immig

June 25, 2024

[1]:
```
import pandas as pd

import matplotlib.pyplot as plt
```

[2]:
```
dfcan = pd.read_excel('/Users/Yedou/Desktop/Immigration_Data/Canada.xlsx',
    sheet_name='Canada by Citizenship',
    skiprows=range(20),
    skipfooter=2)
```

[3]:
```
dfcan # All row and columns
```

[3]:

	Type	Coverage	OdName	AREA	AreaName	REG	\
0	Immigrants	Foreigners	Afghanistan	935	Asia	5501	
1	Immigrants	Foreigners	Albania	908	Europe	925	
2	Immigrants	Foreigners	Algeria	903	Africa	912	
3	Immigrants	Foreigners	American Samoa	909	Oceania	957	
4	Immigrants	Foreigners	Andorra	908	Europe	925	
..	…	…	…	…	…	…	
190	Immigrants	Foreigners	Viet Nam	935	Asia	920	
191	Immigrants	Foreigners	Western Sahara	903	Africa	912	
192	Immigrants	Foreigners	Yemen	935	Asia	922	
193	Immigrants	Foreigners	Zambia	903	Africa	910	
194	Immigrants	Foreigners	Zimbabwe	903	Africa	910	

	RegName	DEV	…	DevName	1980	…	2004	2005	2006	\
0	Southern Asia	902		Developing regions	16	…	2978	3436	3009	
1	Southern Europe	901		Developed regions	1	…	1450	1223	856	
2	Northern Africa	902		Developing regions	80	…	3616	3626	4807	
3	Polynesia	902		Developing regions	0	…	0	0	1	
4	Southern Europe	901		Developed regions	0	…	0	0	1	
..	…	…	…	…	…	…	…	…	…	
190	South-Eastern Asia	902		Developing regions	1191	…	1816	1852	3153	
191	Northern Africa	902		Developing regions	0	…	0	0	1	
192	Western Asia	902		Developing regions	1	…	124	161	140	
193	Eastern Africa	902		Developing regions	11	…	56	91	77	
194	Eastern Africa	902		Developing regions	72	…	1450	615	454	

```
     2007  2008  2009  2010  2011  2012  2013
0    2652  2111  1746  1758  2203  2635  2004
1     702   560   716   561   539   620   603
2    3623  4005  5393  4752  4325  3774  4331
3       0     0     0     0     0     0     0
4       1     0     0     0     0     1     1
..    …   …   …   …   …   …   …
190  2574  1784  2171  1942  1723  1731  2112
191     0     0     0     0     0     0     0
192   122   133   128   211   160   174   217
193    71    64    60   102    69    46    59
194   663   611   508   494   434   437   407

[195 rows x 43 columns]
```

[4]:
```
dfcan.head() # first 5 columns
```

[4]:
```
         Type    Coverage          OdName  AREA AreaName   REG  \
0  Immigrants  Foreigners     Afghanistan   935     Asia  5501
1  Immigrants  Foreigners         Albania   908   Europe   925
2  Immigrants  Foreigners         Algeria   903   Africa   912
3  Immigrants  Foreigners  American Samoa   909  Oceania   957
4  Immigrants  Foreigners         Andorra   908   Europe   925

           RegName  DEV             DevName  1980  …  2004  2005  2006  \
0    Southern Asia  902  Developing regions    16  …  2978  3436  3009
1  Southern Europe  901   Developed regions     1  …  1450  1223   856
2  Northern Africa  902  Developing regions    80  …  3616  3626  4807
3        Polynesia  902  Developing regions     0  …     0     0     1
4  Southern Europe  901   Developed regions     0  …     0     0     1

   2007  2008  2009  2010  2011  2012  2013
0  2652  2111  1746  1758  2203  2635  2004
1   702   560   716   561   539   620   603
2  3623  4005  5393  4752  4325  3774  4331
3     0     0     0     0     0     0     0
4     1     0     0     0     0     1     1

[5 rows x 43 columns]
```

[5]:
```
# Moving column OdName to index

dfcan.set_index('OdName', inplace=True)
```

[6]:
```
dfcan
```

[6]:
```
                    Type    Coverage  AREA AreaName   REG  \
OdName
Afghanistan     Immigrants  Foreigners   935     Asia  5501
Albania         Immigrants  Foreigners   908   Europe   925
Algeria         Immigrants  Foreigners   903   Africa   912
American Samoa  Immigrants  Foreigners   909  Oceania   957
Andorra         Immigrants  Foreigners   908   Europe   925
…                      …         …   …      …   …
Viet Nam        Immigrants  Foreigners   935     Asia   920
Western Sahara  Immigrants  Foreigners   903   Africa   912
Yemen           Immigrants  Foreigners   935     Asia   922
Zambia          Immigrants  Foreigners   903   Africa   910
Zimbabwe        Immigrants  Foreigners   903   Africa   910

                           RegName  DEV             DevName  1980  1981  …  \
OdName                                                                   …
Afghanistan          Southern Asia  902  Developing regions    16    39  …
Albania            Southern Europe  901   Developed regions     1     0  …
Algeria            Northern Africa  902  Developing regions    80    67  …
American Samoa           Polynesia  902  Developing regions     0     1  …
Andorra            Southern Europe  901   Developed regions     0     0  …
…                              …  …                 …   …   …  …
Viet Nam        South-Eastern Asia  902  Developing regions  1191  1829  …
Western Sahara     Northern Africa  902  Developing regions     0     0  …
Yemen                Western Asia  902  Developing regions     1     2  …
Zambia              Eastern Africa  902  Developing regions    11    17  …
Zimbabwe            Eastern Africa  902  Developing regions    72   114  …

                2004  2005  2006  2007  2008  2009  2010  2011  2012  2013
OdName
Afghanistan     2978  3436  3009  2652  2111  1746  1758  2203  2635  2004
Albania         1450  1223   856   702   560   716   561   539   620   603
Algeria         3616  3626  4807  3623  4005  5393  4752  4325  3774  4331
American Samoa     0     0     1     0     0     0     0     0     0     0
Andorra            0     0     1     1     0     0     0     0     1     1
…                …   …     …     …     …     …     …     …     …     …
Viet Nam        1816  1852  3153  2574  1784  2171  1942  1723  1731  2112
Western Sahara     0     0     1     0     0     0     0     0     0     0
Yemen            124   161   140   122   133   128   211   160   174   217
Zambia            56    91    77    71    64    60   102    69    46    59
Zimbabwe        1450   615   454   663   611   508   494   434   437   407

[195 rows x 42 columns]
```

[7]:
```
# convertion: Columns name to string

dfcan.columns =list(map( str,dfcan.columns))
```

[8]:
```
# Info

dfcan.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Index: 195 entries, Afghanistan to Zimbabwe
Data columns (total 42 columns):
 #   Column    Non-Null Count  Dtype
---  ------    --------------  -----
 0   Type      195 non-null    object
 1   Coverage  195 non-null    object
 2   AREA      195 non-null    int64
 3   AreaName  195 non-null    object
 4   REG       195 non-null    int64
 5   RegName   195 non-null    object
 6   DEV       195 non-null    int64
 7   DevName   195 non-null    object
 8   1980      195 non-null    int64
 9   1981      195 non-null    int64
 10  1982      195 non-null    int64
 11  1983      195 non-null    int64
 12  1984      195 non-null    int64
 13  1985      195 non-null    int64
 14  1986      195 non-null    int64
 15  1987      195 non-null    int64
 16  1988      195 non-null    int64
 17  1989      195 non-null    int64
 18  1990      195 non-null    int64
 19  1991      195 non-null    int64
 20  1992      195 non-null    int64
 21  1993      195 non-null    int64
 22  1994      195 non-null    int64
 23  1995      195 non-null    int64
 24  1996      195 non-null    int64
 25  1997      195 non-null    int64
 26  1998      195 non-null    int64
 27  1999      195 non-null    int64
 28  2000      195 non-null    int64
 29  2001      195 non-null    int64
 30  2002      195 non-null    int64
 31  2003      195 non-null    int64
 32  2004      195 non-null    int64
 33  2005      195 non-null    int64
 34  2006      195 non-null    int64
 35  2007      195 non-null    int64
 36  2008      195 non-null    int64
 37  2009      195 non-null    int64
```

```
 38  2010      195 non-null    int64
 39  2011      195 non-null    int64
 40  2012      195 non-null    int64
 41  2013      195 non-null    int64
dtypes: int64(37), object(5)
memory usage: 65.5+ KB
```

[9]:
```
# Sum of immigration from 1980 to 2013

dfcan['Total'] = dfcan[dfcan.columns[8:42]].sum(axis=1)
```

[10]:
```
dfcan
```

[10]:

OdName	Type	Coverage	AREA	AreaName	REG	\
Afghanistan	Immigrants	Foreigners	935	Asia	5501	
Albania	Immigrants	Foreigners	908	Europe	925	
Algeria	Immigrants	Foreigners	903	Africa	912	
American Samoa	Immigrants	Foreigners	909	Oceania	957	
Andorra	Immigrants	Foreigners	908	Europe	925	
…	…	…	…	…	…	
Viet Nam	Immigrants	Foreigners	935	Asia	920	
Western Sahara	Immigrants	Foreigners	903	Africa	912	
Yemen	Immigrants	Foreigners	935	Asia	922	
Zambia	Immigrants	Foreigners	903	Africa	910	
Zimbabwe	Immigrants	Foreigners	903	Africa	910	

OdName	RegName	DEV	DevName	1980	1981	…	\
Afghanistan	Southern Asia	902	Developing regions	16	39	…	
Albania	Southern Europe	901	Developed regions	1	0	…	
Algeria	Northern Africa	902	Developing regions	80	67	…	
American Samoa	Polynesia	902	Developing regions	0	1	…	
Andorra	Southern Europe	901	Developed regions	0	0	…	
…	…	…	…	…	…	…	
Viet Nam	South-Eastern Asia	902	Developing regions	1191	1829	…	
Western Sahara	Northern Africa	902	Developing regions	0	0	…	
Yemen	Western Asia	902	Developing regions	1	2	…	
Zambia	Eastern Africa	902	Developing regions	11	17	…	
Zimbabwe	Eastern Africa	902	Developing regions	72	114	…	

OdName	2005	2006	2007	2008	2009	2010	2011	2012	2013	Total
Afghanistan	3436	3009	2652	2111	1746	1758	2203	2635	2004	58639
Albania	1223	856	702	560	716	561	539	620	603	15699
Algeria	3626	4807	3623	4005	5393	4752	4325	3774	4331	69439
American Samoa	0	1	0	0	0	0	0	0	0	6

```
Andorra              0     1     1     0     0     0     0     1     1     15
…                  …   …   …   …   …   …   …   …   …   …
Viet Nam          1852  3153  2574  1784  2171  1942  1723  1731  2112  97146
Western Sahara       0     1     0     0     0     0     0     0     0      2
Yemen              161   140   122   133   128   211   160   174   217   2985
Zambia              91    77    71    64    60   102    69    46    59   1677
Zimbabwe           615   454   663   611   508   494   434   437   407   8598

[195 rows x 43 columns]
```

[11]:
```
# Created years variable

years = list(map(str,range(1980,2014)))# rang (1980,2014) means between 1980
 ↪and 2013

years
```

[11]:
```
['1980',
 '1981',
 '1982',
 '1983',
 '1984',
 '1985',
 '1986',
 '1987',
 '1988',
 '1989',
 '1990',
 '1991',
 '1992',
 '1993',
 '1994',
 '1995',
 '1996',
 '1997',
 '1998',
 '1999',
 '2000',
 '2001',
 '2002',
 '2003',
 '2004',
 '2005',
 '2006',
 '2007',
 '2008',
 '2009',
```

```
 '2010',
 '2011',
 '2012',
 '2013']
```

[13]:
```
# Ascending list by number of immigrants by country

dfcan.sort_values(['Total'],ascending=False,axis = 0,inplace= True)
```

[14]:
```
# Five top countries with highest number of immigrants

dfcan.head()
```

[14]:

OdName	Type	Coverage	AREA	AreaName	REG	RegName	DEV	DevName	1980	1981	…	2005	2006
India	Immigrants	Foreigners	935	Asia	5501	Southern Asia	902	Developing regions	8880		…		
China	Immigrants	Foreigners	935	Asia	906	Eastern Asia	902	Developing regions	5123		…		
United Kingdom of Great Britain and Northern Ir…	Immigrants	Foreigners	908	Europe	924	Northern Europe	901	Developed regions	22045		…		
Philippines	Immigrants	Foreigners	935	Asia	920	South-Eastern Asia	902	Developing regions	6051		…		
Pakistan	Immigrants	Foreigners	935	Asia	5501	Southern Asia	902	Developing regions	978		…		

```
India                                                 8670  …  36210  33848
China                                                 6682  …  42584  33518
United Kingdom of Great Britain and Northern Ir…     24796  …   7258   7140
Philippines                                           5921  …  18139  18400
Pakistan                                               972  …  14314  13127

                                                      2007   2008   2009  \
OdName
India                                                28742  28261  29456
China                                                27642  30037  29622
United Kingdom of Great Britain and Northern Ir…      8216   8979   8876
Philippines                                          19837  24887  28573
Pakistan                                             10124   8994   7217

                                                      2010   2011   2012  \
OdName
India                                                34235  27509  30933
China                                                30391  28502  33024
United Kingdom of Great Britain and Northern Ir…      8724   6204   6195
Philippines                                          38617  36765  34315
Pakistan                                              6811   7468  11227

                                                      2013   Total
OdName
India                                                33087  691904
China                                                34129  659962
United Kingdom of Great Britain and Northern Ir…      5827  551500
Philippines                                          29544  511391
Pakistan                                             12603  241600

[5 rows x 43 columns]
```

[15]:
```
# The last five countries with low numbers of immigrants

dfcan.tail()
```

[15]:
```
                      Type    Coverage  AREA AreaName  REG          RegName  \
OdName
San Marino        Immigrants  Foreigners   908   Europe  925  Southern Europe
New Caledonia     Immigrants  Foreigners   909  Oceania  928       Melanesia
Marshall Islands  Immigrants  Foreigners   909  Oceania  954      Micronesia
Western Sahara    Immigrants  Foreigners   903   Africa  912  Northern Africa
Palau             Immigrants  Foreigners   909  Oceania  954      Micronesia

                  DEV             DevName  1980  1981  …  2005  2006  2007  \
OdName                                                 …
San Marino        901   Developed regions     1     0  …     0     0     0
```

```
New Caledonia     902  Developing regions     0     0  …     0     0     0
Marshall Islands  902  Developing regions     0     0  …     0     0     2
Western Sahara    902  Developing regions     0     0  …     0     1     0
Palau             902  Developing regions     0     0  …     0     0     1

                  2008  2009  2010  2011  2012  2013  Total
OdName
San Marino           0     0     1     0     0     0      5
New Caledonia        1     0     0     0     0     2      5
Marshall Islands     0     0     0     0     0     0      2
Western Sahara       0     0     0     0     0     0      2
Palau                0     0     0     0     0     0      1

[5 rows x 43 columns]
```

[18]:
```
#Created a dataframe for Five top countries with highest number of immigrants

dfcantop = dfcan.head()
```

[19]:
```
dfcantop
```

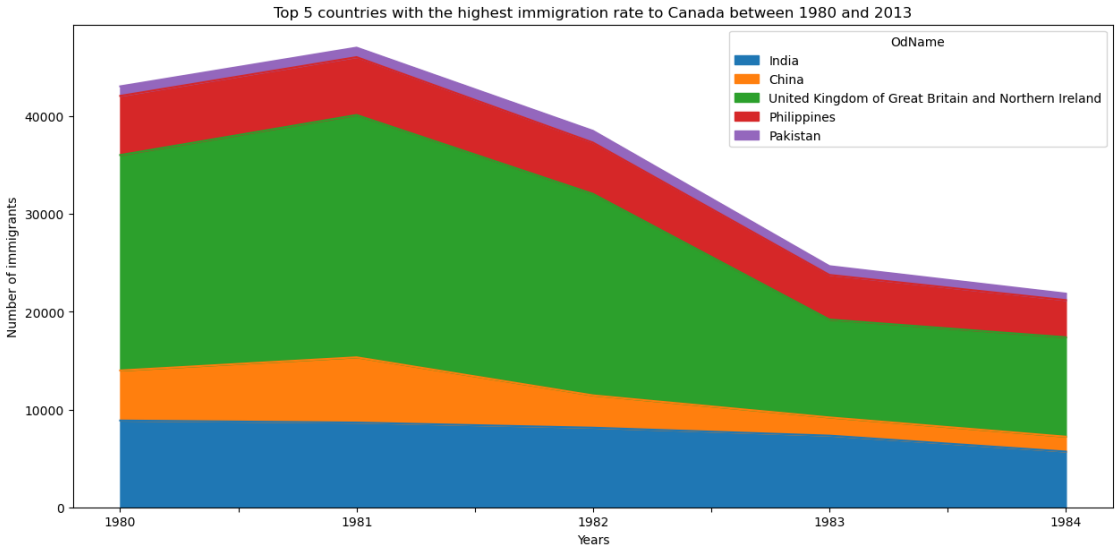
[19]:
```
                                                    Type    Coverage  \
OdName
India                                          Immigrants  Foreigners
China                                          Immigrants  Foreigners
United Kingdom of Great Britain and Northern Ir…  Immigrants  Foreigners
Philippines                                    Immigrants  Foreigners
Pakistan                                       Immigrants  Foreigners

                                                AREA AreaName   REG  \
OdName
India                                            935     Asia  5501
China                                            935     Asia   906
United Kingdom of Great Britain and Northern Ir…   908   Europe   924
Philippines                                      935     Asia   920
Pakistan                                         935     Asia  5501

                                                          RegName  DEV  \
OdName
India                                                Southern Asia  902
China                                                 Eastern Asia  902
United Kingdom of Great Britain and Northern Ir…     Northern Europe  901
Philippines                                        South-Eastern Asia  902
Pakistan                                             Southern Asia  902

                                                          DevName   1980  \
OdName
```

```
India                                                Developing regions   8880
China                                                Developing regions   5123
United Kingdom of Great Britain and Northern Ir…   Developed regions  22045
Philippines                                          Developing regions   6051
Pakistan                                             Developing regions    978

                                                      1981  …   2005   2006  \
OdName                                                      …
India                                                 8670  …  36210  33848
China                                                 6682  …  42584  33518
United Kingdom of Great Britain and Northern Ir…     24796  …   7258   7140
Philippines                                           5921  …  18139  18400
Pakistan                                               972  …  14314  13127

                                                      2007   2008   2009  \
OdName
India                                                28742  28261  29456
China                                                27642  30037  29622
United Kingdom of Great Britain and Northern Ir…      8216   8979   8876
Philippines                                          19837  24887  28573
Pakistan                                             10124   8994   7217

                                                      2010   2011   2012  \
OdName
India                                                34235  27509  30933
China                                                30391  28502  33024
United Kingdom of Great Britain and Northern Ir…      8724   6204   6195
Philippines                                          38617  36765  34315
Pakistan                                              6811   7468  11227

                                                      2013   Total
OdName
India                                                33087  691904
China                                                34129  659962
United Kingdom of Great Britain and Northern Ir…      5827  551500
Philippines                                          29544  511391
Pakistan                                             12603  241600

[5 rows x 43 columns]
```

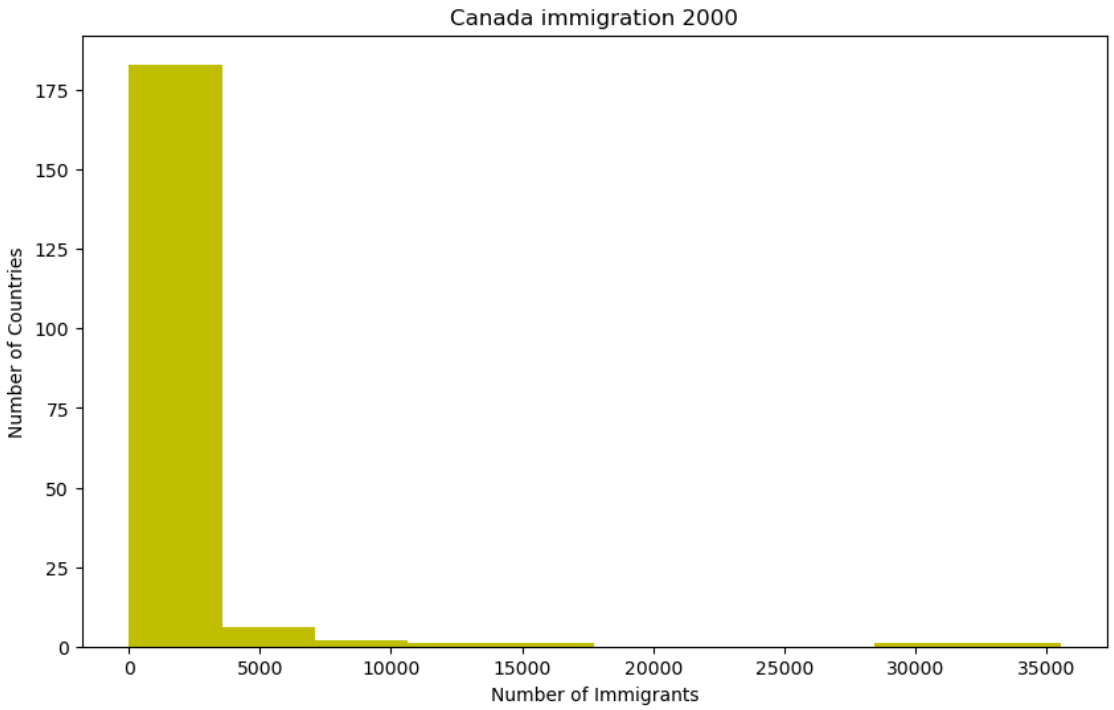
[20]:
```
# Change countries column to index
dfcantop = dfcantop[years].transpose()
```

[21]:
```
# Visualization of Five top countries with highest number of immigrants with␣
 ↪plot(kind = 'area')

dfcantop.head().plot(kind = 'area', figsize = (15,7))
```

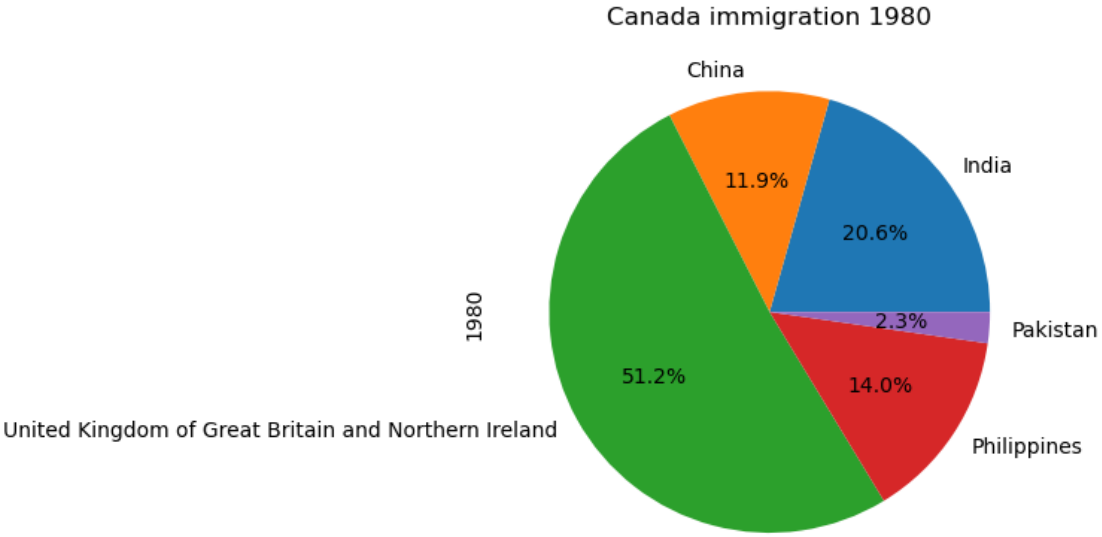
```
plt.title("Top 5 countries with the highest immigration rate to Canada between␣
 ↪1980 and 2013")
plt.xlabel("Years")
plt.ylabel("Number of immigrants")
plt.show()
```



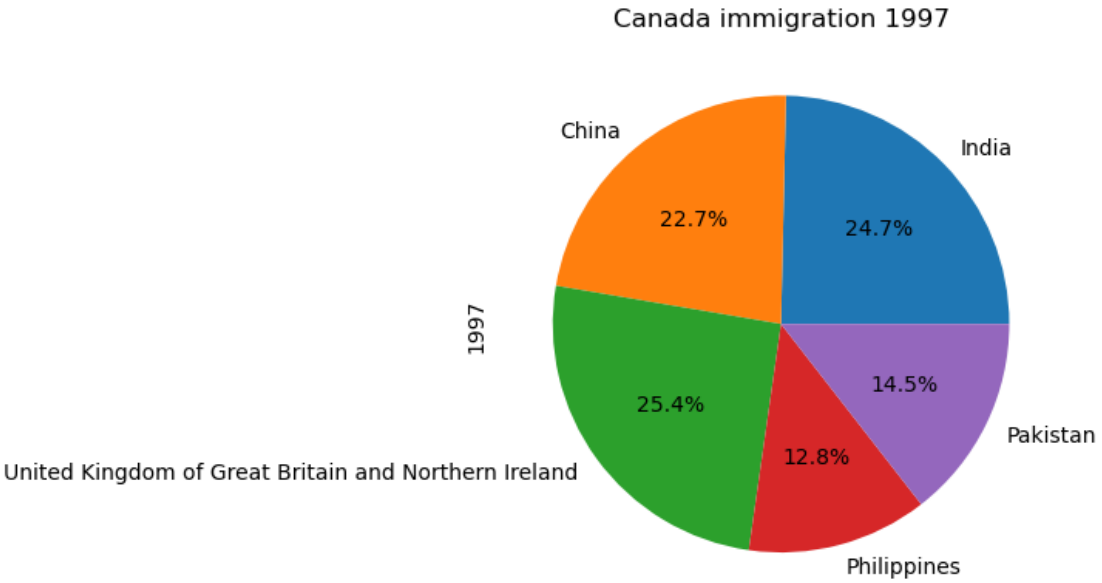
[22]:
```
# Visualization of the number of immigrants at 2000 with graph (kind = 'hist')

dfcan['2000'].plot(kind = 'hist', color = 'y', figsize = (10,6))

plt.title("Canada immigration 2000")

plt.xlabel("Number of Immigrants")

plt.ylabel("Number of Countries")

plt.show()
```



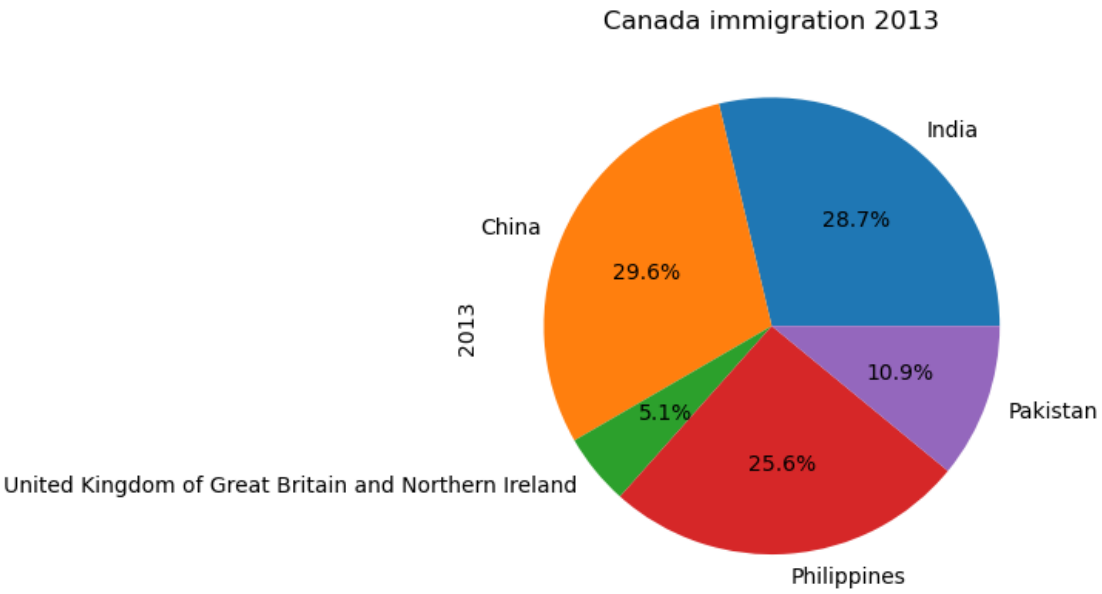
[23]:
```
# 1980 immigration by country
dfcan.head()['1980'].plot(kind = 'pie', autopct='%1.1f%%')

plt.title("Canada immigration 1980")

plt.show()
```



[24]:
```
# 1997 immigration by country

dfcan.head()['1997'].plot(kind = 'pie', autopct='%1.1f%%')

plt.title("Canada immigration 1997")

plt.show()
```

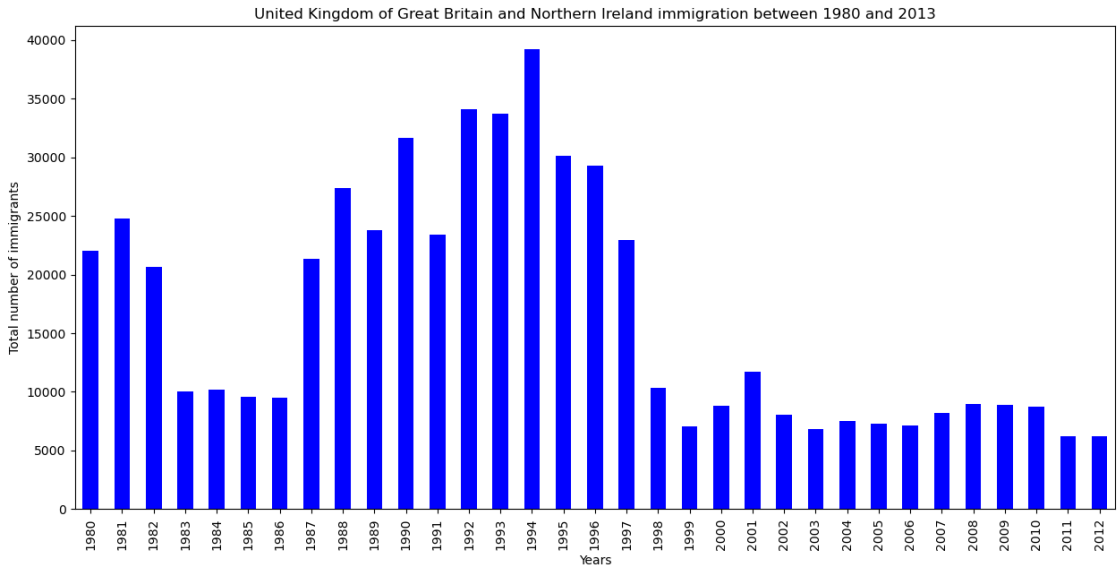


[25]:
```
# immigration by country

dfcan.head()['2013'].plot(kind = 'pie', autopct='%1.1f%%')

plt.title("Canada immigration 2013")

plt.show()
```

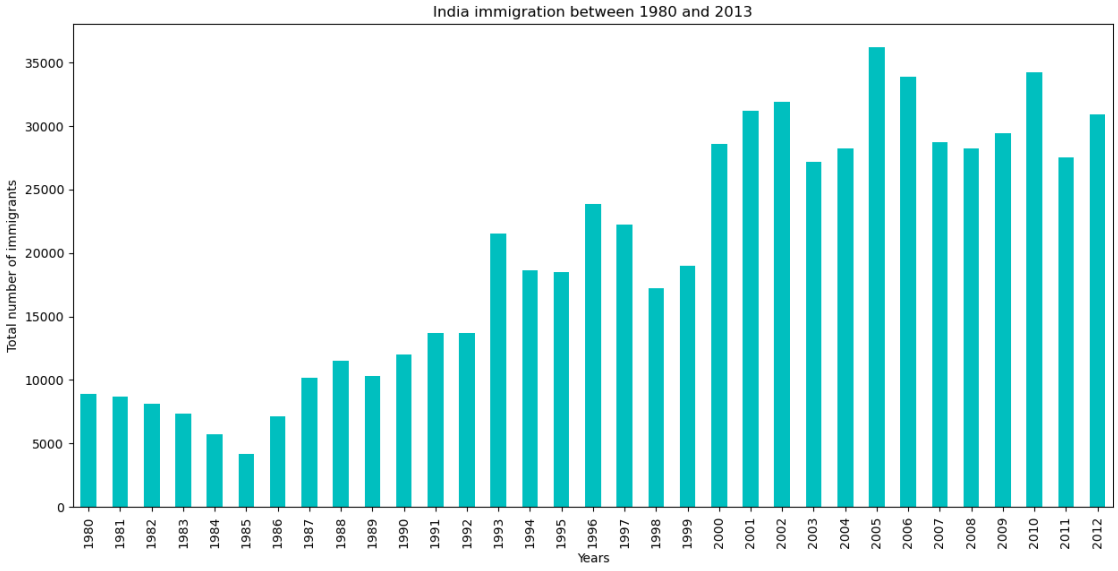


[27]:
```
years2 = list(map(str,range(1980,2013))) #  list creation
```

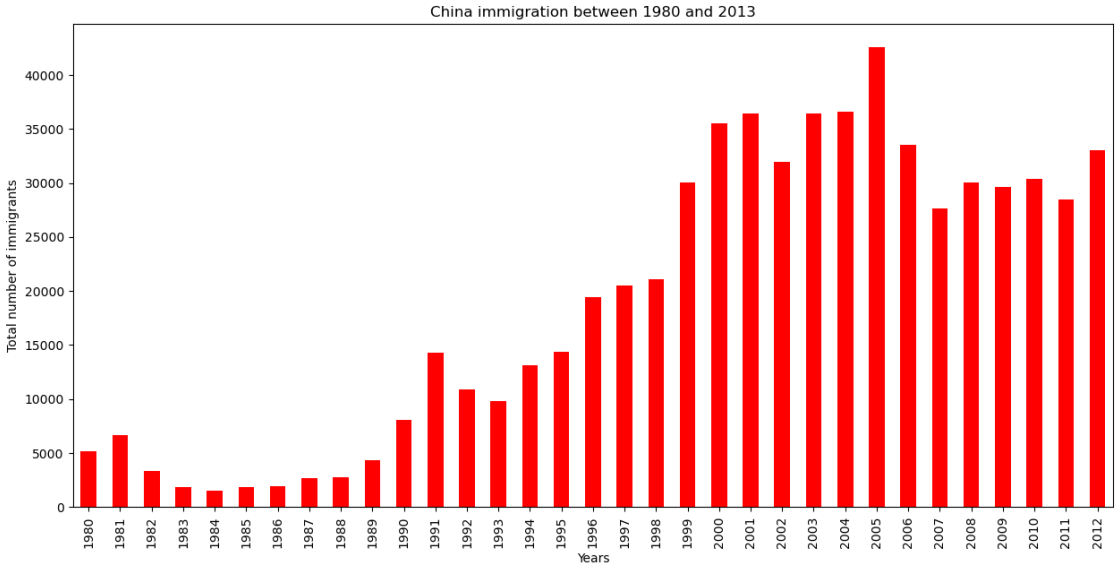
[32]:
```
dfuk = dfcan.loc['United Kingdom of Great Britain and Northern Ireland',years2]

dfuk.plot(kind = 'bar',color ='b',figsize = (15,7))

# Histogram: United Kingdom of Great Britain and Northern Ireland immigration␣
 ↪between 1980 and 2013

plt.title("United Kingdom of Great Britain and Northern Ireland immigration␣
 ↪between 1980 and 2013")

plt.xlabel("Years")

plt.ylabel("Total number of immigrants")

plt.show()
```



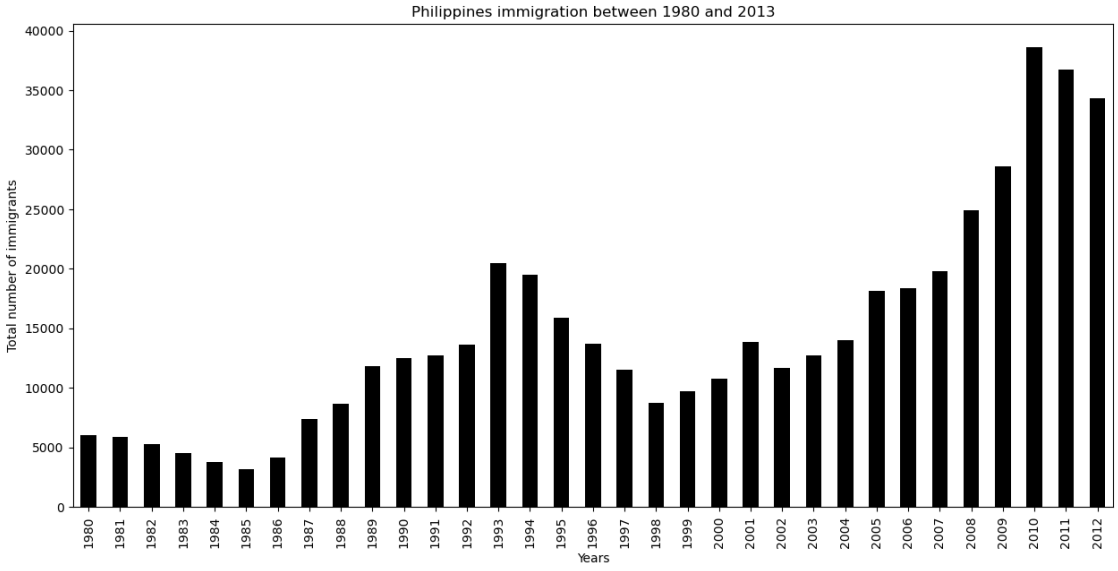
[29]:
```
# Histogram: India immigration between 1980 and 2013

dfindia = dfcan.loc['India',years2]

dfindia.plot(kind = 'bar',color ='c',figsize = (15,7))

plt.title("India immigration between 1980 and 2013")

plt.xlabel("Years")

plt.ylabel("Total number of immigrants")

plt.show()
```



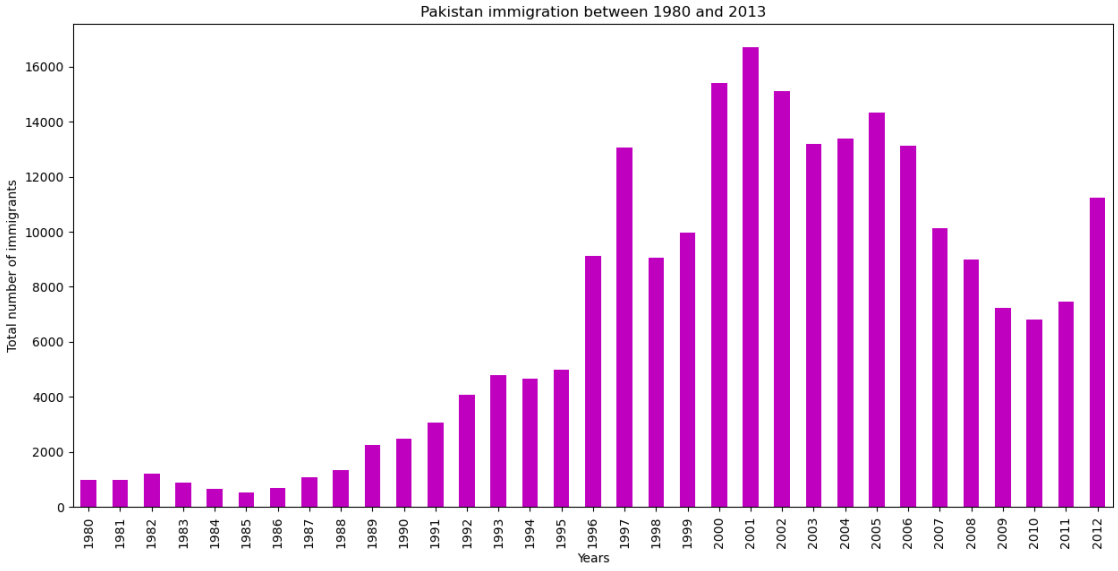
[33]:
```
# Histogram: China immigration between 1980 and 2013

dfchina = dfcan.loc['China',years2]

dfchina.plot(kind = 'bar',color ='r',figsize = (15,7))

plt.title("China immigration between 1980 and 2013")

plt.xlabel("Years")

plt.ylabel("Total number of immigrants")

plt.show()
```



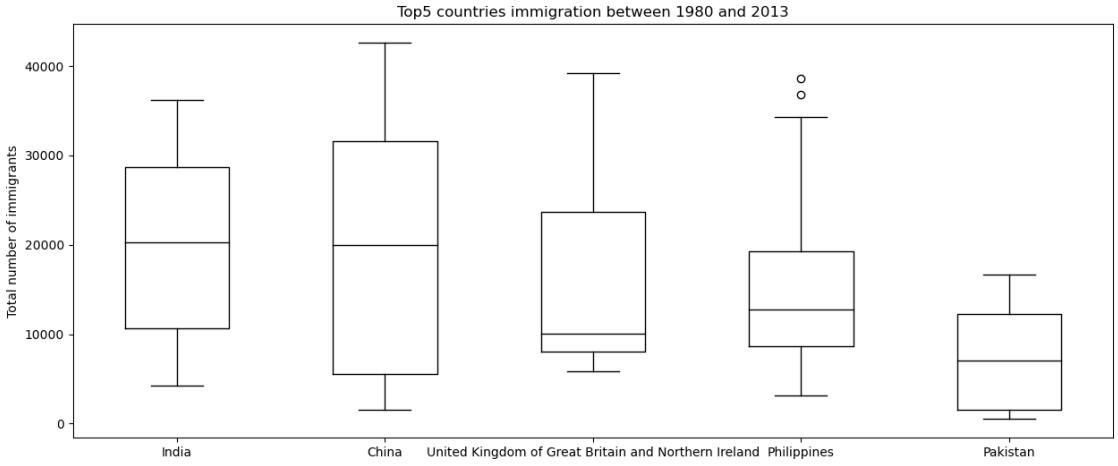
[36]:
```
# Histogram: Philippines immigration between 1980 and 2013

dfphilip = dfcan.loc['Philippines',years2]

dfphilip.plot(kind = 'bar',color ='k',figsize = (15,7))

plt.title("Philippines immigration between 1980 and 2013")

plt.xlabel("Years")

plt.ylabel("Total number of immigrants")

plt.show()
```



[37]:
```
# Histogram: Pakistan immigration between 1980 and 2013

dfpak = dfcan.loc['Pakistan',years2]

dfpak.plot(kind = 'bar',color ='m',figsize = (15,7))

plt.title("Pakistan immigration between 1980 and 2013")

plt.xlabel("Years")

plt.ylabel("Total number of immigrants")

plt.show()
```



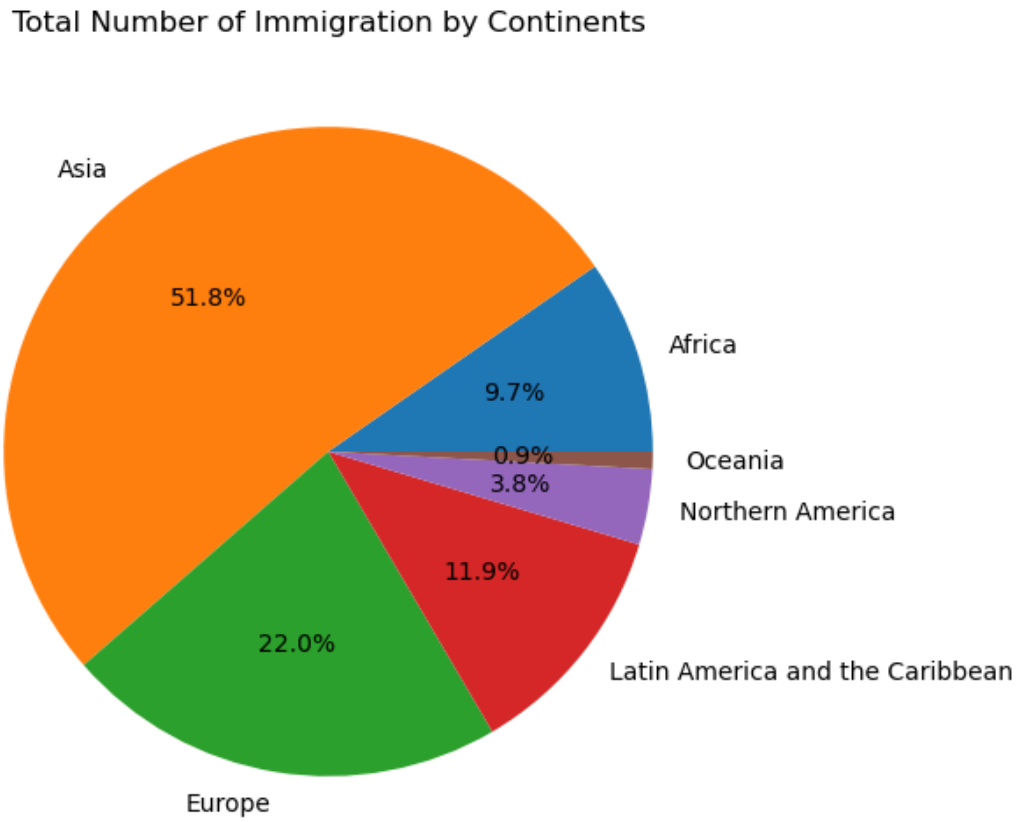
[43]:
```
# Box Plot: Top5 countries immigration between 1980 and 2013

dftop5 = dfcan.loc[['India','China','United Kingdom of Great Britain and␣
 ↪Northern Ireland','Philippines','Pakistan'], years].transpose()

dftop5.plot( kind = 'box', color = 'k', figsize = (15,6))

plt.title("Top5 countries immigration between 1980 and 2013")

plt.ylabel("Total number of immigrants")

plt.show()
```



[48]:
```
# Pie chart: Total Number of Immigration by Continents

df_conti = dfcan.groupby("AreaName",axis = 0).sum()

print(df_conti.head())

df_conti['Total'].plot( kind = 'pie',autopct='%1.1f%%', figsize = (15,6))
plt.title("Total Number of Immigration by Continents")

plt.xlabel("")

plt.ylabel("")

plt.show()
```

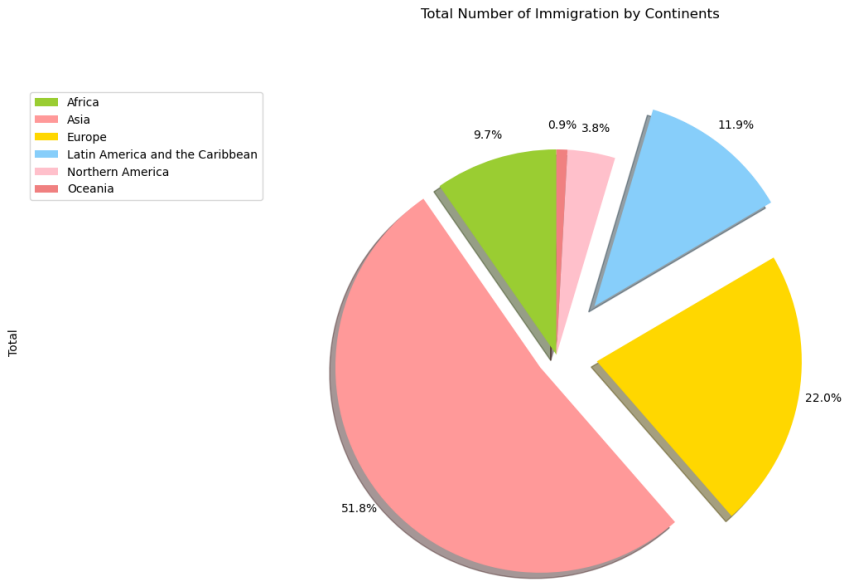
```
                                AREA     REG    DEV   1980   1981   1982  \
AreaName
Africa                         48762   49242  48708   3951   4363   3819
Asia                           45815  109147  44197  31025  34314  30214
Europe                         39044   39754  38743  39760  44802  42720
Latin America and the Caribbean  29832   30395  29766  13081  15215  16769
Northern America                1810    1810   1802   9378  10030   9074

                                1983   1984   1985   1986  …    2005  \
AreaName                                                  …
Africa                          2671   2639   2650   3782  …   27523
Asia                           24696  27274  23850  28739  …  159253
Europe                         24638  22287  20844  24370  …   35955
Latin America and the Caribbean  15427  13678  15171  21179  …   24747
Northern America                7100   6661   6543   7074  …    8394

                                 2006    2007    2008    2009    2010  \
AreaName
Africa                          29188   28284   29890   34534   40892
Asia                           149054  133459  139894  141434  163845
Europe                          33053   33495   34692   35078   33425
Latin America and the Caribbean  24676   26011   26547   26867   28818
Northern America                 9613    9463   10190    8995    8142

                                 2011    2012    2013    Total
AreaName
Africa                          35441   38083   38543   618948
Asia                           146894  152218  155075  3317794
Europe                          26778   29177   28691  1410947
Latin America and the Caribbean  27856   27173   24950   765148
Northern America                 7677    7892    8503   241142
```

```
[5 rows x 38 columns]
```



[49]:
```
##

colors = ['yellowgreen','#ff9999','gold', 'lightskyblue', 'pink','lightcoral']
explode = [0, 0.1, 0.2,0.3,0, 0] # only "explode" the 2nd slice
plt.title("Total Number of Immigration by Continents")

df_conti['Total'].plot( kind = 'pie',
                        figsize = (17,8),
                        autopct ='%1.1f%%',
                        startangle = 90,
                        shadow = True,
                        labels = None,
                        pctdistance = 1.12,
                        colors = colors,
                        explode = explode

                        )
```

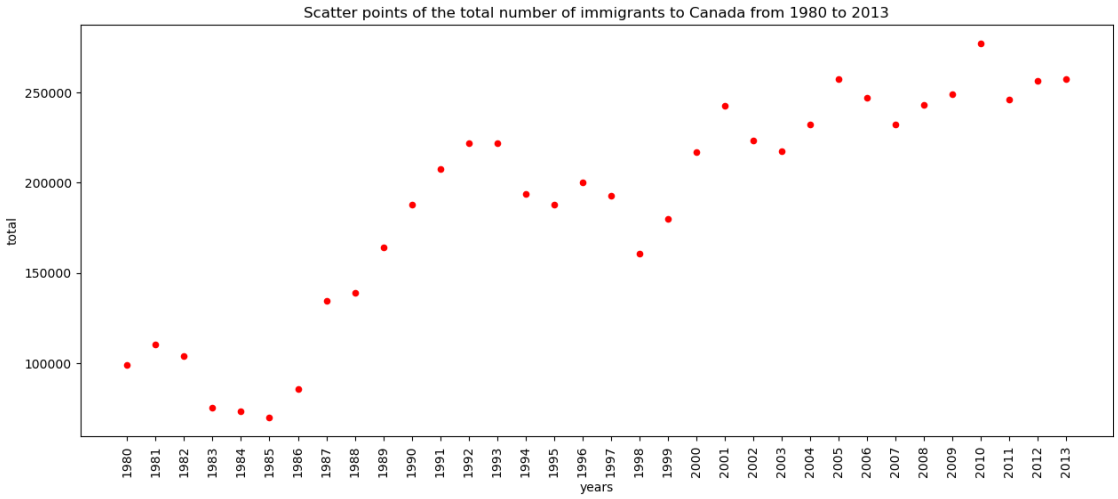
```
plt.title("Total Number of Immigration by Continents", y = 1.12)

plt.axis('equal')

plt.legend(labels = df_conti.index, loc = 'upper left')

plt.show()
```



[52]:
```
dftot = pd.DataFrame(dfcan[years].sum(axis=0))

dftot.head()

# print(dftot.head())

# Created index

dftot.reset_index(inplace=True)

#print(dftot.head())

# Rename columns

dftot.columns = ['years','total']
```

[53]:
```
# Ceated scatter
```

```
dftot.plot(kind = "scatter", x = "years", y = "total",figsize = (15,6), color =␣
 ↪"red")

plt.title("Scatter points of the total number of immigrants to Canada from 1980␣
 ↪to 2013")

plt.xlabel("years")
plt.xticks(rotation = 90)
plt.ylabel("total")

plt.show()
```



[ ]: