第三章 实训

**实训 1 分析1996~2015年人口数据特征间的关系**

import numpy as np

import matplotlib.pyplot as plt

data=np.load('C:/Users/3SXL\_21144/Desktop/populations.npz',allow\_pickle=True)

print(data.files)

print(data['data'])

print(data['feature\_names'])

plt.rcParams['font.sans-serif']='SimHei'

name=data['feature\_names']

values=data['data']

p1=plt.figure(figsize=(12,12))

pip1=p1.add\_subplot(2,1,1)

plt.scatter(values[0:20,0],values[0:20,1],marker='8',color='red')

plt.ylabel('总人口（万人）')

plt.legend('年末')

plt.title('1996~2015年末与各类人口散点图')

pip2=p1.add\_subplot(2,1,2)

plt.scatter(values[0:20,0],values[0:20,2],marker='o',color='yellow')

plt.scatter(values[0:20,0],values[0:20,3],marker='D',color='green')

plt.scatter(values[0:20,0],values[0:20,4],marker='p',color='blue')

plt.scatter(values[0:20,0],values[0:20,5],marker='s',color='purple')

plt.xlabel('时间')

plt.ylabel('总人口（万人）')

plt.xticks(values[0:20,0])

plt.legend(['男性','女性','城镇','乡村'])

p2=plt.figure(figsize=(12,12))

p1=p2.add\_subplot(2,1,1)

plt.plot(values[0:20,0],values[0:20,1],color='r',linestyle='--',marker='8')

plt.ylabel('总人口（万人）')

plt.xticks(range(0,20,1),values[range(0,20,1),0],rotation=45)

plt.legend('年末')

plt.title('1996~2015年末总与各类人口折线图')

p2=p2.add\_subplot(2,1,2)

plt.plot(values[0:20,0],values[0:20,2],'y-')

plt.plot(values[0:20,0],values[0:20,3],'g-.')

plt.plot(values[0:20,0],values[0:20,4],'b-')

plt.plot(values[0:20,0],values[0:20,5],'p-')

plt.xlabel('时间')

plt.ylabel('总人口（万人）')

plt.xticks(values[0:20,0])

plt.legend(['男性','女性','城镇','乡村'])

plt.show()

答：

['data', 'feature\_names']

[['2015年' 137462.0 70414.0 67048.0 77116.0 60346.0]

['2014年' 136782.0 70079.0 66703.0 74916.0 61866.0]

['2013年' 136072.0 69728.0 66344.0 73111.0 62961.0]

['2012年' 135404.0 69395.0 66009.0 71182.0 64222.0]

['2011年' 134735.0 69068.0 65667.0 69079.0 65656.0]

['2010年' 134091.0 68748.0 65343.0 66978.0 67113.0]

['2009年' 133450.0 68647.0 64803.0 64512.0 68938.0]

['2008年' 132802.0 68357.0 64445.0 62403.0 70399.0]

['2007年' 132129.0 68048.0 64081.0 60633.0 71496.0]

['2006年' 131448.0 67728.0 63720.0 58288.0 73160.0]

['2005年' 130756.0 67375.0 63381.0 56212.0 74544.0]

['2004年' 129988.0 66976.0 63012.0 54283.0 75705.0]

['2003年' 129227.0 66556.0 62671.0 52376.0 76851.0]

['2002年' 128453.0 66115.0 62338.0 50212.0 78241.0]

['2001年' 127627.0 65672.0 61955.0 48064.0 79563.0]

['2000年' 126743.0 65437.0 61306.0 45906.0 80837.0]

['1999年' 125786.0 64692.0 61094.0 43748.0 82038.0]

['1998年' 124761.0 63940.0 60821.0 41608.0 83153.0]

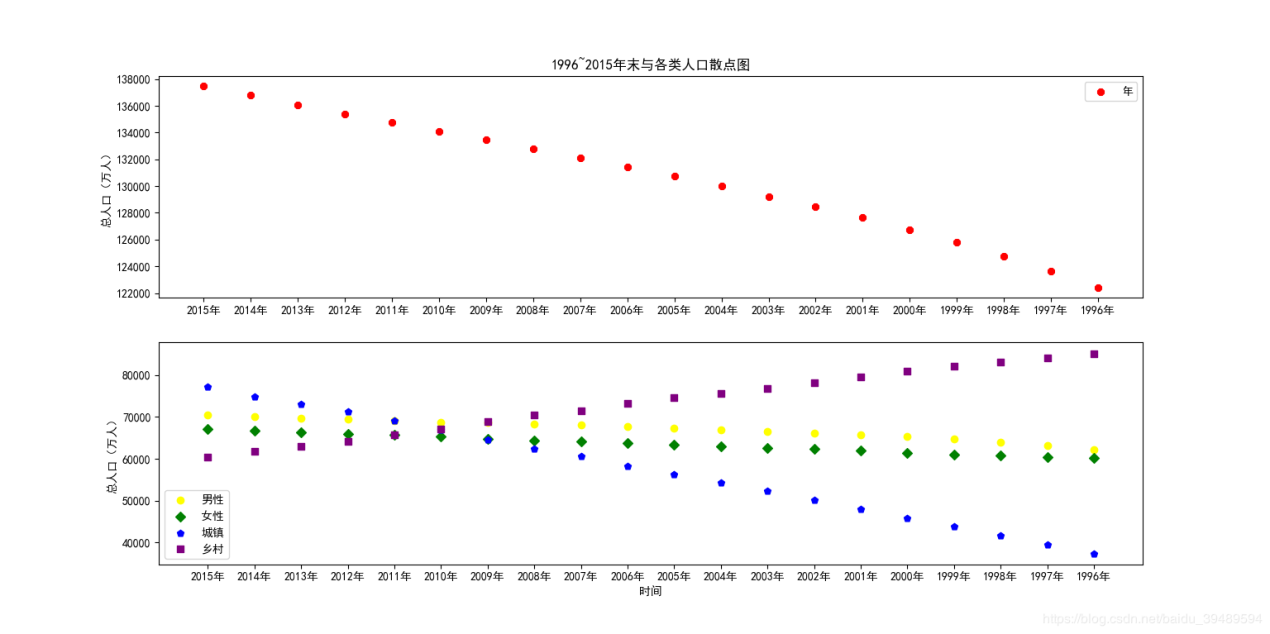
['1997年' 123626.0 63131.0 60495.0 39449.0 84177.0]

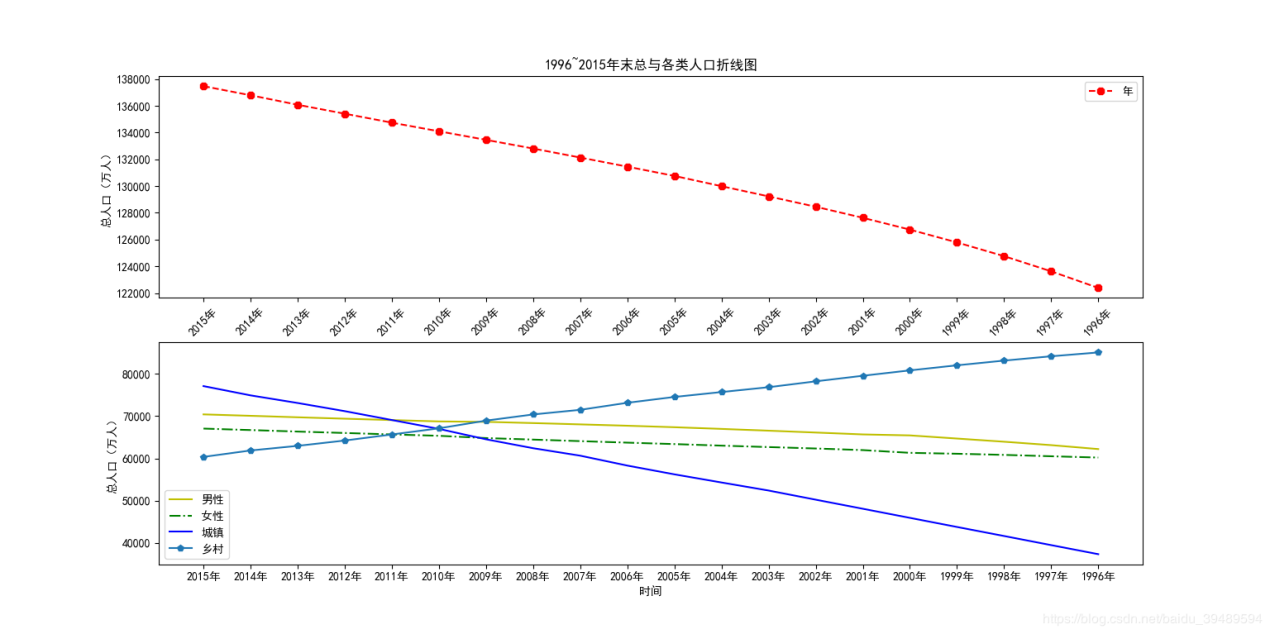
['1996年' 122389.0 62200.0 60189.0 37304.0 85085.0]

[nan nan nan nan nan nan]

[nan nan nan nan nan nan]]

['时间' '年末总人口(万人)' '男性人口(万人)' '女性人口(万人)' '城镇人口(万人)' '乡村人口(万人)']





**实训2 分析1996-2015年人口数据各个特征的发布与分散状况**

import numpy as np

import matplotlib.pyplot as plt

data=np.load('C:/Users/3SXL\_21144/Desktop/populations.npz',allow\_pickle=True)

print(data['data'])

name=data['feature\_names']

values=data['data']

plt.rcParams['font.sans-serif']='SimHei'

label1=['男性','女性']

label2=['城镇','乡村']

ex=[0.01,0.01]

p1=plt.figure(figsize=(12,12))

a1=p1.add\_subplot(2,2,1)

plt.bar(range(2),values[19,2:4],width=0.5,color='orange')

plt.ylabel('人口（万人）')

plt.ylim(0,80000)

plt.xticks(range(2),label1)

plt.title('1996年男、女人口数直方图')

b1=p1.add\_subplot(2,2,2)

plt.bar(range(2),values[0,2:4],width=0.5,color='red')

plt.ylabel('人口（万人）')

plt.ylim(0,80000)

plt.xticks(range(2),label1)

plt.title('2015年男、女人口数直方图')

c1=p1.add\_subplot(2,2,3)

plt.bar(range(2),values[19,4:6],width=0.5,color='orange')

plt.xlabel('类别')

plt.ylabel('人口（万人）')

plt.ylim(0,90000)

plt.xticks(range(2),label2)

plt.title('1996年城、乡人口数直方图')

d1=p1.add\_subplot(2,2,4)

plt.bar(range(2),values[0,4:6],width=0.5,color='red')

plt.xlabel('类别')

plt.ylabel('人口（万人）')

plt.ylim(0,90000)

plt.xticks(range(2),label2)

plt.title('2015年城、乡人口数直方图')

plt.savefig('C:/Users/3SXL\_21144/Desktop/tmp/1996、2015年各类人口直方图.png')

p2=plt.figure(figsize=(8,8))

a2=p2.add\_subplot(2,2,1)

plt.pie(values[19,2:4],explode=ex,labels=label1,colors=['pink','crimson'],autopct='%1.1f%%')

plt.title('1996年男、女人口数饼图')

b2=p2.add\_subplot(2,2,2)

plt.pie(values[0,2:4],explode=ex,labels=label1,colors=['PeachPuff','skyblue'],autopct='%1.1f%%')

plt.title('2015年男、女人口数饼图')

c2=p2.add\_subplot(2,2,3)

plt.pie(values[19,4:6],explode=ex,labels=label2,colors=['pink','crimson'],autopct='%1.1f%%')

plt.title('1996年城、乡人口数饼图')

d2=p2.add\_subplot(2,2,4)

plt.pie(values[0,4:6],explode=ex,labels=label2,colors=['PeachPuff','skyblue'],autopct='%1.1f%%')

plt.title('2015年城、乡人口数饼图')

plt.savefig('C:/Users/3SXL\_21144/Desktop/tmp/1996、2015年各类人口饼图.png')

p3=plt.figure(figsize=(10,10))

plt.boxplot(values[0:20,1:6],notch=True,labels=['年末','男性','女性','城镇','乡村'],meanline=True)

plt.xlabel('类别')

plt.ylabel('人口（万人）')

plt.title('1996~2015年各特征人口箱线图')

plt.savefig('C:/Users/3SXL\_21144/Desktop/tmp/1996`2015年各特征人口箱线图.png')

plt.show()

答：

[['2015年' 137462.0 70414.0 67048.0 77116.0 60346.0]

['2014年' 136782.0 70079.0 66703.0 74916.0 61866.0]

['2013年' 136072.0 69728.0 66344.0 73111.0 62961.0]

['2012年' 135404.0 69395.0 66009.0 71182.0 64222.0]

['2011年' 134735.0 69068.0 65667.0 69079.0 65656.0]

['2010年' 134091.0 68748.0 65343.0 66978.0 67113.0]

['2009年' 133450.0 68647.0 64803.0 64512.0 68938.0]

['2008年' 132802.0 68357.0 64445.0 62403.0 70399.0]

['2007年' 132129.0 68048.0 64081.0 60633.0 71496.0]

['2006年' 131448.0 67728.0 63720.0 58288.0 73160.0]

['2005年' 130756.0 67375.0 63381.0 56212.0 74544.0]

['2004年' 129988.0 66976.0 63012.0 54283.0 75705.0]

['2003年' 129227.0 66556.0 62671.0 52376.0 76851.0]

['2002年' 128453.0 66115.0 62338.0 50212.0 78241.0]

['2001年' 127627.0 65672.0 61955.0 48064.0 79563.0]

['2000年' 126743.0 65437.0 61306.0 45906.0 80837.0]

['1999年' 125786.0 64692.0 61094.0 43748.0 82038.0]

['1998年' 124761.0 63940.0 60821.0 41608.0 83153.0]

['1997年' 123626.0 63131.0 60495.0 39449.0 84177.0]

['1996年' 122389.0 62200.0 60189.0 37304.0 85085.0]

[nan nan nan nan nan nan]

[nan nan nan nan nan nan]]

