Data Visualization and Healthcare: Opportunities and Challenges When Exploring Electronic Health Records

Davey Chafe

York University

In Partial Fulfillment of Psychology of Data Visualization: 6135

Submitted to Dr. Michael Friendly

May 6, 2019

Data Visualization and Healthcare: Opportunities and Challenges When Exploring Electronic Health Records

Brief history

The 21st century is without a doubt the "Digital Age" in which anyone with an internet connection has nearly instant access to an ever-expanding body of data. Gregson, Brownlee, Playforth, and Bimble (2015) state that nearly 90% of the worlds stored data was collected in the preceding two years. The vast expanse of data in recent years has allowed government and researchers to analyze, track, and navigate information like never before, leading to more informed decisions around public policy, economics, and of particular relevance here, healthcare information (SciDev, 2016). However, with the massive amounts of data now circulating the world, making sense of it in a clear, coherent way can be challenging. One way of making large, incomprehensible sets of data more manageable, is through data visualization. The graphical display of data can be traced back to William Playfair in the 18th century. He is credited with the first use of visualizations like line graphs, pie charts, and bar graphs. Playfair believed that graphs were a better way of displaying information that were more understandable and memorable than raw data (Spence, 2006).

Data visualization in healthcare also has a long history. One of the earliest data visualizations related to healthcare is John Snow's map of 1854 London which helped identify the source of a cholera outbreak (Figure 1 – retrieved from Gotz & Borland, 2016). Another early visualization related to medical data is that of Florence Nightingale's coxcomb chart that represented the causes of death from the Crimean war from 1854 – 1856. This visualization showed that mortality rates were largely due to lack of sanitation in the hospital wards, rather than injuries themselves that were sustained (Figure 2 – retrieved from West, Borland, & Hammond, 2014).

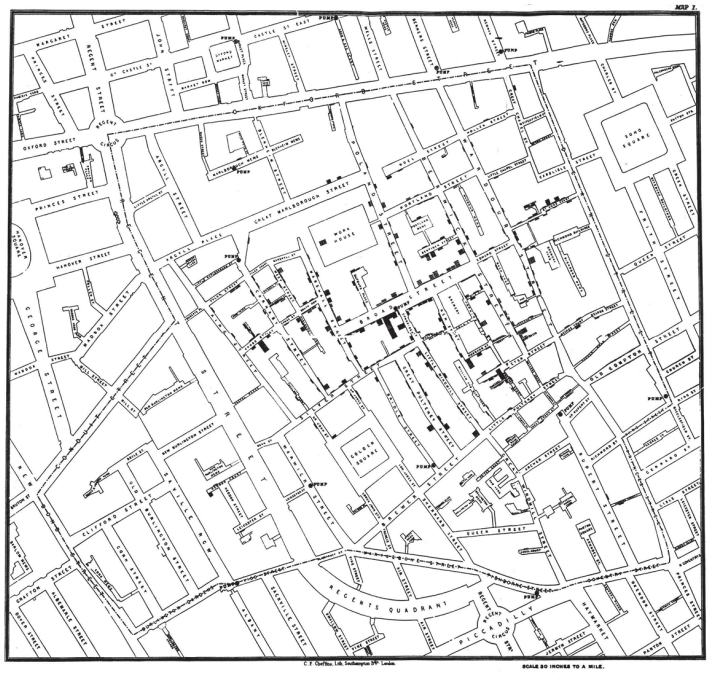

Figure 1. John Snow's map of the cholera outbreak in London, 1854 (above).

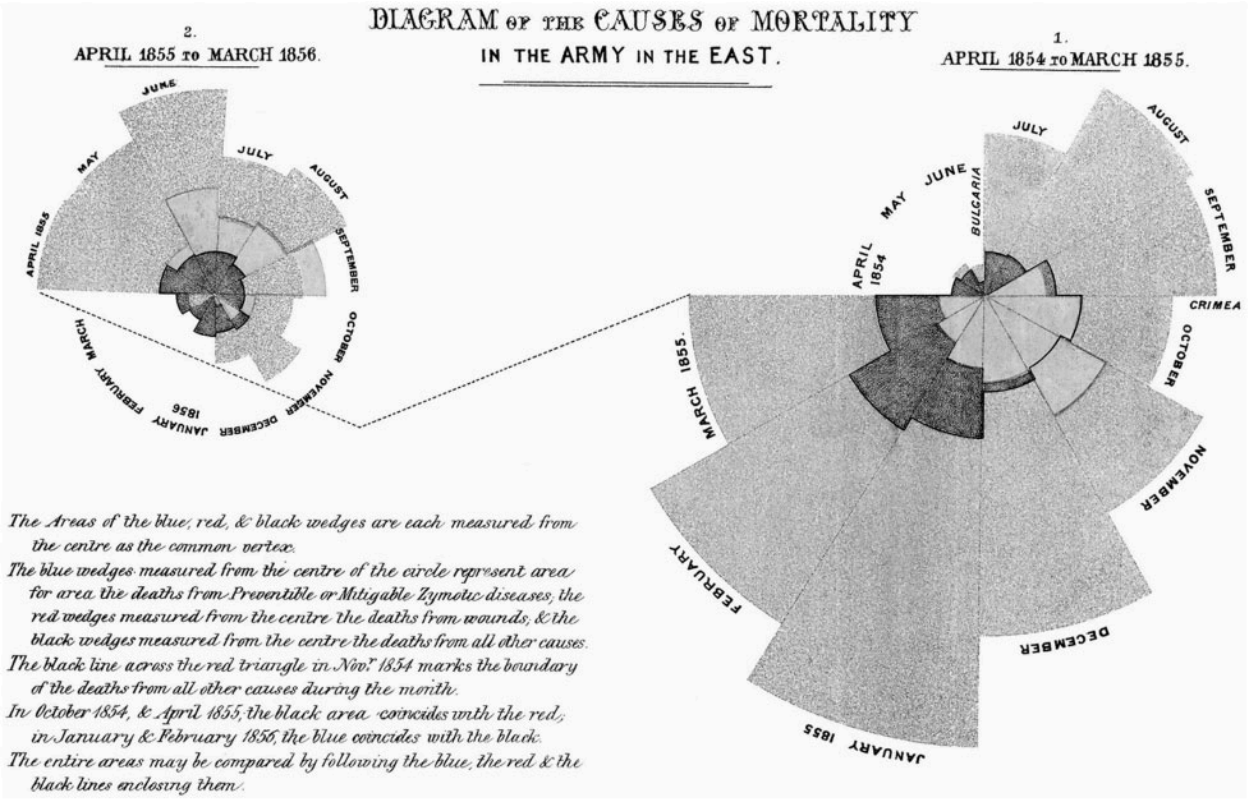

Figure 2: Florence Nightingale's chart of Crimean war mortality between 1854-1856.

These early visual representations of crucial medical data highlight the importance of displaying data in a clear, understandable way. Below I will introduce several current methods that are utilized to visually represent the ever-increasing amount of electronic health records (EHR) in hopes of informing decision making in healthcare and improving access to health-related services. I will conclude by highlighting some of the current challenges in representing healthcare information visually.

Techniques Used

Given the scale and complexity of healthcare information that is now captured electronically, sophisticated data visualization programs are needed to collect and transform the data in such a way that make it understandable and meaningful to researchers, clinicians, and patients alike. A review paper from West, Borland, and Hammond (2014) investigated innovative strategies for data visualization in healthcare settings and described several popular methods that were often cited in the research as being advanced applications for visualizing a variety of patient data. An early application from the 1990's, and one which has received considerable attention is LifeLine (Plaisant et al., 2003). Lifeline uses electronic patient data to create a temporal map of a single patient's health related events. This method was eventually adapted to LifeLine2, which allowed for further refinement of records by looking at numerical and categorical patient data, which allows users to "drill down" into important details concerning their EHRs. Another approach by Gotz and Wongsuphasawat (2012) called Outflow was designed to look at disease progression by offering visual displays outlining events, outcomes, and sequences of certain diseases that allow for individuals to quickly analyze and identify factors related to the progression of the illness.

West and his colleagues identified several other common programs used throughout the literature. These visualization programs all aim to encompass the complex data in EHRs. EHRs

are often longitudinal in nature and try to consider additional factors such as prescribed medications, frequency of visits, date of diagnoses, tests administered, and the like. While there are many factors to consider when analyzing and visualization patient medical data, doing this effectively can vastly improve client care.

Improving Healthcare With Visualizations

The medical field has various technologies that allow for the visualization and interpretation of medical information. Ranging from patient vital signs displayed as lines that allow for quick, easy interpretation to MRI's that provide a wealth of information about potential patient problems. In line with this, advancements in visualizing and understanding health databases and patient information are allowing for more informative care at the level of patient, clinician, and policy makers that allow for broader healthcare related decision-making (Shortliffe & Cimino, 2013).

With the increase of technology available to obtain patient information, there have been useful frameworks developed to manage this information. Hesse et al. (2010) provided the framework called "Health 2.0" which refers to modern strategies of EHR collection that are web based, participatory, and mobile. Strategies under Health 2.0 include things like utilizing social media, smartphones, personal health sensors like fitbits, and visual analytic methods. Hesse and colleagues mentioned three domains under the umbrella of Health 2.0: Personal health information, clinical health information, and public health information.

Personal Health Information

Personal health information refers to individuals who collect data on their own health and lifestyle. As mentioned above, this may include things like body monitoring equipment that allow individuals to better understand their own physical and mental health. These sort of "patient led" health programs may include things like exercise and nutritional programs that

empower people to take initiative and responsibility over their own health behaviours (Shneiderman, Plaisant, & Hesse, 2013). This data can then be shared and visualized with resources like the "PatientsLikeMe" website (see figure 3). This site provides visual patient histories that are informative and easy to understand for non-medical professionals.

Figure 3. The PatientLikeMe website that visualizes thousands of patient's information.

Clinical Health Information

Although patients have ever increasing resources to analyze their own health information, there is a vast amount of information captured by EHR that require specialized knowledge from clinicians to properly utilize and interpret the data. In the past, clinical trials were one of the only ways to conduct sound clinical research. However, the large stores of electronic patient data now provide new opportunities to reveal critical information about patient care. With the help of visualizations, researchers can now gain valuable information with respect to temporal trends,

patterns, and interactions that may go otherwise unnoticed (Wongsuphasawat et al., 2011). In addition, these data can help identify cohorts of patients for future clinical trials (Shneiderman, Plaisant, & Hesse, 2013).

Public Health Information

On a broader scale, public health information has a vast amount of data collected by organizations such as the National Center for Health Statistics, Centers for Disease Control, The World Health Organization, and Census data. Although there is a plethora of information in public health records, navigating these data sets are challenging for many reasons. The data collected often range from representing a single individual, to entire cities, and range in time, from minutes to years. Given the heterogeneous nature of the data, it can often be difficult to find clear patterns, even where some may exist (Shneiderman, Plaisant, & Hesse, 2013). Despite these complexities, there are visualizations such as the The Community Health Map that provides access to health data which is relatively easy to navigate. Public data of this nature can be analyzed visually and used to make strong cases to inform healthcare policy and procedure by showing historic trends and future trajectories (Sopan et al., 2012).

In line with this, social media and public health information provide unique opportunities for data visualization. Research from SciDev.Net showed that scientific articles containing data visualizations received 942% more comments on Twitter than did articles without data visualizations (see figure 4). This shows the impact of visualizations on public uptake of scientific data. Not only are visualizations useful for engaging the public with research, but we now also have the tools necessary to record and visualize data obtained from social media. For example, Hansen et al. (2010) analyzed keywords used in hashtags by region on Twitter. In doing so, they were able to cluster communities of discussion and identify popular discussion

topics based on keywords like “children” and “vaccination” and their relation to diseases such as polio.

Figure 4. Increased comment activity for scientific articles posted which included visualizations.

Decision Making and Visualizations

As has been discussed, the large quantities of data that are collected by medical organizations and public health institutes are a valuable resource to be utilized to enhance the quality of client healthcare. These data also need to be considered when making strategic healthcare plans that affect individuals, communities, and entire regions. Data visualization is uniquely suited for “big data” that is collected in health-related settings as it can take large, complex sets of data and make the message of the data clear. Therefore, data visualization is vital to informing and supporting decision makers in the medical field. A study by Lavrac et al. (2006) investigated the effects of data visualization on decision support for public health care resources. The authors found that using a certain statistical method, they were able to create the most effective and accurate visualization possible by mapping the availability of health services with the access of those services based on geographic location in a clear, easy to interpret visualization. This was well-accepted by the health-care professionals as it gave them a clear interpretation of the data, and using the visualization highlighted the communities that needed improvements with respect to their access to medical care. According to the authors, further expert analysis was motivated based on this new multi-criteria visualization. This research highlights that data visualization not only benefits the individual health care practitioner, but can help enact change in policy and procedure that benefits society as a whole.

Challenges and Recommendations

Although the role of data visualization in healthcare has provided new insights and access to understandable information as never before, it is not without its limitations and challenges. First, the amount, the size, and complexity of EHR data is a challenge. With so much data, it can be difficult to identify meaningful information, even in the best visualizations. Interactive datasets that have tools such as hover options, zoom, highlight, and other interactive features help researchers and clinicians hone in on the relevant information they are interested in. Second, given the amount of data collected in EHRs and the nature of longitudinal data, these data sets are complicated by missing values, heterogenous data coming together, and inaccurate or incomplete data entry (West, Borland, & Hammond, 2014). Although not detrimental, these are all important factors to consider when developing a data visualization. Third, training time to understand and effectively utilize the information in the visualization is crucial. Given that the health care system is generally understaffed and without sufficient resources (Shneiderman, Plaisant, & Hesse, 2013) visualizations displaying patient data must be relatively easy to learn. This not only benefits the time-limited doctors and nurses, but also facilitates learning for paraprofessionals such as family members or assistants that may take over the delivery of care. Lastly, there must be careful attention paid to statistical accuracy when creating visualizations related to healthcare. Conclusions based on information presented to doctors and other healthcare professionals must be accurate and interpreted correctly. Whether it is the medication dosage of a single patient, or a conclusion drawn about an entire medical population, mistakes in computation or interpretation can have life-or-death consequences. Therefore, statistical rigor is of the utmost importance.

Conclusion

Data visualization has been utilized in medical and healthcare related fields for nearly two centuries. With the rise of the "digital age," electronic health records have resulted in a massive amount of medical data being stored. Given this, there have been a multitude of statistical techniques that allow for the sorting, analyzing, and visualization of healthcare data. This provides a unique opportunity to improve access to healthcare information to better understand trends in illness, side effects of medication, and temporal effects of certain health interactions. It can also provide policy makers with clear, tangible information regarding the areas of healthcare that are in most need of assistance or reform. Although there are challenges in successfully utilizing visualizations in healthcare settings, there are a plethora of opportunities for data visualization to enhance and elucidate healthcare policies, procedures, and best practices.

References

Gotz, D., & Borland, D. (2016). Data-driven healthcare: Challenges and opportunities for interactive visualization. *IEEE computer graphics and applications*, *36*(3), 90-96.

Gotz, D., & Wongsuphasawat, K. (2012). Interactive intervention analysis. In *AMIA annual symposium proceedings* (Vol. 2012, p. 274). American Medical Informatics Association

Gregson, J., Brownlee, J. M., Playforth, R., & Bimbe, N. (2015). *The future of knowledge sharing in a digital age: Exploring impacts and policy implications for development* (No. IDS Evidence Report; 125). IDS.

Hesse, B. W., Hansen, D., Finholt, T., Munson, S., Kellogg, W., & Thomas, J. C. Social Participation in Health 2.0. Computer (Long Beach Calif) 2010 Nov 11; 43 (11): 45–52. doi: 10.1109/MC. 2010.326.

Lavrač, N., Bohanec, M., Pur, A., Cestnik, B., Debeljak, M., & Kobler, A. (2007). Data mining and visualization for decision support and modeling of public health-care resources. *Journal of Biomedical Informatics*, *40*(4), 438-447.

NET, S. (2016). Data visualisation: Contributions to evidence-based decisionmaking A SciDev. Net Learning Report.

Plaisant, C., Mushlin, R., Snyder, A., Li, J., Heller, D., & Shneiderman, B. (2003). LifeLines: using visualization to enhance navigation and analysis of patient records. In *The Craft of Information Visualization* (pp. 308-312). Morgan Kaufmann.

Shneiderman, B., Plaisant, C., & Hesse, B. W. (2013). Improving healthcare with interactive visualization. *Computer*, *46*(5), 58-66.

Shortliffe, E. H. and Cimino, J. J. (2013), *Biomedical Informatics: Computer Applications in Healthcare and Biomedicine: 4th Edition,* Springer, New York.

Sopan, A., Noh, A. S. I., Karol, S., Rosenfeld, P., Lee, G., & Shneiderman, B. (2012). Community Health Map: A geospatial and multivariate data visualization tool for public health datasets. *Government Information Quarterly*, *29*(2), 223-234.

Spence, I. (2006). William Playfair and the psychology of graphs. In *Proceedings of the American Statistical Association, Section on Statistical Graphics* (pp. 2426-2436).

West, V. L., Borland, D., & Hammond, W. E. (2014). Innovative information visualization of electronic health record data: a systematic review. *Journal of the American Medical Informatics Association*, *22*(2), 330-339.

Wongsuphasawat, K., Guerra Gómez, J. A., Plaisant, C., Wang, T. D., Taieb-Maimon, M., & Shneiderman, B. (2011, May). LifeFlow: visualizing an overview of event sequences. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 1747-1756). ACM.