

第七课 数据可视化
第12节 数据集分布可视化(1) --
单变量分布、双变量分布
主 讲: Robin

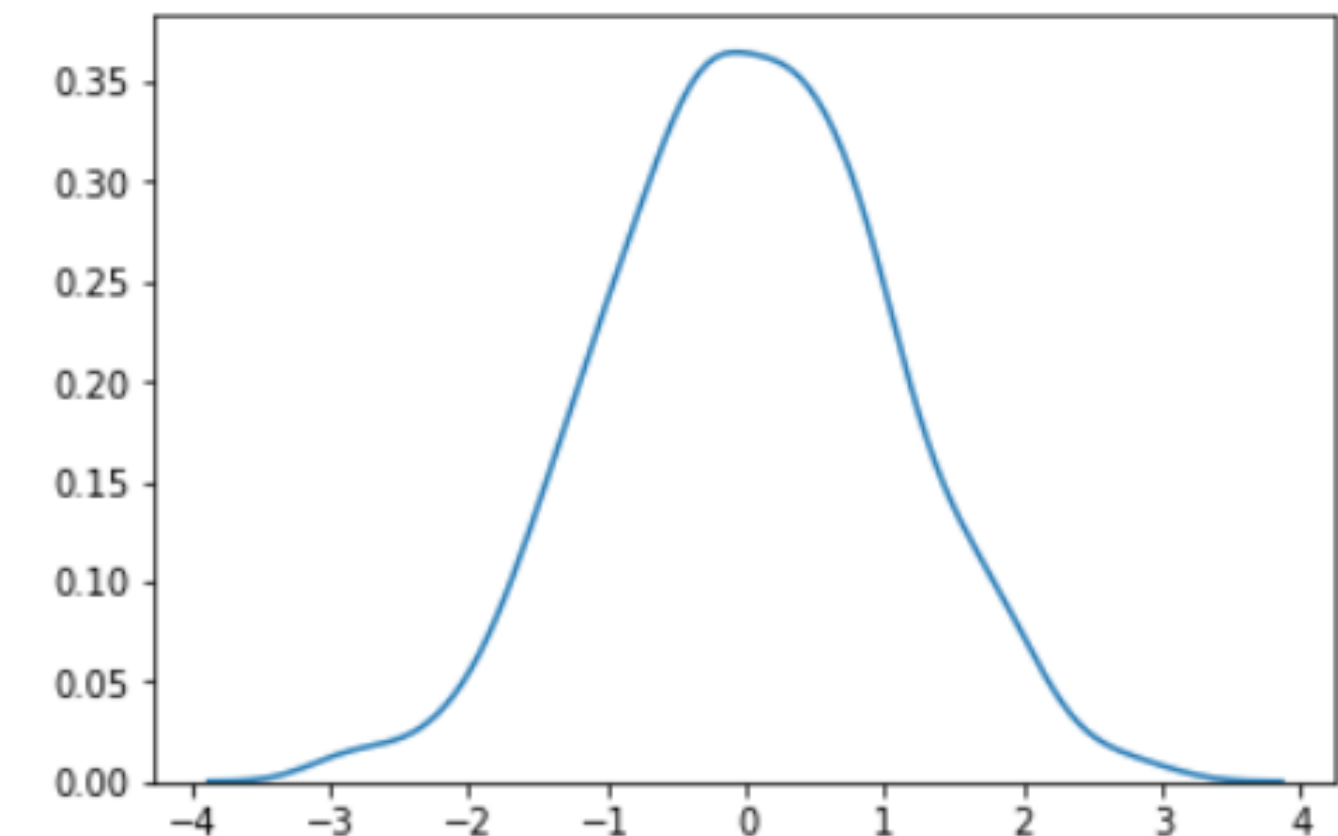
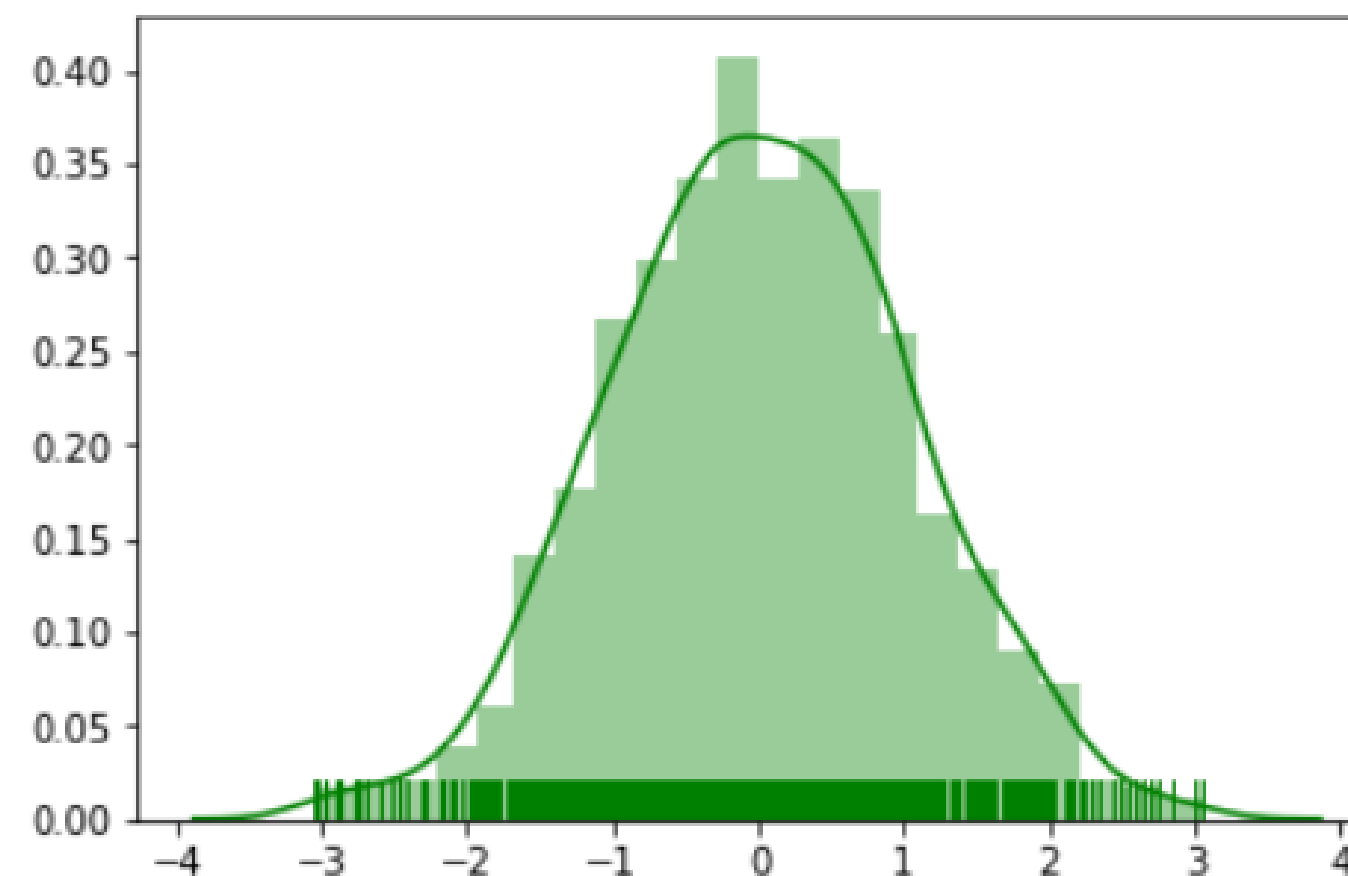
单变量分布

核密度估计图 `sns.kdeplot(data)`

- 核密度估计(kernel density estimation)是在概率论中用来估计未知的密度函数
- 通过核密度估计图可以比较直观的看出数据样本本身的分布特征

`sns.distplot(kde=True, hist=True, rug=False)`

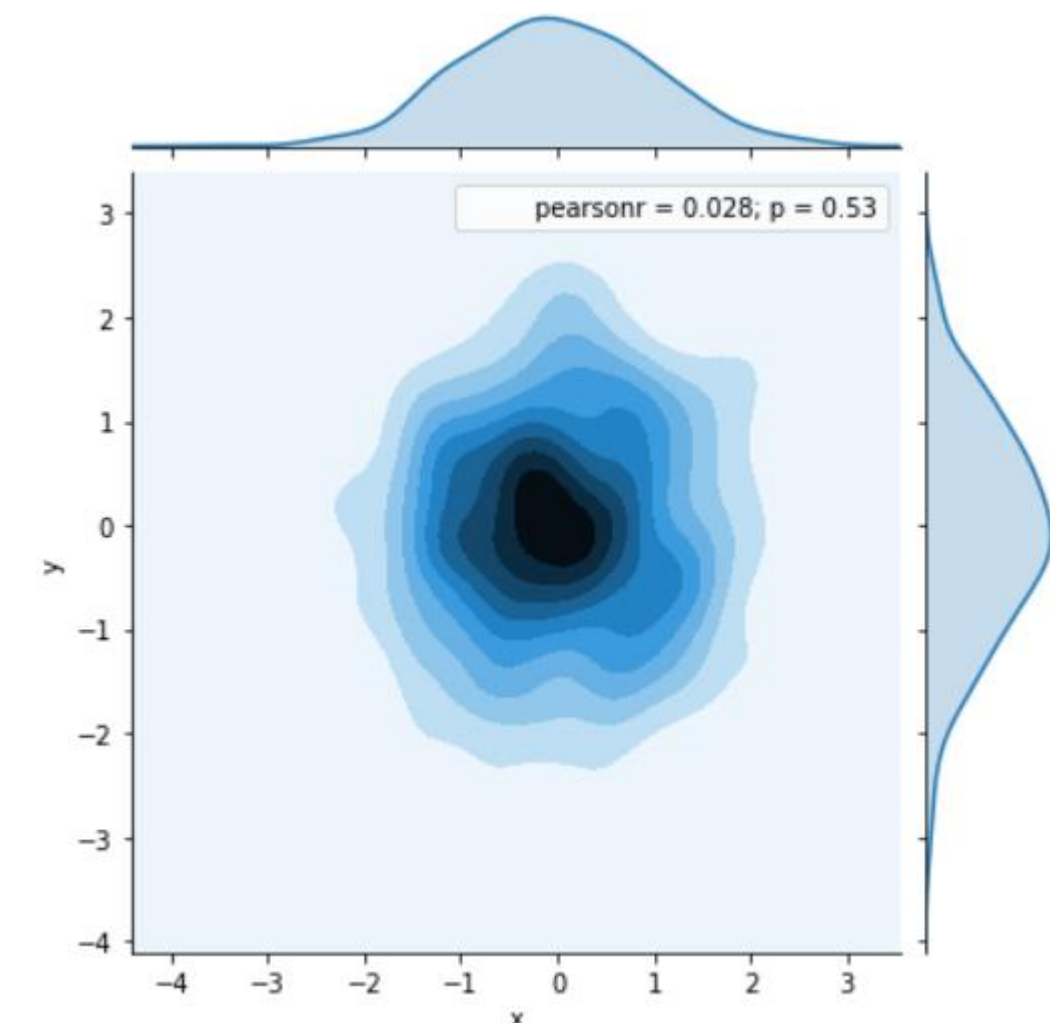
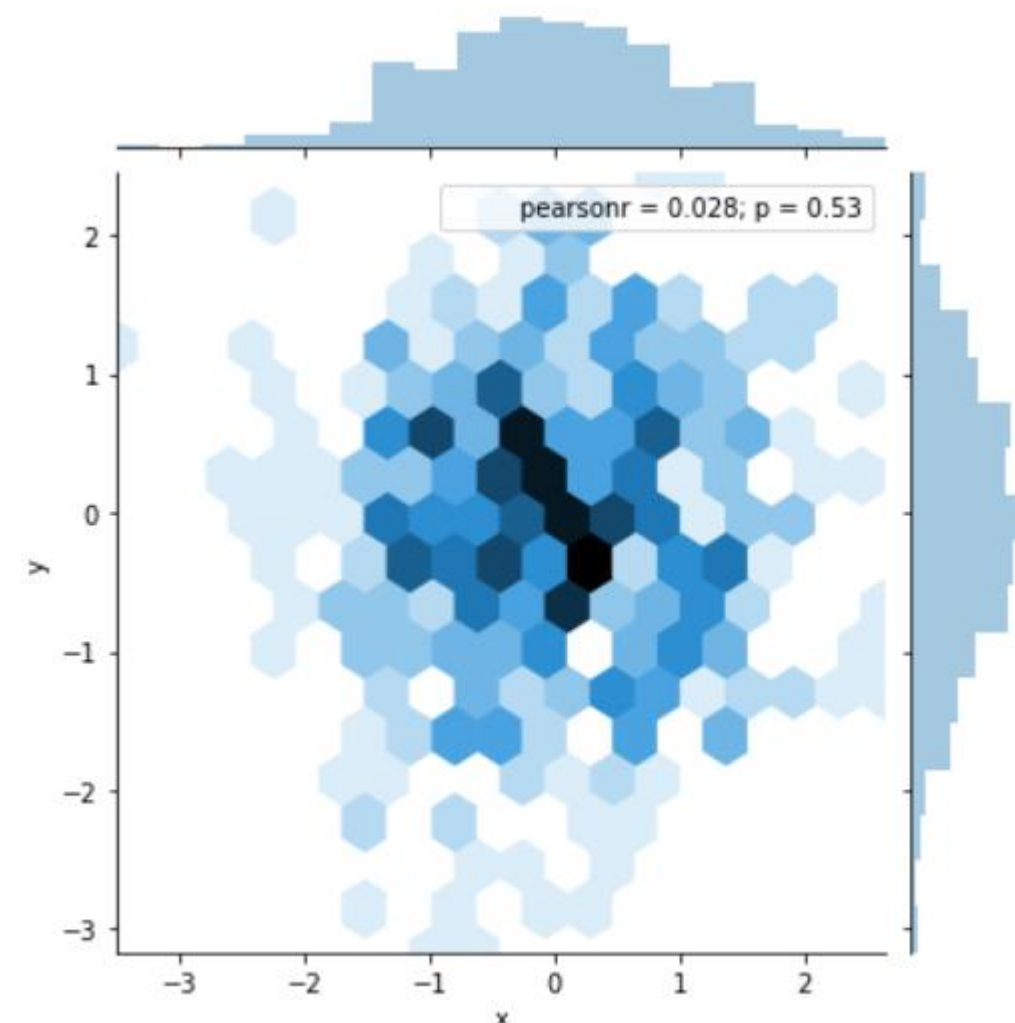
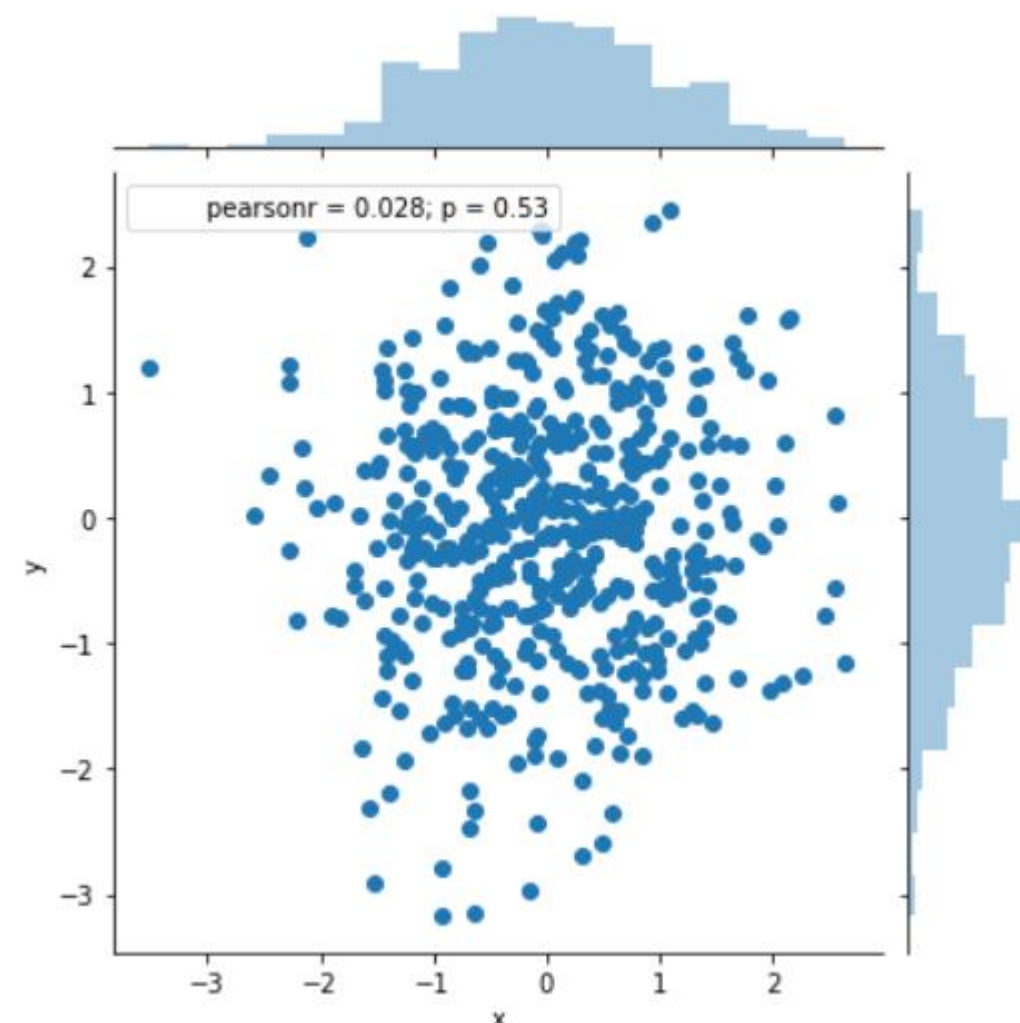
- 集合了直方图与核函数估计的功能
- 只显示直方图 `kde=False`
- 只显示核密度估计 `hist=False`
- 显示观测条 `rug=True`



双变量分布

`sns.jointplot(x, y, data, kind)`

- x, y 二维数据，向量或字符串
- data, 如果 x, y 是字符串 data 应该为 DataFrame
- kind='scatter' 默认，二维散点图
- kind='hex', 二维直方图
- kind='kde', 二维核密度估计图



只为遇见明天更优秀的你！