



【matplotlib绘图库】

绘制箱线图

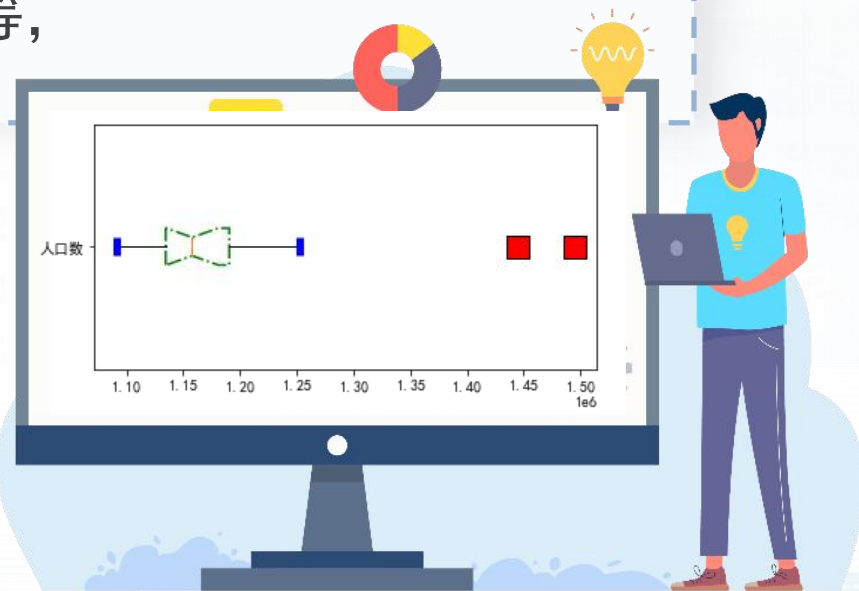
主讲：孙 睿





绘制箱线图

- 箱线图 (boxplot) 又称箱形图，因形状如箱子而得名。
- 箱线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，

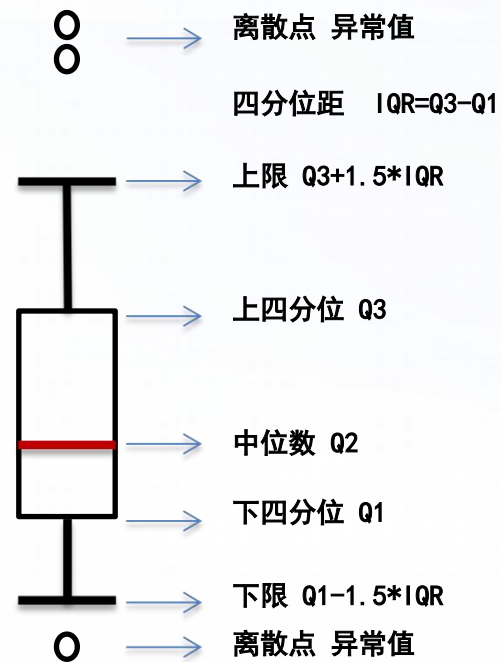




绘制箱线图

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)





>>> 基础语法

```
boxplot( x, notch=None, positions=None, whis=None, widths=None,
         patch_artist=None, bootstrap=None, usermedians=None,
         labels=None, conf_intervals=None, sym=None, vert=None,
         meanline=None, showmeans=None, showcaps=None,
         showbox=None, showfliers=None, boxprops=None,
         flierprops=None, medianprops=None, meanprops=None,
         zorder=None, whiskerprops=None, manage_ticks=True,
         data=None, autorange=False, *, capprops=None )
```

参数	说明
x	接收array或向量序列。表示输入数据。必备参数
labels	接收array。表示每个数据集的标签，默认值为'None'。可选参数
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



绘制箱线图

>>> 基础语法

参数	说明
showfliers	接收bool，表示是否显示离群值。默认值为True。可选参数
sym	接收str，离群点的默认标记符号。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+'
flierprops	接收dict，表示离群点的样式。默认值为None。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
capprops	接收dict，表示箱须横杠的样式。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
autorange	接收bool，默认值为False。可选参数。当取值为True且数据分布满足上四分位数（75%）和下四分位数（25%）相等，whis设置为(0, 100)，即箱须端点为数据的最大值和最小值
showbox	接收bool，表示是否显示箱体。默认值为True。可选参数
notch	接收bool，表示控制箱体中央是否有V形凹槽。当取值为True时，箱体中央有V形凹槽，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
vert	接收bool，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
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。默认值为range(1, N+1)，N为箱线图的个数
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
boxprops	接收dict，表示箱体的样式。默认值为None。可选参数
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。当取值为False时，箱体由Line2D生成，否则，箱体由Patch对象生成



绘制箱线图

基础语法

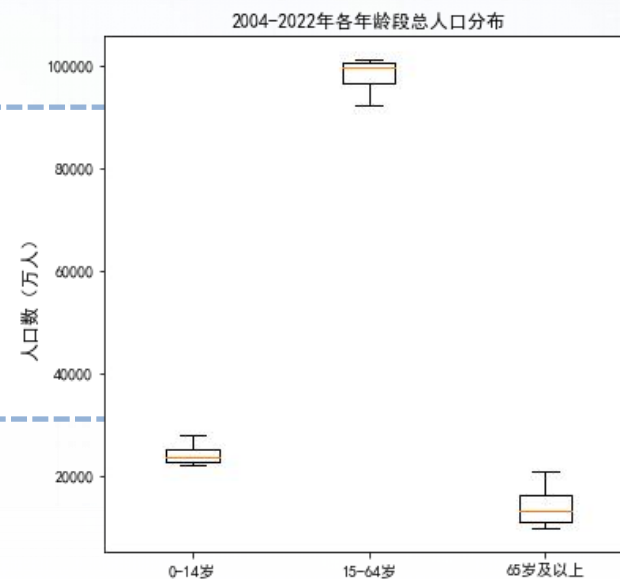
参数	说明
medianprops	接收dict, 表示中位数的样式。默认值为None。可选参数
manage_ticks	接收bool, 表示控制刻度和标签位置, 默认值为True。可选参数。取值为True时, 刻度和标签位置自动匹配箱线图的位置
showmeans	接收bool, 表示是否显示算术平均值。默认值为False。可选参数
medianprops	接收dict, 表示中位数的样式。默认值为None。可选参数
meanprops	接收dict, 表示算术平均值的样式。默认值为None。可选参数
manage_ticks	接收bool, 表示控制刻度和标签位置, 默认值为True。可选参数。取值为True时, 刻度和标签位置自动匹配箱线图的位置
meanline	接收bool, 表示均值显示为线还是点。默认值为False。可选参数。当取值为True, 且showmeans、shownotches参数均为True, 时显示为线, 线条属性受meanprops参数控制; 否则显示为点
meanprops	接收dict, 表示算术平均值的样式。默认值为None。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。当参数取值为None时, 中位数置信区间由某种高斯渐进逼近算法确定。否则, 采用bootstrap方法求中位数95%置信区间, bootstrap 参数定义了抽样次数。建议取值范围为1000-10000
usermedians	表示是否指定中位数。类型为一维类数组结构。可选参数。对于一维数组, 数组元素个数为1, 对于多维数组, 长度等于len(x)。如果元素不为None, 则将该值强制设置为对应数据集的中位数; 如果元素为None, 则由matplotlib生成中位数
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。如果元素不为None, 则将该值强制设置为箱体V型凹槽位置 (只有当notch参数为True时绘制); 如果元素为None, 箱体V型凹槽的位置由其他参数计算, 例如bootstrap
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。浮点数, 则下箱须位于高于 $Q1 - whis * (Q3 - Q1)$ 的最低数据处, 上箱须位于低于 $Q3 + whis * (Q3 - Q1)$ 下方的最高数据处, 其中Q1和Q3分别为下四分位数和上四分位数。默认值whis=1.5对应于箱线图的原始定义。浮点数元组, 则表示要在绘制箱须的百分位数 (例如, (5, 95))。将其设置为 (0, 100) 则箱须覆盖整个数据范围。当 $Q1 == Q3$ 时, 如果autorange为True, whis将自动设置为 (0, 100)。箱须范围之外的数据将被视为异常值, 绘制为点



绘制箱线图

➤➤➤ 绘制箱线图，观察近年来我国人口年龄结构变化趋势

```
plt.boxplot( [df['0-14岁人口(万人)'],  
              df['15-64岁人口(万人)'],  
              df['65岁及以上人口(万人)']],  
            labels=['0-14岁', '15-64岁', '65岁及以上'])
```

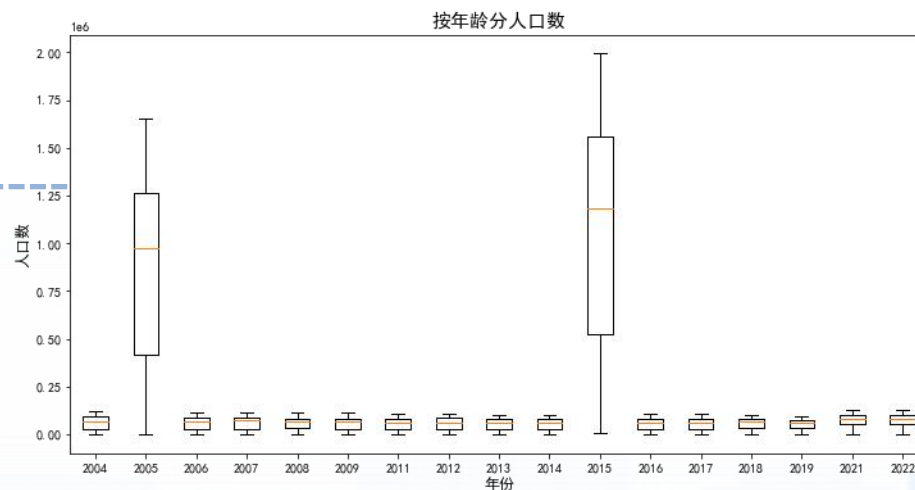




绘制箱线图

>>> 根据采样人口数绘制箱线图

```
df=pd.read_Excel('人口数据.xlsx','按年龄分人口数',index_col=0)
data=df.iloc[:,1:].T
fig=plt.figure(figsize=(12,6))
plt.boxplot(data,labels=data.columns)
plt.title('按年龄分人口数',fontsize=15)
plt.ylabel('人口数',fontsize=12)
plt.xlabel('年份',fontsize=12)
plt.show()
```

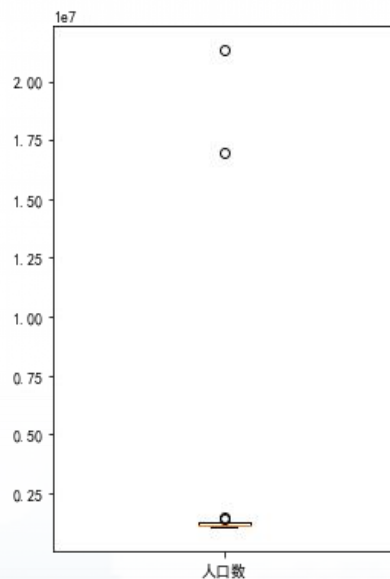




绘制箱线图

>>> 箱线图的离群点（异常值）

```
plt.boxplot(df['人口数'], labels=['人口数'])
```

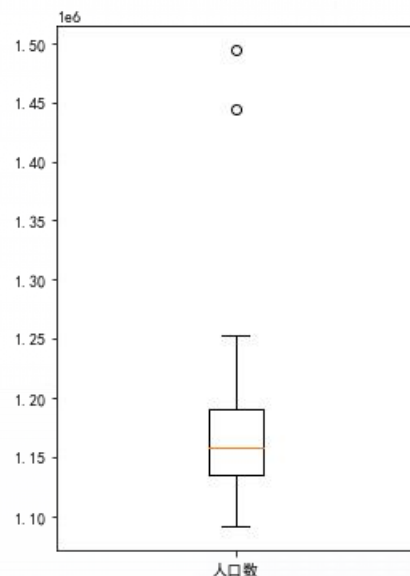




绘制箱线图

>>> 箱线图的离群点（异常值）

```
plt.boxplot(df['人口数'].drop([2005, 2015]), labels=['人口数'])
```

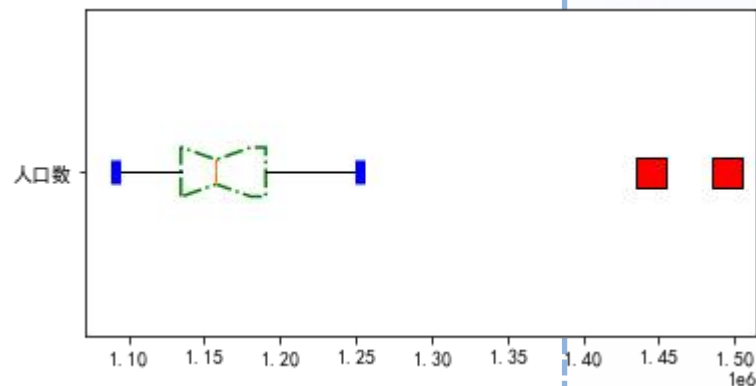




绘制箱线图

>>> 设置箱形图的外观

```
plt.boxplot(df['人口数'].drop([2005,2015]),labels=['人口数'],  
            vert = False,  
            notch = True,  
            flierprops = { "marker": "s",  
                           "markerfacecolor": "red",  
                           "markersize": 15},  
            capprops = {"linestyle": "--",  
                        "color": "b",  
                        "linewidth": 5},  
            boxprops = { 'color': 'g',  
                          'linestyle': '-.',  
                          'linewidth': 1.5},  
            meanline = {'color': 'red',  
                        'linewidth': 1.5}))
```



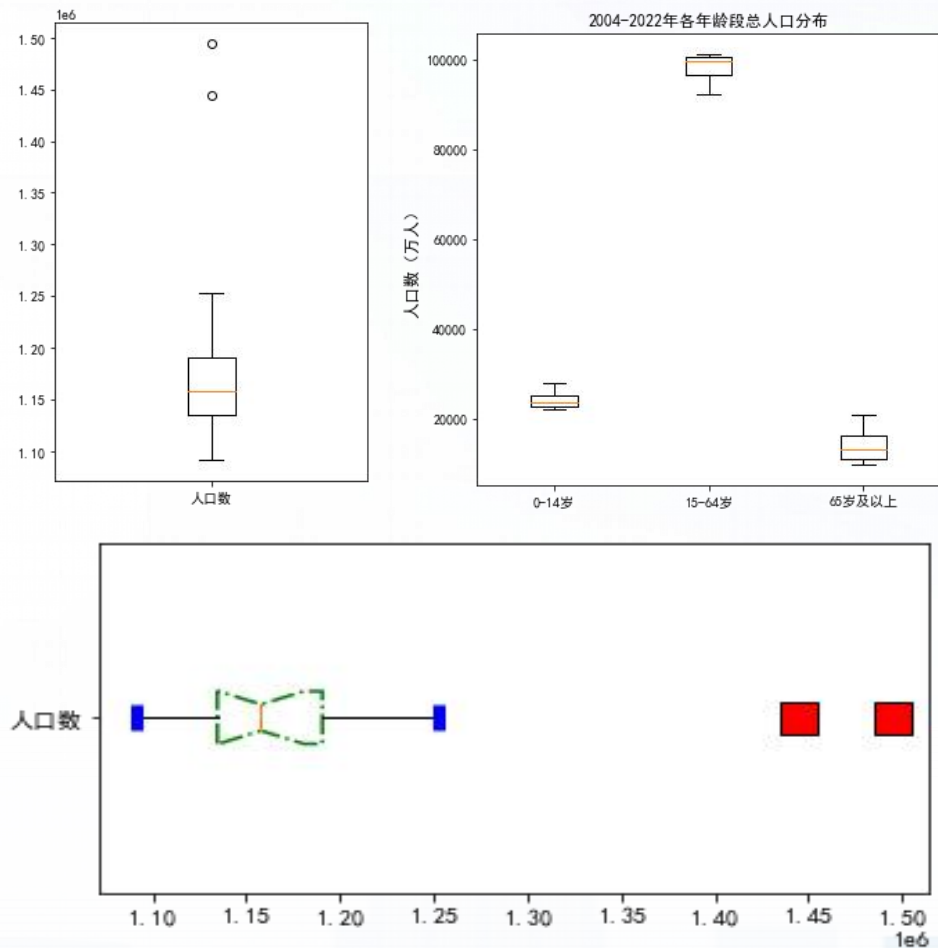


绘制箱线图

>>> 总结

`matplotlib.pyplot.boxplot ()`

x	输入数据。
sym	离群点的默认标记符号。
以及其他参数设置箱形图的外观	





【matplotlib绘图库】

绘制箱线图

主讲：孙 睿

