Visual Analytics for Public Health: Supporting Knowledge Construction and Decision-Making

Samar Al-Hajj  
SCIENCE Lab  
Simon Fraser University  
samara@sfu.ca

Ian Pike  
BCIPRU  
University of British Columbia  
ipike@cw.bc.ca

Bernhard E. Riecke  
Simon Fraser University  
iSpace Lab  
Bern1@sfu.ca

Brian Fisher  
SCIENCE Lab  
Simon Fraser University  
bfisher@sfu.ca

Abstract

Massive and complex data impose a challenge on the public health community to explore, analyze, and synthesize valuable information to make timely informed decisions. This study exploits the use of Visual Analytics (VA) to enable health professionals to understand heterogeneous injury data and decide about dynamic health situations. Visual Analytics is defined as the “science of analytical reasoning facilitated by interactive visual interface” [14]. We conducted collaborative Paired and Group Analytics sessions to examine how VA assists health professionals in investigating injury data as well as in supporting knowledge construction and decision-making. This manuscript reports how stakeholders perceived VA to be usefulness in helping them understand the injury data, get insights and build knowledge that could potentially prompt actions into critical health situations. Future study implications can inform the design of innovative VA tools and techniques that synthesize novel collaborative VA approaches to optimize the decision-making process.

1. Introduction

Advanced technologies enable the production and collection of massive amount of complex and multidimensional data. The public health community is faced with the challenge of synthesizing salient information from complex and dynamic health data to build fundamental knowledge and make informed and timely decisions. Integrating Visual Analytics within the public health sector could facilitate and accelerate health professionals’ knowledge translation and dissemination as well as support their decision-making processes. Visual Analytics (VA), defined as the “science of analytical reasoning facilitated by interactive visual interface” [14], exploits humans’ visual capabilities to amplify their cognitive and perceptual skills and enable them to perceive and analyze large amounts of data [15][5]. Visual Analytics exploits humans’ pattern recognition brain capabilities to enhance the problem solving process [15]. Moreover, Visual Analytics exploits humans’ visual perception capabilities to reduce information overload and build accurate mental models that can leverage humans’ cognitive skills [6]. As Heuer postulated, building accurate mental models have a great influence on the way analysts understand data and make decisions [6]. The accuracy of the analyst’s mental model represents the foundation for accurate analysis and judgment needed to successfully analyze data and make informed decisions. Liu et al. explained how external visualizations of datasets could greatly shape analysts’ internal representations of the data and ultimately help them reason and make sense of events and situations under investigation. In the context of Visual Analytics, “external visualizations are internalized as mental models…which can be constructed and simulated in working memory for reasoning”[10]. Consequently, visualization tools and techniques are designed with an eye towards amplifying end-users’ analytical reasoning and cognitive skills with external visualizations aids that are essential for building accurate mental models and performing reliable data analysis.

Visual Analytics seemed to be an ideal solution for the health information overload problem and for facilitating health professionals’ data analysis and decision-making process. The objective of this current study is twofold. Firstly, it aims at presenting the application of Visual Analytics for exploratory analysis of public health data using injury indicators data. Secondly, it aims at evaluating the effectiveness of using Visual Analytics tool in a collaborative setting to facilitate insight generation, knowledge construction and decision-making.
2. Related Work

In order to substantially improve the outcome of our study, we used Visual Analytics in Paired and Group Analytics sessions to enable VA analysts to collaboratively work with injury stakeholders to effectively explore the designed injury data visualizations and construct knowledge about the injury data and ultimately support decision-making.

Previous studies have reported the successful synergy between advanced visualization tools and human cognitive capabilities that can lead to dramatic improvements in the data analysis and knowledge generation process. Keim et al. emphasized the relevance of integrating the human factor along the computational process of the VA system to improve the efficacy of the hypothesis generation and data analysis process [7][8]. Another study conducted by Heer et al. proposed the integration of social interactions to the VA analysis process and therefore argued that social aspects of Visual Analytics resourcefully contribute to the computational process and therefore advance the Visual Analytics data analytical process [4][5]. Along the same lines, Bernnam et al. presented a successful framework for multiple-analysts Visual Analytics collaboration to boost exploratory data analysis through interactive visual display [2]. Furthermore, many researchers emphasized the use of Visual Analytics for various analytical domains to support the decision making process. For instance, Xang et al. successfully introduced a Visual Analytics system to support bridge managers and domain users' decision-making [16]. Within the medical sector, Mane et al. adopted a Visual Analytics approach, the VisualDecisionLine, to leverage medical professionals' information processing capabilities and facilitate their therapeutic clinical decision-making process [11].

In this spirit, we exploited Visual Analytics tools and techniques to understand public health injury data in Canada. Canadian injuries constitute a major financial and economic burden [18]. They affect the lives of millions of Canadian youth and children every year [19]. Injuries constitute the leading cause of death among youth and children as well as the leading cause of hospitalization among 10-14 years old in Canada [20]. Analyzing the injury data is critical to identify trends and patterns in injuries and to flag common and leading injury causes in order to monitor and improve the health and well being of children and youth in Canada.

Currently, British Columbia Injury system uses the injury Data Online Tool (iDOT) to depict injury data using basic visual displays and tabular views (http://www.injuryresearch.bc.ca/). The current system restricts injury stakeholders’ data exploration process and limits their understanding of the injury situation. Therefore the need arises for innovative and advanced visual approaches that would incorporate temporal and geospatial data analyses to efficiently communicate the injury indicators characteristics and assist stakeholders’ in building essential knowledge and a comprehensive picture of the injury situation, and accurately identifying areas of needed interventions to make decisive actions and policies.

To assist injury stakeholders in improving the injury situations, we collaboratively worked with them and solicited their needs and goals to address the issue. Based on injury stakeholders’ feedback, we identified their needs for an innovative approach that could facilitate and support their decision-making process. The solicited feedbacks were reflected into the design of our injury data visualizations.

In our study, we combined humans’ cognitive and perceptual capabilities with social collaboration to advance the analysis process and improve health professionals’ decision-making process. We hypothesized that an overview of the data using collaborative Visual Analytics would help stakeholders get insights into valuable information about the injury data. We also hypothesized that exploring the data using collaborative Visual Analytics would help stakeholders to construct knowledge of the injury data situation as well as to support injury stakeholder’s decision-making process. In addition, we hypothesized that the knowledge constructed from the visualizations could potentially help stakeholders initiate appropriate actions to improve the injury situation.

The remainder of this manuscript is organized as follows: the next section describes the data used and methods adopted for this study, the following section summarizes the study results and the final section concludes with a discussion of the study results and future work.

3. Data and Methods

Mortality and Morbidity injury data in British Columbia, Canada were used to create the visualizations using Tableau Software. The injury data are complex and heterogeneous. They are massive and segmented into categories including patients’ age, gender, socioeconomic status as well as injuries’ types and geographic locations [17]. A mixed method approach was adopted in our study using semi-structured interaction with the Visual Analytics visualizations in a Paired and Group Analytics sessions as well as a questionnaire to solicit stakeholders’ feedback for later analysis. Our study’s independent variables were the interactive visualizations while the
study’s dependent variables were stakeholders’ perceived usefulness in terms of exploring the injury data, generating insights and knowledge as well as initiating appropriate actions.

3.1. The Visualizations

We designed and created time series Visual Analytics representations [Figure 1, 2] as well as an Interactive Visual Analytics Dashboard [Figure 3] for each of the injury indicators data: Mortality and Hospitalization indicators data. The time series visualizations and the Interactive VA Dashboard were built using Tableau Software and shown to stakeholders. Tableau is a commercially available visualization software that uses the Visual Query Language (VisQL) to visually represent large databases through interactive visual interfaces [21].

As illustrated in Figure 2, the time series Hospitalization visualization incorporates effective use of additional dimensions (i.e. bubble colors and bubble sizes) to enable stakeholders to track injury types and injury costs in addition to the injuries trends over time. Each color refers to an injury cause across different Canadian regions. For instance, Fall related injury (red bubble) has the highest rate and cost followed by the Transport related injury (blue bubble). Moreover, we embedded textual annotations into the Mortality time series visualization [Figure 1] to support stakeholders’ process of understanding the injury data and building a comprehensive picture of the mortality situation.

As shown in Figure 3, the Interactive Visual Analytics Dashboard combines multiple visualization views to efficiently depict the Hospitalization and Mortality indicators. A dashboard is defined as a comprehensive visual representation of the most relevant information required for stakeholders to reach specific goals [3]. The Interactive VA Dashboard encompasses 4 windows [see Figure 3]:

1. The Bar Chart to represent the different causes of injuries.
2. The Stacked Bar to represent the distribution of injuries across Health Authorities (HA).
3. The Map to represent the Geospatial distribution of injuries across the province.
4. The Time Series to represent temporal view of injury trends and pattern changes over time.

The Dashboard incorporates advanced visualization techniques applying Shneiderman’s Visual Information Seeking Mantra: overview first, zoom and filter, then details on demand [12]. Based on Shneiderman’s coordinated visualizations theory, we integrated multiple visualizations into the dashboard system in order to exploit the functionalities and capabilities of various visualizations and enhance end-users’ data exploratory and analysis process. The
Dashboard’s all four views are correlated together in a way that stakeholders’ interactions with one view are automatically reflected in other views.

3.2. Participants

We invited 17 stakeholders (10 Females and 7 Males) from various Canadian Provinces to participate in this study. Stakeholders self-reported their gender, age groups, and types of work. Stakeholders’ ages ranged between 25 and 64. Stakeholders come from diverse backgrounds; they are injury prevention practitioners (53%), researchers (15%), epidemiologists (7%) and public health policy makers (23%). Stakeholders volunteered to participate in the Analytics sessions, which were video recorded for later analysis. We emailed the consent forms to participants along with the Analytics sessions’ information. Stakeholders signed their informed consent forms prior to their participations.

3.3. Procedure: The Analytics Sessions

To explore stakeholders’ collaborative data exploration and knowledge construction process, we conducted two types of Analytics sessions: Group Analytics and Paired Analytics sessions. The Group Analytics session took place first, followed by 6 Paired Analytics sessions. During the Analytics sessions, Subject Matter Experts (SMEs) collaboratively worked with the Visual Analytics Expert (VAE) to solve an analytical problem using the proposed visualization tool [1]. In our case, SMEs were the participating injury stakeholders and the VAE was the manuscript’s first author.

3.3.1. Group Analytics. We conducted one Group Analytics session that lasted 45-50 minutes [Figure 4]. During the Group Analytics session, VAE demonstrated to the focus group, (SMEs) the created time series visualizations as well as the interactive Dashboard for the Hospitalization and Mortality data. Although the Group Analytics session was semi-structured, the main discussion revolved around 4 main topics:

1. Tool Demonstration: VAE and SMEs observed the visualizations together, exchanged expertise and skills, and built knowledge about the injury data.
2. Data Issues: VAE and SMEs discussed data access limitations, data issues, and concerns.
3. SMEs’ perceived usefulness of the visualizations in terms of stakeholders’ needs and goals.


3.3.2. Paired Analytics. We conducted 6 Paired Analytics sessions. Each session lasted for 8-20 minutes [Figure 5]. During the one-on-one Paired Analytics sessions, VAE explored with each SME the time series visualizations as well as the interactive Dashboard for the Hospitalization and Mortality data. VAE answered SME’s specific questions that are tailored to the SME’s work needs and preferences. To attend to stakeholders’ requests and provide further information about particular injury situation, VAE created additional visualizations with further levels of granularities using advanced visualization techniques like drill down, detail-on-demand as well as brushing and linking. During the Paired Analytics session, VAE took notes, and audio/video recorded each semi-structured session. The produced records were stored digitally and analyzed to observe the SMEs’ conversations, hand gestures, and interactions with the visualizations.

Following the Analytics sessions, stakeholders were asked to complete a semi-structured questionnaire. The questionnaire consisted of 20 questions that are scaled on a 7-point Likert scale (1-Strongly Disagree, 7-Strongly Agree) to solicit stakeholders’ feedback about the perceived usefulness of the visualizations and their potential to help stakeholders generate insights, build knowledge, and support decision-making.
4. Results

4.1. Stakeholders Experience

The questionnaire’s response rate was 76% (13/17). We compiled and analyzed the feedback data. The results demonstrated that stakeholders perceived the visualizations to be useful (average rating = 6.384/7, SD = 0.65) and time saving (average rating = 5.692/7, SD = 1.25) [Figure 6].

In Figure 7, the results showed that stakeholders highly ranked the visualizations for facilitating data exploration (average rating = 5.923/7, SD = 0.862) and data overview process (average rating = 6.153/7, SD = 0.688), as well as for helping stakeholders identify peak injury values (average rating = 6.153/7, SD = 0.375), see patterns in injury variables (average rating = 6.076/7, SD = 0.759), discover trends in injury data (average rating = 6.583/7, SD = 0.514), track injury data (average rating = 5.769/7, SD = 0.926), answer questions like “how much” and “how many” for injuries across health service delivery areas (average rating = 6.153/7, SD = 0.898), compare injury rates across regions (average rating = 5.846/7, SD = 1.143), and accurately assess the impact of injuries on provincial health authorities (average rating = 6.583/7, SD = 0.514).

Figure 8 represents a graphical summary of stakeholders’ ranking of the visualizations in terms of assisting stakeholders get insights (average rating = 5.692/7, SD = 1.109), generate knowledge of the injury data (average rating = 5.307/7, SD = 1.182) as well as support the potential initiation of appropriate actions (Average rating = 5.461/7, SD = 1.05).

Using Spearman's rank correlation, we constructed 3 linear models of the compiled feedback data to test the study’s hypotheses and identify correlations between (Data Overview & Getting Insights), (Data Exploration & Generate Knowledge) and (Generate Knowledge & Prompt Action) [Figure 9].

![Figure 6. Average rating for metrics: Usefulness and Time Saving.](image1)

![Figure 7. Average rating for feedback compiled data.](image2)

![Figure 8. Analysis of compiled feedback data.](image3)
The results of the Spearman’s rank correlation coefficient shown in Figure 9 demonstrate strong and significant correlations between data overview and insights into relevant information about the injury data (r = 0.612, df = 11, p = 0.026), stakeholders’ data exploration and knowledge generation about injury situations (r = 0.679, df = 11, p = 0.010) and finally between stakeholders’ knowledge construction and making informed decisions and perceived initiating appropriate actions (r = 0.614, df = 11, p = 0.025). As predicted, these results suggested that there are significant associations between stakeholders’ understanding the injury data and acquiring insights and knowledge about injury situations.

4.2. Group Analytics

During the Group Analytics session, SMEs observed the Interactive VA Dashboard while the analyst (VAE) showed various visualizations of the injury indicators data. Stakeholders explored the injury indicators data and constructed a comprehensive image about injury situations across BC’s provincial health authorities. SMEs, with diverse backgrounds, exchanged expertise and skills, and discussed the possibility of harmonized data collection across health authorities and an effective hospitalization cost calculation. The social aspects of the Group Analytics session improved stakeholders’ analytical reasoning process through interactions and argumentations, as well as data clarification and understanding [9]. During the session, stakeholders drilled down to further levels of details about interested injury types and locations. As stakeholders stated: “Can I check sports related injuries?”... “Can I drill down to compare injuries across regions?”... “Can I see the underlying raw data?”... “Can any one of these visualizations be copied, printed out or shared?” The Group Analytics session enabled stakeholders to pool their expertise, fuse various perspectives and scenarios about the injury situations and discuss data uncertainty and security in order to collaboratively optimize the decision-making process.

4.3. Paired Analytics

During the Paired Analytics sessions, SMEs collaborated with VAE in a think aloud approach [1] to analyze and understand the injury data. Interactive exploration of injury data using created visualizations and VA Dashboard enabled stakeholders to build essential background knowledge about injury situations across various provincial health authorities. The Paired Analytics sessions served to address stakeholders’ particular interests and needs. For instance, epidemiologists were interested in observing how injury rates are trending over time across different regions, as one stakeholders observed: “but look at the suicide rate... low, rise, rise, plateau...it’s very interesting”. On the other hand, policy makers were focusing on specific high injury rates at various local regions to inform regional policy decisions. As one of the stakeholders commented when observing a peak injury rate at a specific health authority: P3: “why the peak? What happened there... that something? Why? What was it? That’s what we (stakeholders) like to drill down to...”

The visualizations empowered stakeholders with the ability to answer their own questions about injury indicators and triggered them to further investigate the data as well as helped them to build knowledge and strengthen personal perspectives about the underlying characteristics of injury indicators.

5. Discussion and Conclusion

In this study, we presented the implications of using collaborative Visual Analytics (VA) to facilitate injury data exploration and support stakeholders’ decision-making process. We adopted Paired and Group Analytics sessions, where stakeholders used VA visualizations and Interactive Visual Analytics Dashboard to explore the massive injury data in order to build knowledge and support health related decision-making.

The injury data are rich with endless information and have a powerful potential to guide informed decision-making. In a nutshell, Visual Analytics helped health stakeholders to reason and make sense of the massive and heterogeneous injury data, to reveal and retrieve salient information and ultimately to construct fundamental knowledge that is vital to make informed decision-making process. The VA visualizations enabled customization to reflect and accommodate stakeholders’ analysis needs and goals. They offered stakeholders advanced visualization techniques such as zooming and filtering, details on demand and brushing and linking capabilities to enhance stakeholders’ cognitive and perceptual skills and enable them to...
competently examine the injury data at various levels of abstractions through the use of interactive visual interfaces. Stakeholders’ reasoning about a specific analytical problem and interpretation of particular injury situations or events play a pivotal role in the way stakeholders draw valid conclusions about the injury situation and make better decisions. According to the decision theory, “logical reasoning and decision making should be closely related topics” [13]. The created visualizations improved stakeholders’ logical reasoning and decision-making capabilities through interactive visual interfaces [8]; they provided stakeholders with analytical, temporal, and geospatial data visualizations, which summarize injuries’ data characteristics, injuries’ trends over time as well as injuries’ distribution across local health authorities.

The Analytics sessions compounded the analyst and stakeholders’ intellect and expertise to the powerful potential of the VA tool to advance the data analysis process. The Analytics sessions empowered stakeholders with foundational knowledge to inform decision-making and potentially prompt appropriate actions. They enabled stakeholders to collaboratively seek answers to their questions (i.e. “How many injuries?”), compare injuries across health authorities, and observe injuries trends and patterns over time in order to build a comprehensive picture of the injury situations.

This study supported our proposed hypotheses. It demonstrates that the use of collaborative Paired and Group Visual Analytics enables knowledge-stakeholders to collaborate together and explore the data as well as to identify and address gaps that could lead to the generation, translation and mobilization of new insights and knowledge, which are essential to support decision-making and initiate appropriate actions. The study’s empirical results support previous research on the effect of collaborative VA in promoting insights and enhancing the decision making process [4][5][8]. Our new findings supplement previous work and show that there are strong correlations between exploring and understanding data and the ability to get insights, generate knowledge and make informed decisions, which is demonstrated by the apparent influence of data exploration and understanding on getting insights and generating knowledge of the injury data and the potential to prompt appropriate actions.

We acknowledge that our current study encounters several limitations. One key constraint is that data were collected from a small focus group. The small sample size limits our findings from being generalizable to the broader public health community. This restriction is imposed by patient information privacy policies that allow only limited number of authorized stakeholders to access and interact with the available public health data. Another constraint is dictated by the patient data confidentiality policies, which impact the design of our VA displays when creating the visualizations. Not all injury data can be displayed in the visualizations. Injury indicators data should be aggregated at the regional level and all injury rates that are less than 5 cases should not displayed in order to protect patients’ identity from being disclosed or revealed.

Furthermore, we acknowledge the constraint imposed by the Paired and Group Analytics methodology as well as the introduction of a new visualization technology into the public health sector. We also understand the need for adequate training to help health stakeholders to comprehend and accurately interpret the injury data visualizations.

In conclusion, the study’s empirical results provide insightful approaches to expand the Paired and Group Analytics methodologies and emphasize the importance of establishing cooperation between the analysts and health stakeholders in the analytical loop to exploit the VA tool and iteratively interact with the data in order to filter and customize the visualizations and assess the quality of the retrieved analytical results. For future work, we need to conduct further heuristic evaluations of the introduction of Visual Analytics tools into the daily analytical activities of health professionals to improve the quality of health decision-making (i.e. inform injury prevention strategies and programs, introduce new injury regulations and policies, initiate appropriate injury prevention actions). Future work also includes the application of Visual Analytics to other domains and areas within the healthcare sector, including Trauma Registry data and Emergency Patients data. Furthermore, from a design perspective, the study’s implications can inform the design of future innovative visualization tools that synthesize the collaborative aspect of Visual Analytics as a key component to enhance knowledge construction and optimize decision-making.

6. Acknowledgment

We would like to thank the team at the BC Child and Youth Injury Research and Prevention Unit for their contribution to this study. We also would like to thank supervisors and colleagues for revising earlier versions of this manuscript.

7. References


