



华北电力大学
NORTH CHINA ELECTRIC POWER UNIVERSITY

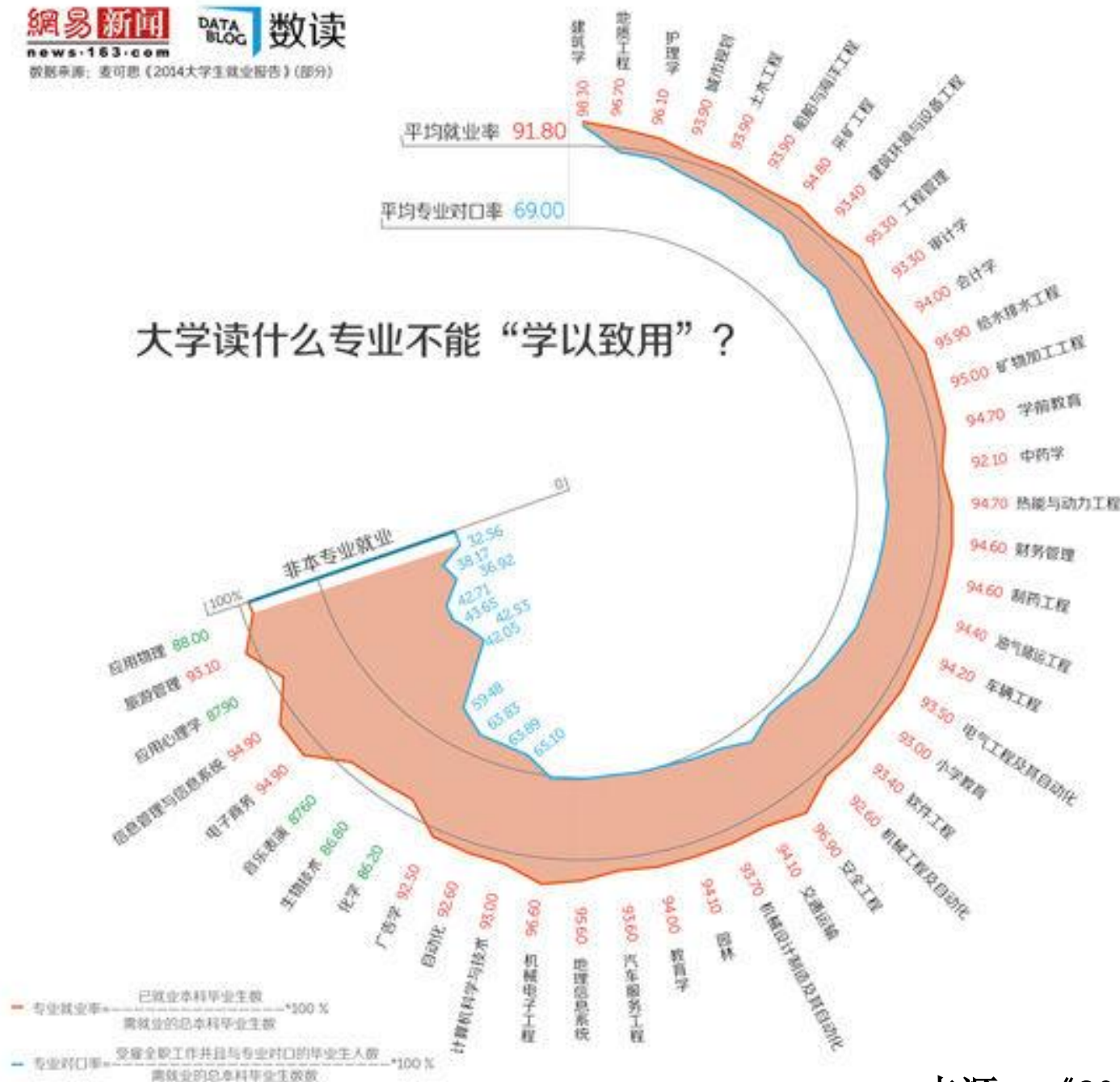
第4章 数据可视化与解释

身处看脸时代 颜值更要爆表

作品赏析

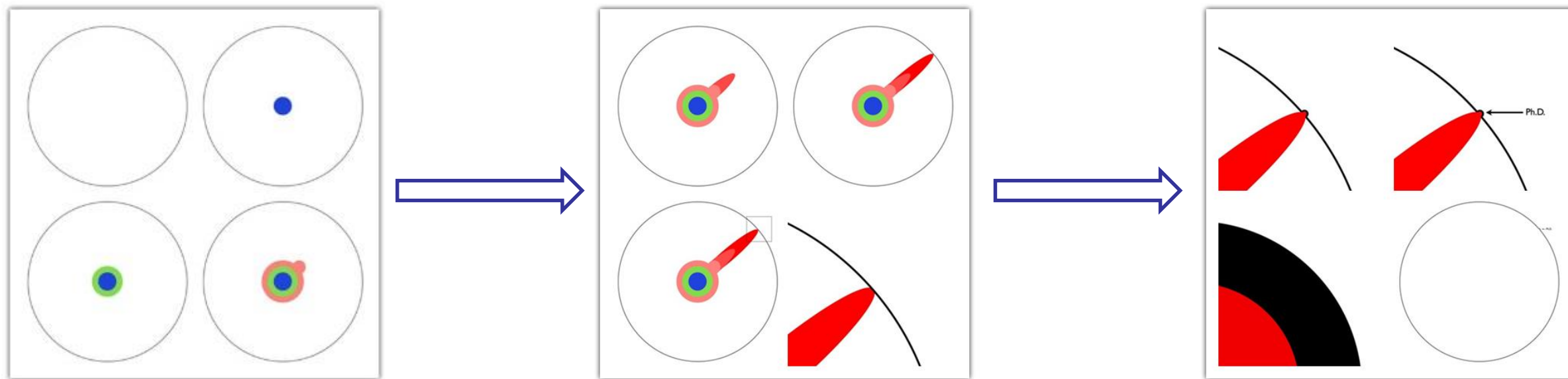
专业该如何选

大学读什么专业不能“学以致用”？



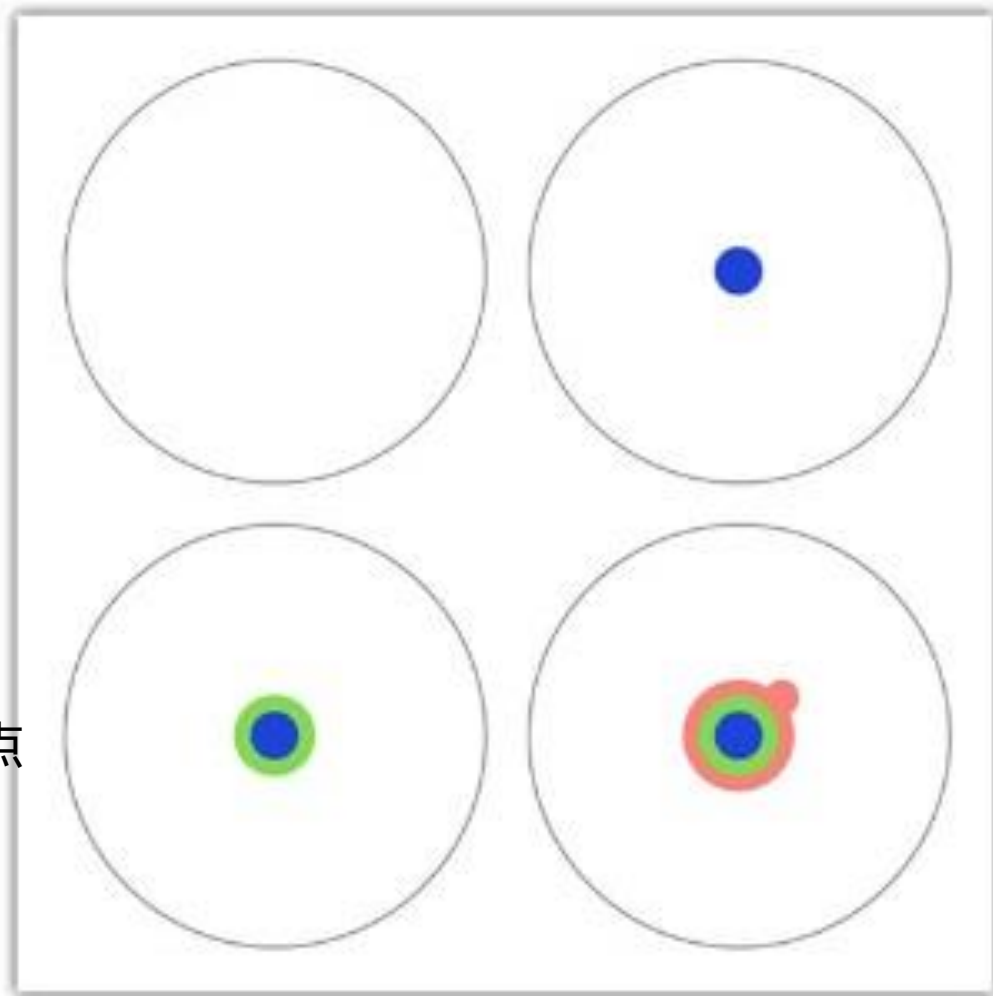
来源: 《2014年大学生就业报告》

你的求学之路



你的求学之路

0.用圈代表人类所有的知识



1.读完小学，你有了一些基础子知识

2.读完中学，你的知识又多了一点

3.读完本科，你有了专业方向

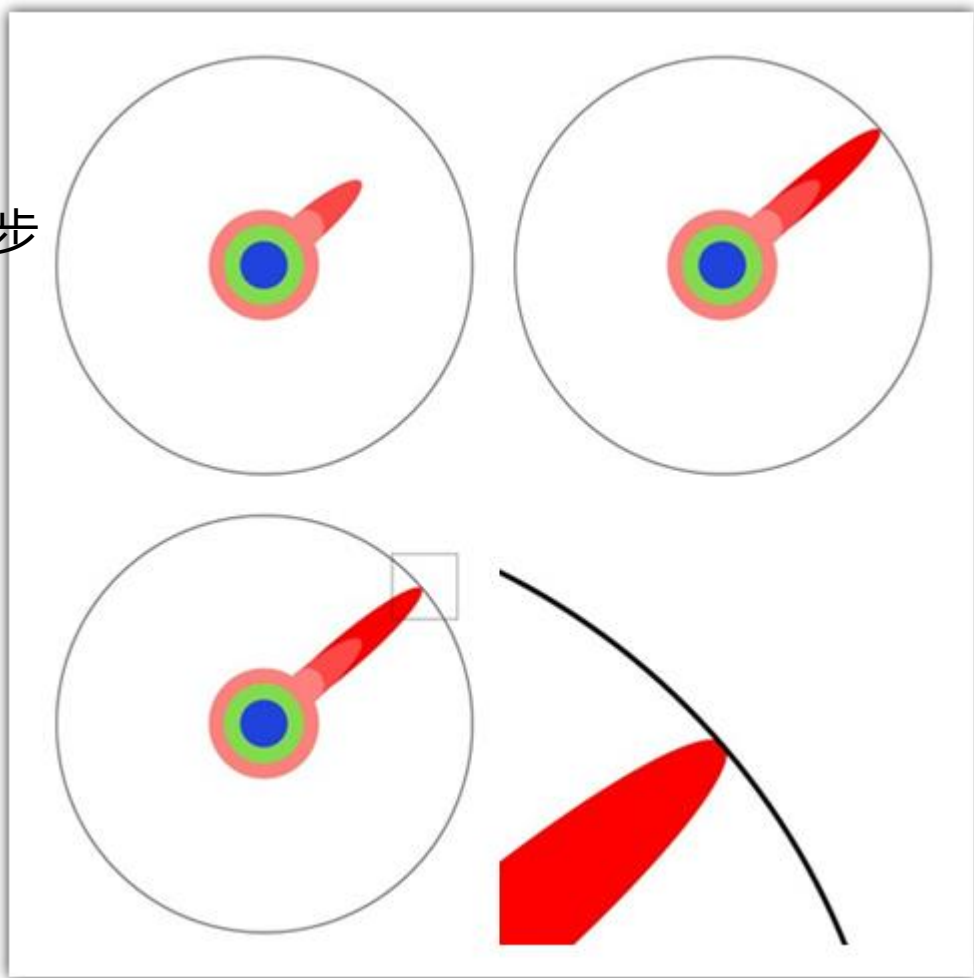
你的求学之路

4.读完硕士，你在专业上又前进了一步

5.阅读大量文献，接触本专业前沿

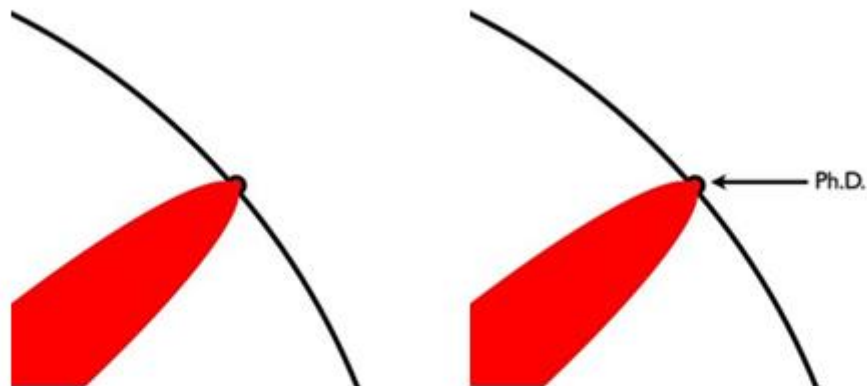
6.选择某一专题，作为主攻方向

7.在主攻专题上潜心研究好几年



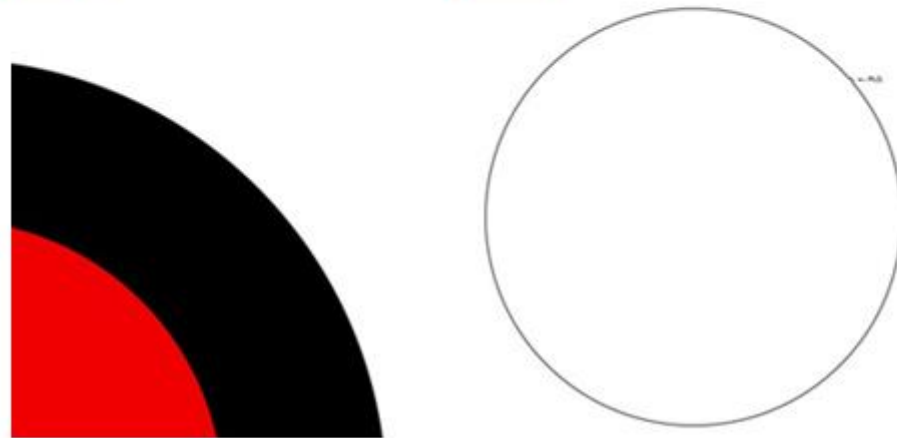
你的求学之路

8.终于取得突破性成绩



9.你把人类知识推进了一步，成博士

10.现在你看待世界的方式已不同



11.但是，不要忘了学无止境



什么是可视化?

说说你心中的可视化

什么是可视化?

计算机学科：利用人眼的感知能力对数据进行交互的可视表达以增强认知的技术；
用图表来tell a story!

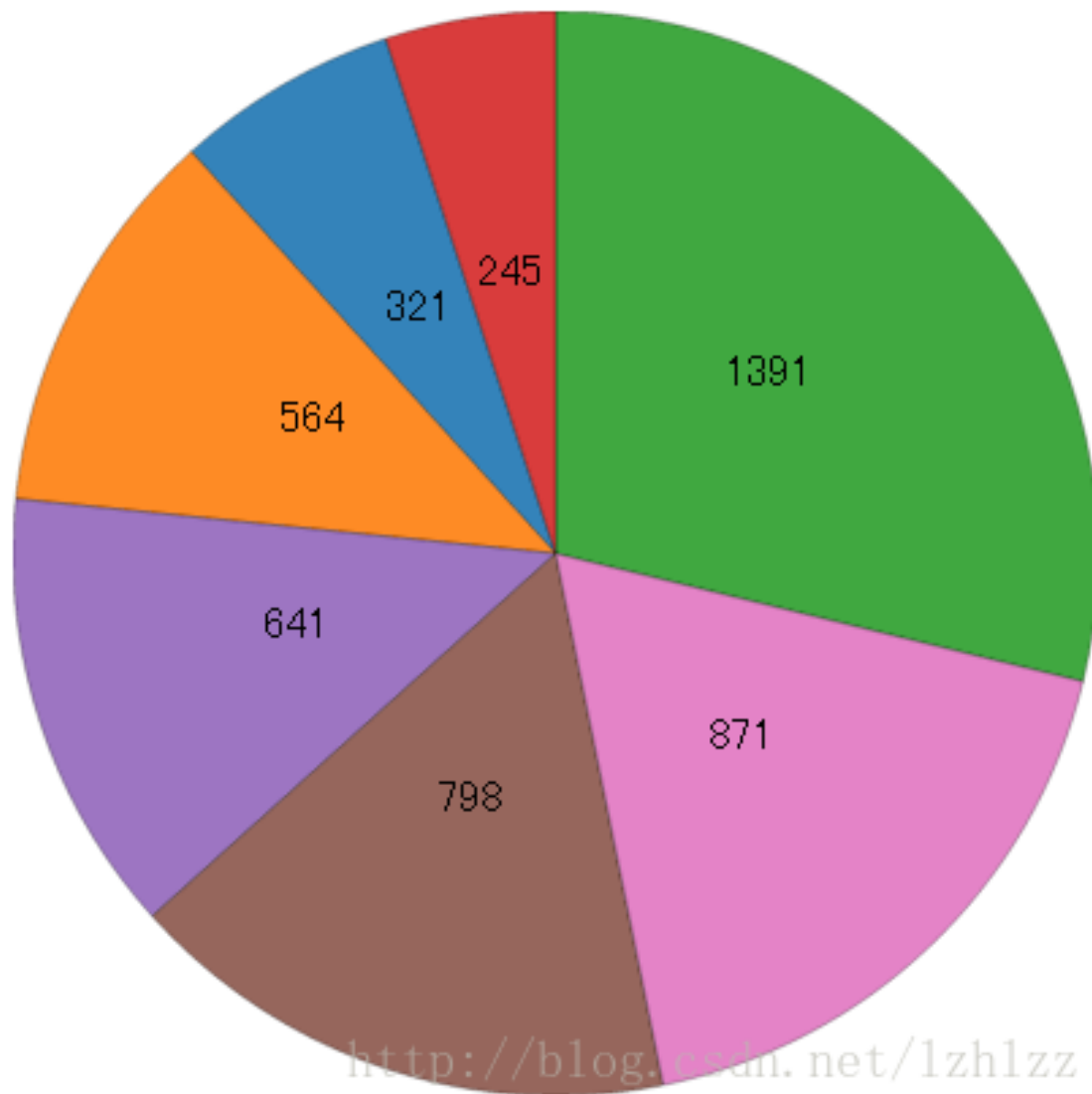
使用图形传达信息

Every picture tells a story. —Rod Stewart, 1971

什么是可视化?

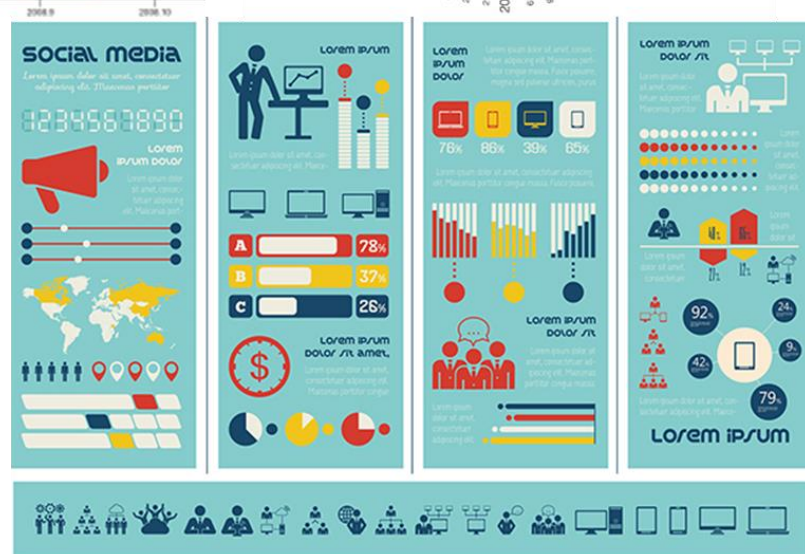
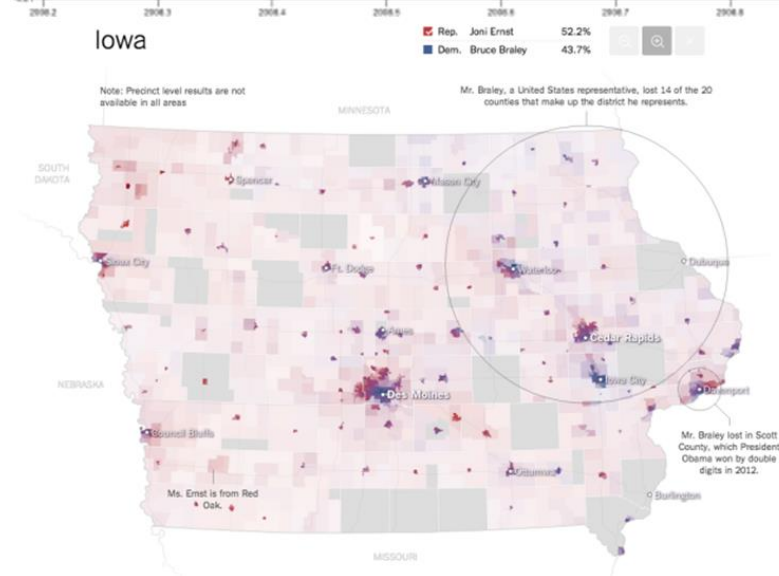
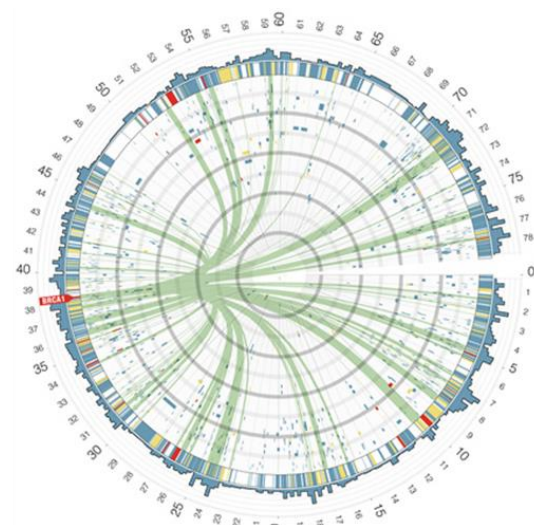
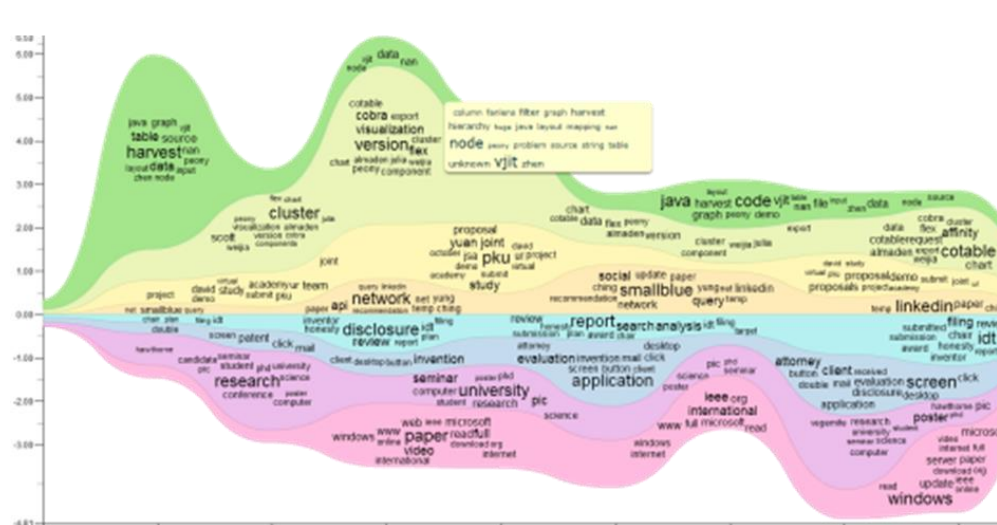
你能一眼看出哪一个最大吗?

【 321, 564, 798, 245, 641, 1391, 871 】



信息可视化：抽象、非结构化化的数据集合的可视化

时空数据、
层次和网络数据、
文本和跨媒体数据
多变量高维数据



这是一个**注意力稀缺**的时代

在互联网时代，可以称得上是信息爆炸。现在的社会不缺信息，缺的是注意力。

人的注意是有限的。

注意力 > 信息
注意力很便宜

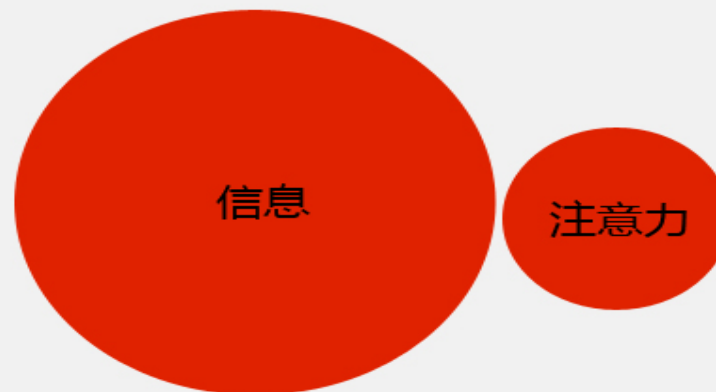
人的注意是有选择的。



人的注意是有竞争的。

工业时期

注意力 < 信息
注意力稀缺

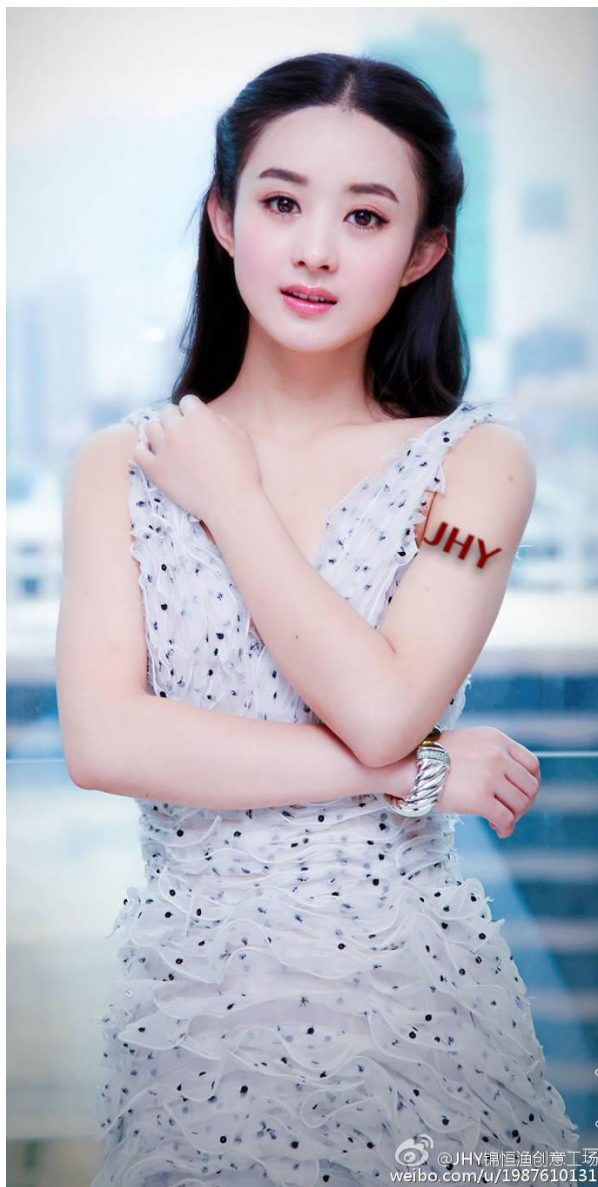
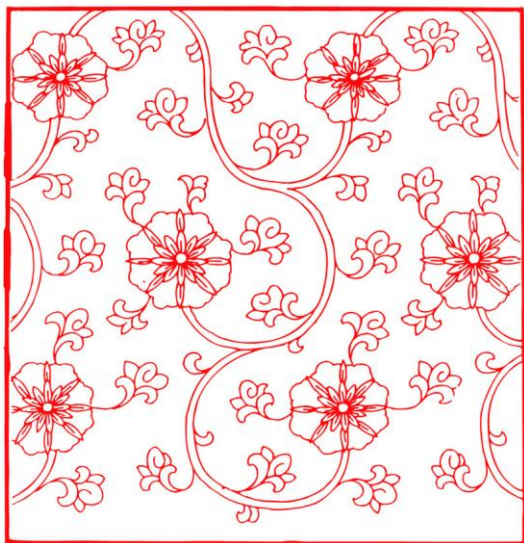


互联网时期



测测你的注意力

你最先看到的是什么？



大家来找茬



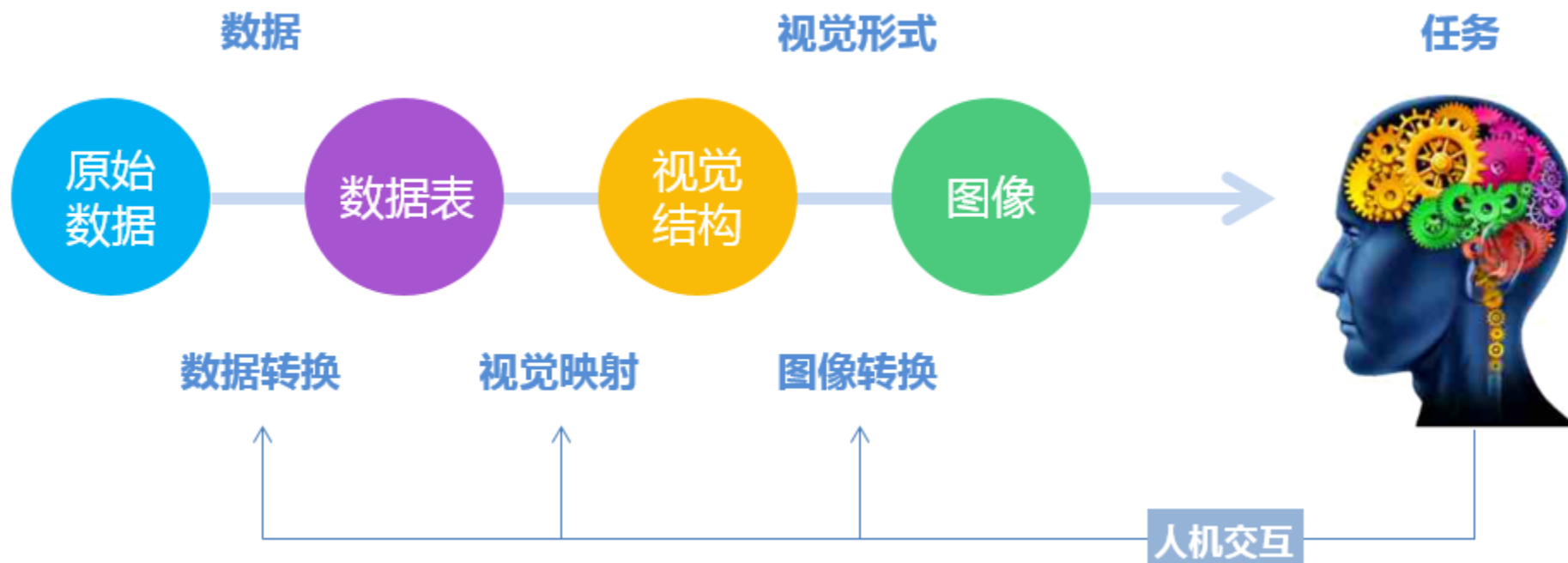
这是一个拼颜值时代





可视化的重要之处

可视化提升了易读性，降低了遗忘性。（脑力革命）



例：中国超级计算机数量全球第二，多闲置少利用

近日有报道称，位于长沙的“天河一号”超级计算机闲置近一年，已成烂尾工程。“天河一号”自2013年9月主机设备安装到位后便一直封存，今年6月才进行加电测试，至今仍未投入运营。这台曾经投资了8.6亿美元的超级计算机，如今竟面临机房天花板因潮气发霉、空调冷凝管漏水等问题，引发了不少人的叹息。

中国的超算研发始于1983年，第一台超级计算机“银河一号”的诞生使中国成为继美国、日本之后第三个能独立设计和研制超级计算机的国家。超级计算机是计算机中功能最强、运算速度最快、存储容量最大的一类计算机，多用于国家高科技领域和尖端技术研究，被认为是一个国家科研实力的体现。

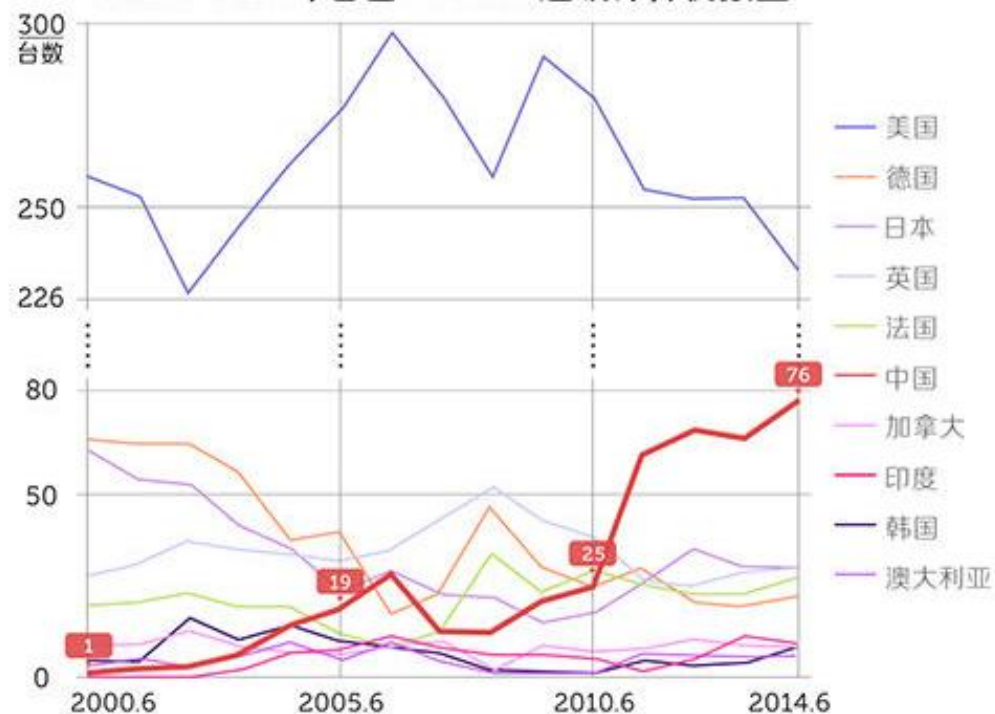
在这近二十年间，中国超级计算机的研发突飞猛进。时至今日，中国的超级计算机数量已居全球第二位。据国际超级计算机大会（ISC）2014年6月公布的最新数据显示：目前全球速度最快的前500个超算设备中，中国有76套占总数15.2%，美国则以232套雄居榜首。同时，不仅在数量上陆续超过了日本、英国、德国等先发国家，更在计算速度上多次赶超并夺魁。6月23日，超级计算机TOP500最新排行榜公布，落户于国家超级计算广州中心的“天河二号”继续领跑，以每秒33.86千万亿次的浮点运算速度第四次问鼎全球最快。

但是，目前国内超算应用效率低是个普遍的问题。中国科学院计算技术研究所2011年的一项关于“中国高性能计算研究与应用”的调查报告显示，国内用于科学计算研究的超级计算机不到20%，用于金融业和制造业的比例也偏少。据2008年12月的世界TOP500统计，全球共58台超级计算机用于金融业，70台用于工业制造，而中国仅1台超过10万亿次计算机用于金融业，制造业几乎是空白。虽有中国11台超级计算机用于信息服务业，但其中6台在游戏公司，而世界上主要用于信息处理与服务和企业信息管理。

另一方面，超算的市场效益问题更为紧迫。超级计算机系统的使用寿命大概是5年左右。根据TOP500的统计，对任何一套系统来说，从第1名落到500名要经历6-8年的时间，而在落到跟笔记本电脑一般的水平，只要经过8-10年的时间。2004年超级计算机“曙光400A”的总投资近1亿元，而今天同样的10万亿次的计算能力只要约100万元。超算系统技术更新周期短，这意味着超级计算机面临一个尴尬的事实：如果5年内机器实用率不高，或者效能不能充分发挥，盈利就很成问题，甚至亏本。

例、中国超级计算机数量全球第二，多闲置少利用

• 2000~2014年各国TOP500超级计算机数量



• 2014年全球超级计算机TOP500中各国所有情况



可视化的重要之处

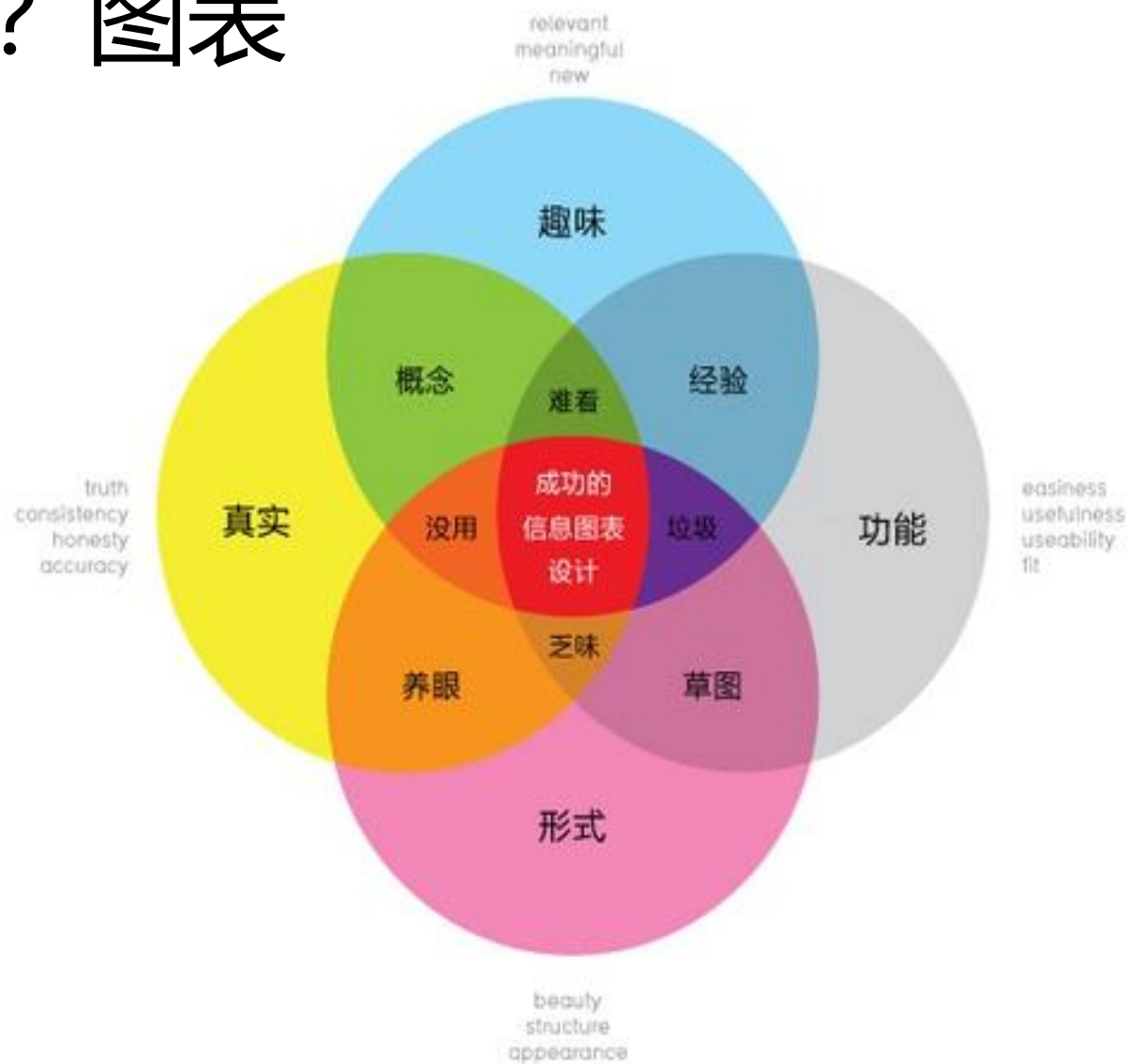
可视分析，快速发现趋势，预测未来，降低决策风险





如何做好可视化?

可视化=? 图表



形式：图形元素

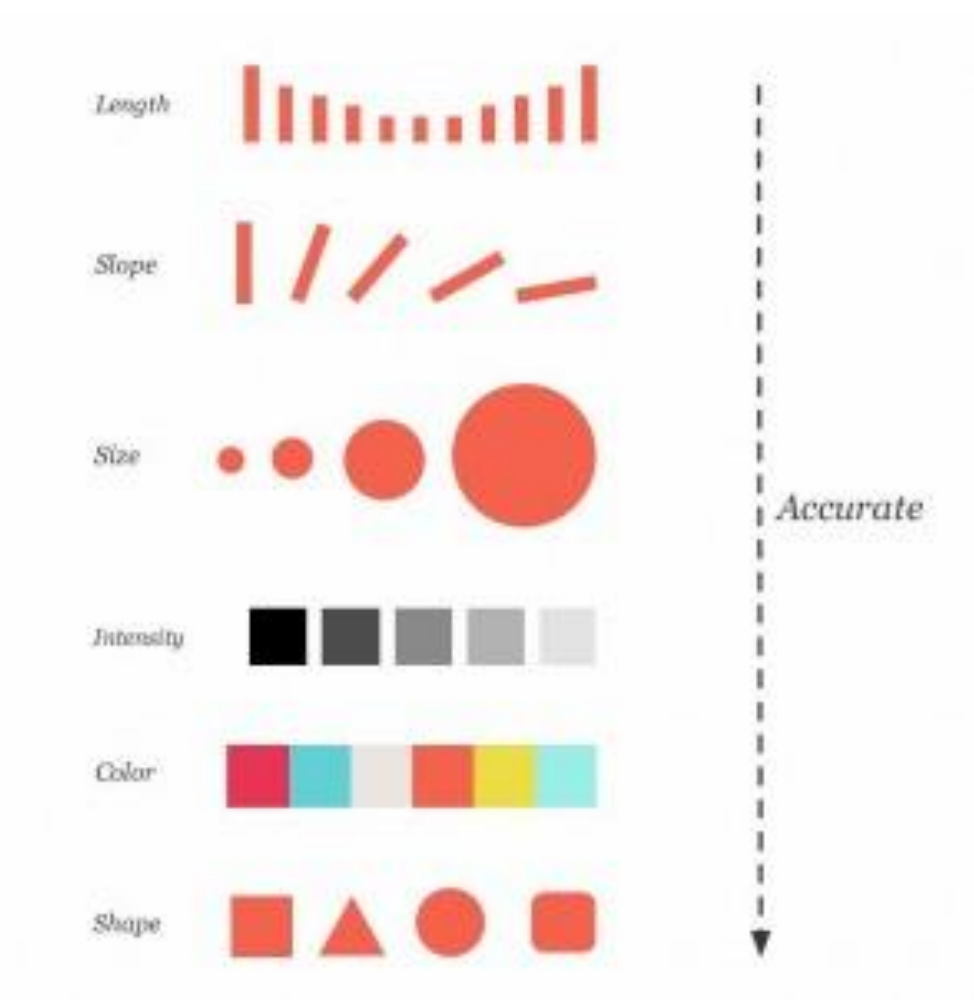
图形形状

图形大小

图形色彩

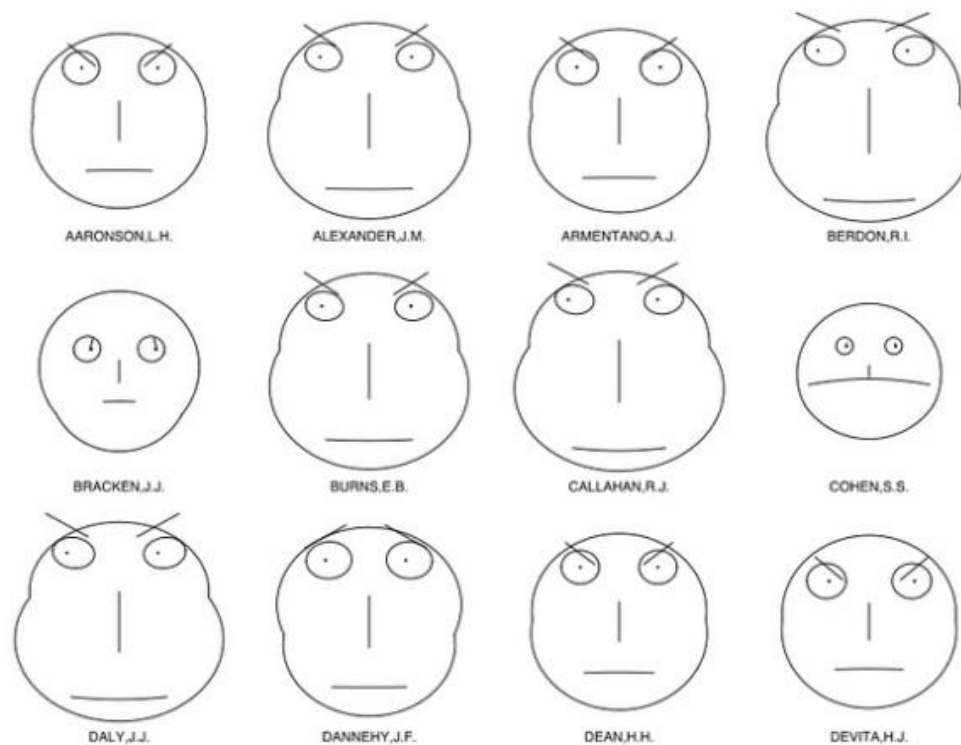
图形明暗

图形.....



趣味

好的可视化要有创造力，迅速抓住人的眼球，具有趣味性

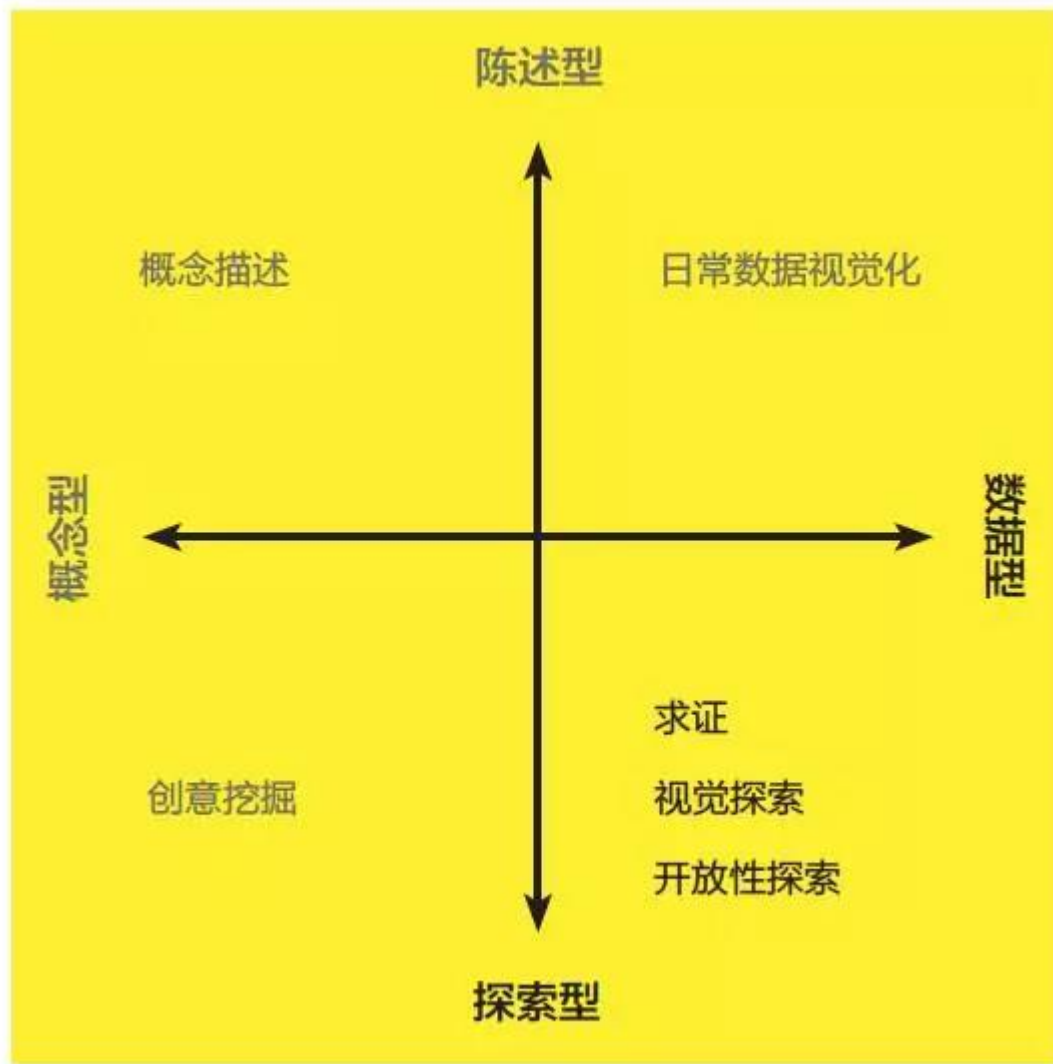


美国法官的评价

美国法官评价
排名（卡通脸
谱图）

功能

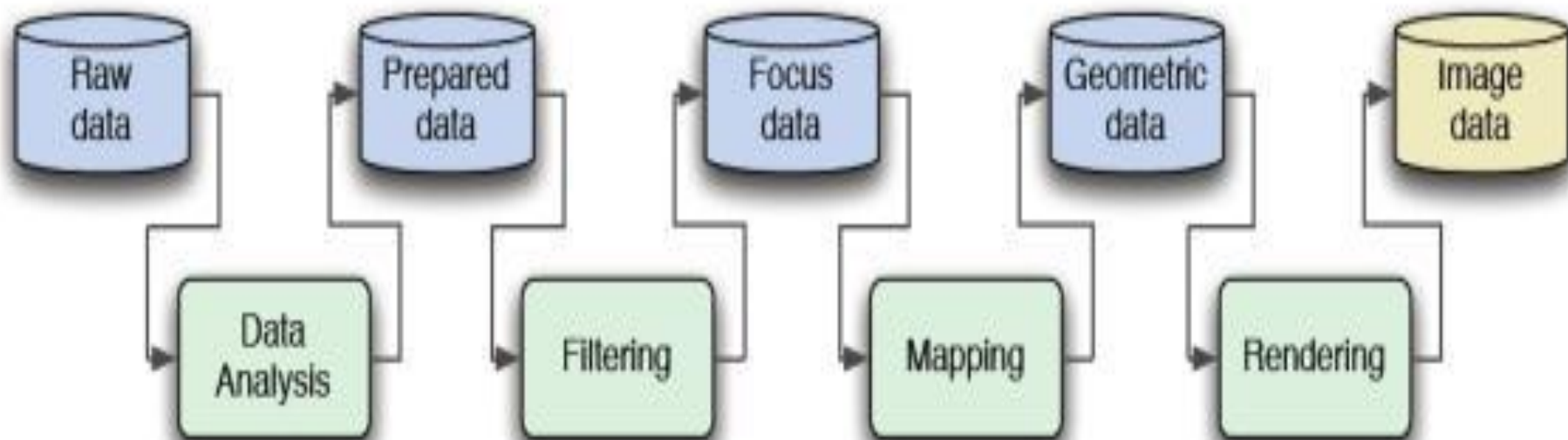
可视化必须针对具体的任务类型
(陈述型、概念型、数据型、探索型) 传达准确、真实的信息。





可视化过程模型有哪些？

流水线模型

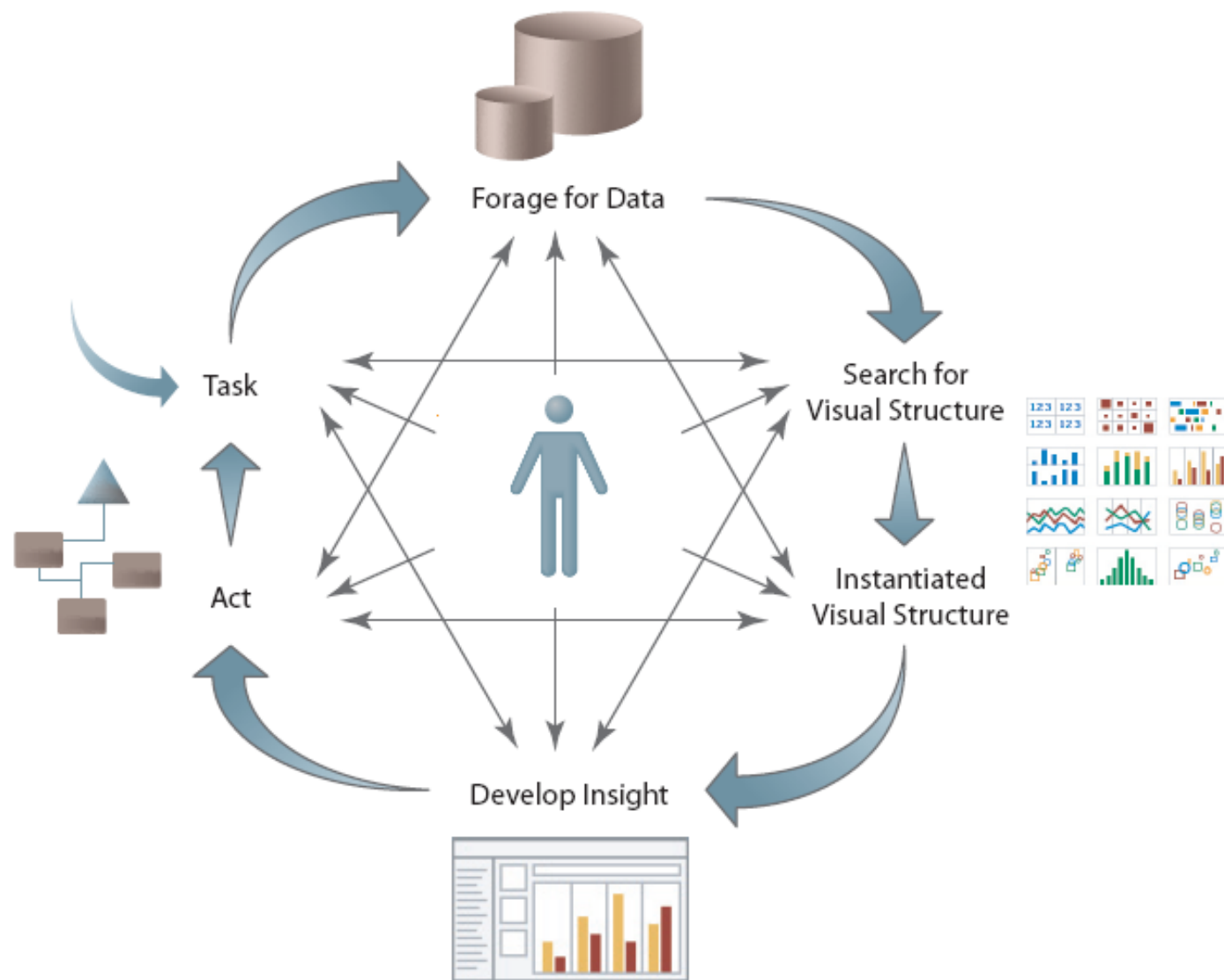


Haber, R. B. and McNabb, D. A. Visualization idioms

该模型由Haber和McNabb于1990年提出，它描述了从数据空间到可视空间的映射关系，体现了数据处理和图形绘制之间的逻辑过程。

循环模型

该模型描述了基于任务的可视化过程



回路模型

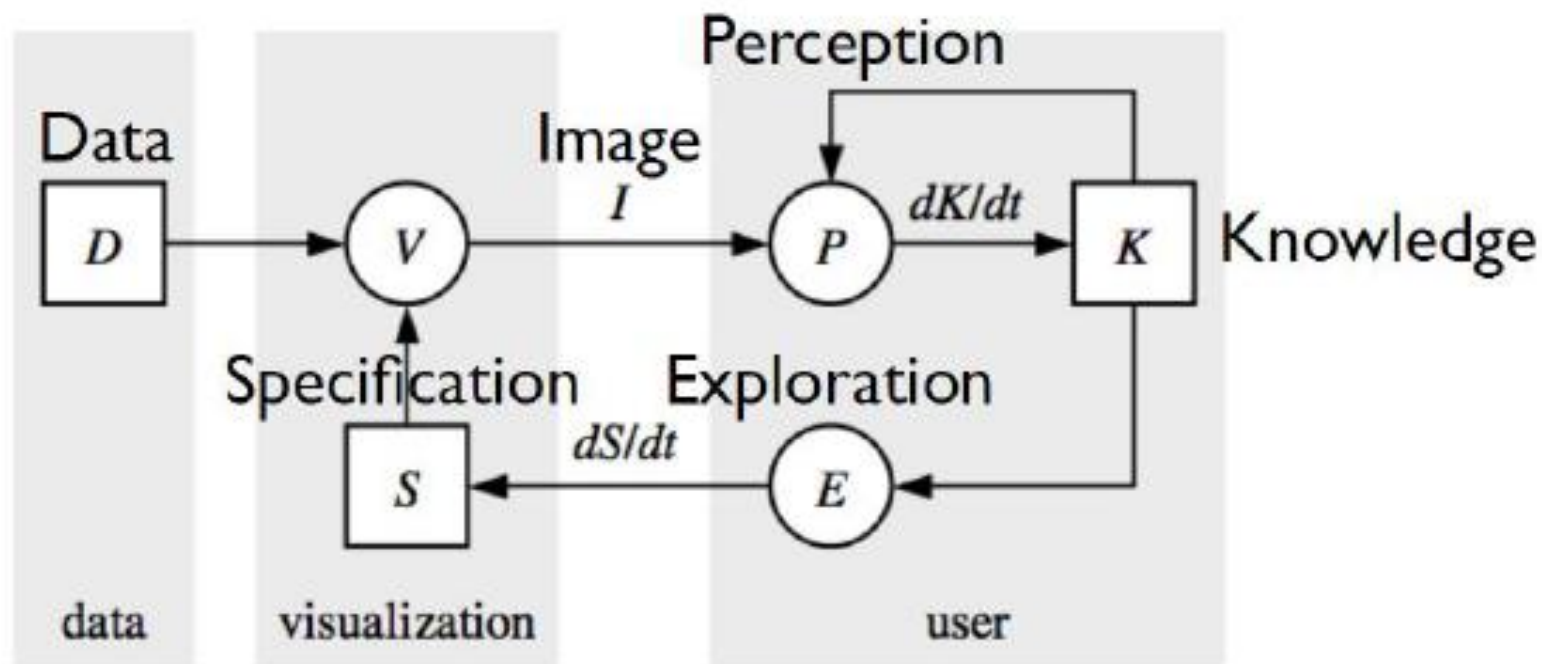
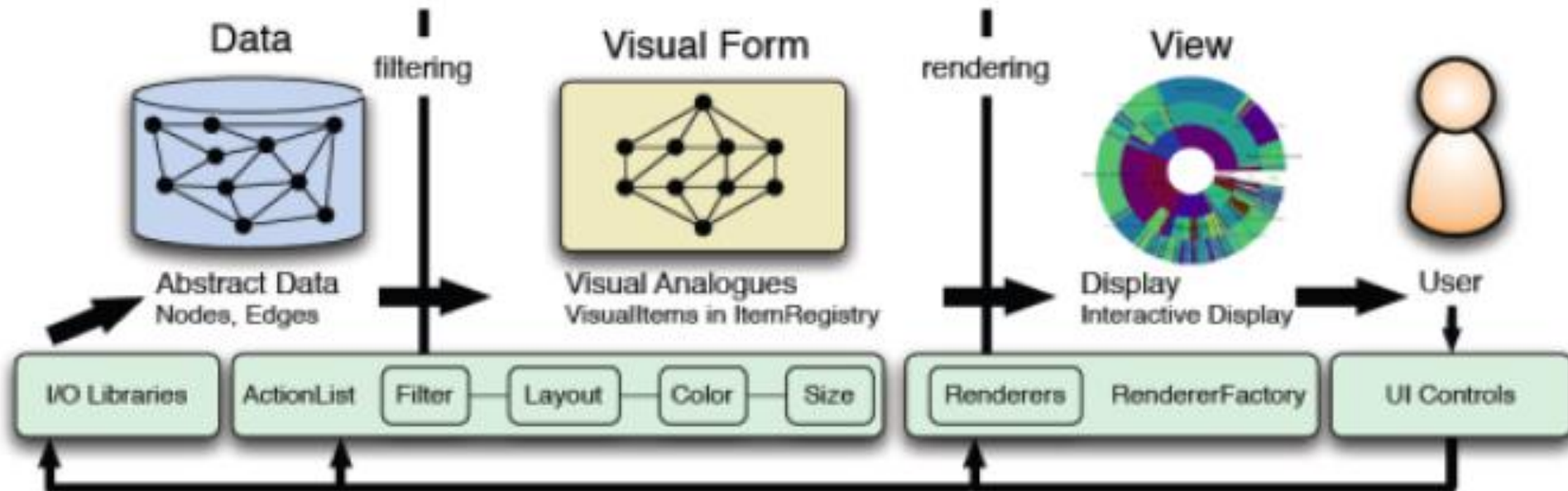


Figure 1: A simple model of visualization

Jark Van Wijk等人在前人的基础上提出的可视化回路模型，注重用户感知和交互探索在回路中的作用。强调了人在可视化过程中的作用。

回路模型



Card, Mackinlay & Shneiderman提出的信息可视化模型：将流水线改成回路，且此模型中用户交互可以出现在任何阶段。更加突出人的作用，后续几乎所有著名的可视化系统和工具均支持这个模型。

注意

可视化流程的三大要素：数据表示与变换、数据可视呈现、用户交互是可视过程中必须考虑的问题。

可视化以任务为出发点，人才是核心。人是最终决策者，是知识的加工者和使用者。可视化目标是增强人的能力而不是替代人的作用。



什么可视化工具好用？

可视化工具 开源软件

1

R、Python、Processing、Gephi、Circos...

2

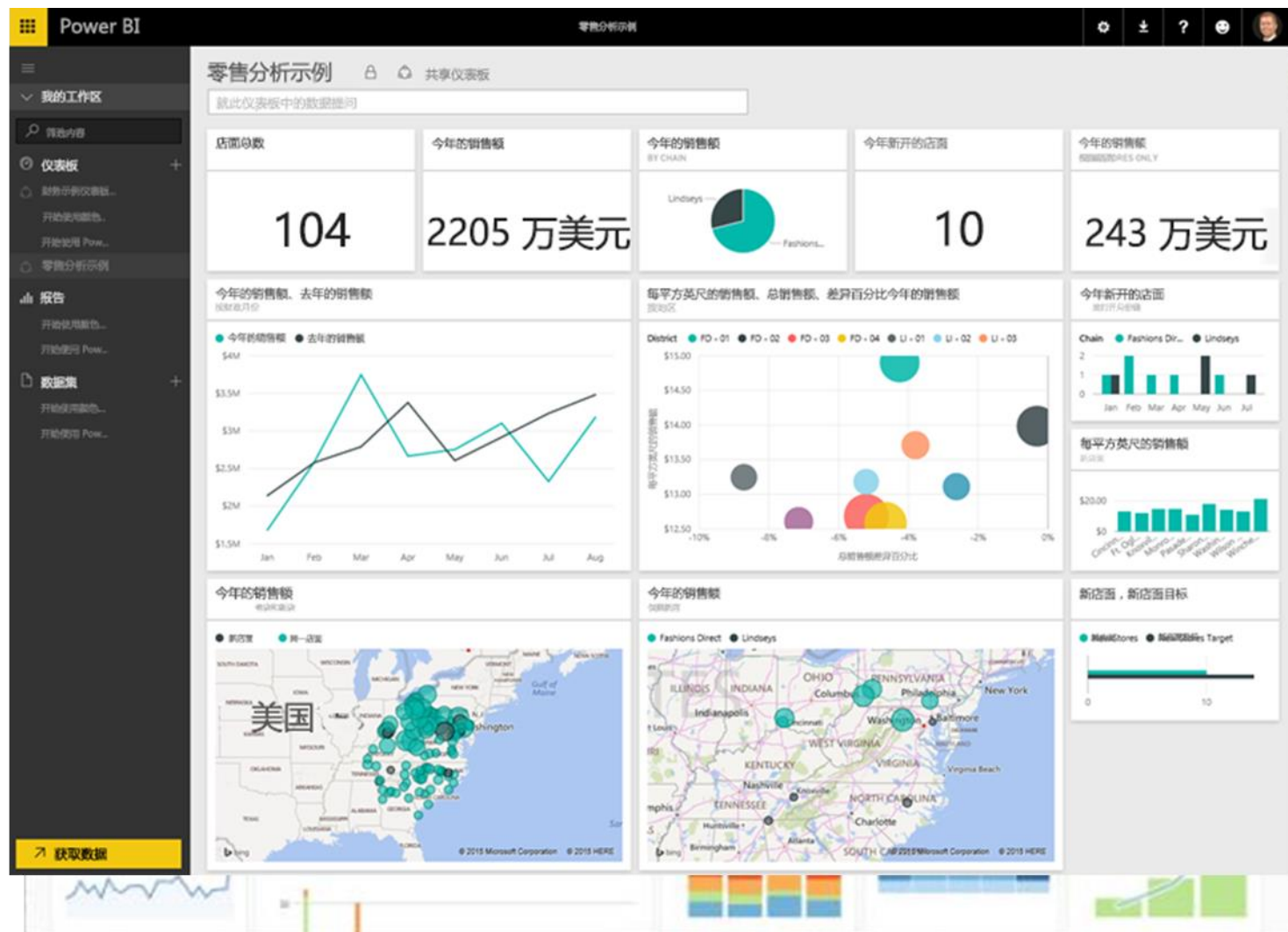
Processing.js、D3.js、Echarts、G2...

3

Plotly、Datavisual、Infogra.am、PiktoChart、Easel.ly、Tableau Public、Visme、Canva...

可视化工具 - 商业软件

- Tableau
- IBM Watson Analytics
- Power BI
- SAP Lumira
- DOMO
-



常见的图表类型



比较类



关系类



趋势类



分布类



路径类



地图类



区间类



时间类



聚合类



流程类

图表类型 - 比较类



柱状图



层叠柱状图



分组柱状图



饼图



环图



直方图



南丁格尔玫瑰图



面积图



层叠面积图



折线图



气泡图



散点图



雷达图



分级统计图



带气泡的地图



漏斗图



马赛克图



色块图



矩形树图



茎叶图



和弦图



等高线

图表类型 - 关系类



层叠柱状图



分组柱状图



面积图



层叠面积图



折线图



气泡图



散点图



雷达图



和弦图



韦恩图



你准备好了吗？



华北电力大学
NORTH CHINA ELECTRIC POWER UNIVERSITY

第4章 数据可视化与解释

身处数据时代，故事更要动听

第一部分 新闻传播中的数据可视化

第一节 数据可视化的新闻叙事

- ☞ 一、“前数据化”的叙事模式
- ☞ 二、数据可视化在叙事模式上的创新
- ☞ 三、数据可视化对传统叙事模式的解构
- ☞ 四、数据可视化的叙事模式

一、“前数据化”的叙事模式

- 叙事学是在结构主义基础上发展起来的对叙事文本进行研究的理论。法国叙事学家热拉尔·热奈特（Gerard Genette）在《叙事话语》一书中，对“叙事”一词所包含的三层含义进行了区分，对故事、叙事和叙述这三个不同的概念做了界定。他认为，故事指真实或虚构的事件，叙事指讲述这些事件的话语或文本，叙述则指产生话语或文本的叙述行为。
- 也有学者将新闻叙事学概括为运用话语分析的研究方法，对新闻的叙事行为和叙事策略进行的科学研究。新闻叙事主要研究新闻是如何通过叙述方式来表现新闻本质的，新闻的本质决定着新闻叙事的基本原理，**其核心是客观真实与目的性的高度统一。**
- 数据可视化是大数据时代新闻传播的新领域，它的出现与发展，推进着新闻叙事从“前数据化”模式向“数据化”模式的转移。

二、数据可视化在叙事模式上的创新

（一）突破了传统的新闻叙事模式

传统的新闻文本是以线性的文字信息为主的，“5W+1H”被奉为传统新闻叙事的经典模式，做报道就是讲故事，时间(when)、地点(when)、人物(who)、事件(what)、原因(why)、怎么样(how)是构成故事的基本元素。传统的叙事方式在结构上形成故事中心化，在叙事过程中体现出明显的因果等逻辑关系。

而数据可视化叙事就是将抽象数据具象化，挖掘出数据之间的相关性，在形成新闻文本的同时，制作出具有高度互文性、动态性、参与性的可视化信息，用动态、直观、互动的文本挖掘出数据背后的深层意义。这种叙事模式既增强了文本的交互性、内容的易读性，还加强了传受双方的互动性，为受众检验和学习数据新闻中所包含的信息提供了更加便捷的渠道。

二、数据可视化在叙事模式上的创新

(二) 弥补了传统新闻宏观叙事的不足

传统的新闻叙事更注重细节，抓典型、找个案，力求以小见大、见微知著。当这种新闻叙事缺乏足够的新闻事实，或者说缺乏典型案例时，媒体往往采访一些权威专家、政府官员、相关行业的专家或负责人，用他们的话语和观点来提高报道的说服力。但是，这种操作手法，一是可能存在“部分真实”和“断章取义”的风险，二是因私利、立场的不同，容易被人利用，使媒体的公共服务性大打折扣。

从统计学的角度看，传统的新闻叙事是基于个案或小样本的报道，数据可视化则是基于大样本的报道。个案报道以典型性、代表性给受众以启发，大样本则呈现出客观的一般性规律。可以说，可视化的叙事模式是通过数据对报道客体的一种宏观叙事。因此，数据新闻的可视化弥补了传统新闻在宏观叙事方面的不足，借助大数据技术，数据可视化表现出一定的宏观解释力。

二、数据可视化在叙事模式上的创新

（三）改变了以文字为中心的传统新闻叙事方式

在以文字为中心的新闻叙事中，文字是承载新闻信息的主体，**即使有图片和数据，也只是文字的“配角”，为文字叙事做辅助。**

数据新闻的生产流程可以划分为三个阶段，即获取数据、处理数据和呈现数据。第一阶段是海量的数据搜集与准备；第二阶段是深入分析、挖掘数据的新闻价值和新闻故事，并确立新闻选题；第三阶段是将数据、故事及深层含义以视觉语言的方式呈现给受众。可视化处于最后一个环节，可以说是报道的“点睛之笔”。



三、数据可视化对传统叙事模式的解构

（一）“讲述者”部分物化为“数据”，“写作者”出现去职业化

大数据作为一种信息资源，不仅蕴含着经济价值，同样蕴含着大量的新闻线索、报道选题，“数据”开始取代“讲述者”，“数据”也在“说话”，当事人的讲述被表现为各种传播符号的数据所替代，采集数据和挖掘数据正在逐渐取代记者的现场采访。

随着政府部门加强数据库建设，加大数据公开力度，数据获取渠道的多元化以及数据分析开源软件的普及推广，原来只有专职“写作者”从事的新闻生产活动，正逐渐有业余的、非专业的传播者参与其中，他们根据自己的兴趣，学习并掌握了数据处理技术，在公共媒体或社交媒体平台进行可视化产品的生产与发布，因而新闻叙事者出现“去职业化”的现象。

三、数据可视化对传统叙事模式的解构

（二）从二维叙事结构到立体叙事结构的嬗变

一般性的单一事件报道多以时间为线索，依次展现新闻要素。重大的多个事件的报道往往采取时空组合或时间延续性的报道方式。

与传统新闻的二维叙事结构不同的是，数据可视化更多地采用立体的叙事结构。这种立体结构体现在两个方面。一是时间线、空间线加上超链接，共同构建了具备互联网传播优势的信息结构。二是时间线的延续。数据可视化不仅表现为对已经发生或正在发生的事态的准确叙述，更表现为通过数据挖掘，掌握事物发展的规律和趋势，对未发生事实的预测性报道。

三、数据可视化对传统叙事模式的解构

(三) 叙事话语的再结构化和可视化

● 再结构化：

- 对同类新闻中不同时期的相关新闻数据进行归类统计、整合比较，更为深入、立体、多元化地揭示新闻；
- 对网络搜索引擎、社交媒体内容、用户数据进行深度挖掘，揭示个别、分散行为中蕴含的共同规律；
- 从政府机构、企业等发布的公开数据中寻找可作为新闻背景的有用信息；
- 通过网络观察、调查或众包的形式收集数据。

● 可视化：

- 一是将数据信息的量与关系等转变为直观的图形；
- 二是看图说话，将文字信息变为形象符号；
- 三是以图整合，在图表中集成多元信息。

第三节 数据可视化的类型

☞ 一、数据地图

☞ 二、时间轴

☞ 三、信息图

☞ 四、气泡图

☞ 五、社会网络关系图

☞ 六、词频图

一、数据地图

数据地图是数据新闻中最具有代表性的一种可视化类型。数据地图就是把数据添置在地图的坐标中，宏观、清晰地揭示地理位置与数据之间的关联。

数据地图通常采用两种呈现方式。一是把数据以小圆点的方式标注出来，受众点击小圆点即可获得具体的信息。二是数据以热力图的方式附着于地图之上，此举是为了使多维度数据同时展现，这种呈现使数据的挖掘和对比更深入，受众不仅能了解整体情况，还可以发现数据背后的深层原因。



二、时间轴

时间轴又称时间线（timeline），就是从时间的维度，系统、完整地记录、展现事物的变化轨迹，是基于时间顺序呈现事物发展的可视化形式。

时间轴的功能主要体现在两方面：一是以时间为依据将相关内容整合起来；二是以时间为线索揭示整个事物的变化过程。当数据地图和时间轴结合在一起的时候，可以为受众提供全景式的信息，这种方式在战争、社会运动、事件性新闻的报道中使用最为普遍。

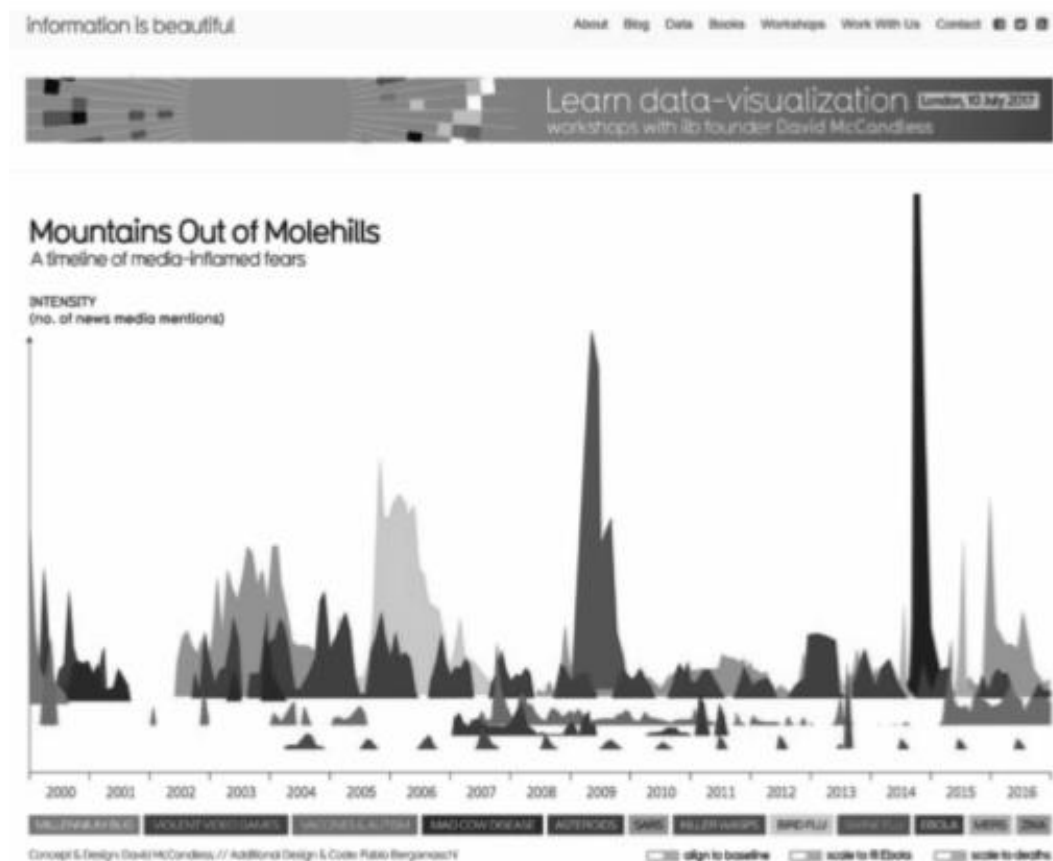
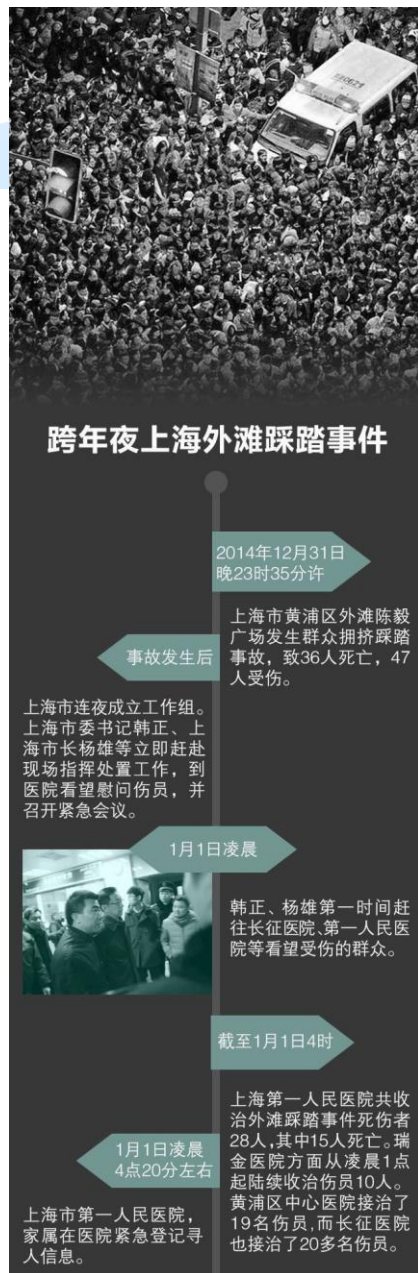


图 2-11 “信息是美的”网站刊发的《山峰的鼠丘：媒体夸大恐慌的时间线》
对时间轴的应用

时间轴



三、信息图

根据自身结构的复杂程度，信息图存在单一型与复合型的区别。**单一型**是指在呈现一组数据时，选择某一维度，使用一幅图表。**复合型**则是从不同维度，制作多幅图表予以呈现，这类图表有助于展示事物的全貌。具体选择何种类型，主要依据数据的复杂程度和数据背后的意义大小而定，有时还要参考制图编辑所追求的传播效果。

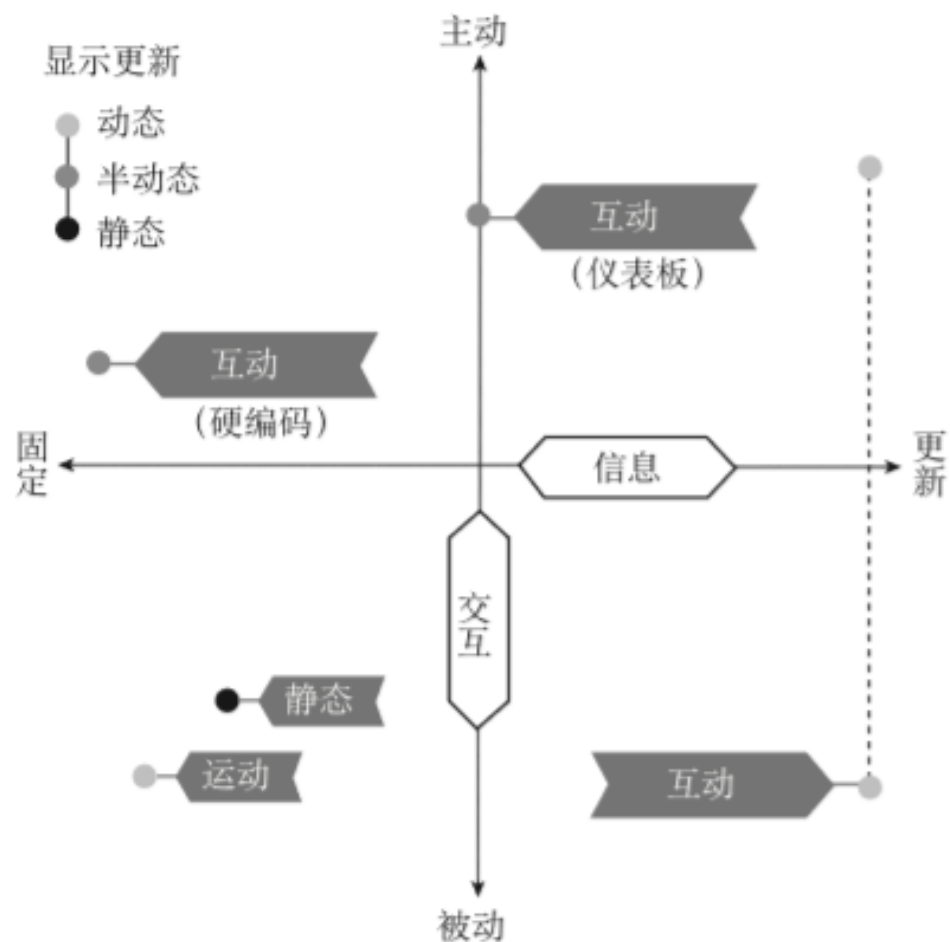


图 2-12 信息图表格式象限图

四、气泡图

气泡图与XY散点图类似，散点图一般表示二维信息，而**气泡图是对成组的三个数值进行比较**。除了X值、Y值外，气泡的颜色和大小也可以表达特定的信息，可以说气泡图能够传达四维信息。

与饼图、条柱图、折线图相比，**气泡图更为简约，既可做静图，也可做动图，视觉化效果更具优势，而且能够运用交互方式调动受众参与的积极性。**

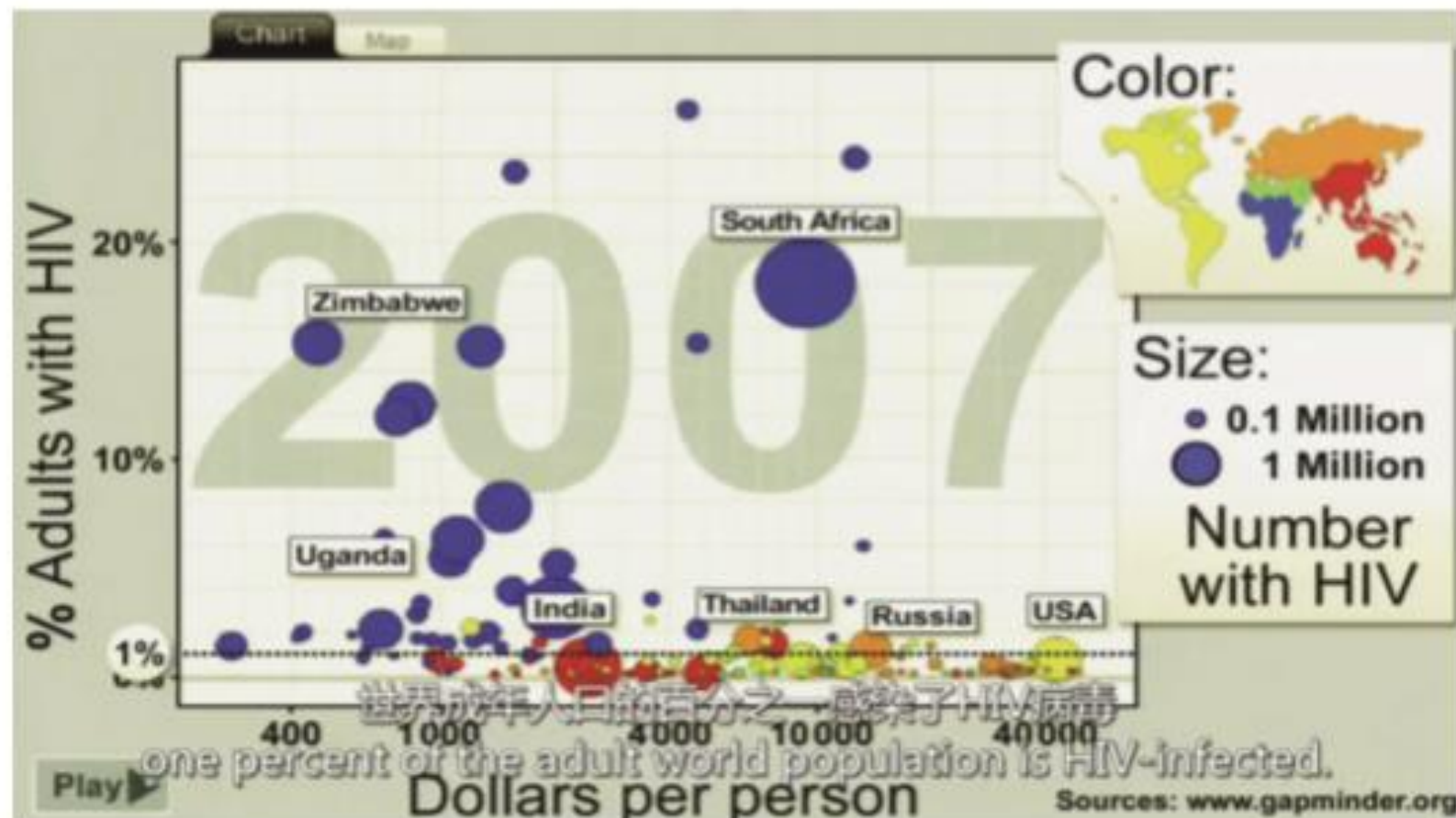


图 2-13 罗斯林在《谈 HIV 新数据和震撼的图表》的演讲中使用了气泡图

五、社会网络关系图

在新闻报道中，人物之间的社会关系是随着事件的发展逐渐显露的，而时间越长，空间越大，社会关系就越复杂，受众梳理起来就越困难，花费较多的心智成本将会严重影响受众持续阅读下去的兴趣。社会网络关系图能够帮助受众更清晰地了解新闻人物之间盘根错节的关系。



图 2-14 财新网制作的《周永康的人与财》

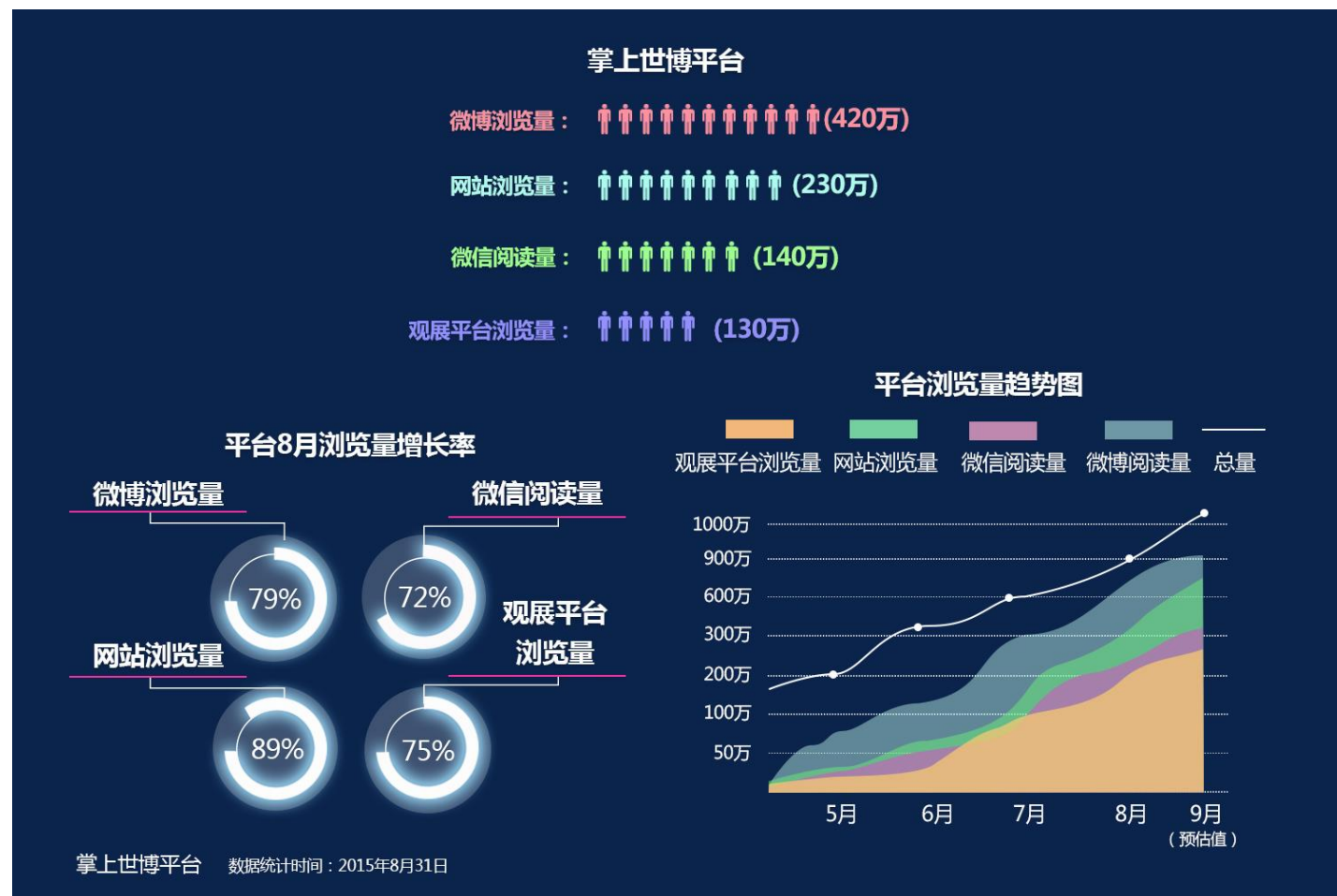
六、词频图

词频指的是某一个给定的词语在该文件中出现的次数。**词频统计是一种用于情报检索与文本挖掘的常用加权技术，用以评估一个词对于一个文件或者一个语料库中的一个领域文件集的重要程度。**对文本中的某些关键词进行词频统计，然后采用可视化的形式呈现出来，将文字信息变为形象化的符号，可以引导受众关注重点信息



总结

数据可视化的价值和意义并不只是用丰富多彩的图形将数据呈现出来，而是对核心信息的深层挖掘。因此，数据可视化不能满足于对浅层信息的梳理和呈现，如果缺乏对数据的深入分析，就很难发现新闻或新信息。数据可视化对新闻行业而言，也是新闻不断被发现和深化的过程。



第二部分 数据新闻可视化的生产实践

第一节 数据可视化的数据与工具

- ☞ 一、可视化中的数据类型分析
- ☞ 二、可视化的逻辑结构
- ☞ 三、数据可视化的工具

一、可视化中的数据类型分析

(一) 分类数据

分类数据 (categorical data) 指的是反映事物类别的数据，是离散数据(discrete data)。分类属性具有有限个（但可能很多）不同值，值之间无序。如人按性别可以分为男女两类，季节可以分为春夏秋冬。人们习惯于对事物进行分类，即所谓“物以类聚，人以群分”。关键是，分类可以带来结构化，否则就只是乱糟糟的一团东西。

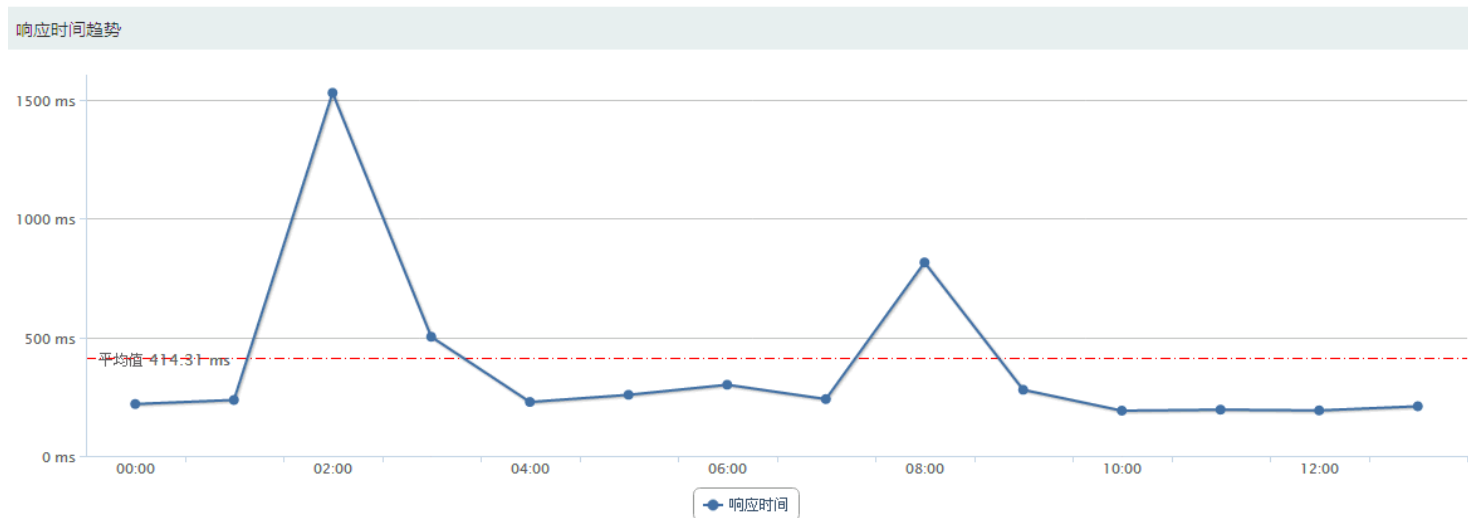


一、可视化中的数据类型分析

(二) 时序数据

时序数据 (timing data) 是指时间序列数据。它是同一统一指标按时间顺序记录的数据列，时间序列分析的目的在于通过找出样本内时间序列的统计特性和发展规律性，构建时间序列模型，进行样本外预测。

折线图和条柱图同样适用于呈现时序数据，只需将坐标轴上的分类更换为时间即可。同样可以采用不同数据粒度来呈现数据的变化。



一、可视化中的数据类型分析

（三）空间数据

空间数据（spatial data）是指凡是带有空间坐标的数据，它用来表示物体的位置、形态、大小分布等各方面的信息，是对现实世界中存在的具有定位意义的事物和现象的定量描述。空间数据是一种用点、线、面以及实体等基本空间数据结构来表示人们赖以生存的自然世界的的数据。



图 3-1 《纽约时报》网站的《2008 年大选结果》数据地图

二、可视化的逻辑结构

（一）相关性的逻辑结构

- 通过大数据技术分析出数据间的关联性，并以恰当的形式呈现出来，正是数据可视化要实现的目标。数据自己不会说话，我们通过运算模型（算法）来分析数据，挖掘数据背后的意义，还原事实或现象，预测未来。
- 相关关系的核心是量化两个数据值之间的数理关系。相关关系强是指当一个数据值增加时，其他数据值很有可能也会随之增加。相关关系弱就意味着当一个数据值增加时，其他数据值几乎不会发生变化。相关关系通过识别有用的关联物来帮助我们分析一个现象，而不是揭示其内部的运作。

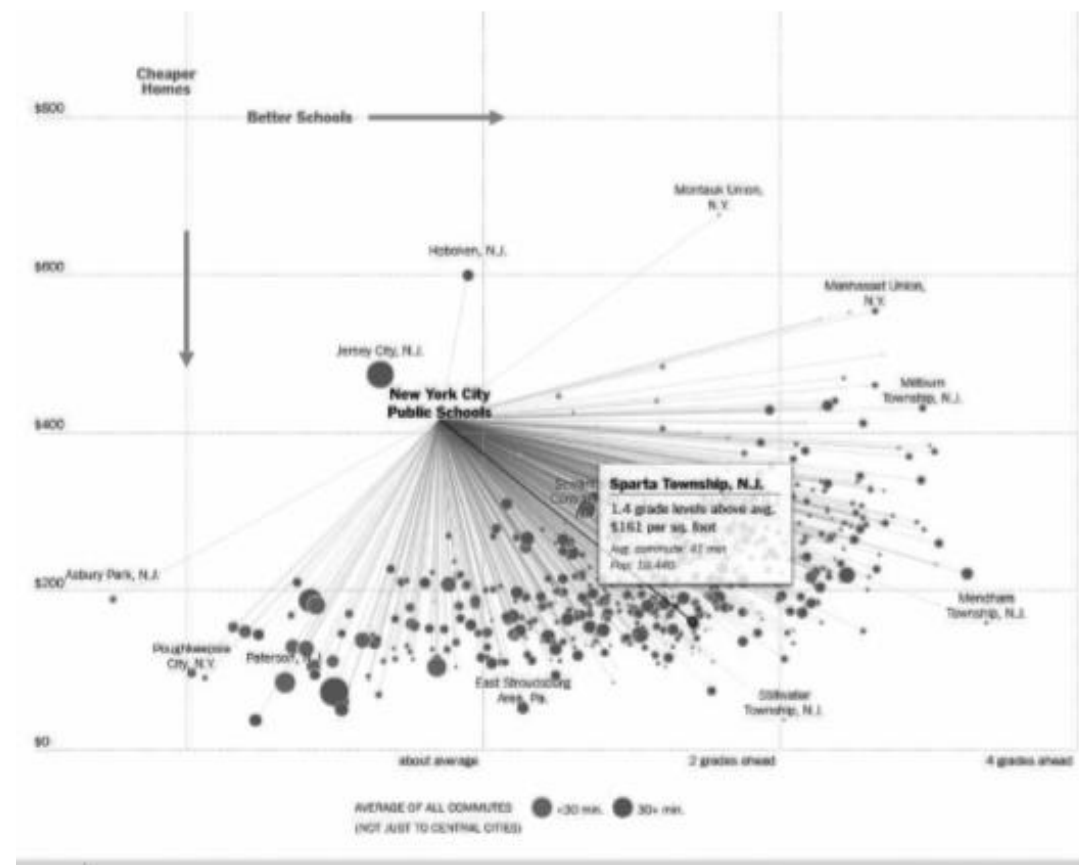


图 3-2 《纽约时报》网站的交互图表《好学校，买得起的房子：
找到城郊的最佳区域》

二、可视化的逻辑结构

（二）对比性的逻辑结构

对比分析是可视化叙事中的常用手法，从简单的条柱图、饼图到较为复杂的、汇集了更多时空维度的信息图，都通过分析、对比各种维度数据，挖掘出数据之间的关系，揭示出深层的原因，最后以数据可视化的方式向用户进行直观的展现。



图 3-3 《纽约时报》网站的交互图表《重塑纽约》

二、可视化的逻辑结构

(三) 演变性的逻辑结构

对事物的发展轨迹进行推演也是常见的一种可视化逻辑。这种演变性的逻辑结构往往基于时间维度或单一变量的比较，借助时间线来揭示事物在更大跨度内的变化。

从事多年数据可视化研究的斯坦福大学的**杰夫·麦吉**（Geoff McGhee）教授认为，现在的新闻越来越和数据有关，媒体有责任向公众解释那些复杂难懂的数据。可视化方面的专家正在开发工具帮助普通人更好地理解数据，记者们则努力做到如何应用数据使新闻报道更具说服力，有经验的数据图表设计师能够把数据引入新闻学。



图 3-4 财新网的交互图表《星空彩绘诺贝尔》

第二节 数据可视化的生产

- ☞ 一、数据可视化的生产流程
- ☞ 二、数据可视化的设计原则
- ☞ 三、数据可视化的应用方式

一、数据可视化的生产流程

- 美国邱南森 (Nathan Yau) 在著作《数据之美》中把数据可视化过程描述为四个步骤：

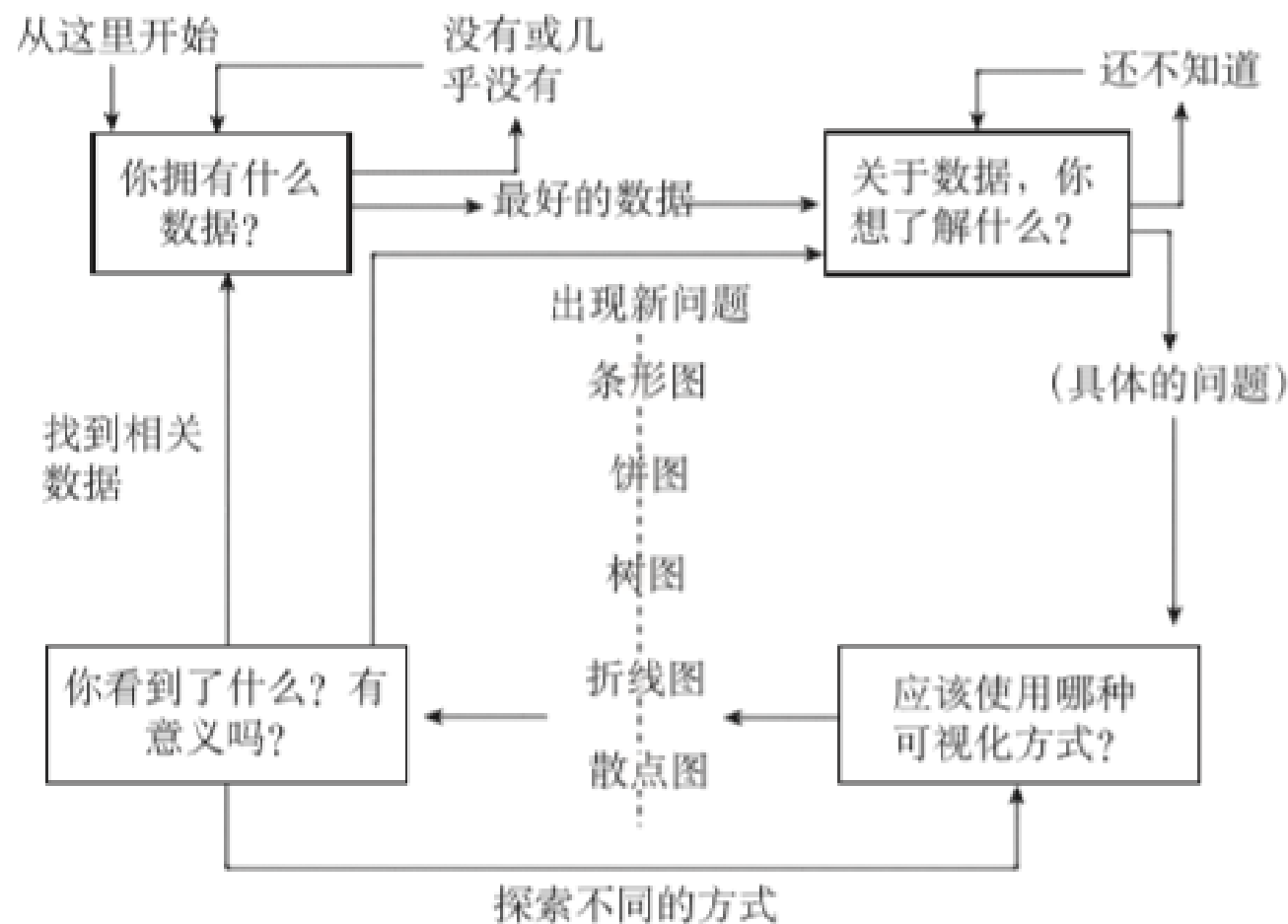


图 3-5 迭代的数据探索过程

一、数据可视化的生产流程

- 本·弗莱 (Ben Fry) 在著作《可视化数据》中把数据可视化的流程分为七步：获取、分析、过滤、挖掘、表示、修饰、交互。

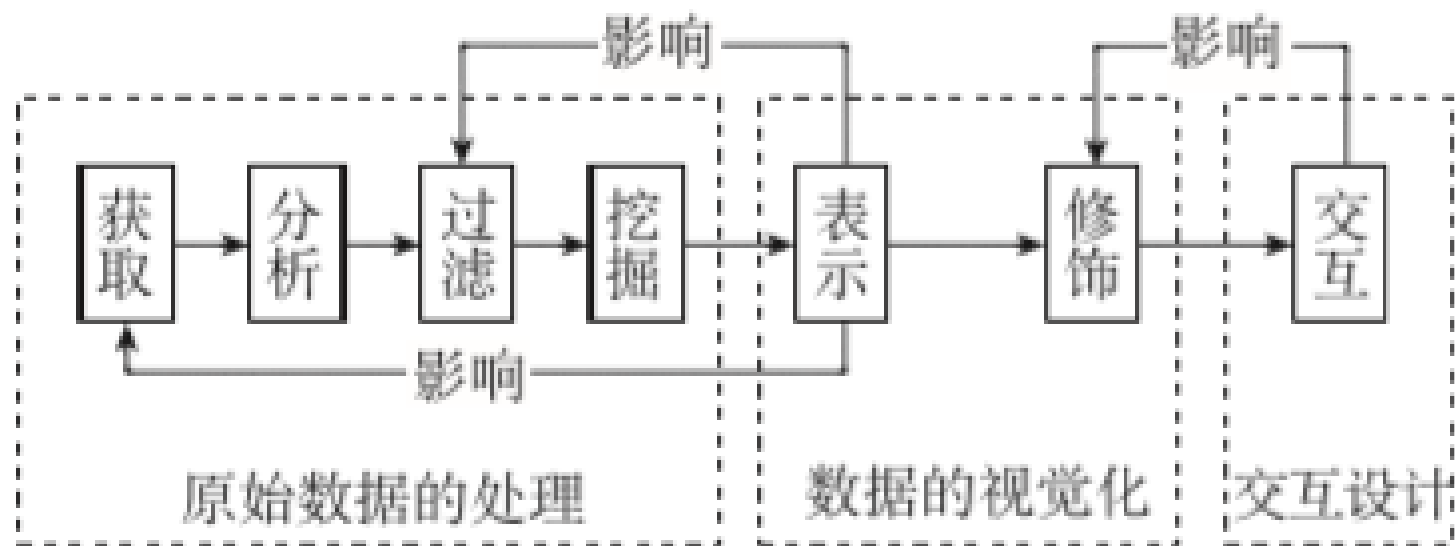


图 3-6 数据可视化的生产流程图

一、数据可视化的生产流程

- 原始数据的处理

- 此部分包括本·弗莱所说的获取、分析、过滤和挖掘四个环节。随着各种数据库的开放，新闻从业者业务素养的不断提高，搜寻数据的方法和途径也日益增多。数据有了，并不意味着可以进入数据分析的环节，为了保证数据的质量和数据新闻的可信度，要对需要加工的素材数据进行清理。清理之后要对数据进行初步分析，洞悉各数据源之间的关联性，识别数据的用途，发现数据变化的趋势等，然后才能将可用数据、参考数据等录入电子表格。数据分析是报道（讲故事）的组成部分，分析数据的过程就是发现新闻点、剖析数据背后的新闻意义的过程，进而预测可能发生的事件。在大数据时代，缺乏的不是数据，而是对数据的准确分析、深入挖掘。

一、数据可视化的生产流程

- 数据的视觉化

- 此部分包括本·弗莱所说的表示和修饰两个环节。数据新闻可视化的方式千差万别，但终极目标是把重要的新闻传播出去，帮助受众迅速发现有用的信息，并更好地理解之。数据可视化的效果取决于所用的数据的质量，包括数据的体量、准确性、关联性等。可视化是手段，是把数据分析的结果、需要向受众阐释的意义，以一种视觉的方式呈现出来。若文字或多媒体能够把故事讲得更精彩，可以考虑不采用可视化的表现手法。正如传统新闻报道“七分采访三分写作”的说法，数据新闻在制作过程中，花费在采集、清理、分析数据上的时间占到了70%，可视化环节只占30%。

一、数据可视化的生产流程

- 交互设计

- 此部分就是本·弗莱所说的交互环节。在这一环节，可视化产品的制作者要完成用户获取内容及其属性路径的设计。交互式的可视化产品之所以为受众所喜爱，是因为它给了受众对于所呈现的数据或可视化的部分操控权，吸引受众停留得更久，并产生分享的欲望。交互设计的目的不仅在于帮助用户寻找数据与每个人的联系，让他们从一般的信息接收者转变成信息利用者，而且通过分析受众的关注细节，为其提供差异化内容。在交互环节，受众从被动接收数据转变为积极主动地去发现、思考数据，从而为受众提供了控制和探索数据的可能。

二、数据可视化的设计原则

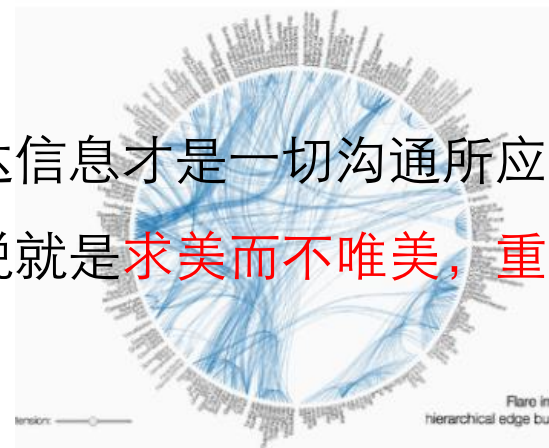
（一）功能性原则

- “功能决定形式” 源于1896年，出自美国建筑师路易斯·沙利文（Louis Sullivan）的文章《经过艺术设计的高层办公建筑》。技术实体的形式必须依赖于其辅助完成的任务。在处理信息图表与可视化时需要记住的重要原则之一是：形式应该受制于你所展示的功能。
- 数据可视化的目的就是满足受众的信息需求，尽管不同受众因个体差异而对可视化产品的认知存在差别，但是，借助视觉化工具，实现意义的建构和传达是数据可视化不可或缺的功能。
- 功能性的实现需要可视化工作者认真考虑三个因素，即目标受众、传播主题、可视化类型之间的平衡与协调。目标受众与传播主题应该是一致的，而这个受众群的认知水平与解读能力决定了可视化呈现类型的选择，在一定程度上也影响到对数据的挖掘深度等。既要考虑目标受众的实际情况，又要以数据所蕴含的价值为选择标准，才能寻找到最恰当的可视化方式。

二、数据可视化的设计原则

(二) 美观性原则

- 数据可视化就是利用最小的空间，高效地视觉化一组繁杂的数据，说到底就是以独特的视觉优势推动内容的传播。在所有分析和传播统计信息的方法中，一张精心设计的数据图表通常是最简单也是最有效的。为了实现数据可视化的传播有效性，适当的美化修饰是必然的。
- 但是，数据新闻可视化所承担的是沟通的任务，而不是艺术，确保所传递的信息被清楚地理解是最紧要的。在进行有效视觉沟通时，如果是具体的事物，能用照片则用照片；如果是抽象的数据，则可以用可视化来呈现它。
- 无论是以文字形式还是采用视觉形式来传递信息，清楚、有效地传达信息才是一切沟通所应该达到的目标，视觉符号仅仅是增强沟通效果的有效手段而已，简单说就是求美而不唯美，重要和实用要胜于好看。



二、数据可视化的设计原则

(三) 启示性原则

- 在数据可视化设计中，**启示性原则的核心就是帮助受众更好地解读数字、图形等符号，提高对数据的分析能力。**根据认知科学原理，人们的注意分为**无意注意与有意注意**。注意是一个互动的过程，只有被注意客体与信息接收主体之间发生接触，而且客体必须具有吸引主体的刺激点，才能引发主体的注意，才能促成信息的传播。**数据可视化的传播属于无意注意，而无意注意离不开刺激，图形元素则发挥着刺激物的作用。**换句话说，设计者可以通过使用特异、新颖的图形元素来营造刺激点，以求达到吸引受众注意的目的。
- 可视化生产者为了实现启示性目标，需要控制设计的形式、创造层次感和意义，通常会参考**格式塔法则**，如接近、相似、主体背景和断点续接等。

三、数据可视化的应用方式

（一）解读文本信息

可视化解读文本信息，一般通过统计关键词出现的频率，或者提取关键信息点来实现。比如，近两年的一些主流媒体在两会期间就通过对报告中关键词的提取来解读其中的重要信息，这些关键词以数据为主，并附以直观图形，便于受众理解和记忆。

（二）提供背景知识

出于为受众解疑释惑，帮助受众全面、深刻理解新闻的需要，可以在适当的时候配发一些背景材料，包括新闻人物与组织、地理气候、科学知识、名词术语等，而这些信息既可以文字、照片的形式出现，也可以数据图表的方式来展示，其目的是帮助受众全面、深入地掌握和理解信息，对新闻价值有更清楚的认识。

三、数据可视化的应用方式

（三）揭示事物之间的联系

清晰地揭示报道客体背后的深层联系有助于受众理解新闻事件的真实情况和发展方向。联系即互动，报道客体间的互动具有复杂性，以抽象的文字来表述恐难以胜任，而可视化的方式则能较为轻松、明晰地表述和解析其中隐藏的联系，从而揭示新闻的深意。

（四）阐述观点

数据可视化是在形象化地展示数据，但是，这种展示遵循着一定的逻辑关系，而非随意性的排列和罗列。在可视化之前的数据分析阶段，梳理数据之间的逻辑关系是其一项重要工作内容。可视化的数据即论据，经过图形化处理的数据，逻辑关系和论证思路明晰，更具易读性、易懂性，得出的结论往往也不证自明、不言而喻，其观点的说服力更强，更容易被接受。

三、数据可视化的应用方式

（五）基于时空维度呈现信息

事物的存在离不开时空。以时间线的方式来呈现不同时间节点产生的信息（数据），不仅能够清晰地叙述故事、聚合新闻，而且有助于通过分析具有历史跨度的信息，挖掘出数据背后的深意。数据地图则是从空间维度来呈现新闻要素的空间信息及其相互间关系的。数据地图的最大优势在于空间信息呈现的直观性和各新闻要素与地理关系的明确性，地图、数据与色彩的组合，以及交互技术的应用，优化了数据可视化的用户体验和传播效果。

数据可视化不仅仅是对数据的视觉化处理，更是对内容建构与形式美感之间的平衡，用巧妙的可视化方式争取比文字更好的叙事效果，便于受众理解复杂的情景。

第三节 数据可视化的设计要点

- ☞ 一、数据与数据源
- ☞ 二、常用图表的设计规范
- ☞ 三、数据可视化中的色彩
- ☞ 四、数据可视化中的图例与图标
- ☞ 五、数据图表的版式设计

一、数据与数据源

判断数据好不好关键要看它的来源。一般而言，来源的权威性、可靠性和中立性是高质量信息源的必要条件，特别是经过第三方审核的数据更能避免偏颇，增强可信度。数据是构建信息图表的内容元素，一个错误的数字可能会毁掉整个图表，即坏数据 + 好的可视化 = 坏图表。数据可以有不同的可视化方式，具体需要用哪种方式，所要传达的信息和新闻工作者所追求的传播效果是做出选择的依据。

在数据收集时，可以说“多多益善”，数据越多越有利于深入分析。为了确保数据来源的质量，需要注意以下几点：一是数据来源可以用来讲故事；二是确保数据来源的可靠性；三是确保数据来源与主题密切相关。

二、常用图表的设计规范

(一) 折线图的设计规范

折线图常用于表现一种或几种事物发展的状态和趋势。

- 合理调整纵坐标上刻度的精度，尽量使线段的高度既能真实地反映事实，也能取得较舒适的视觉效果。
- 适度调整纵坐标刻度的精度，同时把底部的零基线加粗一点。
- 多线折线图用不同的粗细或深浅来区分线段。
- 关键点（重要信息）的标注：一是文字量不要太多；二是标注的位置距离线段，特别是关键点不要太远。可以考虑在线段或关键点上直接标注的方法来解决问題。



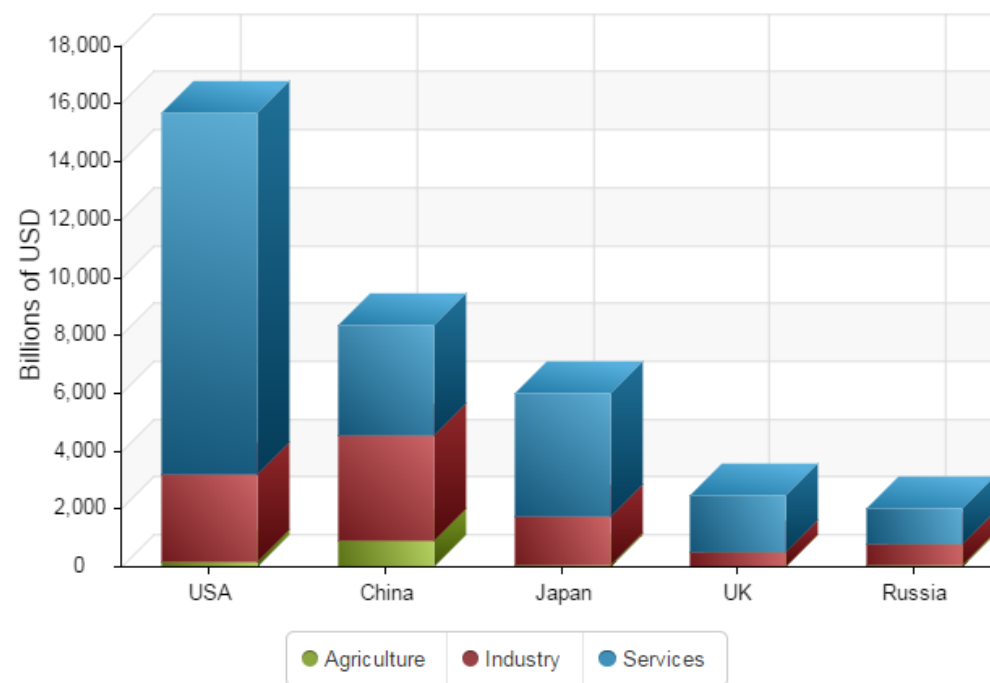
二、常用图表的设计规范

(二) 条柱图的设计规范

条柱图是柱状图与条形图的合称。

- 柱形的宽度等于或略宽于相邻两个柱形之间的空白即可；
- 最好不要给柱形添加过多的装饰，过度装饰不能传达任何信息或数据，只会分散注意力；
- 在绘制柱状图时，零基线不能省略；
- 条形图是按照某一属性，对所有个体进行的排序，最好按照数值从大到小或从小到大的顺序进行排列。随机排列会导致图表参差不齐，使条形图失去原有的功能。

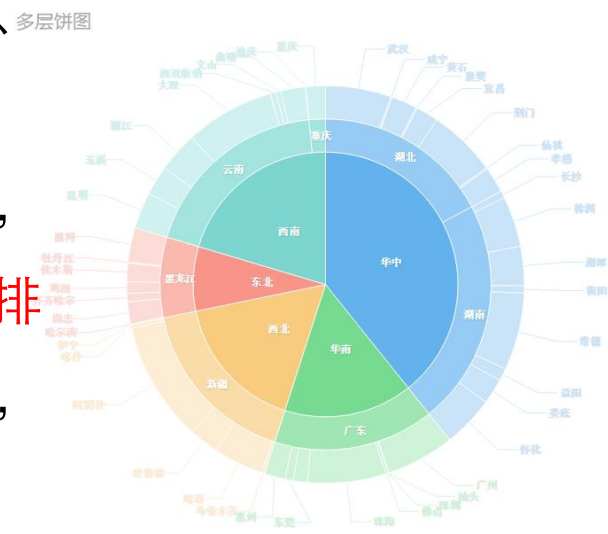
Major economies by GDP sector composition (2011)



二、常用图表的设计规范

(三) 饼图的设计规范

- 一般情况下，切分的份数不要太多。如果份数过多，可以考虑把占比重较小的几个部分合并组成新的一块“饼片”，定名“其他”。
- 如果每一部分确实需要独立呈现，可以考虑进行再切分，即“大图加小图”。再切分后的小图在类型上最好还用饼图，以保持读者的惯性思维，把“小图”中每个“饼片”的数值或在原始“大图”中所占的百分比，以图例方式标注或直接标注出来。
- 饼图是以面积来呈现部分信息的。12点位置是视线的起点，从视觉考虑，最大部分放在12点位置的右边，以突出其重要性。余下部分最好的呈现排序是把第二大部分放在12点位置的左边。剩下部分按照逆时针方向排序，这样的话，最小的部分（最不重要）就会落在饼图的底部。



三、数据可视化中的色彩

色彩是极具表现力的视觉工具，几乎没有什么视觉元素能够像色彩那样具有强烈的表现力。色彩含有丰富的心理信息，可以影响内容的传达。

- 用同一色彩来表示同一变量；
- 不要在有多个柱形的柱状图中使用不同的色彩，强烈的对比同样会分散读者的注意力，使用同一色彩的渐变明度或者同一色系的颜色；
- 解决色盲问题：一是对图表中的重要信息元素进行“双重强调”。二是调整色彩明度上的差异。三是对可视化中的数字、文字采用背景反衬的手法。

四、数据可视化中的图例与图标

图例是集中在图表上的各种符号和颜色所代表内容与指标的说明，有助于更好地认识、理解图表。图例在绘图时是表示图表内容的准绳，在读图时是必不可少的阅读指南。



四、数据可视化中的图例与图标

图标指的是以动物、人物、植物、器物、建筑及自然景观等视觉元素，进行以写实为主的造型表现，是用具象思维设计制作的小图形，一目了然地指代某类事物，排列多个相同的具象标志使其具有数量意义。尽管图标直观、形象，在视觉吸引力上有较强的优势，但是，在呈现大容量数据方面却不如条柱图、折线图有效，图标仅限于传达较少的信息。另外，也不能以艺术的标准来要求具象标志，过多的细节也会分散读者的注意力。图标设计要尽可能简单，即使在大量使用的情况下，依然能够清晰易辨。



五、数据图表的版式设计

（一）依据内容排列组合

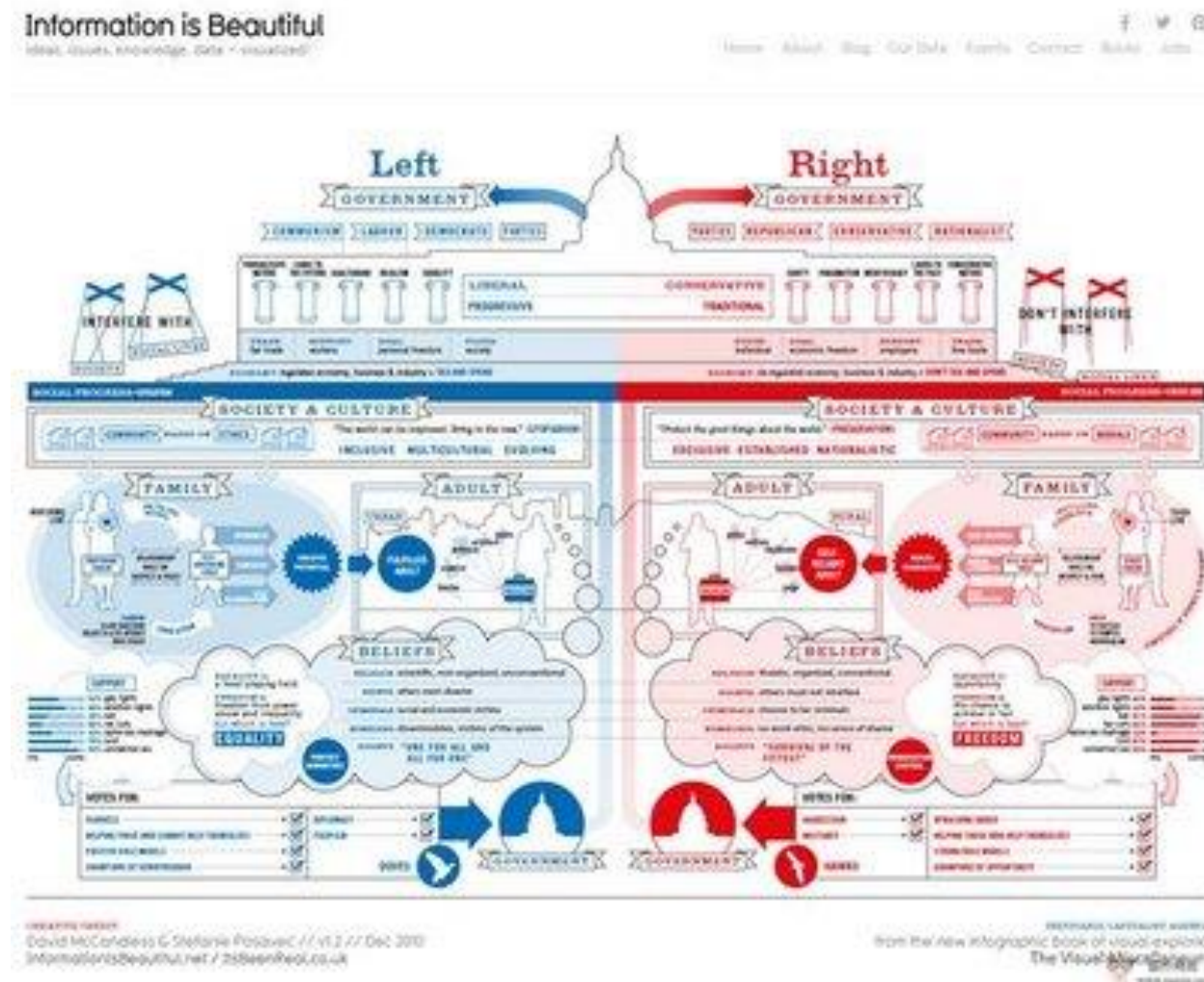
信息量少但内容重要的图表，常常以主题图案为轴心，其他要素环顾四周形成向心力。信息量多的图表往往以各要素为主体，根据内容间的逻辑关系均匀分布。版面自然分割是实现构图均衡的方式之一。有时候还可以采用“品”字形构图，内容主次分明，图形与信息相得益彰。有饼图、有柱状图，各自独立又相互关联。还有些图表可以采用左右或左中右的水平布局方式，尊重人们从左至右的阅读习惯，其信息点之间要么是并列关系，要么是总分关系。



五、数据图表的版式设计

(二) 灵活运用对称版式

采用对称版式的题材往往是有关双方的斗争、竞争、竞赛，或者事物在不同发展阶段的状况对比等，这种版式的优势是逻辑关系清晰明了，设计制作时省时省力；不足是过于平稳，缺乏变化。采用这种构图的图表往往有着明显的表格痕迹，所选用的图形也处于对称的位置。



五、数据图表的版式设计

（三）依据优先律引导阅读顺序

优先率指的是图表中各部分内容所占的尺寸比例。当媒体希望受众注意到某些内容，并遵循一定的读图顺序时，可以通过调整图像或文字的大小来引导受众。一般情况下，可以通过调整图形的大小，营造反“6”字形视线流，这种构图增强了版面的新颖性和变动感。有时也可以通过在页面中加入引导读者视线移动的元素，从而对视线流向进行预设。通过设置小图形，引导读者的视线按照“Z”字形移动。

（四）利用段组分割增强易读性

段组的作用就是保证图表易读、易懂。把关系密切的信息放在一起，并采用同一色彩作为标识，较容易和其他信息相区别。从版式布局来看，段组有时容易和对称结构混合使用。

五、数据图表的版式设计

（五）合理使用留白

犹如中国画讲究阴与阳、明与暗的平衡，图表构图需要考虑留白，像报纸版面一样，需要呼吸感，不能太拥挤，不要把每个空间都放上视觉元素，要通过留白给受众以视觉的缓冲，这样在视觉效果上才能更具有吸引力。在图表中留白的最简便方法是删除仅用于装饰的图形，或者缩小一些图形的尺寸，并不是所有的图形都要大小一致。当然，物极必反，过度留白也会降低信息图表的影响力，当受众看到大片的空白会认为这一部分不重要。

数据可视化有时过于强调视觉效果，经常采用图形的特效，其结果就是过度装饰。适可而止、懂得取舍和注意避让，以最简单的方式来呈现信息，站在艺术性和信息量的角度来制图才是实现最佳传播效果的捷径。

本章结束!