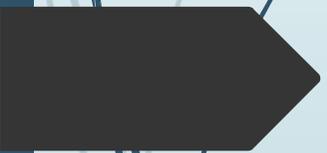




## 第7章 有关关系的可视化

---





# 本章主要内容

---

1. 在关系中寻求什么
2. 关联性 (圆点、探索更多的变量、气泡)
3. 分布 (老式分布图表、有关分布的柱形、延续性的密度)
4. 对照和比较



## 7.1 在关系中寻求什么

---

- ▶ 当某个数量增加时，另一个数量是否会减少？  
它们之间是因果关系还是关联关系？  
群体彼此之间是存在交叠还是互不干涉？
- ▶ 关联性可以帮助我们根据某一已知指标来预测另一指标。

## 7.2 关联性——散点图 (探索更多的变量)

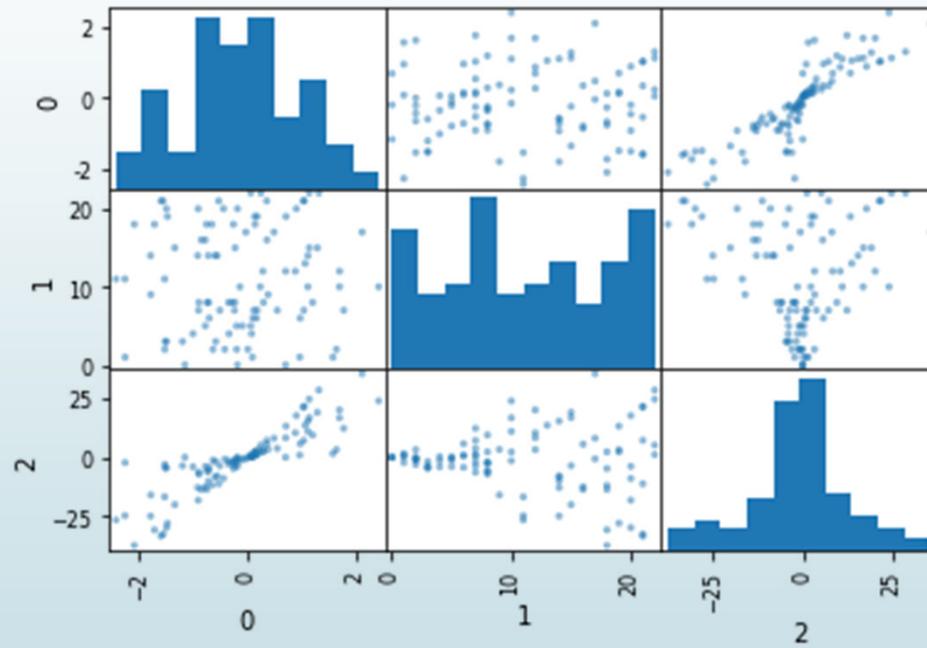


图. 散点图矩阵的基本框架



## 7.2 关联性 — 气泡图

优势在于一次比较3个变量:

➤ x轴

➤ y轴

➤ 气泡的面积

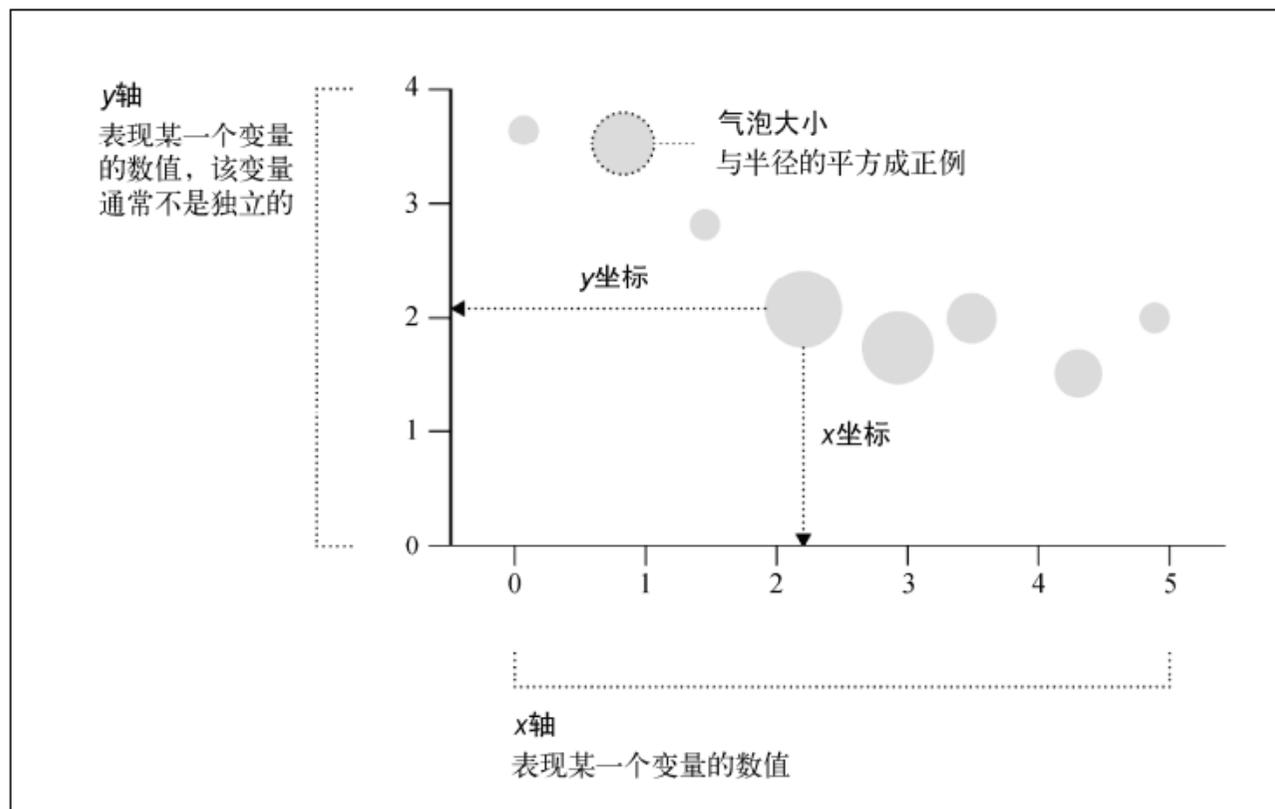


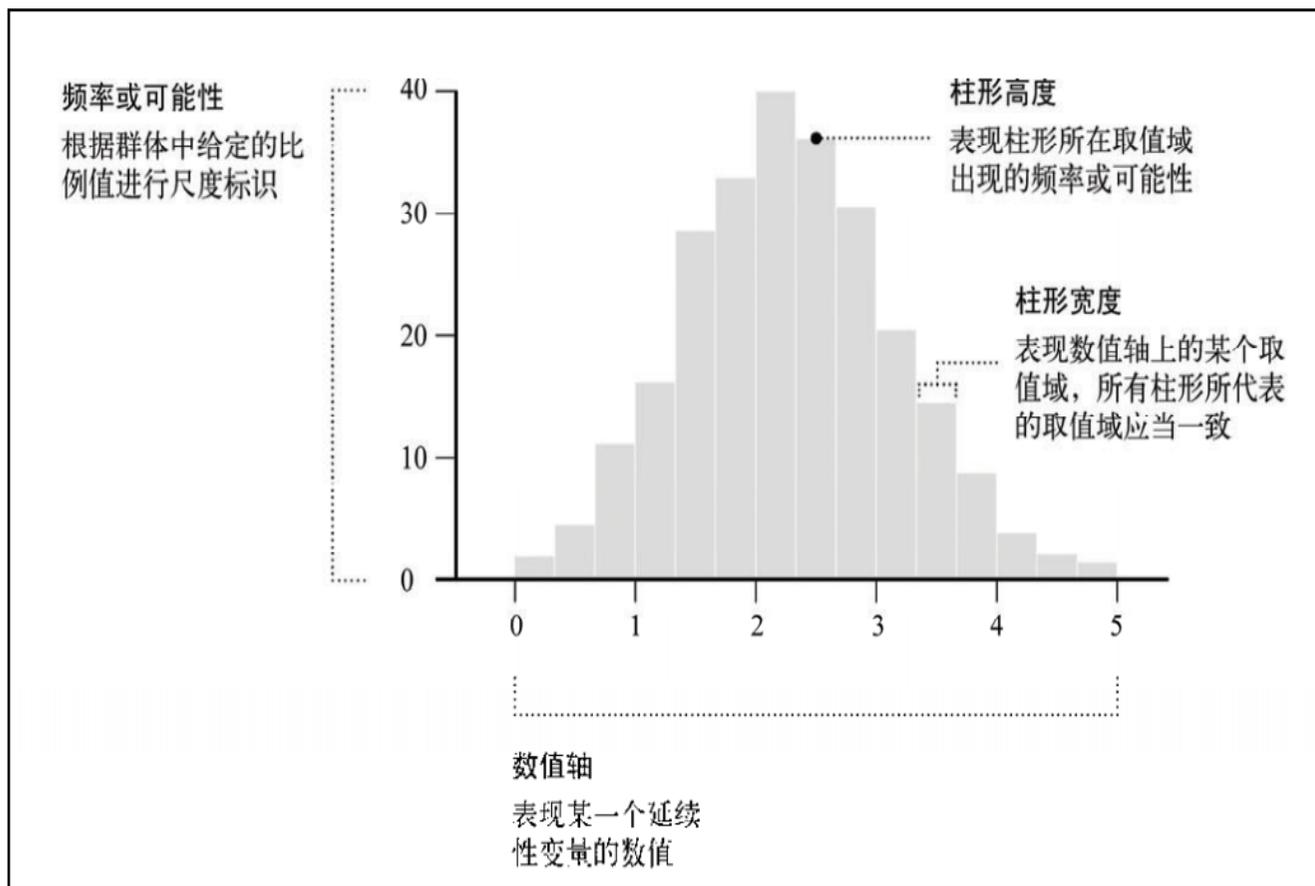
图. 气泡图的基本框架



## 7.3 分布

### 直方图 Hist()

用一系列宽度相等、高度不等的长方形来表示数据分布



**matplotlib.pyplot.hist(x, bins, range, density, weights, cumulative, bottom, histtype, \*\*kwargs)**

参数说明（部分）：

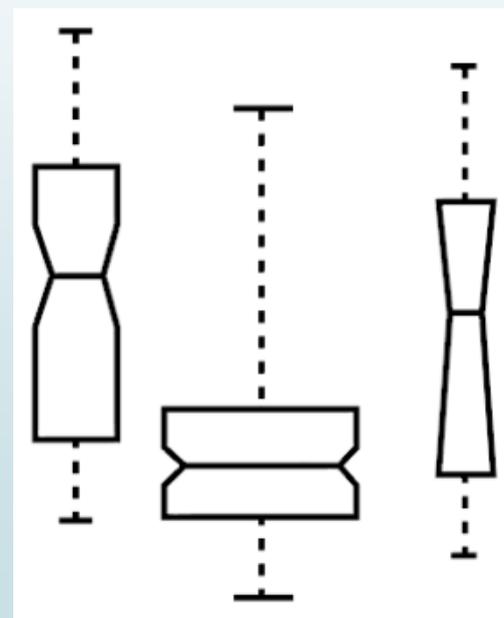
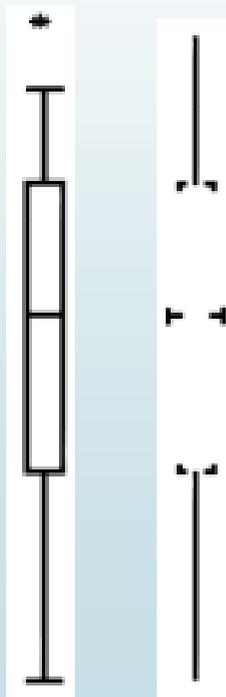
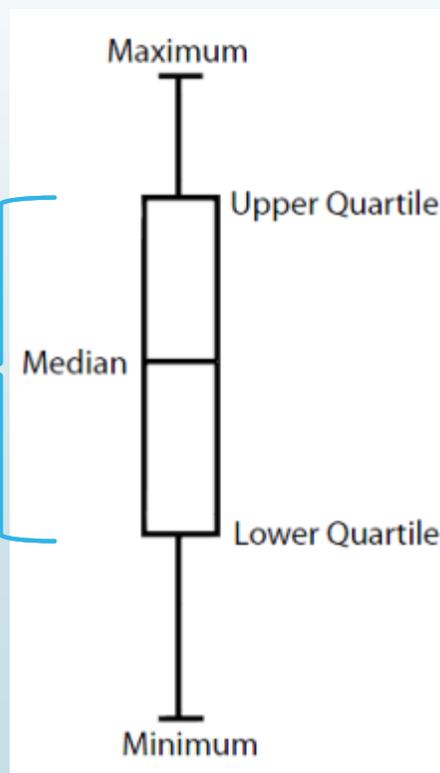
- **x**→作图所用的数据，一维数组，多维数组可以扁平化后再作图
- **bins**→直方图的柱数，可选项，整数、序列或字符串型。默认10
- **range**→指定直方图数据的上下界，默认包含绘图数据的最大值和最小值（范围）
- **density**→布尔值，如果为"True"，将y轴转化为密度刻度，默认为None
- **weights**→为每个数据点设置权重
- **cumulative**→布尔值，是否需要计算累计频数或频率，默认为False
- **bottom**→为直方图添加基准线，数组型，默认为none
- **histtype**→直方图类型，分为bar,barstacked,step,stepfilled。默认为bar



# 盒须图（箱形图）

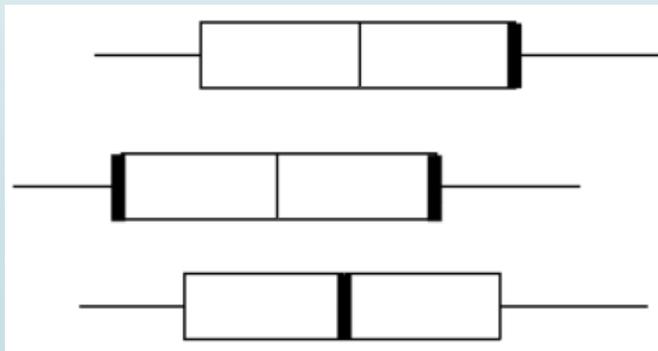
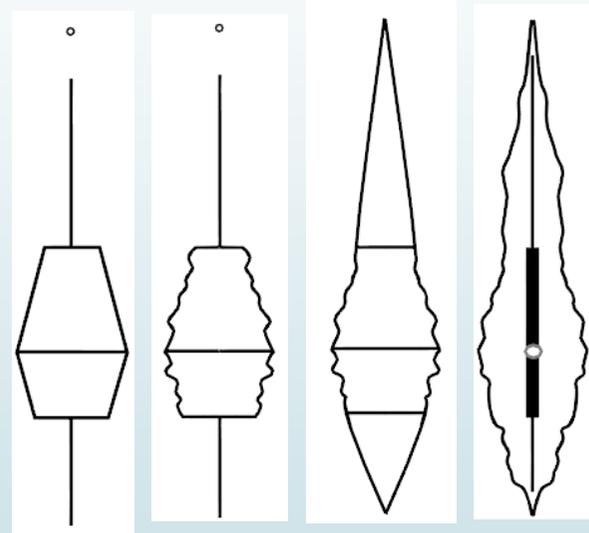
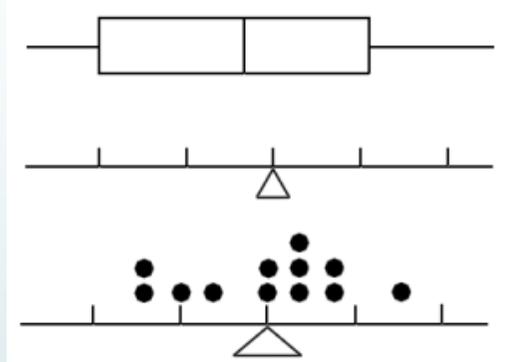
主要用于反映原始数据分布的特征，也可以进行多组数据分布特征的比较。

内距**IQR**  
Inter-  
Quartile  
Range，  
又称为四分位差。





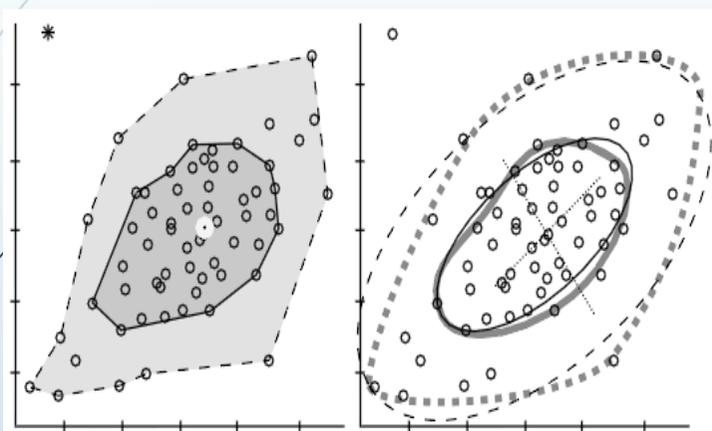
# 盒须图变种



可以水平或者垂直绘制

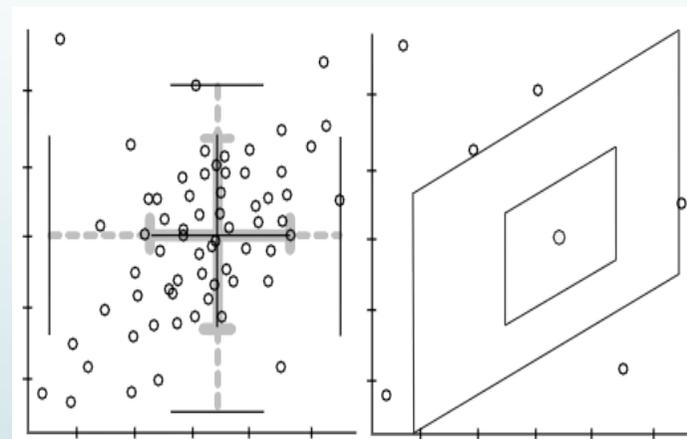


# 盒须图变种



2D Box Plot

Relplot



Rangefinder Box Plot

Bag Plot



`matplotlib.pyplot.boxplot(x, notch, vert, patch_artist, whis, widths, labels, **kwargs)`

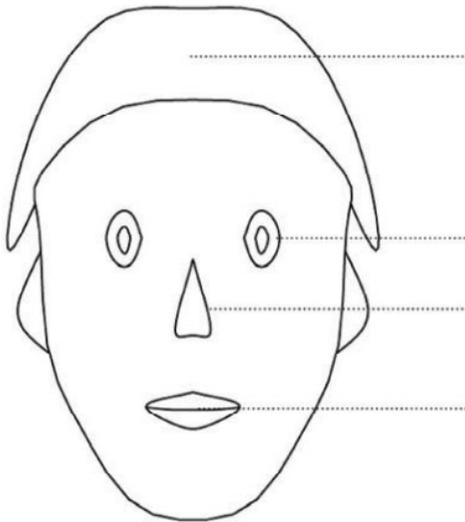
- **x**: 输入数据。类型为数组或向量序列。必备参数。
- **notch**: 控制箱体中央是否有V型凹槽。当取值为**True**时，箱体中央有V型凹槽，凹槽表示中位数的置信区间；取值为**False**时，箱体为矩形。类型为布尔值，默认值为**False**。可选参数。
- **sym**: 离群点的默认标记符号，详解**flierprops**参数。类型为字符串，默认值为'**None**'。可选参数。取值为"隐藏离群点，取值为'**None**'时，取值为'**b+**'。
- **vert**: 箱体的方向，当取值为**True**时，绘制垂直箱体，当取值为**False**时，绘制水平箱体。类型为布尔值，默认值为**True**。可选参数。
- **whis**: 箱须的位置。类型为浮点数或浮点数二元组，默认值为**1.5**。可选参数。

## 7.4 对照和比较 (多变量比较)

### 1. 切尔诺夫脸谱图 (Chernoff Faces)

- 根据数据，将多个变量一次性展现在人脸的各个部位上，例如耳朵、头发、眼睛、鼻子等

面部  
整体表现某个对象，  
或者数据的一整行



特征  
头发高度、眼睛高度、  
鼻子高度以及嘴角上  
扬的曲线等面部特征  
会根据它们代表的数  
据而发生改变

以上内容仅为本文档的试下载部分，为可阅读页数的一半内容。如要下载或阅读全文，请访问：<https://d.book118.com/678015143110006115>