

seaborn-05

用分类数据绘图

类别内的数据分布

箱形图

箱形图 (Box-plot) 又称为盒须图、盒式图或箱线图，是一种用作显示一组数据分散情况资料的统计图。因形状如箱子而得名

它能显示出一组数据的最大值、最小值、中位数、及上下四分位数。

规模

上极限

上四分位数

中位数

下四分位数

晶须

下极限

异常值/单一数据点

API

seaborn.boxplot(x=None, y=None, hue=None, data=None, orient=None, color=None, saturation=0.75, width=0.8)

- (1) palette: 用于设置不同级别色相的颜色变量。---- palette=["r","g","b","y"]
- (2) saturation: 用于设置数据显示的颜色饱和度。---- 使用小数表示

小提琴图

用于显示数据分布及其概率密度

这种图表结合了箱形图和密度图的特征，主要用来显示数据的分布形状。

中间黑色粗条表示四分位数范围，从其延伸的幼细黑线代表 95% 置信区间，而白点则为中位数。

箱形图在数据显示方面受到限制，简单的设计往往隐藏了有关数据分布的重要细节

(规模)

95% 置信区间

密度图宽 = 频率

中位数

四分位数范围

API

seaborn.violinplot(x=None, y=None, hue=None, data=None)

类别散点图

seaborn.stripplot(x=None, y=None, hue=None, data=None, order=None, hue_order=None, jitter=False)

- (1) x, y, hue: 用于绘制数据的输入。
- (2) data: 用于绘制的数据集。
- (3) jitter: 表示抖动的程度(仅沿类别轴)。当很多数据点重叠时，可以指定抖动的数量或者设为True使用默认值

swarmplot0函数绘制散点图

sns.swarmplot(x="day", y="total_bill", data=tips)

绘制成对的双变量分布

sns.pairplot(dataset)

绘制图像

绘制散点图

sns.jointplot(x="x", y="y", data=dataframe_obj)

绘制二维直方图

二维直方图类似于"六边形"图，主要是因为它显示了落在六角形区域内的观察值的计数，适用于较大的数据集

sns.jointplot(x="x", y="y", data=dataframe_obj, kind="hex")

绘制核密度估计图形

sns.jointplot(x="x", y="y", data=dataframe_obj, kind="kde")

绘制双变量分布

seaborn.jointplot(x, y, data=None, kind='scatter', stat_func=None, color=None, ratio=5, space=0.2, dropna=True)

- (1) kind: 表示绘制图形的类型。
- (2) stat_func: 用于计算有关关系的统计量并标注图。
- (3) color: 表示绘图元素的颜色。
- (4) size: 用于设置图的大小(正方形)。
- (5) ratio: 表示中心图与侧边图的比例。该参数的值越大，则中心图的占比会越大。
- (6) space: 用于设置中心图与侧边图的间隔大小

绘制单变量分布

seaborn.distplot(a, bins=None, hist=True, kde=True, rug=False, fit=None, color=None)

- (1) a: 表示要观察的数据，可以是 Series、一维数组或列表。
- (2) bins: 用于控制条形的数量。
- (3) hist: 接收布尔类型，表示是否绘制(标注)直方图。
- (4) kde: 接收布尔类型，表示是否绘制高斯核密度估计曲线。
- (5) rug: 接收布尔类型，表示是否在支持的轴方向上绘制rugplot

安装

pip3 install seaborn

导入

import seaborn as sns

seaborn绘制统计图形

案例：NBA球员数据分析