

6. Matplotlib数据可视化基础

本课程目标



了解绘图基础语法与常用参数



分析特征间关系



分析特征内部数据分布与分散情况

课程目录

Course catalogue

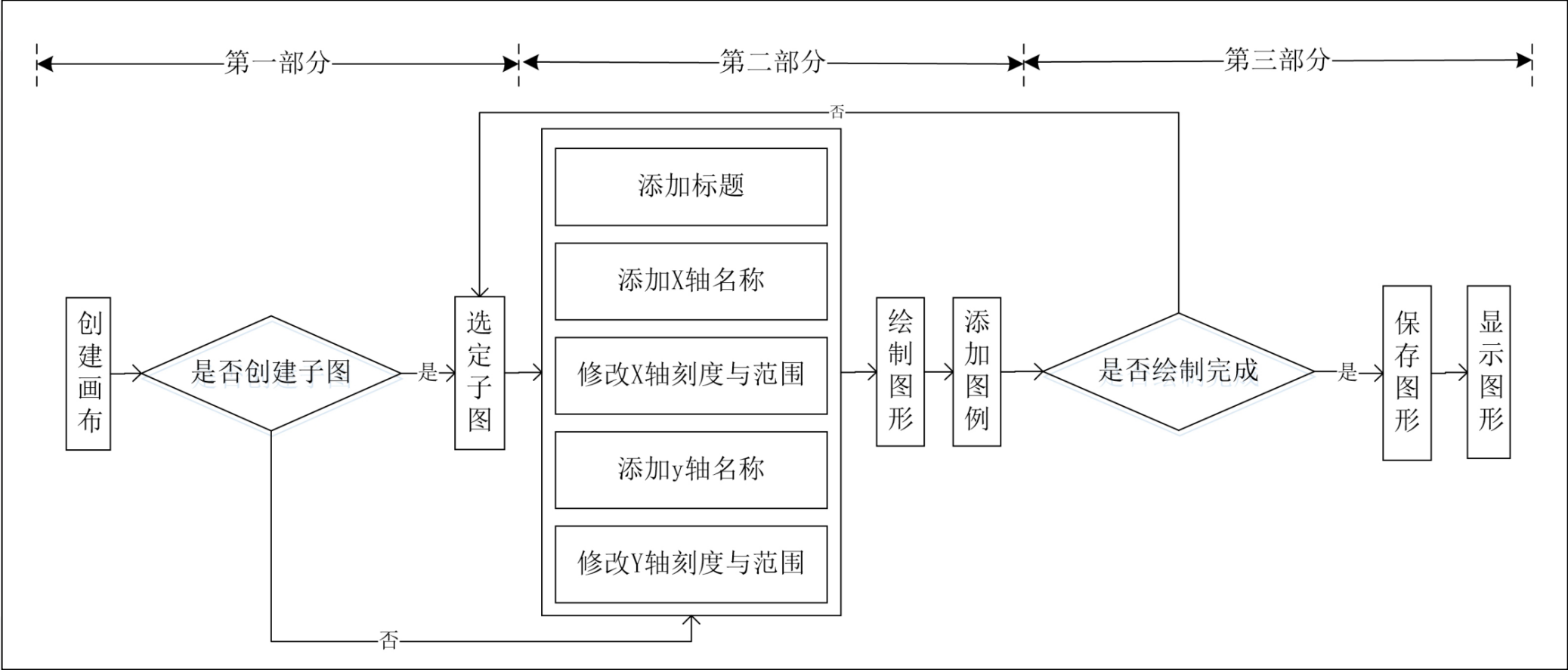
1/ 了解绘图基础语法与常用参数

2/ 分析特征间的关系

3/ 分析特征内部数据分布与分散状况

掌握pyplot基础语法

基本绘图流程



掌握pyplot基础语法

1. 创建画布与创建子图

第一部分主要作用是构建出一张空白的画布，并可以选择是否将整个画布划分为多个部分，方便在同一幅图上绘制多个图形的情况。最简单的绘图可以省略第一部分，而后直接在默认的画布上进行图形绘制。

函数名称	函数作用
plt.figure	创建一个空白画布，可以指定画布大小，像素。
figure.add_subplot	创建并选中子图，可以指定子图的行数，列数，与选中图片编号。

掌握pyplot基础语法

2. 添加画布内容

第二部分是绘图的主体部分。其中添加标题，坐标轴名称，绘制图形等步骤是并列的，没有先后顺序，可以先绘制图形，也可以先添加各类标签。但是添加图例一定要在绘制图形之后。

函数名称	函数作用
plt.title	在当前图形中添加标题，可以指定标题的名称、位置、颜色、字体大小等参数。
plt.xlabel	在当前图形中添加x轴名称，可以指定位置、颜色、字体大小等参数。
plt.ylabel	在当前图形中添加y轴名称，可以指定位置、颜色、字体大小等参数。
plt.xlim	指定当前图形x轴的范围，只能确定一个数值区间，而无法使用字符串标识。
plt.ylim	指定当前图形y轴的范围，只能确定一个数值区间，而无法使用字符串标识。
plt.xticks	指定x轴刻度的数目与取值。
plt.yticks	指定y轴刻度的数目与取值。
plt.legend	指定当前图形的图例，可以指定图例的大小、位置、标签。

掌握pyplot基础语法

3. 存储于展示图形

函数名称	函数作用
<code>plt.savefig</code>	保存绘制的图片，可以指定图片的分辨率、边缘的颜色等参数。
<code>plt.show</code>	在本机显示图形。

设置pyplot的动态rc参数

pyplot使用rc配置文件来自定义图形的各种默认属性，被称为rc配置或rc参数。

在pyplot中几乎所有的默认属性都是可以控制的，例如视图窗口大小以及每英寸点数、线条宽度、颜色和样式、坐标轴、坐标和网格属性、文本、字体等。

设置pyplot的动态rc参数

线条的常用rc参数名称、解释与取值

rc参数名称	解释	取值
lines.linewidth	线条宽度	取0-10之间的数值，默认为1.5。
lines.linestyle	线条样式	可取"-""--""-."": "四种。默认为"-".
lines.marker	线条上点的形状	可取"o""D""h"".", "S"等20种，默认为None。
lines.markersize	点的大小	取0-10之间的数值，默认为1。

设置pyplot的动态rc参数

常用线条类型解释

linestyle取值	意义	linestyle取值	意义
-	实线	-.	点线
--	长虚线	:	短虚线

设置pyplot的动态rc参数

线条标记解释

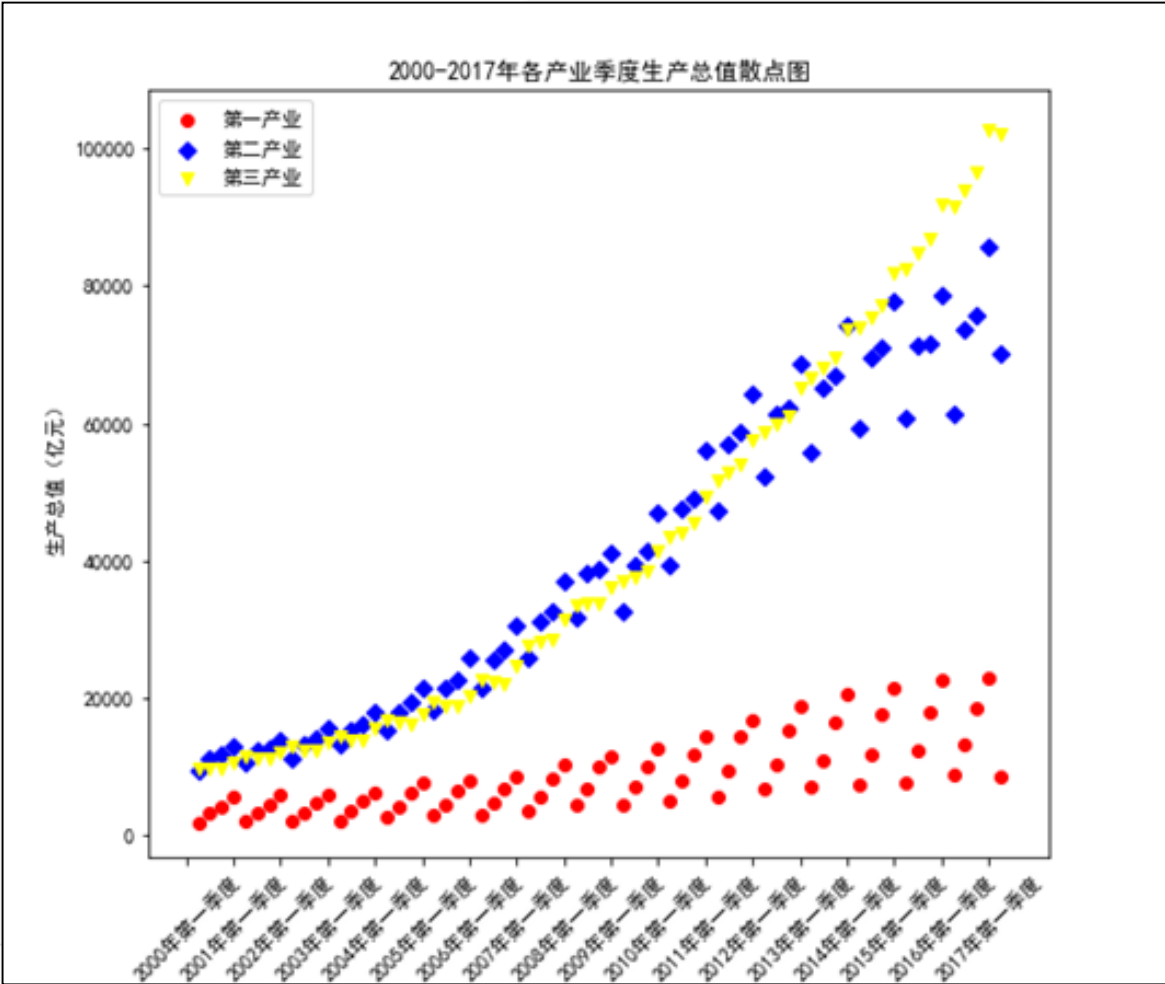
marker取值	意义	marker取值	意义
'o'	圆圈	'.'	点
'D'	菱形	's'	正方形
'h'	六边形1	'*'	星号
'H'	六边形2	'd'	小菱形
'-'	水平线	'v'	一角朝下的三角形
'8'	八边形	'<'	一角朝左的三角形
'p'	五边形	'>'	一角朝右的三角形
','	像素	'^'	一角朝上的三角形
'+'	加号	'\'	竖线
'None'	无	'x'	X

课程目录

Course catalogue

- 1/ 了解绘图基础语法与常用参数
- 2/ 分析特征间的关系
- 3/ 分析特征内部数据分布与分散状况

绘制散点图



散点图（scatter diagram）又称为散点分布图，是以一个特征为横坐标，另一个特征为纵坐标，利用坐标点（散点）的分布形态反映特征间的统计关系的一种图形。

值是由点在图表中的位置表示，类别是由图表中的不同标记表示，通常用于比较跨类别的数据

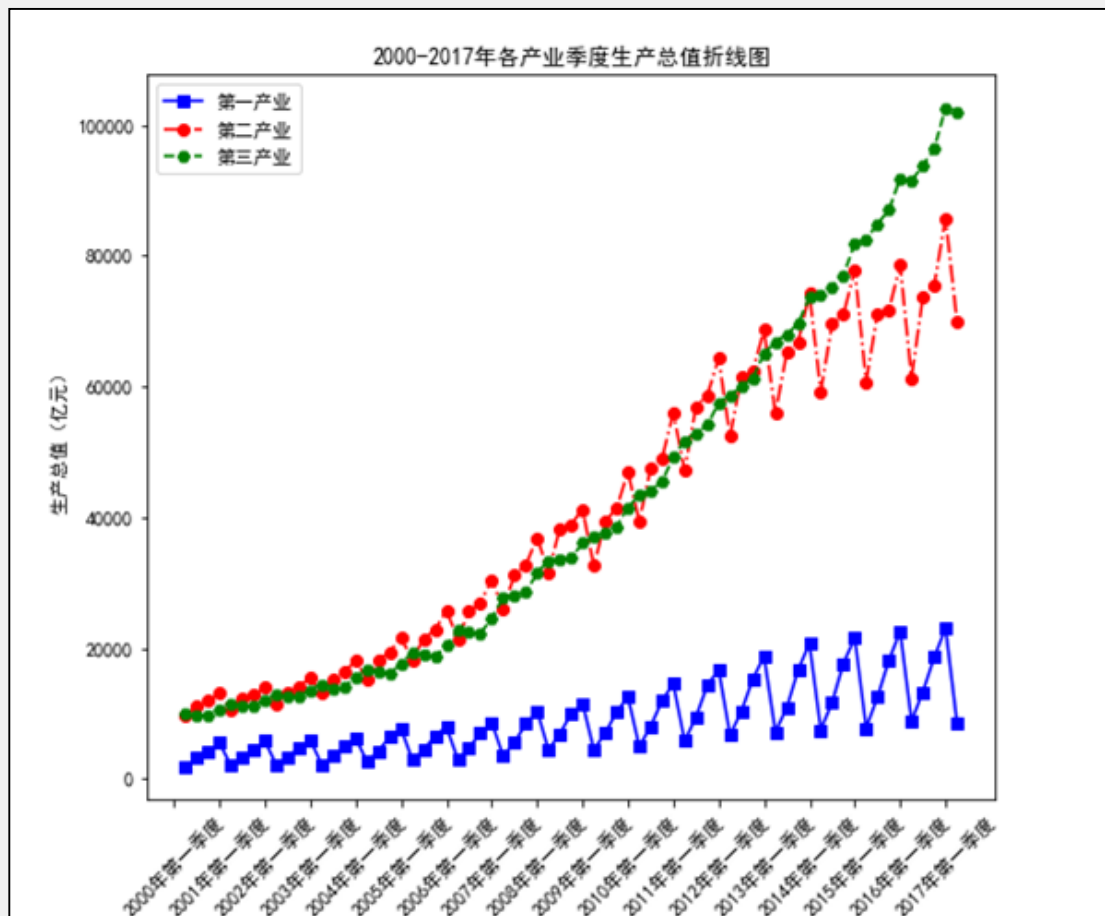
绘制散点图

```
matplotlib.pyplot.scatter(x, y, s=None, c=None, marker=None, alpha=None, **kwargs)
```

- 常用参数及说明如下表所示：

参数名称	说明
x, y	接收array。表示x轴和y轴对应的数据。无默认。
s	接收数值或者一维的array。指定点的大小，若传入一维array则表示每个点的大小。默认为None。
c	接收颜色或者一维的array。指定点的颜色，若传入一维array则表示每个点的颜色。默认为None
marker	接收特定string。表示绘制的点的类型。默认为None。
alpha	接收0-1的小数。表示点的透明度。默认为None。

绘制折线图



折线图 (line chart) 是一种将数据点按照顺序连接起来的图形。可以看作是将散点图, 按照x轴坐标顺序连接起来的图形。

折线图的主要功能是查看因变量y随着自变量x改变的趋势, 最适合用于显示随时间 (根据常用比例设置) 而变化的连续数据。同时还可以看出数量的差异, 增长趋势的变化。

绘制折线图

*matplotlib.pyplot.plot(*args, **kwargs)*

plot函数在官方文档的语法中只要求填入不定长参数，实际可以填入的主要参数主要如下：

参数名称	说明
x, y	接收array。表示x轴和y轴对应的数据。无默认。
color	接收特定string。指定线条的颜色。默认为None。
linestyle	接收特定string。指定线条类型。默认为“-”。
marker	接收特定string。表示绘制的点的类型。默认为None。
alpha	接收0-1的小数。表示点的透明度。默认为None。

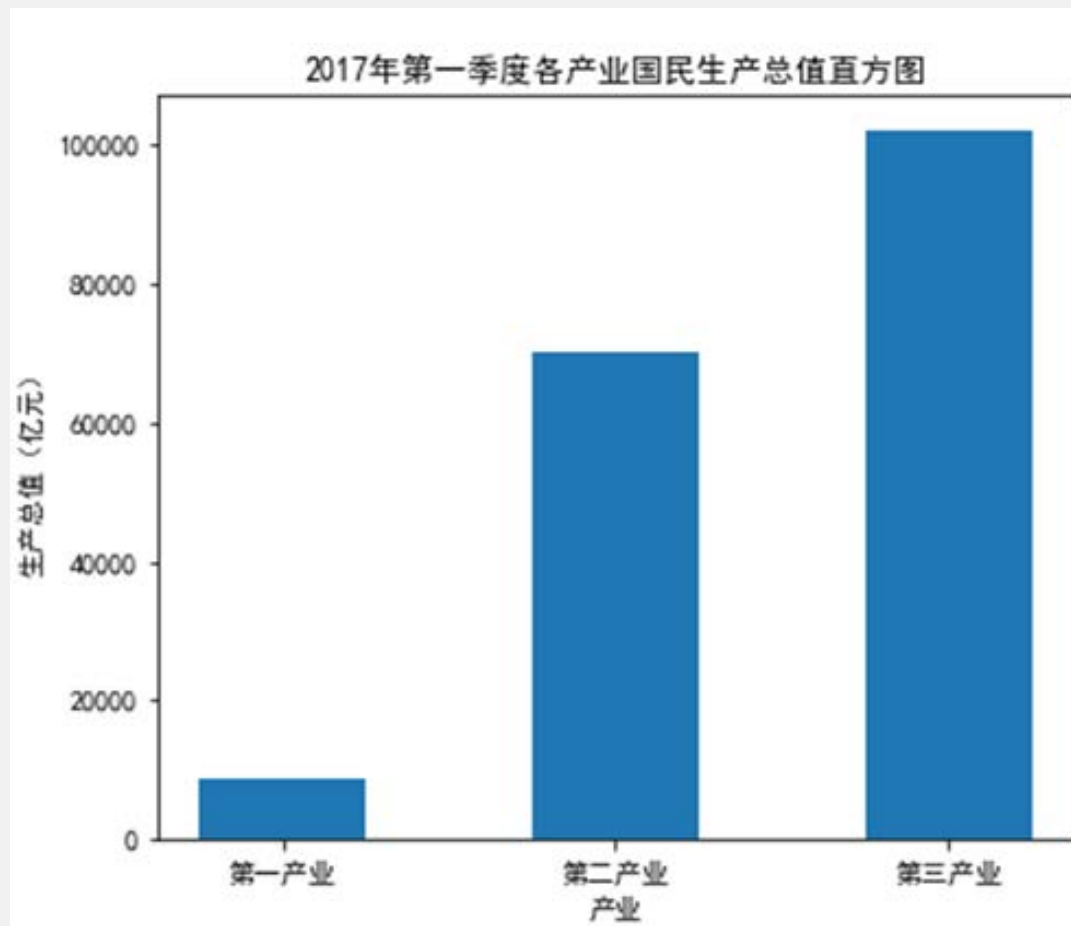
颜色缩写	代表颜色	颜色缩写	代表颜色
b	蓝色	m	品红
g	绿色	y	黄色
r	红色	k	黑色
c	青色	w	白色

课程目录

Course catalogue

- 1/ 了解绘图基础语法与常用参数
- 2/ 分析特征间的关系
- 3/ 分析特征内部数据分布与分散状况

绘制直方图



直方图 (Histogram) 又称质量分布图, 是统计报告图的一种, 由一系列高度不等的纵向条纹或线段表示数据分布的情况, 一般用横轴表示数据所属类别, 纵轴表示数量或者占比。

用直方图可以比较直观地看出产品质量特性的分布状态, 便于判断其总体质量分布情况。直方图可以发现分布表无法发现的数据模式、样本的频率分布和总体的分布

绘制直方图

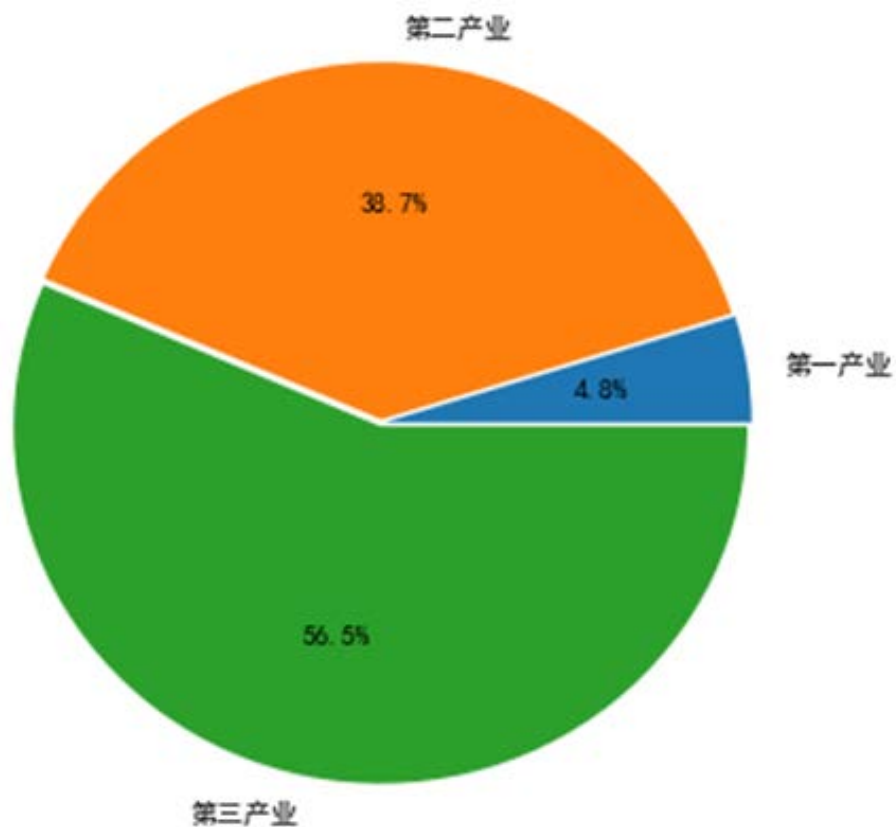
*matplotlib.pyplot.bar (left, height, width = 0.8, bottom = None, hold = None, data = None, **kwargs)*

常用参数及说明如下表所示：

参数名称	说明
left	接收array。表示x轴数据。无默认。
height	接收array。表示x轴所代表数据的数量。无默认。
width	接收0-1之间的float。指定直方图宽度。默认为0.8。
color	接收特定string或者包含颜色字符串的array。表示直方图颜色。默认为None。

绘制饼图

2017年第一季度各产业国民生产总值饼图



饼图 (Pie Graph) 是将各项的大小与各项总和的比例显示在一张“饼”中, 以“饼”的大小来确定每一项的占比。

饼图可以比较清楚地反映出部分与部分、部分与整体之间的比例关系, 易于显示每组数据相对于总数的大小, 而且显现方式直观。

绘制饼图

```
matplotlib.pyplot.pie(x, explode=None, labels=None, colors=None, autopct=None, pctdistance=0.6, shadow=False, labeldistance=1.1, startangle=None, radius=None, ... )
```

• 常用参数及说明如下表所示。

参数名称	说明	参数名称	说明
x	接收array。表示用于绘制撇的数据。无默认。	autopct	接收特定string。指定数值的显示方式。默认为None。
explode	接收array。表示指定项离饼图圆心为n个半径。默认为None。	pctdistance	接收float。指定每一项的比例和距离饼图圆心n个半径。默认为0.6。
labels	接收array。指定每一项的名称。默认为None。	labeldistance	接收float。指定每一项的名称和距离饼图圆心多少个半径。默认为1.1。
color	接收特定string或者包含颜色字符串的array。表示饼图颜色。默认为None。	radius	接收float。表示饼图的半径。默认为1。