Name: Sheng Cheng

Student #: 1004769654

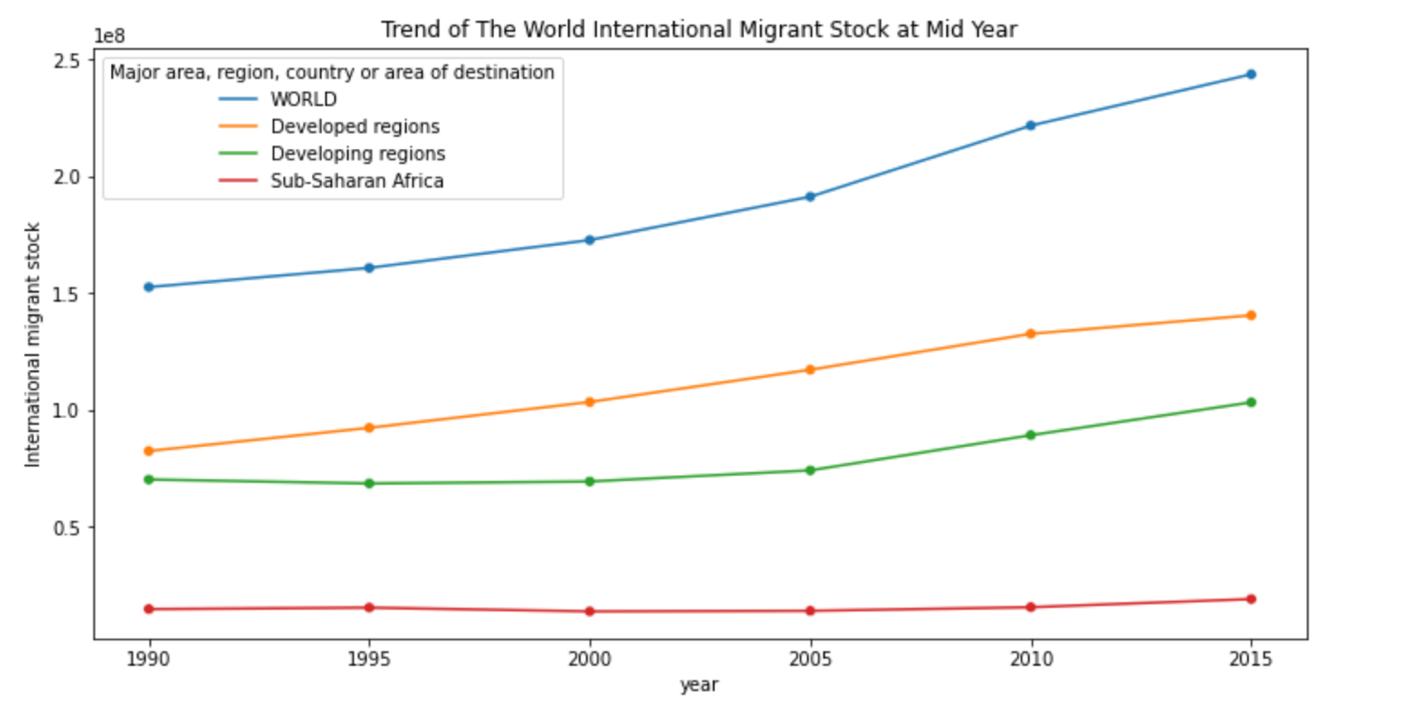
1. Introduction:

In this final project, we need to generate different kinds of figures from the data we cleaned in the midterm project to reveal new insights into our data. This is Data visualization. To analyze vast volumes of data and make data-driven decisions, data visualization tools and technologies are crucial in the world of big data. In this project, Tufte’s principles are the main guideline to build the appropriate figures.

2. Method and Result

Table 1:

Table 1 stores the data on the international migrant stock of many countries, major areas, and regions in mid-years. The world is consisting of three main regions: developed regions, developing regions, and Sub-Saharan Africa. Due to Tufte’s principle three: Show data variation, not the design variation. A line chart is used to show the trend of the international migrant stock of three regions and the world. The chart is listed below.



As is shown in the chart, the international migrant stock of the whole world was keeping increasing from 1990 to 2015. Between 2005 and 2010, the growth rate was relatively fast. From 1990 to 2010, the growth rate of developed regions grew rapidly and steadily. After 2010, it grew relatively slowly. For developing regions, from 1990 to 2005, the growth rate of the migrant stock remained stable. After 2005, it grew rapidly. For Sub-Saharan Africa, the growth rate remained stable.

Gender is another important factor in this table. It can show more variation based on the distribution of males and females in the world's international migrant stock. The bar chart is used to reveal the variation. As shown below, for both males and females, the overall trend was upward. However, as time went on, the gap between males and females had become more pronounced.

Chart, bar chart

Description automatically generated

Next, the boxplot is used to show the difference in the international migrant stock between major areas in 2015. As we can see from this plot, till 2015, the mean value of the international migrant stock in Asia was the highest while that of Oceania was the lowest. At the same time, the data of North America was heavily right skewed, which implied there were some countries in North America that had really high international migrant stock.

Chart

Description automatically generated

Table 2:

Table 2 stores the data on the total population of different countries. The population of major areas and the growth trends are worth analyzing. With the help of Tufte’s principles: sorting rows and columns and small multiplies, I constructed six bar charts with Each chart ranking the six major areas from left to right in order of data size. As we can see, from 1990 to 2015, the total population of Asia maintained the first position and had a big gap with the second place. In 1990 and 1995, the total population of Europe ranked second position. But after 1995, the total population of Africa ranked second position. The total population of Latin America and the Caribbean, North America, and Oceania had been in fourth, fifth, and sixth place respectively.

Chart, bar chart

Description automatically generated

Chart, bar chart

Description automatically generated

Table 3:

Table 3 stores the data on the International migrant stock as a percentage of the total population. Here, I applied Tufte’s principle: sort rows and columns to see which major area had the highest percentage from 1990 to 2015. As shown below, there are two bar charts that show information on the percentage of international migrant stock in 1990 and 2015. The percentage of Oceania was always the highest, at around 20%. Meanwhile, the percentage of Oceania, North America, and Europe was much higher than that of Africa, Latin America and the Caribbean, and Asia.

Chart, bar chart

Description automatically generated

Table 4:

Table 4 stores the information on Female migrants as a percentage of the international migrant stock. To see the variation in this data based on Tufte’s principle 3. A line chart was built to compare the difference in the trend of female migrants as a percentage of the world international migrant stock at mid-year. The female migrant percentage of the developed regions maintained stable from 1990 to 2005. Since 2005, it had risen rapidly. For the female migrant percentage of the world, it kept a downward trend. The trend of female migrant percentage in Sub-Saharan Africa fluctuated, falling from 1990 to 2005, rising to 2010, and then falling again. The trend of female migrant percentage in developing regions was keeping going down. Besides, there is a boxplot generated to see the difference in female migrant percentage between major areas.

Chart, line chart

Description automatically generated

As shown below, the mean value of the female migrant percentage in Europe was the highest. In general, the female migrant percentages of the other five areas were similar. Overall, the volatility was greatest in Asia. Chart, box and whisker chart

Description automatically generated

Table 5:

Table 5 stores the data on the annual rate of change of the international migrant stock at mid-year. Due to Tufte’s principle three: Show data variation. A line chart is used to show the difference in the trend of the annual rate of change in different regions. The trend of the annual rate of Sub-Saharan Africa was the most obvious, the rate of 1995-2000 dropped to the lowest, and then soared all the way, to 2010-2015, the rate had become the highest. The annual rate of developed regions and the whole world had changed a little. The annual rate of developing regions reached peak at 2005-2010, then started to decline.

Chart, line chart

Description automatically generated

The following bar chart is used to compare the difference in the annual rate of major areas. At 1990-1995 and 1995-2000, the annual rate in North America was the highest, And then it went down slowly. The annual rate in Europe went up to 2000-2005, then it went down. The annual rate of Africa, Asia, and Latin America and the Caribbean followed the cosine distribution which presented volatility.

Chart, bar chart

Description automatically generated

Table 6:

Table 6 sores the three different kinds of information which are refugees as a percentage of the international migrant stock, annual rate of change of refugee stock, and estimated refugee stock.

Chart, bar chart

Description automatically generated

The above bar chart is used to show the variation in the refugee percentage. The refugee percentage of Asia was keeping increasing from 1990 to 2015. It had been the highest since 1995. The refugee percentage of Latin America and the Caribbean had been decreasing from 1990. The refugee percentage of Africa was kept in a slightly fluctuating state. The remaining areas remained very low.

The next bar plot is used to compare the annual rate of change of the refugee stock of major areas. The annual rate of Africa fluctuates around zero at very low values. The annual rate of Latin America and the Caribbean was very low which indicated the decreasing state. The rest of the areas fluctuated around zero.

Chart, box and whisker chart

Description automatically generated

The following boxplot is used to compare the estimated refugee stock of different major areas in 2015. As we can see, the mean value of the estimated refugee stock of Africa was the highest. Meanwhile, the estimated refugee stock of Asia and North America was heavily right skewed. By the way, the estimated refugee stock of Oceania and Latin America and the Caribbean was pretty low.

Chart, bar chart, box and whisker chart

Description automatically generated

3. Discussion:

In this project, data visualization plays an important role to reveal the different insights of the data in the table. With the help of the line chart, bar chart, and box plot, we can easily compare the difference in the data at different levels. For example, as I mentioned above, the difference between the major areas, and the major regions. By the way, Tufte’s principles are really helpful to develop clean dashboards. In this process, I once again realized the importance of data cleaning. Due to some incorrect operations in the last project, data visualization has met some troubles.