

数据可视化

来源：

史蒂文·布劳恩：《数据可视化在身边》

数据可视化的定义

在定义什么是和什么不是数据可视化是不同人有不同的意见。

一些定义突出美学，强调数据可视化内在的视觉和传达性质。如

- 海伦·肯尼迪等人：传达精确的信息和有价值的数据以及数据集的视觉表现

有些给出的定义则强调功能。如

- 斯图尔特·K·卡特等人：以电脑辅助、交互式、视觉再现形式表现抽象数据，以提高认知
- 伊莎贝尔·莱斯在《信息设计》：记录信息、传达意义、便于发现和支持感性推理
- 安迪·柯克：便于理解的数据的呈现和再现
- 雅克·贝尔坦在《图形符号学》：信息图形是表现图像和认知关系的认知艺术品，图形表示法是人类创造的基本符号系统之一，目的是为了存储、理解和传达重要信息。作为视觉语言，图像具有视觉认知的普遍特点
- 爱德华·R·塔夫特提出了一种规范定义：将数据可视化定义为具有几种功能的信息图形：显示数据；避免曲解数据试图讲述的故事：从广义的概述到细分结构上分析不同层次的数据细节

信息设计：用于传达仅用数据无法表现的信息的一般设计形式

新体图表：具有说服力的可视化数据的表现形式

不同定义之间的共同点：叙述的中心性和重要性

从广义上讲，这种中心性和重要性提供了集中解释什么是或不是数据可视化的新框架，且与所采用的语言无关。当设计师创作一个可视化作品时，他们会创作一个有关数据背后的故事，而那个故事成为了设计师向用户进行解读的焦点

因此，我们可以将数据可视化理解成有两个互补含义的术语：

- 第一是指将数据转化成视觉表现的创作过程；
- 第二是指在创作过程中产生的作品，包括图形、图表和其他在转化成视觉媒介过程中产生的表现形式。

数据可视化可能存在于我们身边的任何地方，有时候在我们不能立即辨认出或注意到的地方。历史已经证明，关于什么能够或不能够被可视化并没有限制，而该领域的实践者们也在不断地发掘数据可视化设计的新方法。**唯一的限制是人类的想象力，这是信息设计领域本身面临挑战的地方。**可视化以这种方式接触我们的社会、文化、政治和生活，影响我们如何微妙地处理信息。

设计的主要原则

如何确定一个可视化作品是否设计得成功也同样具有很大争议。尽管许多设计师对于有效性或质量有不同的标准，然而，今天却有一些普遍化的原则能够规范大多数设计思想。

最基本和最经典的原则可能来自信息设计师爱德华·R.塔夫特，他因为其设计原则经常被引用而众所周知。在《量化信息的视觉表现》中，塔夫特概述了几种构成可视化优势和图像整体性基础的想法。还有他提出的**数据墨水比例**观点也经常被引用，其定义为一个**可视化中用于表现数据的墨水在用于打印图表的总墨水中的占比**。塔夫特因提出“最重要的是显示数据”的说法而为人所知，这个目的可以通过完成两件简单的任务达成：**在合理的基础上，使数据墨水比例最大化并去除非数据墨水和多余的数据墨水**。

除了塔夫特的作品外，优秀的可视化设计原则分布在整个设计领域的不同方面。比如，

在可视化机械学方面，设计师们通常极其依赖感知组织的格式塔理论，以指导他们在可视化设计中做出具体的选择。这些原则描述了我们在视觉空间中对物体进行分组所遵循的基本规律，包括就近、相似、连续和图形 / 背景。通过这种方式，艺术和设计的基本原则以更加具体的方式表现在可视化作品中。

同时，认知科学的原则也应用在可视化设计中，包括科学家对某些视觉特征是如何以不同的方式在大脑中进行处理的。比如，

- 科林·韦尔提到了“前注意机制”，如颜色、大小和长度，这些都是能够在大脑中快速处理的视觉线索的特征。他还指出，根据这些特征以及它们在视觉工作记忆中的呈现形式而做出的设计选择能够极大地提高一个优秀可视化作品的质量。

然而，在考虑什么元素能够构成一个优秀的可视化作品时，其他框架则充分利用了结构和实际效果。比如，

- 邱南森在《可视化：设计、可视化和统计的流动数据指南》中指出了优秀的可视化设计的几个关键点：鼓励读者解释编码；标记轴线；采用大小和比例合适的几何图形（包含资源）；并考虑受众。这样的框架结合了实际用途和美学设计来评估一个可视化的有效性。

由于数据可视化的多种定义，设计师们在评估一个数据可视化的质量时所采用的多种方法也各不相同，但同时也显示了其中的共同点：

- 第一个共同点是，优秀的数据可视化应该讲述一个故事。仅仅陈述一个数据集的事实是不够的，相反，这些事实应该围绕一个连贯统一的故事进行叙述。
- 可视化应该精确。围绕数据建立一个故事时，非常重要的一点是，这个故事应是对它所基于的数据的精确描述。
- 一个可视化还应该直观，考虑受众，一个优秀的可视化设计不应要求受众花费大量精力来理解，应提供能在其他背景中应用或推算的新信息或知识。
- 最后，一个优秀的可视化设计必须优雅地完成解释数据、探索数据和发现知识的核心功能。

可视化及其客观性

可视化普遍存在于我们的世界中，它们的普遍应用具有几种不同的**核心功能**。

- 其中第一个，也可能是最基本的一个，就是解释数据，或以视觉方式传达信息。可视化最常见的功能是提供一种表达媒介，通过这种媒介可以表现和陈述信息，同时在该过程中吸引我们注意特定的问题或想法。
- 此外，可视化还是一种探索数据的媒介。作为一种视觉分析形式，可视化使得剖析和处理数据集以及与其进行互动成为了可能，而通过其他非视觉媒体可能难以完成这些工作。
- 最后，可视化能帮助用户发现知识。通过可视化，用户可能发现从其他表现形式中未能发现的新观点或知识。

数据可视化的前两个功能在我们的日常生活中以可预测的方式发挥作用。我们期待一个首先能够向我们解释数据的可视化。可视化的第二个核心功能，即探索数据，具有同样的特点。事实上，我们同样习惯于将可视化当作对话和论述去理解和研究，而非静态产品。一个成功的可视化有助于我们以用其他方式无法做到的方式接触、移动或研究数据。

然而，可视化发挥第三大功能——将信息转化成知识——却并非总是那么容易。正如海伦·肯尼迪等人证明的那样，可视化和数据作为实践和学科本身就是复杂的，它们挑战我们概念中的权威和知识，还以我们在十年前甚至都无法想象的方式进行实际展示。在数据不可避免地受限于我们的现代知识概念及其合理性的时候，可视化带来了一个全新时代的表现方法，同时提出了数据本身之于我们对事实的要求具有何种意义的问题。虽然我们将数据视为客观的和完整的，数据在很多情况下却只是如同表现它们的可视化一样，都是自身的展示。在追求客观性的过程中，可视化有时候能给我们这样的印象，即如果以正确的方式探索或设计，那么知识就是绝对的。

一个可视化设计如何完美地完成最后一个功能非常重要，既标志着它作为最终交流媒介的成功，也意味着可视化在很大程度上取决于其作为一门学科对知识表达的理论意义。因此，表现功能性与如何构建知识之间的关系在任何信息设计的学习过程中都不能忽略，而且对可视化的设计工作提出了特殊的跨专业挑战。许多不同领域或方法能告诉我们与可视化建立联系的不同方式，而全面认知可视化及其作用则要求将它们联系起来进行全面考虑。

设计可视化作品几乎无法成为一个中立行为，它的解读也不是一种没有偏见或暗示的行为，这使得用一种中立的方式通过可视化设计去发现新知识变成了一个艰难的过程。由于设计过程包括一系列有关表现形式的选择，而这些选择往往是由一个人或一群人做出的，因此，总是不可避免地出现一些带有偏见的元素，与个人的理想状态无关。保留可视化的知识发现能力要求设计师认识到偏见是如何通过具体的数据选择表现出来的，还要求他们**具有补救的能力**。为了成功地达到该要求，设计师们努力利用可视化来实现“用数据做好事”的宗旨，也就是说，将可视化作为一个中立的传播媒介来利用。

这种补救方法在设计中具体指什么呢？一种努力保持中立性的方法是，思考设计选择与知识的不同概念和知识是如何创造的有什么关联。可视化从业者在设计可视化时通常面临一系列问题，包括：

- 数据集具有什么内在的模式？
- 从这些模式中可以得出哪些结论？
- 如何利用可视化或不用可视化验证这些结论？
- 这些结论含有哪些知识？
- 读者如何根据视觉观察得出结论？
- 读者如何将结论转化成知识？
- 某些特定的设计选择是如何为将数据的内在模式转变成忠实于数据本身的知识提供平台的？

这些问题从两个广义的方面提出了挑战。首先，每个问题无疑都缺少一个清晰、唯一的答案，且它不会随着应用环境的不同而变换。事实上，任何一个问题的提出都可能受到明确和暗示因素的影响，这些因素包括客户的要求和需要，也包括设计师的思想倾向。其次，每个问题本身就带有提出问题之人的偏见，无论是设计师还是用户。另外，设计师和用户如何确定知识的特征对每种特征与某种具体的视觉表现形式有何种关联具有重要影响，尤其是当这些特征具有不可调和的矛盾的时候。

慢慢出现的众多不同的框架为我们对数据及其可视化和研究方式进行更具批判性的理解提供了方便。比如，这些框架之一是女性主义研究，支持所有知识都来源于社会的女性主义观点。凯瑟琳·D·伊格纳齐奥与麻省理工学院公民媒体中心合作进行研究时发现，女性主义研究具有一个独特的优势，那就是提供了一个平台，可用于研究表现在可视化中的权利结构，还用于表现不确定性和模糊性，数据背后的实体经济，以及支持不同意见的新方式。从这个观点看，知识是一个架构产品，而可视化变成了一个构建空间。在这个空间内，设计师和用户对表达方式进行商议。在这样的框架里，可视化被认为是一种客观（或主观）交流的媒介以及其作为一个学科对我们以新方式理解信息、数据和知识提出挑战的重要方法。当然，随着信息设计领域的不断发展和成熟，更多的可视化从业者和用户都将学会如何通过一种更具批判性的视角来评估可视化蕴含的信息。

未来趋势

展望未来，可视化显然不仅将继续在学术理论方面，而且还将在文化、社会和政治领域带来持久的影响。同样的，数据可视化的未来依赖于数据的未来以及我们与数据的关系，就如同它对可视化和设计本身的依赖一样。随着诸如哈佛大学的克莱因互联网与社会中心以及麻省理工学院公民媒体中心的出现和发展，批判探索的新方向将继续影响整个数据可视化领域的发展。这在社会和政治领域尤其明显，因为数据可视化扮演着一个日渐重要的角色，不断改变我们研究数据的方法，而这些数据与影响我们的社

会生活的现象息息相关。克莱因中心和公民媒体中心的研究均推进了从这些方向对数据的理解，提供了一个有用的框架，方便我们理解数据及可视化在未来对我们的生活可能产生影响的方式。

一个尤为重要的表现是数据可视化开始在生活和社会公正方面中扮演重要角色。社会的方方面面都有大量可用的数据，这使得社会公正的提倡者了解个人、政治团体、社会机构和政府的责任成为了可能，促进了提倡信息公开的新运动并引发期待。数据可视化领域的从业者也开始响应这些运动。

可视化在公共事务方面可见性的增加所产生的一个有趣的副作用就是社会信息素养和跨媒体信息素养的逐渐提高。数据世界的另一个重要趋势是自我量化，特别是在物联网背景下。可视化在这个领域也有可能扮演一个更加重要的角色。因为越来越多的设备能追踪与我们日常行为有关的数据，包括我们的身体活动、位置和生活习惯，人们愈加渴望寻找对这个过程中产生的大量数据进行可视化的方法。更广泛地说，大数据将继续向数据可视化提出新的挑战，需要采用新方法可视化大量数据，以解释可能掩埋在其中的模式。同样的，数据可视化的趋势将更多地受到数据趋势的影响，而不是受该领域从业者的工作的影响。

数据可视化具有挑战我们对世界以及我们在其中位置的认知的巨大潜力。设计可视化具有多模式和跨学科的特性，它可能改变我们对学科、方法和认知方式之间的已知差别的认识。但是，该领域在这个方面所面临的主要挑战是有效地利用作为具有改变能力的媒介的可视化。**数据可视化领域的发展一直以来都是由设计师们的智慧和创意促进的，促进这种发展的机会将继续出现前所未有的、令人激动的新的方向。**