

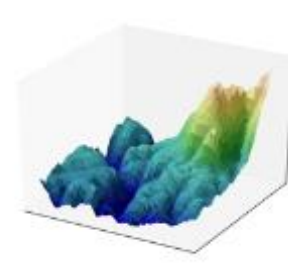
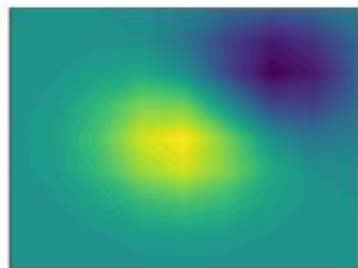
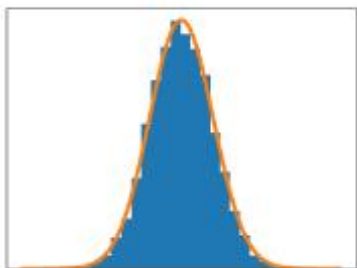
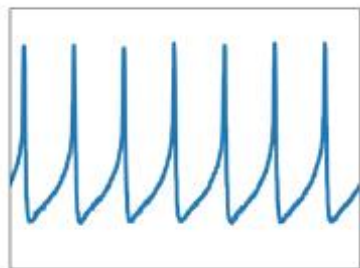


Matplotlib数据可视化基础

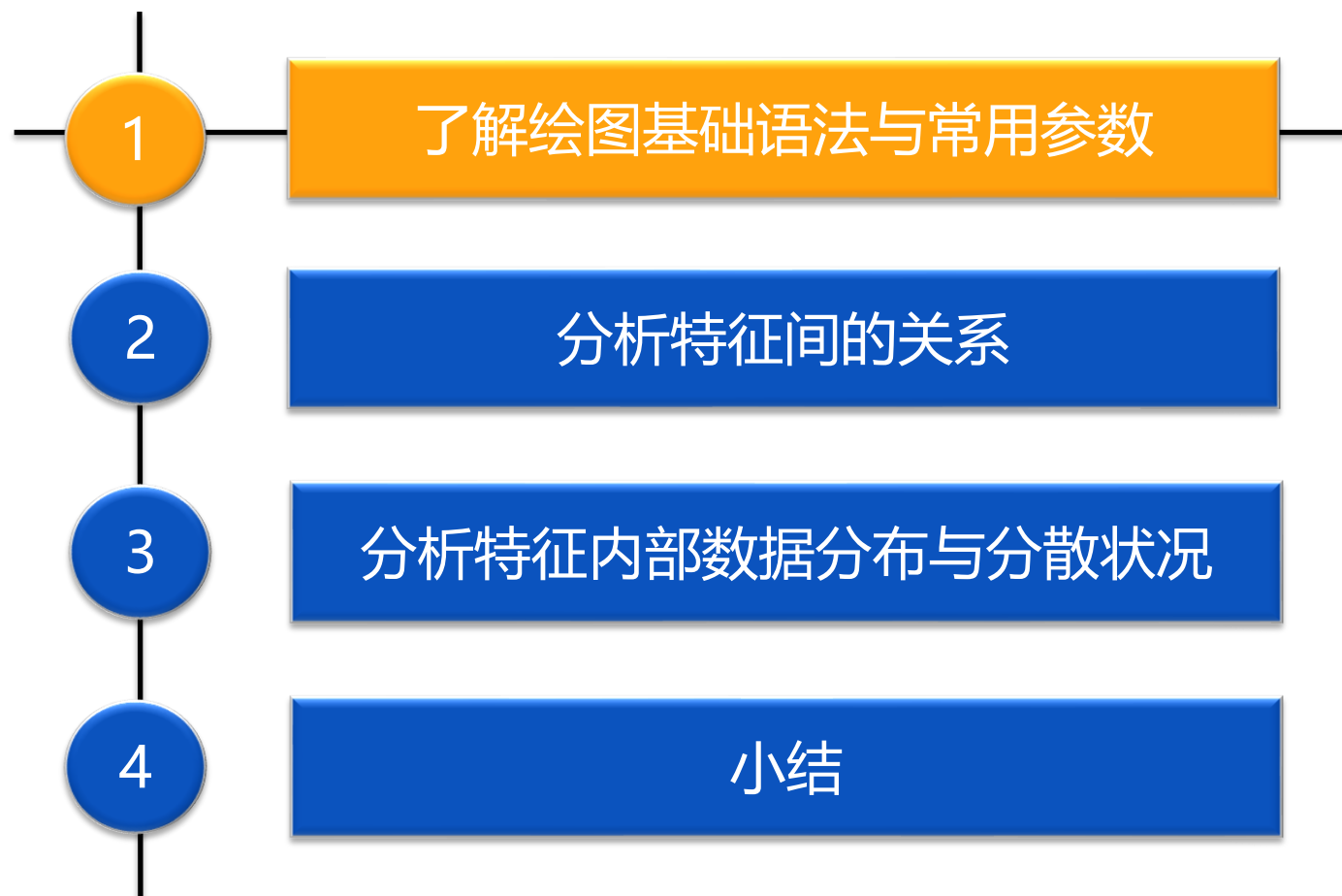
Matplotlib库介绍



- Matplotlib是Python中最常用的可视化工具之一，可以非常方便地创建海量类型的2D图表和一些基本的3D图表，可根据数据集（DataFrame，Series）自行定义x,y轴，绘制图形（线形图，柱状图，直方图，密度图，散布图等等），能够满足大部分需要。
- Matplotlib最早是为了可视化癫痫病人的脑皮层电图相关的信号而研发，因为在函数的设计上参考了MATLAB，所以叫做Matplotlib。
- 官方文档: <https://matplotlib.org/>
- Matplotlib中最基础的模块是pyplot。

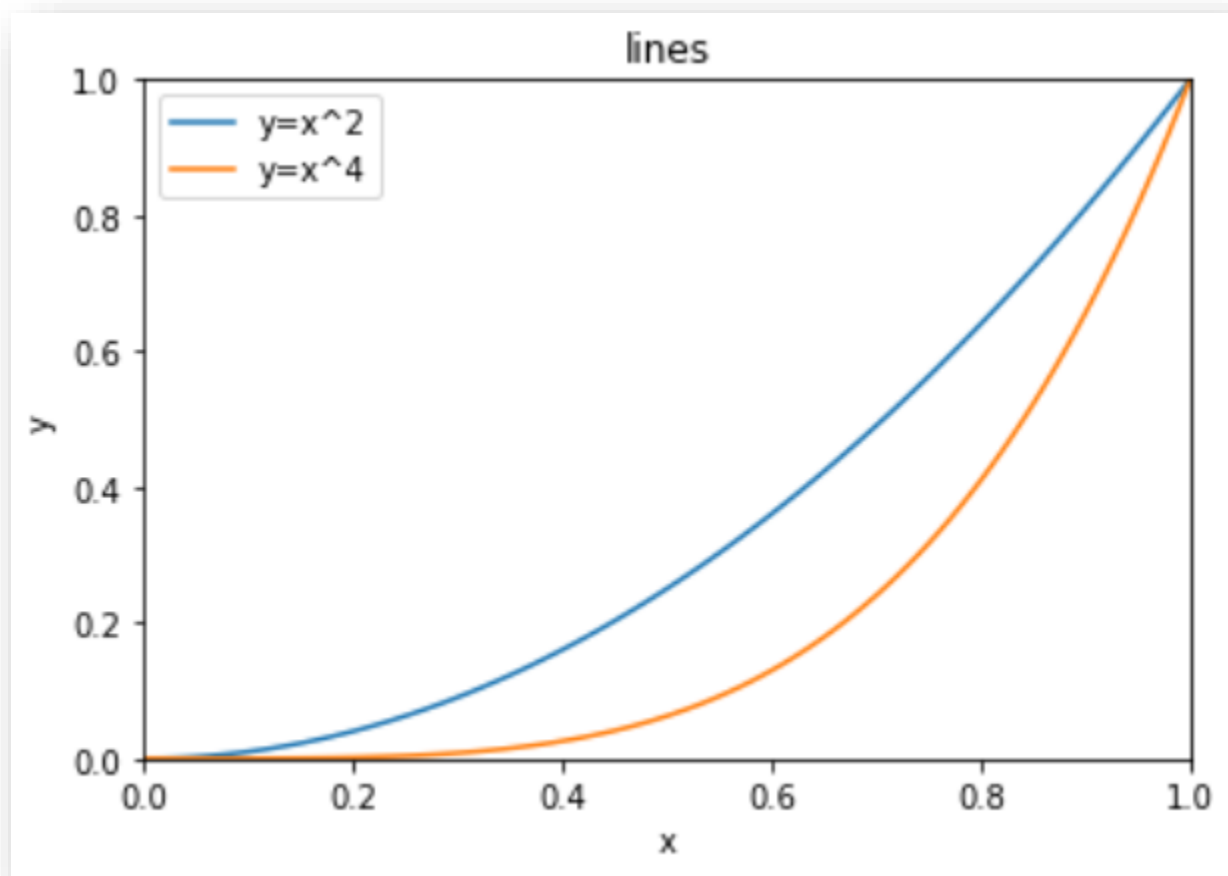


目录



了解绘图基础语法与常用参数

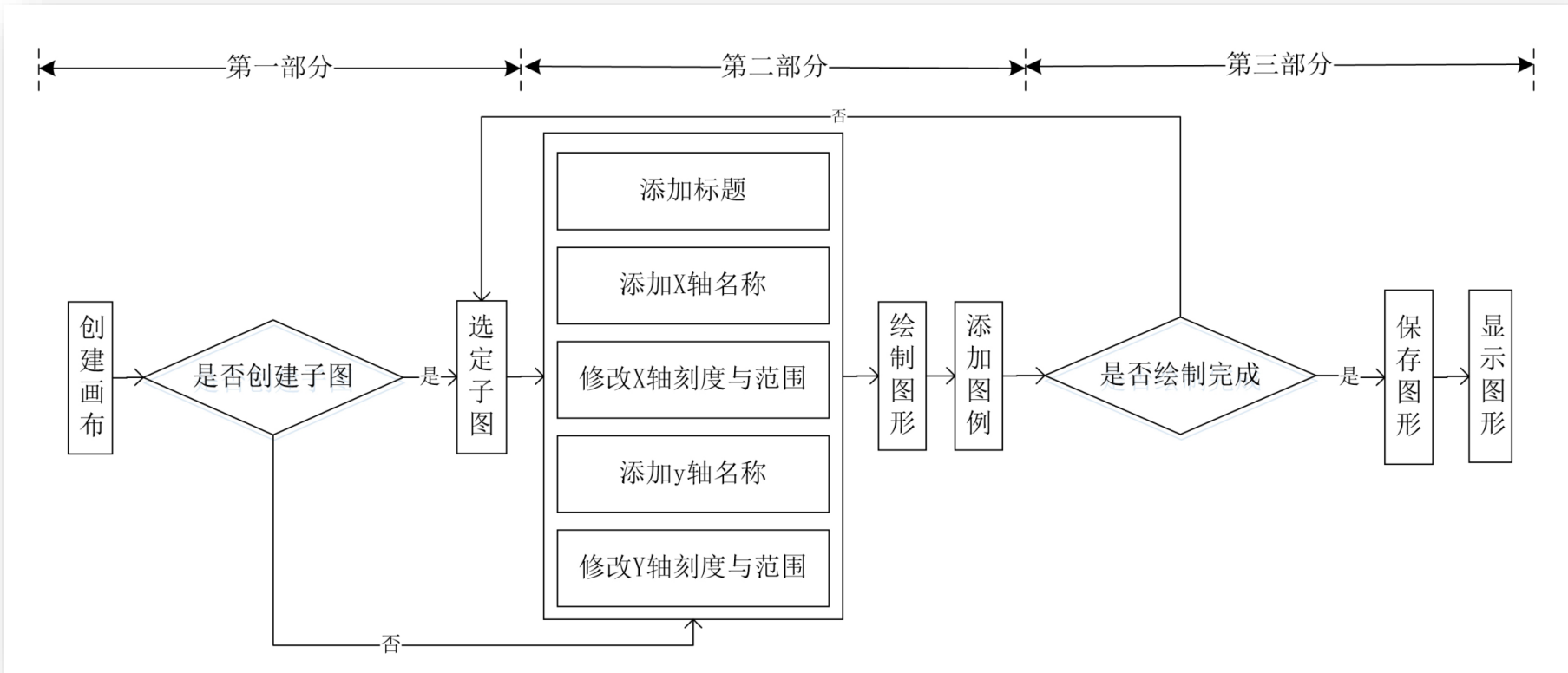
图形示例



了解绘图基础语法与常用参数

掌握pyplot基础语法

基本绘图流程



了解绘图基础语法与常用参数

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 |
|---------------------------------|-------------------------------|
| <code>plt.figure</code> | 创建一个空白画布，可以指定画布大小，像素。 |
| <code>figure.add_subplot</code> | 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。 |

了解绘图基础语法与常用参数

掌握pyplot基础语法

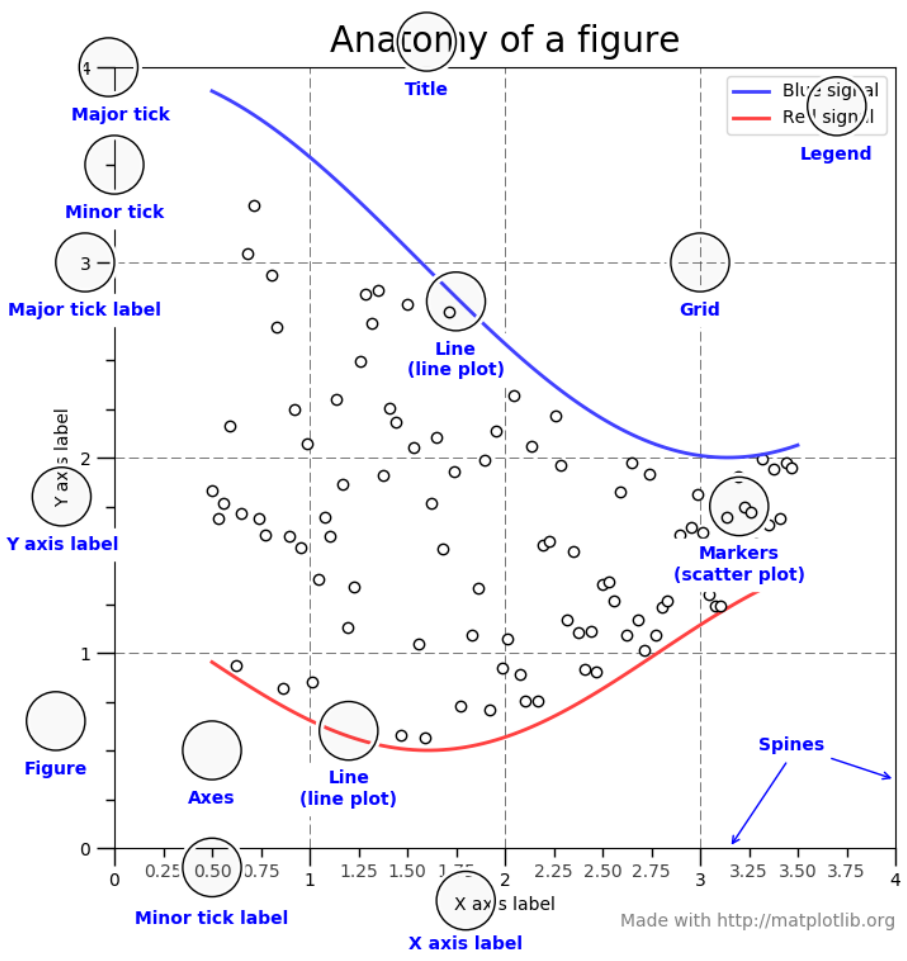
2.添加画布内容

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 |
|-------------------------|-------------------------------------|
| <code>plt.title</code> | 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。 |
| <code>plt.xlabel</code> | 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。 |
| <code>plt.ylabel</code> | 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。 |
| <code>plt.xlim</code> | 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。 |
| <code>plt.ylim</code> | 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。 |
| <code>plt.xticks</code> | 指定x轴刻度的数目与取值。 |
| <code>plt.yticks</code> | 指定y轴刻度的数目与取值。 |
| <code>plt.legend</code> | 指定当前图形的图例，可以指定图例的大小、位置、标签。 |

了解绘图基础语法与常用参数

图参数说明



了解绘图基础语法与常用参数

掌握pyplot基础语法

3.保存与展示图形

第三部分主要用于保存和显示图形。

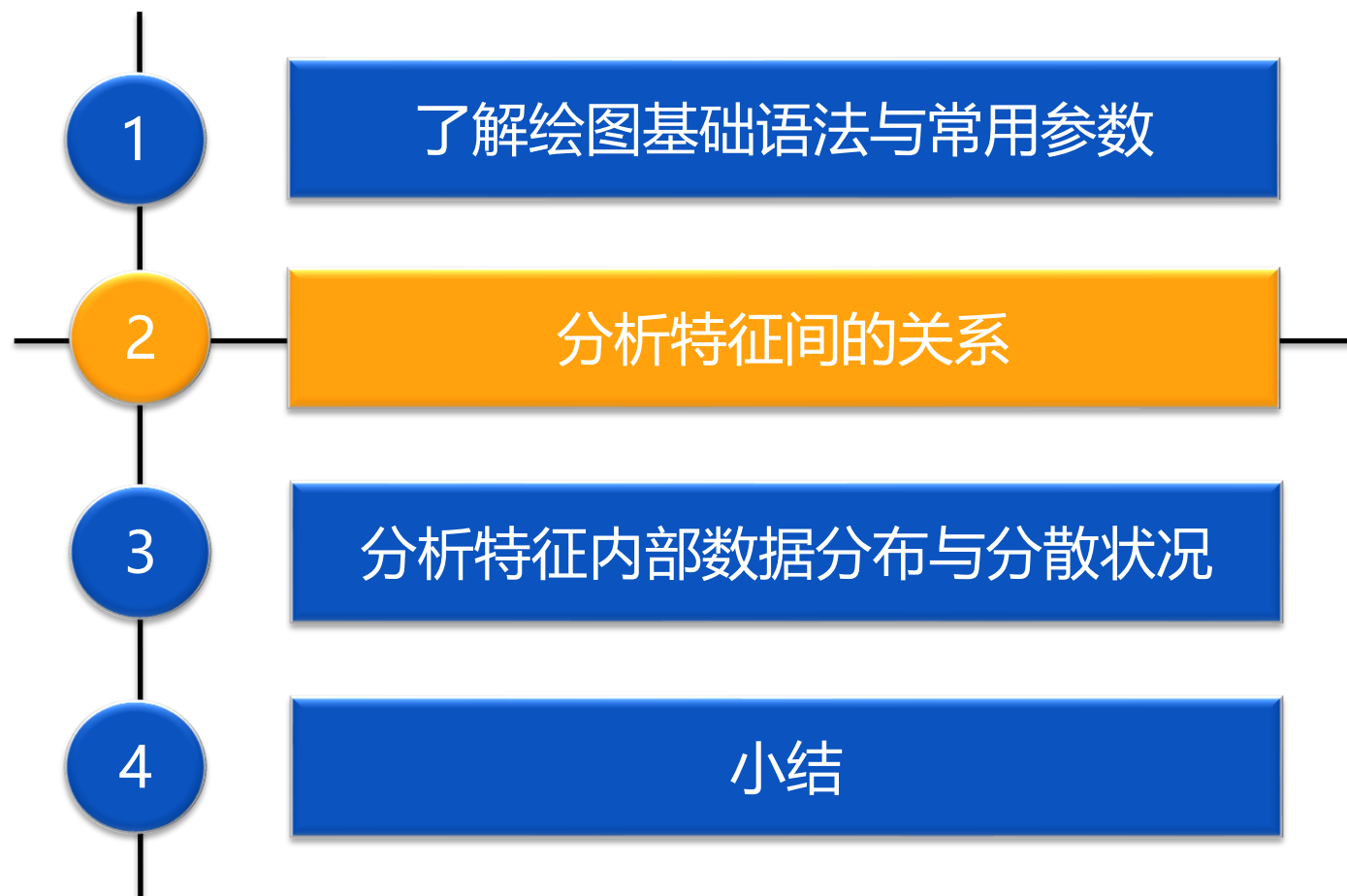
| 函数名称 | 函数作用 |
|-------------|------------------------------|
| plt.savefig | 保存绘制的图片，可以指定图片的分辨率、边缘的颜色等参数。 |
| plt.show | 在本机显示图形。 |

了解绘图基础语法与常用参数

设置pyplot的动态rc参数

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

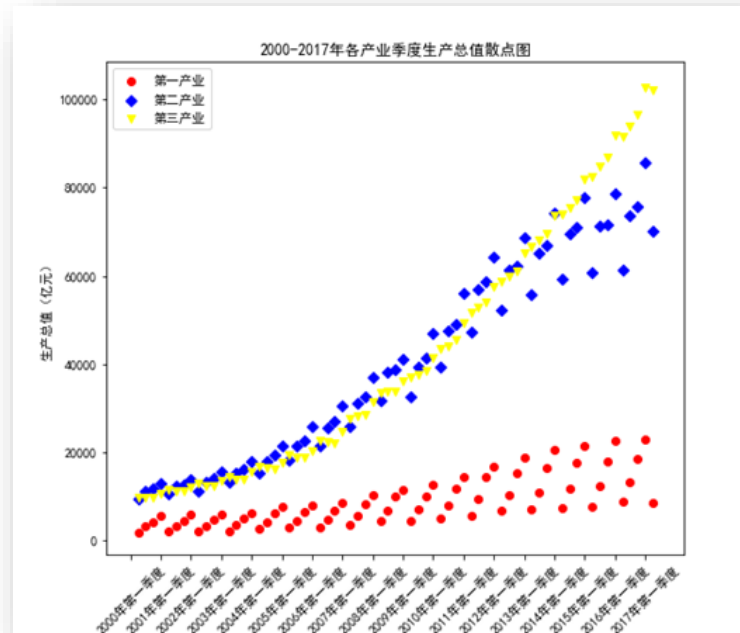
目录



分析特征间的关系

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。



分析特征间的关系

绘制散点图

scatter函数

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

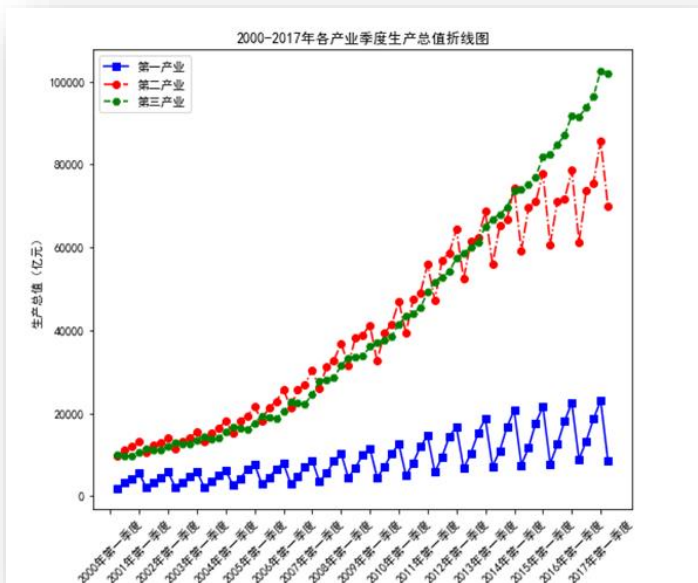
常用参数及说明如下表所示

| 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。 |

分析特征间的关系

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。



分析特征间的关系

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。 |

分析特征间的关系

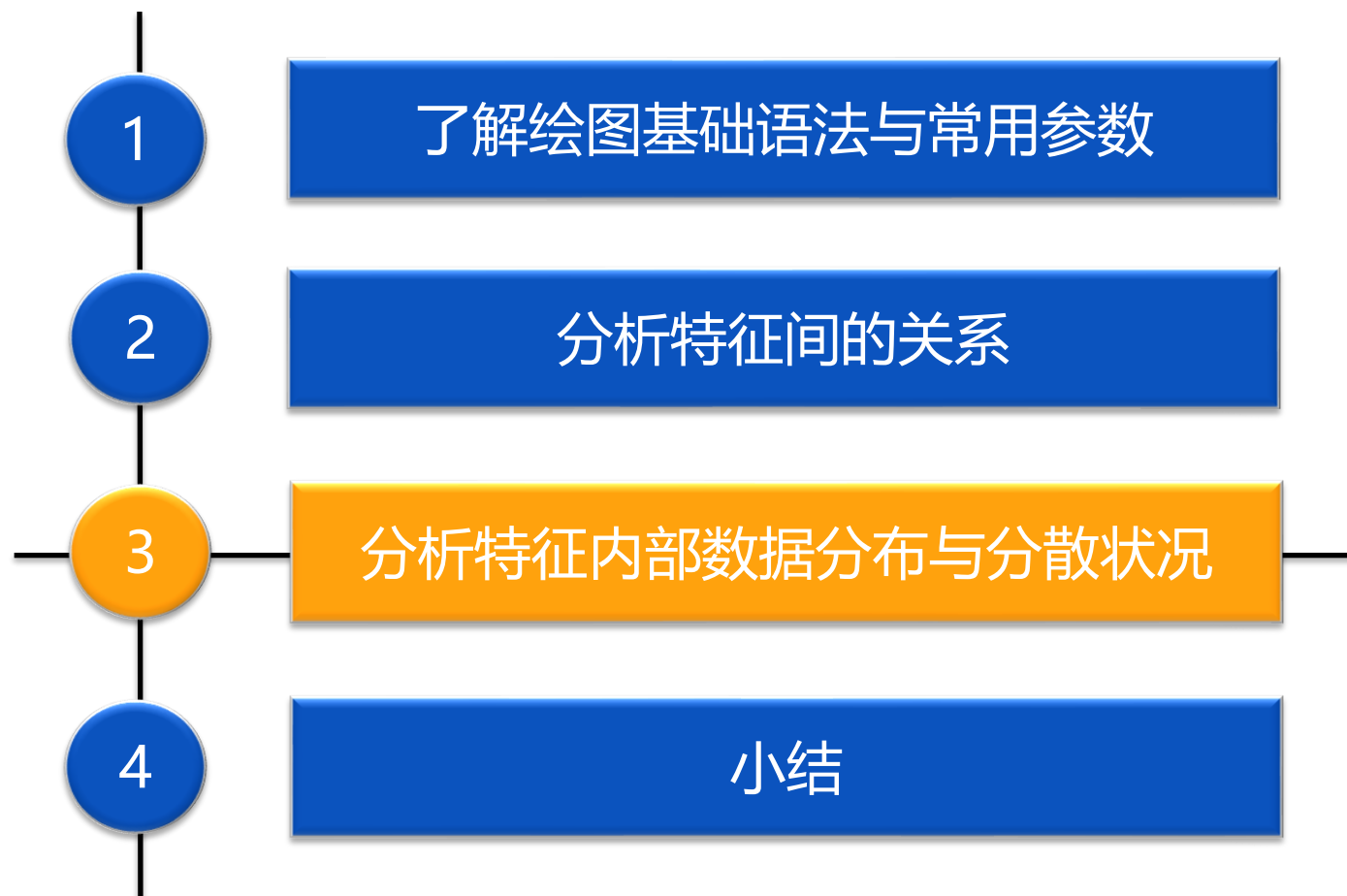
绘制折线图

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 |

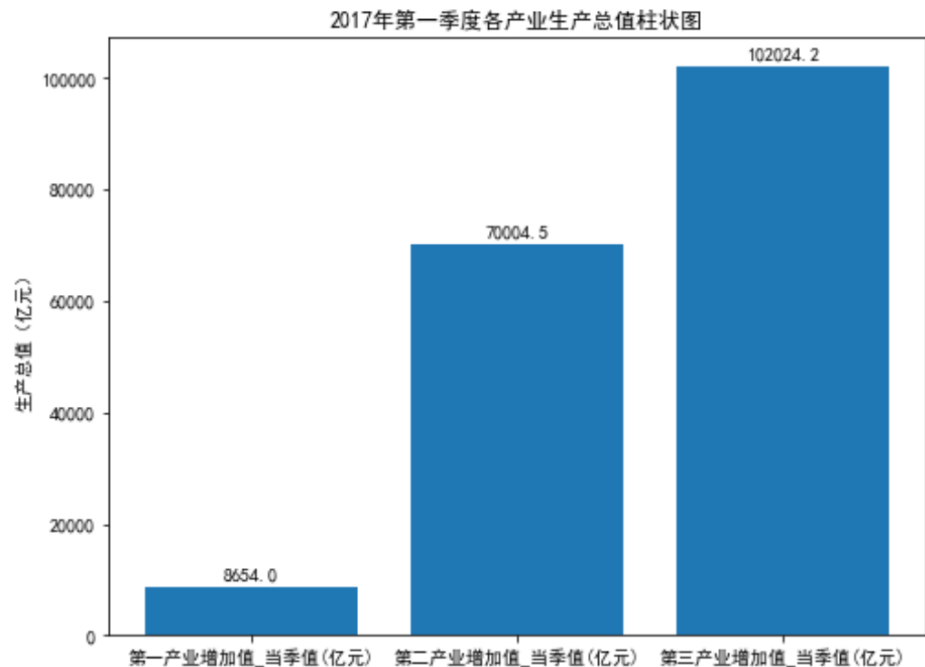
目录



分析特征内部数据分布与分散状况

绘制柱状图

- 柱状图，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。
- 用柱状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。柱状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。



分析特征内部数据分布与分散状况

绘制柱状图

bar函数

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

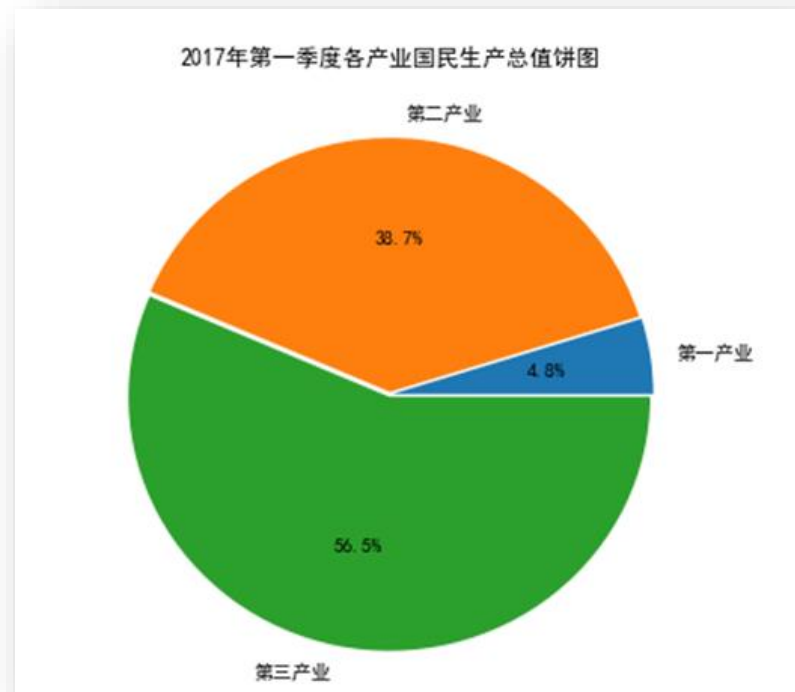
常用参数及说明如下表所示。

| 参数名称 | 说明 |
|--------|--|
| x | 接收array。表示x轴数据或坐标值。无默认。 |
| height | 接收array。决定了柱子高度。无默认。 |
| width | 接收0-1之间的float。指定柱状图宽度。默认为0.8。 |
| color | 接收特定string或者包含颜色字符串的array。表示柱状图颜色。默认为None。 |

分析特征内部数据分布与分散状况

绘制饼图

- 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。



分析特征内部数据分布与分散状况

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, ...)

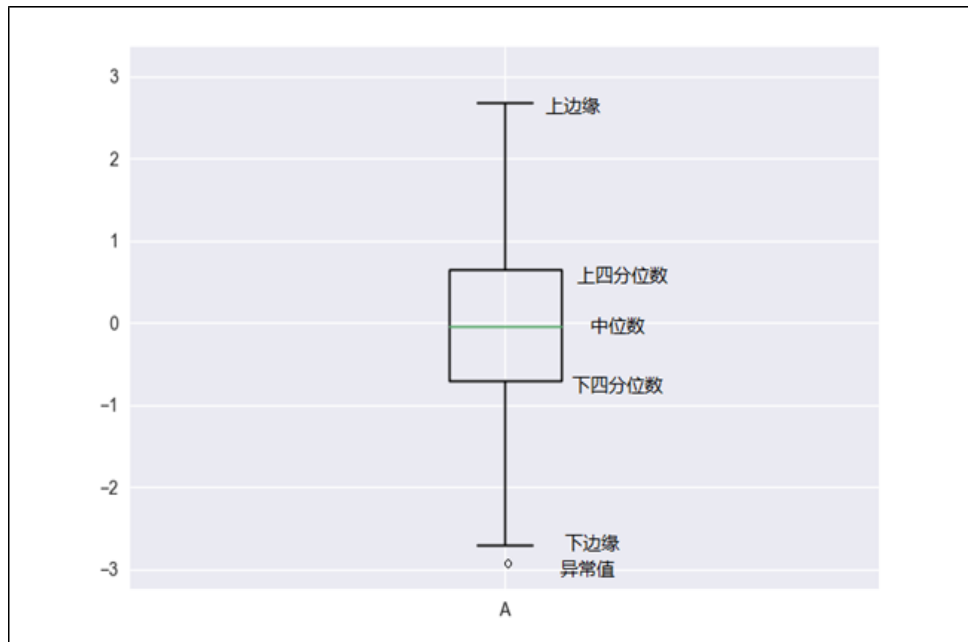
常用参数及说明如下表所示。

| 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。 |

分析特征内部数据分布与分散状况

绘制箱线图

- 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 (下边缘、下四分位数、中位数、上四分位数和上边缘) 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。



分析特征内部数据分布与分散状况

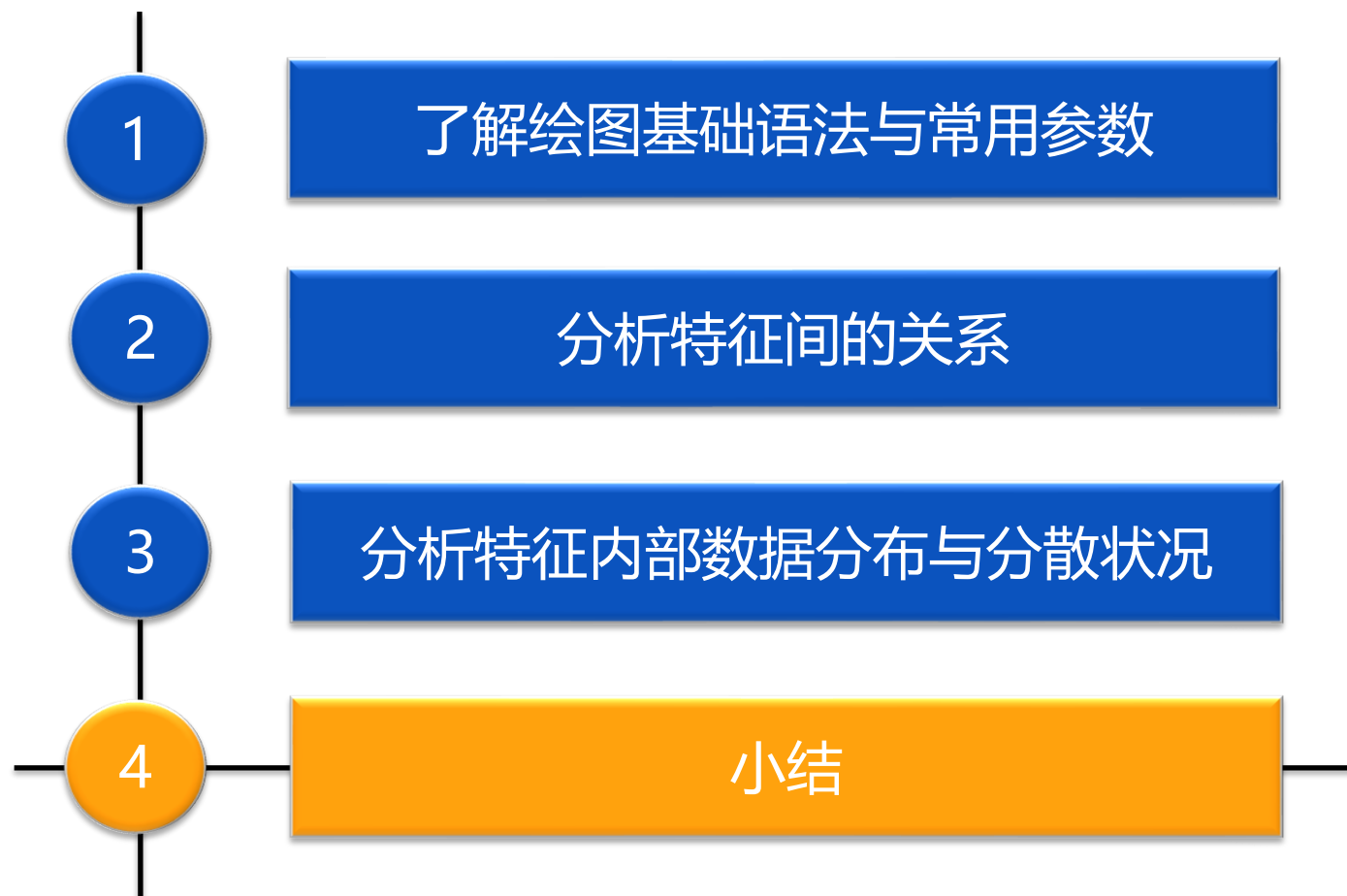
绘制箱线图

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, ...)*

| 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。 |

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。



Thank you!