



Basics of Data Visualization

数据可视化基础

中国地质大学（北京）信息工程学院

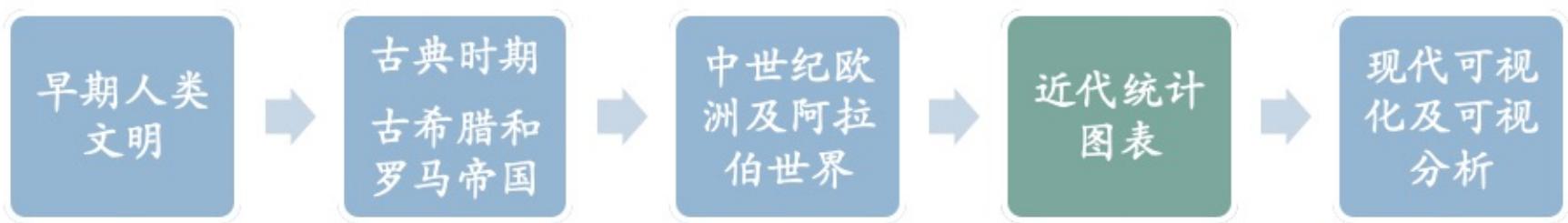
叶山

yes@cugb.edu.cn

可视化简史（继续）

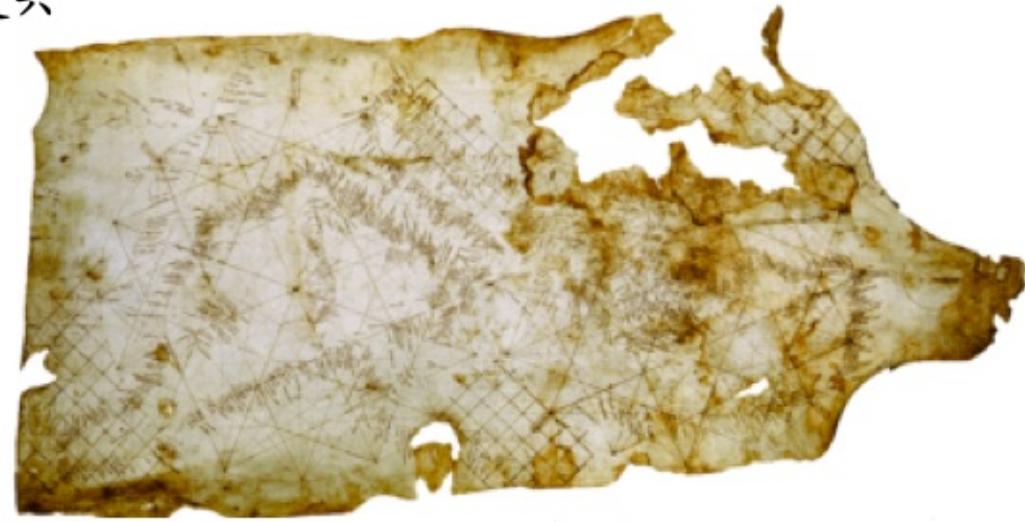


可视化简史

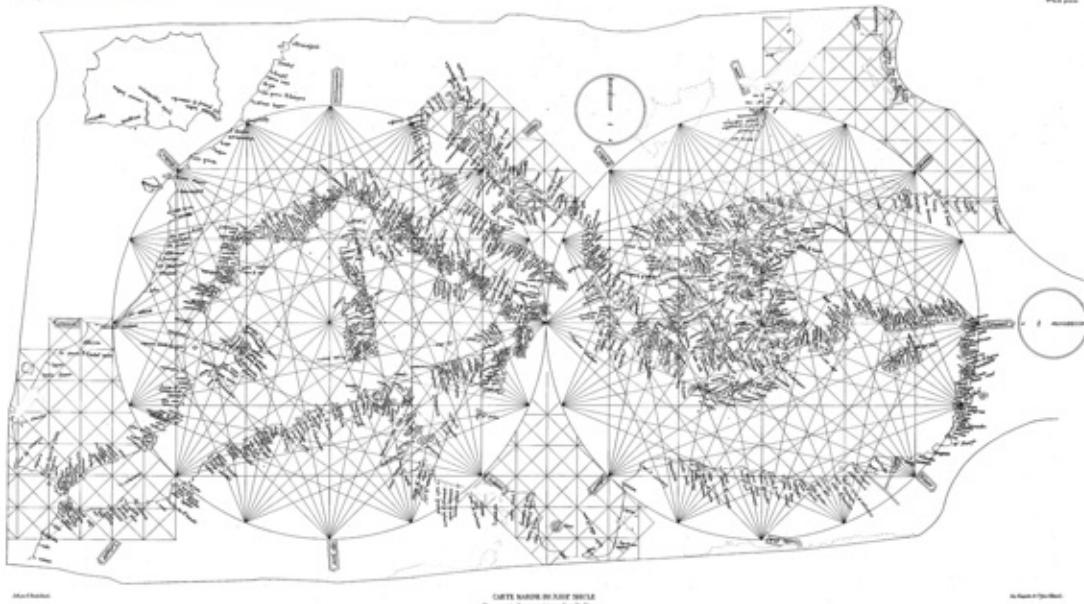


可视化简史

文艺复兴

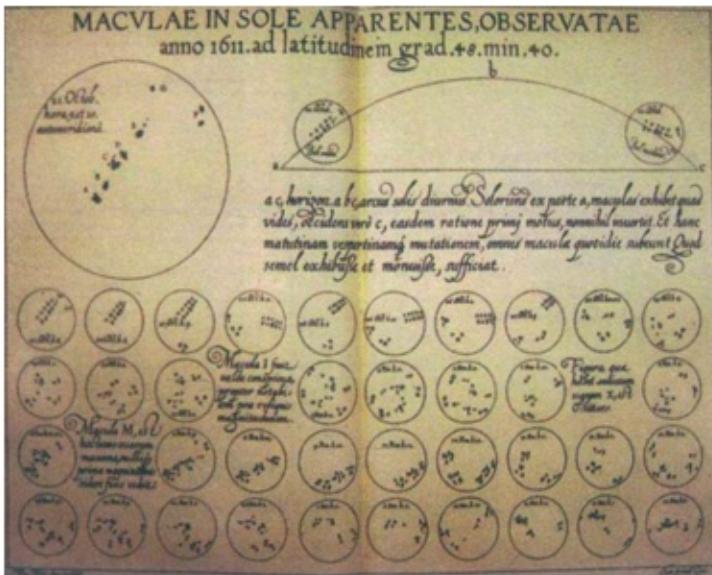


1275年的比萨航海图
(Carta Pisana)

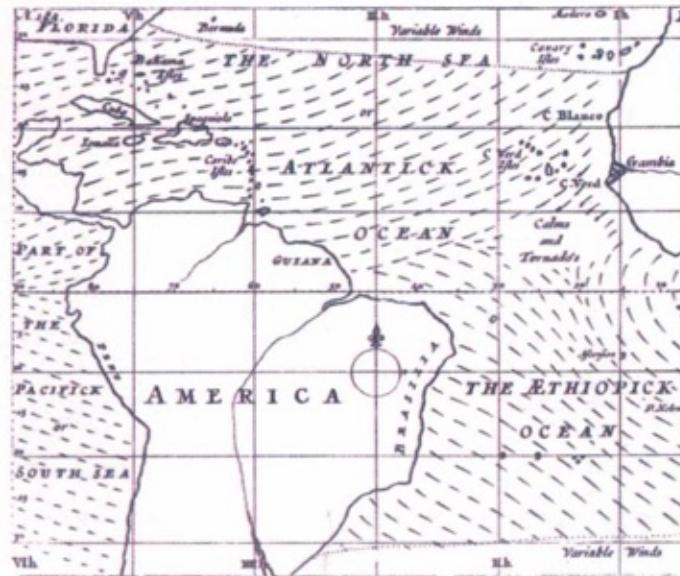


可视化简史

十七世纪



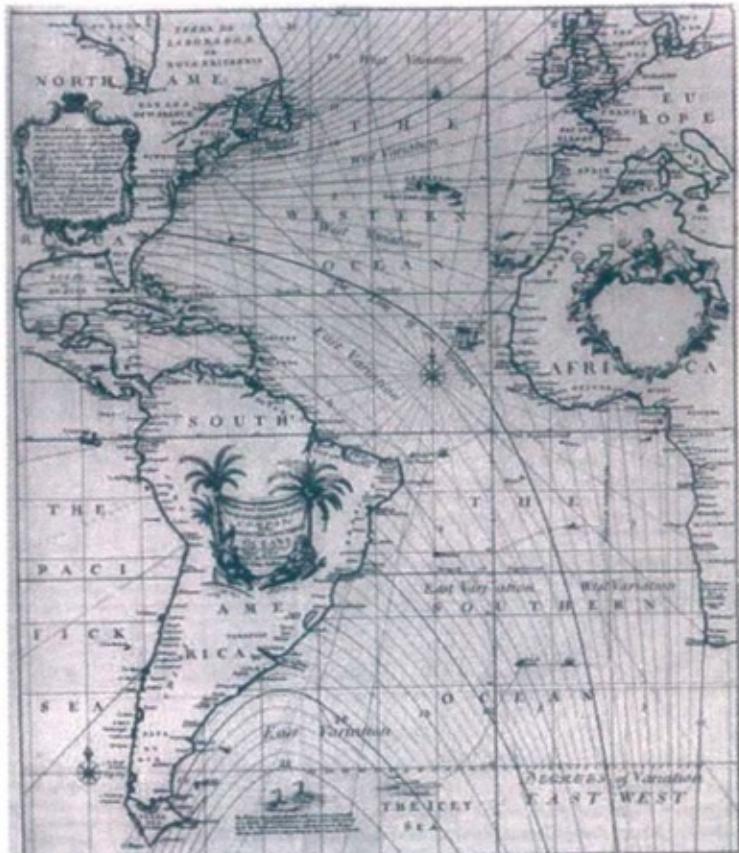
太阳黑子变化示意图 (1626)



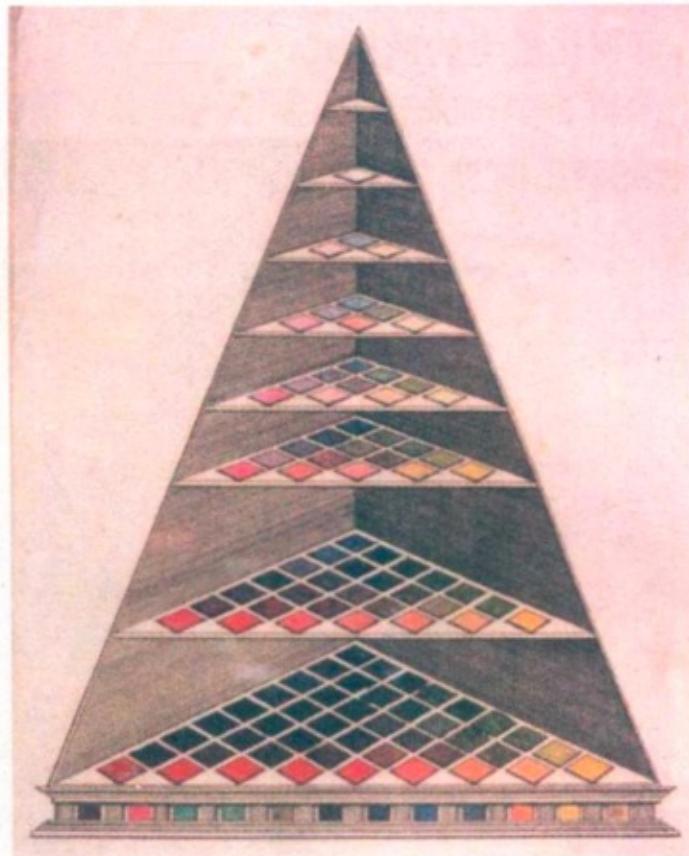
最早的天气预报图 (1686)

可视化简史

十八世纪



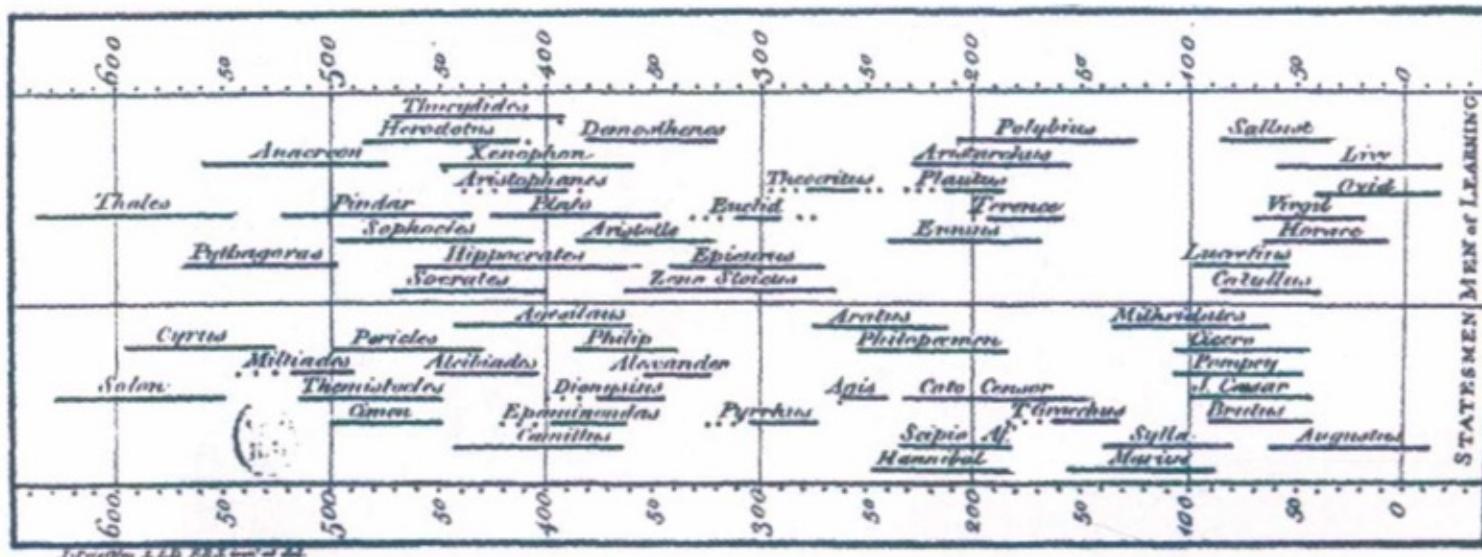
地球等磁线图 (1701)



三维金字塔颜色系统图 (1758)

可视化简史

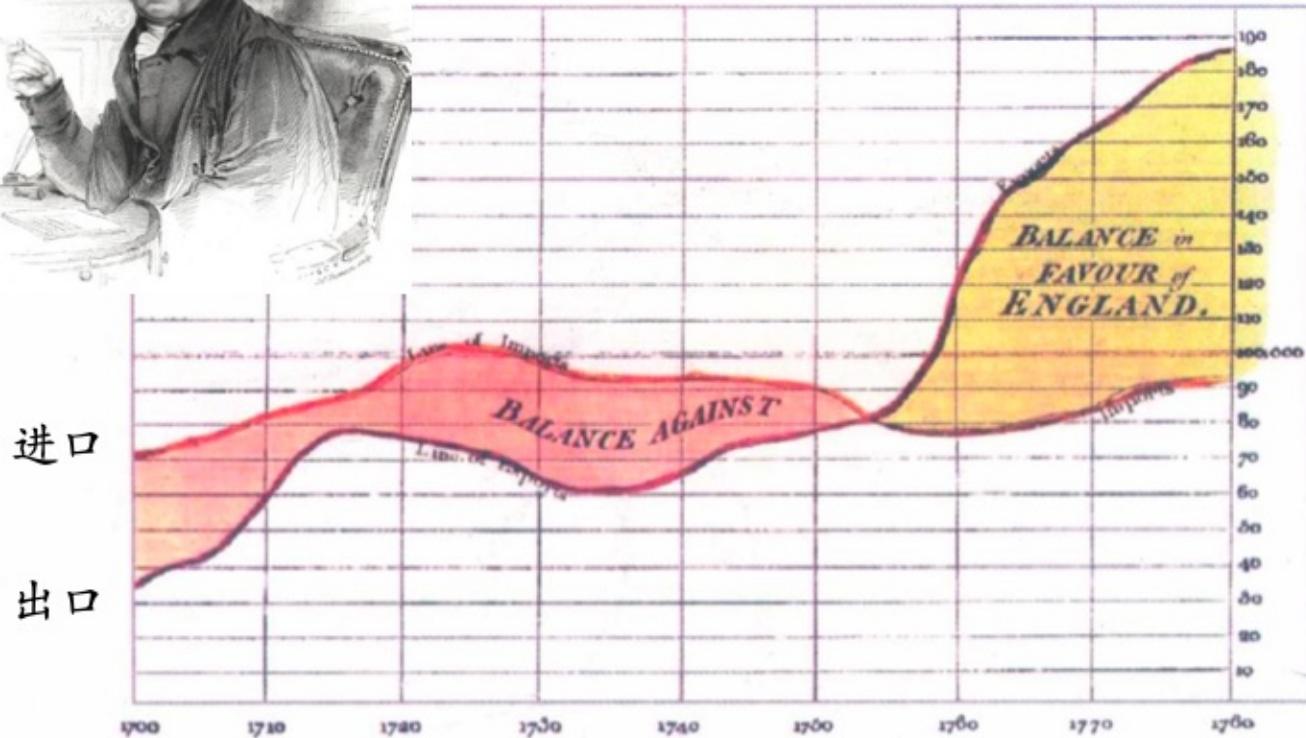
十八世纪



约瑟夫·普雷斯特利发明的时间线图（1765），展示了著名人物的生平

可视化简史

十八世纪

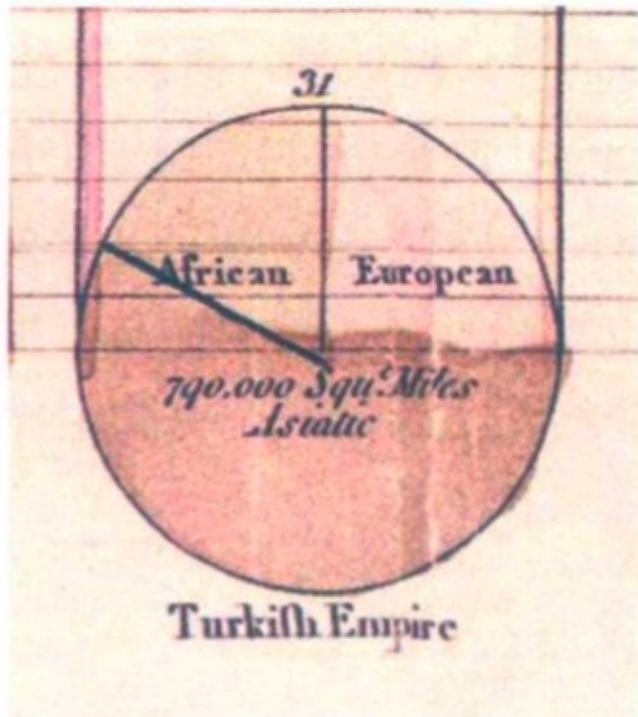


丹麦挪威联合王国 - 英国之间的贸易

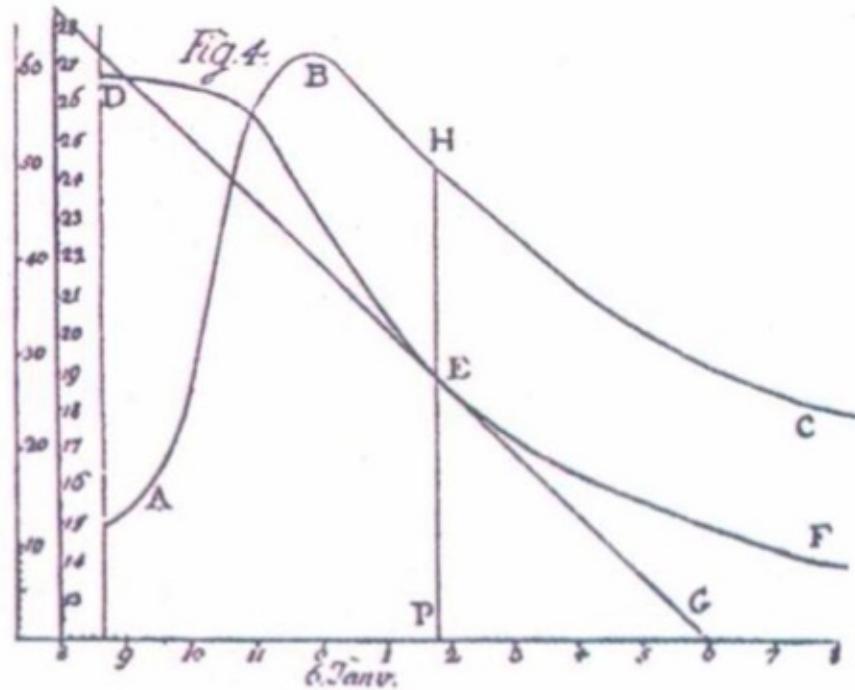
威廉·普雷菲尔 (W. Playfair)
现代可视化的奠基人，发明了折线图、柱状图、饼图

可视化简史

十八世纪



普雷菲尔绘制的第一幅饼图
1789年土耳其帝国在亚、欧、非的领土比例



水的蒸发与时间的关系（时间序列）

可视化简史

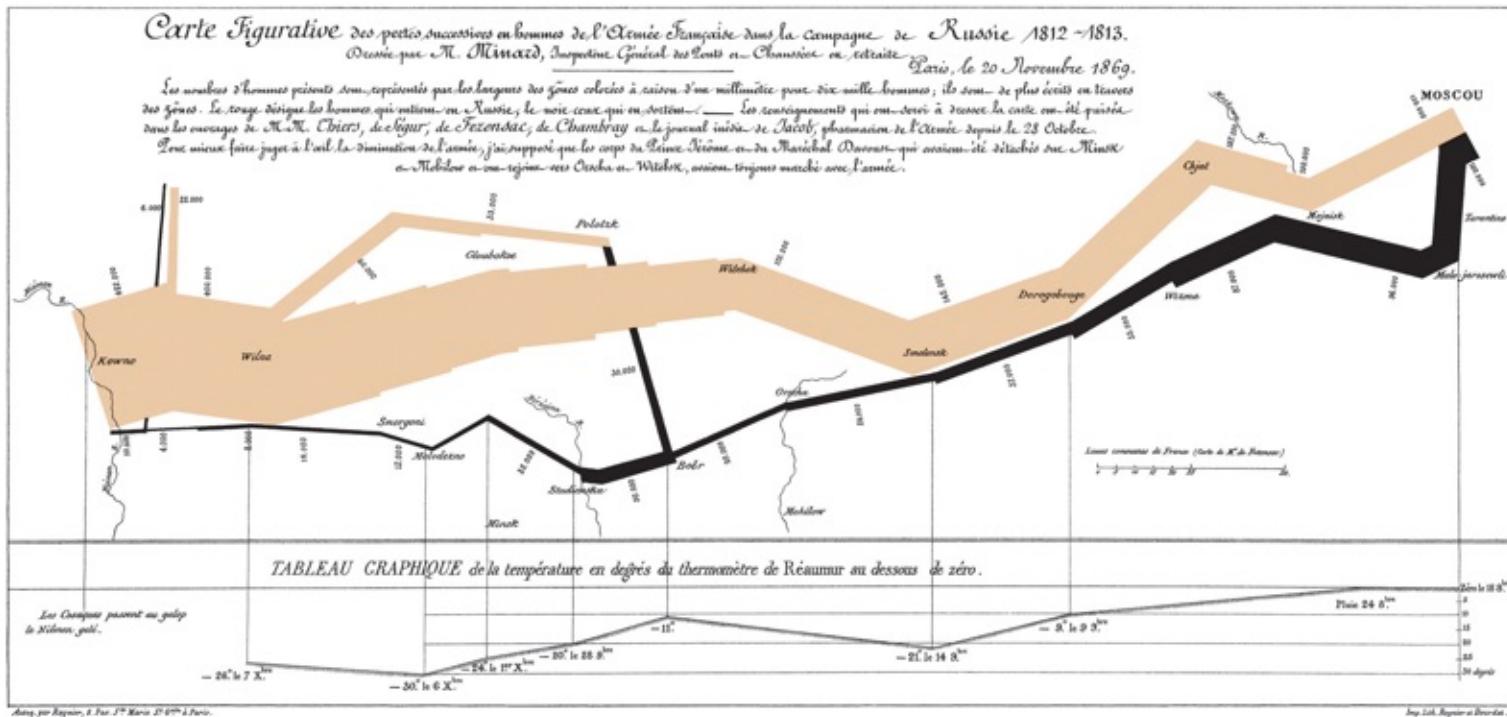
十九世纪



历史上第一幅流图
(flow chart)，用可
变宽度的线段显示爱
尔兰铁路交通运输轨
迹和乘客数量
(1837)

可视化简史

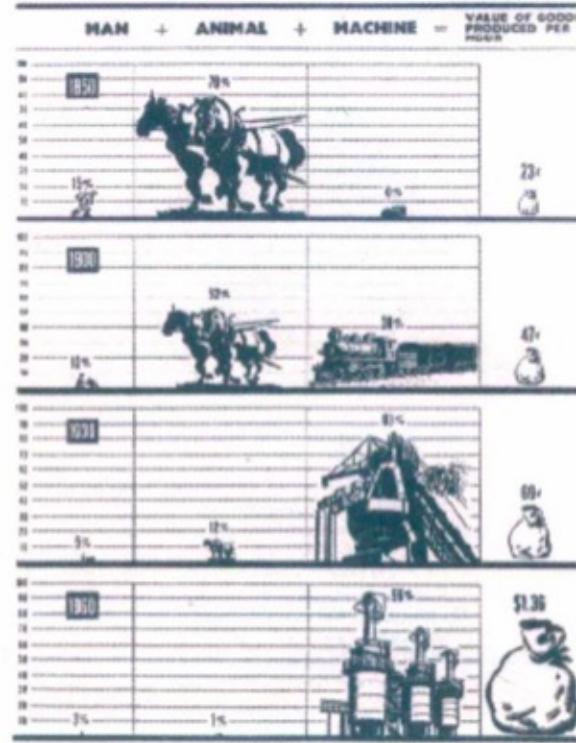
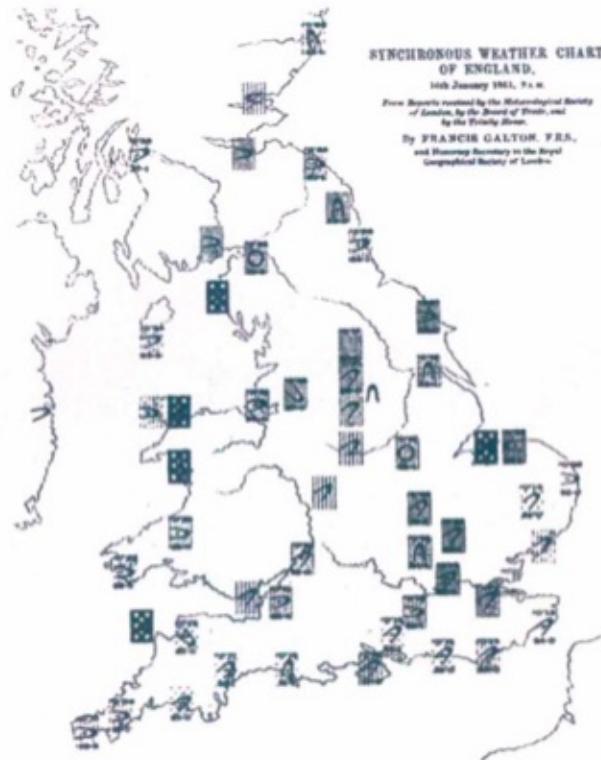
十九世纪



Charles Joseph Minard
拿破仑行军图 (1869)

可视化简史

十九世纪

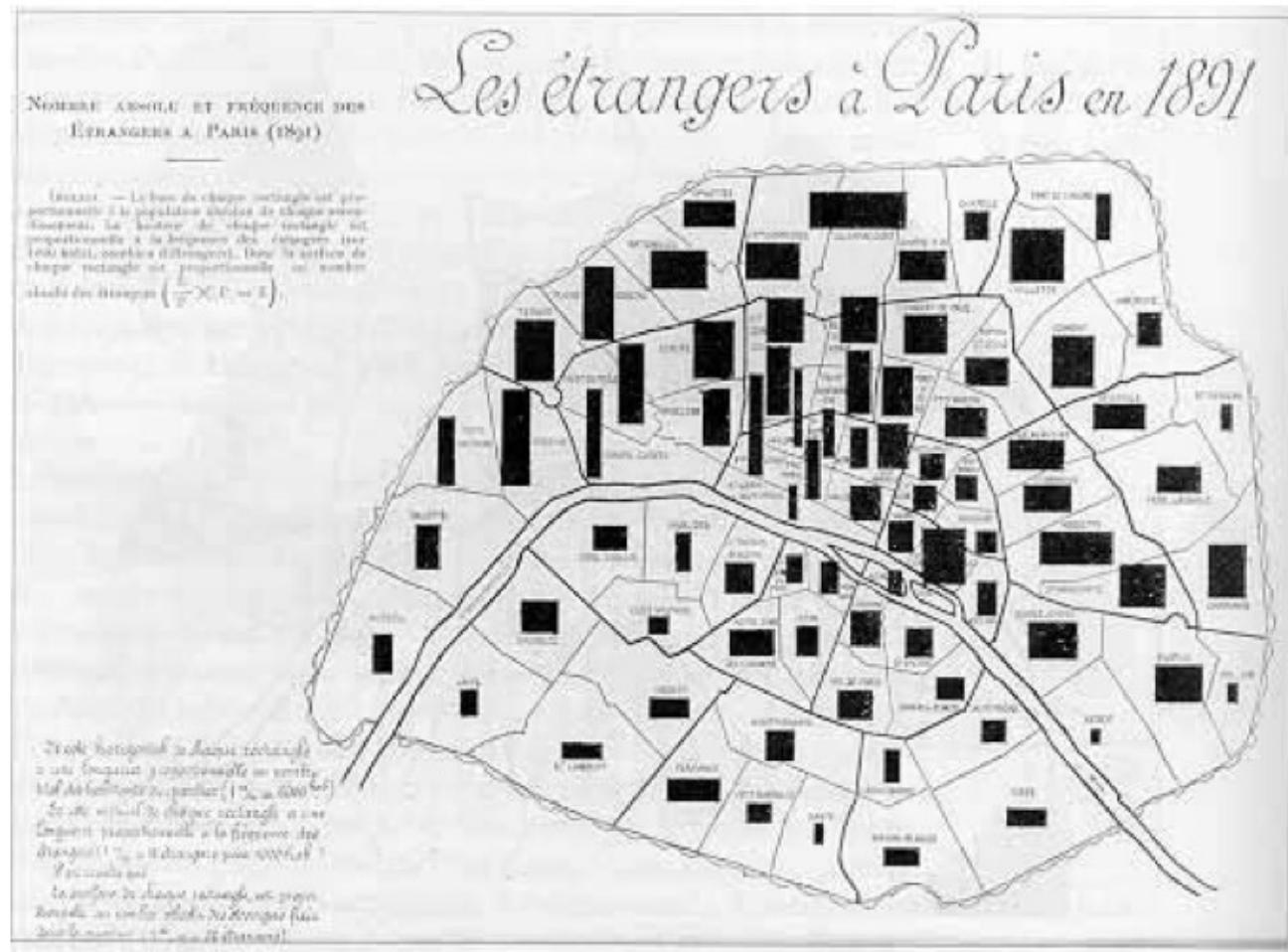


左：1861年的英国天气图（各地气压变化），帮助发现了低压区域中的风场反气旋移动现象

右：1884年的图标可视化作品

可视化简史

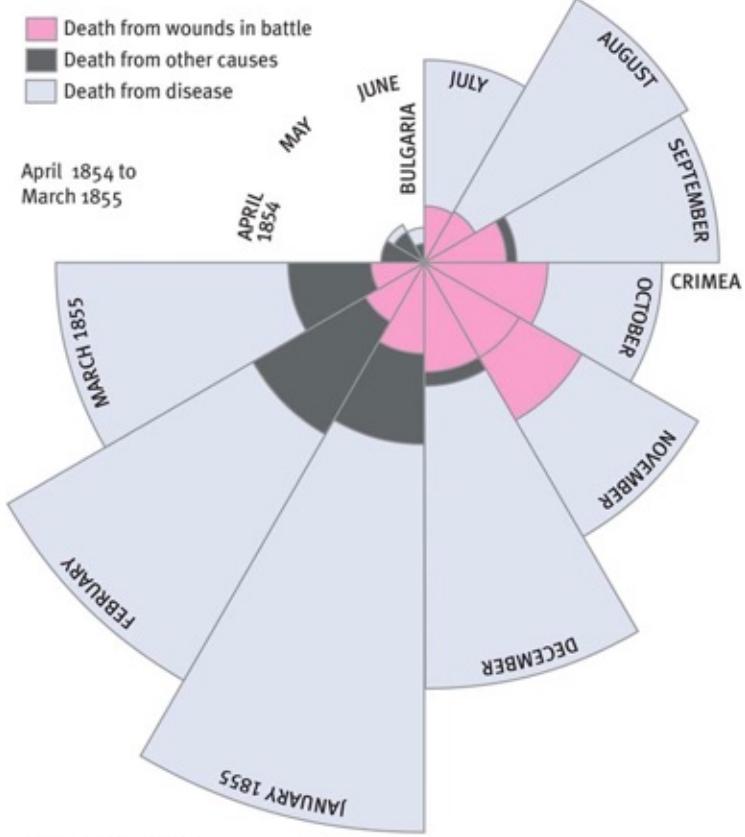
十九世纪



Jacques Bertillon 巴黎地图，采用长方形图标显示两个变量：该区人口占巴黎人口的比例、外国人占该区人口的比例。
长方形的面积：法国巴黎各区的外国人数目。

可视化简史

十九世纪



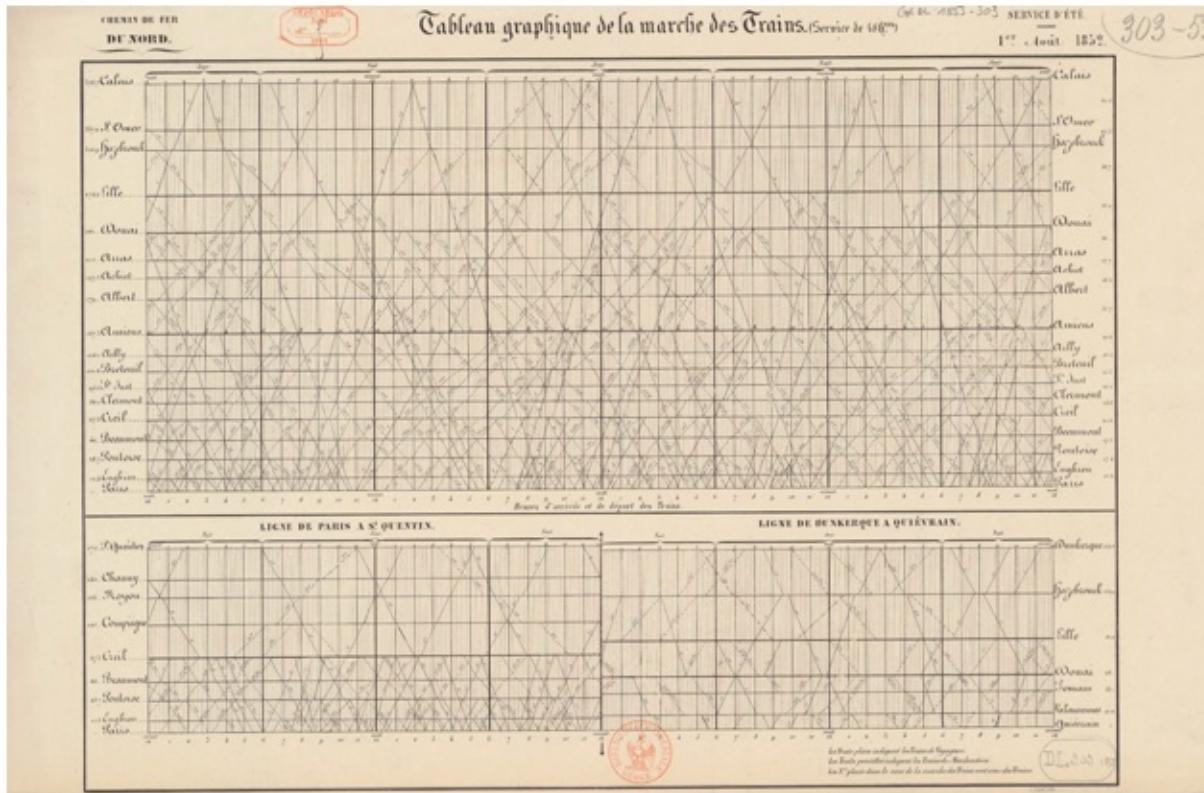
© 2015 Worth Publishers



南丁格尔玫瑰图（克里米亚战争）

可视化简史

十九世纪



图形化列车时刻表（1852 法国）
Étienne-Jules Marey (通常认为的绘制人)
Charles Ibry (实际上的绘制人)

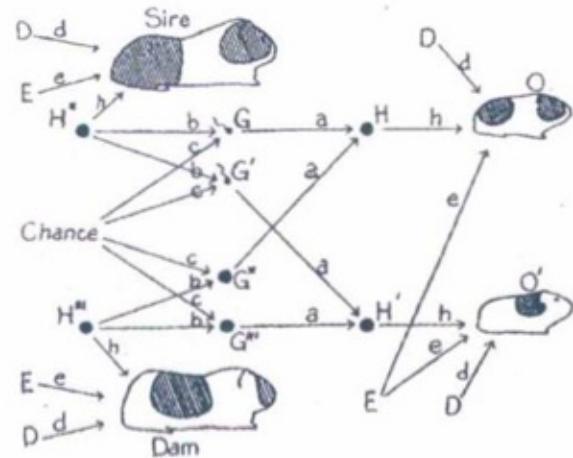
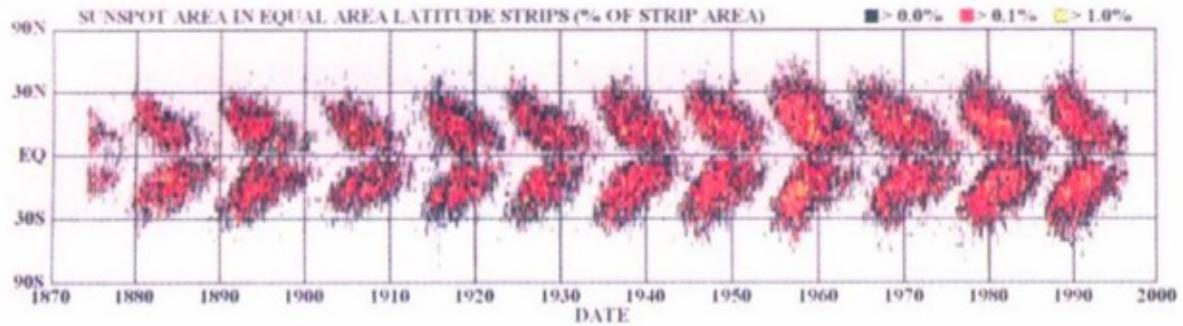
可视化简史



可视化简史

二十世纪

更大的影响力、更多的应用范围、更多维的数据



左：太阳黑子隋时间扰动 蝴蝶图（1904）

右：结构方程系统路径图（1920）

可视化简史

二十世纪



Henry Beck伦敦地铁图（1933）

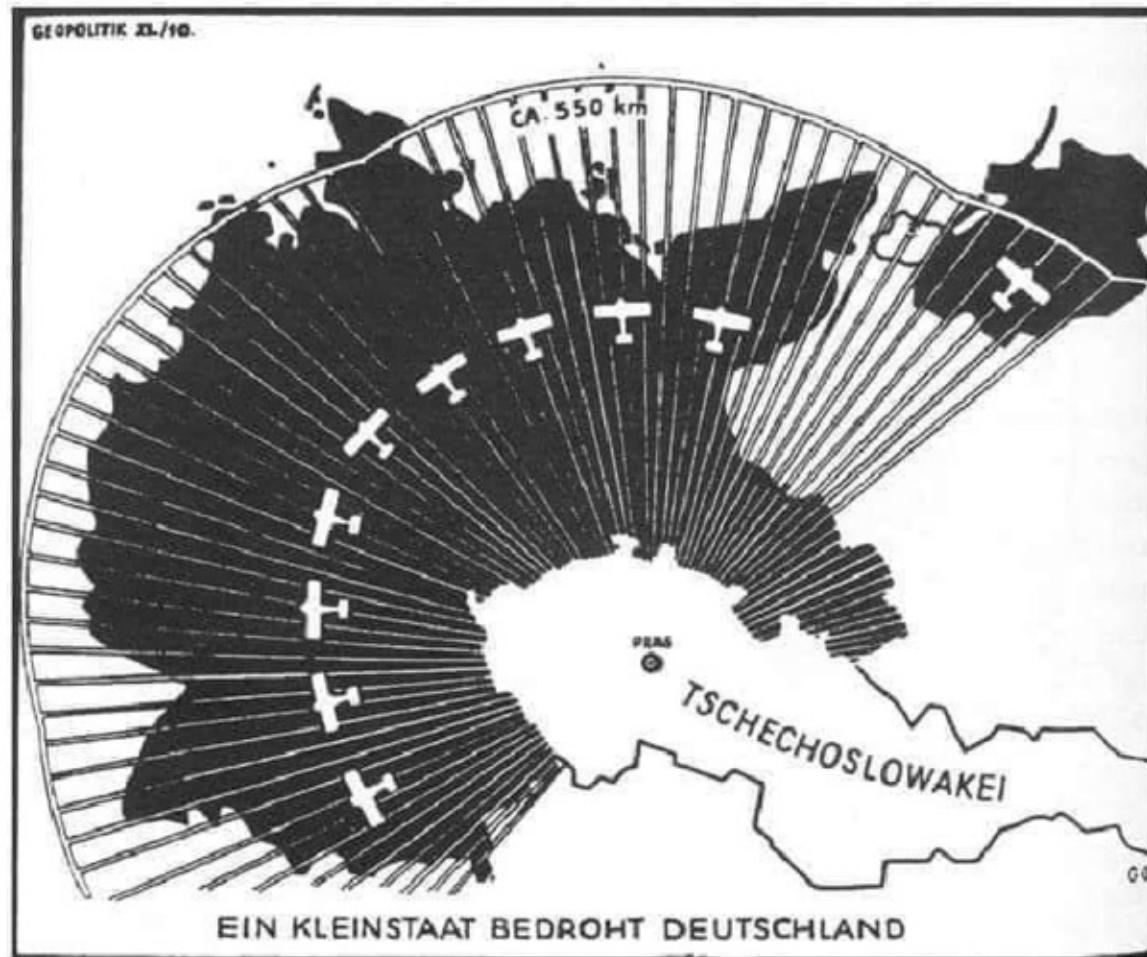
奠定了Schematic Map的标准绘制法

把所有线条按拓扑关系抽象为水平、竖直和45°斜线

可视化简史

二十世纪

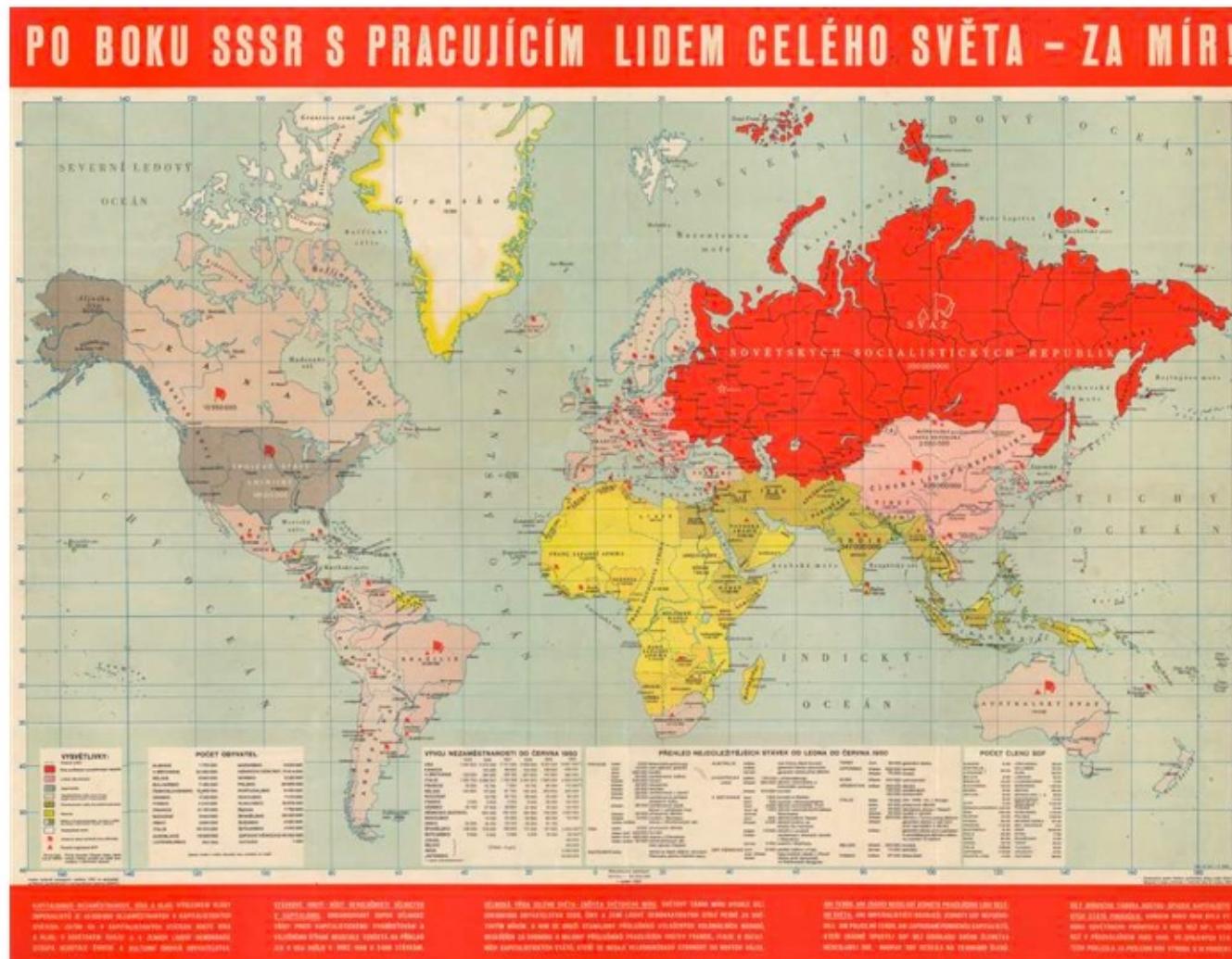
心理学介入可视化



二战前德国为进占苏台德地区而发行的宣传地图

可视化简史

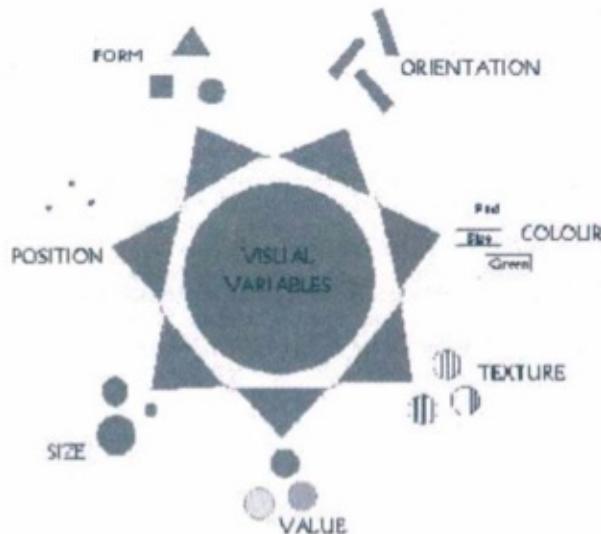
二十世纪



麦卡托投影：保持陆地轮廓、放大高纬度地区

可视化简史

- 1967年，法国人Jacques Bertin出版了Semiology of Graphics（《图形符号学》），确定了构成图形的基本要素，包括图形符号、视觉通道、色彩空间、纹理、大小、位置等内容，从理论层面上确立了可视化的基本框架。



Position
Size
(Grey)Value
Texture
Color
Orientation
Shape

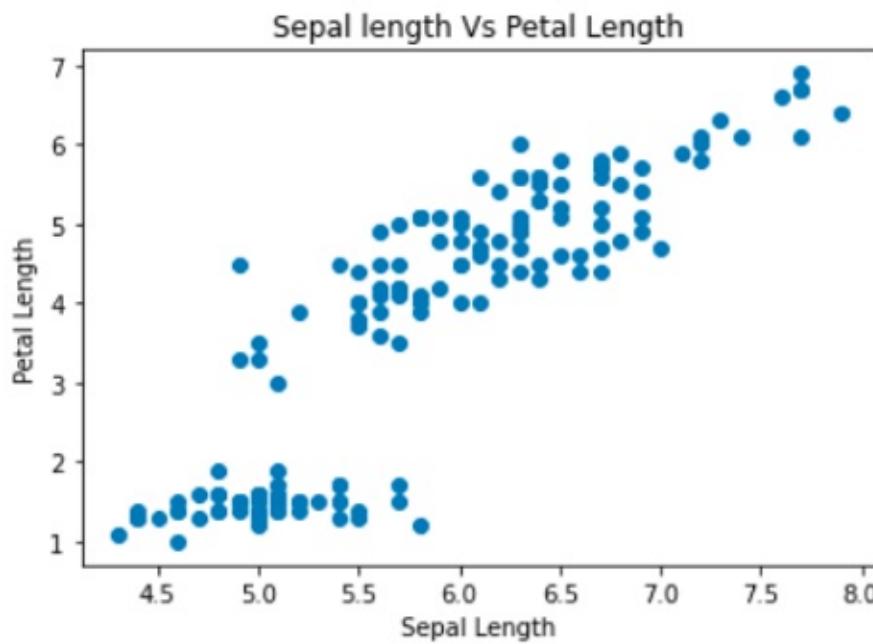
	Marks	Points	Lines	Areas
LES VARIABLES DE L'IMAGE				
XY 2 DIMENSIONS DU PLAN	POINTS	LIGNES	ZONES	
Z TAILLE				
VALEUR				
LES VARIABLES DE SÉPARATION DES IMAGES				
GRAIN				
COULEUR				
ORIENTATION				
FORME				

可视化简史

二十世纪

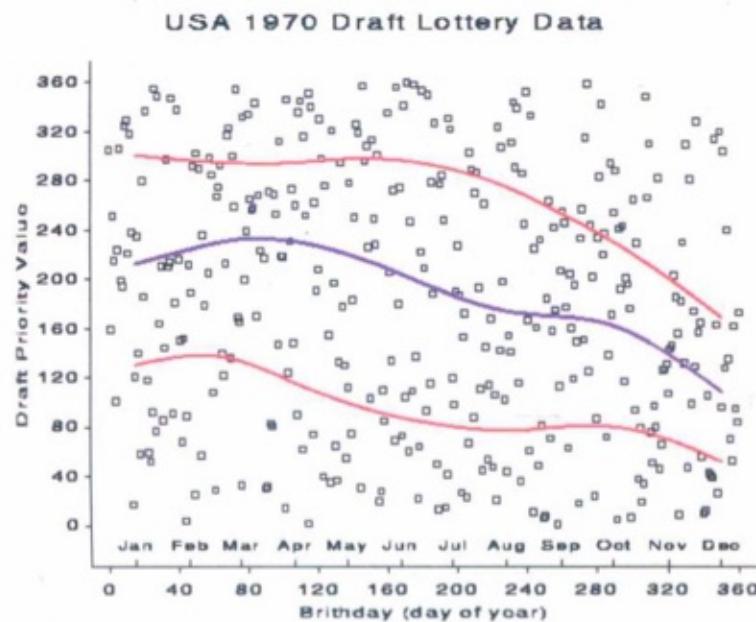


```
plt.scatter(data.sepal_length, data.petal_length)
plt.title("Sepal length Vs Petal Length")
plt.xlabel("Sepal Length")
plt.ylabel("Petal Length")
plt.show()
```

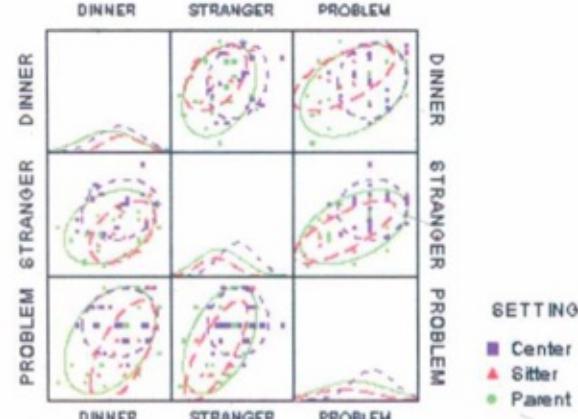


可视化简史

二十世纪



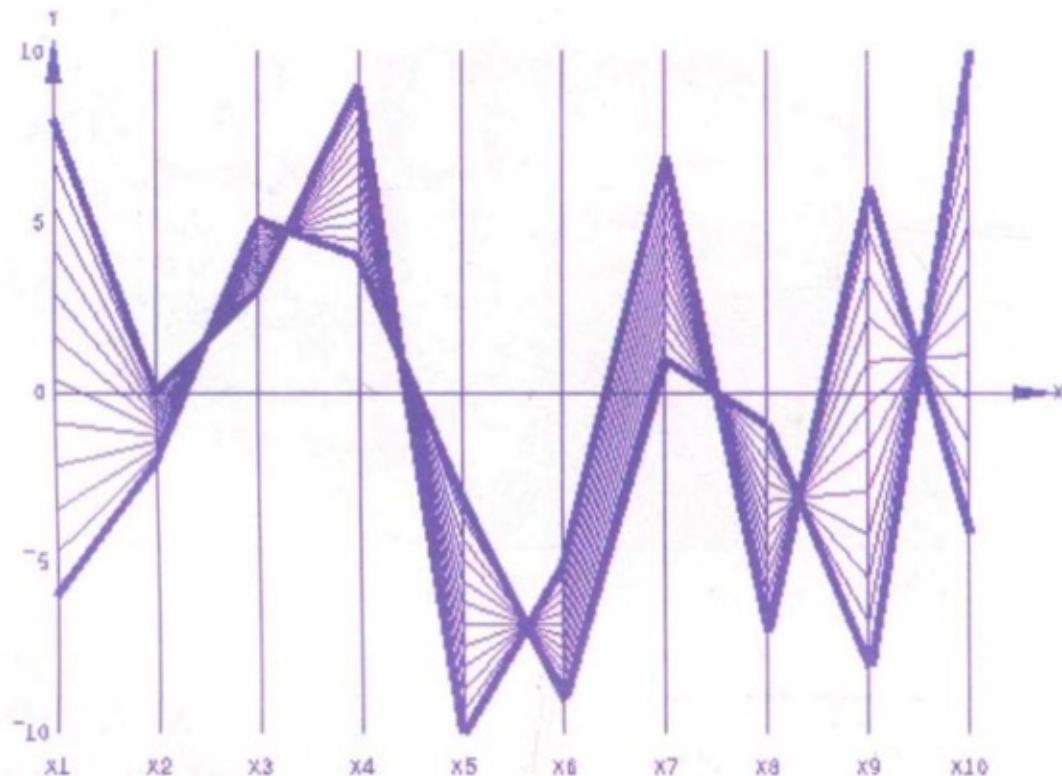
Social Competence Measures Across Settings



左：1975年发明的移动统计均线（增强散点图的表达能力）
右：John Hartigan发明的散点图矩阵

可视化简史

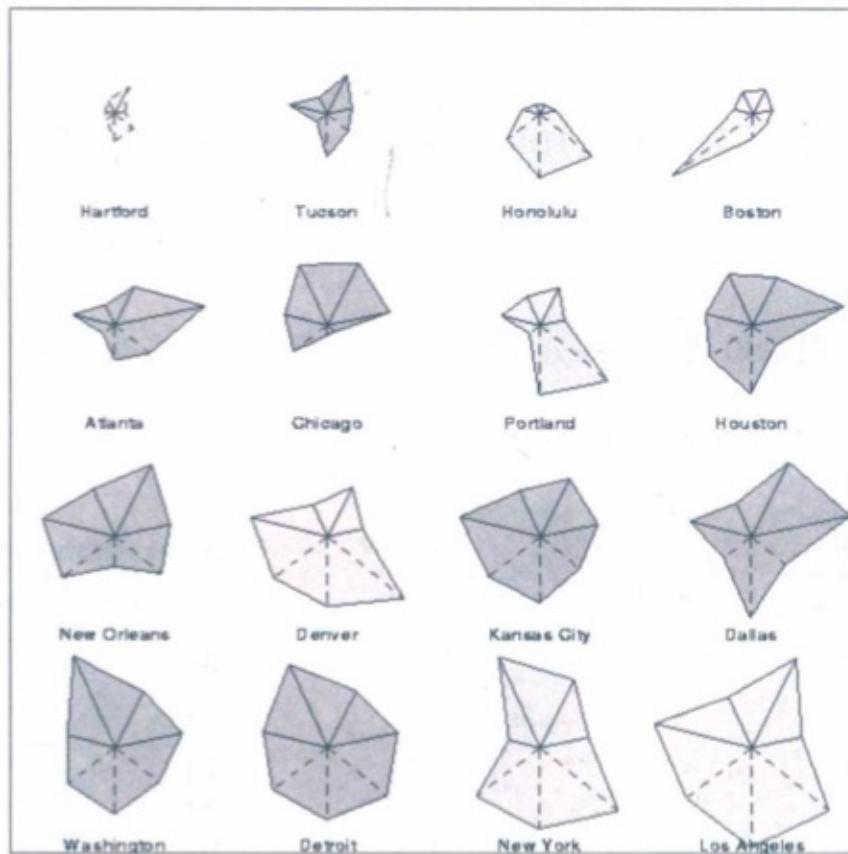
二十世纪



1985：平行坐标体系（用于表达高维数据）

可视化简史

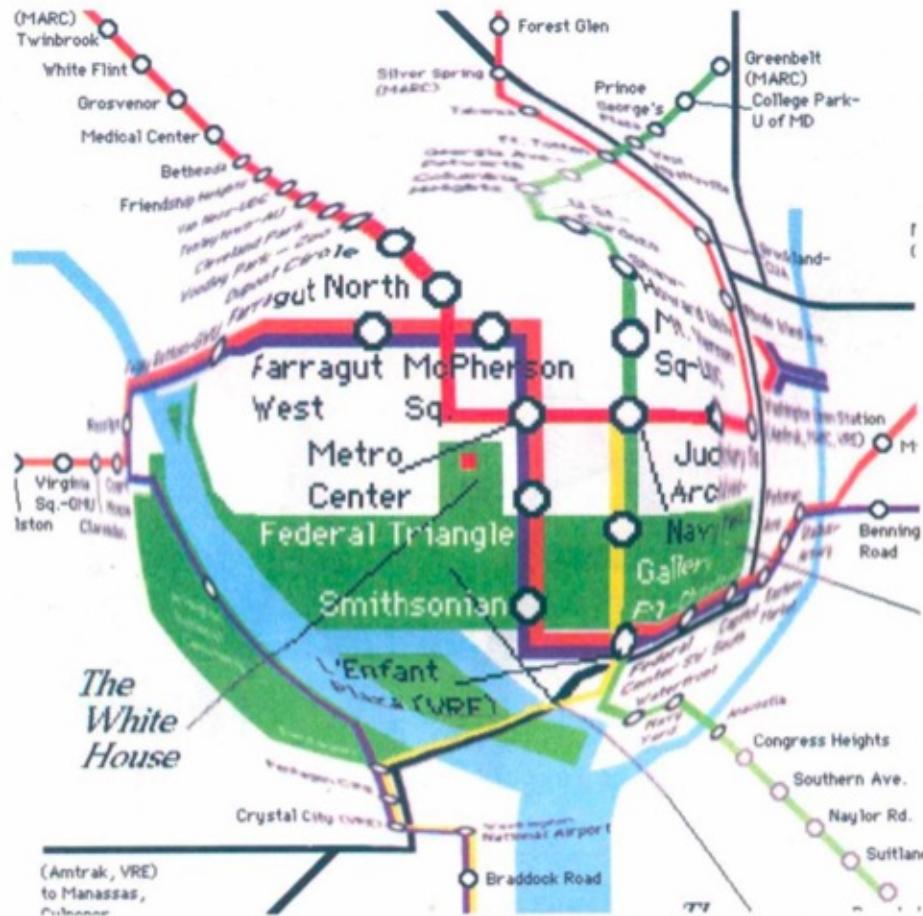
二十世纪



表达多变量数据的不规则多边形图，或称星形图（1971年发明）

可视化简史

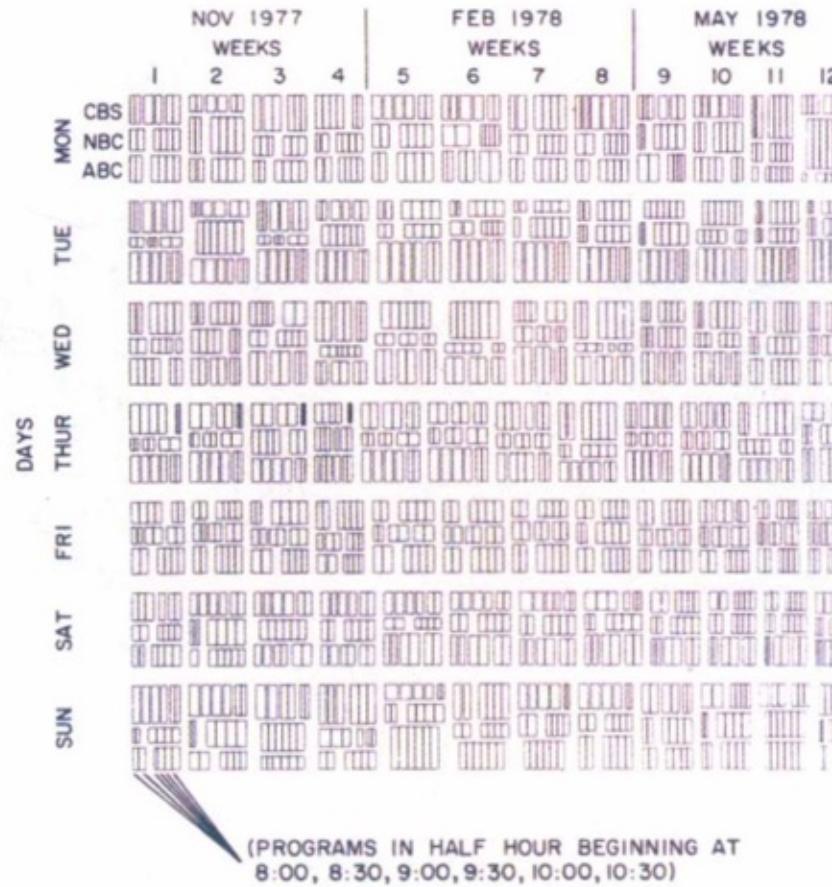
二十世纪



1981：鱼眼图法，对重要细节提供更高的专注度

可视化简史

二十世纪



John Hartigan发明的表达多维类别型数据的马赛克图（1981）

可视化简史

1986年

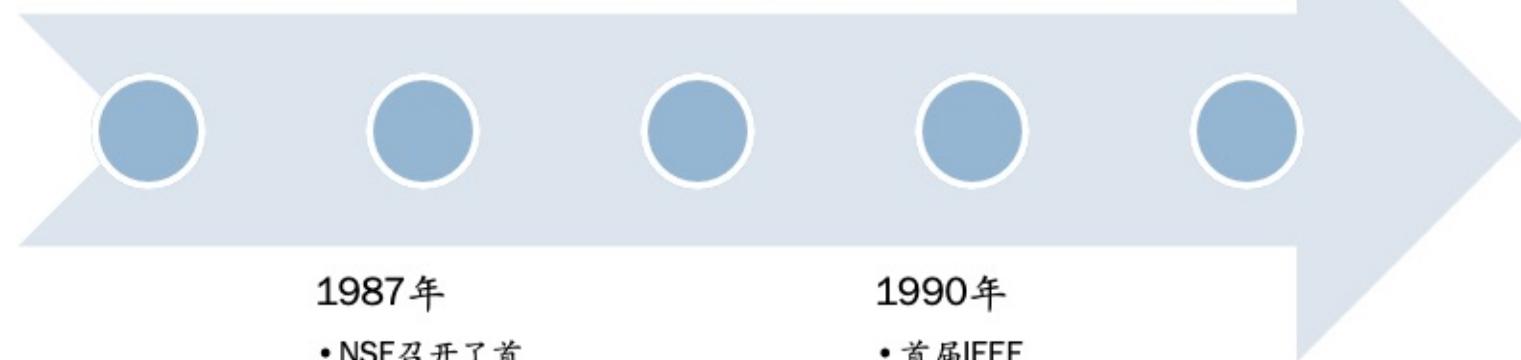
- ViSC（科学计算中的可视化）

1988年

- 面向多变量统计数据的动态可视化手段

21世纪

- 可视化分析



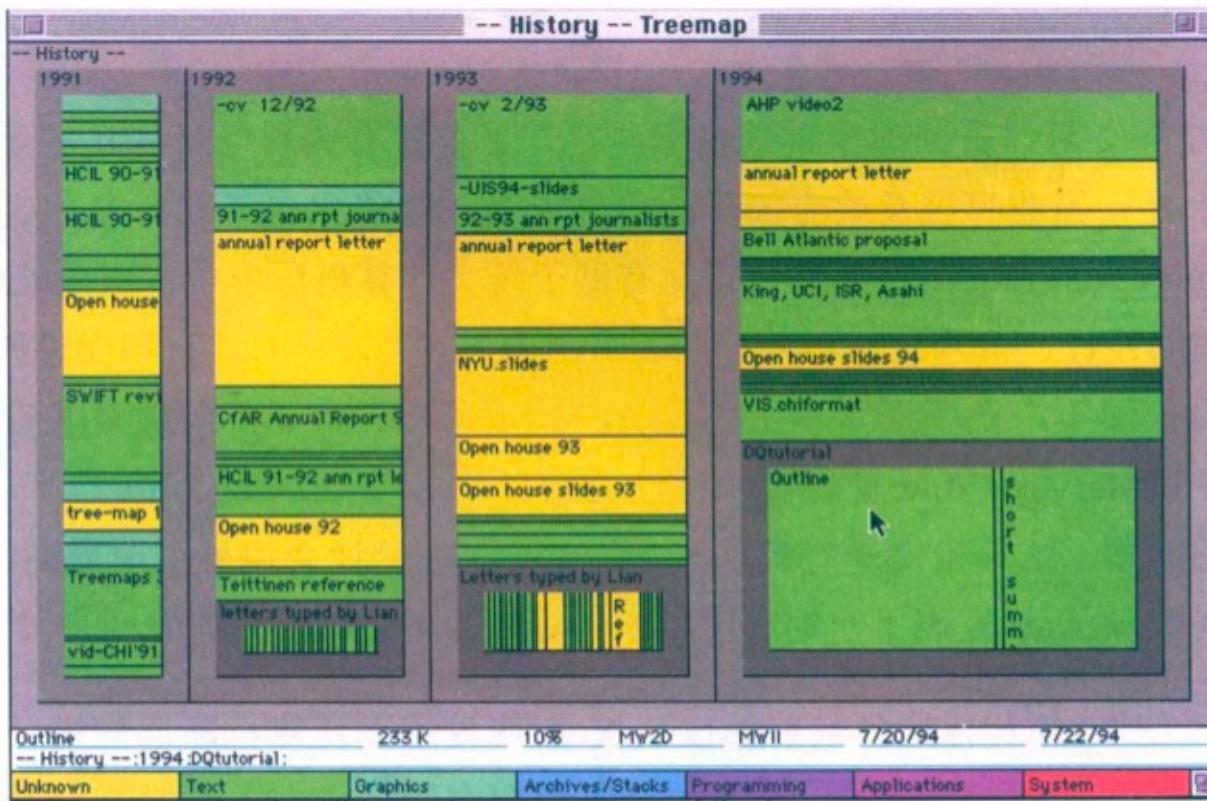
1987年

- NSF召开了首次有关科学可视化的会议
- ACM SIGGRAPH：“移动立方体法”
(Marching Cubes)

1990年

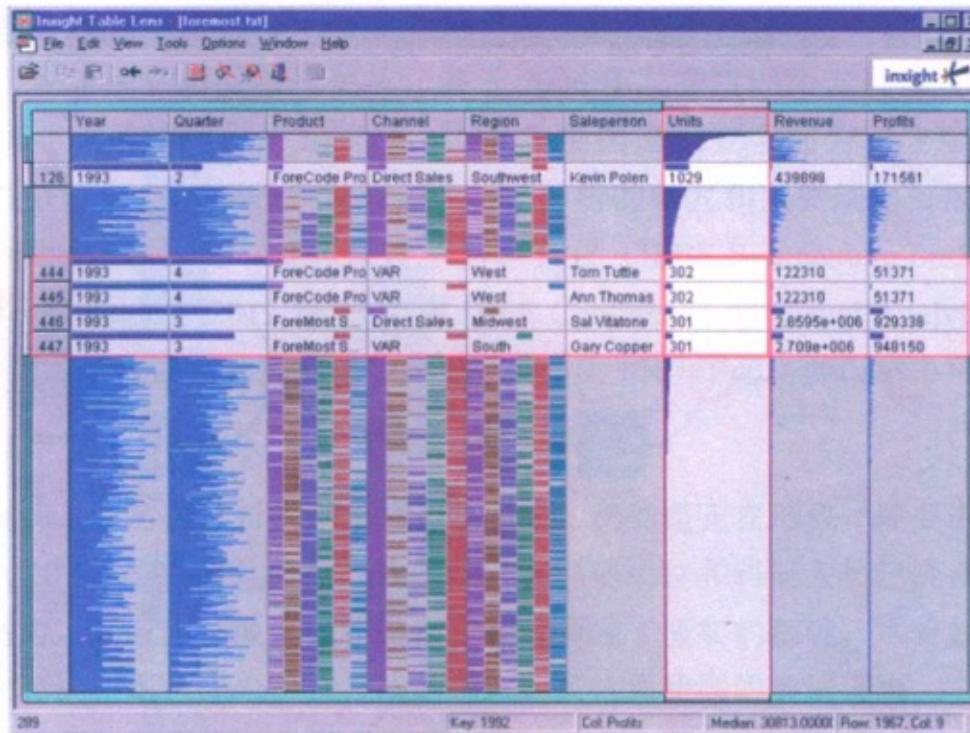
- 首届IEEE Visualization Conference

可视化简史



树图 (1991)

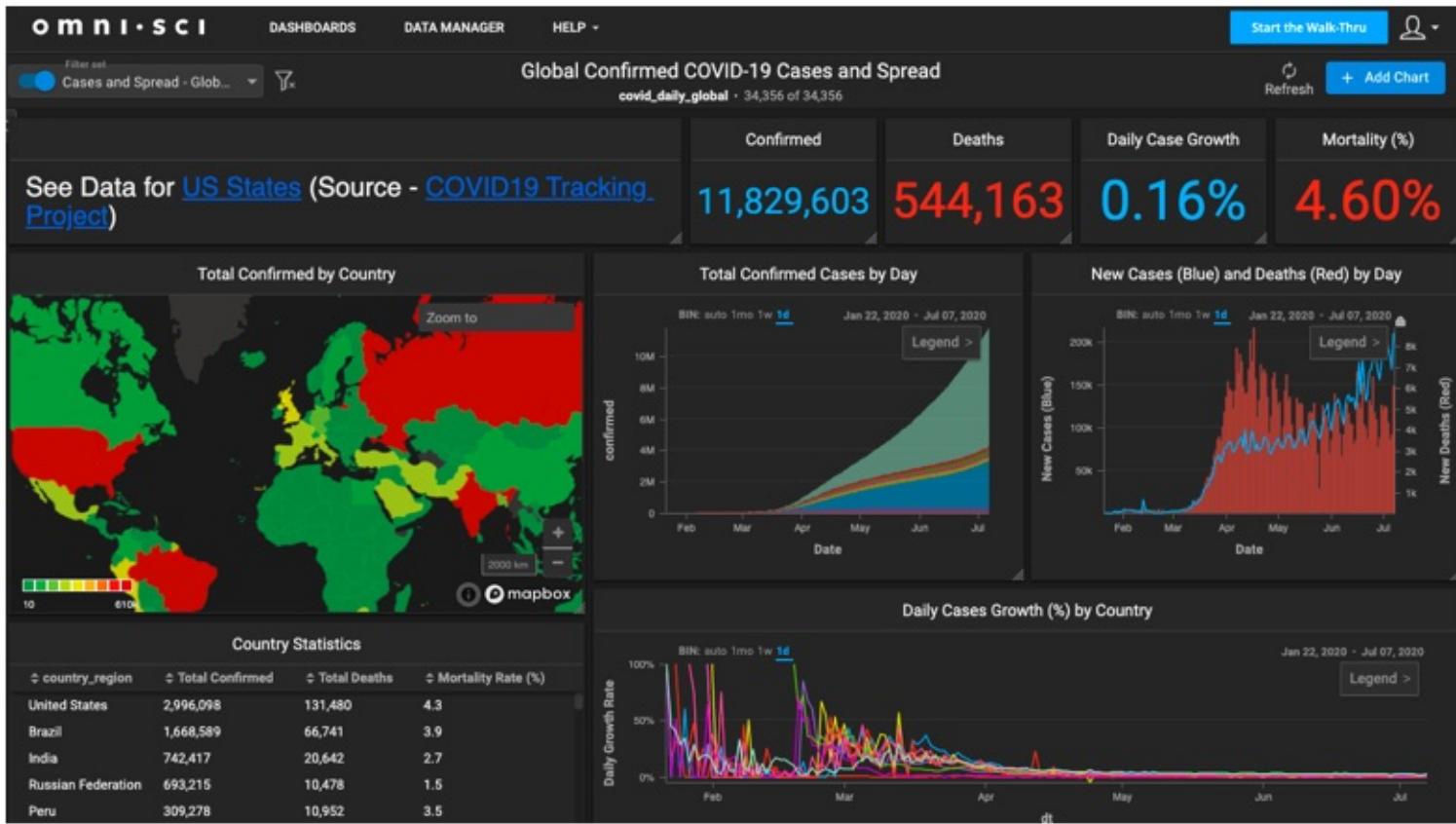
可视化简史



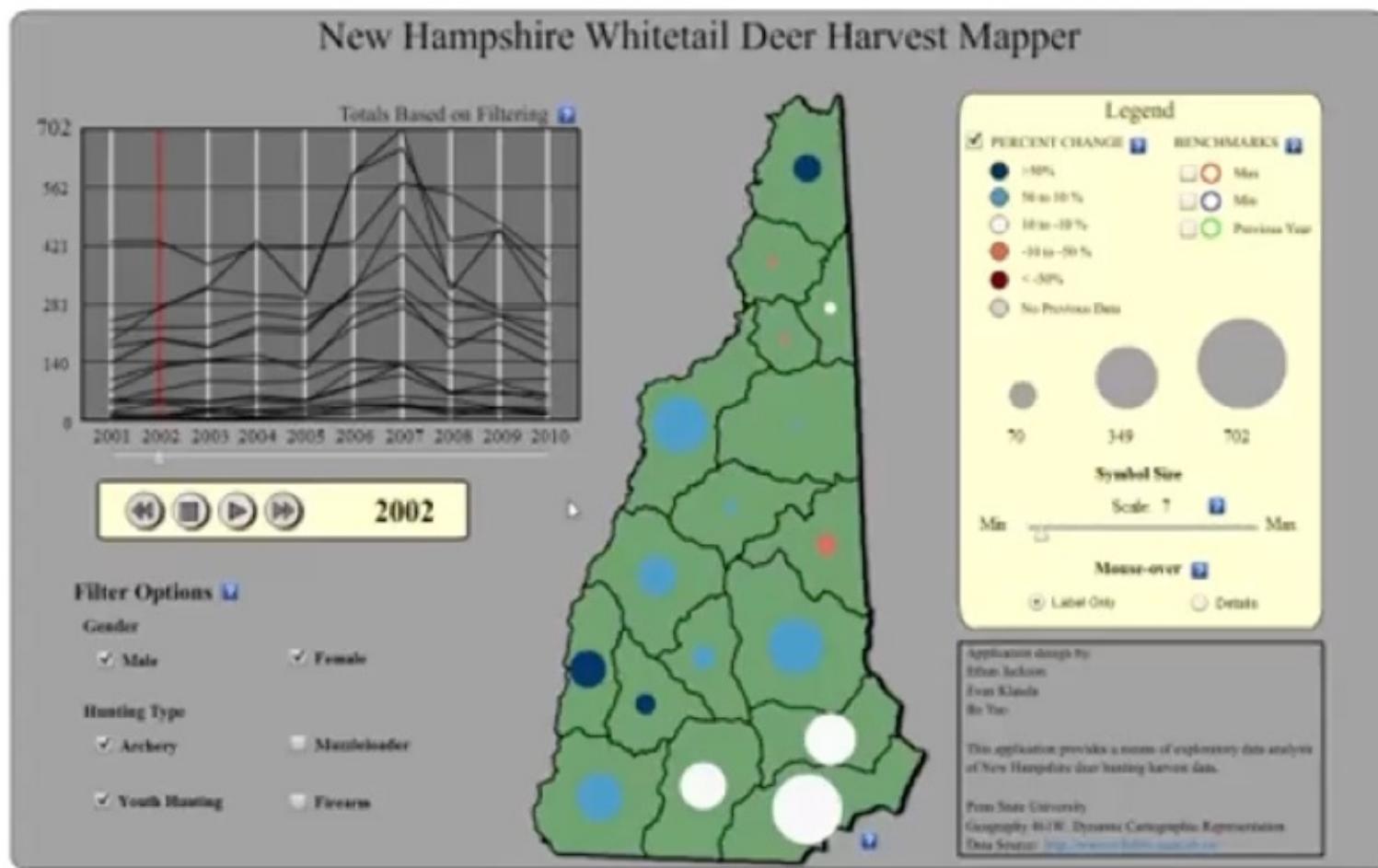
1994：美国施乐公司发明的表格透镜技术
提取表格中的某些行，并与可视化图像交互

可视化分析

- 进入21世纪，现有的可视化技术已经难以应对海量、高维、多源和动态的数据分析需求。
- 这促成了可视化分析（visual analytics）的兴起。



可视化分析



扩展资料

Milestones in the History of Data Visualization

FROM THE ORIGINAL

"Milestones in the History of
Thematic Cartography,
Statistical Graphics,
and Data Visualization"

An illustrated chronology of innovations
by Michael Friendly and Daniel J. Denis, York
University, Canada

Organization by Mario Kanno
(www.infografie.com.br)



见“学习通”



Visual Perception and Color Space

视觉感知和色彩空间

中国地质大学（北京）信息工程学院

叶山

yes@cugb.edu.cn

Pop Quiz

回答正确得2分
回答错误得1分
缺席不得分

1. 从宏观角度看，可视化的三个功能不包括哪一个？

- A. 信息记录 | B. 修饰数据 | C. 辅助分析和推理 | D. 信息传播与协同

2. 以下哪种说法不正确？

- A. 可视化必须有目的性，能协助人们完成某项任务。
- B. 可视化是通过视觉观感来传递信息的。
- C. 可视化作品一定是美观的。
- D. 在设计可视化时，要考虑它的受众群体。

3. 以下哪种说法正确？

- A. 在整个视野中，视觉处理是并行的。
- B. 人眼的信息输入最高带宽大约为每秒10MB。
- C. 大约30%的人脑功能被用于视觉的感知。
- D. 人脑从外界获取的信息中，大约有50%是通过视觉输入的。

4. 可视化技术的演化和以下哪门技术的发展无关？

- A. 地图制图 | B. 科学制图 | C. 统计制图 | D. 仿真制图

5. 欧洲中世纪时期最具代表性的地图类型是？

- A. 波尔图兰海图 | B. T-O地图 | C. 绳结地图 | D. 地图册

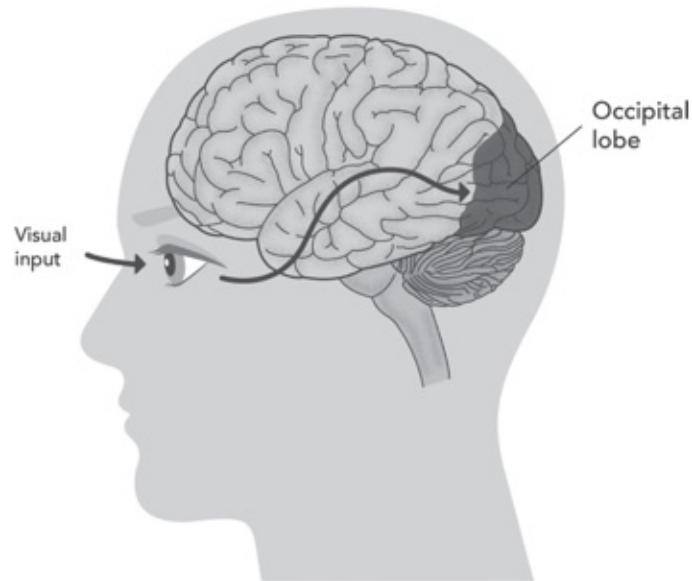


视觉感知 格式塔理论

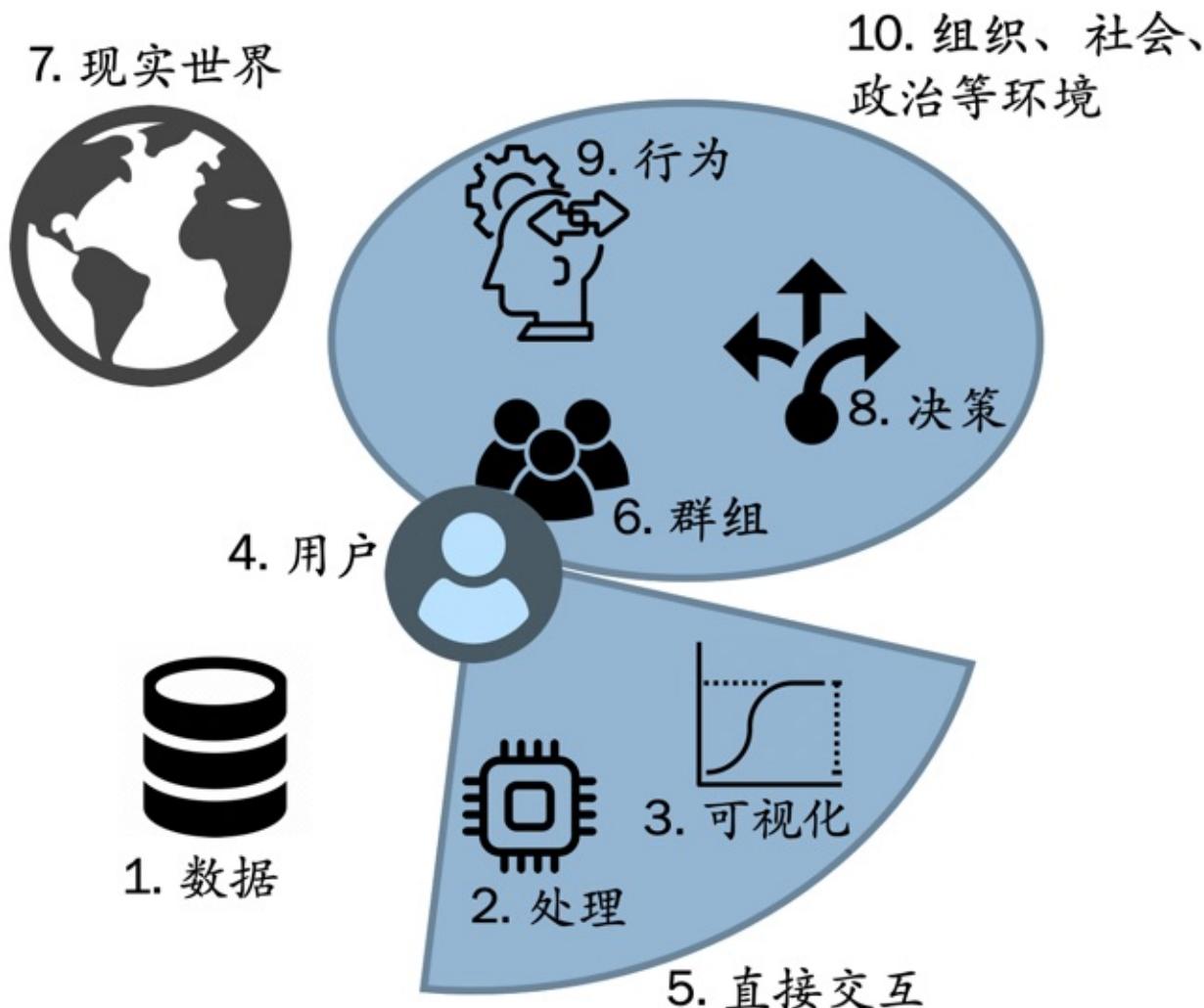


视觉感知

- 在可视化与可视分析过程中，用户是所有行为的主体。
- 用户通过视觉感知（Visual Perception）器官（眼睛）获取可视信息、形成认知（Cognition），从而在交互分析过程中获取解决问题的方法。
- 在这个过程中，感知和认知能力，直接影响信息的获取、进程的处理，进而影响用户对外在世界环境做出的反应。



视觉感知



视觉感知

低阶视觉

- 与物体的物理性质相关的视觉信息。
- 包括物体的形状、大小、颜色、表面材质等。

高阶视觉

- 识别出物理性质及现象背后的原因、逻辑。

视觉感知

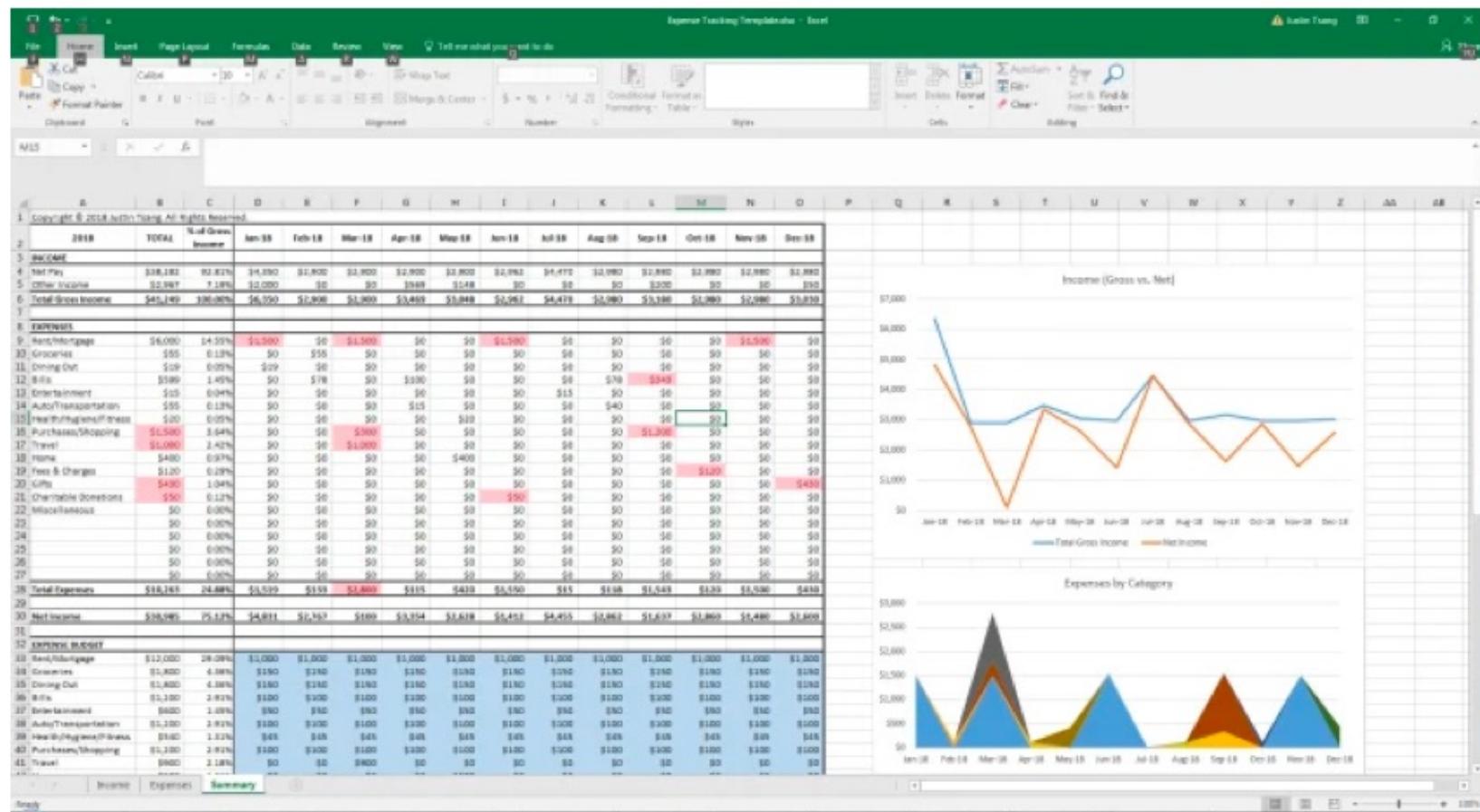
- 前注意视觉：对于突出的要素进行快速辨认。

有多少个“3”？

129456780823071201302016262684879805613618498101049062618090941040914
063648077988938092866587091642980.

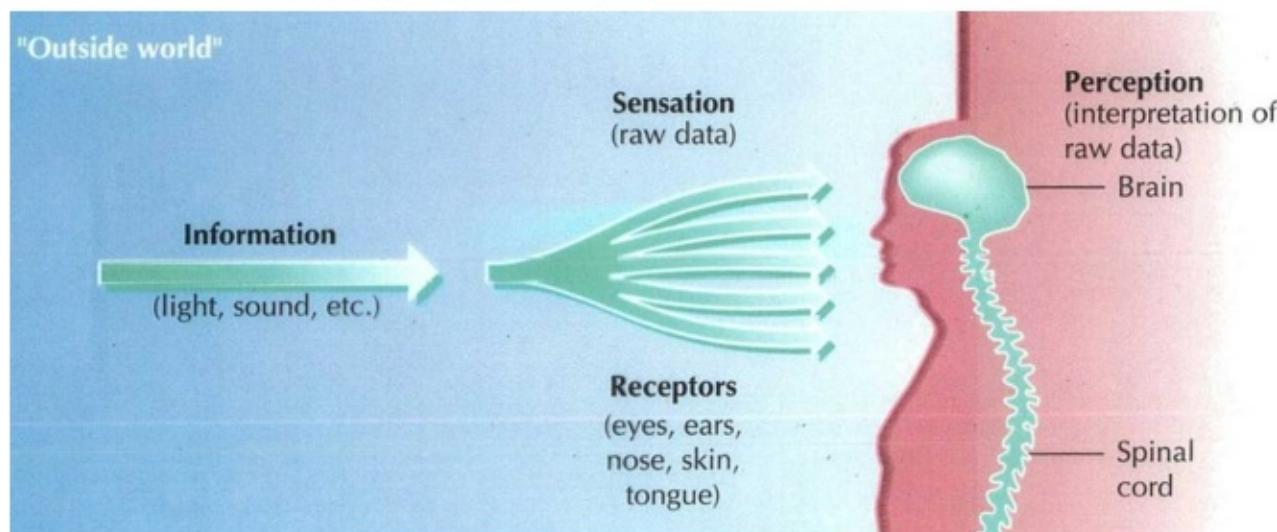
129456780823071201302016262684879805613618498101049062618090941040914
063648077988938092866587091642980.

视觉感知



视觉感知

- 感知 (sensing/sensation)：客观事物通过感觉器官，在人脑中的直接反映。
 - 五种感知：视觉、嗅觉、听觉、味觉、触觉。
 - 感知器官 (receptors)：眼、鼻、耳、舌、皮肤。
 - 可类比为数据接收的硬件装置（传感器、摄像头、温度计等）。
- 认知 (perception)：一个人在感知的过程中，对感觉信号进行加工处理的过程。
 - 加工方式：进行接收、检测、转换、简约、合成、编码、储存、提取、重建、概念形成、判断等。
 - 可类比为数据处理的软件和算法。



视觉感知

认知心理学将认知过程看成由信息的获取、编码、储存、提取和使用等一系列认知阶段按一定程序进行信息加工的系统。

获取

- 感觉器官接受客观世界的刺激，通过感觉的作用获得信息。

编码

- 以利于后续认知阶段的进行。

储存

- 信息在大脑里的保持。

提取

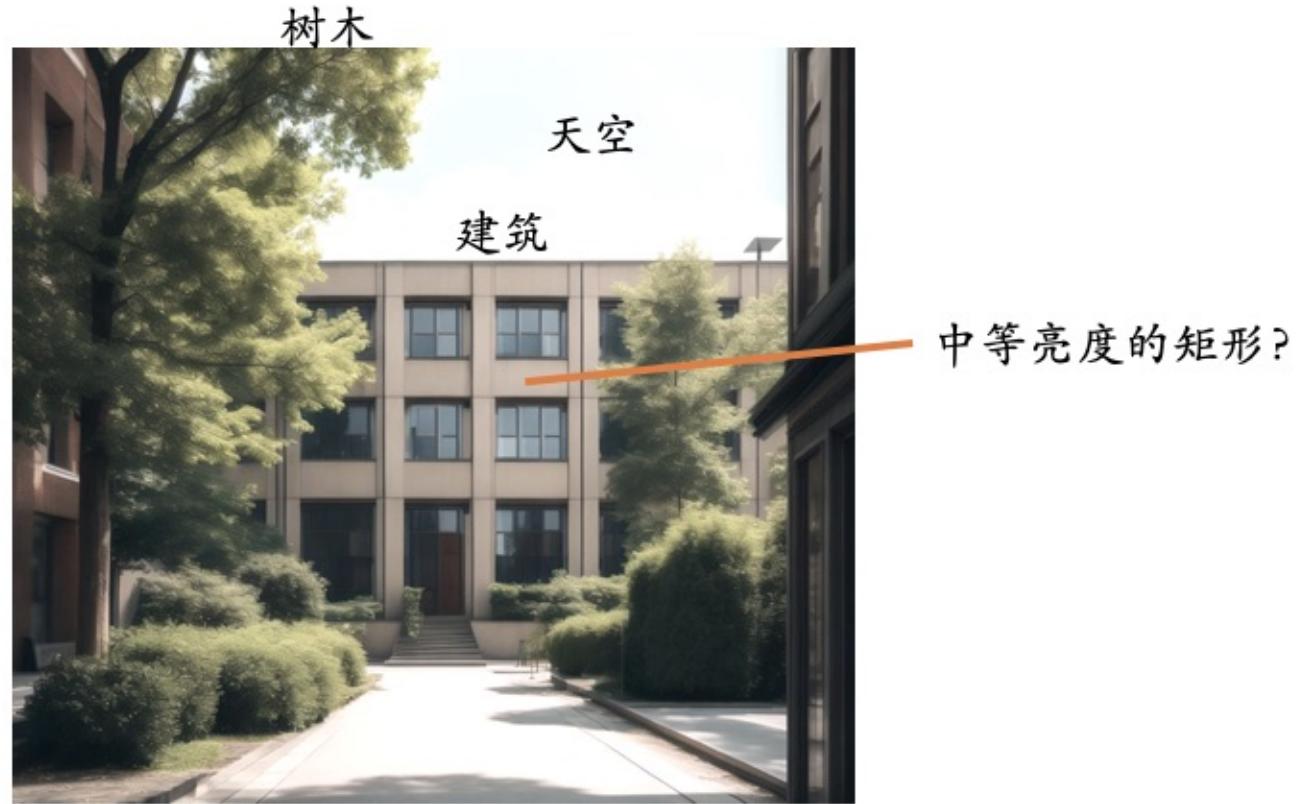
- 依据一定的线索，从记忆中寻找并获取已经储存的信息。

使用

- 利用提取的信息，对信息进行认知加工。

格式塔理论

- 格式塔 (Gestalt) : Shape (形状) / Form (构成)

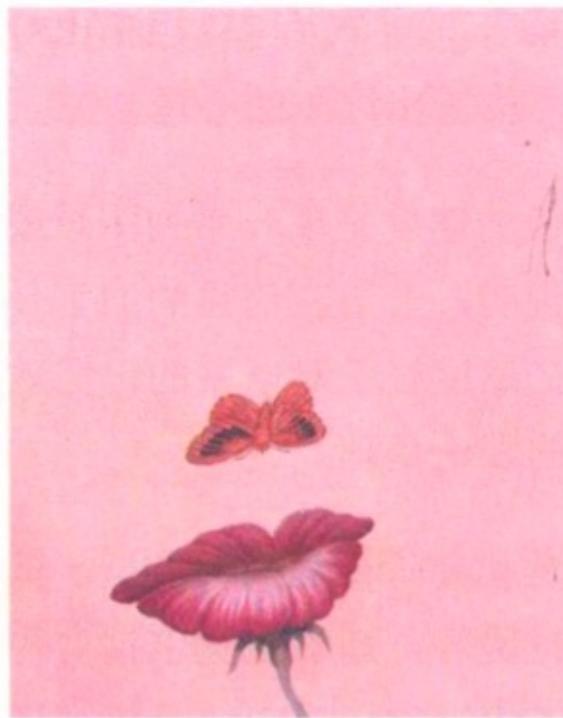


视觉感知 > 看到的东西

格式塔理论



格式塔理论



格式塔理论

- 简单精炼法则：人在进行观察的时候有两个倾向
 - 1. 将视觉感知到的内容理解为一些基本的结构（常规的、简单的、相连的、对称的、有序的结构）。
 - 2. 将事物理解为一个整体，而不是各个组成部分的集合。所以格式塔理论又称为完图法则理论。

贴近原则
(proximity)

相似原则
(similarity)

连续原则
(continuity)

闭合原则
(closure)

共势原则
(common fate)

好图原则
(good figure)

对称原则
(symmetry)

经验原则
(past experience)

突出原则
(figure/background)

格式塔理论

- 贴近原则：当视觉元素在空间上距离较近的时候，人们更倾向于把它们归为一组。



格式塔理论



格式塔理论



格式塔理论

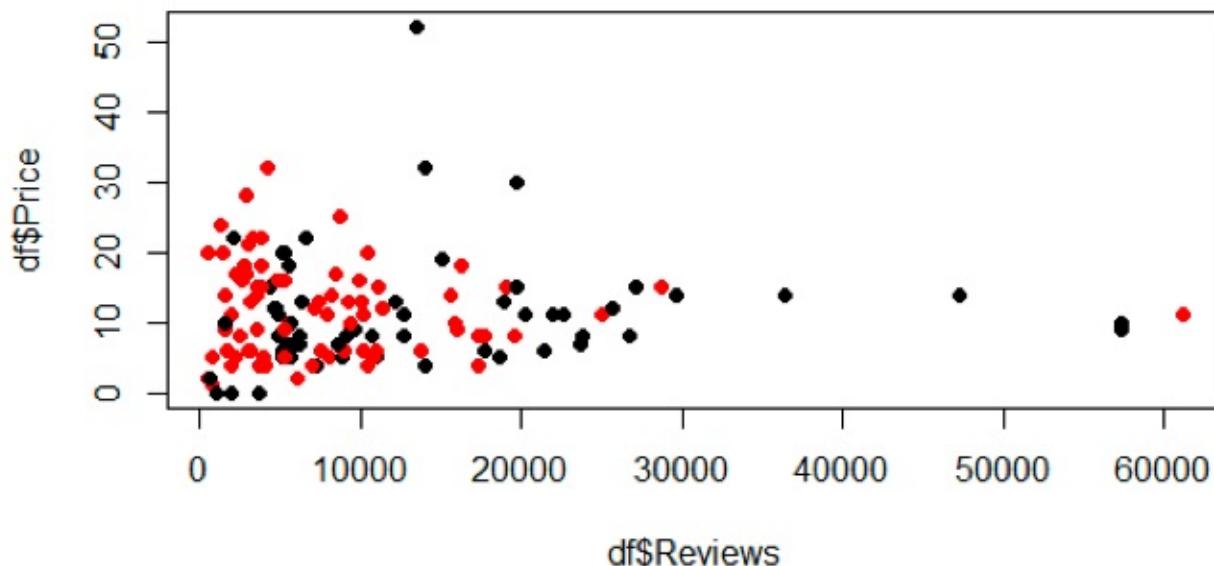
- 贴近原则与两个物体之间的距离有关，说明图像显示的方式会影响我们对事物之间关系的感知。
- 相互靠近的物体看起来像是一个群体的一部分。那么相反，相距较远的物体会被自动地理解为不同的东西。

App界面设计：



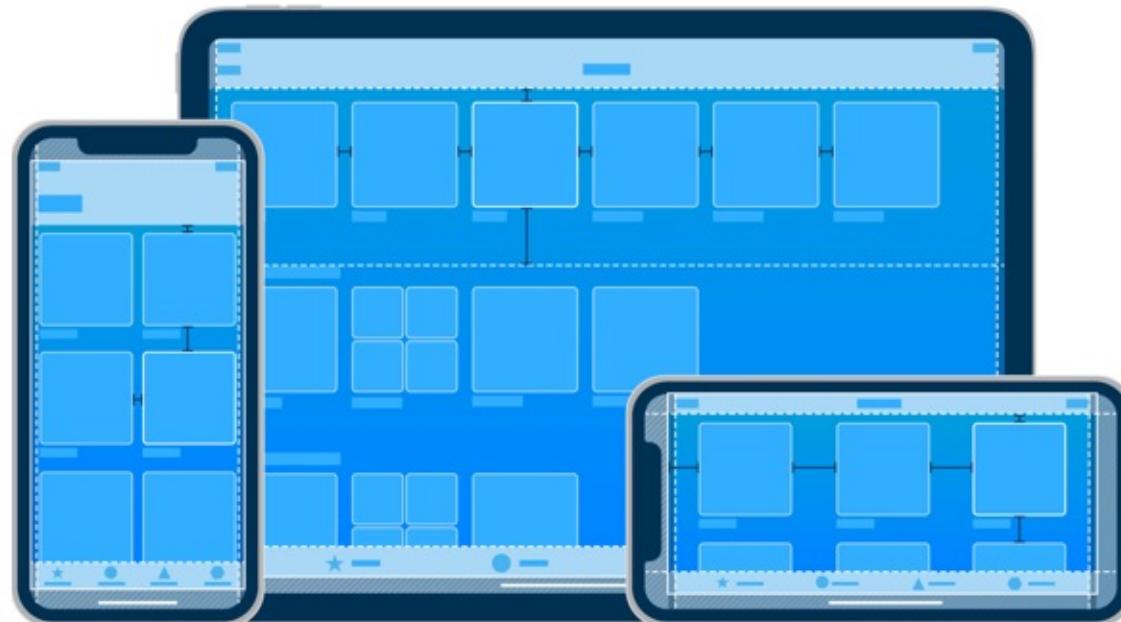
格式塔理论

- 相似原则：在观察事物的时候，人会自然地根据事物在某一个或某一些感知要素上的相似性进行感知分组，哪怕实际上事物本身并不存在分组的现象。



格式塔理论

- 人脑会努力地简化图像。



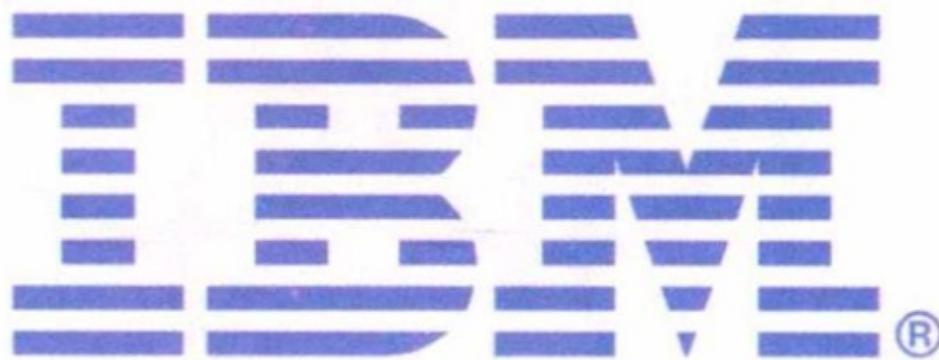
格式塔理论

- 连续原则：人在观察事物的时候，倾向于把按照线条和轮廓排列的独立要素视为连续的要素。



格式塔理论

- 闭合原则：视觉中的物体可能是不完整的，然而只要物体的形状足以表征物体本身，人们就能感知出整个物体并忽视未闭合的特征。



中国地质大学

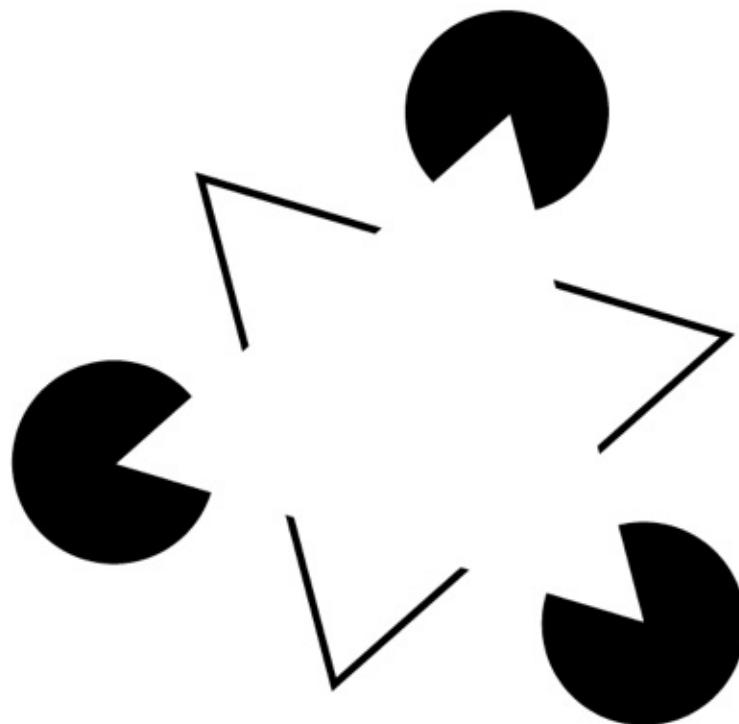
格式塔理论

0123456789



格式塔理论

- Kanizsa Triangle 凯尼萨三角

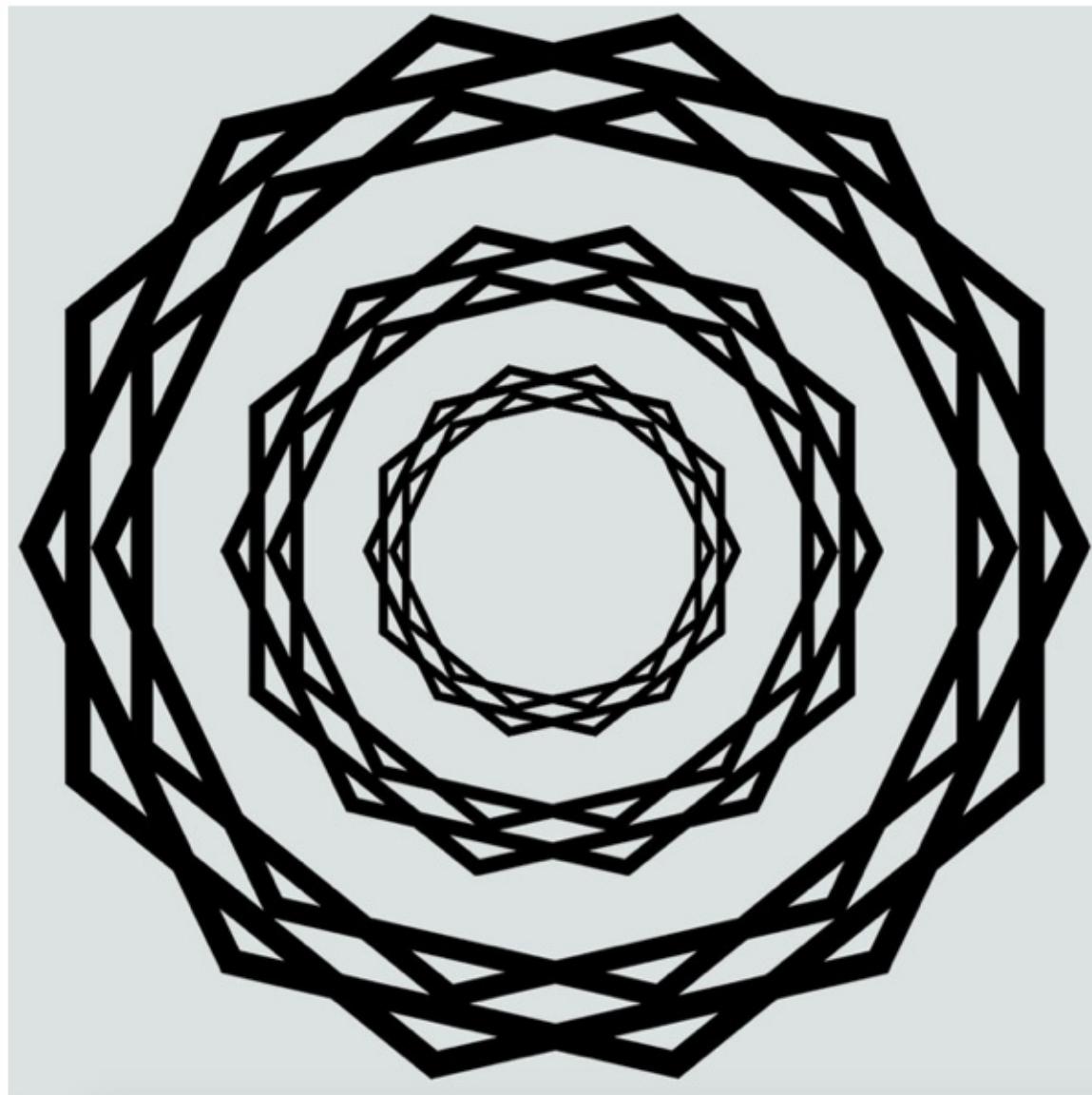


格式塔理论



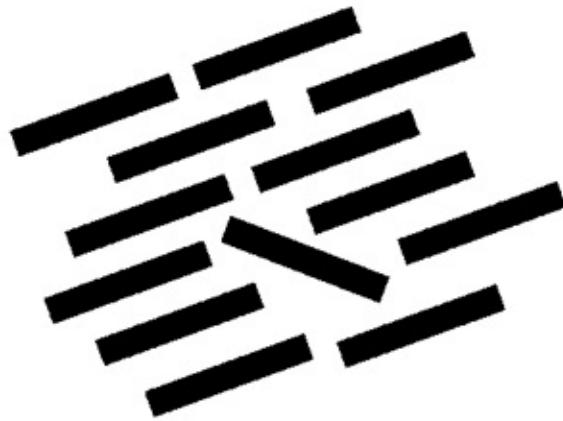
WWF

格式塔理论



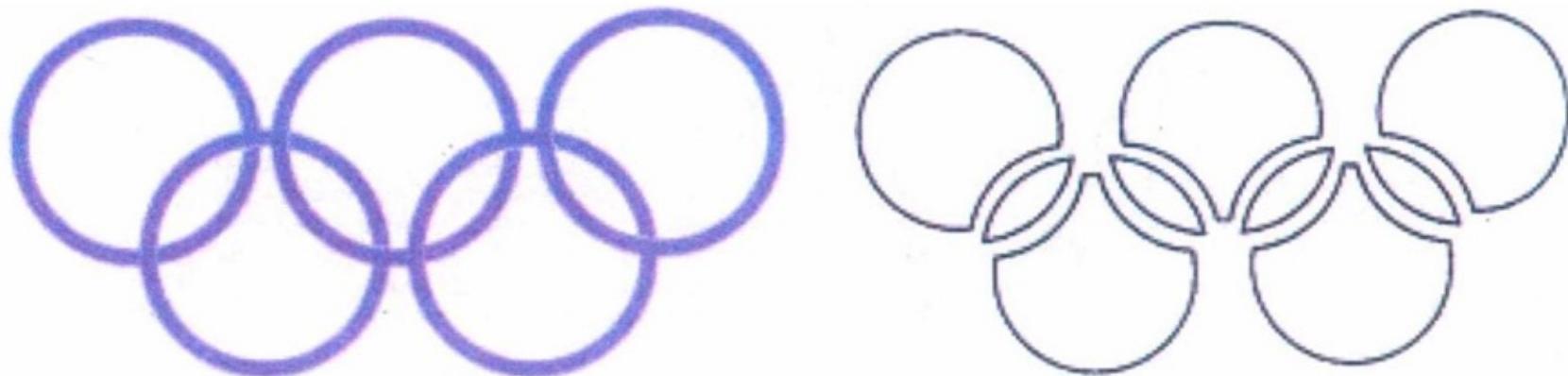
格式塔理论

- 共势原则：格式塔理论里唯一涉及运动、动画的原则。
- 人们倾向于认为有共同运动方式（移动趋势）的物体是相似或相关的



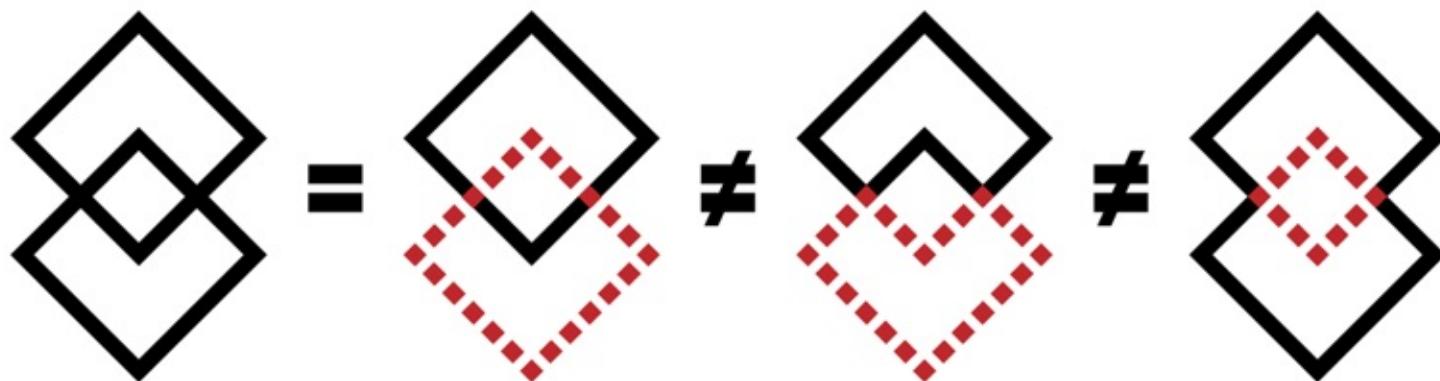
格式塔理论

- 好图原则：人眼通常会自动将一组物体按照简单、规则、有序的元素排列方式进行识别。



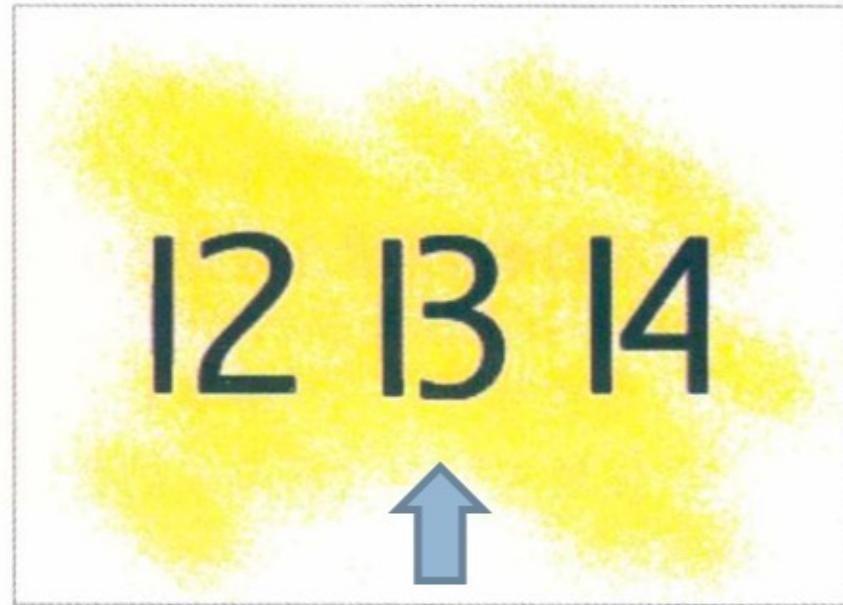
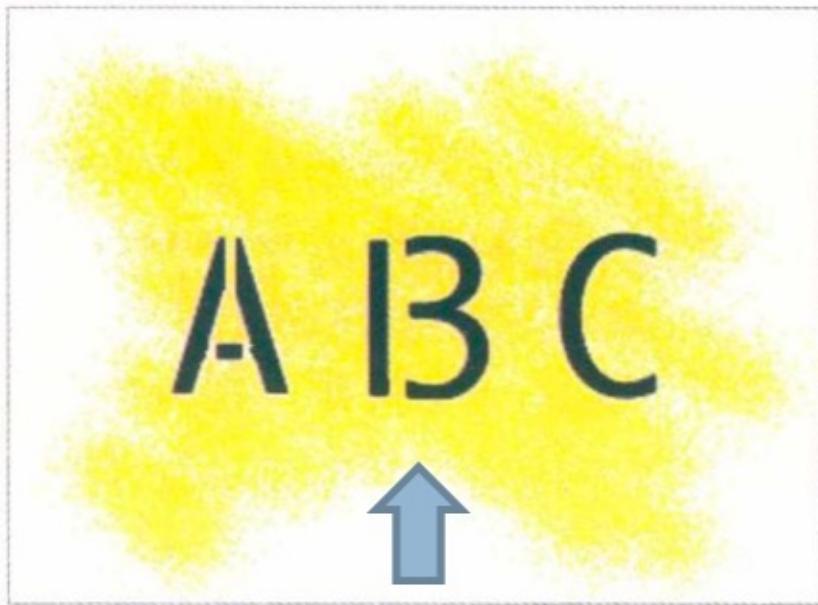
格式塔理论

- 对称原则：人的意识倾向于将物体识别为沿某点（或某个轴线）对称的形状。因此，将图像按照对称原则分为偶数个对称的部分，则各个对称的部分会被识别为相连的形状，从而增强认知的效率和强度。
- 如果两个对称的形状彼此相似，它们更容易被认为是一个整体。



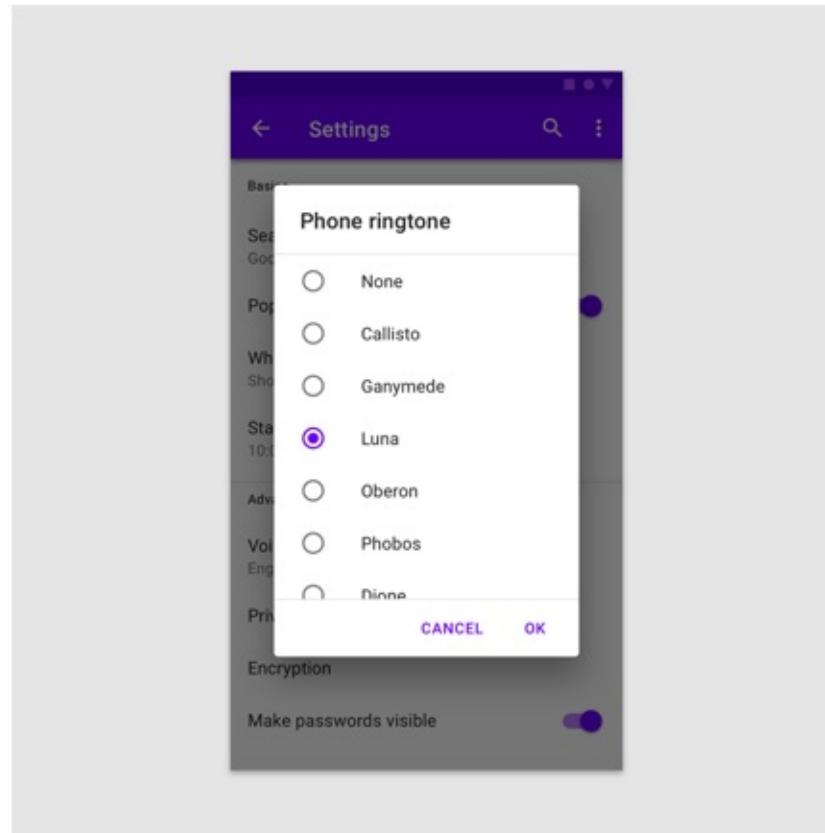
格式塔理论

- 经验原则：在某些情形下，视觉感知的结果与过去的经验有关，或者与图像中其他元素的暗示有关。

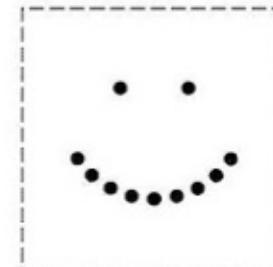
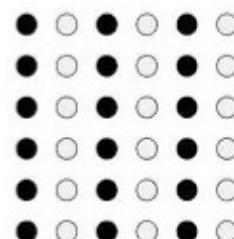
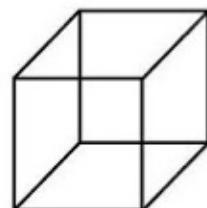
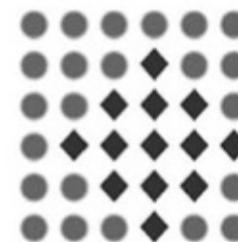
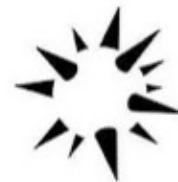
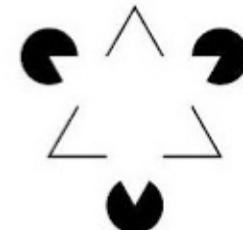
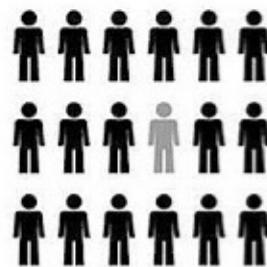
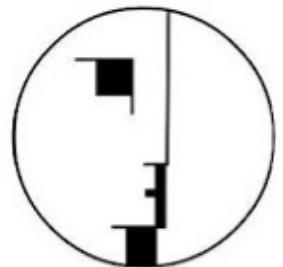


格式塔理论

- 突出原则（图像背景原则/主次原则）：人在观察物体时，总是倾向于把观察对象分为主要和次要的关系。图像（前景）比其背后的背景元素更重要。



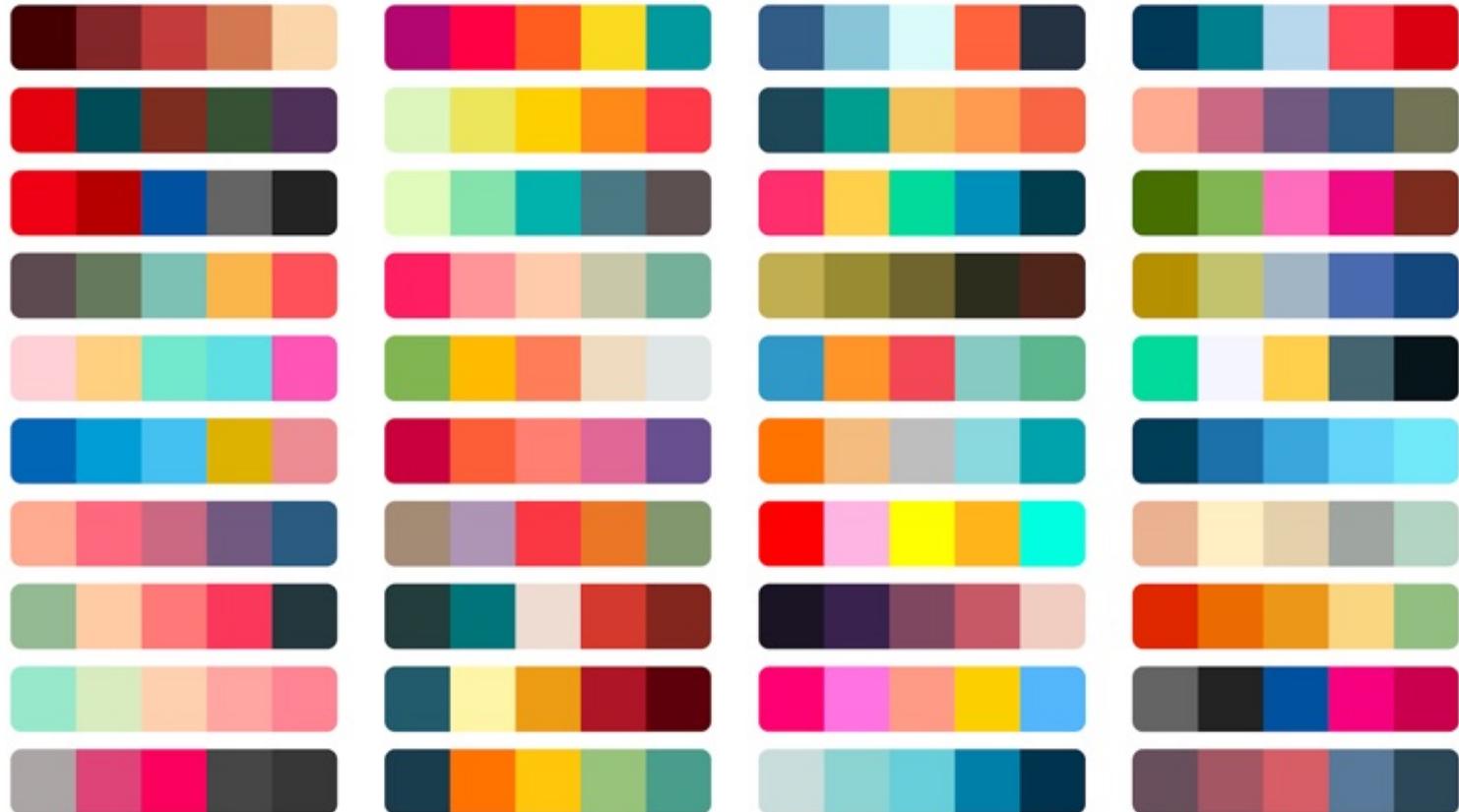
格式塔理论



颜色和色觉



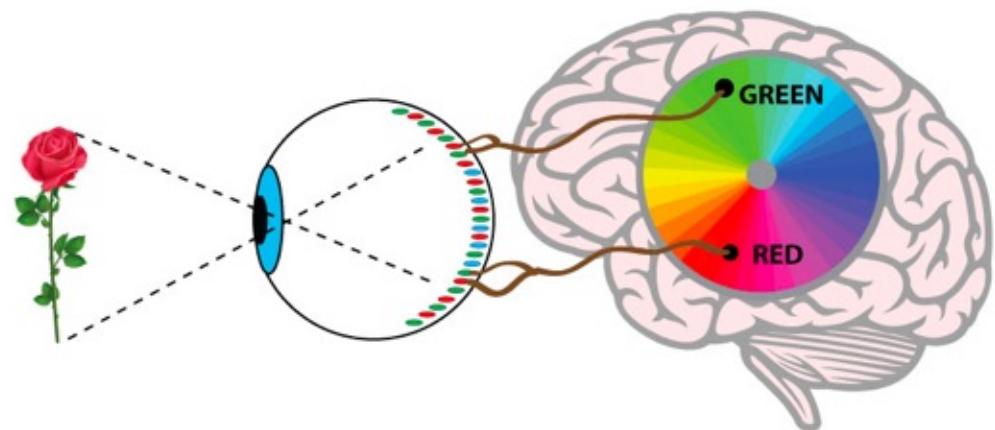
颜色和色觉



颜色、形状和布局构成了最基本的数据编码手段。

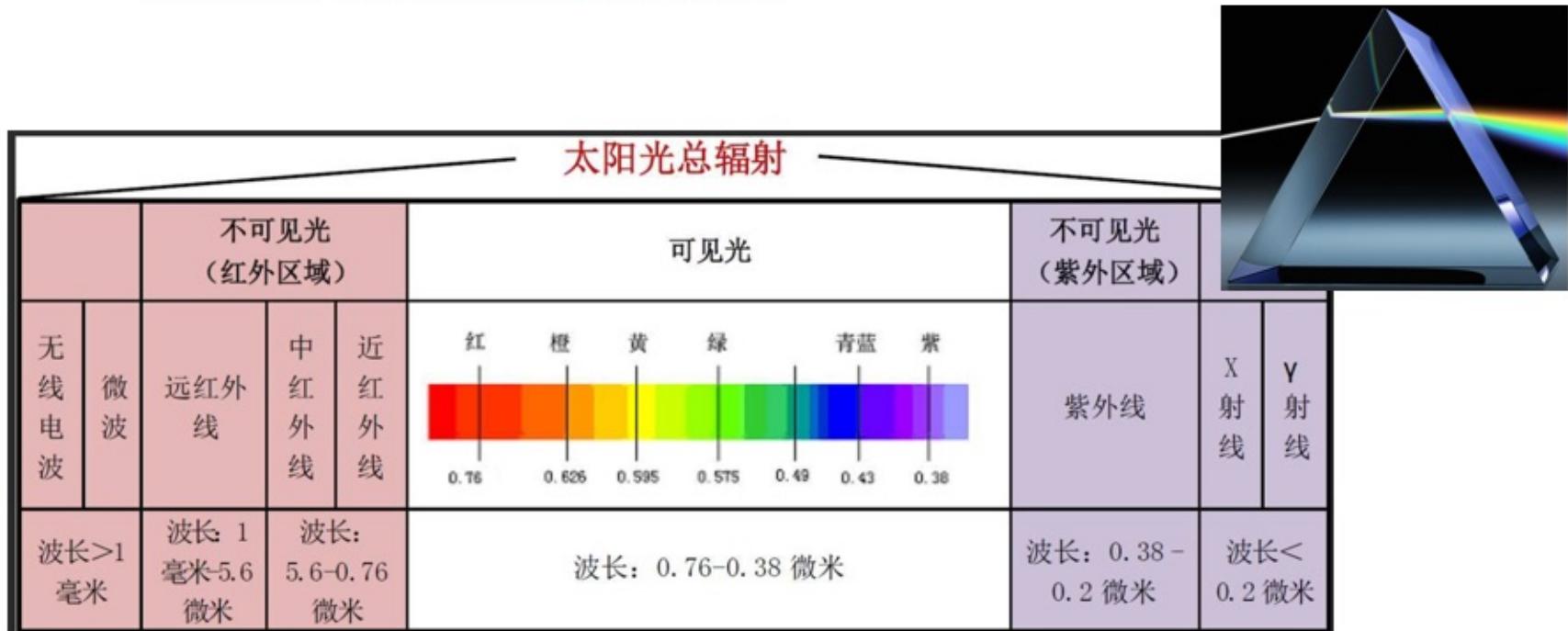
颜色和色觉

- 由可见光（电磁能）经过周围环境的相互作用后，到达人眼，并经过一系列生理学和神经科学的作用转化为人脑所能处理的电脉冲信号，最终形成对颜色的感知。



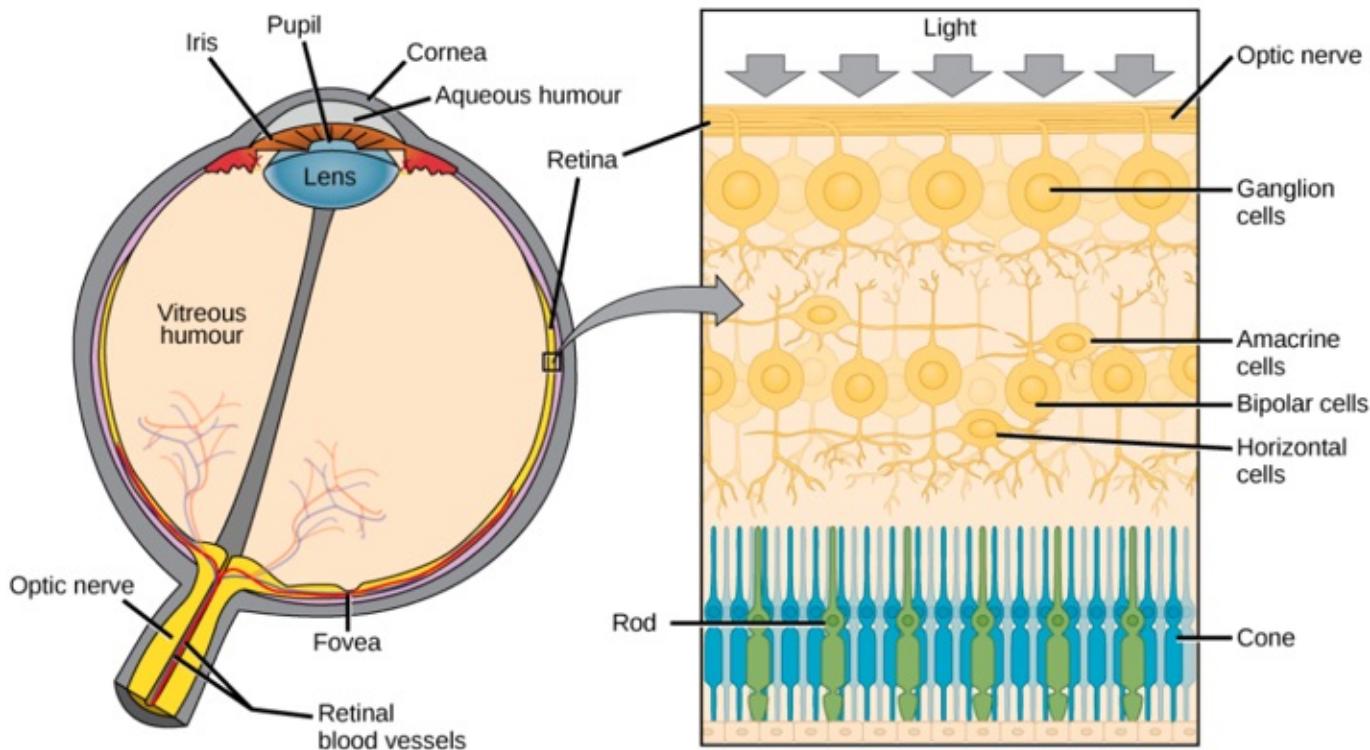
颜色和色觉

- 颜色刺激理论：人眼和可见光
- 可见光：能被人眼捕获，并且在脑中形成颜色感知的电磁波，它在整个电磁波波谱上只占了很小一部分。
- 复色光经过色散装置（比如棱镜）进行分光后，会按照光的波长大小顺序排列成彩色图案



太阳光色散实验（牛顿 1665）

颜色和色觉

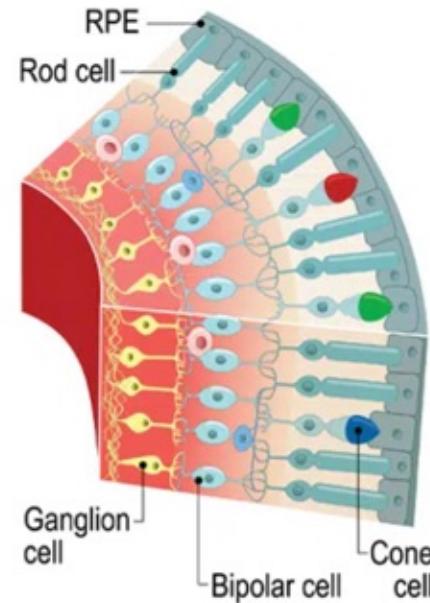


视网膜上的光感细胞

- 杆状细胞（视杆细胞/rod）：负责低照明情况下的视觉，能感觉物体的大小和形状
- 锥状细胞（视锥细胞/cone）：负责感知物体的颜色

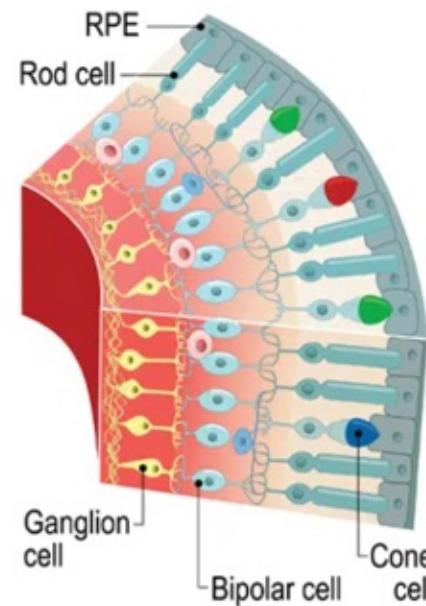
颜色和色觉

- 杆状细胞是视网膜上受光刺激最敏感的光感细胞，比锥状细胞对光的刺激要敏感10倍以上，通常具有很强的暗视觉（Scotopic Vision），能在低照明的情况下感知到物体的大小和形状。
- 杆状细胞几乎无法感知颜色信息，因此它的视觉一般是灰度视觉，所能感应的可见光范围为400~700nm



颜色和色觉

- 锥状细胞只对明亮的光线产生刺激反应，主要负责明视觉（Photopic vision）。
- 锥状细胞虽然没有杆状细胞多，但每个锥状细胞单独和一条视觉神经相连，因此具有更加清晰的视觉能力。
- 锥状细胞分为三种：感知不同波长的光（红、绿、蓝）
 - 长波长可见光：L细胞
 - 中波长可见光：M细胞
 - 短波长可见光：S细胞



颜色和色觉

- 物体所呈现的颜色由物体的材料属性、光源中各种波长分布、人的心理认知有关
- 色觉存在个体差异，所以颜色既是一种生理和物理现象，也是一种心理现象。

