

大数据时代数据关系的视觉构建

萧冰

(上海交通大学, 上海 200240)

摘要: **目的** 分析大数据背景下,数据关系的视觉构建。**方法** 结合图形设计和认知心理学相关理论,归纳总结大数据可视化的设计原则。**结论** 大数据可视化应遵循准确性、直观化、生动性、读图化、简约性、动态分层次信息呈现6大设计原则,未来的大数据信息可视化将向交互、动态的方向发展。

关键词: 大数据; 视觉构建; 信息可视化

中图分类号: TB472 **文献标识码:** A **文章编号:** 1001-3563(2016)08-0005-04

Visual Construction of Data Relationships in the Era of Big Data

XIAO Bing

(Shanghai Jiao Tong University, Shanghai 200240, China)

ABSTRACT: Under the background of big data, it analyzes the visual construction of the data relationship. According to the theory of cognitive psychology and graphic design, it summarizes the design principle of data visualization. Big data visualization should follow six design principles: accuracy, intuitive, vivid, image, simplicity and dynamic hierarchical information presentation, and the future of big data information visualization will be interactive, dynamic direction of development.

KEY WORDS: big data; visual construction; information visualization

自从大数据时代的概念被提出以来,大数据科学作为一个横跨多领域的新兴交叉学科正逐渐成为研究的热点。众所周知,大数据具有海量、多样、高速和价值的特征^[1]。大数据处理的核心是数据分析,但对于用户来说,更关心的往往是数据分析的结果展示。不恰当展示方法很可能让用户难以理解甚至产生误导,而可视化技术作为最佳的展示方式,能够直观地将数据分析的结果展示出来^[2]。要在这高速变化的海量数据中迅速寻找它的价值,大数据信息可视化是非常有效的发掘工具。可视化是从哪些方面帮助研究者建立数据关系的,在可视化过程中应当遵循怎样的原则,这都是需要思考的问题。

1 大数据信息可视化的学理基础

1.1 信息可视化的基础

从学科支撑的角度来讲,认知心理学和图形设计是信息可视化的两大基础学科。认知心理学是当今心理学发展的主流方向,是用信息的加工来解释人的心理的过程,它将人的内部心理机制用信息加工的理论和方法表现出来;图形设计是指以图像、文字、符号、色彩等多种手段进行创造,以此来传达想法或信息的视觉设计形式。图形设计涉及对视觉信息的抽

收稿日期: 2015-12-16

基金项目: 教育部青年基金项目(13YJCZH205);上海交通大学文理交叉认知专项(13JCRY03)

作者简介: 萧冰(1976—),男,山东人,上海交通大学副教授,主要研究方向为新媒体艺术和视觉传播。

象提炼、归纳整合、变形夸张等艺术手段,可综合运用点、线、面、体、色彩、光线、肌理、透视等视觉元素,以及疏密、节奏、构图、比例等艺术法则来进行艺术表现。图形与文字的组合编排是图形设计的有机组合部分。

图形设计与心理学一直有着密切的联系。鲁道夫·阿恩海姆对形状、色彩、位置、空间、光线等视觉元素的心理效能进行了系统地分析后指出:与科学家一样,艺术家用类似于形和色彩的要素去猎取那些具有普遍意义的东西^[3]。毋庸置疑,设计师在通过图形将信息传递给观者时,必须要考虑观者的心理认知因素。

1.2 大数据可视化的分类

大数据可视化可分为文本可视化、网络可视化、时空数据可视化和多维数据可视化4种类型^[4]。典型的文本可视化技术是指标签云和语义结构的可视化;网络可视化技术是基于节点和边的可视化,包括H状树、圆锥树、气球图等;流式地图和时空立体是时空数据可视化的常用方式;多维数据可视化的表现方式是散点图、投影法和平行坐标。

1.3 可视化研究的变量

可视化变量选取的恰当与否会直接影响信息展示的质量。有些学者将可视化变量分为位置、形状、方向、纹理、色彩、灰度等级与尺寸7种,把色彩分成色相、明度和饱和度^[5],可视化变量的感知性质见表1,表现了可视化变量的属性。

表1 可视化变量的感知性质
Tab.1 Perceptual properties of visual variables

	联系性	选择性	次序性	数量性
位置	+	—	—	—
形状	+	—	—	—
方向	+	0	—	—
颜色	+	++	—	—
纹理	0	+	0	—
灰度等级	—	+	++	—
尺寸	—	+	+	+

注:++很好;+好;0中等;—差

色彩在表达选择性方面是最突出的,纹理同样在选择性方面表现突出,比如可以把某些图形用斜线表示,而另外一些用方格予以区分,但这种表示方法没有色彩的差异那么显著。灰度等级是指将某色彩混以不同比例的黑色或白色,使之呈现出明度阶梯差

异,使其在表达次序性方面有明显优势。位置等变量在表达联系性方面具有优势,比如在结合地图的可视化设计中,信息与其所出现的位置具有天然的联系。

2 大数据可视化的设计原则

Andreea提出了一套信息可视化应该遵守的一般规则:(1)展示数据,诱导观众思考内容而不是方法论;(2)图形设计或技术化的图形制作;(3)避免扭曲数据要说的内容;(4)在较小的空间展示大数据;(5)引导视线在不同数据间进行比较;(6)在不同细节层次显示数据^[6]。胥琳佳认为,普通的图表无法适应大数据视觉展示的需求,可视化展示应当通过清晰美观的图示,揭示数据的真正含义^[7];王斌则认为信息可视化的核心是把数据放入一定的情境中进行展示^[8];而高立伟提出应当在信息可视化中,运用虚拟现实技术,为大数据的探索和认知提供便利^[9]。

以上观点从可视化技术或新闻报导的角度对大数据可视化的原则进行了讨论,但却忽视了信息可视化的支柱,即对图形和认知的视角阐述。结合图形设计与认知理论,大数据的信息可视化设计应当遵循以下6项原则。

2.1 准确性原则

准确性是大数据可视化最基本的原则。庞杂的信息转换为视觉信息时,难免会有偏差或解读上的歧义。准确性原则首先要求其在可视化的过程中不要扭曲数据,其次要求其清晰地描绘数据。如果图像中的信息较少会比较容易理解,如果信息太多可能会导致理解困难,那么就需要通过确立信息的维度,使画面信息条理清晰。对于更复杂的信息而言,可通过确立坐标、刻度和范围等来增加数据的准确性。

2.2 直观化原则

如果大数据图表上充斥着太多的线条和色彩,用户往往会感到不解和茫然。大数据信息可视化不仅是将信息转换成抽象的几何图形与色彩,而且是要以用户能够理解的方式将信息之间的视觉化关系表达出来。关系是大数据的核心要素,也是一个充满变数的要素。尤其是在将抽象的关系用具象的视觉要素落实到画面时,设计师所选择的图形、色彩、表述方式等应当直观、便于理解,这也是为什么地图、几何形、色彩等最基本的元素在可视化设计中被应用得最多的原因。

2.3 生动性原则

在庞大的数据库中提炼出若干信息,并通过视觉化的方式将之呈现在用户面前时,需要注意设计的生动性。用户不能直接接触到具体鲜活的信息,对所呈现的宏观信息有隔阂和距离感,呆板的画面会让用户很快失去进一步阅读的兴趣,而生动且富有创意的画面不但能够吸引用户深入阅读,而且还能够帮助用户更快、更好地理解设计意图。

生动性原则要求可视化信息具有艺术性,以富有创意的图形和表现手法来展现可视化信息。表现力是评价可视化有效性的重要标准,实现生动性的有效方式是赋予信息情境,让读者在特定的情景中理解信息。地图、具象图标、实景照片等都有助于营造情境^[10],具有很强的代入感。

2.4 读图化原则

可视化就是读图化吗?其实不然。人在阅读文字与读图时,视线运动的方式和掌握信息的方式都是不同的。在阅读文字时,视线运动遵循文字的书写方向,顺序且线性地移动,不太会发生跳跃性观看的情况;而在读图时,视线在画面中是任意跳跃的,它没有固定的规律和方向,可以从画面的任意位置开始,不仅可在画面中随意跳跃观看,而且还不影响对画面的理解。

有些信息可视化设计虽然将文字转换成了图表,但是在阅读时仍需按照由左到右、由上到下的规律阅读,这种方式或许对理解既有的逻辑信息有所帮助,但不利于创新性地发现,无益于用户在比较不同数据中发掘其间的新关系。

2.5 简约性原则

面对大数据,无论多大的画面都显得很小时,无法将所有数据一一呈现,因此就要求画面中的视觉元素尽可能简约,使每个元素都有出现的必要性,冗余的元素会使画面看起来更加繁复,不利于人们发现数据间有价值的关系。George指出,人们短时间内能够接受、处理和记忆的信息量非常有限,特别是对抽象概念或不常用的数据^[11]。少即是多是设计领域的重要原则,只有直截了当的可视化设计才具有高效性。

系统的图形构建是实现可视化的必要因素,坐标轴和风格统一的图标与合理的布局都能使信息更加清晰易懂,图形的序列、比例、量级等有助于建立数据之间的关系。色彩是画面中系统性最强的视觉元素,

依据光谱解析出的色彩除固有的物理属性之外,还具有色彩心理与色彩联想的意义,比如具有前进感的暖色比具有后退感的冷色更能引起读者的注意,大面积邻近色中的少量对比色更容易被关注。

2.6 动态分层次的信息呈现原则

大数据需要呈现海量的数据信息,然而并不是所有用户对数据都具有相同层次的需求。从宏观方面或从信息的某个具体侧面需要来了解信息,按照不同层次进行信息的动态展现是非常必要且有效的。

让用户通过与计算机的简单交互实现不同层次之间信息的转换,就如同Google Earth可以让用户通过鼠标的滚轮,在任意比例尺下观看地图,既可以整体观察整个地球的样貌,又可以不断聚焦城市甚至建筑,使不同层级的信息可以根据用户的需要逐次呈现。

《谁引爆了击毙本·拉登的消息》(图片摘自SocialFlow官网)见图1,充分体现了以上几项原则。SocialFlow公司在分析了近15 000 000条与本·拉登相关的Twitter消息后,创造了这个银河系一般的流程图。该图准确地将Twitter中信息传递的过程展现了出来,黑色背景下的蓝白色名字汇聚成大小疏密不均的光团,如同爆发的星系,不但新颖独特,而且表现出了数字艺术的魅力。读者的视线可以在任何一个星团上跳跃,发现其中最引人注目的还是Keith Urbahn和Branstelter两个集群。图1表达的仅仅是本·拉登信息的传递关系,与之无关的信息均被省略了,以用户名和其相互关系构成的抽象的点、线、面组合,将画面元素精简到了极致。

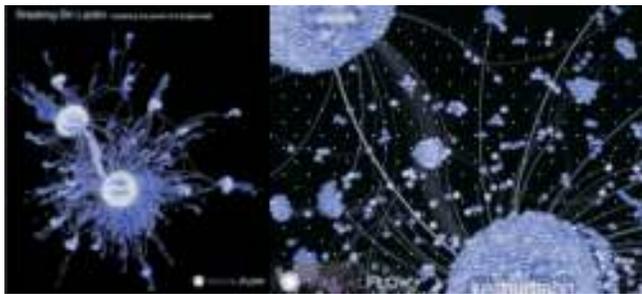


图1 《谁引爆了击毙本·拉登的消息》
Fig.1 "Who blew up the news of Ben Laden"

3 大数据信息可视化的发展趋势

当前,结合了大数据的信息可视化研究少之又少,大多数的研究成果也产生于大数据概念出现之前。究其原因,一方面可能是因为大数据概念爆发迅速,相关研究尚集中于技术层面,对数据可视化的研

究还未上升到社会研究的高度,还未能从信息接受者的角度研究信息的可视化;另一方面,信息可视化的另一支柱,即图形设计一向是以实践先行的,目前处于大数据信息可视化的个人设计体验的总结阶段,还不太可能针对此产生系统性的理论成果。随着大数据可视化设计的发展,未来数据可视化将注重从接受者的角度来阐释大数据,将更注重对图形的选择和运用,可视化也将呈现出更加多样化的面貌。

4 结语

未来的大数据信息可视化必然具有交互性和动态性特征。一方面,大数据时时刻刻都在更新着,静态的图形已经不能完成这一任务,必须随着数据的发展而变化;另一方面,将人纳入大数据的交互体验中,在互动中了解、参与分析和脑机接口等生物科技的介入^[13],给人们提供了更多观察、发现新的相互关系的机会。

参考文献:

- [1] 刘智慧,张泉灵.大数据技术研究综述[J].浙江大学学报,2014(6):96—97.
LIU Zhi-hui, ZHANG Quan-ling. Research Overview of Big Data Technology[J]. Journal of Zhejiang University, 2014(6): 96—97.
- [2] 孟小峰,慈祥.大数据管理:概念、技术与挑战[J].计算机研究与发展,2013,50(1):146—149.
MENG Xiao-feng, CI Xiang. Big Data Management: Concepts, Techniques and Challenges[J]. Journal of Computer Research and Development, 2013, 50(1): 146—149.
- [3] 阿恩海姆·鲁道夫.艺术与视知觉[M].成都:四川人民出版社,1998.
ARNHEIM R. Art and Visual Perception[M]. Chengdu: Sichuan People's Publishing House, 1998.
- [4] 任磊,杜一,马帅.大数据可视分析综述[J].软件学报,2014(9):190—193.
REN Lei, DU Yi, MA Shuai. Visual Analytics Towards Big Data[J]. Journal of Software, 2014(9): 190—193.
- [5] 宋绍成,毕强,杨达.信息可视化的基本过程与主要研究领域[J].情报科学,2004,22(1):13—18.
SONG Shao-cheng, BI Qiang, YANG Da. Basic Process of Information Visualization and Primary Research Field[J]. Information Science, 2004, 22(1): 13—18.
- [6] ANDREEA M. Scientific Images and Visualizations in Digital Age, from Science to Journalism[J]. Journal of Media Research, 2012, 14(3): 10—12.
- [7] 胥琳佳.大数据对于传播学研究内容和方法的影响[J].中国出版,2013(9):3—6.
XU Lin-jia. The Impact of Big Data on the Content and Methods of Communication[J]. China Publishing Journal, 2013(9): 3—6.
- [8] 王斌.大数据与新闻理念创新——以全球首届“数据新闻奖”为例[J].编辑之友,2013(6):16—19.
WANG Bin. Big Data and News Concept Innovation: the First Global "Data Journalism Award" as an Example[J]. Editors Friend, 2013(6): 16—19.
- [9] 高立伟.关于大数据时代数据信息可视化的研究[J].电子世界,2013(8):19.
GAO Li-wei. Research on Visualization of Data Information in the Era of Big Data[J]. Electronics World, 2013(8): 19.
- [10] 杨彦波,刘滨,祁明月.信息可视化研究综述[J].河北科技大学学报,2014,35(1):91—92.
YANG Yan-bo, LIU Bin, QI Ming-yue. Summary of Information Visualization Research[J]. Journal of Hebei University of Science and Technology, 2014, 35(1): 91—92.
- [11] GEORGE A. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information[J]. Psychological Review, 1994(2): 34.
- [12] 陈志刚,鲁晓波.大数据背景下信息与交互设计的变革和发展[J].包装工程,2015,36(4):6—9.
CHEN Zhi-gang, LU Xiao-bo. Reformation and Development of Information and Interaction Design Based on the Big Data[J]. Packaging Engineering, 2015, 36(4): 6—9.