Antonietta Fracassi

INF 1340 F

Final Assignment

Introduction

This final paper will provide an analysis of the data presented by the United Nations "Trends in International Migrant Stock: The 2015 Revision" information. The analysis will include various visualization techniques that will offer insight into the data collected. Tufte's data design principles emphasizes "Graphical Excellence" which will be further explained in the results and discussion sections. The graphs presented will answer some questions about International Migrant Stock, which includes, what are the top countries that most of the migrants settled in, in what year was the greatest amount of migrants and were there more male or female migrants.

Method

The visualization method used consists of graphs such as a bar plot, line plot, box plot, scatterplot and a histogram. For a couple of the tables, the mean (average) was used to depict the trend of migration. All the graphs are clearly labelled with the relevant heading for the graph, the x-axis and the y-axis. The purpose of the graphs is to represent the data in a clear and concise way that provides insight into the data. Utilizing "groupby" functionality and matplotlib and seaborn libraries to perform the visualization tasks offered great insight into the data.

Results & Discussion

Table 1 Analysis

For Table 1 which represents "International Migrant Stock at mid-year by sex and by major area, region, country or area for the years 1990-1995", a bar plot is used. The following question was asked: "What were the top 10 countries with the highest amount of migrants in the year 2015 using the data for "both sexes"? Prior to running the code to create the bar plot, I began by excluding data related to "Major Area" or "Region" as I wanted to capture and sort data based on country. The bar plot is depicted below:

International migrant stock at mid-year, 2015, Top 10 Countries
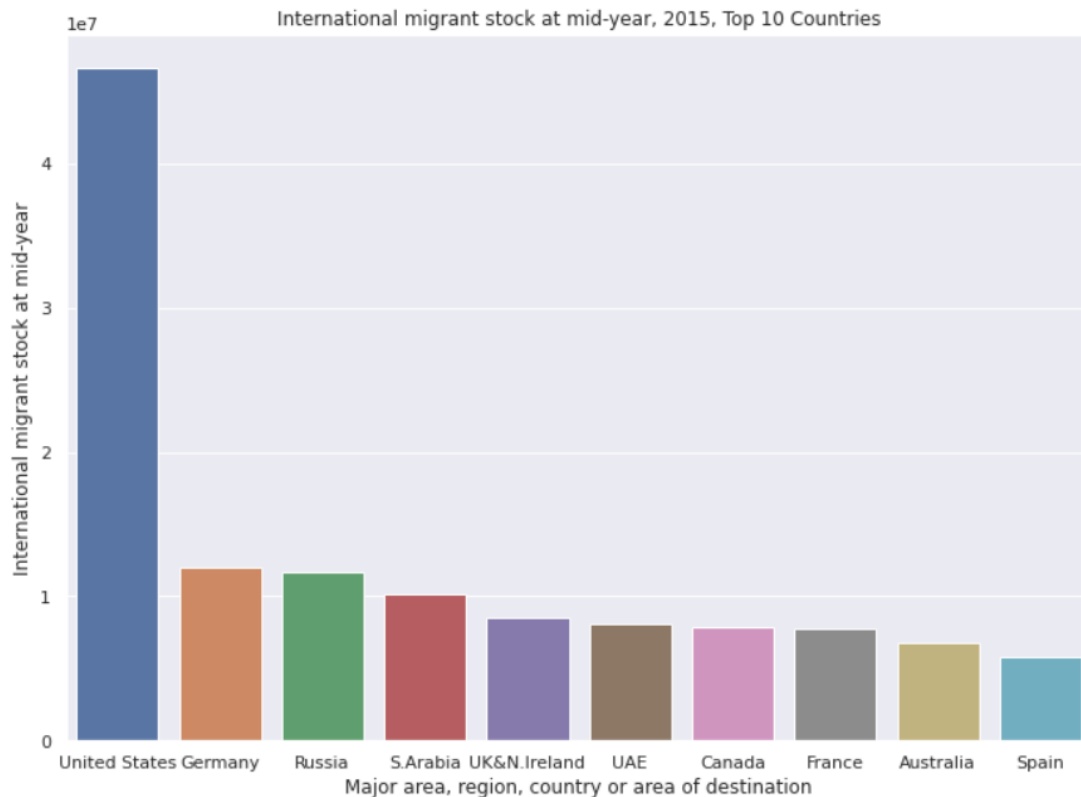
As noted by the graph, United States received the highest amount of Migrants in 2015 followed by Germany, Russia, Saudi Arabia, United Kingdom, United Arab Emirates, Canada, France, Australia and Spain. Sorting the data gives useful insight. In this case, we see from the graph above, that International migrant stock was greatest in the US in 2015.

Table 2 Analysis:

For Table 2 which represents "Total population at mid-year by sex and by major area, region, country or area, from 1990-2015 (thousands)", a line plot was used. In addition, the Total Population mean was calculated. The following question was asked: "On average, were there more male than female migrants for the years 1990-2015?" Prior to running the code to create the bar plot, I began by excluding data related to "Major Area" or "Region" as I wanted to capture and sort data based on country.

The line plot clearly shows that there were more male than female migrants but the lines are very close which allows one to conclude that there isn't a material difference between the amount of male and female migrants.

In addition, this graph also indicates the increase in migrants from 1990 to 2015 has increased significantly. The line graph is a simple way of visualizing the data.
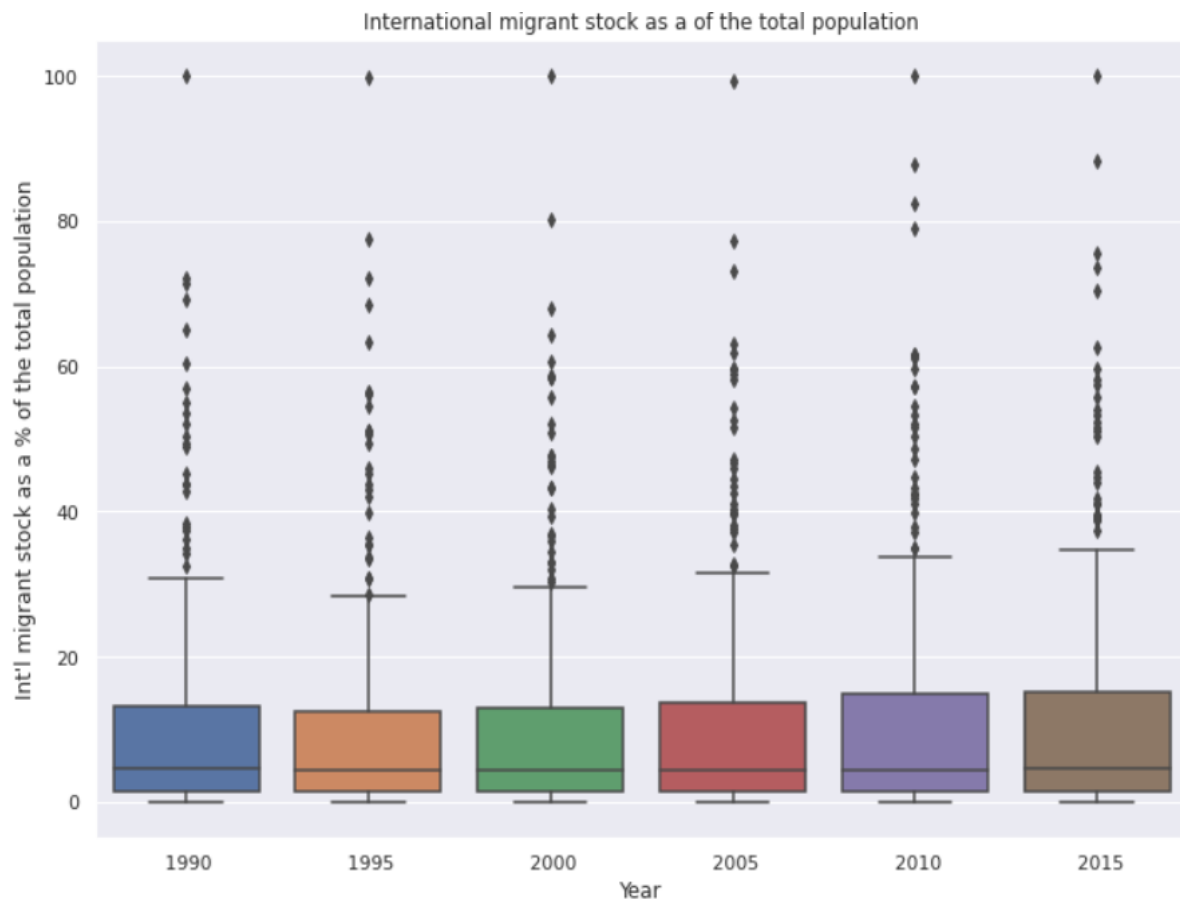

Table 3 Analysis:

For Table 3 which represents "International migrant stock as a percentage of the total population by sex and by major area, region, country or area, from 1990-2015", a box plot was used. The following question was asked: "In which year was the International Migrant stock as a % of the total population the greatest?" Prior to running the code to create the box plot, I began by excluding data related to "Major Area" or "Region" as I wanted to capture and sort data based on country.

The box plot below shows 2015 represents the greatest percentage of International Migrant Stock as a % of total population. This is consistent to Table 2 which shows a growth in migrant stock from 1990 to 2015.

In addition, this graph also demonstrates that there are outliers, specifically, between 80-100%.

International migrant stock as a of the total population

Table 4 Analysis:

For Table 4 which represents "Female migrants as a percentage of the International migrant stock by major area, region, country or area, 1990-2015" a scatter plot was used. The following question was asked: "In which year was there more female migrants as a percentage of the International Migrant stock"?" Prior to running the code to create the scatter plot, I began by excluding data related to "Major Area" or "Region" as I wanted to capture and sort data based on country.

The box plot below shows that in the Year 2000, female migrants as a % of the International migrant stock was greatest and it was the least in 2005.

Female migrants as a % of the Int'l migrant stock by Year
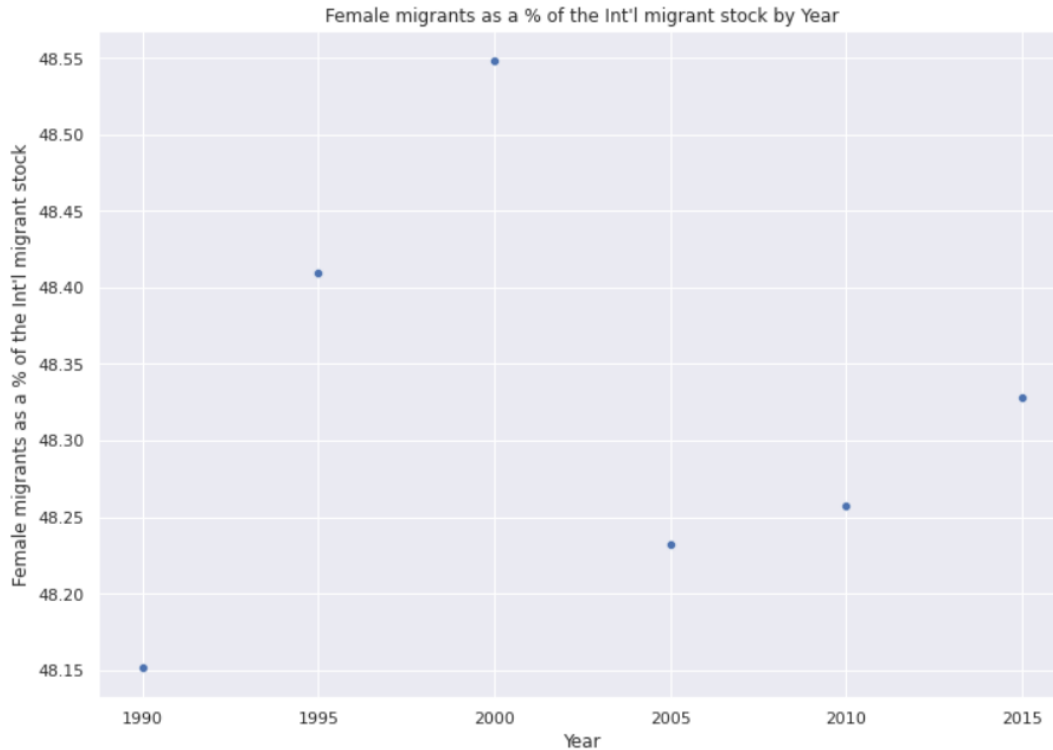
Table 5 Analysis:

For Table 5 which represents "Annual rate of change of the migrant stock by sex and by major area, region, country or area, 1990-2015 (percentage)" a histogram was used. The following question was asked: "What was the highest annual rate of change of International Migrant stock"?" Prior to running the code to create the scatter plot, I began by excluding data related to "Major Area" or "Region" as I wanted to capture and sort data based on country.

The histogram below demonstrates that the highest annual rate of change of International Migrant stock is close to 50. This view also demonstrates the outliers on the left and right of the histogram.

Annual rate of change of the migrant stock, 2005-2010, Both Sexes
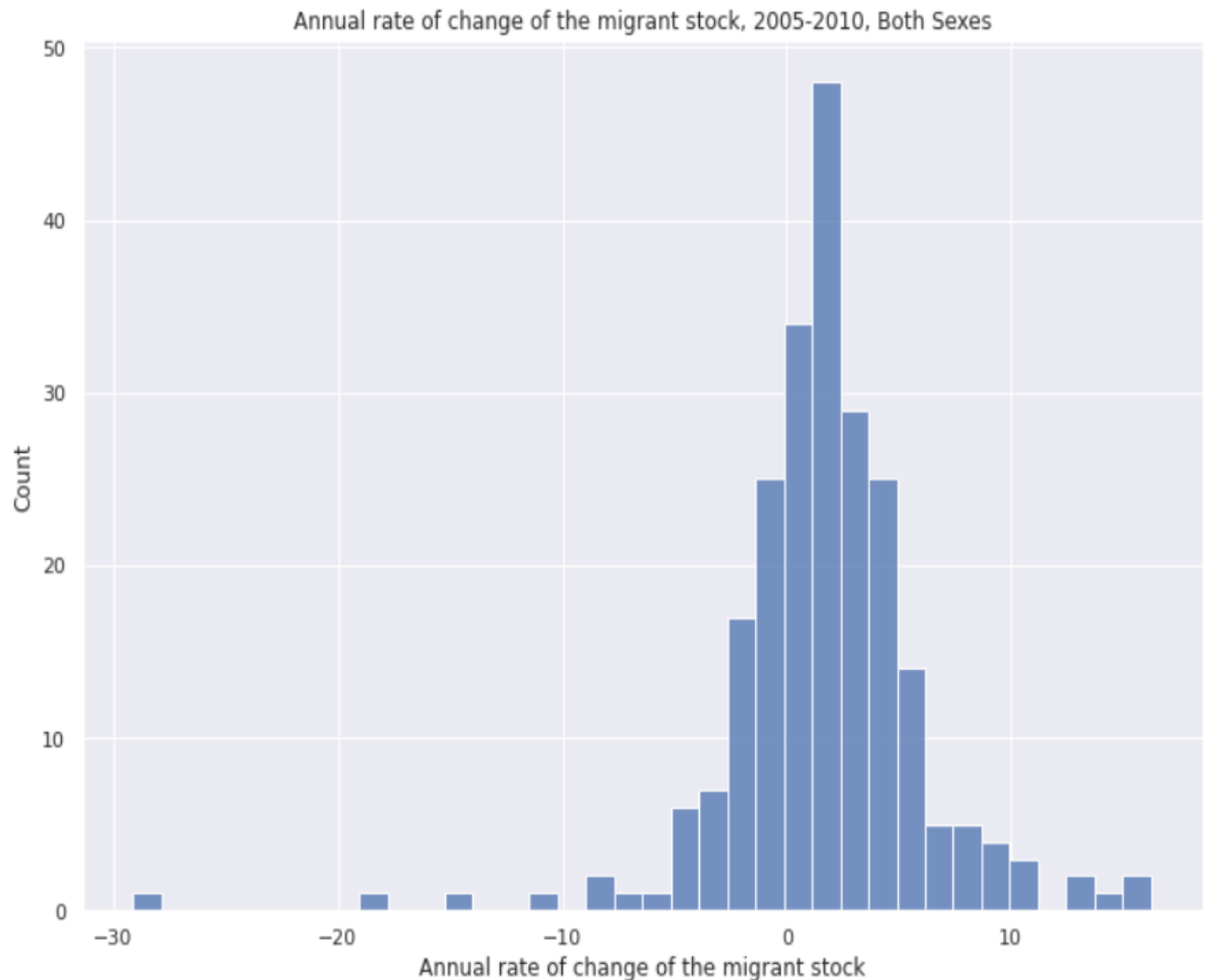
Conclusion

- Visualization techniques allow you to answer your research questions and allows you to explore and gain further insight into the data. As shown above, depicting information in a graph provides a layer of analysis into the data. Being able to tell the story with few words is one of Tufte's principles of graphical excellence. He states that "graphical excellence is that which gives to the viewer the greatest number of ideas in the shortest time with the least ink in the smallest space" (Pursey, 2020). This exercise was challenging, however, it depicted the importance of data visualization and awe inspiring. Being able to present the data in graphical form highlights the message behind the numbers.

References:

Pursey, G., "Tufte's data design principles and insights";   4th January 2020

Link: Tufte's data design principles and insights – Guy Pursey

United Nations, Population Division, Department of Economic and Social Affairs; "Trends in International Migrant Stock: The 2015 Revisions"; United Nations database, 2015.