数据分析常见图像

1. Seaborn
   1. 简介

Seaborn 是基于matplotlib的python 数据可视化库，提供一些画图的接口从而画出数据分析图像

* 1. 安装方式

基于Python: pip install seaborn

基于anaconda: conda install seaborn

* 1. 常见图像

1.3.0 图像归纳： lmplot, ScatterPlot

1.3.1 lmplot 回归图

seaborn.lmplot(x, y, data, hue=None, col=None, row=None, palette=None, col\_wrap=None, height=5, aspect=1, markers='o', sharex=True, sharey=True, hue\_order=None, col\_order=None, row\_order=None, legend=True, legend\_out=True, x\_estimator=None, x\_bins=None, x\_ci='ci', scatter=True, fit\_reg=True, ci=95, n\_boot=1000, units=None, order=1, logistic=False, lowess=False, robust=False, logx=False, x\_partial=None, y\_partial=None, truncate=False, x\_jitter=None, y\_jitter=None, scatter\_kws=None, line\_kws=None, size=None)

回归图的作用是显示每个数据集内线性回归的结果，可以通过Implot直观地总览数据的内在关系。

常见参数

X,Y:分别是该数据集的两列，这个函数可以通过绘制一条函数曲线来拟合数据集

Hue：用于分类，根据hue指定的这列的分类，把数据集分成多类并分别拟合

Order：控制进行回归的幂次(一次以上就是多项式回归)

Col & Row： Col是按属性在列上分类，简单来说就是横着一行；Row相反是按行分，最后效果是竖着一列。（可以自由组合）

Share：共享坐标轴尺度

1.3.2 ScatterPlot 散点图

散点图能够显示2个维度上2组数据的值。每个点代表一个观察点。X（水平）和Y（垂直）轴上的位置表示变量的值。研究这两个变量之间的关系是非常有用的。

seaborn.scatterplot(x=None, y=None, hue=None,

style=None, size=None, data=None,

palette=None, hue\_order=None, hue\_norm=None,

sizes=None, size\_order=None, size\_norm=None,

markers=True, style\_order=None, x\_bins=None,

y\_bins=None, units=None, estimator=None,

ci=95, n\_boot=1000, alpha='auto', x\_jitter=None,

y\_jitter=None, legend='brief', ax=None, \*\*kwargs)

常见参数：

X，Y：同上；

Hue： 同上

Style：根据这一列生成不同的标记点

* + 1. Lineplot 折线图

常见参数

X,Y:分别是该数据集的两列，这个函数可以通过绘制一条函数曲线来拟合数据集

Hue：用于分类，根据hue指定的这列的分类，把数据集分成多类并分别拟合

Size:在图像的绘制中会根据指定列的分类生成不同的大小

Markers：可以把数据点进行标注

err\_style：{"bars","bands"}，能够在曲线上画线段，并且用线段的宽度来显示计算的误差

units：基于这一列分组和采样，然后每一个unit会以独立的line画出来。

* + 1. relplot (关系图)

replot分两种：kind=’scatter’是散点图，kind=’line’是曲线图

X,Y，Hue: 同上

Kind = {‘line’, ’scatter’}

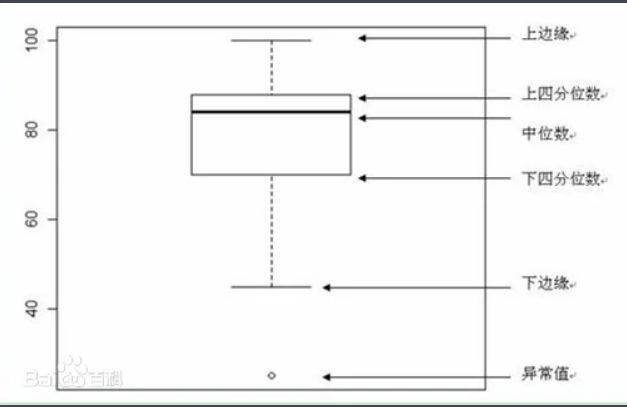
* + 1. Catplot (分类型数据关系图)

Catplot分两种：kind=‘strip‘(默认) 分类散点图；kind = ’swarm’ 蜂群图；kind = ‘box’ 箱型图; kind = ‘boxen’ 箱型图进阶版；kind = ’violin’ 小提琴图；kind = ‘bar’ 条形图; kind = ‘count’ (分类变量而非连续变量的)直方图;

略写，主要方法都可以参考我介绍的其他图的含义和作用。这个对象的价值主要在于对关系型数据的单独的归类上。

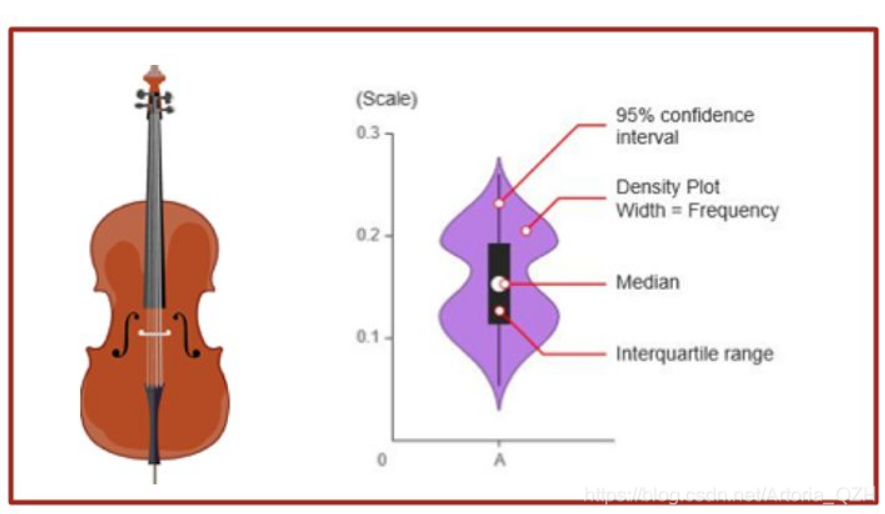
* + 1. boxplot（箱型图）

箱形图（Box-[plot](https://so.csdn.net/so/search?q=plot&spm=1001.2101.3001.7020)）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。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**。因形状如箱子而得名。在各种领域也经常被使用，常见于品质管理。图解如下：



* + 1. Violinplot (小提琴图)

小提琴图是[箱线图](https://so.csdn.net/so/search?q=%E7%AE%B1%E7%BA%BF%E5%9B%BE&spm=1001.2101.3001.7020)与核密度图的结合，箱线图展示了分位数的位置，核密度图则展示了任意位置的密度，通过小提琴图可以知道哪些位置的数据点聚集的较多，因其形似小提琴而得名。



seaborn.violinplot(\*, x=None, y=None, hue=None, data=None,

order=None, hue\_order=None, bw='scott', cut=2,

scale='area', scale\_hue=True, gridsize=100, width=0.8,

inner='box', split=False, dodge=True, orient=None,

linewidth=None, color=None, palette=None, saturation=0.75,

ax=None, \*\*kwargs)

* + 1. Jointplot ()

seaborn.jointplot(\*, x=None, y=None, data=None, kind='scatter', color=None,

height=6, ratio=5, space=0.2, dropna=False, xlim=None,

ylim=None, marginal\_ticks=False, joint\_kws=None,

marginal\_kws=None, hue=None, palette=None, hue\_order=None,

hue\_norm=None, \*\*kwargs)

在[seaborn](https://so.csdn.net/so/search?q=seaborn&spm=1001.2101.3001.7020)中，jointplot函数提供了一种综合的可视化方案，可以同时绘制两组变量的散点图以及各自对应的直方图。

