Advances in Informatics, Management and Technology in Healthcare
J. Mantas et al. (Eds.)

doi:10.3233/SHTI220680

Towards the Adoption of Novel Visualizations in Public Health

Simon R. MINSHALL[a,1], Helen MONKMAN[a], Andre KUSHNIRUK[a] and Luca CALZONI[b]
[a] *School of Health Information Science, University of Victoria, Canada*
[b] *National Institute on Minority Health and Health Disparities (NIMHD), National Institutes of Health, Bethesda, Maryland, USA*

Abstract. Visualizations form an important part of public health informatics (PHI) communications. Visualizing data facilitates discussion, aids understanding, makes patterns apparent, promotes analysis, and fosters recall. How rare are novel visualizations in the PHI literature? In Phase 1, we used a rapid review methodology to test the commonness of the Sankey diagram in the PHI theory literature via an automated text search for key terms. In Phase 2, we prototype an uncommon chart type. A total of 27 relvant papers were searched and a computer-generated Sankey diagram was prototyped. PHI professionals have access to visualization tools emerging from social media and niche systems. PHI literature underutilizes uncommon visualizations requiring programming expertise. The authors advocate for: multi-disciplinary teamwork, technical education, the use of open visualization tools, and further adoption of visualization for public health professionals.

Keywords. data visualization, public health informatics, expertise, teams

1. Introduction

Visualizations form an important part of our cognitive processes because the human brain absorbs more through vision than through all other senses combined [1]. Interpreting a complex visualization is accomplished with vast amounts of information processing, a feat most humans perform unconsciously [1]. The interpretation of visual data occurs during day-to-day activity (e.g., pedestrians avoiding collisions in crowds).

In public health informatics (PHI), visualizations can emerge as artifacts of data analytics. Visualizing data facilitates discussion, aids understanding, promotes pattern recognition, assists analysis, and fosters recall [2]. Visualizations serve as tools of communication to describe complex datasets and can bridge medical professionals, civil policy experts and members of the public [3]. PHI professionals have access to tools used to construct novel visualizations such as choropleth maps, trees, and graphs [4]. These tools tend to be created for siloed applications and used by experts in other fields, which make them difficult to adapt to other disciplines [5]. Because a paucity of literature exists on visualization in PHI, the adoption of these tools, to date, has not been widespread [4].

In this paper, we explore how Sankey diagrams [6] —alluvial representations of proportional flows (e.g., matter, patients) as graphical tributaries—chosen subjectively for its novel appearance and its potential to visualize patient journey mapping and population flow. Sankey diagrams suit aggregate data rather than data for a single entity.

[1] Corresponding Author, Simon Minshall, School of Health Information Science, University of Victoria, British Columbia, Canada; E-mail: sminshall@uvic.ca.

Specifically, we ask the following: How commonly are Sankey diagrams reported in PHI literature? How challenging are they to construct and apply to PHI? How might we better utilize more complex modes of diagraming for PHI?

2. Methods

2.1. Phase 1 – Rapid Scoping Review

A challenge exists in rapidly assessing a body of literature for graphical content as their proper names vary (i.e., “pie chart” or “map”) and are seldom explicitly mentioned. To surmount the challenge, we used a rapid review methodology to test the commonness of the Sankey diagram in the PHI literature by examining visualization theory papers and performing an automated text search for key terms. The PubMed database was searched with the query: ‘("Public health") AND ("visualization") AND (theory)’ between 2017–2022, including the quotation marks, and with no additional filters. Data analysis was performed by systematically searching the selected PDF files for the specific names of a subset of those visualizations listed in Chistie [4]. The case-insensitive whole-word search was performed using the *pdfgrep* utility [7].

2.2. Phase 2 – Prototyping the Approach

To explore the difficulty of creating a Sankey diagram we constructed a scenario mapping the flow of patients on their journeys, having self-tested positively for COVID symptoms. The diagram, shown in Figure 1, describes aggregate patient journeys and visualizes their various stages of hospital treatment. Origins are on the left; destinations are on the right. The pipe-like links between stages represent patients’ progression from the initial referral on the left-most vertical bar to the final outcomes on the right-most bars. The demonstration of a Sankey diagram, shown in Figure 1, was created using the Data-Driven Documents (D3) [8] software toolkit and a small fabricated dataset.

Table 1. Showing the frequency of the terms listed found in papers and the total occurences within the text of all papers.

Viz. name	Papers	Within papers	… continued		
“graph”	18	273	“line graph”	1	7
“map”	14	70	“bar chart”	1	19
“network”	12	132	“pie chart”	1	15
“chart”	7	57	“flow chart”	1	15
“pleth”	2	2	“Sankey”	0	0

3. Results

3.1. Phase 1 – Rapid Scoping Review Results

A total of 27 papers were identified. On examining titles and abstracts, we rejected papers which were concerned with: scientific visualization (n=1), biomedical subjects (n=5), or were inaccessible (n=3). A total of 18 papers remained for analysis. The frequency of names for specific visualization methods referenced in papers and their frequency within papers is shown in Table 1. For example, the term "line graph" occurred in one paper and was repeated seven times within that paper. More specific terms such as "bar chart" are subsumed in the simpler terms such as "chart." The top three most common terms for visualization mentioned in the retrieved papers were: (a) graph, (b) map and (c) network. The least common were the compound phrases "pie chart" and "flow chart," and the term "Sankey" occurred zero times.

3.2. Phase 2 – Prototyping Results

It was still challenging to create a Sankey diagram once we produced the dataset. Knowledge of software development tools such as the programmer's notebook [8] and of object-oriented development techniques were required to adapt and customize an example Sankey diagram, shown in Figure 1, to utilize our own dataset.

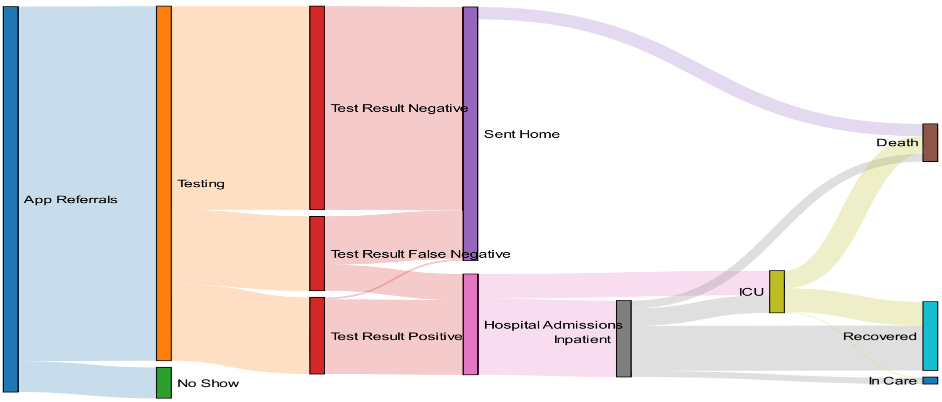

Figure 1. A computer-generated Sankey diagram created using the Observable notebook [8].

4. Discussion

Graphs and bar charts are featured much more frequently in the PHI literature than are chloropleths or Sankey diagrams. This is unsurprising given that commonly-used office software facilitates only the construction of basic visualizations (e.g., scatter plots, bar graphs) but not more unique or specific-purpose ones (e.g., choropleth, Sankey). Currently, creating more complex visualizations of data requires a programmatic skillset. For example, constructing Sankeys requires the complicated rearrangement of data queries to supply the algorithms for rendering a diagram. Once made, Sankey diagrams are visually compelling and seem well suited to PHI. For example, they may be used to explain the flow of pandemic patients within a community and could be animated to visualize patterns changing in interactive displays.

There are limitations to this rapid review. For example, the terms chosen are not exhaustive. Many visualization types have synonomous names, and most of the terms searched for are synonyms for items not considered to be visualizations (e.g., flow is not related to fluid dynamics). Despite these limitations, this work shows the underuse of some visualizations, likely due to challenges associated with their creation.

Software packages offer visualization toolkits that can meet many PHI professionals' requirements; however, users are limited to the set of configurable options.

Difficulty increases when a dataset calls for repurposing an uncommon visualization which is beyond its intended use case. The transition from pre-packaged to bespoke visualization requires skill. Examples of this include Koch and Denike's 2009 reconstruction of John Snow's Cholera visualization [9], and disaster data overlaid on existing maps [10].

The abstraction offered by visualization packages shields the user from the complexities of programming and data wrangling; however, the trade-off limits the visualizations one may create. Creating uncommon visualizations, such as choropleths and isopleths, and network graphs [5], requires programming skill, data manipulation and knowledge of visualization libraries. Software libraries such as those offered by D3, can be used to construct representations with relative ease for a skilled practitioner [8]. Since its inception in 2011, software developers and graphic designers have used D3 to create visualizations often published in the press and interactively online [11].

The barrier to greater adoption of novel and uncommon visualization remains. Collaboration is now easier thanks to the increased availability of computational notebooks and toolkits which enable the intermediately skilled researcher to begin constructing complex visualizations, thanks to the availability of software libraries. Three skillsets combined are required to produce uncommon or novel visualizations: public health expertise, informatics data-wrangling, and computer programming.

The collection of new kinds of data may stimulate the requirement for novel computational visualizations, such as the ones described in this paper, requiring specific skills. To facilitate new visualizations, we advocate for: (a) the formation of multi-disciplinary teams, (b) the requisite skills to be taught at universities, (c) the use of open access and customizable tools, such as D3, for creating visualizations, and (d) continued efforts by PHI researchers to adopt complex visual artifacts.

References

[1] Ware C. Information visualization: perception for design. Morgan Kaufmann; 2019 Dec 19.

[2] McElroy PD, Rothenberg RB, Varghese R, Woodruff R, Minns GO, Muth SQ, Lambert LA, Ridzon R. A network-informed approach to investigating a tuberculosis outbreak: implications for enhancing contact investigations. The International Journal of Tuberculosis and Lung Disease. 2003 Dec 1;7(12):S486-93.

[3] Rohwer A, Taylor M, Ryan R, Garner P, Oliver S. Enhancing Public Health Systematic Reviews With Diagram Visualization. Am J Public Health. 2021 Jun;111(6):1029-1034.

[4] Chishtie JA, Marchand JS, Turcotte LA, Bielska IA, Babineau J, Cepoiu-Martin M, Irvine M, Munce S, Abudiab S, Bjelica M, Hossain S, Imran M, Jeji T, Jaglal S. Visual Analytic Tools and Techniques in Population Health and Health Services Research: Scoping Review. J Med Internet Res. 2020 Dec 3;22(12):e17892. doi: 10.2196/17892. PMID: 33270029; PMCID: PMC7716797..

[5] Carroll LN, Au AP, Detwiler LT, Fu TC, Painter IS, Abernethy NF. Visualization and analytics tools for infectious disease epidemiology: a systematic review. J Biomed Inform. 2014 Oct;51:287-98. doi: 10.1016/j.jbi.2014.04.006. Epub 2014 Apr 16. PMID: 24747356; PMCID: PMC5734643..

[6] Sankey HR. The Thermal Efficiency of Steam Engines, *Minutes Proc. Inst. Civ. Eng.* **134** (1898) 278–312. doi:10.1680/imotp.1898.19100.

[7] H.-P. Deifel, pdfgrep, 2021. https://pdfgrep.org/ (accessed March 12, 2022).

[8] Bostock M, Meckfessel M. Observable, (2021). https://observablehq.com (accessed July 14, 2021).

[9] Koch T, Denike K. Crediting his critics' concerns: remaking John Snow's map of Broad Street cholera, 1854. Soc Sci Med. 2009 Oct;69(8):1246-51. doi: 10.1016/j.socscimed.2009.07.046. Epub 2009 Aug 27. PMID: 19716638..

[10] Nacenta M, Hinrichs U, Carpendale S. FatFonts, in: Proc. Int. Work. Conf. Adv. Vis. Interfaces, Association for Computing Machinery, New York, NY, USA, 2012: pp. 407–414.

[11] Bostock M, Carter S, Tse A. Is It Better to Rent or Buy?, *N. Y. Times*. (2014). https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html (accessed July 16, 2021).