

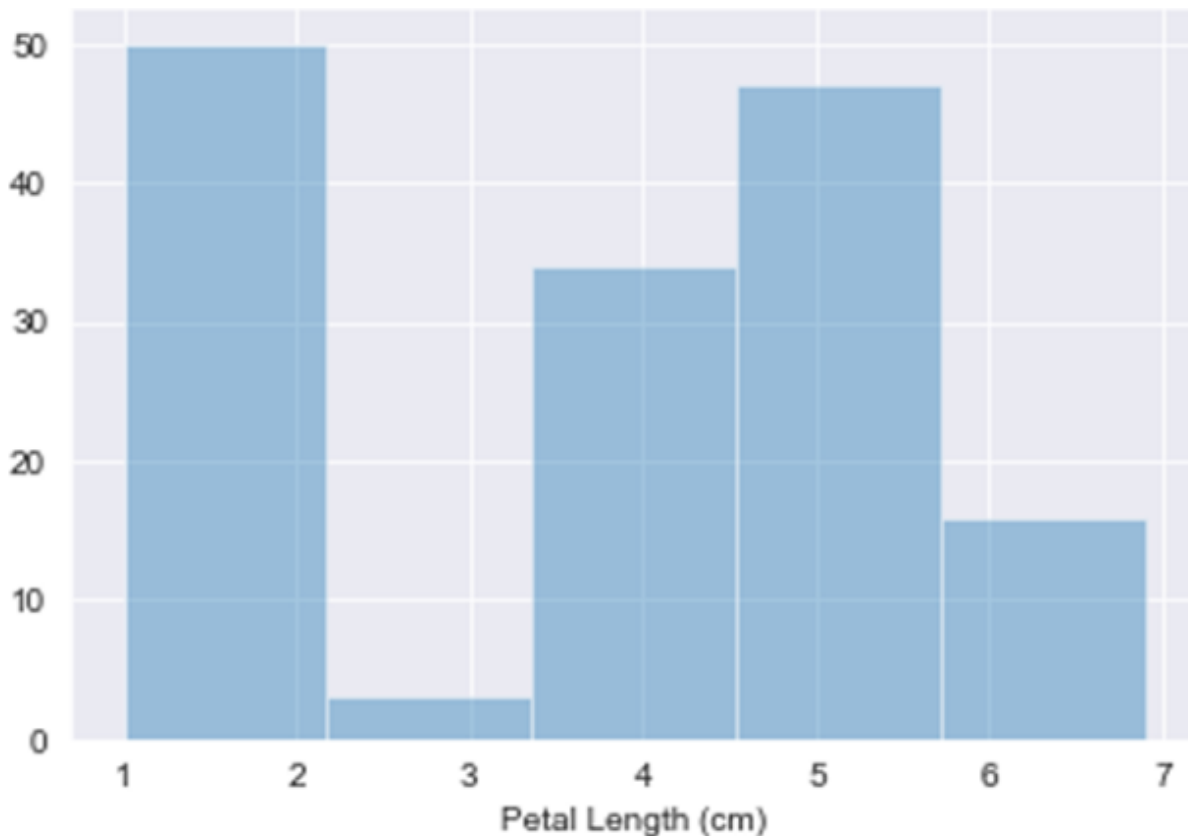
常用可视化图表

该课程主要为大家讲授如下内容：

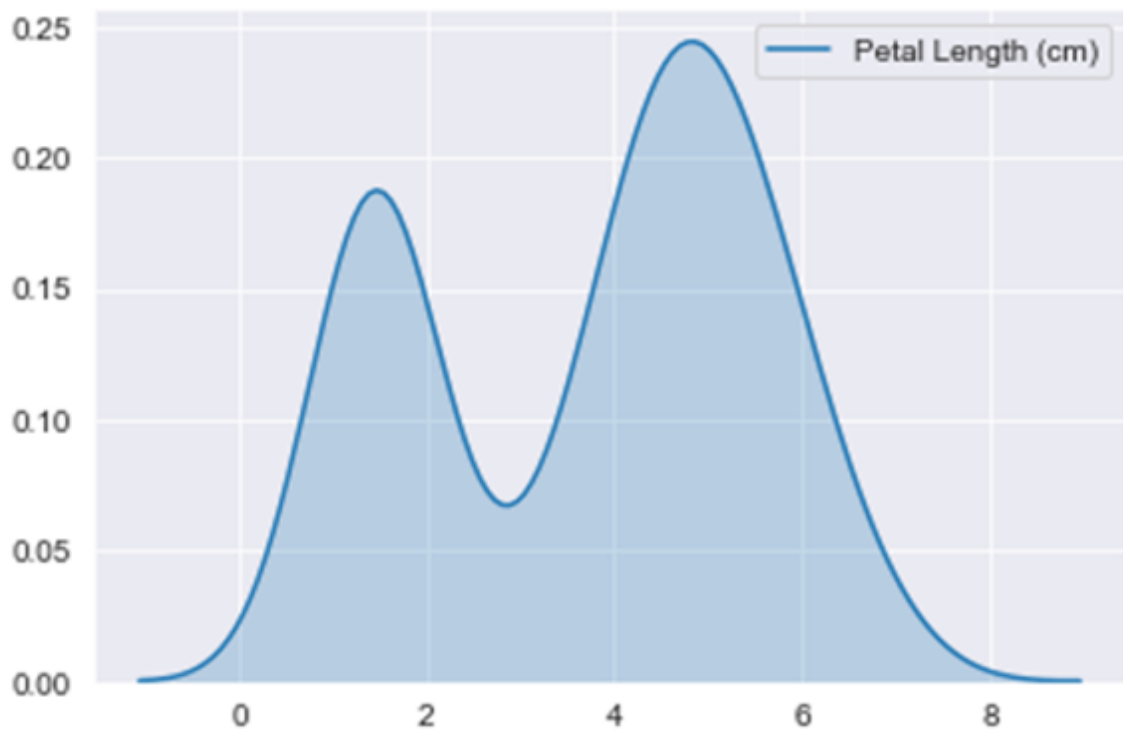
- 单变量图表
- 多变量图表
- 复合图表

1. 单变量图表

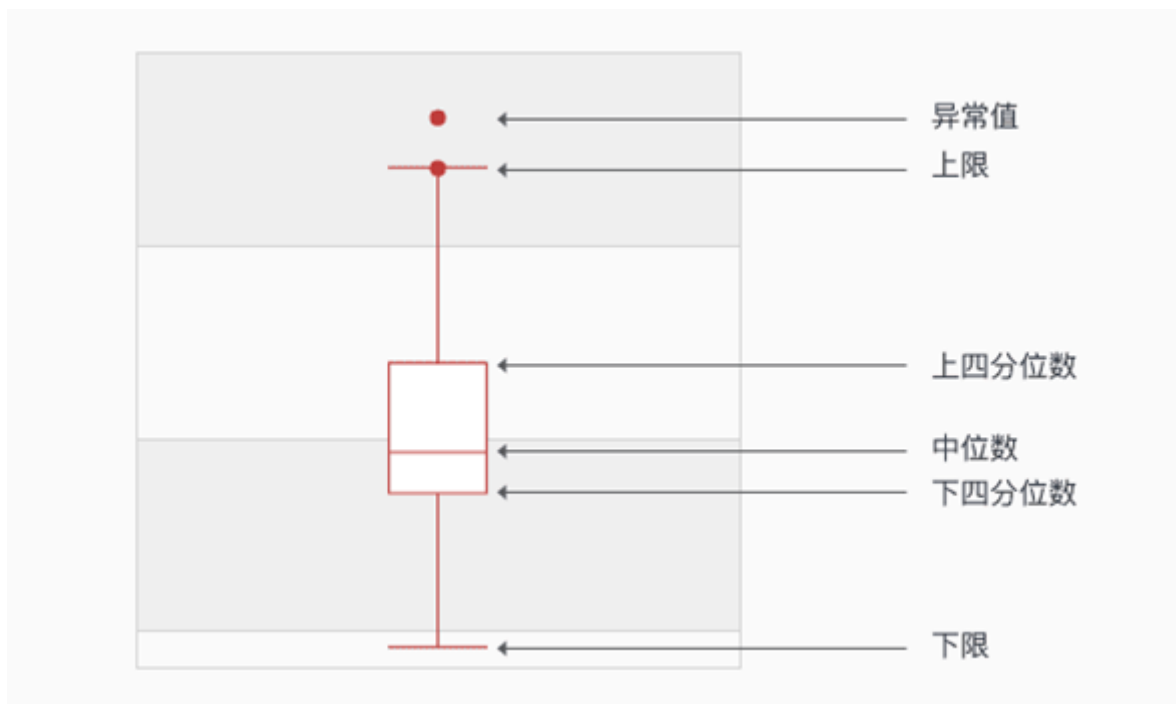
直方图是一种对区间内数据进行频次信息展示的图表。直方图的每一个个矩形代表参数在这个区间内出现的频次。



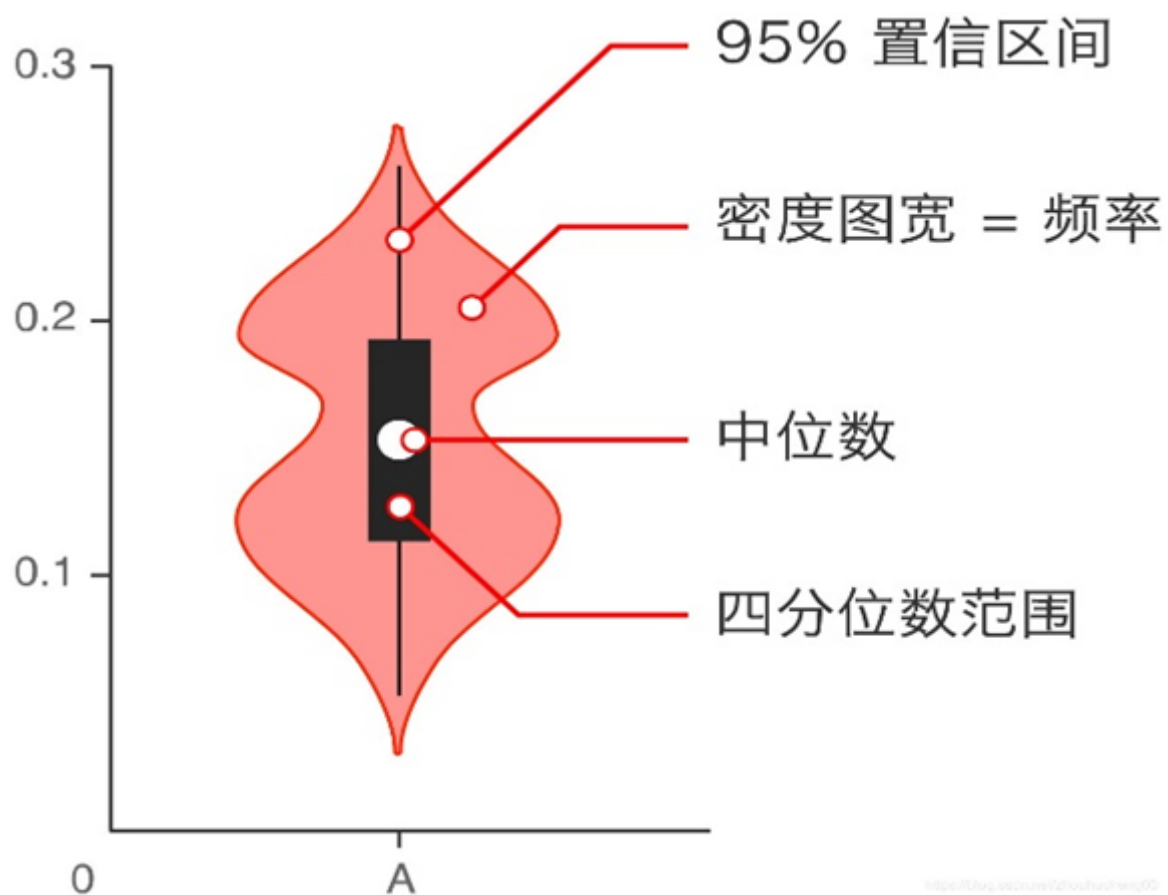
另一种单变量图表是KDE图，KDE全称Kernal Density estimation，也就是核密度估计。KDE图也是用于展示一维特征分布信息的图表，但与直方图不同，它是一条平滑的曲线，kde曲线的平滑程度可以通过参数进行调节。



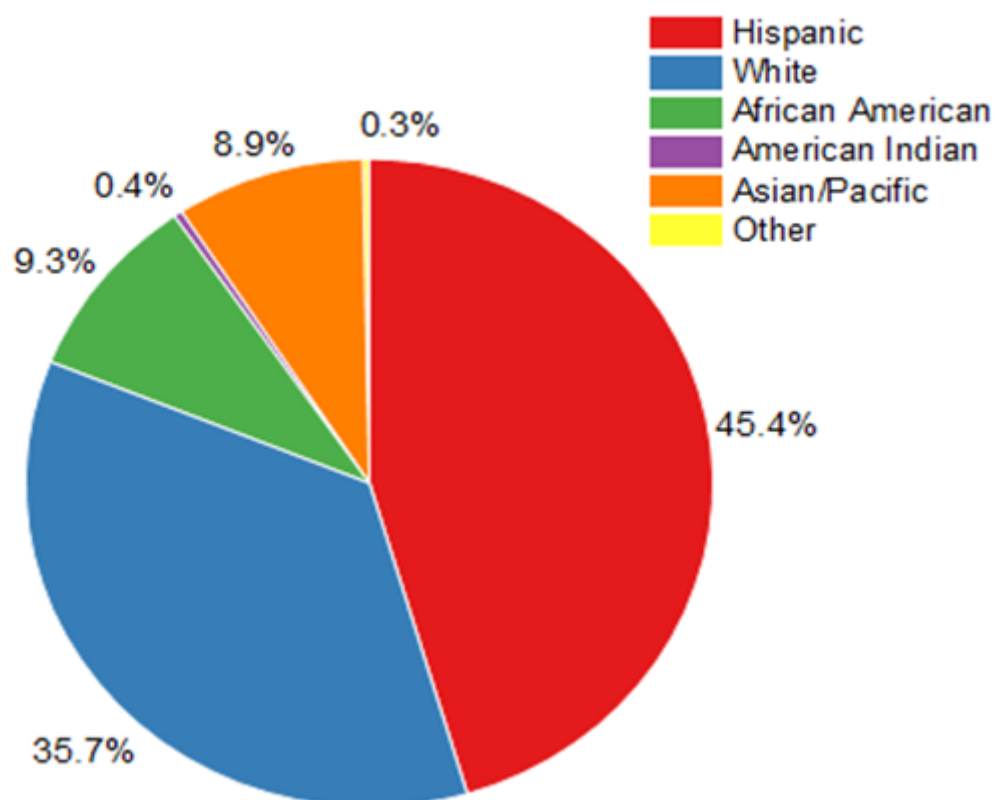
箱型图，也叫做盒型图，英文是 box plot，它是一种用来显示一组数据分散情况的统计图。



小提琴图 and 箱型图类似，但它还包含了纵向的kde曲线来展示特征的分布。实际使用中小提琴图的用量较少，主要原因还是小提琴图包含了很多信息，比较复杂。

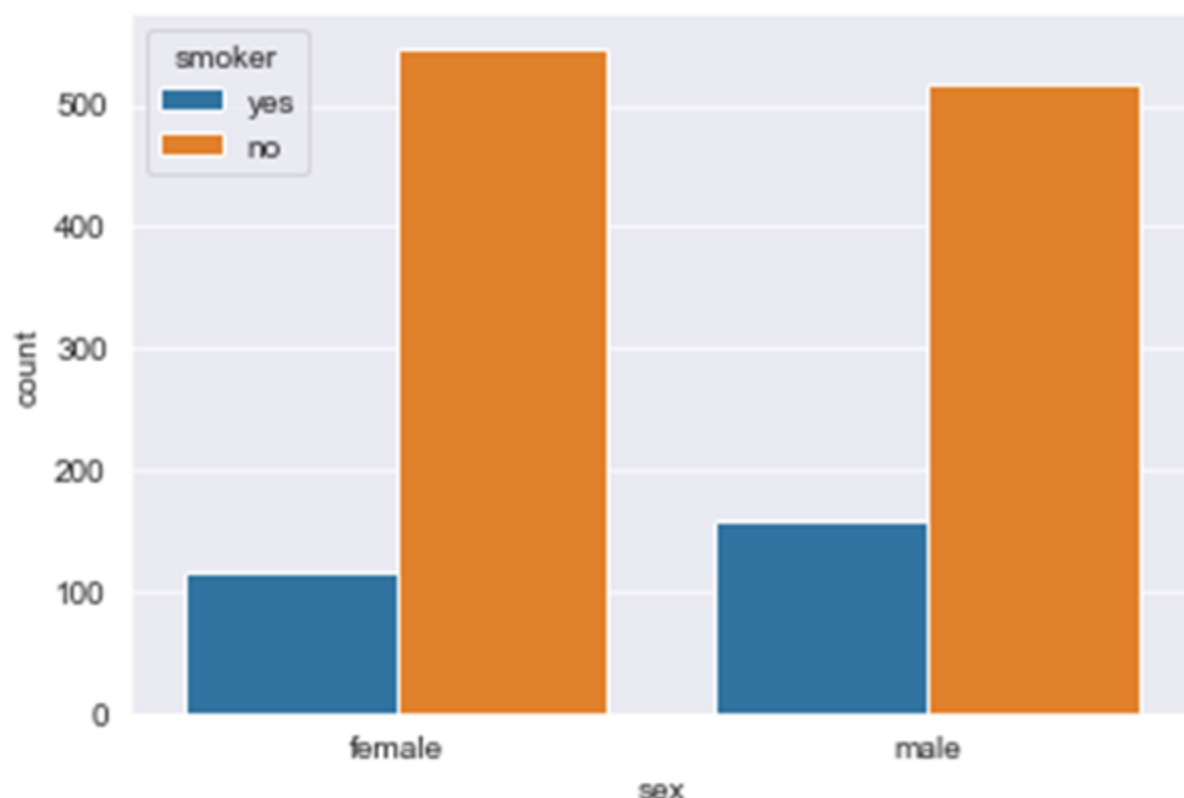


最后一个单变量图表是饼图，饼图通常用来展示分类数据的占比情况。饼图是一种应该避免使用的图表，因为人眼对面积大小并不特别敏感。

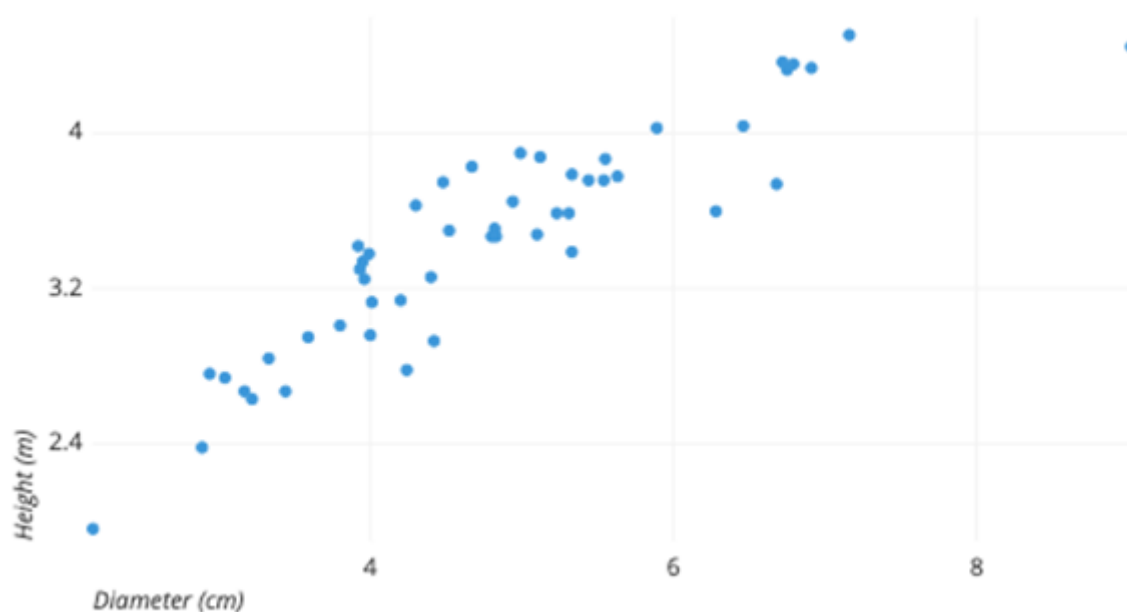


2. 多变量图表

条形图，从名字可以知道只要使用矩形绘制的图表都可以称为条形图，前面介绍的直方图也是一种条形图。

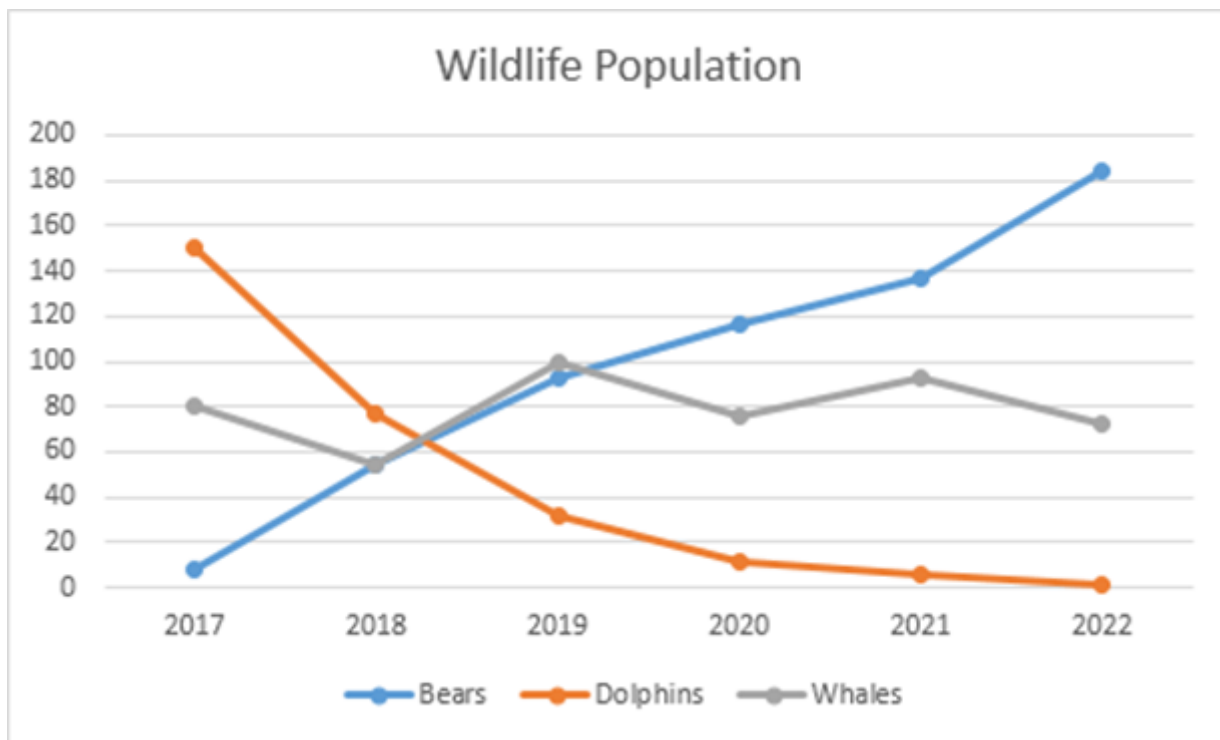


散点图是非常常用的一种图表，可以很好的可视化两个连续变量之间的关系，从而发现某种相关性，此外它对于查找异常值或理解数据分布也很有效。



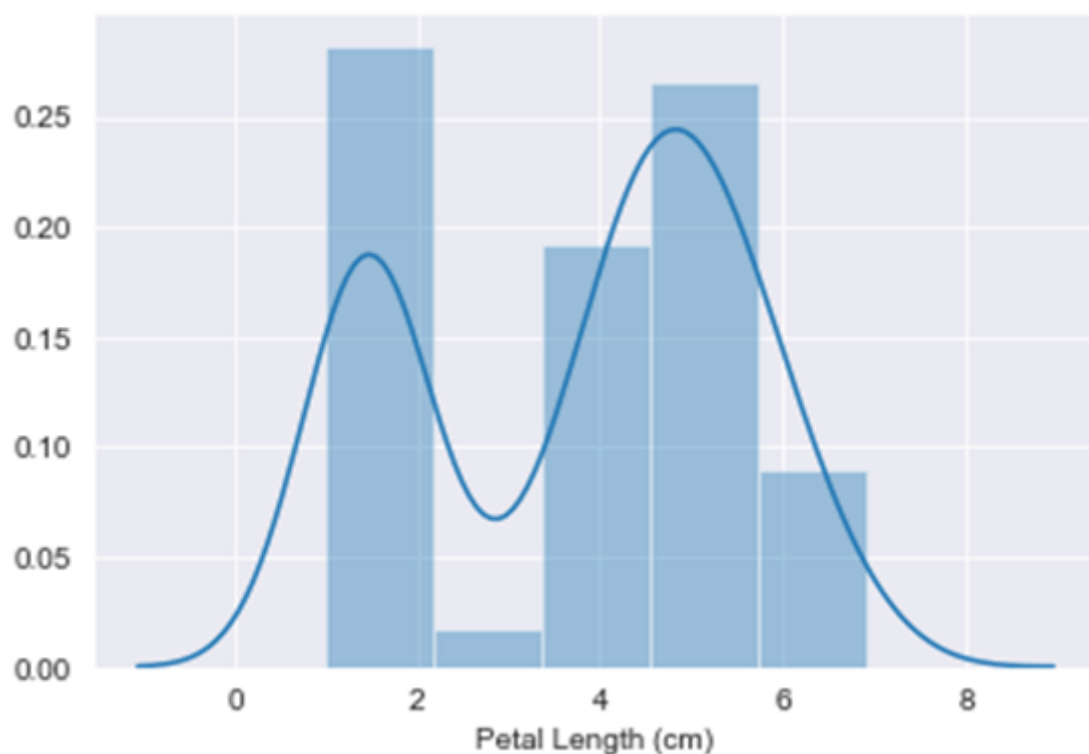
折线图主要是用来展示一维数据随着另一维数据的变化。X轴通常是连

续数据，最常见的是时间。



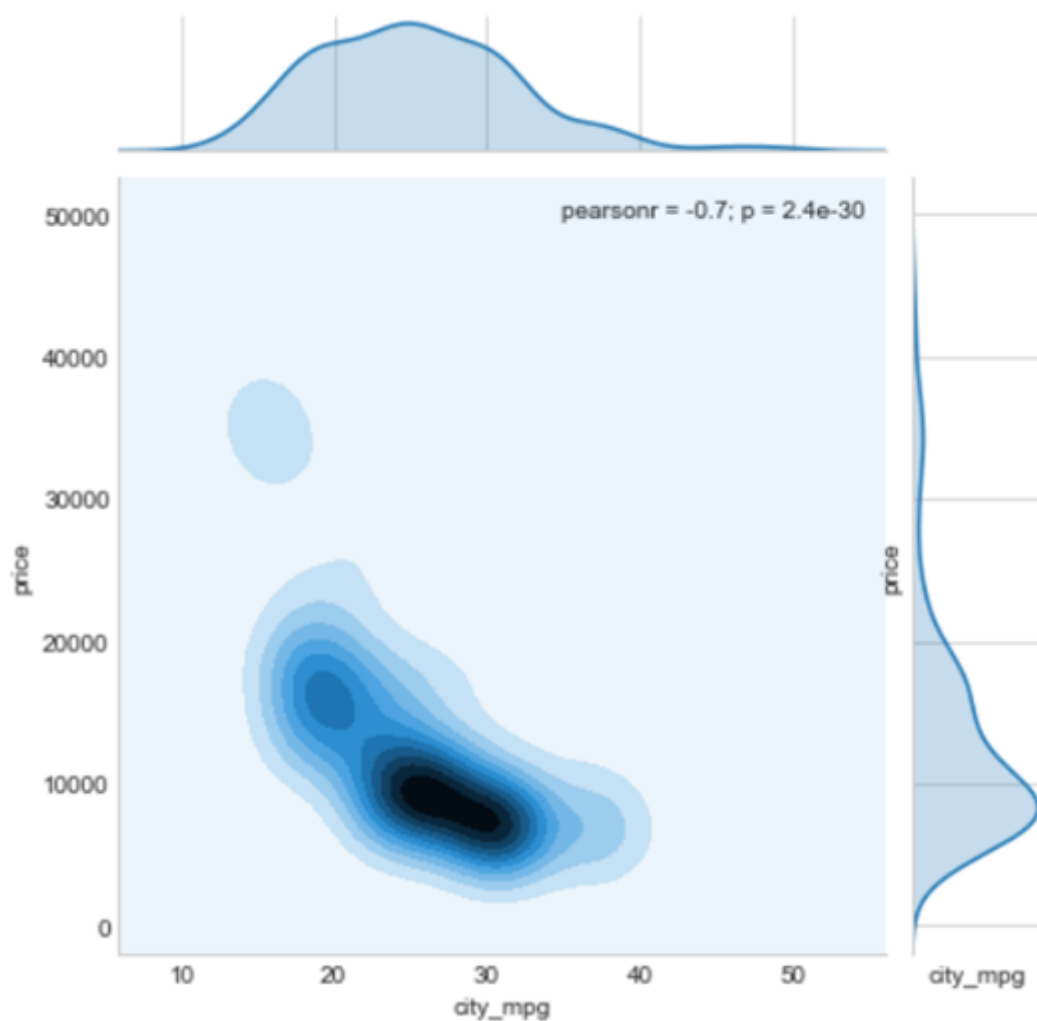
3. 复合图表

第一种复合图表是直方图和KDE图的组合，从当前的图表我们既可以看到在每一个区间内样本的频次也可以看得到样本的变化的趋势。



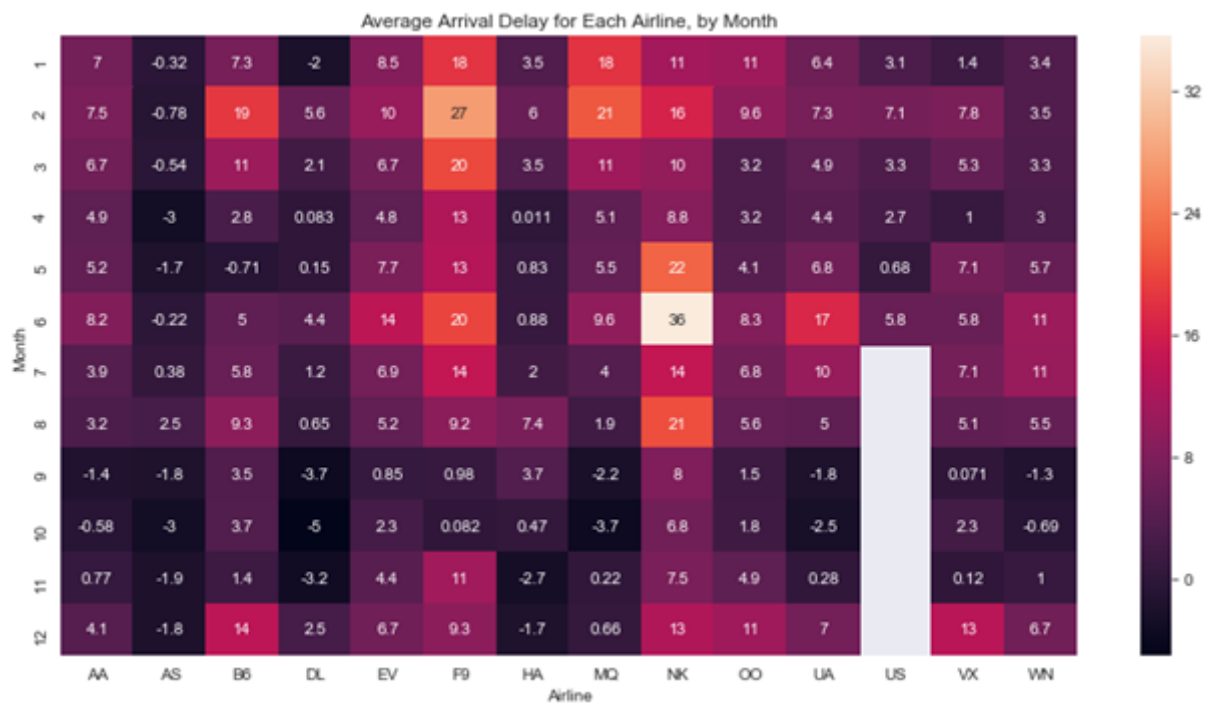
下图是另外一种联合图表，它可以绘制一个二维图表和两个一维图

表。



4. 高维数据可视化

热力图主要是用于三维数据可视化，这种图表还是比较常用的,它可以很好的表示和地理位置相关的一些信息，比如说天气预报或者是道路拥堵情况等等。



平行坐标图可以可视化任意多的特征，它把特征都罗列在X轴，纵轴代表每一个特征的这个取值。

