

Import Libraries

In [1]:

```
import numpy as np
import pandas as pd
from pandas import DataFrame as df
import matplotlib.pyplot as plt
import seaborn as sns
```

Load Dataset

In [10]:

```
canada = pd.read_excel('https://s3-api.us-gio.objectstorage.softlayer.net/cf-courses-data/CognitiveClass/DV0101EN/labs/Data_Files/Canada.xlsx',
    sheet_name='Canada by Citizenship',
    skiprows=range(20),
    skipfooter=2)
```

In [12]:

```
canada.head()
#spare df
canada1 = canada
```

Rename columns

In [13]:

```
canada.rename(columns={'OdName': 'Country', 'AreaName': 'Continent', 'RegName': 'Continent-Region'}, inplace=True)
canada.drop(['AREA', 'REG', 'DEV', 'Type', 'Coverage'], inplace=True, axis='columns')
canada.isnull().sum().sum()
```

Out[13]:

0

In [14]:

```
canada.head(5)
```

Out[14]:

	Country	Continent	Continent-Region	DevName	1980	1981	1982	1983	1984	1985	...
0	Afghanistan	Asia	Southern Asia	Developing regions	16	39	39	47	71	340	...
1	Albania	Europe	Southern Europe	Developed regions	1	0	0	0	0	0	...
2	Algeria	Africa	Northern Africa	Developing regions	80	67	71	69	63	44	...
3	American Samoa	Oceania	Polynesia	Developing regions	0	1	0	0	0	0	...
4	Andorra	Europe	Southern Europe	Developed regions	0	0	0	0	0	0	...

5 rows × 38 columns

In [15]:

```
canada.index.values
```

Out[15]:

```
array([ 0,  1,  2,  3,  4,  5,  6,  7,  8,  9, 10, 11,
12,
      13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24,
25,
      26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37,
38,
      39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50,
51,
      52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63,
64,
      65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76,
77,
      78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89,
90,
      91, 92, 93, 94, 95, 96, 97, 98, 99, 100, 101, 102,
103,
      104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115,
116,
      117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128,
129,
      130, 131, 132, 133, 134, 135, 136, 137, 138, 139, 140, 141,
142,
      143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154,
155,
      156, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167,
168,
      169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180,
181,
      182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193,
194])
```

Reassign Index

In [16]:

```
canada.set_index('Country', inplace=True)
```

In [18]:

```
canada.head()
```

Out[18]:

	Continent	Continent-Region	DevName	1980	1981	1982	1983	1984	1985	1986
Country										
Afghanistan	Asia	Southern Asia	Developing regions	16	39	39	47	71	340	496
Albania	Europe	Southern Europe	Developed regions	1	0	0	0	0	0	1
Algeria	Africa	Northern Africa	Developing regions	80	67	71	69	63	44	69
American Samoa	Oceania	Polynesia	Developing regions	0	1	0	0	0	0	0
Andorra	Europe	Southern Europe	Developed regions	0	0	0	0	0	0	2

5 rows × 37 columns

Convert columns to str

In [154]:

```
canada.dtypes
canada.columns = list(map(str,canada.columns));
```

Subsetting

In [23]:

```
canada[canada['Continent']=='Asia'].head(4)
```

Out[23]:

	Continent	Continent-Region	DevName	1980	1981	1982	1983	1984	1985	1986
Country										
Afghanistan	Asia	Southern Asia	Developing regions	16	39	39	47	71	340	496
Armenia	Asia	Western Asia	Developing regions	0	0	0	0	0	0	0
Azerbaijan	Asia	Western Asia	Developing regions	0	0	0	0	0	0	0
Bahrain	Asia	Western Asia	Developing regions	0	2	1	1	1	3	0

4 rows × 37 columns

Viewing Line chart

In [320]:

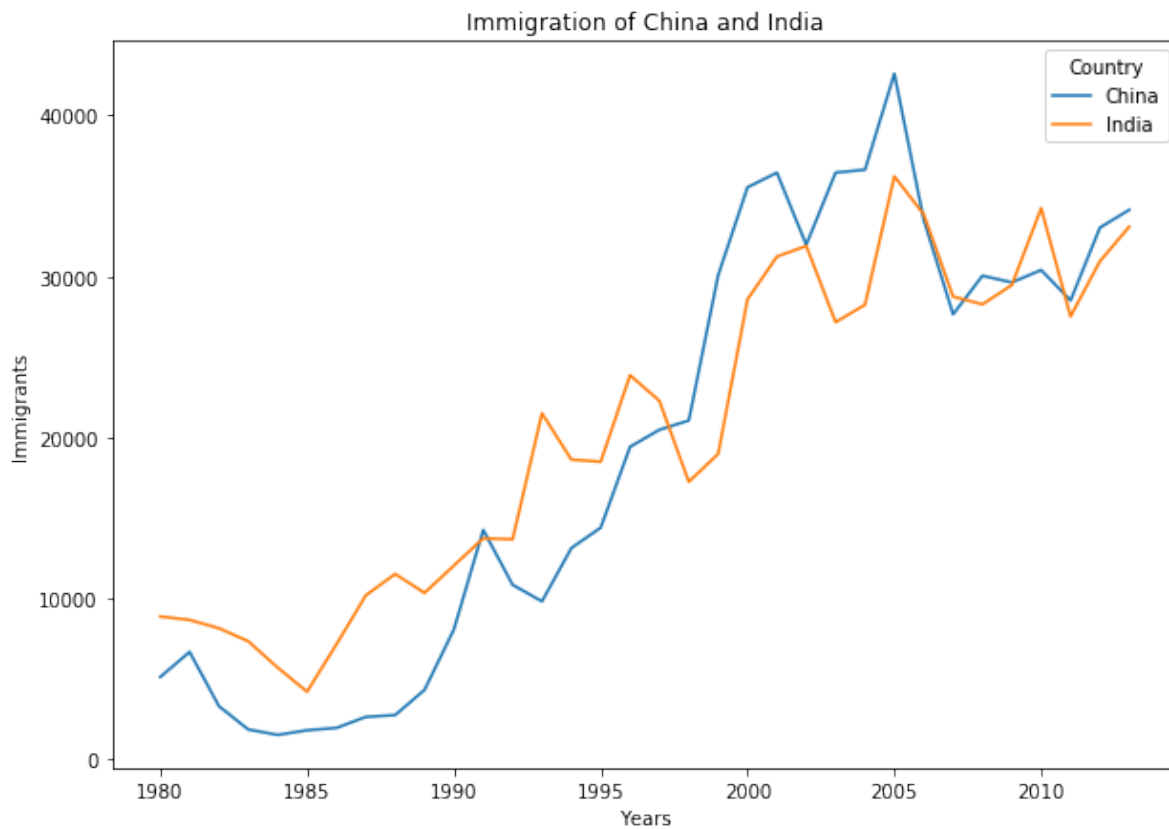
```
years = list(map(str, range(1980, 2014)))
india_imgt = canada.loc['India', years]
india_imgt.plot(figsize=(10, 6));
india_imgt.plot(kind='line');
plt.title('Immigration from India');
plt.ylabel('Number of immigrants');
plt.xlabel('Years');
```



Plotting two countries and comparing

In [321]:

```
canada.loc[['China','India'], years].transpose().plot(figsize=(10,7));  
plt.title('Immigration of China and India');  
plt.xlabel('Years');  
plt.ylabel('Immigrants');
```

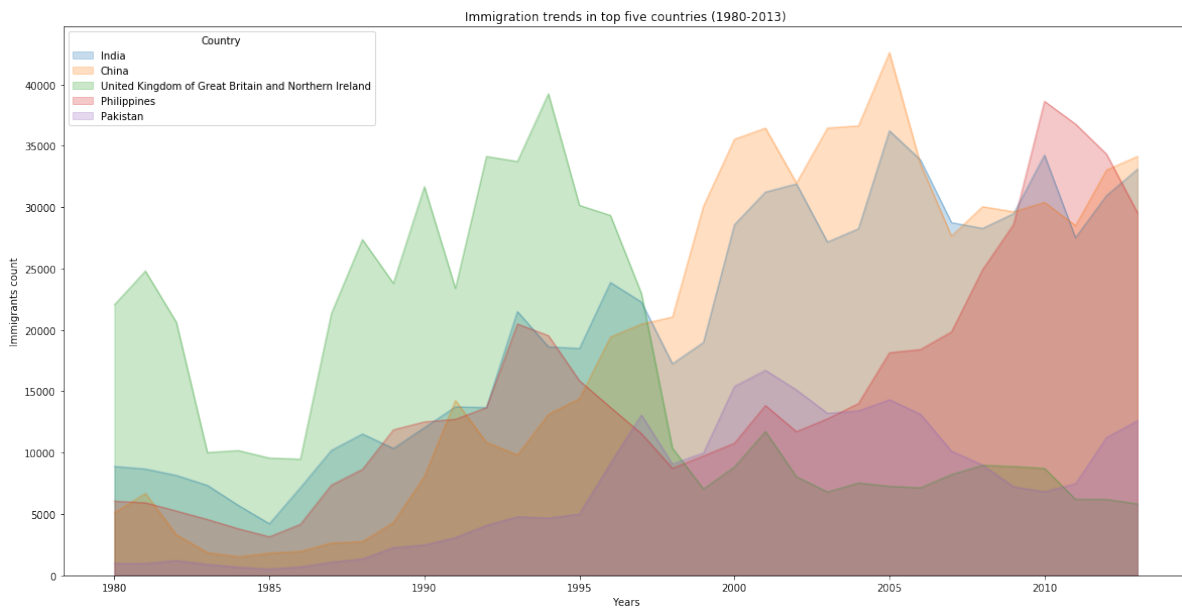


Area Plots

Unstacked

In [456]:

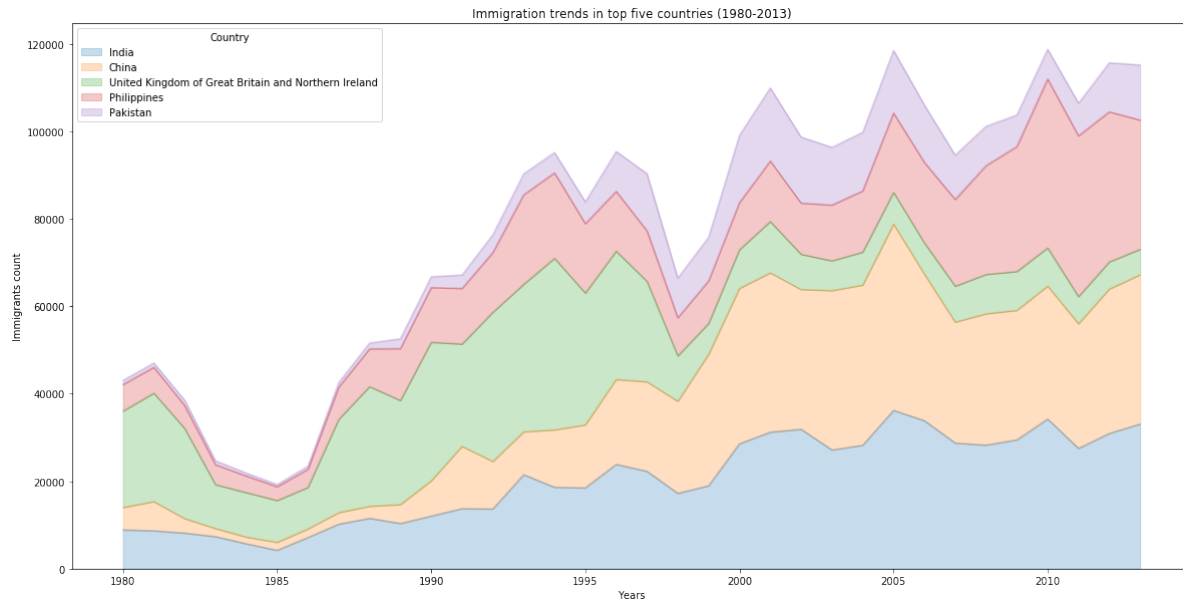
```
top5_clean = canada.head(5)[years].transpose()  
top5_clean.index  
top5_clean.index = top5_clean.index.map(int)  
top5_clean.plot(kind="area", stacked=False, figsize=(20,10), alpha=0.25)  
plt.title('Immigration trends in top five countries (1980-2013)')  
plt.xlabel('Years')  
plt.ylabel('Immigrants count')  
plt.show()
```



Stacked

In [457]:

```
top5_clean.plot(kind="area", stacked=True, figsize=(20,10), alpha=0.25)
plt.title('Immigration trends in top five countries (1980-2013)')
plt.xlabel('Years')
plt.ylabel('Immigrants count')
plt.show()
```



Bar PLots and Histograms

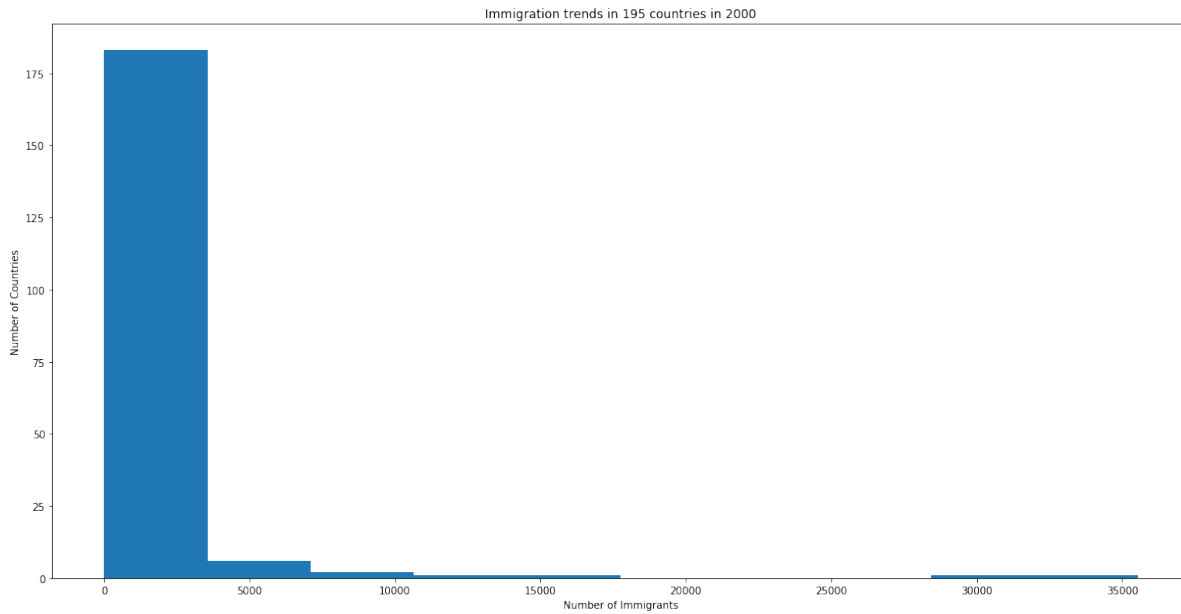
In [459]:

```
df_can = canada
```

Bar

In [460]:

```
df_can['2000'].plot(kind='hist', figsize=(20,10))  
plt.title('Immigration trends in 195 countries in 2000')  
plt.xlabel('Number of Immigrants')  
plt.ylabel('Number of Countries')  
plt.show()
```



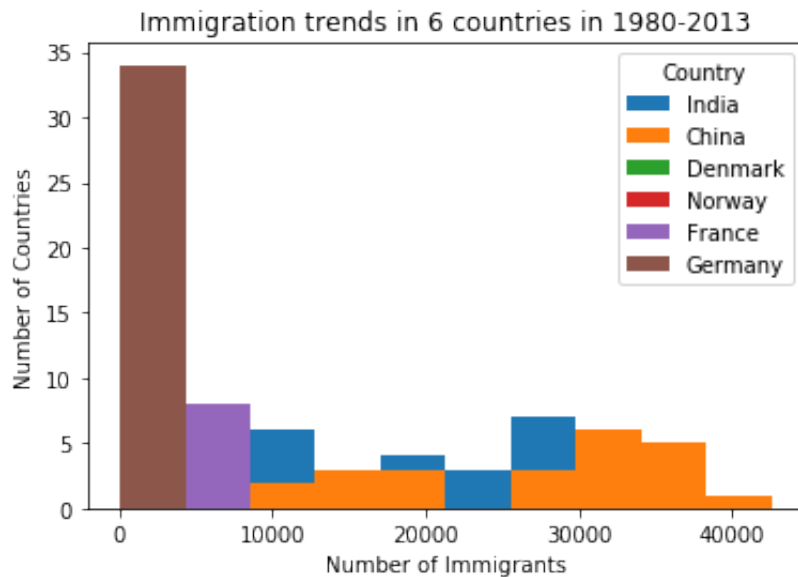
Histogram

In [467]:

```
df_can.loc[['India', 'China', 'Denmark', 'Norway', 'France', 'Germany'], years].transpose().plot.hist()

plt.title('Immigration trends in 6 countries in 1980-2013')
plt.xlabel('Number of Immigrants')
plt.ylabel('Number of Countries')

plt.show()
```



Vertical Bar Plots

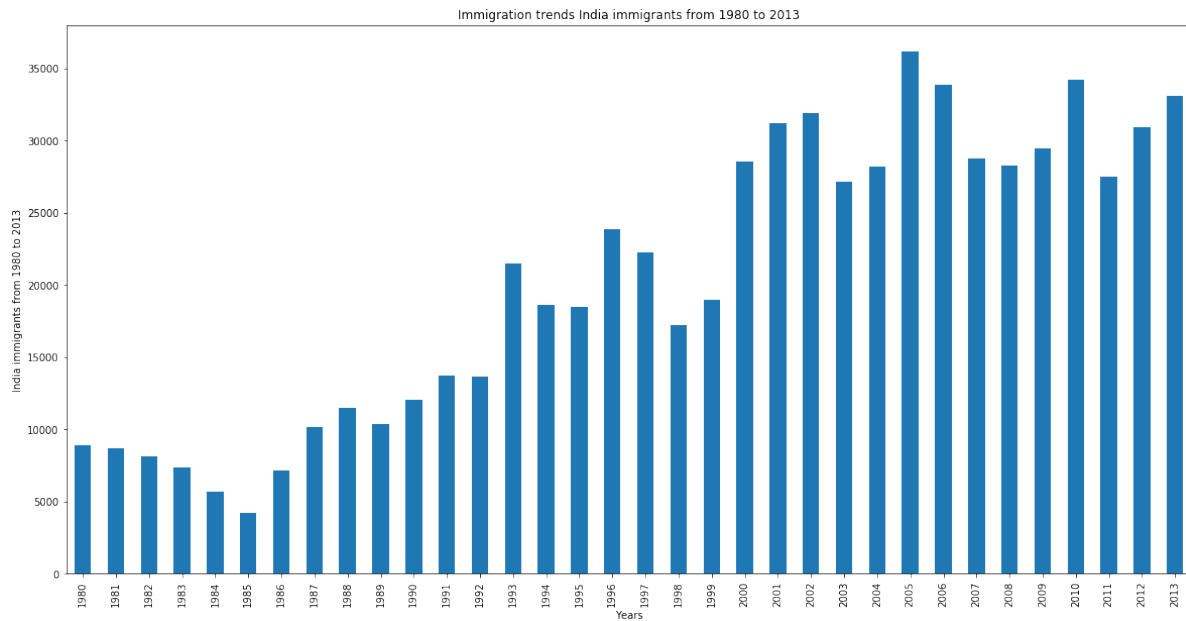
In [464]:

```
india = df_can.loc['India', years]

india.plot(kind='bar', figsize=(20,10))

plt.title('Immigration trends India immigrants from 1980 to 2013')
plt.xlabel('Years')
plt.ylabel('India immigrants from 1980 to 2013')

plt.show()
```



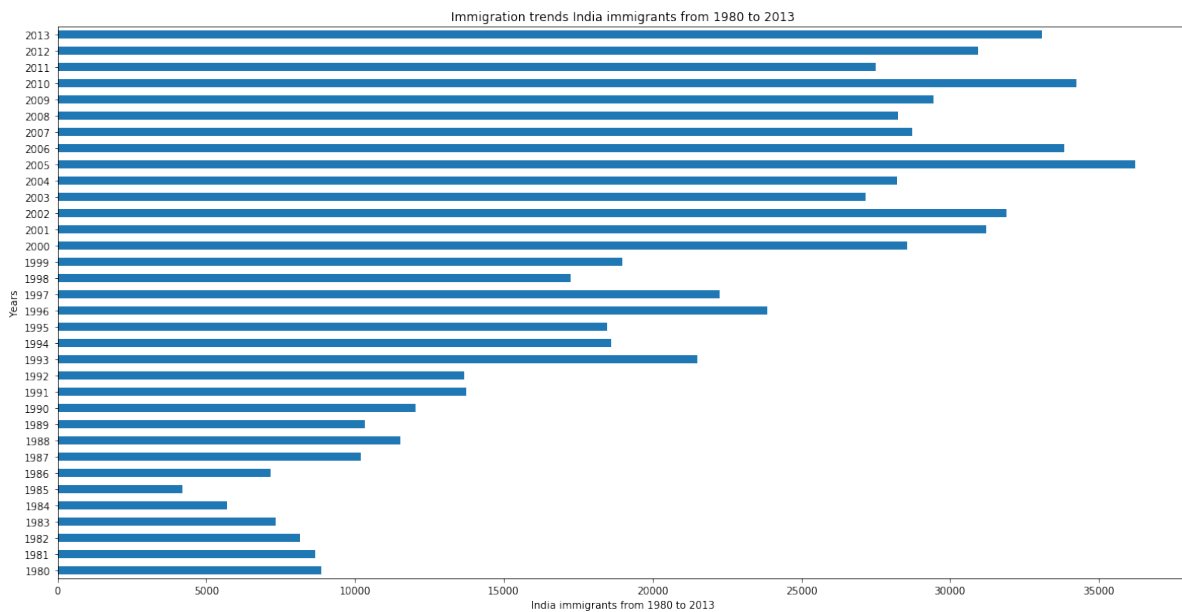
Horizontal Bar Plots

In [465]:

```
india = df_can.loc['India', years]

india.plot(kind='barh', figsize=(20,10))

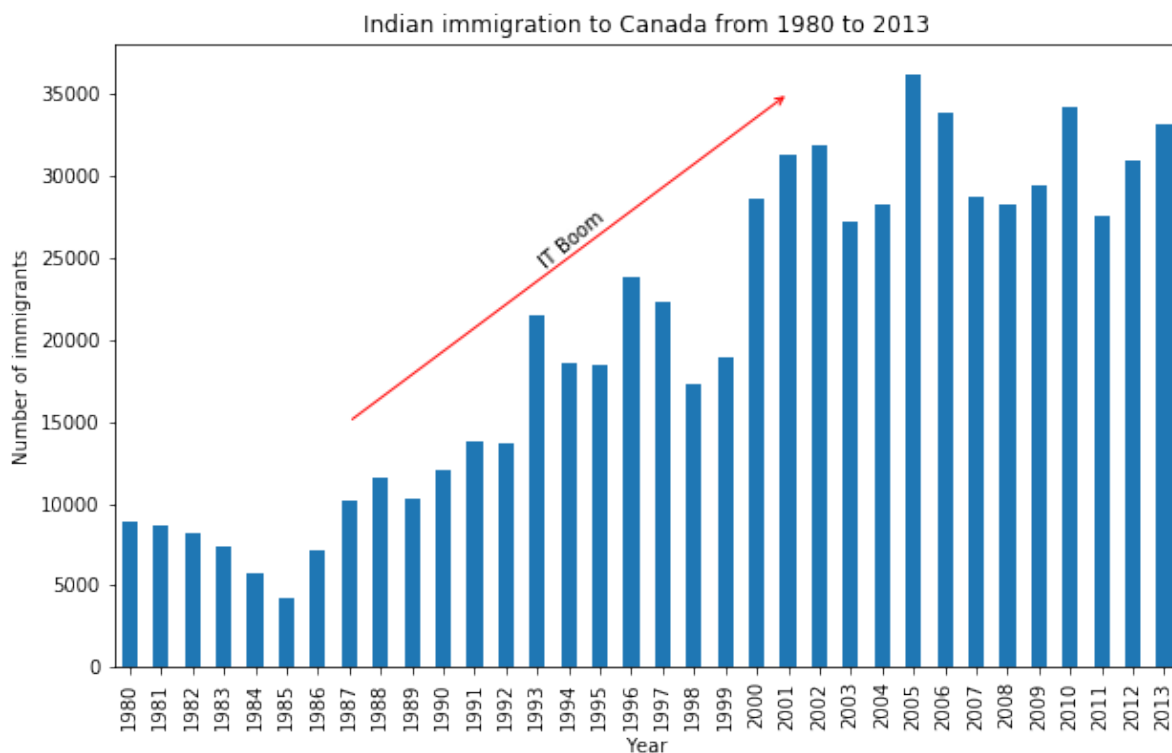
plt.title('Immigration trends India immigrants from 1980 to 2013')
plt.xlabel('India immigrants from 1980 to 2013')
plt.ylabel('Years')
plt.show()
```



Using annotate

In [466]:

```
df_india = df_can.loc['India',years]
df_india.plot(kind='bar', figsize=(10,6))
plt.title("Indian immigration to Canada from 1980 to 2013")
plt.xlabel("Year")
plt.ylabel("Number of immigrants")
plt.annotate('', #arrow title
             xy=(21,35000), #x,y of arrow head
             xytext=(7,15000), #x, y of arrow tail
             xycoords='data', #keep unchanged
             arrowprops=dict(arrowstyle='->',color='red') #arrow style with color
)
plt.annotate('IT Boom', #add text to arrow
             xy=(13,28000), #x, y of text position
             rotation=40, # counter clockwise rotate text by angle
             xycoords='data', #keep unchanged
             va='top', #position text
             ha='left') #position text
plt.show()
```



Assignment

Which two countries have similar immigration trends over the years 1980-2013?

In [322]:

```
canada['total_immigration'] = canada.sum(axis='columns')
canada.sort_values(by='total_immigration', ascending=False, axis='index', inplace=True)
canada.head(10)
```

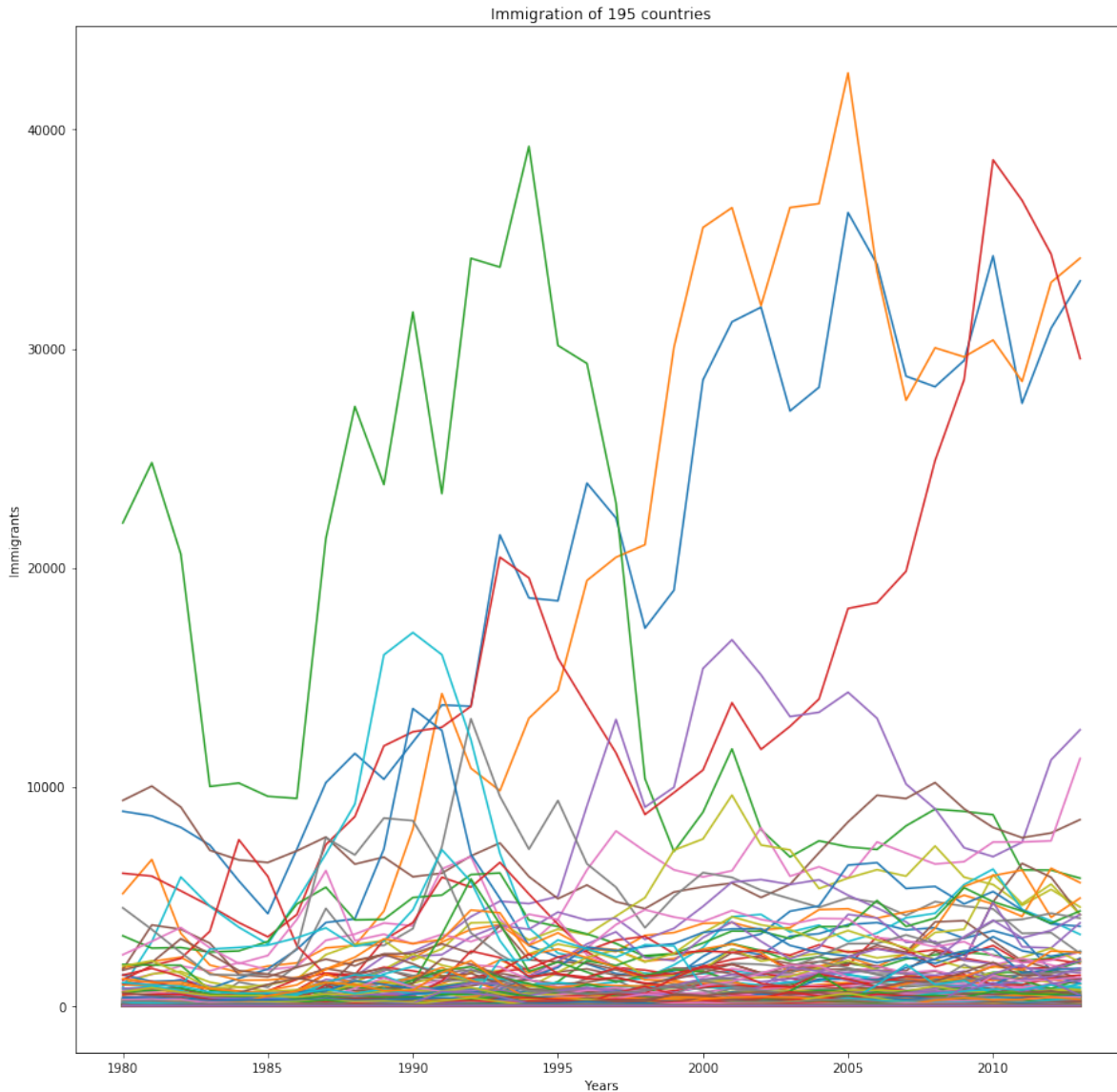
Out[322]:

	Continent	Continent-Region	DevName	1980	1981	1982	1983	1984	1985	
Country										
India	Asia	Southern Asia	Developing regions	8880	8670	8147	7338	5704	4211	7
China	Asia	Eastern Asia	Developing regions	5123	6682	3308	1863	1527	1816	1
United Kingdom of Great Britain and Northern Ireland	Europe	Northern Europe	Developed regions	22045	24796	20620	10015	10170	9564	9
Philippines	Asia	South-Eastern Asia	Developing regions	6051	5921	5249	4562	3801	3150	4
Pakistan	Asia	Southern Asia	Developing regions	978	972	1201	900	668	514	
United States of America	Northern America	Northern America	Developed regions	9378	10030	9074	7100	6661	6543	7
Iran (Islamic Republic of)	Asia	Southern Asia	Developing regions	1172	1429	1822	1592	1977	1648	1
Sri Lanka	Asia	Southern Asia	Developing regions	185	371	290	197	1086	845	1
Republic of Korea	Asia	Eastern Asia	Developing regions	1011	1456	1572	1081	847	962	1
Poland	Europe	Eastern Europe	Developed regions	863	2930	5881	4546	3588	2819	4

10 rows × 39 columns

In [445]:

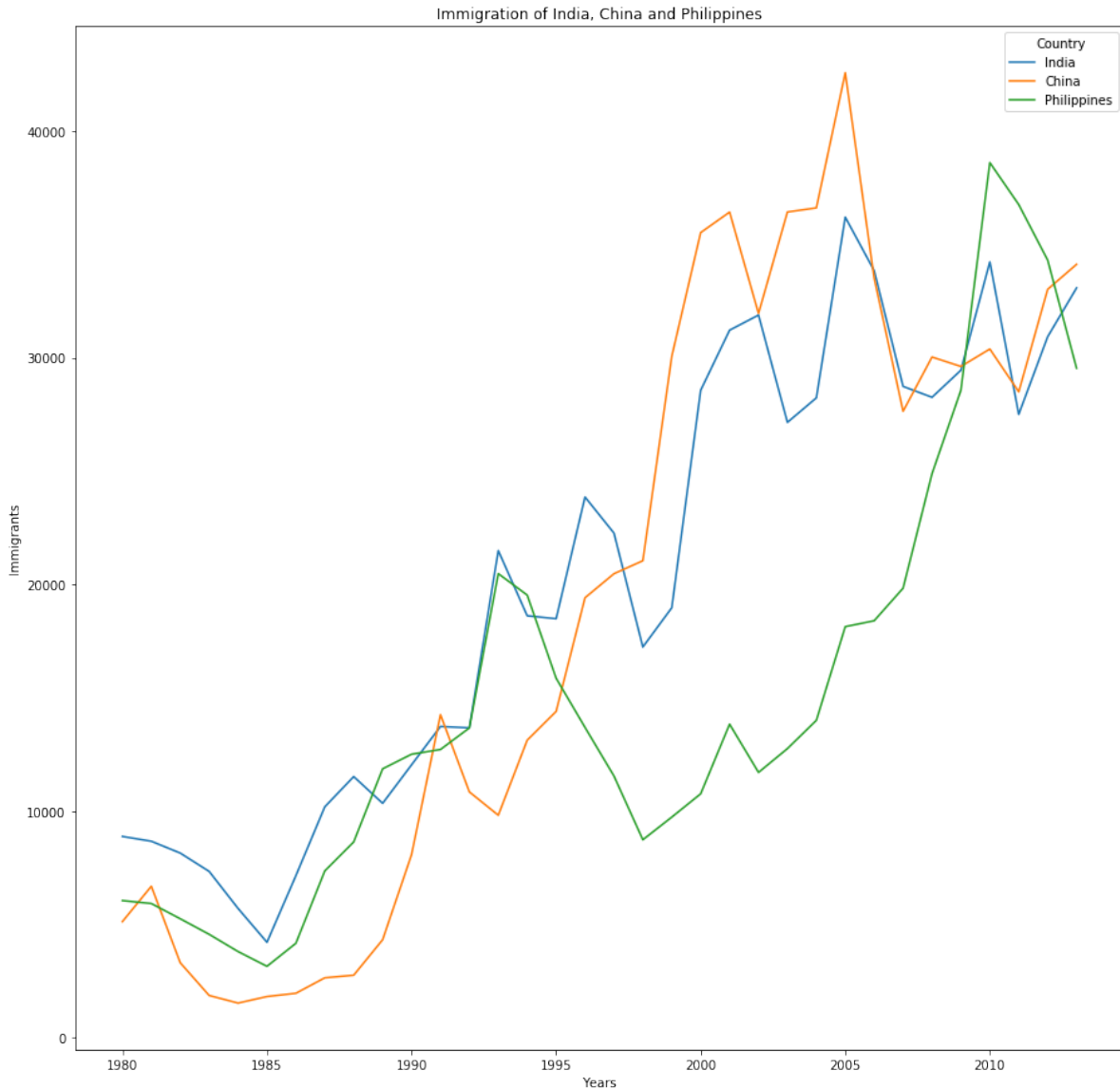
```
canada.loc[canada.index.tolist(), years].transpose().plot(figsize=(15,15),legend=False);
plt.title('Immigration of 195 countries');
plt.xlabel('Years');
plt.ylabel('Immigrants');
```



Plotting three similar countries after year 2009 in the top 10 list.

In [323]:

```
canada.loc[['India', 'China', 'Philippines'], years].transpose().plot(figsize=(15, 15));  
plt.title('Immigration of India, China and Philippines');  
plt.xlabel('Years');  
plt.ylabel('Immigrants');
```



Since 2009, India China and Philippines have shown similar trends.

As it is impossible to compare 195 countries and their charts, Last 3 year's average is taken and compared in the following steps. The selection of n years is random.

Taking avg. of last 3 years and cleaning the Series

In [324]:

```
canada['avglast3'] = (canada['2011']+canada['2012']+canada['2013'])/3
l3 = canada['avglast3']
l3 = l3.sort_values(ascending = False )
l3 = l3.round()
l3 = l3[l3>2]
```

In [325]:

```
l3.head(10)
```

Out[325]:

Country	
Philippines	33541.0
China	31885.0
India	30510.0
Pakistan	10433.0
Iran (Islamic Republic of)	8768.0
United States of America	8023.0
United Kingdom of Great Britain and Northern Ireland	6075.0
Haiti	5508.0
France	5328.0
Iraq	5052.0

Name: avglast3, dtype: float64

Removing outliers

In [326]:

```
l5 = l3[l3<11000];
```

Calculating Standard Deviation for bins

In [327]:

```
np.std(l5)
```

Out[327]:

```
1629.5934270734047
```

Creating bins

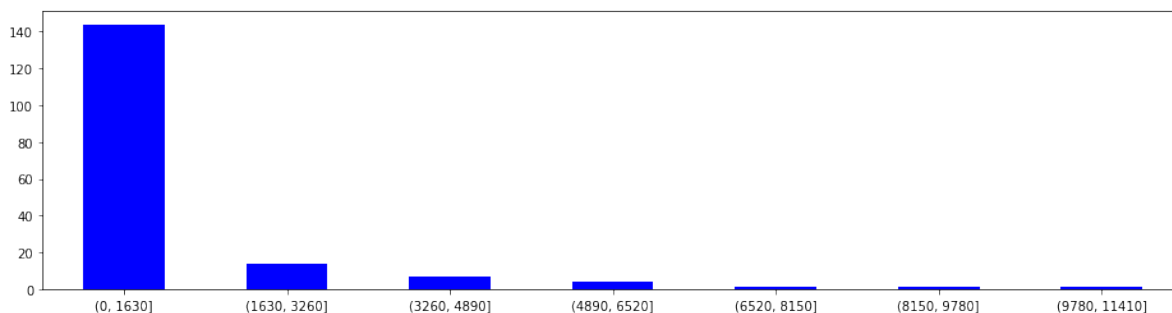
In [328]:

```
type(l5)
dl5 = df(l5)
dl5
ahg = pd.Series(range(0,12000,1630))
bins = ahg
dl5['binned'] = pd.cut(dl5['avglast3'], bins)
```

Plotting the bar graph

In [329]:

```
out = dl5['binned']
ax = out.value_counts(sort=False).plot.bar(rot=0, color="b", figsize=(16,4))
plt.show()
```



Bin values

In [330]:

```
dl5['binned'].value_counts(sort=True)
```

Out[330]:

```
(0, 1630]          144
(1630, 3260]       14
(3260, 4890]        7
(4890, 6520]        4
(9780, 11410]        1
(8150, 9780]         1
(6520, 8150]         1
Name: binned, dtype: int64
```

Countries showing similar trends (2011-2013), grouped

In [331]:

```
print('(9780, 11410]')
print(' {ho} {jo}'.format(ho = 15.index[0], jo = 15[0]))
print('')
print('')
print('(8150, 9780]')
print(' {ho} {jo}'.format(ho = 15.index[1], jo = 15[1]))
print('')
print('')
print('(6520, 8150]')
print(' {ho} {jo}'.format(ho = 15.index[2], jo = 15[2]))
print('')
print('')
print('(4890, 6520]')
print(' {jo}'.format(jo = 15[3:7]))
print('')
print('')
print('(3260, 4890]')
print(' {jo}'.format(jo = 15[7:14]))
print('')
print('')
print('(1630, 3260]')
print(' {jo}'.format(jo = 15[14:28]))
print('')
print('')
print('(0, 1630]')
print(' {jo}'.format(jo = 15[28:172]))
```

```
(9780, 11410]
Pakistan 10433.0
```

```
(8150, 9780]
Iran (Islamic Republic of) 8768.0
```

```
(6520, 8150]
United States of America 8023.0
```

```
(4890, 6520]
Country
United Kingdom of Great Britain and Northern Ireland    6075.0
Haiti                                                    5508.0
France                                                    5328.0
Iraq                                                      5052.0
Name: avglast3, dtype: float64
```

```
(3260, 4890]
Country
Republic of Korea    4804.0
```

```

Egypt          4794.0
Algeria        4143.0
Mexico         4057.0
Colombia       3913.0
Morocco        3846.0
Nigeria        3573.0
Name: avglast3, dtype: float64

```

```
(1630, 3260]
```

```

Country
Bangladesh          3041.0
Sri Lanka           3014.0
Ukraine             2422.0
Lebanon             2286.0
Afghanistan         2281.0
Jamaica             2240.0
Cameroon            2195.0
Russian Federation  2169.0
Israel              2016.0
Ethiopia            1878.0
Viet Nam            1855.0
Somalia             1715.0
Democratic Republic of the Congo 1663.0
Germany             1657.0
Name: avglast3, dtype: float64

```

```
(0, 1630]
```

```

Country
Romania            1625.0
Brazil             1621.0
Tunisia            1524.0
Eritrea            1412.0
Jordan             1365.0
...
Equatorial Guinea    6.0
Mozambique           6.0
Brunei Darussalam    5.0
Cabo Verde           4.0
Tonga                3.0
Name: avglast3, Length: 144, dtype: float64

```

Plots

Comparison on basis of average of immigrants between years 2011 and 2013

Top 3 - excluding India, China and Philippines

In [332]:

```
a1 = 15[0:3].index.tolist()
canada.loc[a1, years].transpose().plot(figsize=(15,15));
plt.title('Immigration of {as1}'.format(as1 = a1));
plt.xlabel('Years');
plt.ylabel('Immigrants');
```



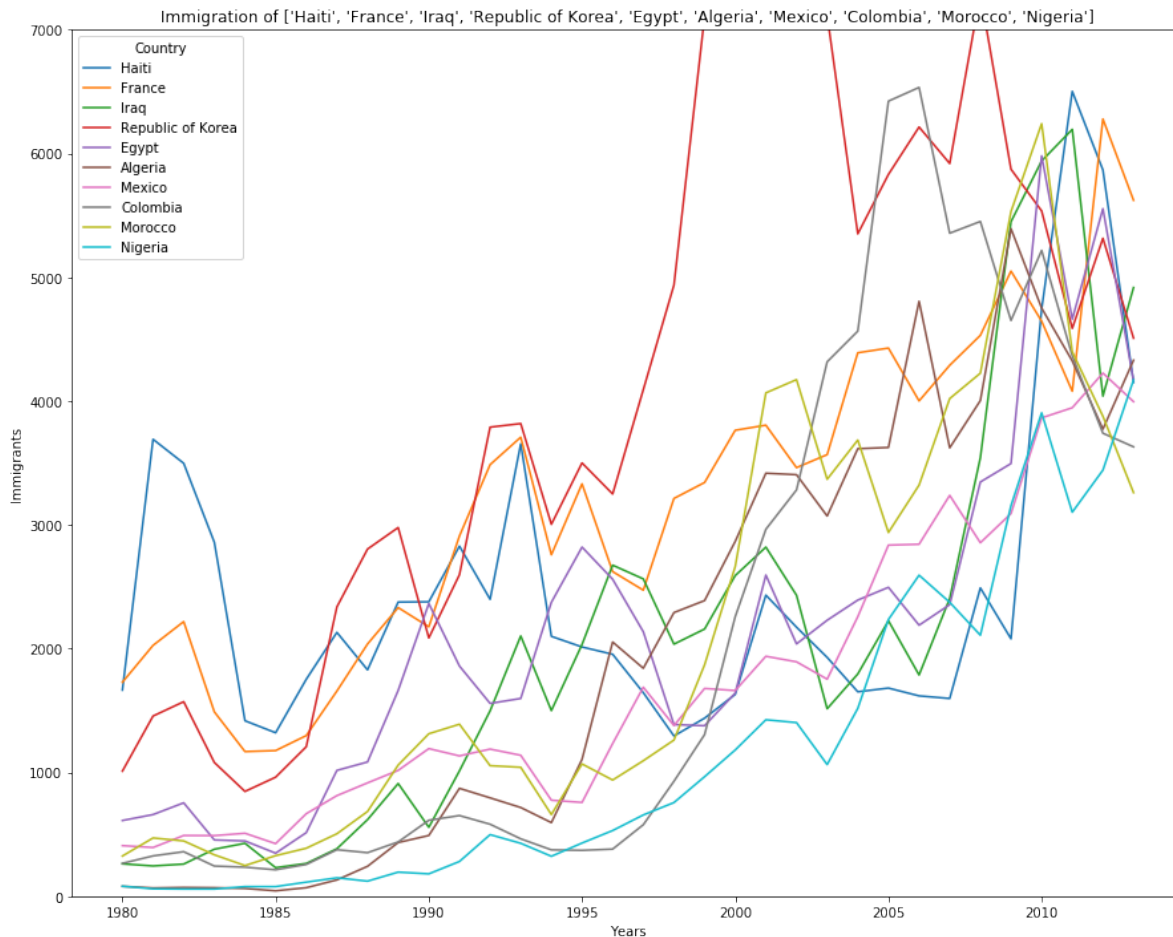
Inference

We can see that Pakistan and Iran show similar trend 2011 onwards, while on the other hand USA and Pakistan show similar trends between the years 2007 and 2011

Countries in Range - (3260, 6520]

In [375]:

```
years = list(map(str, range(1980, 2014)))
a2 = 15[4:14].index.tolist()
canada.loc[a2, years].transpose().plot(figsize=(15, 12));
plt.title('Immigration of {as2}'.format(as2 = a2));
plt.xlabel('Years');
plt.ylabel('Immigrants');
plt.ylim((0, 7000));
```



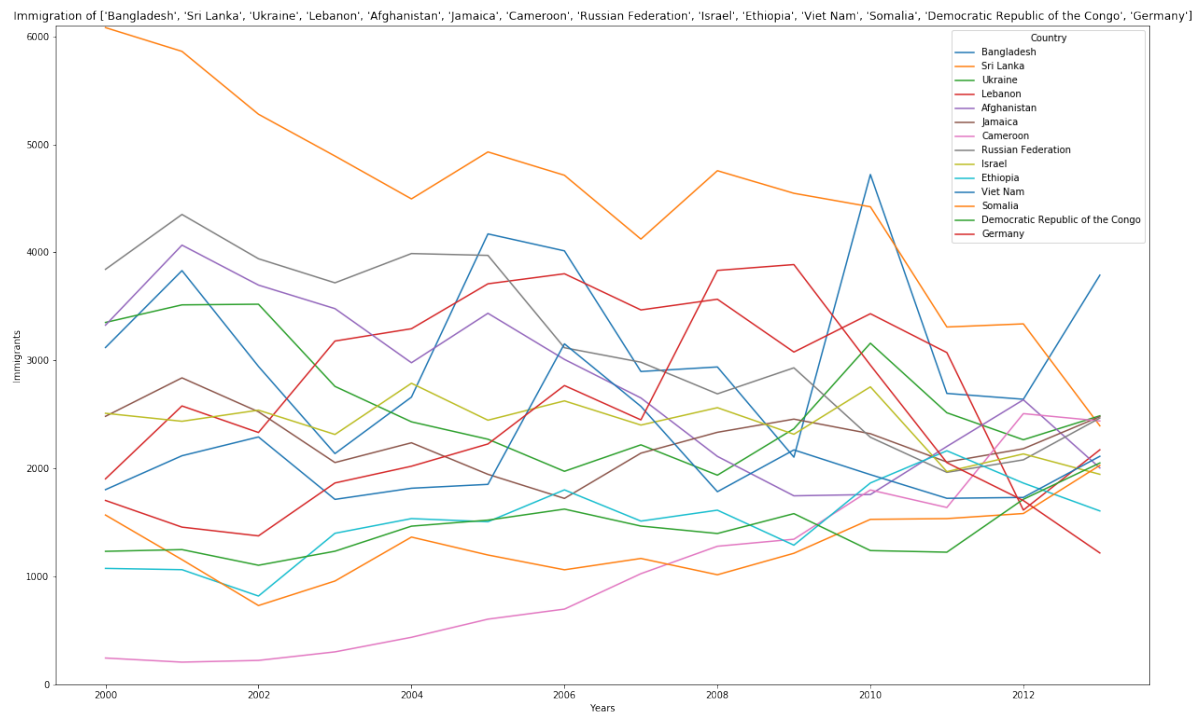
Inference

We can see that all countries except South Korea show almost similar trend throughout from 1980 to 2013.

Countries in Range - (1630, 3260]

In [410]:

```
years1 = list(map(str,range(2000,2014)))
a4 = 15[14:28].index.tolist()
canada.loc[a4, years1].transpose().plot(figsize=(21,13));
plt.title('Immigration of {as4}'.format(as4 = a4));
plt.xlabel('Years');
plt.ylabel('Immigrants');
plt.ylim((0,6100));
```



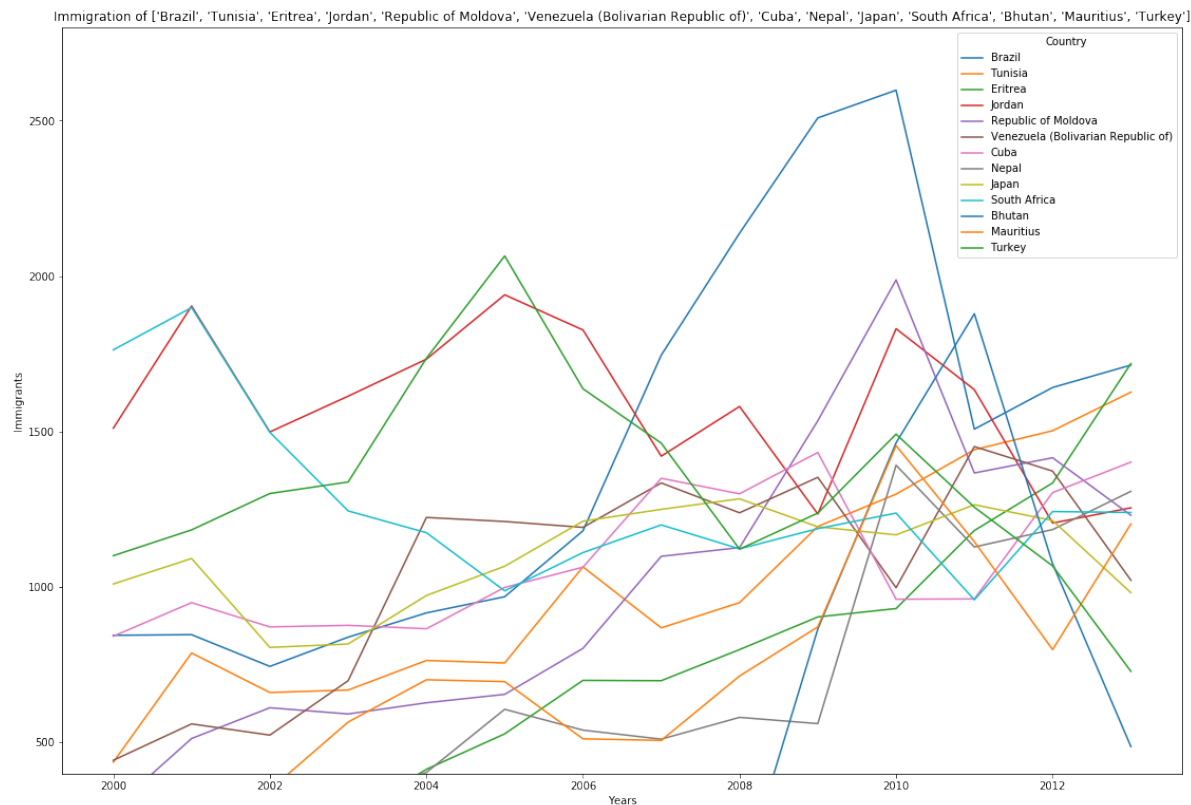
Inference

For ease of comparison, timeline has been reduced to 14 years. Other than Sri Lanka (highest value, 2000) and Cameroon (lowest value, 2000) show similar trends throughout from 2000 to 2013.

Countries in Range - (1000, 1630]

In [429]:

```
a4 = 15[29:42].index.tolist()
canada.loc[a4, years1].transpose().plot(figsize=(19,13));
plt.title('Immigration of {as4}'.format(as4 = a4));
plt.xlabel('Years');
plt.ylabel('Immigrants');
plt.ylim((400,2800));
```



Inference

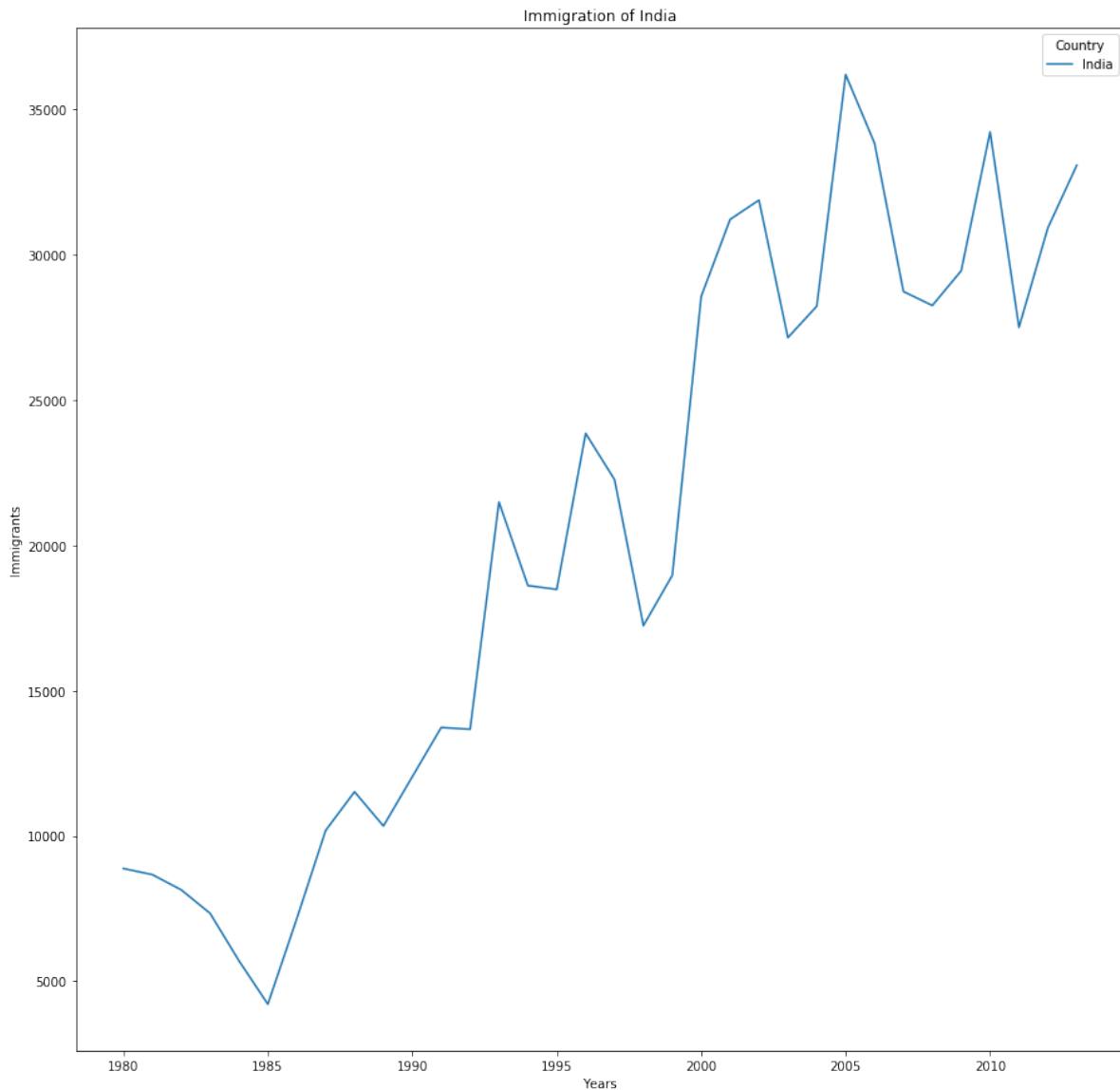
For ease of comparison, timeline has been reduced to 14 years. Except Brazil, which shows a sudden spike after 2006 and Bhutan, which is low throughout and values only spike in year 2011, all other countries are in the 500 - 2000 bracket with inconsistent trends.

Key comparisons and analysis

India

In [451]:

```
canada.loc[['India'], years].transpose().plot(figsize=(15,15));  
plt.title('Immigration of India');  
plt.xlabel('Years');  
plt.ylabel('Immigrants');
```



Inference

Fist spike is observed during FY 1985-1986. Next spike is observed in 1990-1991. This can be attributed to Liberalization. Following that in 1995-1996 india became the fifth largest economy in world with 3.9% share in world GDP. We can see that 1999-2000 there is a spike. This can be attributed to the Y2k bug solution.

India and China

In [416]:

```
canada.loc[['China','India'], years].transpose().plot(figsize=(15,15));  
plt.title('Immigration of China and India');  
plt.xlabel('Years');  
plt.ylabel('Immigrants');
```



Inference

Both countries show a steady growth rate, which starts to plateau after 2001.

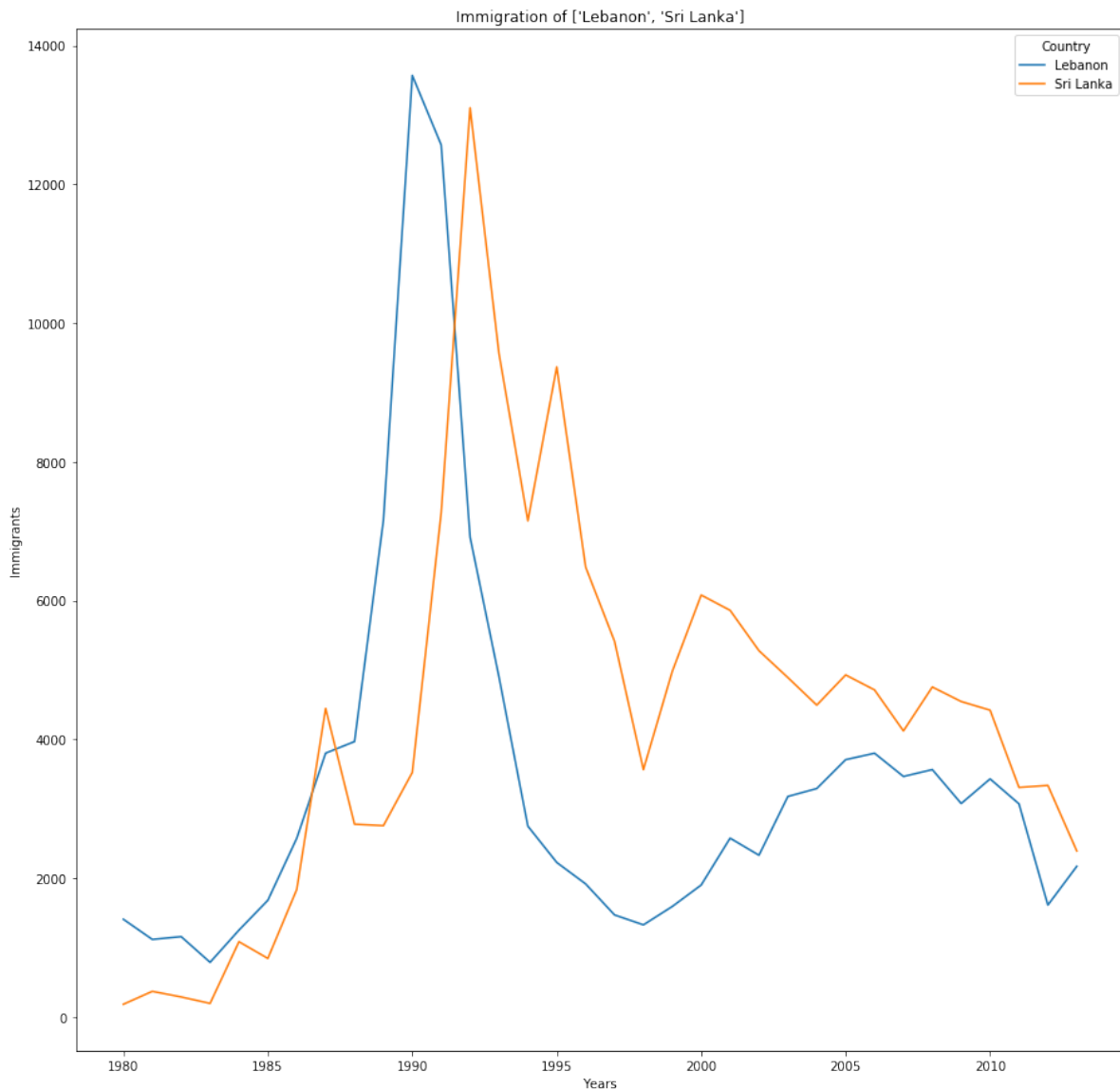
China

China has strengthened its export sector, which has attributed to a decline in the increasing immigration rate.

Lebanon and Sri Lanka

In [447]:

```
a9 = ['Lebanon', 'Sri Lanka']  
canada.loc[['Lebanon', 'Sri Lanka'], years].transpose().plot(figsize=(15,15));  
plt.title('Immigration of {as9}'.format(as9 = a9));  
plt.xlabel('Years');  
plt.ylabel('Immigrants');
```



Inference

Sri Lanka

Increased privatization, economic reform, and a stress on export-oriented growth helped improve the economic performance, increasing GDP growth to 7% in 1993. This attributes to the drop in immigrants from Sri Lanka.

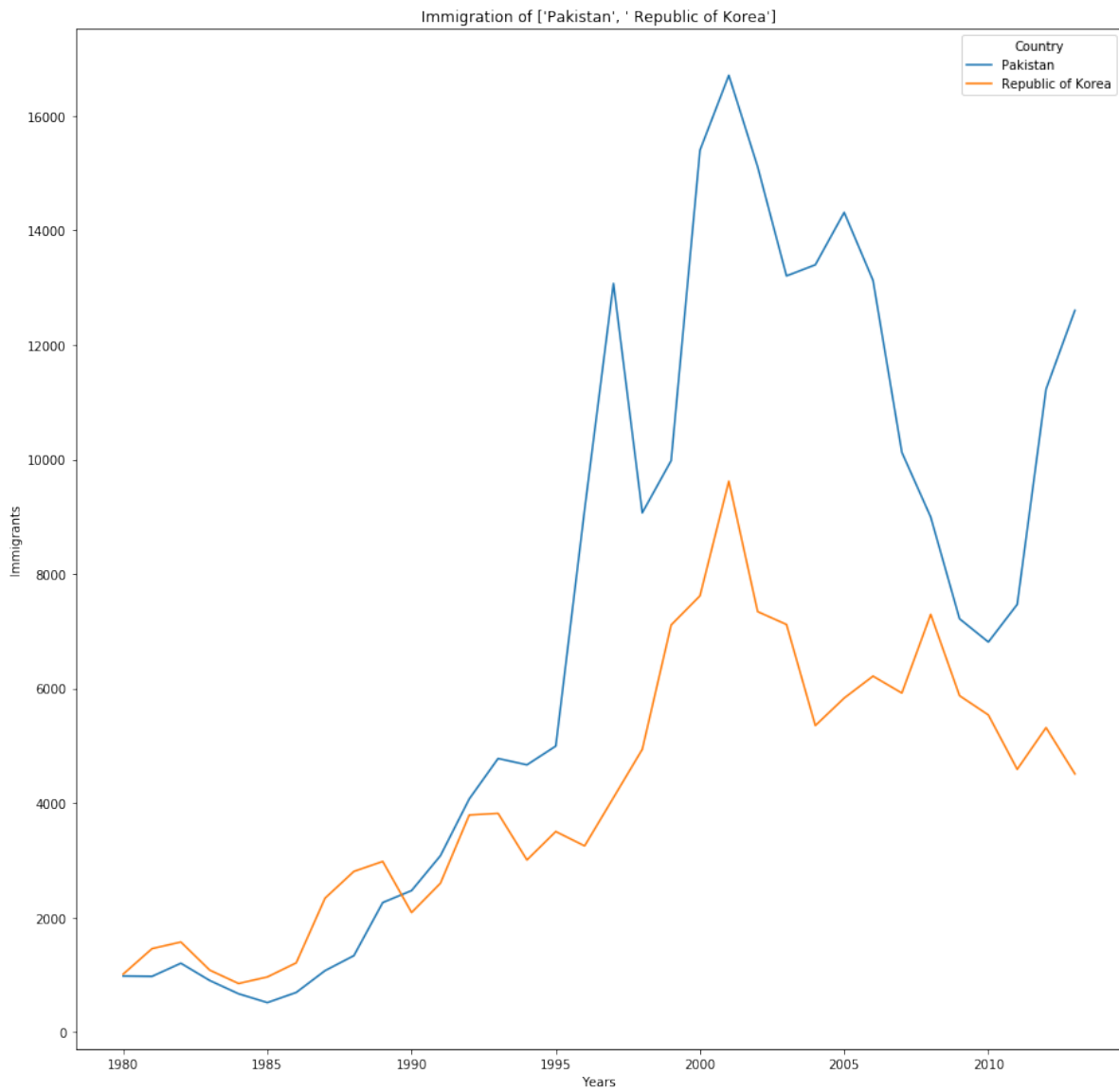
Lebanon

The Lebanese Civil war which ended in 1990, severely damaged the Lebanese economy. This dropped continuously for four years until 1994, when the Real GDP rose to 8%. This was possible because of the government's 20 billion dollar reconstruction program, which managed to create new jobs, making way for immigrants to return back

Pakistan and South Korea (Republic of Korea)

In [424]:

```
a9 = ['Pakistan', ' Republic of Korea']
canada.loc[['Pakistan', 'Republic of Korea'], years].transpose().plot(figsize=(15,15));
plt.title('Immigration of {as9}'.format(as9 = a9));
plt.xlabel('Years');
plt.ylabel('Immigrants');
```



Inference

Both countries follow a similar trendline. The graph shows steady increase till about the year 1995. Both countries show a drop in the 2000-2002 period.

Pakistan:

In 1994 inflation rate in Pakistan went to a whopping 14.5%. The following year the GDP growth rate dropped from 4.5% to 1.7%, lowest in decades. Pakistan's GDP growth rate dropped from 7.68% to 1.7% in a period of 4 years.

In 1999 Pervez Musharraf lead a millitary coup and overthrew the Nawaz Sharif government. This resulted in Liberalization. During Musharraf's term (1999-2008), around 11,9 million new jobs were created. Therefore this can be attributed to the drop in immigrants. Also, immigration was affected due to various events that happened in USA during 2001.

South Korea:

Korea shows a spike in the year 1997, this indicates the Asian Financial Crisis. The Korean Won began to depreciate heavily. Within months, a third of Korea's merchant banks were closed. Korean economy shrunked at an average of -6.65% per year. Conglomerates became a casualty of this.

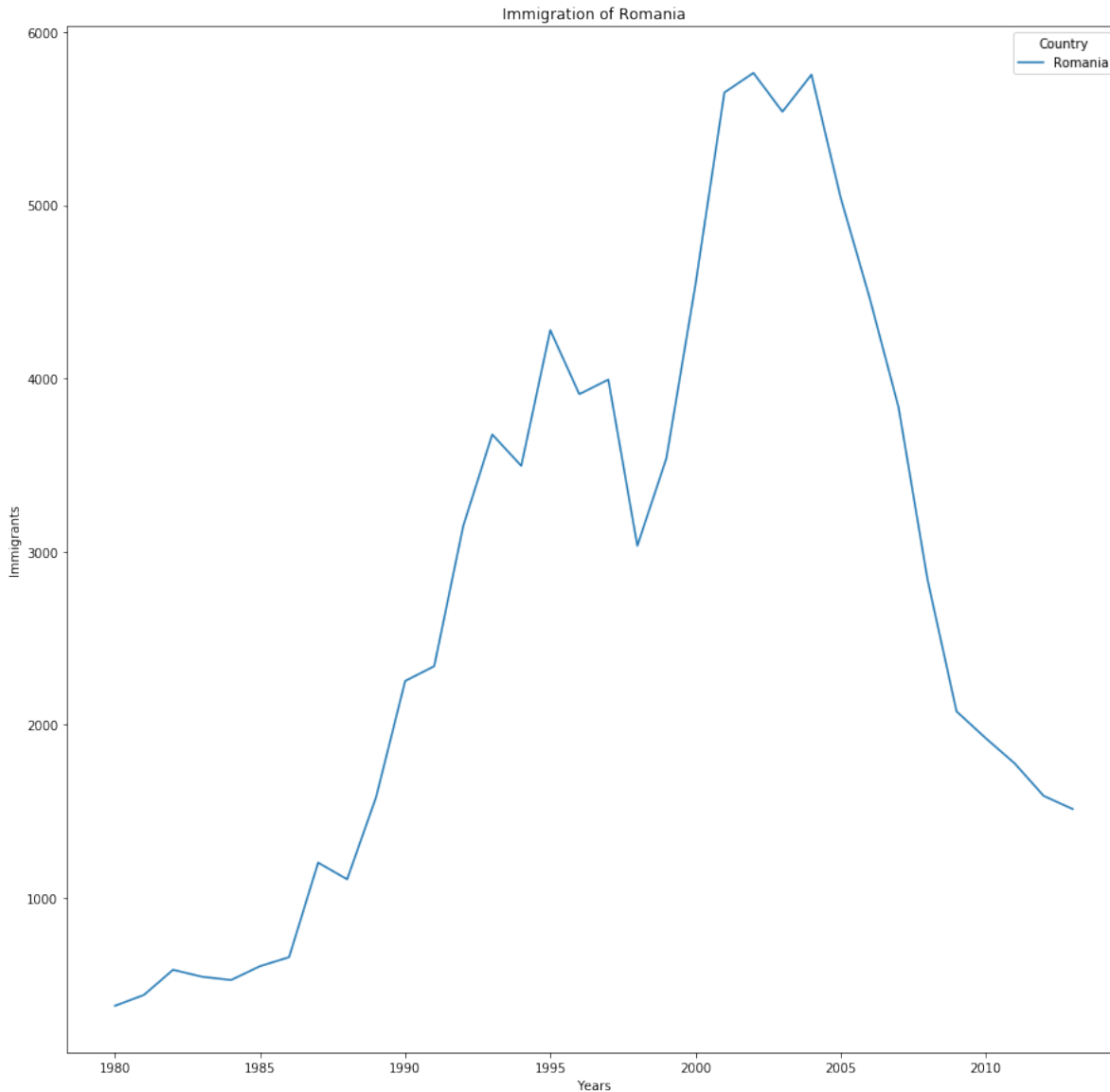
In 2000, the Korean economy shifted to a more market oriented economy. Growth rates touched closer to 11%, highest in decades. This resulted a drop in immigrants.

Individual trend analysis

Romania

In [425]:

```
years = list(map(str, range(1980, 2014)))  
canada.loc[['Romania'], years].transpose().plot(figsize=(15, 15));  
plt.title('Immigration of Romania');  
plt.xlabel('Years');  
plt.ylabel('Immigrants');
```



Inference

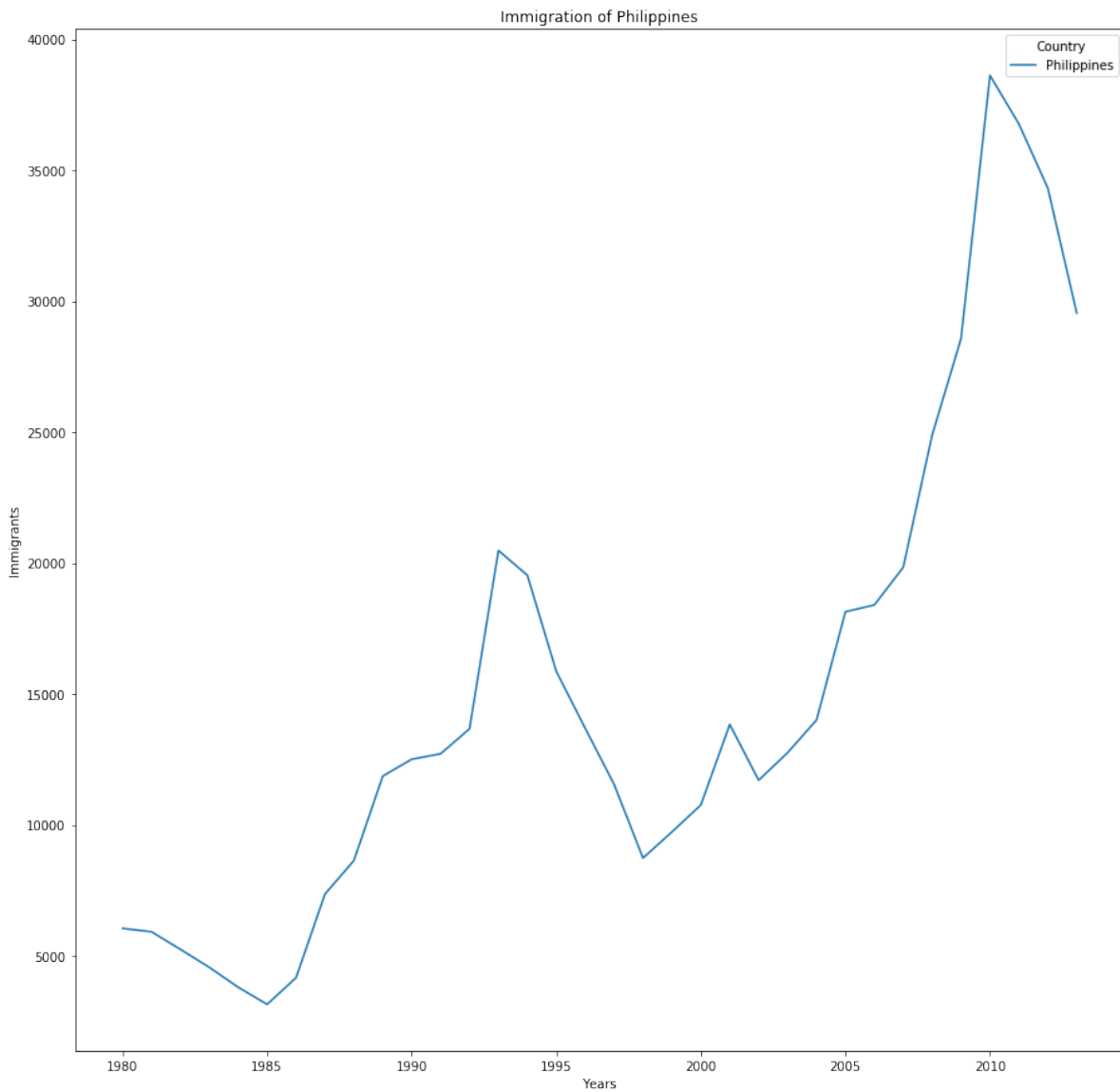
Point of Interest: 2004-2006 period.

This was the period of FDI in Romania. Demand of Romanian products increased in EU. Also, as Romania closed in towards joining EU, immigration to other countries substantially reduced,

Philippines

In [426]:

```
years = list(map(str, range(1980, 2014)))  
canada.loc[['Philippines'], years].transpose().plot(figsize=(15, 15));  
plt.title('Immigration of Philippines');  
plt.xlabel('Years');  
plt.ylabel('Immigrants');
```



Inference

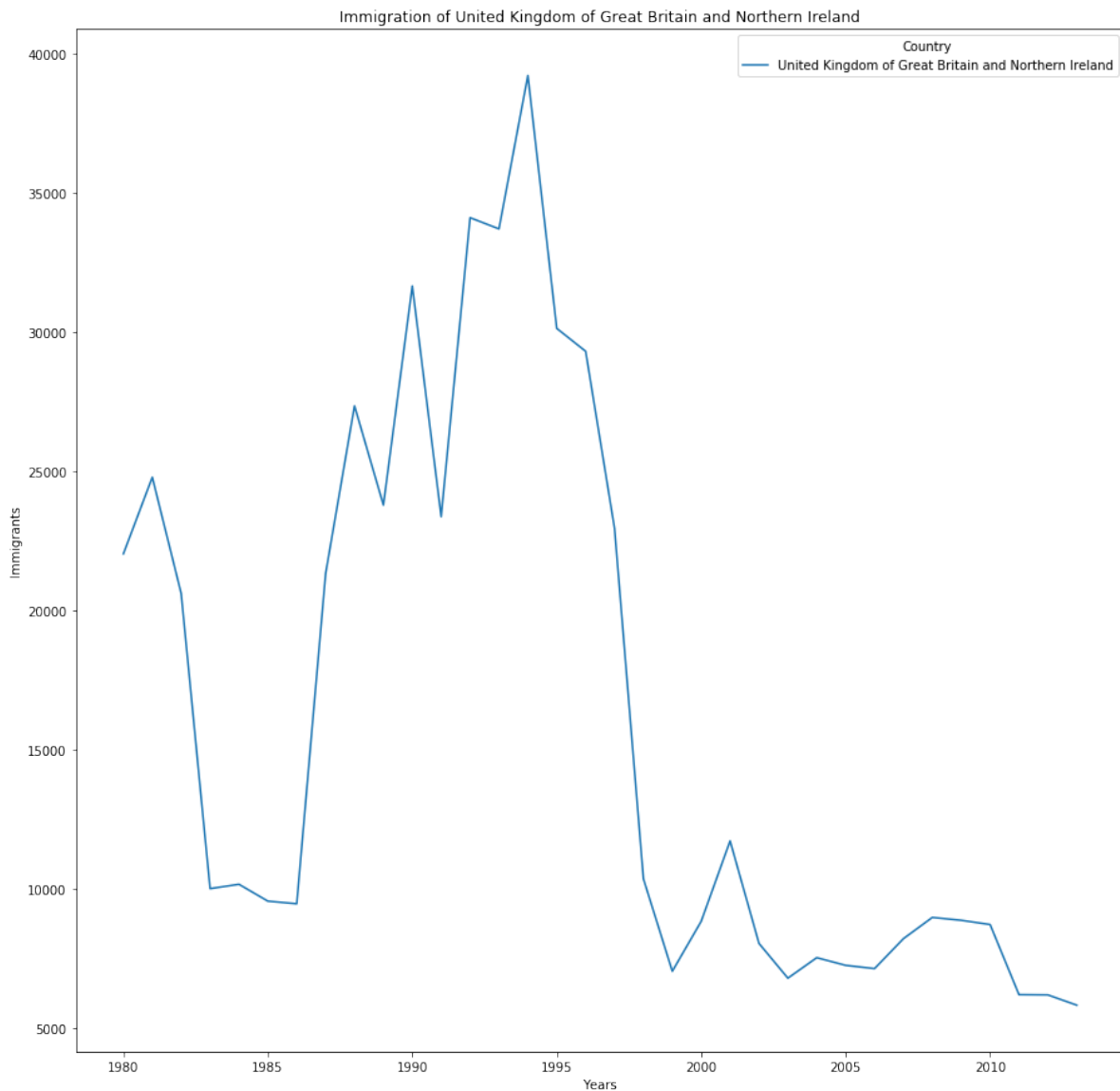
Points of Interests: 1985 and 1998

In the 1980s, annual GDP growth rate dropped to 2%, which was the lowest in decades. This resulted in people immigrating for jobs. Just like South Korea, the Asian Financial Crisis affected Philippines, which explains the spike after 1998,

United Kingdom

In [427]:

```
years = list(map(str, range(1980, 2014)))
canada.loc[['United Kingdom of Great Britain and Northern Ireland'], years].transpose().plot(figsize=(15, 15));
plt.title('Immigration of United Kingdom of Great Britain and Northern Ireland');
plt.xlabel('Years');
plt.ylabel('Immigrants');
```



Inference

Points of interests: 1984-1986 and 1994

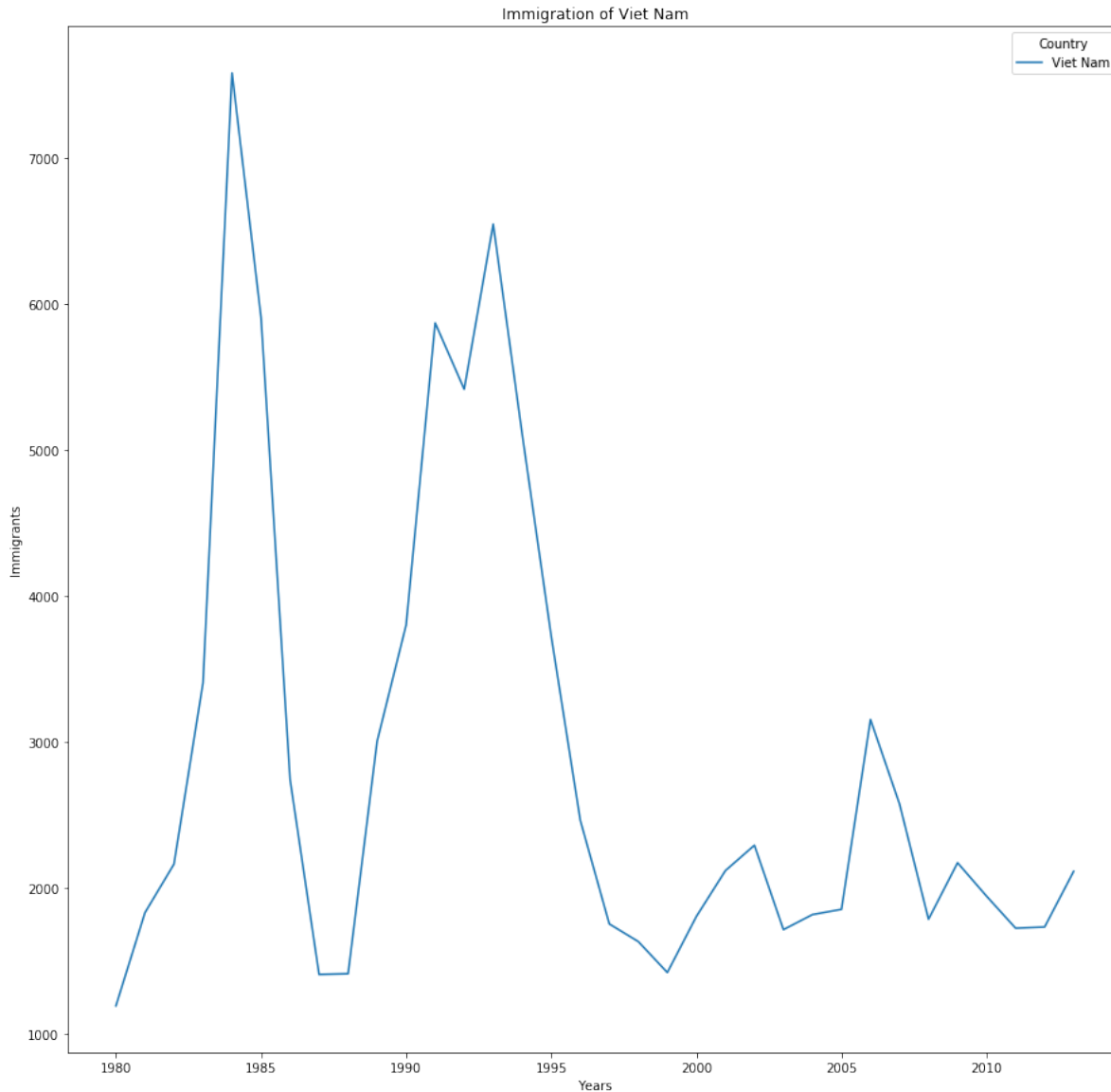
In 1984, an all time high unemployment rate of 11.9% was witnessed in the UK, forcing people to search for jobs in other countries. This attributes to the spike in 1984-1985.

The John Major government in 1990 focussed to creating new jobs, which substancially rediced the unemployment rate. This resulted in tyhe immigrants returning back to the UK, which can be seen as the graph is seen dipping post 1994.

Viet Nam

In [428]:

```
years = list(map(str, range(1980, 2014)))
canada.loc['Viet Nam', years].transpose().plot(figsize=(15, 15));
plt.title('Immigration of Viet Nam');
plt.xlabel('Years');
plt.ylabel('Immigrants');
```



Inference

Points of interest: 1983-1998 and 1993,

Vietnam entered its Third Five-Year-Plan scheme in 1981. This resulted in a massive economic failure, slowing down the economic growth heavily. The government swiftly corrected its path. Increase in exports is the main reason for immigrants to drop post 1993. This was done to bring back the economy on track.

In []: