Yasmin Sawan
Student ID: 1004500505
Final – Write-Up

Introduction:

In this final assignment, I first focused on generating visual graphs depicting data for the world for all six sheets/tables and the corresponding trends between sexes that were grouped on the x-axis for each year interval (note: table six has three clean data tables). I then moved on to visualize the data for developing and developed regions for every sheet. I was able to call on specific country codes through "iloc[]". Tufte's visualization principles indicates that minimal design is key to visualizing data. To minimize design, I did not include all data for the visuals to limit overload of ink that can create confusion. Another principle is small multiples, where you design a series of graphs to compare variables across several categories. To meet this principle, I chose to compare visuals of data on developed and developing regions side by side for some tables and compared Canada to the United States as neighboring developed countries. I also made sure there was a symbiosis between the data (quantification) and design, by consistently labeling the axis and title for each graph, as well as providing interactive visuals through plotly express functions. The plotly express functions can empower the viewer to have the autonomy and control to explore the graph to engage and discover the data presented to them visually. Lastly, I chose to focus on contextual socio-political issues for country codes and their corresponding years to critically understand data trends. Using a historical and socio-political lens allowed me to narrow down my focus on what I was trying to visualize from the data to further interpret it descriptively.

Methods:

Visual 1: All 6 tables

To start, I relied on different visual libraries from python. For all tables, I visualized the data for the world between sexes to show trends over time as a global unit. It is important to display trends between sexes over the years as that avoids the Simpson's Paradox. This was easily visualized through indicating colour codes for sexes. For this, I relied on Matplotlib to create static visuals and defined the data frame. I ensured that the bars were grouped under a given year between sexes, and they are colour coded with a legend for reference. However, some tables only looked at both sexes or females, where I adjusted the code accordingly or resorted to using a line graph instead. This is explained further in the methods sections under the specific sheet name.

Sheet1(IMS2015): International migrant stock 1990-2015

Visual 1: plotly express bar blot of world data between sexes over the years.

Visual 2: I relied on plotly express to visualize all country codes in the dataframe for my own reference (IMS2015_barfig). The x-axis is indicated by sex, and the y-axis is the International Migrant Stock (IMS). This graph may be too complex for viewers and against Tufte's principles, but I kept it to demonstrate my thinking process. This function allows me to hover over the bars on the graph to see their individual legends depicting the full data, which aggregates multiple data point in a rectangular mark, indicating country code, IMS count, year. I can avoid analyzing a large dataset from a table. Color codes and thickness indicates trend in migrant stock in relation to the years to visualize largest IMS.

Visual 3: I created side-by-side box plots using the seaborn library to compare developed vs. developing regions throughout the years and by sex, narrowing the focus from the large dataset. I was able to do so with plt.subplot. The country codes are 901 for developed regions by calling on

"IMS2015.iloc[18:36]", and 902 for developing using IMS2015.iloc[37:54]." I defined the legend using "hue="Countrycode." I also defined the main title and indicated the subtitles for each graph to clarify what each is representing ("Developed", "Developing").
Visual 4: For the last two graphs, I used what I learned from IMS2015_barfig (second chart) by visualizing the continents of Europe and Asia using a bar chart through plotly express. Using the same axes labels, I visualized their trend of IMS over the years by sex.

Sheet2(Pop2015): Total population at mid-year for Both Sexes

Visual 1 and 3: I repeated visuals 1 and 3 from Sheet1.

Sheet3 (IMSP2015): International migrant stock as %of the total population

Visuals 1-2: Repeated visuals 1 and 3 from Sheet1.
Visual 3: I created a line plot to further represent a side-by-side trend of IMS as a percentage of the total population over the years for developed regions. I chose to also do a line plot because the y-axis shows continuous variable, the rate of IMS is a percentage. This time, I wanted to show a 10-year trend that compared the year 2005 and 2015. I found it important to visualize the data in a different format, and to narrow down the focus of the visual from a range of 25 years to 10 years. I ensured the title is clear and included subtitles for each graph to indicate the year.

Sheet4(FM_IMSRate2015): Female migrants as % of the international migrant stock

Visual 1: While this step has the main idea as the first visual for sheets 1 and 3, I used a line graph instead of a grouped bar graph. This dataset looked at the female sex only. I also realized it includes a continuous variable as a percentage ("FMigrants_IMS_Rate"). Therefore I used plotly express (px.line) to visualize the world data for female migrant rates as a percentage of international migrant stock. Data for the world is indicated by the code FM_IMSRate2015.iloc[0:6]. I also renamed the y-axis label for clarity.

Sheet 5 (AnnualRateChange_MS2015): Annual rate of change of the migrant stock

Visual 1: repeated visual 1 from sheets 1-3.

Sheet 6 (ERS_2015)(RPIMS_2015)(ARCRS_2015): Estimated refugee stock by country

Visual 1 ((ERS_2015),(RPIMS_2015)(ARCRS_2015)): I aimed do the same concept for the visual for step 1 in sheets 1 to 3 and 5. However, the sixth sheet only looks at both sexes. I generated three plotly express bar plots that has only one colour code (blue) indicated for both sexes. The "Year" is on the x-axis and the "Estimated Refugee Stock" on the y-axis. I included a title.

(ERS_2015)

Visual 2: To look at the estimated refugee stock trend for Canada and the United States, I used plotly express bar charts. For Canada, I called on the country code using "ERS_2015.iloc[198:204]," and for the United States I used "ERS_2015.iloc[1332:1338]." I indicated the x="Year", and y="Estimated_Refugee_Stock." I then changed the label name of the y-axis using "labels= {'Estimated_Refugee_Stock' : 'Estimated Refugee Stock'}."
Visual 3(2 sub visuals): with the same concept for visual 3 in sheet 1 and beyond, I wanted to plot change in estimated refugee stock over years in developed and developing regions indicated by country code in one graph. However, I chose to do a line plot instead of a bar plot with both country codes (901, 902) indicated by the code"ERS_2015.iloc[1398:1410]." The y axis is the "year," the x axis is the estimated refugee stock. Two lines represent country code by the color indicated by the legend, referring to developed as blue and developing as red. I then created a stacked bar plot by using plotly express bar plot function to display this concept in a different design format. This design is more interactive for the viewer to engage with the visual and hover

over the bars to get the full legend of each region. The two regions are colour coded as yellow for 902, and blue for 901.

(RPIMS_2015)

Visual 2: I wanted to display the trend of refugee rates in the country Guinea over the years. I called on the country code using "RPIMS_2015.iloc[504:510]." Since the refugee rate is a percentage and I am looking at a range of years, I decided to do a plotly express line plot. The x axis indicated the year range 1990-2015, and the y axis is the refugee rate renamed using "labels={'Refugee_Rate_IMS':'Refugees Rate'}."

(ARCRS_2015)

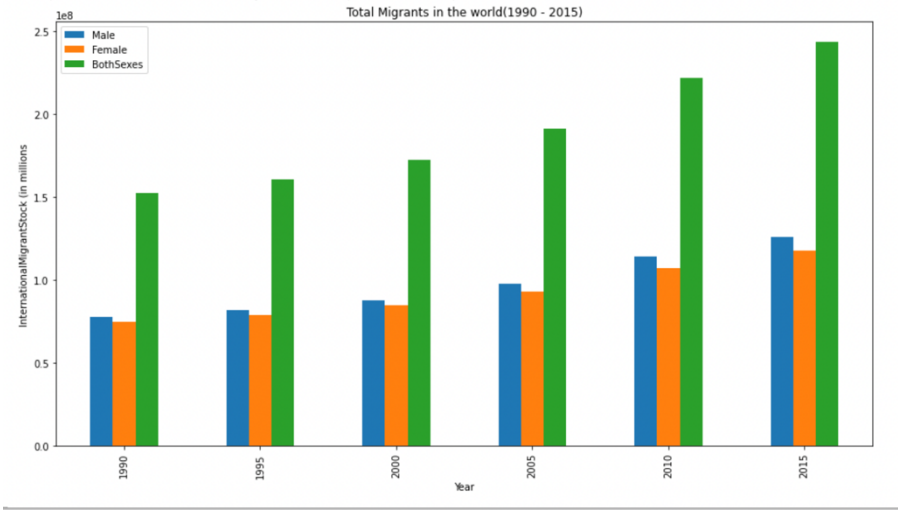
Visual 2: I repeated the same line plot and labels as visual for (RPIMS_2015) above. This time, I plotted for country code 4 which represents Afghanistan by using "RPIMS_2015.iloc[0:6]." I adjusted the title accordingly.

Visual 3: by checking country code 4 for Afghanistan using "ARCRS_2015.iloc[0:5]," there is 0 in the years 1995-2005 but went from 120% between 2005-2010 to 50% between 2010-2015. With existing knowledge on Afghanistan's historical socio-political struggles, I decided to only plot the years 2005-2010 and 2010-2015 using "ARCRS_2015.iloc[3:5]" as it represents a major event occurring.

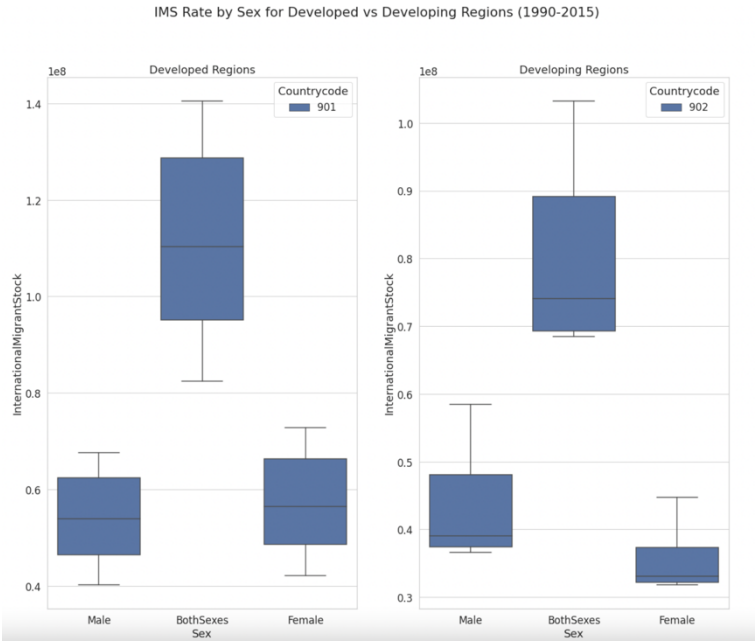
Results:

Sheet1(IMS2015): International migrant stock 1990-2015

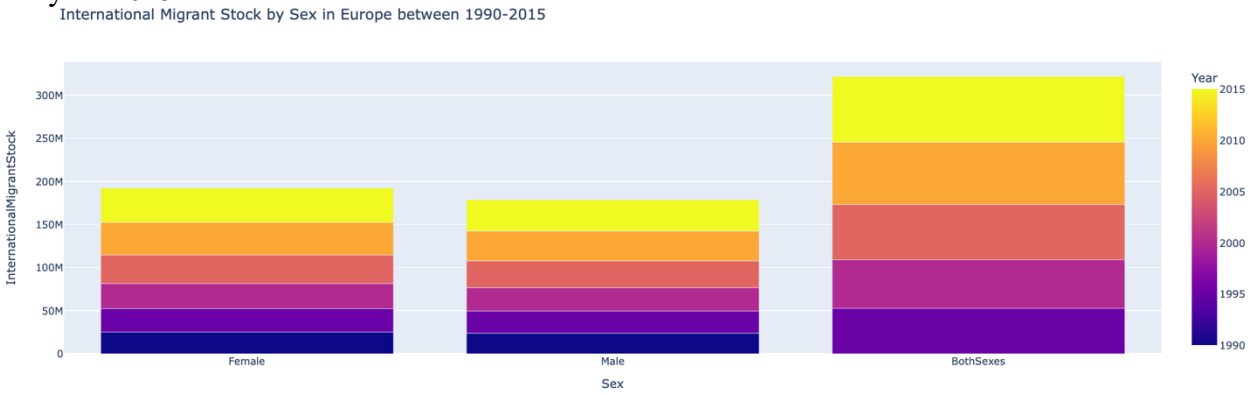
Visual 1: there are more male migrants than females over the years in the world. Both sexes increase over the years*(graph on right)*



Visual 2: Color codes and thickness indicates increase in migrant stock towards the year 2015, with Europe and Asia being the largest. I later visualize these continents separately. Europe with country code 908 has about 76.15 million IMS in 2015, and Asia with country code 935 has about 75.08 million in 2015.

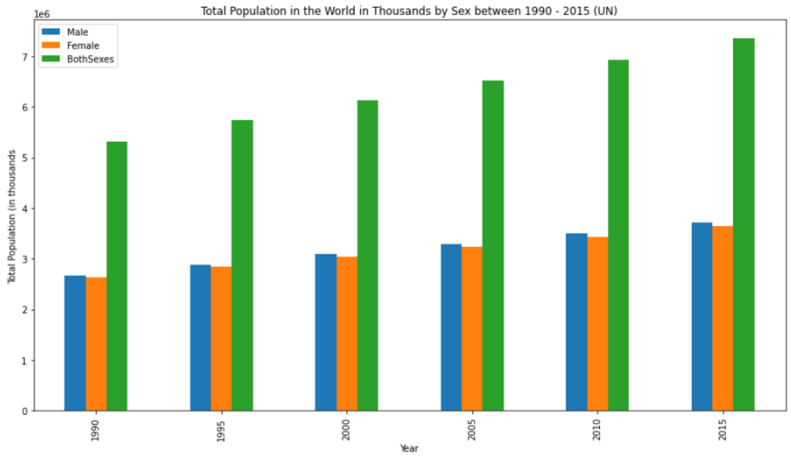


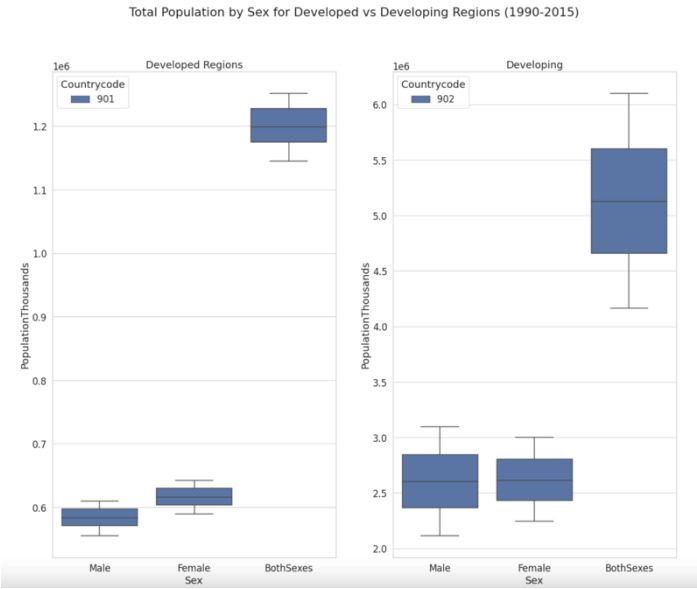
Visual 3: The box plot shows majority of data points within 'box' representing country code (development status), to be able to look at data symmetry and skewness, and outliers. Developed countries with country code 901 displays symmetric distribution and longer boxes than developing regions, meaning more dispersed data over years. Box plots for developing regions, country code 902, shows the distribution is positively skewed for all sexes (skewed right). Higher IMS scores for developing countries than the median for all sexes is also indicated. The long upper whiskers for all sexes in developing regions indicate high variation amongst the positive quartile group (possible outliers). There is also a difference between males and female IMS, with males have a higher IMS count. Median line for male box plot slightly higher than for female, meaning difference between the two indicates most IMS rate are males.
Visual 4: thicker rectangles and colour code yellow indicates the most IMS in Europe has been in the year 2015 for all sexes.



Sheet2(Pop2015): Total Population

Visual 1: population increase per sex over the years, males indicating higher than females.

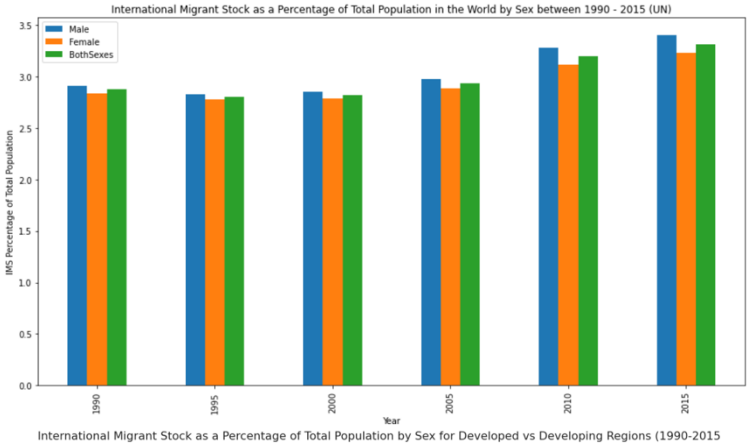


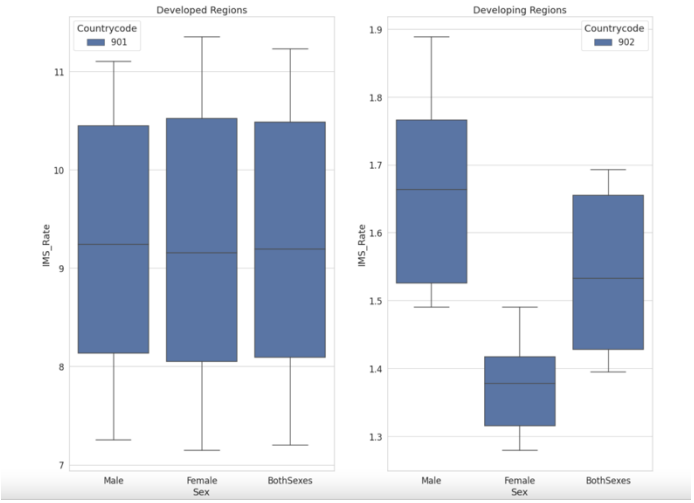


Visual 2: box plots for developed and developing regions depict symmetric distribution of population for all countries with developed or developing statuses. There are longer upper and lower whiskers for males for developing regions compared to females indicating greater range in the sample (greater IQR). There is a greater distribution and IQR for male population in developing countries than developed.

Sheet3(IMSP2015): IMS % total population

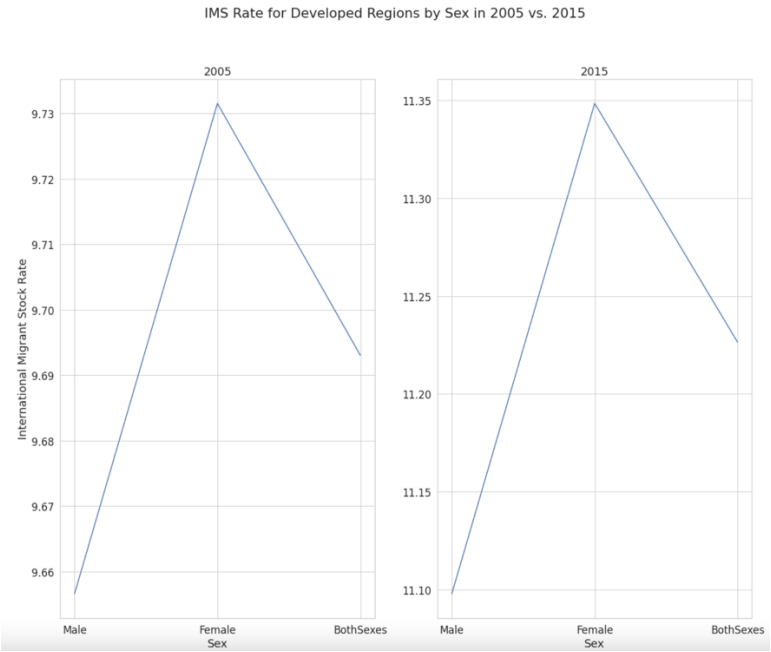
Visual 1: male and female migrants in the world decrease from 1990 to 1995 and increase after 2000 to 2015. Number of migrants for both sexes is less than males, which is useful to visualize the Simpson's paradox. When data is aggregated, it can produce misleading interpretations, however separating them into different bars and colour codes tells us about the actual trend within each sex category.





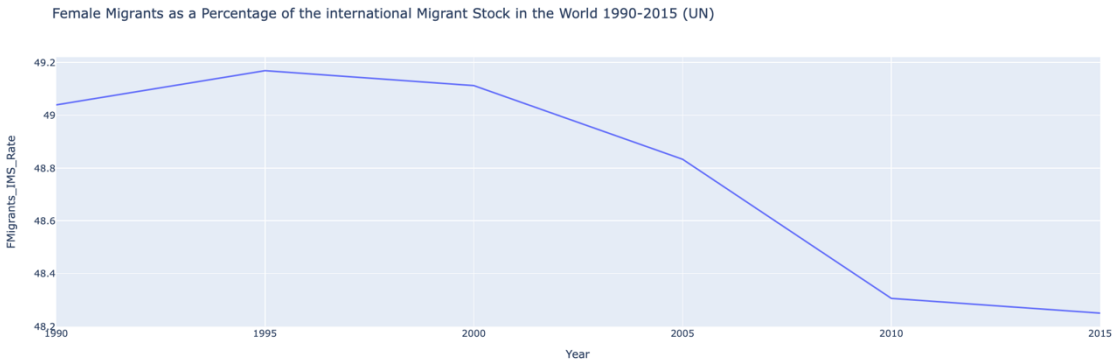
Visual 2(right graph): the boxplot for developed regions indicates symmetric distribution for all sexes and depicts females with a wider distribution and a median less than males. For developing regions, the female and male box are negatively skewed, meaning lower IMS rates than the median. (right graph). Long whisker upper tale for male may indicate outliers from sample.

Visual 3(below): line plots indicated peak trend in IMS rate for developed regions in the year 2005 and 2015 by sex. Peak is seen for females for 2005 about 9.73, and for 2015 it is 11.35.



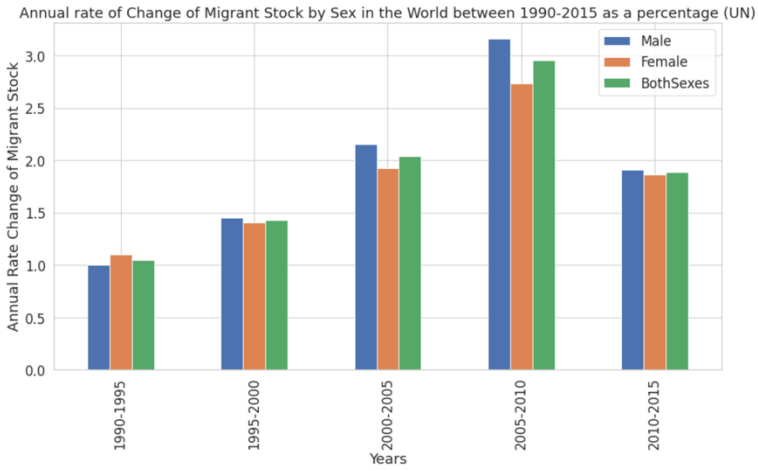
Sheet4(FM_IMSRate2015): Female migrants *(below)*

Visual 1: the peak of female migrants in the world was in 1995, with 49.17% compared to in 2015 with 48.25%. There is an upward trend from 1990 to 1995, then afterwards the female migrant stock decreases in the world at a rather exponential rate (concave down).



Sheet 5 (AnnualRateChange_MS2015): Annual rate of change of the migrant stock

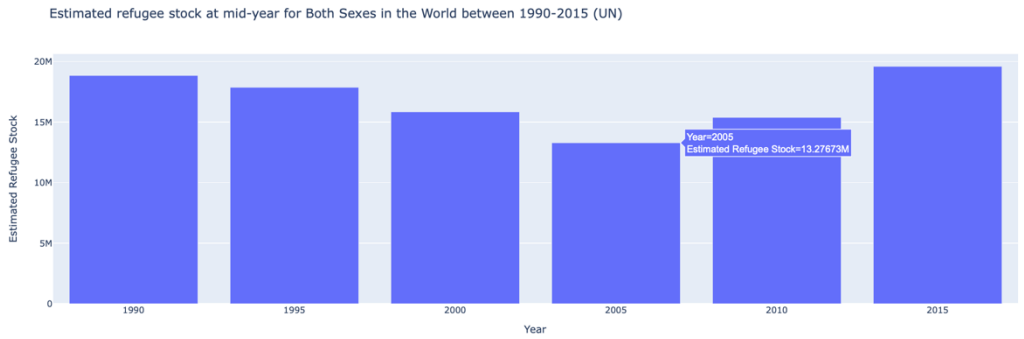
Visual 1: annual rate of change of migrant stock is generally highest for males. The peak for male annual rates is in the years 2005-2010. The trend is depicted by an increase in the annual rate of males from 1990 to 2010, with a significant drop in the years 2010-2015. Additionally, there is consistently higher rates of change for males than females, except in the years 1990-1995. Lastly, rate for both sexes is less than males for the years 1995-2015.

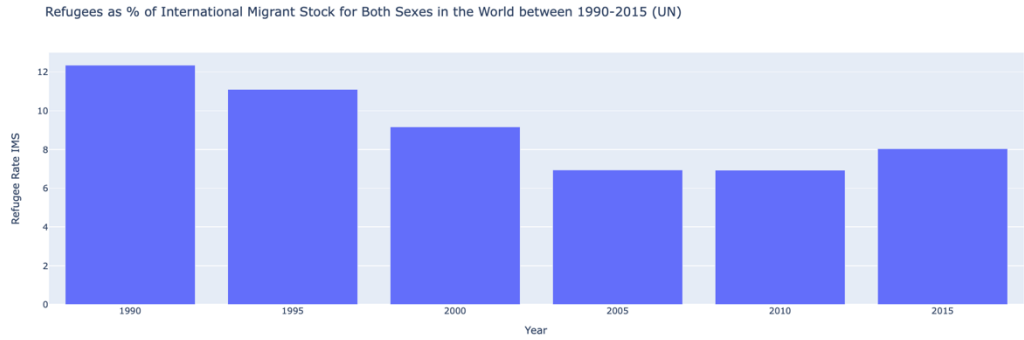


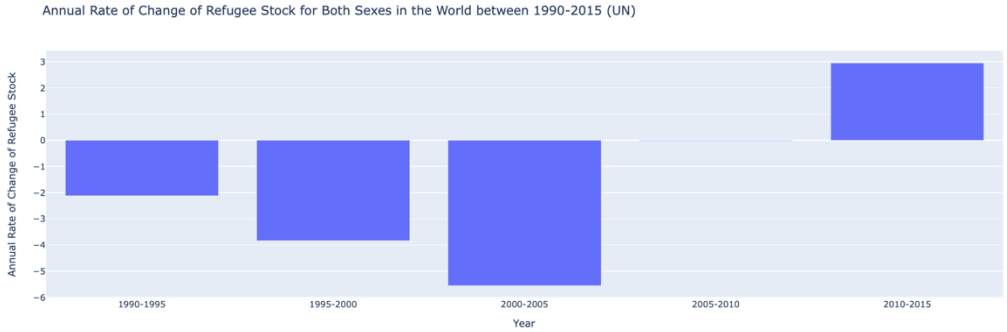
Sheet 6:Estimated refugee stock

(ERS_2015) (RPIMS_2015) (ARCRS_2015):

Visual 1: For the first world visual of *(ERS_2015),* there appears to be a decrease in estimated refugee stock for both sexes 1990 and 2005, the lowest being about 13 million. The estimation increases to 15 million in 2010 to 19.6 million in 2015. For *(RPIMS_2015),* the highest refugee as a percentage of international migrant stock for both sexes were in 1990 at 12.34 million. The percentage decreases throughout the years to 2010 which is the lowest percentage at 6.93 million. There is an increase in 2015 with the percentage being about 8 million. For *(ARCRS_2015),* there is a negative annual rate of change for both sexes in the world from the range years 1990 to 2010. From 2005 to 2010, there was only a 0.03 percent change. The percentage change increases in the years 2010 to 2015 that is about 2.9 percent.

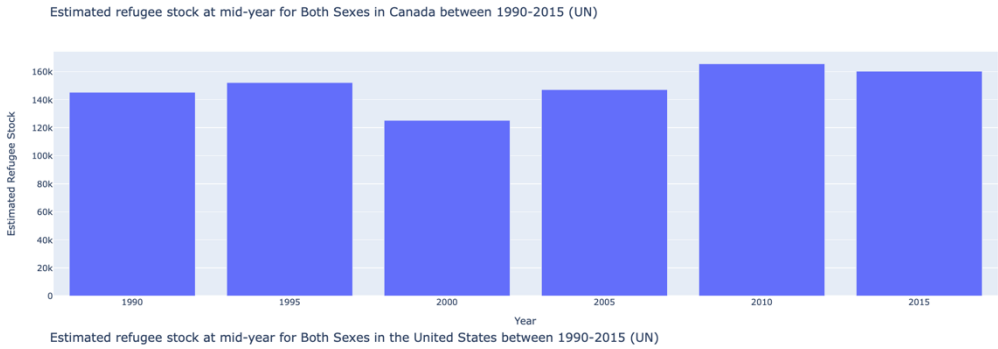


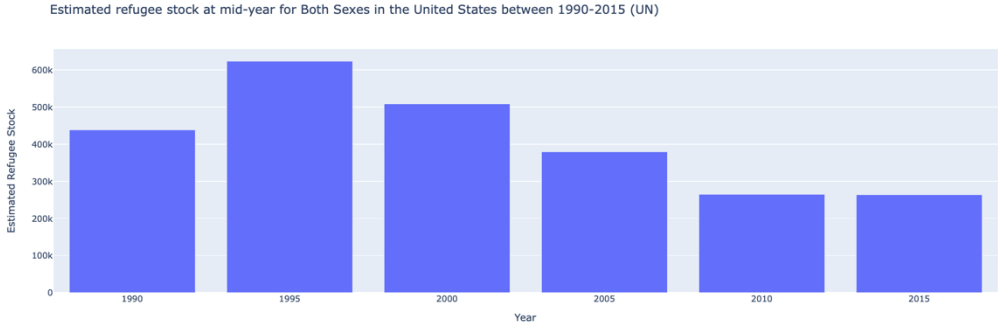




(ERS_2015):

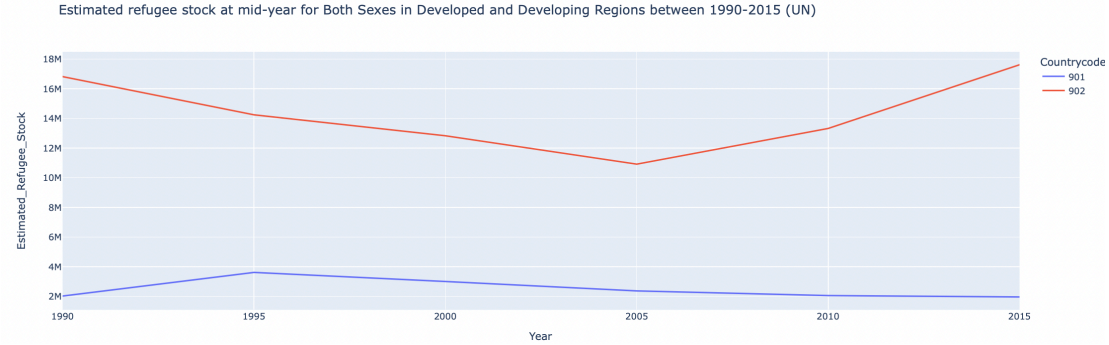
Visual 2: The estimated refugee stock for Canada ranges in thousands from 145 to 165 thousand across the years. However, the dip is seen in the year 2000 being at about 125 thousand. Compared to 2010, which was 165.549k, the year 2015 shows a dip in estimated refugee stock being at 160.29k. For the U.S., the estimated refugee stock went from 438.193k in 1990 to 623.294k





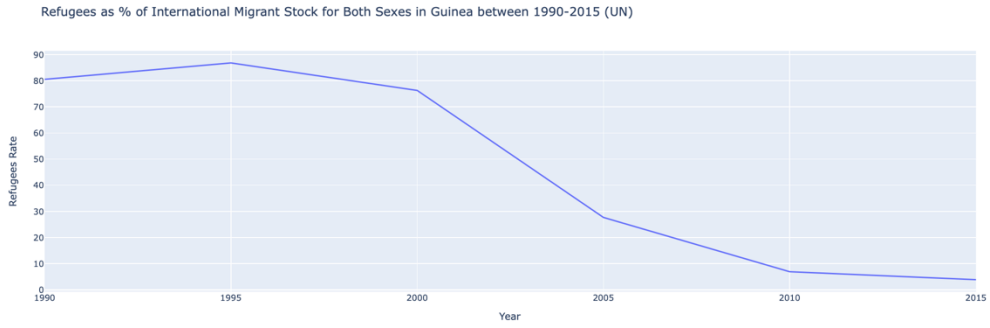
in 1995. The estimation decreases to 263.662k in 2015.

Visual 3: for ERS_2015Devfig, the line plot between developed and developing regions demonstrates higher estimated refugee rates for developing regions than developed from 1990 to 2015. However, while developing regions saw an increase in estimated refugee stock from about 2 million in 1990 to about 3.6 million in 1995, the estimation decreases throughout the years to about 1.95 million in 2015. Comparatively, developing regions experienced a decreased in estimated refugees from about 16.8 million in 1990 to about 10.9 million in 2005, where a sharp increase occurred in the next 10 years to about 17.6 million in 2015. The plotly express bar plot depicts the same trend, but it is not as obvious when quickly analyzing the general trends.

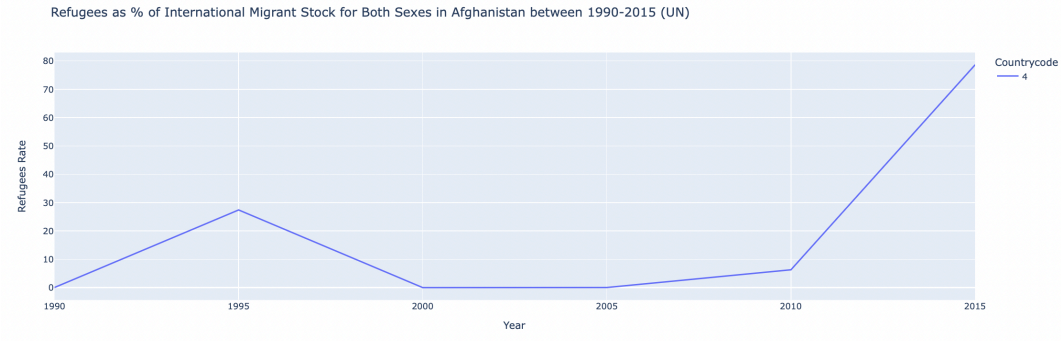


(RPIMS_2015)

Visual 2: the refugee rate for Guinea reveals a downward exponential plot (concave down). The highest rate was in 1995 at about 86.8%. Afterwards, the plot concaves downwards with the lowest rate in 2015 at 3.83%.

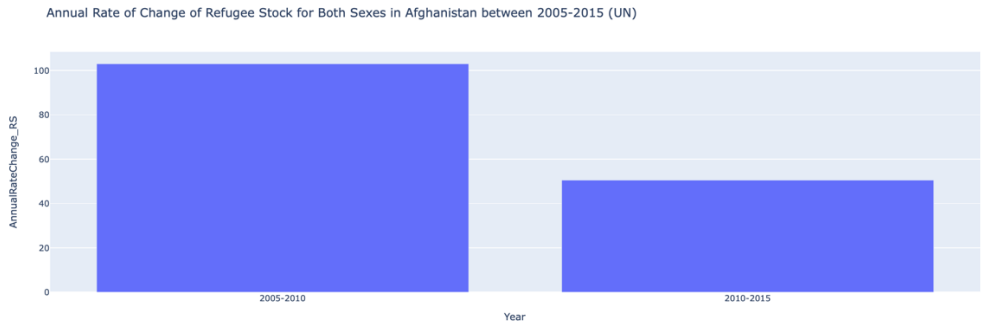


Visual 3: the annual refugee rate for Afghanistan reveals an exponential growth rate from the year 2005 at 0.03% to 78.6% in 2015. In the year 2000, the rate was 0 compared to 27.4% in 1995.



(ARCRS_2015):

Visual 2: looking at Afghanistan from the years 2005 to 2010 and 2010 to 2015, there was a larger rate of change per year in refugee stock in the year at about 103% between 2005 to 2010 compared to 50.5% in the years 2010 to 2015.



Discussion:

Visual1: All tables. While trends or data for both sexes may say high for a given year, the proportion of females may be different than males. Separating the data for sexes can tell viewers, for example, if there were more male international migrant stock rates (% of population) than females in a particular year in the world. This is shown in (IMSP2015) and (AnnualRateChange_MS2015). These graphs depicting world trends possibly signifies the global socio-cultural positionality of males empowered to migrate where women face more difficulties. Females may be socialized to rely on family for migrations and their outcomes (safety, money, stability) (Zlotnik, 1995).

Sheet1 (IMS2015):

Visual 1: there are more male international migrants (those living in a country they were not born in) than females in the world over the 25-year period.

Visual 2: Color codes and thickness indicates increase in migrant stock towards the year 2015, with Europe and Asia being the largest. These continents are geo-politically sensitive to migrant influxes due to being surrounded by many countries with socio-political conflict zones.

Visual 3: the long upper whiskers for all sexes in developing regions indicates high variation amongst the positive quartile group, meaning a lot of developing countries have an IMS above the median. The long whiskers can indicate outliers in the distribution of data.

Visual 4: Europe is in a geo-political position for incoming migrants in surrounding regions. The year 2015 is indicated as Europe's migrant crises, such as events of the Syrian war that year where Syrians fled to Greece, Hungary, Turkey, Germany, and more (United Nations High Commissioner for Refugees). This provides context to the large IMS that year for the continent.

Sheet2(Pop2015): Total Population

Visual 1: there are more males than females in the world over over the 25-year period.

Visual 2: the long box plot for males for developing regions indicates the data may be unreliable.

Sheet3(IMSP2015): IMS % total population

Visual 1: there are more male IMS rates than females across the world over over the 25-year period. This makes sense relative to the previous 2 sheets on data about the world.

Visual 3: peak in 2005 and 2015 indicates IMS rate for females is higher than males in those years.

Sheet4(FM_IMSRate2015): Female migrants (below)

Visual 1: there has a been an exponential decrease in the percentage of female migrants from IMS in the world beginning in 1995.

Sheet 5 (AnnualRateChange_MS2015): Annual rate of change of the migrant stock

Visual 1: annual rate of change of migrant stock in the world for both sexes separately decreased significantly after the year 2010. It is most close to the years 2000 to 2005, signifying a decrease in how frequent people migrate/become migrants over the world.

Sheet 6: Estimated refugee stock *(ERS_2015) (RPIMS_2015) (ARCRS_2015):*

Visual 1: the least amount of estimated refugee stock in the world for both sexes were in 2005 while the highest was in 2015. The percentage of refugee stock for IMS has also increased towards 2015. Lastly, a significant increase in the annual rate of change of refugee stock occurred after the year 2010. There were many regional political conflicts in different countries across the world during the year of 2015, including but not limited to Afghanistan, the aftermath of the Syrian civil war from 2011 that created a ripple effect into the Jordan-Syria border, Yemeni civil war, Burkina Faso, Burundi. These conflicts forced people to seek refuge in neighboring nations.

(ERS_2015)

Visual 2: the decline in estimated refugee stock for both sexes in the U.S. after the year 1995 is significantly different than its northern neighbor Canada, which saw an increase throughout the years. The U.S. experienced a terrorist attack on 9/11 which enforced stricter national security measures and immigration and refugee policies (Martin and Yankay, 2012, p.3). Canada is known to implement more pro-immigration policies, such as for economic reasons.

Visual 3: the trend indicates that there are more estimated refugee stocks for developing regions than developed. This may be due to more flexible borders flows to travel between developing countries compared to stricter security protocols for developed ones.

(RPIMS_2015)

Visual 2: Guinea was positioned as a refugee host country for those fleeing from conflict in Liberia from 1995 to 2007, and those from Sierra Leone, and Côte d'Ivoire (United Nations High Commissioner for Refugees).

Visual 3: Afghanistan experiences a continuous political conflict and is a source country of refugees or those seeking asylum. It also experienced a significant deterioration in security and economy in 2010.

(ARCRS_2015): Visual 2: Bilateral agreements and improved measures for Afghani refugees saw a decrease in the rate of change, thereby for the first time in 30 years Afghanistan lost its status as world's largest refugee population (Willner-Reid, 2017). However, this rate has increased since 2015 (Willner-Reid, 2017).

References

United Nations High Commissioner for Refugees. (n.d.). *UNHCR CDR Background Paper on Refugees and asylum seekers from Afghanistan*. Refworld. Retrieved December 20, 2022, from https://www.refworld.org/docid/3ae6a6440.html

United Nations High Commissioner for Refugees. (n.d.). *2015: The Year of Europe's refugee crisis*. UNHCR. Retrieved December 20, 2022, from https://www.unhcr.org/news/stories/2015/12/56ec1ebde/2015-year-europes-refugee-crisis.html

United Nations High Commissioner for Refugees. (n.d.). *World Refugee Survey 2008 - Guinea*. Refworld. Retrieved December 20, 2022, from https://www.refworld.org/docid/485f50d72.html

Willner-Reid, M. (2020, October 22). *Afghanistan: Displacement challenges in a country on the move*. migrationpolicy.org. Retrieved December 20, 2022, from https://www.migrationpolicy.org/article/afghanistan-displacement-challenges-country-move

Zlotnik H. (1995). Migration and the family: the female perspective. Asian and Pacific migration journal : APMJ, 4(2-3), 253–271. https://doi.org/10.1177/011719689500400205

Appendix I

Midterm Clean Code:

I made some iterations to all the clean tables I submitted for the midterm.

- I dropped the index from “set.index()” I indicated previously as the ‘Sort Order,” to be able to sort through columns and rows. Setting the index could still allow you to do that but I felt more comfortable navigating the coding process for the visuals by removing it.
- I also removed “Sort Order,” and “Location” from every clean data-frame to resort to the country codes that is standardized throughout the datasets.
- I saved the dataframes as a pickle file for each clean table