

**Data Visualization with Python**

**Lab Report**

**学号 2017007809**

**姓名 卫万灵**

**专 业 软件工程(信息与大数据工程方向)**

**授课老师 韩晓红**

**日 期 2019.5.**

### Line Pots(Series/Dataframe)

1. **Question: plot a line graph of immigration from Haiti using df.plot()**

haiti = df\_can.loc['Haiti', years]

haiti.index = haiti.index.map(int)

haiti.plot(kind='line')

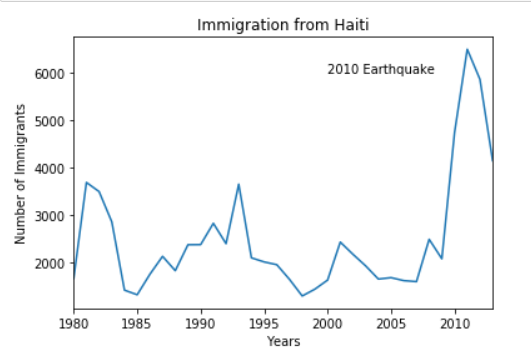
plt.title('Immigration from Haiti')

plt.ylabel('Number of Immigrants')

plt.xlabel('Years')

plt.text(2000, 6000, '2010 Earthquake')

plt.show()



**2. Question: Let’s compare the number of immigrants from india and China from 1980 to 2013**

df\_CI = df\_can.loc[['India', 'China'], years]

df\_CI = df\_CI.transpose()

df\_CI.index = df\_CI.index.map(int)

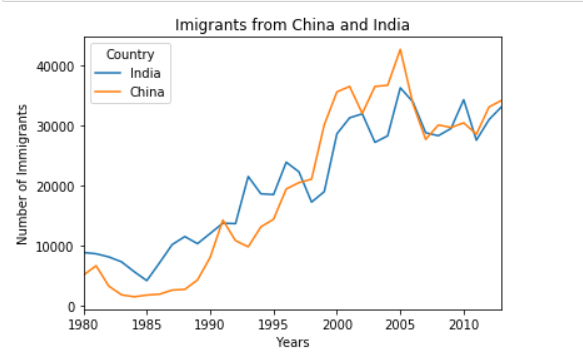
df\_CI.plot(kind='line')

plt.title('Imigrants from China and India')

plt.ylabel('Number of Immigrants')

plt.xlabel('Years')

plt.show()



1. **Question: Compare the trend of top 5 countries that contributed the most to immigration to Canada**

df\_can.sort\_values(['Total'],ascending=False,axis=0,inplace=True)

df\_top5 = df\_can.head(5)

df\_top5 = df\_top5[years].transpose()

df\_top5.index = df\_top5.index.map(int)

df\_top5.plot(kind='line', figsize=(14, 8))

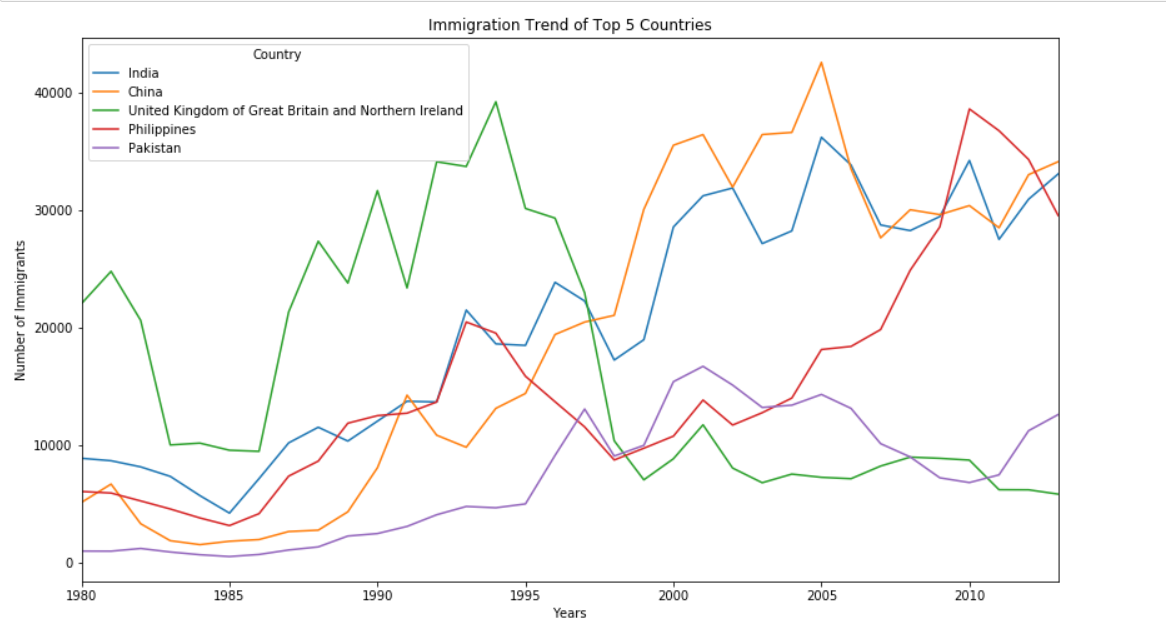
#print(df\_top5)

plt.title('Immigration Trend of Top 5 Countries')

plt.ylabel('Number of Immigrants')

plt.xlabel('Years')

plt.show()



# Area Plots, Histograms, and Bar Plots

## Area Plots

1. **Question: Use the scripting layer to create a stacked area plot of the 5 countries that contributed the least to immigration to Canada from 1980 to 2013.Use a transparency value of 0.45**

df\_least5=df\_can.tail(5)

df\_least5=df\_least5[years].transpose()

#print(df\_least5)

df\_least5.index=df\_least5.index.map(int)

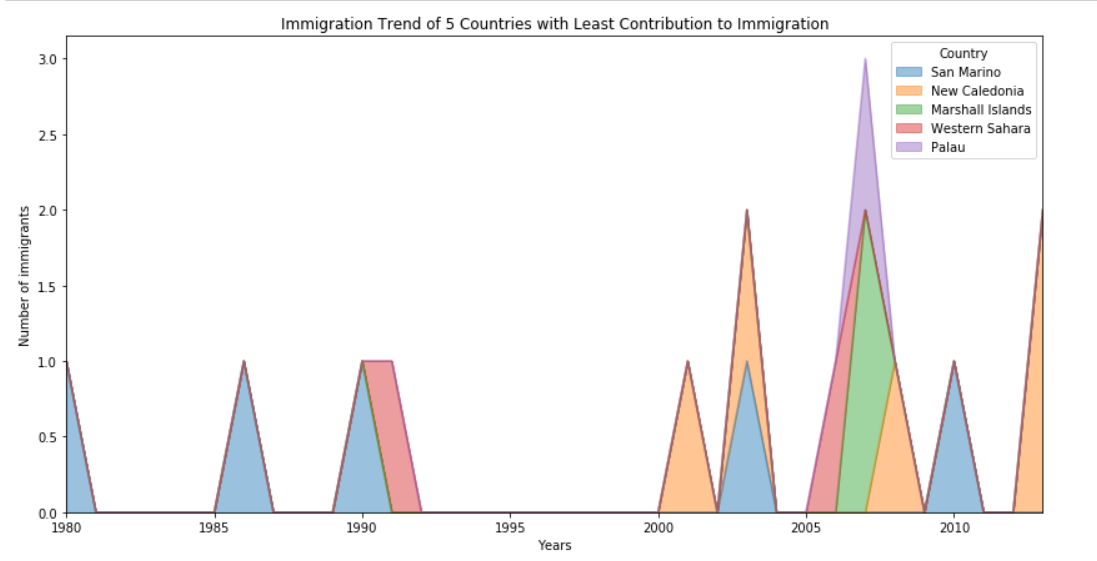
df\_least5.plot(kind='area', alpha=0.45, figsize=(14,7),stacked=True)

plt.title('Immigration Trend of 5 Countries with Least Contribution to Immigration')

plt.ylabel('Number of immigrants')

plt.xlabel('Years')

plt.show()



1. **Question: Use the artist layer to create an unstacked area plot of the 5 countries that contributed the least to immigration to Canada from 1980 to 2013 Use a transparency value of 0.55**

df\_least5 = df\_can.tail(5)

df\_least5 = df\_least5[years].transpose()

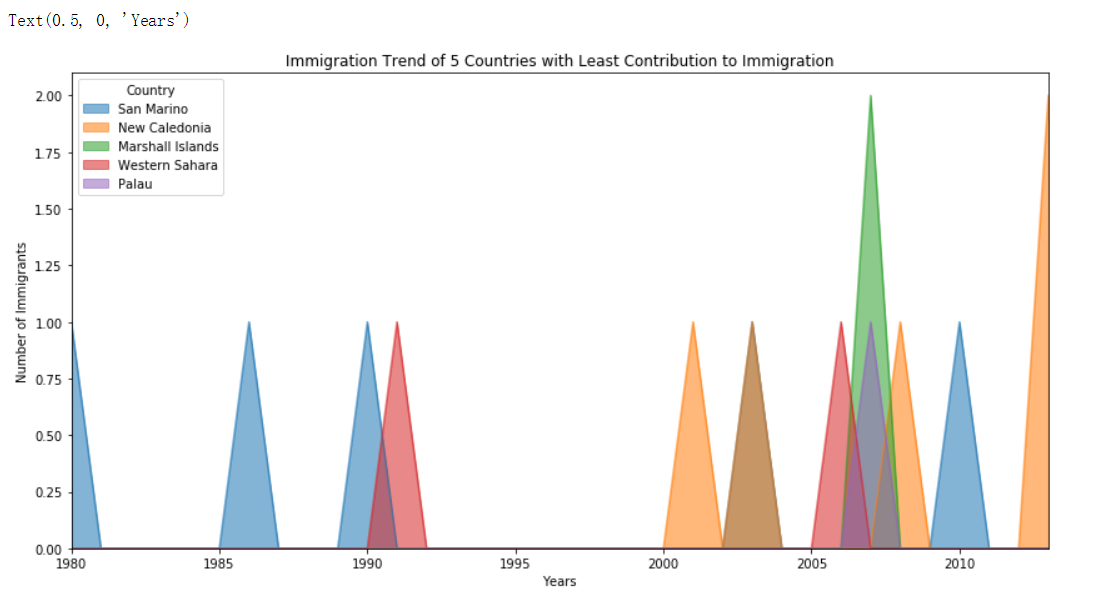
#print(df\_least5.head())

ax = df\_least5.plot(kind='area', alpha=0.55, figsize=(14,7 ),stacked=False)

ax.set\_title('Immigration Trend of 5 Countries with Least Contribution to Immigration')

ax.set\_ylabel('Number of Immigrants')

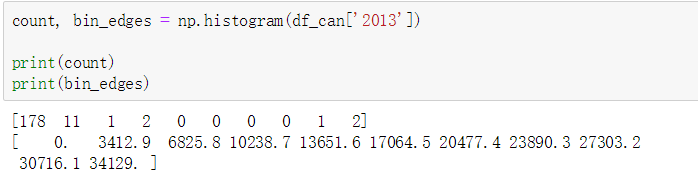
ax.set\_xlabel('Years')



## Histograms

**1. Question: What is the frequency distribution of the number(population) of new immigrants from the various countries to Canada in 2013?**

**Before we proceed with creating the histogram plot let’s first examine the data split into intervals ,To do this we will use numpy’s histogram method to get the bin ranges and frequency counts as follows**



**2. Question: What’s the immigration distribution for Denmark. Norway and Sweden for years 1980-2013**

df\_dns\_t = df\_can.loc[['Denmark', 'Norway', 'Sweden'], years].transpose()

count, bin\_edges = np.histogram(df\_dns\_t, 15)

df\_dns\_t.plot(kind ='hist',

figsize=(10, 6),

bins=15,

alpha=0.6,

xticks=bin\_edges,

color=['coral', 'darkslateblue', 'mediumseagreen']

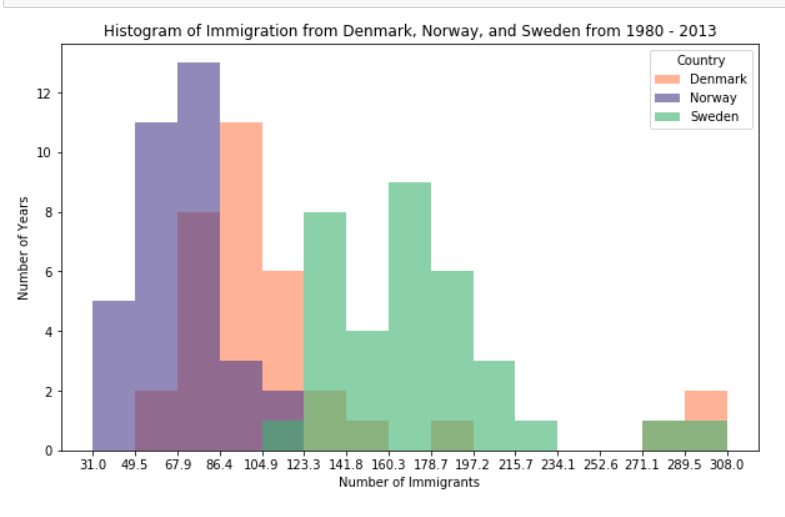
)

plt.title('Histogram of Immigration from Denmark, Norway, and Sweden from 1980 - 2013')

plt.ylabel('Number of Years')

plt.xlabel('Number of Immigrants')

plt.show()



**3. Question: Use the scripting layer to display the immigration for Greece Albania and Bulgaria for years 1980-2013? Use an overlapping plot with 15 bins and a transparency value of 0.35**

df\_gab = df\_can.loc[['Greece', 'Albania', 'Bulgaria'], years]

df\_gab = df\_gab.transpose()

count, bin\_edges = np.histogram(df\_gab, 15)

df\_gab.plot(kind ='hist',

figsize=(10, 6),

bins=15,

alpha=0.35,

xticks=bin\_edges

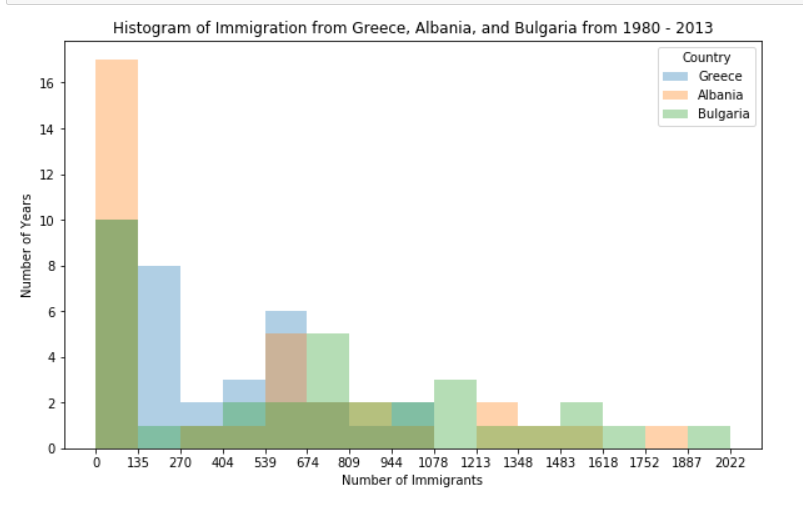
)

plt.title('Histogram of Immigration from Greece, Albania, and Bulgaria from 1980 - 2013')

plt.ylabel('Number of Years')

plt.xlabel('Number of Immigrants')

plt.show()



## Bar Chart

**1. Question: Let’s compare the number of lcelandic immigrants(country=lceland)to Canada from year 1980 to 2013**

df\_iceland = df\_can.loc['Iceland', years]

df\_iceland.plot(kind='bar', figsize=(10, 6),rot=90)

plt.xlabel('Year')

plt.ylabel('Number of Immigrants')

plt.title('Icelandic Immigrants to Canada from 1980 to 2013')

plt.annotate('',

xy=(32, 70), #head

xytext=(28, 20), #base

xycoords='data',

arrowprops=dict(arrowstyle='->', connectionstyle='arc3', color='blue', lw=2)

)

plt.annotate('2008 - 2011 Financial Crisis',

xy=(28, 30), #text start

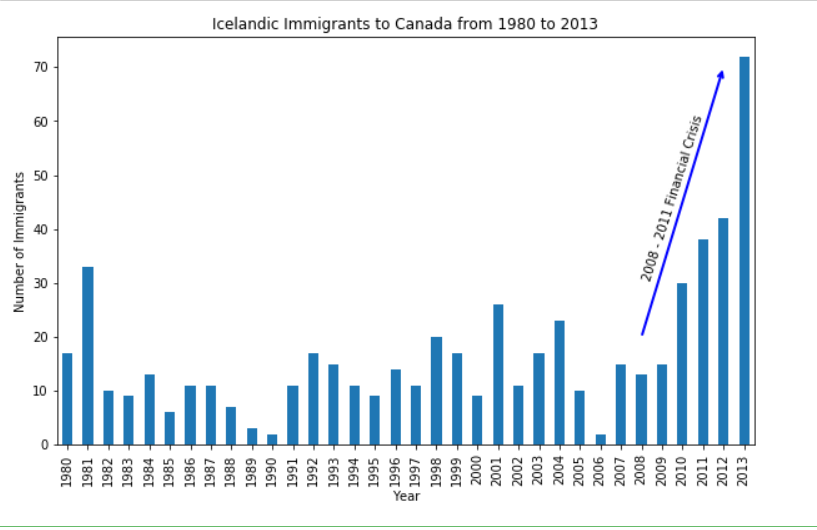
rotation=72.5,

va='bottom',

ha='left',

)

plt.show()



### Horizontal Bar Plot

**1. Question: Using the scripting layer and the df\_can dataset create a horizontal bar plot showing the total number of immigrants to Canada from the top 15 countries for the period 1980-2013 label each country with the total immigrant count**

df\_can.sort\_values(['Total'],ascending=True,inplace=True,axis=0)

df\_top15=df\_can.tail(15)

df\_top15=df\_top15['Total']

df\_top15.plot(kind='barh',figsize=(12,12),color='steelblue')

plt.xlabel('Number of Immigrants')

plt.title('Top 15 Conuntries Contributing to the Immigration to Canada between 1980 - 2013')

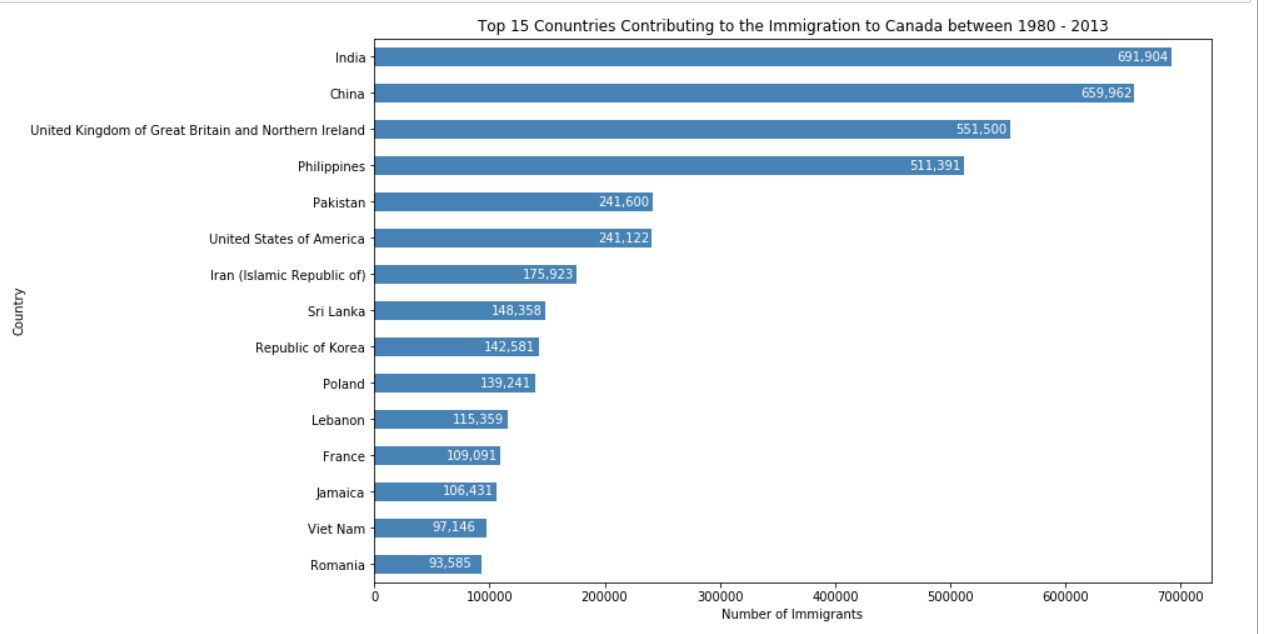
for index,value in enumerate(df\_top15):

#print(index,' ',value)

label=format(value,',')

plt.annotate(label,xy=(value-47000,index-0.1),color='white')

plt.show()



# Pie Charts Box Plots Scatter Plots and Bubble Plots

## Pie charts

1. **Question: using a pie chart explore the proportion(percentage)of new immigrants grouped by continents in the year 2013**

df\_continents = df\_can.groupby('Continent', axis=0).sum()

explode\_list = [0.1, 0, 0, 0, 0.1, 0.2]

#print(df\_continents['2013'])

df\_continents['2013'].plot(kind='pie',

figsize=(15, 6),

autopct='%1.1f%%',

startangle=90,

shadow=True,

labels=None,

pctdistance=1.12,

explode=explode\_list

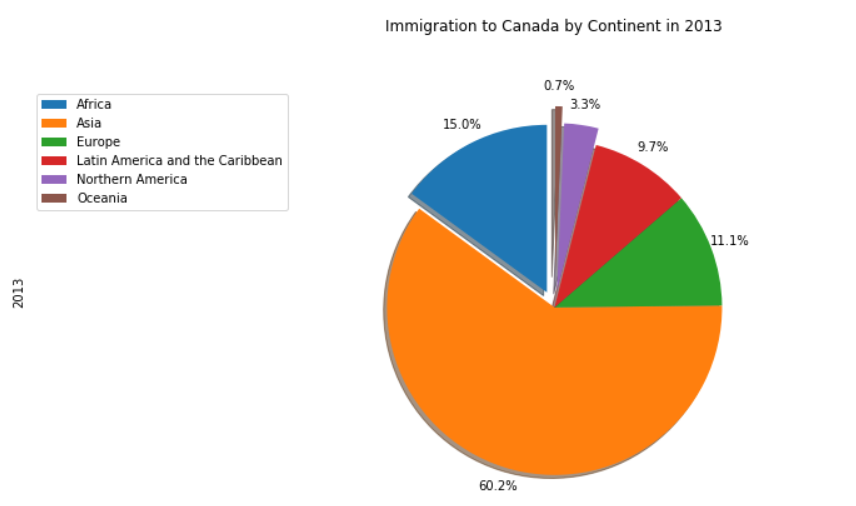
)

plt.title('Immigration to Canada by Continent in 2013', y=1.12)

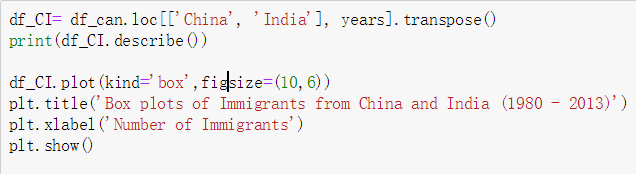
plt.axis('equal')

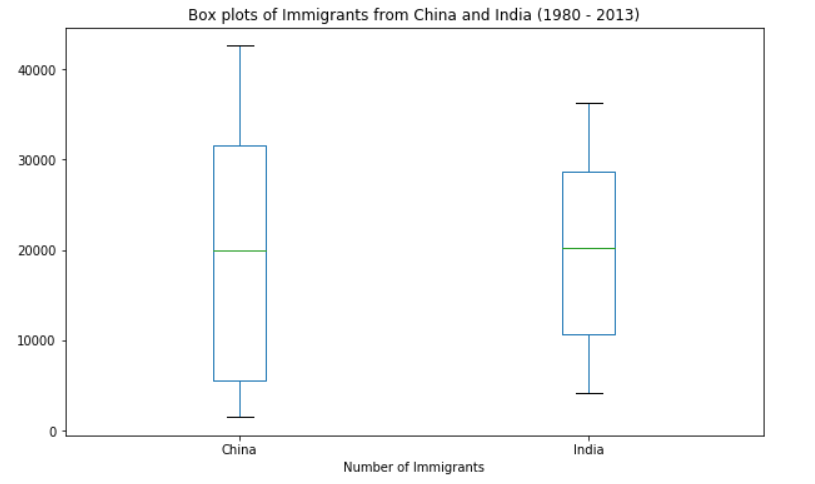
plt.legend(labels=df\_continents.index, loc='upper left')

plt.show()



**2. Question: Compare the distribution of the number of new immigrants from india and china for the period 1980-2013**





**3. Question: create a box plot to visualize the distribution of the top 15 countries (based on total immigration) grouped by the decades 1980s and 2000s**

df\_top15=df\_can.sort\_values(['Total'],ascending=False,axis=0).head(15)

years\_80s = list(map(str, range(1980, 1990)))

years\_90s = list(map(str, range(1990, 2000)))

years\_00s = list(map(str, range(2000, 2010)))

df\_80s = df\_top15.loc[:, years\_80s].sum(axis=1)

df\_90s = df\_top15.loc[:, years\_90s].sum(axis=1)

df\_00s = df\_top15.loc[:, years\_00s].sum(axis=1)

new\_df = pd.DataFrame({'1980s': df\_80s, '1990s': df\_90s, '2000s':df\_00s})

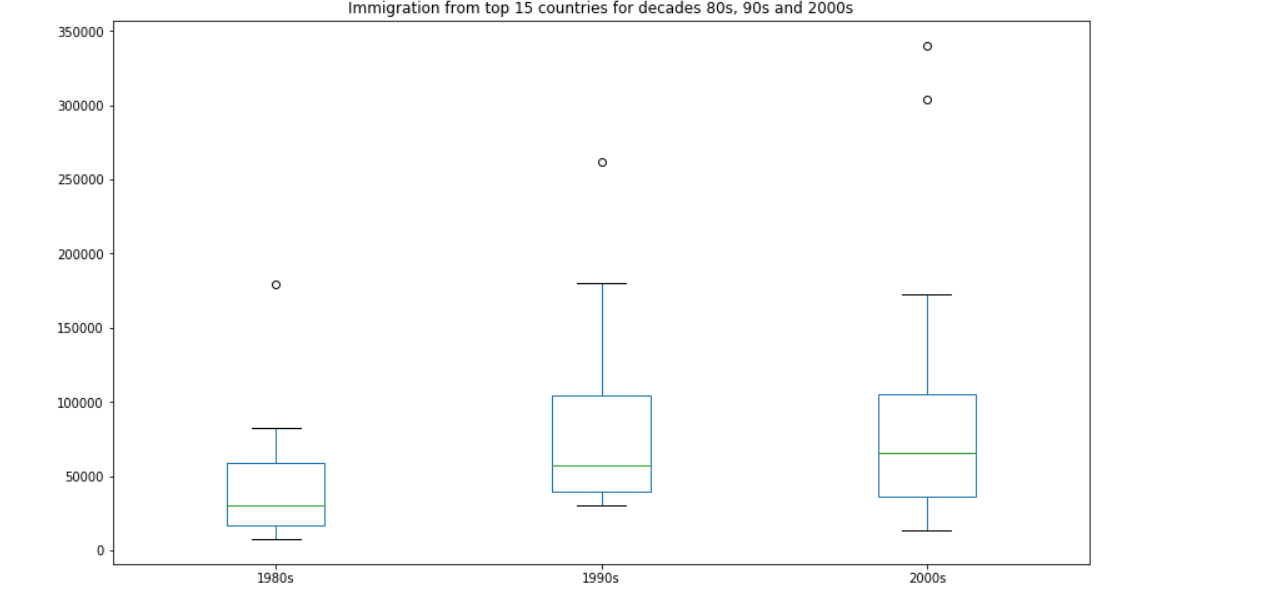
print(new\_df.head())

print(new\_df.describe())

new\_df.plot(kind='box',figsize=(14,8))

plt.title('Immigration from top 15 countries for decades 80s, 90s and 2000s')

plt.show()



## Scatter plots

**1. Question: create a scatter plot of the total immigration from Denmark Norway and Sweden to Canada from 1980 to 2013**

df\_countries = df\_can.loc[['Denmark', 'Norway', 'Sweden'], years].transpose()

print('\*'\*30,df\_countries,sep='\n')

df\_total = pd.DataFrame(df\_countries.sum(axis=1))

print('\*'\*30,df\_total,sep='\n')

df\_total.reset\_index(inplace=True)

print('\*'\*30,df\_total,sep='\n')

df\_total.columns = ['year', 'total']

print('\*'\*30,df\_total,sep='\n')

df\_total['year'] = df\_total['year'].astype(int)

print('\*'\*30,df\_total,'\*'\*30,sep='\n')

print(df\_total.head())

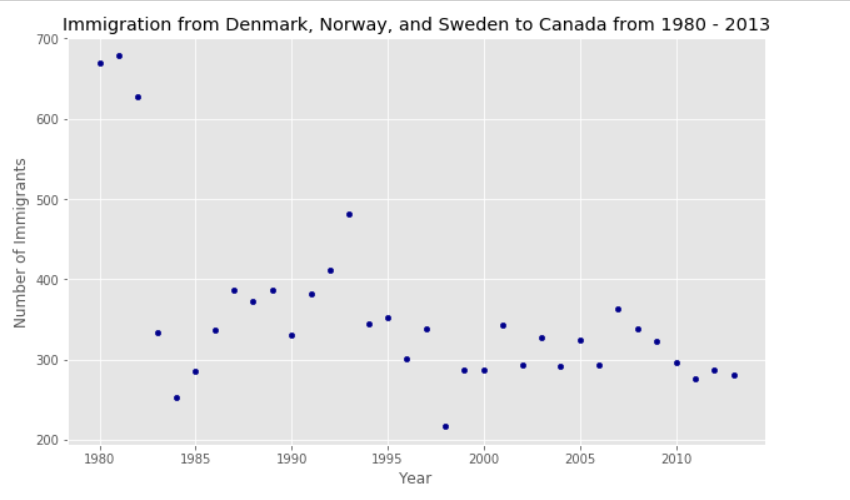
df\_total.plot(kind='scatter', x='year', y='total', figsize=(10, 6), color='darkblue')

plt.title('Immigration from Denmark, Norway, and Sweden to Canada from 1980 - 2013')

plt.xlabel('Year')

plt.ylabel('Number of Immigrants')

plt.show()



## Bubble plots

1. **Question: previously in this lab we created box plots to compare immigration from China and India to Canada create bubble plots of immigration from china and india to visualize any differences with time from 1980 to 2013 you can use df-can-t that we define and used in the previous example**

df\_can\_t = df\_can[years].transpose()

df\_can\_t.index = map(int, df\_can\_t.index)

df\_can\_t.index.name = 'Year'

df\_can\_t.reset\_index(inplace=True)

norm\_china = (df\_can\_t['China'] - df\_can\_t['China'].min()) / (df\_can\_t['China'].max() - df\_can\_t['China'].min())

norm\_india = (df\_can\_t['India'] - df\_can\_t['India'].min()) / (df\_can\_t['India'].max() - df\_can\_t['India'].min())

ax0 = df\_can\_t.plot(kind='scatter',

x='Year',

y='China',

figsize=(14, 8),

alpha=0.5, # transparency

color='green',

s=norm\_china \* 2000 + 10, # pass in weights

xlim=(1975, 2015)

)

# India

ax1 = df\_can\_t.plot(kind='scatter',

x='Year',

y='India',

alpha=0.5,

color="blue",

s=norm\_india \* 2000 + 10,

ax = ax0

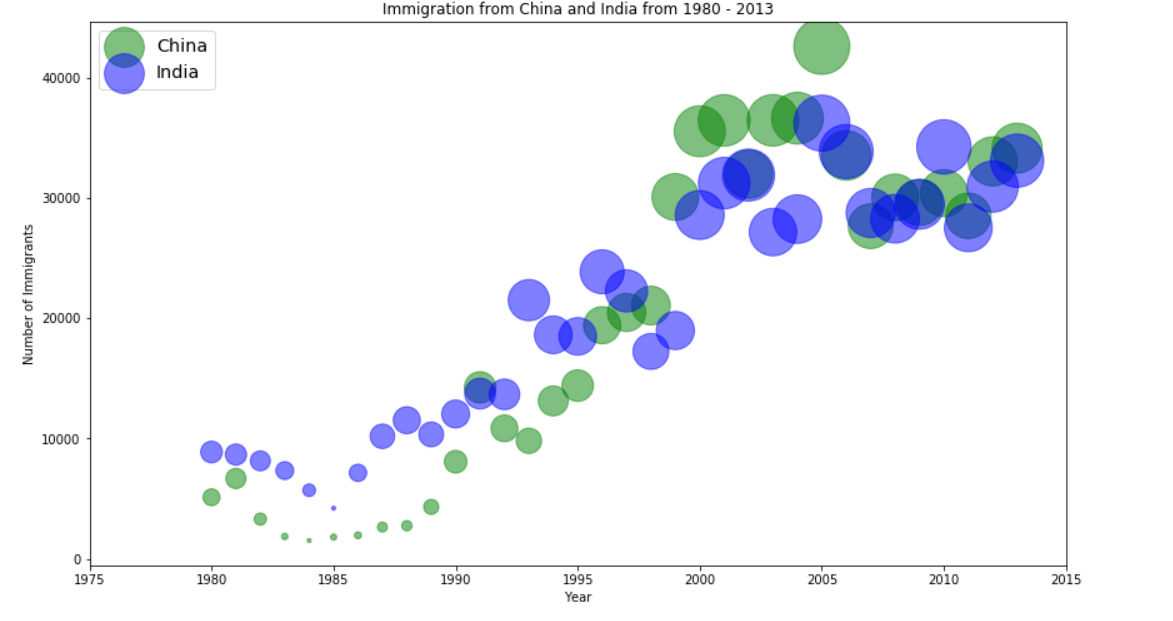
)

ax0.set\_ylabel('Number of Immigrants')

ax0.set\_title('Immigration from China and India from 1980 - 2013')

ax0.legend(['China', 'India'], loc='upper left', fontsize='x-large')

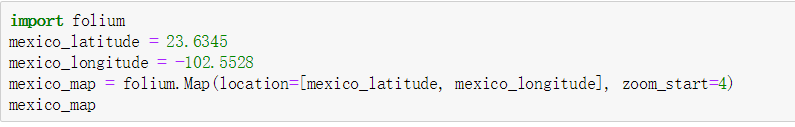
plt.show()

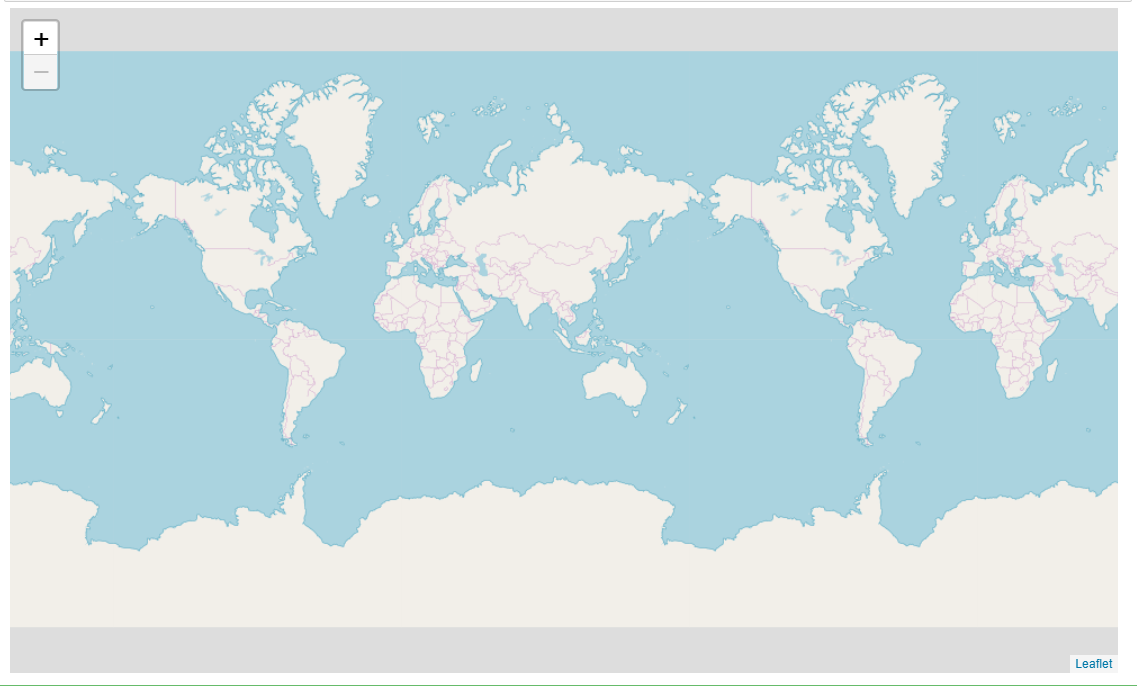


# Generating maps with python

## Introduction to folium

**1. Question: create a map of Mexico with a zoom level of 4**





### Map box bright maps

**1. Question: create a map of Mexico to visualize its hill shading and natural vegetation use a zoom level of 6**

