Geoffrey Crowther Final Report

Introduction

When attempting to visualize the UN Migrant Total 2015 dataset I began by sorting the first two excel sheets in descending order to see if there were any general trends that became apparent in the data. After this, I made a line graph for both the migrant and total populations of the world as well as the developed (according to the notes comprising: "Europe, Northern America, Australia/New Zealand and Japan.") and developing regions (according to the notes comprising: "all regions of Africa, Asia (except Japan), Latin America and the Caribbean plus Melanesia, Micronesia and Polynesia.") alongside them. This showed a general increase in population for all of them although not at the same rate, with the developed regions growing at a greater rate than the developing regions and the world growing the most being the aggregate of them all. Next, I subset the migrant and total population data into the major regions (continents) as defined by the dataset, namely: Africa, Asia, Europe, Latin America & the Caribbean, Northern America, and Oceania. After performing subsetting the data in accordance with Tukey's EDA principles, I then made a small multiple of line graphs in accordance with Tufte's principles of visualization to show the demographic trends over the years measured in the dataset. This produced 2 different sets of small multiples showing demographic trends over time (1990-2015 specifically) for the migrant and total population subsets respectively. These relationships were also displayed in the form of a horizontal bar graph for each respective subset to allow for easier comparisons than is allowed by the small multiples of line graphs. The migrant and total populations for these major regions were also graphed in boxplots and delineated based on whether the sex of the migrant or population was male, female, or both (an aggregate of the other two). The next relationship graphed was a horizontal bar plot of which percentage of the total population of an area was migrants at the time of measurement. The last relationship graphed was another horizontal bar plot of the percentage of refugees of the migrant stock in each major area. The last two graphs were especially pertinent in drawing conclusions from the dataset as Northern America, Europe, and Oceania had the greatest portion of migrants in their total population. As well Africa, Asia, and Europe had the greatest portion of refugees as their migrants over the time period in the dataset. This indicates that along with the generally increasing population and migrants over time

there is a greater amount of migration to those former areas and a greater amount of refugees migrating to the latter areas.

Methods

The first thing that was done to the previously cleaned UN Migrant Total 2015 dataset was to convert the migrant population column to a numeric variable so that further sorting and graphing could be done to it via the following code.

```
#first convert variables to numeric
UN1['Migrant Population']=pd.to_numeric(UN1['Migrant Population'], errors=
'coerce')
#errors just ignores Nan values
```

Once this was done an initial sort was done for the population column by descending order which mainly highlights the “WORLD”, “developing regions”, and “developed regions” as the largest values. This was done in accordance with Tukey's principles of exploratory data analysis namely the "sort" principle as the data was sorted in a way that made further patterns apparent.

```
#a quick sort by highest migrant population down
UN1sort=UN1.sort_values('Migrant Population', ascending=False)
```

As those values were the most immediately visible for discovering general population trends they were further subsetted and then graphed in a basic point-line graph to illustrate the demographic trends worldwide through the following code. This also met Tukey’s principle of ‘subset’ as the most pertinent data was selected

```
#sort the populations by World, Developing, Developed region
UN1MJpop=UN1.loc[UN1['Country Code'].isin([900,901,902])]
UN1MJPop=UN1MJpop.loc[UN1['Sex'].isin(['Both'])]
#graph of the world, developed, developing subset
plot1=sns.catplot(data=UN1MJPop, kind="point", x="Year of Measure", y="Mig
rant Population", hue = "Geographic Area",height=8).set(title="Migrant Sto
ck (1990-2015) Worldwide")
```

This code produced the following visualization:

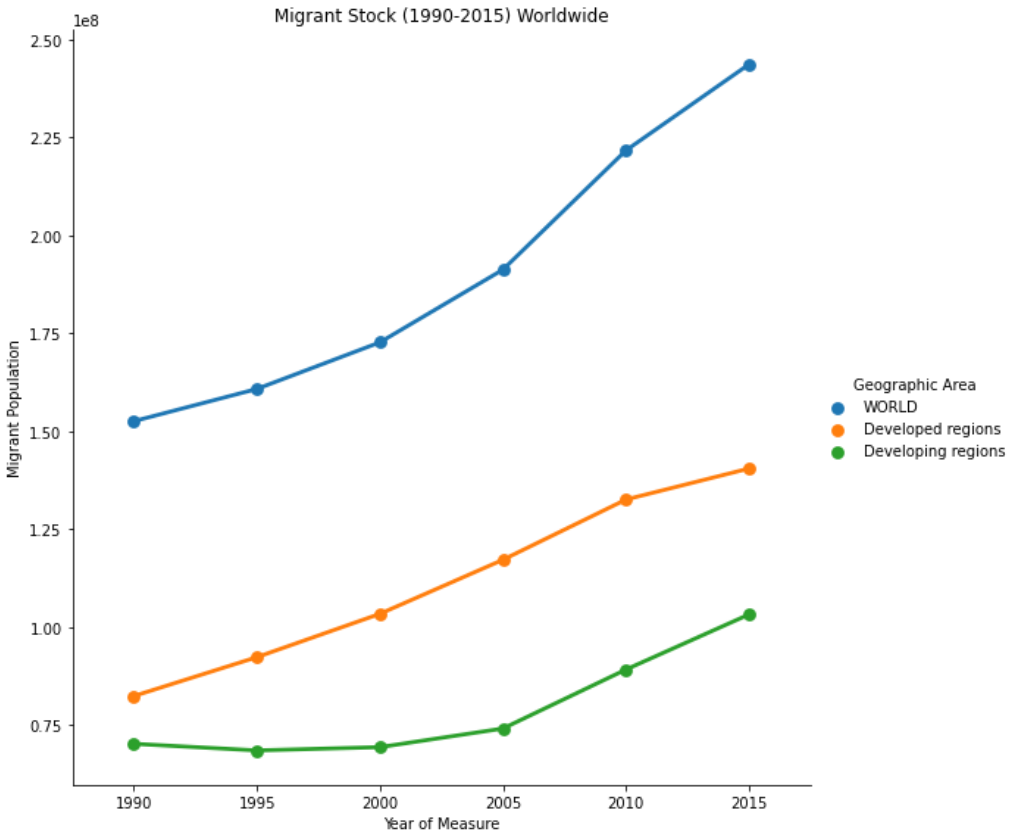

Figure 1: Migrant Stock Worldwide

After a general demographic trend was captured in Figure 1, a more granular approach was taken by focusing on the major continental regions (as defined in the dataset) inhabited by humans including Africa, Asia, Europe, Latin America & the Caribbean, Northern America, and Oceania. The demographic trends of these areas were discovered through soring and then graphed in a set of line graphs as a set of small multiples in accordance with Tufte's principle of "small multiples" and Tukey's "sort" principle through the following code.

```
#make a new dataframe to hold each of the major regions, Africa, Asia, Lat
in America and Caribbean, North America, Oceania, while signalling out onl
y Both sex values
UN1Rpop=UN1.loc[UN1['Country Code'].isin([903,935,908,904,905,909])]
UN1RPop=UN1Rpop.loc[UN1['Sex'].isin(['Both'])]
#Sort of major region migrant pop from top down
UN1sort2=UN1RPop.sort_values('Migrant Population', ascending=False)
display(UN1sort2)
#plot 2 small mutli of migrant pop in major areas
```

```
fig, ((ax1, ax2, ax3), (ax4, ax5, ax6)) = plt.subplots(nrows=2, ncols=3, s
harex=True, sharey=True, figsize=(15,15))
#Africa graph
UN1RPop[UN1RPop['Geographic Area']=='Africa'].plot(x='Year of Measure', y=
'Migrant Population',legend=False,ax=ax1)
ax1.set_title("Migrant Pop. Africa")
ax1.set_ylabel("Migrant Population")
#Asia graph
UN1RPop[UN1RPop['Geographic Area']=='Asia'].plot(x='Year of Measure', y='M
igrant Population',legend=False,ax=ax2)
ax2.set_title("Migrant Pop. Asia")
ax2.set_ylabel("Migrant Population")
#Europe graph
UN1RPop[UN1RPop['Geographic Area']=='Europe'].plot(x='Year of Measure', y=
'Migrant Population',legend=False,ax=ax3)
ax3.set_title("Migrant Pop. Europe")
ax3.set_ylabel("Migrant Population")
#Latin America and Caribbean graph
UN1RPop[UN1RPop['Geographic Area']=='Latin America and the Caribbean'].plo
t(x='Year of Measure', y='Migrant Population',legend=False,ax=ax4)
ax4.set_title("Migrant Pop. Latin America & the Caribbean")
ax4.set_ylabel("Migrant Population")
#North America graph
UN1RPop[UN1RPop['Geographic Area']=='Northern America'].plot(x='Year of Me
asure', y='Migrant Population',legend=False,ax=ax5)
ax5.set_title("Migrant Pop. Northern America")
ax5.set_ylabel("Migrant Population")
#Oceania graph
UN1RPop[UN1RPop['Geographic Area']=='Oceania'].plot(x='Year of Measure', y
='Migrant Population',legend=False,ax=ax6)
ax6.set_title("Migrant Pop. Oceania")
ax6.set_ylabel("Migrant Population")
#Figure title

plt.suptitle('Migrant Pop. in each Major Region',fontweight='bold',fontsiz
e=12,y=1.02)
plt.tight_layout()
```

This code produced the following set of plots:

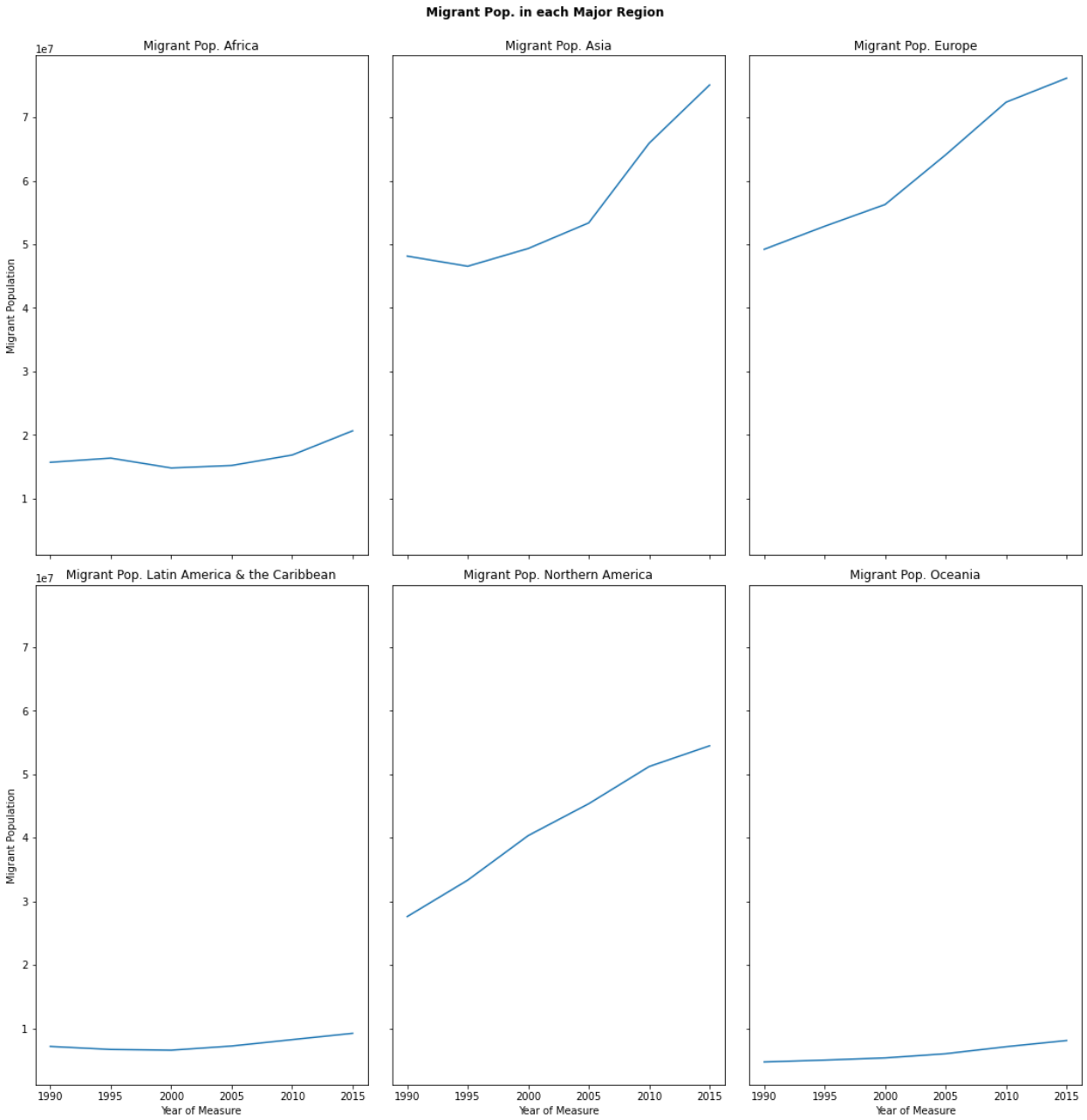

Figure 2: small multiples of major regions

The same relationship was also graphed as a horizontal bar chart in order to allow for easier comparison of the various regions and their trends through the following code:

```
#plot 3 shows count of migrant stock in major regions
plot3=sns.catplot(data=UN1RPop, kind="bar", y="Year of Measure", x="Migrant Population", hue = "Geographic Area",orient='h',height=8).set(title="Migrant Stock in Major Regions")
```

This code produced this plot:

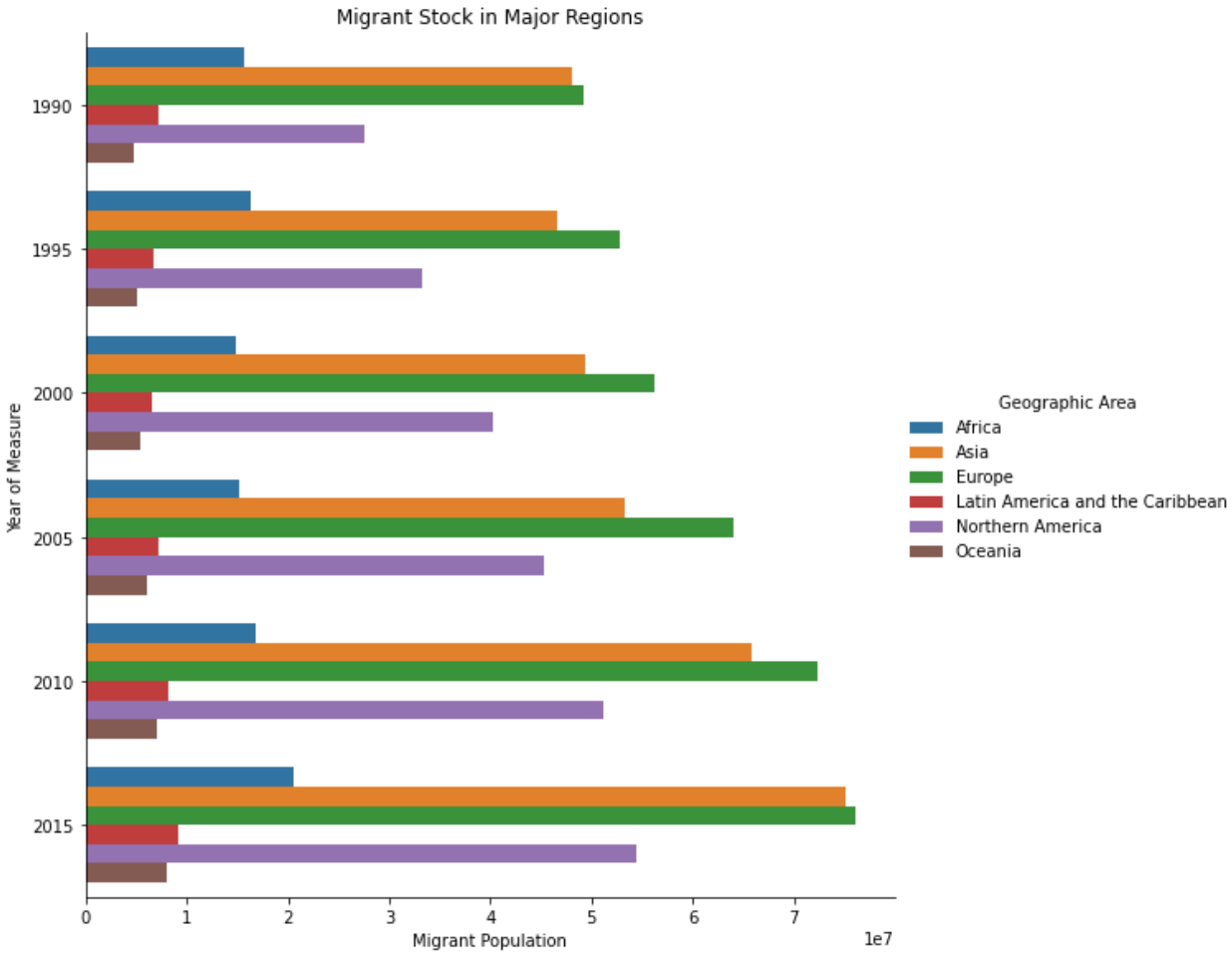

Figure 3: bar plot of migrant demographic trends in Major Regions

Lastly, the migrant data was also displayed in a boxplot and this time subset to include the sex of the particular migrant as well to take a more granular view of the overall demographic trend through the following code.

```
#plot 4 shows count of migrant stock in major regions by sex
plot4=sns.catplot(data=UN1Rpop, kind="box", y="Year of Measure", x="Migran
t Population", hue = "Sex",height=8).set(title="Migrant Stock in Major Reg
ions by sex")
```

This code produced this plot:

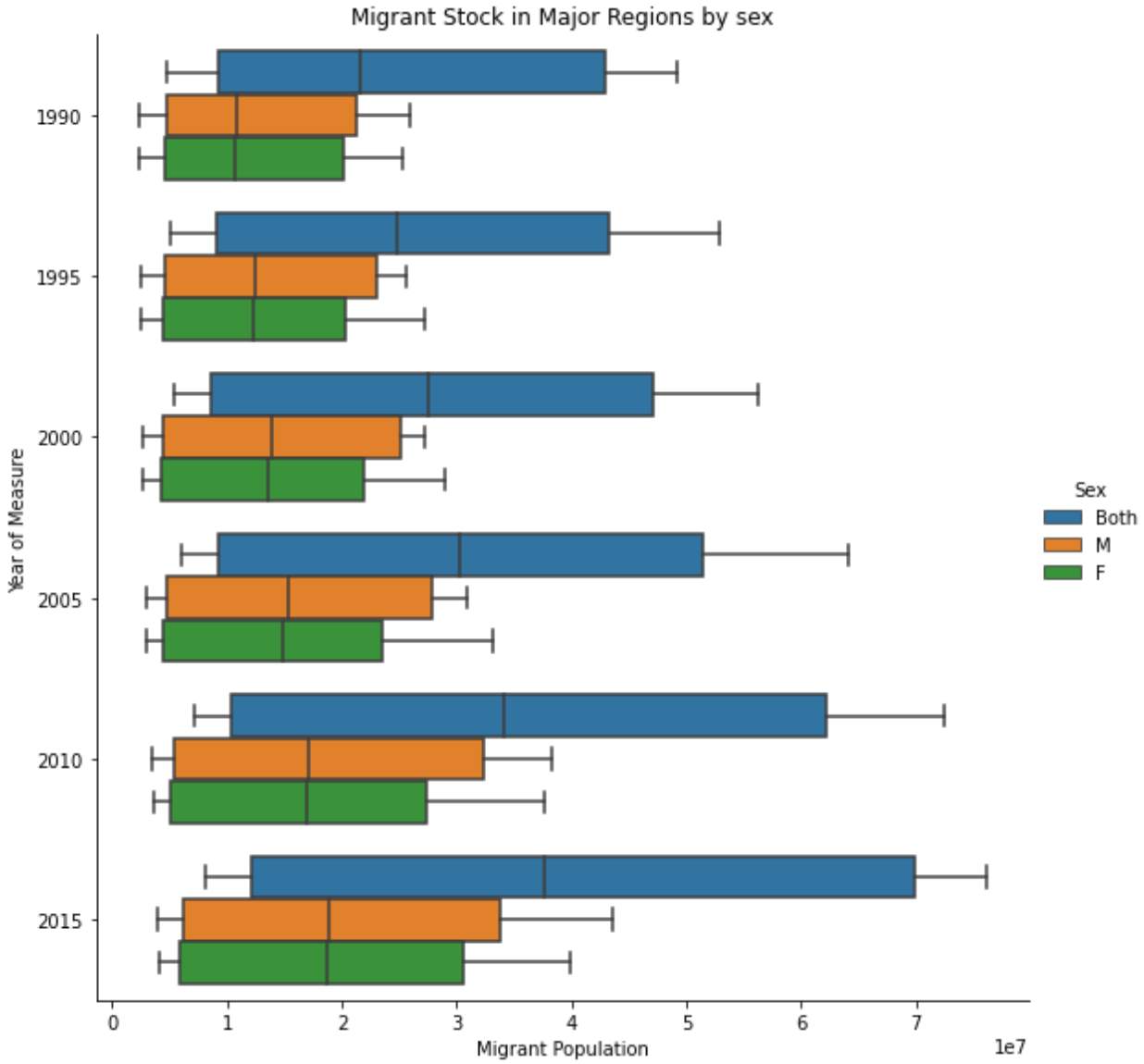

Figure 4: Boxplot of demographic trends in each sex for Major Regions

The same approach was then applied to the Total population portion of the dataset which produced the following graphs with essentially the same code.

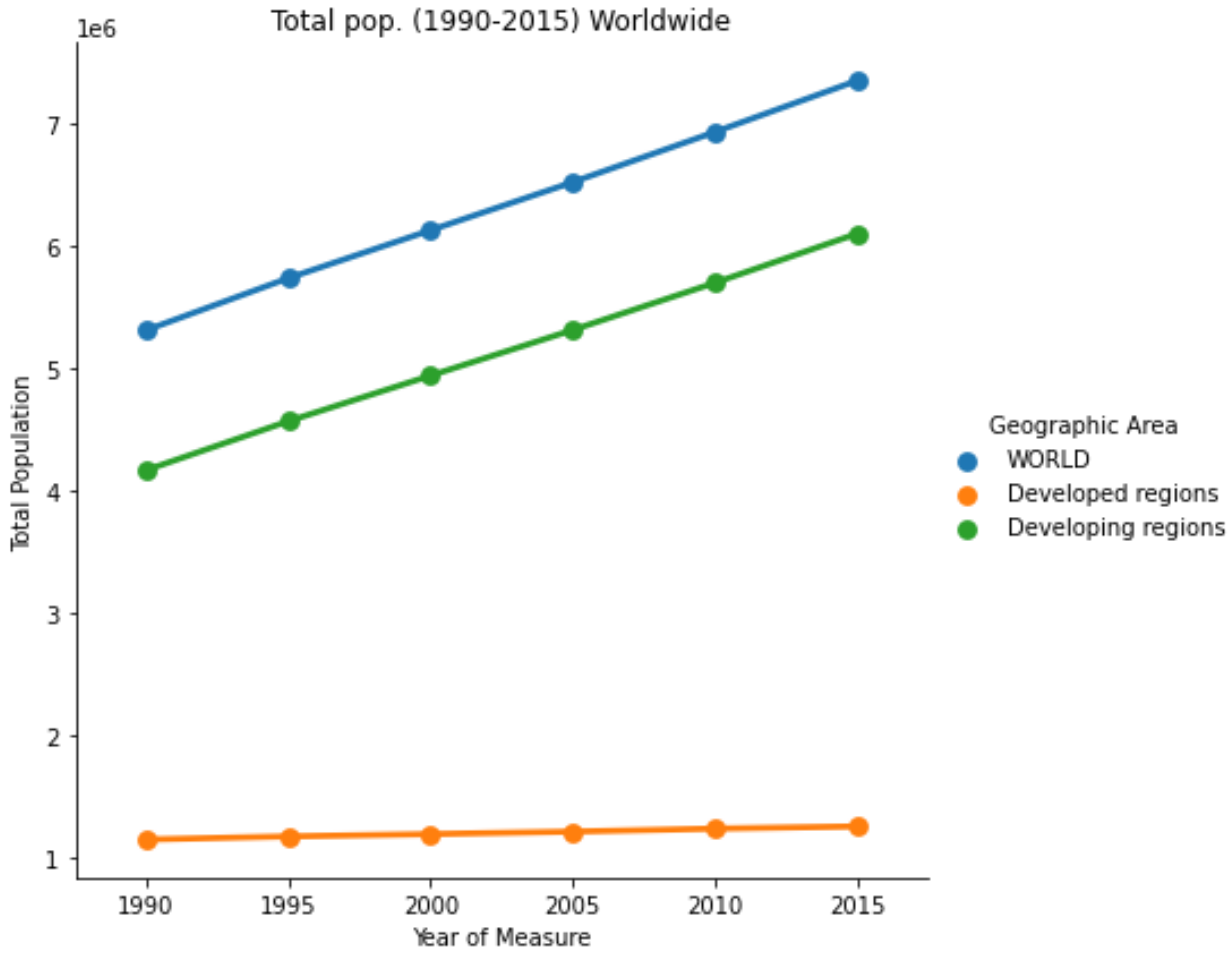

Figure 5: Line graph of trends in total population worldwide

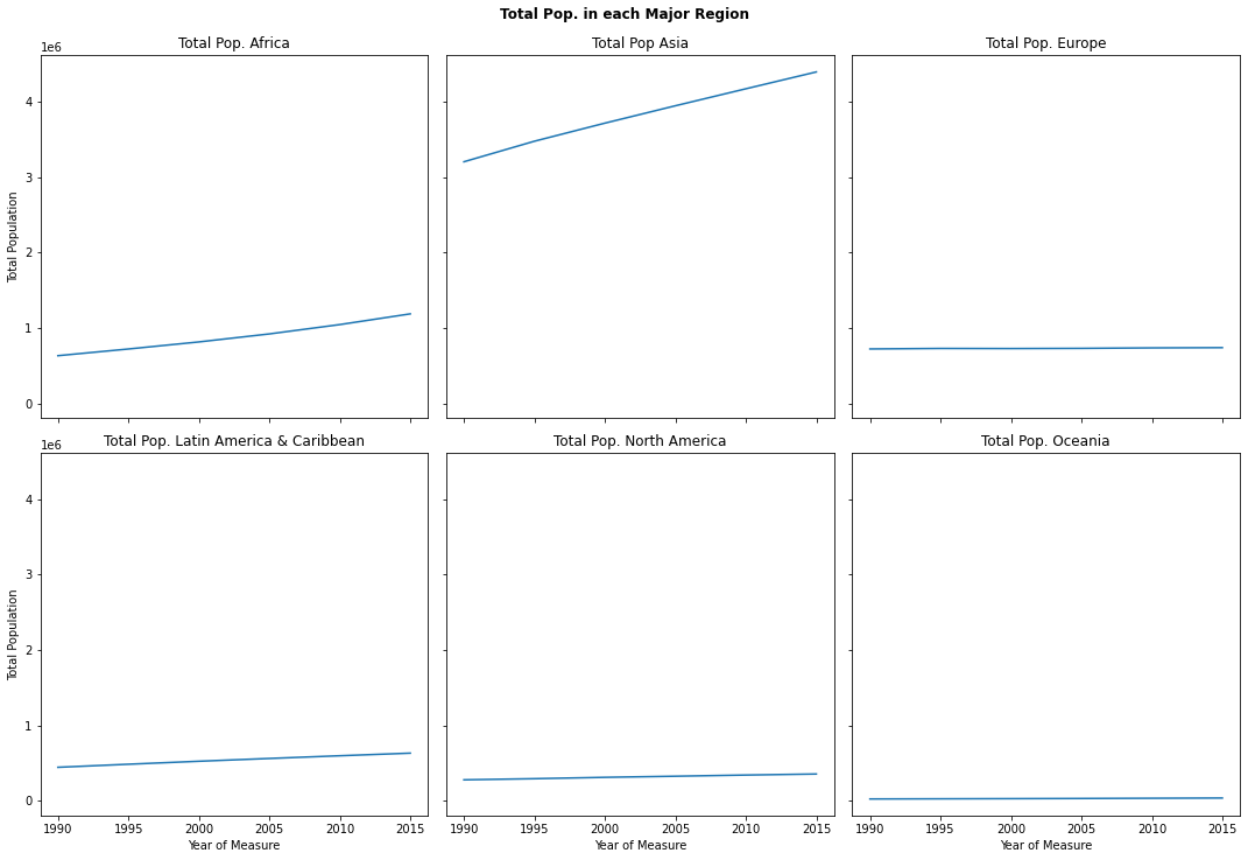

Figure 6: small multiples of total population trends over Major Regions

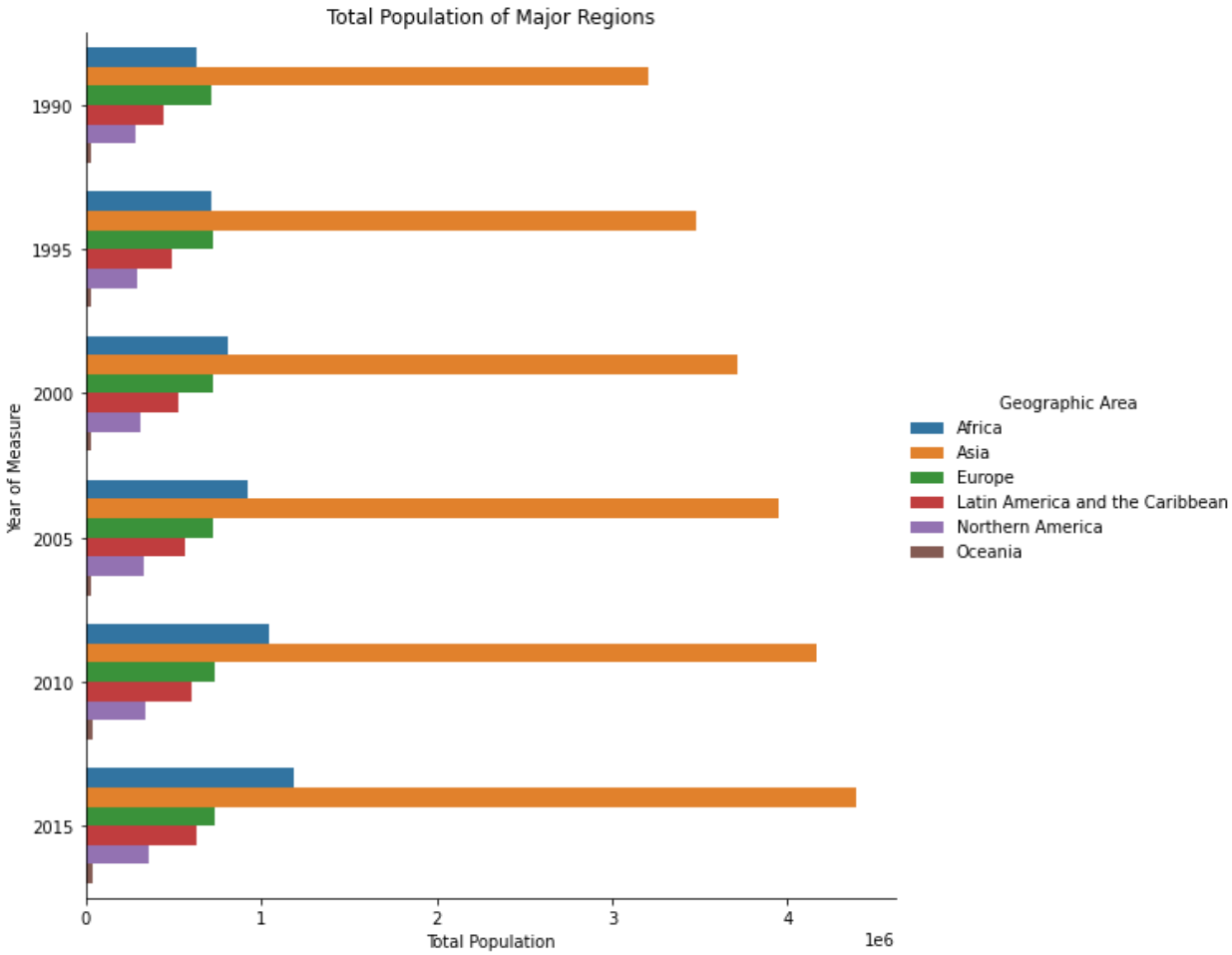

Figure 7: Bar Graph of Total population trends over Major Regions

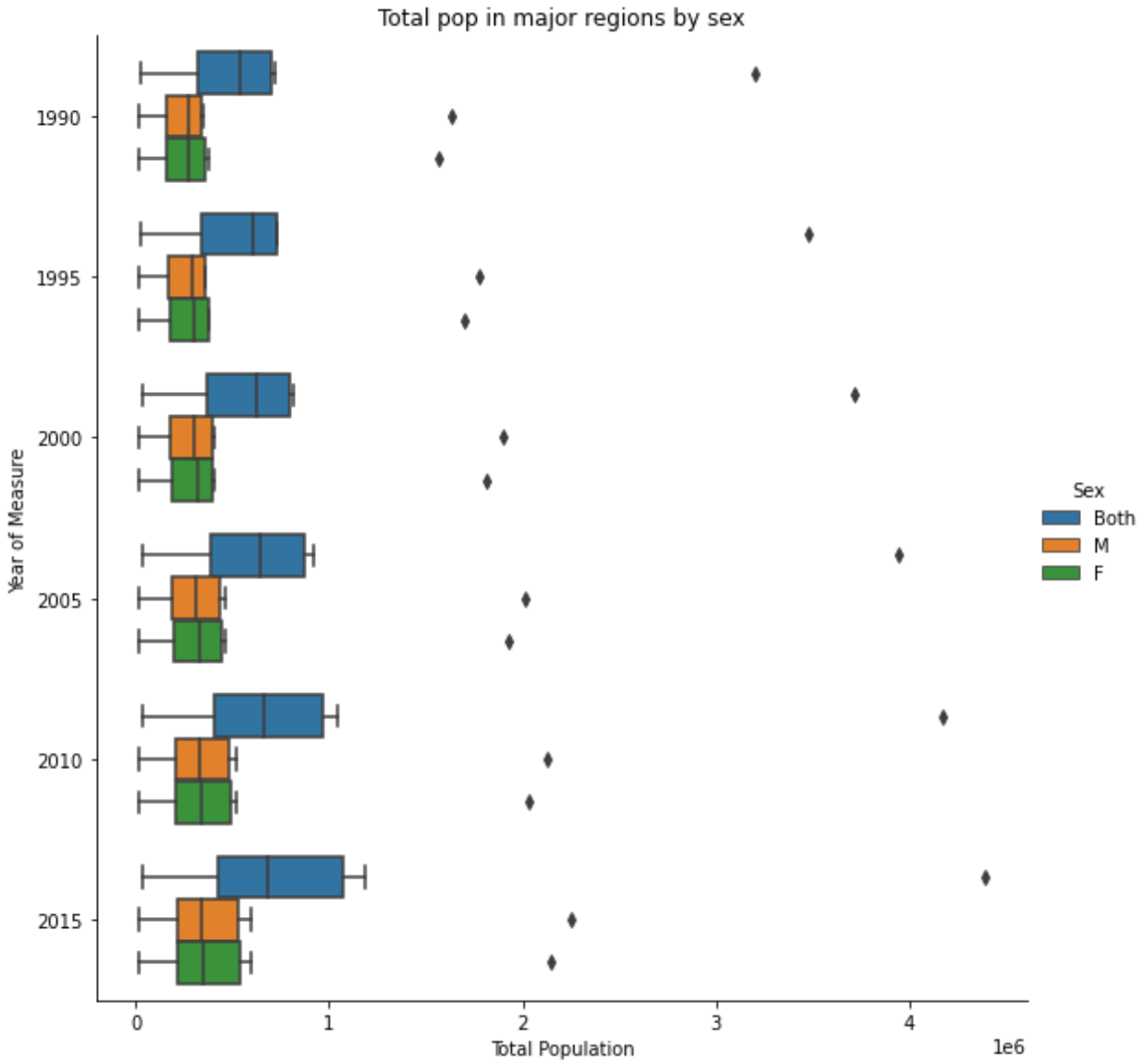

Figure 8: Trends in total population over Major Regions by Sex

The last two relationships graphed were the percentage of migrants of the total population of the same areas as above, and the percentage of refugees of the migrants in the same areas as above. Once again the respective relevant data points were subset in accordance with Tukey's principles and then graphed through the following code. As well for all plots, the axis titles were simply taken from the column names and the plot titles were a description of the relationship on display in the respective plot.

```
#Convert Percentage of Migrants of Total Populations into numeric
```

```
UN3['Percentage of Migrants of Total Population']=pd.to_numeric(UN3['Perce
ntage of Migrants of Total Population'], errors='coerce')
#make a dataframe of same values for areas and sexes as we did for migrant
 population

UN3Rpop=UN3.loc[UN3['Country Code'].isin([903,935,908,904,905,909])]
UN3RPop=UN3Rpop.loc[UN3['Sex'].isin(['Both'])]
#plot 9 bar graph of % of migrants of total pop
sns.catplot(data=UN3RPop, kind='bar', x="Percentage of Migrants of Total P
opulation",y="Year of Measure", hue = "Geographic Area",orient="h",height=
8).set(title="% of Migrants of Total Population of Major Regions")

plt.show()
```

This code produced the following plot:

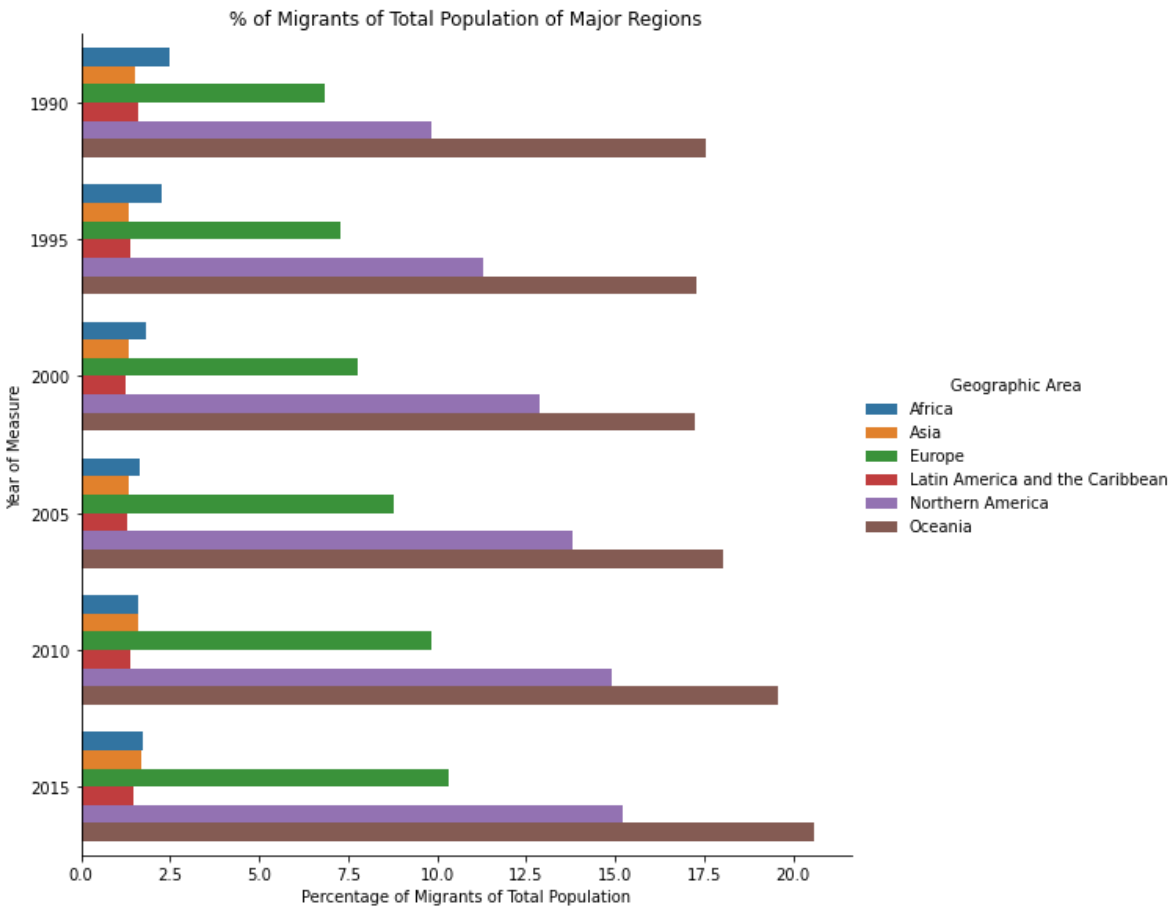

Figure 9: Percent of migrants of the total population in Major Regions

```
#Convert Refugee stock as Percentage of Total Migrants into numeric
UN6t2['Refugee stock as Percentage of Total Migrants']=pd.to_numeric(UN6t2
['Refugee stock as Percentage of Total Migrants'], errors='coerce')
#make a dataframe of same values for areas and sexes as we did for migrant
 population
UN6t2RPop=UN6t2.loc[UN6t2['Country Code'].isin([903,935,908,904,905,909])]
#plot 10, bar graph of refugees as % of Total Migrants
sns.catplot(data=UN6t2RPop,kind='bar' ,x="Refugee stock as Percentage of T
otal Migrants",y="Year of Measure", hue = "Geographic Area",orient="h",hei
ght=8).set(title="% of Refugees of Migrant stock of Major Regions")

plt.show()
```

The above code created this plot:

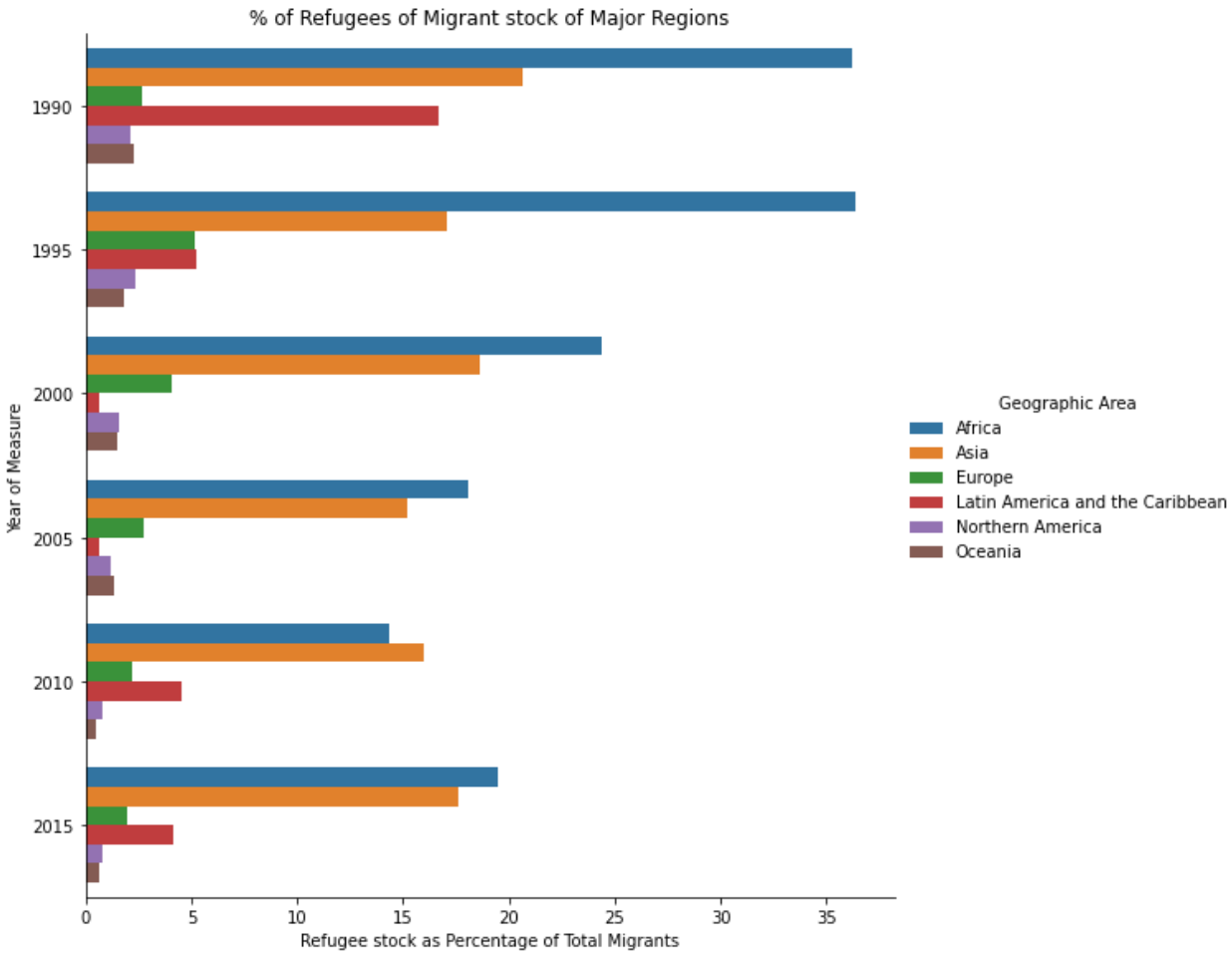

Figure 10: The percent of Refugees of Migrants in Major Regions

Results and Discussion

To first look at Figures 1 and 5 we can see that generally there has been an overall increase in the number of migrants over time for all regions and an increase in population for the world and developing regions but not for the developed regions over the time period of the study. Next, look at Figures 2 and 3 as well as Figures 6 and 7 which graph the same relationship for migrants and total population respectively. Figures 2 and 3 show that the majority of migrants go to Europe, Northern America, and Asia respectively which could indicate that countries in these regions or the regions, in general, are seen as desirable places to emigrate to. Figures 6 and 7 show that the greatest domestic population growth is in Asia by the greatest margin followed by Africa and Europe. Figures 4 and 8 show that both for the migrant population and total domestic population the growth per sex is about even in the major regions. Most crucial however are Figures 9 and 10, the former shows that the major regions with the greatest number of migrants comprising their population are Northern America, Oceania, and Europe over the entire time period of study. The latter Figure however shows the largest portion of refugees in the migrant population are in Africa, Asia, and Europe with the Caribbean heavily decreasing in refugees over time. These plots especially indicate which regions, in particular, could use assistance in dealing with large numbers of refugees as well as what countries could supplement their populations further with migrants as well as just general population trends over the past 25 years.

Sources

"eda.ipynb" lectures

"Tukey and Tufte's Principals" lectures

Pandas documentation

Python basic functions documentation

Seaborn documentation

Matplotlib documentation