



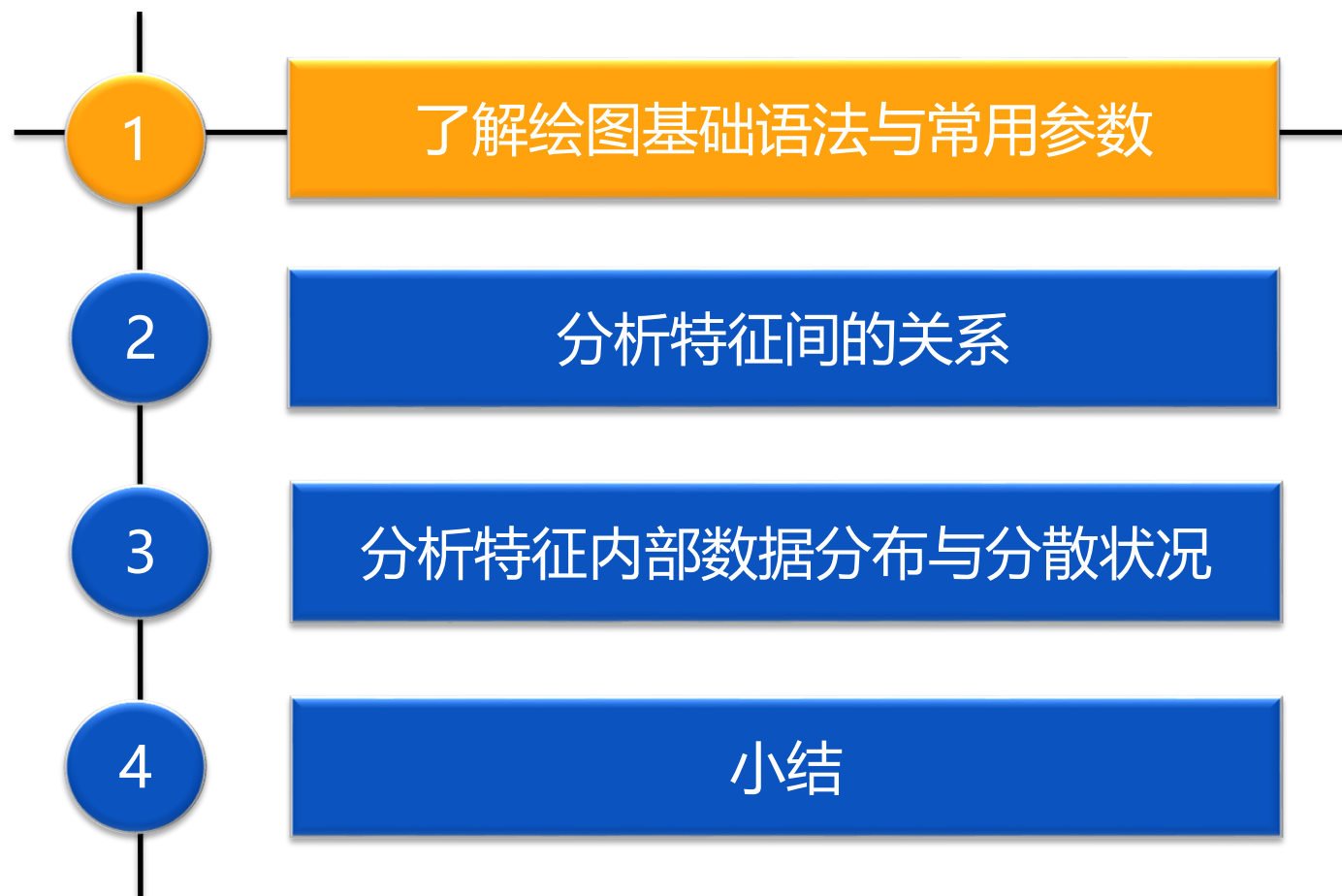
大数据，成就未来



Matplotlib数据可视化基础

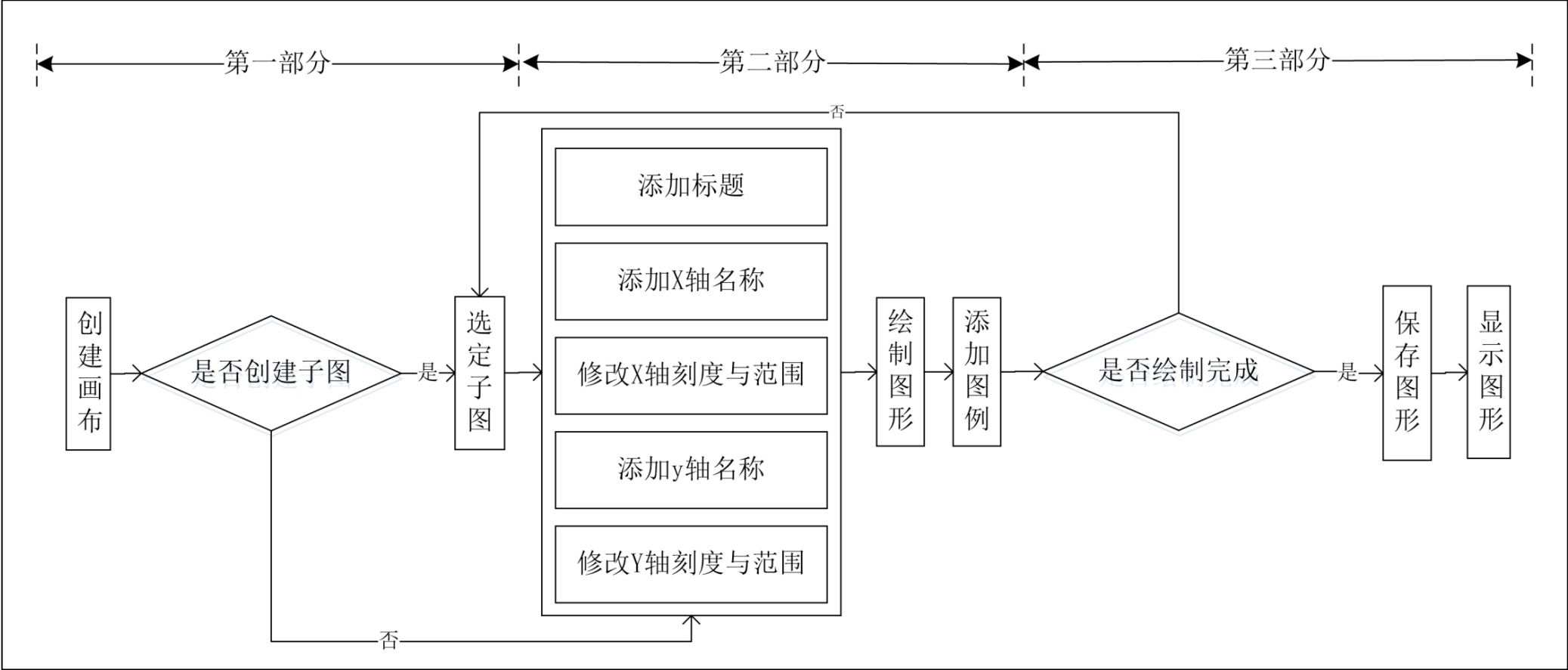
2019/9/8

目录



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.存与展示图形

第三部分主要用于保存和显示图形。

函数名称	函数作用
plt.savefig	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。
plt.show	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

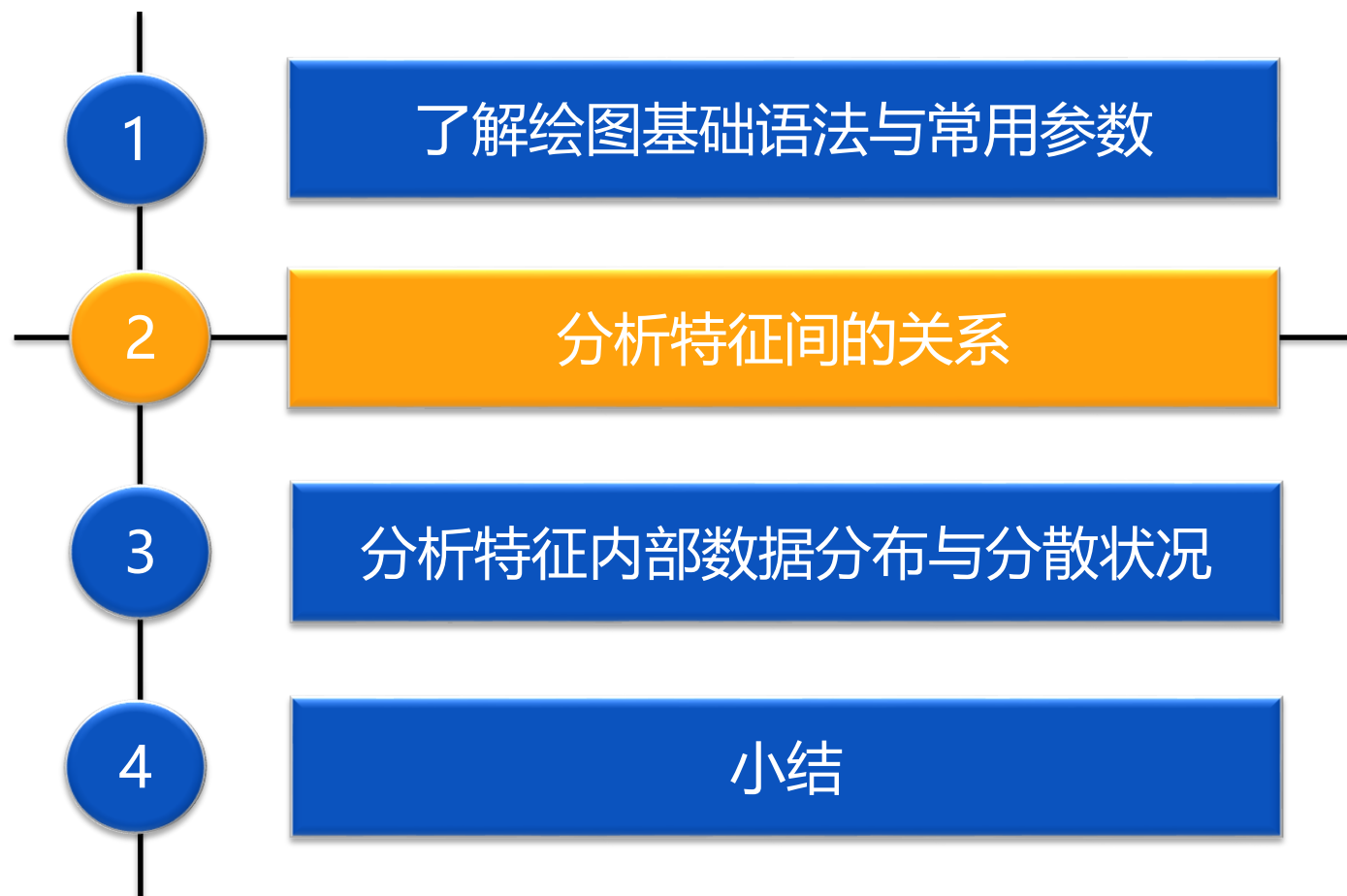
marker取值	意义	marker取值	意义
'o'	圆圈	'.'	点
'D'	菱形	's'	正方形
'h'	六边形1	'*'	星号
'H'	六边形2	'd'	小菱形
'-'	水平线	'v'	一角朝下的三角形
'8'	八边形	'<'	一角朝左的三角形
'p'	五边形	'>'	一角朝右的三角形
' , '	像素	'^'	一角朝上的三角形
'+'	加号	'\'	竖线
'None'	无	'x'	X

设置pyplot的动态rc参数

注意事项

- 由于默认的pyplot字体并不支持中文字符的显示，因此需要通过设置font.sans-serif参数改变绘图时的字体，使得图形可以正常显示中文。同时，由于更改字体后，会导致坐标轴中的部分字符无法显示，因此需要同时更改axes.unicode_minus参数。
- 除了设置线条和字体的rc参数外，还有设置文本、箱线图、坐标轴、刻度、图例、标记、图片、图像保存等rc参数。具体参数与取值可以参考官方文档。

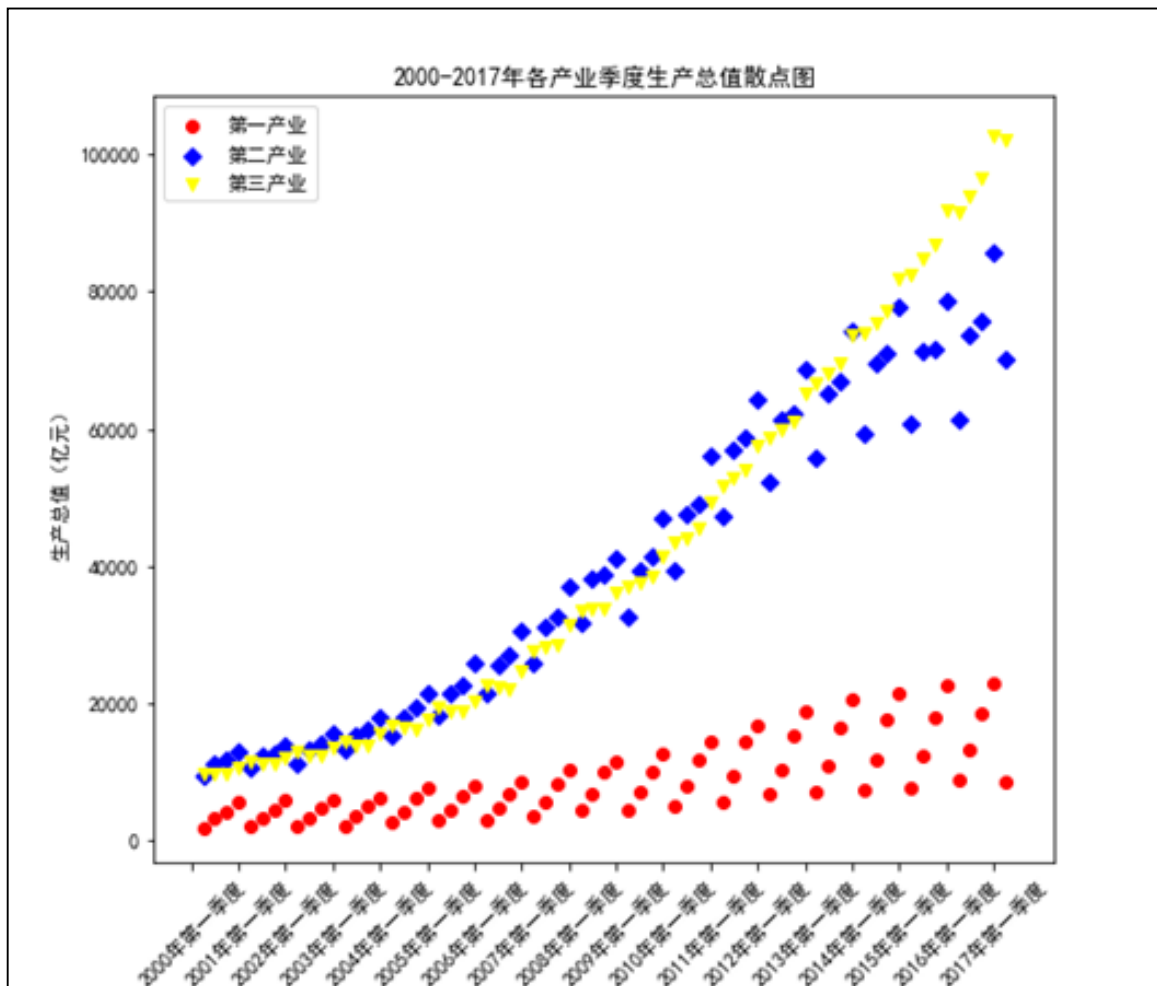
目录



绘制散点图

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

scatter函数

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

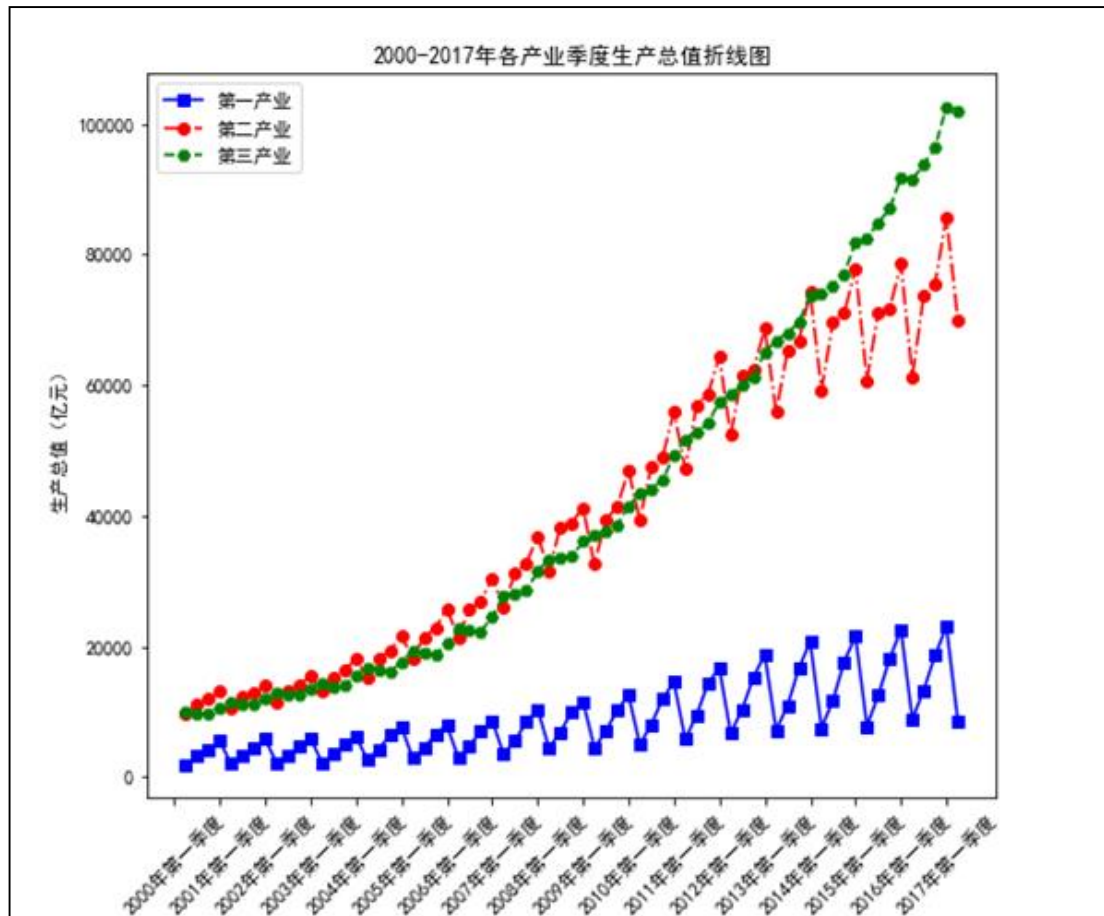
➤ 常用参数及说明如下表所示。

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

- 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

plot函数

```
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。

绘制折线图

plot函数

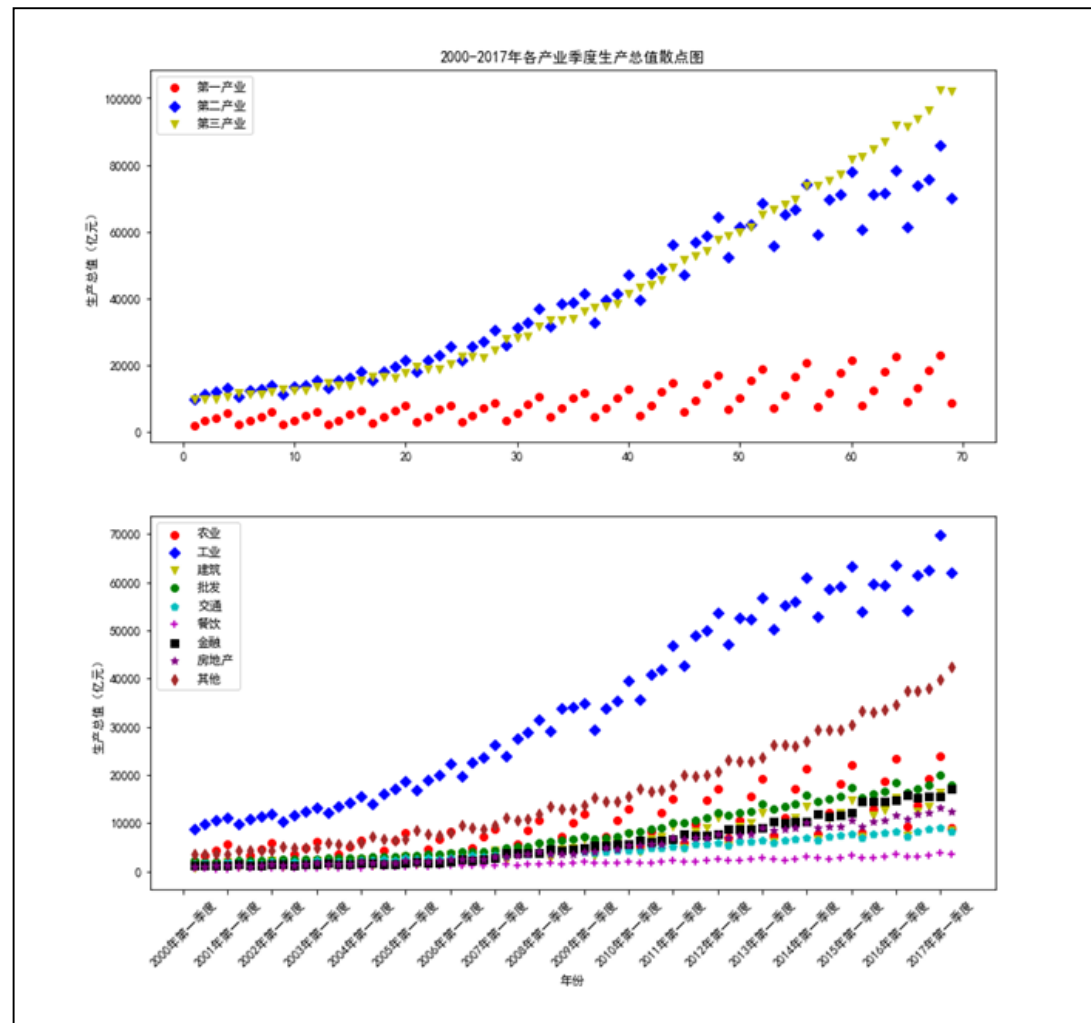
color参数的8种常用颜色的缩写。

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

任务实现

1.绘制2000-2017各产业与行业的国民生产总值散点图

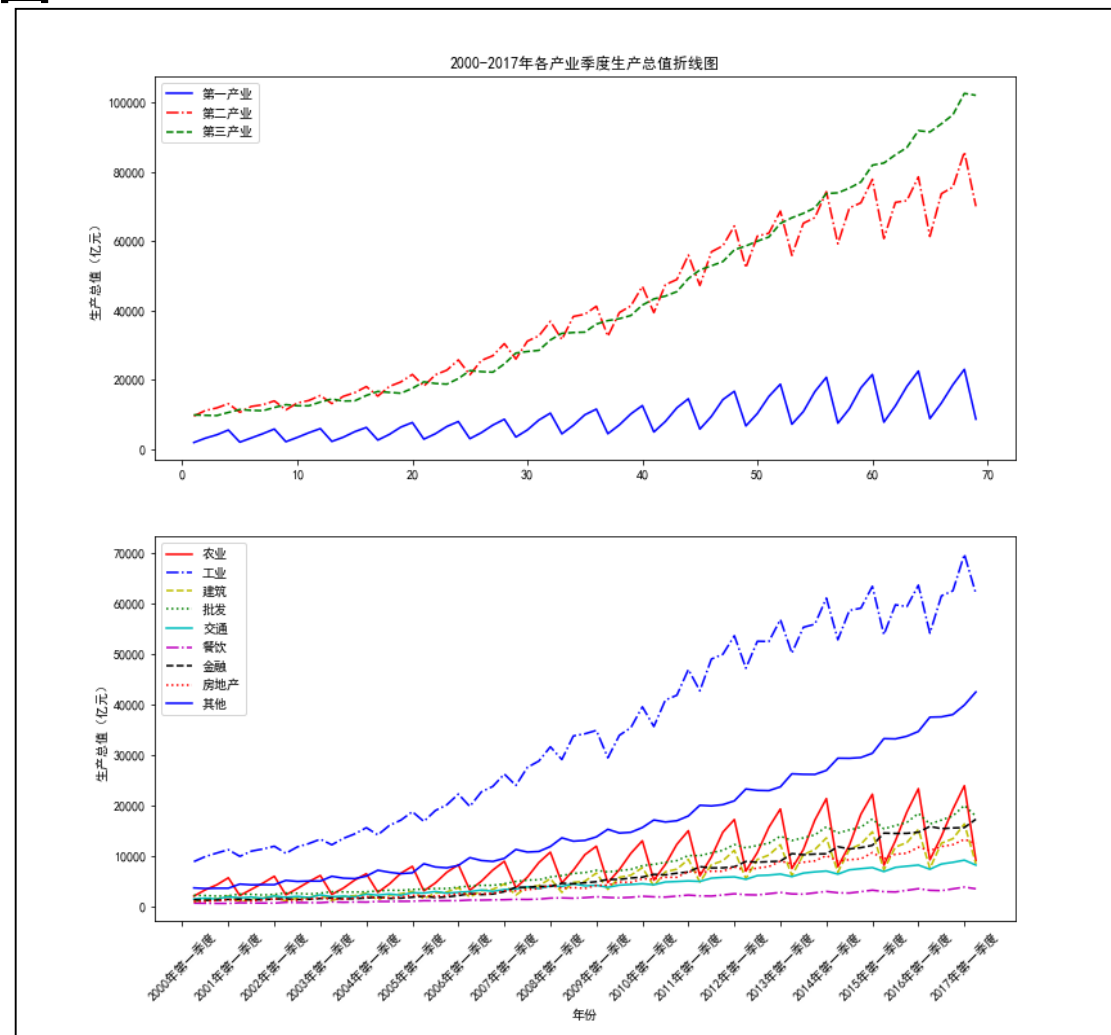
- 国民生产总值数据总共有三大产业的国民生产总值，以及农业、工业、建筑、批发、交通、餐饮、金融、房地产和其他行业各个季度的增加值。
- 通过散点图分析三大行业的国民生产总值可以发现我国产业结构。通过比较各行业间季度的增加值则可以发现国民经济的主要贡献行业。



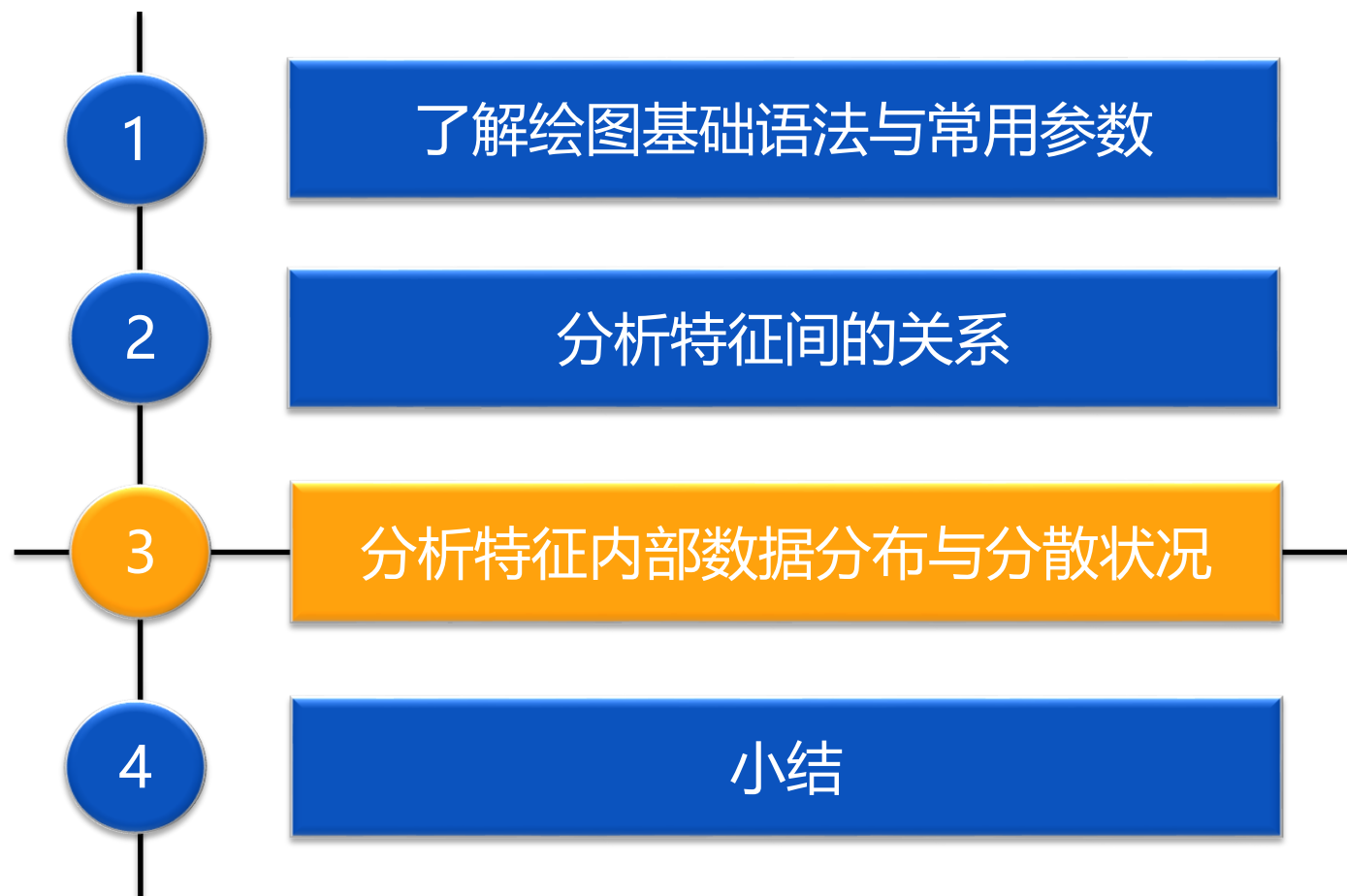
任务实现

2.绘制2000-2017各产业与行业的国民生产总值折线图

- 通过绘制2000-2017各产业与行业的国民生产总值折线图，分别能够发现我国经济各产业与各行业增长趋势。



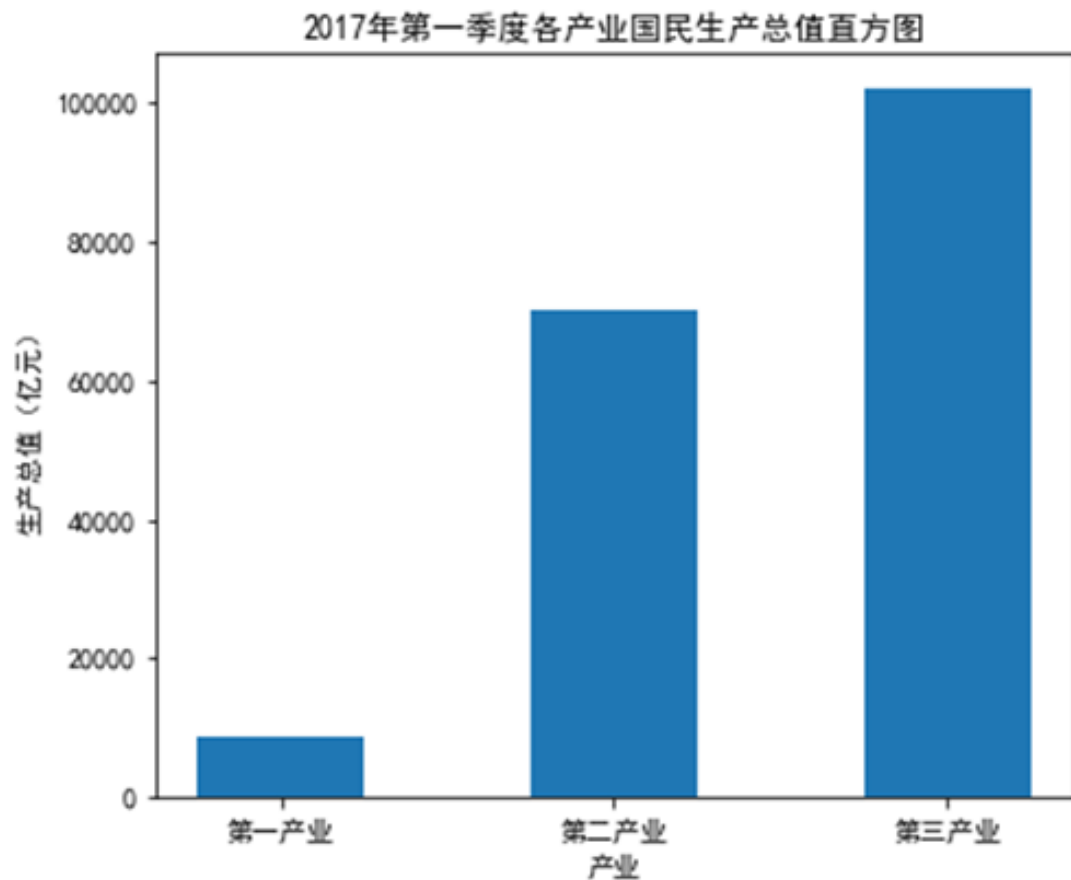
目录



绘制直方图

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

bar函数

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

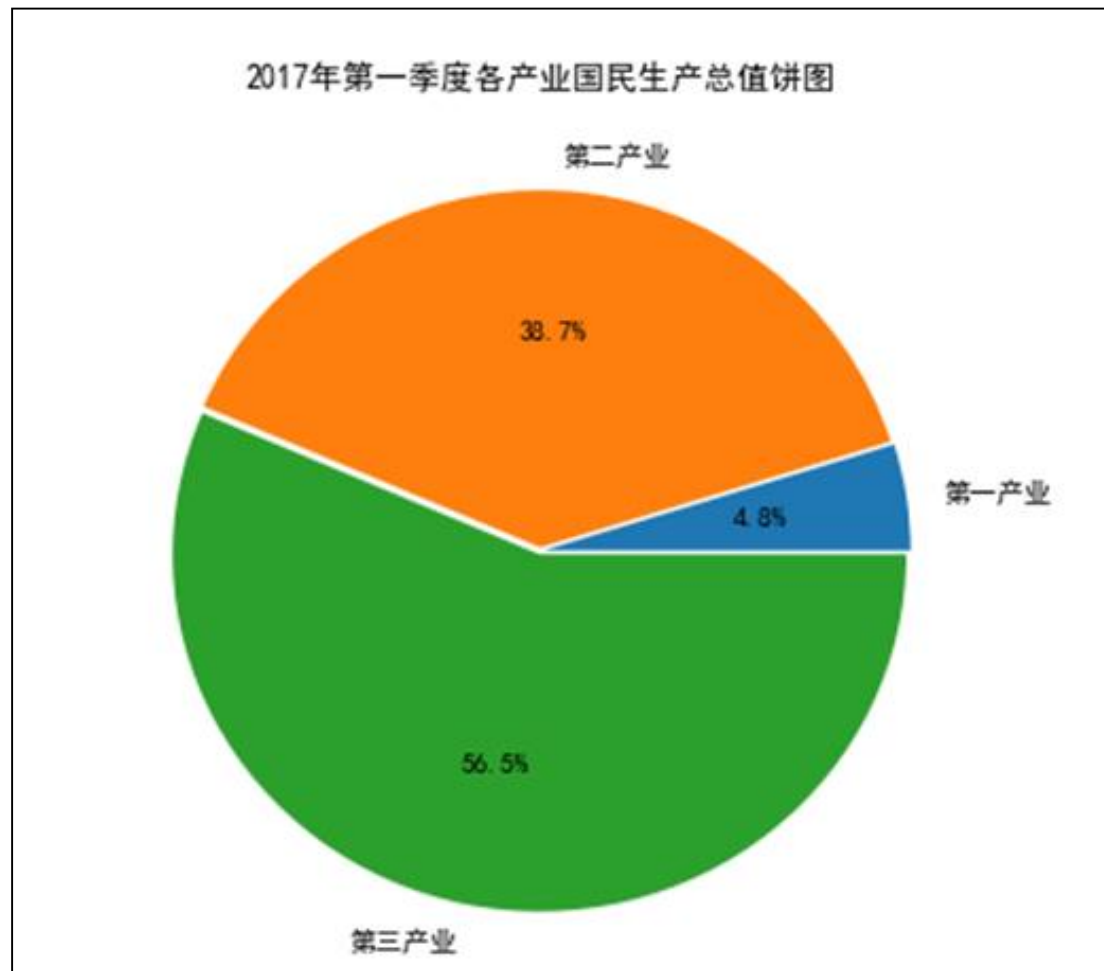
➤ 常用参数及说明如下表所示。

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

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

pie函数

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

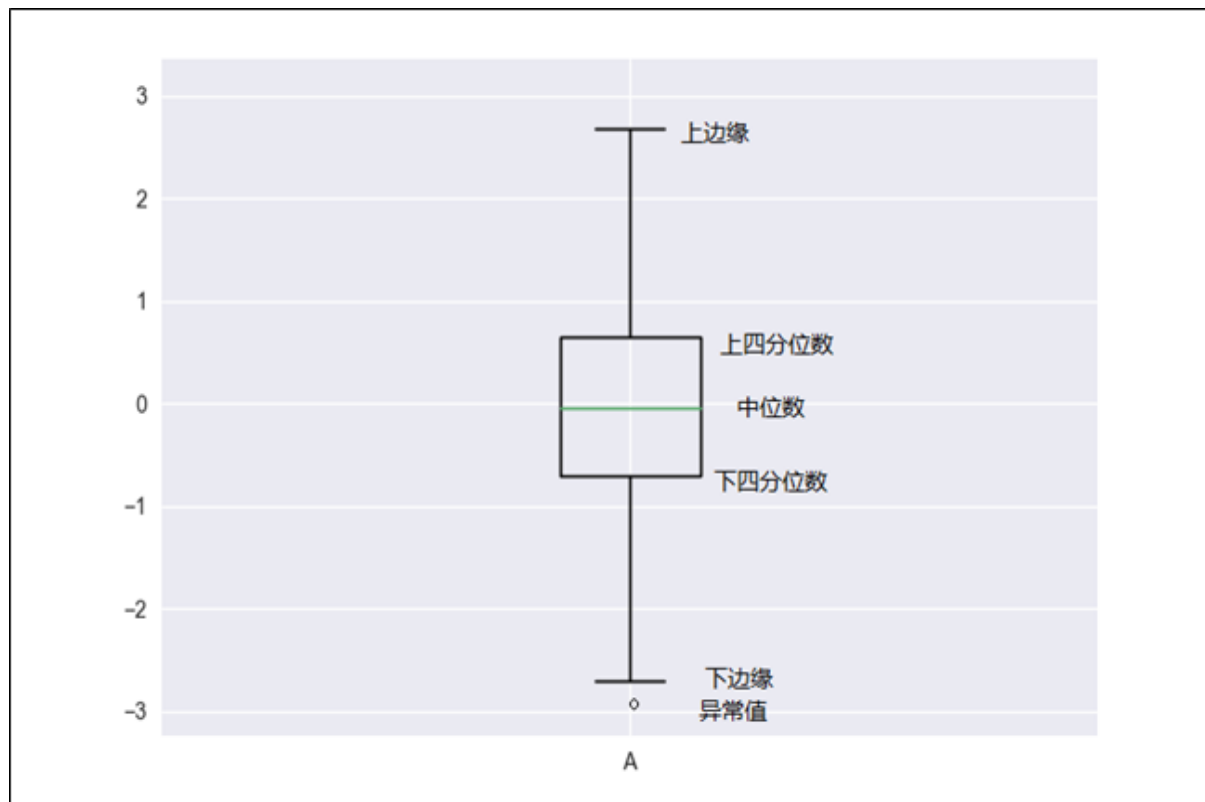
➤ 常用参数及说明如下表所示。

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。

绘制箱线图

箱线图

- 箱线图 (boxplot) 也称箱须图，其绘制需使用常用的统计量，能提供有关数据位置和分散情况的关键信息，尤其在比较不同特征时，更可表现其分散程度差异。
- 箱线图利用数据中的五个统计量（最小值、下四分位数、中位数、上四分位数和最大值）来描述数据，它也可以粗略地看出数据是否具有对称性、分布的分散程度等信息，特别可以用于对几个样本的比较。



绘制箱线图

boxplot函数

matplotlib.pyplot.**boxplot**(*x*, *notch*=None, *sym*=None, *vert*=None, *whis*=None, *positions*=None, *widths*=None, *patch_artist*=None,*meanline*=None, *labels*=None, ...)

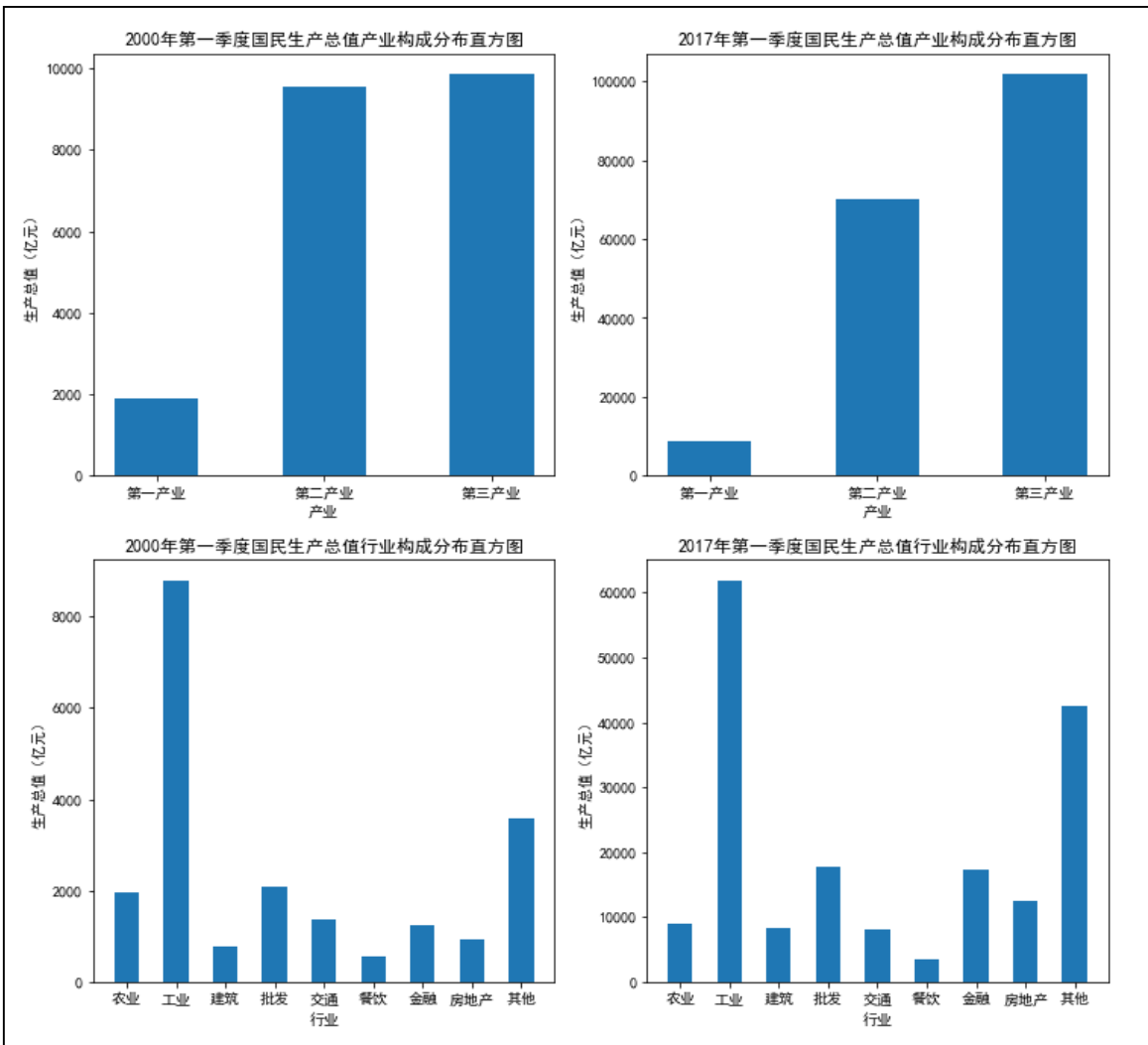
➤ 常用参数及说明如下表所示。

参数名称	说明	参数名称	说明
x	接收array。表示用于绘制箱线图的数据。无默认。	positions	接收array。表示图形位置。默认为None。
notch	接收boolean。表示中间箱体是否有缺口。默认为None。	widths	接收scalar或者array。表示每个箱体的宽度。默认为None。
sym	接收特定sting。指定异常点形状。默认为None。	labels	接收array。指定每一个箱线图的标签。默认为None。
vert	接收boolean。表示图形是横向纵向或者横向。默认为None。	meanline	接收boolean。表示是否显示均值线。默认为False。

任务实现

1. 绘制国民生产总值构成分布直方图

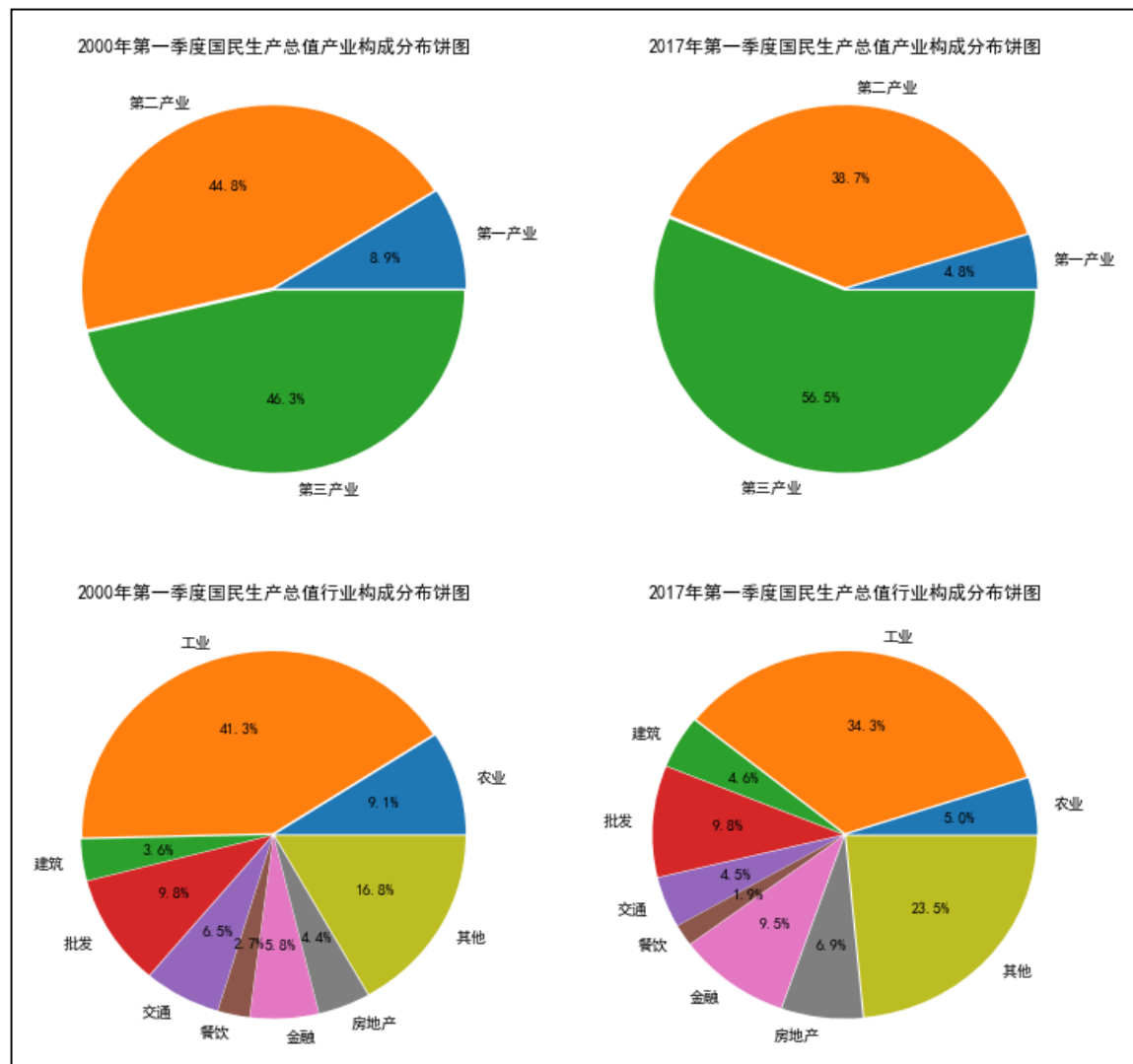
- 通过直方图分析2000年第一季度和2017年第一季度之间的三大产业的国民生产总值，可以发现各产业绝对数值之间的关系，并通过对比发现产业结构的变化。
- 同理可以得出行业间的绝对数值关系以及17年来行业发展状况。



任务实现

2.绘制国民生产总值构成分布饼图

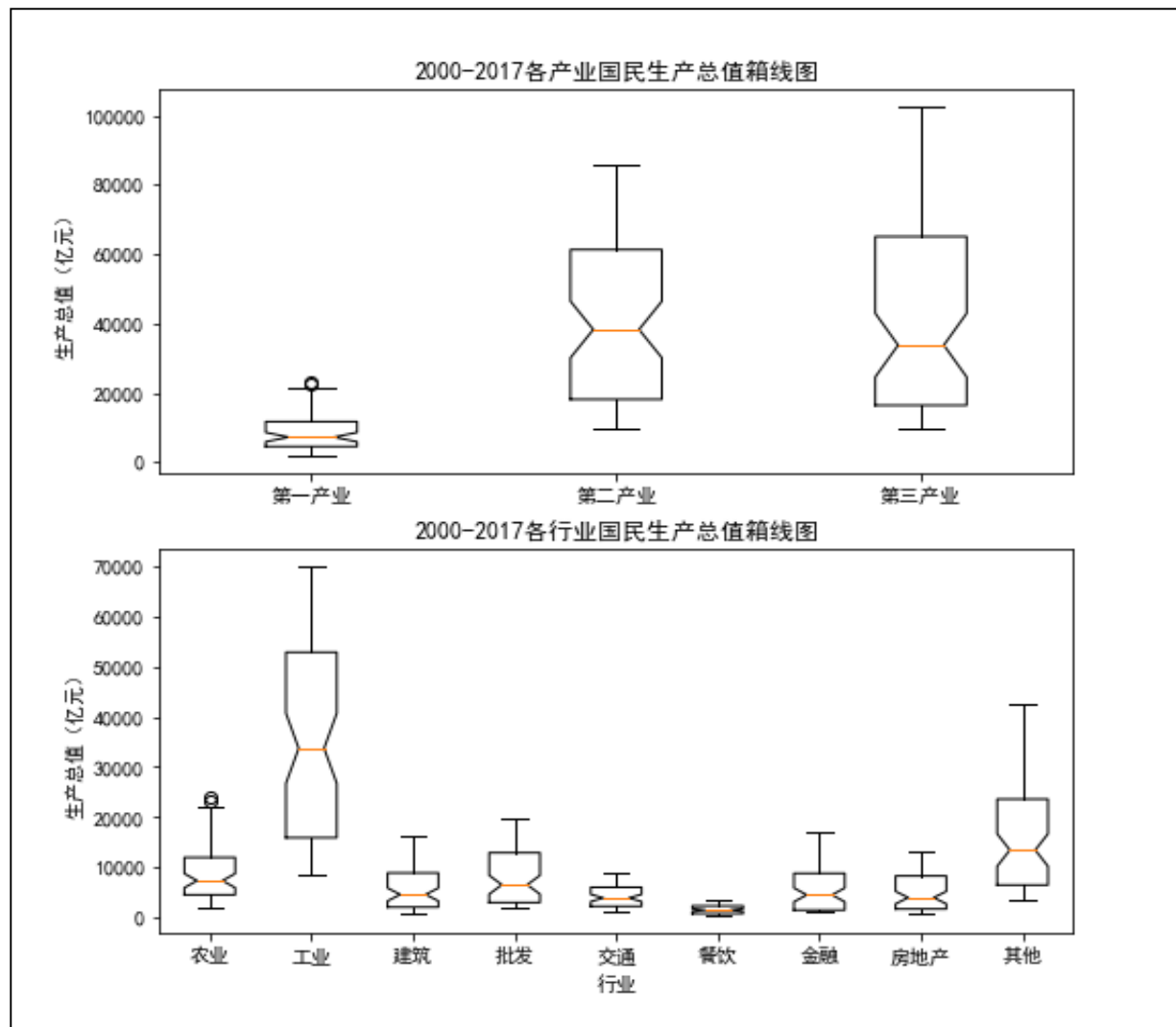
- 通过分析2000年与2017年不同的产业和行业在国民生产总值中的占比，可以发现我国产业结构变化和行业变迁。



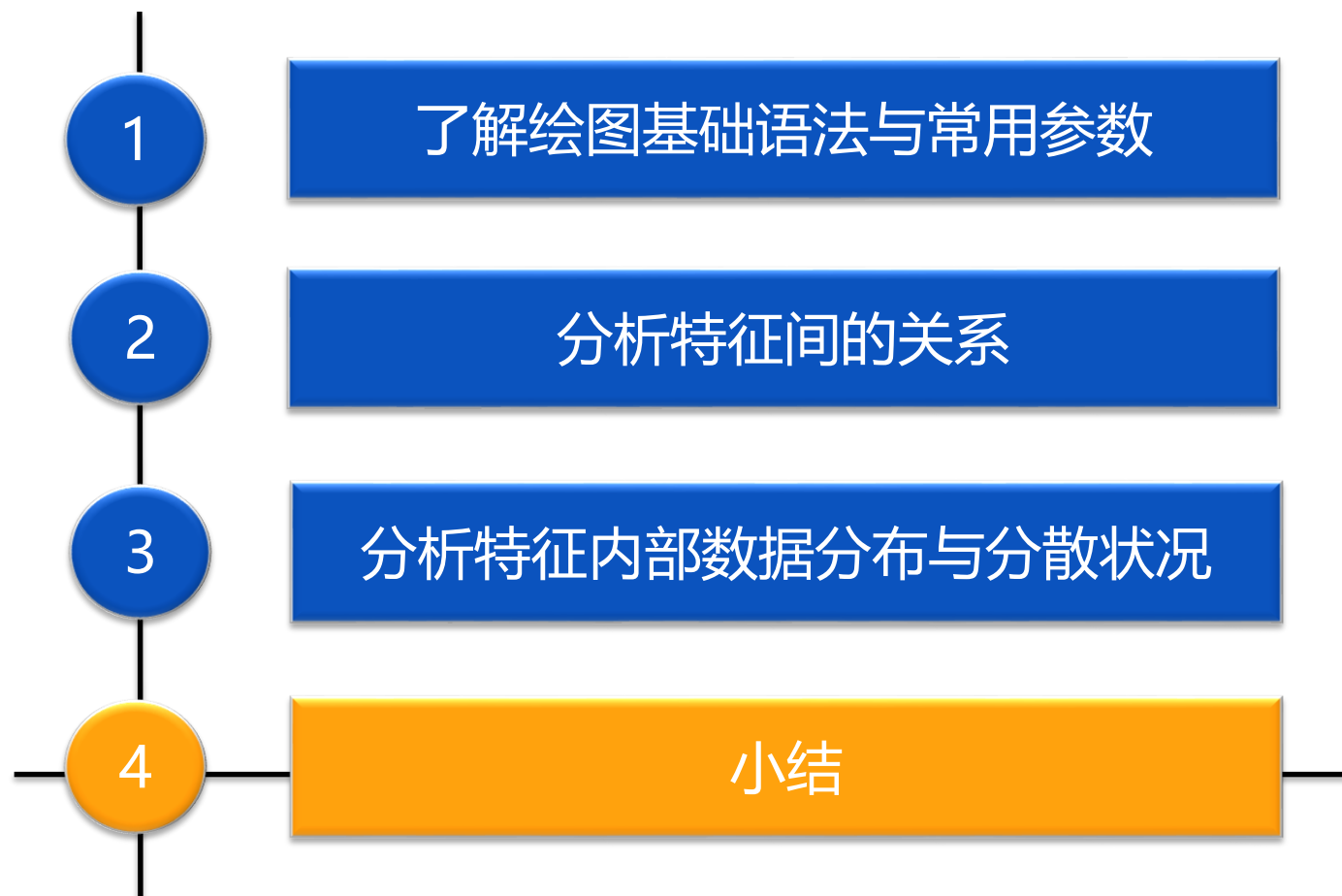
任务实现

3.绘制国民生产总值分散情况箱线图

- 通过分析2000年至2017年不同的产业和行业在国民生产总值中的分散情况，可以发现整体分散情况，从而判断整体增速是否加快。



目录



小结

本章以2000至2017年各季度国民生产总值数据为例，介绍了pyplot绘图的基本语法，常用参数。

- 介绍了分析特征间相关关系的**散点图**。
- 分析特征间趋势关系的**折线图**。
- 分析特征内部数据分布的**直方图**和**饼状图**。
- 以及分析特征内部数据分散情况的**箱线图**。

为读者后续深入学习Matplotlib数据可视化打下了深厚的基础。



大数据，成就未来



Thank you!