UCSanDiegoX: DSE200x
Python for Data Science
Week 6: Mini Project

Does the aging population drag down the GDP growth?

Y. YANG
Oct. 2019

Dataset

- World Development Indicators Dataset

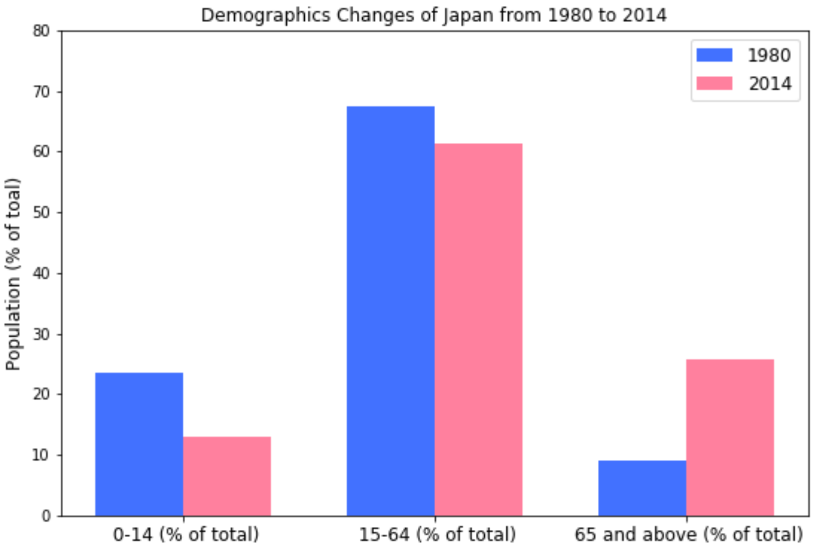
Research Question

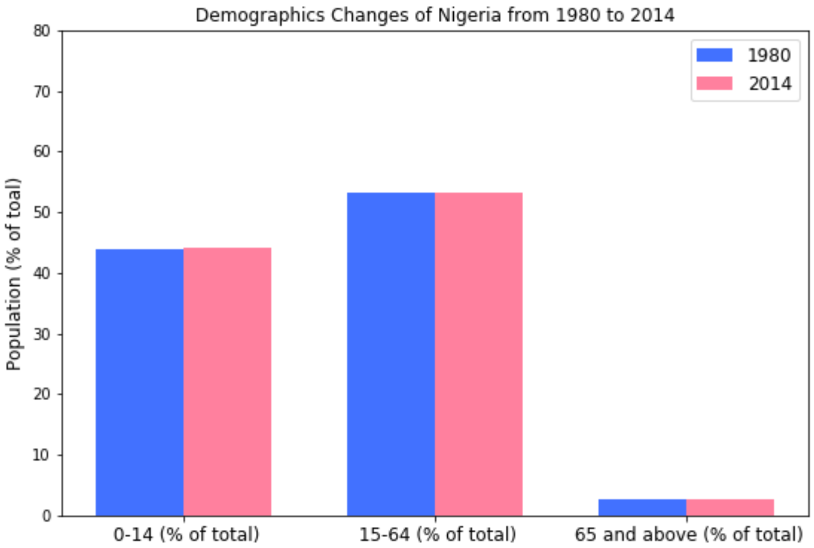
- Does the aging population drag down the GDP growth?

Steps

In order to answer this question effectively, I picked up two representative countries considering their dramatic demographic change across the time – Japan and Nigeria.
I will firstly have a look on how the demographic distribution and the annual GDP growth of these two countries change from 1980 to 2014. Specifically, I will compare the elderly people population of Japan with that of Nigeria, to show how different their societies are.
Finally, I will evaluate the correlation between the number of seniors and the growth of economy and draw a conclusion.

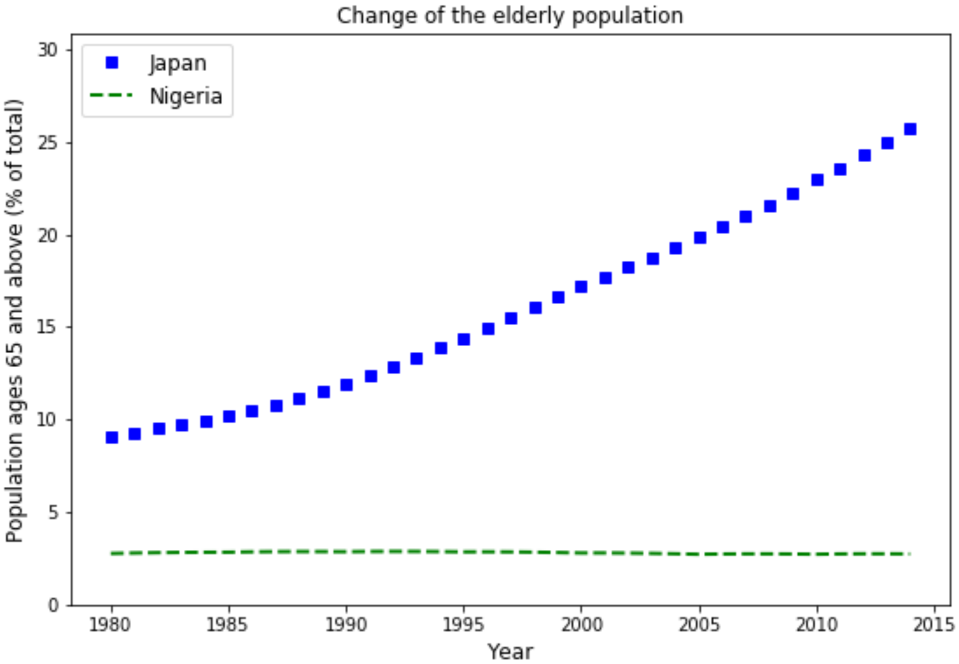
Finding 1-1: Demographic changes





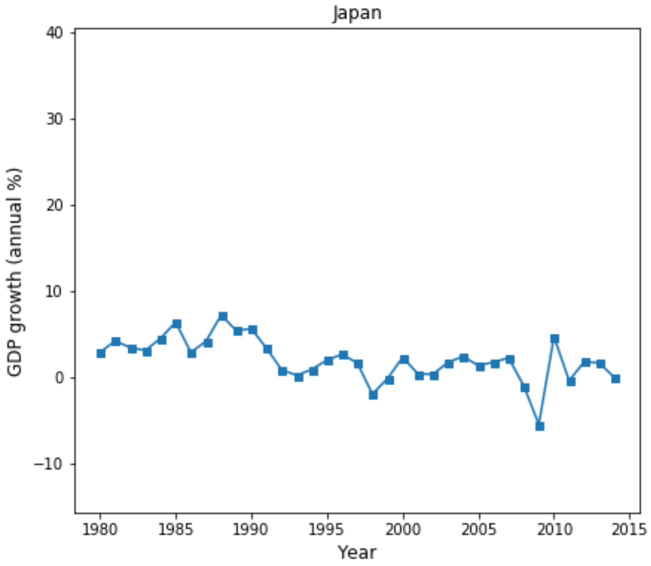
Japan saw a dramatic increase of elderly population as it accounted for a quarter of its total population by the end of 2014, while the constitution of Nigeria's population remained almost the same as what it was 25 years ago.

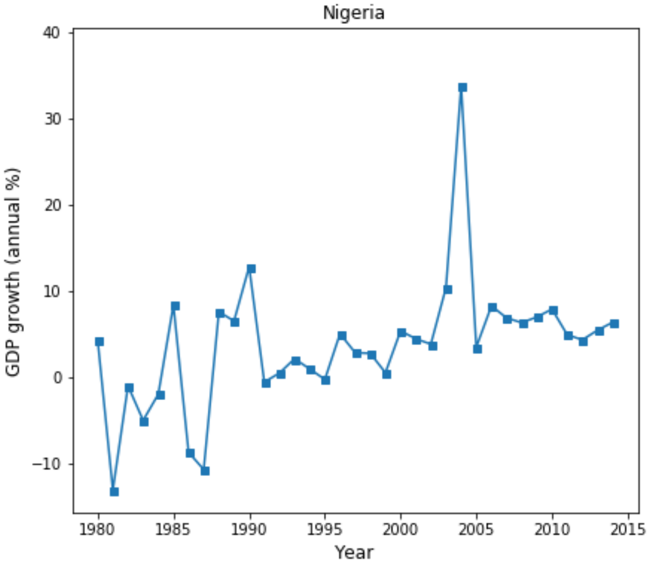
Finding 1-2: The elderly population changes



People aged 65 and older in Japan reached a quarter of its total population, up more than 15% 25 years ago. Japan had a post-war baby boom but followed by a prolonged period of low fertility.

Finding 2: Annual GDP growth

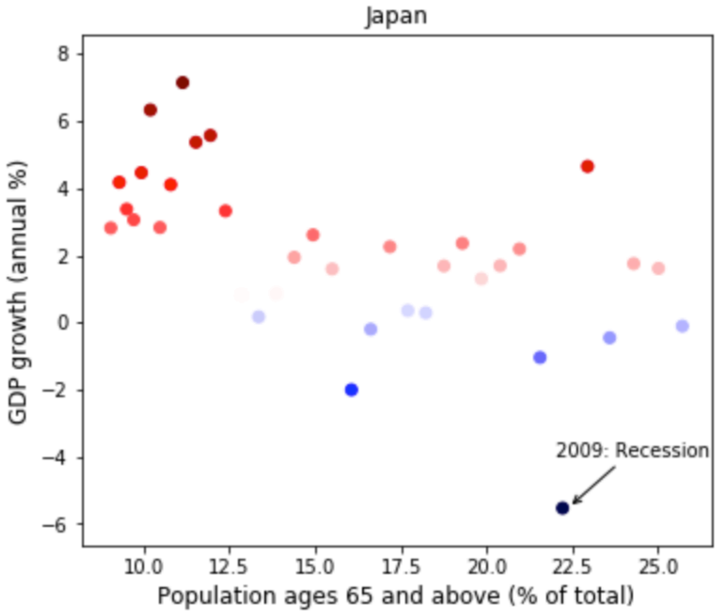


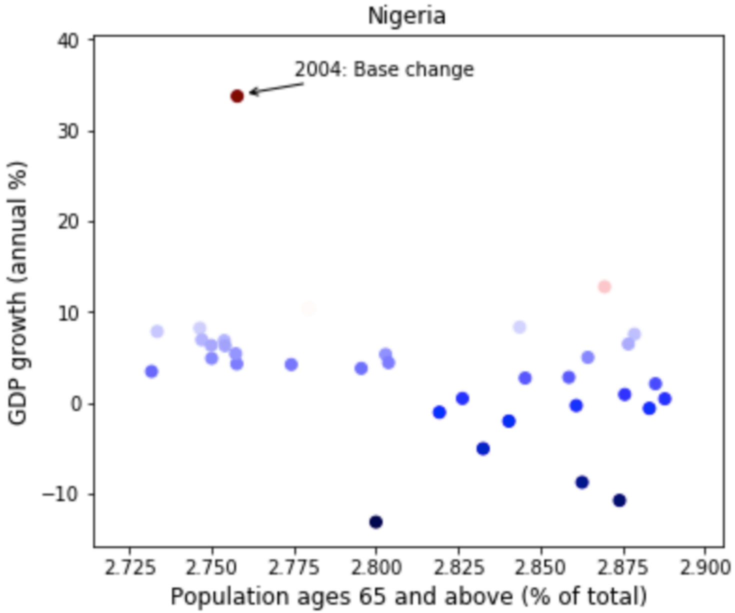


Year 2009 in Japan: The Japanese economy has been hit hard by the slump in global demand for advanced manufacturing products such as cars since the start of the 2008 Great Recession.

Year 2004 in Nigeria: Nigeria left the base year used to evaluate the real GDP far too long; as a result, its old GDP figures were inaccurate. In 2004, the new figures used 2010 as the base year.

Finding 3: Aging population and GDP growth





Correlation	Japan	Nigeria
Full data	-0.554	-0.365
Withdraw the outlier	-0.535	-0.560

Conclusion

The calculated result of Japan shows there is a strong ($-1 < r < -0.5$) inverse linear relationship between the two variables - the aging population and the GDP growth. In Nigeria, this potential correlation becomes solid after withdrawing the outlier year.

We may conclude that the growing elderly population is likely to create economic slowdown in either Japan or Nigeria. In addition, while high population growth in Nigeria may offset this negative effect and boost its GDP at some level. Broadly speaking, Japan may represent developed countries and Nigeria stands for the developing. But to support this general claim, more data explorations are needed.

References

[1] Kelley, Allen C., and Robert M. Schmidt. "Aggregate population and economic growth correlations: the role of the components of demographic change." *Demography* 32.4 (1995): 543-555.

Question: How does the changing population affect the economy?

I will be using "The World Development Indicators" dataset obtained from the World Bank containing over a thousand annual indicators of economic development from hundreds of countries around the world.
Specifically, I will study "how does the changing population affect the economy" by comparing two representative countries in demographic change - Japan and Nigeria, through 1980 - 2015.

In [1]:
```
import pandas as pd
import numpy as np
import random
import matplotlib.pyplot as plt
```

Data Exploration and Cleaning

In [2]:
```
data = pd.read_csv('/Users/yang/Downloads/world-development-indicators/Indicators.csv')
data.head()
```

Out[2]:

	CountryName	CountryCode	IndicatorName	IndicatorCode	Year	Value
0	Arab World	ARB	Adolescent fertility rate (births per 1,000 wo...	SP.ADO.TFRT	1960	1.335609e+02
1	Arab World	ARB	Age dependency ratio (% of working-age populat...	SP.POP.DPND	1960	8.779760e+01
2	Arab World	ARB	Age dependency ratio, old (% of working-age po...	SP.POP.DPND.OL	1960	6.634579e+00
3	Arab World	ARB	Age dependency ratio, young (% of working-age ...	SP.POP.DPND.YG	1960	8.102333e+01
4	Arab World	ARB	Arms exports (SIPRI trend indicator values)	MS.MIL.XPRT.KD	1960	3.000000e+06

Investigate the time period of the data:

In [3]:
```
print (data.Year.max())
print (data.Year.min())
```

```
2015
1960
```

How does the demographics distribution change from 1980 to 2014?

In [4]:
```
mask_all = data.IndicatorName.str.contains('Population, ages 0-14|Population, ages 15-64|Po
pulation ages 65 and above')
mask_y1= data.Year == 1980
mask_y2= data.Year == 2014
mask_A= data.CountryName == 'Japan'
```

In [5]:
```
demo_A1= data[mask_all & mask_y1 & mask_A][['CountryName','IndicatorName','Year','Value']]
```

In [6]:
```
def extract_AgeNo(df):
    new = df['IndicatorName'].str.split('ages', expand=True)
    df['Age Index']=new[1]
    return df
```

In [7]:
```
def sort_indexAge(df):
    df['sort'] = df['IndicatorName'].str.extract('(\d+)',expand=False).astype(int)
    df.sort_values('sort',inplace=True, ascending=True)
    df = df.drop('sort', axis=1)
    return df
```

In [8]:
```
demo_A1=sort_indexAge(extract_AgeNo(demo_A1))
demo_A1.head()
```

Out[8]:

	CountryName	IndicatorName	Year	Value	Age Index
1055266	Japan	Population, ages 0-14 (% of total)	1980	23.561734	0-14 (% of total)
1055267	Japan	Population, ages 15-64 (% of total)	1980	67.392501	15-64 (% of total)
1055259	Japan	Population ages 65 and above (% of total)	1980	9.045765	65 and above (% of total)

In [9]:
```
demo_A2= data[mask_all & mask_y2 & mask_A][['CountryName','IndicatorName','Year','Value']]
demo_A2=sort_indexAge(extract_AgeNo(demo_A2))
demo_A2.head()
```

Out[9]:

	CountryName	IndicatorName	Year	Value	Age Index
5589433	Japan	Population, ages 0-14 (% of total)	2014	12.939212	0-14 (% of total)
5589434	Japan	Population, ages 15-64 (% of total)	2014	61.355365	15-64 (% of total)
5589426	Japan	Population ages 65 and above (% of total)	2014	25.705422	65 and above (% of total)

In [10]:
```
mask_B= data.CountryName == 'Nigeria'
```

In [11]:
```
demo_B1= data[mask_all & mask_y1 & mask_B][['CountryName','IndicatorName','Year','Value']]
demo_B1=sort_indexAge(extract_AgeNo(demo_B1))
demo_B1.head()
```

Out[11]:

	CountryName	IndicatorName	Year	Value	Age Index
1069692	Nigeria	Population, ages 0-14 (% of total)	1980	43.978688	0-14 (% of total)
1069693	Nigeria	Population, ages 15-64 (% of total)	1980	53.247057	15-64 (% of total)
1069685	Nigeria	Population ages 65 and above (% of total)	1980	2.774255	65 and above (% of total)

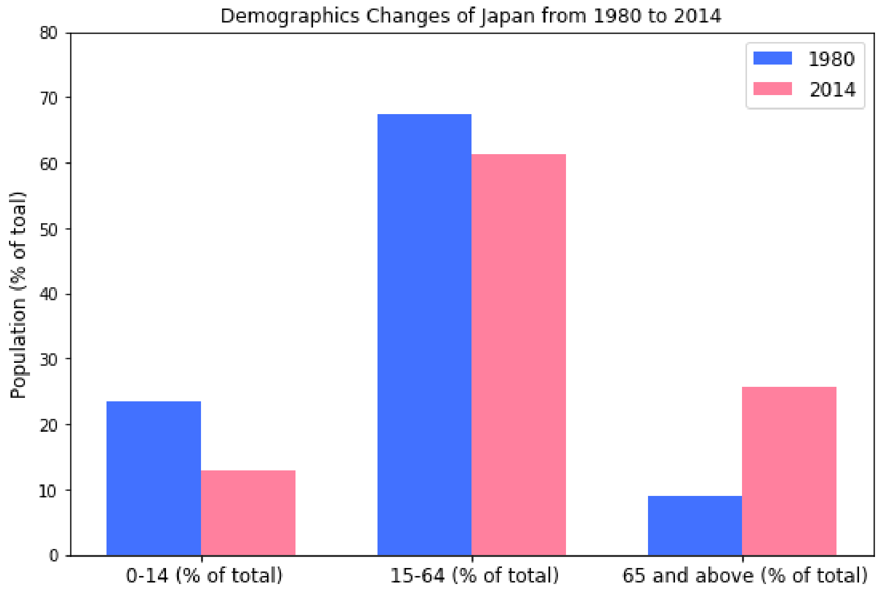
In [12]:
```
demo_B2= data[mask_all & mask_y2 & mask_B][['CountryName','IndicatorName','Year','Value']]
demo_B2=sort_indexAge(extract_AgeNo(demo_B2))
demo_B2.head()
```

Out[12]:

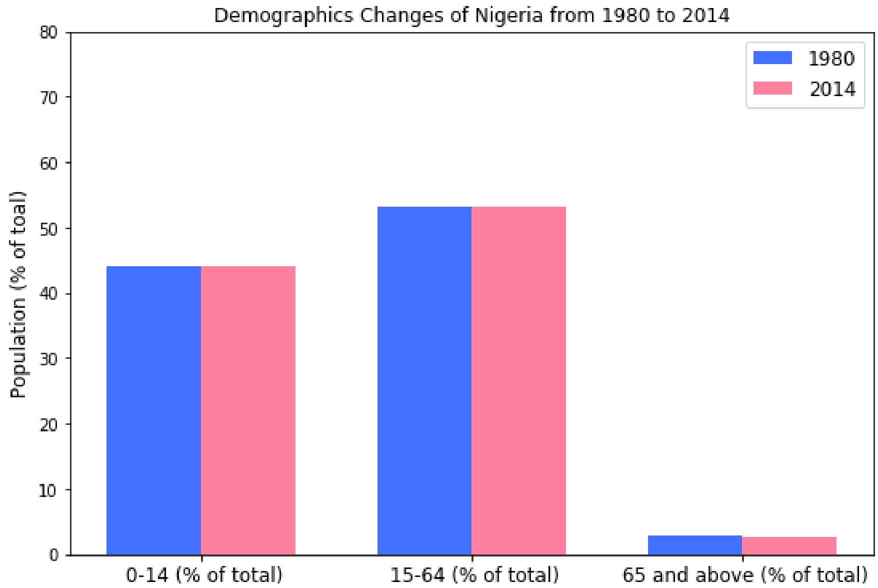
	CountryName	IndicatorName	Year	Value	Age Index
5610777	Nigeria	Population, ages 0-14 (% of total)	2014	44.071179	0-14 (% of total)
5610778	Nigeria	Population, ages 15-64 (% of total)	2014	53.178905	15-64 (% of total)
5610770	Nigeria	Population ages 65 and above (% of total)	2014	2.749916	65 and above (% of total)

In [13]:
```
def demo_bar(df1, df2):
    assert len(df1) == len(df2)
    N = len(df1)
    ind = np.arange(N)
    width = 0.35

    fig, ax = plt.subplots(figsize=(9, 6))
    ax.bar(ind, df1.Value, width, label = df1.Year.iloc[0], color=(0.2588,0.4433,1.0))
    ax.bar(ind+width, df2.Value, width, label = df2.Year.iloc[0], color=(1.0,0.5,0.62))
    ax.set_ylabel('Population (% of toal)', fontsize=12)
    ax.set_ylim(0, 80)
    ax.set_xticks(ind + width / 2)
    ax.set_xticklabels(df1['Age Index'].values, fontsize=12)
    ax.set_title('Demographics Changes of '+ df1.CountryName.iloc[0] +
                 ' from '+ str(df1.Year.iloc[0]) + ' to ' + str(df2.Year.iloc[0]), fontsize
=12)

    ax.legend(loc = 'best', fontsize='large', fancybox=True)
    fig.savefig('/Users/yang/Desktop/graph.png')
    return plt.show()
```

In [14]:
```
demo_bar(demo_A1, demo_A2)
```



In [15]:
```
demo_bar(demo_B1, demo_B2)
```



How about the GDP growth of each country in this period?

In [16]:
```
mask_G = data.IndicatorName.str.contains('GDP grow')
mask_y= data.Year >= 1980
GDP_A= data[mask_G & mask_y & mask_A][['CountryName','IndicatorName','Year','Value']]
GDP_A.head()
```

Out[16]:

	CountryName	IndicatorName	Year	Value
1055047	Japan	GDP growth (annual %)	1980	2.817591
1138191	Japan	GDP growth (annual %)	1981	4.176844
1222888	Japan	GDP growth (annual %)	1982	3.376608
1308181	Japan	GDP growth (annual %)	1983	3.060738
1393773	Japan	GDP growth (annual %)	1984	4.463899

In [17]:
```python
GDP_B= data[mask_G & mask_y & mask_B][['CountryName','IndicatorName','Year','Value']]
GDP_B.head()
```

Out[17]:

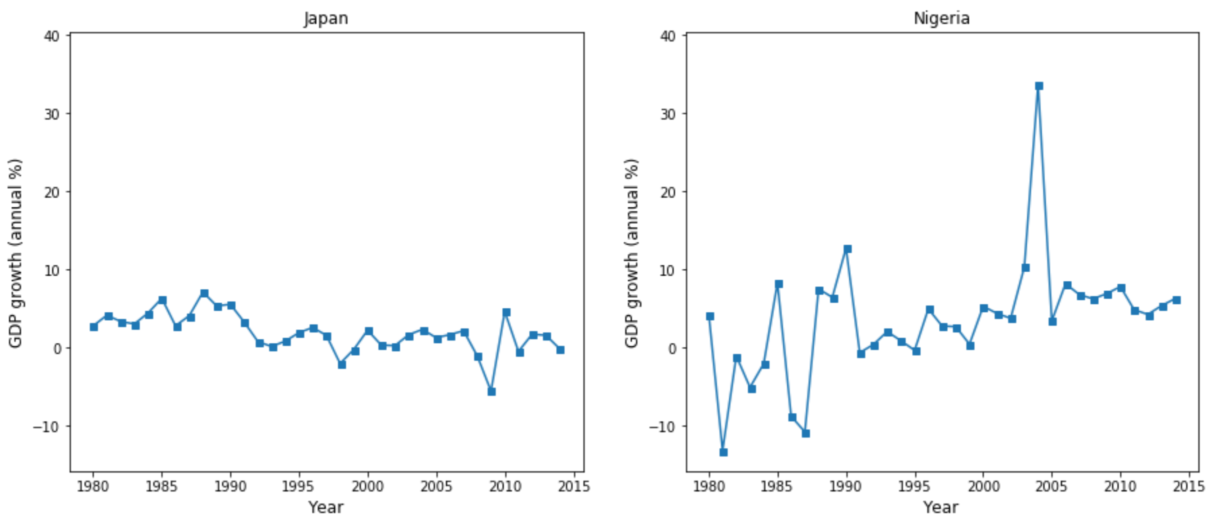
	CountryName	IndicatorName	Year	Value
1069527	Nigeria	GDP growth (annual %)	1980	4.204831
1152958	Nigeria	GDP growth (annual %)	1981	-13.127880
1237831	Nigeria	GDP growth (annual %)	1982	-1.053186
1323313	Nigeria	GDP growth (annual %)	1983	-5.050451
1409009	Nigeria	GDP growth (annual %)	1984	-2.021538

In [18]:
```python
for c in (GDP_A, GDP_B):
    print (c.CountryName.iloc[0], len(c))
```

```
Japan 35
Nigeria 35
```

In [19]:
```python
def Tosee(df1, df2, row, col):
    fig, ax = plt.subplots(nrows=row, ncols=col, figsize=(15, 6))
    MAX = max (df1.Value.max(),df2.Value.max())
    MIN = min (df1.Value.min(),df2.Value.min())
    for a, d in zip(ax,(df1, df2)):
        X = d.Year.values
        Y = d.Value.values
        a.set_ylim(1.2*MIN, 1.2*MAX)
        a.plot(X, Y, marker='s', markersize=5)
        a.set_xlabel('Year', fontsize=12)
        a.set_ylabel(d.IndicatorName.iloc[0], fontsize=12)
        a.set_title(d.CountryName.iloc[0], fontsize=12)
        fig.savefig('/Users/yang/Desktop/graph2.png')
    return plt.show()
```

In [20]:
```python
Tosee(GDP_A, GDP_B, 1, 2)
```



How does the elderly population change across this period?

In [21]:
```python
mask1 = data.IndicatorName.str.contains('Population ages 65 and above')
Old_A= data[mask1 & mask_y & mask_A][['CountryName','IndicatorName','Year','Value']]
Old_A.head()
```

Out[21]:

	CountryName	IndicatorName	Year	Value
1055259	Japan	Population ages 65 and above (% of total)	1980	9.045765
1138399	Japan	Population ages 65 and above (% of total)	1981	9.293358
1223101	Japan	Population ages 65 and above (% of total)	1982	9.508684
1308394	Japan	Population ages 65 and above (% of total)	1983	9.712745
1393986	Japan	Population ages 65 and above (% of total)	1984	9.936894

In [22]:
```python
Old_B= data[mask1 & mask_y & mask_B][['CountryName','IndicatorName','Year','Value']]
Old_B.head()
```

Out[22]:

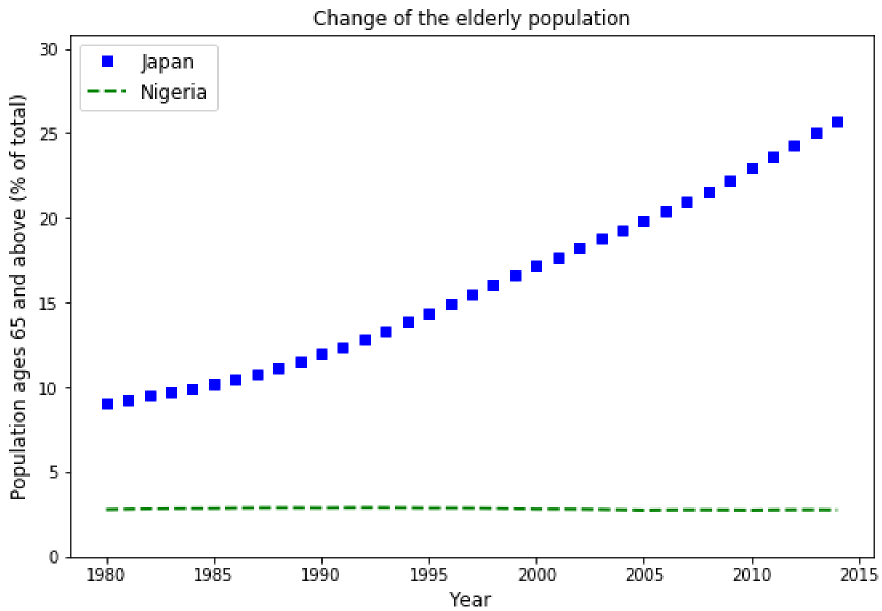
	CountryName	IndicatorName	Year	Value
1069685	Nigeria	Population ages 65 and above (% of total)	1980	2.774255
1153199	Nigeria	Population ages 65 and above (% of total)	1981	2.800028
1238081	Nigeria	Population ages 65 and above (% of total)	1982	2.819332
1323568	Nigeria	Population ages 65 and above (% of total)	1983	2.832640
1409262	Nigeria	Population ages 65 and above (% of total)	1984	2.840471

In [23]:
```python
for c in (Old_A, Old_B):
    print (c.CountryName.iloc[0], len(c))
```

```
Japan 35
Nigeria 35
```

In [24]:
```python
def Tosee_together(df1, df2):
    fig, ax = plt.subplots(figsize=(9, 6))
    MAX = max (df1.Value.max(),df2.Value.max())
    MIN = min (0,df1.Value.min(),df2.Value.min())
    ax.set_ylim(MIN, 1.2*MAX)
    ax.plot(df1.Year, df1.Value, 'bs', label = df1.CountryName.iloc[0])
    ax.plot(df2.Year, df2.Value, 'g--', linewidth=2, label = df2.CountryName.iloc[0])
    ax.set_xlabel('Year', fontsize=12)
    ax.set_ylabel(df1.IndicatorName.iloc[0], fontsize=12)
    ax.set_title('Change of the elderly population', fontsize=12)
    ax.legend(loc = 'upper left', fontsize='large', fancybox=True)
    fig.savefig('/Users/yang/Desktop/graph22.png')
    return plt.show()
```

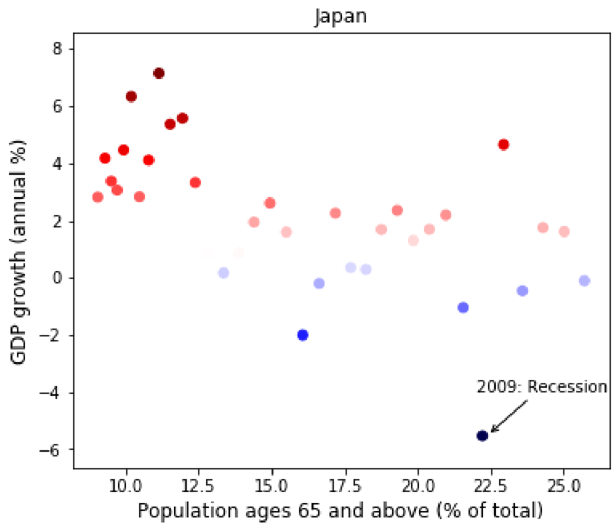
In [25]:
```python
Tosee_together(Old_A, Old_B)
```



Does the aging population of Japan drag down the GDP growth?

In [26]:
```
def Scat(df1, df2): # x, y
    fig, ax = plt.subplots(figsize=(6, 5))
    X = df1.Value.values
    Y = df2.Value.values
    #area = 40*df2.Value.values
    ax.set_ylim(1.2*df2.Value.min(), 1.2*df2.Value.max())
    ax.scatter(X, Y, c=Y, cmap=plt.cm.seismic)
    ax.set_alpha(1)
    ax.set_xlabel(df1.IndicatorName.iloc[0], fontsize=12)
    ax.set_ylabel(df2.IndicatorName.iloc[0], fontsize=12)
    ax.set_title(df1.CountryName.iloc[0], fontsize=12)

    fig.savefig('/Users/yang/Desktop/graph_cor.png')
    return ax
```

In [27]:
```
cor_J = Scat(Old_A, GDP_A)
cor_J.annotate("2009: Recession",
            xy=(22.4, -5.5), xycoords='data',
            xytext=(22, -4), textcoords='data',
            arrowprops=dict(arrowstyle="->",
                            connectionstyle="arc3"),
              )
plt.show()
```



In [28]:
```
np.corrcoef(Old_A['Value'], GDP_A['Value'])
```

Out[28]:
```
array([[ 1.        , -0.55388617],
       [-0.55388617,  1.        ]])
```

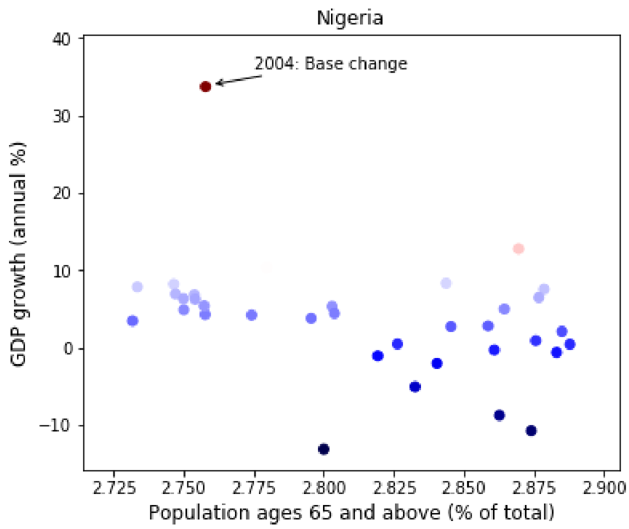
Background: The Japanese economy has been hit hard by the slump in global demand for advanced manufacturing products such as cars since the start of the 2008 Great Recession.
Let's truncate the outlier.

In [29]:
```
Old_A_trim = Old_A[Old_A.Year != 2009]
GDP_A_trim = GDP_A[GDP_A.Year != 2009]
assert len (Old_A_trim) == len (Old_A) -1 == len (GDP_A_trim) == len (GDP_A) -1
```

In [30]:
```
np.corrcoef(Old_A_trim['Value'], GDP_A_trim['Value'])
```

Out[30]:
```
array([[ 1.       , -0.5350806],
       [-0.5350806,  1.       ]])
```

In [31]:
```python
cor_N = Scat(Old_B, GDP_B)
cor_N.annotate("2004: Base change",
            xy=(2.76, 34), xycoords='data',
            xytext=(2.775, 36), textcoords='data',
            arrowprops=dict(arrowstyle="->",
                            connectionstyle="arc3"),
            )
plt.show()
```



In [32]:
```python
np.corrcoef(Old_B['Value'],GDP_B['Value'])
```

Out[32]:
```
array([[ 1.        , -0.36539494],
       [-0.36539494,  1.        ]])
```

Background: Nigeria left the base year used to evaluate the real GDP far too long; as a result, its old GDP figures were inaccurate. In 2004, the new figures use 2010 as the base year and thus the figure in 2004 upgraded so big. Let's see what will be after removing the outlier year.

In [33]:
```python
Old_B_trim = Old_A[Old_A.Year != 2004]
GDP_B_trim = GDP_A[GDP_A.Year != 2004]
assert len (Old_B_trim) == len (Old_B) -1 == len (GDP_B_trim) == len (GDP_B) -1
```

In [34]:
```python
np.corrcoef(Old_B_trim['Value'], GDP_B_trim['Value'])
```

Out[34]:
```
array([[ 1.        , -0.55988178],
       [-0.55988178,  1.        ]])
```

Conclusion

The calculated result of Japan shows there is a strong (-1 < r < -0.5) inverse linear relationship between the two variables - the aging population and the GDP growth. In Nigeria, this potential correlation becomes solid after withdrawing the outlier year. We may conclude that the growing elderly population is likely to create economic slowdown in either Japan or Nigeria. In addition, while high population growth in Nigeria may offset this negative effect and boost its GDP at some level. Broadly speaking, Japan may represent developed countries and Nigeria stands for the developing. But to support this general claim, more data explorations are needed.

In [ ]:
```python

```