



【Seaborn绘图库】

绘制箱线图

主讲：韩敏





绘制箱线图

>>> 箱线图概述

- 箱线图，也被称为盒须图、盒式图，是一种用于显示数据分布情况的统计图。
- 它主要展示了一组数据的中位数、上下四分位数、最大值、最小值以及异常值（离群点）。



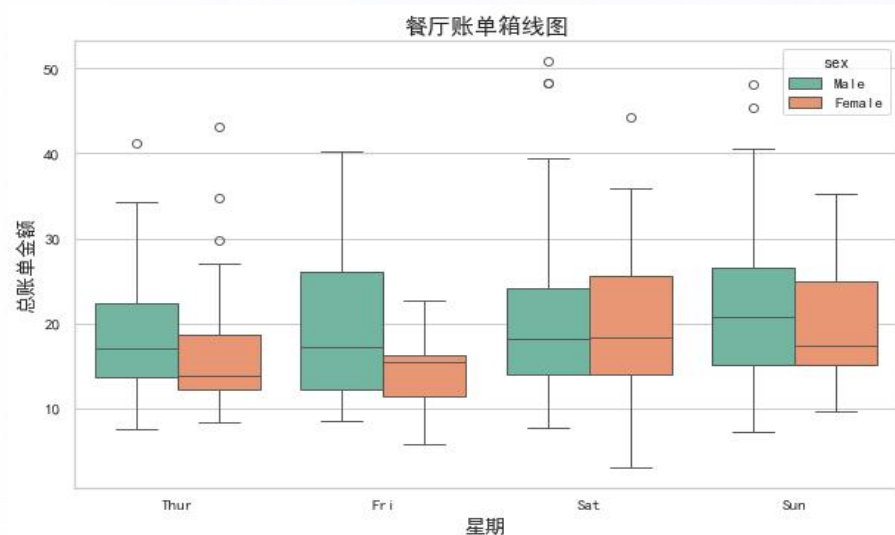


绘制箱线图

>>> 箱线图要素

箱线图的绘制过程包括以下几个要素：

- 箱体： 画出上下四分位数（Q1和Q3）之间的矩形，中间的线表示中位数（Q2）。
- 须： 从箱体延伸出来的线段，表示数据的范围。通常，须的长度由数据的分布情况确定，可以是数据的最小值和最大值，也可以根据一定的标准计算。
- 异常值： 在图表上以点的形式标出的数据点，它们偏离了常规的数据分布。异常值的定义通常是根据一定的统计学标准。





绘制箱线图

>>> 箱线图语法

Seaborn 中用于绘制箱线图的主要函数是 `boxplot()` 函数，以下是 `boxplot` 函数的完整形式及一些常用参数：

```
seaborn.boxplot(x=None, y=None, hue=None, data=None, order=None, hue_order=None, orient=None, color=None, palette=None, saturation=0.75, width=0.8, dodge=True, fliersize=5, linewidth=None, whis=1.5, notch=False, ax=None, **kwargs)
```

参数	说明
x	指定X轴数据
hue	可选参数，用于对数据进行分组，可以是另一个分类变量，会以不同颜色表示
data	数据集，可以是 DataFrame 或其他数据结构
order, hue_order	可选参数，用于指定X轴或 hue 分组的顺序
orient	可选参数，控制绘图的方向，可以是 "v"（垂直）或 "h"（水平）
color	可选参数，用于设置箱线图的颜色
palette	可选参数，用于设置颜色调色板
saturation	控制颜色的饱和度
width	控制箱线图的箱体宽度

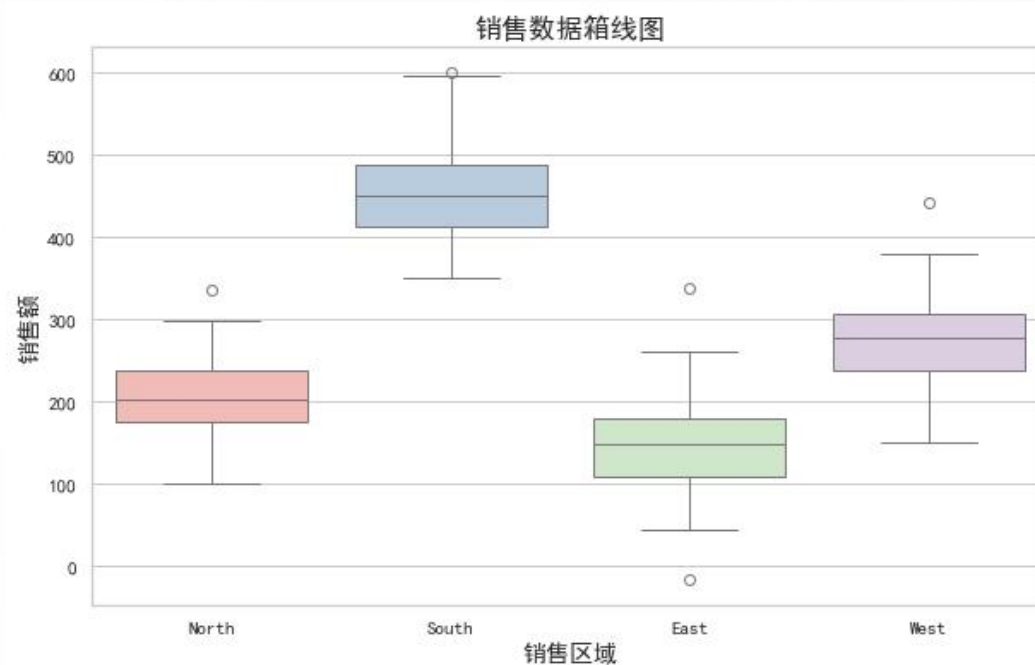


绘制箱线图

>>> 箱线图案例实战

案例介绍:

- 模拟包含不同销售区域的销售数据，并使用 Seaborn 的 boxplot 函数绘制了箱线图。箱线图展示了每个销售区域销售额的中位数、分布范围和可能的异常值。





绘制箱线图

>>> 箱线图案例实战

案例代码

```
import seaborn as sns
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
np.random.seed(42)
regions = ['North', 'South', 'East', 'West']
sales_data = {region: np.random.normal(np.random.randint(100, 500), 50, 100) for region in regions}
df_sales = pd.DataFrame(sales_data)
plt.figure(figsize=(10, 6))
sns.boxplot(data=df_sales, palette="Pastel1")
plt.title("销售数据箱线图", fontsize=16)
plt.xlabel("销售区域", fontsize=14)
plt.ylabel("销售额", fontsize=14)
plt.show()
```

