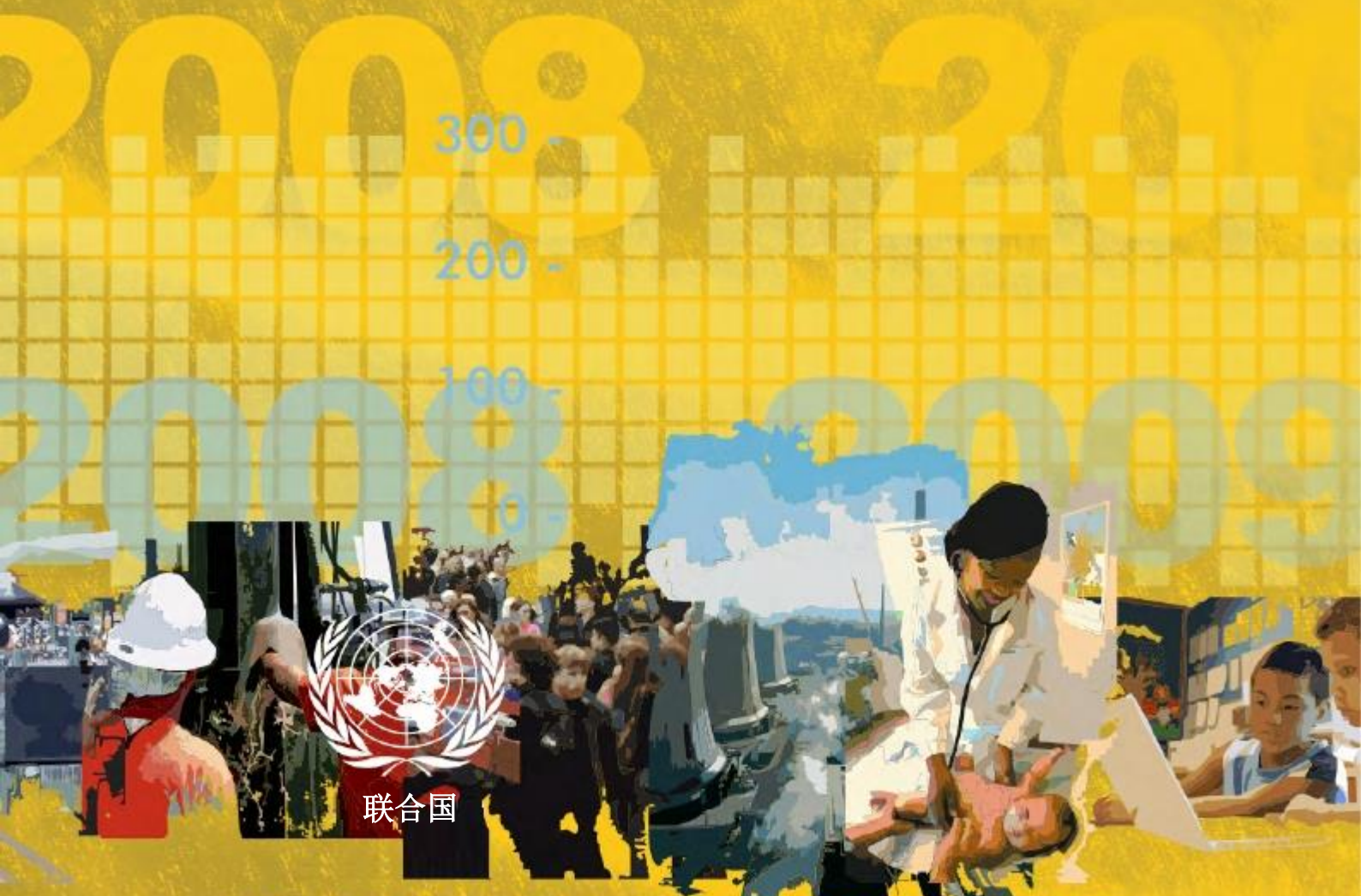


联合国欧洲经济委员会

让数据有意义

第二部分：统计数据展示指南



联合国

联合国欧洲经济委员会

让数据有意义

第二部分：统计数据展示指南



联合国
日内瓦，2009年

说明

本出版物中所使用的名称和展示的材料不代表联合国秘书处针对任何国家、领土、城市或地区、或其主权、边境或疆界划定等的法律地位持任何意见。

中文版由中国国家统计局社会科技和文化产业统计司、联合国儿童基金会驻中国办事处、联合国人口基金驻华代表处于 2015 年通过三方共同合作的数据项目组织翻译。

目录

简介.....	1
1. 传递信息.....	2
2. 统计数据可视化.....	7
3. 表格.....	12
4. 图表.....	17
5. 地图.....	31
6. 新兴的可视化技术.....	42
7. 易用性问题.....	46
8. 参考文献与延伸阅读.....	51

简介

《让数据有意义》是根据欧洲统计学家大会¹工作方案，在联合国欧洲经济委员会（UNECE）统计数据传播与发布工作会议²框架内编写的。

本指南作为一个实用工具，主要面向统计机构，特别是那些正在完善其传播策略的统计机构的管理人员、统计人员和媒体公关人员。本指南提供了关于文字、表格、图表、地图，以及其他工具的使用建议，以使统计数据走进非统计专业人员的生活中。本指南包含有建议、指导性意见和实例，但并非金科玉律。

有效的数据发布通过结合使用文本、表格和图表，尽可能有力地传递各种类型的信息。《让数据有意义第一部分：统计故事写作指南》（2006年发布）强调写作技巧的有效使用。《让数据有意义第二部分：统计数据展示指南》旨在帮助读者从诸多传播方法中找到最合适的一套工具和技巧，以最佳的方式将资讯传递给非专业人士。

本指南承认各统计机构之间存在诸多实际做法和文化上的差异，以及各国之间的方法差异。

本指南由统计传播与发布专家组编写。参加编写的人员有（排名不分先后）：

- Petteri Baer，联合国欧洲经济委员会
- Colleen Blessing，美国能源情报署
- Eileen Capponi，经济合作与发展组织
- Jérôme Cukier，经济合作与发展组织
- Kerrie Duff，澳大利亚统计局
- John Flanders，加拿大统计局
- Colleen Flannery，美国人口普查局
- Jessica Gardner，联合国欧洲经济委员会
- Martine Grenier，加拿大统计局
- Armin Grossenbacher，瑞士联邦统计局
- David Marder，英国国家统计局
- Kenneth Meyer，美国人口普查局
- Terri Mitton，经济合作与发展组织
- Eric St. John，加拿大统计局
- Thomas Schulz，瑞士联邦统计局
- Anne-Christine Wanders，联合国欧洲经济委员会

非常感谢 Martin Lachance（加拿大统计局），Andrew Mair（澳大利亚统计局），Alan Smith（英国国家统计局）和 Steven Vale（UNECE）为本指南的撰写所做的贡献。

¹ 有关欧洲统计学家大会的信息，参见 UNECE 网站：<http://www.unece.org/stats/archive/act.00.e.htm>。

² 有关联合国欧洲经济委员会统计传播与发布工作会议的信息，参见 UNECE 网站：<http://www.unece.org/stats/archive/04.05.e.htm>。

1. 传递信息

1.1 书面文字

新闻稿往往是统计机构藉以向目标受众（通常是社会公众）传播主要统计发现和分析结果的工具。文字是用于解读发现、概述趋势和提供背景信息的主要工具。

本章中，我们将针对准备“有效的”新闻稿或其他文件（诸如报告或分析性文章等）提供一些建议。

怎样才能使新闻稿、报告或分析性文章有效呢？也许最好的解释来自于《让数据有意义第一部分：统计故事写作指南》，其中初步提供了一整套促进信息传递的建议。有效的新闻稿具有以下特点：

- 讲述关于数据的故事；
- 与公众息息相关，并且回答了“为什么读者愿意读下去？”这一问题；
- 用标题或图片迅速吸引读者；
- 易于理解、有意思，往往具有趣味性；
- 鼓励包括媒体在内的其他机构以恰当的方式使用统计数据，为其所传播的信息增加影响力。

下面有一些小建议来帮助你开始进行文字写作。

1.2 目标受众定位：你的第一个决定

你必须做出的第一个重要决定是要锁定受众：你是为谁写作？简单地说，受众就是掌握方向盘的人。总的来说，受众想要什么，你就应当给他什么。你必须听从他们的意见，发现并选择恰当的叙述、语言、视觉和图形工具，以吸引他们的注意。

现在由于互联网的存在，受众的定位更加复杂。大多数统计机构都有向社会公众进行信息传播的义务，这里的公众指的是非专业但都受到良好教育的外行人士。在纸质新闻稿的时代，主要目标受众可能是媒体，统计机构都依赖媒体来向公众传播主要的发现。

然而今天，统计机构已经通过其网站、电子邮件，以及基于互联网的其他信息发布方式，形成了相当数量的直接读者群体。这意味着他们要与大量受众同时进行沟通：公众、数据用户、银行家，金融分析师、大学教授、学生等，其中每类用户都有自己特定的数据需求。

传媒世界也在不断发展。成功的商业媒体都知道这一点，并持续（通常是实时）

监控哪些故事最受关注。然后，他们使用各种工具，诸如视频、更多照片或更多分析等，针对他们手中的资源来创造出更为丰富的内容，以鼓励与每位受众进行更多互动。

任何情况下，在投入宝贵的资源开展任何形式的传播(新的或已经建立的)之前，首要的是确定你的受众或利益相关者是谁，他们想从你这里得到什么，以及他们希望怎么来获取。

如果是同时针对几类受众，必须选择最合适的方法将信息传递给他们每一位，即通过合适的渠道来进行信息传送，并且使用合适的传播技巧。

不过，往往由于时间和资源的不足，无法在有的时候能将信息传递给所有受众。你有一个选择：可以进行优先排序，或者如果想实现受众面最大，则应设法找到不同群体需求最清晰的共同点。

这是许多统计机构做的事。他们的受众是社会公众，但是通过把记者作为一种“管道”，齐心协力把信息传递给受众。亦即，目标受众是大众，但是记者是与这些受众进行沟通的渠道。专家和专业人士也能从这种做法中受益。通常，那些适用于大众的简单明了的技术是最受欢迎的，甚至也受到最专业的受众的欢迎。

1.3 了解你进行信息传播的背景

统计信息传播不是孤立的。因此，了解你进行信息传播的背景非常重要。受众消费媒体的方式是不断变化的。代际之间在技术能力和统计数据的理解方面也存在明显差异。

在规划统计信息传播时，应当牢记网络媒体消费的四点特别趋势，这既是机遇，也是挑战。

1. 万维网正日益成为一种娱乐性媒介。如果不用有趣的方式来呈现信息，就可能无法引起年轻用户的兴趣。
2. 社会上关于信息的消费已经发展成为一种“快餐文化”。用户越来越希望信息消费短小、快捷。
3. 互联网用户易于“满足”：他们找到勉强相关的信息之后就会停下来，而不是去继续寻找更为相关的信息。
4. 针对不同的受众和展现风格，在使统计信息传播更为有趣或更易消费的过程中，尽量不要丢掉重要的受众。

因此，为最大限度利用好互联网，你该做些什么？您必须使用最恰当的语言、结构和展现工具来传播你的信息。下面的章节将解释该如何去做。

1.4 叙事：讲故事

首要的任务是发现故事

要想数据对于一般受众具有意义，找到数字所蕴含的意义非常重要。统计界和科学界人士对“故事”一词往往比较警惕，因为这个词本身带有虚构或修饰的色彩，可能会导致数据的误读。如果数据分析人员不能够认真谨慎地对待数据，这种观点可能还真有一定道理。

不过，换一种做法，也就是完全不使用故事，情况可能会更糟。人们通常会因为自己无法理解数据而不信任统计数据，感觉它们是在误导。出现这种情况，是因为我们作为生产数据的一方，没有让人们感觉到数据与他们有关系，也没能用人们能理解的方式去解读。新闻稿如果没有故事线，就变成了纯粹的数字描述。

统计故事必须以对数据和所研究现象的足够了解为基础。否则只是有趣，其实都是错误的。在撰写统计故事时，还必须要谨记“官方统计的基本原则”³：

1. 公正无偏；
2. 专业精神；
3. 元数据；
4. 对错误解读发表评论；
5. 多种来源；
6. 保密性；
7. 透明度；
8. 全国协调一致；
9. 国际标准；
10. 国际合作。

统计机构要保持公正无偏，确保被访者和少数子群体信息的保密性，这一点至关重要。

行文时应在短期和长期趋势的背景下讨论最重要和最突出的发现，在循证基础上探索相互关系、成因和影响。应当向读者展示最新信息的重要意义。

以新闻风格来写作

使用记者的“倒金字塔”写作风格。首先呈现最重要的事实，然后按照重要性递减的顺序摆放其他支持性要点。读者的兴趣很短暂，因此最关键的信息必须放在文章的开头。

³ 这些原则于 1994 年经联合国统计委员会通过，其具体内容详见 UNECE 网站 <http://www.unece.org/stats/archive/docs.fp.e.htm>。

避免以方法开头，以结论结束。可以方法的要点小结以注释的形式提供给读者——对方法的解释越简单越好。结论应放在前面或作为开头段。

开头是文章最重要的部分，它应当讲述关于数据的故事，并联系故事所处的背景，简明、扼要、直白地小结故事线。应当专注于一条信息或主题，并包含最少量的数据。

开头段范例：

荷兰非金融类公司 2008 年第二季度的净利润达到 190 亿欧元，创三年来新低。利润额较 2007 年二季度下降 11%。净利润的下降有两个主要原因：一是利息成本提高，使公司需要支付更多净利息；二是海外子公司的利润下降。

资料来源：荷兰统计局

不要把读者淹没在一大堆数字里，正文中只摆放取整后的关键数字。不重要的数字应当放到附表里。文字用来给出分析、趋势和背景，而不要重复表中的数字。

注重结构

使文章结构化，以便每一部分都能发挥自身作用，同时还能辅助支撑要讲述的整个故事。子标题是强化文稿结构的有效工具，可以将文章分解成可管理和有意义的小节。

子标题简单总结每一小节的主要发现。包含动词的子标题会更具吸引力，更能被理解。

子标题范例：

“库存水平略有缓和”

“能源产品增长导致进口增加”

针对互联网传播方式，每一小节都应能发挥其自身的作用。这意味着应当规范术语拼写并标明来源。搜索引擎往往会指引用户直达网站内的深层链接，而不是通过网站首页或你所创建的其他页面引导用户访问目标地址。

信息还应当分出层次，以满足受众的不同信息需求。每一小节以主题句开始，主题句明确给出本小节的主要发现。后续段落中可以对发现进行详细阐述。

这种结构会提供一条清晰的路径，以方便关注细节的受众深入了解更多数据、分析和技术信息。此外，电子出版物还可以通过使用超链接的方式，为读者提供更为复杂的分析资源。

1.5 语言：做到简明扼要

简单的语言是所有成功沟通的核心。有句老话叫“简短至上”（KISS），这句话任何时候都不过时，不仅适用于文字，而且也适用于表格、视频和图形。既然有许多办法能做到通俗易懂，就不要把事情复杂化。

语言力求朴实。很多时候，我们拿到的材料是“官方”写作风格，读起来既低效，往往也不好懂或不清晰。

语言朴实并不意味着屈尊俯就、平淡琐碎或过度简化，也不意味着不讲语法。朴实的语言所传递的信息清晰简洁，将读者放在心中，并使用恰当的口吻。

朴实的语言阅读起来更快，让信息传播得更频繁、更容易和更友好。

记住：简单明了的信息不等同于“弱智化”。

简明写作的几个要点：

- 使用短句；
- 每句话只包含一个意思；
- 拆分长句；
- 每段以最重要信息开头；
- 段落简短；
- 文字干脆。

动词避免使用被动语态，要使用主动语态。动词被动式容易造成混淆，使文章冗长、不直接。

不好的例子：

“失业率上升是由经济危机所导致的。”

好的例子：

“经济危机导致失业率上升。”

尽可能避免术语。术语是对特定人群具有特定含义的语言，基本上不适合更大范围的受众。要使用适于信息传递和读者理解的最简单的语言。不过，如果受众完全由特定领域的专家组成，则术语可能是最适合使用的语言。

1.6 评估影响：媒体分析

受众对你的数据是否做出了准确而睿智的解读？他们能否通过分析得出关于经

济、环境和社会的合理判断并制定相应对策？

除非通过分析，否则你不会清楚受众使用你发布的数据做了什么。所以，统计机构对媒体和其他受众如何使用其信息进行监测是一个不错的想法。有如下几种方法：

- 建立客户反馈信息收集机制，诸如“客户之声计划”；
- 通过由关键利益相关者组成的用户组和焦点小组，评估公众的反应；
- 进行用户调查，或“一对一”地咨询重要客户；
- 如需深入了解敏感性问题或难以接触到的受众，可购买市场研究服务；
- 监测网络访问量，借以确定哪些数据和故事是受众最需要的；
- 分析关键字搜索模式和术语，找到简单、有力的技巧，帮助理解受众究竟需要哪些信息。

可直接与媒体合作，确保信息报道的准确、公正。媒体是向广大受众传递统计信息的最简单、最经济和最有效的渠道。

不幸的是，一些媒体有自己的动机，可能导致对数据的误用、误读和误解。因此，对信息通过媒体的传达情况、以及报道的语气和覆盖的质量进行监测至关重要。如果确定数据被误用，你应当迅速、一贯和坚决地做出反应。

现在我们已经讲解了与文字有关的关键要素，下面让我们把重点转到可视化部分，帮助我们更好地解释数据。

2. 统计数据可视化

2.1 为何一图胜千言

我们都听过一句老话：“一图胜千言”。理解数据的最佳技巧之一就是以图形的形式来展示数字。这样，就能更容易看清那些可能被隐藏的数据模式。

数据可视化方式有很多种，从简单的柱状图，到更为复杂的散点图、专题地图和动态的人口金字塔。这方面的技术资源非常丰富，包括数据可视化的书籍，大量相关的主题网站，以及大量为此目的而开发的可下载的软件和程序。

本章将简要回顾数据可视化的历史，并就如何制作好的数据可视化产品提供指导原则。

2.2 可视化是统计业务流程的组成部分

数据的有效展示应当成为统计生产过程的组成部分。而数据可视化作为展示统计趋势和关系的重要方面，是必须坚持去做的工作，不能把它当成可有可无的事情。在 UNECE/Eurostat/OECD 统计元数据指导小组正在开发的通用统计业务流程模型⁴中，可视化属于“发布”阶段的内容。

读者可以很容易地理解可视化展示。通过报纸、电视、互联网和书本，图表和地图对几乎每个人都产生影响。利用图表或地图而非一长串数字来展现统计数据更容易被理解——当然，这是在假定可视化展示使用无误的情况下。

这种展示应能快速而简单地说明趋势和关系，这是一种从数据库中提取信息并输入读者大脑的有效方式。

但是要注意，糟糕的统计信息可视化是会产生误导的。不管是有意，还是多数情况下的无意，有许多方式会提供误导性信息。因此，设计性和功能性之间要保持平衡。复杂的可视化往往不利于沟通。由于解读图表本身要求很高，因此不要强迫读者从图表中“挖掘”信息。

不同的文化传统也可能导致误解和误读。例如，颜色在世界各地有不同的象征意义。

2.3 苏格兰人对数据可视化的历史性影响

根据我们现在的了解，统计图表的历史还不到 300 年。尽管古罗马人和阿拉伯人善于运用数字，但是他们并没有使用柱状图或曲线来直观地展现数字。

法国百科全书派学者和哲学家笛卡尔（René Descartes, 1596-1650）是第一个用直角坐标系来对其观察结果进行可视化的人。苏格兰工程师和经济学家威廉·普莱费尔（William Playfair, 1759-1823），是第一个以我们今天熟悉的方式对统计数据图形展示的人。

普莱费尔出版了两本关于 18、19 世纪之交（1786-1801）世界贸易发展的书，并在书中使用了他所谓的“图解法”。他热衷于推销这种新方法，并利用以下观点来宣传可视化展示的优势：

1. 图形展示有助于简化难以观察到的复杂关系：
“我们希望在这些图表的辅助下，可以不需要费力地研究细节特点，就能得到想要的信息。”
2. 企业主、政治家和决策者需要快速、直观地掌握统计信息中的精华，因为他们没有时间研究细节：

⁴ <http://www1.unece.org/stat/platform/display/metis/The+Generic+Statistical+Business+Process+Model>

“地位高的人或生意人只关注事情的要点[...]。”

3. 表格，尤其是大量表格，会比较枯燥，难以理解。图形展示能帮助用户：
“[...]在五分钟的时间里获得尽可能多的信息，并给读者留下深刻的印象，而不需要花一整天的时间去研究数字表格来达到同样的效果。”
4. 图形吸引眼球。普莱费尔强调了他的方法在理解关系和数量方面的有效性。
5. 图形方法吸引和挑战大脑，不仅有助于大脑感知，而且有助于理解结构和关系。

虽然自普莱费尔时代以来可视化的方法已经发生了革命性变化，但是仍有改进和发展的空间（参见第 6 章）。不过，重要的是要避免过于草率的展示和扭曲内容的展示。记住，**技术只是工具**。不要因为你有技术能力，就添加无用的注释和晦涩的元素。要让信息简单示人。

推动图形展示和人类感知发展的新生代大师有：Jacques Bertin, William Cleveland, Jan-Erik Kristiansen, Vesa Kuusela, Hans Rosling, Edward Tufte和Howard Wainer。尽管这些现代专家们做了很多工作，但是统计图形的基本原理仍然并将继续建立在威廉·普莱费尔的 innovation 之上。

2.4 关于人类感知的基本事实

我们能快速而轻松地进行视觉观察，是基于大脑感知规律性和无规律性的能力。这些能力大部分是在不知不觉中工作的。几乎就在我们开始思考之前，对事物的比较就完成了。

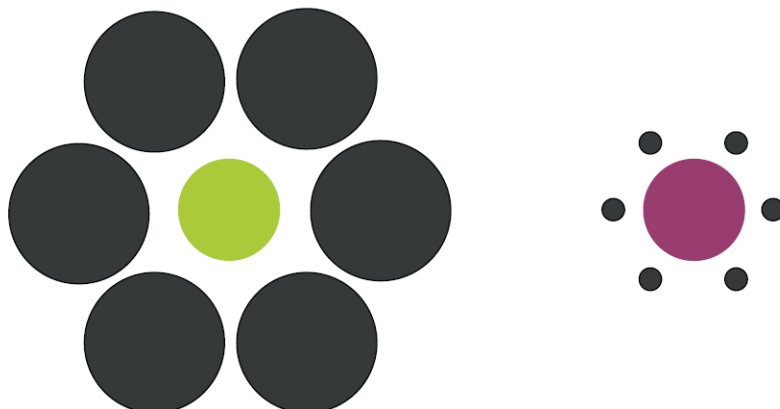
请看下图，中间的横条哪一端颜色较暗？



资料来源：Helmholtz, H. (1821-1894), “同时对比的错觉”，见“维基百科，免费的百科全书”，
http://en.wikipedia.org/wiki/Optical_illusion，检索时间：2009年7月。

横条两端的色度其实是相同的，但是由于背景的色差而使人产生错觉。拿张纸盖住背景，你会很容易发现这一点。

现在，看看下面这两幅图中哪一个圆更大。是左边图中央的圆大，还是右边图中央的圆大？



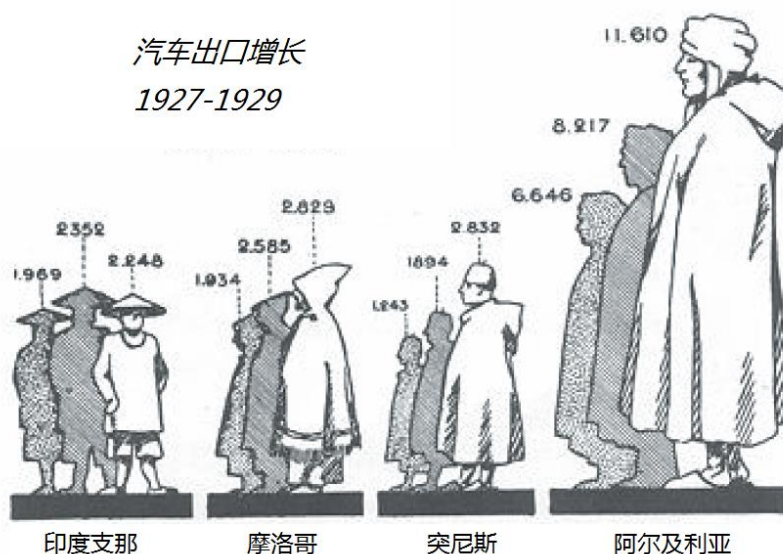
资料来源：Ebbinghaus, H. (1850-1909), “艾宾浩斯错觉”, 见“维基百科, 免费的百科全书”, http://en.wikipedia.org/wiki/Hermann_Ebbinghaus, 检索时间: 2009年7月。

左右两幅图中央的两个圆大小相同。如果不信，可以测量一下！

这两个例子说明人类大脑如何自动依据周围环境来观察物体。正是由于与周围事物的对比，导致你得出关于横条明暗和两个圆形大小的错误结论。消除环境影响后，更能获得正确的结论。

该信息对统计人员的启示是：在生成统计信息的视觉展示产品时要务必小心。展示的背景可能会扭曲用户的感知。

让我们看一下第三个例子。下图的目的是展示1927-1929年印度支那、摩洛哥、突尼斯和阿尔及利亚的汽车出口增长信息。图中人物的大小如何能很好地反映数据的真实变化？



资料来源：Satet, R. (1932), *Les Graphiques*, 巴黎。引自Tufte (2001)。

在观察图片时，我们的大脑会比较每一个物体的相对大小。该图中，1927-1929年间汽车出口数据由四幅人物图绘的高度和体积来体现。图中相对数值可能会被误读，体现在以下两个方面：

1. 比例不准确。图中阿尔及利亚人物的身高应当高出很多。数值大4倍，反映在人物的身高和大小上也应该大4倍。
2. 将近期的人像重叠到早期的人像上，更难计算年份之间的数值差异。从数值上看，前面人像对应的数值比后面的大很多，应该用他们的相对身高才能显示不同年份之间的变化。

2.5 感知也基于经验

经验对如何感知图形也有影响。因此要了解你的受众以及他们的能力、经验和可能产生的差别。无论是关于统计数据还是关于背景知识，不要认为你知道的他们都知道。

统计人员必须认识到，自己在识别数字模式方面有一定经验，而他们的读者则可能缺乏相关经验。作为分析统计数据的专业人士，统计人员可能比一般人更能发现潜在信息。由于广大受众需要更多帮助来认识统计信息，因此统计人员需要利用可视化方法来解释其发现。数据的视觉展示应当使主要发现更容易被观察和理解。

2.6 不要滥用工具

有众多的技术工具可以帮助实现数字信息的可视化。要将重点放在图表的内容和信息上，而不是放在图形展现的方法、设计或技术上。

基本的电子表格软件中有超过70个标准图表类型可供选择，另外还有生成自定义图表的能力。由于选择余地很大，往往会引发用户使用多种不同的方式来对数据进行可视化。然而，结果往往是复杂过了头，或者扭曲了统计信息。

随着越来越多、越来越炫的的新工具的出现，上述风险也在增加。使用这些工具时应当牢记，技术只是实现结果的一种手段而非目的。统计数据中包含的信息远比展示你知道使用多少种炫酷的工具更为重要。

2.7 开发好的数据可视化产品的要点

在进行可视化展示时，应当考虑：

- **目标群体：**不同的受众（例如，企业界或学术界，专家或一般公众）可能需

要不同的展现形式。

- **整体演示中图形的作用：**分析整体情况和聚焦关键点，可能需要不同的视觉展示类型。
- **如何和通过何种形式来传递信息：**使用较长而详细的分析还是快速的幻灯放映。
- **可能引起曲解的应用情境：**是面对专业的数据用户还是新手。
- **文字分析和数据表格哪个解决方案更好。**
- **关于易用性的考量：**
 - 为图表和图像等非文本元素提供文本替代方案。
 - 不要仅仅依靠颜色。如果去掉颜色，受众是否还能理解演示内容？色彩组合是否有足够的对比度？颜色对色盲（红/绿）有影响吗？
 - 确保用户能够控制时间敏感的内容（例如，暂停动画图形的播放）。
- **数据可视化之间的一致性：**确保可视化元素设计的一致性，并且尽可能使用约定俗成的做法（例如，蓝色代表地图上的水）。
- **大小、时长和复杂性：**演示内容容易理解吗？是否因为某个部分内容太多而使受众一时无法领会？
- **误读的可能性：**在同事、朋友或目标群体中的某些人面前进行演示，看看他们是否得到了你想要传达的消息。

3. 表格

3.1 为什么表格很重要

无论是新闻稿，还是分析性文章或研究论文，好的表格都是其中的重要组成部分。使用表格有助于最大限度减少文字中的数据使用量，而且还无需讨论对故事情节不是很重要的非关键变量。

Miller（2004）在她关于数字写作的书中，针对如何设计好的表格给出以下指导原则：

- 让受众能够容易地找到和理解表格中的数据。
- 以直接和自然的方式设计表格的布局和标签，将注意力放在数据要传递的基本意思上，而不是放在表格的结构上。

本章将讲述如何制作有效的表格。

3.2 两种类型的表格

有两种类型的表格值得关注。第一种是小表，称为演示（或展示）表格，用于突出新闻稿、网页或分析性刊物中的关键数字。

第二种是大表，称为参考表格。它们渐渐被交互式数据库所取代，这些数据库允许用户在线生成自定义表格。由于参考表格更多是一种分析工具，因此不在这里进行讨论。

在演示表格中，数据的展示应简洁，组织有序，以支持相应的分析。简短且格式良好的表格可以提供大量信息供读者迅速吸收。

不管是通过报告、文章、出版物，还是通过网页发布，表格都应当能够独立存在。每个表格应当包含足够的元数据，诸如描述性标题和数据来源等，以使其能够被拷贝和粘贴到其他文档中并仍然具有意义。如果表格具备了独立性，则无论有没有原来的上下文背景，都更有可能得到正确解读。

3.3 好表格的设计要点

要描述表格中包含的数据，需要有以下五个支持元素：

- **表格标题**应当针对数据给出清晰而准确的描述。标题应当回答三个问题，即“何事”、“何地”和“何时”。标题应简明扼要，避免使用动词。

表格元素范例
“2007年加拿大分职业的计划退休年龄。”
正文中给出正确理解和使用数据所需的所有信息。

- **列标题**，位于表格上部，用于标识表格中每一列的数据和提供相关的元数据（例如，计量单位、时间段或地理区域）。
- **行标题**，位于表格第一列，用于标识表格中每一行的数据。
- **脚注**，位于表格底部，提供所需的任何补充信息（例如定义），以正确理解和使用数据。
- **数据来源**，位于表格底部，用于提供数据来源，即生产数据的机构和数据采集方法（例如人口普查或劳动力调查）。

下图给出了表格的各个元素都是如何显示的。

表标题	
行标题	列标题
	数据
脚注	
资料来源	

演示表格中所呈现的，应当仅是对数据经过挑选后，最适合传播信息的一个小的子集，如下面的例子所示。

演示表格范例

2008年6-7月加拿大各省和各地区的制造业销售额 经过季节调整

	2008年 6月 ^r	2008年 7月 ^p	2008年6-7月 百分比变化 ¹
	百万美元		
加拿大	52 685	54105	2.7
纽芬兰和拉布拉多	692	674	-2.5
爱德华王子岛	123	115	-6.1
新不伦瑞克	1914	1872	-2.2
魁北克省	13019	13280	2.0
安大略省	23902	25015	4.7
马尼托巴省	1360	1445	6.2
萨斯喀彻温省	1079	1108	2.8
阿尔伯塔	6298	6316	0.3
不列颠哥伦比亚省	3347	3306	-1.2
育空地区	3	4	45.5
西北地区和努纳武特地区	4	3	-27.4

^r 修订数

^p 初步数

¹ 百分比变化根据以“千加元”为量纲的数据计算得到

数据来源：加拿大统计局

数值的摆放应能方便关键信息的提取。用户逐列浏览还是逐行浏览更加容易，取决于所要传递的信息。当决定是纵向还是横向来展示表格时，要考虑到这一点。线条或细微阴影也有助于用户进行横向或纵向阅读。间距和阴影可以改变读表的方式。

列应该均匀分布，不要相距太远。表格宽度应恰好满足数据内容的需要。

为确保表格易于理解，应考虑以下几条原则：

- 避免不必要的文字。
- 显示数据时，要么按时间序列中的时间顺序，要么使用标准分类。针对较长的时间序列数据，有时使用逆向的时间排列更为合适（即先从最近的时间段开始，依次倒排），例如月度失业情况。
- 使用最少的小数位。
- 使用千位分隔符。使用空格代替符号可以避免不同语种间翻译时可能出现的问题。
- 对齐小数点（或在没有小数位的情况下向右对齐），以便能一目了然地看出数值的相对大小。不要让数据居中，除非它们具有相同的量级。
- 不要使任何数据单元为空。缺失值应标识为“无法获得”或“不适用”。由于这两种情况都适合用英文缩写“NA”代替，因此需要对其具体含义加以明确。

本文下一节将对上述部分原则进行讨论和说明。

当为出版物或网站制作系列表格时，所有表格应使用相同的布局。要考虑在表格标题中需要包括多少信息（哪些信息是显而易见的，哪些不是）。缩写要保持一致。

3.4 数字取整和小数

当数据中有三位或更多位数字不一样时，许多非统计专业的用户就很难看出数据值之间的差别。你可以通过对表格中的数据进行取整来帮助他们读懂数字。当数据不需要具有较高精度时也可以采用取整的做法。某些情况下，只有取整的数据是可靠的，因此在表格中显示时也应该取整。不过，应防止在数据取整时丢失太多信息。

好的例子	不好的例子
1 320 000	1324567
1 670 000	1673985
1 830 000	1829456

在上述例子中，左边取整后的数字比右边精确的数据更易于理解和记住。这个例子还显示出了如何使用空格作为千位分隔符。

如果需要显示具有不同小数位数的数值，应当先进行小数点对齐而非向右对齐。下面的例子中，左边的值比右边的值更容易阅读。这个例子也说明，对所有数值，显示相同的小数位数会更方便阅读。

好的例子	不好的例子
93.2	93.2
1045.0	1045
385.6	385.63

数值应当右对齐。请注意，使用同样的例子，当将数据调整为左对齐时，阅读数字的难度会增加，如下图所示。

好的例子	不好的例子
93.2	93.2
1045.0	1045.0
385.6	385.6

3.5 如何完善表格的示例

为了说明3.3节所述原则的有效性，下面我们提供一个不好的表格例子，并介绍如何轻松地对其进行改进。

不好的例子

分行业的最终能源消耗——百分比

	1980年	1985年	1990年	1995年	2000年	2002年	2003年
交通运输	27.81	27.92	28.24	31.12	36.82	39.48	39.13
住宅	31.11	33.91	30.41	27.61	24.33	23.71	23.97
工业	31.47	27.21	23.86	22.11	21.41	19.53	18.78
农业	n/a	n/a	3.51	3.7	3.11	2.91	2.82
服务业	9.61	10.96	13.98	15.46	14.33	14.37	15.3
合计	100	100	100	100	100	100	100

上面这个表有哪些不妥之处？

- 不知道数据指的是哪个地理区域。
- 没有说明数据来源。
- 数据居中而非右对齐。
- 数据不应当显示两位小数（信息太多）。
- 合计数应当与其他数具有相同的小数位。
- 没有对缩写“n/a”进行解释。
- 使用交替行灰色阴影以及格式相同的表格线对于理解表格中展示的不同数据并无帮助。
- 没必要将表扩展到整个页面的宽度。

好的例子

分行业的总能源消耗份额（%）

爱尔兰，1980-2003年

	1980年	1985年	1990年	1995年	2000年	2002年	2003年
交通运输	27.8	27.9	28.2	31.1	36.8	39.5	39.1
住宅	31.1	33.9	30.4	27.6	24.3	23.7	24.0
工业	31.5	27.2	23.9	22.1	21.4	19.5	18.8
农业	n/a ¹	n/a ¹	3.5	3.7	3.1	2.9	2.8
服务业	9.6	11.0	14.0	15.5	14.4	14.4	15.3
合计	100.0	100.0	100.0	100.0	100.0	100.0	100.0

¹ 农业部门能源消耗数据从1990年才开始采集。

数据来源：爱尔兰公共企业部

表格在哪些地方得到了改进？

- 在标题和副标题中提供了用于理解数据所需的所有信息。
- 指明了数据来源。
- 所有的数值都是右对齐，并统一保留一位小数。
- 缩写“n/a”在脚注中做了解释。
- 只显示用于区分表的不同部分（表头、数据、脚注和来源）的线条，去除无用的阴影。
- 表格的宽度根据所有标题和数据的显示需要而定。

4. 图表

4.1 为何使用图表？

相比表格，图表能够使统计数据变得更容易被理解。图表是统计数据的视觉展现，其中数据通过诸如条状或折线等符号来表示。图表能方便快捷地显示数据、便于比较和揭示数据内部的趋势和关系，是非常有效的可视化工具。

图表一般为一维或二维图形，如柱状图或折线图等。虽然也有三维图表，但是通常认为其过于复杂而不易于理解。

图表可用于展示大量数据所具有的模式，或者用于传递重要的发现或信息。如果想要展示以下信息，应当考虑使用图表：

- **比较：**有多少？哪个更大或更小？
- **随时间推移而发生的变化：**变量如何演变？
- **频率分布：**各项是如何分布的？有什么差别？
- **相关：**两个变量之间有关系吗？
- **整体中的相对份额：**子项与整体相比是怎么样的？

本章中我们将讨论最常见的图表类型，并就如何制作好的图表提供指导。

4.2 好图表的设计要点

如果认为图表最合适展现某些数据，则无论使用何种类型的图表，都需要牢记以下三条原则：

1. **明确目标受众：**他们对该问题有何了解？
2. **确定要传递的信息：**数据说明了什么？需要传达多个信息吗？
3. **确定信息的特点：**是要进行项目比较、显示时间趋势，还是分析数据关系？

好图应当做到：

- 抓住读者注意力；
- 简洁明了、准确无误地展现信息；
- 不误导读者；
- 以集中方式显示数据（例如一幅折线图，而不是多幅饼图）；
- 方便进行数据比较，突出趋势和差别；
- 所附文字说明了信息、主题或故事情节。

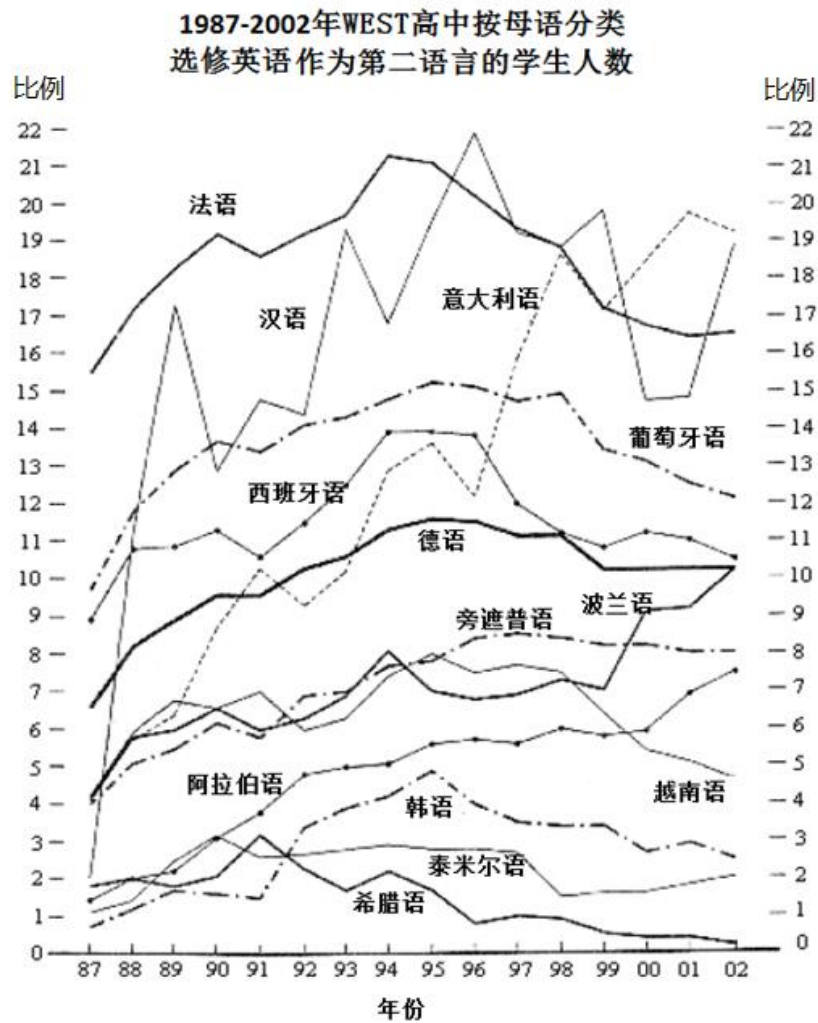
4.3 不适合使用图表展现数据时如何处理

图表并非总是展示统计信息的最佳工具。有时文本和（或）数据表格可以更好地向受众解读数据，并且能替你节省很多时间和精力。

出现下述几种情况时，应当重新考虑是否使用图表：

- 数据值很分散；
- 数据值太少；
- 数据值太多；
- 数据值变化很小或者没有变化。

不好的折线图例子



资料来源：加拿大统计局，《学习资源：使用图形》⁵。

应当避免任何类似上面的折线图。由于图中数据过于庞杂，分析人员想阐明的任何故事线都迷失在折线丛中。

4.4 选择合适的图表类型

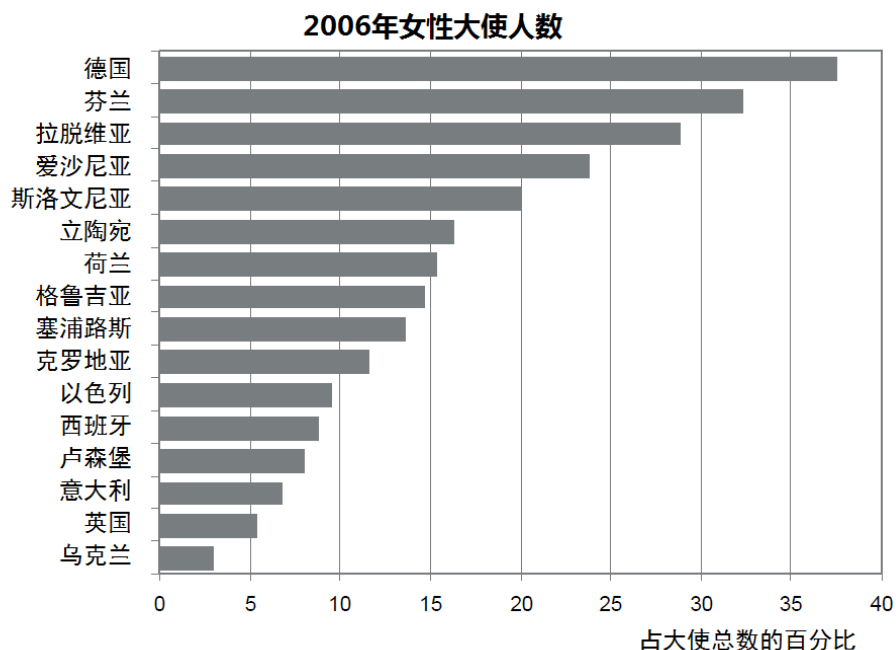
知道哪种图表类型适合哪类信息至关重要。数据的特点决定了有些图表比另一些图表更合适。本节中，我们将提供最常见图表类型的使用原则，包括柱状图和人口金字塔、折线图、饼图和散点图。

⁵ <http://www.statcan.gc.ca/edu/power-pouvoir/ch9/using-utilisation/5214829-eng.htm>

柱状图

柱状图是最简单的图表类型，方便制作和阅读。柱状图用于对不同类别或组别的频数或数值大小进行比较。

好的柱状图例子



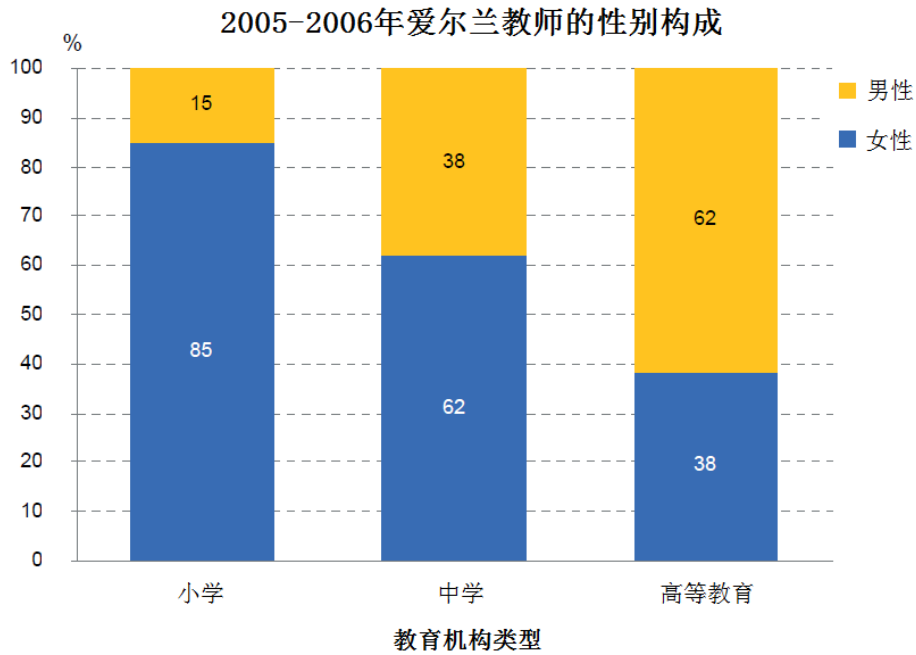
资料来源：UNECE统计数据库

柱状图的柱子可以是垂直或水平方向。当为水平方向时，文字更容易读认，如上例所示。当按照柱状图中各值大小由小到大依次排列时，更容易对不同的值进行比较，而不是杂乱无章地显示。

柱状图中柱子的宽度应当远大于它们之间的间隔。间隔不应超过柱宽的40%。

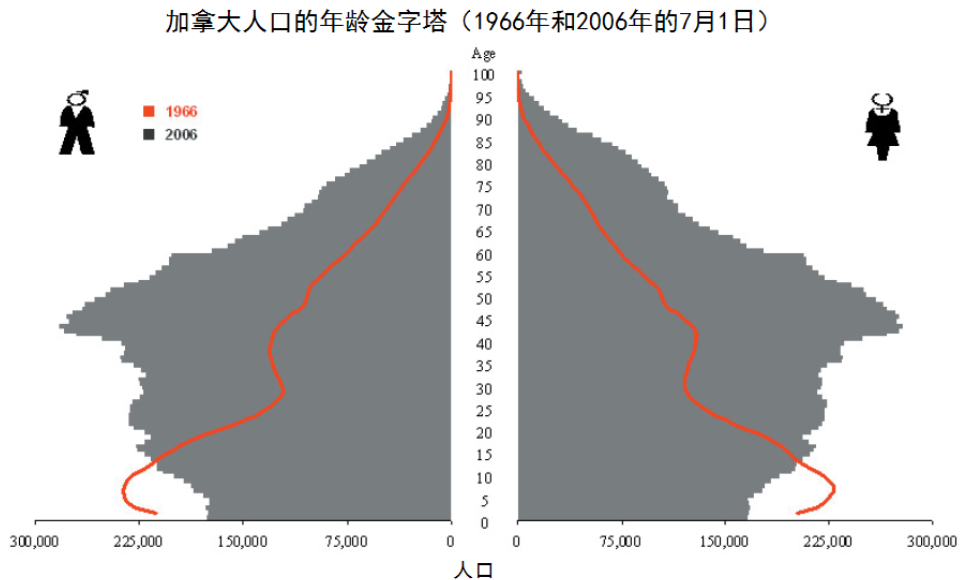
堆积柱状图可用于显示和比较总体中的不同部分。使用这种类型的图表时应当格外小心。如果每个堆积中有太多项，或许多项的大小相当接近的话，则难以利用该图进行分析和比较。

好的堆积柱状图例子



人口金字塔是两个横向柱状图的组合,代表着国家或地区的女性和男性人口的年龄结构。男性通常显示在左边,女性通常显示在右边。当比较不同人口金字塔时,通常最好是展现男性和女性人口在总人口中的比例,而不是数量。

好的人口金字塔例子



数据来源: 加拿大统计局⁶

大多数欧洲国家的人口金字塔已不再呈现金字塔的形状,但是人口金字塔仍然是

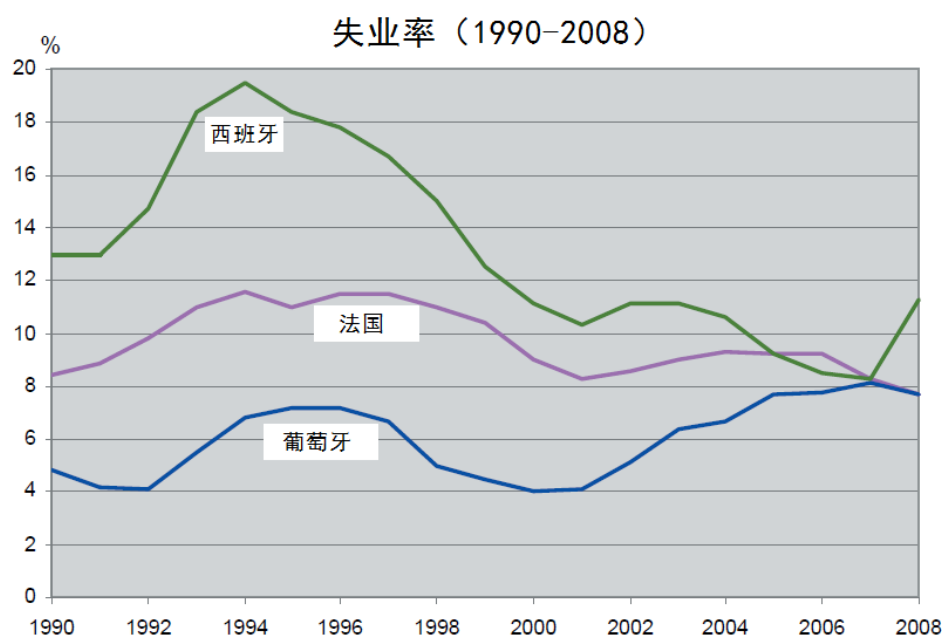
⁶ <http://www.statcan.gc.ca/daily-quotidien/061026/figure.htm>

非常有效的方法，能展示大量关于人口年龄和性别结构的信息。如果进行“动态”演示（即随时间的推移而变化），效果会更好。

折线图

折线图是对数据的时间趋势变化进行可视化展示的有效工具，因此是展示时间序列数据的最佳图表类型。可以通过调整图表参数来更好地传达信息，但是注意不要扭曲数据，第4.6节将针对该问题进行讨论和说明。

好的折线图例子



资料来源：UNECE统计数据库

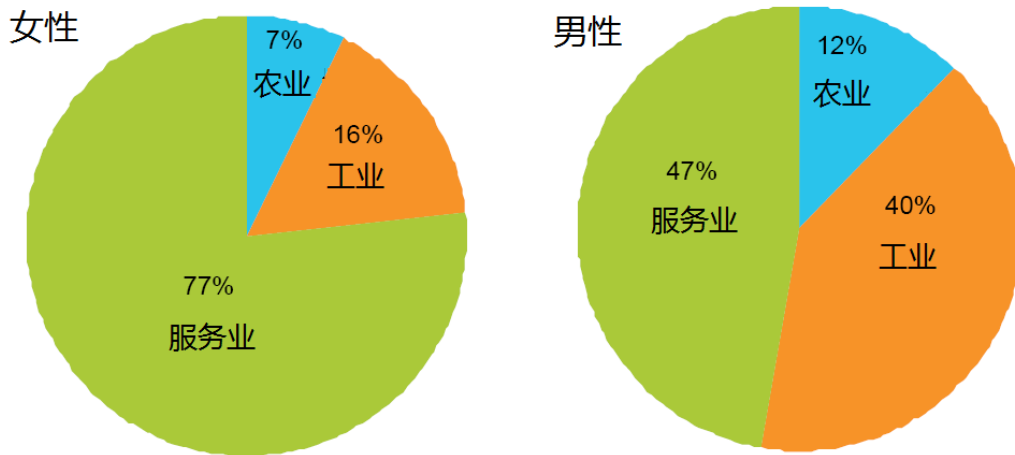
饼图

饼图可用于显示同一个变量各类别的百分比分布，但是分类个数不能太多，通常不超过6个类别。由于饼图难以对“饼”的不同扇区进行比较，更难以在不同饼图之间进行数据比较，因此许多统计人员不推荐使用这种图表类型。为克服这一问题，可以在饼图的各个扇区标上实际数值。有些情况下，还将类别名称作为标签写在图表上，这样就不需要使用图例。饼图的各个扇区最好按照从小到大的顺序排列，而不是大小穿插。

大多数情况下，使用其他图表类型（例如柱状图）更为合适。不过，由于饼图能有效展现某一类别在总体中的相对重要程度，因此也不应完全排除饼图的使用。饼图很适于提供情况概览，如下面的例子所示。

好的饼图例子

2007年拉脱维亚主要行业的就业情况



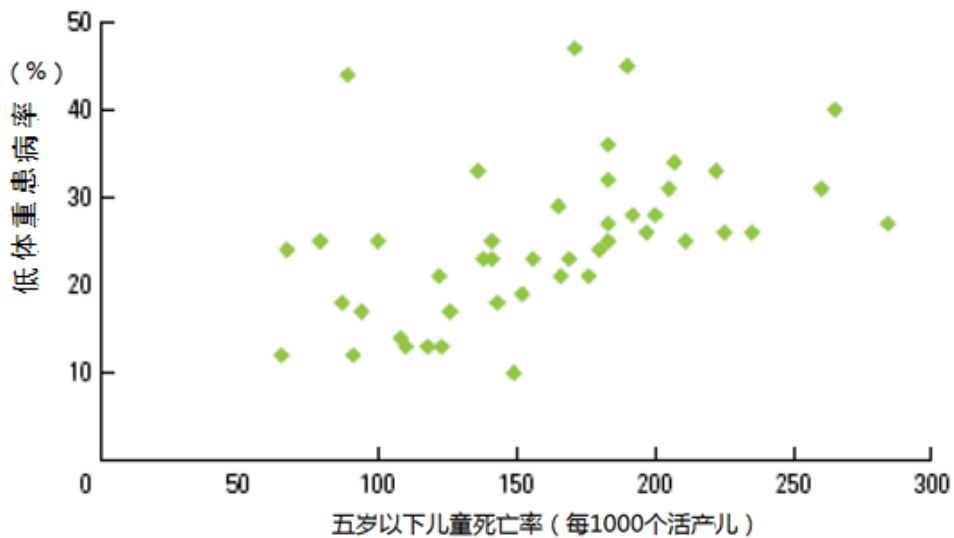
资料来源：UNECE统计数据库

散点图

散点图用于显示两个变量之间的关系，是显示相关性的最准确方法，如下面的例子所示。不过，由于散点图不易解读，有些分析人员更倾向于使用柱状图。

好的散点图例子

2003年撒哈拉以南非洲国家五岁以下儿童死亡率和低体重患病率



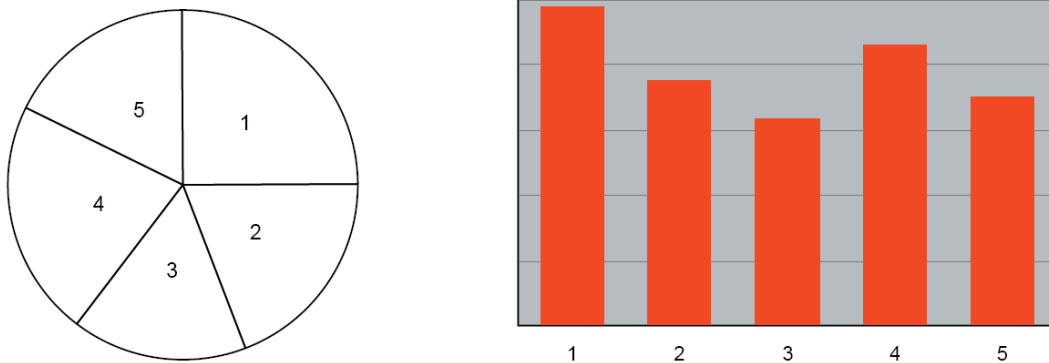
资料来源：Jamison et al. (2006)，《撒哈拉以南非洲国家的疾病和儿童死亡率（第二版）》，华盛顿特区，世界银行⁷。

⁷ <http://www.dcp2.org/file/66/Disease and Mortality in SSA.pdf>

尝试不同的图表类型

应当使用哪一种图表类型？你需要做出决定。好的做法是尝试不同的图表类型，并从中挑选出最合适的信息传递工具。

下面是对同一数据进行图形展示的两种不同方法。哪一个更清楚？



你能区分出饼图中哪个扇区最大吗？一些读者通常认为相对于柱形或折线而言，饼图角度比较的难度更大。饼图中扇区 1 和 4 看起来大小差不多，但是放到柱状图中，相对大小的差别立刻明显起来。

4.5 怎样使图表高效

图表组件

图表中的不同组件会竞相吸引读者的注意。图表中包含的特征越多，读者就越难以看到你的重点。

图表组件分为以下三类：

1. **数据组件**用于展现数据：条、线、面或点。
2. **辅助组件**用于辅助理解数据：标题、图例、数据标签、网格线、脚注和数据来源。
3. **与数据无关的装饰物**。

只有数据组件是不够的。为了确保读者对图表的正确理解，还需要包含下述辅助组件：

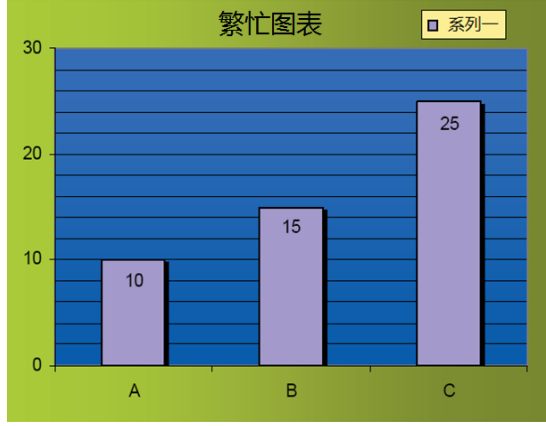
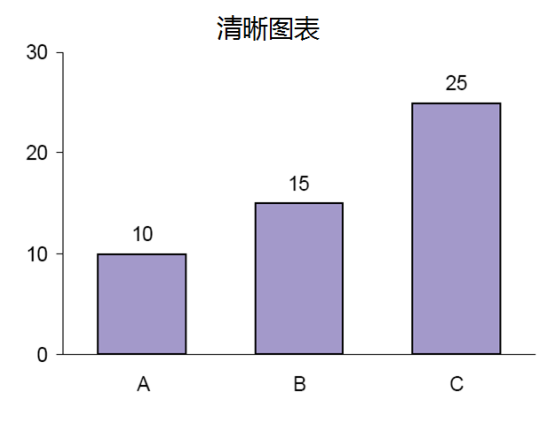
- **图表标题**，应该清晰地告诉读者图表是在描述什么。图表标题应当简明扼要，有两种标题类型供参考：
 - **信息类标题**，提供理解数据所需的所有信息。这类标题应能回答三个问题，即“何事”、“何地”和“何时”。

- **描述类标题**，突出图表中显示的主要数据模式或趋势，寥寥数词描绘出图表所要讲述的故事。
- **坐标轴标签**，用于标识图表中展示数值的内容。两个坐标轴上的标签都水平显示。
- **坐标轴标题**，用于标识数据的度量单位（例如，“千”、“%”、“年龄（岁）”或“\$”）。如果度量单位显而易见（例如，作为时间序列的“年”），则不需要在图表中包含坐标轴标题。
- **网格线**，可以添加到柱状图和折线图中，帮助用户读图和进行数值比较。
- **图例和数据标签**，用于标识图表中展示数据所用到的符号、图案或颜色。如果图表中只有一个数据系列，则不用显示图例。可能情况下，建议使用数据标签，而不是图例。数据标签显示在数据组件（条、面、线）的上方或一侧，以辅助进行标识和理解。
- **脚注**，可用于提供有关定义或方法的信息。
- **数据来源**，应置于图表底部。

一切以数据为中心

为最大化图表使用效果，应当以数据为中心。辅助组件应做到：

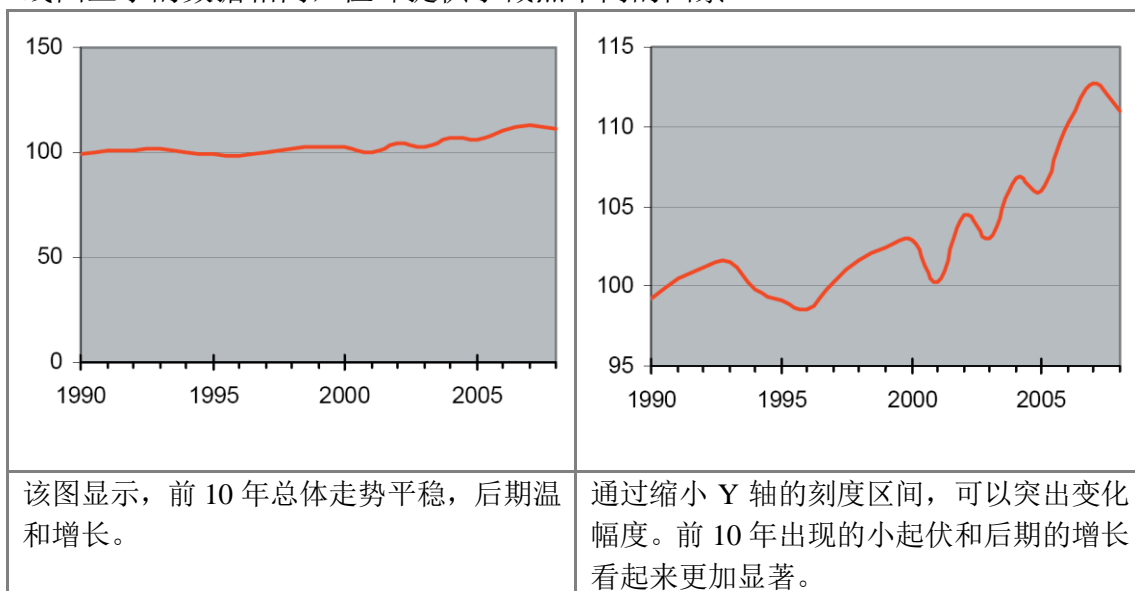
- **需要时才出现**。坐标轴标题、图例和数据标签可能对于正确理解图表至关重要，也可能根本就不需要，具体取决于数据的特点。
- **不要喧宾夺主**。坐标轴和网格线的颜色应当比数据组件的颜色浅。不要让装饰物分散了读者的注意力。

不好的例子	好的例子
	
<p>所有组件都被最大限度地突出，结果得到了一张非常繁忙的图表，尽管展示的只有三个数值，但却难以阅读。</p>	<p>阅读起来容易得多。最低限度使用辅助组件，确保了数据的中心地位。</p>

数据组件之间也会相互打架。要展示的变量和数值越多，就越难以把数据展示清楚。有效的图表应具有清晰和可视的信息。如果图表承载内容太多，就会变成需要费力去理解的难题，甚至会造成误导。

4.6 调整图表参数

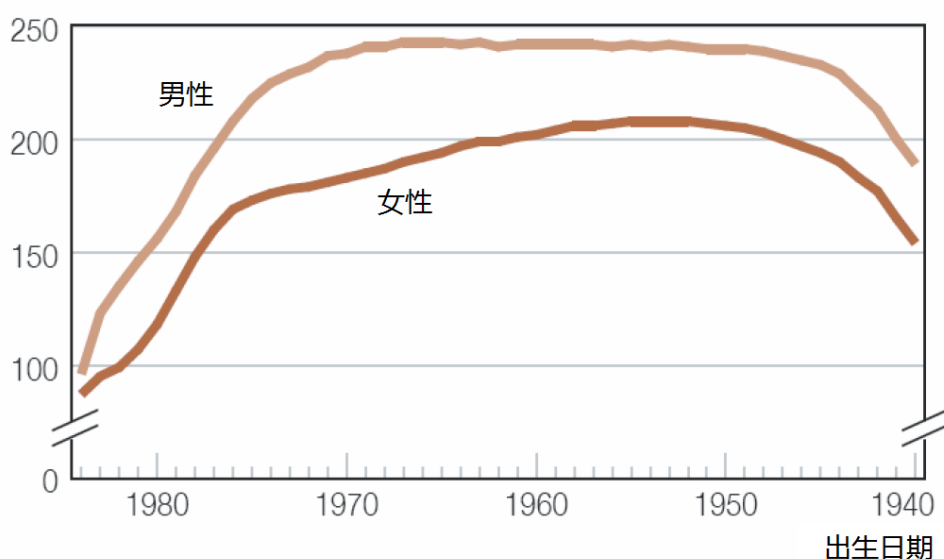
设计图表时，可以对比例尺进行调整，以使信息传递效果达到最佳。下面两幅折线图显示的数据相同，但却提供了截然不同的图景。



当坐标轴上的数值刻度不是从 0 开始时（如上图右侧例子中的 Y 轴），建议使用一些符号进行标识。最好的做法是让坐标轴显示从 0 开始，但是标记出一个折线或间断，举例如下。

图表中 Y 轴不从 0 开始的好的例子

2004 年瑞典 20-64 岁人口的退休金收入
以千瑞典克郎为单位的平均收入



资料来源：瑞典统计局（2006），《瑞典的女人和男人：事实和数据 2006》⁸

⁸ http://www.scb.se/statistik/_publikationer/LE0202_2006A01_BR_X10ST0602.pdf

4.7 控制图表的认知负荷

数据中可能包含多条希望通过图表来展现的信息。图表与出版物中的其他元素一样，被赋予“认知负荷”。通俗地说，认知负荷就是读者需要费多大力气才能理解你要表达的意思。具有高认知负荷的图表不容易被领会和记住，其信息也很难传递。具有低认知负荷的图表会让读者一目了然，其信息显而易见。关于有效图表设计的大多数原则都认为应该让认知负荷保持在较低水平。

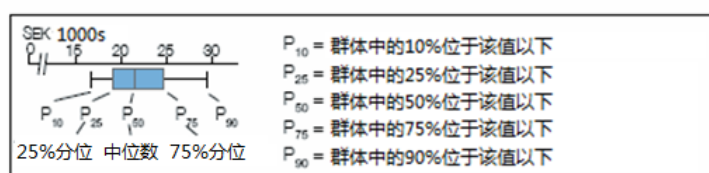
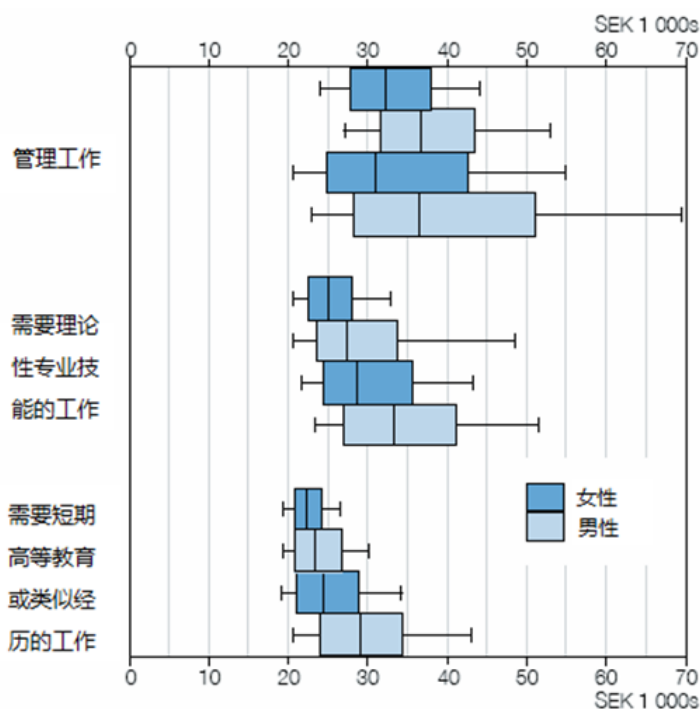
设计图表时，要控制其认知负荷。可以通过使用约定俗成的做法和格式来降低认知负荷和传递明确信息。如果想让受众关注数据的细微之处，也可以有意尝试提高图表的认知负荷。提高认知负荷可以迫使读者从不同的角度来就图表进行思考。下面是具有较高认知负荷的图表例子。

具有高认知负荷的图表范例

2004 年瑞典要求高等教育的职业群体的工资分布

以瑞典克郎（SEK）为单位的月工资

每个职业群体中上面两个箱线分别表示公共部门的女性和男性，下面两个箱线表示私营部门。



数据来源：瑞典统计局（2008），《瑞典的女人和男人：事实和数据 2008》⁹

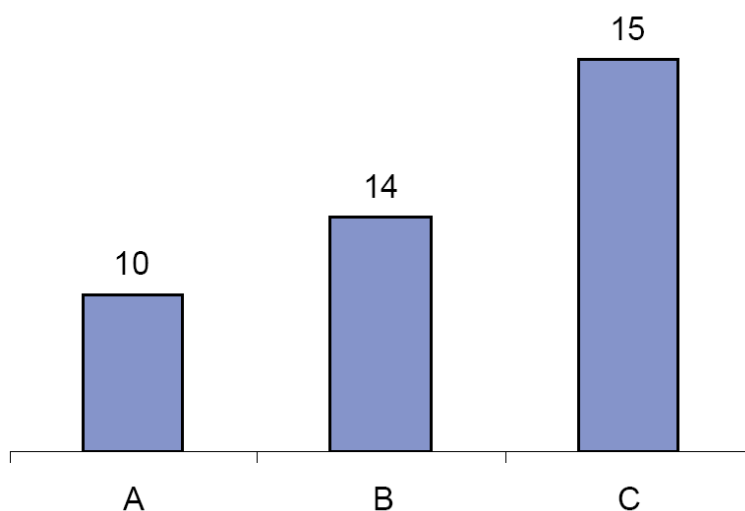
⁹ http://www.scb.se/statistik/_publikationer/LE0202_2008A01_BR_X10BR0801ENG.pdf

4.8 完善图表的若干建议

务求准确

应调整图形对象的大小，以准确显示其比例。如下例所示，在图表中将数据显示为大小不成比例的对象后，会造成误导。

不好的例子：图表对象相对大小不成比例



将数据排序

使用柱状图或饼图时，为便于比较，应从小到大对数据值进行排序。

不好的例子	好的例子																												
<p>2006年青少年生育率</p> <table border="1"><thead><tr><th>国家</th><th>生育率</th></tr></thead><tbody><tr><td>阿尔巴尼亚</td><td>13</td></tr><tr><td>格鲁吉亚</td><td>36</td></tr><tr><td>希腊</td><td>11</td></tr><tr><td>匈牙利</td><td>20</td></tr><tr><td>罗马尼亚</td><td>35</td></tr><tr><td>塞尔维亚</td><td>23</td></tr></tbody></table>	国家	生育率	阿尔巴尼亚	13	格鲁吉亚	36	希腊	11	匈牙利	20	罗马尼亚	35	塞尔维亚	23	<p>2006年青少年生育率</p> <table border="1"><thead><tr><th>国家</th><th>生育率</th></tr></thead><tbody><tr><td>希腊</td><td>11</td></tr><tr><td>阿尔巴尼亚</td><td>13</td></tr><tr><td>匈牙利</td><td>20</td></tr><tr><td>塞尔维亚</td><td>23</td></tr><tr><td>罗马尼亚</td><td>35</td></tr><tr><td>格鲁吉亚</td><td>36</td></tr></tbody></table>	国家	生育率	希腊	11	阿尔巴尼亚	13	匈牙利	20	塞尔维亚	23	罗马尼亚	35	格鲁吉亚	36
国家	生育率																												
阿尔巴尼亚	13																												
格鲁吉亚	36																												
希腊	11																												
匈牙利	20																												
罗马尼亚	35																												
塞尔维亚	23																												
国家	生育率																												
希腊	11																												
阿尔巴尼亚	13																												
匈牙利	20																												
塞尔维亚	23																												
罗马尼亚	35																												
格鲁吉亚	36																												
<p>数据按照国家的首字母顺序排列，导致数值大小难以比较。注意力放在了第一个值和最后一个值上，但二者之间并无特殊关联。</p>	<p>数据按照从小到大的顺序排列，容易进行比较。注意力集中在了数据集的最小值和最大值上。</p>																												

避免谬误相关

在同一幅图表上绘制具有不同比例尺刻度的变量容易导致错误的结论。两条曲线走势相同的事实并不足以建立二者的相关性。

下面的图表试图将日本的自杀率和长期失业率联系起来。这两条曲线走势相同，但是对应完全不同的两个变量。一个是每 10 万人口中的自杀人数，另一个是失业 12 个月及以上的失业人数占总失业人数的百分比。两个变量在整个周期内的取值碰巧都在 10 至 35 之间，但是稍微改变一下定义或比例尺，产生的图表就会迥然不同。下图成功地给人以两个变量相关的假象，但是却不能证明这一点。

不好的例子：两个变量之间产生谬误相关



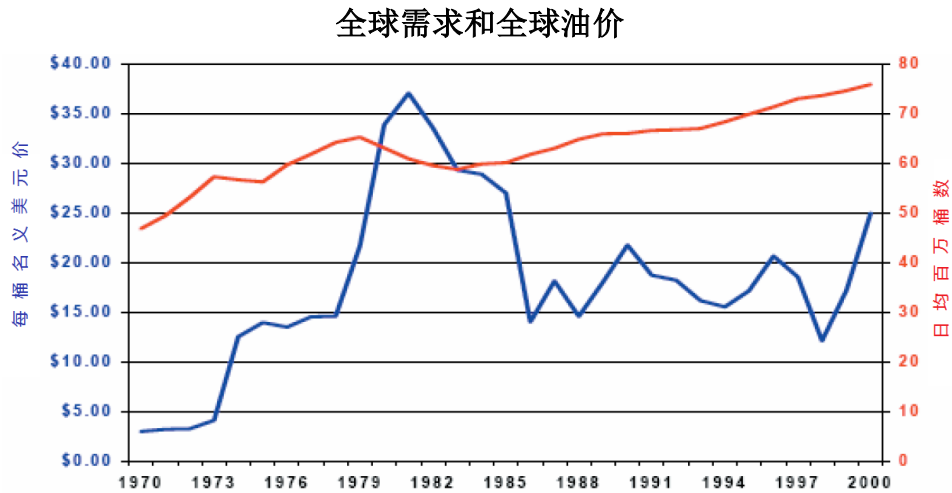
资料来源：Swivel¹⁰

谨慎使用双 Y 坐标轴

双 Y 坐标轴可能会造成混乱。你也许能用它成功地展示两个不同的变量（例如，价格和数量），但是要格外注意坐标轴的标签，要使每条数据线的颜色与相应坐标轴的颜色相同，方便用户使用。如下面的例子所示。

¹⁰ <http://www.swivel.com/graphs/show/28847825>

双 Y 坐标轴图标的好的例子



数据来源: Blessing et al. (2003), 《统计图表的认知测试: 方法和结果》¹¹

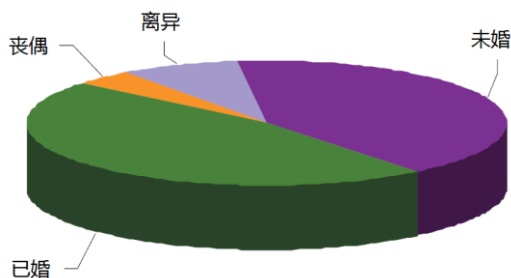
如果同一张图表中的两个 Y 轴使用相同的单位, 但用不同的比例尺, 即使该领域的专家也会被弄晕。双 Y 轴图表中貌似相交的两条线, 如果以同样的比例绘制, 他们可能永远也不会靠近彼此。

避免多余的图形特征

任何与当前数据集展示无关的图形特征都会妨碍对图表的理解。使用三维图表来展现简单数据集时尤其如此。虽然现有软件能够轻松实现三维图表, 而且有些人会认为三维图表更具吸引力, 但是三维图表经常让数据失真。如下图所示, 以简单的二维格式进行数据展示, 会使信息容易理解得多。

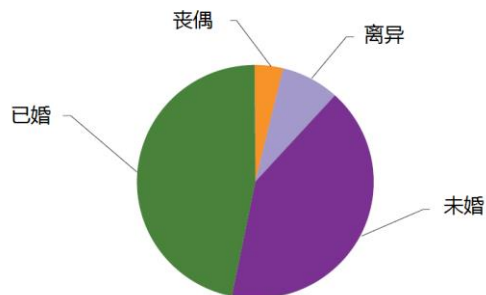
不好的例子

2004年冰岛18岁及以上人口的法定婚姻状况



好的例子

2004年冰岛18岁及以上人口的法定婚姻状况

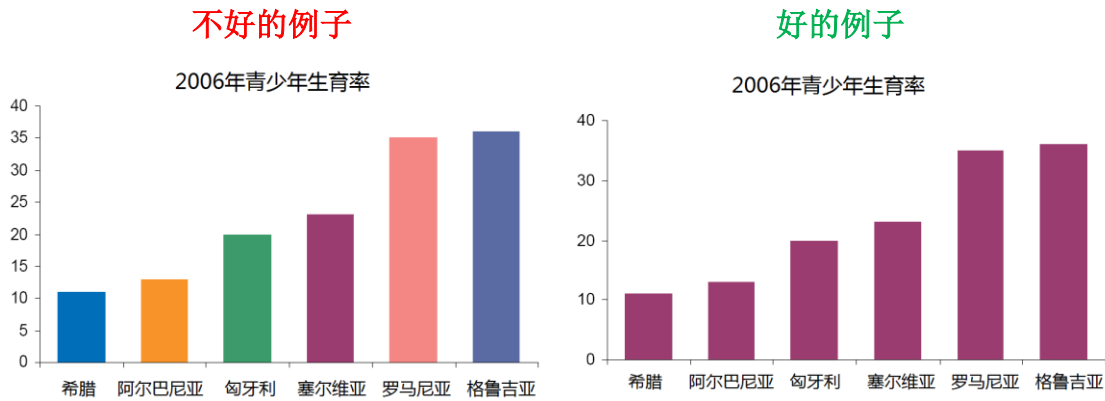


三维图表很少能带来更多好处, 而且容易迷惑读者。由于三维纵深的关系, 使得图表看起来有些部分离得近, 有些部分离得远。人的大脑会进行相应补偿, 认为

¹¹ <http://www.fcsm.gov/03papers/BlessingBradsher.pdf>

远处的东西实际上要比看起来的大一些。不过，当展示的数据具有两个及以上的属性时，通过使用纵深能让读者更容易理解你想传递的信息。

一般情况下，图表中应避免添加与数据无关的特征。例如，在柱状图中为每个值设置不同的颜色会加大阅读难度，如下面的例子所示。这种情况下应坚持使用一种颜色。



5. 地图

5.1 为何一图胜千数

地理信息是所有统计数据的必要组成部分。地理区域具有边界、名称及其他信息，从而方便对其进行定位并且将统计信息与之关联。这种空间关系对于普查数据而言尤为重要。

地图是对空间布局进行可视化的最有效工具。地图经过精心设计和展示后，不仅起到统计展示中的装饰功能，还能帮助人们识别和突出显示在表格和图表中可能无法清晰展示的分布和模式。

如果说“一图胜千言”，那么也可以说“一图胜千数”。在当今的可视化时代，地图是强有力的信息媒介，是专家、官员和社会公众的有力决策工具，满足了社会各界不断增长的信息需求。

地图的力量

容易理解的地图都是经过精心设计的，能使人对大量信息做到一目了然。地图能概括大量的数据表格或长而复杂的文字。当需要展示全国所有地区的统计信息时，

你可以制作一整套图表，也可以将所有信息显示在一张地图里。

现在许多制图技术都很成熟，从提供了大量分析功能和集成了地图组件的地理信息系统（GIS），到统计部门中用于专业地图和地图集制作的高端制图信息系统（CIS）。在纸质出版物的时代，统计人员往往没有充分利用地图，因为地图不能显示准确的数据值。交互式绘图工具的出现，消除了这一缺点。这种工具允许用户检索“地图背后”的实际数据。

地图在统计中的应用

无论是在普查和抽样调查的准备过程中，还是在调查结果的分析和报告中，地图都能发挥重要作用。下述情形应当考虑使用地图：

- 显示数据的地理位置和空间分布；
- 进行不同区域的比较；
- 概括大量数据并降低其复杂度；
- 传递明确信息；
- 对有关发现进行验证；
- 吸引人们的注意；
- 在地理信息系统中存储空间信息。

本章中，我们将讨论几种最常见的地图类型，并就如何制作出好的地图提供指导。

5.2 好地图的设计要点

地图制作集艺术、科学和技术于一体。由于其组织布局无一定之规，因此是一项比较复杂的任务。

随着 Web 2.0 技术的发展（见 6.2 节和 6.4 节），网上出现了许多交互式绘图工具，允许用户即时上传数据和检索专题地图。专题地图的制作成本大为降低、速度大为提升，但是还不能自动生成设计精巧的、能够准确传递信息的地图。

要设计出好的地图，需要考虑以下四个原则：

1. **明确目标受众：**如何以及在什么情况下使用地图？是否有访问方面的限制？
2. **明确要传达的信息：**数据有什么含义？要传达的信息多不多？
3. **明确数据特点：**需要在地图上绘制多少变量？是否有时间维度？
4. **明确合适的绘图技术、颜色和符号：**数据有什么特点（定量或定性，绝对值或相对值）？有任何技术限制吗（例如，格式要求或黑白片）？颜色或分类的惯用做法是什么？

好的地图应该做到：

- 简单易懂；
- 信息清楚客观；
- 准确展现数据，不误导；
- 将读者的注意力引导到最重要的信息上；
- 设计精巧，引人注目；
- 支持输出格式，符合受众需求；
- 能够独立存在而无需进一步解释；
- 色盲人群也能使用。

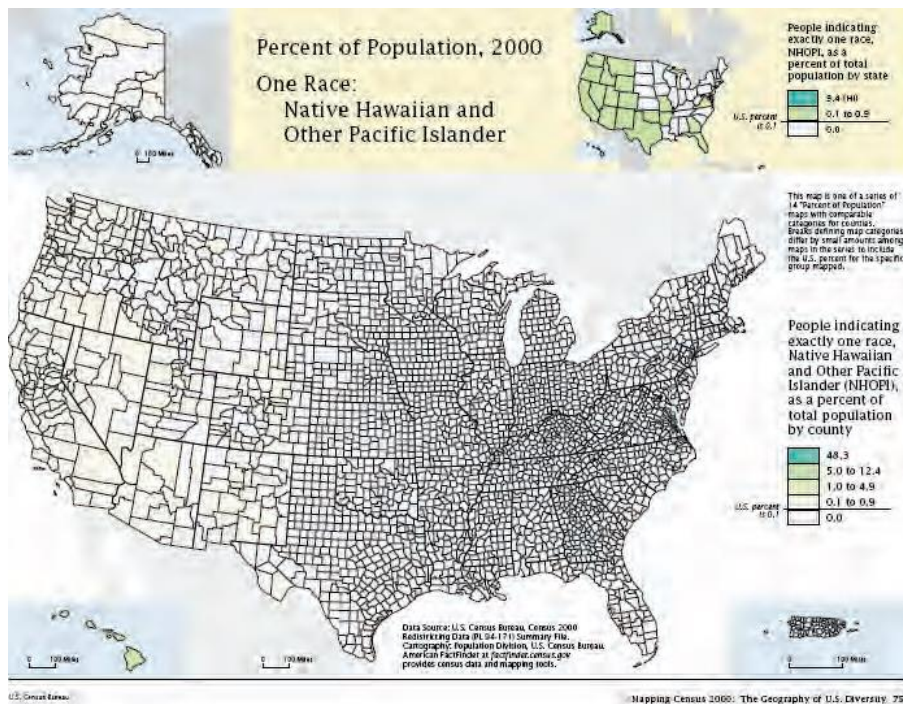
5.3 当不适合制作地图时

在开始制作地图之前，应当考虑地图是否是展现数据的最佳工具。如果图表或数据表格能够更好地传递信息，就不要浪费时间和精力去制作地图。

下列几种情况不适合使用地图来展示数据：

- 数据没有按地理区域进行细分；
- 数据点之间没有表现出明显的变异；
- 目标受众可能对地图理解有难度；
- 没有足够空间来展现地图，使得地图无法被正确地阅读和理解。

不好的地图例子



数据来源：Brewer, C.A.和 Suchan, T.A., 美国人口普查局（2001《2000年普查地图：美国多样性地理》，美国政府印刷局，华盛顿特区¹²。

¹² <http://www.census.gov/population/www/cen2000/atlas/index.html>

上面的例子表明，当没有足够的信息需要绘制时，使用地图就是浪费。上图是想展示美国一个少数民族群（夏威夷的原住民和其他太平洋岛民）的空间分布。整幅地图几乎是空的，因为在大多数县，这个族群不到全县总人口的1%，即使是该族群人口超过1%的少数几个县，在地图上也不容易看到他们。

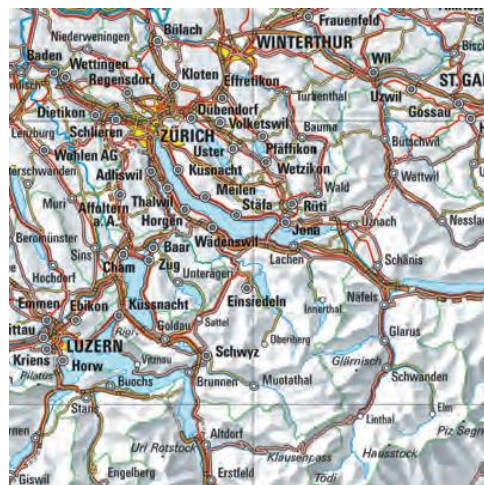
5.4 不同类型的地图

地图可根据规模、功能、设计、生产工艺或它们在出版物中使用的方式来进行分类。

一般地说，有两种地图类型：

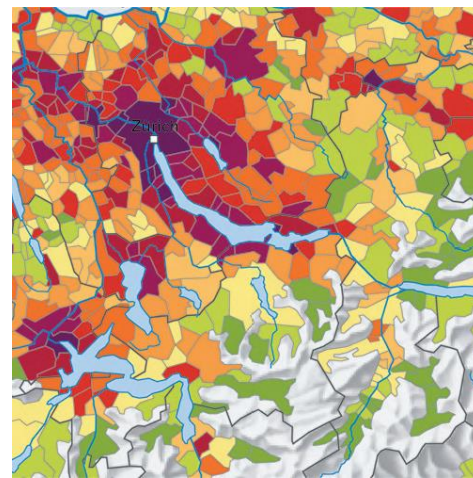
- 一般参考性（地形）地图，用于辨认空间方位和显示诸如湖泊、河流、山脉、海岸线、道路等大量不同地貌的位置，帮助用户识别地理区域的边界。
- 专题（统计）地图，用于显示一个或多个统计属性的空间分布。专题地图的设计总是服务于一个目的，并且回答有关政治、社会、文化、经济、农业或自然现象的具体问题。

一般参考性地图的例子



数据来源：swisstopo，瑞士联邦地理信息中心¹³

专题地图的例子



数据来源：瑞士联邦统计局¹⁴

静态或交互式地图

地图可以是静态的或交互的。用户无法对静态地图进行编辑。交互式地图为用户提供了灵活性，使用户能够改变设计，选择和检索数据，动态播放地图，改变主题或者聚焦自己最感兴趣的方面。

¹³ <http://www.swisstopo.admin.ch/internet/swisstopo/en/home.html>

¹⁴ http://www.bfs.admin.ch/bfs/portal/en/index/regionen/thematische_karten/maps.html

地图既可以基于图像，也可以基于数据。如果地图完全基于图像，它是事先制作好并且静态显示的。制作基于数据的地图时，需要在数据库中存储生成地图所需的所有信息（包括数据和元数据）。地图本身是在用户从网上发出请求时才生成的。利用这种技术，用户能轻松地对数据和地图参数进行更新和更改，而无需重新生成地图。

专题地图集

地图集的最简单形式是将多张地图合成一本册子。专题地图集还集成文字、图表和表格等其他形式的有价值的信息，极大地丰富了地图集的内容，能够全面展示统计数据。大多数统计机构认识到地图集在发布数据方面的潜力，并且正在制作广受欢迎的普查地图集或专题地图集（例如有关人口、卫生或经济数据的地图集）。

地图集技术过去十年已有显著提高。现代的在线地图集信息系统（AIS）允许用户浏览地图背后的数据，点击选中区域，定制自己的地图，结合自己的数据，并且与地图制作者或制作机构进行沟通。在幕后，新出现的制作技术推动了诸如地图制作、图形设计、数据分析、写作和翻译等不同类型的专业知识的整合。

好的在线专题地图集例子



数据来源：国际货币基金组织，*IMF Data Mapper*¹⁵

¹⁵ <http://www.imf.org/external/datamapper/index.php>

5.5 选择合适的地图类型

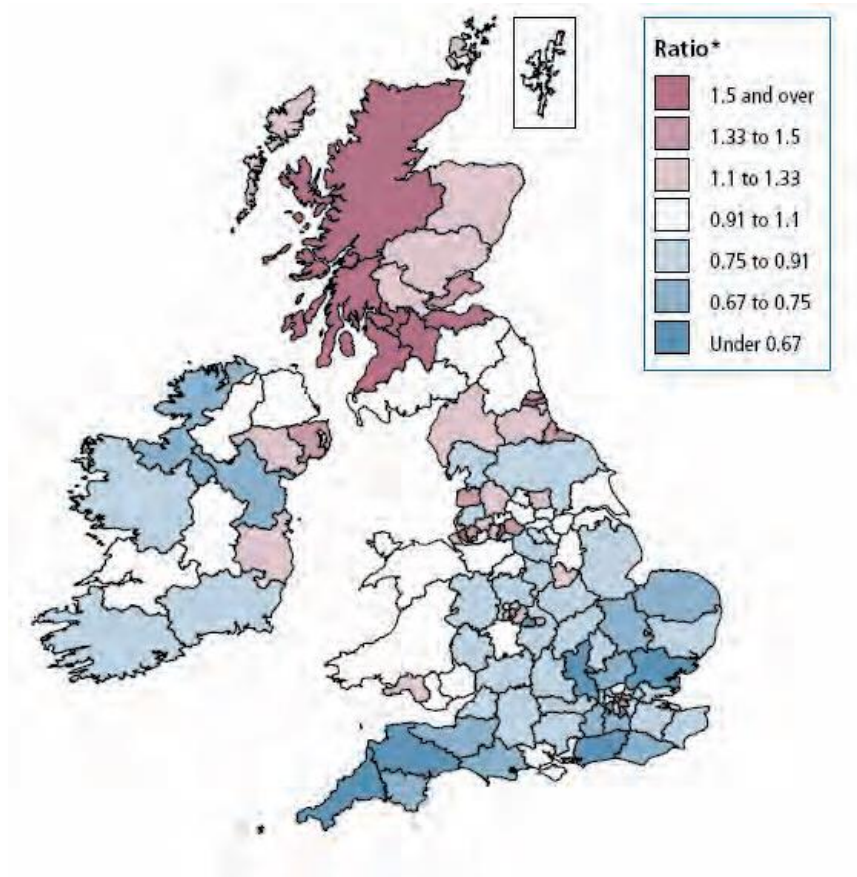
针对图表给出的建议也同样适用于地图：关键是根据信息类型确定适用的地图类型。选择的绘图技术是否合适也取决于数据特点。本节将针对三种最常用的专题地图类型提供了指导建议，这三种地图分别是等值区域地图、点图和比例符号地图。

等值区域地图

最常见的地图类型是等值区域地图，其中每块区域根据所示变量值的大小涂上相应颜色。这种地图对空间布局的可视化提供了一种简易方法，如下面的例子所示。

不幸的是，丰富的等值区域线数据和使用 GIS 设计的便捷导致等值区域图绘制技术常常被滥用。只有比值（即比例、比率或密度）可以使用这种绘制技术。千万不要用它来展现诸如人口规模等绝对值指标。

好的等值区域地图例子
喉癌：卫生局公布的发病率
英国和爱尔兰的男性人口，1991-1999 年



数据来源：英国国家统计局（2005年），《英国和爱尔兰癌症图谱（1991-2000）》，伦敦¹⁶。

等值区域地图的绘制基于预定区域的汇总数据，例如上面关于英国和爱尔兰卫生当局例子。这种地图类型非常适合针对这些卫生辖区进行讨论。不过，真实数据的分布模式往往与预定空间区域的分布模式不一致，从而可能导致严重误读。因此，等值区域地图比较适合展现每个预定空间单元内部数据均匀分布的现象。

使用等值区域技术时，数据分组很关键。地图上显示的空间模式是由数据值的分组来确定的。同样的数据，应用不同的分组方法，选取不同的分组个数，采用不同的组区间取值范围，可以生成不同的地图。这里有很多技术可以考虑，但是并没有一定之规，因此应当根据具体情况确定最适当的数据分组方法。

点地图

点地图使用符号来显示某个人群或现象的位置和密度，使用户能够快速掌握数据的大小，以及它们的集中或分散情况。每个点代表一个离散值，其中通常包含大量实体，如下面的例子所示。

好的点地图例子



数据来源：美国人口普查局《普查数据和应急准备》¹⁷

比例符号地图

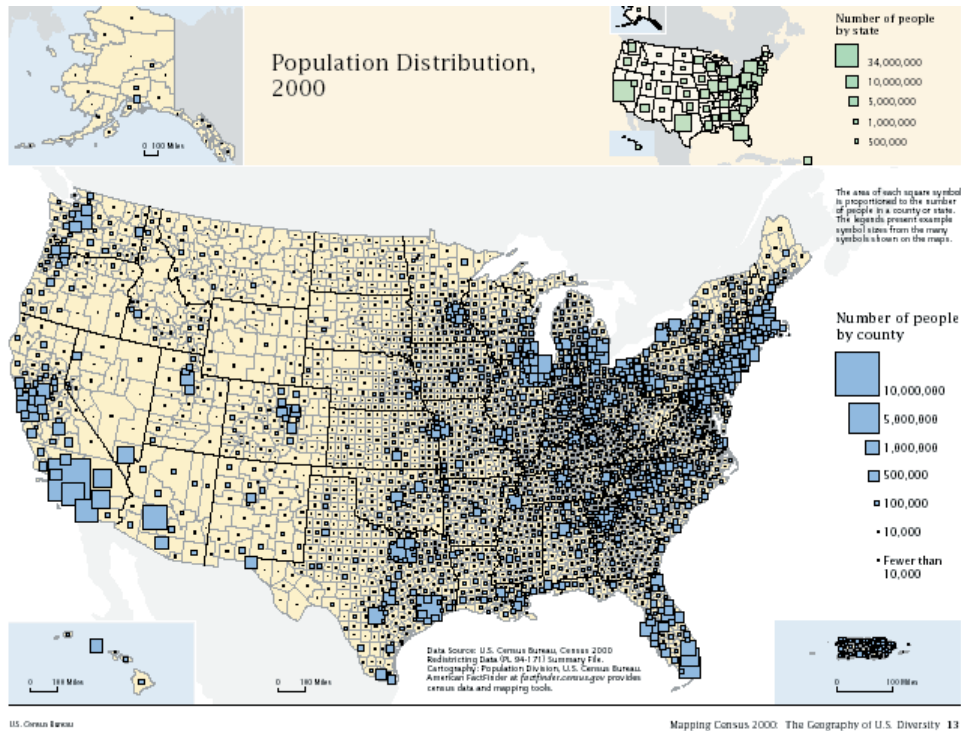
比例（或等级）符号地图用于显示绝对值。符号大小与所展现的人群或现象的大小成比例。每个符号被粘贴到空间单元中的特定点上，这个特定点通常是该空间中心点或首都的位置。

圆形结构紧凑且容易按比例绘制，是最常使用的符号。其它如正方形或三角形等几何形状也可以使用，如下面的例子所示。

¹⁶ <http://www.statistics.gov.uk/statbase/Product.asp?vlnk=14059>

¹⁷ <http://www.census.gov/Press-Release/www/emergencies/>

好的比例符号地图的例子



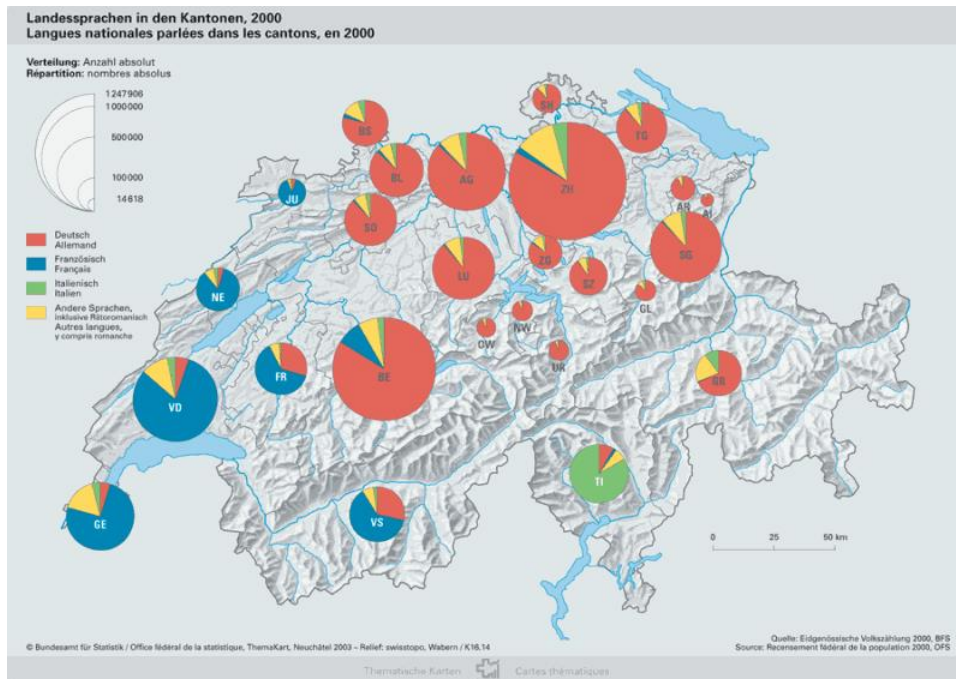
数据来源：Brewer, C.A. 和 Suchan, T.A., 美国人口普查局 (2001), 《2000 年普查地图：美国多样性地理》，美国政府印刷局，华盛顿特区¹⁸。

上述例子表明，制作设计巧妙的比例符号地图需要一定的绘制技巧。若符号的尺寸比相应空间单元的尺寸大，则难以辨认符号所指向的单元。如果有许多符号发生重叠，读图会更加困难，正如上面这幅美国人口分布地图所显示的那样。

可以将图表和地图技术相结合，以便在同一幅地图上显示不同人口类别的分布。在复杂的符号地图中，使用饼图或柱状图作为符号。在下面的瑞士地图中，饼图表示 26 个行政区中说主要官方语言（德语、法语或意大利语）或其他语言的人口比例。

¹⁸ <http://www.census.gov/population/www/cen2000/atlas/index.html>

好的复杂符号地图的例子



数据来源：瑞士联邦统计局，《瑞士图册——语言和宗教》¹⁹

使用复杂技术需要格外小心，因为此时地图和图例很容易造成信息超负荷。这种地图只能绘制到国家层面（例如，欧洲各国）或地区层面（例如，上图中的瑞士各行政区）。而且，饼图或柱状图中显示的类别不能超过 5 个。

5.6 设计小提示：简单至上！

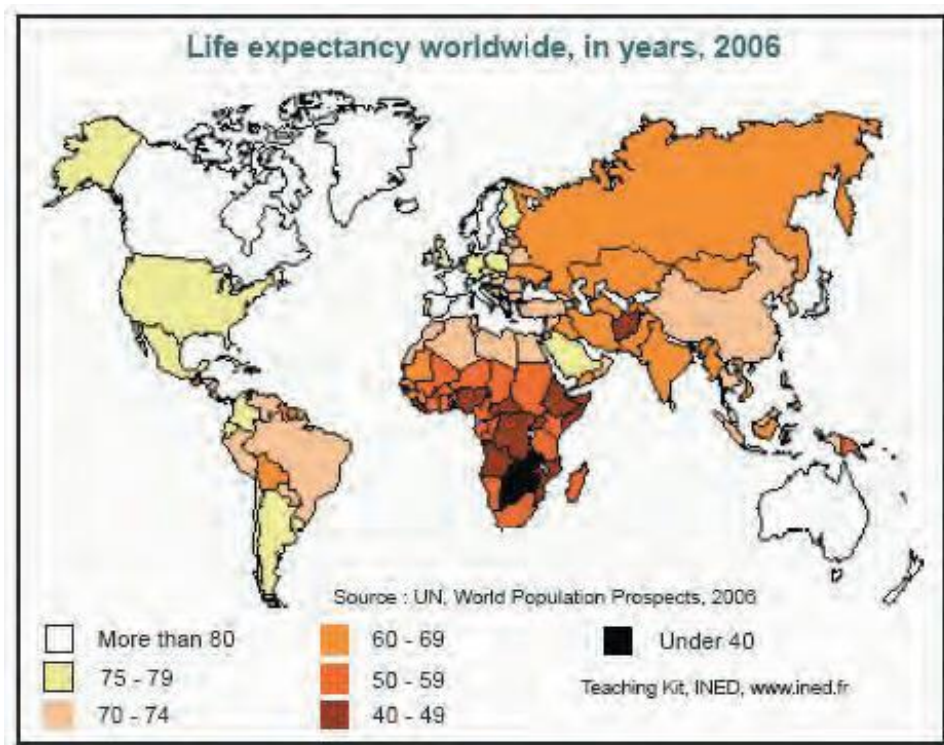
对于所有的数据可视化，最重要的建议就是保持**简单**，以确保信息的传递。注意不要显示太多信息或太多视觉上相互冲突的元素，以免分散用户注意力或把他们弄晕。

了解用户也至关重要。他们是什么背景？他们在日常生活中对地图是否熟悉？地图内容有没有冒犯他们的地方？应当对目标受众的敏感性有所了解。地图拥有强大的视觉冲击力，某些颜色或符号对有些人而言可能有负面含义。

设计的地图任何时候都应能**独立于故事文本或数据表格**。与图表一样，地图自身应当能解释清楚问题，而无须再去参阅其他文本或注释。地图一旦发布，就应当能独立于文本内容而被阅读或下载使用。因此，必须将下述不同的地图组件包括进来。

¹⁹ http://www.bfs.admin.ch/bfs/portal/en/index/regionen/thematische_karten/maps.html

无须进一步参阅其他资料的好的地图例子



数据来源：法国人口学研究所（INED）

地图组件

地图组件彼此会竞相吸引读者的注意力。为最大限度地提高信息传送效率，须确保一切以数据为中心进行布局，尤其是当还有其他信息（水、海拔高度等）要展现的时候。地图应当占总的展现空间的 80-85%。

为帮助用户理解地图，需要以下组件：

- **地图标题**，指明地图要描述的是什么。标题应当简明扼要，若要提供更为详细的信息（例如度量单位），可以添加副标题。
- **图例**，标识地图中用来展现数据的所有符号、图案和颜色。
- **地理单位**，这是地图中数据展示所对应的区域，应当在标题（或副标题）或图例中予以明确。
- **文本标签**，可以添加在地图上，用于标识重要或相关的地名或者其他信息。
- **地图比例尺**，用于帮助用户测量距离和对不同地图进行比较。
- **脚注**，用于提供关于定义或其他方法的信息。
- **数据来源**，应当在地图底部进行标明。
- **版权信息**，应当在地图底部标明对地图内容负责的作者信息。

有些地图中还可以包括其他组件，但并非必需：

- **指北箭头**，仅当地图上方不指向正北时才需要。
- **经纬度**，仅在世界地图或大洲地图上才需要。

- **定位地图**，是底图的小型副本，它将所绘制的区域放置到更大的地域背景中。如果用户不熟悉该区域的地理情况，则使用这种地图比较有用。
- **图表**，在有助于加强对地图理解时进行添加。

标题和图例中仅使用关键字词

虽然地图传递的是可视消息，相关的文字同样重要。标题及图例的措辞决定了用户如何理解和解读地图，因此应仔细斟酌。下面给出地图中使用文字的几点准则：

- 务求准确，简单至上。
- 只使用关键字词，避免在标题、图例或脚注中重复相同的词。
- 使用中性语言。
- 避免使用缩略语。
- 图例字号要小于标题字号，脚注字号要再小些，但要确保所有文字清晰可辨。

精心设计图例

图例设计必须确保对地图的正确理解和解读。每一种地图类型都需要不同类型的图例。但是对于等值区域地图和比例符号地图，有以下几条基本原则：

- 所有分组应当有明确的上下限：避免出现诸如 100-200、200-300、300-400 这样的范围划分。
- 组别之间不应当存在间隔：避免出现诸如 1.0-1.5、2.0-2.5、3.0-3.5 这样的范围划分。
- 哪些区域没有可用数据，应当标示清楚。
- 在地图中展示多个变量时，图例应当按照变量的重要性进行降序排列。

颜色：另一项重要选择

颜色是最强大的图形特征之一。在地图上选择颜色要非常小心，因为颜色既有助于读图，有时也会产生误导。

专题地图中的颜色选择取决于数据和地图类型，但是还应考虑以下三个方面：

第一，要了解某种颜色是否存在约定俗成的用法，以及可能存在的正面或负面含义。

第二，要确保每个人都能通过地图中所使用的颜色来理解你的信息。例如，色盲人群区分某些颜色会有困难，最常见的是红绿色盲，如果使用红色和绿色来显示增长和下降这两种区域类型之间的差异，色盲人员就无法分辨。简单的解决方案是用紫色取代红色，这样色盲人群也能够区分。

第三，对于分组不多且取值连续的数据（例如人口密度），应考虑使用同一颜色的不同深度而非不同的颜色来表示。如果是离散数据，或者数据有正值和负值之分，使用不同颜色则往往更为合适。

6. 新兴的可视化技术

6.1 为何说可视化不只是图片

新兴工具和技术为实现数据可视化以及面向用户进行更有趣的展示提供了新的机遇。动态表格、图表和地图生成器允许用户处理数据并自己进行可视化。动画和视频形式引人入胜，有点像看电视，能较好地解释随时间而发生的变化，并且利用口头或文字描述来解释数字背后的含义。诸如迷你图和标签云（见 6.4）等新型可视化工具也不断涌现，为信息图解提供了更多选择。

最新的网络技术以及用户群体中对其产生的期待，正在改变统计机构传播统计数据的方式。互联网现在是双向的交流论坛，用户借此可以分享他们自己的数据可视化，讨论他们的研究结果。例如，Many Eyes²⁰、Swivel²¹和 Data Place²²等网站都是讨论和分享数据和图形的在线社区。

新的可视化技术和更多互动网站的发展，在提高用户使用灵活度的同时，也可能给统计机构带来麻烦。不管是有意还是无意，用户越来越容易曲解统计数据，然后将这些被曲解的结果广泛传播给他人。因此，统计机构有必要针对如何应用和提供新的可视化技术制定明确的政策。

本章将对这些新的可视化工具和技术进行概述。

6.2 动态可视化

随着互联网的出现和 Web 2.0 技术²³的发展，用户可以与数据进行交互并创建自己的可视化产品。现在许多统计机构通过其网站对外开放他们的数据库访问，方便用户自行查询和下载统计信息；同时配套了越来越多的可视化工具，允许用户在线创建表格、图表或地图，而无需将数据和结果下载到其他应用程序中。

²⁰ Many Eyes 网站允许用户上传数据、生成图表及进行其他可视化，并对有关发现进行讨论（参见 <http://manyeyes.alphaworks.ibm.com/manyeyes/>）。

²¹ Swivel 网站提供的功能与 Many Eyes 类似（参见 www.swivel.com）。

²² Data Place 网站为用户提供关于美国各城市、乡镇和州的统计数据（参见 www.dataplace.org/）。

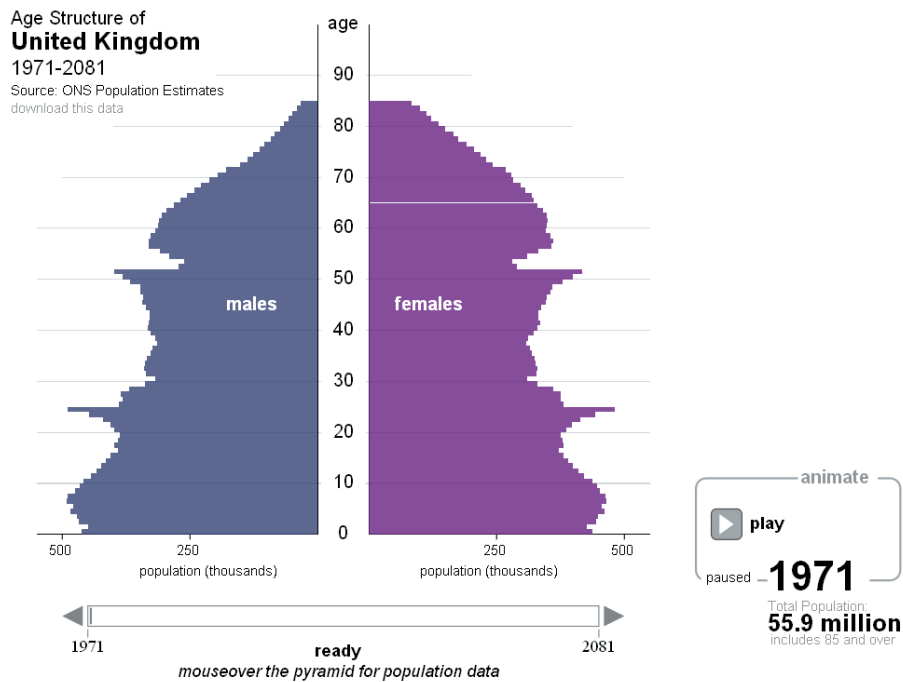
²³ “Web 2.0”一词用于描述新一波互联网技术发展。该技术允许用户从事更多网上操作，例如可以添加、更改或影响网页内容，而不只是在在线访问信息。维基网站的例子有维基百科、博客，社交网站的例子有 Facebook 或 LinkedIn。

目前对于向用户开放这种级别数据控制权的后果存在一些担心。用户可能会生成毫无意义的图表或得出不恰当的相关关系。不过，让用户能够访问和处理数据肯定是第一位的。清晰显见的方式提供关键的元数据，向缺乏经验的用户提供支持，以及监督和纠正任何对数据的滥用，可以将潜在的麻烦减至最小。

6.3 动画和视频

动画和视频是新出现的两种重要的数据可视化技术。用户喜欢这种通过动态图像来接收信息的方式，想想电视和电影的受欢迎程度，就不会对此感到诧异。这种形式通过将音频或文字描述与图解说明结合起来，传递数字背后的含义，从而使讲故事变得更加容易。

动画传播统计数据的好的例子



数据来源：英国国家统计局²⁴

由英国国家统计局和加拿大统计局等多家统计机构制作的动态人口金字塔，是在简单界面中将动画与交互相结合的范例。用户可以点击“play”（播放）来观看人口金字塔的形状如何随着时间的变化而变化。用户可以选择年龄组，查看详细数值以及各组在总人口中的比例，与图表进行交互。

Gapminder²⁵的共同创始人 Hans Rosling 在以动画形式图解和传播统计数据方面取得了巨大成功。Rosling 已经通过在线视频这个日益受到欢迎的互联网功能吸

²⁴ http://www.statistics.gov.uk/populationestimates/svg_pyramid/uk/index.html

²⁵ <http://www.gapminder.org/>

引了大量的用户，他在 2006 年 TED 会议²⁶上的演讲视频已有成千上万次在线观看和下载。基于这种方式在传播统计数据方面的普及，Rosling 还进一步开发了名为“gapcasts”的视频课件短片，内容涵盖孕产妇死亡率、全球化、能源和人类发展趋势等诸多话题。

动画和新型网络技术相结合传播统计数据的好的例子



数据来源：Gapminder²⁷

6.4 Web 2.0 和围绕数据开展社区建设

Many Eyes、Swivel 和 Data Place 等网站通过围绕数据可视化和共享来创建网络社区，正在为可视化展示增加新的维度。这些网站允许用户上传数据集和创建图形，便于与其他用户进行分享和讨论。许多应用以通常被称为“混搭应用程序”的方式将来自两个或多个来源的数据或功能结合起来，从而产生新的服务。混搭的一个例子是将统计数据和地图数据结合起来，生成数据的地理视图²⁸。

一些官方统计数据生产者正尝试使用 Web 服务与更广泛的受众进行沟通。例如，联合国欧洲经济委员会（UNECE）已向 Swivel 上传 4 个数据集，不过效果喜忧参半。可喜的是这几个数据集在头 18 个月内的浏览次数均超过 5000 次，表明许多用户已经看到了数据。然而同一时期，网站上仅有 2 次用户评论，Swivel 上的

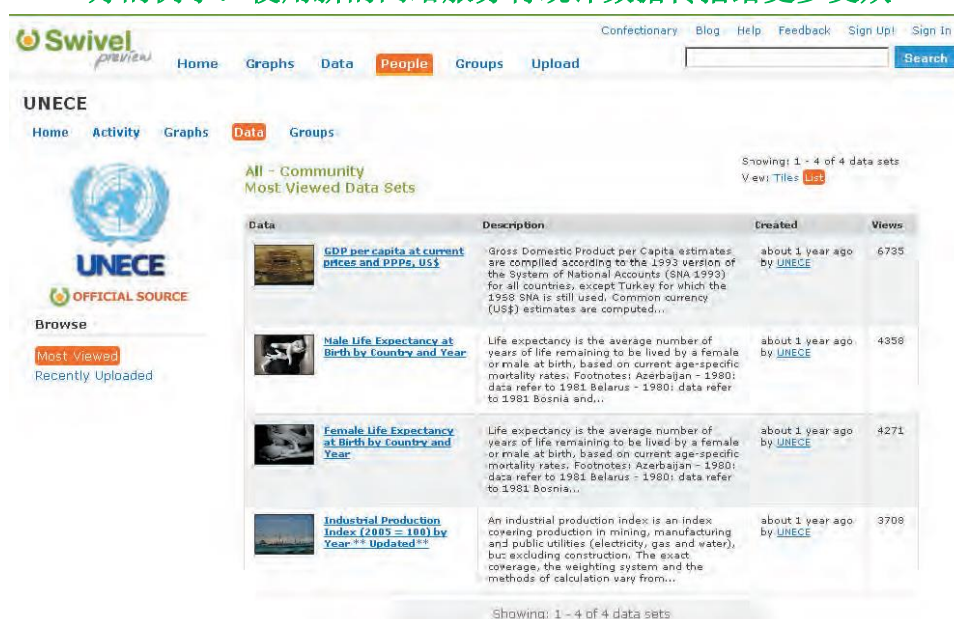
²⁶ TED 是一种吸引潮流中人在各自领域就当前话题进行演说的年度会议，其中 T 代表技术，E 代表娱乐，D 代表设计。这些演讲在 TED 的网站 www.ted.com 上有视频发布。

²⁷ <http://www.gapminder.org/video/gap-cast/>

²⁸ 例如，最新版 PC-Axis 套装统计传播软件（<http://www.pc-axis.scb.se/>）允许将数据与谷歌地图和谷歌地球相结合。

UNECE 统计数据库网站链接仅被关注 10 次，因此可以说，事实上并没有获取有关这些新用户的任何信息。

好的例子：使用新的网络服务将统计数据传播给更多受众



数据来源：Swivel²⁹

目前为止，尽管诸如此类的网上社区不全是成功的案例，但是这种形式显然为覆盖更多受众提供了一种相对容易的方法，因此是数据可视化值得跟进的新领域³⁰。

6.5 其他新的可视化技术

迷你图

迷你图顾名思义个头较小，形式上可表现为与文字大小相当的折线图，能显示出时间趋势。迷你图的好处是能够一览无余地显示大量信息，并且可以在旁边通过文字解释其含义。

迷你图由 Tufte 在 2006 年时首次提出。下面例子中的迷你图用于展示欧元对其他货币的汇率波动。这些“密集、简单、如字词般大小的图形”增强了表格数据的视觉展示效果但又不会太占地方。

²⁹ <http://www.swivel.com/users/show/1005968>

³⁰ 更多参考资料请见 IAOS 统计学报特刊 (Vol. 25, nos 3-4, 2008): “Web 2.0 和官方统计数据”，材料链接：<http://iospress.metapress.com/content/v0376364l348/?p=fc2e171758ee4053a01be16bbb ae10eb&pi=0>

迷你图的好的例子



数据来源：Tufte, E.R. (2006), 《美丽的证据》, Cheshire CT, 图形出版社。

标签云

标签云（有时也叫文字云）是对字词或标签在特定文本或数据集中出现频率的可视化展示。不同的标签大小和颜色代表标签的受欢迎程度或重要性。网站上经常能看到关于标签的类别列表，其中每个字词都是可以点击的链接，点击后提供给用户更多与该字词有关的信息。

标签云是标示常用的文本术语和建立关键词分类的有用方式。下例中的标签云是通过统计本章文字生成的，清晰地显示了关键词。

标签云的好的例子



利用 Tag Cloud Builder 生成的标签云³¹

7. 易用性问题

若要基本统计信息得到广泛和方便的使用，必须使其处处可及。这意味着每个人都必须能够查询和理解信息，而不论他们使用何种技术或可能具有何种残障。因此，传播策略的一个重要目标就是生成容易使用的信息。

³¹ <http://tagcloud.oclc.org/tagcloud/TagCloudDemo>

为了吸引并拥有大量用户，信息展示方式应结构化，条理清楚，且符合已建立的公认标准。本章中考虑了易用性的三个方面，其中存在部分重叠。

- 提供合适的元数据，以帮助用户理解数据；
- 提供多种格式的数据，包括通过新媒介如移动电话或其他手持设备提供数据；
- 确保残障人士能够最大限度地获取信息，通常应遵守特定的法律或政策要求。

让所有人都能访问到统计信息，需要加强意识，多做一点工作，多投入一些资源。不过，在基本统计资料的获取上，平等原则十分重要。应在深思熟虑的基础上制定访问策略，以惠及所有人。

本章中，我们将讨论优质信息的主要组成要素，这些要素在统计信息传播时应当加以考虑。

7.1 文字

为使用户容易地找到他们想要的信息，行文必须做到简明扼要、连贯通顺且条理清楚。把文章划分为若干层次分明、逻辑清晰的章节，加上标题和副标题，会使文字搜索和格式转换变得更加容易。

为了有效服务于所有目标群体，文字应当以多种格式存在，例如：盲文、音频或超大字体。当创建原始文本时，应当预先想到可能需要的格式转换，以尽量减少由此带来的影响。例如，为每条图形信息添加文本描述将会减少在网页上将其转换为盲文、音频甚或网站 HTML³²格式的工作量。

由国际万维网联盟（W3C）³³设计的易用性规则，旨在确保使用自适应技术来获取电子格式信息。包括屏幕阅读器、字符放大系统和便携式数据接入设备（PDA）等多种措施。

7.2 表格

使用表格展示数据时，同样应当注意易用性。如果展示的数据没有包括用于理解和解读数据所需的全部信息，数据则不可用或者可能产生误导。

例如，考虑表格跨页显示的情况。如果在页面上看不到表格的列标题或行标题，用户将很难读懂表格。因此，在每一页上重复显示列标题和行标题非常重要。

³² 超文本标识语言。

³³ <http://www.w3.org/>

好的例子：满足易用性要求的表格

2007 和 2008 年加拿大牲畜存栏量

牲畜种类	2007		2008	
	千头	%	千头	%
牛	15 885	50.2	15195	52.0
猪	14 690	46.4	12985	44.4
羊	1 096	3.5	1062	3.6
总计	31 671	100	29242	100

数据来源：加拿大畜牧业统计，2008 年 8 月³⁴

每个数值不仅与牲畜种类和年份有关，还要涉及牲畜头数（单位：千头）或百分比。显示屏阅读器必须针对表格中的每个数值给出全部信息。因此，用户将听到：“2007 年，肉牛所占比例为 50.2%”。

对于 HTML 这样的电子格式，列标题和行标题中可以包括标签，必要时还可附带关于标签内容的简短说明。使用标签对大家都有好处，当下载表格后，对列标题和行标题直接进行转换，避免发生混乱。此外，如果表格结构比较复杂，一定要准确描述数据的组织方式，以使用户能通过最小的努力获得最多的信息。

7.3 图表

制作图表时，要考虑到并非所有用户都能够正常访问图片。可以按照图表所提供的信息，生成文字说明。下面是来自《全国盲人协会磁带录制手册》中的例子。

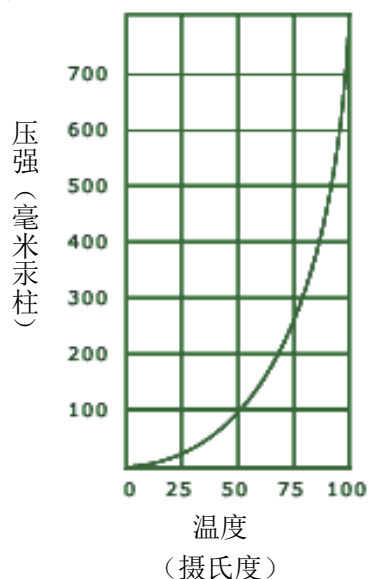
在网站上，“alt text”标签（替代文本）或“longdesc”属性（存储于其他页面上并能通过电子邮件地址进行访问的详细描述）的使用是 W3C 中所认可和介绍的技术。这些描述也可用于音频、盲文或其他格式文件的制作中。

可能你有太多的图表都需要创建替代文本，如果人工来做会很麻烦。一些能够从图表中自动提取信息的设计工具正在开发之中。这些工具在生成图表的同时，也能生成描述性页面。

³⁴ <http://www.thedairysite.com/articles/1613/canadian-cattle-statistics-august-2008>

好的例子：带有替代文本的图表

水的蒸气压与其温度之间的关系



“水的蒸气压与其温度之间的关系”。这是一幅曲线图，X轴是温度，单位为摄氏度，取值范围从0度到100度；Y轴是压强，单位是毫米汞柱，取值范围从0到800毫米汞柱。曲线从原点开始逐渐上升，当X为25度时，Y大约为40毫米汞柱；当X为75时，Y正好位于300下方；当X为100时，Y大约为760。

数据来源：《易于使用的数据媒体：电子出版物、多媒体和网站的设计指南》³⁵

7.4 地图

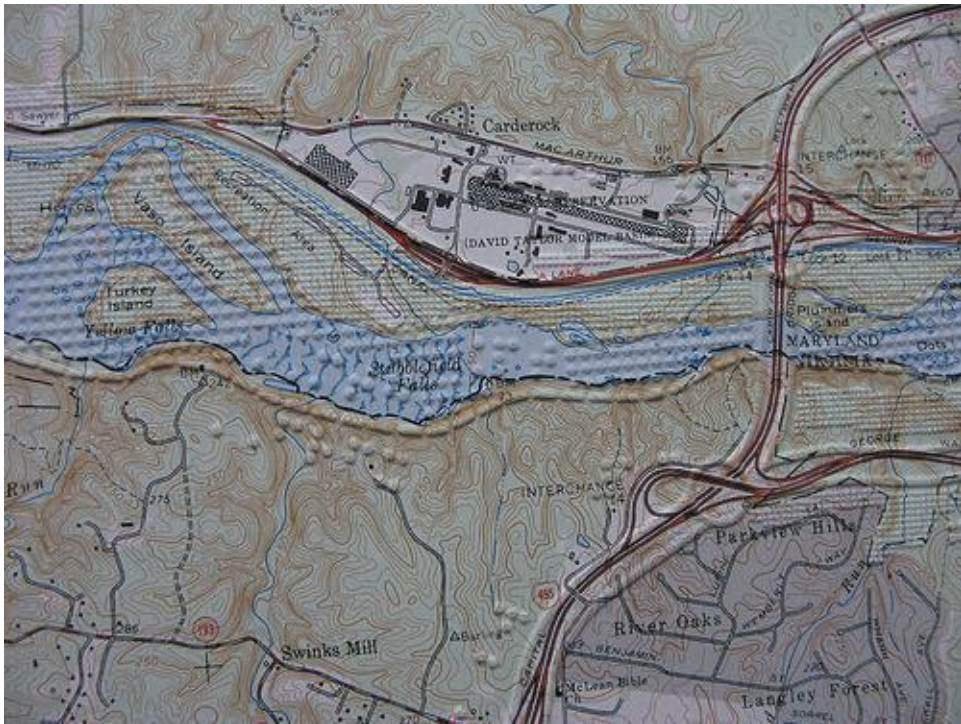
地图遇到的主要技术挑战是如何让某些残障人士也能进行访问。信息通过图像和颜色的结合进行传递，不过这两种方法的访问标准实际上并不相容。因此，我们应当考虑为地图附带一份替代文本，以提供地图中所展示的信息，并且（或者）允许访问数据表格。

另一种做法是提供搜索工具，允许用户选择诸如区域、乡镇或街道等参数。数据库搜索可以类似这样：“搜索年龄在15-49之间、在指定区域居住并从事农业的男性人数”。生成的查询结果应当是统计表格的形式，提供的信息则与下面的可视化显示相同。

³⁵ http://ncam.wgbh.org/publications/adm/guideline_f.html

好的例子：易于盲人使用的地图

华盛顿特区（部分）的盲文地图



数据来源：美国地质勘测图书馆，Flickr 公司提供³⁶

7.5 元数据

统计机构应确保为用户提供用于理解数据所需的元数据，以及元数据的优势和局限性。必须及时对元数据进行更新，以体现定义、分类和方法的最新变化。

为保证所有人都能访问和使用元数据，应当考虑以下几点建议³⁷：

- 允许通过诸如纸质出版物、CD-ROM 等多种不同媒介来访问元数据。由于用户通常通过互联网来查找最新元数据，因此建议所有元数据都应当能够从网上获取。
- 元数据的展现方式应能满足各类用户的不同需求和（或）统计专家的需要。
- 在互联网上免费发布元数据，即使纸质版本仍要收费，或者元数据所描述的统计数据需要收费才能获取。
- 确保元数据和它所描述的表格和图表的之间的相互链接有效。
- 元数据不仅以本国语言提供，如果可能的话，可另外提供英语等通用语言版本。
- 提供能够进行免费文本搜索的本地搜索引擎。
- 采取措施确保 URL（统一资源定位器）网址的稳定性，或提供新旧 URL 地

³⁶ <http://www.flickr.com/photos/98169608@N00/3296197787/>

³⁷ 元数据的报告和发布准则节选自 2007 年 OECD 出版物《数据和元数据报告及展示手册》，具体参见 <http://www.oecd.org/bookshop?9789264030329>。

址之间的切换链接，把用户重新指向新的地址。鉴于网站之间链接的重要性，这个问题很关键。

- 提供联系人姓名或电子邮件地址，以方便获得关于概念、定义和统计方法的详细信息。有些机构提供的“联系人”可能是机构通用的联系方式或客户咨询的转介服务。

有些统计机构参照金字塔的层级方式提供元数据。通过这种方法，随着用户从金字塔顶端向下走，元数据也更加详细：

- **金字塔顶端：**大致了解统计数据所需的关键元数据，包括与数据状态（是初步数还是核定数）有关的信息。大众用户所需的元数据基本仅限于此。
- **金字塔中部：**一些与数据使用有关的解释性说明，提供关于统计数据的简短描述（定义、关键问题、局限性等）。中级用户可能会查询这类元数据。
- **金字塔基部：**提供最详细的方法性信息，通常以方法手册或类似文档的形式存在。资深用户以及使用数据进行详细研究的用户可能会对这类元数据感兴趣。

8. 参考文献与延伸阅读

Bertin, J. (1981), *Graphics and Graphic Information-processing*, New York, de Gruyter.

Bertin, J. (1983), *Semiology of Graphics: Diagrams Networks Maps*, Madison Wis., University of Wisconsin Press.

Blessing, C., Bradsher-Fredrick, H., Miller, H., Miller, R. and Rutchik, R. (2003), *Cognitive Testing of Statistical Graphs: Methodology and Results*, Washington D.C., U.S. Energy Information Administration.

Bosch ten, O. and Jonge de, E. (2008), “Visualising official statistics”, in *Statistical Journal of the IAOS*, vol. 25, nos 3-4, pp. 103-116, available at: <http://iospress.metapress.com/content/v03763641348/?p=fc2e171758ee4053a01be16bbbae10eb&pi=0>.

Cleveland, W.S. and McGill, R. (1984), “Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods”, in *Journal of the American Statistical Association*, 79, pp. 531-554, available at: <https://secure.cs.uvic.ca/twiki/pub/Research/Chisel/ComputationalAestheticsProject/cleveland.pdf>.

Cleveland, W.S. and McGill, R. (1987) “Graphical Perception: The Visual Decoding of Quantitative Information on Statistical Graphs (with Discussion)”, in *Journal of the Royal Statistical Society Series A*, 150, pp. 192-229, available at: http://www.wjh.harvard.edu/~kwn/Kosslyn_pdfs/1987Cave_JRoyStatSocA_CommentaryClevelandArticle.pdf

Few, S. (2004), *Show Me the Numbers: Designing Tables and Graphs to Enlighten*, Oakland CA, Analytics Press.

Gardner, J. (2008), "Blogs, wikis and official statistics: New perspectives on the use of Web 2.0 by statistical offices", in *Statistical Journal of the IAOS*, vol. 25, nos 3-4, pp. 81-92, available at:
<http://iospress.metapress.com/content/v03763641348/?p=fc2e171758ee4053a01be16bbbae10eb&pi=0>

Harris, R.L. (2000), *Information Graphics*, New York and Oxford, Oxford University Press.

Kennedy, D. (2007), *Research Paper: Data Visualization*, Canberra, Australian Bureau of Statistics, available at:
[http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/7F8E375FC22D26A5CA25731C0022DF1B/\\$File/1211055001_jul%202007.pdf](http://www.ausstats.abs.gov.au/ausstats/subscriber.nsf/0/7F8E375FC22D26A5CA25731C0022DF1B/$File/1211055001_jul%202007.pdf).

Miller, J.E. (2004), *The Chicago Guide to Writing About Numbers*, Chicago, University of Chicago Press.

Playfair, W. (1786), *The Commercial and Political Atlas: Representing, by Means of Stained Copper-Plate Charts, the Exports, Imports, and General Trade of England, at a Single View*, London.

Playfair, W. (1801), *Statistical Breviary: Shewing, on a Principle Entirely New, the Resources of Every State and Kingdom in Europe*, London, Wallis.

Robbins, N.B. (2005), *Creating More Effective Graphs*, Hoboken NJ, John Wiley & Sons.

Schulz, T. (2009), *Guidelines on the Presentation of Statistical Maps*, UNECE Work Session on the Communication and Dissemination of Statistics, Warsaw, available at:
<http://www.unece.org/stats/documents/ece/ces/ge.45/2009/crp.1.e.pdf>.

Smith, A. and Rogers, S. (2008), "Web 2.0 and official statistics: The case for a multi-disciplinary approach", in *Statistical Journal of the IAOS*, vol. 25, nos 3-4, pp. 117-123, available at:
<http://iospress.metapress.com/content/v03763641348/?p=fc2e171758ee4053a01be16bbbae10eb&pi=0>.

Tufte, E.R. (2001), *The Visual Display of Quantitative Information (2nd ed.)*, Cheshire CT, Graphics Press.

Tufte, E.R. (2006), *Beautiful Evidence*, Cheshire CT, Graphics Press.

Wainer, H. (1984), "How to Display Data Badly?", in *The American Statistician*, vol. 38, no. 2, pp. 137-147, available at:
http://www.soc.washington.edu/users/bp Pettit/soc504/wainer_display.pdf.

Ware, C. (2004), *Information Visualization: Perception for Design*, San Francisco CA, Morgan Kaufmann Publishers.

Further useful information and discussion on the presentation of statistics can be found at: <http://blogstats.wordpress.com/>.

让数据有意义

统计数据展示指南

一图胜千言。以图片形式展示数据通常会使数据形态更加清晰可见。展示数据有许多方法，从简单的柱状图到较为复杂的散点图、专题地图和动态图形。

本指南作为实用工具，旨在帮助统计数据生产方采用清晰和有意义的方式展示数据。本指南针对如何制作高效的表格、图表和地图，以及如何使用其他形式的可视化工具以使统计数据走进生活给出建议。本指南还就如何避免不好的或误导性的可视化展示提出建议。清晰的数据展示，明确的目标受众，都将增加统计数据的使用效果，揭示其所含信息的价值。

