



专注于商业智能BI和大数据的垂直社区平台

# 常见图形的绘制

Allen

[www.hellobi.com](http://www.hellobi.com)

## 课程目录

- 直方图
- 茎叶图
- 箱线图
- 五数总括
- 小结

## 直方图

- 直方图：由一系列高度不等的纵向条纹或线段表示数据分布的情况，一般用横轴表示数据类型，纵轴表示分布情况
- 将数据的取值范围分成若干区间，每个区间长度称为组距
- 组距对直方图的形态有影响，组距太小，每组的频数较少，组距太大，每组的频数太多，直方图反映的形态就不灵敏

# 直方图

- R语言：hist()画直方图
- 参数：
  - freq：是否绘制频率直方图
  - breaks：规定直方图的组距
  - col：表示直方图中填充的颜色

## 直方图

- R语言：核密度估计函数density()，目的是用已知样本，估计其密度
- 参数：
  - x：样本向量
  - bw：带宽
  - col：表示颜色

## 茎叶图

- 茎叶图：将数组中的数按位数进行比较，将数的大小基本不变或变化不大的位作为一个主干（茎），将变化大的位的数作为分枝（叶）
- 列在主干的后面，这样就可以清楚地看到每个主干后面的几个数，每个数具体是多少
- 对比直方图，能够更细致的看出数据分布的结构

## 茎叶图

- R语言：stem()画茎叶图
- 参数：
  - scale：控制绘制茎叶图的长度
  - width：绘图的宽度

## 箱线图

- 箱线图：用作显示一组数据分散情况资料的统计图
- 能提供有关数据位置和分散情况的关键信息，尤其在比较不同的母体数据时更可表现其差异
- 提供了中位数、Q1、Q3、最大值和最小值等五个数据



## 箱线图

- R语言：boxplot()画箱线图
- 参数：
  - formula：公式
  - range：触须范围
  - col：表示颜色

## 五数总括

- 反映数据特征：中位数，下四分位数，上四分位数，最小值，最大值
- R语言：使用fivenum()函数

## 小结

- 直方图
- 茎叶图
- 箱线图
- 五数总括
- 小结