Gabrielle Woodside
1009676780

Final Project Write-Up

Introduction

Through this project, I gained experience with Exploratory Data Analysis (EDA) and Data Visualization. The EDA principles of John Tukey that we learned in class guided my approach. They include:

- **Sorting** – seeing the dataset in order of a specific variable (either ascending or descending) to easily find the maximum or minimum values
- **Grouping** – comparing larger scope variables by combining all instances of them and either picking out the standout datapoint from each group, or comparing the sums of their data
- **Sub setting** – focusing only on variables of highest interest, to keep the data exploration process simple
- **Comparing** – looking at two variables (which may or may not be related) side-by-side to understand any differences or relationships between them

Principles from Edward Tufte also inspired my data visualization process. In particular, I aimed to use "Small Multiples" to compare variables across different categories. Considering the data in this dataset has multiple aspects and facets to it, the small multiples was essential in making sure the various variables were considered in the visualizations.

Methods and Results

Step 1 – Exploring the WORLD data

I took baby steps while exploring this data. To begin, I narrowed in on one of my sixteen tables – "df1_world". This had the variables from the original Tables 1, 2, 3, 4, and 6 that were measured at specific year marks (1990, 1995, etc.). I chose to focus on this table to start so that one: I could use it to get my feet wet. However, I also wanted to start with the broadest view possible of these variables. And what is broader than the world?

To begin, I turned the dataframe into a *datascience* Table, so that I could use the *sort()* functions from that package. From there, I sorted the table by the first variable ("International Migrant Stock at Mid-Year") (or IMS for short), in descending order so that I could see the highest values first. Then I grouped the data by Sex ("Both sexes", "Male", or "Female"), so that I could specifically see in which year the international migrant stock was highest for each level of sex. These are examples of Tukey's EDA principles, as I sorted and grouped the data in order to explore it further.

Level	Sex	Year	International migrant stock at mid-year	Total population at mid-year (in thousands)	International migrant stock as percentage of total population
WORLD	Both sexes	2015	243700236	7.34947e+06	3.31589
WORLD	Female	2015	117584801	3.64227e+06	3.22834
WORLD	Male	2015	126115435	3.70721e+06	3.4019

Figure 1 - Table of the Years in which International Migrant Stock was the highest (per Sex). Other unneeded columns were not included in this screenshot.

I ended up with a small table with just three rows (*Figure 1*). It revealed that for all levels of sex, the international migrant stock is the highest in the most recent year (2015). However, that does not reveal how the IMS increased. Was it steady or not? In order to answer that question, I needed to graph the variable over time. I used a vertical bar plot, with different coloured bars for each sex, which had a few

useful purposes. First, it allowed me to easily see how the variable changed over time. Secondly, it allowed me to easily compare the Male and Female numbers.

Figure 2 demonstrates that the IMS has indeed been increasing somewhat steadily over the years. Plus, the numbers for the males and females are fairly close throughout the years, which I was somewhat surprised to see, even though I don't know that much about international migration.

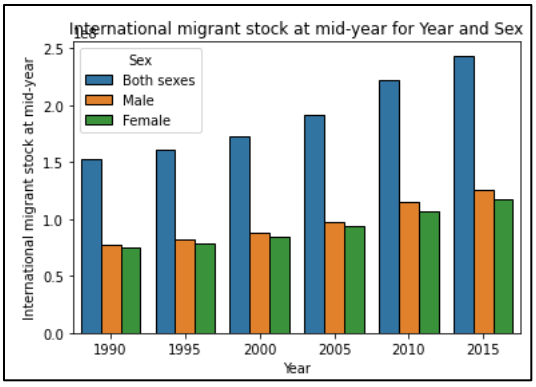

Figure 2 - International Migrant Stock at Mid-year, Per Year and Sex - This reveals the steady increase of IMS, and the relative equality between Males and Females.

Next, I did the exact same steps for the "Total Population" variable. I was curious to know how the trends in this variable compared to those in the IMS. As *Figure 3* and *Figure 4* reveal, the trends were very similar. The Total Population for all sex levels was highest in 2015, and the population has grown steadily over time. Males and Females are also fairly equal in their population.

Level	Sex	Year	International migrant stock at mid-year	Total population at mid-year (in thousands)
WORLD	Both sexes	2015	243700236	7.34947e+06
WORLD	Female	2015	117584801	3.64227e+06
WORLD	Male	2015	126115435	3.70721e+06

Figure 4 - Table of the Years in which Total Population was the highest (per Sex). Other unneeded columns were not included in this screenshot.

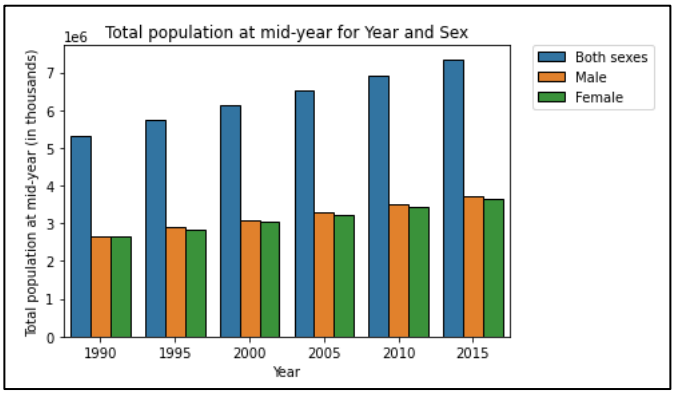

Figure 5 - Total Population at mid-year, by Year and Sex. It reveals the steady increase of population and relative equality between males and females

Now, I knew that there was already a variable demonstrating the percentage of the population that comes from the migrant stock. However, I wanted to further explore Tukey's comparison principle, and visualize the relationship between the two variables. In order to do this, I subsetted the data to only include rows from "Both Sexes". I thought that would make the comparison less confusing, especially since the individual sexes follow very similar patterns as the total. I also narrowed the table down to only include the year, IMS and Total Population. Further, I rescaled the "Total Population" variable, because it was original measured in thousands, which made it look much smaller than IMS when they were originally put on the same scale.

I used a horizontal barchart in this situation so that I could have the bars from the two different variables right next to each other. As can be seen in *Figure 6*, the IMS is very, very small in comparison to the total population. This drove home for me that a better way to focus on their relationship would be to examine the percentage variable.

Gabrielle Woodside
1009676780

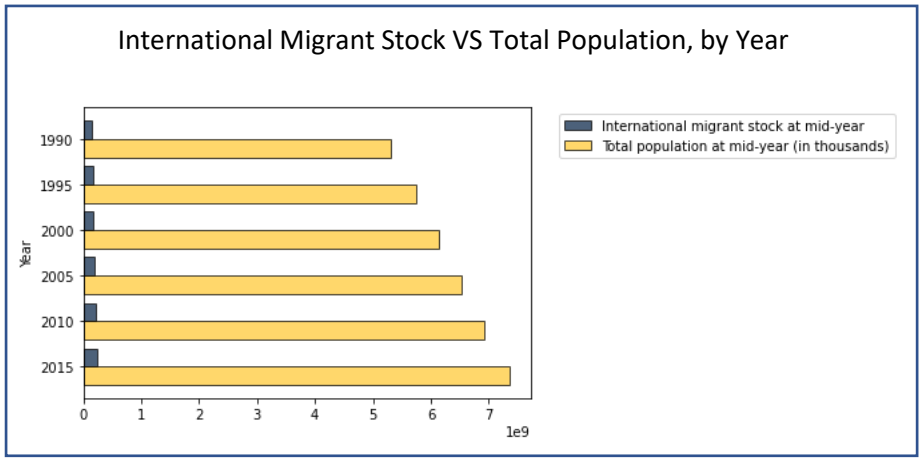

Figure 6 - International Migrant Stock VS Total Population, by Year

Step 2 – Exploring the "Extra" Areas using Small Multiples

Now that I had started to learn a little bit about the world, I wanted to add another variable into the mix. I had originally split all of the unusual data groups into separate dataframes (ex. "Developed Regions", "Sub-Saharan Africa", etc.). However, I grouped them back together so that I could examine how they compared to each other.

I sorted and grouped them by each of the variables again (IMS, Total population, Percentage of total population that is the migrant stock, Estimated refugee stock, and percentage of migrants that are refugees). My findings are below.

Figure 7 reveals how most of the other areas followed the same pattern as the World, in that their migrant stock rose steadily. I had a zoomed in look at the Least Developed Countries and Sub-Saharan Africa, because they were almost too small too see. It revealed that the Least Developed Countries followed a slightly different "U" shaped pattern, where the migrant stock dipped a little in the early 2000's before rising again in 2015.

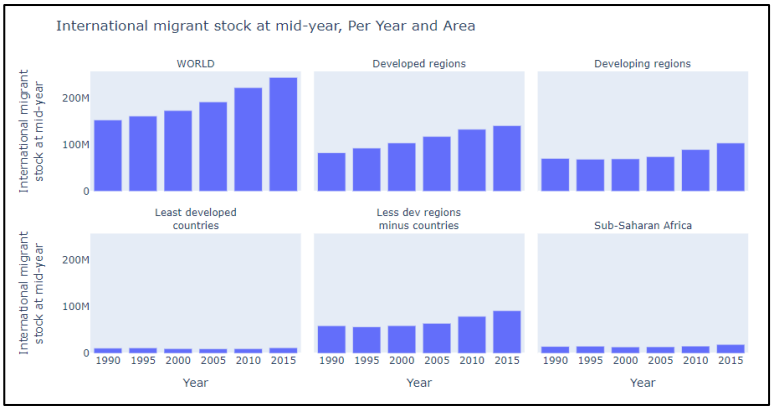

Figure 7 - International Migrant Stock at mid-year, per Year and Area

Figure 8 below demonstrates how across the world, in all other regions too, the total population has been increasing fairly steadily between 1990 and 2015. As far as I am aware, this trend has been continuing even past 2015.

Gabrielle Woodside
1009676780

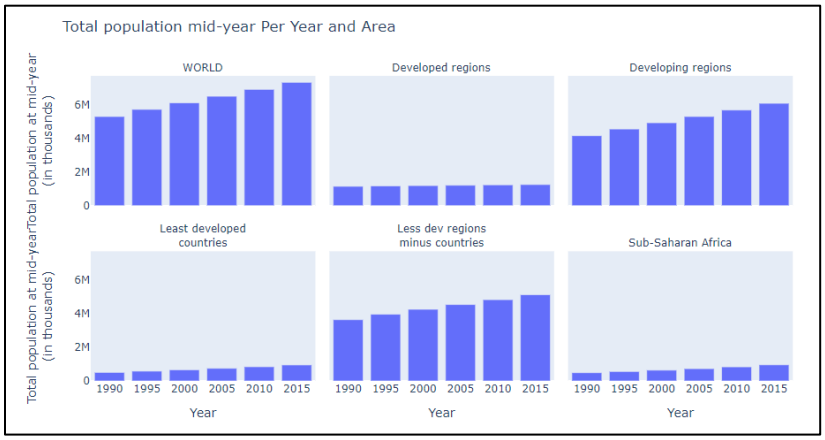

Figure 8 - Total Population mid-year, per Year and Area

Things get a little more interesting in *Figure 9*, which demonstrates the international migrant stock as a *percentage* of the total population. First off, I noticed that the percentage is significantly higher in areas that are more developed. Secondly, it is one of the only areas where the percentage is increasing. In most other areas, the percentage is decreasing.

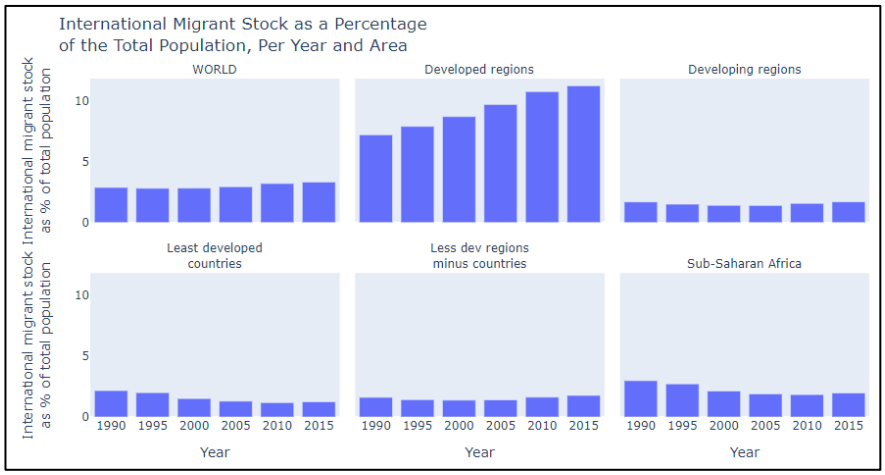

Figure 9 - International Migrant Stock as a Percentage of the Total Population, by Year and Area

There were also some interesting trends in *Figure 10*, which represents the estimated refugee stock per area, over the years. Considering the sample size in each area is quite different, we cannot compare the height of their bars. However, we can compare the "shapes" of their changes. It looks like in almost all areas, it is a "U" shape, where there was a decrease in refugee stock in the early 2005-2010's. But it's gone back up again since then, with some of the highest estimates being most recently.

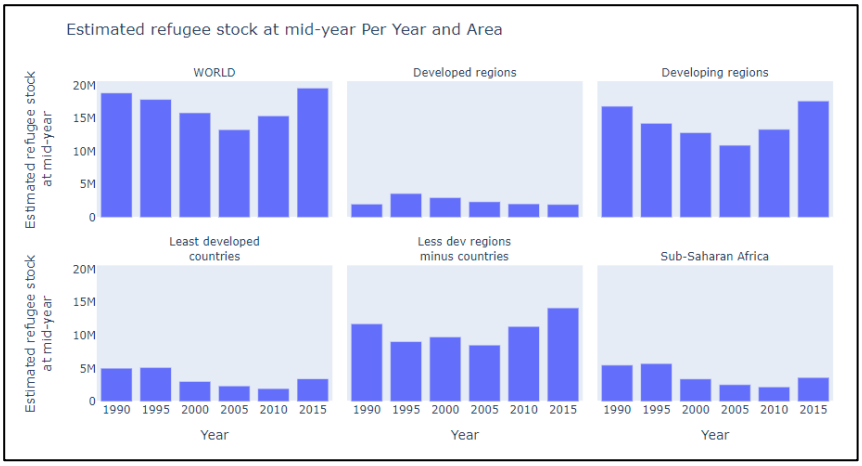

Figure 10 - estimated Refugee stock at mid-year, per Year and Area

Gabrielle Woodside
1009676780

Figure 11 below shows the last of the variables measured per year: refugee stock as a percentage of migrant stock. This is another interesting finding, where it has gone down in most places, with a slight increase in 2015 again.

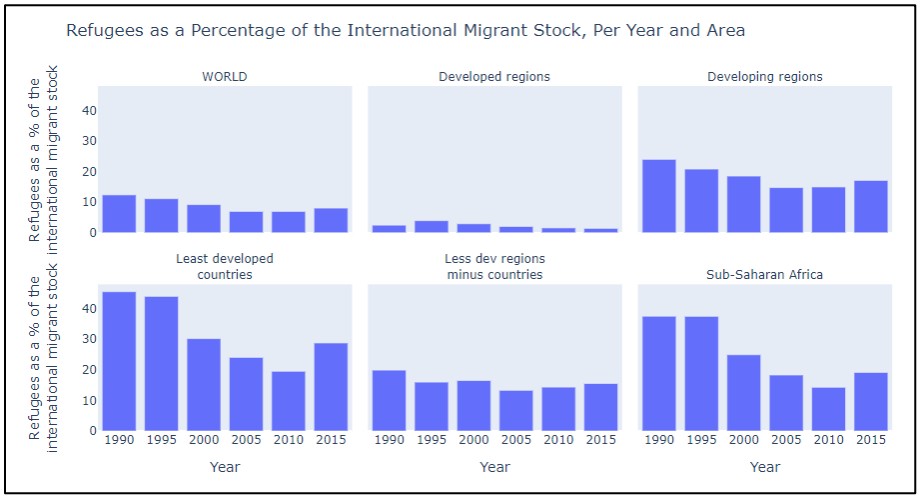

Figure 11 - Refugees as a Percentage of the International Migrant Stock, per Year and Area

Discussion

In my assignment, I examined many aspects of the migration and refugee stocks over the years. I used boxplots and Small Multiples to compare the results over the different areas and years.

One finding that stood out to me was from Figure 9, which showed how the percentage of migrants in the population has increased significantly in the most developed regions. Thinking about this though, that makes sense, as I always hear people talking about starting a “new life” in the developed areas of the world.

It was also very interesting to see how the refugee stock has been going down, but also shooting back up a bit in 2015. That is concerning, if people are again being forced out of their countries and homes, that is not a good sign. It provides a new perspective on the growing migration to developed regions. Is that migration forced, or of their own choice?

Looking at the decreasing percentage of migrants who are refugees, it does appear that more less people are migrating due to being forced out of their homes. So again, this perspective adds another caveat. With immigration increasing and refugees decreasing (somewhat, and maybe not as much anymore), it suggests that there was a time where people were mostly able to *choose* to come to another country, rather than being forced out of their homes.