

数据可视化——图

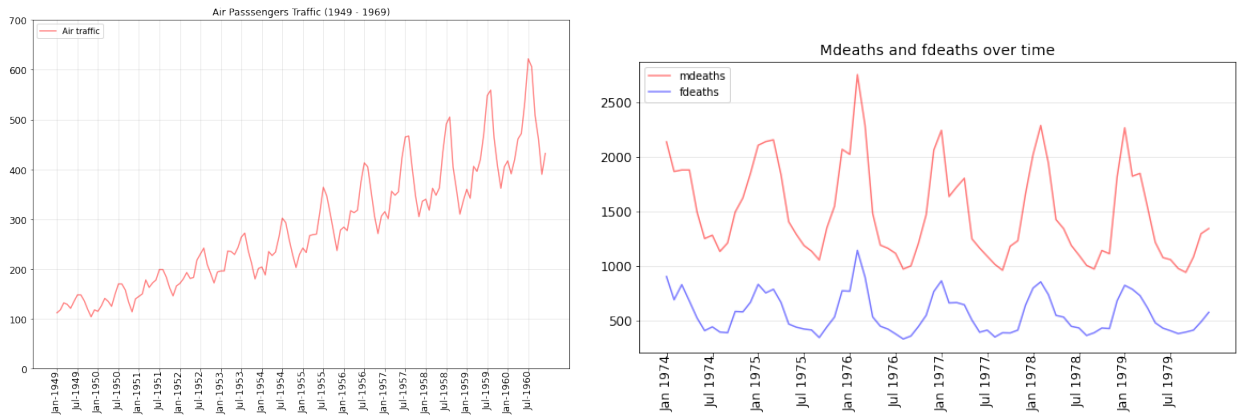
武汉理工大学数学建模协会
阮滨

目录

1 折线图	2
2 条形图	2
3 柱状图	3
4 直方图	3
5 堆叠图	4
6 饼图	4
7 箱型图	5
8 小提琴图	5
9 热力图	6
10 散点图	7
11 雷达图	7
12 山脊图	8
13 矩形树图	8
14 词云	9
15 桑基图	9
16 时序面积堆	10
17 华夫图	10

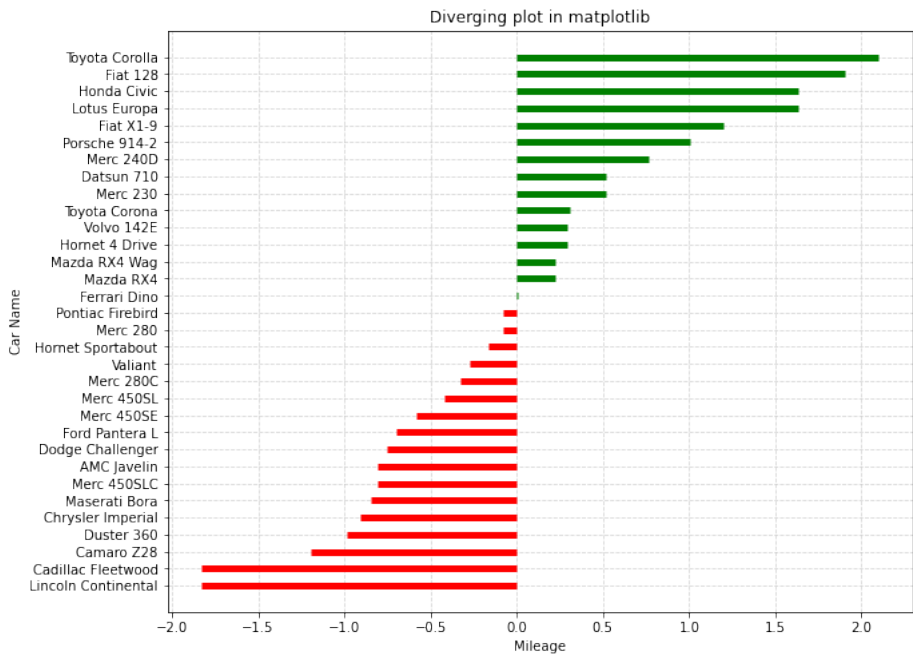
折线图

折线图利用折线，多用于表现时序数据（随时间变化）。每一个数据点表示一个状态的数据，通过直线将点连起来则得到折线图，利用折线图可以直观地看出数据的变化趋势。同时，在同一张图上绘制多条折线可以同时展示不同的发展趋势且可以进行对比。



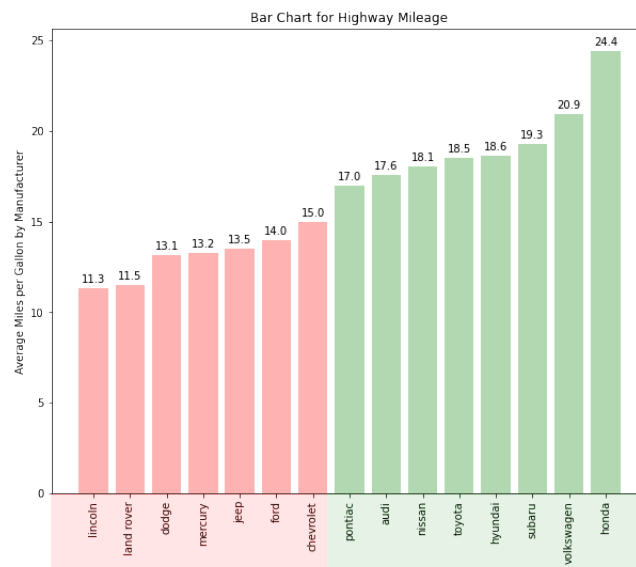
条形图

条形图将数据用矩形表示，通过长度表示数据的大小。条形图多用于对比，所以条形图需要提供不止一种信息，比如进行不同商品之间的售价需要提供多个商品的数据，通过比较长度可以直观地看出数据之间的差异。一般的条形图足以看出数据之间的大小比较，如果想一眼得出比较之后的顺序关系也可以先按大小进行排序之后再作出条形图。



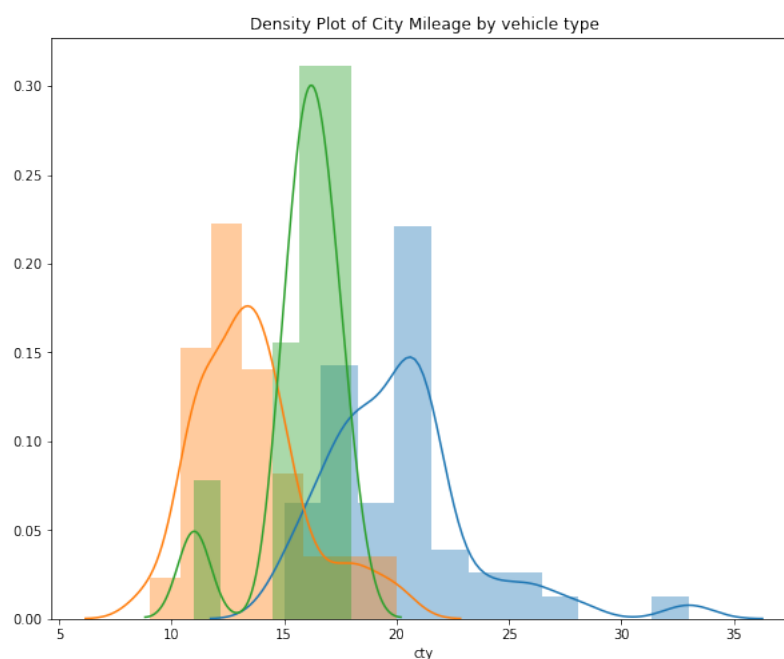
柱状图

柱形图可以视作垂直方向的条形图，功能和条形图类似。使用条形图的情况多可以使用柱状图替代，若 x 轴刻度值过长导致图片文字重叠时，可使用条形图表示（令 x 轴旋转 90° 也可以解决重叠问题，但是不方便图形查看，如下图所示）。



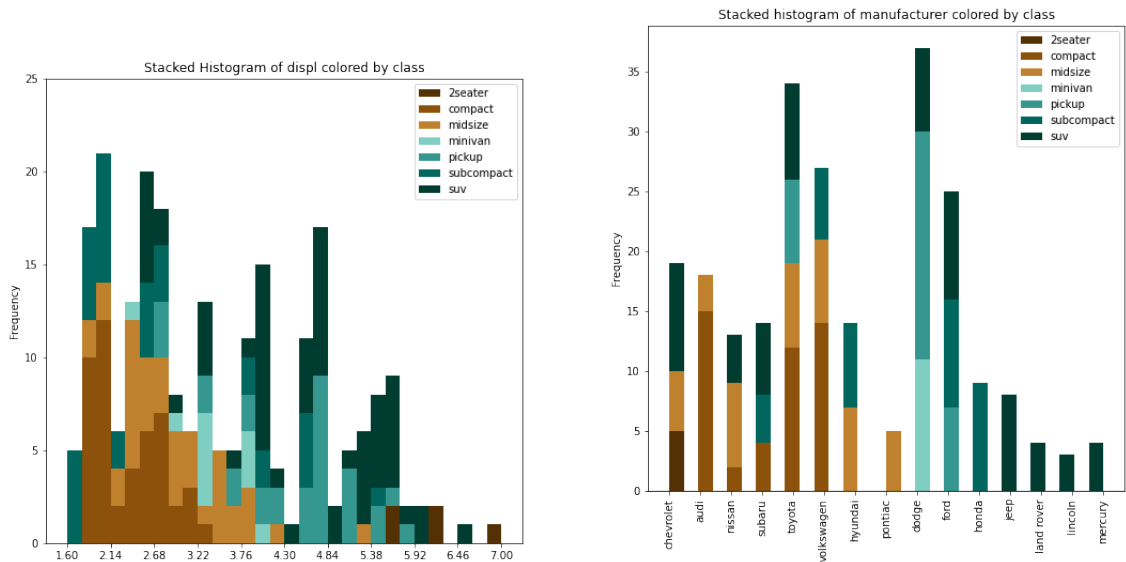
直方图

直方图也是将数据用矩形进行表示，而直方图多用于表示数据的分布。对于直方图而言，数据多具有连续性，因而直方图的矩阵多为连续排序。利用直方图展示数据分布时，也可以拟合出概率密度曲线添加图片信息。



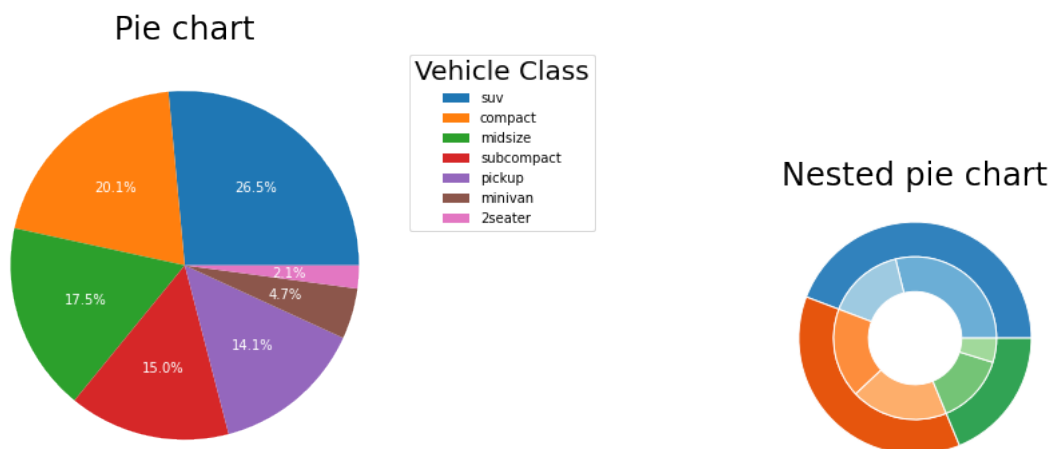
堆叠图

堆叠图在直方图和柱状图的基础上进行改进，不仅能表示某一个维度数据中不同类型的数据的差异，同时也可以比较类别总数的差距。



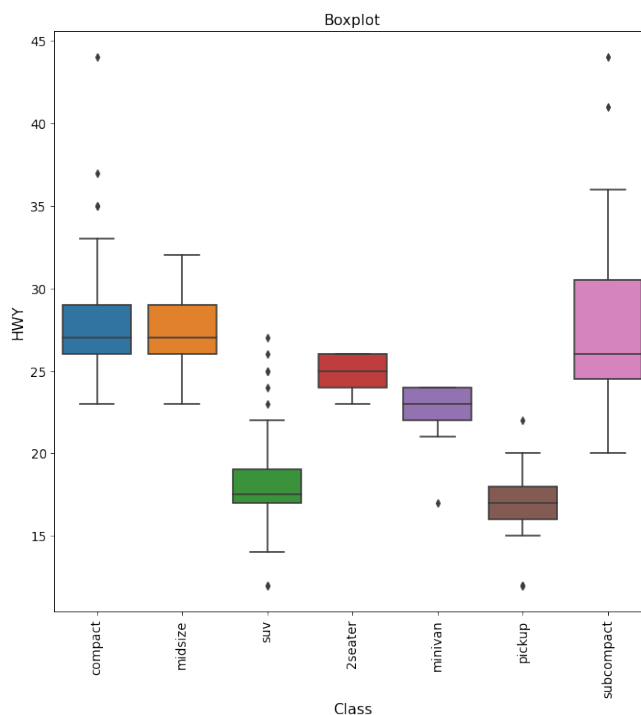
饼图

饼图多用于展示数据的分布占比，饼图每一部分的大小就代表某一类别数据占总量的多少。除了基本的饼图之外，还可以做嵌套饼图，除了显示每个大类别的占比情况，也可以显示每个大类别下的小类别的占比情况。



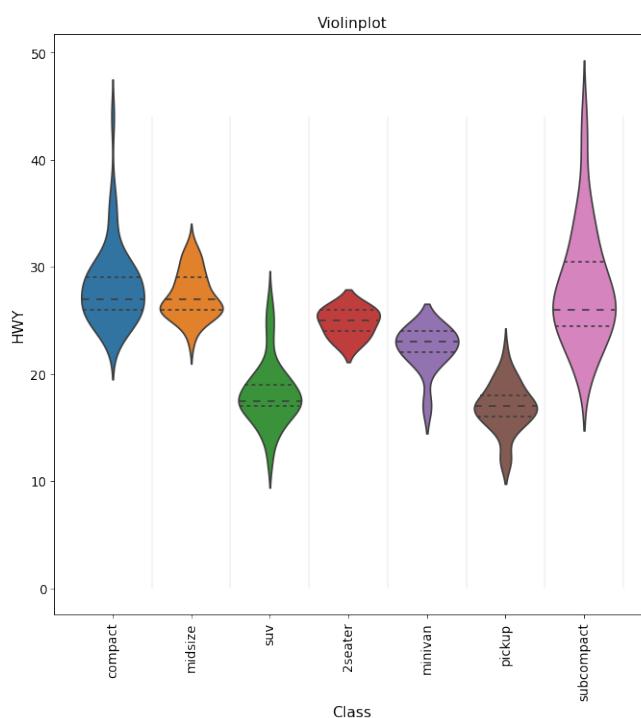
箱型图

箱型图用来表示数据的分布情况与部分统计数据。部分统计数据包括数据的上下四分位数、最大值、最小值和中位数。对于箱型图而言，其具有六个重要数据点，即上文提到的五个数据点，第六个数据点为异常点。在一张图上可以同时做出多种数据的箱型图进行数据分布的比较。

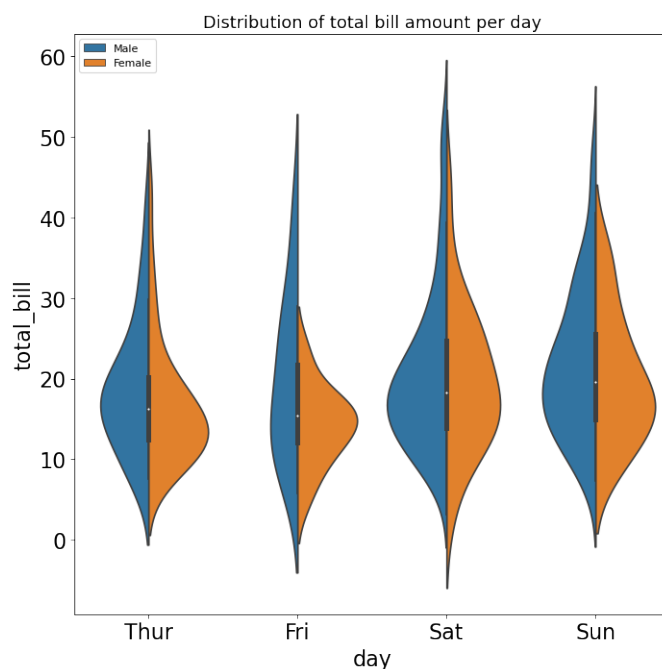


小提琴图

小提琴图用来展示数据的分布状态与概率密度，该图标同时结合箱型图和密度图的特征。在表现不同种类数据的分布时，可以将多个小提琴绘制同一张图上。



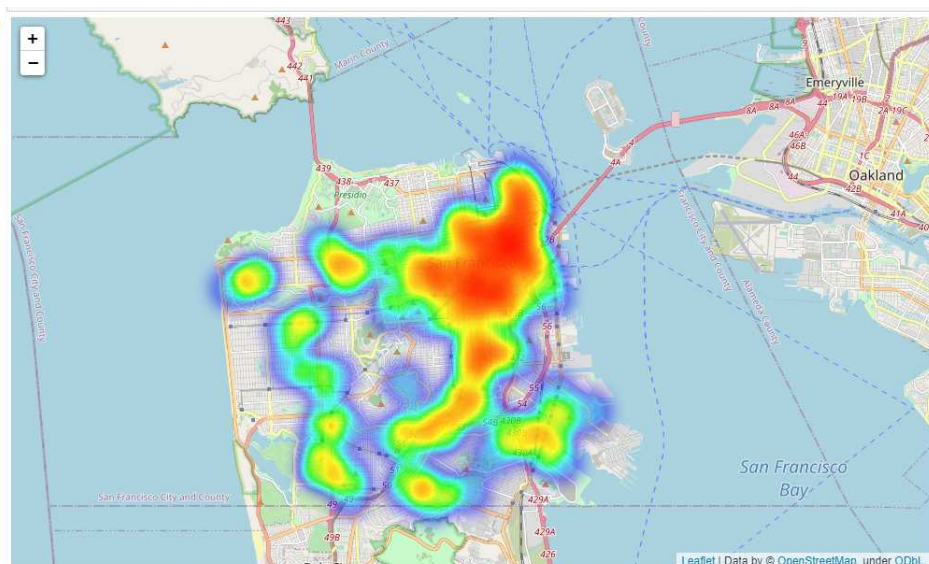
同时，考虑到小提琴图的对称性（图形信息重复），可以将小提琴图的每一半设置为不同种类的数据在第三个维度上进行对比，如男性和女性，如下图所示。



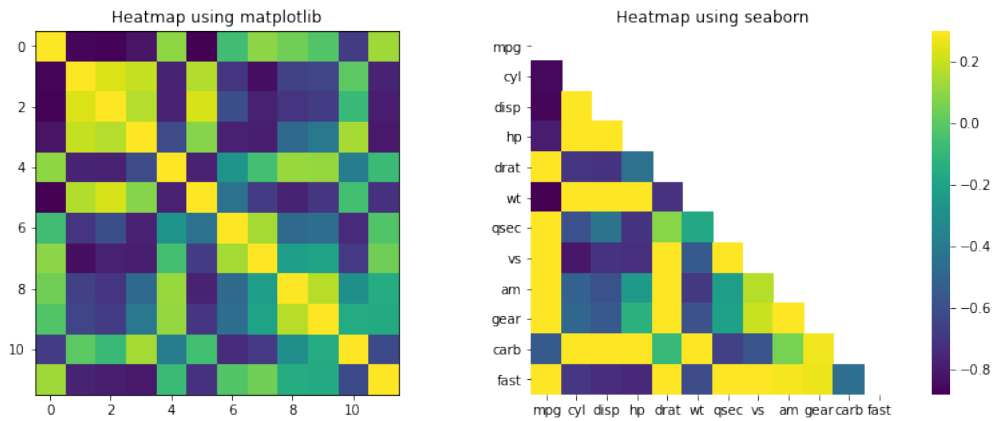
热力图

热力图通过可视化密度在图中直观地表示出点的密集程度或者数据的大小。经常应用在地图的数据可视化或者系数矩阵的可视化。

利用热力图在地图中的可视化可以清楚地表示出地图上的火势大小，城市的犯罪密度（MCM2010B-犯罪学问题）等等。

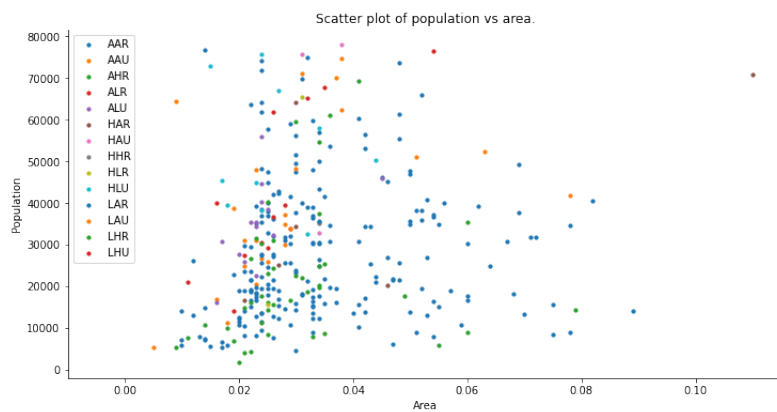


此外，对于系数矩阵，同样可以用热力图表示不同块的取值，从直观上获取大小信息（左下图）。对于相关系数矩阵而言，矩阵具有对称性，因而会导致图形信息具有冗余，所以在作图时可以屏蔽一半的信息（右下图）。



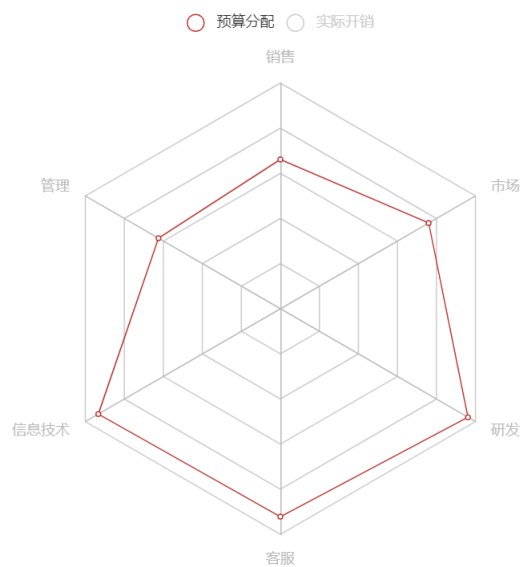
散点图

散点图，此处表示二维散点图（无特殊说明，本文所有图形默认为二维图形）。散点图多用于表示二维数据的分布情况。可对不同种类的点利用不同的颜色进行区分。



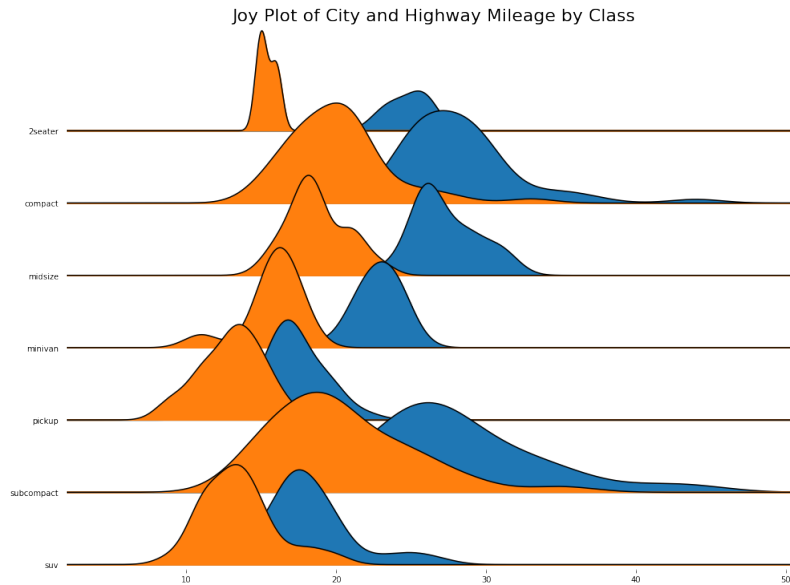
雷达图

当需要进行多变量之间的指标评价时，普通的图无法满足要求，此时可以使用雷达图。雷达图给予多个维度（多变量）之间的指标展示。



山脊图

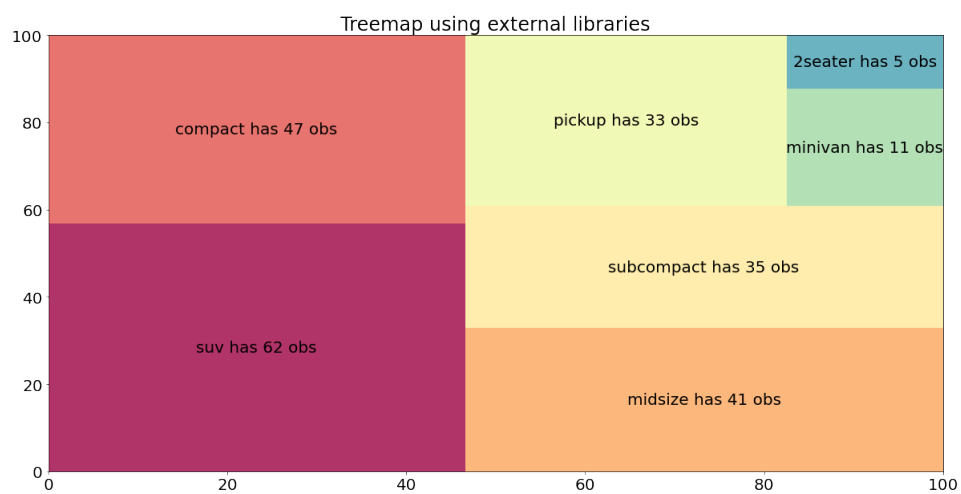
山脊图用来显示数据的分布密度，功能和直方图类似，在图形显示上表现为多座山的形式因此得名为山脊图，具有较好的美观性。



矩形树图

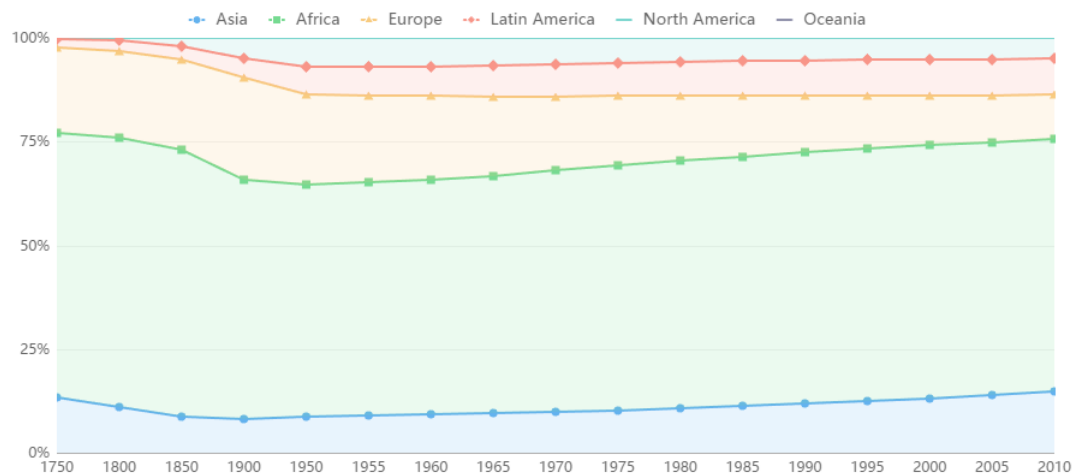
矩形树图将层次结构（树状结构）的数据显示为一组嵌套矩形。树的每个分支都有一个矩形，然后用较小矩形代表每个分支的子分支并平铺在其中。

叶子节点的矩形面积与数据占比成比例。通常叶节点会使用不同的颜色来显示数据的关联维度。



时序面积堆

若要展示某类数据占额时序数据可以使用时序面积堆，该图相比于饼图引入了时序信息，同时也可以表示各类数据的占比。



华夫图

华夫图具有两种功能。一种为显示占比类似饼图（下图一）表示类别的占比，另一种引入时序信息类似于 GitHub 的活动展示图（下图二），这种图有时候也称为日历图，主要表现每天的活跃度。

