

疫情的流图

——以设计实验理解可视化技术应用方法

向帆 / 清华大学美术学院

朱舜山 / 清华大学美术学院智能交互系统设计研究所

摘要：“看见”传染病发展的趋势是每一个人的愿望。作为对主流媒体疫情数据信息传播的一种补充，我们向公众提供一种具有流动感的交互可视化程序——“疫情的流图”，让人们可以通过移动终端看到各地的每日新增病例的流向：地区与数量的关系。从视觉设计的角度，我们考察了四种堆叠图形式在疫情数据表达方面的可用性，并在开发中进行扩展性设计，最终实现了“疫情的流图”。这个可视化项目成了一个不断自我更新的、生长的、吸引大量公众访问和参与的公共信息媒介。结果证明，“疫情的流图”的外部轮廓能够非常概括性地表达新冠疫情传染病暴发和发展的总趋势，但并不适合表达数据量微弱的地区信息，而动画、交互和地区信息分类则可能对其加以弥补，用户反馈则显现了项目的有效性。由此表明，引用成熟的可视化形式并针对使用目标而扩展开发是视觉化设计的基本原则和有效工作方法。

关键词：流图 堆叠图 新冠疫情 数据可视化

引言

如何将疫情数据呈现在时空关系之中，并让大众能够看得清楚、看得懂并愿意信赖，是新冠肺炎疫情暴发期间最亟需解决的信息传播问题，无论主流媒体还是可视化领域，在这个方面都没有成熟经验。

哪一种可视化的方法才最有效呢？通过对现有所有媒介播报的疫情数据新闻的收集、整理和研究，并不能回答这个问题。因为几乎所有的可视化成果都是类

似的，不是地图就是折线图，要不就是以上两种形式的组合或变异。就连当前日本、美国等对本国疫情的数据新闻也大多是如此，鲜有更多模样。从视觉化的有效性而言，这些成果有以下这些问题：就地图而言，以红色深浅板块构成省级地图并不具有信息的精度，会让人产生全国所有地区都已经出现感染病例的误解，也无法锁定病情严重区域；同时行政区域的边界会让人们误读疾病传播的空间特点——其传播并不会在人为的国界、省界戛然而止。其次，用传统的折线图来表达疫情数据中死亡病例、新增病例、疑似病例这三个具有包涵属性的维度，则不能满足人们对深度信息的需求，即发现更多的相关性。

传染病信息的独特性在于时空性，其中，每日新增病例是疫情中最重要的指标，它的变化能够准确地描述病情的暴发、增长、僵持、缩减和长尾等状态，甚至让读者在观察中快速发现事件焦点。因此，我们认为以时间线为骨骼分布每日新增病例数据，可以将传染病数据视觉化为一种具有流态的形象，应对人们对疫情信息认知的体验。

一、相关技术背景

具有流动感的图形源自堆叠图，为了理解流动形式与机能，本文使用“流图”来定义这一类堆叠图，并在下文统称为“流图”。

（一）堆叠图时代

早在19世纪，可视化实践者威廉·普莱费尔(William Playfair)就使用了一种近似今天人们所见到的流图的形式，他在时间线上堆叠代表不同类别信息的层，这种形式被称为“堆叠图”(Stackedgraph)。

例如，普莱费尔所做的英国对丹麦和挪威的进出口图(图1:1)，时间线上设置了两层，一层是英国的出口，另一层是英国的进口。当英国的出口曲线大大高于进口曲线时，这个进出口的差额区域就被标注为“对英国有利的平衡”。^①读者对这种图形的理解还是有些费力，因为有些堆叠图层的做法会让人们的视觉关注度聚焦在图层与图层的高差区间，当多个图层的宽度接近时就难以辨认；各层必须透明，否则就会出现某些层可能被前面的层完全覆盖的情况；它最根本的特征在于，整体河流的宽长比例会强烈地影响图中曲线的陡缓程度。可能正是由于以上原因，流图在过去的200年间并没有被非常广泛地应用。

^①Edward Tufte, *The Visual Display of Quantitative Information*, Warwickshire: Graphics Press, 2009, p. 92.

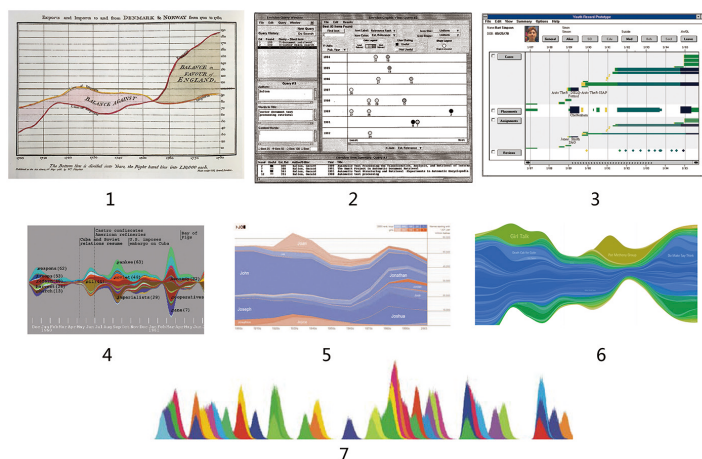


图1 堆叠图案例

注：例1为威廉普莱费尔的英国进出口堆叠图；例2为“展望”项目；例3为“生命线”项目；例4为“主题河流”项目；例5为“名字旅行者”项目；例6为“聆听历史”项目；例7为“模因跟踪”项目。

（二）流动的意象发现

时间一直是最重要的信息维度，如何基于计算机程序将信息按照时间序列呈现的研究从未停止。1990年代后期，弗吉尼亚科技大学开发的“展望”（Envision, 1996）数字图书查询系统，让用户基于时间维度自由查询书籍；^①马里兰大学开发的个人历史可视化项目“生命线”（LifeLines, 1998），将个人健康数据和犯罪记录按照时间线展示。这些项目大多将对象按照时间线分散排列：或从上向下，或从左向右。^②

2000年，第一个将计算技术运用于堆叠图形式的可视化项目“主题河流”，通过创建平滑曲线，以X轴为中心上下对称布局的技术方案，生成了形态优美、信息丰富的堆叠图，看起来如同彩色的水流。^③它针对美国1960年代新闻文本中的高频词进行表达，水流变窄变快表示该层主题在当时的强弱程度，在宽窄变化的曲线中，我们可以快速、简单地识别其中的趋势、关系、异常和逻辑结构。例如，人们可以清晰地看到美国媒体在古巴导弹危机事件期间所使用的新闻高频词及其频率变化，从“石油”转变到“美国佬”“苏联”“帝国主义”，再发展到“美

①L. T. Nowell, R. K. France, D. Hix, L. S. Heath and E. A. Fox, “Visualizing Search Results: Some Alternatives to Query-Document Similarity,” Proceedings of SIGIR'96, Zurich: ACM Press, 1996, pp. 67-75.

②C. Plaisant, B. Milash, A. Rose, S. Widoff and B. Shneiderman, “Lifelines: Visualizing Personal Histories,” Proceedings of CHI'96, Association.

③S. Havre, E. Hetzler, P. Whitney and L. Nowell, “ThemeRiver: Visualizing thematic changes in large document collections,” IEEE Transactions on Visualization and Computer Graphics, vol. 8, no. 1, 2002, pp. 9-20.

国佬”“肯尼迪”“合作”和“甘蔗”。这个过程生动勾勒出新闻中所隐藏的跌宕起伏之历史景象——从对峙、僵持到解决的剧烈变化过程。

笔者认为“主题河流”最大的价值在于其有机的律动形式、流畅的曲线以及对称的结构恰好符合传统的审美标准。“主题河流”的作者也找到了这种审美在心理学理论方面的支持，认为它符合创立于德国的格式塔心理学派关于视觉识别的原则，即“整体大于局部之和”的观点，也就是说人们在感知过程中，不会依靠单个、细节的元素，总会自然而然地追求结构的整体性，先“看见”一个图像的整体，然后才“看见”整体中的各个部分。

（三）流动的审美环境

2000年之后，随着网络技术的快速发展，堆叠图走出了实验室、专业分析领域，以在线交互的方式获得公众的欢迎。马丁·瓦滕伯格（Martin Wattenberg）开发了“名字旅行者”（NameVoyager，2005），^①让人们可以在时间线上观察6,000个人名之间的关系，并快速发现不同时代热门的婴儿名字。“名字旅行者”的画面用粉蓝色代表男性、粉红色代表女性，简单的信息结构和流畅的曲线视觉上极具好感度，程序发布两周之后就吸引了超过50万访问量。它针对大众的共同兴趣提供了独特的视野，激发每个人在数据观察中不断提出问题：我的名字非常80年代吗？当年最流行的名字是什么？我朋友的名字如何？一年后，瓦滕伯格在论文《为社会数据分析而设计》中提出，信息可视化不仅是一个工具，而是在线社会环境的一部分。^②

李·拜伦（Lee Byron）在瓦滕伯格之后权衡审美与认知的需求，调整了算法和设计策略，生成了极具音乐气质的彩色图表，并将自己开发的新型叠加图称为“流图”（Streamgraph）。拜伦认为可视化应该动态地权衡美学和功能的优先关系，若没有必然的阅读对象，可视化必须考虑审美以吸引观看。“美学上的愉悦、甚至是迷人的品质，可能与易读性相冲突。《纽约时报》的票房图表看起来并不像一个标准的统计图表，这却很可能是其吸引注意力的原因。但是，如果图形普遍使用，它的吸引力真的会减少吗？另一方面，如果这种吸引力更持久，也许是一个很好的例子，说明赏心悦目的外观会吸引读者深入探索。”^③

“主题河流”“名字探索者”和“聆听历史”项目激发了大量研究者对流图的

①NameVoyager (2005), <http://babynamewizard.com/namevoyager/>.

②M. Wattenberg and J. Kriss, "Designing for social data analysis," *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, no. 4, 2006, pp. 549-557.

③L. Byron and M. Wattenberg, "Stacked graphs-geometry and aesthetics," *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 6, 2008, pp. 1245-1252.

兴趣,例如针对新闻的分析项目“模因跟踪”(Meme Tracker, 2009)揭示了新闻的传播渠道主要是从主流媒体流向博客,仅仅3.5%的博客内容可能逆流。^①从堆叠图到流图,人们对这种可视化形式的理解也如流图一般高低起伏,从低认知度、高美感发展到权衡思考,这种趋势倒并不令人惊异,反倒与设计艺术学相一致,因为设计活动本身就是一项平衡功能与审美的作品。

二、可视化设计

(一) 疫情数据

在疫情暴发初期,官方数据源并不充足。本文第二作者从2020年1月23日开始手动下载、整理和规范数据,当时微信公众号“丁香医生”发布的官方新闻是一个重要信息来源,我们1月10日至26日的数据均来自于此;1月27日起,腾讯新闻开始与我们合作,我们1月27日至29日的数据采集于腾讯新闻网页;1月30日起腾讯新闻开始向我们提供数据接口,从此“疫情的流图”每日可以自动获得更新的官方数据。需要说明的是,腾讯新闻的数据来自不同渠道,境内疫情数据来源为“国家及各地卫健委”,而境外部分则是“各国(地区)官方通报和权威媒体报道”。截至2020年3月7日中国国内地点信息也有一些特殊性:直辖市与省关系并行,有些单列单位,如新疆建设兵团、监狱;海外部分数据按国家或地区分类,并未精确到下辖地区,其中包含并非地理名称的“钻石公主号邮轮”。

(二) 形态的选择

为快速开发出“疫情的流图”,我们采用了由拜伦和瓦滕伯格所贡献的代码,^②然后用疫情发生以来的数据进行比较和观察(图2)。

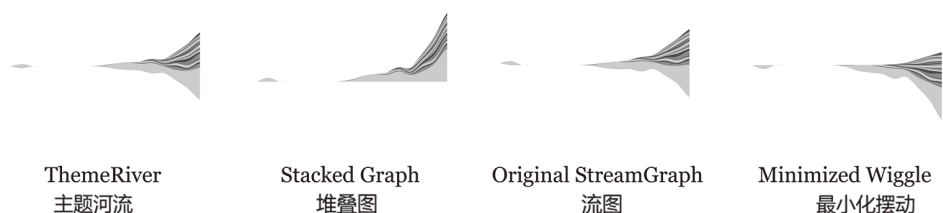


图2 四种“堆叠图”形式的疫情可视化比较

①Jure Leskovec, Lars Backstrom and Jon Kleinberg, “Meme-tracking and the dynamics of the news cycle,” In Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD’09), Association for Computing Machinery, New York, NY, 2009, pp. 497-506. DOI: <https://doi.org/10.1145/1557019.1557077>.

②<http://github.com/leebyron/streamgraph>.

通过对比,具有流动感的“主题河流”“流图”看起来更像是一条自然流淌的河流,虽然这四种形式都源自叠加图的核心构造方式。“主题河流”的外轮廓和“流图”近似,但前者在X轴的左端起点是水平对称的,相对于后者歪扭的起点,似乎更庄重地暗示了疫情的登场。经历38天之后,为了观察表现性,我们再次比较了以上四种流图的形态(图3),看起来“主题河流”和“流图”非常近似,“最小化摆动”反倒呈现出了一种意外的效果:湖北省以外的区域自成一个别样的外轮廓,并没有跟随湖北省的摆动而失去个性。

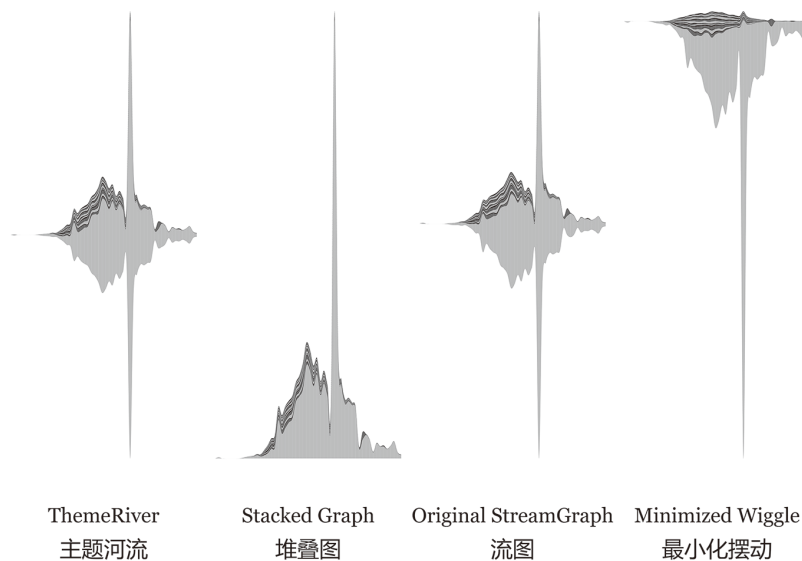


图3 截至2020年3月3日的堆叠图方式比较

虽然雅克·贝尔坦(Jacques Bertin)早就指出过堆叠图长宽比例的问题,^①我们却既不能决定整个图的长度(代表疫情的长短);也无法决定河流的最大宽度(代表每日新增病例的最大数量),被动地每日手动调整的方式成了唯一选择。我们选择以清晰易懂为目的,外轮廓曲线不因过分舒缓而淡化严重性,也不过分陡增而激发人们的惊慌。手动调整宽度的方式在疫情猛烈暴发阶段仿佛是一种艺术性安慰手法,虽然调整的幅度非常微小。2月13日湖北省出现了一次当日新增病例14,823例的陡增波动,代表其他省的流层高度已经显得微乎其微。为了保持与2月13日之前的流图在比例上的一致,我们停止了比例微调。

^①M. Wattenberg and J. Kriss, “Designing for social data analysis,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, No. 4, 2006, pp. 549-557.

（三）色彩决定

大量的流图案例中，可视化者们都使用鲜艳的、随机的色彩定义每个流层以示区别。唯有“聆听历史”项目建立了一个二维矩阵，用以映射音乐家出现的时间及其流行度，这几乎是所有流图项目中最具色彩系统性思考的作品。但“出现时间”的维度对疫情来说并不适当，所有省份的疫情数据几乎同时出现，这也恰是所谓暴发性事件的重要视觉特征。我们在项目初期使用了红紫色调进行主观表达（图4），但从色彩学的基本原理考虑，不同色彩因其固有色相、明度的差异会导致视觉认知性上的强弱效果，例如人们总是会更加注意那些最鲜艳、明亮的色彩，为了避免这些干扰，最后我们把“疫情的流图”统一调为灰色系。

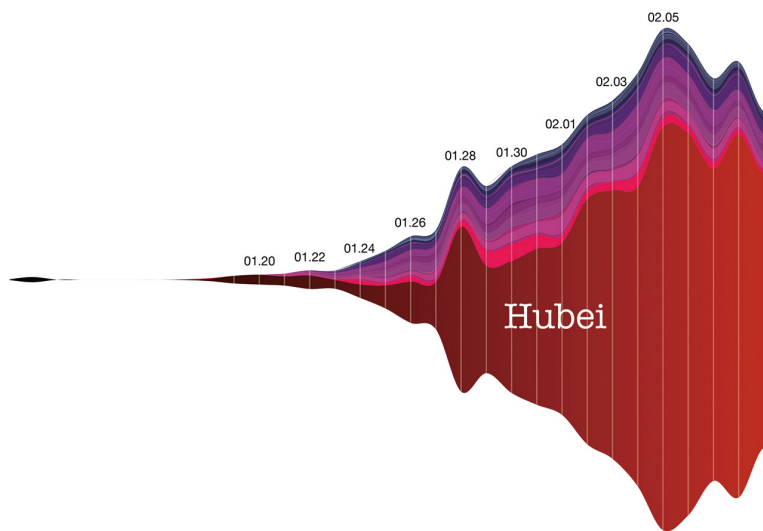


图4 疫情的流图（2020年2月9日）

（四）多层的信息结构

在疫情中，湖北省的新增病例占绝大多数，当其他省份的数据与其并置时便显得不容易“阅读”。因此，我们决定分别呈现中国全国和中国除湖北省外地区两种流图，之后又增加了全球、中国之外全球等两种流图（图5），共计四种类别，以弥补无法深度观察的问题。

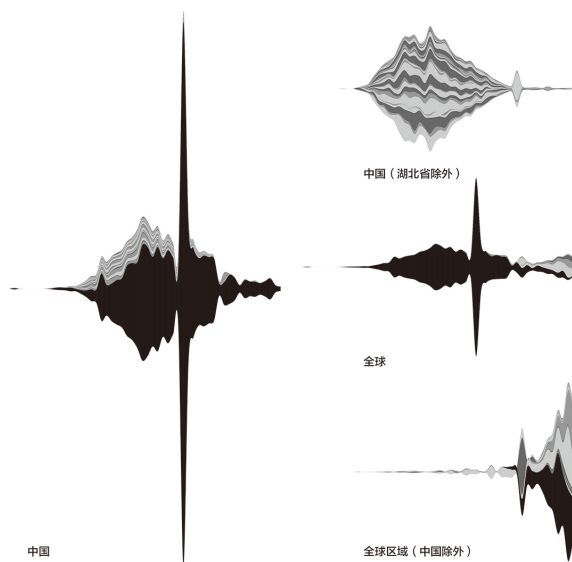


图5 区域的可视化分类方式(数据截至2020年3月3日)

(五) 动画与交互

当用户打开“疫情的流图”时,^①会看到一条自动从左向右流淌的流图(图6),流图的最右端就是当天的新增病例总量,这种逐渐揭开展趋势的方式让用户的视觉体验更加生动。在动画停止之后,用户可以自主拉动时间滑条观看;用户也可以点击流图中的任何一层,激活一个弹出方窗了解该层所代表的省份目前有多少总病例数、当日新增病例等信息。



图6 “疫情的流图”的界面与使用场景

^①<https://zeelab.cn/WuhanThemeRiver/>.

三、用户反馈

（一）流量与评论

“疫情的流图”一直以三种形式发表：微信朋友圈、公众号和独立网页，内容因这三种媒介的特性而不同。在2020年1月28日至3月5日期间，17篇公众号文章获得超过18万次阅读；独立网页得到了29,223独立IP访问，访问者中90%来自中国，重复访问者占10%，手机用户占87%。微信朋友圈也留下了约500条评论。在整个疫情发展过程中，用户对每日发布的流图评论涉及担忧、猜测、调侃等情绪表达，被我们视为“疫情可视化”过程中的宝贵部分。

（二）用户评价

正如我们期待的，人们很愿意看到“疫情的流图”。评论中有不少内容都充满了对未来的猜测，例如：“超级期待早日变成菱形”“等到变成菱形就能出门”“我想问终点将会在哪里呢”。当我们暂时中断发布，人们会主动私下发来信息提醒；恢复之后，人们的评论是：“等了好几天了，这样的数据图看起来乐观”“终于更新了”。当然也有不少人留下建议，例如：“我能说这个图很美吗？虽然现实如此残酷和严峻”“可否做一个与SARS的对比”。从数字媒体艺术的角度来观察，笔者认为最有意思的在于人们在观察“疫情的流图”时的猜测，他们会想象当疫情结束的时候这张图会像个瓶子、坛子、天坛、帽子、喇叭花、蝙蝠、鱼、跳蚤、苍蝇……这些反馈并不能成为定量判断的标准，但是如此充满形象感的词汇，折射出可视化作为信息媒介的意义。

媒体学家马诺维奇曾经直言不讳地指出：即便我们可以在可视化网站上看到上百种可视化形式，但是如果按照可视化方法进行分类，只有很少的几个类别。^①大众会越来越对这种地图、折线图产生厌倦，有限的可视化形式尤其需要精心挑选、巧妙扩展以准确表达信息。“疫情的流图”，就是在寻找通用的可视化套路之外的解决方式，用户的反馈证明它赋予了一种让人们可以深度阅读数据信息的新颖体验，这一定是接近10%的用户愿意多次返回观看的原因。

结 论

“疫情的流图”作为一个短时间内运用开源代码开发的可视化工具，让用户

^①Lev Manovich, “Museum Without Walls, Art History Without Names: Visualization Methods for Humanities and Media Studies” (2012), <http://manovich.net/index.php/projects/museum-without-walls-art-historywithout-names-visualization-methods-for-humanities-and-media-studies>.

每天都能观察到实时新冠疫情变化。它被人们转发、推荐，成了在主流媒体之外的一种独立媒介。

从数据可视化应用的角度来说，我们发现流图的外轮廓特别适合表现疫情暴发突然、曲折发展的特点，并在疫情过程中具有强烈感染力——吸引人们主动猜测未来的走向。针对本次新冠疫情中绝大部分病人集中于湖北地区，导致其他地区的信息并不容易被识别和观察的情况，我们通过点击交互、时间滑条、地区分级等可视化方式进行了弥补。最终的结论是：流图在表现重大、突发疫情方面具有鲜明的视觉表现力，但是它并不适合表达地区等分级概念。

根据数据特性、认知需求，在文献中寻找适合的视觉形式与开源技术——在用户反馈中逐渐迭代——快速开发，这样的“疫情的流图”的路径，可能是在模板范式与信息特性之间找到的一种可视化创新方式。

（感谢开发团队的成员清华大学美术学院交叉学科研究生任祚承同学、江南大学数字媒体专业王瑞同学、清华大学美术学院马菡璐同学、清华大学校友姜慎薇。感谢腾讯新闻魏传举先生促成了腾讯新闻对我们的数据支持；感谢清华大学彭刚老师的支持，令本项目获得有力的媒介推广，让我们积累了有利于技术优化的大量用户反馈；感谢广州美术学院钱磊老师精心校对。）

CaseFlow: Understanding Stream Graph in Visualization

Xiang Fan, Zhu Shunshan

Abstract: The map and line charts are seemingly the universal approaches to communicate in a complex situation and have become ubiquitous in the mainstream media to express the COVID-19 outbreak. However, these conventional methods are less appealing to panic-stricken people who are eager to understand more from the epidemic data. Based on the ThemeRiver visualization, we developed a mobile visualization applet, CaseFlow, to enrich people's perspective—the space and time distributions of confirmed the daily new COVID-19 cases. This paper describes the literature surveying and design decisions in the process of the development and discusses the reaction from the internet users. We identify the stream graph is well suitable for sketching the contour of the overall quantitative changes, observe the demerit in detail information display and suggest approaches such as interaction, animation, and space classification to make up for the deficiency of the Theme River Layout. Finally, we share our thoughts on the comments from our users. Throughout the paper, we emphasize the importance of visual thinking in applying stereotyped forms of visualization.

Keywords: Stream Graph; Stackedgraph; COVID-19; Data Visualization

（编辑：封帆）