 2, 5, and 7), but it also means removing "chart junk" and shrinking margins between small multiples.

I removed chart junk by removing the grids from Figure 1 and Figure 2. I also shrunk the margins between the facet columns to a 0.001 measurement, so that we can still register a division between continents, but without the whitespace in between as visual clutter.

3) Communicating

Data visualization design should facilitate accessible and clear communication. In order to make these visualizations as accessible as possible, I used the colorblind palette in seaborn plots, and the “ggplot2” palette in plotly, since the muted red tones are easier for colorblind folk to interpret. The muted red color used on the grey background aligns with Tufte's principles for using color in data visualization. In *Envisioning Information*, Tufte suggests avoiding the use of bright, clashing colors on a harsh white background whenever possible (2018). By using a muted red on a grey background, I aim to create a more cohesive and visually appealing visualization that is in line with Tufte's principles.

Using the appropriate charts is also part of communicating effectively (Tufte, 2018). I used line charts in figures 3, 4, 5, 6, and 7 to show how trends develop over time. I used box plots in Figures 1 and 2 to have a bird’s eye view on patterns and variability across different continents. Boxplots also provide us with summary statistics, which is a helpful way to orient ourselves in the dataset during the first phases of exploratory data analysis. In Figure 8 and Figure 9, I used bar charts to see whether both sexes were equally impacted by the Bosnian War in the year 1995. Since I was only looking at one year, and not a span of years, bar charts were a better choice than line plots.

It is important to recognize that each visualization presents only one aspect of the data, and that further analysis may be necessary to fully understand the underlying structures and trends. For example, the box plots in Figures 1 and 2 provide an overview of the variability, means, and ranges in data across different continents, but do not show density. (For instance, high variability and many outliers could also be due to how the data is aggregated across categories. I explore this further in Results). Subsetting and visualizing the data in different ways can provide additional insights that the boxplots obscure.

Results

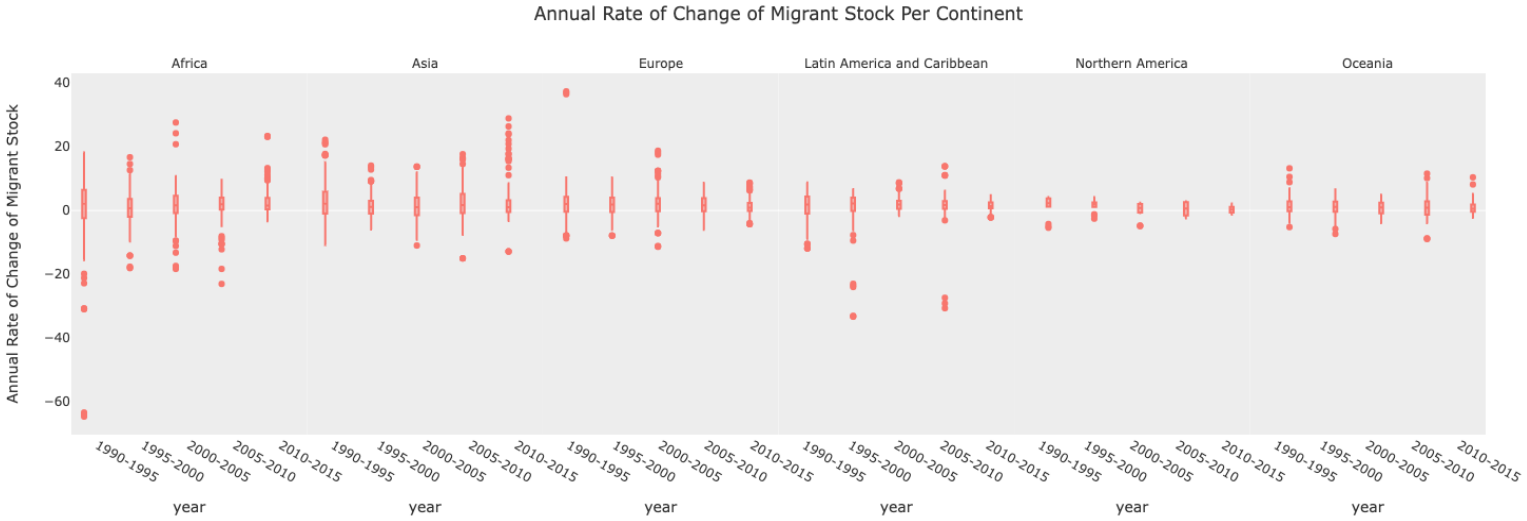

Figure 1. Annual Rate of Change of Migrant Stock Per Continent.

In Figure 1, we can observe that Asia and Africa have the highest variance in annual rates of change in migrant stock, with the most outliers. Northern America has the least amount of variance. This could be attributed to how the data is aggregated across the categories -- Asia and Africa both have more countries (which are the observational units of this table) than North America.

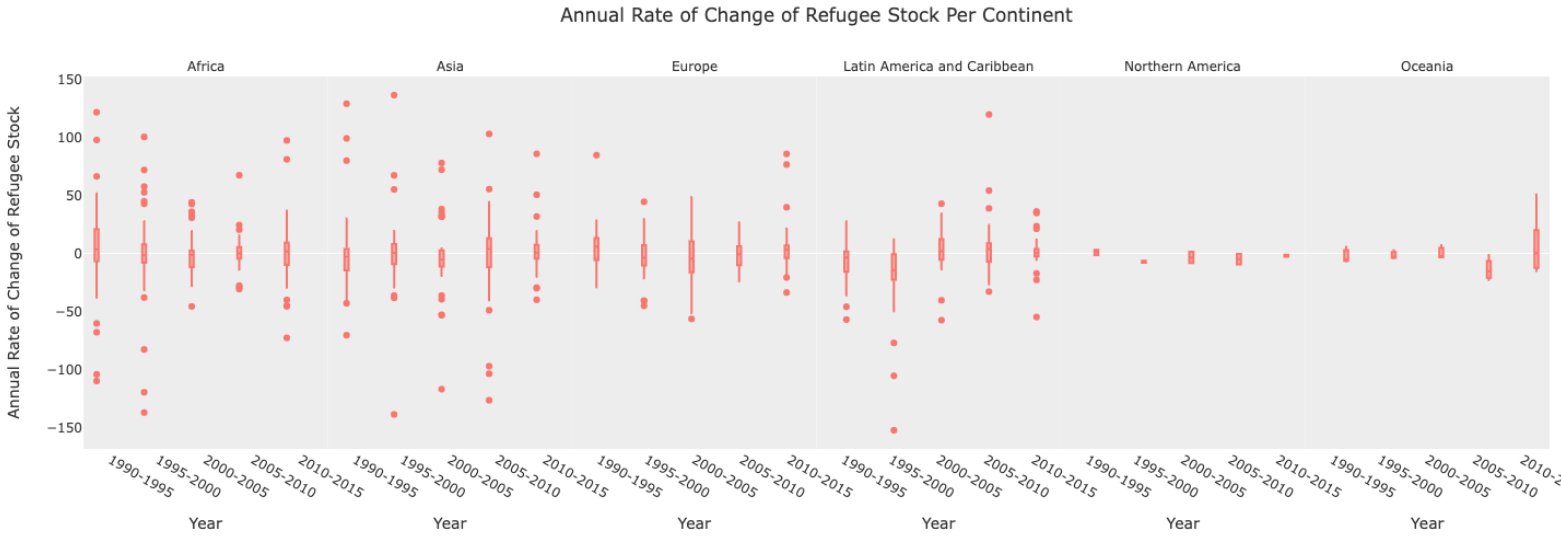

Figure 2. Annual Rate of Change of Refugee Stock Per Continent

Again, Africa and Asia have the most variance in the annual rate of change of refugee stock. Northern America and Oceania have the smallest variance, which could again be due to how the data is aggregated across these categories. It could, though, also suggest something about migratory patterns. Northern America is further, geographically, from these places, which makes it a more expensive destination to reach for people seeking asylum.

I was drawn to Europe's trends in this table -- it has far less variance than Asia and Africa, but similar means and range (especially from 1995-2000, which has a significant range).

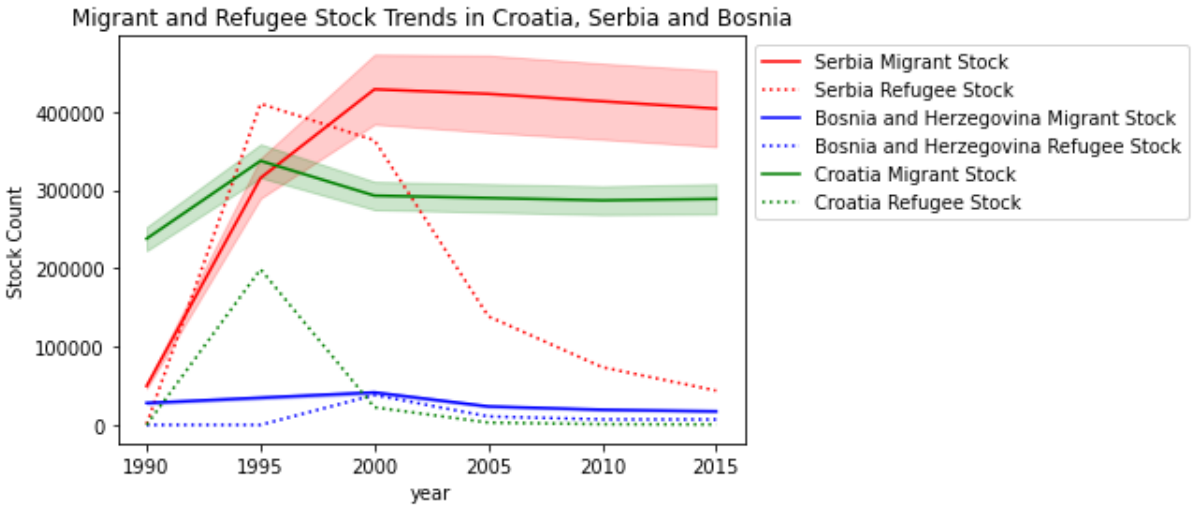

Figure 3. Migrant Stock and Refugee Stock in Croatia, Serbia, and Bosnia

Croatia and Serbia experienced a spike in migrant and refugee counts from 1990 until 1995, with all activity smoothing out around 2000. Bosnia did not experience an increase in migrant and refugee counts.

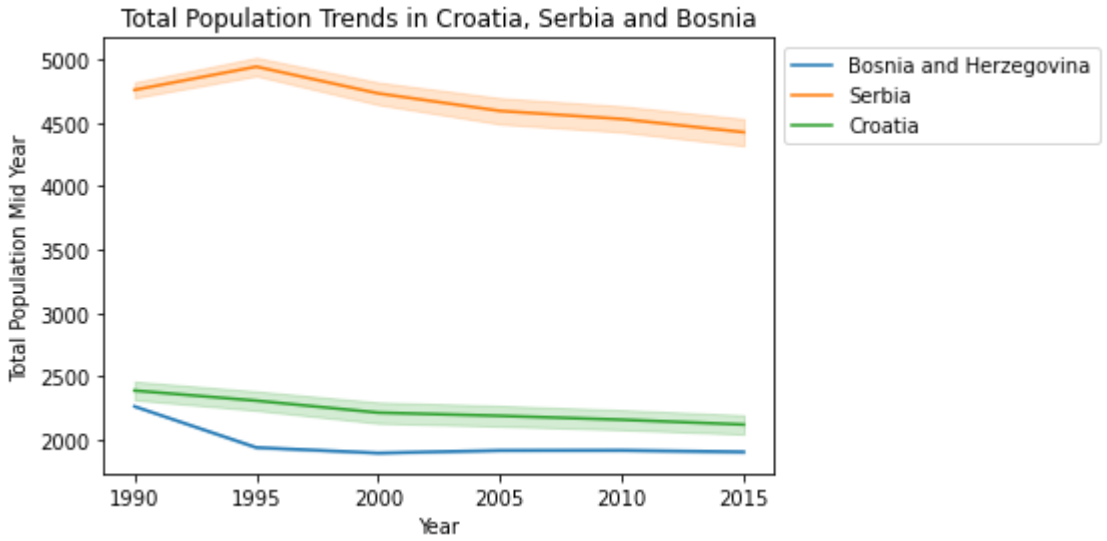

Figure 4. Total Population Mid Year in Croatia, Serbia, and Bosnia

Bosnia experienced a decline between 1990 and 1995 in total population, while Serbia experienced a small increase. Croatia has a gradually small decline over the entire timespan visualized, with no significant peaks.

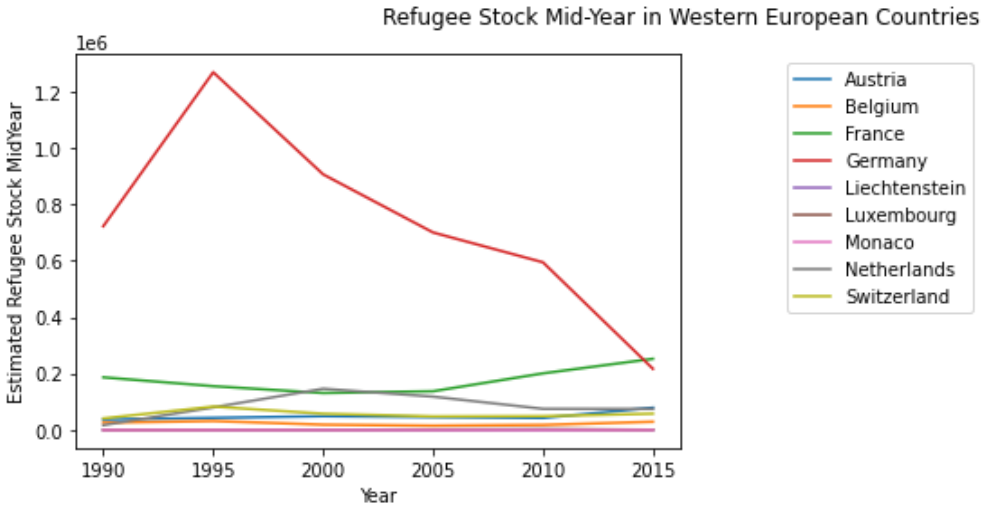

Figure 5. Estimated Refugee Stock Mid-Year in Western European Countries (1 = 1 million)

Germany has a significantly higher estimated stock of refugees than other Western European countries, especially in 1995. This peak coincides with the migration patterns observed in the Balkans in Figure 4.

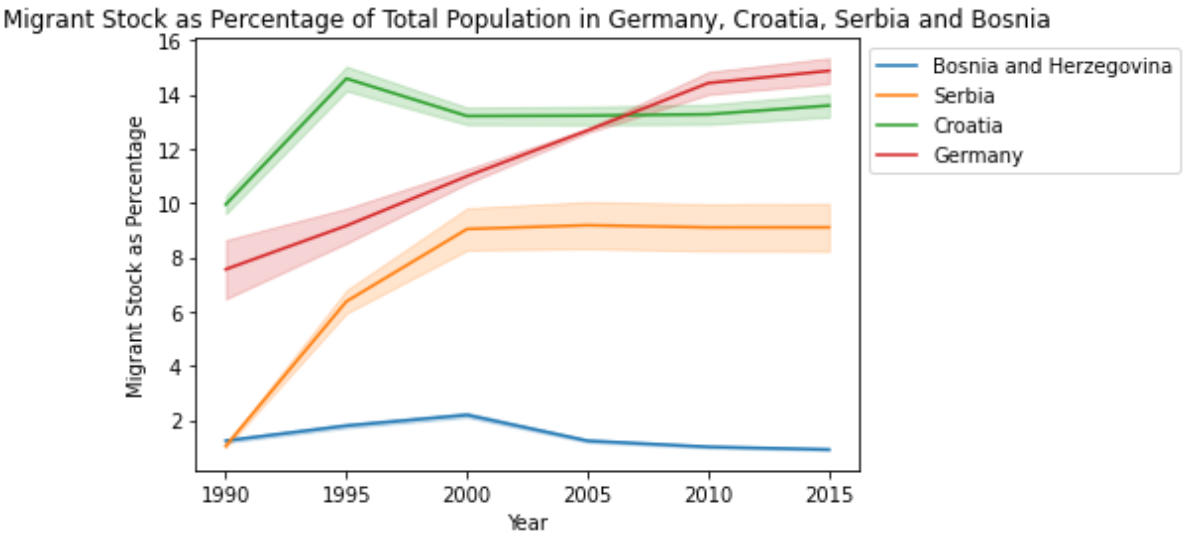

Figure 6. Migrant Stock as Percentage of Total Population in Germany, Croatia, Serbia and Bosnia

Serbia and Croatia both experienced growth from 1990-1995 in the migrant stock as percentage of total population. Germany experienced a more gradual growth over time, reaching its highest in 2010 and 2015.

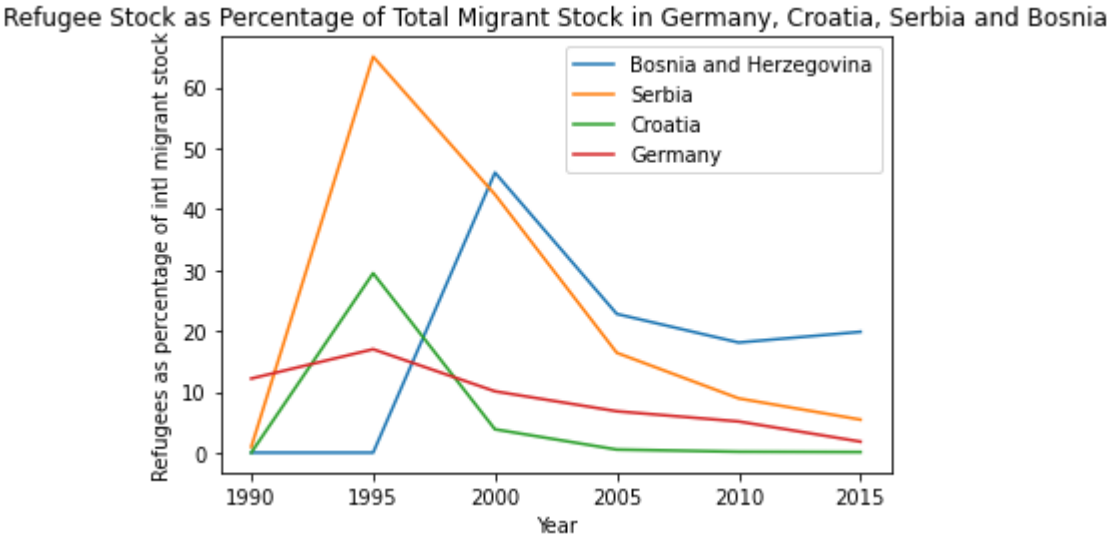

Figure 7. Refugee Stock as Percentage of Total Migrant Stock in Germany, Croatia, Serbia and Bosnia

Serbia, Croatia, and Germany all experienced spikes in refugee stock as percentage of total migrant stock between 1990-1995. Bosnia experienced a spike between 1995-2000.

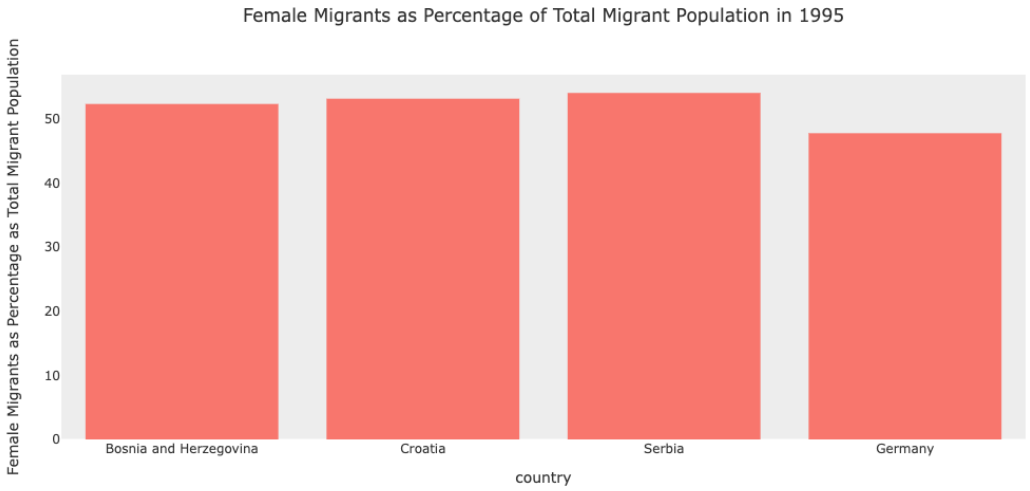

Figure 8. Female Migrants as Percentage of Total Migrant Population

Female migrants make up around half of the percentage of the total migrant population in Bosnia, Croatia, Serbia, and Germany in 1995.

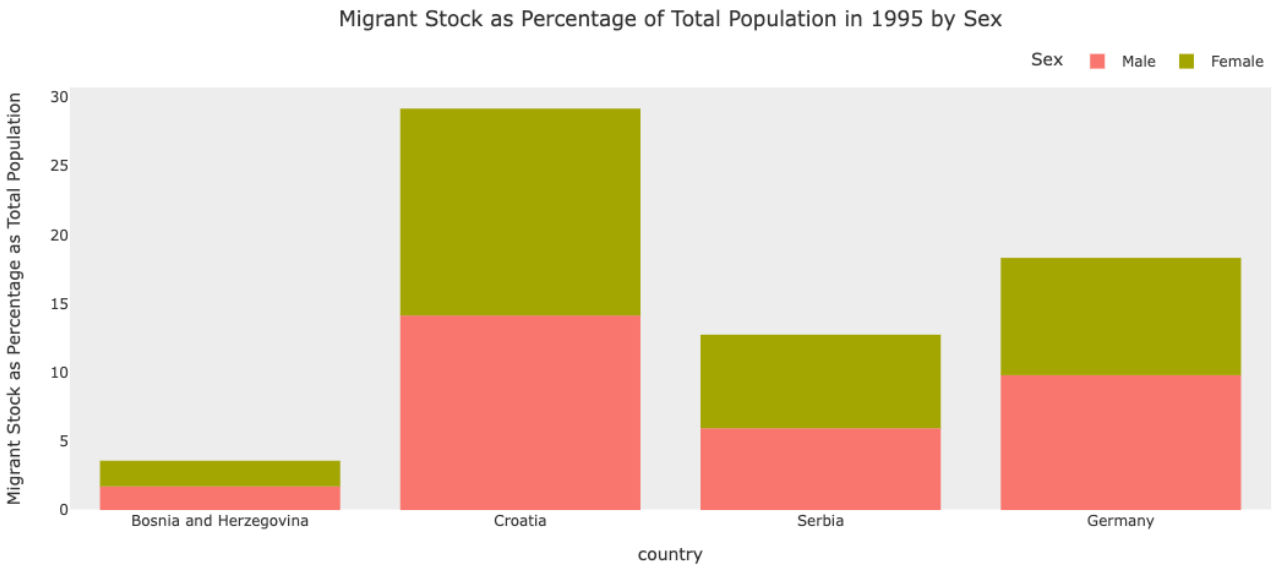

Figure 9. Migrant Stock as Percentage of Total Population in 1995 by Sex

In 1995, the distribution of sexes among the migrant stock from Bosnia, Croatia, Serbia, and Germany was approximately equal.

Discussion

Europe displays interesting patterns in the annual rate of change in refugee stock and migrant stock. While it has less variance than Asia and Africa in both tables, it has similar ranges and means. This means that, overall, the degree to which rates are changing appear similar overall to Asia and Africa. This is interesting, because Europe has a smaller number of countries (and thus, observational units) than Asia and Africa, which suggests that these means reflect some sort of real, concentrated change, and not just a coincidence of data aggregation.

Europe's ranges and means in Figure 1 and Figure 2 are also higher than those in North America and Oceania, which would be regions to compare Europe to if one were to look for trends across "developed" areas. While many factors, most of which are beyond the scope of this study (i.e. immigration policy), contribute to these kinds of trends, one significant factor could be the geographic one: Oceania and North America are located further from most areas that experienced civil unrest during this time period (i.e. MENA during the Arab Spring, the Bosnian War, the Rwandan genocide). This would make these two continents expensive and difficult places to get to. It makes more sense that the regions immediately surrounding zones of conflict have higher rates of refugees.

Since Europe has a mix of both developed and developing countries, and it is located in closer proximity to zones of conflict, this makes it an interesting site for studying migration trends. The wider variance in both the annual rate of change of refugee stock and migrant stock from 1995-2000 in Europe reflects a time of rapid change and fluctuation within the continent. One could infer that, among other factors, this pattern in the data might reflect migratory patterns that were brought about by the after effects of the fall of the Soviet Bloc and the Bosnian War.

In order to understand my data from another angle, I sorted the rate of change in migrant stock in descending order, and found that Serbia ranked the highest globally in rate of change in Migrant Stock from 1990-1995. In order to see if my hypothesis that geographic proximity influences the counts of migrant and refugee stock, I visualized migrant and refugee stock over time in Croatia, Serbia, and Bosnia (Figure 3). Serbia and Croatia both show an increase in migrant and refugee stock from 1990-1995. Not only does Bosnia not experience a spike in refugee and migrant stock, but Figure 4 shows that Bosnia experienced a total population decline. From this, we could infer that Bosnians might have been fleeing

through and to neighbouring countries, which could be the reason for Serbia's ranking the highest globally between 1990-1995 for rate of change in migrant stock.

A 2017 report titled "Integration of Refugees: Lessons from Bosnians in Five EU Countries" by Mikkel Barslund et. al. provides supporting evidence for this hypothesis. According to the report, "approximately half of displaced Bosnians fled to Serbia, Montenegro, and Croatia" (2017). The next most popular destination for registered Bosnian refugees was Germany, with a total of 320,000 Bosnian refugees relocating there. Germany offered social assistance similar to natives, but granted limited access to the labour market and education, with a residency status that enforced repatriation once the Bosnian war ended.

The data does confirm something interesting happening in Germany, which could be related to the pro-refugee temporary immigration policy articulated in Barslund et al (2017). Compared to other Western European countries in Figure 5, Germany does have the highest amount of refugee stock, with a significant spike from 1990-1995 (which could be related to the Bosnian War, in addition to incoming refugees from other conflicts). The spike then dwindles over time, which would reflect the "repatriation" clause in its temporary residency status.

While, again, many other factors can contribute to Germany's steady growth in migrants, Figure 6 shows how civil unrest directly impacts the demographic distribution in small, neighbouring countries, while having a more subtle impact on larger countries with open immigration policies.

Being countries with smaller populations, Serbia and Croatia have more pronounced increases in migrant stock as percentage of total population than Germany (there are fewer people already living there, so an influx of migrants has more of an effect on the migrant:native ratio) (Figure 6). Bosnia doesn't increase nearly as much, since we might assume that people, overall, are leaving Bosnia. Germany steadily increases over time, with another peak that might coincide with other global migration events between 2010 and 2015 (such as the Arab spring).

In other words, we can infer that areas and populations beyond the Balkans are also undergoing change and forced relocation, which is accounted for in Germany's steady growth as it absorbs new kinds of migrant populations over time (which could be attributed to lenient immigration policies).

Figure 7 explores this dynamic further: Serbia and Croatia have the most pronounced increase in refugees between 1990 and 1995. From this, we can infer that while Germany's rapid growth in migrant stock visualized in Figure 6 may be mediated by other kinds of migrants (i.e. people immigration for work), the increase in migrants in the Balkans around that time is more directly linked to the forced displacement of Bosnian Refugees.

Figure 8 and 9 explore whether there was a difference in how each sex was impacted by the displacement that peaked in 1995. In both charts, we see that migrant populations are just about equally split between men and women.

Conclusion

In conclusion, using Tukey's principles of exploratory data analysis and Edward Tufte's principles of data visualization facilitated an exploration of the 2015 UN Trends in International Migrant Stock dataset. This led to a case study analysis of migration patterns within Europe during the 1990s, specifically examining the impact of the Bosnian war on refugee migration. By sorting and grouping the data, and incorporating secondary sources, a clear picture emerged of the factors influencing migration patterns in Europe during this time period. Both policy and geographic proximity seemed to have an influence on migration data, although further statistical testing and historical analysis would be required to make any causal inferences.

References

Barslund, M., Busse, M., Lenaerts, K., Ludolph, L., & Renman, V. (2017). Integration of Refugees: Lessons from Bosnians in Five EU Countries. *Intereconomics*, *2017*(5), 257–263.

Tukey, J. W. (1977). Exploratory Data Analysis. Addison-Wesley.

Tufte, E. R. (2018) Envisioning Information. Cheshire, Ct., Graphics Press,.