Jiayuan Guo
1340 Final Project
Dec 14, 2022

Introduction

According to the data cleaning results of the last international migration trend report, explore the migration trend, population and the proportion of migrant population in the total population in all regions of the country, as well as the relationship between the three and the regional development level.

Method/Result

The final result table after data cleaning is shown as follows.

	Region	Code	Developed region	year	migrant_stock	total population	percentage
0	Southern Asia	5501	No	1990	19436343	1189260506	1.634322
1	Southern Asia	5501	No	1995	15343019	1319395960	1.162882
2	Southern Asia	5501	No	2000	15278020	1451932761	1.052254
3	Southern Asia	5501	No	2005	13722011	1581124102	0.867864
4	Southern Asia	5501	No	2010	14326591	1702990822	0.841261
5	Southern Asia	5501	No	2015	14103682	1822974074	0.773663
6	Southern Europe	925	Yes	1990	4340454	143403793	3.026736
7	Southern Europe	925	Yes	1995	5985675	144146548	4.152493
8	Southern Europe	925	Yes	2000	7517054	145058070	5.1821
9	Southern Europe	925	Yes	2005	11974334	149735395	7.996996
10	Southern Europe	925	Yes	2010	16154120	153360036	10.533461
11	Southern Europe	925	Yes	2015	15747685	152347892	10.336661
12	Northern Africa	912	No	1990	2403200	140116613	1.715143
13	Northern Africa	912	No	1995	2081640	157438162	1.322195
14	Northern Africa	912	No	2000	1885650	171890851	1.097004
15	Northern Africa	912	No	2005	1782054	186917286	0.953392
16	Northern Africa	912	No	2010	1921613	203716733	0.943277
17	Northern Africa	912	No	2015	2159048	223891528	0.964328
18	Polynesia	957	No	1990	58419	547869	10.66295
19	Polynesia	957	No	1995	64801	579483	11.182554

First of all, we rank the migrant population in each region in descending order according to the total migrant population from 1990 to 2015.
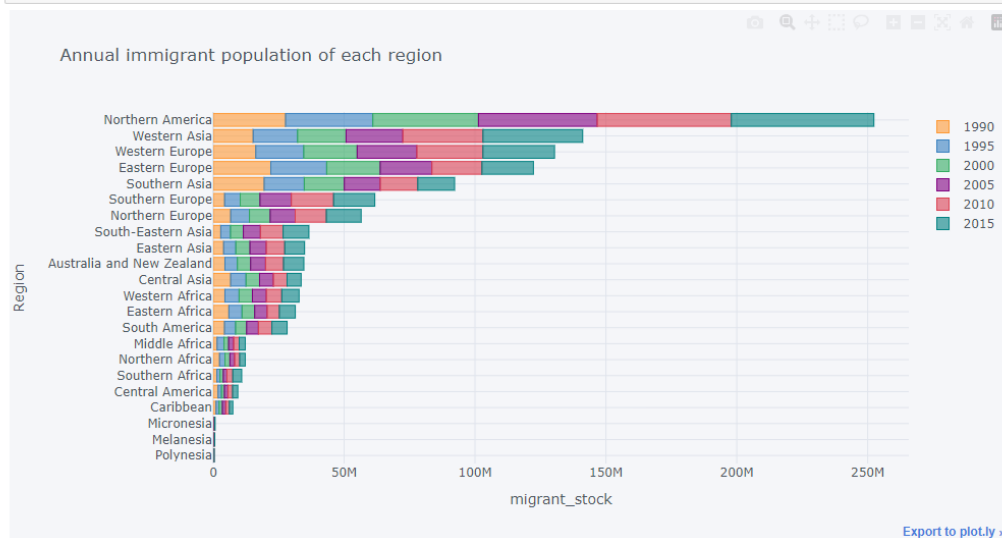
```
In [33]: picture_data1 = region_migrant[['Region','year','migrant_stock']].drop_duplicates().pivot(index='Region',columns='year',values='migrant_stock')
         picture_data1['SUM'] = np.sum(picture_data1,axis=1)
         picture_data1.head(20)
```

Out[33]:

year	1990	1995	2000	2005	2010	2015	SUM
Region							
Australia and New Zealand	4473260	4741947	5065063	5717982	6830423	7803399	34632074.0
Caribbean	1056555	1154700	1255647	1328740	1345732	1367407	7508781.0
Central America	1829911	1298916	1107577	1385713	1749940	2040151	9412208.0
Central Asia	6630683	5890035	5183872	5238699	5262414	5393504	33599207.0
Eastern Africa	5964031	5022742	4844795	4745792	4657063	6129113	31363536.0
Eastern Asia	3959345	4658475	5393081	6229524	7061814	7596693	34898932.0
Eastern Europe	21995312	21344124	20428480	19747392	19116929	19684357	122316594.0
Melanesia	90210	96828	94078	103878	105684	109595	600273.0
Micronesia	108969	118687	130186	128390	116245	115772	718249.0
Middle Africa	1460530	2646108	1756687	1928828	2139979	2307688	12239820.0
Northern Africa	2403200	2081640	1885650	1782054	1921613	2159048	12233205.0
Northern America	27610542	33341147	40351848	45363387	51220996	54488725	252376645.0
Northern Europe	6645605	7194824	7900973	9588814	11858391	13331442	56520049.0
Polynesia	58419	64801	70700	73171	73012	72120	412223.0
South America	4283262	4241024	4215204	4518645	5143123	5826431	28227689.0
South-Eastern Asia	2876616	3700057	4926833	6522343	8661853	9867722	36555424.0
Southern Africa	1392359	1191582	1222314	1439426	2203306	3435194	10884181.0
Southern Asia	19436343	15343019	15278020	13722011	14326591	14103682	92209666.0
Southern Europe	4340454	5985675	7517054	11974334	16154120	15747685	61719322.0
Western Africa	4470503	5410742	5090860	5295046	5918053	6618514	32803718.0

Observe the annual migration of each region by using a stacked bar chart. Bar charts allow one to see the size of individual data at a glance and to easily compare the differences between data.

```
In [35]: #Draw a picture in descending order of the number of immigrants
         picture_data1.iplot(kind='barh',barmode='stack',xTitle='migrant_stock', yTitle='Region',title='Annual immigrant population of each region')
```


Secondly, we use the bubble chart to observe the annual population of each region since the bubble chart is used to show the relationship between three variables
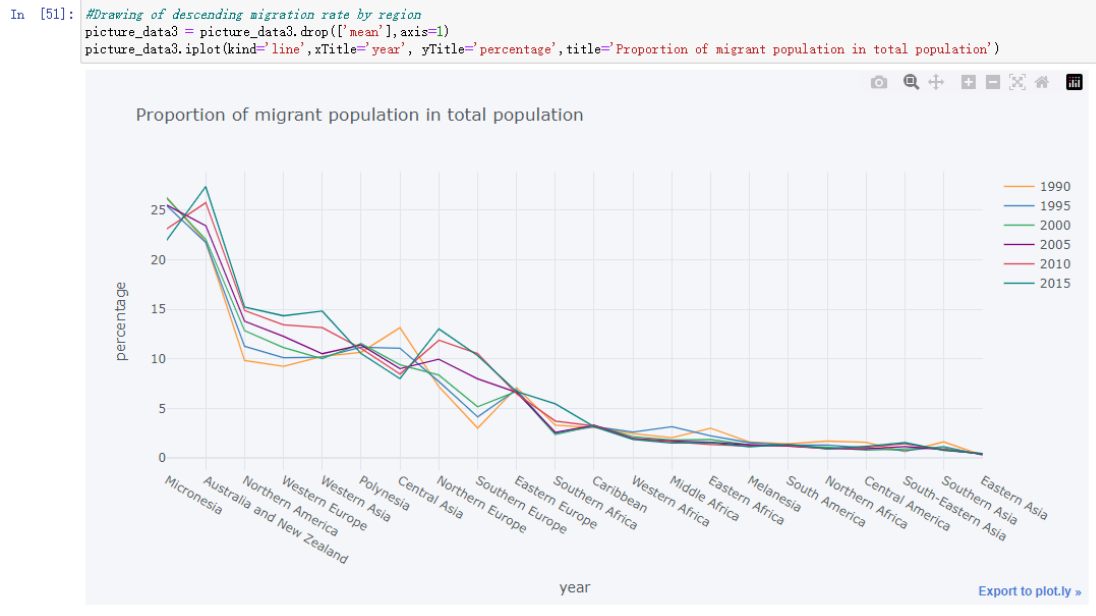
Annual total population of each region
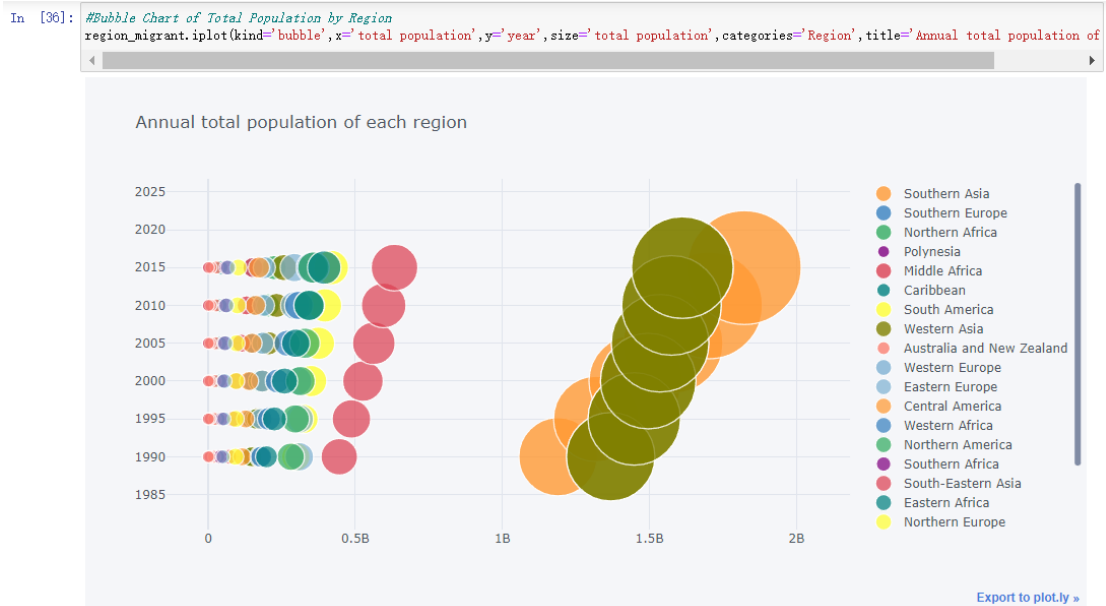
Next, we rank the average proportion of migrant population in the total population from 1990 to 2015 in descending order.

```
In [50]: #Observe the immigration rate of each region
         picture_data3 = region_migrant[['Region','year','percentage']].drop_duplicates().pivot(index='Region',columns='year',values='percentage')
         picture_data3['mean'] = np.mean(picture_data3,axis=1)
         picture_data3 = picture_data3.sort_values('mean',ascending=False)
         picture_data3.head(20)
```

Out[50]:

year	1990	1995	2000	2005	2010	2015	mean
Region							
Micronesia	26.273925	25.436126	26.195684	25.505582	23.133794	21.995501	24.756769
Australia and New Zealand	21.82674	21.752394	22.05511	23.42573	25.744201	27.38275	23.697821
Northern America	9.838663	11.275336	12.862208	13.808229	14.884238	15.227203	12.982646
Western Europe	9.246246	10.125542	11.151449	12.274909	13.445415	14.351878	11.765906
Western Asia	10.25855	10.174655	10.034239	10.523799	13.150533	14.819178	11.493492
Polynesia	10.66295	11.182554	11.542951	11.414741	11.060598	10.536774	11.066761
Central Asia	13.154796	11.076107	9.405144	9.025555	8.468826	8.012451	9.857147
Northern Europe	7.220322	7.723637	8.369902	9.963591	11.8962	13.024358	9.699668
Southern Europe	3.026736	4.152493	5.1821	7.996996	10.533461	10.336661	6.871408
Eastern Europe	7.094647	6.894779	6.724573	6.638181	6.489308	6.719523	6.760169
Southern Africa	3.311276	2.515236	2.37567	2.604143	3.730193	5.484577	3.336849
Caribbean	3.089559	3.17444	3.277269	3.31951	3.233308	3.165345	3.209905
Western Africa	2.47906	2.621241	2.164158	1.9721	1.921556	1.873744	2.171977
Middle Africa	2.06038	3.15482	1.827726	1.723508	1.638597	1.518698	1.987288
Eastern Africa	3.008616	2.229263	1.86789	1.594493	1.358764	1.55373	1.935459
Melanesia	1.636152	1.559705	1.345381	1.329028	1.212488	1.138906	1.370277
South America	1.43797	1.309287	1.205048	1.205586	1.29522	1.392393	1.307584
Northern Africa	1.715143	1.322195	1.097004	0.953392	0.943207	0.964328	1.165890
Central America	1.593687	1.022507	0.798078	0.93008	1.086129	1.181053	1.101922
South-Eastern Asia	0.645466	0.759951	0.936342	1.158176	1.451607	1.557676	1.084869

Line chart can display continuous data over time and are therefore suitable for showing trends in data over equal time intervals. Then we make a broken line chart for the descending results to observe the proportion of migrants in the total population in each region.

```
In [51]: #Drawing of descending migration rate by region
         picture_data3 = picture_data3.drop(['mean'], axis=1)
         picture_data3.iplot(kind='line', xTitle='year', yTitle='percentage', title='Proportion of migrant population in total population')
```


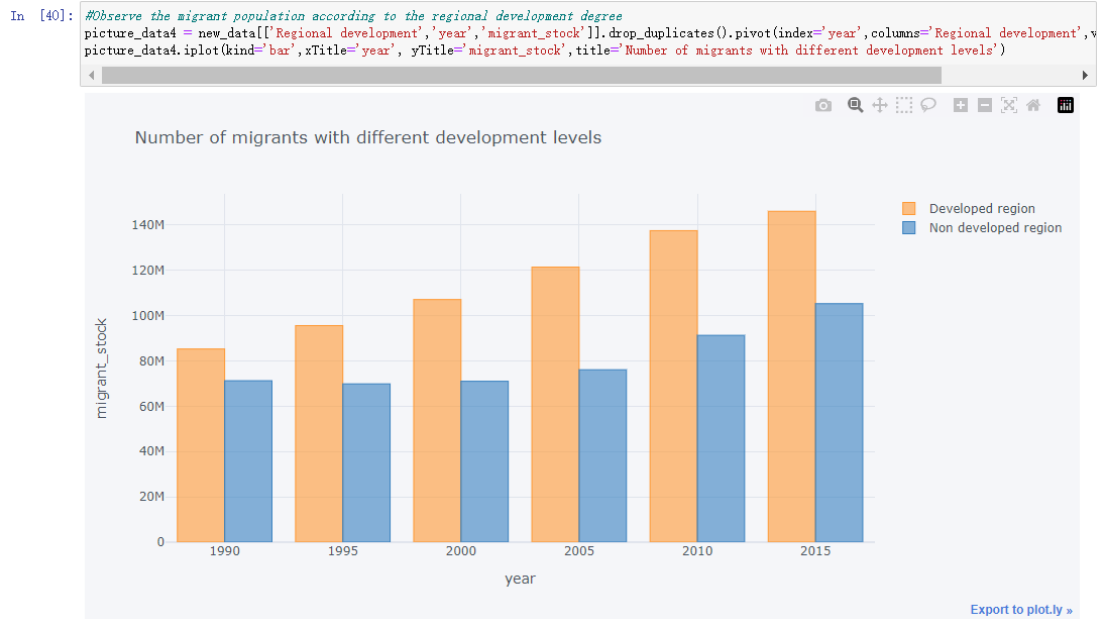Proportion of migrant population in total population

Let's explore whether the number of immigrants is related to the degree of regional development, and calculate the migrant population for each region divided into developed regions and underdeveloped regions.

```
In [39]: #The relationship between migrant population and regional development
         region_migrant['Regional development'] = region_migrant['Developed region'].apply(lambda x:'Developed region' if x=='Yes' else 'Non developed
         new_data = pd.DataFrame(region_migrant[['Regional development','year','migrant_stock']].groupby(['Regional development','year']).sum('migrant_s
         new_data.head(20)
```
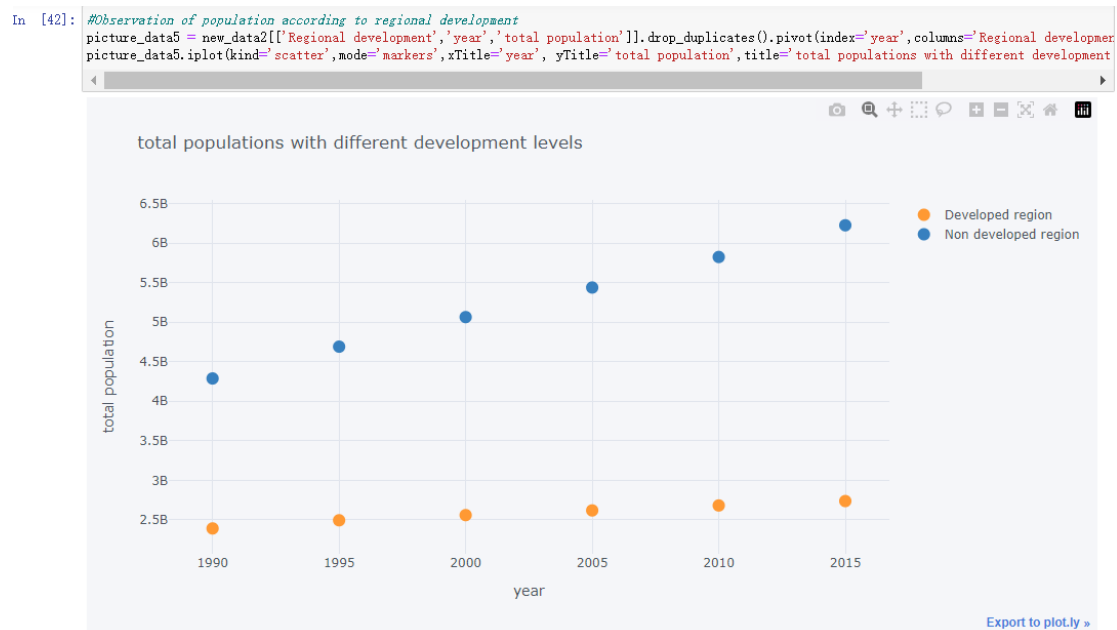
Out[39]:

	Regional development	year	migrant_stock
0	Developed region	1990	85262347
1	Developed region	1995	95584232
2	Developed region	2000	107081877
3	Developed region	2005	121397717
4	Developed region	2010	137487988
5	Developed region	2015	146034771
6	Non developed region	1990	71260210
7	Non developed region	1995	69875995
8	Non developed region	2000	71014513
9	Non developed region	2005	76100907
10	Non developed region	2010	91288069
11	Non developed region	2015	105262158

Use the bar chart to observe the development degree of the region and the number of immigrants.

```
#Observe the migrant population according to the regional development degree
picture_data4 = new_data[['Regional development','year','migrant_stock']].drop_duplicates().pivot(index='year',columns='Regional development',v
picture_data4.iplot(kind='bar',xTitle='year', yTitle='migrant_stock',title='Number of migrants with different development levels')
```


The scatter plot determines whether there is some association between two variables or summarizes the distribution pattern of coordinate points. So we use the scatter chart to observe the population situation in developed and non-developed areas.

```
#Observation of population according to regional development
picture_data5 = new_data2[['Regional development','year','total population']].drop_duplicates().pivot(index='year',columns='Regional developmen
picture_data5.iplot(kind='scatter',mode='markers',xTitle='year', yTitle='total population',title='total populations with different development
```


Lastly, using boxplot and line chart to observe the proportion of migrant population to the total population in developed and non-developed areas from 1990 to 2015.

```
#Observe the proportion of migrant population in the total population according to the regional development degree
picture_data6 = new_data3[['Regional development','year','percentage']].drop_duplicates().pivot(index='year',columns='Regional development',val
picture_data6.iplot(kind='box', yTitle='percentage',title='Proportion of immigrants with different development levels')
```


Proportion of immigrants with different development levels

(Developed region, max: 0.05335452)
(Developed region, q3: 0.05127516)
(Developed region, median: 0.04410043) Developed re...
(Developed region, q1: 0.03832543)
(Developed region, min: 0.0356626)

Developed region
Non developed region

Discussion

To sum up, we can see that in areas with a large population, the number of immigrants is not necessarily large, and in areas with a small population, the number of immigrants is not necessarily small. This is different from our subjective judgment. So what is the main factor affecting immigration?

Conclusion

1、 From the accumulation histogram, we can see that the top five regions of the total immigrant population from 1990 to 2015 were Northern America, Western Asia, Western Europe, Eastern Europe and Southern Asia.

2. From the bubble chart, we can see that the top three populations are Eastern Asia, Southern Asia and South Eastern Asia.

3. From the broken line chart, we can see that Micronesia and Australia and New Zealand always rank first in the proportion of immigrants to the total population.

4. Through the bar chart, we can see that the number of migrants in developed areas has always been greater than that in underdeveloped areas, which indicates that the number of migrants is related to the degree of regional development.

5. Through the scatter chart, we can see that the population of the underdeveloped areas is always more than that of the developed areas.

6. From the box and line diagram, we can see that the proportion of migrants in the total population in developed regions from 1990 to 2015 was higher than that in non developed regions.