

第五讲：可视化的评估

信息可视化

曹楠（教授），石洋（研究员）

<https://idvlab.com>

同济大学

常用的可视化评估方法

评估方法

设计准则比对 (Design Principles)

非正式用户调研 (Informal Study)

受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)

评估方法

- 设计准则比对 (Design Principles)
- 非正式用户调研 (Informal Study)
- 受控用户调研 (Controlled User Study)
- 实际应用及案例分析 (Case Study)
- 理论分析 (Theoretical Analysis)
- 基准化分析法 (Benchmark)

检验可视化的设计
是否符合一般性的设计准则

评估方法

设计准则比对 (Design Principles)

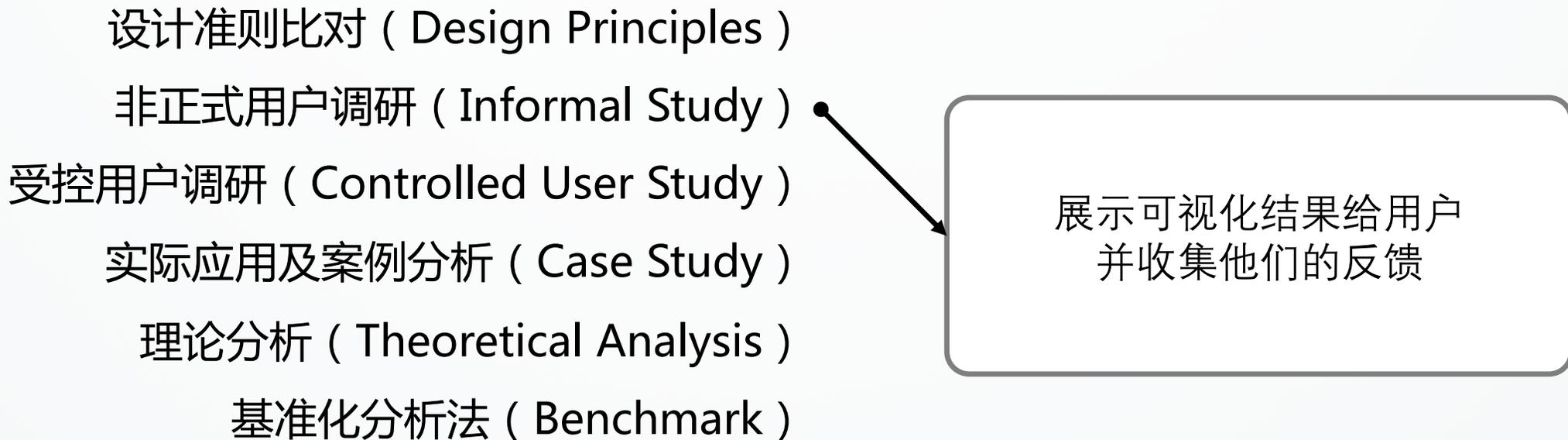
非正式用户调研 (Informal Study)

受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)



展示可视化结果给用户
并收集他们的反馈

评估方法

设计准则比对 (Design Principles)

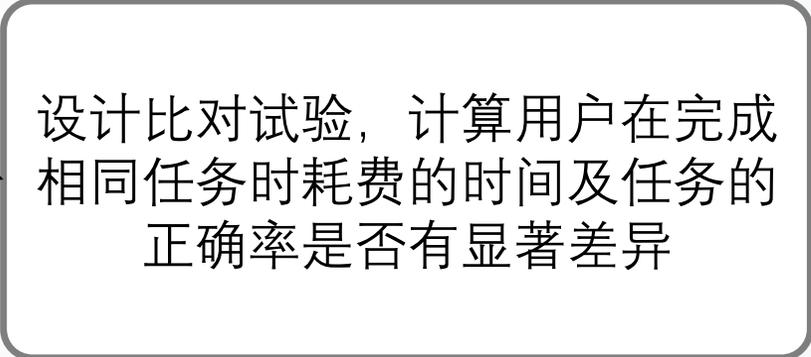
非正式用户调研 (Informal Study)

受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)



设计比对试验，计算用户在完成相同任务时耗费的时间及任务的正确率是否有显著差异

评估方法

设计准则比对 (Design Principles)

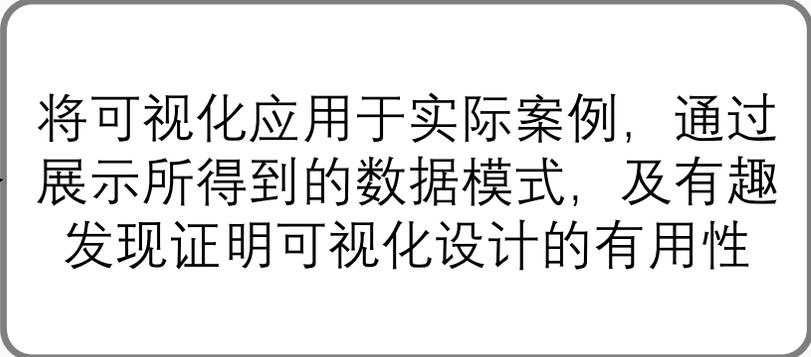
非正式用户调研 (Informal Study)

受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)



将可视化应用于实际案例，通过展示所得到的数据模式，及有趣发现证明可视化设计的有用性

评估方法

设计准则比对 (Design Principles)

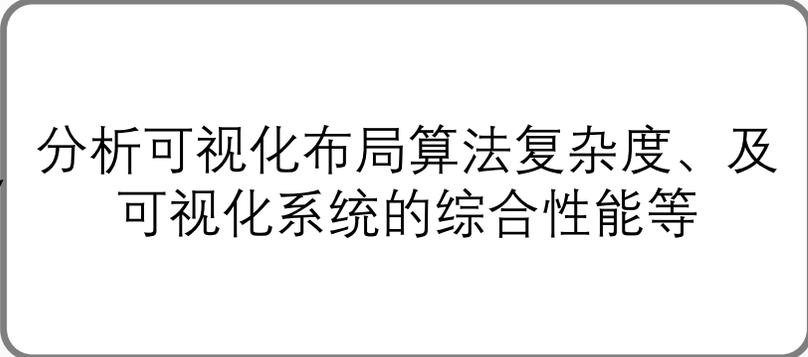
非正式用户调研 (Informal Study)

受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)



分析可视化布局算法复杂度、及
可视化系统的综合性能等

评估方法

设计准则比对 (Design Principles)

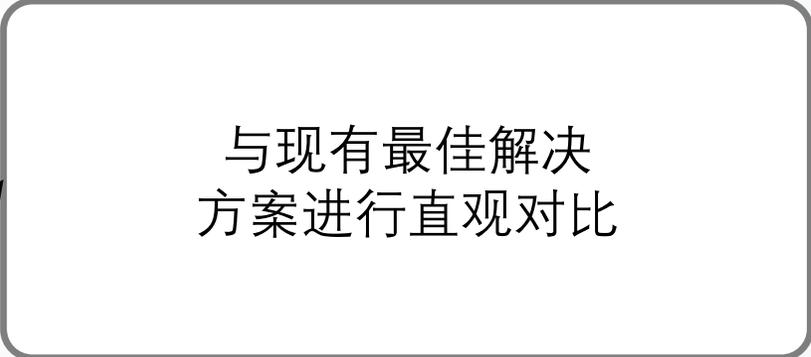
非正式用户调研 (Informal Study)

受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)



与现有最佳解决方案进行直观对比

用户研究（对比实验）

人机交互

概览

- 什么是用户研究?
 - 以经验检验假设
 - 评价界面
- 为什么要进行用户研究?
 - 确定“真实性”
 - 判断一个陈述是否正确

用户研究的组成部分

- 设计
 - 假设
 - 任务
 - 度量
 - 用户
- 步骤
- 数据分析
- 结论和讨论

设计

- 我们如何评估一个工具?
 - 假设
 - 你想发现什么?
 - 任务
 - 帮助你检验假设的指导方针
 - 度量
 - 你将如何评估工具的效率?
 - 用户
 - 谁将参加这项研究?

设计 – 假设

- 你要评估的陈述
- 通常用以下形式书写
 - “就**给定任务**而言，设计A比设计B**好**”
- 例如，**在寻找异常点的相关任务中，散点图矩阵比平行坐标更快**

设计 – 假设

- 确定自变量和因变量
 - **自变量** —— 可被实验者操纵 (例如, *不同的设计A或B*)
 - **因变量** —— 由自变量引起的变量 (例如, *响应时间或错误率*)

设计 – 假设

- **有待验证的假设**
 - 在寻找异常点的相关任务中，散点图矩阵比平行坐标更快
- **零假设:**
 - 基于散点图矩阵和平行坐标查找异常点时所耗时间相同
- **你的工作证明是不同！你需要拒绝零假设**

设计 - 用户

- 将要参加这项研究的人
- 选择一组合适的参与者
 - 他们的年龄、性别、文化、背景、专业.....
 - 越多样化，越平衡越好
 - 然而，这通常是很难实现的
- 数量
 - 越多越好
 - 招募（数量大于20人即可达到不错的效果）

设计 - 任务

- 任务被定义来指导用户研究
 - 例如，找出散点图视图中显示的所有异常点
- 在研究中，任务通常是在不同的条件下执行，这是由测试数据决定的
 - 例如，数据项数量，数据维度数量

设计 - 指标

- 在执行任务时的度量什么?
- 客观度量
 - 完成任务的时间
 - 错误率
 - ...
- 主观度量
 - 满意度
 - ...

两种研究类型

- 组内研究
 - 所有的用户都需要使用所有正在测试的工具来完成所有的任务
- 组间研究
 - 用户被分成多个小组，每个小组只使用一种测试工具
- 问: 利弊?

步骤

- 邀请用户
- 介绍不同的技术/设计，研究目的
- 简单教学系统的使用，让用户做一些练习
- 执行任务并记录结果
- 结束后，发放问卷/进行访谈

分析

- 需要对研究结果进行分析，从而判断这两种设计是否确实存在不同
- 需要解决的两个关键问题：
 - 当测量平均值相等时，是否意味着两种测试设计是相同的？
 - 当平均测量值是“不同的”，这种差异是源于偶然吗？
- 我们寻找的是“统计上不同”的结果

分析：T检验 (T-test)

- t检验是一种广泛用于比较两组样本均值的统计检验方法
- t检验用来评估两组数据的均值是否存在统计学上的显著差异

分析：T-test

- 单一样本t检验
 - 将总体的平均值与理论值进行比较
- 非配对双样本t检验
 - 比较两个独立样本的均值（用于分析组间研究的结果）
- 配对t检验
 - 比较两组相关样本的均值（用于分析组内研究的结果）

当实验条件多于两种时？

- 方差分析 (ANOVA)
- 阅读作业
 - <http://www.statisticssolutions.com/manova-analysis-anova/>

SPSS教程

- 独立t检验
 - <https://www.youtube.com/watch?v=l0TMKRkpuNU>
- 配对样本t检验
 - <https://www.youtube.com/watch?v=eVZi-62uTTg>
- 单向ANOVA
 - <https://www.youtube.com/watch?v=jYn5Jv7Gh4s>
- 重复检验ANOVA
 - https://www.youtube.com/watch?v=6T6dvrwDe_U

第五讲：可视化的评估

信息可视化

曹楠（教授），石洋（研究员）

<https://idvlab.com>

同济大学