Association for Information Systems
AIS Electronic Library (AISeL)

MWAIS 2015 Proceedings

Midwest (MWAIS)

Spring 5-14-2015

Visual Storytelling: Impact of Data visualization on citizens' health behaviors

Dawei Wang
University of Oklahoma, dawei@ou.edu

Radhika Santhanam
University of Oklahoma, radhika@ou.edu

Visual Storytelling: Impact of Data visualization on citizens' health behaviors
Research in Progress

Dawei Wang
University of Oklahoma
dawei@ou.edu

Radhika Santhanam
University of Oklahoma
radhika@ou.edu

ABSTRACT

Although data visualization is gaining in popularity due to its ease of use and learning, and superiority of presenting data in a pleasing manner, healthcare sector has lagged behind other industries in the use of data visualization. In order to understand the appropriateness of data visualization technique, especially storytelling, in improving public health, it is necessary to empirically investigate whether and how this technique could help general public understand complex healthcare datasets and gain insights. In this study, we present a preliminary conceptual framework and our proposed research design to conduct this study. Specifically, we employ the stimulus-organism-response (S-O-R) model as a framework, and Tableau storytelling features to undertake this investigation. By doing so, we will be able to further our understanding of data visualization with storytelling component, and of whether and how data visualization influence common citizens' understanding about proper healthcare behaviors.

Keywords

Data visualization, healthcare, big data, storytelling, Tableau.

INTRODUCTION

An increasing number of organizations are using analytics and Business Intelligence (BI) techniques to gain insights to improving decision making (Chaudhuri, Dayal and Narasayya, 2011). Among these techniques, data visualization is gaining in popularity because of its ease of use and learning, ability to present data in a pleasing manner, and data interaction capability. Although the superiority of data visualization is recognized in communicating data to decision makers in practice, empirical evidence to support this, especially with new types of visualization techniques and when users are general public, is sparse.

It was argued that data visualization can serve the needs of policy makers to make informed decision about nationwide health care policy (Sopan, Noh, Karol, Rosenfeld, Lee and Shneiderman, 2012). However, it may be more appropriate to target the general public, rather than policy makers, in order to obtain the greatest benefits of data visualization techniques in the healthcare context. Empirical research has revealed that unhealthy lifestyles increase financial burden on national healthcare systems, approximately $34 billion to the expenses ("Research," 2013). By using data visualization techniques, it could be expected that the general public will be better educated about proper health behavior, improving quality of life nationwide, and reducing national healthcare spending. Further, scant literature devoted to exploring this issue, and no empirical research has ever tested this idea. Therefore, there is a strong need to investigate whether the benefits of data visualization techniques could be gained by the general public in the healthcare sector.

The intent of this study is to empirically examine whether or not, and to what extent, if any, data visualization techniques improve or diminish learning outcomes for general public. In particular, this study is designed to address the following research questions: When the general public is exposed to different types of healthcare data visualization (e.g., traditional spreadsheet, static visualization, and interactive data visualization), are there differences in their engagement (i.e., meaningful interaction, or active cognitive processing)? Does task complexity influence the relationship between visualizability and engagement? To what extent is the general public's engagement related to learning outcomes?

THEORETICAL BACKGROUND AND HYPOTHESIS DEVELOPMENT

Why is data visualization appropriate for citizen healthcare?

In the big data era, the volume of data grows by 30% to 50% each year across most organizations (Beath, Becerra-Fernandez, Ross and Short, 2012). To handle this data explosion, various BI tools have been introduced with goal of dealing with such processes as data management, data analysis and data visualization. In this study, data visualization is defined as "the use of computer-supported, interactive, visual representations of abstract data to amplify cognition" (Card et al., 1999, p.7). Due to the fact that information visualization and data visualization are frequently used interchangeably, we use the term data visualization in this study. Data visualization techniques have been increasingly employed in organizations, including banking ("Who," 2015), medical sector, and government agencies (Heise and Naumann, 2012). Because a lot of healthcare information is formatted as huge, multi-level and discrete datasets, use of data visualization by policy and decision makers is increasingly seen in practice (Sopan et al., 2012). However, a much larger population, in other words the general public, have yet to realize the benefits of healthcare data visualization. Some attempts to inform citizens using data visualization are seen in magazines, such as the New York Times, where visual reports are provided. We argue that data visualization's ability to influence the health practices of citizens is an opportunity that needs investigation. Visualization techniques are not new to health data analysis, but hitherto the focus has been on developing appropriate presentations for public policy decision makers (Afzal et al., 2011) rather than the general public. Unlike public policy decision makers, who can understand results of statistical models and quantitative data, the general public is less likely to be skilled in understanding statistical model results; hence easy-to-understand methods must be adopted. A possible solution to this situation is delivery of a content via the relatively new technique of visual storytelling, which uses the art of imposing narrative patterns to data. Recently, Tableau developed a storytelling feature, using deep interactivity with data, which allows viewers to manipulate visualization and make sense of the information. Storytelling is comprised of a series of interactive story points, by which viewers can read visual presentations about data (Kosara and Mackinlay, 2013). A good story conveys rich information in relatively few words in a format that can be easily assimilated by viewers. Besides, "People usually find it easy to understand information integrated into stories than information spelled out in serial lists, such as bullet points" (Gershon and Page, 2001). Therefore, storytelling opens the door to influencing the behaviors of citizens.

In order to understand the appropriateness of storytelling in improving public health, it is necessary to conduct a study of whether and how this technique could help the general public understand complex healthcare datasets and gain insights. Our study focuses primarily on the presentation dimension (i.e., storytelling) of data visualization as opposed to data preparation dimension.

RESEARCH MODEL AND PROPOSITION DEVELOPMENT

We draw on stimulus-organism-response (S-O-R) model as our holistic framework (Mehrabian and Russel, 1974), because prior environment psychology and IS research has successfully documented the S-O-R paradigm to model how stimulus influences information receivers' behaviors (Jiang et al., 2010; Sherman et al., 1997). In this study, visualizability serves as stimulus, engagement serves as organism, and learning outcomes serves as response. Visual representation can have a lasting impact because it is a way by which the brain easily comprehends information and immerses in the narrative in an interactive manner. Interactivity refers to "the extent to which user can participate in modifying the form and content of the mediated environment in real time" (Steuer, 1992). Visualizability is defined as "the quality of being visualizable" ("Visualizability," 2015), which can be conceptualized as different levels of visually and interactively presenting data (i.e., traditional spreadsheet, static visualization, and interactive visualization). We incorporate interactivity as a component in our construct *visualizability*. Static data visualization has limited capacity to change display or data selection, whereas interactive data visualization provides information users with "desired presentation format form alternatives provided by an information preparer" (Janvrin et al., p.45).

Storytelling is very interactive and engrossing, because the users can interact and move forward in a sequential manner through the story, go backward, and navigate at their own pace. Hence, the level of interactivity is more personalized, narrative in form, and meaningful, and thus could potentially enhancing viewers' engagement with data. According to engagement theory, *engagement* is defined as the meaningful interaction (active cognitive processing) that leads to active learning (Kearsley and Schneiderman, 1998). Similar to engagement, involvement can display focused attention as well. However, involvement may occur due to task demands or deadlines and thereby might not be enjoyable (Sandelands and Buckner, 1989). Engagement, in contrast, involves intrinsic interest. The existing e-learning literature supports that engagement serves as the intermediate variable between external intervention and consumer attitude and behaviors (Rappaport, 2007; Wang, 2006). In parallel, we conceptualize engagement as mediating variable.

The outcome variable we aim to investigate is *learning outcomes* (i.e., the amount of learning), which are comprised of three indicators (i.e., understanding, long-term memory, and intentions to change behaviors). Relevant to story-telling, the first indicator *long-term memory* represents "a relatively permanent memory store, from which information is not lost" (Atkinson and Shiffrin, 1971, p. 4). Long-term memory is highlighted in this study, because proper health behaviors have to be remembered for a long time in order to improve health quality. Empirical research supports the relationship between visual representation of a scene and long-term scene memory (Hollingworth, 2005). Thus, it could be expected that the increase in vividness and interactivity of representation would positively relate to long-term memory through the impact of engagement. However, whether there is partial or full mediation between the two constructs needs to be examined. The second indicator of learning outcomes is *understanding,* which refers to the degree of people's understanding of the displayed information. Understanding can be accessed by accuracy and time. The better the general public understand the complex healthcare data presentation, the more likely that they intend to change their unhealthy behaviors. Understanding, in this regard, reflects the degree to which different types of data visualization technique effectively communicate information. The last component of learning outcomes is *intentions of changing behaviors*, which refers to the extent to which participants intend to change their behavior after viewing different types of healthcare data presentation. According to Webb and Sheeran (2006), large changing behavioral intentions will lead to a small-to-medium behavior change. Accordingly, the overall impact of change in behavioral intention matters to whether proper health behavior would ultimately be realized. Therefore, it is incorporated as an indicator of learning outcomes.

Complexity of task derives from "task attributes that increase information load, diversity, or rate of change" (Campbell 1988, p. 43). *Task complexity* reflects the volume of content and number of pages of data presentation, which is a "central feature in determining the task performance and consequent information needs" (Vakkari 1991, p.825). It has been found that task complexity plays a moderating role in a variety of contexts and at both individual- and team-level research (Liu and Li, 2012), including software development (Balijepally et al., 2009), web browsing and searching (Adipat et al., 2011), and surgical tasks (Vashdi, Bamberger and Erez, 2013). Since much more information is contained in complex tasks than in simple tasks, as task complexity increases, presentation viewers' cognitive load and efforts increase. When facing high task complexity, presentation viewers would exert more mental effort and pay more attention to the presentations so as to solve problems and make decisions. Human beings, by nature, are cognitive misers, so more cognitive processing load a viewer experiences, the more mental stress and less intrinsic enjoyment the viewer would feel. As such, lower-level engagement would arise in the presence of high task complexity, and the need for visualizability increases in order to reduce cognitive efforts. In contrast, when facing low task complexity, presentation reviewers would experience reduced cognitive load and efforts. As task complexity decreases, the benefits of visualizability become more salient. Therefore, we expect that task complexity moderates the relationship between visualizability and engagement.

Based on the aforementioned rationale, the following hypotheses are developed.

Hypothesis 1: Visualizability is positively related to engagement.

Hypothesis 2: Task complexity moderates the relationship between visualizability and engagement, such that as task complexity decreases, visualizability will have an increasingly positive effect on engagement.

Hypothesis 3: The relationship between visualizability and learning outcomes is moderated by task complexity through the mediating effects of engagement, such that visualizability will have a positive indirect relationship with learning outcomes (via high engagement) when task complexity is low, and a negative indirect relationship with learning outcomes (via low engagement) when task complexity is high.

Hypothesis 4: Engagement will be positively related to the learning outcomes (i.e., understanding, long-term memory, and intentions of changing behaviors).

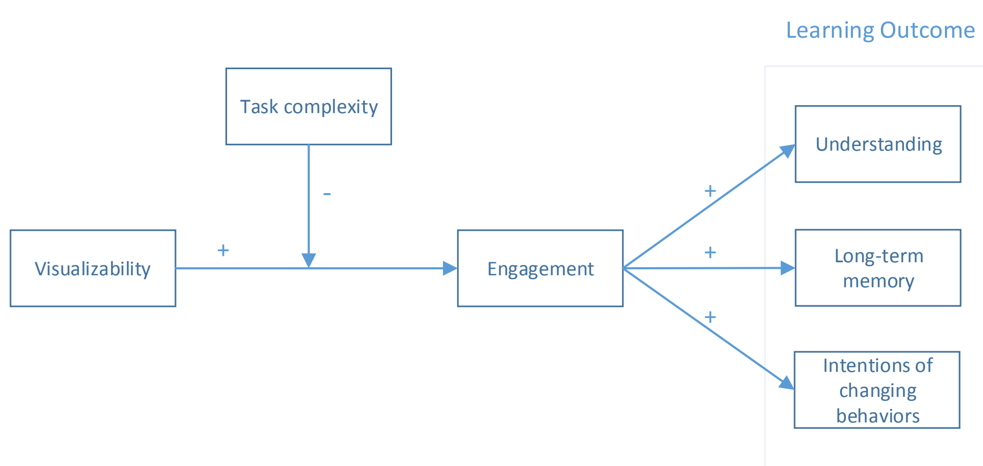

Figure 1. Research Model

CONTROLLED EXPERIMENTAL DESIGN

The expected sample size will involve 240 graduate and undergraduate students at a North American university (80 subjects in each group). In order to draw a samples that is representative of the general public, participants have to be strategically selected by excluding students majoring in certain majors (e.g., health care and statistics). For instance, students majoring in healthcare would not be appropriate for the sample because they have extra expertise and background knowledge in answering the questionnaire questions, thereby biasing the results. Each participant will be paid $15 for participation. The hypotheses will be tested through a controlled laboratory experiment. Participants will be asked to answer public health questions based on the presentations they are shown. Each question will have three different formats of presentations. One week later, the same participants will take a test to determine how much information they remember under different settings. We choose Tableau to display the visualization representation of healthcare data, because Tableau's ability-to-execute topped Gartner's (2015) magic quadrant for existing business intelligence and analytics tools. We will control for GPA, age, gender, race, and task complexity to ensure the same distribution of each for all three groups. For example, by ensuring all participants fall into the same range of GPA, it is reasonable to expect no variations in learning ability between the three groups. Then, random assignment of participants (i.e., probability sampling) will be ensured to handle external validity before the experiment.

OPERATIONALIZATION OF VARIABLES

A closed-ended questionnaire will be used to measure participants' learning outcomes. Engagement will be measured using a modified version of Webster and Ho's (1997) 7-point Likert scale, where 1 indicates "strongly disagree" and 7 indicates "strongly agree." The seven-item measure of engagement showed "high reliability and a uni-dimensional factor structure" with Cronbach's α larger than .9 (Webster and Ho, 1997 p. 71). Sample items are: (1) the visual presentation kept me totally absorbed in the browsing (navigating); (2) held my attention; (3) was fun. Understanding refers to the degree to which information receivers accurately understand the displayed information, which will be operationalized through two measures: the accuracy (i.e., percentage of correctly answered questions) (Adipat et al., 2011), and average completion time of each question measured (Murray and Häubl, 2011). The average completion time will be automatically recorded through JavaScript. Long-term memory will be operationalized as the percentage of correctly answered questions in the delayed test, which will also be automatically recorded using JavaScript. Visualizability will be operationalized by three distinct levels of interactivity: traditional presentation (text and spreadsheet), static visualization presentation, and interactive visualization presentation, with each coded as 0, 1, 2, respectively. Intentions of changing behaviors will be developed using advice from an expert panel. A pilot study will be conducted to check the internal consistency and reliability (e.g. using Cronbach's α). To ensure measurement validity, the self-developed instrument will be revised until Cronbach's α is larger than the widely accepted value 0.7 (Nunnaly and Bernstein, 1967).

DISCUSSION

To our knowledge, this is the first study that attempts to empirically investigate the impact of visualization techniques on citizen engagement with and learning about healthy behaviors. In doing so, we expect such visualizations to provide the general public the ability to adopt healthy behaviors. By introducing a conceptual framework that aims to empirically examine whether and how data visualization influence common citizens' understanding about proper healthcare behaviors, we will be able to further our understanding of data visualization with storytelling component.

REFERENCES

1. Adipat, B., Zhang, D., and Zhou, L. (2011) The effects of tree-view based presentation adaptation on mobile web browsing. *MIS Quarterly*, *35*(1), 99-122.
2. Afzal, S., Maciejewski, R., and Ebert, D. S. (2011) Visual analytics decision support environment for epidemic modeling and response evaluation, *Visual Analytics Science and Technology (VAST), 2011 IEEE.*
3. Atkinson, R. C. and R. M. Shiffrin (1971) The control processes of short-term memory, Citeseer.
4. Balijepally, V., Mahapatra, R., Nerur, S., and Price, K. H. (2009) Are two heads better than one for software development? The productivity paradox of pair programming, *MIS Quarterly*, 91-118.
5. Beath, C., Becerra-Fernandez, I., Ross, J., and Short, J. (2012) Finding value in the information explosion, *MIT Sloan Management Review*, *53*(4), 18.
6. Campbell, D. J. (1988) Task complexity: A review and analysis, *Academy of management review*, *13*(1), 40-52.
7. Card, S. K., Mackinlay, J. D., and Shneiderman, B (1999) *Readings in information visualization: using vision to think*. Morgan Kaufmann.
8. Chaudhuri, S., Dayal, U., and Narasayya, V. (2011) An overview of business intelligence technology, *Communications of the ACM*, *54*(8), 88-98.

9. D'Innocenzo, L., Mathieu, J. E., and Kukenberger, M. R. (2014) A meta-analysis of different forms of shared leadership–team performance relations, *Journal of Management*, 0149206314525205.
10. Gershon, N. and W. Page (2001) What storytelling can do for information visualization, *Communications of the ACM* 44(8): 31-37.
11. Heise, A. and F. Naumann (2012) Integrating open government data with Stratosphere for more transparency, Web Semantics: Science, Services and Agents on the World Wide Web 14: 45-56.
12. Hollingworth, A. (2005) The relationship between online visual representation of a scene and long-term scene memory, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3): 396.
13. Janvrin, D. J., Raschke, R. L., and Dilla, W. N (2014) Making sense of complex data using interactive data visualization, *Journal of Accounting Education*, *32*(4), 31-48.
14. Jiang, Z., Chan, J., Tan, B. C., & Chua, W. S. (2010) Effects of interactivity on website involvement and purchase intention, *Journal of the Association for Information Systems*, *11*(1), 1.
15. Kearsley, G. and B. Shneiderman (1998) A framework for technology-based teaching and learning, *Education Technology*, Sept/Oct: 20-37.
16. Kosara, R. and J. Mackinlay (2013) Storytelling: The next step for visualization, *Computer,* (5): 44-50.
17. Liu, P., and Li, Z. (2012). Task complexity: A review and conceptualization framework. *International Journal of Industrial Ergonomics*, *42*(6), 553-568.
18. Mehrabian, A. and J. A. Russell (1974). *An approach to environmental psychology*. The MIT Press.
19. Murray, K. B. and G. Häubl (2011) Freedom of choice, ease of use, and the formation of interface preferences. *MIS Quarterly,* 35(4): 955.
20. Nunnally, J. C., Bernstein, I. H., and Berge, J. M. T. (1967), *Psychometric Theory* (Vol. 226), New York: McGraw-Hill.
21. Rappaport, S. D. (2007) Lessons from online practice: New advertising models, *Journal of Advertising Research,* 47(2): 135-141.
22. Research Reveals Unhealthy Lifestyles Takes a Financial Toll on National Healthcare Systems. Retrieved from: http://newsroom.gehealthcare.com/research-reveals-unhealthy-lifestyles-takes-a-financial-toll-on-nation.
23. Roediger, H. L. and J. D. Karpicke (2006). Test-enhanced learning taking memory tests improves long-term retention, *Psychological Science,* 17(3): 249-255.
24. Segel, E. and J. Heer (2010) Narrative visualization: Telling stories with data, Visualization and Computer Graphics, *IEEE Transactions on Visualization and Computer Graphics*. 16(6): 1139-1148.
25. Sherman, E., Mathur, A., and Smith, R. B. (1997) Store environment and consumer purchase behavior: mediating role of consumer emotions, *Psychology & Marketing,* 14(4): 361-378.
26. Sopan, A., Noh, A. S. I., Karol, S., Rosenfeld, P., Lee, G., and Shneiderman, B. (2012) Community Health Map: A geospatial and multivariate data visualization tool for public health datasets, *Government Information Quarterly*, *29*(2), 223-234.
27. Steuer, J. (1992) Defining virtual reality: Dimensions determining telepresence, *Journal of communication* 42(4): 73-93.
28. Vakkari, P. (1999) Task complexity, problem structure and information actions: Integrating studies on information seeking and retrieval, *Information Processing & Management,* 35(6): 819-837.
29. Vashdi, D. R., Bamberger, P. A., and Erez, M. (2013) Can surgical teams ever learn? The role of coordination, complexity, and transitivity in action team learning, *Academy of Management Journal*, *56*(4), 945-971.
30. Visualizability. (n.d.) In Wikionary. Retrieved March 1st, 2015 from http://en.wiktionary.org/wiki/visualizability.
31. Wang, A. (2006) Advertising engagement: A driver of message involvement on message effects, *Journal of Advertising Research*-New York- 46(4): 355.
32. Webster, J., and Ho, H. (1997) Audience engagement in multimedia presentations, *ACM SIGMIS Database*, *28*(2), 63-77.
33. Who's Using Tableau? Retrieved from: http://www.tableau.com/about/customers