



数据可视化基础

王叙萌

wangxumeng@nankai.edu.cn

课程安排

第1周	第1讲：数据可视化简介	第2讲：数据
第2周	第3讲：感知和认知	第4讲：数据可视化基础（1）
第3周	第5讲：数据可视化基础（2）	实验1：可视化图表研究
第4周	第6讲：可视化生成方法（1）	实验2：制作网页（1）
第5周	第7讲：可视化生成方法（2）	实验2：制作网页（2）
第6周	第8讲：可视化生成方法（3）	实验3：使用可视化工具
第7周	第9讲：跨媒体数据可视化	实验4：可视化流程复现（1）
第8周	第10讲：地理数据可视化	实验4：可视化流程复现（2）
第9周	第11讲：时间数据可视化	实验4：可视化流程复现（3）
第10周	第12讲：空间数据可视化	实验4：可视化流程复现（4）
第11周	第13讲：高维数据可视化	实验5：可视化创作（1）
第12周	第14讲：层次数据可视化	实验5：可视化创作（2）
第13周	第15讲：图数据可视化	实验5：可视化创作（3）
第14周	实验5中期汇报	实验5：可视化创作（4）
第15周	第16讲：交互	实验5：可视化创作（5）
第16周	实验5展示	实验5展示

目录

- 可视化生成模型
- 可视映射的选择
- 可视化设计模型

目录

- 可视化生成模型
- 可视映射的选择
- 可视化设计模型

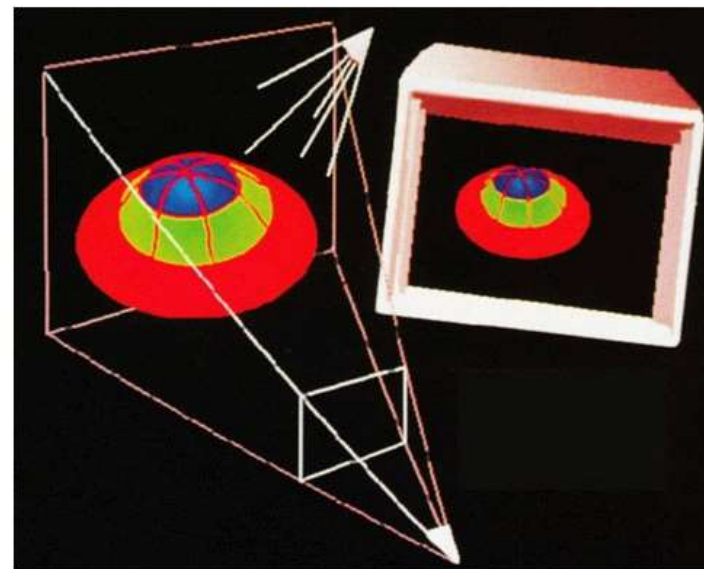
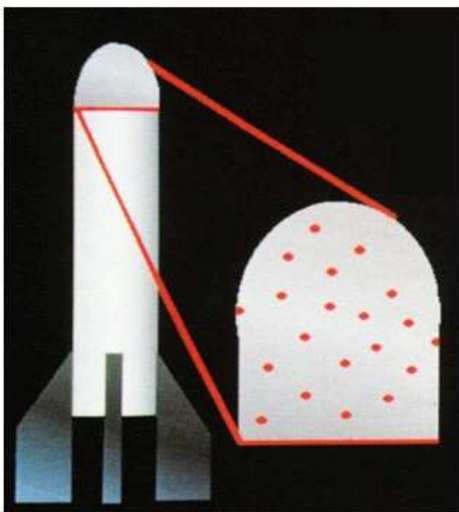
可视化生成模型



Haber and McNabb, 科学可视化概念模型, 1990.

可视化生成模型

模拟数据



可显示图片

派生数据



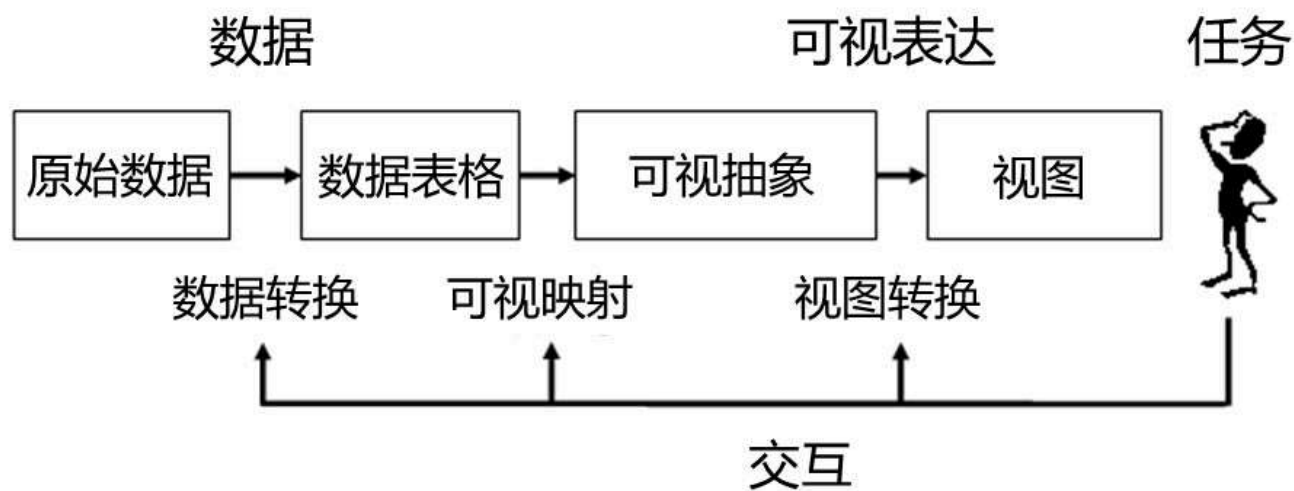
抽象可视化对象

可视化生成模型



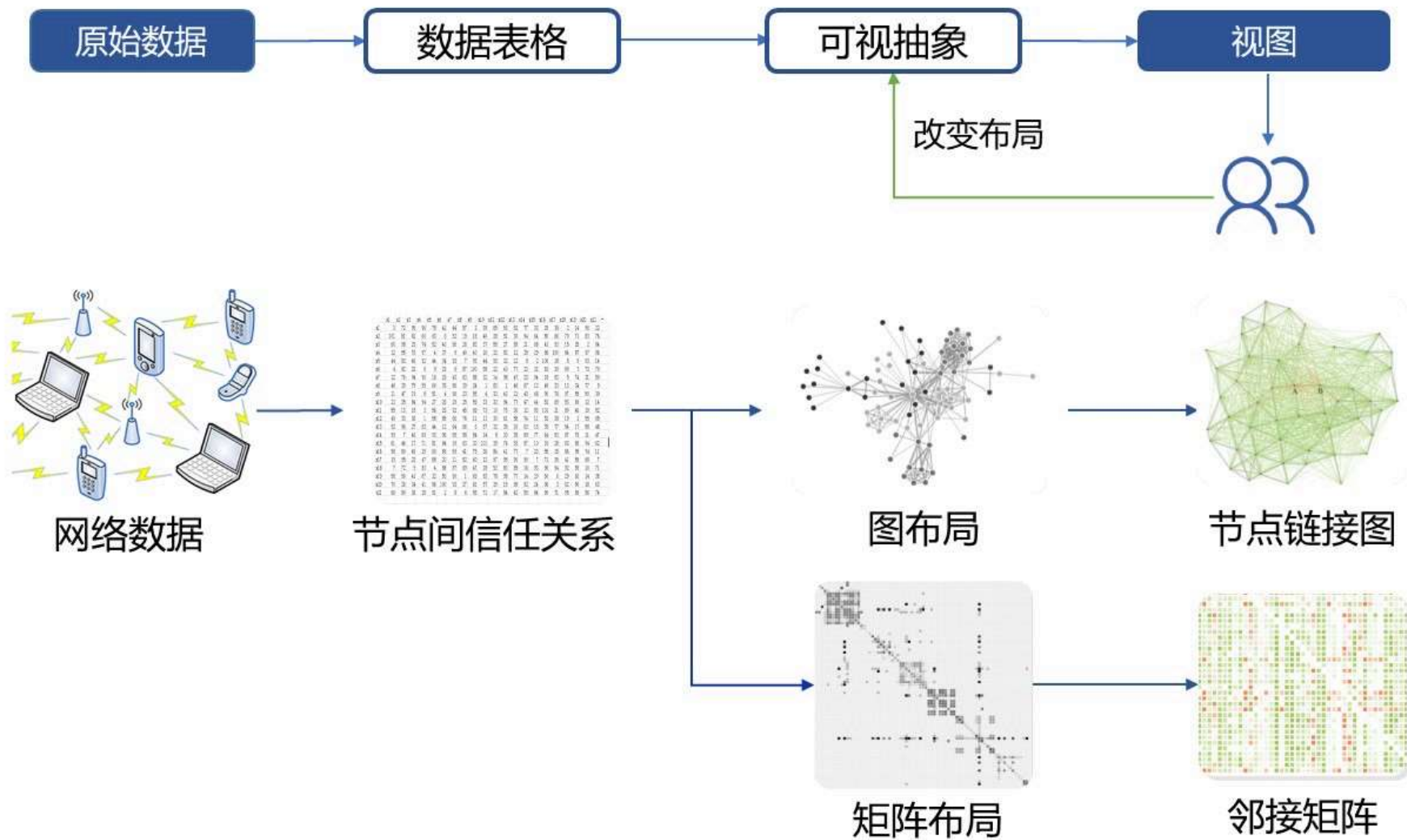
Chi, 数据状态参考模型, 2000.

可视化生成模型

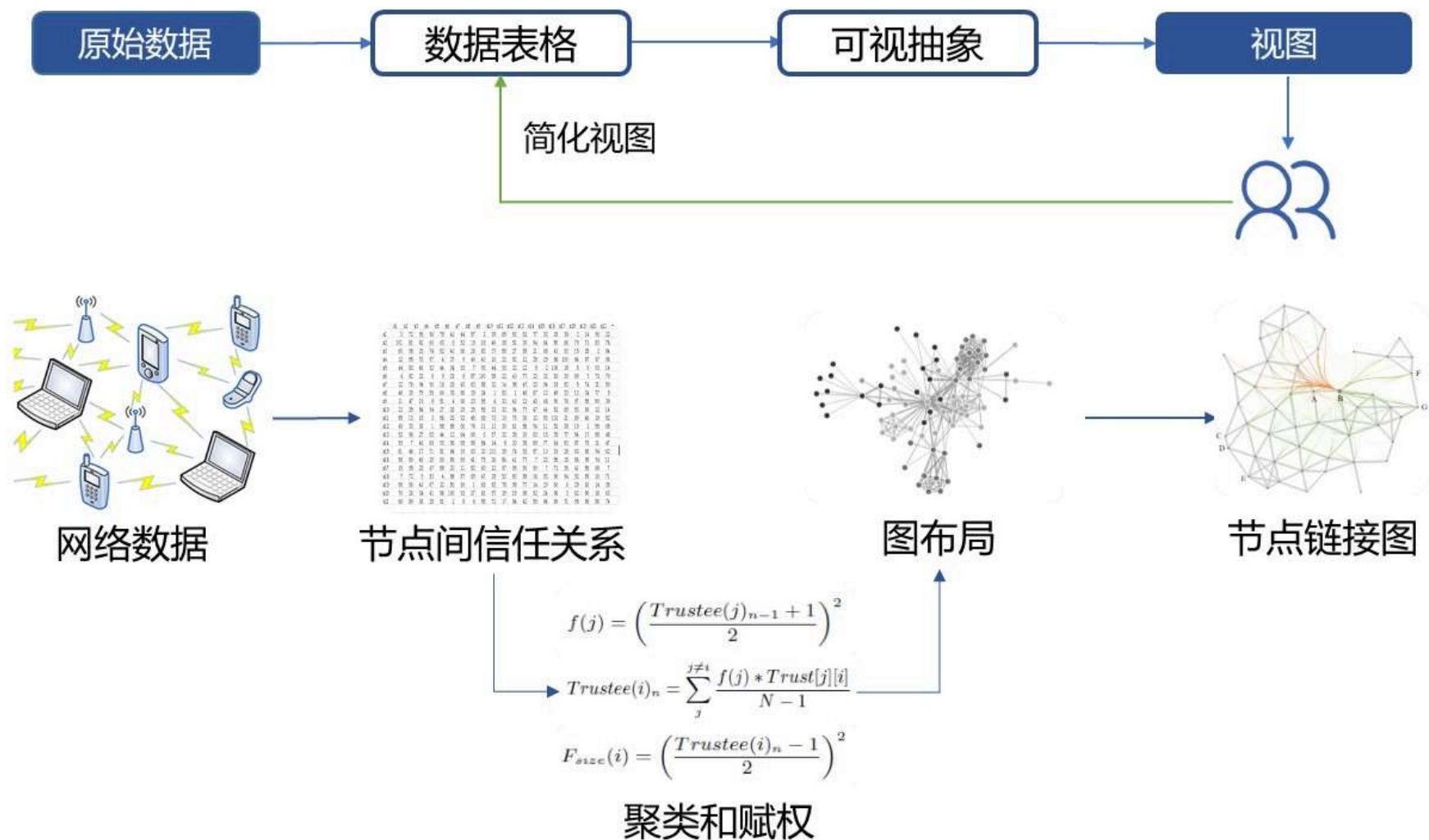


JAVA可视化工具包Prefuse中使用的可视化参考模型

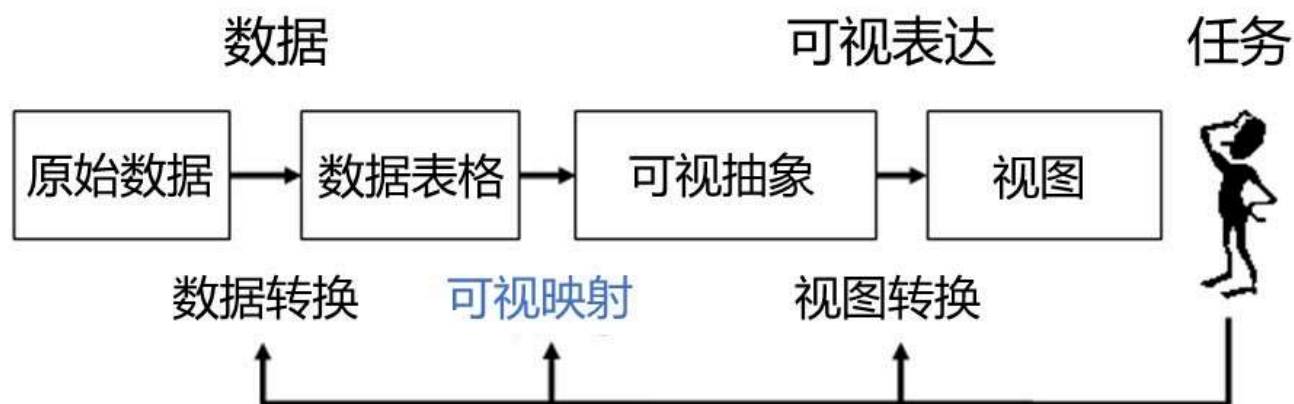
案例



案例



可视化生成模型



JAVA可视化工具包Prefuse中使用的可视化参考模型

目录

- 可视化生成模型
- 可视映射的选择
- 可视化设计模型

可视映射



Jacques Bertin

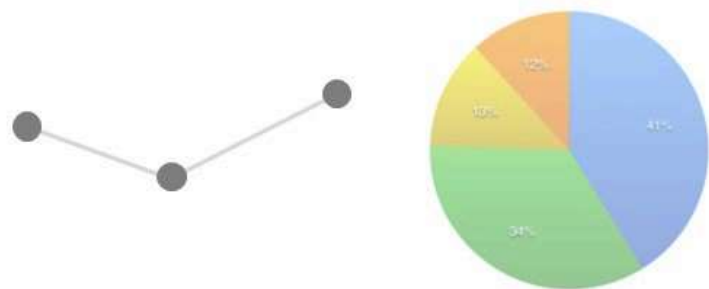
- French Cartographer [1918]
- [Semiology of Graphics](#) [1967]
- Theoretical principles for visual encodings

图符+编码

	点	线	面
位置			
大小			
灰度			
纹理			
颜色			
方向			
形状			

• 一些问题

- 点线面的区分不方便
 - 三者存在组成关系
- 总结不全面
 - 角度等可视映射没被提到



根据属性类型选择

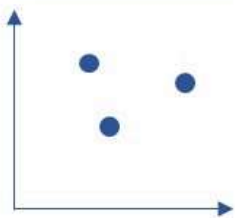


Jock Mackinlay

- 信息可视化专家
- Tableau副总裁
- Automating the Design of Graphical Presentations of Relational Information [1986]

可视映射

位置



长度



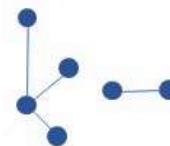
形状



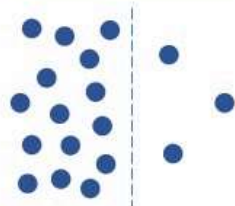
饱和度



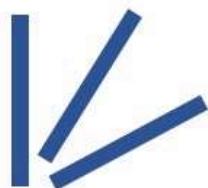
连接



密度



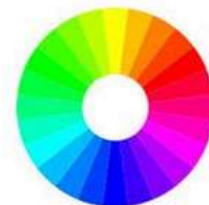
斜度



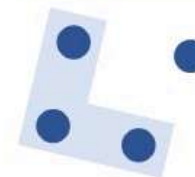
面积



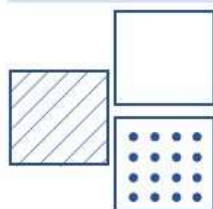
色调



包含



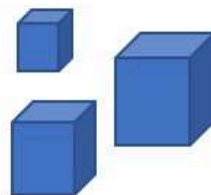
纹理



角度



体积

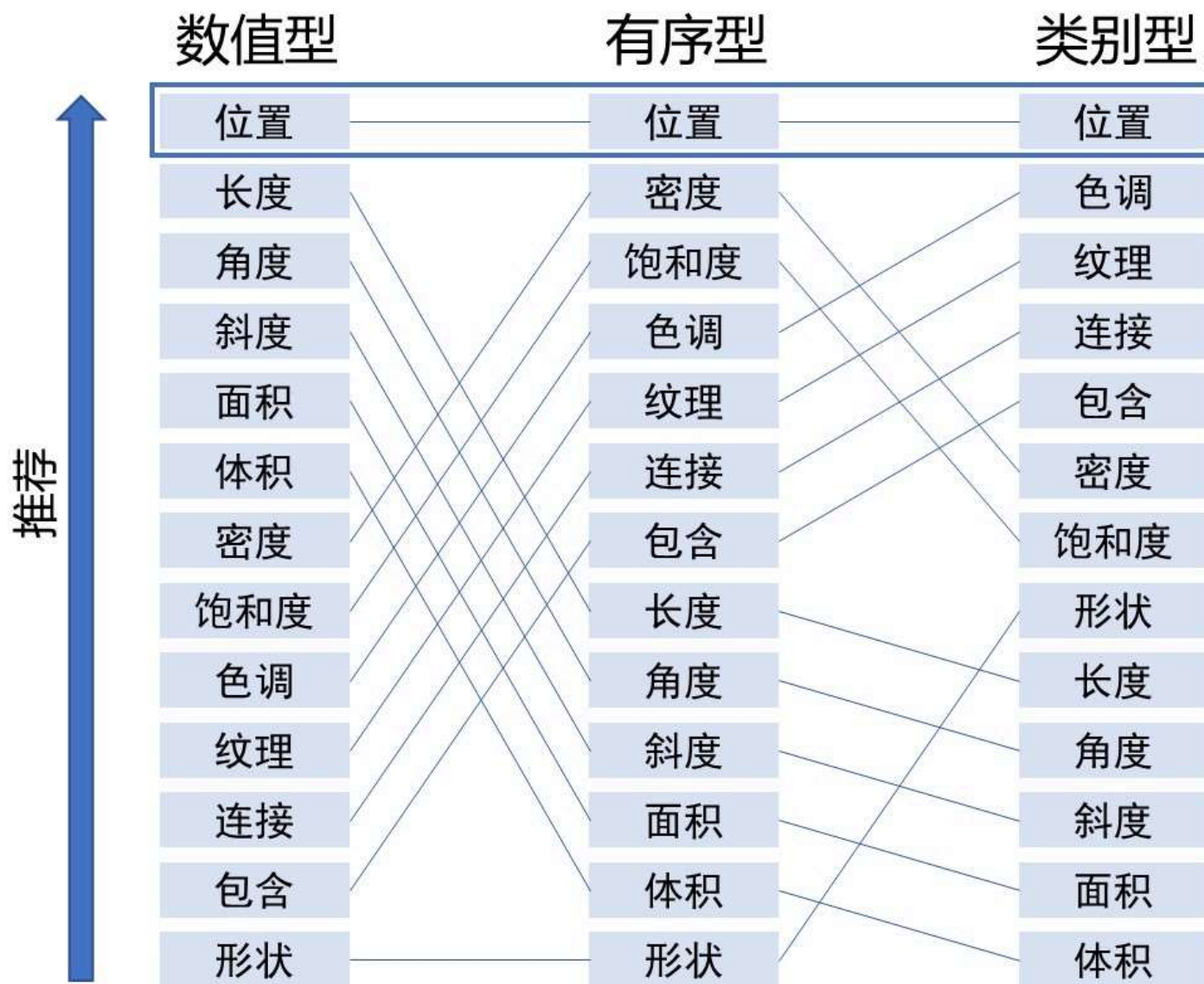


如何选用?

可视映射的选择

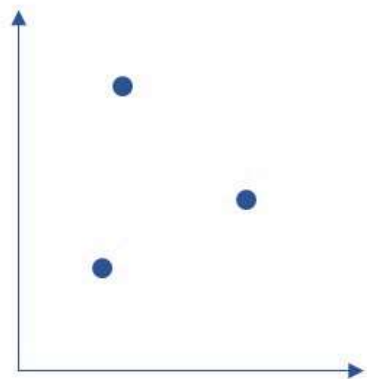
- 被映射的数据是什么
 - 属性类型
 - 属性组合
- 映射后用来干什么
 - 感知
 - 认知（支持分析任务）

根据感知效果按属性类型推荐

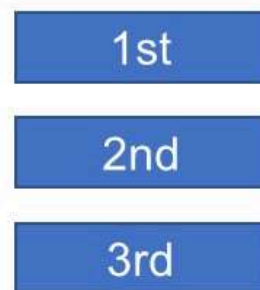


位置

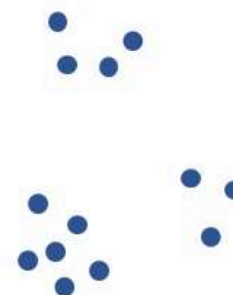
数值型



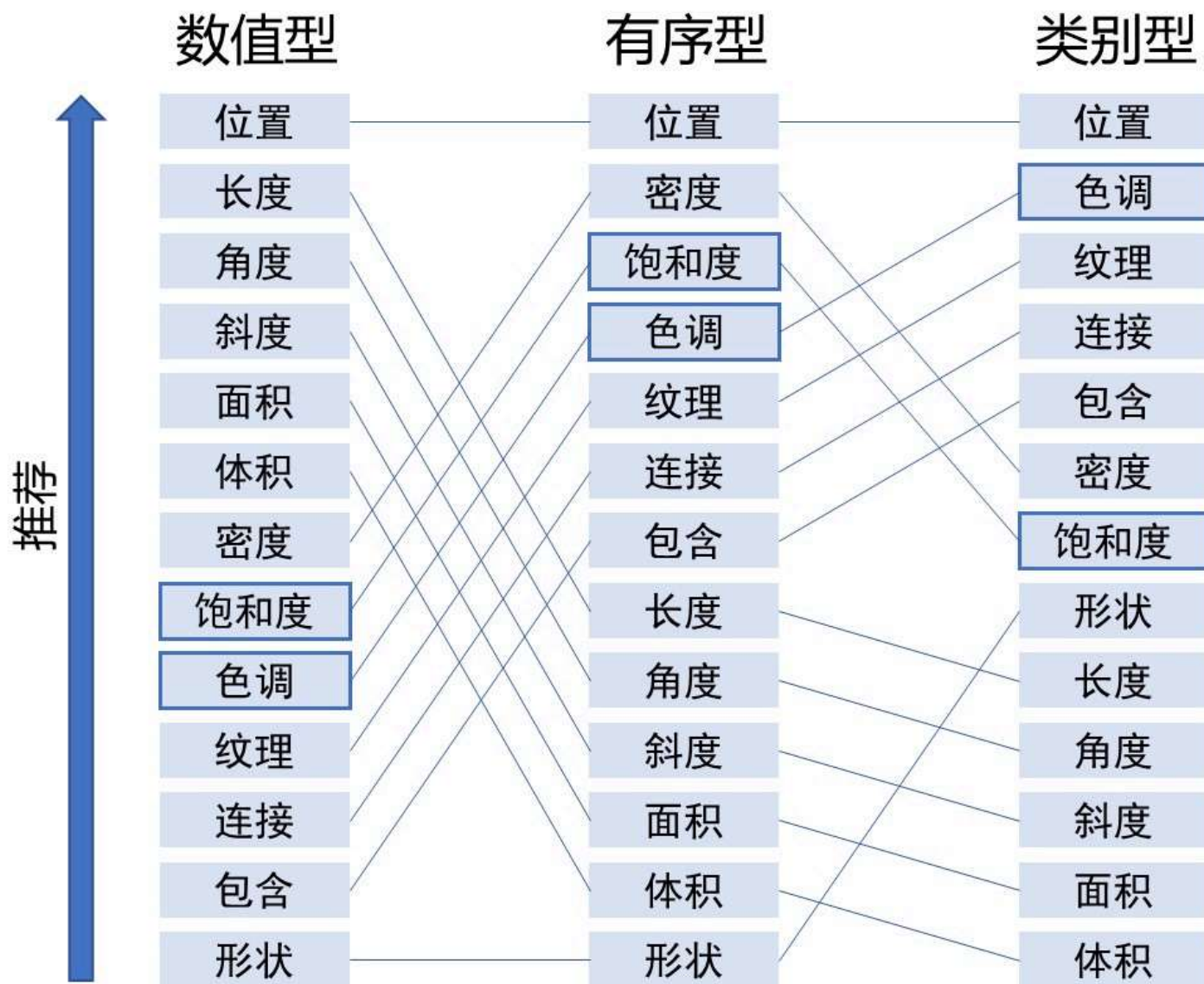
有序型



类别型

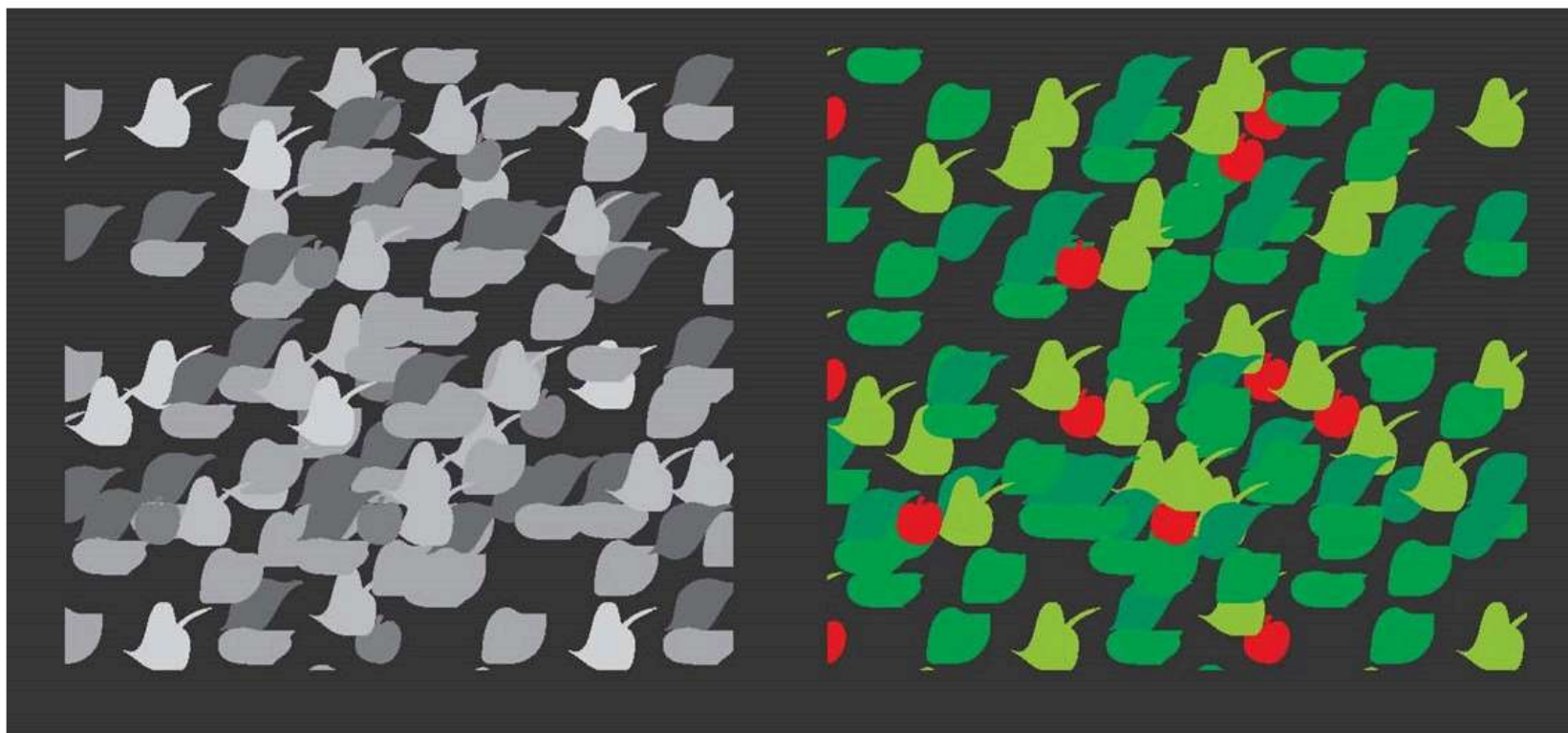


根据感知效果按属性类型推荐



颜色

- 颜色存在感知区分度和语义认知差异



颜色



数值型



有序型



类别型

数值型属性的颜色映射

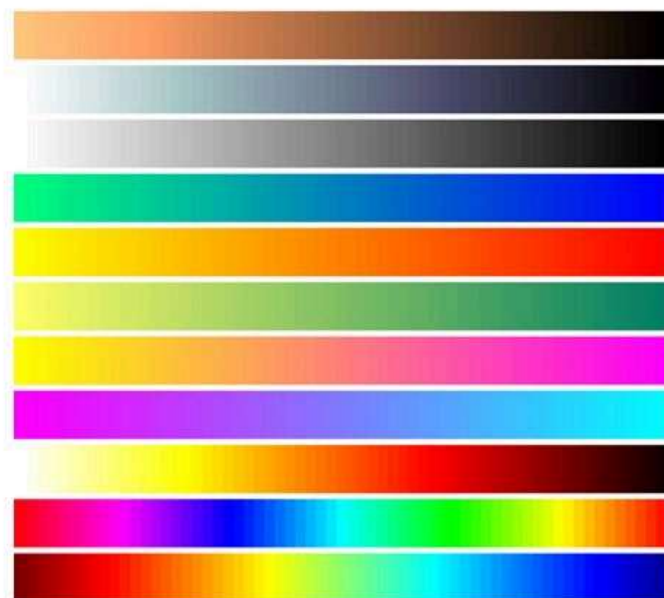
- 数值范围-颜色渐变

单色调渐变

多色调渐变

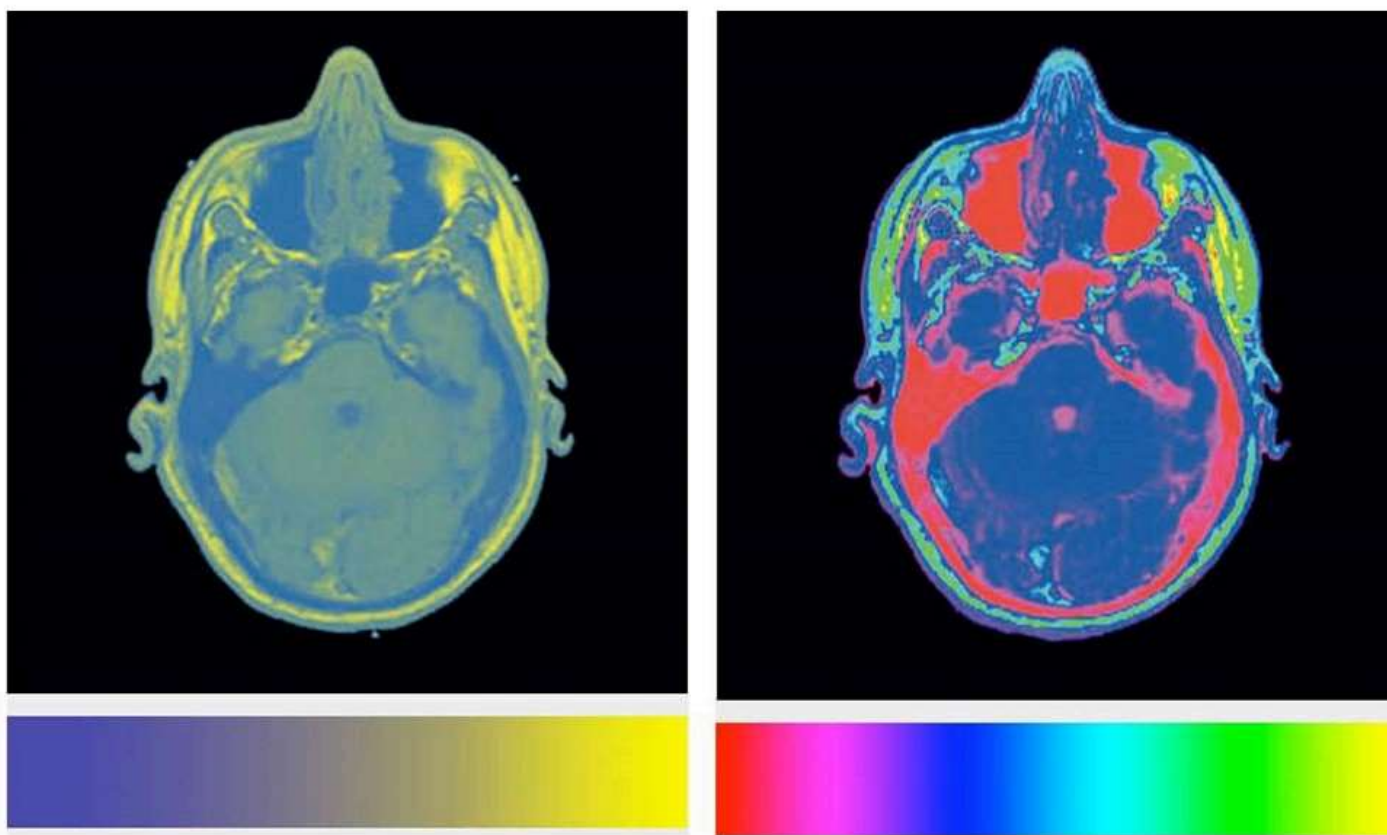
对比色渐变

选择时需考虑感知认知效果
与数据分布是否匹配



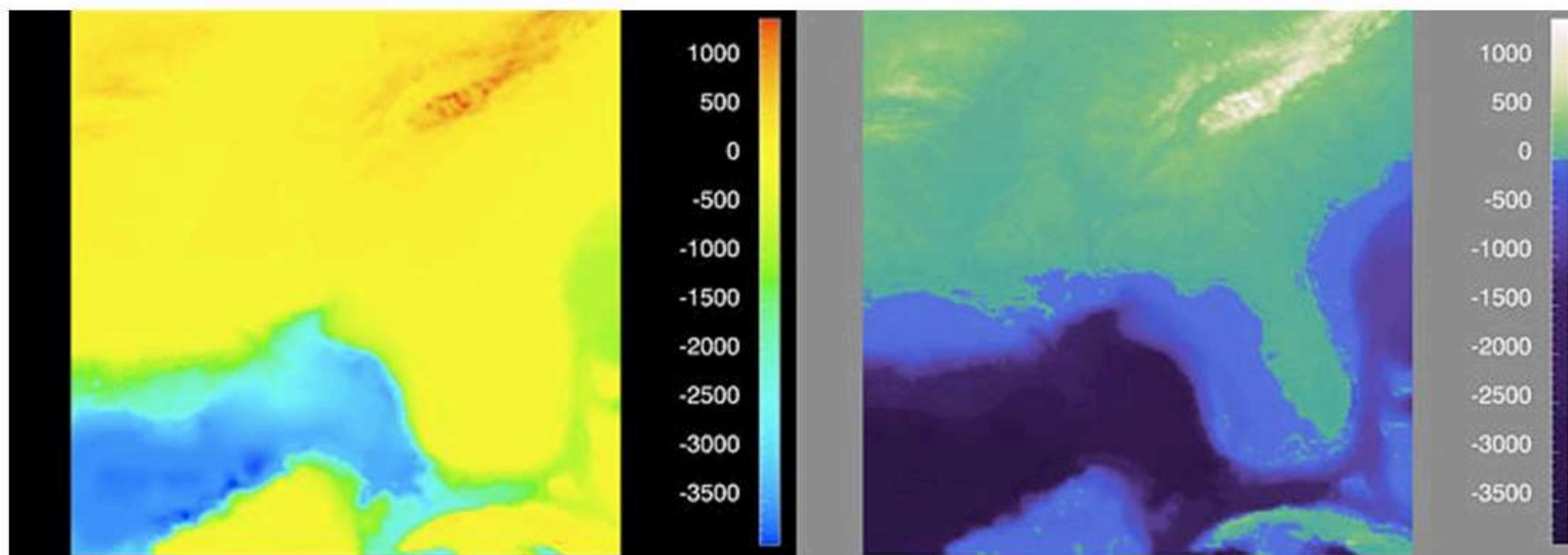
数值型属性的颜色映射

- 感知：数值分布是否需要细致区分



数值型属性的颜色映射

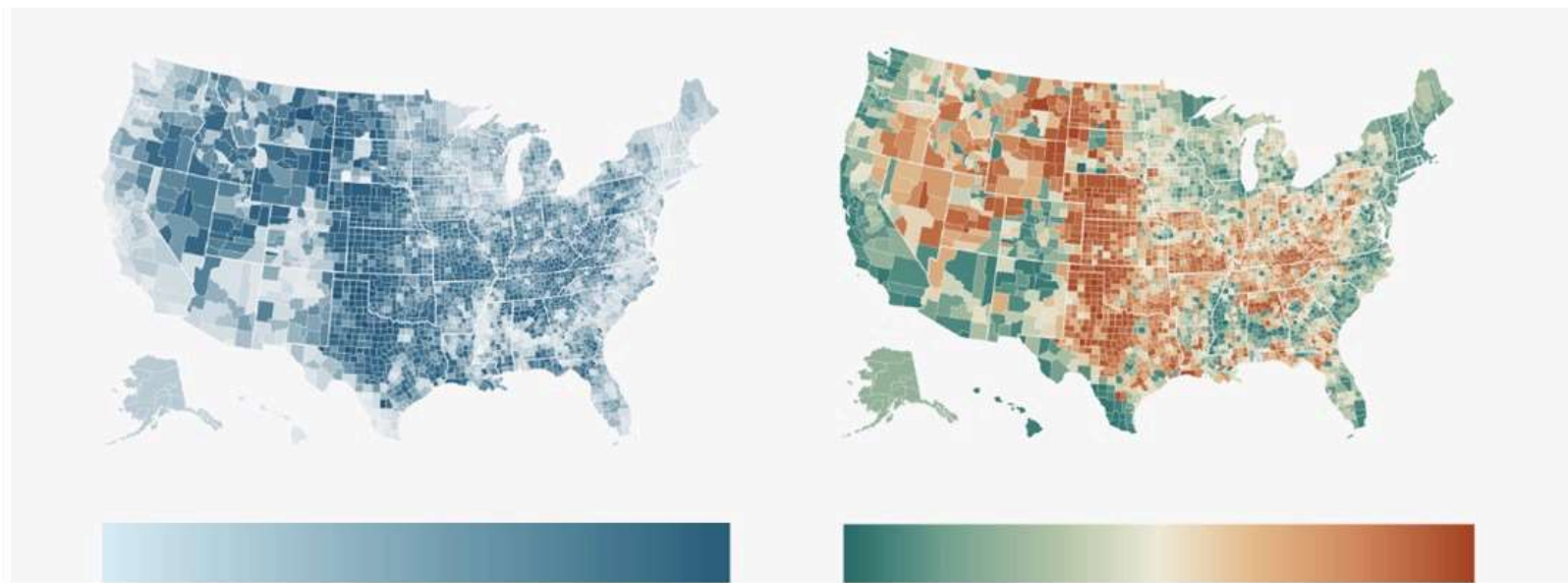
- 感知+认知：分布是否均匀，区分需求是否一致



调整映射方式可以显示更多细节

数值型属性的颜色映射

- 认知：数值范围是否区分正负

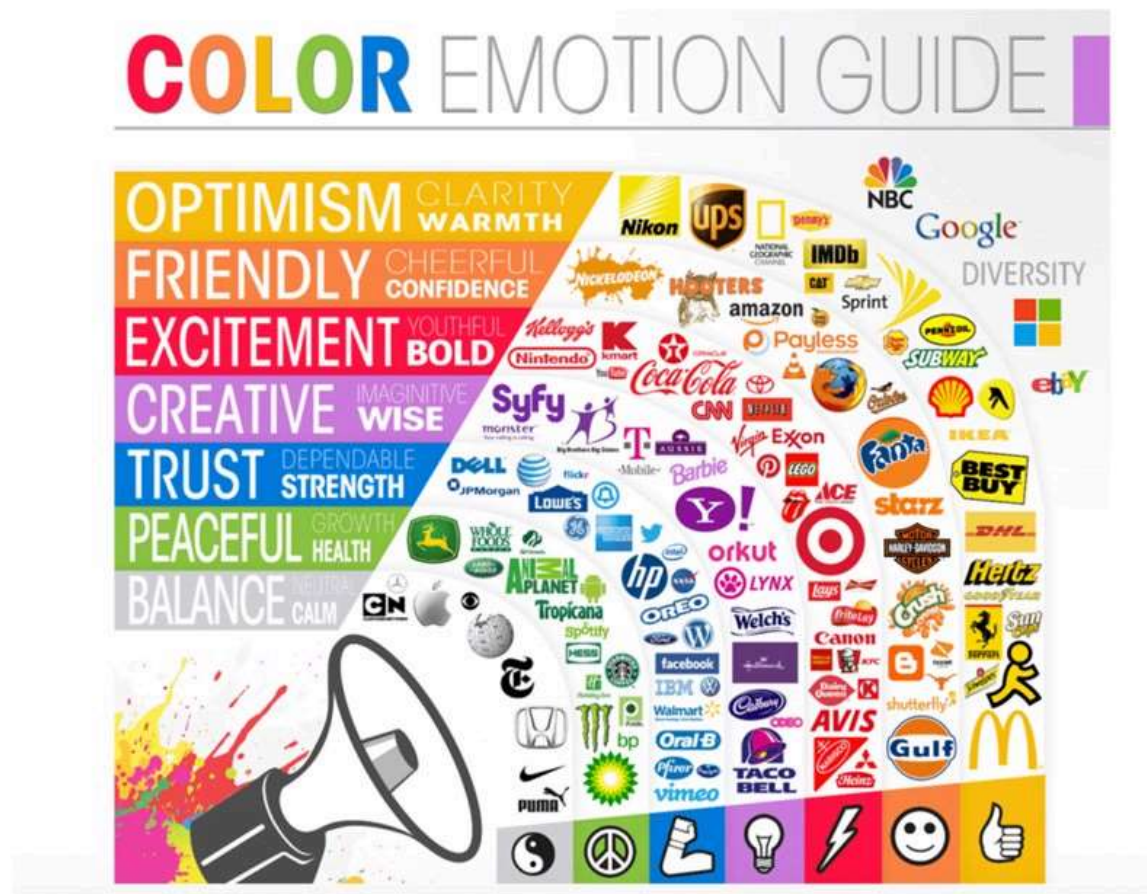


单色渐变：适用于由小到大

对比渐变：适用于由负到正

类别型属性的颜色映射

- 认知：颜色的语义和要表达的语义是否一致



类别型属性的颜色映射

- 分析主体对颜色语义的理解不同

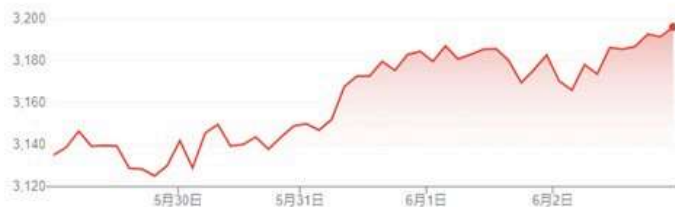
市场概况 > 上海证券交易所综合股价指数

3,195.46

+60.43 (1.93%) ↑ 过去5天

6月2日 GMT+8 15:00 • 免责声明

1天 5天 1个月 6个月 YTD 1年 5年 最大



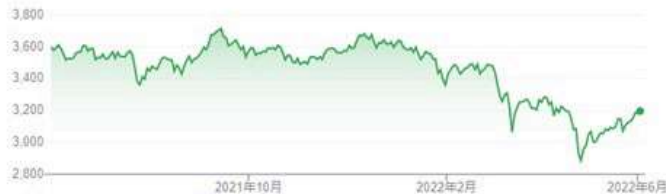
市场概况 > 上海证券交易所综合股价指数

3,195.46

-404.08 (-11.23%) ↓ 过去1年

6月2日 GMT+8 15:00 • 免责声明

1天 5天 1个月 6个月 YTD 1年 5年 最大



Market Summary > Nasdaq Composite

12,012.73

-3,212.42 (-21.10%) ↓ past 6 months

Jun 3, 17:15 EDT • Disclaimer

1D 5D 1M 6M YTD 1Y 5Y Max



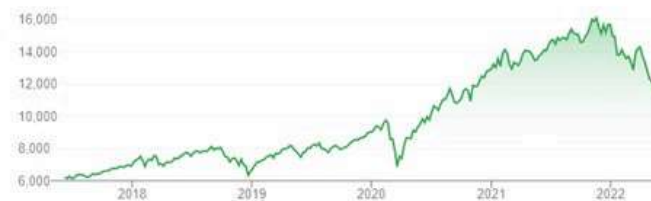
Market Summary > Nasdaq Composite

12,012.73

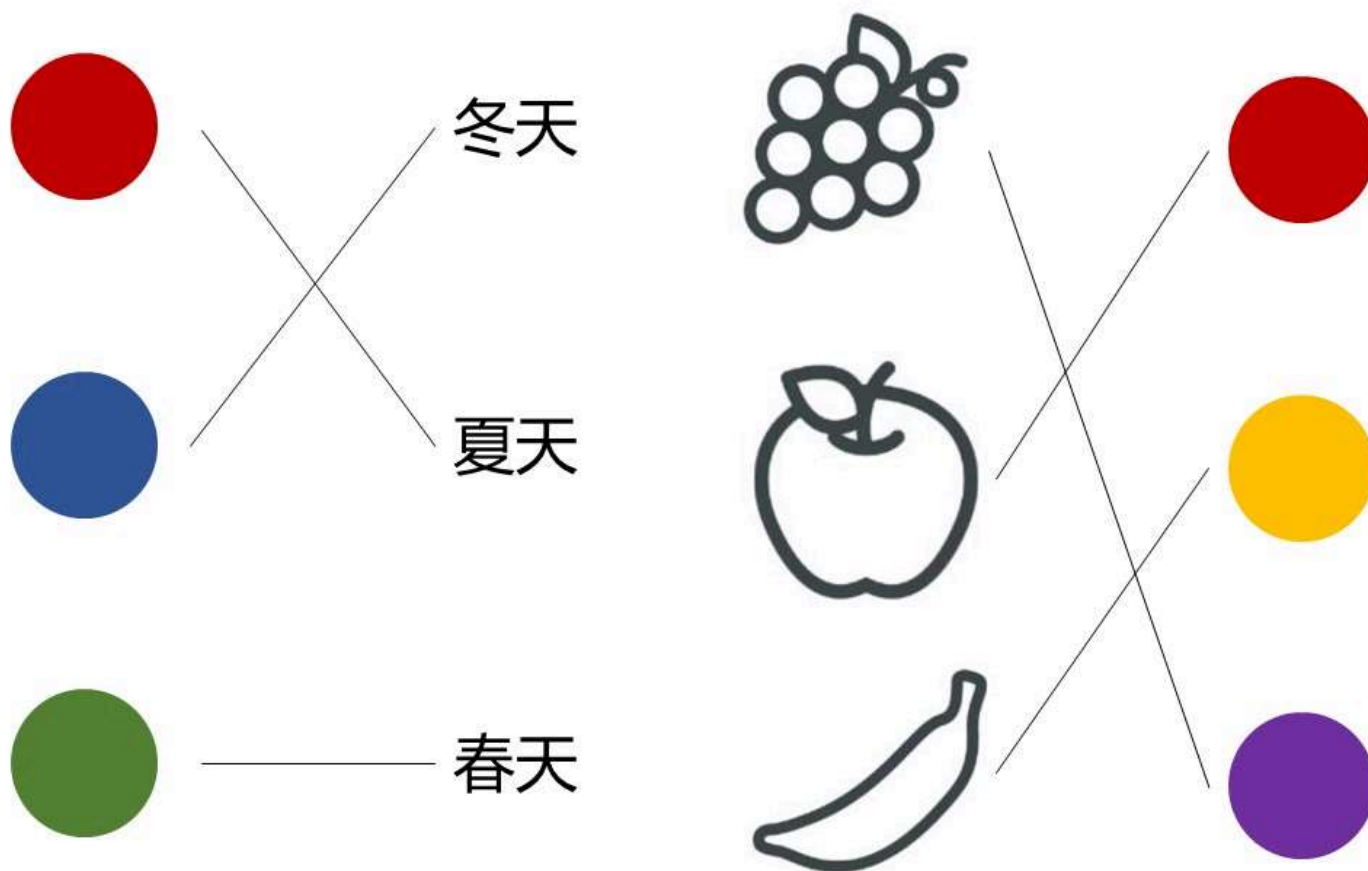
+5,804.81 (93.51%) ↑ past 5 years

Jun 3, 17:15 EDT • Disclaimer

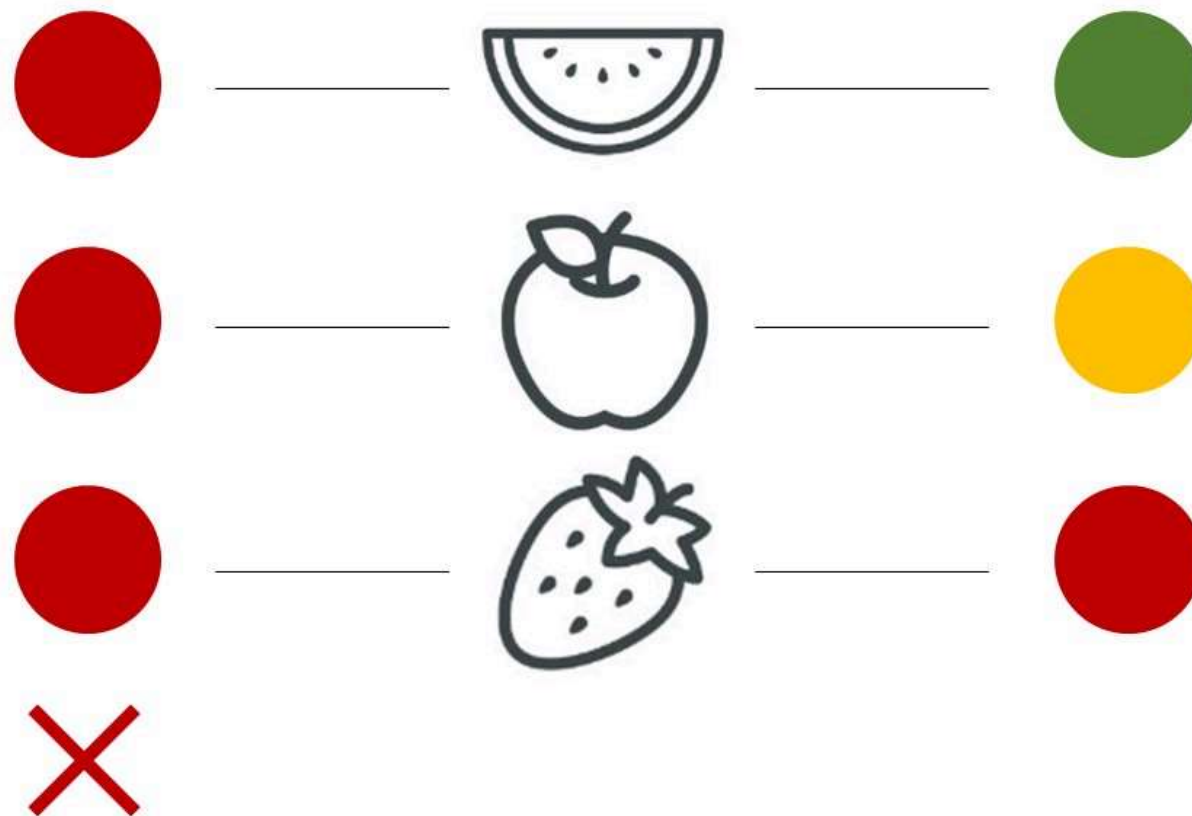
1D 5D 1M 6M YTD 1Y 5Y Max



类别型属性的颜色映射



类别型属性的颜色映射



类别型数据的图符映射

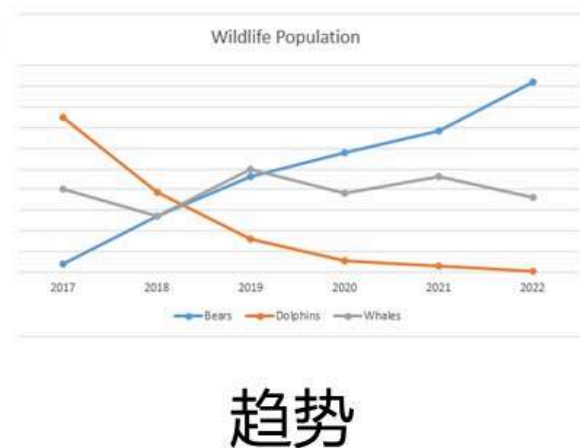
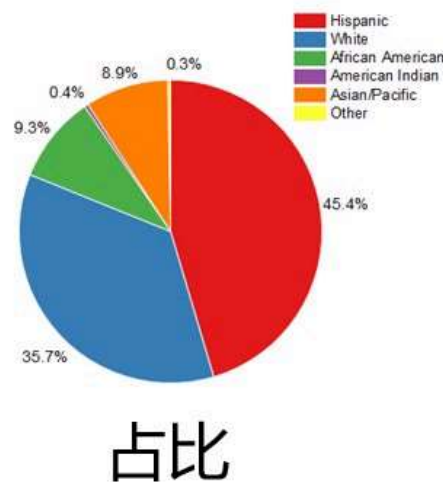
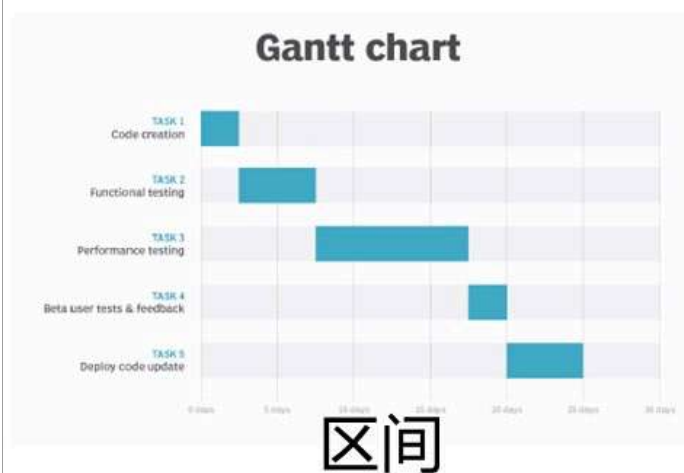
1. 足球 (2)  
2. 篮球 (2)  
3. 排球 (2)  
4. 三人篮球 (2)  
5. 沙滩排球 (2)  
6. 手球 (2)  
7. 曲棍球 (2)  
8. 七人制橄榄球 (2)  
9. 水球 (2)  
10. 乒乓球 (5)     
11. 羽毛球 (5)   TPE  
12. 网球 (5)     
13. 高尔夫 (2)  
14. 游泳 (34)         
          
        
15. 跳水 (8)        
16. 花样游泳 (2)  
17. 马拉松游泳 (2)  
18. 赛艇 (14)         
  
19. 皮划艇静水 (10)     
  

- 20皮划艇激流回旋 (6) 
- 21帆船 (10) 
- 22冲浪 (2) 
- 23滑板 (4) 
- 24攀岩 (4) 
- 25霹雳舞 (2) 
- 26田径 (48) 
- 27竞技体操 (14) 
- 28艺术体操 (2) 
- 29蹦床 (2) 
- 30公路自行车 (4) 
- 31山地自行车 (2) 
- 32场地自行车 (12) 
- 33竞速小轮车 (2) 
- 34自由式小轮车 (2) 
- 35射击 (15) 

- 36射箭 (5) 
- 37击剑 (12) 
- 
- 38马术 (6) 
- 39举重 (10) 
- 40拳击 (13) 
- TPE 
- 41摔跤 (18) 
- 
- 42柔道 (15) 
- 
- 43跆拳道 (8) 
- 44铁人三项 (3) 
- 45现代五项 (2) 

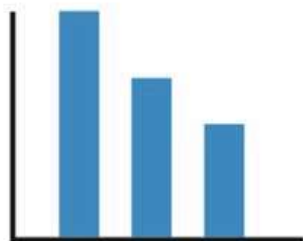
其他可视映射的语义性

- 长度越长往往对应**区间**越大
- 角度越大往往对应**占比**越高
- 斜度上扬代表**走势**上涨

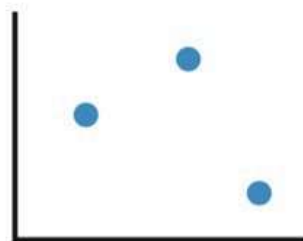


检验映射组合效果

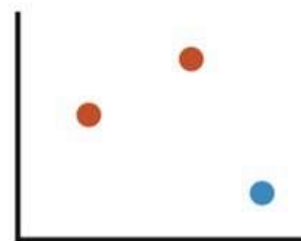
- 数据中往往包含多个属性



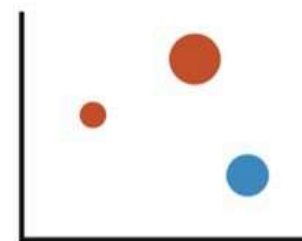
位置+长度



位置+位置



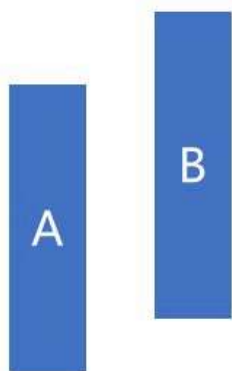
位置+颜色



位置+色调+面积

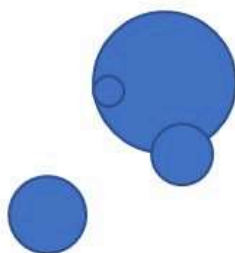
检验映射组合效果

难以比较



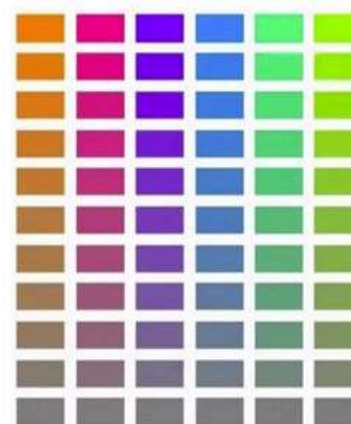
长度+位置

出现遮挡



面积+位置

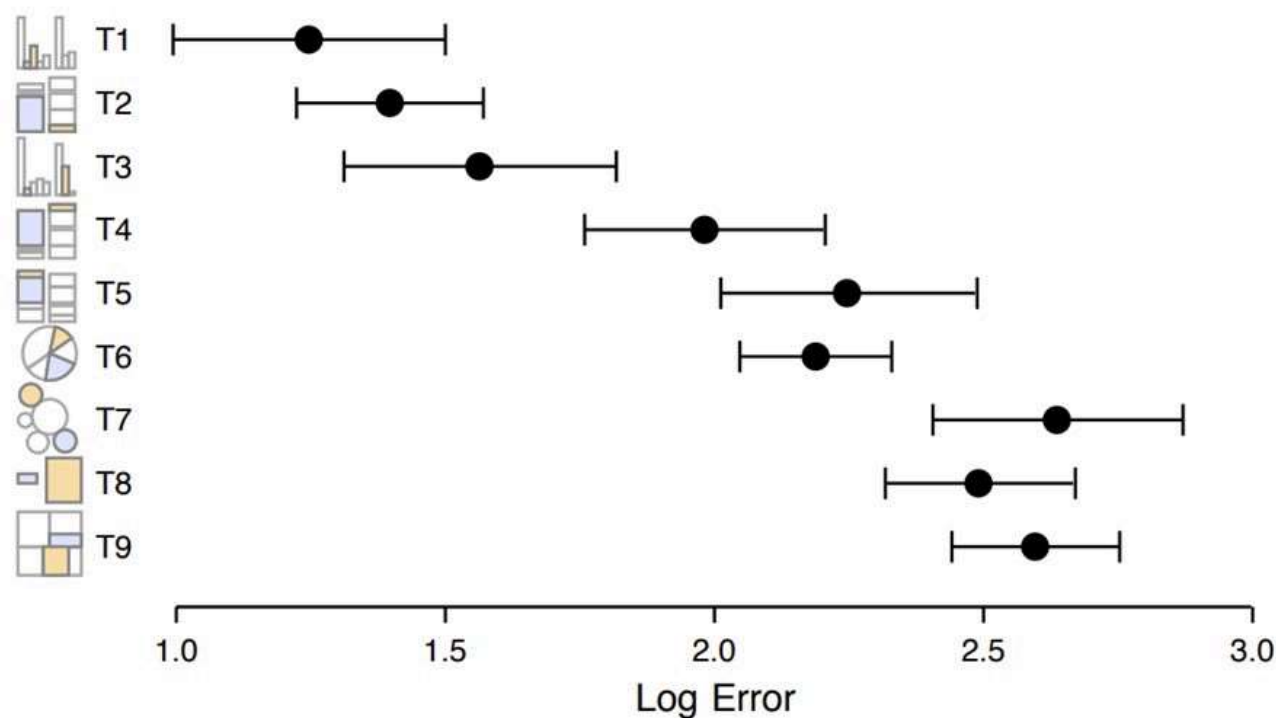
难以区分



色调+饱和度

检验映射组合效果

- 不同形状下的差别感知精确度



基于众包实验研究不同可视设计的感知精确度
Heer and Bostock 2010

可视映射的选择

- 被映射的数据是什么
 - 属性类型
 - 属性组合
 - 映射效果如何
 - 感知
 - 认知
- 

试一试

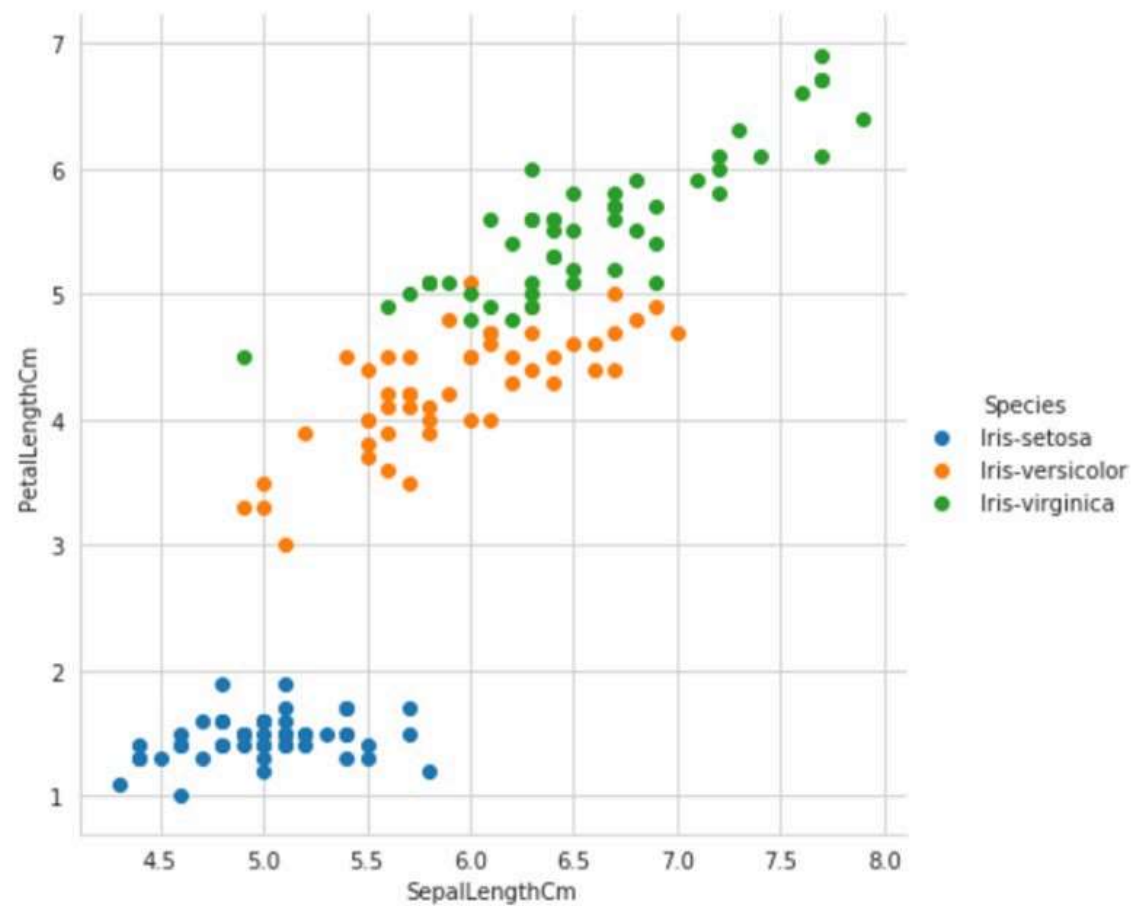
如何同时展示每朵花的全部属性，以分析三类花的形态差别？



Sepal.Length	Sepal.Width	Petal.Length	Petal.Width	Species
5.1	3.5	1.4	0.2	setosa
4.9	3.0	1.4	0.2	setosa
4.7	3.2	1.3	0.2	setosa
4.6	3.1	1.5	0.2	setosa
5.0	3.6	1.4	0.2	setosa
5.4	3.9	1.7	0.4	setosa
4.6	3.4	1.4	0.3	setosa
5.0	3.4	1.5	0.2	setosa
4.4	2.9	1.4	0.2	setosa
4.9	3.1	1.5	0.1	setosa
5.4	3.7	1.5	0.2	setosa
4.8	3.4	1.6	0.2	setosa
4.8	3.0	1.4	0.1	setosa
4.3	3.0	1.1	0.1	setosa
5.8	4.0	1.2	0.2	setosa
5.7	4.4	1.5	0.4	setosa
5.4	3.9	1.3	0.4	setosa
5.1	3.5	1.4	0.3	setosa
5.7	3.8	1.7	0.3	setosa
5.1	3.8	1.5	0.3	setosa
5.4	3.4	1.7	0.2	setosa
5.1	3.7	1.5	0.4	setosa

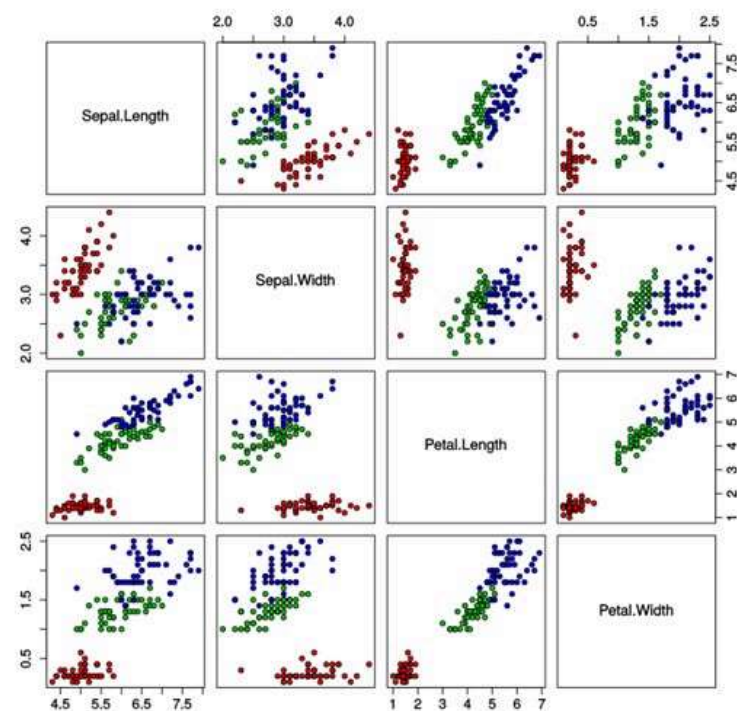
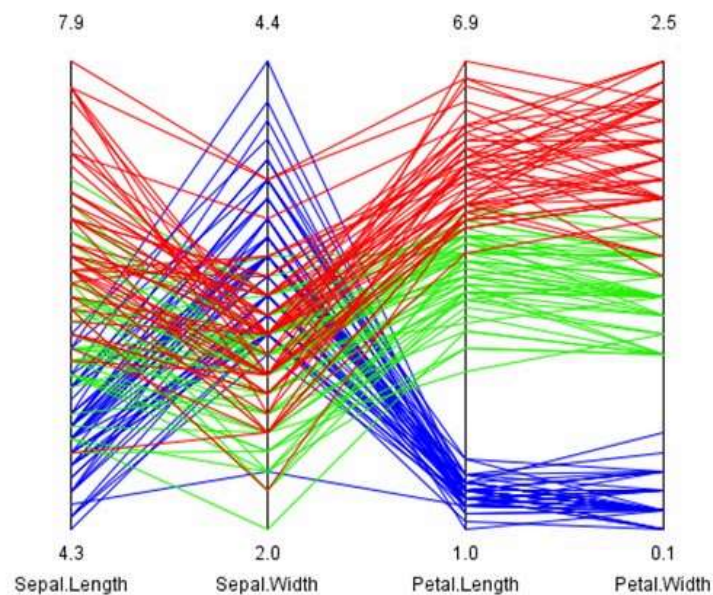
4个数值型+1个类别型（3类）

组合效果



组合效果

优先为重要信息使用排名高的编码



位置+色调的运用

目录

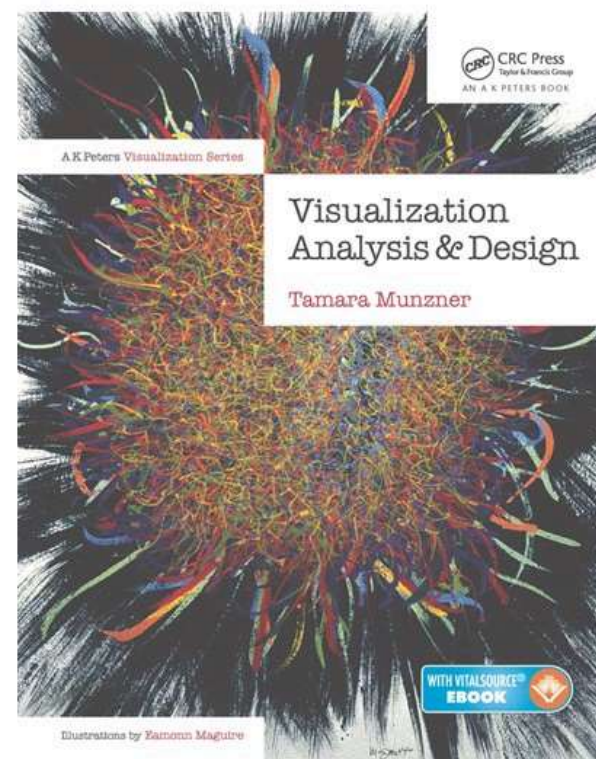
- 可视化生成模型
- 可视映射的选择
- 可视化设计模型

设计模型



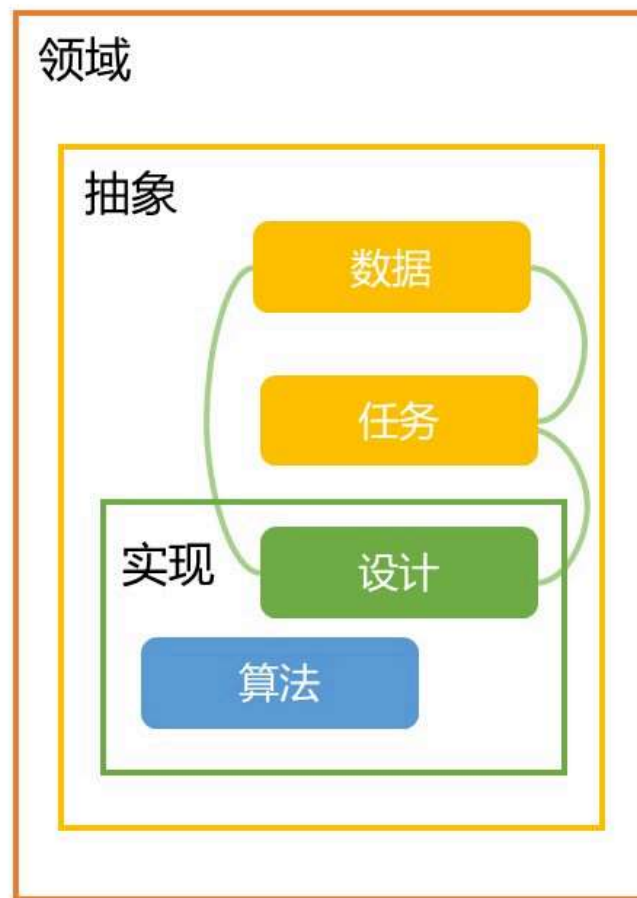
Tamara Munzner

University of British Columbia



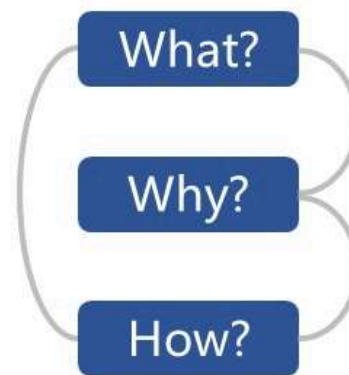
嵌套模型

- 领域-问题
- 抽象-需求
 - 可视化什么数据?
 - 支持什么分析任务?
- 实现-解决方案
 - 可视映射
 - 交互设计
 - 算法优化



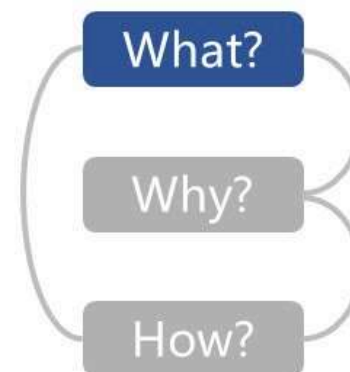
What-Why-How

- 作用
 - 确定自己的设计思路
 - 向别人介绍设计理念



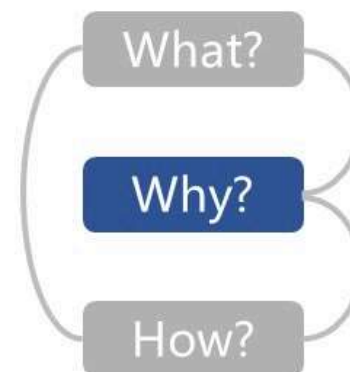
What-Why-How

- What-数据-可视化什么?
 - 什么数据
 - 有哪些信息
 - 数据记录
 - 数据属性
 - 数据特征



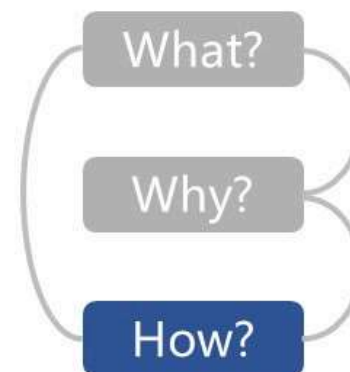
What-Why-How

- Why-任务-为什么要这样可视化?
 - 要做什么任务?
 - 帮谁做任务?



What-Why-How

- How-实现-怎么可视化?
 - 编码
 - 组合
 - 优化



嵌套模型：验证设计效果

领域

抽象

数据

任务

实现

设计

算法

威胁：错误的领域问题

验证：观察或采访目标用户

威胁：不好的数据或不好的分析过程抽象

威胁：无效的编码/交互技术

验证：论证编码/交互设计有效

威胁：低效算法

验证：分析计算复杂度

实现系统

验证：测量系统时间/内存

验证：效果分析、小范围实验

验证：对目标用户测试，在实际应用中部署

验证：观察采用率

总结

- 可视化生成模型
- 可视映射的选择
 - 考虑数据内容和表达效果
- 可视化设计模型
 - 嵌套模型
 - What-Why-How

试一试

- 为以下数据中的信息设计可视映射方案

地区	地区生产总值（亿元/年）和排名									
	2022		2021		2020		2019		2018	
北京市	41610.9	#13	41045.6	#12	35943.3	#13	35445.1	#12	33106.0	#12
天津市	16311.3	#24	15685.1	#24	14008.0	#23	14055.5	#23	13362.9	#23
河北省	42370.4	#12	40397.1	#13	36013.8	#12	34978.6	#13	32494.6	#13
山西省	25642.6	#20	22870.4	#20	17835.6	#20	16961.6	#21	15958.1	#21
内蒙古自治区	23158.6	#21	21166.0	#21	17258.0	#21	17212.5	#20	16140.8	#20
辽宁省	28975.1	#17	27569.5	#17	25011.4	#17	24855.3	#15	23510.5	#17
吉林省	13070.2	#26	13163.8	#26	12256.0	#26	11726.8	#26	11253.8	#26
黑龙江省	15901.0	#25	14858.2	#25	13633.4	#25	13544.4	#25	12846.5	#24
...

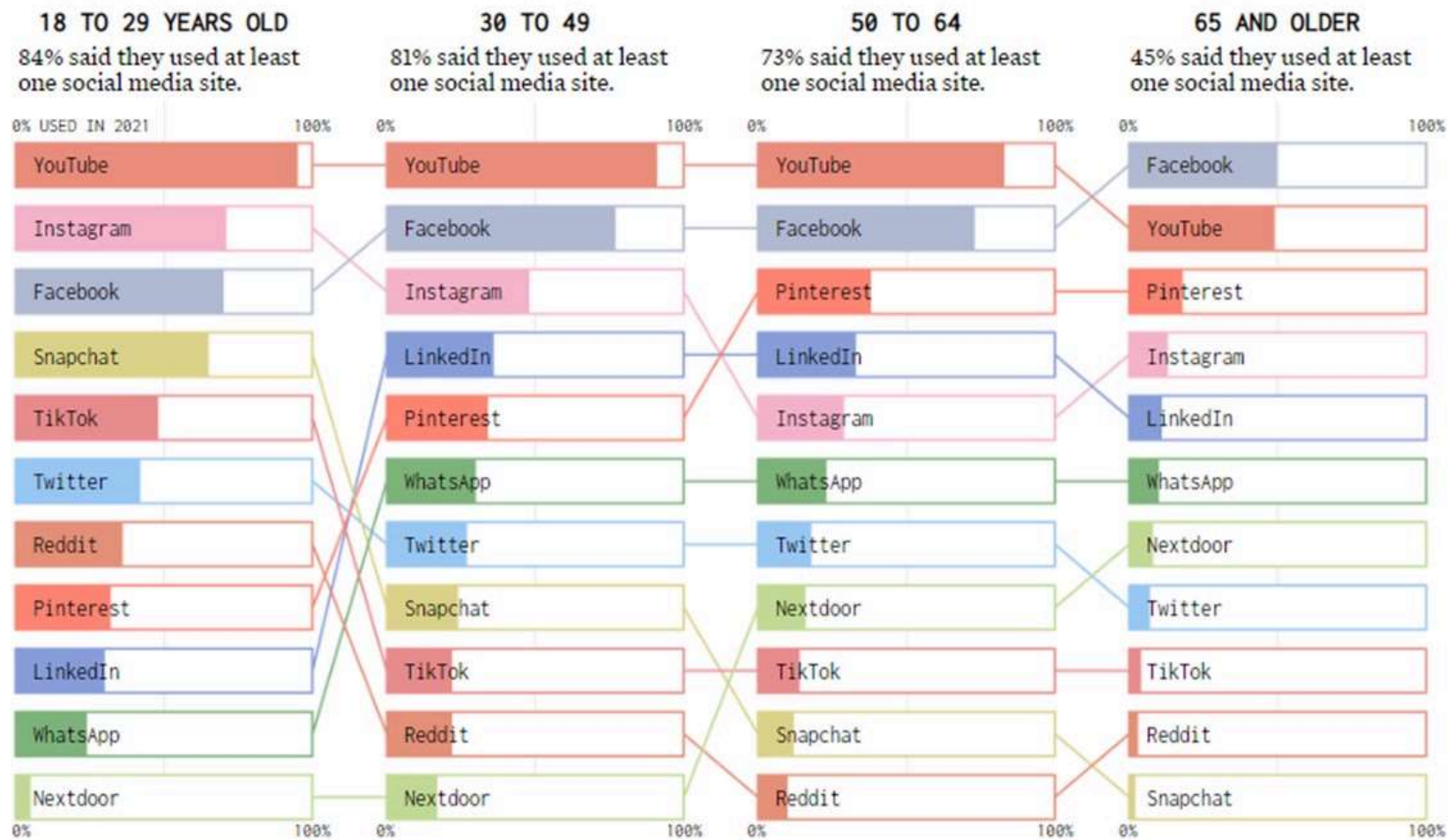
<https://data.stats.gov.cn/easyquery.htm?cn=E0103>

参考答案

App\Age	18-29		30-49		50-64		65+	
Facebook	70%	#3	77%	#2	73%	#2	50%	#1
Instagram	71%	#2	48%	#3	29%	#5	13%	#4
LinkedIn	30%	#9	36%	#4	33%	#4	11%	#5
Twitter	43%	#6	27%	#7	18%	#7	7%	#8
Pinterest	32%	#8	34%	#5	38%	#3	18%	#3
Snapchat	65%	#4	24%	#8	12%	#10	2%	#11
YouTube	95%	#1	91%	#1	83%	#1	49%	#2
WhatsApp	24%	#10	30%	#6	23%	#6	10%	#6
Reddit	36%	#7	24%	#10	10%	#11	3%	#10
TikTok	48%	#5	22%	#9	14%	#9	4%	#9
Nextdoor	5%	#11	17%	#11	16%	#8	8%	#7

<https://www.pewresearch.org/internet/fact-sheet/social-media/?menulitem=4abfc543-4bd1-4b1f-bd4a-e7c67728ab76>

参考答案



Source: Pew Research Center / By: FlowingD