



单独录屏、
录音

Tableau超越你的业务理解

样本控制（初级）

喜乐君
2020/4/16 第二天AM-1

1

业务分析与Tableau实现——2021年版本

- Part 1 可视化分析与业务决策 | 分析面向决策 | 分析是两步走 | 可视化的价值
- Part 2 从Excel、SQL到Tableau：本质与差异
 - 问题分析方法：结构与过程 | 问题的三种结构 | 聚合的过程 + 聚合函数 (+计算)
 - 可视化分析方法：结构 | 六种问题类型 | 空间、标记与意义
 - 字段分类 | 第一字段分类 | 第二字段分类
- Part 3 从问题到图形（上）
 - 初级可视化：图形的构成、选择与基本图形
 - 中级可视化：基于标记、坐标轴、参考线和简单计算的可视化改进
 - 高级可视化：{实例}绝对坐标轴到相对坐标轴的转化
- Part 4 样本控制与交互设置
- Part 5 函数使用与问题分析
- Part 6 数据准备、数据合并和数据模型

© 2020 喜乐君

Tableau

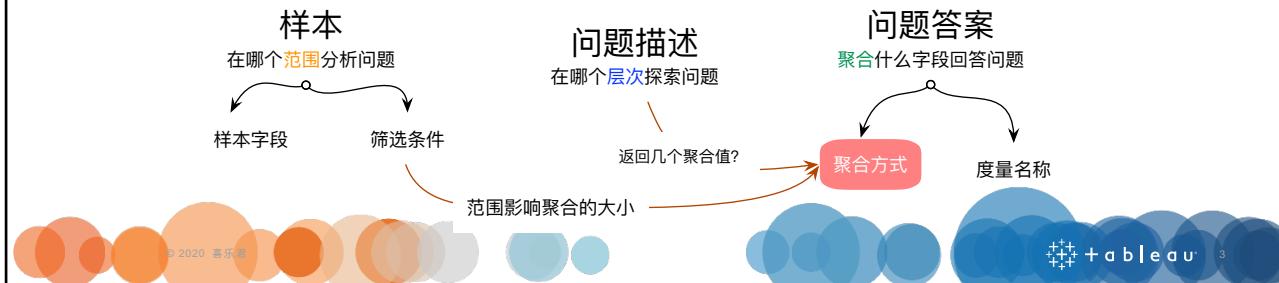
2



Part 2 从Excel、SQL到Tableau：本质与差异

- 1. 分析的结构：任何问题都是**样本范围、问题描述和问题答案**（聚合）三个部分构成
- 2. 分析的过程：分析即聚合，聚合即分析；聚合是从行级别到视图级别的过程
- 3. 差异：Tableau可视化 VS Excel/SQL 交叉表——关键是字段分类
- 4. 字段分类：第一字段分类（维度与度量）描述问题，第二字段分类（连续/离散）构建视图

“在消费者细分市场中，各产品类别的销售额总和与平均折扣”

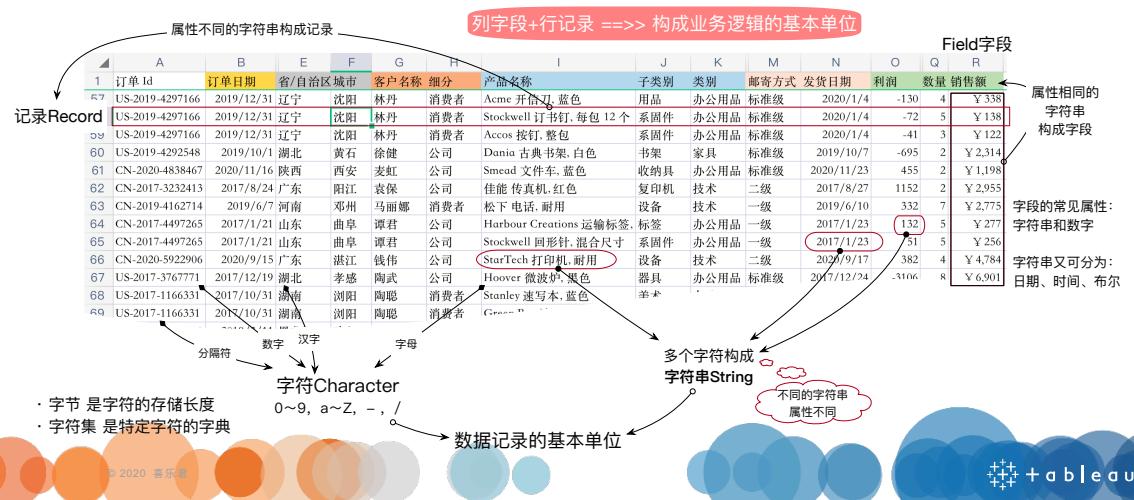


3

【基础】第一字段分类看数据的构成：第一字段分类即数据属性

业务逻辑是由字段列和记录行交叉构成的，前者描述问题，后者描述业务逻辑，合二为一就是事实

记录的常见属性构成：“谁(who)在何时(when)何地(where)，向谁(whom)，如何(how)提供了什么产品(what)，并使用哪些指标量化 (how much)这个行为。”

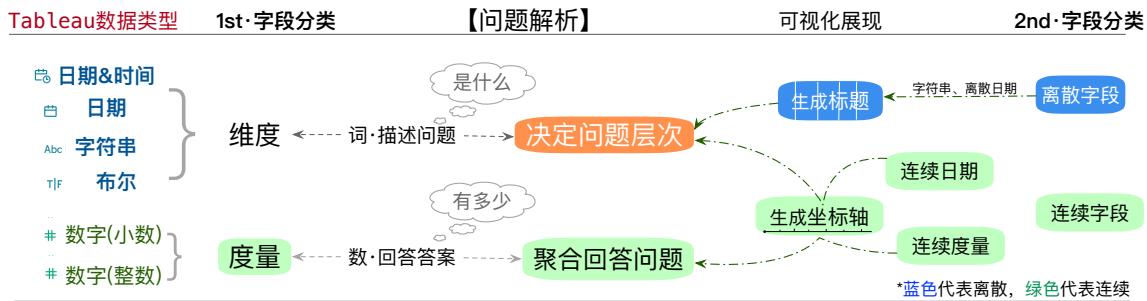


4

第一字段分类与第二字段分类关系图

【第二章最重要的图】

字段的分类、特征与拖曳逻辑（202009新）

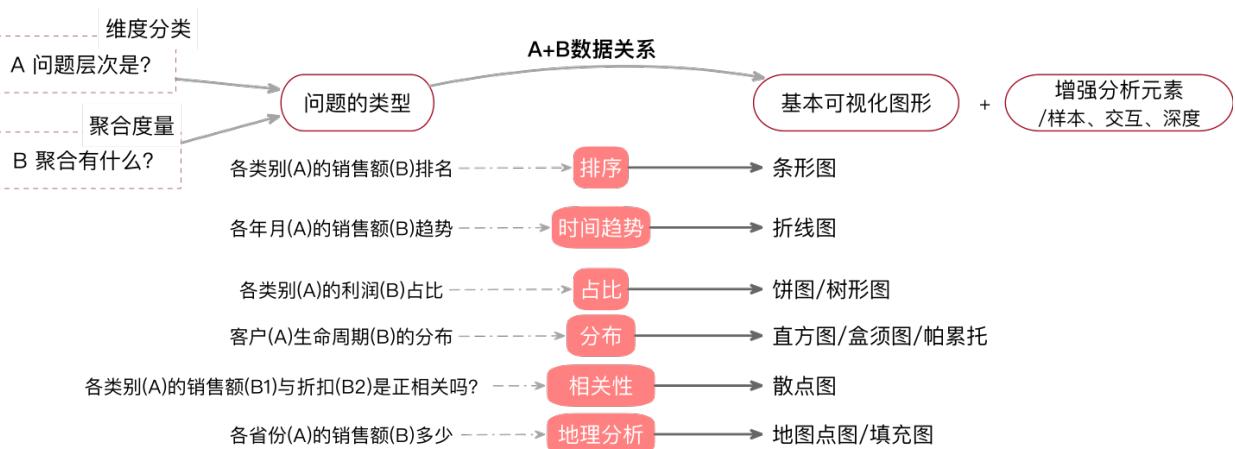


维度描述问题（是什么），聚合度量回答答案（有多少）
维度决定层次、度量默认聚合；离散生成标题、连续创建轴

5

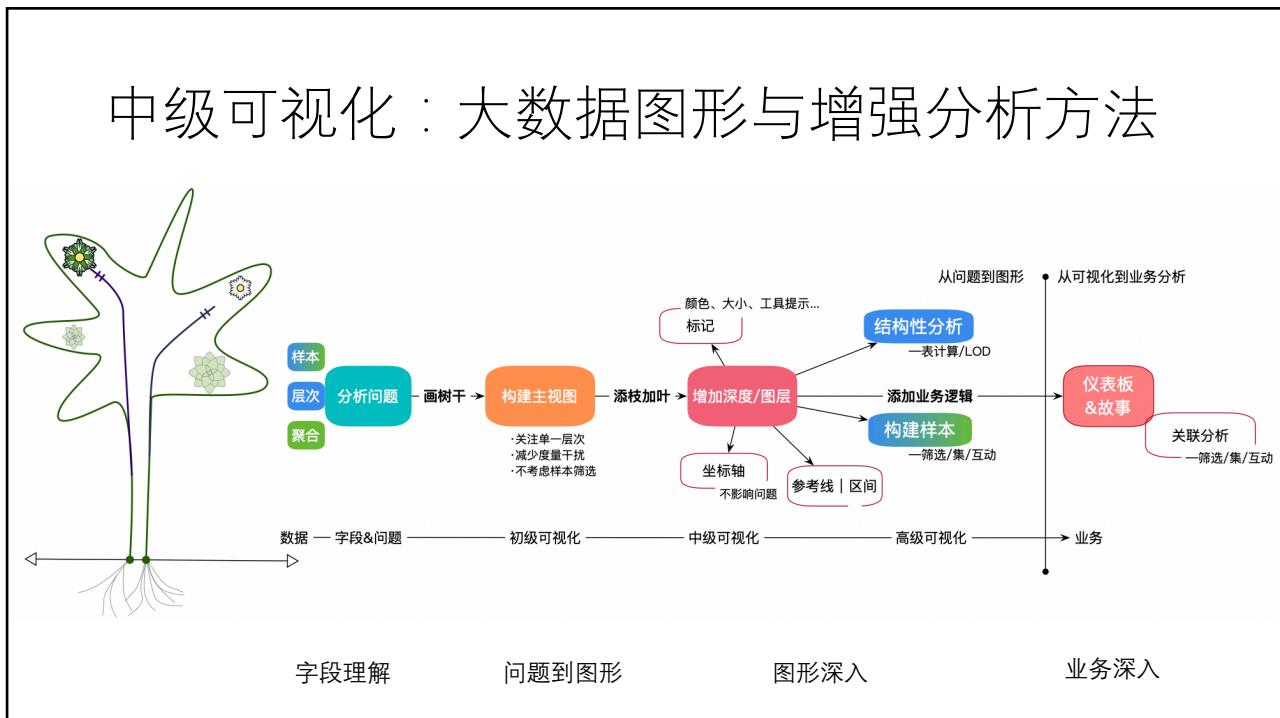
【从问题分类】问题决定图形，而非字段本身

7

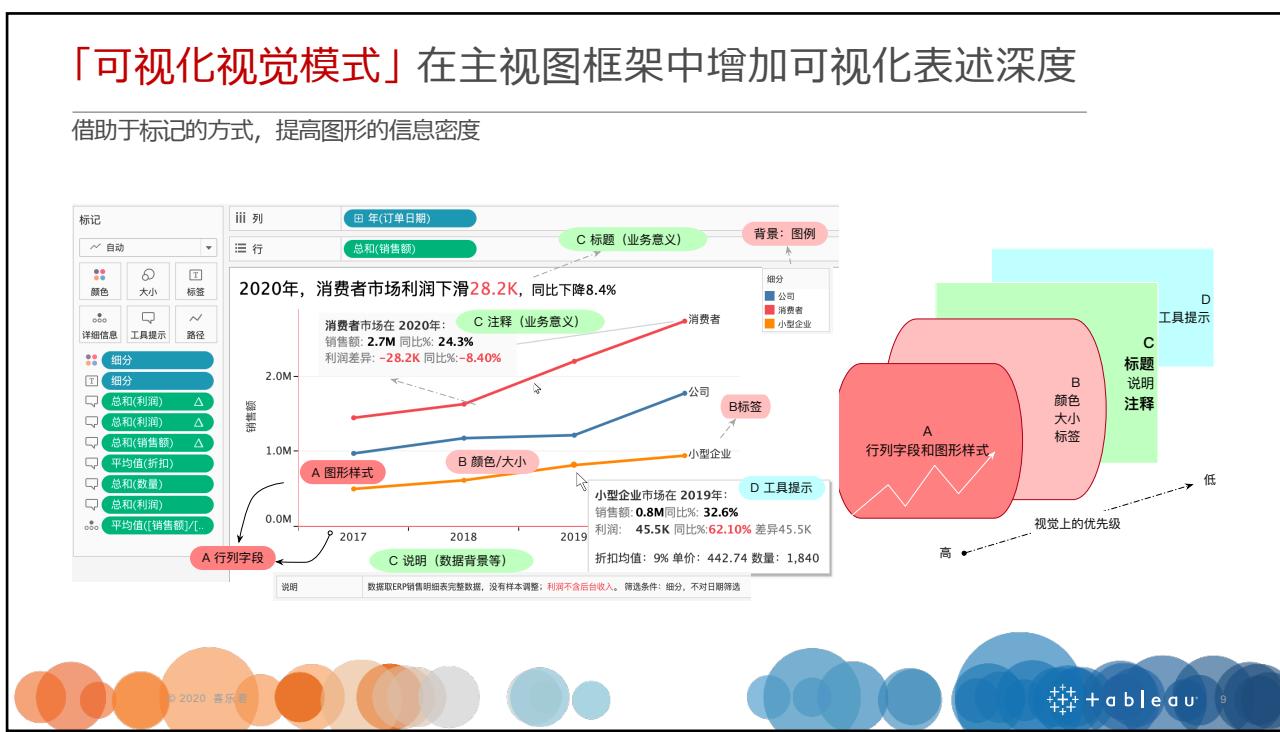


7

3

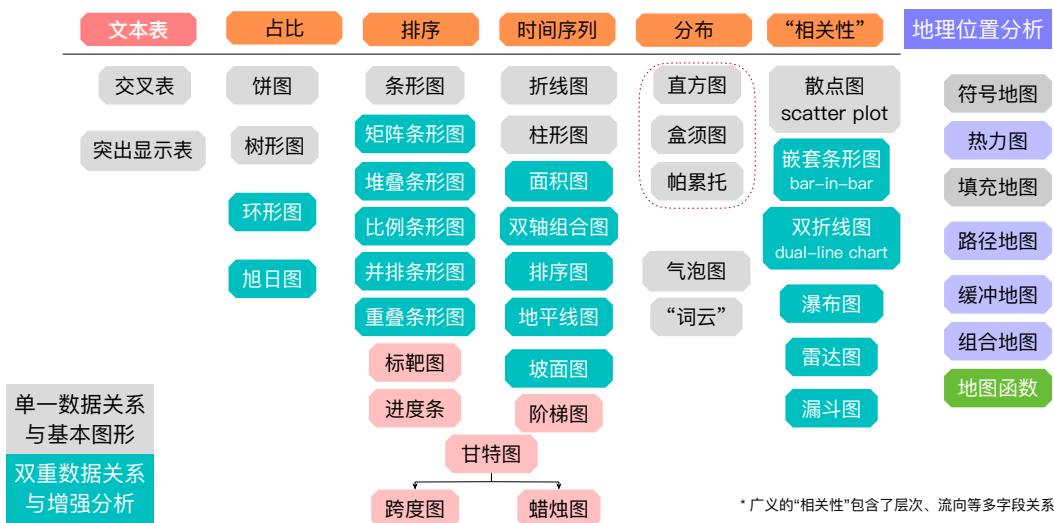


8



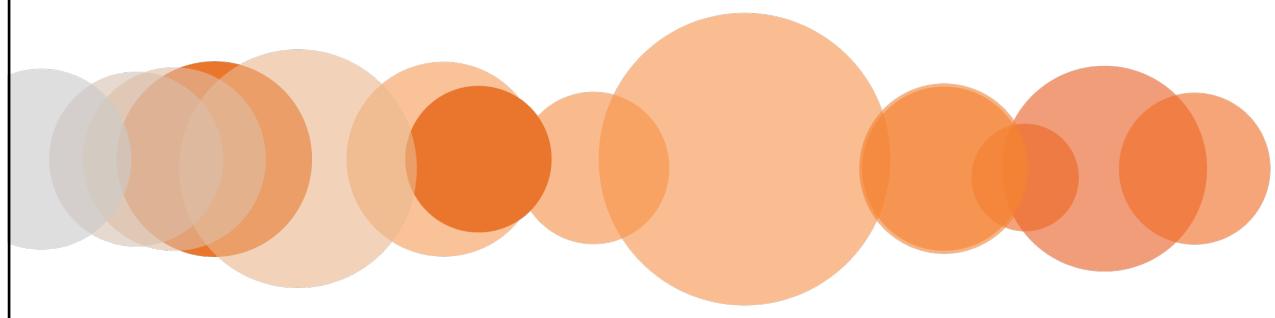
9

初级图形+标记/坐标轴/参考线/字段/计算.. >>



10

Part 4 样本控制与字段操作：可视化增强分析技术



11

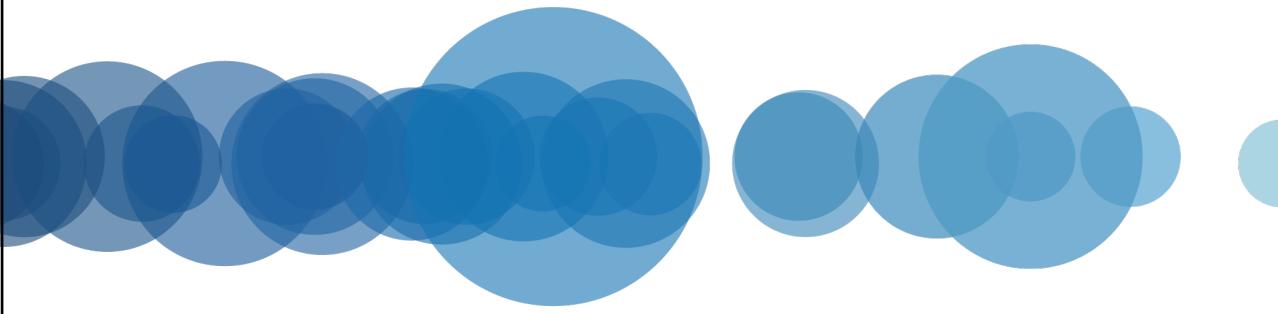
样本控制与交互设置

- Part 4 样本控制与交互设置
 - 最快捷的样本控制：快速筛选器
 - 按照字段分类的筛选器：维度筛选器、度量筛选器
 - 按照应用阶段的筛选器：数据提取筛选器、数据源筛选器、（视图筛选器）
 - 专门用于调整优先级的筛选器：上下文筛选器
 - 参数控制样本范围：参数 和 **参数动作**
 - 样本保存为模型：集 和 **集动作**
 - 有助于交互的：分层结构、分组等
- Part 5 函数使用与问题分析
 - 行级别函数与聚合函数的区别
 - 行级别函数
 - 聚合函数
- Part 6 数据准备、数据合并和数据模型



12

4.1 快速筛选器与动作筛选器

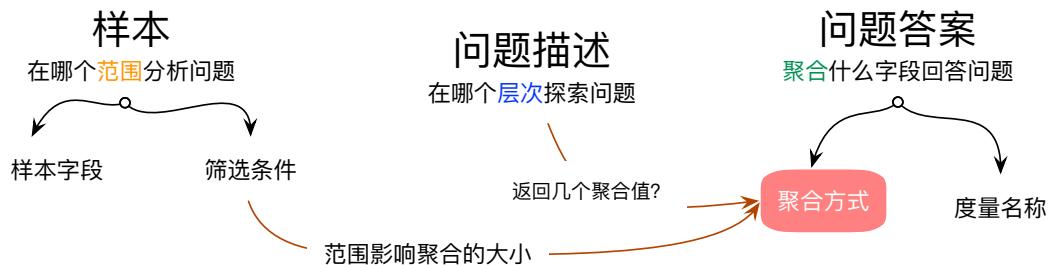


13

「样本控制」是问题分析中最千变万化的环节

样本控制，和关联交互，是敏捷BI的重要特征——反复的假设验证

“在消费者细分市场中，各产品类别的销售额总和与平均折扣”



14

理解明细的筛选：基于Excel的筛选器

在明细表中的筛选，都是基于明细（行级别、数据表层次、聚合度最低、颗粒度最高的层次）的筛选

An Excel screenshot showing a data table with columns: G (国家/地), H (地区), I (省/自治区), J (城市), K (类别), L (子类别), R (销售额), S (利润), T (利润率), U (记录数), V (行). The I column is highlighted with a red box. A dropdown menu is open over the I column header, showing a filter for "省=安徽" (Province = Anhui). Another dropdown menu is open over the T column header, showing a filter for "利润<=0" (Profit ≤ 0). The bottom right corner of the screenshot has a watermark: "© 2020 喜乐君".

15

使用SQL对明细筛选，和建立在明细聚合上的二次筛选

SQL 使用where对行级别做筛选，使用having对聚合后的问题聚合做二次筛选

The screenshot shows the Tableau Data Source interface on the left and the SQL Editor on the right.

Data Source:

- SCHEMAS: SP_deal
- Columns: 已退货? (checkbox), 发运天数 (checkbox), 子类别 (checkbox), 折扣 (checkbox), 国家/地区 (checkbox), 细分 (checkbox), 数量 (checkbox), 销售额 (checkbox), 销售区域 (checkbox).

SQL Editor:

```

1 • select * from tableau.SP_deal
2 where 类别='办公用品' and 利润<0
3 -- 查找 办公用品类别，且利润亏损的交易
    
```

Result Grid:

国家/地区	细分	数量	销售额	销售区域	利润
Panama	公司	2	401.76	LATAM	-50.64
Panama	公司	3	87.84	LATAM	-48.36
Mexico	消费者				

Annotations:

- A red box highlights the WHERE clause in the SQL code.
- A red box highlights the '查询所有明细, 筛选where' (Filter all details, filter where) text above the result grid.
- A red box highlights the '查询结果预览 (累计5800行)' (Preview results (5800 rows)) text below the result grid.
- A red box highlights the HAVING clause in the SQL code.
- A red box highlights the '先筛选where, 再聚合group by, 再筛选having' (First filter where, then aggregate group by, then filter having) text above the result grid.
- A red box highlights the '查询结果 (仅四行)' (Query results (only 4 rows)) text below the result grid.

18

在SQL中完成筛选的多种方法 | 优先级

对应 where 和having 子句—— having 中，维度是聚合的依据

The diagram illustrates the execution order of SQL clauses for filtering:

- 维度 (Dimension):** Category = '办公用品' (highlighted in red dashed box).
- 度量 (Measure):** 利润 < 0 (highlighted in red dashed box).
- 运算顺序 (Execution Order):** select, where, group by, having.
- 筛选逻辑 (Filtering Logic):** 先筛选where, 再聚合group by, 再筛选having (highlighted in red box).

SQL Editor:

```

1 • select 类别,子类别,sum(销售额),sum(利润) from tableau.SP_deal
2 where 类别='办公用品' and 利润<0
3 group by 类别,子类别
4 having sum(销售额)>200000 -- 销售额总和大于20万的子类别
    
```

Result Grid:

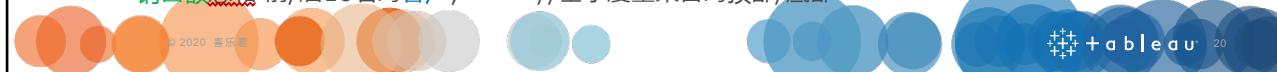
类别	子类别	sum(销售额)	sum(利润)
办公用品	收纳具	1625534.4287...	-590727.7811999992
办公用品	美术	294933.93230...	-115662.99770000014
办公用品	纸张	556479.79130...	-177088.40870000017
办公用品	用品	339311.52890...	-103610.4511000002

19

基于Excel和SQL总结：常见的几种样本控制的类型

广义的筛选器都是条件筛选器，根据字段的第一/第二分类又可以做进一步细分

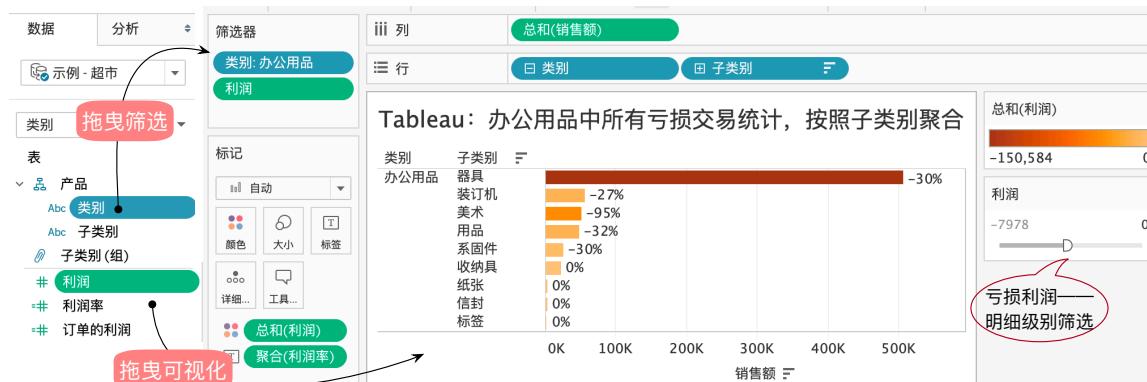
- in/equal (基于相等/不相等的判断)
 - 在 2020年 中, // 离散的日期筛选器
 - 在 西北 地区中, // 维度筛选器
 - 在 办公用品和家具两个类别中, //
 - 在 员工身份证号码开头为1101的样本中, // 基于离散维度的条件筛选
- From To /between (基于范围的判断)
 - From 2021年1月1日 to 4月1日, // 范围, 基于连续的日期
 - 利润总和 from 1万 to 10万的客户中, // 范围, 基于连续的聚合度量
- 条件 (基于范围的判断)
 - 利润小于0的 交易明细, // profit < 0 连续的利润 (度量)
 - 利润总和小于1万的 客户, //sum(profit)<1w 连续的聚合利润 (度量)
 - 销售额总和 前/后10名的 客户, //基于度量聚合的顶部/底部



20

Tableau敏捷BI：拖曳同时实现查询筛选和分析聚合

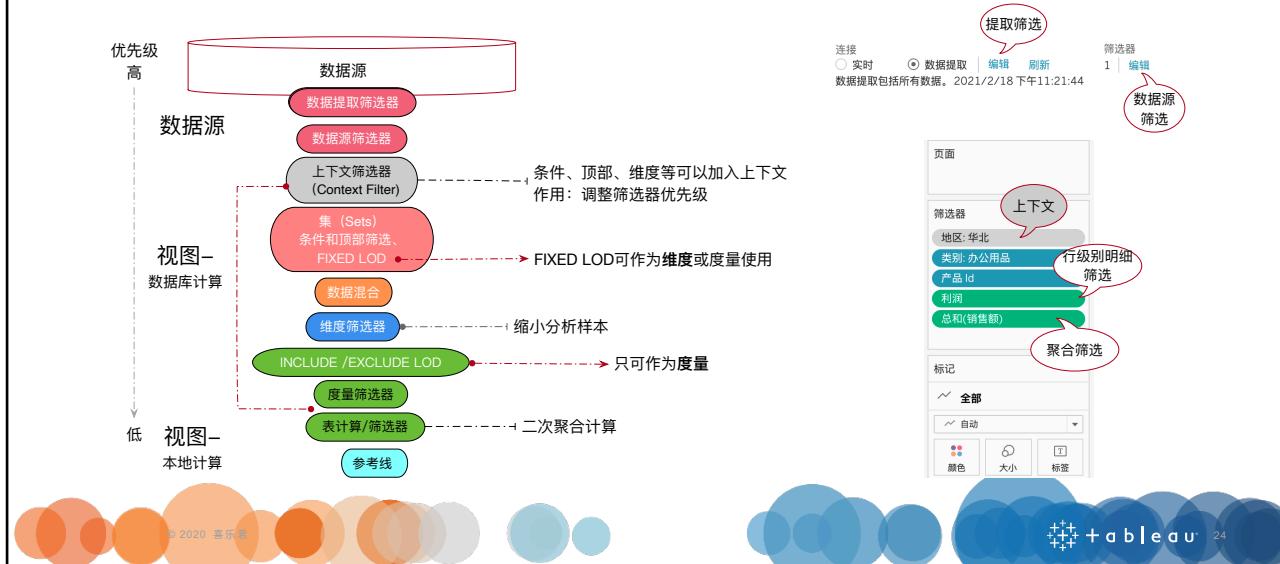
样本范围：样本字段和判断条件，对应的字段拖入筛选器



22

筛选器的优先级：上下文、条件、行级别、聚合

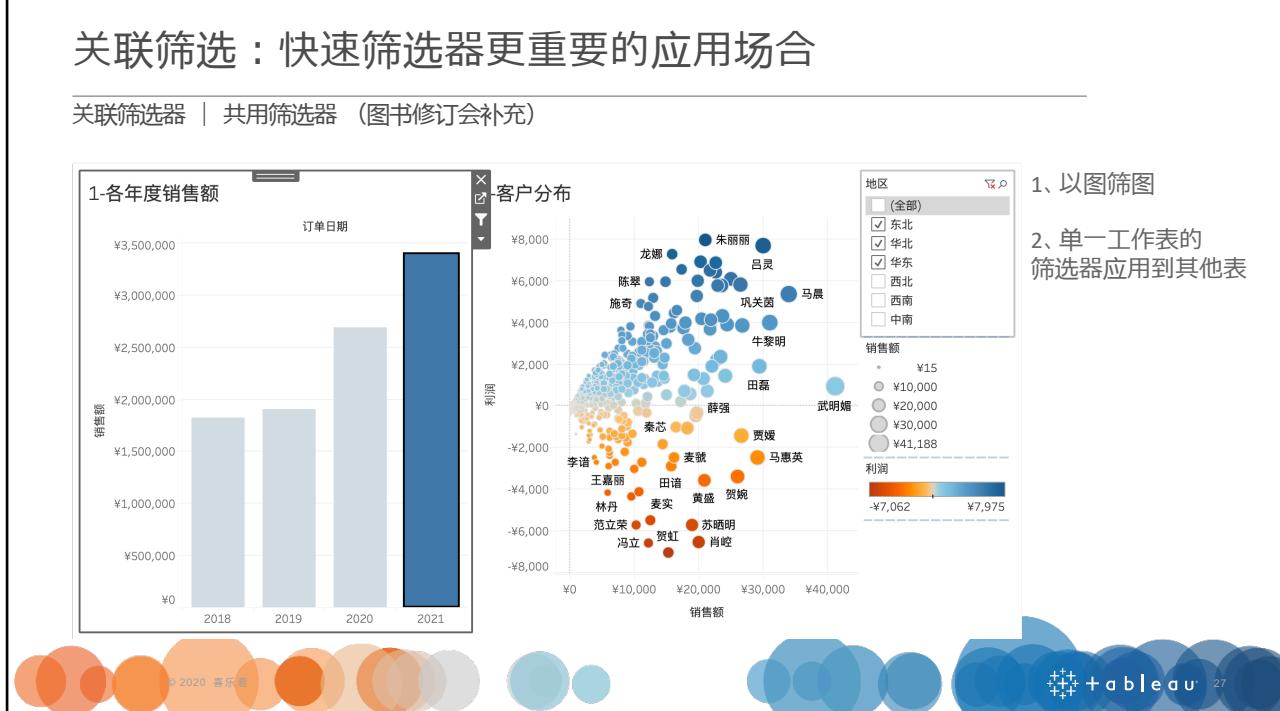
上下文用于处理 筛选器之间优先级的问题



24

关联筛选：快速筛选器更重要的应用场合

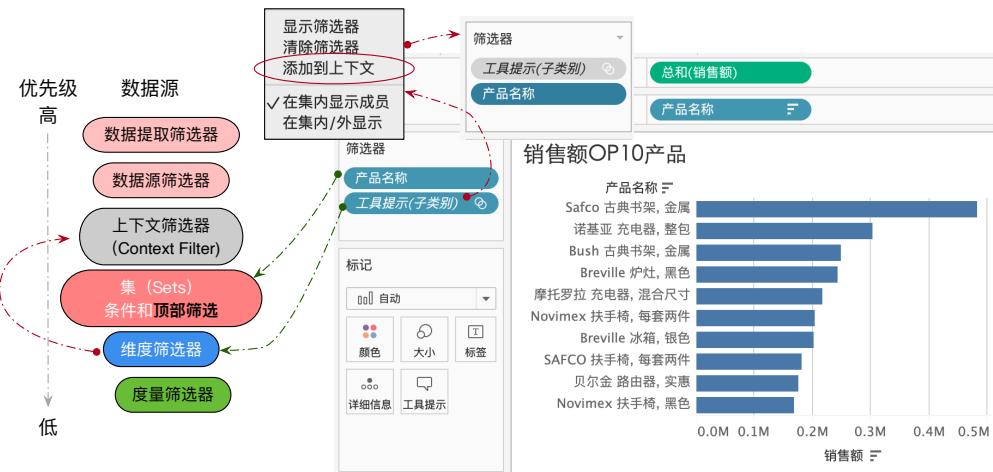
关联筛选器 | 共用筛选器 (图书修订会补充)



27

筛选器/操作筛选器/工具提示筛选器——代表不同创建方式

筛选器的创建方式，不影响优先级的设置



28

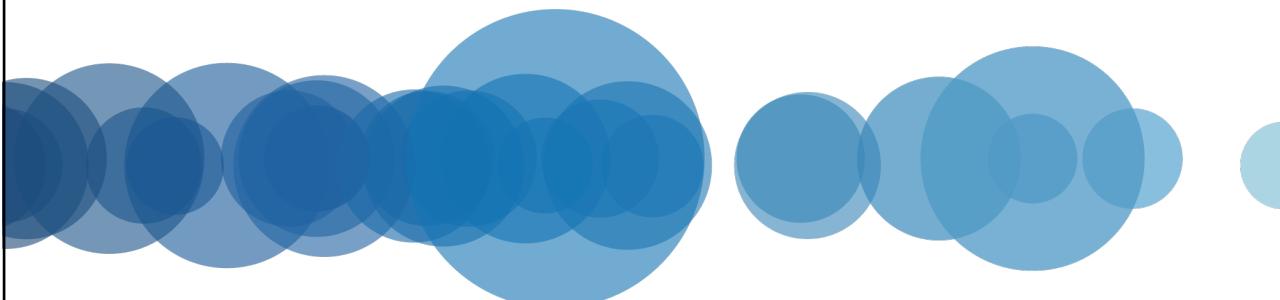
快速筛选器 总结

分析“样本范围”的关键，是拆分问题中不同筛选器的类型与优先次序

- 维度筛选器优先于度量筛选器
 - 基于维度的聚合条件筛选器 优先于 普通维度筛选器
 - 行级别筛选器 优先于聚合筛选器 (利润<0 sum利润<0)
- 上下文筛选器专门用于调整视图中多个筛选器的优先级
- 数据提取筛选器 和 数据源筛选器 是数据准备阶段的筛选器，优先于视图筛选器
- **快速筛选器**是直接添加的筛选器，**动作筛选器**是基于交互的筛选器
 - 在仪表板中，可以以图筛选（表A用作筛选）
 - 在仪表板中，单一工作表的筛选器可以扩展到其他表甚至整个数据源
 - （后续补充）筛选时指定筛选样本

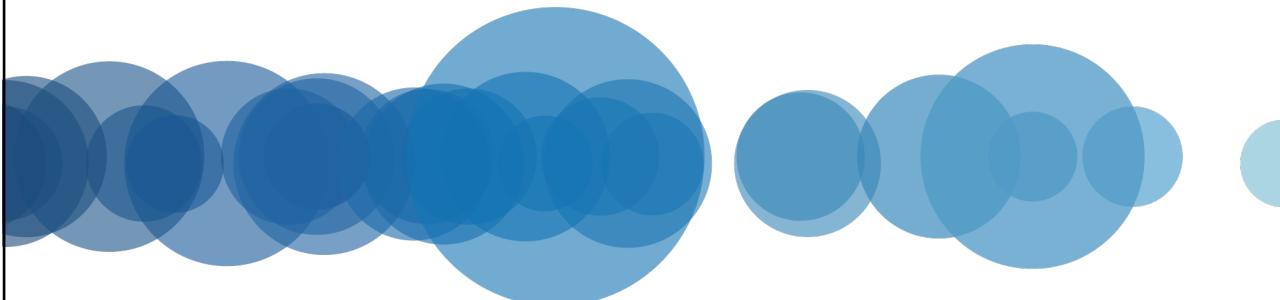
29

4.2【集】将筛选范围保存为 样本 反复使用 ——下次课程



30

4.3 参数：控制变量的重要工具



39

场景：一个视图应该满足不同人的不同样本需求

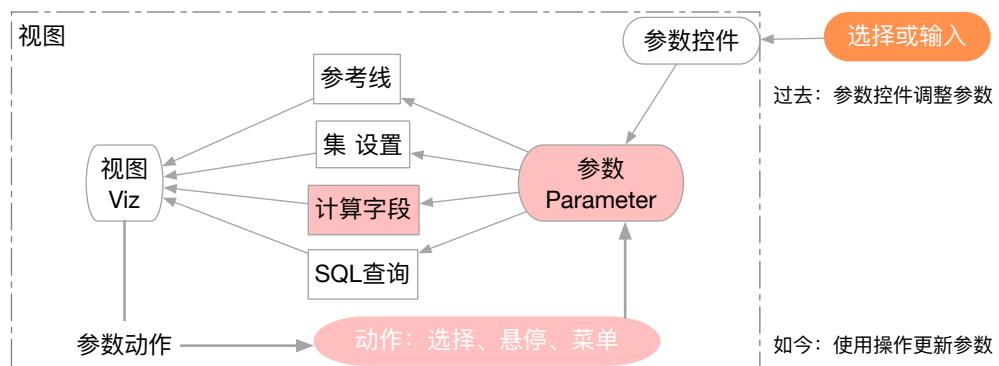
- 条件/TOP分析：
 - A：我要查看top 20 的客户及其占比
 - B：我要查看top 50 的客户及其占比
 - C：我要查看利润总和大于1万的客户的各月趋势，或者查看利润总和大于2万的各月趋势
 - 我希望可以认为控制样本的大小
- 参数控制视图
 - 数据桶：以200为间隔，以500为间隔，以参数为间隔
 - 参数传递 字段
 - 通过更改参数，查看销售额总和/利润总和 / 利润率的各季度波动（参数控制视图）

40

40

参数 和视图的结合，可以实现非常多的效果

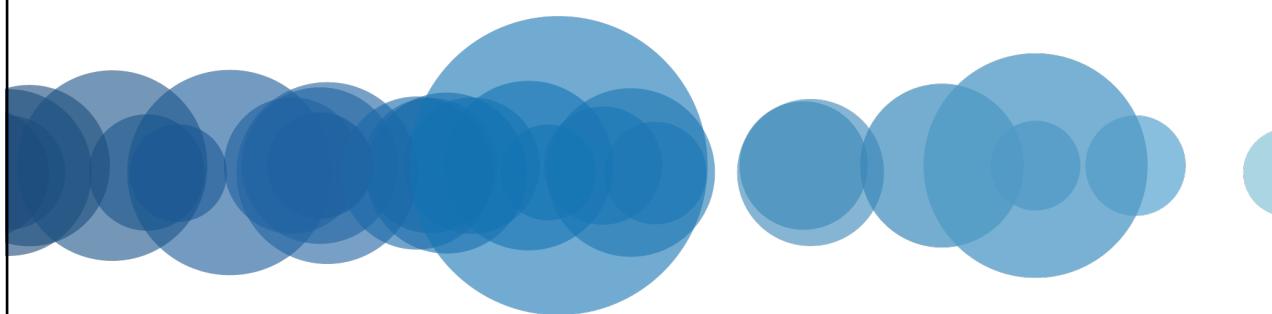
实例：使用参数更改视图的聚合值



41

13

4.4 字段操纵：分层结构、拆分、合并



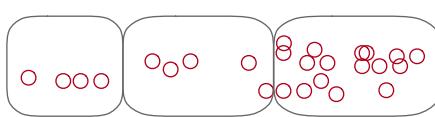
44

分层结构：特指字段之间有明确的一对多关系

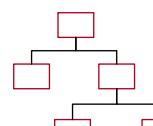
分层结构可以视为对多个字段的数据准备；而分组则是对单一字段内 数据值的数据准备



○ 代表字段内的一个数据值



分层结构



【实例】每个类别、子类别、产品销量最高

45

拆分split 和合并merge ——本质上都是计算

- 实例：从订单id 中拆分 “年”
- 实例：由于不同省份的订单id可能有重复，创建“省份+订单ID”作为分析的 订单id使用
- Split函数
- “merge”函数
 - 使用 + 加号，等同于 Excel中的 连字符&



46

「一切数据处理，皆是计算的化身」

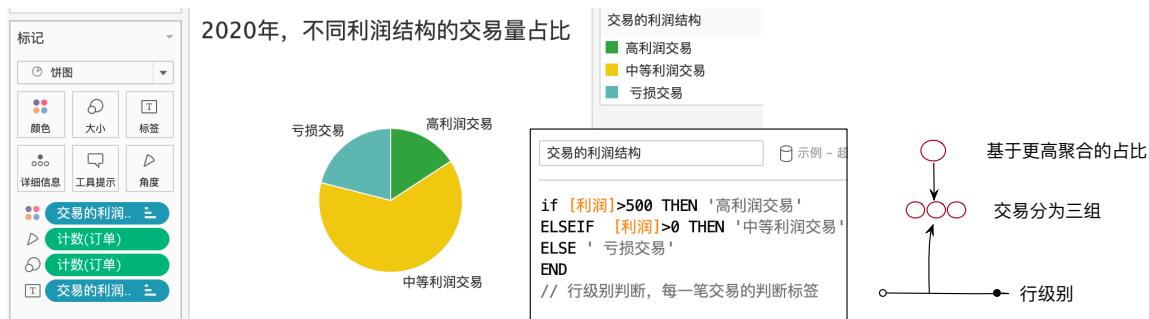
筛选
集
组
分层结构
合并
拆分
.....



48

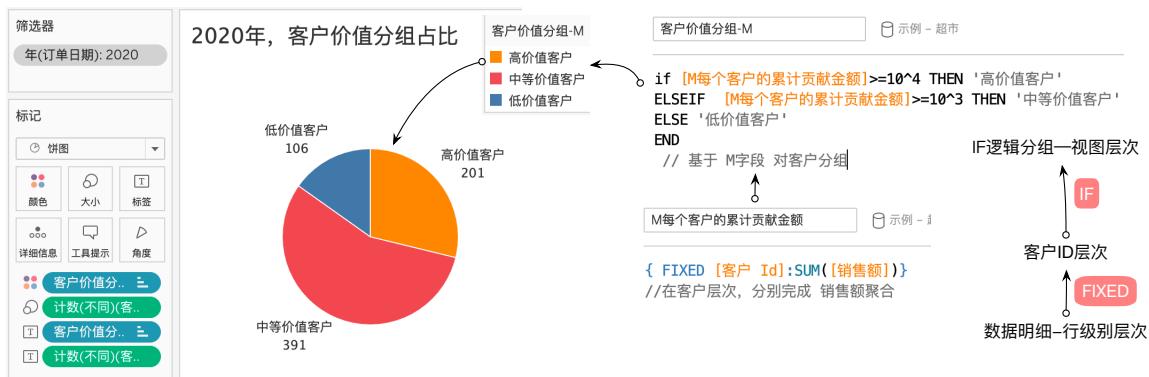
一切皆计算：计算是有限字段满足无限业务场景的关键

实例：单笔交易利润大于500为高利润交易，2020年，公司高利润、低利润、亏岁利润交易的占比分别多少？



49

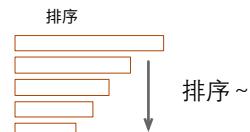
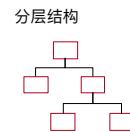
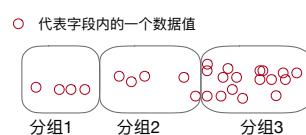
计算是有限字段 满足无限业务场景的关键



50

总结：样本控制与字段操作：可视化增强分析技术

- ok • 筛选 (Filter): 缩小可视化数据的数据范围。
- 集 (Set): 按照一定条件，把数据分为两个部分。
- ok • 参数 (Parameter): 用户可以输入从而与视图交互的入口。
- 排序 (Sort): 对数据按照规则排序。
- 分层结构 (Hierarchy): 多个字段按照层次结构组织架构。
- 分组 (Group): 给定字段内的数据分组，合并相似数据，减少字段内数据量。
- 参考线、趋势线和预测线: 提供模型化的辅助线，帮助理解数据背后的规律。
- 其他: 别名 (alias)、字段合并 (merge)、字段拆分 (split) 等。



51