

VIS投稿经验分享

王叙萌

南开大学 计算机学院

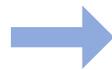


Wang X, Chou J-K, Chen W, et al. A Utility-aware Visual Approach for Anonymizing Multi-Attribute Tabular Data[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2018, 24(1): 351-360.

之前

2015年

- 每周读5篇VIS论文
- 参与工程项目



收获

- 理论知识积累
- 前端技术积累

之前

2015年末-2016年初

- 参与学姐的VIS投稿
 - 基于轨迹的人际关系分析



收获

- VIS投稿全流程经验积累

2015年

- 每周读5篇VIS论文
- 参与工程项目

之前

2015年末-2016年初

- 参与学姐的VIS投稿
 - 基于轨迹的人际关系分析

2015年

- 每周读5篇VIS论文
- 参与工程项目

2016年下半年

- 思考分析和隐私之间的矛盾
- 主导EuroVis的投稿
 - 不同地图信息对轨迹分析和位置隐私暴露的作用

收获

- 实践出困难

之前

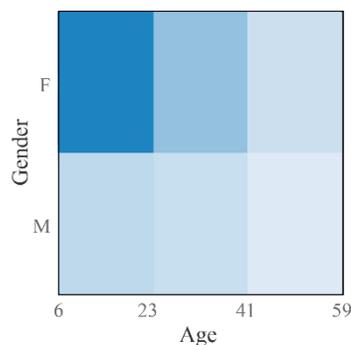


Idea初始化过程

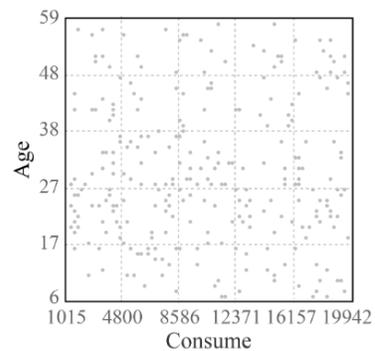


相关工作调研

- 隐私保护领域
 - 语义匿名模型 (k -anonymity, l -diversity, t -closeness, ...)
 - 差分隐私模型 (differential privacy)



语义匿名合并等价类



差分隐私加噪音

相关工作调研

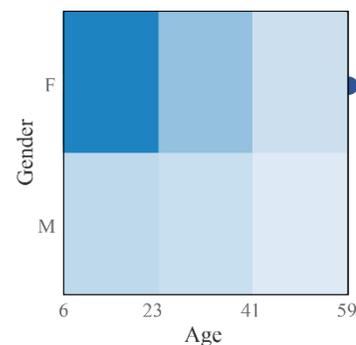
- 隐私保护领域

- 语义匿名模型
- 差分隐私模型

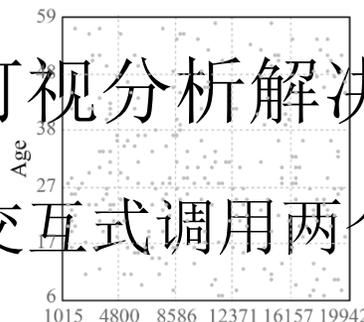


- 问题

- 两类方法各有优劣，需要互补
- 基于量化指标的实用性评估缺乏有效性



语义匿名合并等价类



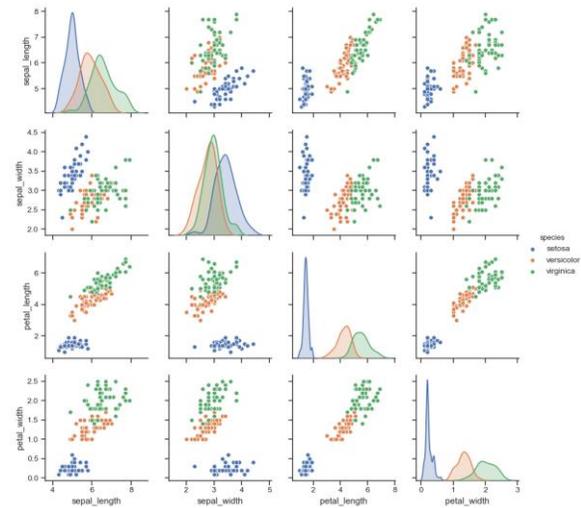
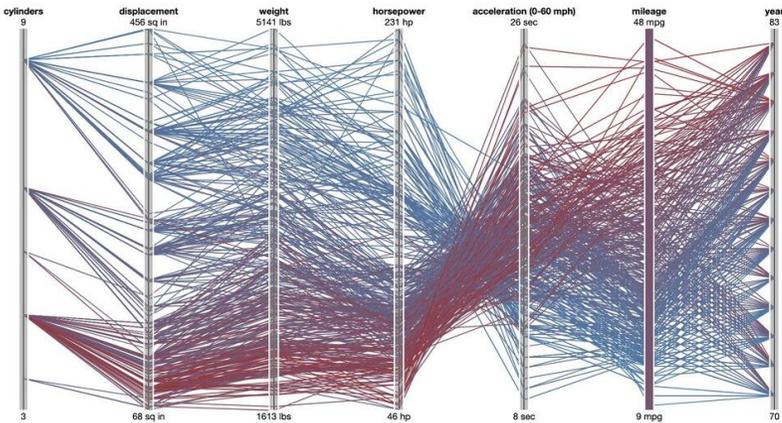
差分隐私加噪音

基于可视分析解决隐私和实用性的矛盾

- 可交互式调用两个隐私保护技术
- 可基于可视表达评估数据实用性

相关工作调研

- 可视分析领域
 - 多维数据（平行坐标，散点矩阵，…）



相关工作调研

- 可视分析领域

- 多维数据表达



- 问题

- 隐私：不能分析大量等价类
- 实用性：无法对比不同粒度的多维数据

相关工作调研

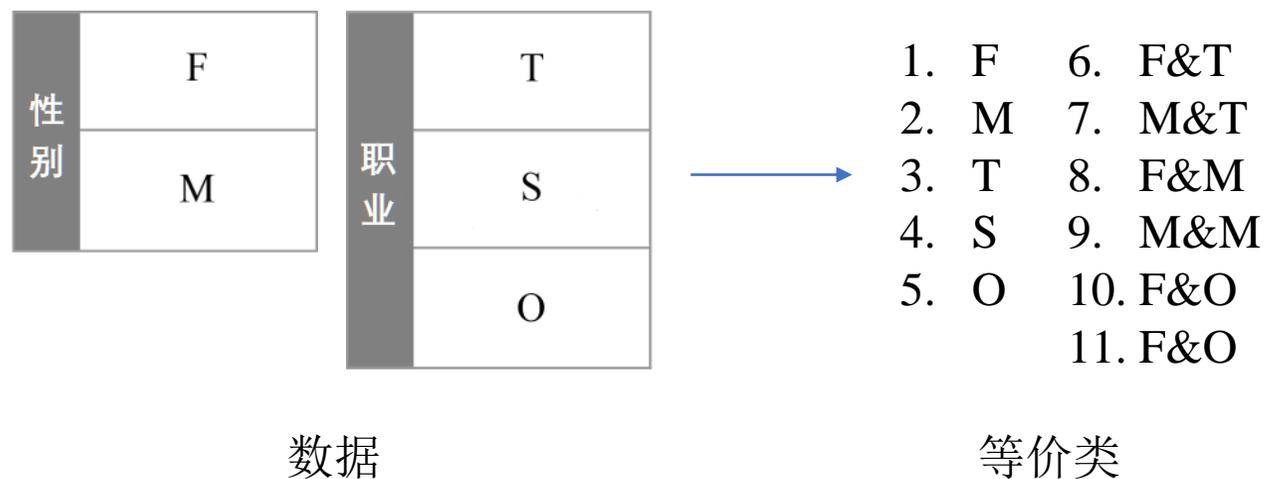
- 可视分析领域

- 多维数据表达



- 问题

- 隐私：不能分析大量等价类
- 实用性：无法对比不同粒度的多维数据



相关工作调研

- 可视分析领域

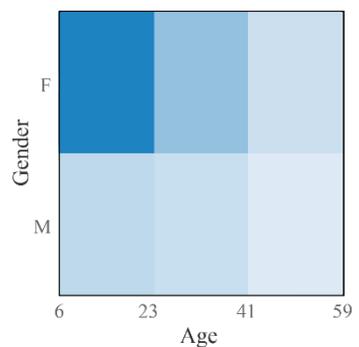


- 问题

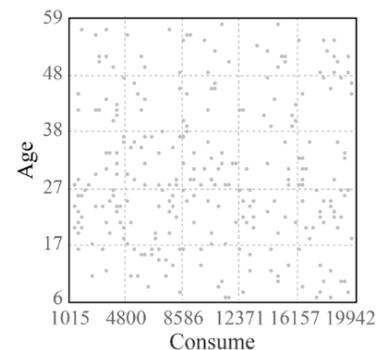
- 多维数据表达

- 隐私：不能分析大量等价类

- 实用性：无法对比不同粒度的多维数据

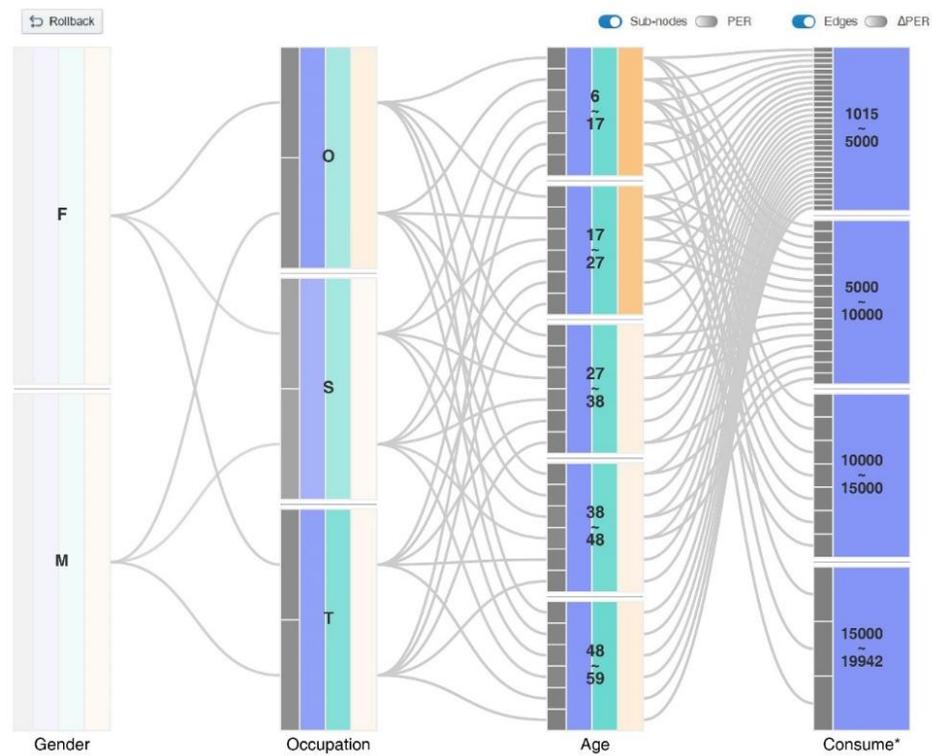


语义匿名合并等价类



差分隐私加噪音

隐私暴露风险树

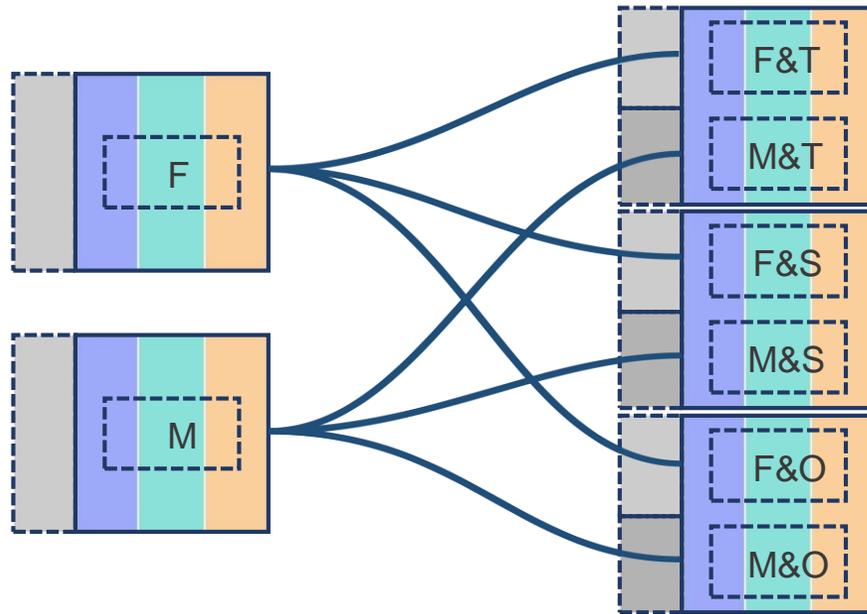


- 基于树结构，有层级地展示等价类

隐私暴露风险树-构造过程

性别	F
	M
职业	T
	S
	O

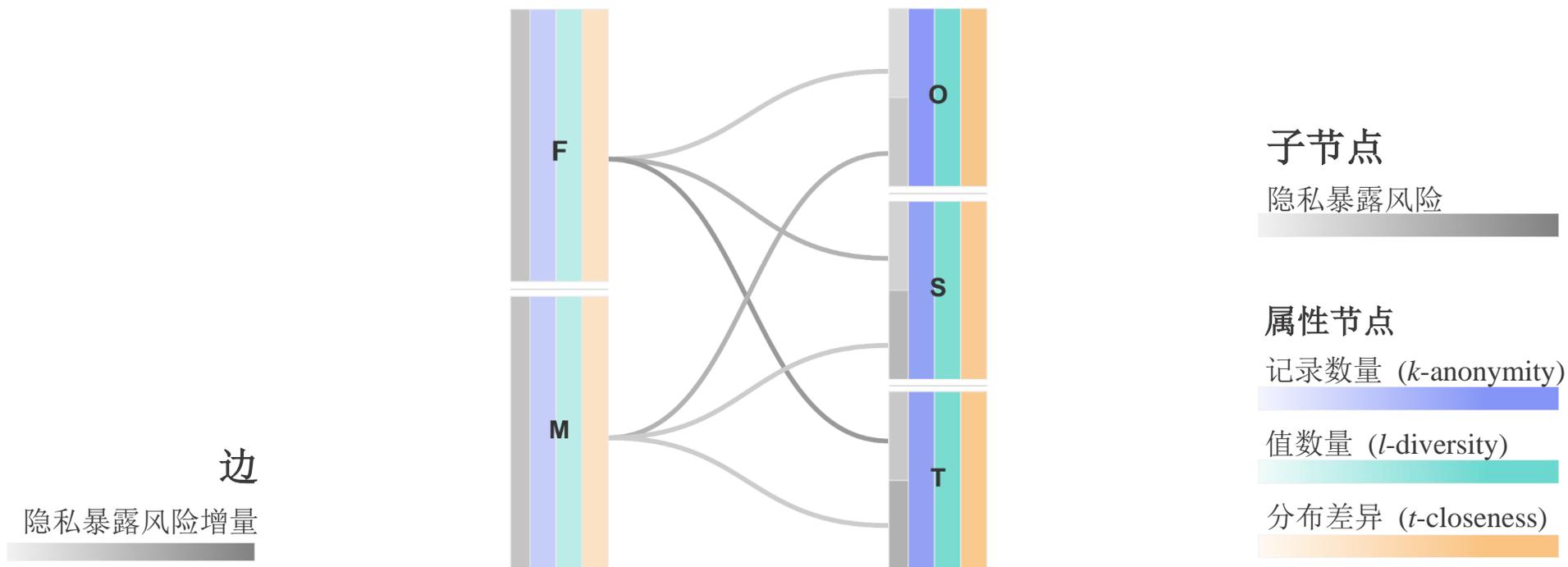
数据



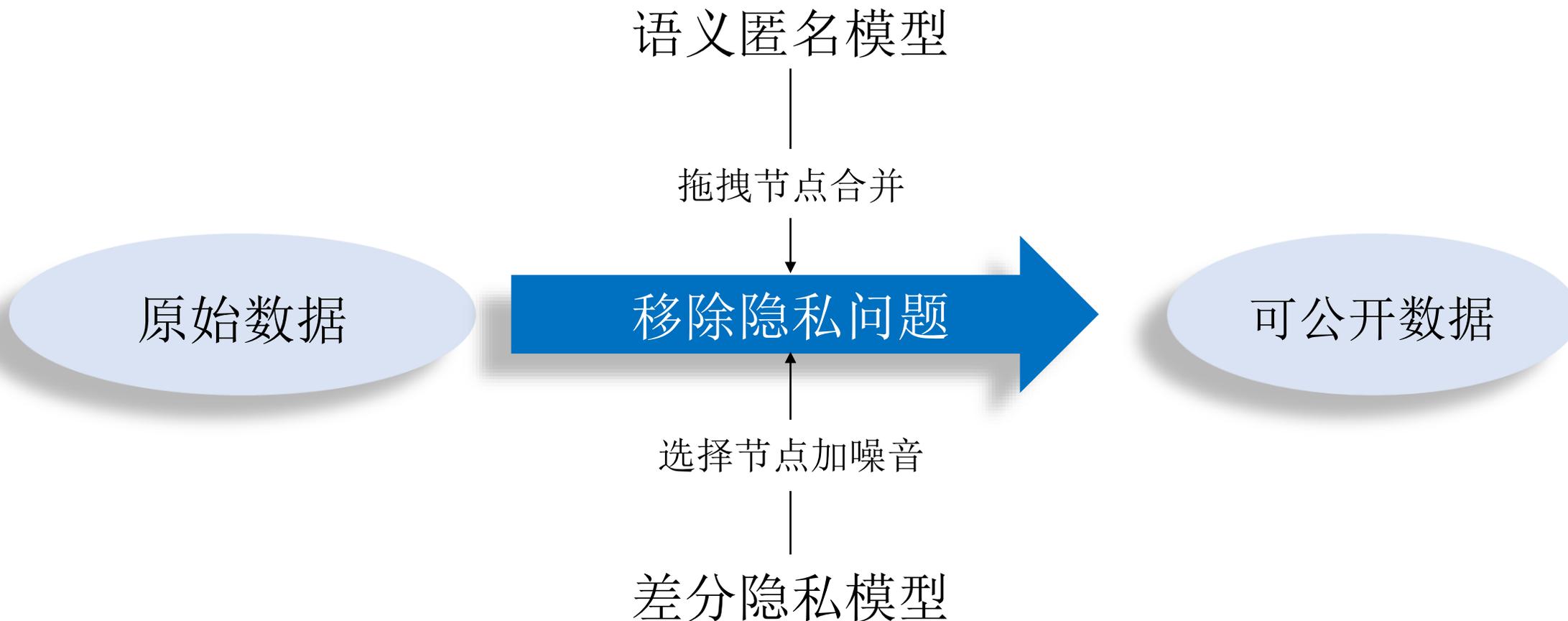
构造过程

节点
性节点

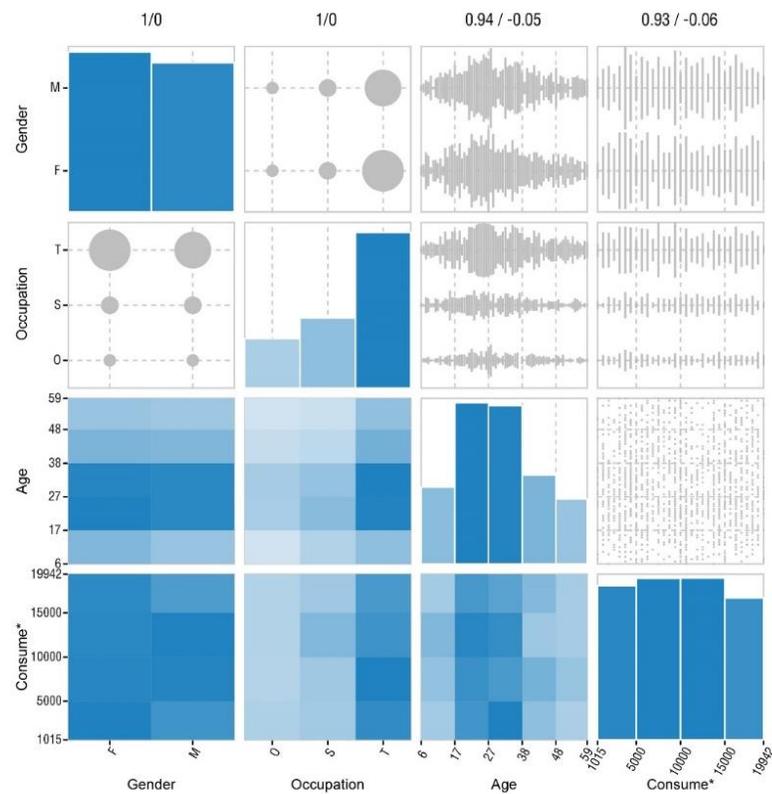
隐私暴露风险树-颜色编码



隐私暴露风险树-交互

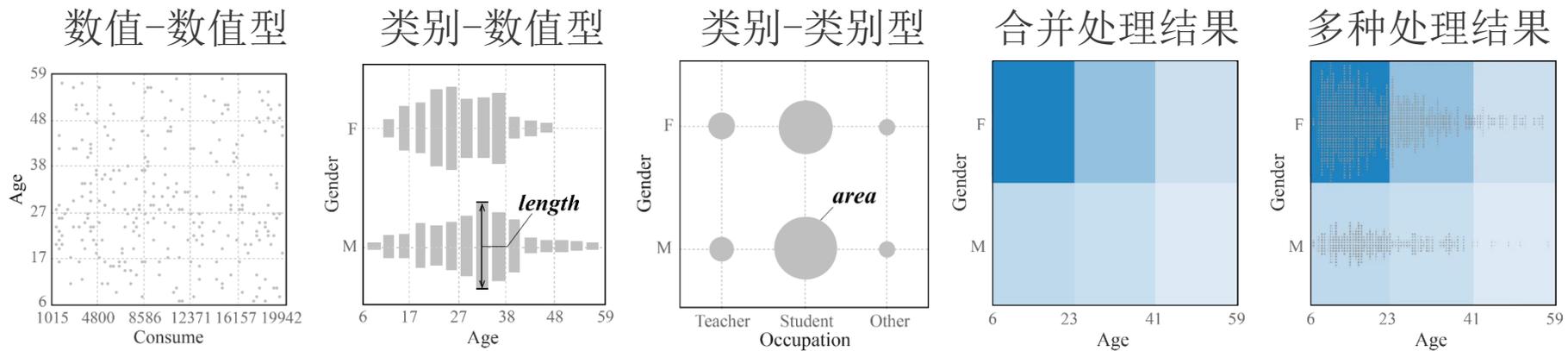


实用性度量矩阵



- 基于矩阵评估实用性损失

实用性度量矩阵-编码



原始数据——“散点图”

处理后数据——矩阵图

实用性比较

- 选择一个格子
- 在自定义的粒度下比较分布差异



隐私保护流程

导入数据

属性	是否必要	是否敏感	属性描述
年龄	✓	🔒	取值范围0~100
支出	✓	🔒	每月
性别	✓	🔒	男/女
姓名	+	🔒	姓, 名
职业	✓	🔒	老师/学生/其他



构建隐私暴露风险树

属性分组

层次排序

定义风险阈值

✓ 记录数量: 0 [slider] 100
✓ 属性值多样性: 0 [slider] 10

观察和调整

观察

调整

合并

添加噪音



实用性检验

比较结果

原始数据

处理后数据

差别步长 [slider]

数量 [slider]

导出数据

可视化

合并处理结果

年龄	支出	性别	职业
0-14	2k-3k	F	S
26-30	5k-8k	M	T
40-50	3k-4k	F	T or O
40-50	3k-4k	F	T or O

加噪音处理结果

年龄	支出	性别	职业
5	2109	F	S
28	8674	M	T
30	3300	F	O
56	4100	F	T

表格数据

提出方法的流程

- 发现问题，验证问题有解决的价值
- 调研已有工作，是否已有完善解决方案
- 结合问题/挑战，提出解决方法
- 验证解决方法有效

写作问题

- Abstract和Intro改重写了无数次

解决方案：先通过框架理逻辑，提炼问题和贡献

- 方法讲不明白，大佬无从改起

解决方案：不要从实现的角度描述方法，要按what-why-how的顺序解释方法

《Visualization Analysis & Design》——Tamara Munzner

设计问题

- 测试不同样式效率低

解决方案：使用css技巧，通过AI等设计软件快速迭代设计

- 颜色搭配找不好

解决方案：将好看保存图片存下来，从中提取主题色/直接套用经典配色

其他收获

- 多和其他人交流
 - 扩充发现问题的渠道
 - 学习别人技能的长处
 - 了解非作者的读后感

招收专硕

- 招收学院：南开大学计算机学院
- 研究方向：可视化与可视分析
- 毕业时间：2023年本科毕业的同学
- 联系方式：wangxumeng@nankai.edu.cn