

# 可视化设计的评估方法

## 信息可视化

石洋 副教授

<https://idvxlabs.com>

同济大学

# 课程大纲

- 可视化设计的评估方法
- 受控用户调研
- 工作坊

# 可视化设计的评估方法

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)

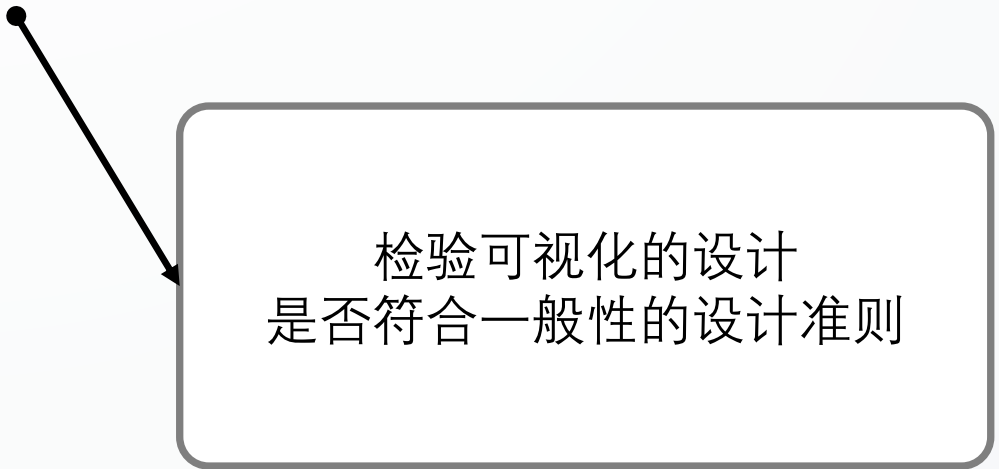
理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)

工作坊 (Workshop)

# 可视化设计的评估方法

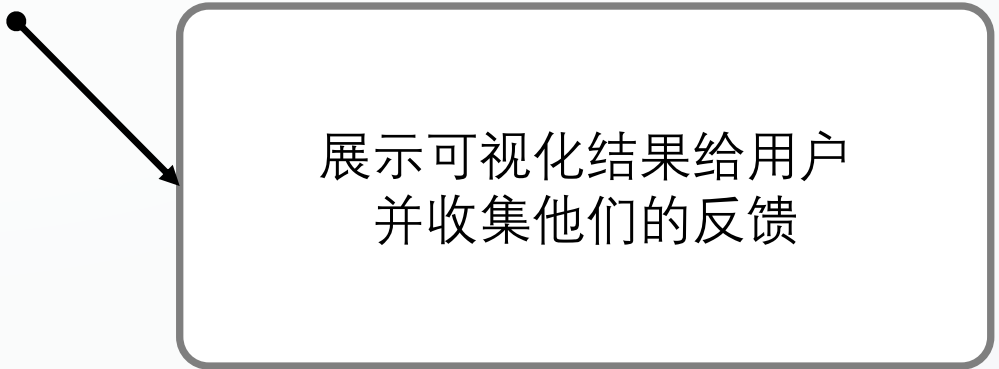
设计准则比对 (Design Principles)  
非正式用户调研 (Informal Study)  
受控用户调研 (Controlled User Study)  
实际应用及案例分析 (Case Study)  
理论分析 (Theoretical Analysis)  
基准化分析法 (Benchmark)  
工作坊 (Workshop)



检验可视化的设计  
是否符合一般性的设计准则

# 可视化设计的评估方法

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)  
工作坊 (Workshop)



展示可视化结果给用户  
并收集他们的反馈

# 可视化设计的评估方法

设计准则比对 (Design Principles)

非正式用户调研 (Informal Study)

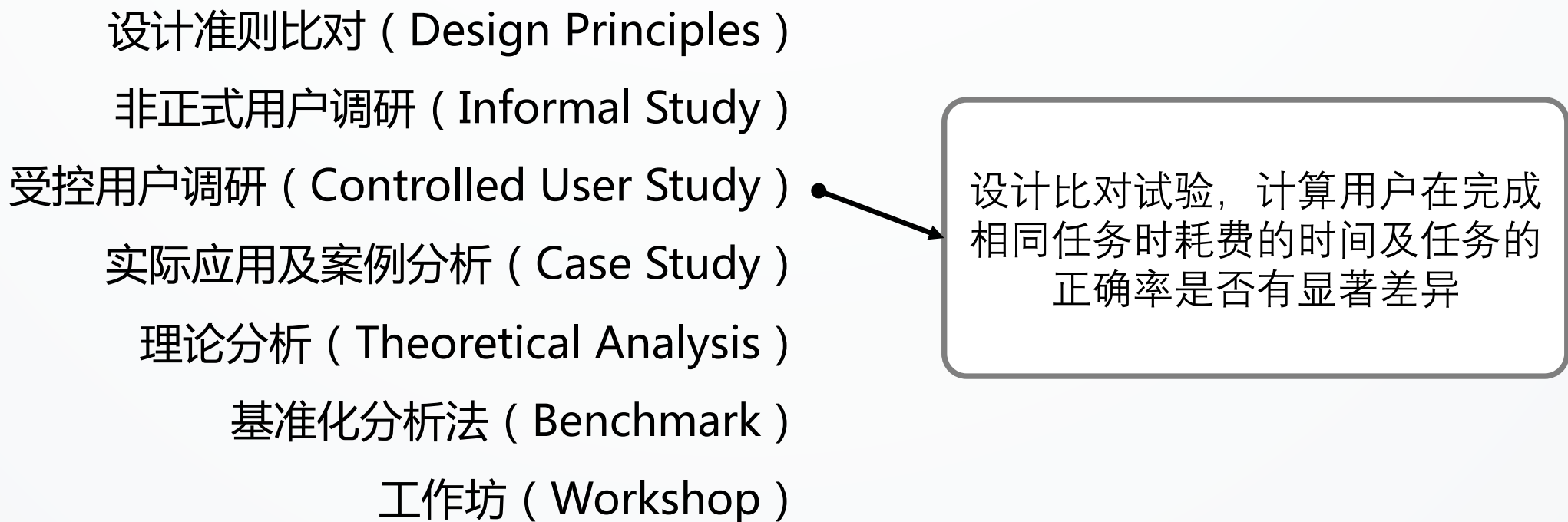
受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)

工作坊 (Workshop)



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

# 可视化设计的评估方法

设计准则比对 (Design Principles)

非正式用户调研 (Informal Study)

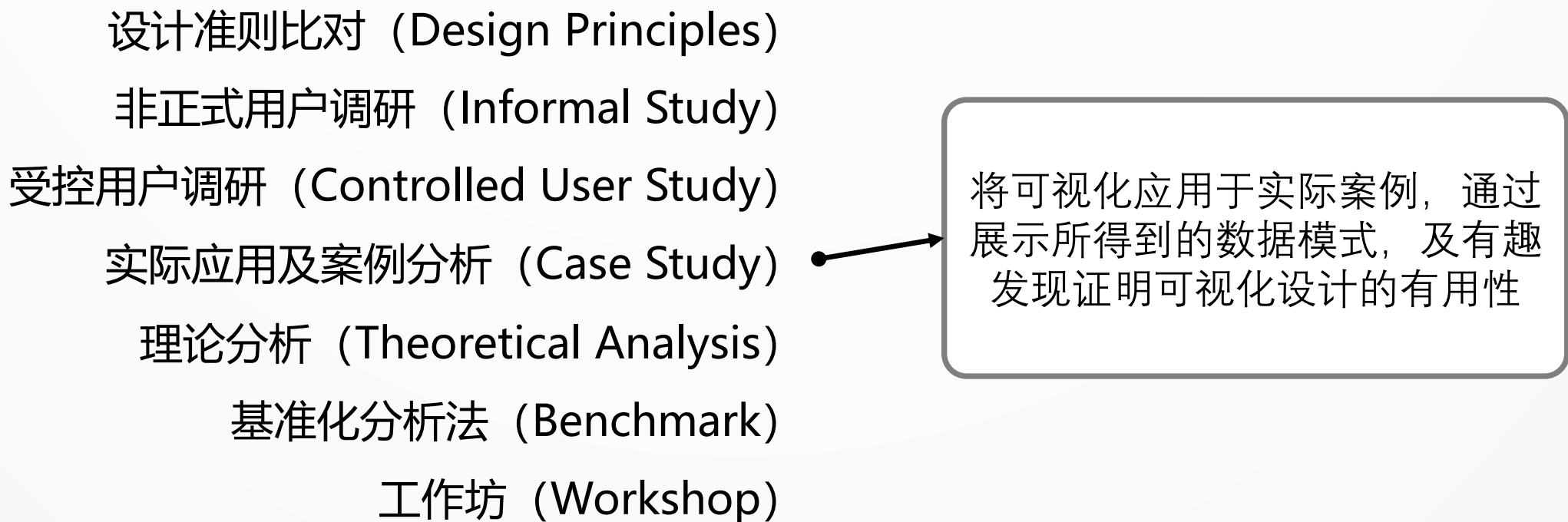
受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)

工作坊 (Workshop)



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

# 可视化设计的评估方法

设计准则比对 (Design Principles)

非正式用户调研 (Informal Study)

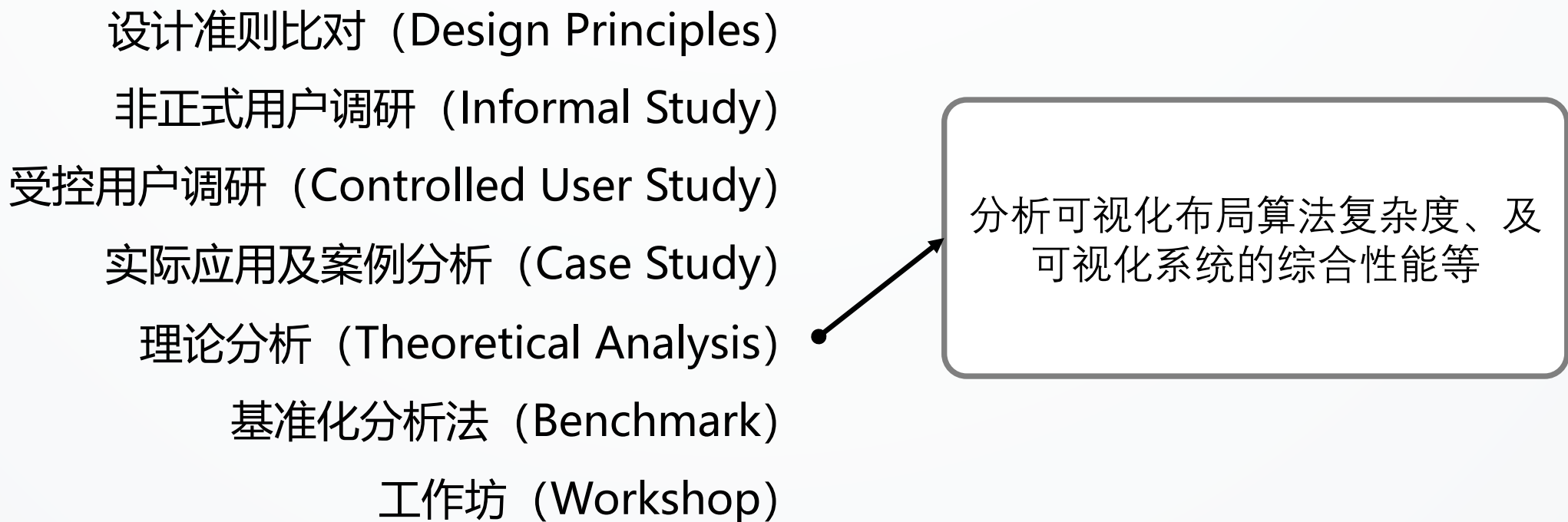
受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)

工作坊 (Workshop)



分析可视化布局算法复杂度、及  
可视化系统的综合性能等



# 可视化设计的评估方法

设计准则比对 (Design Principles)

非正式用户调研 (Informal Study)

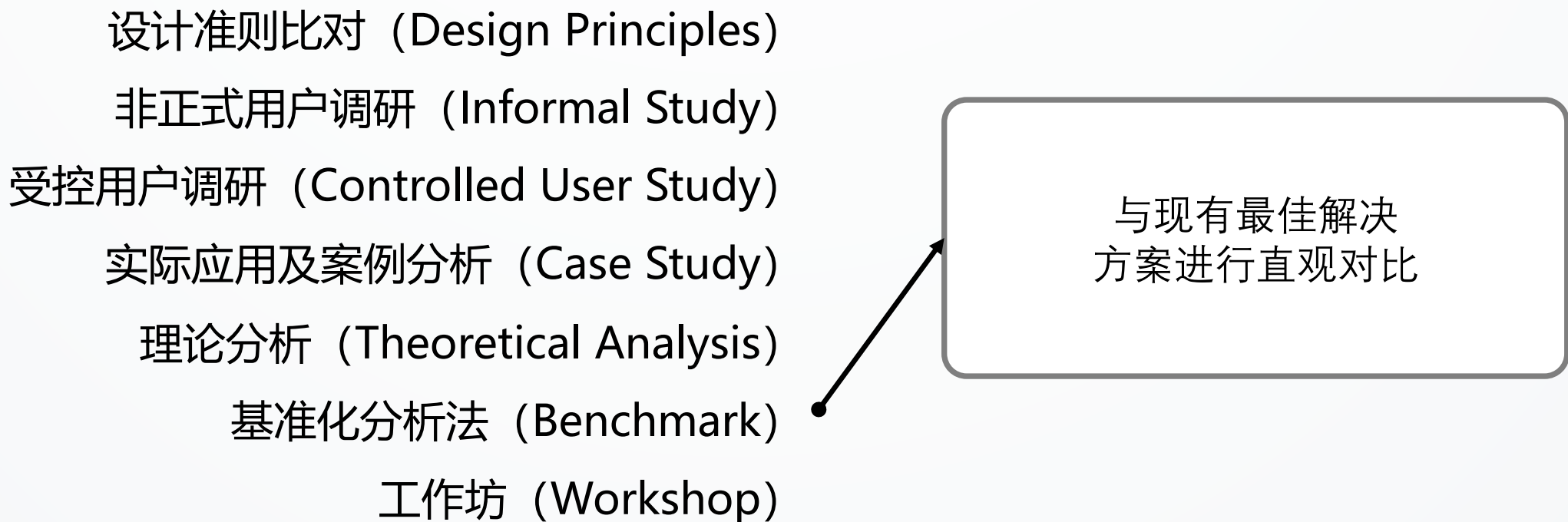
受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)

工作坊 (Workshop)



与现有最佳解决方案进行直观对比

# 可视化设计的评估方法

设计准则比对 (Design Principles)

非正式用户调研 (Informal Study)

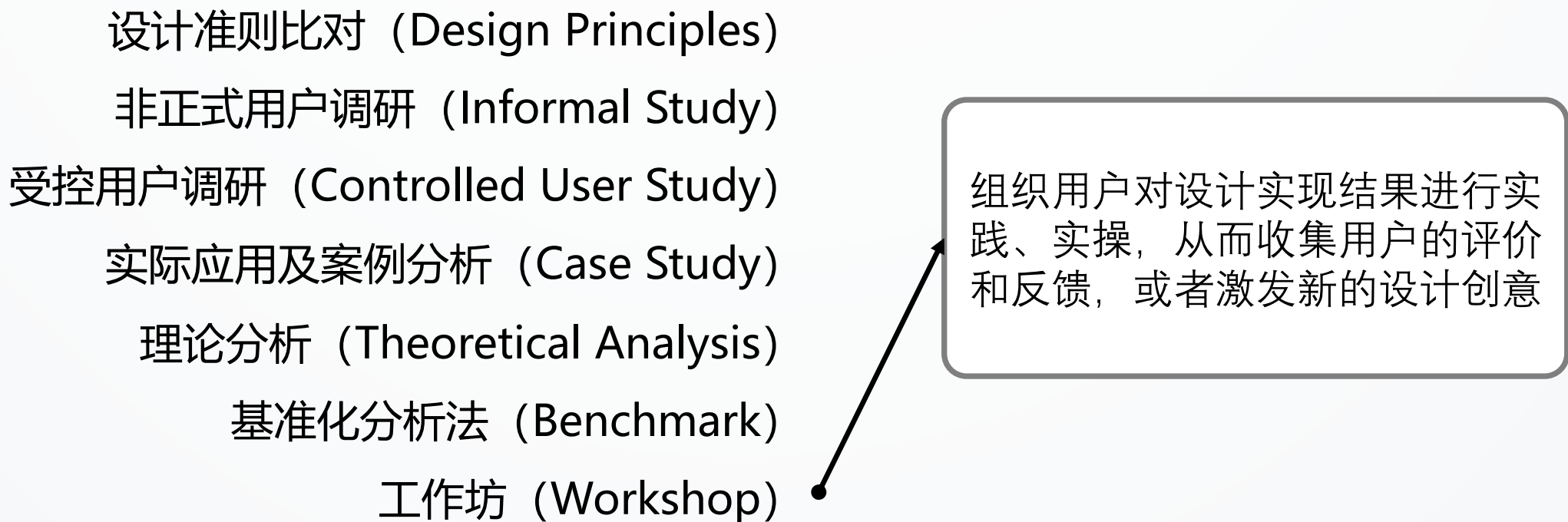
受控用户调研 (Controlled User Study)

实际应用及案例分析 (Case Study)

理论分析 (Theoretical Analysis)

基准化分析法 (Benchmark)

工作坊 (Workshop)



组织用户对设计实现结果进行实践、实操，从而收集用户的评价和反馈，或者激发新的设计创意

# 课程大纲

- 可视化设计的评估方法
- 受控用户调研
- 工作坊

# 概览

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

# 用户调研的组成部分

- 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费时间相同
- 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调研
  - 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)

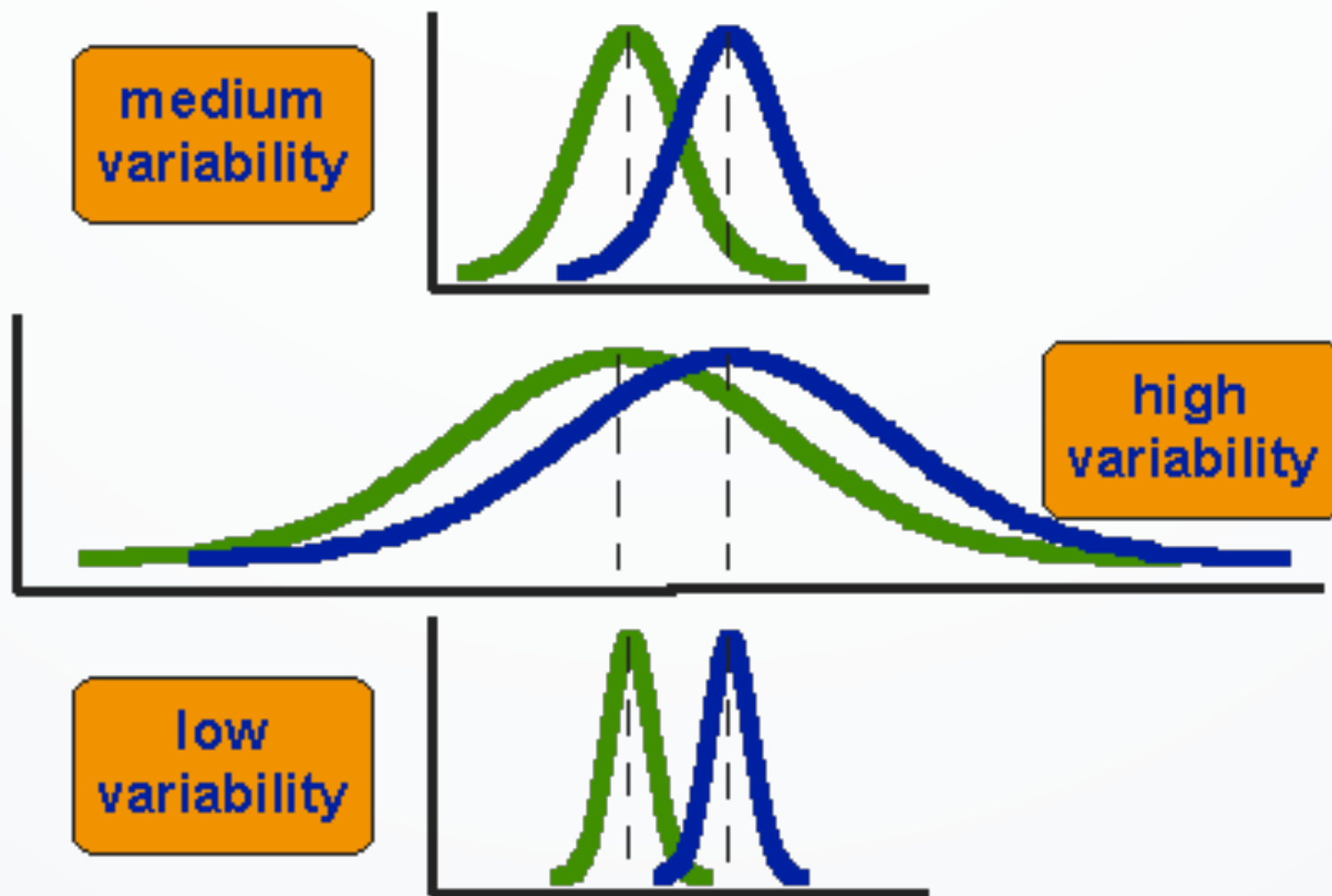
- t检验是一种广泛用于比较两组样本均值的统计检验方法
- t检验用来评估两组数据的均值是否存在统计学上的显著差异



## 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）

# 非配对双样本t检验



## 非配对双样本t检验

- A 和 B 表示需要比较的两组样本

$$t = \frac{m_A - m_B}{\sqrt{\frac{S^2}{n_A} + \frac{S^2}{n_B}}}$$

两组样本平均值之间的差异

两组样本的变化  
(数据是如何分散的)

- $m_A$  和  $m_B$  分别表示组 A 和 B 的平均值

$$S^2 = \frac{\sum (x - m_A)^2 + \sum (x - m_B)^2}{n_A + n_B - 2}$$

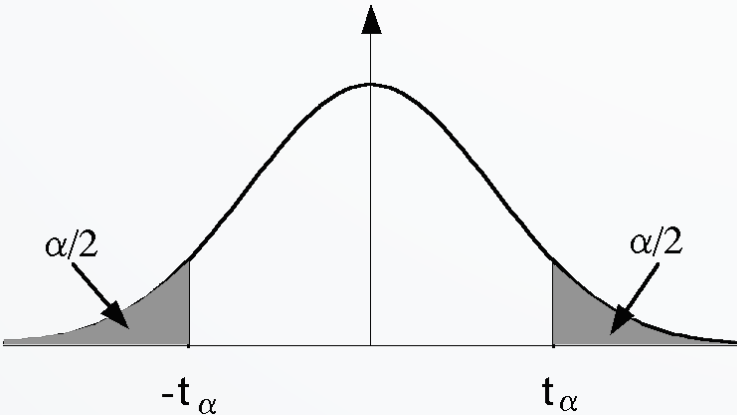
两组样本方差的估计

- $n_A$  和  $n_B$  分别表示组 A 和 B 的大小

自由度  $df = n_A + n_B - 2$

# 非配对双样本t检验

一旦确定了 t 检验统计值，就需要查找 t 检验表中与选择的显著性水平 alpha (0.5%) 相对应的 t 分布临界值。

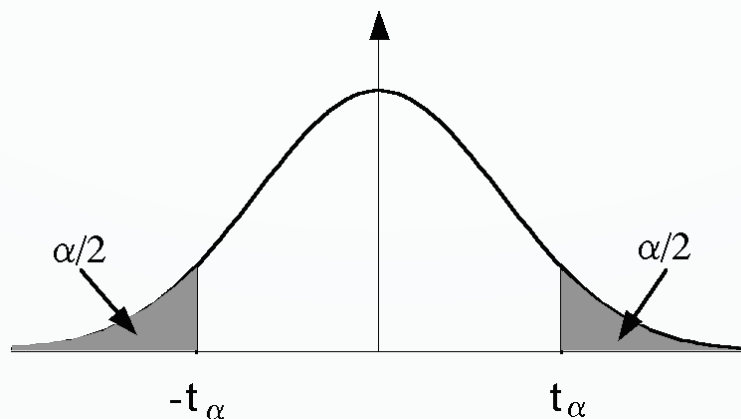


df/a	0.9	0.5	0.3	0.2	0.1	0.05	0.02	0.01	0.001
1	0.158	1	2	3.078	6.314	12.706	31.821	64	637
2	0.142	0.816	1.386	1.886	2.92	4.303	6.965	10	31.598
3	0.137	0.765	1.25	1.638	2.353	3.182	4.541	5.841	12.929
4	0.134	0.741	1.19	1.533	2.132	2.776	3.747	4.604	8.61
5	0.132	0.727	1.156	1.476	2.015	2.571	3.365	4.032	6.869
6	0.131	0.718	1.134	1.44	1.943	2.447	3.143	3.707	5.959
7	0.13	0.711	1.119	1.415	1.895	2.365	2.998	3.499	5.408
8	0.13	0.706	1.108	1.397	1.86	2.306	2.896	3.355	5.041
9	0.129	0.703	1.1	1.383	1.833	2.263	2.821	3.25	4.781
10	0.129	0.7	1.093	1.372	1.812	2.228	2.764	3.169	4.587
11	0.129	0.697	1.088	1.363	1.796	2.201	2.718	3.106	4.437
12	0.128	0.695	1.083	1.356	1.782	2.179	2.681	3.055	4.318

例如，当 df=10, t=2.228, 概率是 alpha=0.05

## 非配对双样本t检验

- 如果  $t$  检验统计值的绝对值 ( $|t|$ ) 大于临界值, 那么差异是显著的, 否则就不显著。



- 显著性水平 (p值) 对应了  $t$  检验表中  $|t|$  值所指向的概率。

# 配对t检验

- 比较两个相关样本的均值
- d是两组样本中每对样本之间的差
- 差异 d 的均值会和 0 比较。如果两组样本之间有显著差异，那么就期望 d 的均值远离 0。

$$t = \frac{m}{s/\sqrt{n}}$$

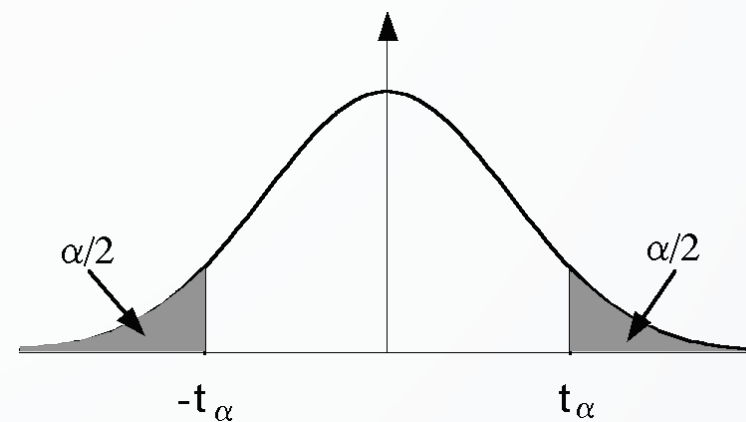
m 和 s 分别是差异 d 的均值和标准差。  
n 是 d 的样本大小。

$$df = n - 1$$

配对 t 检验的自由度为样本数量减一。

# 显著性

- 显著的含义是什么？
  - 你可以相对有信心地认为这并不是偶然发生的。
- 体现在：
  - 样本的个数 ( $n$ )
  - P 值 (例如,  $< 0.05$  或者  $> 0.05$ )
  - 方差/标准差
  - 均值



# 分析-样例

完成任务的时间 (秒)

VIS 1	VIS 2
1.0	2.0
1.2	1.8
1.1	2.3
1.5	2.7
0.9	1.6
1.14	2.08

VIS1 是否显著优于 VIS2?

1. 在组间研究的情况下?
2. 在组内研究的情况下?

df/a	0.9	0.5	0.3	0.2	0.1	0.05
1	0.158	1	2	3.078	6.314	12.706
2	0.142	0.816	1.386	1.886	2.92	4.303
3	0.137	0.765	1.25	1.638	2.353	3.182
4	0.134	0.741	1.19	1.533	2.132	2.776
5	0.132	0.727	1.156	1.476	2.015	2.571
6	0.131	0.718	1.134	1.44	1.943	2.447
7	0.13	0.711	1.119	1.415	1.895	2.365
8	0.13	0.706	1.108	1.397	1.86	2.306



# 组间研究的解答

$$\begin{aligned}
 S^2 &= \frac{\sum (x - m_A)^2 + \sum (x - m_B)^2}{5 + 5 - 2} \\
 &= \frac{0.14^2 + 0.06^2 + 0.04^2 + 0.36^2 + 0.26^2 + 0.08^2 + 0.28^2 + 0.22^2 + 0.62^2 + 0.48^2}{8} \\
 &= \frac{0.96}{8} = 0.2 \\
 t &= \frac{m_A - m_B}{\sqrt{\frac{S^2}{n_A} + \frac{S^2}{n_B}}} \\
 &= \frac{1.14 - 2.08}{\sqrt{\frac{0.04}{5} + \frac{0.04}{5}}} = \frac{-0.94}{0.1265} \approx -7.431 > 2.306 \text{ (significant)}
 \end{aligned}$$

$$df = 5 + 5 - 2 = 8$$

df/a	0.9	0.5	0.3	0.2	0.1	0.05
1	0.158	1	2	3.078	6.314	12.706
2	0.142	0.816	1.386	1.886	2.92	4.303
3	0.137	0.765	1.25	1.638	2.353	3.182
4	0.134	0.741	1.19	1.533	2.132	2.776
5	0.132	0.727	1.156	1.476	2.015	2.571
6	0.131	0.718	1.134	1.44	1.943	2.447
7	0.13	0.711	1.119	1.415	1.895	2.365
8	0.13	0.706	1.108	1.397	1.86	2.306

两组样本之间有显著的差异

# 组内研究的解答

VIS 1	VIS 2	d
1.0	2.0	1.0
1.2	1.8	0.6
1.1	2.3	1.2
1.5	2.7	1.2
0.9	1.6	0.7
1.14	2.08	0.94

$$t = \frac{m}{s/\sqrt{n}} = \frac{0.94}{0.25/2.24} = 8.4224 > 2.776$$

$$df = 5 - 1 = 4$$

df/a	0.9	0.5	0.3	0.2	0.1	0.05
1	0.158	1	2	3.078	6.314	12.706
2	0.142	0.816	1.386	1.886	2.92	4.303
3	0.137	0.765	1.25	1.638	2.353	3.182
4	0.134	0.741	1.19	1.533	2.132	2.776
5	0.132	0.727	1.156	1.476	2.015	2.571
6	0.131	0.718	1.134	1.44	1.943	2.447
7	0.13	0.711	1.119	1.415	1.895	2.365
8	0.13	0.706	1.108	1.397	1.86	2.306

两组样本之间有显著的差异

## 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](https://www.youtube.com/watch?v=6T6dvrwDe_U)

# 课程大纲

- 可视化设计的评估方法
- 受控用户调研
- 工作坊

# 概览

- 什么是工作坊?
  - 将参与者放置于比较真实的体验空间中
  - 组织参与者进行实践、实操，并鼓励放大参与者的主观能动性
- 为什么要举办工作坊?
  - 收集用户的评价和反馈
  - 激发新的设计创意

# 工作坊的组成部分

- 工作坊设计
  - 人员招募
  - 主要任务
  - 统筹分工
- 工作坊流程
- 结果分析方法

# 设计

- 我们如何设计一个工作坊？
  - 人员招募
    - 希望招募的参与者是谁？
  - 主要任务
    - 活动的主要任务是什么？
  - 统筹分工
    - 如何统筹和分工，来保证工作坊顺利进行？



# 设计 – 人员招募

- 人员数量
  - 一场比较活跃且高效的工作坊，人数应当控制在 30 人以内
- 目标用户
  - 需要在招募要求中写明对应的要求
    - 比如，要求设计专业背景，擅长某某工具等
  - 在任何的用户招募中，招募到过于类似的用户群体都是不利的
    - 比如，即使是招募专业的设计师，我们也需要尽量保证他们来自不同的学校、公司，有着不同的年龄和性别。

# 设计 – 主要任务

典型的工作坊任务包括：小组头脑风暴、动手绘图、讨论和辩论等等

- 准备 “预案”
  - 例如，希望参与者评估哪些作品?讨论哪些关键问题?哪些问题先讨论，哪些问题后讨论?
- 时间
  - 多数工作坊持续时间为2-4小时
  - 活动中加入一些 “刺激物”，如休息时间、社交时间
  - 活动现场精准控制每个环节时间

# 设计 – 统筹分工

- 为了保证以上环节的顺利进行，活动的组织者需要提前做好规划和分工
- 活动前：确定场地，发布活动招募信息，采购物料，布置现场，测试设备，联系参与者，准备好待评估素材，小组讨论提纲等
- 活动中：分配主持人、摄影、场务、运营等多个角色，确保每个环节都有人负责。

## 工作坊活动组织和执行清单示例

### 1. 确认场地

- ☐ 寻找可使用场地
- ☐ 确定场地大小、可容纳人数、可用设备
- ☐ 确定活动时间，租借当日场地
- ☐ 提前测试设备可用性（如投影仪、音响、网络等）
- ☐ 提前确定需要自带的设备（如转接头、激光笔等）

### 2. 活动招募

- ☐ 制作宣传海报、文案
- ☐ 制作报名问卷
- ☐ 在微信公号、微博等渠道发布活动信息
- ☐ 审核报名人员，分组（如需要）
- ☐ 通知入选人员
- ☐ 对活动当日的交通信息、需携带的设备等进行通知

### 3. 现场执行

- ☐ 主持人PPT
- ☐ 破冰游戏
- ☐ 热身任务（例如一个数据故事片段）
- ☐ 需要评估的数据故事（如需用电脑打开，测试打开速度）
- ☐ 小组讨论提纲

核心物料

- ☐ 现场签到表
- ☐ 草稿纸、彩色笔、便利贴等文具
- ☐ 零食、饮料等（如需要）
- ☐ 相机/摄像机、三脚架/手机支架
- ☐ 奖品/纪念品（如需要）

辅助物料

- ☐ 主持人/协调人（负责主持、串场，控制总体流程）
- ☐ 前期场务（负责布置现场，如摆放桌椅、物料、张贴海报等）
- ☐ 中期场务（负责补充物料、提供工具支持等）
- ☐ 后期场务（负责清理现场、回收物料等）
- ☐ 摄影（负责拍照、摄像、录音等）
- ☐ 社群运营（负责在群聊中发布通知、反馈活动照片等）
- ☐ 机动人员（负责答疑、计时等）

人员分工

工作坊活动组织和执行清单示例

# 工作坊流程

- 开场（主持人暖场）
  - 介绍主办方
  - 介绍活动主题，确保所有的参与者都真正理解了工作坊的核心议题
  - 引导所有参与者进行破冰
  - 引导参与者形成小组

# 工作坊流程

- 小组讨论
  - 每个小组的组员人数不超过 12 个人
  - 由一名主持人来组织和协调
  - 主持人副手帮忙记录时间和讨论关键词

(在讨论过程中，主办方可以提供一些纸笔、便利贴等用品，帮助参与者捋清自己的想法和观点)

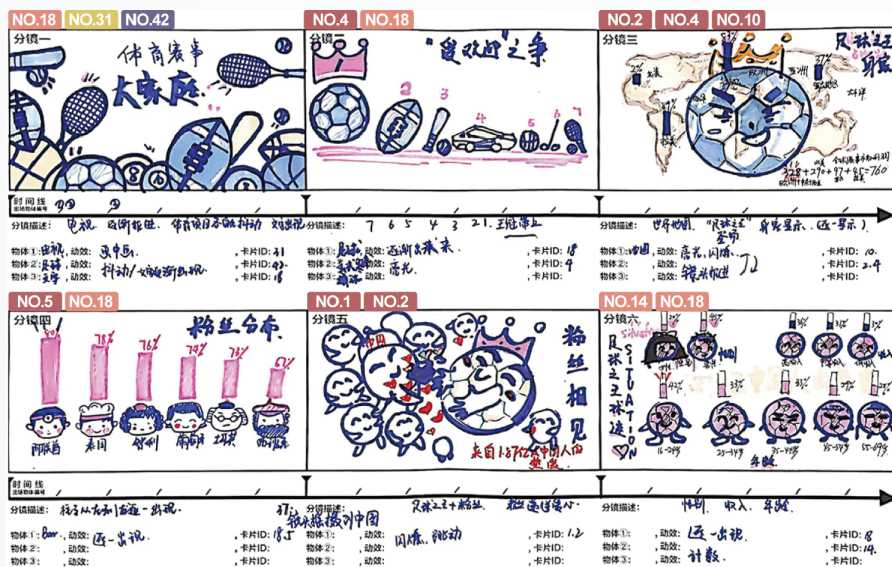


# 工作坊流程

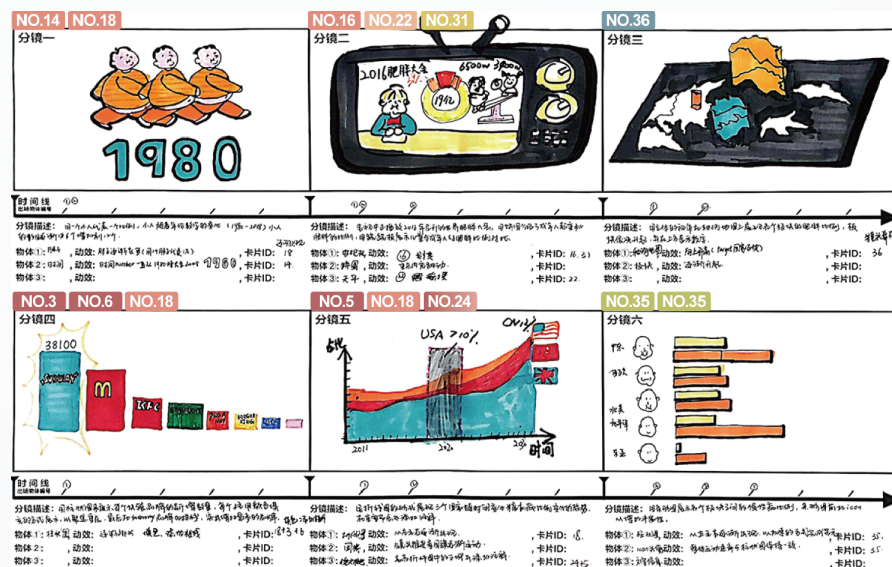
- 分享&总结
  - 召集所有小组，由各组推举代表，对刚刚的讨论做一个分享
  - 主持人进行总结，并宣布活动结束。
  - 整个活动过程需要全程录音/录像。

# 工作坊案例

- 此工作坊包括两个阶段：
  - 设计阶段，参与者创建数据视频的故事板。
  - 制作阶段，根据他们所画的故事板制作数据视频。



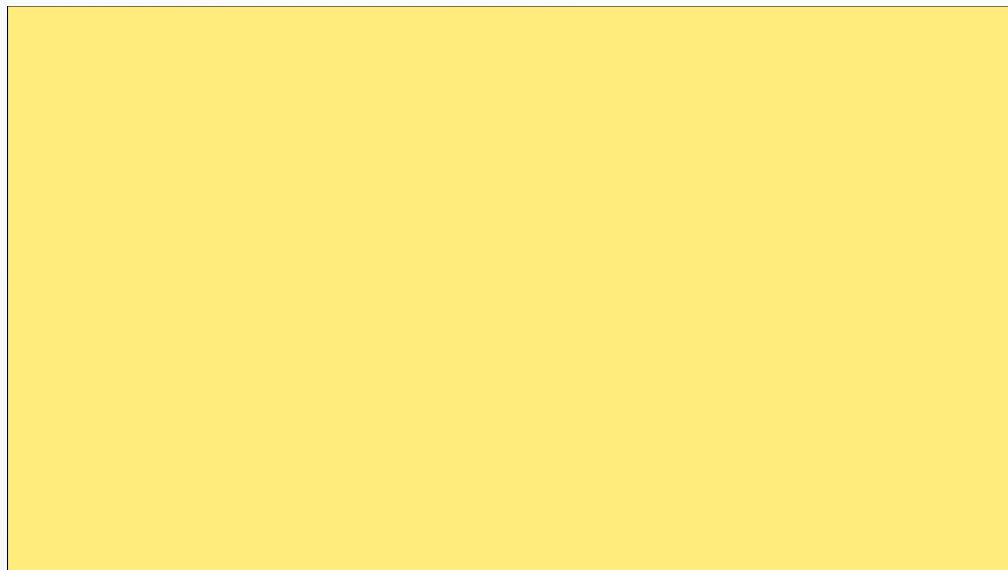
The persona of football fans by Lin Wu and Quan Yuan



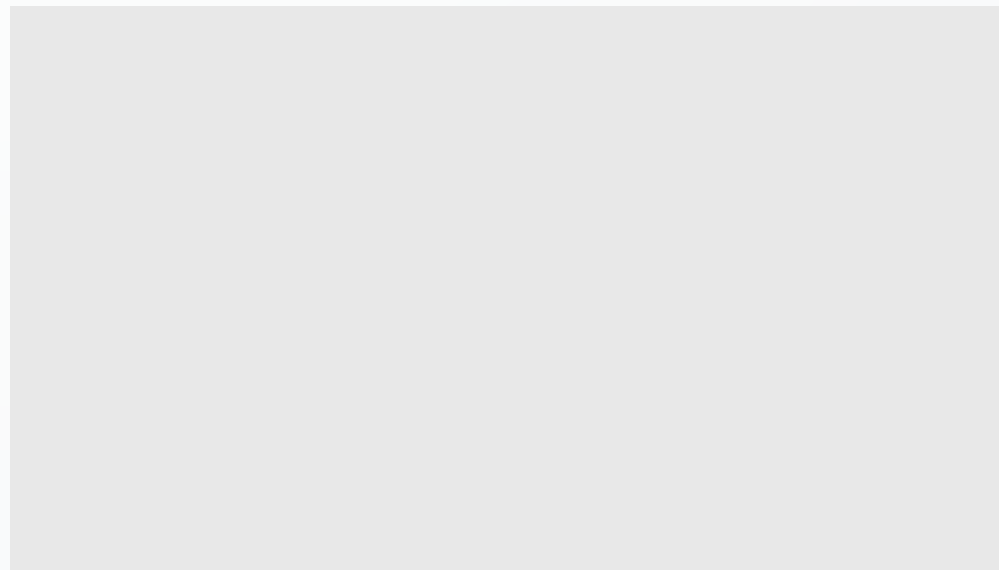
Global trends in obesity by Mingxue Ma and Hanlin Yu

# 工作坊案例

- 此工作坊包括两个阶段:
  - 设计阶段, 参与者创建数据视频的故事板。
  - 制作阶段, 根据他们所画的故事板制作数据视频。



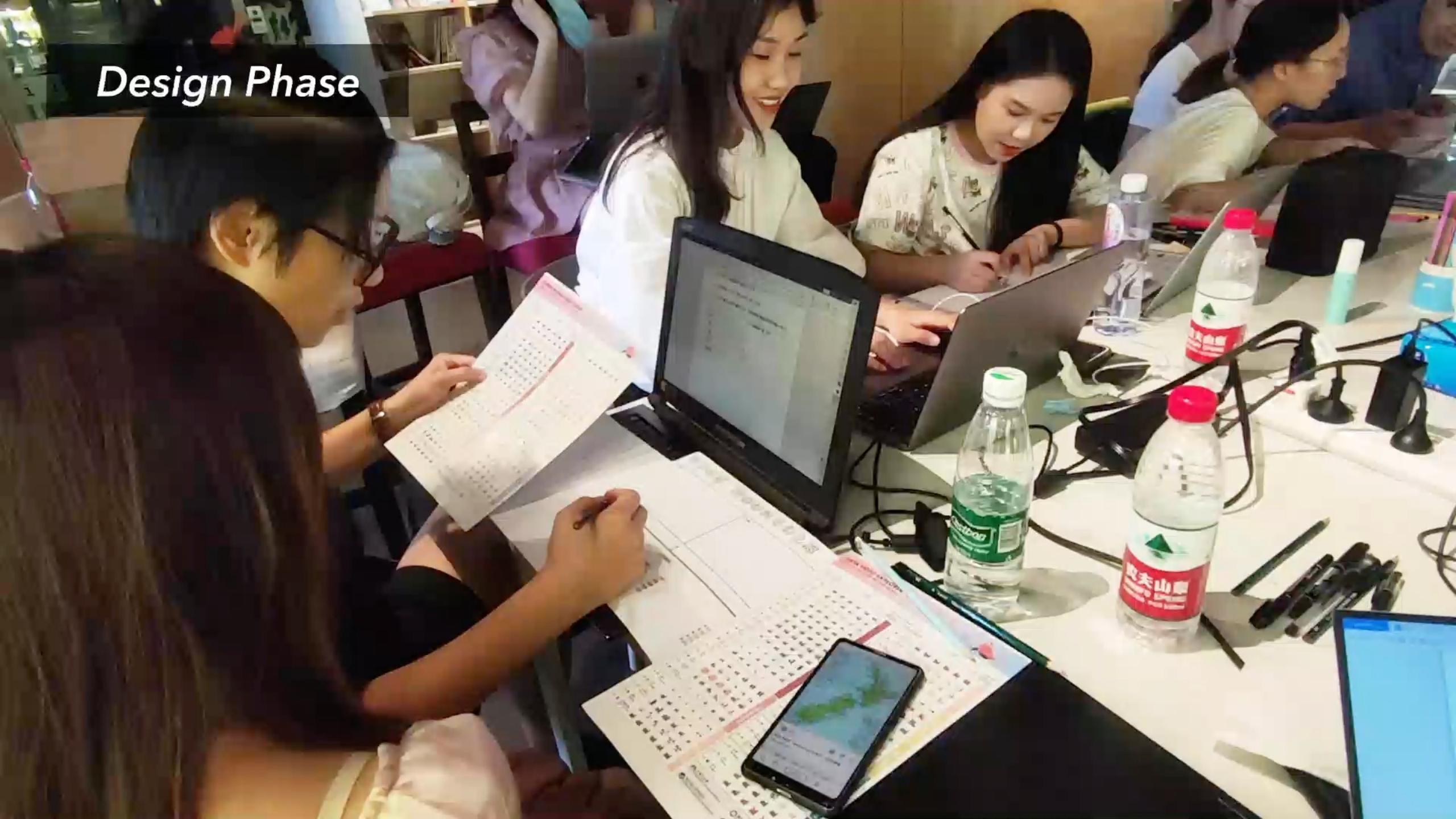
The persona of football fans by Group 3 Participant 4



Global trends in obesity by Group 4 Participant 7



## Design Phase



# 结果分析方法

- 获得的数据为定性数据（包括用户的评价、态度和观点等等）
  - 常用“开放编码”的分析方法
    - 现场录音转换为文本
    - 阅读文本标记处重要信息
    - 归纳、整合信息
- 编码要求
  - 至少2人负责编码（先独立编码，后讨论统一）
  - 审慎和客观
  - 注意语言的上下文环境

# 课程总结

- 本节课
  - 可视化设计的评估方法
  - 受控用户调研
  - 工作坊