Visualization Prototype Development to Demonstrate the Breadth and Depth of Electronic Health Record Data

Ahmad Aljadaan
University of Washington - Biomedical Health Informatics and Medical Education
aljadaan@uw.edu

Ross Lordon
University of Washington - Biomedical Health Informatics and Medical Education
rlordon@uw.edu

Jacob Olsufka
University of Washington - Mechanical Engineering and Applied Mathematics
olsufj@uw.edu

ABSTRACT

In this project we assessed the current state of medical data visualization practices and standards. Limited research has been conducted in regards to medical data visualization techniques, standards, or applications. Subsequently we created prototype data visualizations depicting population level electronic health record data for the UW ITHS Data QUEST team. The Data QUEST team has an immediate need for improved data visualization and interactions on their website. In light of this, we opted to develop three different visualization prototypes. The purpose of each is to allow researchers to quickly and effectively explore the data, while informing potential research questions or projects. In the future, these prototypes will be assessed and refined by the Data QUEST team. Ultimately the visualizations will be incorporated into their redesigned website this fall.

Author Keywords

Data visualization, electronic health record

ACM Classification Keywords

Human-centered computing, applied computing

INTRODUCTION

Since 2004, multiple presidential administrations have passed legislation to improve the quality and efficiency of healthcare. A major focus of the over $19 billion investment have been electronic health record (EHR) systems. This software has become the new standard for the majority of clinical tasks such as medical records, prescription ordering, and requesting lab tests. One of the benefits of using this system is the ability to reuse the data for clinical research. Clinics or patients can be identified for research studies by querying the data (1). An example is the University of Utah's FURTHeR system, which is a tool that queries multiple disparate EHR systems allowing researchers to identify cohorts and then request the corresponding data needed for these patients (2). The University of Washington (UW) Institute of Translational Health Sciences (ITHS) Data QUEST team is offering similar information obtained from clinics located in five different states for researchers to analyze. However, their goal is to give researchers the ability to interact with the data on their website as opposed to querying. The purpose is to allow them to explore the breadth and depth of data available to them in an interactive manner. The Data QUEST team desires to improve the current visualizations, which poorly depict the data. Our motivation was to help them by creating three different interactive prototypes utilizing different methodologies to display the demographics data from the clinics. These prototypes will be formally evaluated by the Data QUEST team, refined, and incorporated into their website redesign in the fall of 2015.

CHI 2009, April 4–9, 2009, Boston, Massachusetts, USA.

RELATED WORK

Foundational EHR Visualization Methods

Within the scientific literature, limited research has been performed to assess data visualization methodologies, standards, or potential applications. In Edward Shortliffe's and Jim Cimino's most recent (4th) edition of their seminal textbook *Biomedical Informatics Computer Applications in Health Care and Biomedicine*, three different data visualization methods for EHRs were outlined. The first visualization technique is called a timeline graph (figure 1). These graphs are generally used to depict longitudinal data such as blood pressure on the y-axis against time on the x-axis.

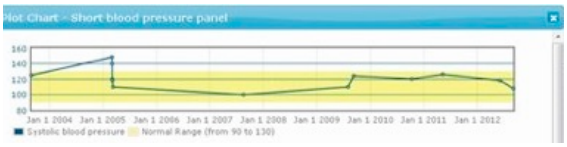

Figure 1. Example of a blood pressure timeline graph.

The second EHR visualization is called a timeline flowsheet (figure 2). While used in various clinical applications, an example timeline flowsheet used by radiologists consists of two major components. A grid depicting the results of various tests on the y-axis and dates on the x-axis is placed above the images obtained during the examinations.

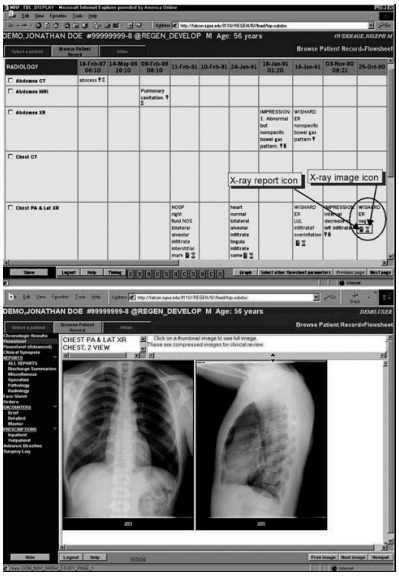

Figure 2. Example of a timeline flowsheet.

The third and final visualization is called a summary and snapshot (figure 3). It is an EHR specific instance of a dashboard. This tool can be used to display various types of information such as clinical summaries regarding patient problems, medication lists, medication allergies, vital signs, and health maintenance reminders (1).

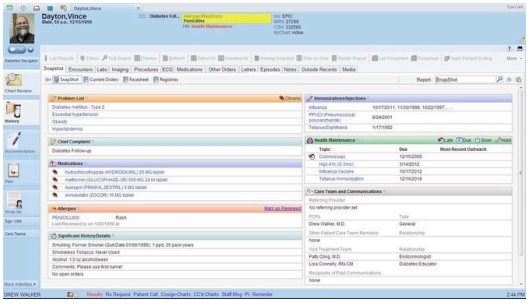

Figure 3. Example of an EHR summary (patient dashboard).

Systematic Review of EHR Data Visualization Methods

Expanding on this information is West et al.'s systematic review published in October 2014 focusing on *Innovative information visualization of electronic health record data*. After applying the exclusion and inclusion criteria, 18 articles were included in their qualitative synthesis. The results identified four methodologies described in table 1. Overall, the authors found there are few innovative EHR data visualization techniques. Despite the ever-increasing amount of information being stored in the EHR. The biggest challenge is making it easier for clinicians and researchers to identify meaningful patterns in the data. This can be challenging to accomplish because the target users desire both macro and detailed micro level data both present in the visualization. One of the limitations in the review concerned inability of the authors to discern if the data was real-time, retrospective, or utilized predetermined datasets. Additionally, the authors did a poor job of describing the visualizations and failed to include screenshots or pictures (3).

Visualization methodology	Description
LifeLines	Provide a timeline of a patient's events such as problems, diagnoses, or medications.
VISITORS	Give clinicians access to temporal data from disparate sources in a single view. Focuses on quality or clinical results
Dynamic Icons (DICON)	Allows users to explore clusters of patients similar to a pre-determined target patient. There was significant amount of training required for users to understand how to use the tool.
Outflow	Provides users with the ability to assess disease progression paths by giving them access to multiple events, the sequence of events, and patient outcomes.

Table 1. Visualization methods identified in West et al.'s systematic review.

Systematic Review of EHR Data Visualization Standards

Kopanitsa et al. published an article regarding the results of a literature reviewing the *Visualization of Medical Data on EHR Standards*. Techniques that currently exist in literature were evaluated based on the following criteria: limits of application, customizability, and re-usability. After applying two rounds of exclusion and inclusion criteria, 15 papers were analyzed for best practices and 4 different methodologies were described (table 2). The results of this study unfortunately revealed none of them meet all of the necessary criteria spelt out in their methodology for a re-usable standard based visualization method. However, each of these contributes to the ideal standard as envisioned by the authors (4).

Visualization Methodology	Limits, customization, and re-usability
LifeLines	No general approach, can only customize colors, well suited for re-use with temporal data.
PropeR	Little to no limitations because it is based on ISO 13606 standard, no tool to customize, flexible GUI allows for reuse with numerous data types.
GastrOS	Little no limitations because it is based on openEHR standards, GUI is very customizable, well suited for creating structured data entry forms
MUDR EHR	Initially developed for a specific EHR but has potential to be adopted by other EHRs, GUI is easily adjustable by the user, well suited to be re-used in other GUI research studies.

Table 2. Visualization methodologies and their characteristics in regards to potentially being incorporated into an EHR visualization standard.

METHODS

Overview

After initially meeting with Dr. Kari Stephens and the Data QUEST team, the design and development process of the three visualizations was iterative throughout the whole project. Kari initially supplied us with some example D3 visualizations, on YouTube (5), a colleague of hers created. Using this as inspiration, we as a team each decided to pursue a different visualization methodology. The reasoning behind this decision is due to the fact we nor the Data QUEST team knows which approach would be best. In the end it would be more useful to quickly develop three different prototypes, which will be formally assessed for usefulness in user interaction studies. The ultimate goal is to utilize the results of the user interaction studies to refine the prototypes. In order to facilitate the rapid development of these tools, the same sub-sample of the dataset provided was used for each prototype.

These visualizations were all built using D3.js. Entering into the class none of the group members had experience using D3, so this class provided a welcome but challenging learning experience. Some inspiration for these visualizations were found on the internet (6, 7). Each one is intended to provide intuitive and engaging interaction for the user, while providing useful information about the available data. The hope was to create methods that can be easily integrated for future use by the Data QUEST team in their website redesign process.

Small multiple bar chart visualization

One visualization method we pursued consisted of a small multiple approach to compare between clinics using bar charts. The purpose of this visualization is to allow researchers the ability quickly assess the demographics of all the clinics combined and for each individual clinic selected for analysis. The demographics included in the prototype are gender, race, ethnicity, and age. One larger multiple is displayed on the left hand side depicting the total for all the clinics currently being analyzed to the right of the larger multiple are the smaller multiples depicting the breakdown of the smaller clinics. Mouse over is incorporated into each bar in each chart to provide details on demand to the user.

Force bubble diagram visualization

The force bubble diagram allows the user to explore the breadth and depth of the data in regards to the patients' demographics. The purpose is to facilitate exploratory data analysis in a fun and engaging experience. It is our hope that this tool creates increased interest from a researcher looking to learn more about the dataset available. This interactive chart allows for two simultaneous user inputs for analysis. First, the bubbles can be sorted by the same demographic categories listed above for the small multiples visualization. Second, the user can color the dots to depict the distribution of a second demographic variable within the clusters of the first demographic variable selected. The size of the bubbles represents the longevity of each patient's time seen at the clinic. Details on demand are provided to the user through mouse over.

Stacked bar chart visualization

The stacked bar chart visualization allows the user to navigate the clinical data for specific disease for different years. This display allows the user to navigate through the aggregate data by number of patients. This is accomplished by selecting from the same demographic options used in the previous examples. Users may find patterns that occurred during a specific time frame for a certain demographic diagnosed with a specific disease. For example, if a user wanted to see if more males than females were diagnosed with depression from 2012-2014, this visualization could help researchers answer the question.

RESULTS

Overview

Our data visualization prototypes have not been formally assessed yet. However each will be in the coming months this summer. Overall our visualizations were well received by both the Data QUEST team and the participants in the poster session. Generally the users enjoyed interacting with each of the visualizations. Anecdotally, the force bubble

diagram was a favorite amongst the majority of the users. This is due to the observed higher than average amount of interaction time with the force bubble diagram compared to the other two visualizations.

Small multiple bar chart visualization

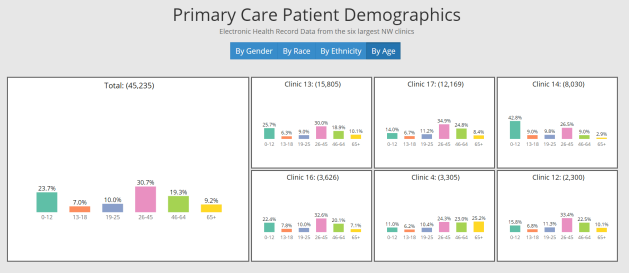

Figure 4. Screenshot of the small multiple visualization

Force bubble diagram visualization

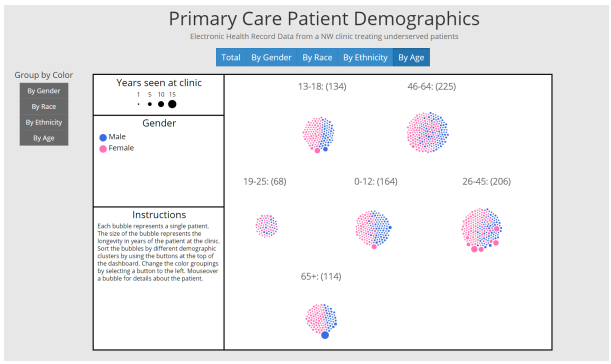

Figure 5. Screenshot of the force bubble diagram visualization

Stacked bar chart visualization

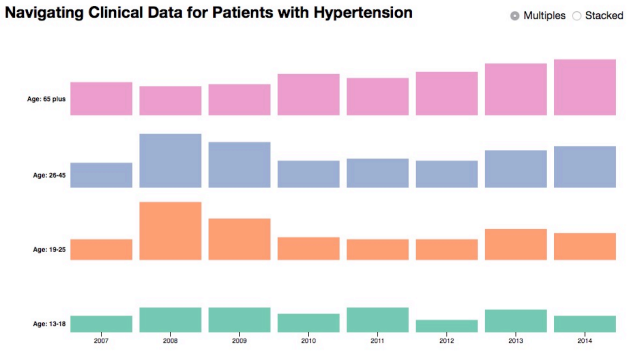

Figure 6. Screenshot of the stacked bar chart visualization

DISCUSSION

During the development process, we had three formal meetings with the Data QUEST team. The first meeting was a basic introductory meeting where we were able to discuss what their needs were and the best way that we could choose a project that was appropriate for our limited time frame. In the second meeting the team was provided with several wireframes of visualization concepts that we had. Once we had some basic prototypes working, a design review meeting was held to obtain feedback on both the design and the usefulness of the ideas being brought forward.

Throughout the design and development process, it was paramount for us to always be thinking about the most important question we were trying to help solve: *how is this technique going to help a researcher learn about the data?*

The team seemed very excited about the engaging experience that the bubble diagram provided for the user. The biggest challenge for this particular visualization was designing the UI. In particular, this visualization technique is most likely unfamiliar to the user. The design process involved restructuring the size and layout of the buttons. Also effectively using the space on the page proved to be a challenge. In regards to the small multiples visualization, the team requested the ability to view overall numbers of each bar in addition to percentages. These values were incorporated using a mouseover feature.

Feedback from poster session

Overall our visualizations were well received by the participants in the poster session. The force bubble diagram was the users favorite because transitions and animations grabbed the users attention quite easily and readily. The small multiples visualization was well received but some users had a major concern, which will need to be addressed in future work. Their main concern was scalability if additional clinics were added to the dataset. This could be addressed by incorporating a map with a brush. A user could select a target population by brushing and then the small multiples for each clinic could be generated alongside the total calculation. The stacked bar chart also received a great deal of praise from the users as well. The main feedback we received was the desire for the addition of a brush to view data aggregated by year.

There were also two high level pieces of feedback we consistently received during the poster session. First, it would be good to combine these visualizations in a meaningful way. As of now each stands on its own in no relation to the other. The second piece of feedback is the addition of a narrative could be very useful and powerful. Adding this functionality would give the users more context to the tools and data available to them. Additionally, the narrative could be used to tie the visualizations together more cohesively.

LIMITATIONS

The biggest limitation our team encountered related to the logistics required to get access to the data. Overall it took about two weeks and multiple in person or phone meetings to finally get a usable dataset. This significantly decreased the amount of time available for the team to code the visualizations.

FUTURE WORK

There are a great number of opportunities to expand upon the foundation we created during this project. First, formal user assessments need to be conducted in order to determine the benefits, if any, of each visualization methodology. Second, there is a multitude of additional data available for researchers that need to be visualized on their website. Additional subsets include vital signs, diagnoses, and billing information. Third, the addition of a narrative would add a lot to the user experience and make it more cohesive. Finally fourth, the Data QUEST team mentioned presented many great ideas during their meetings. Yet currently there aren't enough to make these ideas a reality. For example, the team would like to emphasize the historical aspects of the data in future visualizations. This is just one of many future opportunities to visualize this data set.

ACKNOWLEDGMENTS

We would like to acknowledge Dr. Kari Stephens and the Data QUEST team. Their support and feedback throughout the design and development process was invaluable.

We would also like to thank Dr. Jeff Heer, Jeffrey Snyder, and Dominik Moritz for leading such a fantastic course.

REFERENCES

1. Shortliffe E, Cimino J. Biomedical Informatics: Computer Applications in Health Care and Biomedicine 4th ed: Springer; 2013.

2. FURTHeR | Federated Utah Research and Translational Health e-Repository: University of Utah; [6/9/2015]. Available from: http://www.further.utah.edu/.

3. West VL, Borland D, Hammond WE. Innovative information visualization of electronic health record data: a systematic review. Journal of the American Medical Informatics Association. 2015.

4. Kopanitsa G, Hildebrand C, Stausberg J, Englmeier KH. Visualization of Medical Data Based on EHR Standards. Methods of Information in Medicine. 2012.

5. Lee A. AMD Bubbles. 2015.

6. Force Bubble Charts in D3: Delimited Technologies [6/10/2015]. Available from: http://www.delimited.io/blog/2013/12/19/force-bubble-charts-in-d3.

7. Vallandingham J. Small Multiples with Details [6/10/15]. Available from: http://vallandingham.me/small_multiples_with_details.html