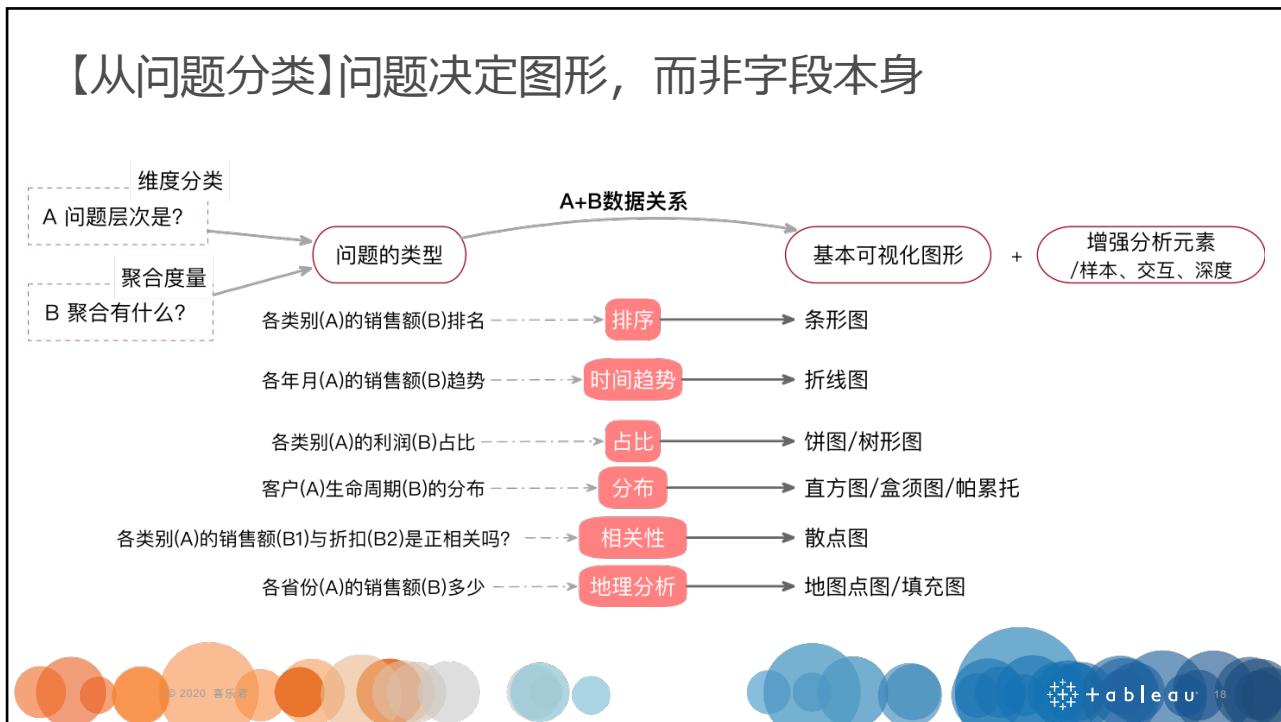
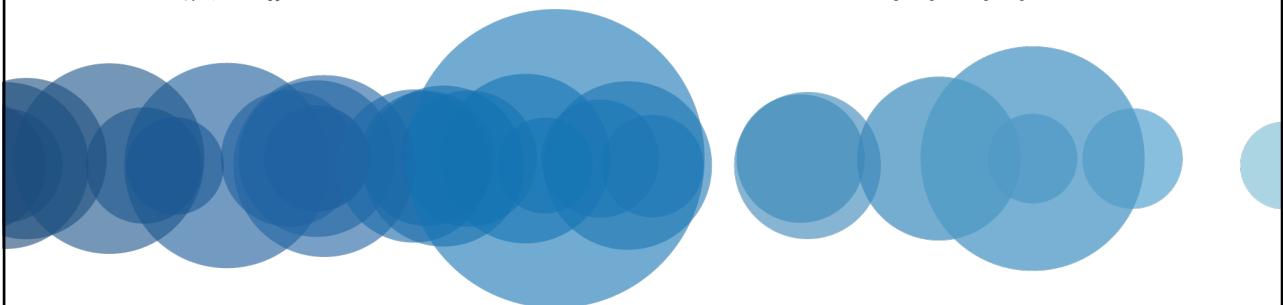


1



18

Part 3  
从问题到图形（下）  
中级可视化：相关性、分布、交叉表、地图与图形增强



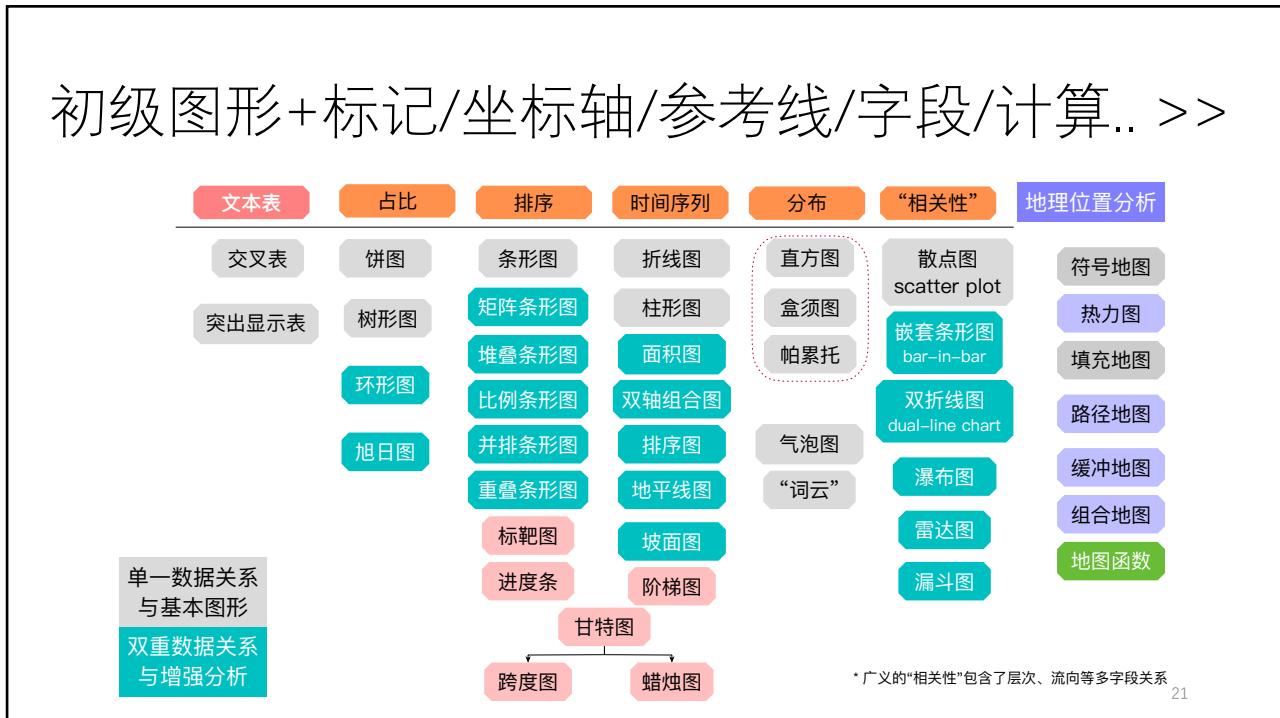
19

- **一、分布分析与相关性/初级展示**
  - 区间分布/直方图      离散分布/盒须图                    集中分布/帕累托
  - 相关性分析（散点图与双轴）
  - 地理空间分析
- **二、聚合方式与层次分析（重新讲解）**
  - 聚合度：衡量层次的尺度（理解层次）
  - 聚合方法：总和、均值/方差、标准差/百分位
  - 坐标轴的使用
  - 参考线的使用
- **三、样本控制功能**

20

20

## 初级图形+标记/坐标轴/参考线/字段/计算.. >>



21

## 区间分布：直方图

使用数据桶创建直方图 | 直接把度量转维度做直方图



22

## 离散分布：盒须图/箱线图

大数据三大图之盒须图

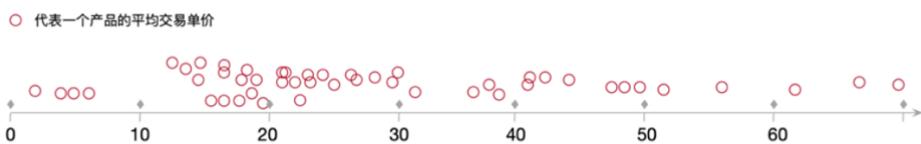


图 5-29 盒须图是对点的分布分析，无须拆分区间

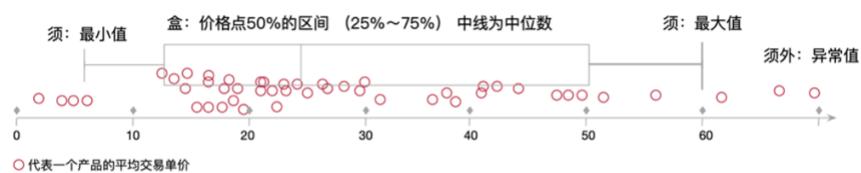


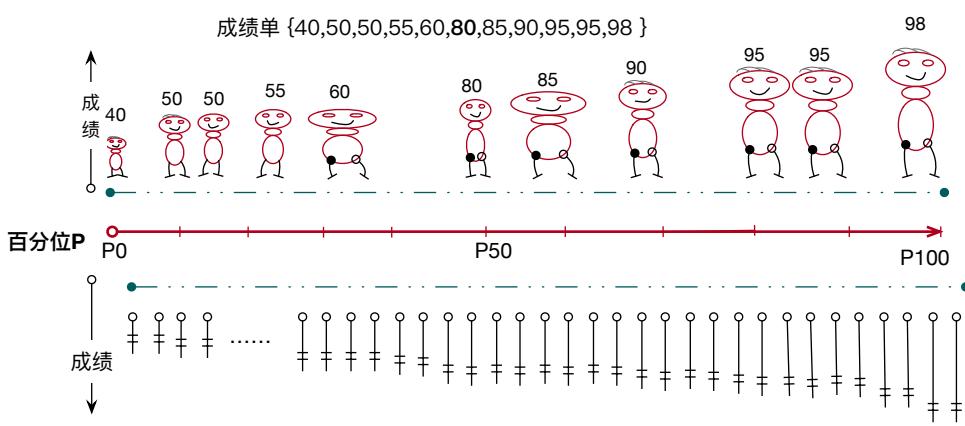
图 5-30 盒须图的可视化原理



23

## 关于聚合：什么是百分位？

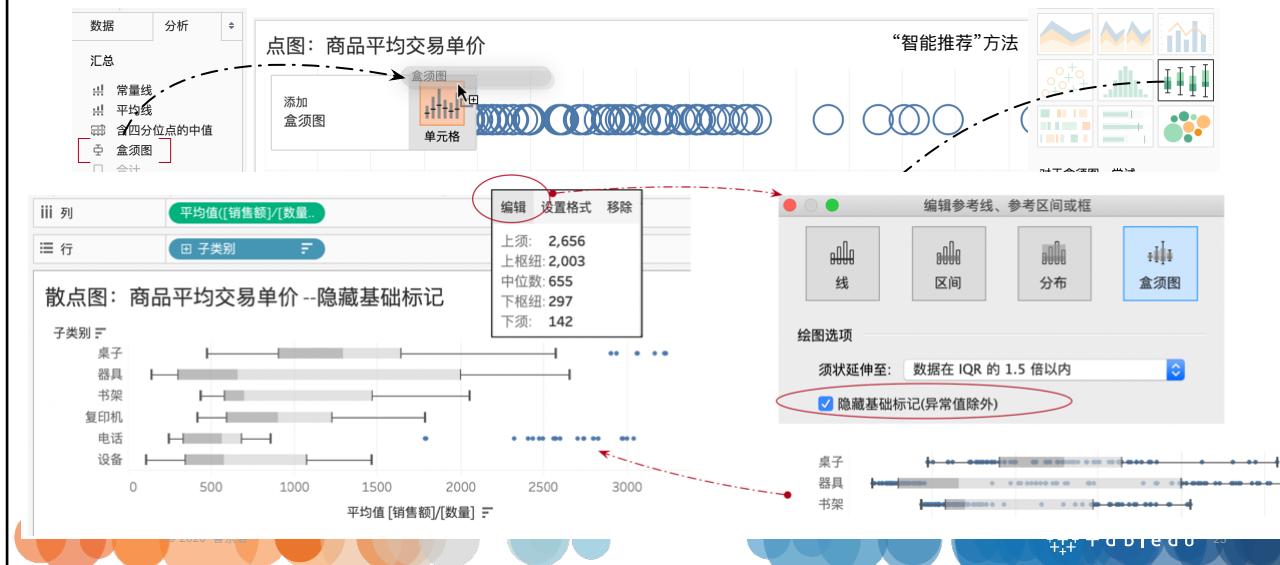
百分位是多个数取一个数的过程，Percentile —— **中位数/最大值/最小值是典型的百分位函数**



24

## 离散分布：盒须图

大数据三大图之盒须图



25

## 百分位与百分比

百分数 %百分比是算数计算，百分位是聚合。

%百分比



马拉松，我到终点了，你在哪？

路标提示，刚刚过“一半”呢 (50%\*X)

P百分位



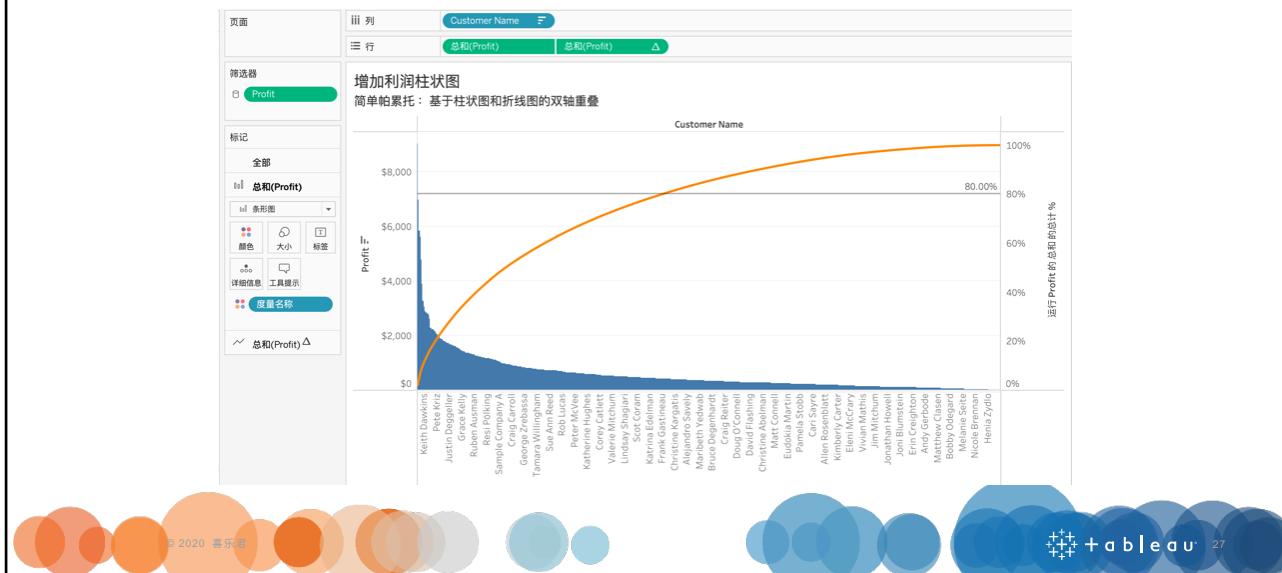
我跑了第1名 (P100)，你呢？

我是第520名... (所有人里) 排60%吧 (P60)

26

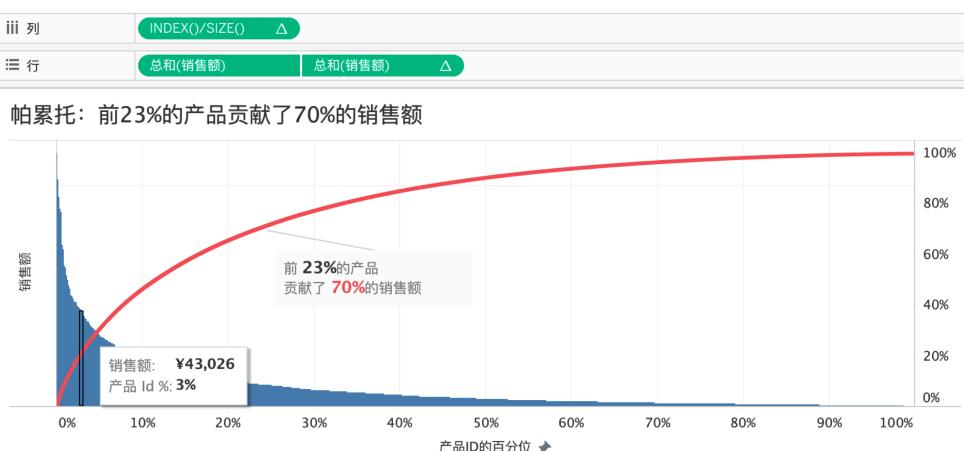
## 集中分布：简单帕累托与高级帕累托 Pareto Chart

### 简单帕累托



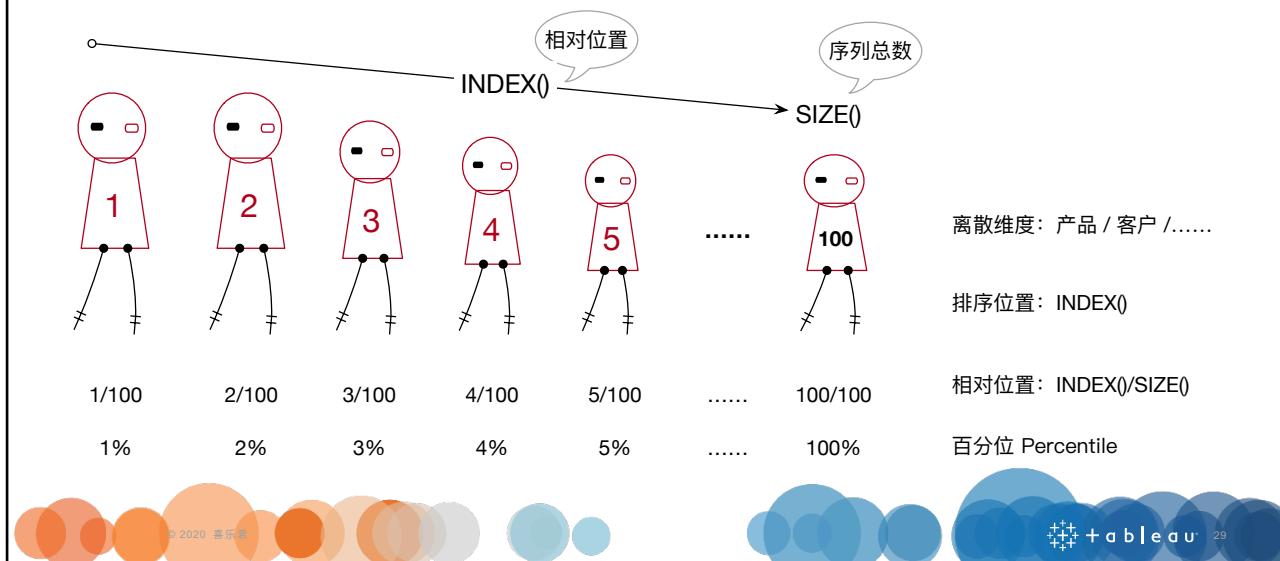
27

## 集中分布：简单帕累托与高级帕累托 Pareto Chart



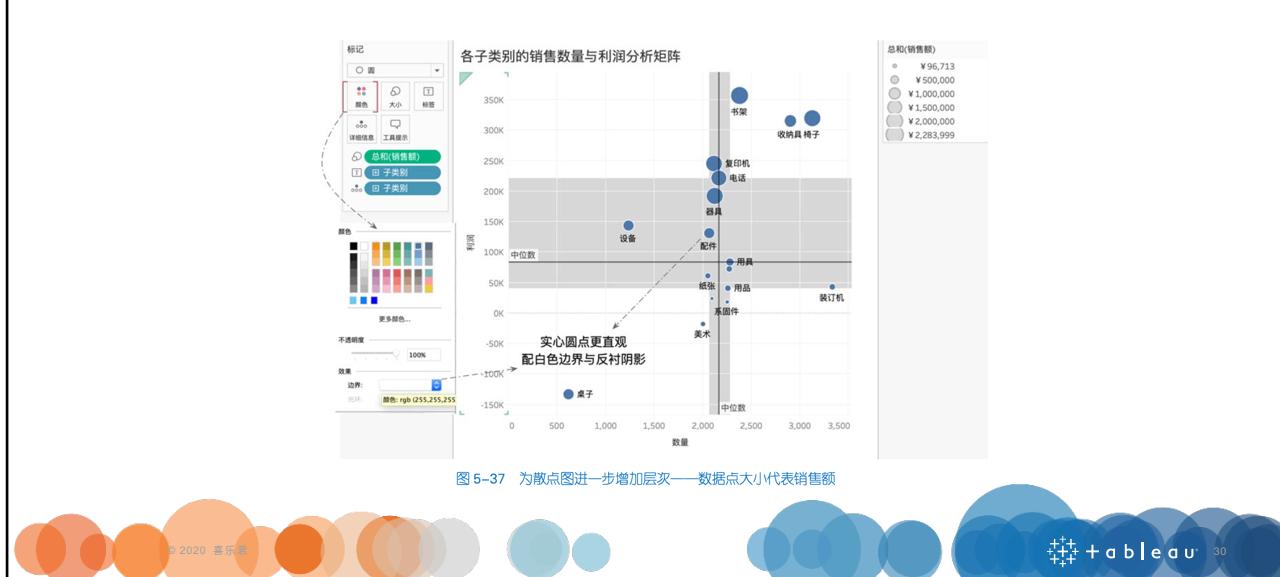
28

## 高级帕累托中需要的百分位计算 —— 把离散转化为连续



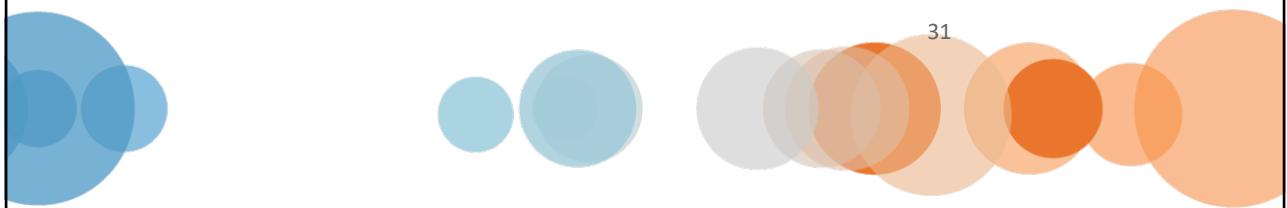
29

## 大数据三大图之散点图



30

## Part 3 (补充) 地理空间可视化



31

### 地图，就是带有背景的XY坐标系

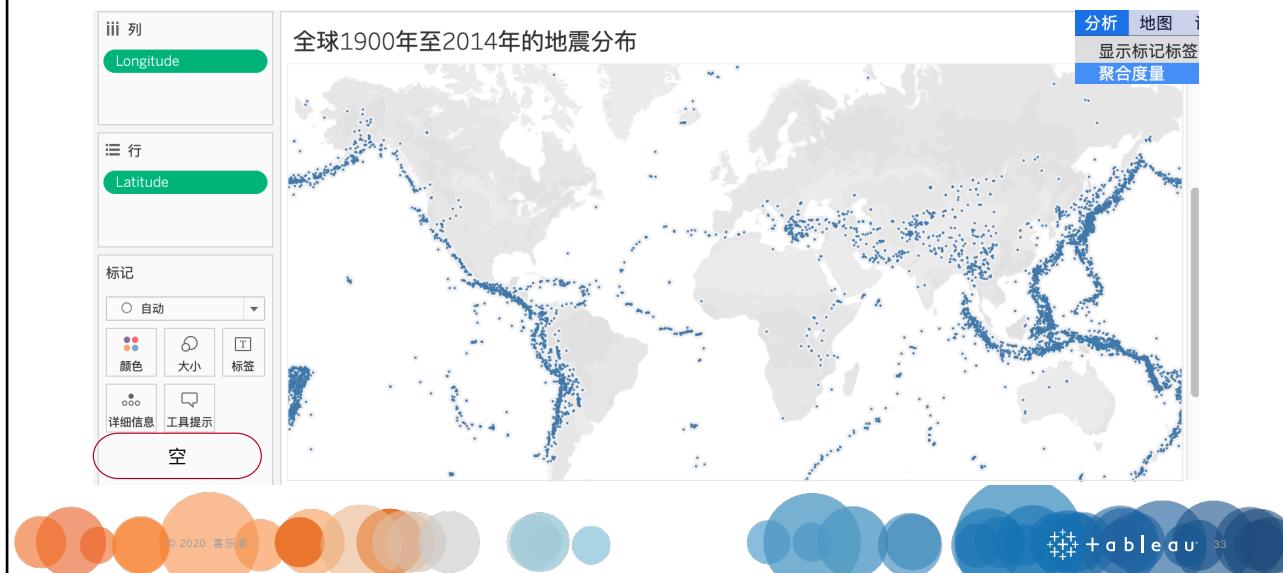
空间信息可视化 = 多种 地图层+ 数据层的合并

地图层
<input checked="" type="checkbox"/> 底图
<input checked="" type="checkbox"/> 地区覆盖
<input type="checkbox"/> 地形
<input type="checkbox"/> 海岸线
<input type="checkbox"/> 街道, 高速公路, 路线
<input type="checkbox"/> 国家 / 地区边界 (浅灰)
<input type="checkbox"/> 国家 / 地区名字 (浅灰)
<input checked="" type="checkbox"/> 国家/地区边界
<input checked="" type="checkbox"/> 国家/地区名称
<input type="checkbox"/> 州/省/市/自治区边界 (浅灰)
<input type="checkbox"/> 州/省/市/自治区名称 (浅灰)
<input checked="" type="checkbox"/> 州/省/市/自治区边界
<input checked="" type="checkbox"/> 州/省/市/自治区名称
<input type="checkbox"/> 县市边界
<input type="checkbox"/> 邮政编码标签
<input type="checkbox"/> 地区代码边界
<input type="checkbox"/> 地区代码标签
<input type="checkbox"/> 美国都市边界 (CBSA)
<input type="checkbox"/> 美国都市名称 (CBSA)
<input type="checkbox"/> 水域标签
<input type="checkbox"/> 城市
<input type="checkbox"/> 景点
<input type="checkbox"/> 小区

32

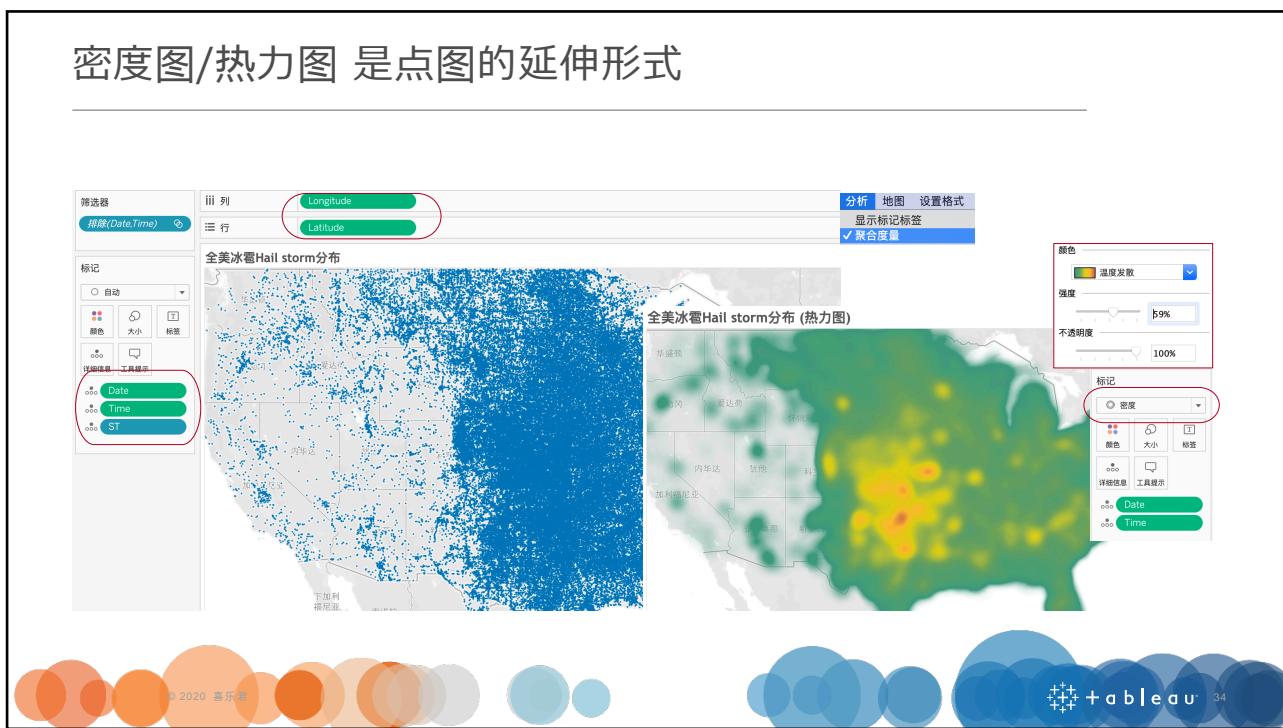
## 最简单的地图：点图 dot map

行级别的数据，在地图上直接分布，没有其他度量值和维度——解聚



33

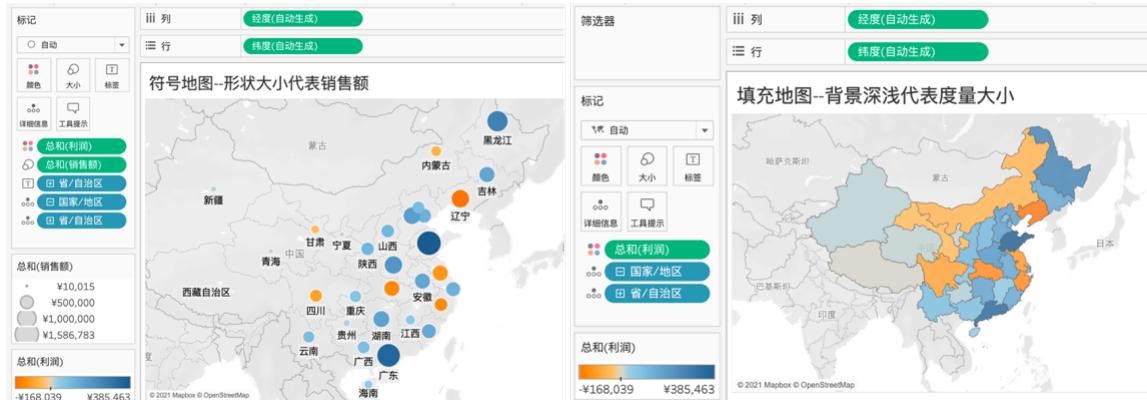
## 密度图/热力图 是点图的延伸形式



34

## 符号地图 VS 背景地图

符号地图可以同时表达两个聚合



35

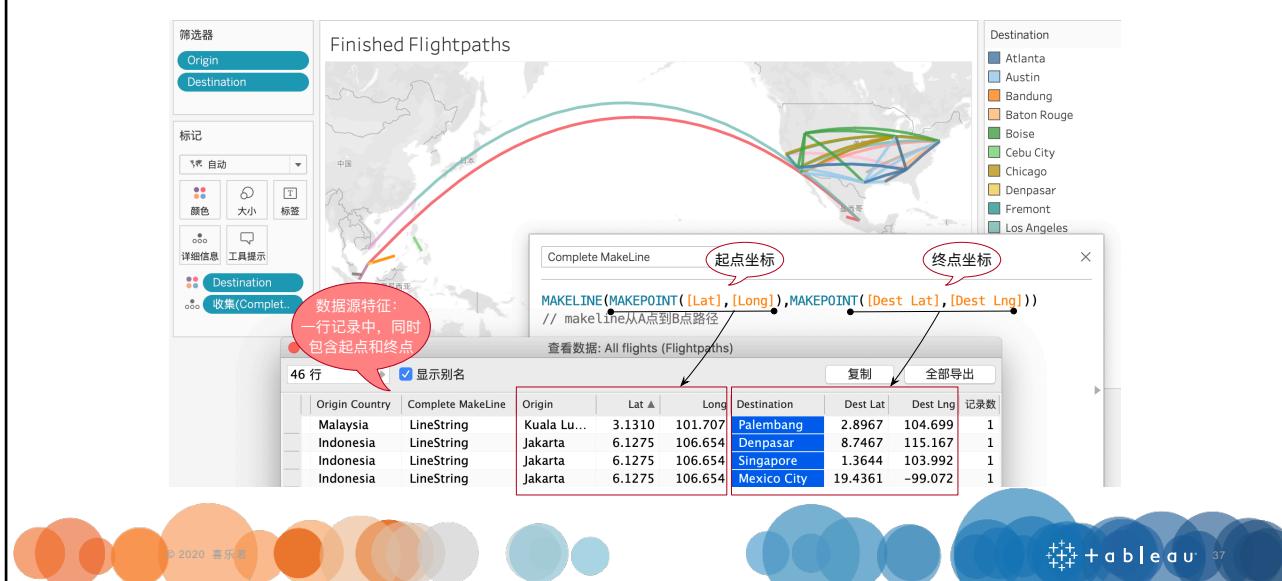
## 路径地图

路径地图，就是点图在地图中的连线——注意，连线的 路径字段 必须是维度



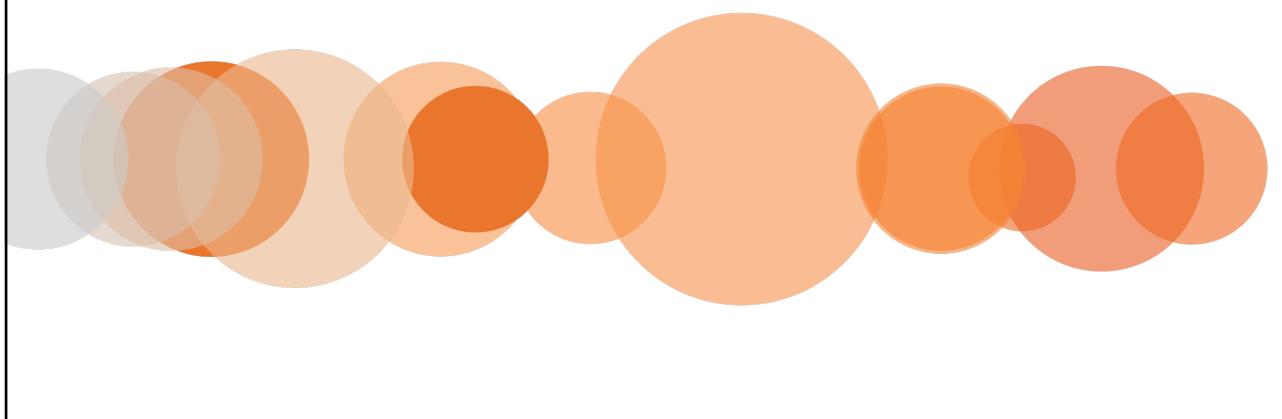
36

## 空间函数：makepoint和makeline函数



37

## 基本交叉表：使用标记和坐标轴改造



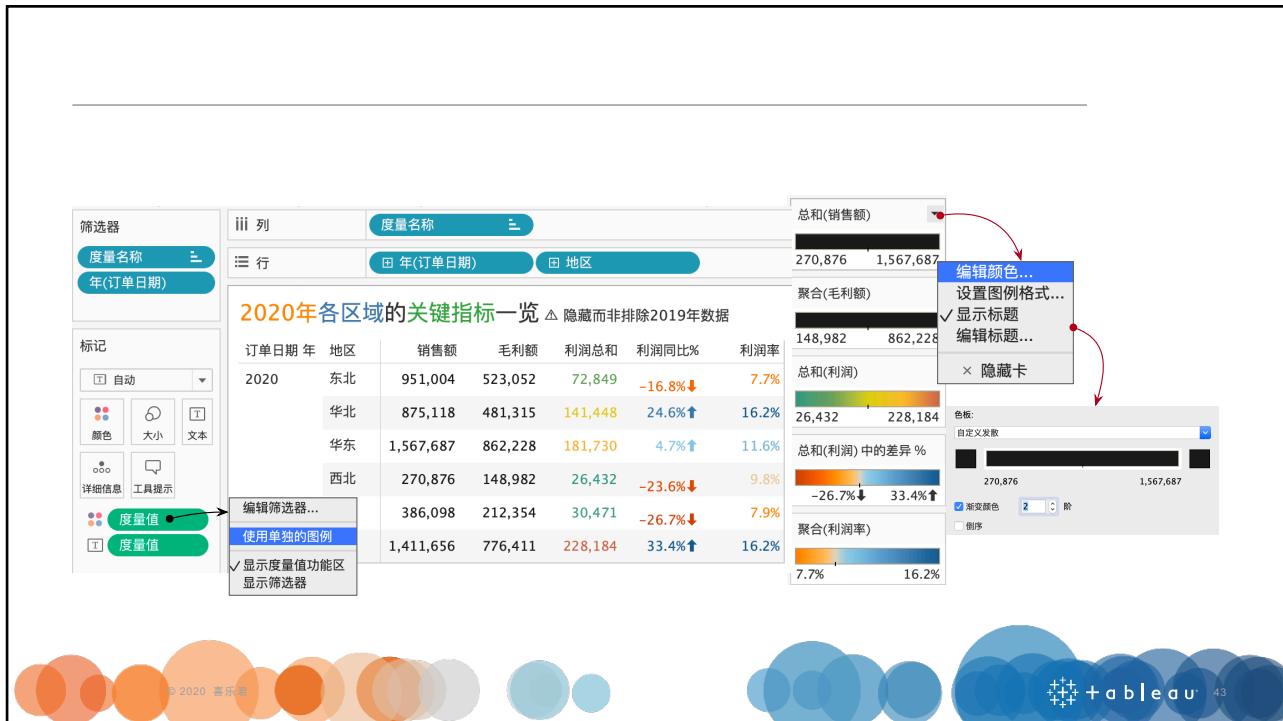
40



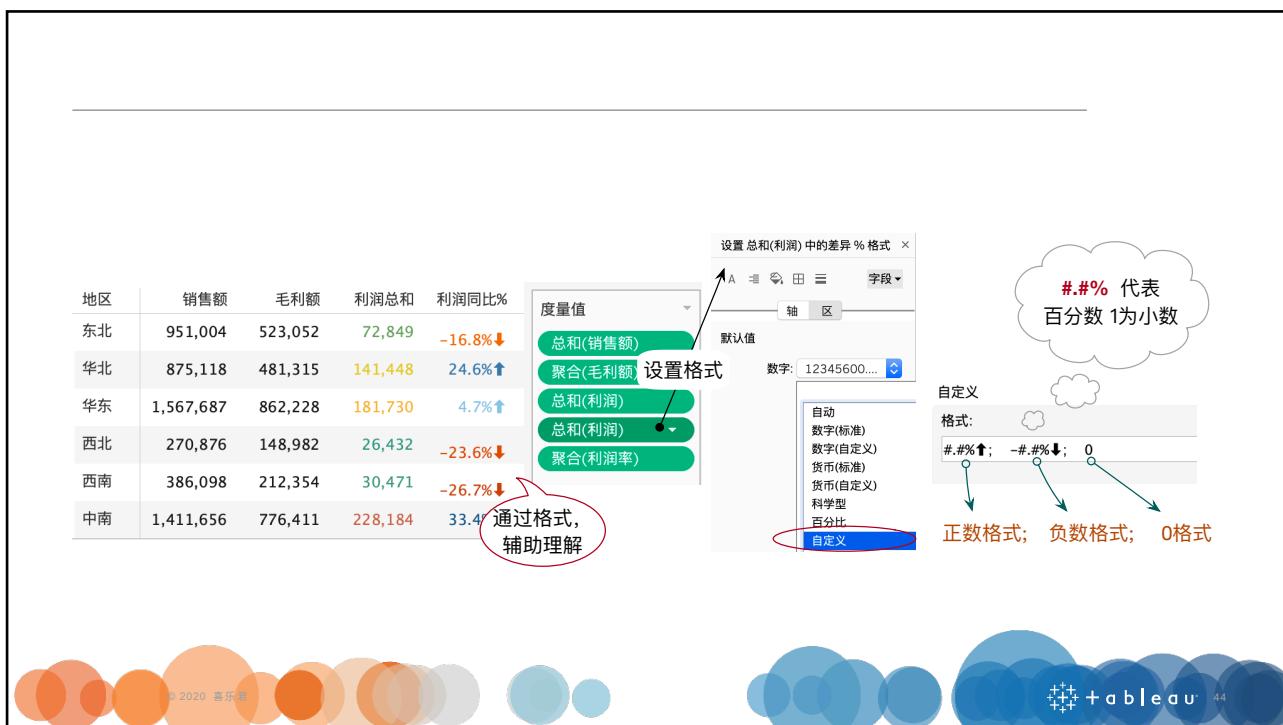
41



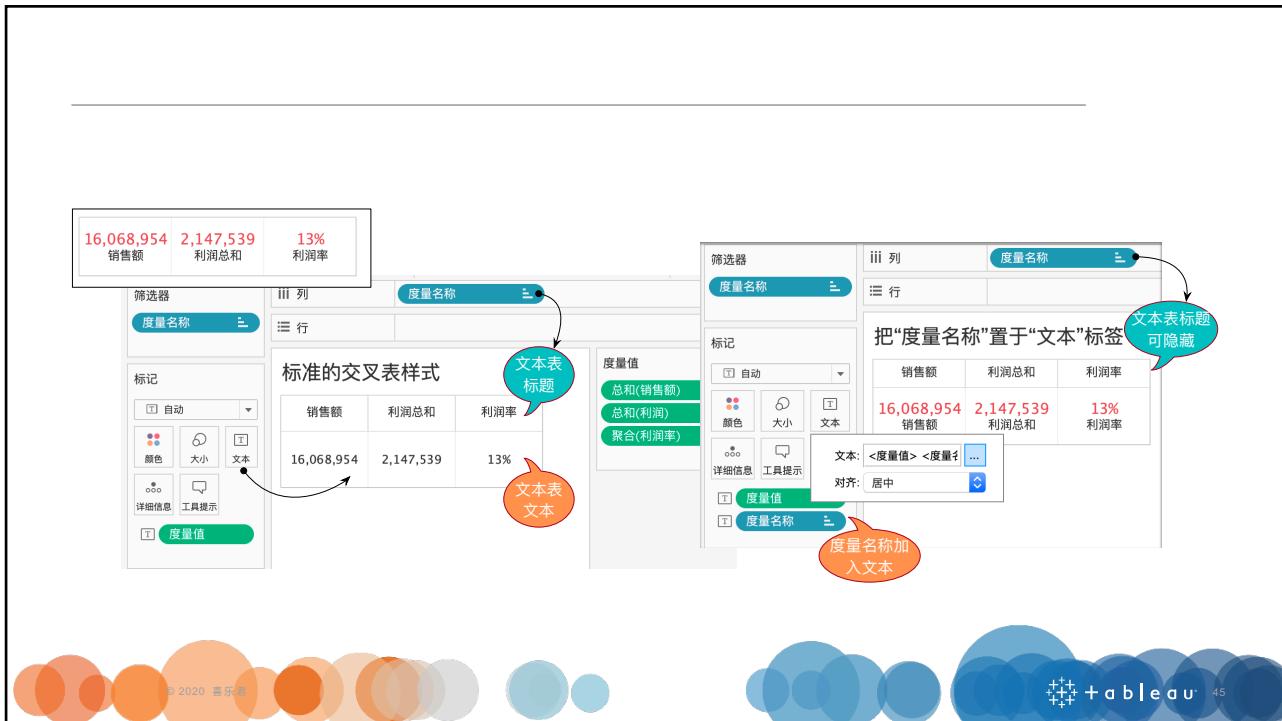
42



43

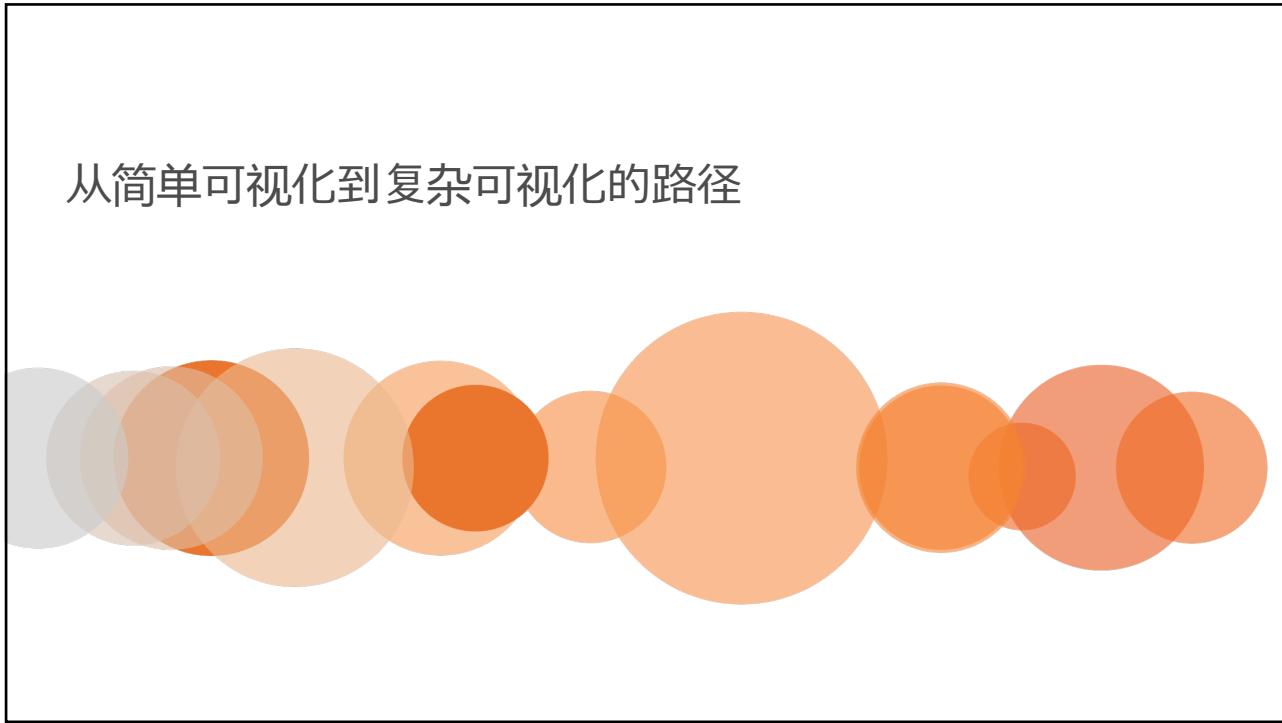


44

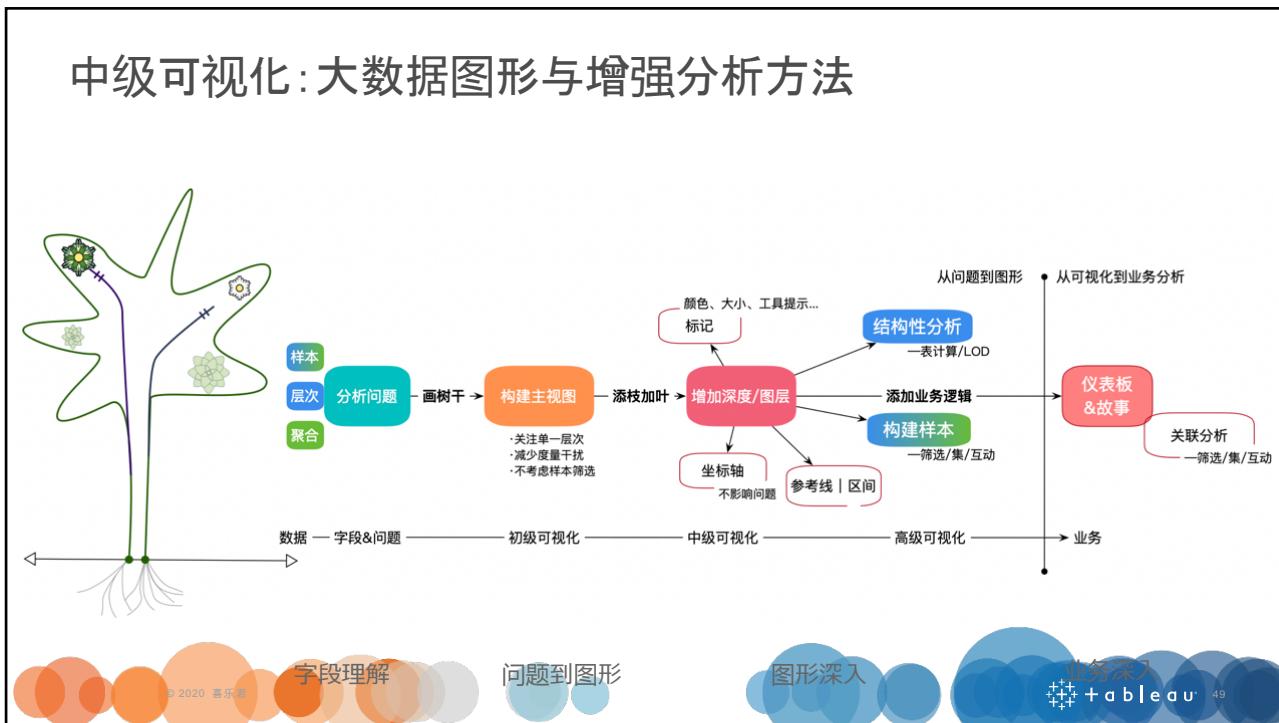


45

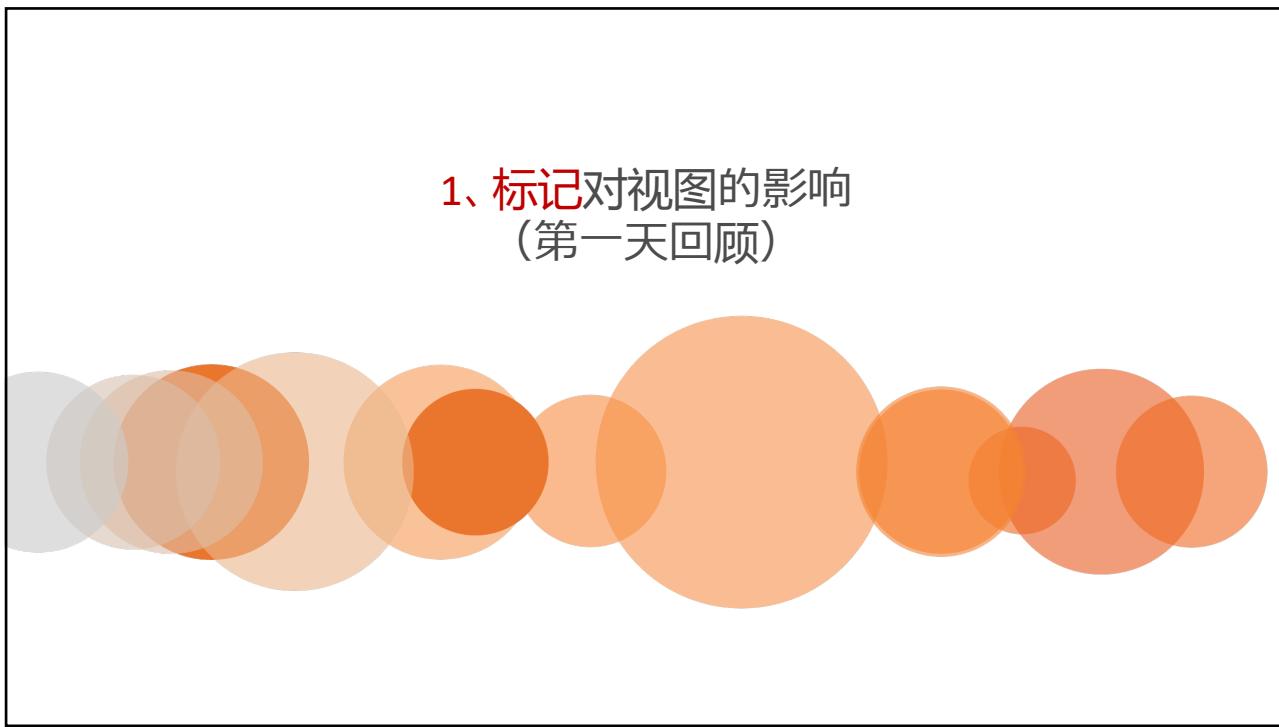
## 从简单可视化到复杂可视化的路径



48



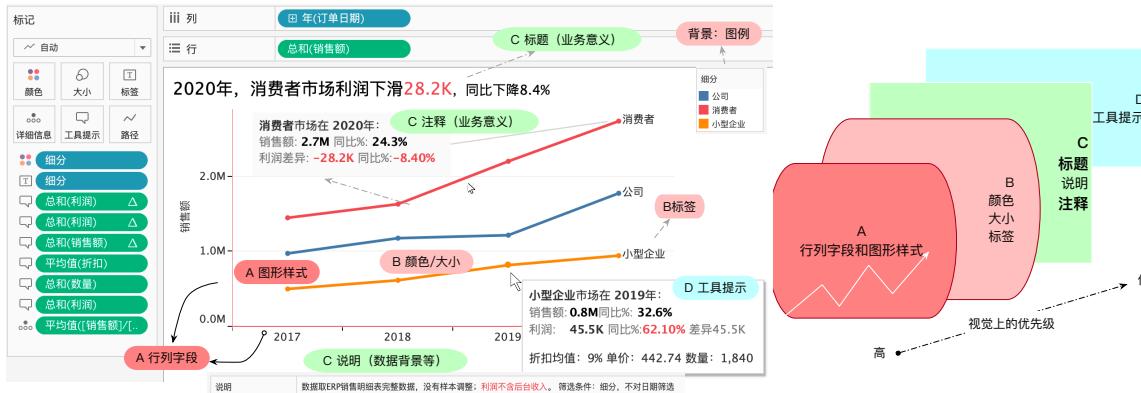
49



50

## 「可视化视觉模式」在主视图框架中增加可视化表述深度

借助于标记的方式，提高图形的信息密度



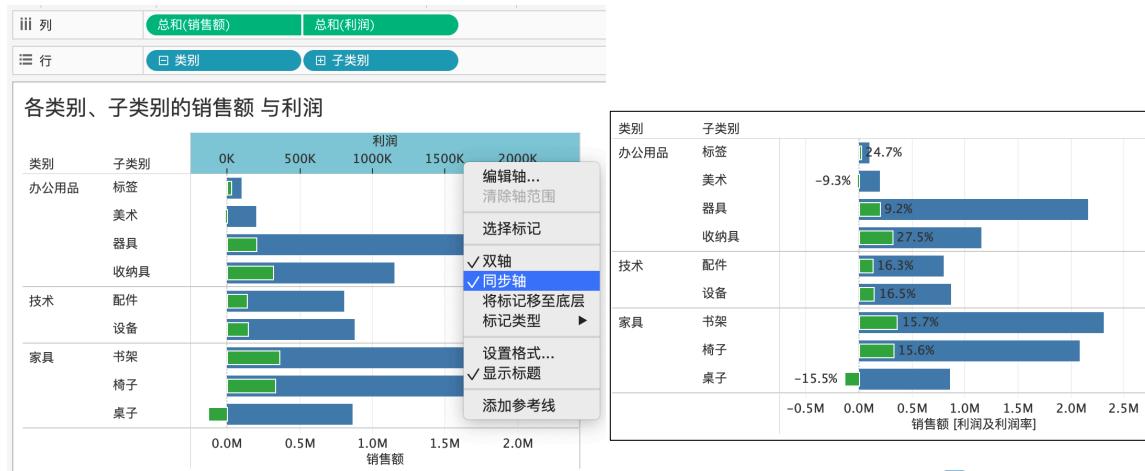
51

## 2、坐标轴、参考线和聚合对视图的影响（上）

52

## 坐标轴对可视化分析的影响——双轴

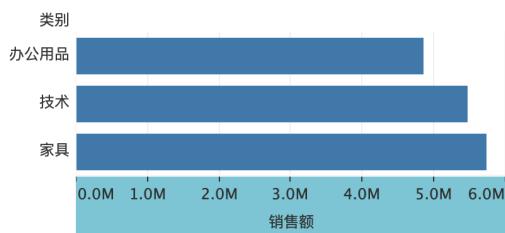
双轴、同步轴、多度量轴、度量名称、交叉表



53

## 双轴代表：泡泡糖图

### 最简单的条形图



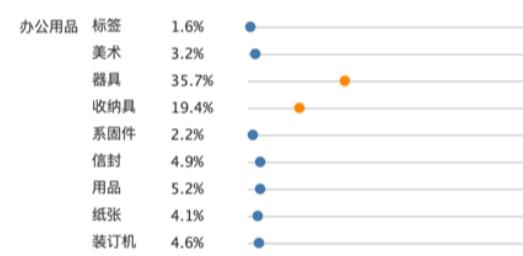
54

## 双轴图的改进：克利夫兰点图

克利夫兰点图-进度比例穿线而过



克利夫兰点图- 数据比较多时，文本加入标题



55

## 使用双轴图做时间序列分析

iii 列

田 月(订单日期)

iii 行

总和(销售额) 总和(利润)

2020年，各月的销售额与利润趋势 (双轴-折线图)



iii 列

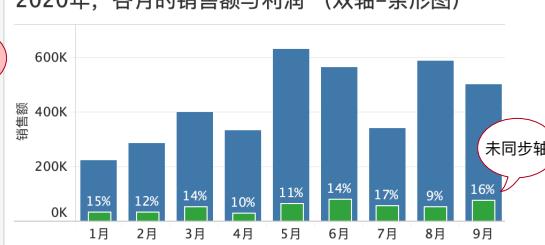
田 月(订单日期)

度量名称

iii 行

总和(销售额) 总和(利润)

2020年，各月的销售额与利润 (双轴-条形图)



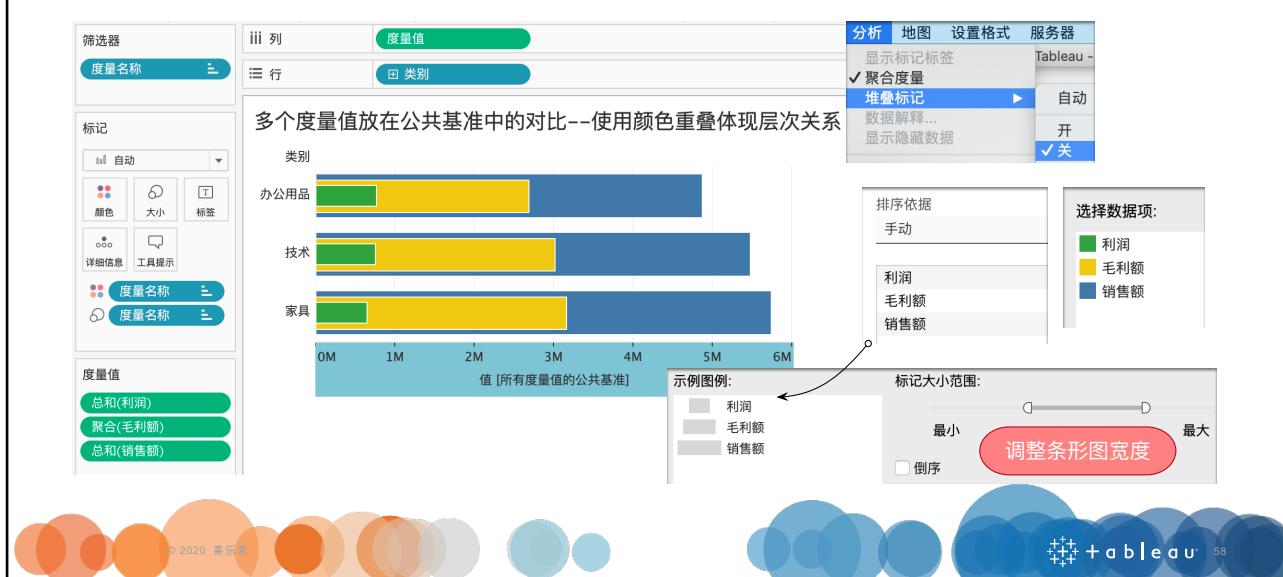
56

## 两个度量放在相同的位置——重叠构成“度量值”公共轴



57

## 多个度量名称借助于度量轴构成重叠组合

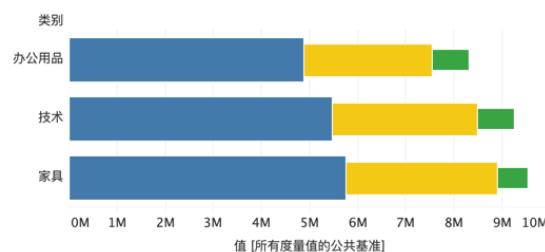


58

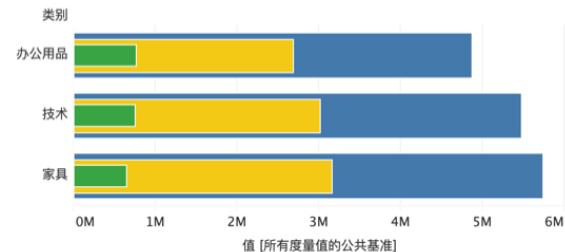
## 重叠和堆叠的区别

Tableau默认堆叠，通过分析改为重叠

多个度量值放在公共基准中的对比--默认堆叠体现累计关系



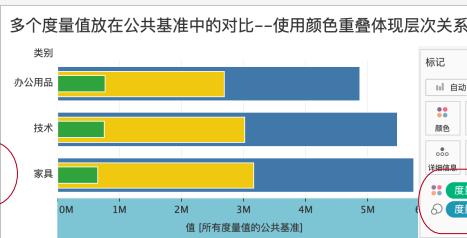
使用颜色「重叠」体现包含关系



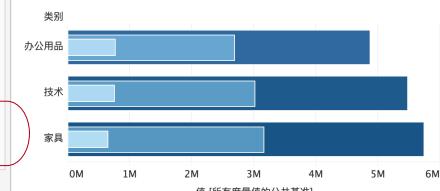
59

## 度量名称和度量值的差异——通过颜色来描述

离散与连续

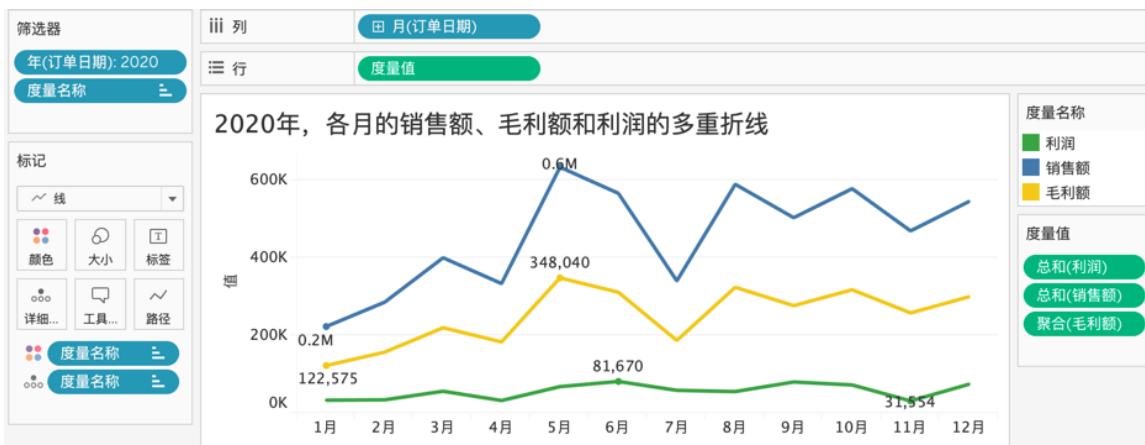


使用颜色「重叠」体现包含关系



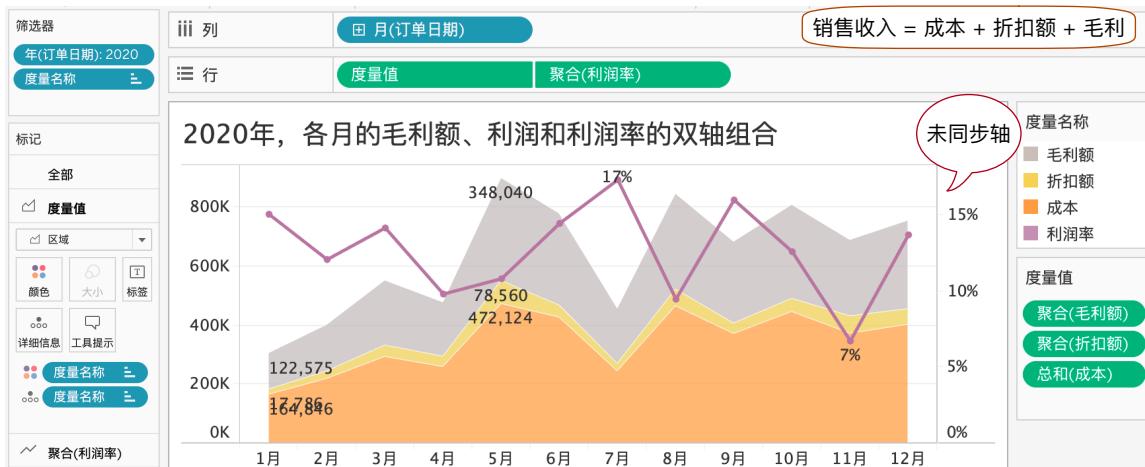
60

## 时间序列中增加多个度量



61

## 度量值 坐标轴与单一聚合的双轴合并



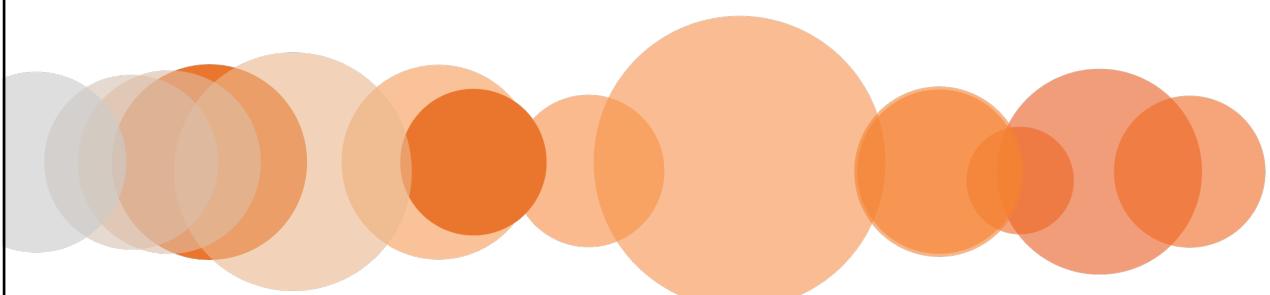
62

## 时间序列中 柱状图和折线的结合



63

## 2.2 高级处理：坐标轴的高级处理 日期和度量，在绝对和相对的转化



68

## 绝对坐标轴与相对日期轴的转化

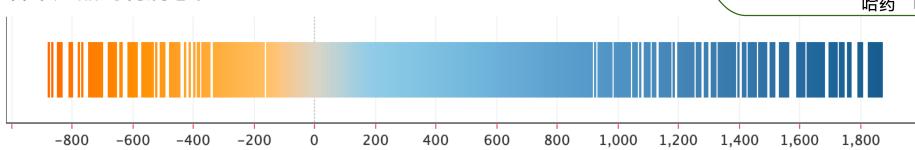
—本质上是基于原点的计算

日期是连续 计算机普遍设定 0=1900-1-1日零点

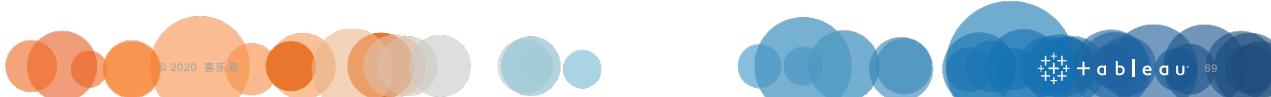


昨天、今天、明天

各个产品的利润总和



第1  
第2  
第3  
.....



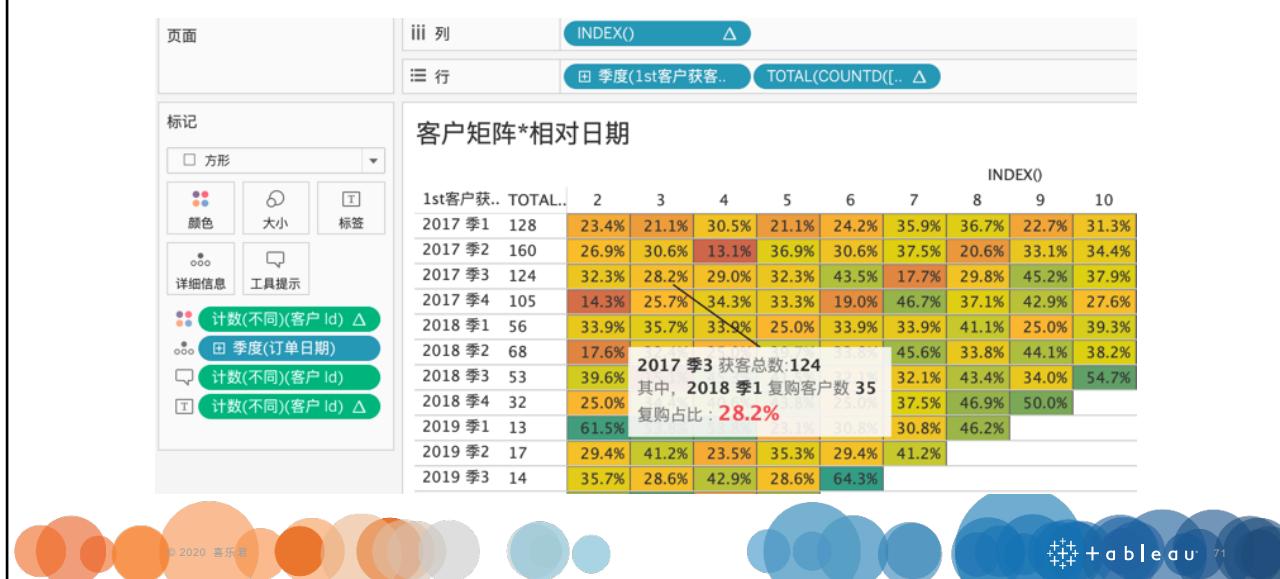
69

## 绝对日期向绝对日期的转化



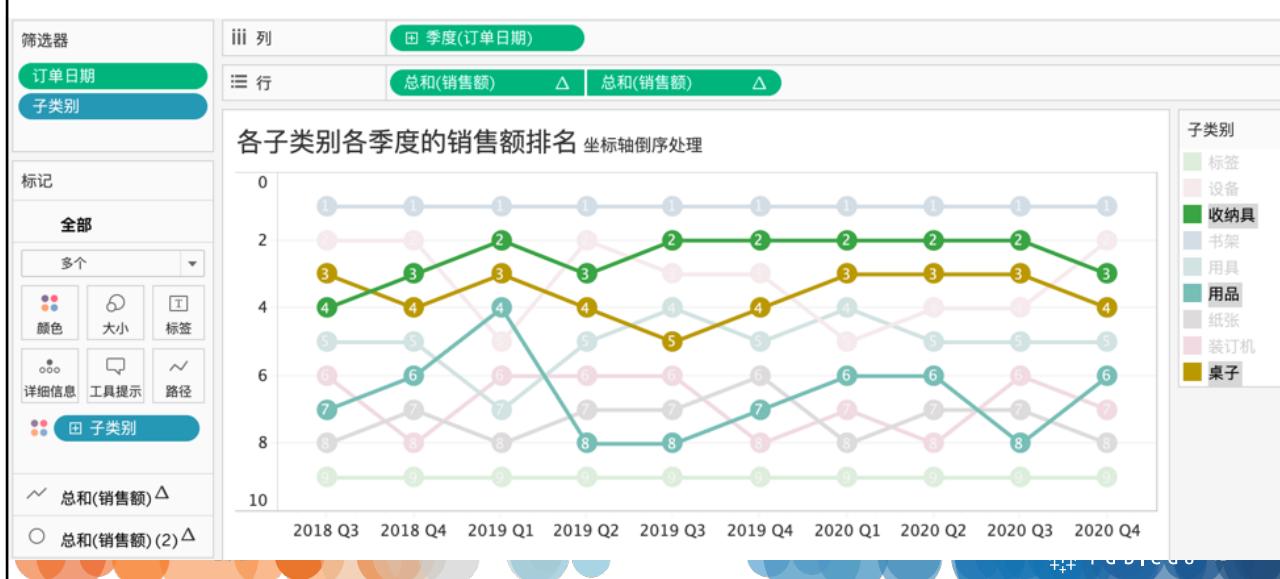
70

## 从绝对日期转向相对日期轴



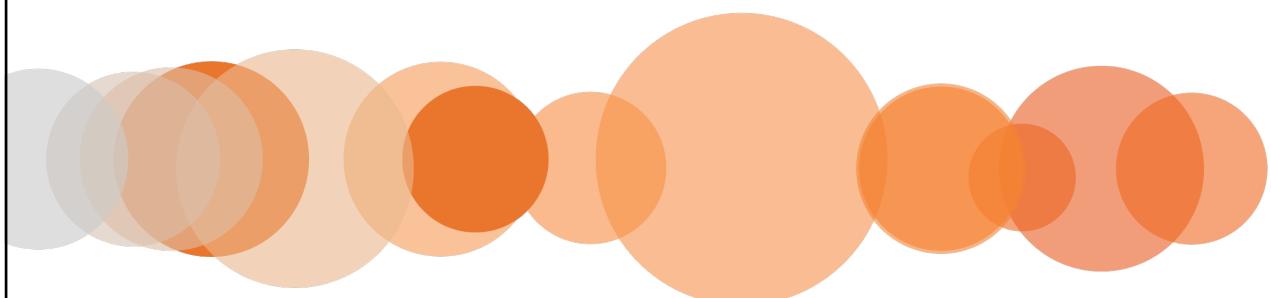
71

## 绝对度量轴到相对度量轴的转化



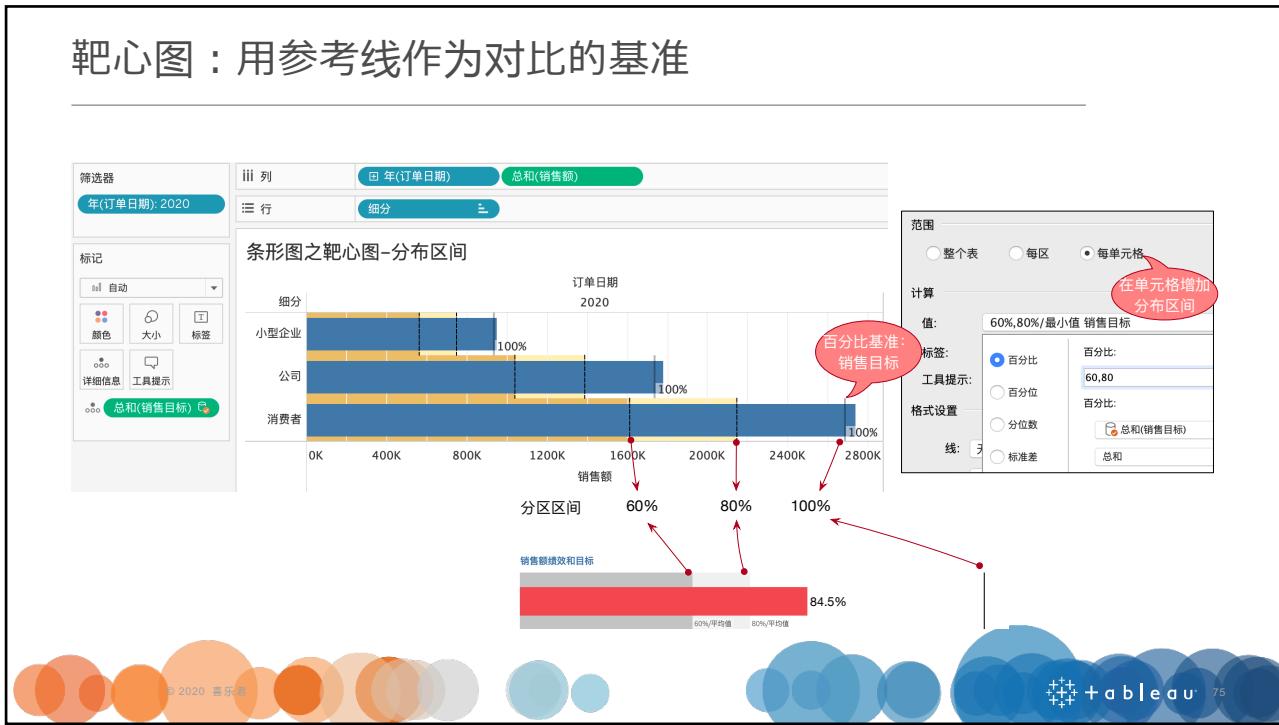
72

### 3 坐标轴、参考线和聚合对视图的影响（中）



74

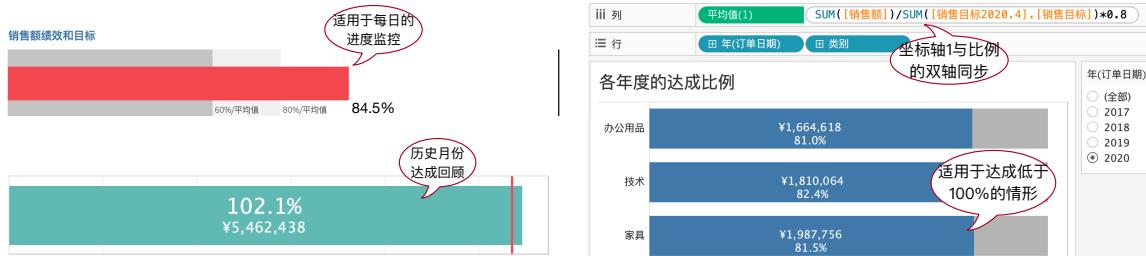
### 靶心图：用参考线作为对比的基准



75

## 双轴图和参考线 做对比的差异

参考线可以作为对比， 双轴图可以转化为背景



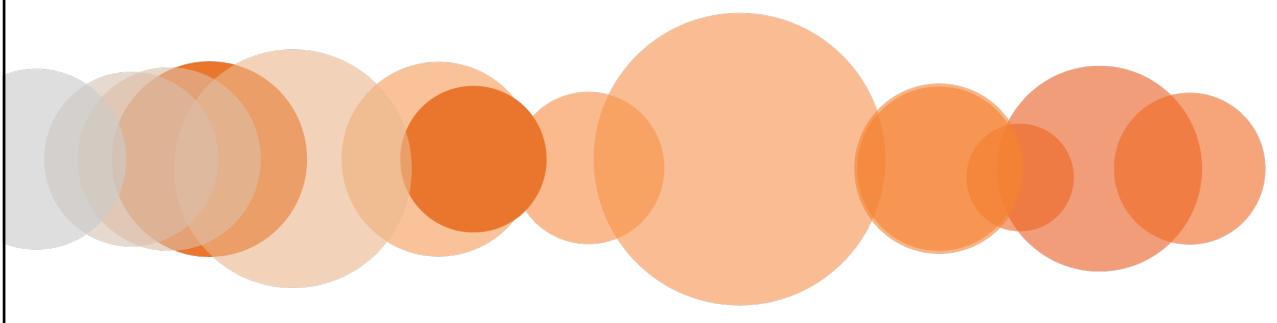
76

## 散点图——波士顿矩阵



77

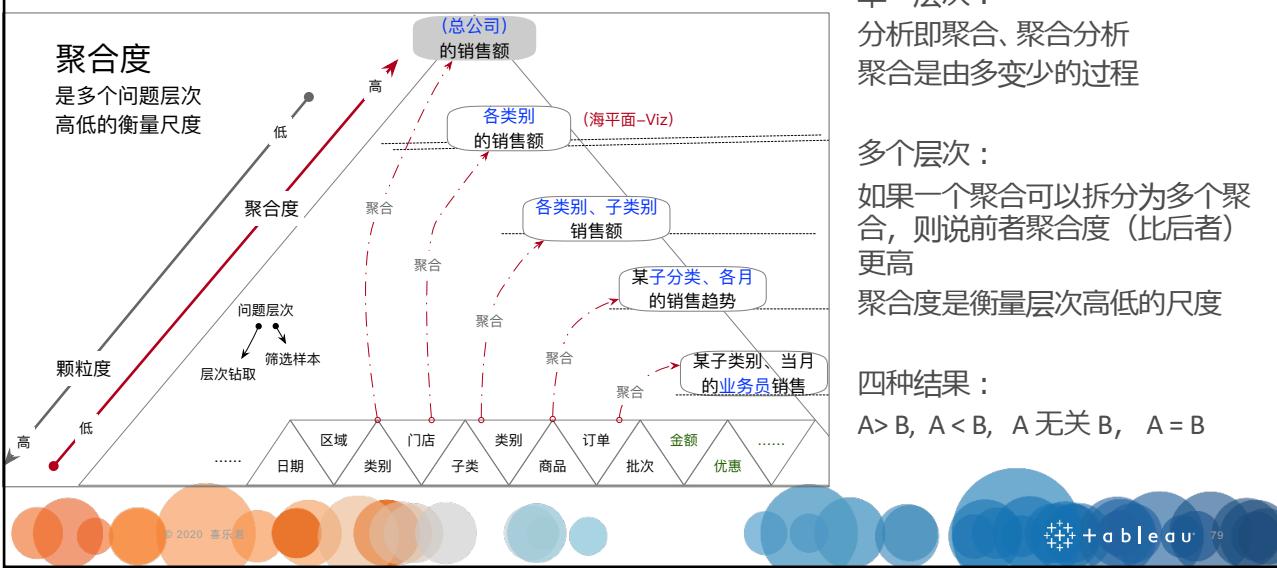
## 4 坐标轴、参考线和聚合对视图的影响（下） 简单OLAP分析入门



78

聚合度：衡量多个问题层次差异的尺度

2021新框架



79

## 聚合的方式

- LOD
- SUM AVG COUNT COUNTD ——描述总体的绝对特征
  - SUM/SUM
- STDEV、STDEVP ——描述波动性，基于AVG的计算转化
- PERCENTILE、MEDIAN、MAX、MIN ——描述个体的相对位置
  
- LOD+
  - WINDOW\_SUM
    - WINDOW\_SUM/WINDOW\_SUM
  - WINDOW\_STDEV
  - WINDOW\_PERCENTILE

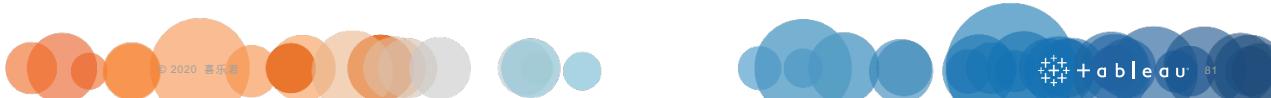
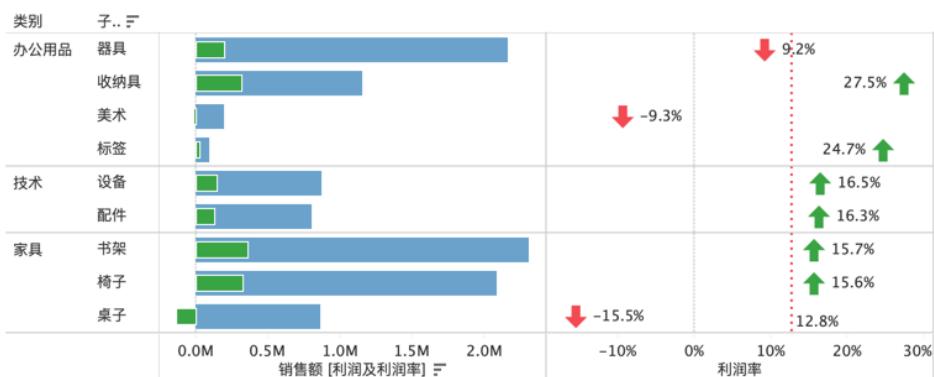


80

## 聚合和聚合的计算（比值）的差异 ——以 利润率 为例

SUM(销售额) 与 SUM(利润) /SUM(销售额)

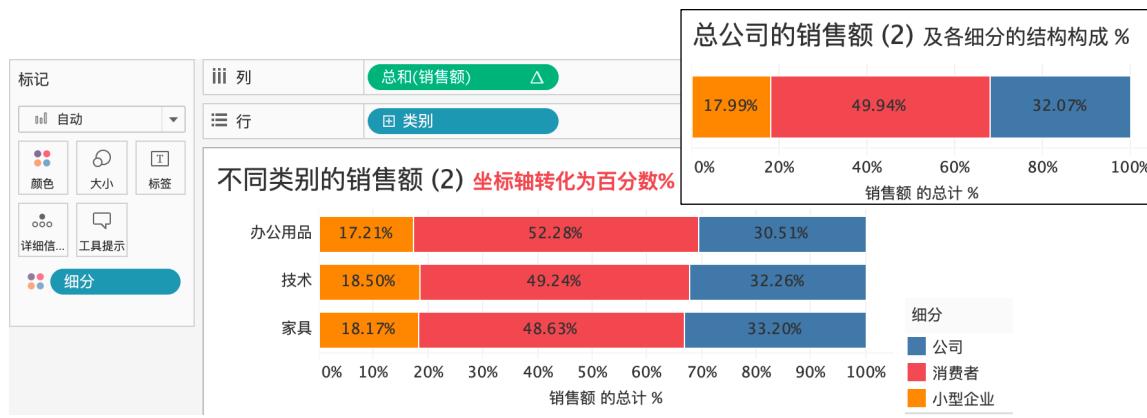
各类别、子类别的销售额、利润与利润率 - 可视化表达方法 (2)



81

## 表计算：聚合的二次聚合 ——以合计百分位为例（显性）

通过层次理解每个聚合所对应的问题位置



82

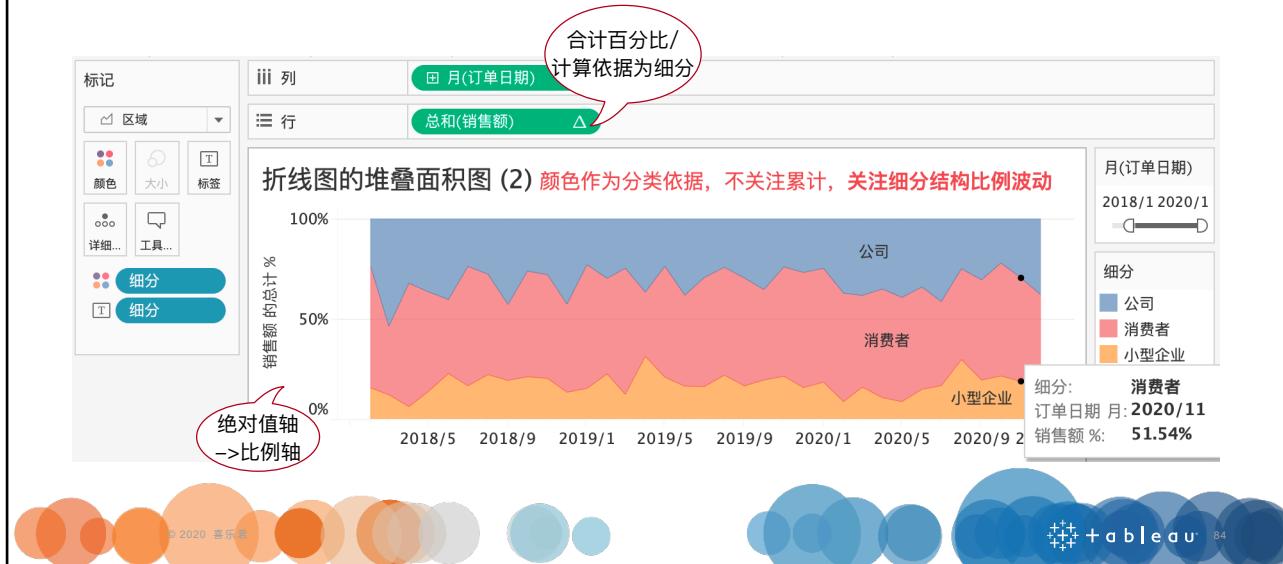
## 聚合的二次聚合：环形图是最简单的高级图形



83

## 表计算：聚合的二次聚合（隐形）

使用合计百分位作为主视图度量——从绝对值趋势转化为构成的趋势



84

## 独立的聚合/预先的聚合

两个聚合是相互独立的, 后者构成前者的结构——2021.1 升级了Fixed LOD计算创建方法



85

## 预先聚合的二次聚合///第四天会专门讲解计算

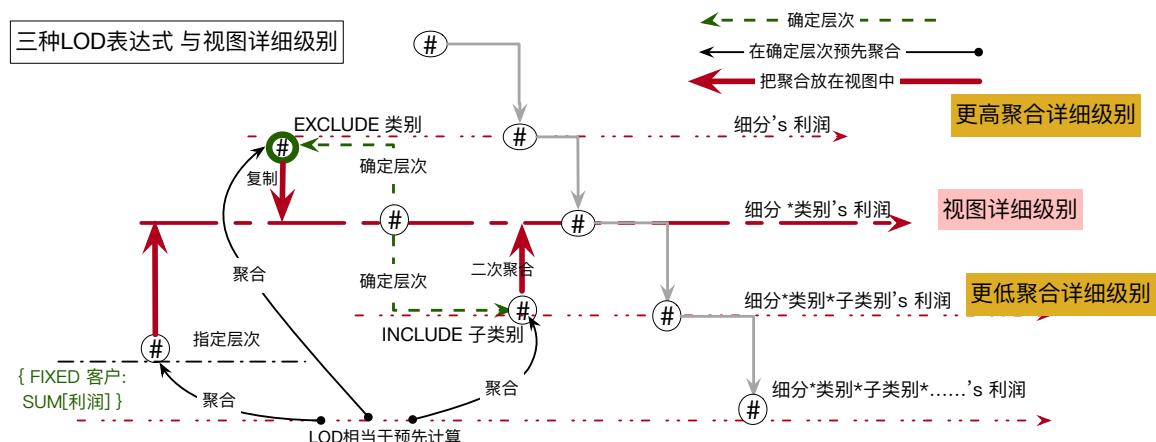
客户购买力分析  $\text{AVG}(\{\text{INCLUDE : MAX()}\})$



86

## 关键：用层次图理解聚合，是理解高级问题的关键//新修改

///第四天会专门讲解计算



88

谢谢

