

第三部分 Matplotlib数据可视化基础（国民生产总值数据分析）

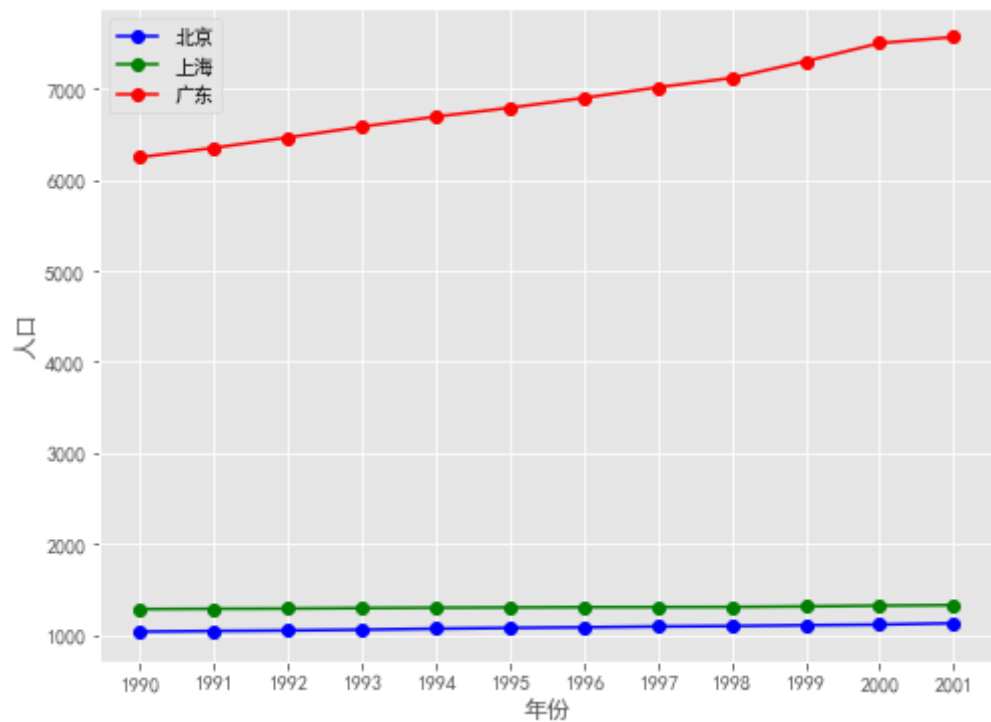
3-1 matplotlib绘图基础

项目步骤：绘制人口数据的折线图

```
# 以人口数据为例
import numpy as np
import matplotlib.pyplot as plt
# 读入全国各省市1990年到2001年的人口数据，成为二维数组（跳过前两行和第一列）
peoples = np.loadtxt("data/人口.csv", skiprows=2, delimiter=",", usecols=
(range(1,13)), encoding="utf8")
# 读入全国各省市名称，文件的第一列（跳过前两行）
names = np.loadtxt("data/人口.csv", skiprows=2, delimiter=",", usecols=
(0), encoding="utf8", dtype=str)
```

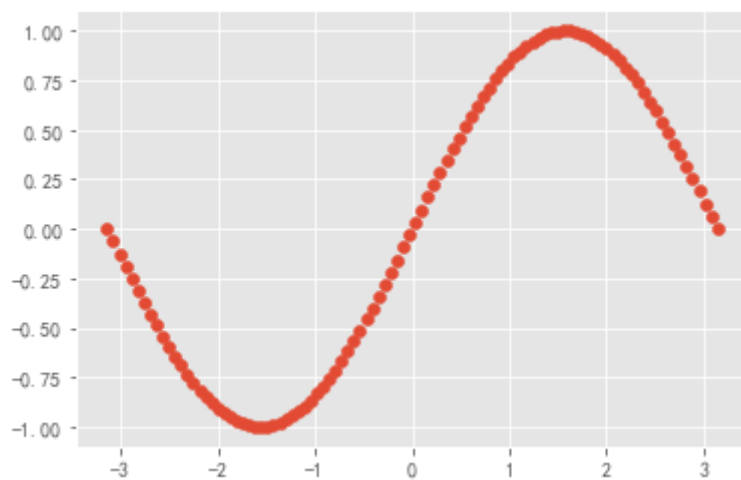
```
# 设置中文显示
plt.rcParams["font.sans-serif"] = "SimHei"
plt.rcParams["axes.unicode_minus"] = False
```

```
# 生成画布并设置大小
fig = plt.figure(figsize=(8,6))
# 绘制北京人口1990-2001年折线图（知识点：绘图函数、点、线型、颜色、标记）
plt.plot(peoples[0,:], marker="o", linestyle="-", color="blue", label=names[0])
# 绘制上海人口1990-2001年折线图（知识点：绘图函数、点、线型、颜色、标记）
plt.plot(peoples[8,:], marker="o", linestyle="-", color="green", label=names[8])
# 绘制广东人口1990-2001年折线图（知识点：绘图函数、点、线型、颜色、标记）
plt.plot(peoples[18,:], marker="o", linestyle="-", color="red", label=names[18])
# 设置x和y轴标签
plt.xlabel("年份")
plt.ylabel("人口")
# 设置x轴刻度为1990-2001年
plt.xticks(range(12), range(1990,2002))
# 设置图例
plt.legend()
# 显示图形
plt.show()
```

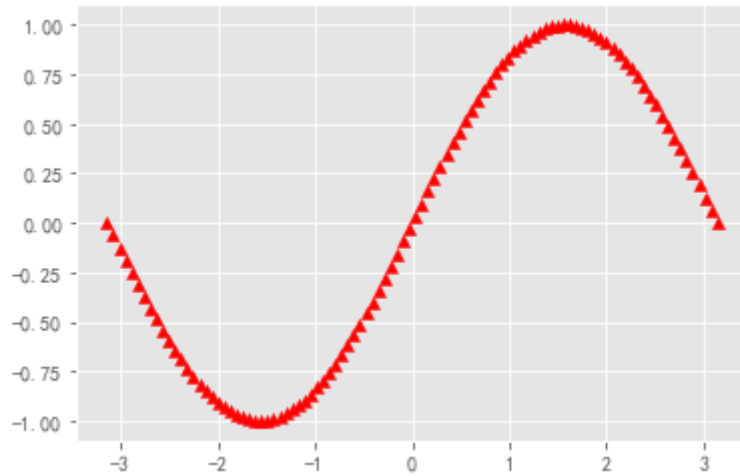


3-2 绘制散点图

```
# 绘制散点图
# 引入相关库
import numpy as np
import matplotlib.pyplot as plt
# 生成x, y轴数据
x = np.linspace(-np.pi, np.pi, num=100)
y = np.sin(x)
# 绘制
plt.scatter(x, y)
plt.show()
```

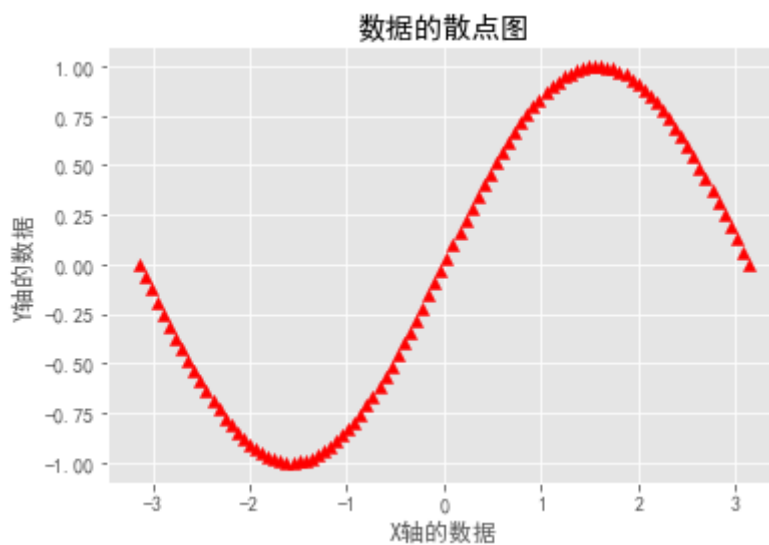


```
# 加入参数
plt.scatter(x, y, c="red", marker="^")
plt.show()
```



```
# 中文显示
plt.rcParams['font.sans-serif'] = 'SimHei'
plt.rcParams['axes.unicode_minus'] = False
```

```
# 加入图形元素
plt.scatter(x, y, c="red", marker="^")
plt.title("数据的散点图")
plt.xlabel("X轴的数据")
plt.ylabel("Y轴的数据")
plt.show()
```



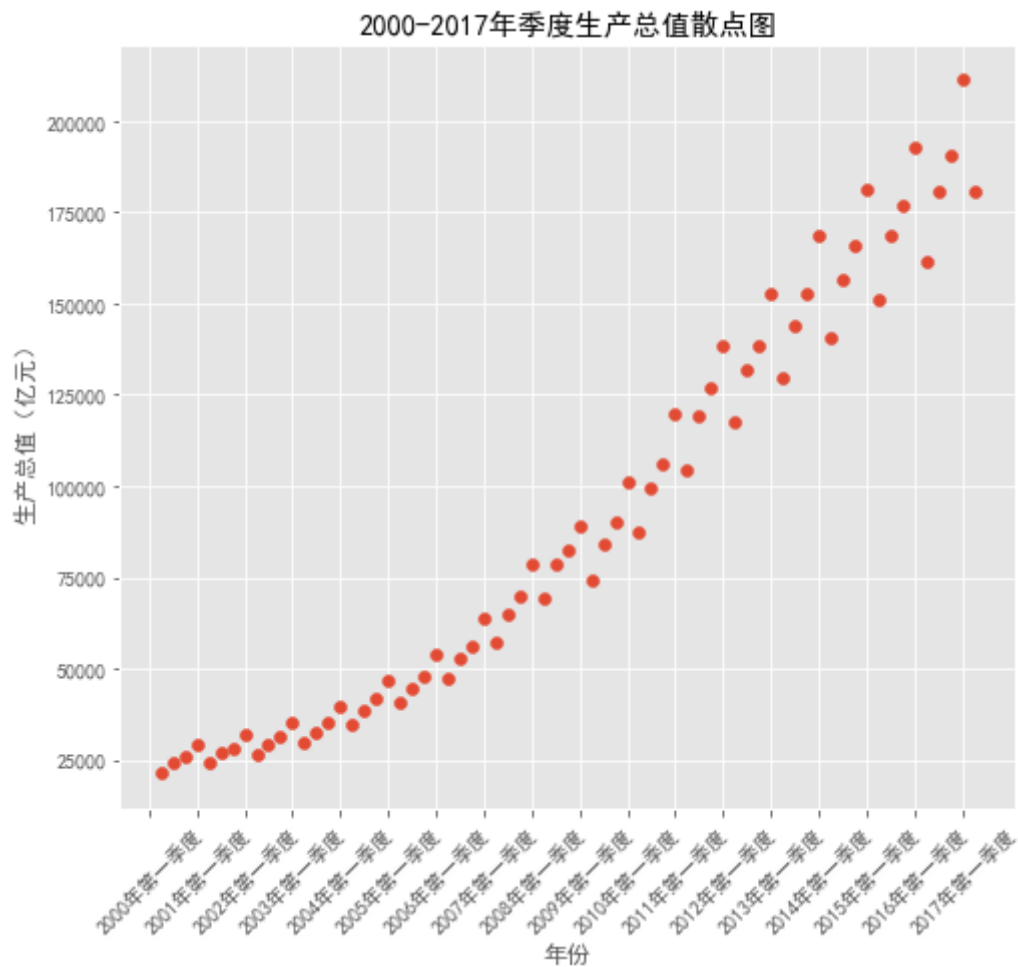
项目步骤：2000-2017年季度生产总值散点图

```
# 读入数据，数据在“data/gdp.csv”中
import numpy as np
import matplotlib.pyplot as plt
plt.rcParams['font.sans-serif'] = 'SimHei' ## 设置中文显示
plt.rcParams['axes.unicode_minus'] = False
```

```
# 这里先用一下pandas库，下一部分会讲解
import pandas as pd
gdp = pd.read_csv("data/gdp.csv", encoding="gbk")
# 转换成数组
datas = gdp.values
names = gdp.columns
print(names)
print(datas)
```

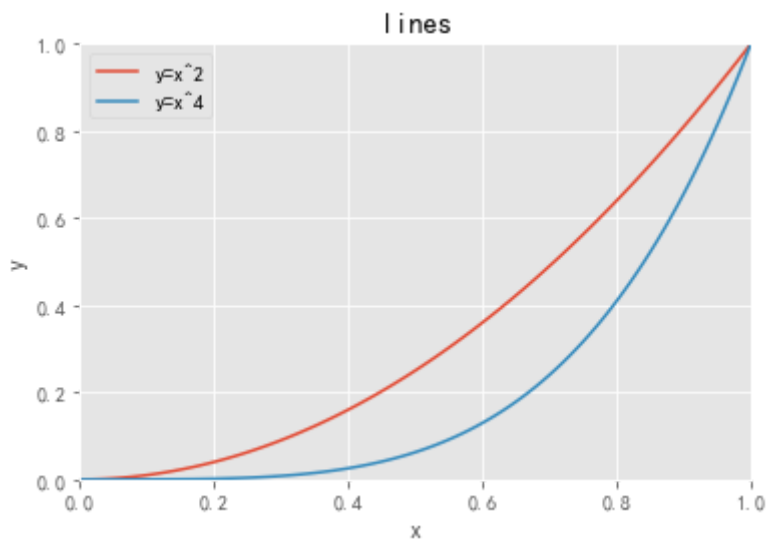
```
Index(['序号', '时间', '国内生产总值_当季值(亿元)', '第一产业增加值_当季值(亿元)', '第二产业增加值_当季值(亿元)',
      '第三产业增加值_当季值(亿元)', '农林牧渔业增加值_当季值(亿元)', '工业增加值_当季值(亿元)',
      '建筑业增加值_当季值(亿元)', '批发和零售业增加值_当季值(亿元)', '交通运输、仓储和邮政业增加值_当季值(亿元)',
      '住宿和餐饮业增加值_当季值(亿元)', '金融业增加值_当季值(亿元)', '房地产业增加值_当季值(亿元)',
      '其他行业增加值_当季值(亿元)'],
      dtype='object')
[[1 '2000年第一季度' 21329.9 ... 1235.9 933.7 3586.1]
 [2 '2000年第二季度' 24043.4 ... 1124.0 904.7 3464.9]
 [3 '2000年第三季度' 25712.5 ... 1170.4 1070.9 3518.2]
 ...
 [67 '2016年第三季度' 190529.5 ... 15472.5 12164.1 37964.1]
 [68 '2016年第四季度' 211281.3 ... 15548.7 13214.9 39848.4]
 [69 '2017年第一季度' 180682.7 ... 17213.5 12393.4 42443.1]]
```

```
# 绘制2000-2017年季度生产总值散点图
plt.figure(figsize=(8,7))## 设置画布
plt.scatter(datas[:,0],datas[:,2], marker='o')## 绘制散点图
plt.xlabel('年份')## 添加横轴标签
plt.ylabel('生产总值（亿元）')## 添加y轴名称
plt.xticks(range(0,70,4),datas[range(0,70,4),1],rotation=45)
plt.title('2000-2017年季度生产总值散点图')## 添加图表标题
plt.savefig('tmp/2000-2017年季度生产总值散点图.png')
plt.show()
```



3-3 绘制折线图

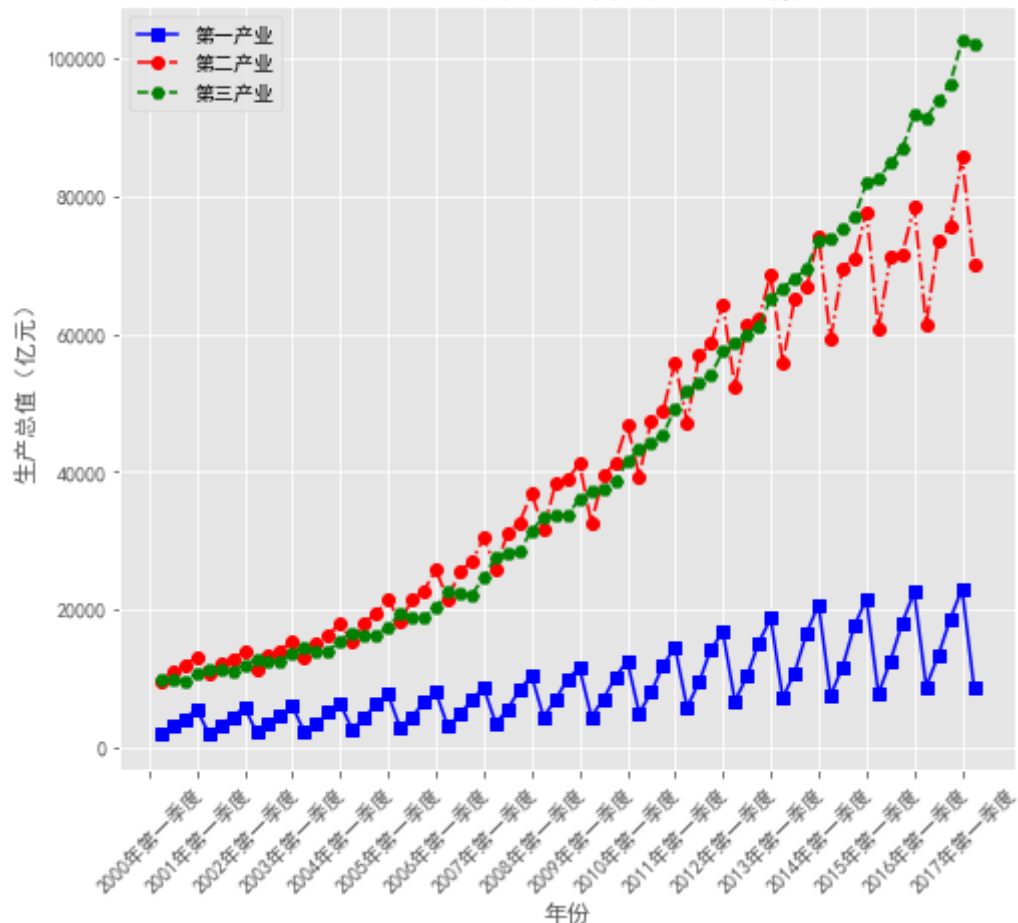
```
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
data = np.arange(0,1.1,0.01)
plt.title('lines') ## 添加标题
plt.xlabel('x')## 添加x轴的名称
plt.ylabel('y')## 添加y轴的名称
plt.xlim((0,1))## 确定x轴范围
plt.ylim((0,1))## 确定y轴范围
plt.xticks([0,0.2,0.4,0.6,0.8,1])## 规定x轴刻度
plt.yticks([0,0.2,0.4,0.6,0.8,1])## 确定y轴刻度
plt.plot(data,data**2)## 添加y=x^2曲线
plt.plot(data,data**4)## 添加y=x^4曲线
plt.legend(['y=x^2', 'y=x^4'])
plt.show()
```



项目步骤：2000-2017年各产业季度生产总值折线图

```
import numpy as np
import matplotlib.pyplot as plt
%matplotlib inline
plt.style.use("ggplot")
plt.figure(figsize=(8,7))## 设置画布
## 绘制折线图
plt.plot(datas[:,0],datas[:,3], 'bs-')
plt.plot(datas[:,0],datas[:,4], 'ro-.')
plt.plot(datas[:,0],datas[:,5], 'gh--')
plt.xlabel('年份')## 添加横轴标签
plt.ylabel('生产总值（亿元）')## 添加y轴名称
plt.xticks(range(0,70,4),datas[range(0,70,4),1],rotation=45)
plt.title('2000-2017年各产业季度生产总值折线图')## 添加图表标题
plt.legend(['第一产业','第二产业','第三产业'])
plt.show()
```

2000-2017年各产业季度生产总值折线图



3-4 绘制条形图

```
import matplotlib.pyplot as plt
import matplotlib
# 设置中文字体和负号正常显示
matplotlib.rcParams['font.sans-serif'] = ['SimHei']
matplotlib.rcParams['axes.unicode_minus'] = False

label_list = ['2014', '2015', '2016', '2017'] # 横坐标刻度显示值
num_list1 = [20, 30, 15, 35] # 纵坐标值1
num_list2 = [15, 30, 40, 20] # 纵坐标值2
x = range(len(num_list1))
"""
绘制条形图
left: 长条形中点横坐标
height: 长条形高度
width: 长条形宽度, 默认值0.8
label: 为后面设置legend准备
"""

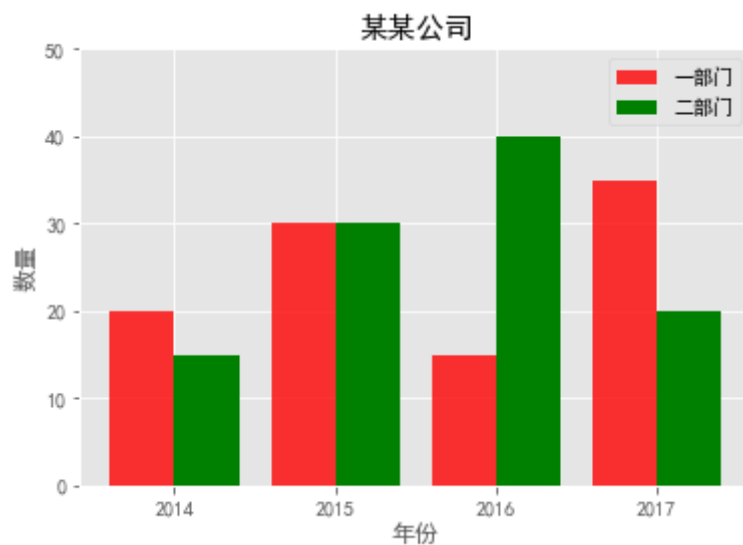
rects1 = plt.bar(x, num_list1, width=0.4, alpha=0.8, color='red', label="一部门")
rects2 = plt.bar([i + 0.4 for i in x], num_list2, width=0.4, color='green',
label="二部门")
plt.ylim(0, 50) # y轴取值范围
plt.ylabel("数量")
"""
设置x轴刻度显示值
参数一: 中点坐标
```

参数二：显示值

"""

```
plt.xticks([index + 0.2 for index in x], label_list)
plt.xlabel("年份")
plt.title("某某公司")
plt.legend()
```

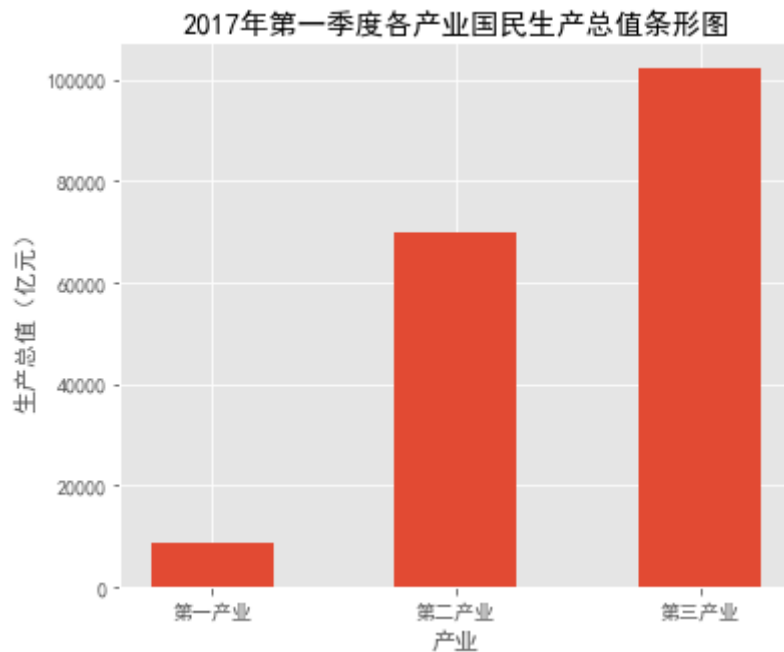
<matplotlib.legend.Legend at 0x2c4db421160>



项目步骤：2017年第一季度各产业国民生产总值条形图

```
import numpy as np
import matplotlib.pyplot as plt
plt.rcParams['font.sans-serif'] = 'SimHei'## 设置中文显示
plt.rcParams['axes.unicode_minus'] = False

labels = ['第一产业', '第二产业', '第三产业']## 刻度标签
plt.figure(figsize=(6,5))## 设置画布
plt.bar(range(3),datas[-1,3:6],width = 0.5)## 绘制条形图
plt.xlabel('产业')## 添加横轴标签
plt.ylabel('生产总值（亿元）')## 添加y轴名称
plt.xticks(range(3),labels)
plt.title('2017年第一季度各产业国民生产总值条形图')## 添加图表标题
plt.show()
```

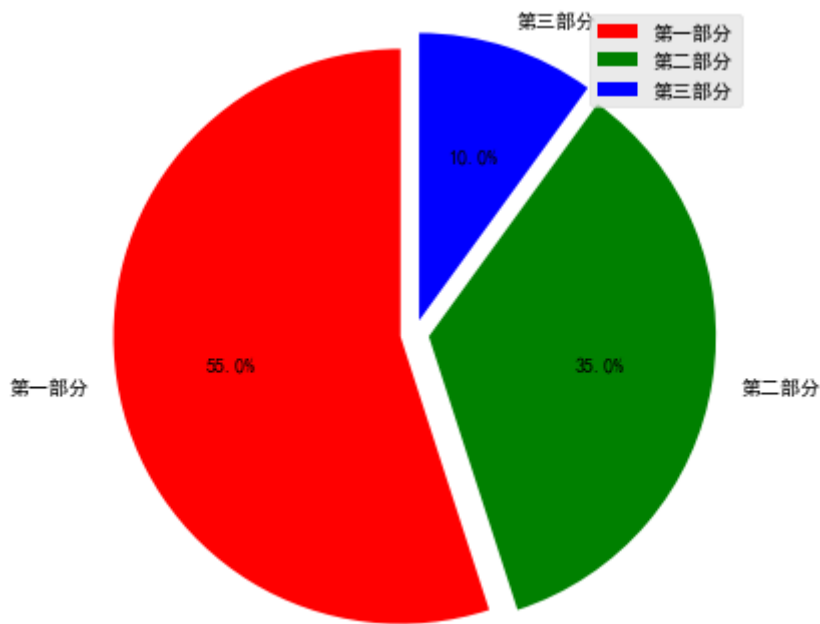
3-5 绘制饼图

```
import matplotlib.pyplot as plt
import matplotlib

plt.figure(figsize=(6,6))
matplotlib.rcParams['font.sans-serif'] = 'SimHei'
matplotlib.rcParams['axes.unicode_minus'] = False

label_list = ["第一部分", "第二部分", "第三部分"]    # 各部分标签
size = [55, 35, 10]    # 各部分大小
color = ["red", "green", "blue"]    # 各部分颜色
explode = [0.05, 0.05, 0.05]    # 各部分突出值
"""
绘制饼图
explode: 设置各部分突出
label: 设置各部分标签
labeldistance: 设置标签文本距圆心位置, 1.1表示1.1倍半径
autopct: 设置圆里面文本
shadow: 设置是否有阴影
startangle: 起始角度, 默认从0开始逆时针转
pctdistance: 设置圆内文本距圆心距离
"""

plt.pie(size, explode=explode, colors=color, labels=label_list,
labeldistance=1.1, autopct="%1.1f%%", shadow=False, startangle=90,
pctdistance=0.6)
plt.axis("equal")    # 设置横轴和纵轴大小相等, 这样饼才是圆的
plt.legend()
plt.show()
```

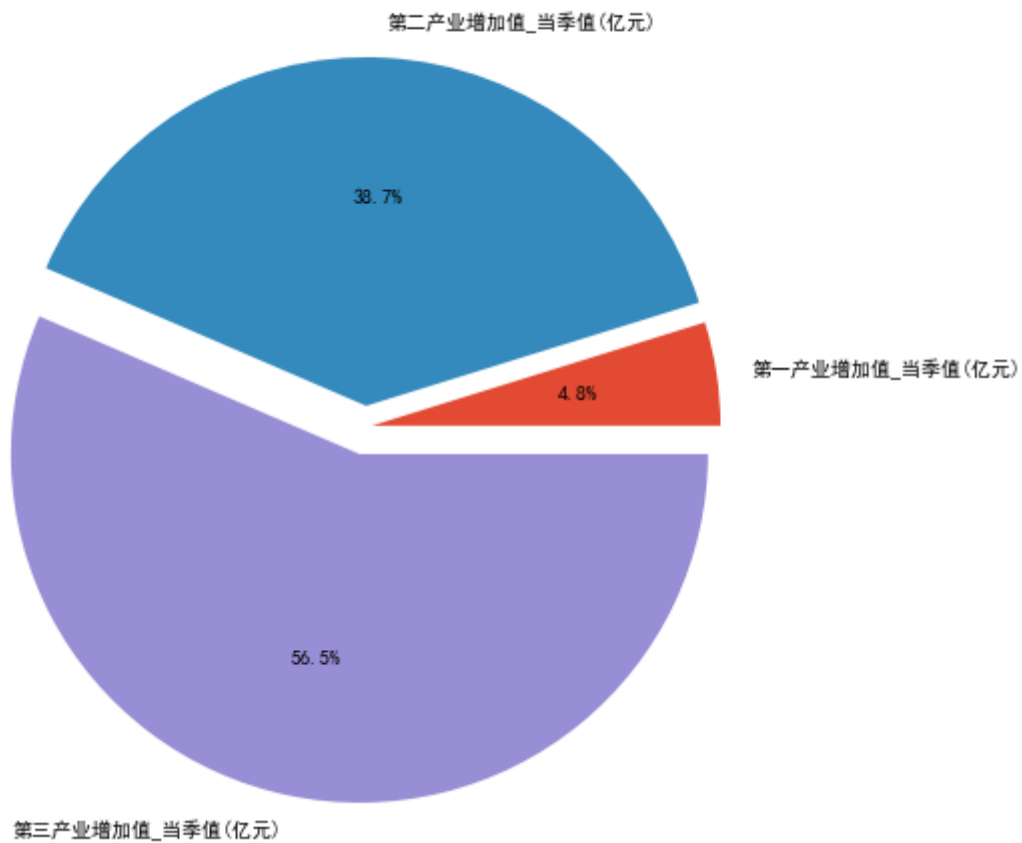


项目步骤：2017年第一季度各产业GDP数据饼图

```
import numpy as np
import matplotlib.pyplot as plt
## 设置中文显示
plt.rcParams["font.sans-serif"] = "SimHei"
plt.rcParams["axes.unicode_minus"] = False

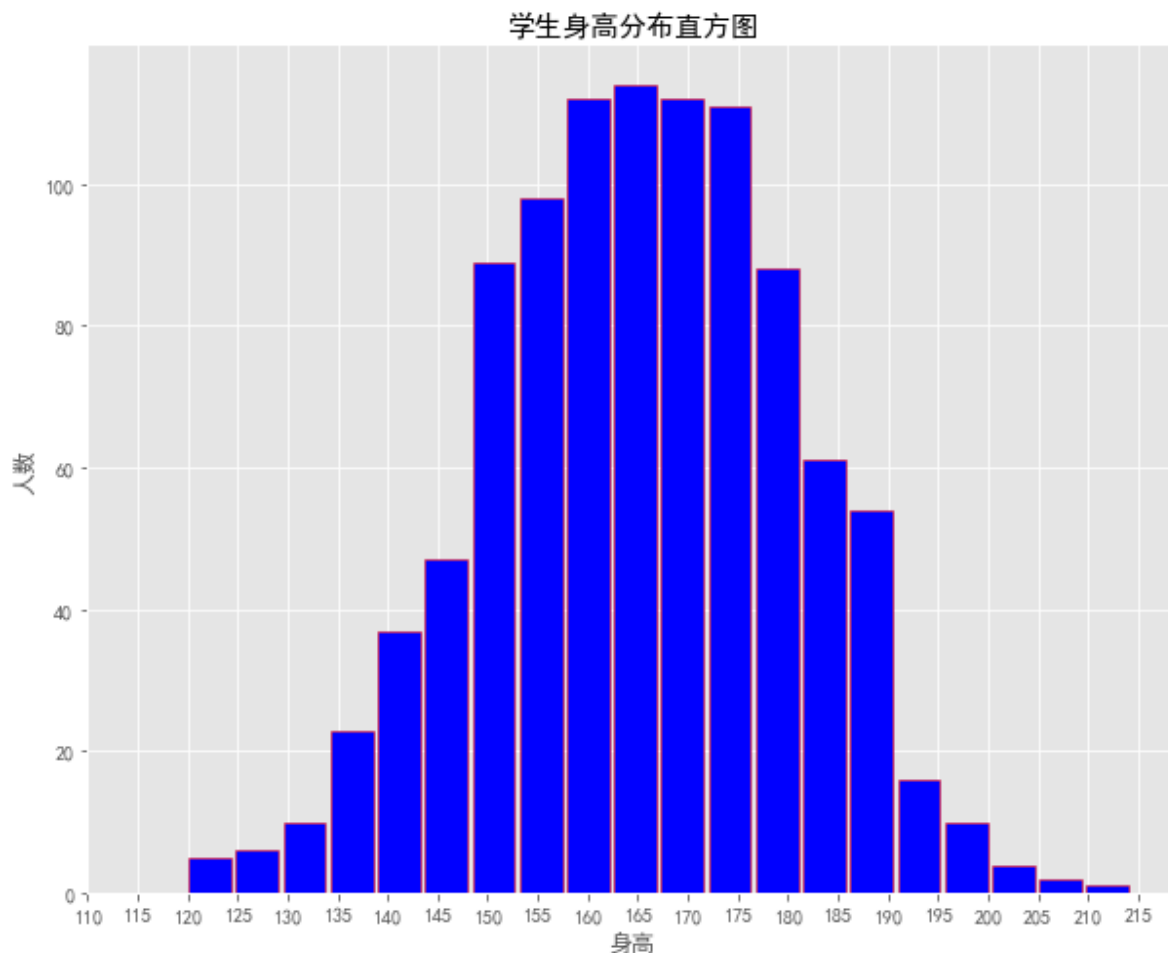
## 将画布设定为正方形，则绘制的饼图是正圆
plt.figure(figsize=(8,8))
## 定义饼状图的标签，标签是列表
label = ["第一产业", "第二产业", "第三产业"]
## 设定各项离心n个半径
explode = [0.02, 0.06, 0.08]
## 绘制饼图
plt.pie(datas[-1,3:6], labels=names[3:6], autopct="%.1f%%", explode=explode)
## 添加图表标题
plt.title("2017年第一季度各产业GDP数据饼图")
## 显示图形
plt.show()
```

2017年第一季度各产业GDP数据饼图



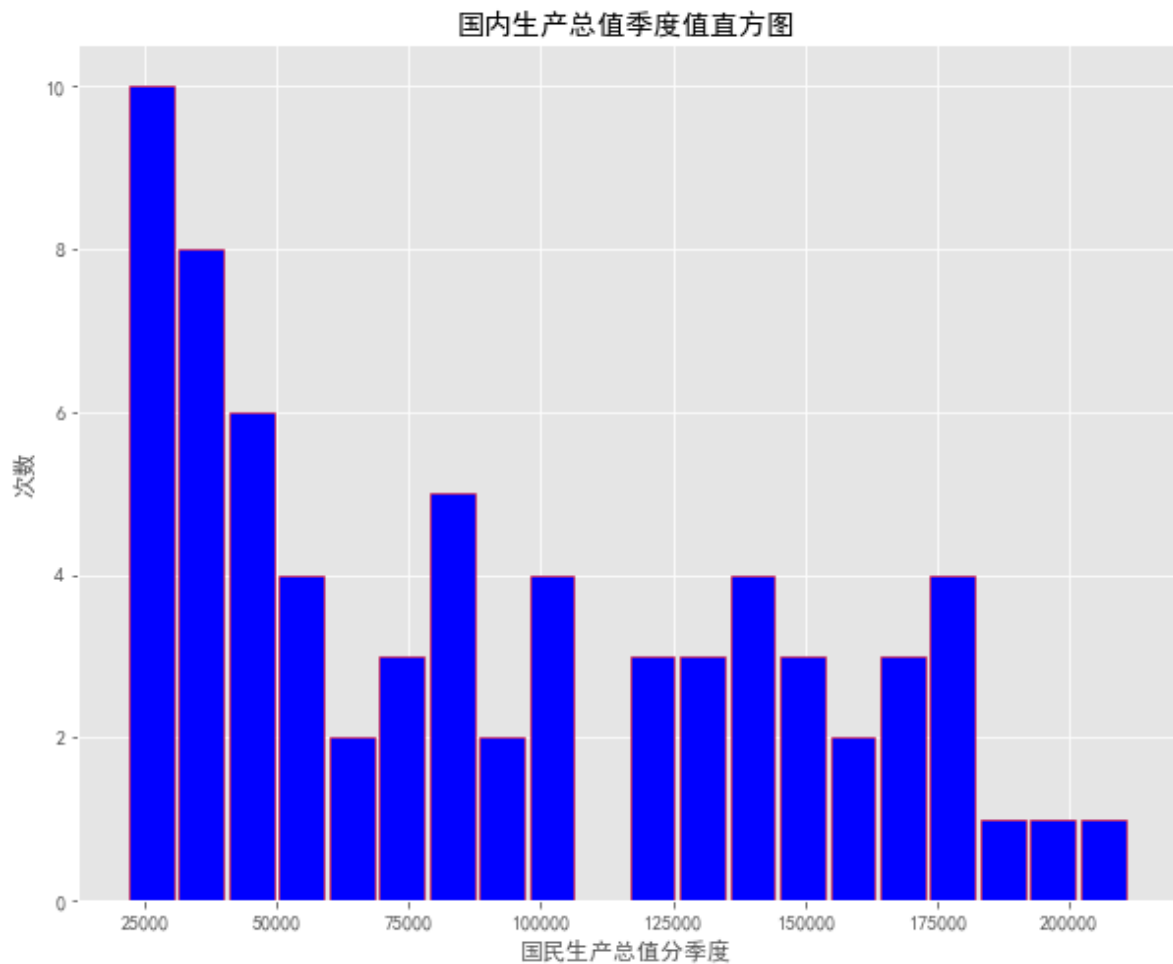
3-6 绘制直方图

```
import numpy as np
import matplotlib.pyplot as plt
plt.figure(figsize=(10,8))
## 设置中文显示
plt.rcParams["font.sans-serif"] = "SimHei"
plt.rcParams["axes.unicode_minus"] = False
## 数据准备
heights = np.random.normal(165, 15, 1000)
#绘制直方图
plt.hist(heights,color="blue",edgecolor="red", rwidth=0.9, bins=20)
plt.xlabel("身高")
plt.ylabel("人数")
plt.title("学生身高分布直方图")
plt.xticks(range(110,220,5))
plt.show()
```



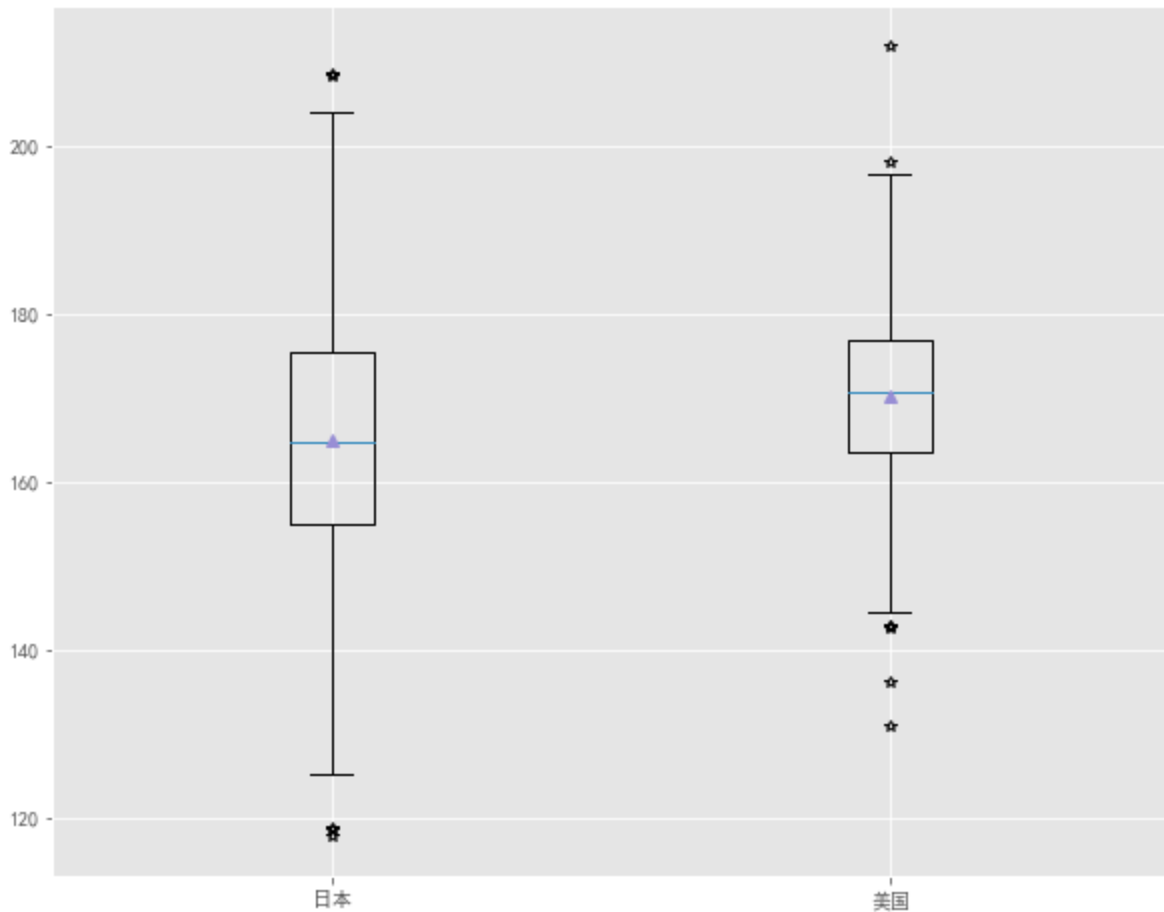
项目步骤：国内生产总值季度值直方图

```
import numpy as np
import matplotlib.pyplot as plt
plt.figure(figsize=(10,8))
## 设置中文显示
plt.rcParams["font.sans-serif"] = "SimHei"
plt.rcParams["axes.unicode_minus"] = False
## 数据准备,取第三列国内生产总值季度值
gdp_sum = list(datas[:,2])
#print(heights)
#绘制直方图
plt.hist(gdp_sum,color="blue",edgecolor="red", rwidth=0.9, bins=20)
plt.xlabel("国内生产总值分季度")
plt.ylabel("次数")
plt.title("国内生产总值季度值直方图")
#plt.xticks(range(110,220,5))
plt.show()
```



3-7 绘制箱型图

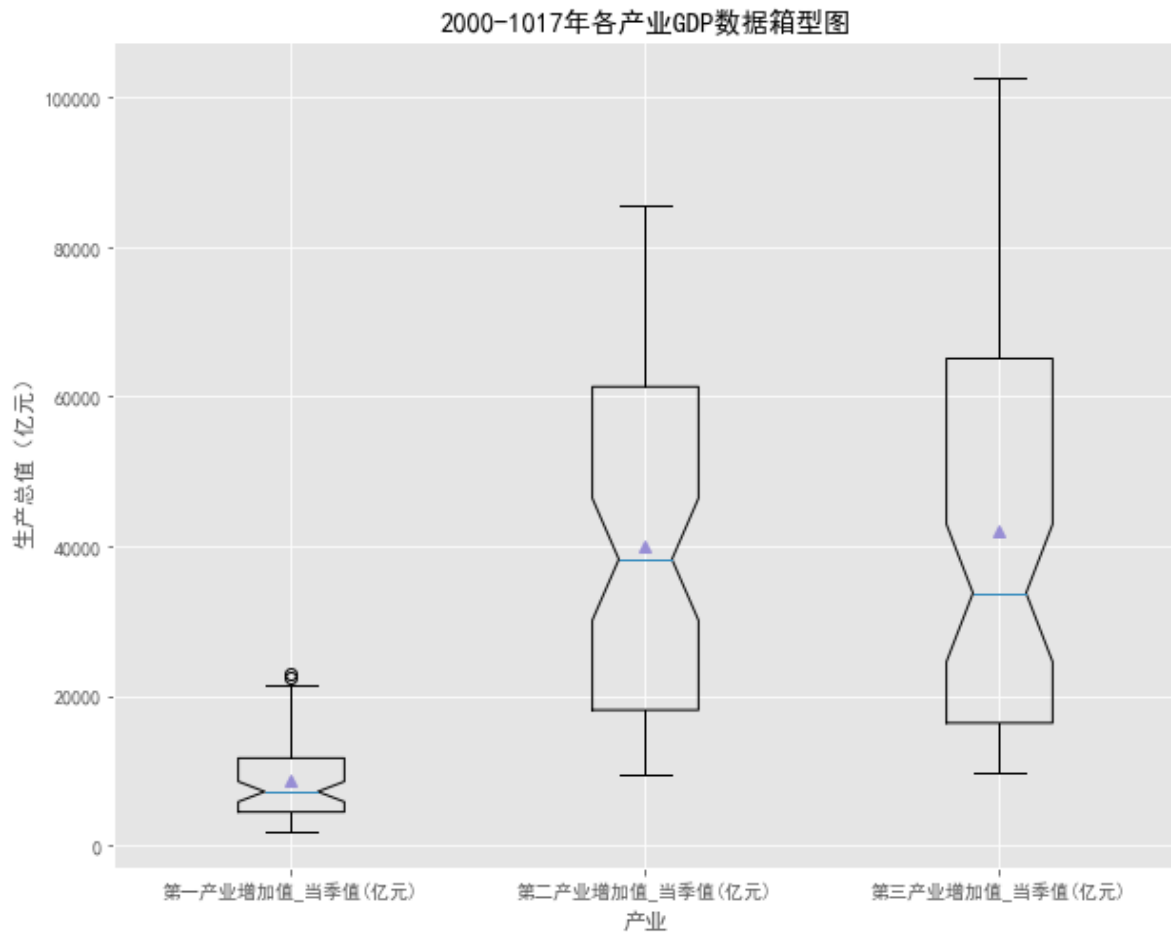
```
# 模拟两个国家的身高数据
plt.figure(figsize=(10,8))
h1 = np.random.normal(165, 15, 1000)
h2 = np.random.normal(170, 10, 1000)
plt.boxplot((h1,h2), showmeans=True, showfliers=True, sym="*", labels=["日本","美国"])
plt.show()
```



项目步骤：2000-2017年各产业GDP数据箱型图

```
import numpy as np
import matplotlib.pyplot as plt
## 设置中文显示
plt.rcParams["font.sans-serif"] = "SimHei"
plt.rcParams["axes.unicode_minus"] = False

## 准备数据
gdp = datas[:,3], datas[:,4], datas[:,5]
## 设置画布
plt.figure(figsize=(10,8))
## 绘制箱型图
plt.boxplot(gdp, labels=names[3:6], showmeans=True, notch=True)
plt.xlabel("产业")
plt.ylabel("生产总值（亿元）")
## 添加图表标题
plt.title("2000-2017年各产业GDP数据箱型图")
## 显示图形
plt.show()
```



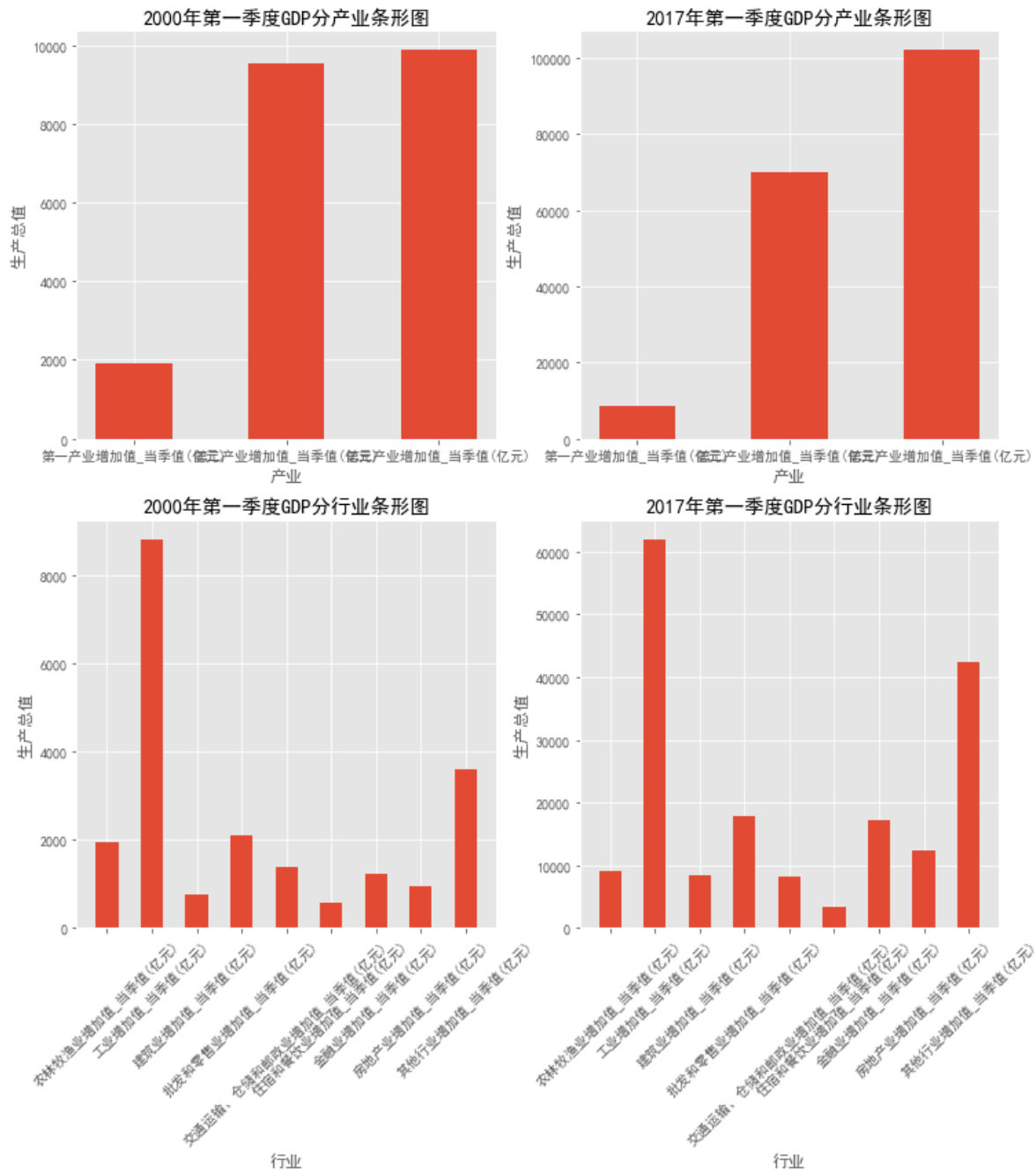
3-8 绘制子图

```
import pandas as pd
gdp = pd.read_csv("data/gdp.csv", encoding="gbk")
# 转换成数组
values = gdp.values
names = gdp.columns
```

```
import matplotlib.pyplot as plt
## 设置中文显示
plt.rcParams["font.sans-serif"] = "SimHei"
plt.rcParams["axes.unicode_minus"] = False
```

```
fig, ax = plt.subplots(2,2,figsize=(12,12))
## 子图1
#plt.bar(range(3), values[0,3:6], width=0.5)
## 绘制条形图
ax[0,0].bar(range(3), values[0,3:6], width=0.5)
## 添加横轴标签
ax[0,0].set_xlabel("产业")
## 添加y轴名称
ax[0,0].set_ylabel("生产总值")
## 设置横轴刻度
ax[0,0].set_xticks(range(3))
ax[0,0].set_xticklabels(names[3:6])
## 设置图形标题
```

```
ax[0,0].set_title("2000年第一季度GDP分产业条形图")
## 子图2
## 绘制条形图
ax[0,1].bar(range(3), values[-1,3:6], width=0.5)
## 添加横轴标签
ax[0,1].set_xlabel("产业")
## 添加y轴名称
ax[0,1].set_ylabel("生产总值")
## 设置横轴刻度
ax[0,1].set_xticks(range(3))
ax[0,1].set_xticklabels(names[3:6])
## 设置图形标题
ax[0,1].set_title("2017年第一季度GDP分产业条形图")
## 子图3
## 绘制条形图
ax[1,0].bar(range(9), values[0,6:15], width=0.5)
## 添加横轴标签
ax[1,0].set_xlabel("行业")
## 添加y轴名称
ax[1,0].set_ylabel("生产总值")
## 设置横轴刻度
ax[1,0].set_xticks(range(9))
ax[1,0].set_xticklabels(names[6:15], rotation=45)
## 设置图形标题
ax[1,0].set_title("2000年第一季度GDP分行业条形图")
## 子图4
## 绘制条形图
ax[1,1].bar(range(9), values[-1,6:15], width=0.5)
## 添加横轴标签
ax[1,1].set_xlabel("行业")
## 添加y轴名称
ax[1,1].set_ylabel("生产总值")
## 设置横轴刻度
ax[1,1].set_xticks(range(9))
ax[1,1].set_xticklabels(names[6:15], rotation=45)
## 设置图形标题
ax[1,1].set_title("2017年第一季度GDP分行业条形图")
## 保存并显示图形
plt.show()
```

```
# 饼图
## 各个模块离圆心的距离1
explode1 = [0.01] * 3
## 各个模块离圆心的距离2
explode2 = [0.01] * 9
##画布
plt.figure(figsize=(12,12))
## 子图1
plt.subplot2grid((2,2),(0,0))
## 绘制饼图
plt.pie(values[0,3:6], explode=explode1, labels=names[3:6], autopct="%.1f%%")
## 标题
plt.title("2000年第一季度GDP分产业饼图")
## 子图2
ax2 = plt.subplot2grid((2,2),(0,1))
```

```

## 绘制饼图
plt.pie(values[-1,3:6], explode=explode1, labels=names[3:6], autopct="%.1f%%")
## 标题
plt.title("2017年第一季度GDP分产业饼图")
## 子图3
ax3 = plt.subplot2grid((2,2),(1,0))
## 绘制饼图
plt.pie(values[0,6:15], explode=explode2, labels=names[6:15], autopct="%.1f%%")
## 标题
plt.title("2000年第一季度GDP分行业饼图")
## 子图4
ax4 = plt.subplot2grid((2,2),(1,1))
## 绘制饼图
plt.pie(values[-1,6:15], explode=explode2, labels=names[6:15], autopct="%.1f%%")
## 标题
plt.title("2017年第一季度GDP分行业饼图")
## 保存并显示图形
plt.show()

```

