

# Seaborn

xbZhong

2024-04-11

[本页PDF](#)

## 导包

```
import seaborn as sns
```

## 数据集

seaborn内置了多个数据集，可以用`sns.load_dataset()`来读取

- `sns.load_dataset(name, cache=True, data_home=None, **kwargs)`
  - `name` : 数据集的名称，字符串类型。例如，可以是 `'tips'`、`'iris'`、`'flights'` 等。
  - `cache` : 布尔值，指定是否将下载的数据集缓存到本地（默认值为 `True`）。
  - `data_home` : 数据集缓存的位置，默认为 `~/.seaborn-data`。
  - `**kwargs` : 其他参数，可选，用于传递给数据加载函数。

```
1 import seaborn as sns
2 ## 加载名为car_crashes的数据集
3 data = sns.load_dataset('car_crashes')
```

## 分布图

- `sns.displot(data=None, x=None, y=None, kind='hist', ...)` : 绘制分布图
  - `data` : 输入数据，可以是 `DataFrame`、数组或列表。
  - `x` : 需要绘制分布的变量名（x 轴）。
  - `y` : 选择的变量名（可选），用于绘制散点图等。
  - `kind` : 指定绘制的图形类型，`'hist'`（默认）绘制直方图，`'kde'` 绘制核密度估计图。

```
1  ## 绘制分布图
2  import seaborn as sns
3  import matplotlib.pyplot as plt
4  import numpy as np
5
6  ## 生成随机数据
7  np.random.seed(0)
8  data = np.random.randn(1000)
9
10 ## 使用 displot 绘制分布图
11 sns.displot(data, kde=True) # kde=True 添加核密度估计
12
13 ## 设置标题
14 plt.title('Distribution of Random Data')
15
16 ## 显示图形
17 plt.show()
18
```

## 联合图

- `sns.jointplot(x, y, data, kind)`
  - `x` :
    - 类型: `str`
    - 描述: 要在 x 轴上绘制的变量的名称。
  - `y` :
    - 类型: `str`
    - 描述: 要在 y 轴上绘制的变量的名称。
  - `data` :
    - 类型: `DataFrame`
    - 描述: 要使用的数据集，通常是 Pandas 的 `DataFrame`。
  - `kind` :
    - 类型: `str`
    - 描述: 指定要绘制的图的类型。可选值包括：
      - `'scatter'` : 散点图。
      - `'kde'` : 核密度估计图。
      - `'hist'` : 直方图。
      - `'hex'` : 六边形热图。

```
1 import seaborn as sns
2 import matplotlib.pyplot as plt
3
4 ## 加载示例数据集
5 tips = sns.load_dataset("tips")
6
7 ## 创建 jointplot
8 sns.jointplot(x='total_bill', y='tip', data=tips, kind='scatter')
9
10 ## 显示图形
11 plt.show()
```

## 核密度图

### 生成概率密度图，平滑曲线

- `sns.kdeplot(x,y,)`
  - **data** :
    - 类型: `array-like` 或 `DataFrame`
    - 描述: 用于计算核密度估计的输入数据。
  - **x** :
    - 类型: `str` 或 `array-like`
    - 描述: 要绘制的变量名称。如果提供 `data`，则此参数应为变量的名称。
  - **bw\_adjust** :
    - 类型: `float`
      - 描述: 用于控制核密度估计的带宽的调整因子。带宽越大，曲线越平滑；带宽越小，曲线越细致。默认值为 `1`
  - **fill** :
    - 类型: `bool`
      - 描述: 是否填充 KDE 曲线下方的区域，默认值为 `False`
  - **color** :
    - 类型: `str`
    - 描述: 指定绘图的颜色。
  - **alpha** :
    - 类型: `float`
    - 描述: 填充区域的透明度，取值范围为 `0` 到 `1`，默认值为 `1`。
  - **ax** :
    - 类型: `matplotlib.axes.Axes`

- 描述：可选参数，指定在哪个坐标轴上绘制图形。
- **legend** :
  - 类型： `bool`
  - 描述：是否显示图例，默认值为 `True`。

```
1 import seaborn as sns
2 import matplotlib.pyplot as plt
3
4 ## 加载示例数据集
5 tips = sns.load_dataset("tips")
6
7 ## 绘制总账单的核密度估计图
8 sns.kdeplot(data=tips['total_bill'], fill=True, color='blue', bw_adjust=0.5)
9
10 ## 添加标题
11 plt.title('Kernel Density Estimate of Total Bill')
12
13 ## 显示图形
14 plt.show()
```