As big data and data analysis become the hot trends in these days, there are many publications regarding exploring the interests between big data mining and data visualization. As computer scientists, we develop and optimize the algorithms to make the data models more accurate and efficient, where the below two articles showed data visualization can play a key role in data analysis as well.

The first article: “Progressive Visual Analytics: User-Driven Visual Exploration of In-Progress Analytics” was published in 2014 by Charles D. Stolper from Georgia Institute of Technology. The article was published in IEEE Transactions on visualization and computer graphics (TVCG). The newer article: “Data-Driven Healthcare: Challenges and Opportunities for Interactive Visualization" was published in 2016 by David Gotz and David Borland from University of North Carolina at Chapel Hill. The new article was published in IEEE Computer Graphics and Applications (C&G). The first article that was published in TVCG focuses more on the development of data visualization tools. The authors develop their own data visualization system to help the experts in healthcare domains. Unlike the first article, the second article focus on challenges and opportunities that occur in data visualization in healthcare domains. It focuses on up-to-date data visualization opportunities and challenges in healthcare industries and data visualization tools that we applied in healthcare. The article can provide general guidelines to later on researchers. Both articles bring us concerns and excitements about big data implementation in healthcare domains: What we can improve in later data visualization tools and What we need to take care of in later research.

Big data in healthcare is a new and hot trend in these days. In fact, there are dozens of big data research focusing on healthcare industries. I previously had a research experience in data analysis in Infusion Pump. The fascinating things regarding data analysis in healthcare domains are: it gathered a large amount of data (usually high dimensions as well), apply the appropriate algorithms and generate results that can show or imply data patterns that we haven’t been discovered. Unlike traditional diagnosis by doctors, big data in healthcare can provide a general, efficiency and accurate diagnosis over a larger population in a relatively short time. For instance, in these days, you can use your blood samples to test your gene in order to know in advance which parts of your organs have a higher chance of getting cancer. All of these was never heard before. However, then a new problem arrives: due to the large amount of data and complexity of the data models, analyst usually waits in days to see the running results of the data models. Thus, Charles D. Stopler from Georgia Institute of Technology introduces a new data analysis system that can overcome this disadvantage.

The data analysis system that they developed is called Progressive Visual Analytics. The main advantage of Progressive Visual Analytics is: it allows analysts to view partial results before the whole data models finished running. It increases the computation speed without interfering the cognitive workflow. [Charles D. Stopler 2014] Differ from the previous studies regarding solving the data complexity in the test sample, the Progressive Visual Analytics allows the users to explore of the new data patterns during the running time. Even more, the new data patterns will update the system to allow users to see the changes. The second advantage of

Progressive Visual Analytics is: it doesn't interfere with the analyst cognitive workflow. That is, the partial results will not cause strong bias to the users before the final output.

To effectively design the Progressive Visual Analytics, the authors applied Hetzler et al rules during the decision making for designing the data visualization tool. The interfaces of the data visualization tools need to make sure that the results are easy to interpret and reasoning. A good interactive visualization tool needs to inform the analysts of new information through subtle, ambients cues as well as allow the analyst to update the information as needed during computation time. [Charles D. Stopler 2014] In this experiments, they used the SPAM algorithm as the algorithm for this system. The main concept of SPAM algorithm is to find the most common sequence in a large and subsequence of events. The algorithm can implement well with the healthcare datasets. For instance, for each patient, his/her had a medical history in a timely sequence (such as when they have their flu shot or drugs doctors give to them from time to time). Finding most common sequence over a large population may provide a hint for diagnosis or prediction for certain disease. By running the algorithm over an enormous dataset that contains the population's medical records, the results can assist analysts to find data patterns that never found before or allow doctors to make more accurate prescriptions of their current patients.

To be visually efficiency, Progressive Visual Analytics provides List View, Scatterplot View, and Tree View to assist analysts examining and exploring dataset's pattern. In order to separate the partial results from the future results, the system uses color coding and a dynamic top-n list. For instance, the system uses purple to marked the results that haven't been finished running and uses orange to show the partial results that have been finished running at that time. The differ is, partial results generated by samples derived from the total population where the final results generated from the total population. At the current running time slot, the top-n most common sequence will be in the top column and marked orange. Once a new partial results coming out, the previous top-n most common sequence will become prefix and marked as blue. Besides, the system also limited the number of information displayed on the interfaces. Thus, the viewer doesn't get confused if there are too many information on the interfaces. By applying the color coding, dynamic visualization, and limited elements display, the Progressive Visual Analytics provides partial results to the analysts before the models finished running.

To examine how effective the system is, the early prototype of the system was applied at University Hospital of North Norway (UNN) by Dr. Stein Olav Skrøvseth over a 2-month period case study. The system was analyzed over 87,000 patients in the surgical department at UNN. Before using the Progressive Visual Analytics, Dr. Stein Olav Skrøvseth needs to wait for days before the technology department running the data models and send him the reports. By using the Progressive Visual Analytics, Dr. Stein Olav Skrøvseth can visualize the partial results earlier and change the parameters in the models according to his need. Dr. Stein Olav Skrøvseth noticed by using the Progressive Visual Analytics, he found it was easier to interpret the results in the earlier stage of runtime and enable him more flexibility of conducting the experiments. However, one limitation of the system is the dataset is only limited from UNN

databases since dataset from other hospitals may have different metrics and will need extra data clean process.

Overall, Progressive Visual Analytics seems to be a promising method that could implement later on in the healthcare industries. Even though there are many features that need to improve, such as using different algorithms, data formats editing...etc.

Then these bring another question, how do we measure our current data visualization tools and what obstacles will we later encounter if we apply more data visualization tools into the healthcare industries? The newer articles by David Gotz from University of North Carolina at Chapel Hill mainly discuss all these issues. The article summarizes several healthcare data visualization systems including Progressive Visual Analytics and discusses current healthcare-data interaction trends from different aspects. In some cases, the bias coming from the data has been overlooked in many researches.

One of the great database storage systems that have been adopted in the United States is the EHR system. According to the article, over 76 percent of US hospitals have been adopted with the EHR system. [David Gotz 2016] The EHR system captures all the information about the patients, whether when they went to the hospital or how many drugs the doctors prescribe the patients over time. As the EHR system containing a large amount of data that never occur in the history before, different aspects or approaches can be analyzed from the EHR databases. The article mainly summarizes the three aspects derived from current data-rich healthcare industries: 1). Breadth of Use 2). Data Complexity 3). Statistical Rigor.

Breadth of Use summarizes two types of applications, one is individually point-of-care application and the second is large-scale population application. By applying data models and data visualization, healthcare applications can solve individual diagnosis to large-scale geographic visualization. For example, an early use is John Snow's map showing clusters of cholera cases in 1854 London. [David Gotz 2016] Then the authors discuss the data complexity causing by the big data trend, where he illustrated the Progressive Data Analytics is a way to solve the data complexity. Later on, the author also addresses the data variety issues(high dimension data), data quantity issues(is the measuring data doesn't have any bias) and the data heterogeneity issues(different metrics from different hospitals). Furthur more, since there are not perfect models to predict the perfect results. One mistake from the models can cause a more serious issue due to the sensitivity in Healthcare industries, which known as Statistic Rigor. In summary, the authors concluded there still exists a lot of potentials with Interactive data visualization in healthcare that accompanied with lots of challenges as well. But overall, there are way more potentials in this industry that allowed the researchers to explore.

Both articles related to data visualization in healthcare industries. And both authors believe there are a lot more potentials in healthcare and big data interaction. The first articles propose a new solution to solve the data complexity and the second articles provide a broader

vision about data visualization in healthcare industries, but both articles agree that there are a lot more on data visualization in healthcare domains that allow us to explore.

Reference:

1)
@ARTICLE{6876049,
author={C. D. Stolper and A. Perer and D. Gotz},
journal={IEEE Transactions on Visualization and Computer Graphics},
title={Progressive Visual Analytics: User-Driven Visual Exploration of In-Progress Analytics},
year={2014},
volume={20},
number={12},
pages={1653-1662},
keywords={data analysis;data visualisation;learning (artificial intelligence);analytic algorithms;computational speed;dataset grow;electronic medical records;event sequences;in-progress analytics;information visualization techniques;progressive insights;progressive visual analytic systems;user-driven visual exploration;Algorithm design and analysis;Data visualization;Heuristic algorithms;Unsolicited electronic mail;Visual analytics;Progressive visual analytics;electronic medical records;information visualization;interactive machine learning;1},
doi={10.1109/TVCG.2014.2346574},
ISSN={1077-2626},
month={Dec},}

2)
@ARTICLE{7466736,
author={D. Gotz and D. Borland},
journal={IEEE Computer Graphics and Applications},
title={Data-Driven Healthcare: Challenges and Opportunities for Interactive Visualization},
year={2016},
volume={36},
number={3},
pages={90-96},
keywords={data visualisation;health care;medical administrative data processing;patient care;patient diagnosis;advanced data visualization tools;complex patient populations;data-driven healthcare;healthcare industry;interactive visualization;patient risk identification;personalized treatment plans;precise diagnosis;Data visualization;Medical diagnostic imaging;Medical services;Sociology;Statistics;Visualization;computer graphics;data visualization;healthcare visualization;visualization},
doi={10.1109/MCG.2016.59},
ISSN={0272-1716},

month={May},}

APA Citations:

1) C. D. Stolper, A. Perer and D. Gotz, "Progressive Visual Analytics: User-Driven Visual Exploration of In-Progress Analytics," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 20, no. 12, pp. 1653-1662, Dec. 31 2014.
2) D. Gotz and D. Borland, "Data-Driven Healthcare: Challenges and Opportunities for Interactive Visualization," in *IEEE Computer Graphics and Applications*, vol. 36, no. 3, pp. 90-96, May-June 2016.