

# 项目 2 人口普查数据图形展示

## 一、课程基本信息

1. **课程名称:** Python 数据可视化
2. **授课章节:** 项目 2 人民生活数据图形展示
3. **授课对象:** [具体专业与年级]
4. **授课时间:** [X] 学时
5. **授课地点:** [机房名称]

## 二、教学目标

### (一) 知识目标

1. 认识 matplotlib, 可以在 Python 环境中安装 matplotlib;
2. 掌握 matplotlib.pyplot 模块常用的绘图函数, 可以绘制一些简单的图表;
3. 掌握绘制子图的几种方式, 可以共享子图之间的坐标轴;
4. 掌握图表常用辅助元素的设置;
5. 掌握默认的图表样式和修改图表样式的方法。

### (二) 能力目标

1. 具有软件安装、调试、维护的实践能力;
2. 具有自主学习的能力、终身学习的理念;
3. 具有较强的口头与书面表达能力、人际沟通能力;
4. 具有团队精神和协作精神;
5. 具有良好的心理素质和克服困难的能力;
6. 具有良好的思想政治素质、行为规范和职业道德。

### （三）素质目标

1. 思想积极上进，热爱祖国、热爱人民，拥护中国共产党；
2. 具有正确的职业道德与行为规范；
3. 具有信息处理能力，能够搜索、甄别信息并应用；
4. 具有独立思考，能够分析并处理问题；
5. 具有创新能力。

## 三、教学重难点

### （一）教学重点

1. Matplotlib 库的安装与基础使用，包括快速绘图方法以及各类配置项的设置。
2. 不同类型图表（任务一和任务二中涉及的所有图表）的绘制方法与关键参数调整，以准确展示人民生活数据特征。
3. 多图组合的实现方式与应用场景，如何根据数据逻辑关系选择合适的多图展示形式。

### （二）教学难点

1. 针对不同人口数据特点，精准选择最恰当的图表类型进行可视化，避免因图表选择不当导致数据信息传达不准确。
2. 深入理解并灵活运用图表的全局配置项和系列配置项，实现复杂图表样式与交互效果的定制，满足实际项目需求。
3. 在多图组合时，确保各图表之间的数据关联性与逻辑一致性，以及整体布局的合理性与美观性。

## 四、教学方法

1. **讲授法**：讲解项目背景、学习目标、matplotlib 库的理论知识、图表类型的特点及适用场景等内容，使学生对课程有初步的系统认识。
2. **演示法**：在机房通过实际操作演示 matplotlib 库的安装过程、各类图表的绘制代码实现、配置项的调整效果等，让学生更直观地理解和掌握操作步骤。

3. **实践法**：安排学生在机房进行大量的实践练习，根据给定的人民生活数据完成各个任务中的图表绘制，通过实际操作加深对知识的理解与运用能力。
4. **小组讨论法**：组织学生进行小组讨论，针对任务中的数据处理、图表选择、多图组合等问题交流思路与解决方案，培养学生的团队协作与思维碰撞能力。
5. **案例分析法**：引入实际的人民生活数据可视化案例，分析其图表选择、设计思路、呈现效果等，引导学生学习优秀案例的经验，提升自身的可视化设计水平。

## 五、教学过程

### （一）课程导入

#### 1. 项目背景介绍

人口问题关系国家发展、民族未来。全面认识、正确看待我国人口发展新形势，既要看到世界人口发展的大趋势也要看到我国国情的特殊性，既要看到人口总量也要看人口质量，既要看到人口也要看人才，不能简单把人口数量的增减、年龄结构的变化等同于人口优势或劣势。人口数量、人口素质、人口结构是反映人口形势的重要指标。

#### 2. 学习目标阐述

将根据我国人口普查数据使用 Matplotlib 库生成图形，直观地展示近年来我国人口数量增长趋势、人口结构变化趋势、人口素质变化趋势等。数据来源：国家统计局发布年度数据，历年人口普查数据及人口抽样调查推算数据。

### （二）技术准备

#### 1. Matplotlib 简介

Matplotlib 是一个用于绘制图表和可视化数据的 Python 库，是数据科学、机器学习和科学计算领域中最流行的绘图库之一，支持 numpy、pandas 等数据结构。它主要用于绘制 2D 图表，具有丰富的绘制图表、定制图表元素或样式的功能，可以用于生成各种静态、交互式 and 动画图表。matplotlib 也可用于绘制一些 3D 图表。

#### 2. 环境搭建

使用 pip 方式安装

```
pip install matplotlib
```

使用 conda 方式安装

```
conda install matplotlib
```

#### 3. 创建画布

#### 4. 编辑图表组件

- (1) 添加图表标题
- (2) 设置坐标轴
- (3) 添加网格线
- (4) 添加参考线
- (5) 图形展示与保存

### (三) 任务实践

#### 任务 2 - 1: 人口增长趋势图形展示

任务描述：本任务要求选用 `pyplot` 模块中合适的图形函数，展示该人口年度数据各特征值的变化趋势。

##### 2.1.1 使用 `plot()` 函数绘制折线图

折线图 (Line Chart) 是将数据标注成点，并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。折线图一般以时间序列为轴，适用于展示固定时间间隔的数据变化趋势。它以折线的倾斜程度来形象地反映数据沿某一维度的变化趋势，能够清晰地展示数据增减的趋势、速率、规律及峰值等特征。

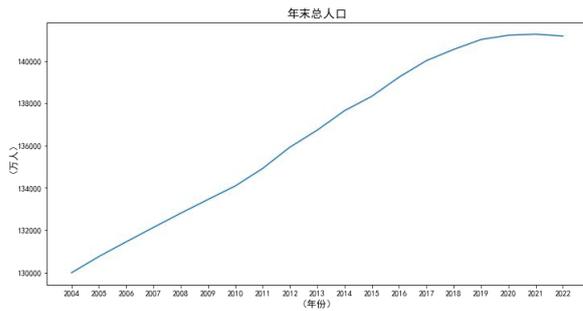
`matplotlib.pyplot.plot ( x, y, fmt, scalex=True, scaley=True, data=None, label=None, *args, **kwargs)`

`plot()` 函数常见参数及说明

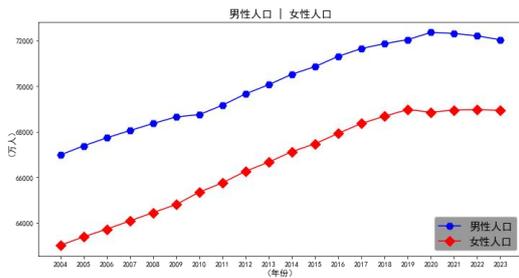
参数	说明
<code>x</code>	接收 <code>array</code> 。表示 X 轴对应的数据。默认值为 <code>range(len(y))</code>
<code>y</code>	接收 <code>array</code> 。表示 Y 轴对应的数据。无默认值
<code>fmt</code>	表示快速设置线条样式的格式字符串
<code>label</code>	表示应用于图例的标签文本
<code>scalex, scaley</code>	接收 <code>bool</code> 。表示这些参数确定视图限制是否适合于数据限制。默认为 <code>True</code>
<code>data</code>	接收可索引对象。表示具有标签数据的对象。默认为 <code>None</code>
<code>color</code>	接收特定 <code>str</code> 。表示指定线条的颜色。默认为 <code>None</code> 颜色取值可以使用单词或其缩写，如 <code>blue</code> 或 <code>b</code>
<code>linewidth</code>	接收 <code>int</code> 或 <code>float</code> ，指定线条宽度
<code>linestyle</code>	接收特定 <code>str</code> 。表示指定线条类型。默认为 <code>"-"</code> 。
<code>marker</code>	接收特定 <code>str</code> 。表示绘制的点的类型。默认为 <code>None</code>
<code>markerize</code>	接收 <code>int</code> ，指定标记点大小
<code>alpha</code>	接收 <code>float</code> 。表示点的透明度。默认为 <code>None</code>

任务案例：绘制折线图展示人口增长趋势

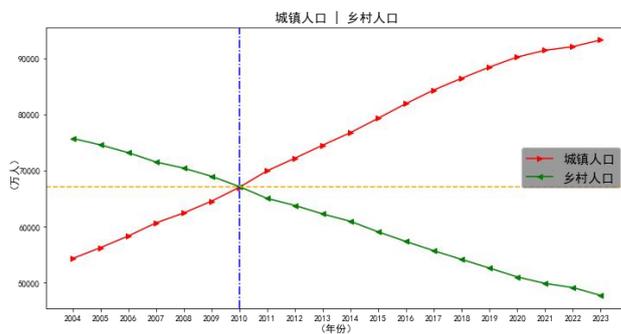
1、读取“人口数据.xlsx”中“总人口”表单数据，并根据历年“年末总人口(万人)”列数据绘制折线图。



2、读取“人口数据.xlsx”中“总人口”表单数据，绘制折线图比较“男性人口”和“女性人口”数据。



3、城镇化是现代的必由之路。有序推进城镇化是我国经济发展和扩大内需最主要的潜力所在。国家统计局发布的人口抽样调查数据中，提供了历年来“城镇人口”和“乡村人口”的数据。读取“人口数据.xlsx”中“总人口”表单数据，绘制折线图比较近年来“城镇人口”和“乡村人口”数据变化。



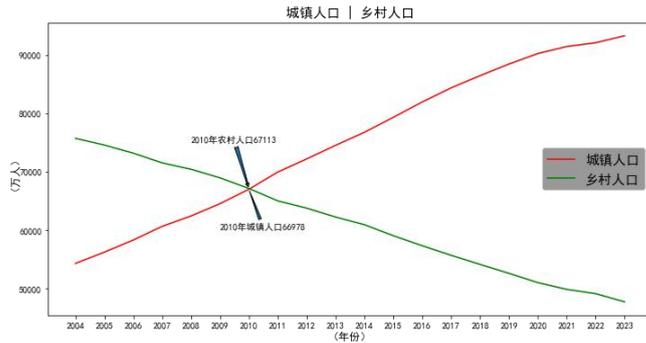
多学一招：

使用 `pyplot` 模块的 `annotate()` 函数可以为图表添加指向型注释文本，语法如下：

`matplotlib.pyplot.annotate(s,xy, *args,**kwargs)`

`annotate()` 函数常见参数及说明

参数	说明
<code>s</code>	接收 <code>str</code> ，表示注释文本的内容
<code>xy</code>	接收元组 <code>(x,y)</code> ，表示被注释的点的坐标位置
<code>xytext</code>	接收元组 <code>(x,y)</code> ，表示注释文本所在的坐标位置
<code>arrowprops</code>	接收 <code>dict</code> ，表示指示箭头的属性字典。包括 <code>width</code> 、 <code>headwidth</code> 、 <code>headlength</code> 、 <code>shrink</code> 、 <code>arrowstyle</code> 等。 <code>arrowstyle</code> 表示箭头类型，可取值： <code>-&gt;</code> 、 <code>&lt;-&gt;</code> 、 <code>&lt;-</code> 、 <code> -&gt;</code> 、 <code>- &gt;</code> 、 <code>&lt; -</code> 、 <code>&lt; &gt;</code> 、 <code>fancy</code> 、 <code>simple</code> 、 <code>wedge</code>
<code>bbox</code>	接收 <code>dict</code> ，表示注释文本的边框属性字典



### 2.1.2 使用 scatter () 函数绘制散点图

散点图 (scatter diagram) 又称为散点分布图, 由若干数据点组成主要用于判断两变量之间是否存在某种关联, 或者总结数据点的分布模式。每个数据点有位置特征: 横坐标和纵坐标。利用坐标点 (散点) 的分布形态反映特征间的统计关系。

散点图中数据点的分布情况可以体现变量之间的相关性。若所有的数据点在一条直线附近呈波动趋势, 说明变量之间是线性相关的; 若数据点在曲线附近呈波动趋势, 说明变量之间是非线性相关的; 若数据点没有显示任何关系, 说明变量之间是不相关的。数据值由数据点在图表中的位置表示, 类别由图表中的不同标记表示, 通常用于在跨类别的数据之间作比较。

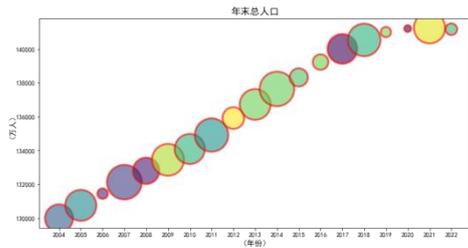
`matplotlib.pyplot.scatter(x, y, s=None, c=None, marker=None, cmap=None, norm=None, vmin=None, vmax=None, alpha=None, linewidths=None, verts=None, edgecolors=None, *, data=None, **kwargs)`

scatter () 函数常见参数及说明

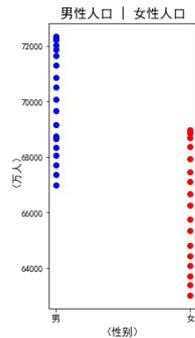
参数名	说明
x, y	表示数据点的位置 (x,y)。必备参数。x,y 都不能为空。x, y 支持多维数组, 但是 x, y 的维度必须一致, 多维数组会被展平 (flatten) 为一维数组
s	表示数据点标记的大小。可选参数。默认值为 <code>rcParams['lines.markersize'] ** 2</code>
c	表示数据点标记的颜色。 可选参数。默认是蓝色'b',表示的是标记的颜色, 或者可以是一个表示颜色的字符, 或者是一个长度为 n 的表示颜色的序列等等如'b'=blue, 'y'=yellow, 'k'=black 等
marker	表示数据点标记的样式, 默认为圆形'o': <code>rcParams["scatter.marker"] ('o')</code>
alpha	表示数据点标记的透明度。取值范围为[0,1], 0 为透明, 1 为不透明。默认值为 None
linewidths	表示数据点标记的描边 (边缘线) 宽度。浮点数或类数组结构。默认值为 <code>rcParams["lines.linewidth"] (1.5)</code>
edgecolors	表示数据点标记的描边 (边缘线) 颜色。取值为{'face', 'none', None}、颜色、颜色序列。默认值为 <code>rcParams["scatter.edgecolors"] ('face')</code> 。 'face': 标记边缘颜色与标记颜色相同 'none': 没有边线
vmin,vmax	表示亮度的最小值和最大值。当 norm 存在的时候忽略

任务案例: 绘制散点图展示人口增长趋势

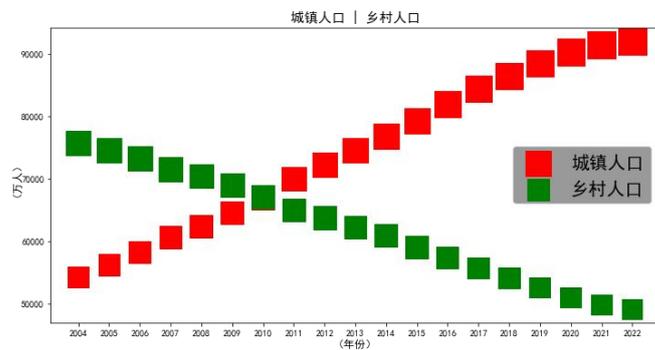
1、读取“人口数据.xlsx”中“总人口”表单数据, 并根据历年“年末总人口(万人)”列数据绘制散点图。



2、读取“人口数据.xlsx”中“总人口”表单数据，绘制散点图比较“男性人口”和“女性人口”数据。



3、读取“人口数据.xlsx”中“总人口”表单数据，绘制气泡图比较近年来“城镇人口”和“乡村人口”数据变化。设置参数改变数据标注气泡的面积大小，使之随着数据的变化而变化。



### 2.1.3 使用 bar () 函数绘制柱形图

柱形图 (Bar Chart) 是由一系列宽度相等的纵向矩形条 (柱体) 组成的图表，它使用柱体的高度表示数值，以此反映不同分类数据之间的差异，是数据分析中最常用、最灵活的图表之一。柱形图的核心思想是对比，常用于显示一段时间内的数据变化或显示各项之间的比较情况。

柱形图一般以数据分类为轴，每类数据一个柱体。如果数据过多，柱形图的美观性和易读性都会受到较大的影响，因此柱形图适用于展示中小规模数据集各分类之间的比较。

`matplotlib.pyplot.bar(x,height,width=0.8,bottom=None,align='center',data=None,tick_label=None,xerr=None,yerr=None,**kwargs)`

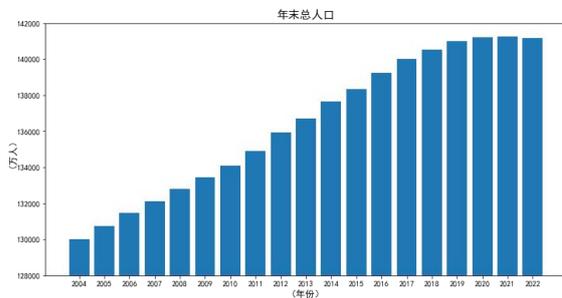
bar()函数常见参数及说明

参数名	说明
x	接收 array 或 float。表示柱形的 X 轴坐标值。无默认值
height	接收 array 或 float。表示柱形的高度。无默认值

width	接收 array 或 float。表示柱形的宽度。默认为 0.8
align	接收 str。表示整个柱形图与 x 轴的对齐方式，可选“center”和“edge”。默认为“center”，表示将柱形与刻度线居中对齐；“edge”。表示将柱形的左边与刻度线对齐
color	接收特定 str 或包含颜色字符串的 list。表示柱形图颜色。默认为 None
bottom	表示柱形底部的 y 坐标值，默认为 0
tick_label	表示柱形对应的刻度标签
xerr, yerr	若未设为 None，则需要为柱形图添加水平/垂直误差棒
error_kw	表示误差棒的属性字典

任务案例：绘制柱形图展示人口增长趋势

1、读取“人口数据.xlsx”中“总人口”表单数据，并根据历年“年末总人口(万人)”列数据绘制柱形图。



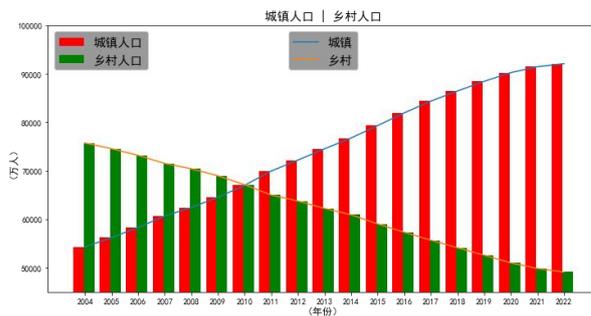
2、读取“人口数据.xlsx”中“总人口”表单数据，绘制柱形图比较“男性人口”和“女性人口”数据

3、读取“人口数据.xlsx”中“总人口”表单数据，绘制柱形图比较近年来“城镇人口”和“乡村人口”数据变化

4、绘制柱形堆叠图

5、绘制有误差棒的柱形图

6、读取“人口数据.xlsx”中“总人口”表单数据，同时使用折线图和柱形图，比较近年来“城镇人口”和“乡村人口”数据变化，并分别添加图例



#### 2.1.4 使用 barh() 函数绘制条形图

条形图是横置的柱形图，由一系列高度相等、长短不一的横向矩形条组成。更适合分类数量较多的数据。

matplotlib.pyplot.barh(y, width, height=0.8, left=None, align='center', \*, \*\*kwargs)

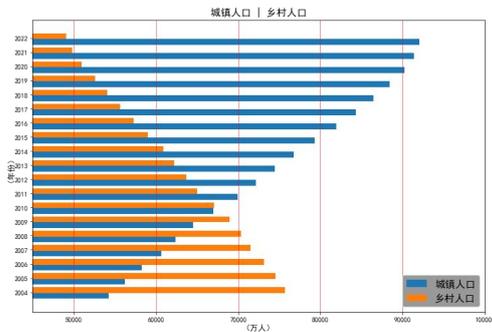
barh()函数常见参数及说明

参数名	说明
y	表示条形的 y 值

width	表示条形的宽度
height	表示条形的高度，默认值为 0.8
color	接收特定 str 或包含颜色字符串的 list。表示柱形图颜色。默认为 None
left	条形左侧的 x 坐标值，默认值为 0
align	表示条形的对齐方式，， 可选“center”和“edge”。默认值为“center”，即条形与刻度线居中对齐；“edge”表示将条形的底边与刻度线对齐
tick_label	表示条形对应的刻度标签
xerr, yerr	若未设为 None，则需要为条形图添加水平/垂直误差棒

任务案例：绘制条形图展示人口增长趋势

- 1: 读取“人口数据.xlsx”中“总人口”表单数据，并根据历年“年末总人口(万人)”列数据生成条形图。
- 2、读取“人口数据.xlsx”中“总人口”表单数据，生成柱形图比较“男性人口”和“女性人口”数据。
- 3、读取“人口数据.xlsx”中“总人口”表单数据，绘制条形图比较近年来“城镇人口”和“乡村人口”数据变化



### 2.1.5 使用 stem()函数绘制茎叶图

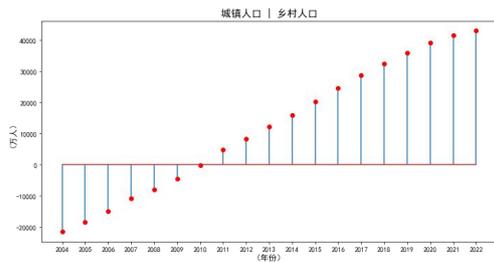
茎叶图从外观来看更像是火柴，由基线、茎线、茎头三部分构成，根据基线（baseline）上的位置（locs）绘制从基线到杆头（heads）的茎线，并在茎头（heads）处放置标记，又称为棉棒图、火柴杆图、大头针图等。茎叶图是柱形图或条形图的变形。

matplotlib.pyplot.stem([loc, head, linefmt=None, markerfmt=None, basefmt=None, bottom=0, label=None, use\_line\_collection=True, orientation='vertical', data=None)

参数	说明
loc	接收 array，表示茎线的位置，对于水平茎叶图，该参数即茎线的 Y 轴位置。默认值为 range(len(heads))。可选参数
head	接收 array，表示茎头的位置，对于垂直茎叶图，该参数即茎头的 y 值，对于水平茎叶图，该参数即茎头的 x 值。必备参数
linefmt	接收 str，表示定义茎线的颜色和线形，取值同 plot 函数。默认值为'C0-'。可选参数
markerfmt	接收 str，表示定义茎头标记的形状和颜色。默认值为'C0o'。可选参数
basefmt	接收 str，表示定义基线的格式字符串。默认值为'C3-'。可选参数
bottom	接收 float，表示基线的位置。默认值为 0。可选参数
label	接收 str，表示图例中使用的标签。默认值 None。可选参数
use_line_collection	接收 bool，表示控制线条对象类型。默认值 True。可选参数

任务案例：绘制茎叶图展示人口增长趋势

- 1: 读取“人口数据.xlsx”中“总人口”表单数据，并根据历年“年末总人口(万人)”列数据绘制茎叶图。
- 2、读取“人口数据.xlsx”中“总人口”表单数据，绘制茎叶图比较“男性人口”和“女性人口”数据，并设置参数 `linefmt='r--'` 指定茎叶图的线段线型，设置参数 `markerfmt='ro',markerfmt='s'`指定茎叶图的标记符号，并设置 X 轴坐标: 男性人口茎线向左偏移 0.15、女性人口茎线向右偏移 0.15。
- 3、读取“人口数据.xlsx”中“总人口”表单数据，绘制茎叶图比较近年来“城镇人口”和“乡村人口”数据变化，通过两个数据的差值观察农业转移人口市民化的进程。



### 2.1.6 使用 `errorbar()`函数绘制误差棒图

误差棒图是使用误差棒注明被测量数据的不确定度大小的图表,用于表示测量数据中客观存在的测量偏差。误差棒是在表示测量值大小的方向上的一条线段,它以被测量数据的平均值为中点,线段长度的一半为不确定度。

`matplotlib.pyplot.errorbar(x, y, yerr=None, xerr=None, fmt="", ecolor=None,*, data=None, **kwargs)`

`errorbar()`函数常见参数及说明

参数名	说明
<code>x, y</code>	表示数据点的位置
<code>xerr, yerr</code>	表示数据的误差范围
<code>fmt</code>	表示数据点的标记样式和数据点之间连接线的样式。默认为‘-’实线。若设置为标记点样式,则不再显示折线
<code>marker</code>	表示数据点的标记样式
<code>markersize</code>	表示数据点的标记样式大小
<code>elinewidth</code>	表示误差棒的线条宽度
<code>ecolor</code>	表示误差棒的颜色
<code>capsize</code>	表示误差棒边界横杆的大小。默认无
<code>capthick</code>	表示误差棒边界横杆的厚度
<code>uplims,lolims</code>	表示 Y 轴误差棒上下限
<code>xuplims,xlolims</code>	表示 X 轴误差棒上下限

任务案例: 绘制误差棒图展示人口增长趋势

- 1: 读取“人口数据.xlsx”中“总人口”表单数据，并根据历年“年末总人口(万人)”列数据绘制误差棒图。假设数据误差在 1%以内，设置参数 `yerr=df['年末总人口(万人)']*0.01`。
- 2、读取“人口数据.xlsx”中“总人口”表单数据，绘制条形图比较近年来“城镇人口”和“乡村人口”数据变化，假设数据误差在 2%以内。

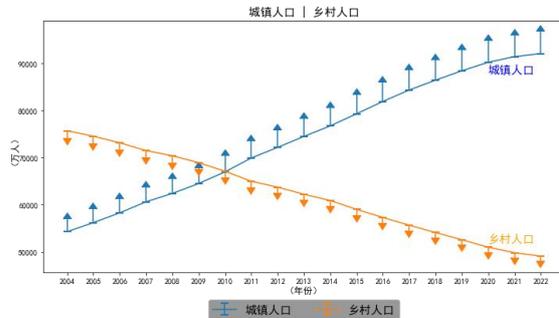
多学一招:

使用 `pyplot` 模块的 `text()`函数可以为图表添加无指向型注释文本,使用文字的注释方式对绘图区域的图形进行说明,语法如下:

matplotlib.pyplot.text(x,y,s,fontdict=None,\*\*kwargs)

text()函数常见参数及说明

参数	说明
x,y	接收 float，表示注释文本的 X 轴和 Y 轴坐标位置。
s	接收 str，表示注释文本的内容。
fontdict	接收 dict，表示字体格式



### 2.1.7 使用 polar()函数绘制雷达图

雷达图又称蜘蛛网图、星状图、极区图，由一组坐标轴和多个等距同心圆或多边形组成，是一种表现多维（4 维以上）数据的图表。

雷达图的坐标轴起始于圆心，每个坐标轴代表一个指标，连接多个维度的数据点，围成一个多边形，适用于对多指标对象做出全局性、整体性评价。

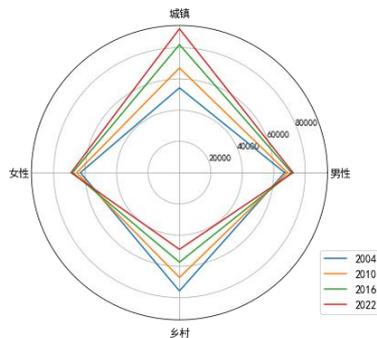
matplotlib.pyplot.polar(theta, r, \*\*kwargs)

参数名	说明
theta	表示每个数据点所在射线与极径的夹角。
r	表示每个数据点到原点的距离。

任务案例：绘制雷达图展示 2004 年、2010 年、2016 年、2022 年城镇人口、农村人口、男性人口、女性人口。

多学一招：

雷达图的数据点可以使用 marker 参数设置标注类型，并同时设置标注的其他属性参数；也可以使用 linestyle 设置线型并设置线段其他属性。



### 任务 2-2 人口年龄结构变化图形展示

**任务描述:** 人口结构是人口形势的重要指标之一, 本任务要求根据国家统计局发布的相关人口取样数据, 选择合适的图形函数绘制图形, 直观地展示近年来我国人口政策的调整对人口年龄结构的影响。

### 2.2.1 使用 stackplot () 函数绘制堆叠区域图

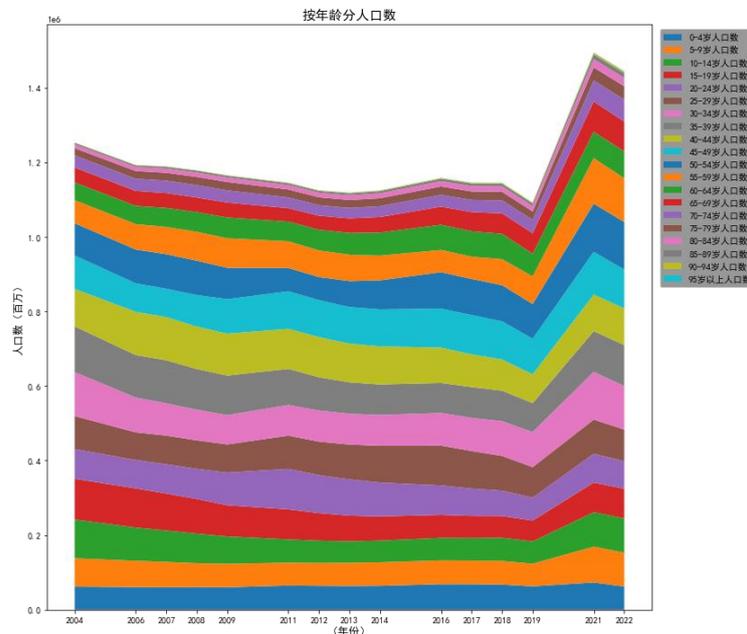
堆叠图可分为堆叠区域图、堆叠柱形图和堆叠条形图。其中堆叠柱形图和堆叠条形图在任务 1 中已学习。堆叠区域图是由若干条折线与折线或水平坐标轴之间的填充区域组成的图表, 用于强调每个部分变化的趋势; 堆叠柱形图和堆叠条形图是由若干个以颜色或线条填充、高度不一的纵向矩形条或横向矩形条堆叠而成的图表, 主要用于反映各构成部分在总体中的比重。

`matplotlib.pyplot.stackplot(x,y,labels=(),baseline='zero',data=None,*args,**kwargs)`

参数	说明
x	形状为(N,)的类数组结构, 即尺寸为 N 的一维数组。必备参数。
y	形状为(M,N)的类数组结构, 即尺寸为(M,N)的二维数组。必备参数。
labels	为每个数据系列指定标签。长度为 N 的字符串列表。表示每个填充区域的标签。
baseline	字符串, 取值范围为{'zero', 'sym', 'wiggle', 'weighted_wiggle'}。默认值为 'zero'。可选参数。表示计算基线的方法。
colors	颜色列表或元组。每组区域所使用的颜色, 循环使用

**任务案例:** 绘制堆叠区域图, 观察近年来我国人口年龄结构的变化趋势。

- 1、读取“人口数据.xlsx”中“人口年龄结构抚养比”表单数据, 并根据历年各年龄段列数据绘制堆叠区域图。
- 2、读取“人口数据.xlsx”中“按年龄分人口数”表单数据, 并根据历年各年龄段列数据绘制堆叠区域图。



### 2.2.2 使用 pie () 函数绘制饼图和圆环图

饼图是由若干个面积大小不一、颜色不同的扇形组成的圆形图表, 它使用圆表示数据的总量, 组成圆的每个扇形表示数据中各项占总量的比例大小, 主要用于显示数据中各项大小与各项总和的比例。

饼图中的圆与扇形分别代表整体与部分，适用于展示数据整体与各类数据的占比关系。

`matplotlib.pyplot.pie(x, explode=None, labels=None, colors=None, autopct=None, pctdistance=0.6, shadow=False, labeldistance=1.1, startangle=0, radius=1, counterclock=True, wedgeprops=None, textprops=None, center=0, 0, frame=False, rotatelabels=False, *, normalize=None)`

参数名	说明
<code>x</code>	接收类似一维数组结构数据，指定用于绘制饼图各个饼块的尺寸
<code>explode</code>	接收类似一维数组结构数据，指定每个饼块相对于饼图圆心的偏移距离，取值为浮点数，默认值为 <code>None</code>
<code>labels</code>	字符串列表。指定每个饼块的标签。默认值为 <code>None</code>
<code>colors</code>	接收类似一维数组结构数据，指定每个饼块的颜色。颜色会循环使用。默认值为 <code>None</code> ，使用当前色彩循环
<code>autopct</code>	接收特定 <code>str</code> 或可调用对象，指定饼块内显示的标签字符串。默认值为 <code>None</code> 。如果值为格式字符串，标签将被格式化，如果值为函数，将被直接调用。可通过格式字符串指定小数点后的位数
<code>pctdistance</code>	接收 <code>float</code> 。指定饼块内标签与圆心的距离。默认值为 <code>0.6</code> ， <code>autopct</code> 不为 <code>None</code> 则该参数生效
<code>shadow</code>	接收 <code>bool</code> ，指定饼图下是否显示阴影。默认值为 <code>False</code>
<code>labeldistance</code>	接收 <code>float</code> 。指定饼块外标签与圆心的距离（相对于半径的比例）。默认值为 <code>1.1</code> 。如果设置为 <code>None</code> ，标签不会显示，但是图例可以使用标签
<code>rotatelabels</code>	接收 <code>bool</code> ，指定饼块外标签是否按饼块角度旋转。默认为 <code>False</code>
<code>startangle</code>	接收 <code>float</code> 。指定饼块起始绘制角度。默认值为 <code>0</code> ，即从 <code>x</code> 轴的正方向逆时针旋转绘制
<code>counterclock</code>	接收 <code>bool</code> ，指定角度是否逆时针旋转。默认值为 <code>True</code>
<code>radius</code>	接收 <code>float</code> 。指定圆形半径。默认值为 <code>1</code>
<code>wedgeprops</code>	接收 <code>dict</code> 。指定饼块属性。例如，通过 <code>wedgeprops={'width':0.5}</code> 将宽度设为 <code>0.5</code> 。默认值为 <code>None</code>
<code>textprops</code>	接收 <code>dict</code> 。指定图表中的文本属性。默认值为 <code>None</code>
<code>center</code>	接收浮点数元组( <code>float, float</code> )。指定饼图中心点位置坐标。默认值为( <code>0,0</code> )
<code>frame</code>	接收 <code>bool</code> ，指定是否绘制子图边框。默认为 <code>False</code>
<code>normalize</code>	接收 <code>bool</code> ，指定是否归一化。默认值为 <code>None</code> 。 <code>True</code> : 完全饼图，对 <code>x</code> 进行归一化， <code>sum(x)==1</code> 。 <code>False</code> : 如果 <code>sum(x)&lt;=1</code> ，绘制不完全饼图。如果 <code>sum(x)&gt;1</code> ，抛出 <code>ValueError</code> 异常

任务案例：绘制饼图和圆环图，观察近年来我国人口年龄结构的变化趋势。

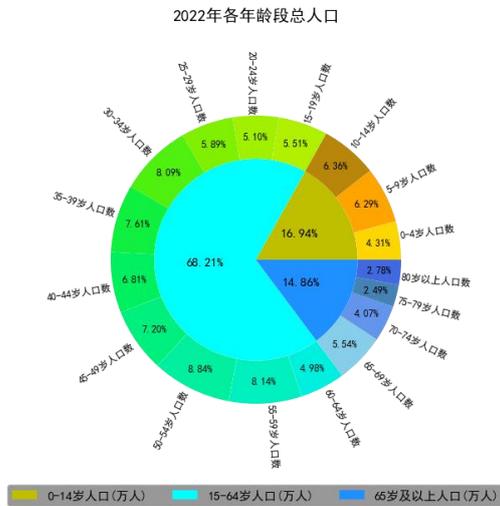
读取“人口数据.xlsx”中“人口年龄结构抚养比”表单数据，并根据 2022 年各年龄段列数据绘制饼图。



2、读取“人口数据.xlsx”中“按年龄分人口数”表单数据，并根据 2022 各年龄段人口数绘

制饼图。

3、可以将“2022年人口结构抚养比”与“2022年各年龄段人口数”两个图形绘制在一起，类似于同心圆。



### 2.2.3 使用 boxplot () 函数绘制箱线图

箱线图 (boxplot) 又称箱形图, 用来展示数据分布的分散程度, 还可以展示多组数据分布特征的比较, 也可以粗略地看出数据是否具有对称性等, 因形状如箱子而得名。

箱线图由六种数值组成: 异常值 (outlier/flier)、最小值 (minimum)、下四分位数 (Q1, 即第 25%分位数)、中位数 (median, 即第 50%分位数)、上四分位数 (Q3, 即第 75%分位数)、最大值 (maximum)。如图 2-71 所示。也可以粗略地看出数据是否具有对称性, 等信息。

四分位数: 就是把一组数据按照从小到大的顺序进行排列, 然后分成四等份, 处于三个分割点位置的数字就是四分位数:

Median (中位数): 即 Q2 (第二四分位数) 等于该样本中所有数值由小到大排列后第 50% 的数字,  $Q2$  的位置 =  $1+(n-1)*0.5$ 。将所有数值从小到大排列, 如果是奇数个数值则取最中间一个值作为中位数, 之后最中间的值在计算 Q1 和 Q3 时不再使用; 偶数个数值则取最中间两个数的平均数作为中位数, 这两个数在计算 Q1 和 Q3 时继续使用。

Q1 (第一四分位数): 又称“较小四分位数”或“下四分位数”, 中位数将所有数据分成两部分, 最小值到中位数的部分按取中位数的方法取中位数作为 Q1, 等于该样本中所有数值由小到大排列后第 25% 的数字,  $Q1$  的位置 =  $1+(n-1)*0.25$ 。

Q3 (第三四分位数), 又称“较大四分位数”或“上四分位数”, 中位数将所有数据分成两部分, 中位数到最大值的部分按取中位数的方法取中位数作为 Q3, 等于该样本中所有数值由小到大排列后第 75% 的数字。Q3 的位置 =  $1+(n-1)*0.75$ 。

四分位间距 (InterQuartile Range, IQR): Q3 与 Q1 的差距(Q3-Q1);

maximum 上限 (大于该值即为异常值):  $Q3$  数 +  $1.5*IQR$ , (1.5 表示超过的比例, 是一个系数, 可根据实际情况调整);

minimum 下限 (小于该值即为异常值):  $Q1$  数 -  $1.5*IQR$ 。

异常值又称离群值, 指大于 1.5 倍的四分位数间距的值。处于 1.5 倍~3 倍四分位数间距的值默认使用空心圆圈表示。极端值是指大于 3 倍的四分位数间距的值, 属于异常值中的一种。

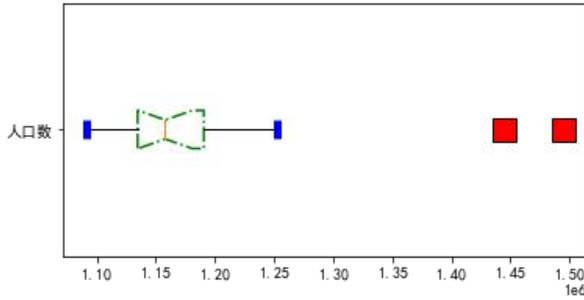
`matplotlib.pyplot.boxplot( x, notch=None, sym=None, vert=None, whis=None, positions=None, widths=None, patch_artist=None, bootstrap=None, usermedians=None, conf_intervals=None,`

meanline=None, showmeans=None, showcaps=None, showbox=None, showfliers=None, boxprops=None, labels=None, flierprops=None, medianprops=None, meanprops=None, capprops=None, whiskerprops=None, manage\_ticks=True, autorange=False, zorder=None, \*, data=None )

参数名	说明
x	接收 array 或向量序列。表示输入数据。必备参数
notch	接收 bool, 表示控制箱体中央是否有 V 形凹槽。
sym	接收 str, 离群点的默认标记符号。默认值为'None'。可选参数。
vert	接收 bool, 箱体的方向, 可选参数
whis	接收 float 或 float 二元组,表示箱须的位置。默认值为 1.5。 可选参数。
bootstrap	接收 int, 表示是否使用 bootstrap 方法计算中位数置信区间。可选参数。
usermedians	表示是否指定中位数。类型为一维类数组结构。可选参数。
conf_intervals	接收 array。表示是否指定置信区间。形状为 (len(x), 2)。可选参数。
positions	接收 array。表示箱体的位置。刻度和极值会自动匹配箱体位置。可选参数。
widths	接收 float 或 array, 表示箱体的宽度。默认值为 0.5 或 0.15*极值间的距离
patch_artist	接收 bool, 表示控制箱体的生成对象。默认值为 False。可选参数。
labels	接收 array。表示每个数据集的标签, 默认值为'None'。可选参数
manage_ticks	接收 bool, 表示控制刻度和标签位置, 默认值为 True。可选参数。
autorange	接收 bool, 默认值为 False。可选参数。
meanline	接收 bool, 表示均值显示为线还是点。默认值为 False。可选参数。
zorder	接收 float, 表示箱线图的叠放次序。默认值为 Line2D.zorder = 2。可选参数
showcaps	接收 bool, 表示是否显示箱须两端的横杠。默认值为 True。可选参数
showbox	接收 bool, 表示是否显示箱体。默认值为 True。可选参数
showfliers	接收 bool, 表示是否显示离群值。默认值为 True。可选参数
showmeans	接收 bool, 表示是否显示算术平均值。默认值为 False。可选参数
capprops	接收 dict, 表示箱须横杠的样式。默认值为 None。可选参数
boxprops	接收 dict, 表示箱体的样式。默认值为 None。可选参数
whiskerprops	接收 dict, 表示箱须的样式。默认值为 None。可选参数
flierprops	接收 dict, 表示离群点的样式。默认值为 None。可选参数
medianprops	接收 dict, 表示中位数的样式。默认值为 None。可选参数
meanprops	接收 dict, 表示算术平均值的样式。默认值为 None。可选参数

任务案例：绘制箱线图，观察近年来我国人口年龄结构的变化趋势。

- 1、读取“人口数据.xlsx”中“人口年龄结构抚养比”表单中历年来各年龄段列数据，并绘制箱线图。
- 2、读取“人口数据.xlsx”中“按年龄分人口数”表单中“人口数”列数据，并绘制箱线图。
- 3、读取“人口数据.xlsx”中“按年龄分人口数”表单中历年各年龄段列数据，并绘制箱线图。



### 2.2.4 使用 hist () 绘制直方图

直方图又称质量分布图，是由一系列高低不等的纵向矩形条或线段组成的图表，通常将连续的数据分割成若干区间，然后计算每个区间内数据点的数量，反映数据的分布和波动情况。数据点数量通常用柱形图表示，柱子的高度表示该区间内数据点的数量。

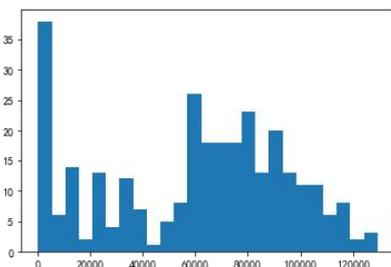
`pyplot.hist(x,bins=None,range=None,density=None,weights=None,bottom=None)`

参数	说明
x	表示 X 轴的数据
bins	表示矩形条的个数，默认为 10
range	表示 X 轴数据的范围，若未设置范围，默认数据范围为(x.min(),x.max())
bottom	表示 Y 轴数据刻度的起始位置
cumulative	表示是否计算累计频数或频率
histtype	表示直方图的类型
align	表示矩形条边界的对齐方式，可设置为'left'、'mid'或'right'，默认为'mid'。
orientation	表示矩形条的摆放方式，默认为'vertical'，即垂直方向。水平方向：horizontal
rwidth	表示矩形条宽度的百分比，默认为 0。
stacked	表示是否将多个矩形条以堆叠形式摆放
facecolor	直方图颜色
edgecolor	直方图边框颜色
alpha	透明度

任务案例：绘制直方图，观察近年来我国人口数据分布情况。

读取“人口数据.xlsx”中“人口年龄结构抚养比”表单中的'年末总人口(万人)'列数据，并绘制直方图。

2、读取“人口数据.xlsx”中“按年龄分人口数”表单中的各年龄段列数据，并去除取样数据较多的 2005 年和 2015 年数据，绘制直方图。



### 任务 2-3 受教育程度数据图形展示

任务描述：在国家统计局公布的人口普查年度数据中，包含了“接受教育程度分人口数”和“15 岁及以上文盲人口数”，读取相关数据并绘图形，观察我国人口素质提升的变化趋势。

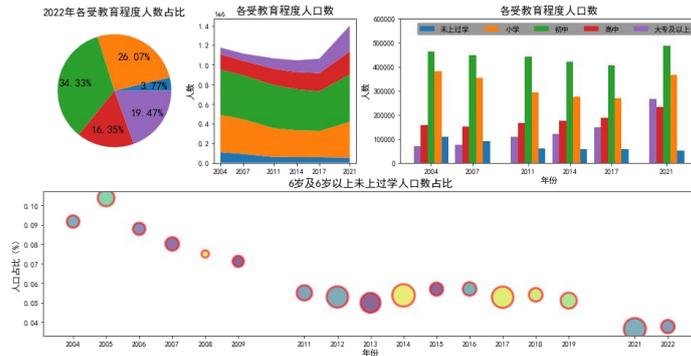
### 2.3.1 使用 subplot()函数绘制单子图

使用 pyplot 模块的 subplot()函数可以在画布上指定位置创建一个子图。

subplot(nrows, ncols, index, \*\*kwargs)

参数	说明
nrows, ncols	接收 int, 表示子区网格的行数、列数
index	接收 int, 表示矩阵区域的索引

任务案例: 读取“人口数据.xlsx”中“按受教育程度分人口数”表单数据, 提取各受教育程度人数的列数据, 并在画布上绘制各种合适的图形进行展示。

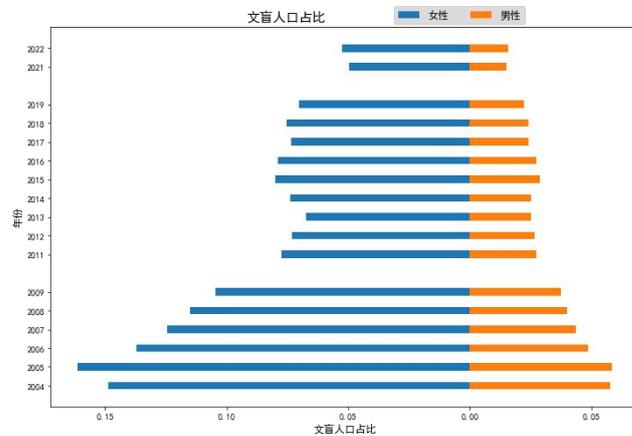


### 2.3.2 使用 add\_subplot()函数添加子图

Figure 类的 add\_subplot()方法可以添加和选中子图, 参数与用法与 subplot()函数相同, 返回一个 axes 类的子对象, 可以详细设置坐标轴各参数。语法如下:

add\_subplot(\* args, \*\* kwargs )

任务案例: 读取“人口数据.xlsx”中“15岁及以上文盲人口数”表单数据, 提取历年统取样数据中 15 岁以上男性与女性文盲数, 并绘制金字塔图展示历年普及教育扫盲的成果。



### 2.3.3 使用 subplots()函数绘制多子图

subplots()函数可以在画布上一次性平均划分出多个子图区域, 语法如下:

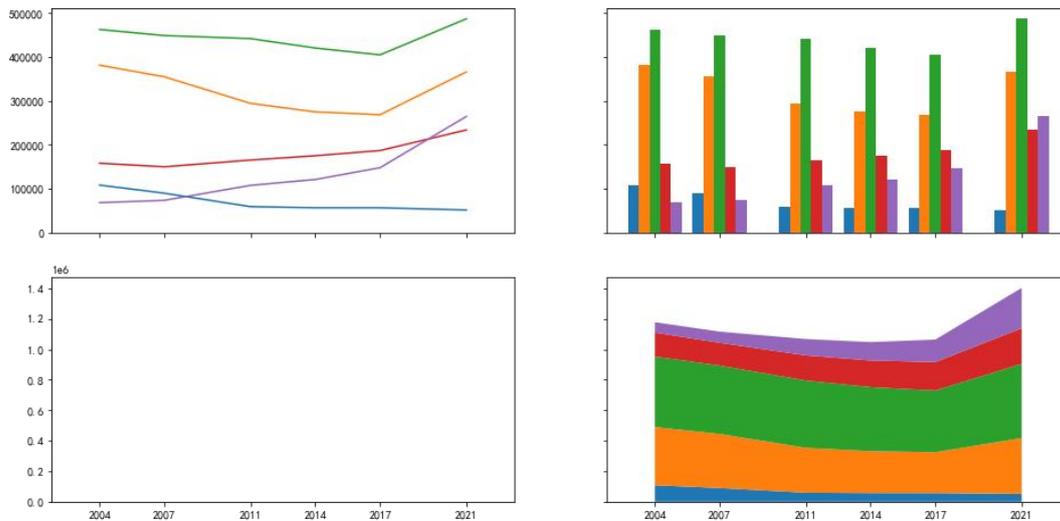
subplots(nrows = 1,ncols = 1,sharex = False,sharey = False,squeeze = True,subplot\_kw = None,gridspec\_kw = None,\*\* fig\_kw)

subplots()函数常见参数及说明见表 2-22。

表 2-22 subplots()函数常见参数及说明

参数	说明
nrows, ncols	接收 int, 表示子网格的行数、列数
sharex, sharey	接收 bool, 表示是否共享 X 轴或 Y 轴, 默认 False

任务案例：读取“人口数据.xlsx”中“接受教育程度分人口数”表单数据，提取各受教育程度人数的列数据，并在画布上绘制各种合适的图形进行展示。

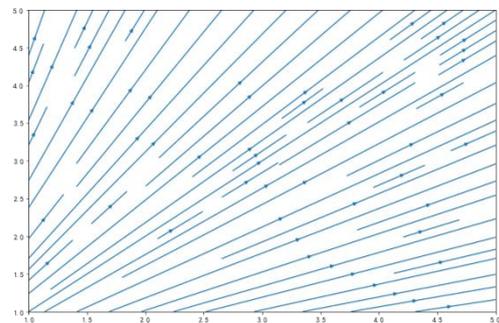


## 知识拓展

使用 `matplotlib` 还可以绘制很多其他高级图表。

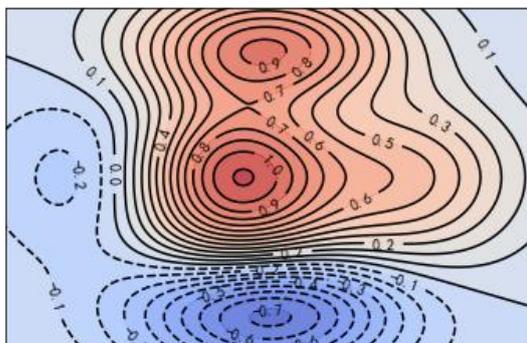
### 1、矢量场流线图

`pyplot` 模块的 `streamplot()` 函数可以绘制矢量场流线图。



### 2、等高线图

`pyplot` 模块的 `contour()`、`contourf()` 函数可以绘制和填充等高线图。



### 3、3D 图表

`matplotlib` 不仅可用于二维图表的绘制，也可以绘制 3D 图表、统计地图等。

