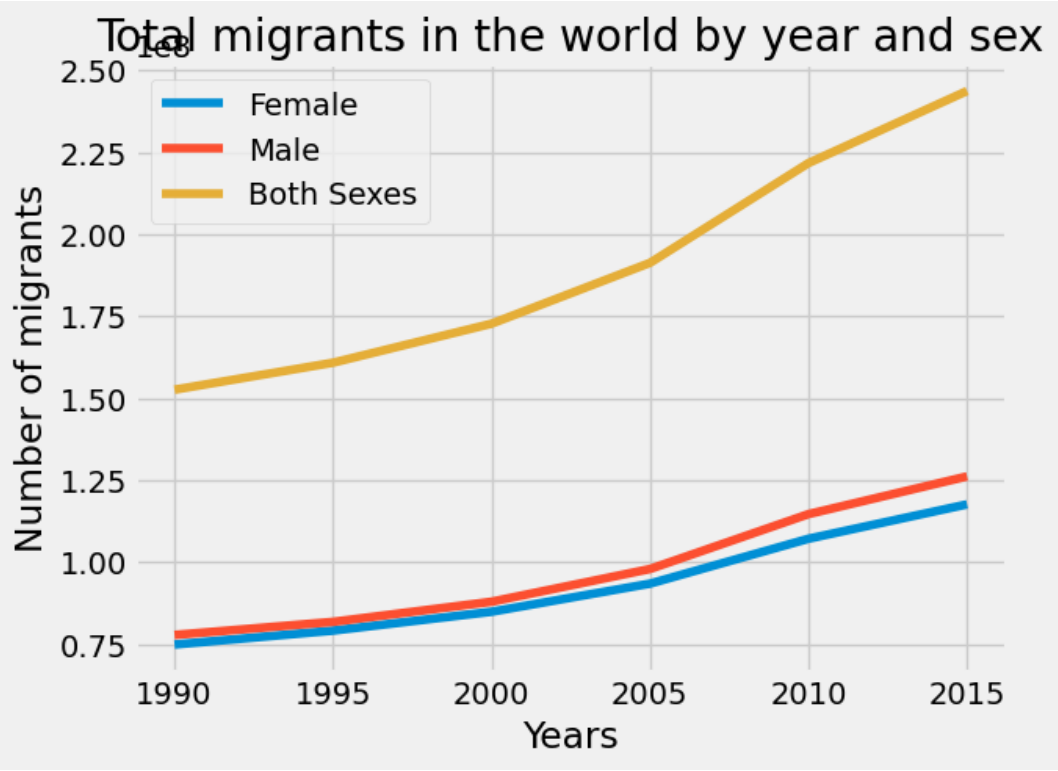
The purpose of this assignment was to create visualisations based on cleaned data from the document "Trend in International Migrant Stock: The 2015 Revision" published by the United Nations. Having previously cleaned this data set in a previous assignment using tidy data principles, I could now move onto visualizing the data using principles of exploratory data analysis such as those outlined by John Tukey and Edward Tufte. The former created steps to help sort, group, subset, and compare data whereas the second used concepts like chart junk and small multiples to create visualization that are minimal and accessible to viewers. Using these ideas, I proceeded to visualize each of the tables from the UN dataset in specific ways.

Methods and Results

Graph 1:

For each of the tables, graph 1 has the same purpose: to chart migrants in the world over a series of time and by sex. To accomplish this, I simply locked onto the World row and separated the data according to sex. I then plotted a line graph based on the migrants by year. This gave me a graph similar to this for each table:

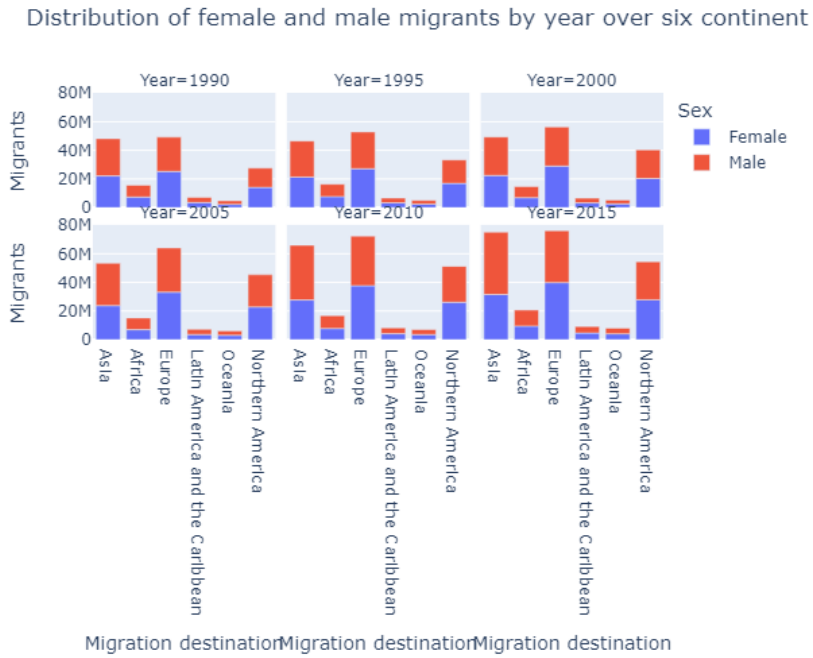


Graph 1: Table 1 – Graph 1, Total migrants in the world by year and sex.

To create this graph, I took into account the concepts of sorting, comparison, and accessibility. First, I sorted the dataset to find a specific row ("World") and organize this dataset by gender and year. Then I compared all the data for migrants who were male, female, or of both sexes. Lastly, I ensured that the colours chosen were accessible to all viewers by using the plyplot function matplotlib which uses primary colours in their graphs.

Graph 2 and Graph 3:

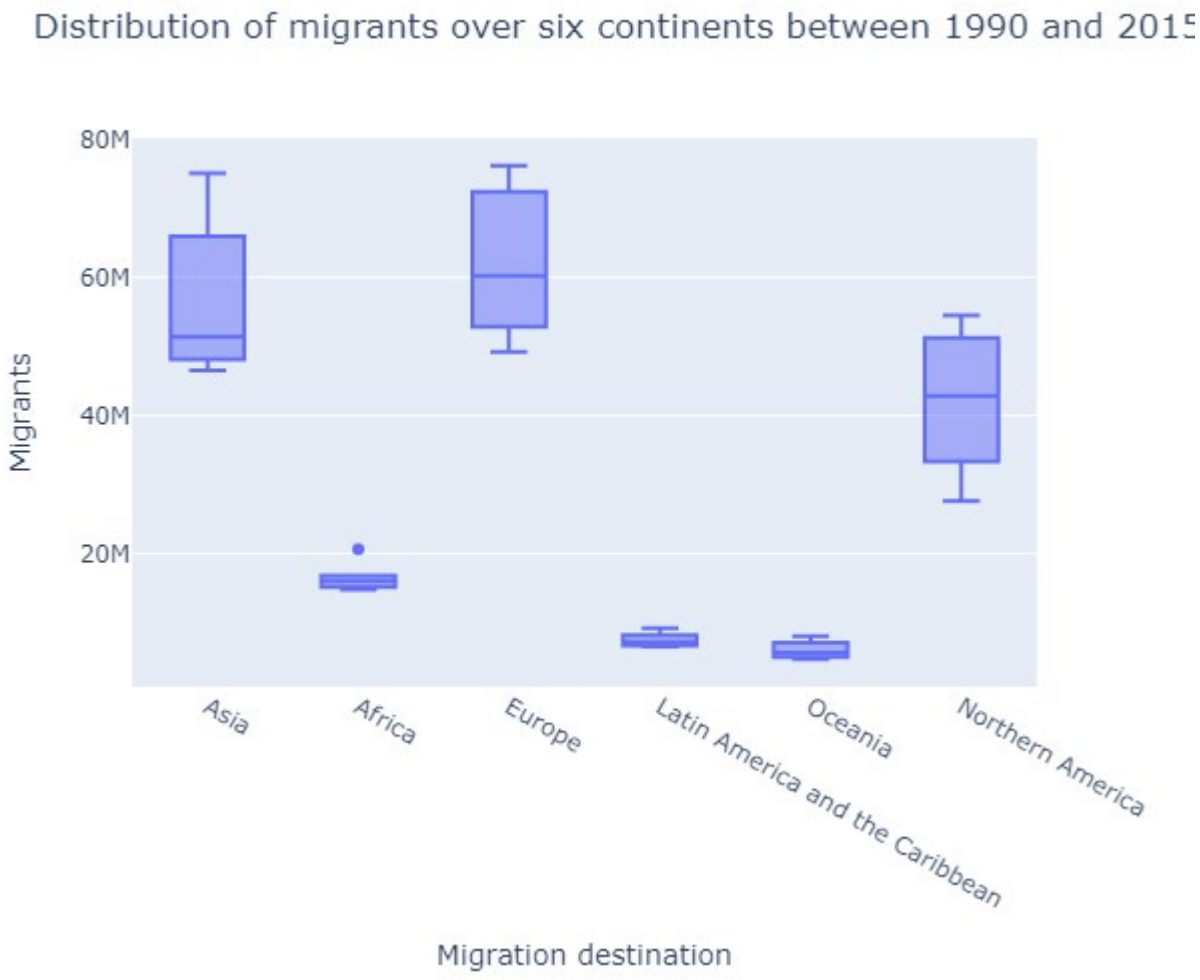
For graph 2[1], I then wanted to create a more specific graph to show the differences between male and female refugees for different regions of the world based on years. To accomplish this, I created a new dataset containing only male and female entries from the originally cleaned table. Based on this new dataset, I once again created a new dataset which isolated the six inhabited continents: Asia, Africa, Europe, Oceania, North America, and South America. With this new dataset of female and male migrants over six continents, I created multiple stacked bar graphs. These graphs show the evolution of the number of female and male migrants over the six continents by year. They look as follows:



Graph 2: Table 1 – Graph 2, Distribution of female and male migrants by year over six continents.

[1] Note that Graph 2 is not present for Tables 6.1, 6.2, and 6.3 as they do not track female and/or male migrants.

Graph 3[2] also shows data over six continents but this time I used a boxplot to present averages, median, maximums, minimums, and other such statistical data for both sexes. These look as follows:



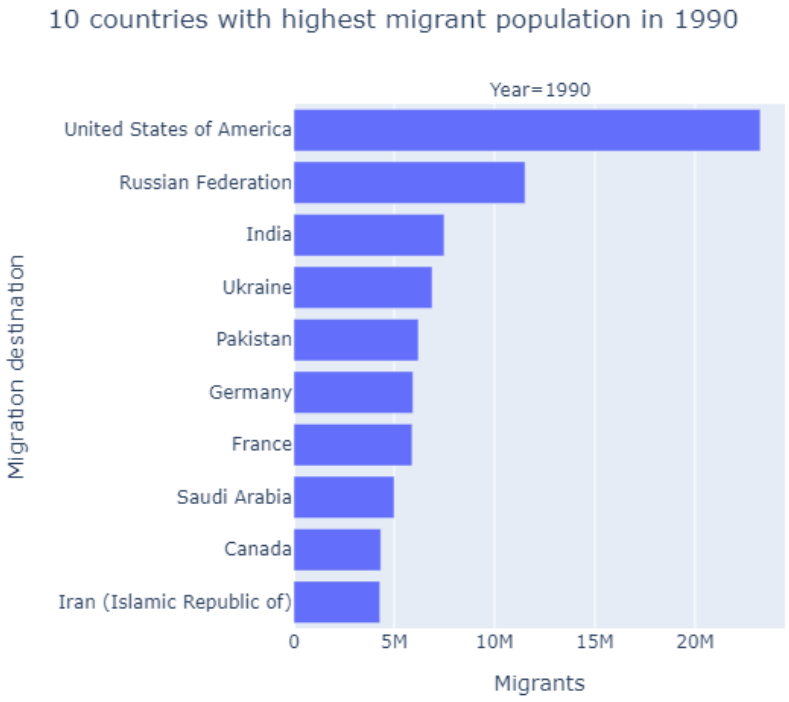
Graph 3: Table 1 – Graph 3, Distribution of female and male migrants by year over six continents.

For these graph I once again used the concepts of sorting, comparison, and accessibility as mentioned in graph one. I also used concepts of small multiples and minimalism to create a readable graph. Rather than creating only one separate graph for each year, I created one graph with six subgraphs to cover the period of time covered by the dataset (Graph 2). I also chose only to use continents rather than countries to make the graphs smaller and minimal (Graph 2 and 3). Finally, I also chose to create a graph with presented statistical data about both sexes which made things even more minimal (Graph 3).

[2] Note that Graph 3 is Graph 2 for tables 6.1, 6.2, and 6.3 to make up for the lack of Graph 2

Graph 4:

As with Graph 2 and 3 which added a new layer of specificity, moving from the world to six continents, Graph 4 now moves even closer to look at countries.[3] This was the longest graph to make but very repetitive. First, I compared the list of migrant destinations from the dataset to the list in the Annex of the original document so that I could select only countries and areas. I then took this new array and created a 2d dataset. I then created a series of six graphs. These were based on year, and number of migrants by country. I then selected the ten countries with the highest number of migrants and plotted it using a horizontal bar graph. This created a list of six horizontal bar graphs which show the ten countries with the largest number of migrants in ascending order over the period of study of the dataset. The purpose of these graphs is to show the variance between the top ten countries. These look as thus:



Graph 4: Table 1 – Graph 4, 10 countries with highest migrant population in 1990.

[3] Note that Graph 4 is Graph 3 for tables 6.1, 6.2, and 6.3 to make up for the lack of Graph 2

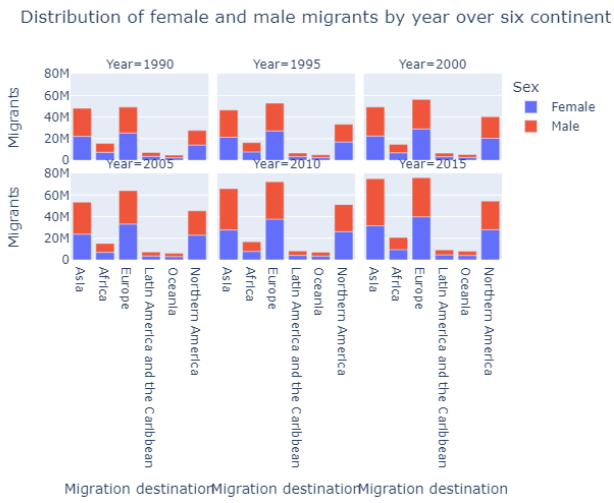
Naturally, I could not show every single country because this would create a massive graph. So, to follow the principles of minimalism and subsetting, I chose only to present ten countries. I also wanted to present all of the graphs in one graph (like for Graph 2), but I could not find a way to merge existing graphs. This would have made the resulting graph better and allowed for a more minimal final graph.

Discussion

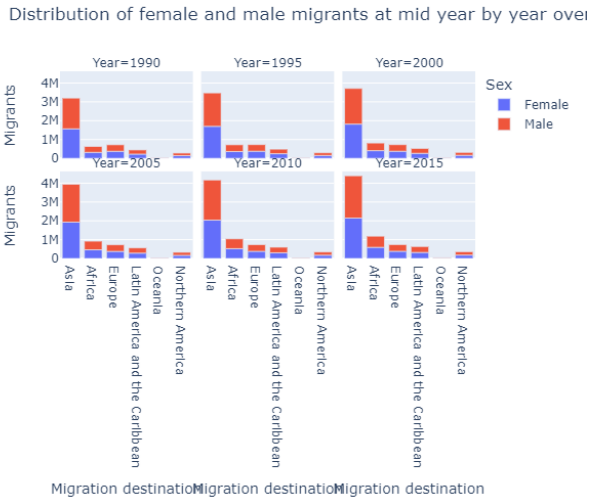
The results of this visualization process created a total of 29 graphs which allow us to see a number of interesting information about the UN dataset: 1) the variance of female and male migrants and refugees and of those of both sexes in the world from 1990 to 2015 (Graph 1); 2) the variance of male and female migrants and refugees in six continents from 1990 to 2015 (Graph 2); 3) the statistical variance of migrants and refugees of both sexes in six continents from 1990 to 2015 (Graph 3); and, 4) the top ten countries with the highest number of migrants and refugees from 1990 to 2015 (Graph 4). By putting all the graphs together, we get comparative visualizations in numbers, in percentages, and in rates of change of some of the important information contained with the UN 2015 migrant dataset.

When using Tukey's and Tufte's principles of exploratory data analysis, I found that they were useful for two different aspects of the project. As discussed in class, Tukey's principles were most useful to take the cleaned data and separate it into easily digestible chunks of information which could then be visualized at a later stage. Tufte's principles aided in this later visualization. The graphs I found most useful are graphs 2 and 3. I believe that by grouping all the graphs (Annex 1) we can track important data based on migrant or refugee sex. This data includes: the distribution by year of female and male migrants and refugees over six continents, the percentage of male and female migrants by year based on total population, the annual rate of change of male and female migrants by year, the distribution of migrants and refugees over six continents between 1990 and 2015, the percentage of migrants and refugees based on total population between 1990 and 2015, the percentage of female migrants based on total population between 1990 and 2015, as well as the annual rate of change of migrants and refugees between 1990 and 2015. We can compare all these graphs to present a comprehensive analysis of the movement and change of migrant and refugee populations over the six inhabited continents between 1990 and 2015.

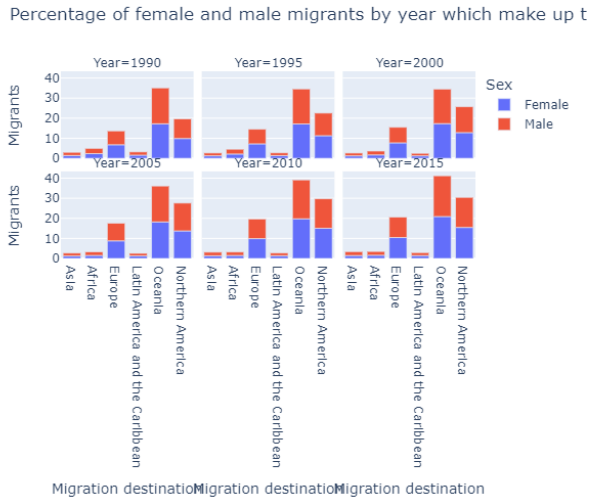
Annex 1



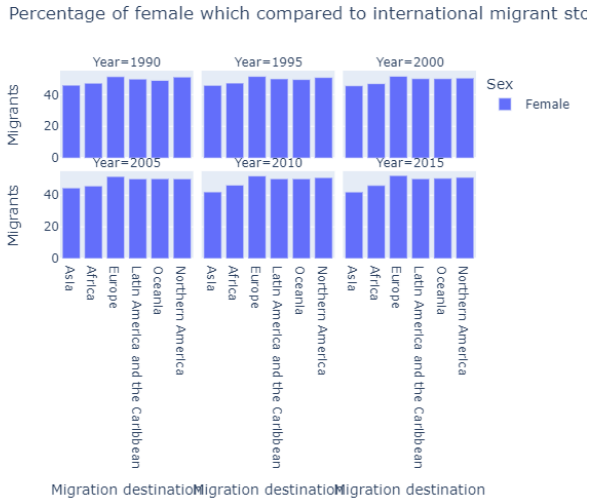
Graph 5: Table 1 – Graph 2



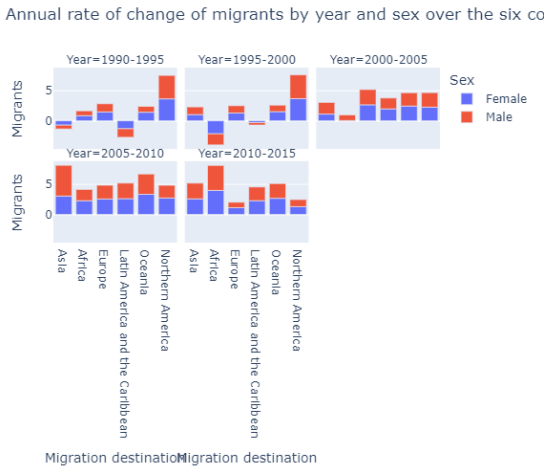
Graph 6: Table 2 – Graph 2



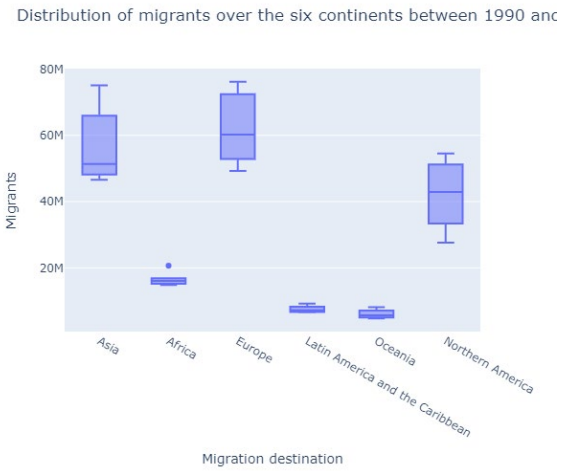
Graph 7: Table 3 – Graph 2



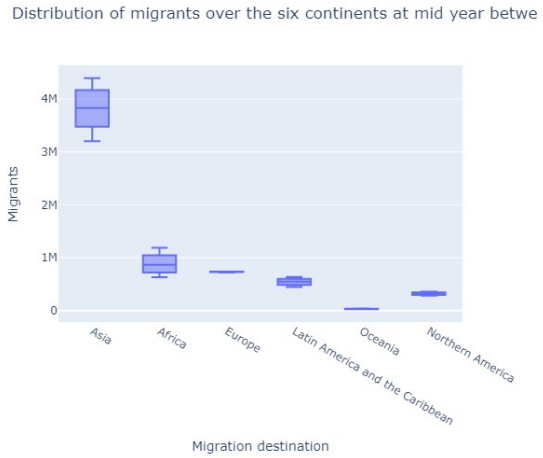
Graph 8: Table 4 – Graph 2



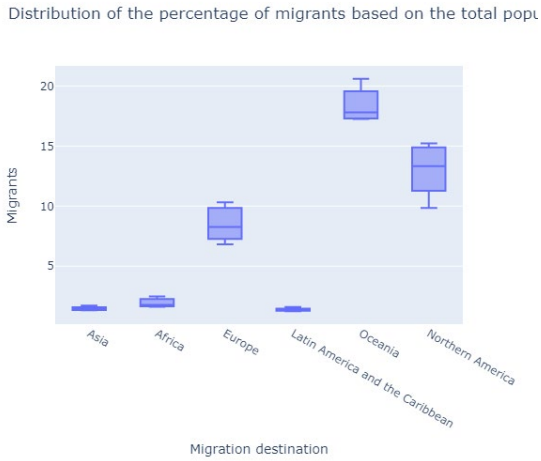
Graph 9: Table 5 – Graph 2



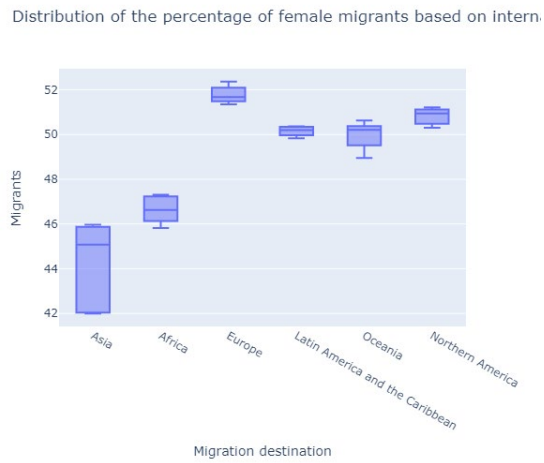
Graph 10: Table 1 – Graph 3



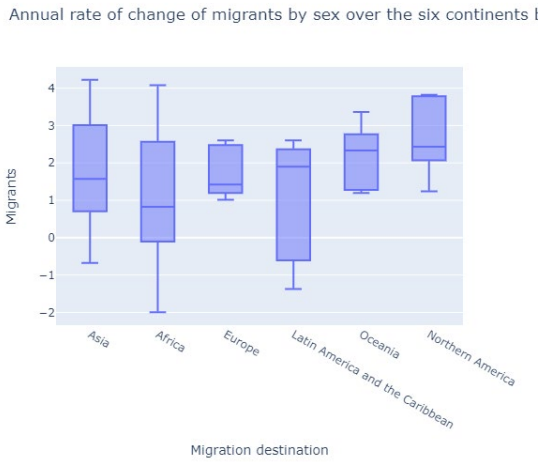
Graph 11: Table 2 – Graph 3



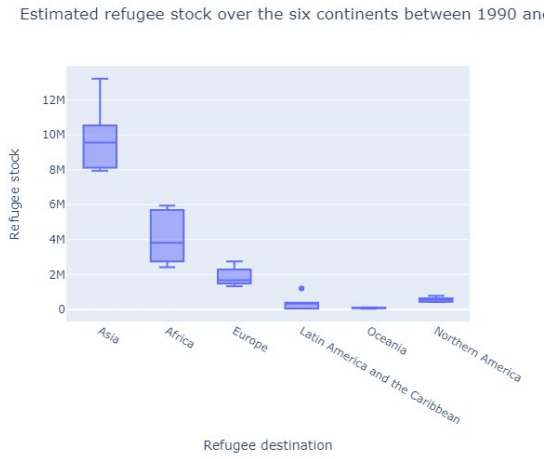
Graph 12: Table 3 – Graph 3



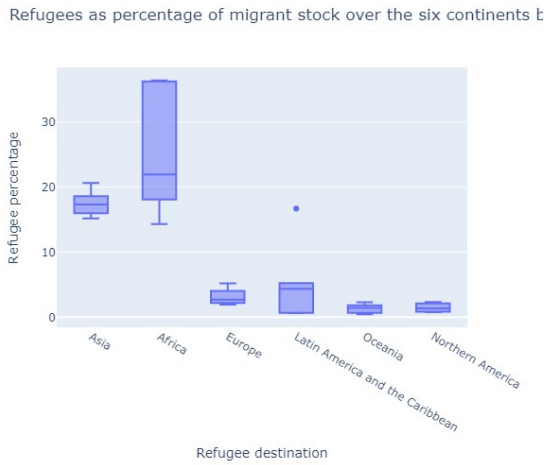
Graph 13: Table 4 – Graph 3



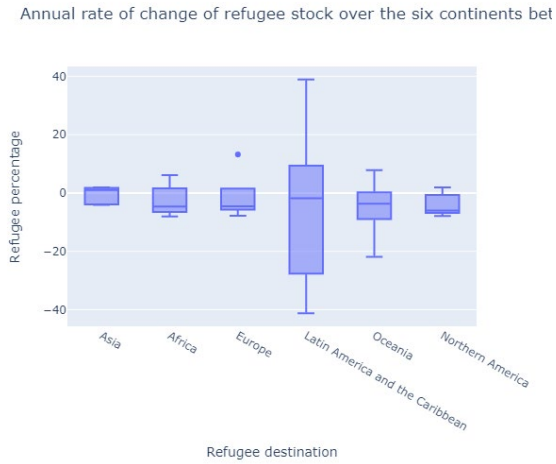
Graph 14: Table 5 – Graph 3



Graph 15: Table 6.1 – Graph 2



Graph 16: Table 6.2 – Graph 2



Graph 17: Table 6.3 – Graph 2