Comparison of Two Visualization Tools in Supporting Comprehension of Data Trends

Chen Ling[1(✉)], Julie S. Bock[2], Leslie Goodwin[2], G. Cole Jackson[2], and Molly K. Floyd[2]

[1] University of Akron, Akron, OH, USA
cling@uakron.edu
[2] University of Oklahoma, Norman, OK, USA
juliesbock@gmail.com, gcjackson02@gmail.com, mollykfloyd@gmail.com, lesliemgoodwin@yahoo.com

Abstract. This study compares two commonly used visualization tools, Gapminder and Tableau Public at simple and integrated level, to assess which tool better support user comprehension of data trends in a large dataset. Forty-seven participants were presented with a data set through Gapminder or Tableau Public at either a simple or integrated level. Then each participant was asked to answer questions about what they observed, showing their understanding of the information as well as rate ease-of-use of the tool. The results show that use of animation in Gapminder helped users comprehend data trends, whereas designs that assume prior geographical knowledge of users hampered their performance. Participants achieved higher accuracy with the graphs rated as easier to use. The results suggest that design of visualization tools should use familiar visual features and consider ease-of-use factor to assist user's comprehension of data trend.

Keywords: Data visualization · Display complexity · Data trend comprehension · Ease of use

1 Introduction

With the developments in technological innovation, vast amounts of data are being collected, which creates a need for an effective method of data presentation. In order to meet this need and present data in an effective way, companies around the world have turned to new data visualization tools. Unfortunately, many of these new data visualization software packages are not effective at communicating trends to people without advanced training in their use. There are often multiple views of a dataset or too many variables to show that users cannot distinguish the trends in data easily.

In an effort to simplify these complex datasets, many visualization tools are now using visualization or interactive charts in their presentations. The technique is to engage the viewer and provide statistical analysis using animation to unveil important statistical trends in large datasets over a period of time (Rosling 2007). For example, a tool called Gapminder animates trends that present information throughout time. The graph moves as time is played. This study investigates the effectiveness of two specific,

S. Yamamoto (Ed.): HIMI 2016, Part I, LNCS 9734, pp. 158–167, 2016.
DOI: 10.1007/978-3-319-40349-6_16

commonly used data visualization software packages, Gapminder and Tableau Public. By understanding what techniques work better, software developers can focus on the aspects of the tool that need to be simplified or made more interactive for the user to most effectively interpret the data.

Good visualization tools allow users to filter, sort, and visualize large datasets and then derive new data from the input data (Heer and Shneiderman 2012). Both Gapminder and Tableau Public allow users to interpret and draw conclusions to see trends forming just by looking at the graphs. These systems not only present a picture of the information, but also animate the information to add another dimension of information to be observed and understood by the viewer. Many tools limit the number of items in a display but allow users to select different filters while on the screen to analyze differences between variables (Heer and Shneiderman 2012). Kienle and Muller (2007) found that there should be multiple views present of the data in order to satisfy different stakeholders. It is worthwhile to compare effectiveness of finding data trend by viewing multiple graphs each of which only present trend on a single variable, or viewing a single graph presenting trends from multiple variables.

The goal of this study is to compare the effectiveness of two popular visualization tools-Tableau Public and Gapminder in terms assisting users in understanding data trends. In particular, we are interested in learning how the design of these two tools affect user's comprehension of data trend using multiple simple graph and one integrated graph. Ease of use of the visualization tool is also studied.

2 Literature Review

The goal of data visualization is to present users with data where trends can be detected and relationships can be established. Good visualization tools allow individuals to understand data in order to make hypotheses, look for patterns, and notice exceptions (Rosling 2007). Therefore, each visualization tool has unique design aspects to relay critical information to users.

Each visualization tool takes a distinct approach to aid the understanding of the data to the user. These tools focus on making the important information noticeable (O'Hare and Stenhouse 2008), while allowing users to more easily see the important trends and relationships within the data. While most experts agree that these tools do not intend to bombard the user with repetitive information and visuals, Heer and Shneiderman (2012) believe it can be useful to coordinate multiple or moving views to show different correlations of variables. Combining a clean layout with multiple views, the user can grasp bits of information at a time in order to understand the overall trends in the data. This experiment takes these varying opinions into account by showing the same information using visualization software on an animated display, on multiple simple displays, or all integrated on one display.

In an educational attainment case study, Heer and Shneiderman (2012) combined a bar chart, map, list, and scatter plot to display the ages, locations, education history, and income to help facilitate comparison, which could ultimately lead to a government decision (Heer and Shneiderman 2012). By compiling this information in a visual manner, the user can fully understand the scope of the data and piece together separate

information to see the bigger picture. This is just one example of how integrating data with an effective design can minimize distortion, ensure correspondence, and create consistency (O'Hare and Stenhouse 2008).

In order for data trends in the moving display, multiple simple displays, and integrated display to be consistent and represent the global picture, the display could not skew the user's view toward a specific relationship, which would misrepresent the data. For example, if in one view the movement from left to right represents increasing movement and correlates positively, then left to right should represent increasing movement and correlate positively in all views. Inconsistency can cause the user to misinterpret the data and have poor recall and interpretation of the information (O'Hare and Stenhouse 2008). One view may display the data differently from another. It is important to display the data so that users can make correct and unbiased inferences.

In order to visualize useful information, the design of the presented data is vital. Good visualization tools portray the data to aid user comprehension of data trends. For example, the visualization software tool Tableau Public allows users to drop data variables onto "shelve" categories, such as size, spatial position, etc. This particular software then determines a visualization design based on these three categories (Heer and Shneiderman 2012). Another important function of these tools is to allow users to filter through data dimensions. It allows users to look at subsets of data to find trends and correlations without having such broad overviews. In fact, many tools limit the number of items in a display but allow users to select different filters while on the screen to analyze differences between variables (Heer and Shneiderman 2012). By filtering, sorting, selecting, and highlighting different data, the user is able to produce various views with the ultimate goal of finding trends and new conclusions from the dataset. These types of dynamic and moving software tools can be represented in Gapminder software, one of the newest visualization tools.

Gapminder aims to engage the viewer and provide statistical analysis using animation to unveil statistical importance in large datasets over a time period (Rosling 2007). Often these trends and details are hidden under the excess information in traditional data visualization tools, but interactive, moving, or dynamic visualization software has made it possible to present this data in a way that is easier to understand (Battista and Cheng 2011).

3 Methodology

3.1 Participants

A total of 47 participants were recruited from the university. There were about twelve participants for each factor-level combination. Randomness was implemented by randomly assigning a visualization tool and complexity combination to each participant until each group had at least ten different participants. The age range of the participants was between 18 and 22 years old. Overall, 68 % of the participants were in STEM (science, technology, engineering, and math) majors, while the remaining 32 % were non-STEM majors.

3.2 Experimental Design

This study examined the effectiveness of data visualization tools, specifically Gapminder and Tableau Public. Participant's comprehension of trends in data presented by visualization software at different levels of complexity was compared. The independent variables were the visualization tool and the level of complexity. The simple level for a visualization tool presented two dimensions of data on one graph and had a total of three graphs. One graph would have life expectancy across time, another had average income across time, and the last one displayed population across time. All three graph conveyed data for four countries, Brazil, India, China, and Russia. The integrated level for a visualization tool conveyed the same four dimensions, but only on one graph. Information on life expectancy, income, population and time for the different countries are all presented on one graph for the integrated level and relied on more advanced visualization techniques. Tableau Public shows integrated graph in an interactive way, whereas Gapminder shows an animated graph.

The dependent variables were the accuracy of a basic comprehension test, and ease-of-use rating of the tools. The accuracy was a tally of correctly answered questions in a 15-question comprehension questionnaire to evaluate the effectiveness of the tool in assisting participants comprehend trends of data, and the ease-of-use ratings were given by participants at the end of each session.

The experiment used a between-subject design. No participant experienced more than one experimental condition to avoid any learning effect.

3.3 Equipment and Software

Two visualization software- Tableau Public and Gapminder software were used as well as a relevant dataset. An existing dataset already on Gapminder was used; this dataset displayed the average life expectancy, average income, and population size of a particular country per year all across time (year). Data from Brazil, Russia, India, and China was displayed because there was a noticeable trend in the data for each country.

3.4 Hypothesis

The Gapminder/integrated level was expected to yield the highest score due to the use of animation as a visual technique as well as the intuitive nature of the software. The animation utilized the flow of time to show trends. The animation visualization technique effectively sort, filter, and organize multiple dimensions of data onto one cohesive graph (Heer and Shneiderman 2012). Therefore, better performances were anticipated for participants experiencing animation technique. Furthermore, there was expected to be no significant difference between the two tools at the simple level because each used familiar techniques that did not differ greatly.

3.5 Procedure

The participants sat in front of the computer screen and observed the data displayed. The experimenter read instruction to the participant from a script and verbally explained the procedures for the experiment. The participants were not given a tutorial on the software to ensure that the results were reflective of the intuitive nature of the visual display and the tool presented. No one has prior experience with either tool. The proctor experimenter then passed out the paper comprehension test. Then the visualization software was run on a MacBook Pro, and the participants completed the test. The participants were allowed to interact with the software to answer the questions, but a time limit of ten minutes was set as a control. No incentives were used to motivate the participants.

4 Results

4.1 Scores of Comprehension Questionnaire

Based on the ANOVA results of the overall scores of the comprehension test, there was not a significant difference between the two visualization tools ($p = 0.33$) nor the levels of complexity ($p = 0.31$). However, a significant interaction effect was found between the complexity level and the visualization tool ($p = 0.013$; see Fig. 7).

The comprehension scores were affected by the combination of level and visualization tool. The average score for participants using the Tableau Public/Simple was higher than those using the Gapminder/Simple condition, but the Gapminder/integrated average scores were higher than those using Tableau Public/Integrated condition (see Table 1). Note that the Tableau Public/Simple participants answered the questions most accurately overall. Tableau Public/Simple condition had the highest average score along with the lowest standard deviation, meaning that the scores were consistently higher with less variance.

The interaction effect showed that participants scored less accurately with the Gapminder/simple combination, but scored higher using Tableau Public/Simple display. The main difference between the visualization tools at the simple level was the representation of the line that depicts the points between the presented variables (one of life expectancy, income, population) and time. In Tableau Public, the line is continuous, and the colors representing the countries are labeled along the right hand side of the line. In Gapminder, the numerical values of variable versus time were plotted with individual dots and users can click "turning on the trails" to produce a continuous lines. In Gapminder/Simple, the countries were labeled at the beginning of that country's color-coded line on the graph. But because the lines were so close together, it was hard to associate the label with the correct line (see Fig. 2). A map of the countries is located at right of the graph with the appropriate colors, but the software made the assumptions that participants were able to locate the country on the map without labels. This was not as effective and can be confirmed by the number of participants who correctly answered the question related to the identifying countries.

One question in the comprehension test asked what color is correlated to each country. The Tableau Public/Simple had the highest number of participants who

Table 1. Comprehension score for experimental conditions

Tools	Levels	Mean score	SD
Gapminder	Integrated	11.500	3.920
	Simple	10.250	2.491
Tableau Public	Integrated	10.091	2.663
	Simple	13.167	1.992

answered the question correctly, with 12, 10, 11, and 9 participants correctly identifying the colors for countries Brazil, China, India, and Russia, respectively. In comparison, Gapminder/Simple participants answered correctly questions 9, 4, 8, and 3 for the questions correlating to the colors for Brazil, China, India, and Russia. This indicated that participants had a difficult time identifying China and Russia, which could have been due to the close proximity of the countries on the map or the overlapping labels of the countries on the graph. Because of this, answering the questions related to Russia and/or China could also have been compromised if the participant could not correctly identify which trend line represented which country.

4.2 Ratings of Ease of Use

Participants gave ratings of ease-of-use at the end of each session where 1 corresponds to "cannot use", 3 corresponds to "neutral" and 5 corresponds to "very easy". ANOVA analysis on the ratings showed that there was no significant difference between the two visualization tools. But the simple level ($M = 3.13$, $SD = 0.99$) was easier to use ($p = 0.03$) than integrated level ($M = 2.55$, $SD = 1.10$). There was also significant interaction ($p = 0.0001$) between the visualization tools and the complexity level.

Figure 6 shows the interaction plot. It has similar trends as Fig. 5, where Tableau Public/simple had the highest rating of ease-of-use, and Tableau Public/integrated had the lowest rating. Gapminder/integrated condition had slightly higher rating than Gapminder/simple.

4.3 Overall Discussion

The results of this study demonstrate the importance of design features used in visualization tools. At the simple level where users only viewed data for one variable of four countries against time, users comprehended data trends better with Tableau Public. This may be due to the use of simple design features that is familiar to most users (see Fig. 1). In contrast, some users struggled with Gapminder, because it was difficult to tell the color-coded lines for four countries apart. The color coded map at the upper right corner did not have countries name marked, and the tool assume users know where each country is located (see Fig. 2).

At the integrated level, where users view all variables of four countries at the same time, users comprehended data trends better with Gapminder tool. This is mostly due to the use of animation in the tool. The animation utilized the flow of time to show trends.

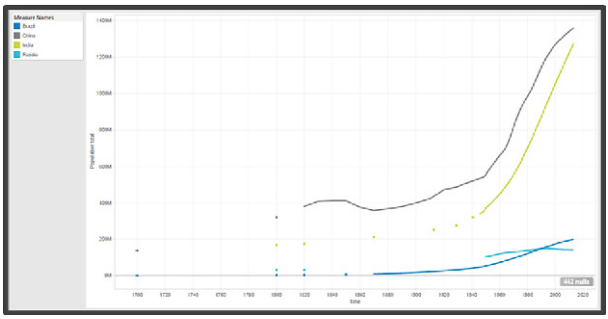

Fig. 1. Tableau Public/simple showing population of countries in a single graph

Fig. 2. Gapminder/simple showing income per person of countries in a single graph

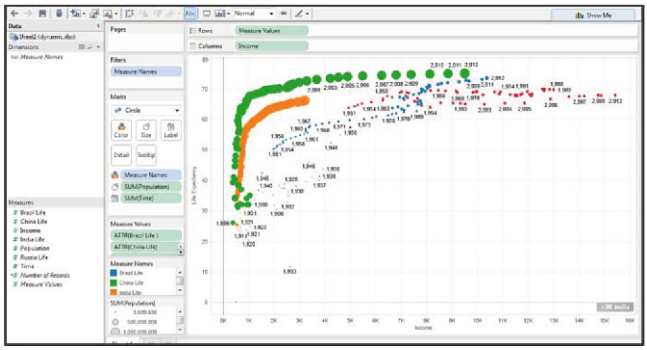

Fig. 3. Tableau Public/integrated showing four variables in an interactive way

Fig. 4. Gapminder/integrated showing all four variables in an animated format

In Gapminder, the year is shown as a light background on each of the graph (see Fig. 4). Users can play animation or slide across time to view changes in other variables across time. The animation utilized the flow of time to show trends. Essentially, this reduced the complexity and clutter of the display by separate out the "year" variable to be handled by animation. For both tools, the x axis was used to show income, y axis was used to show life expectancy, and the size of the data bubble denotes the population. In Gapminder, data of a single year is shown on each graph, but in Tableau Republic, data across all years are shown on one graph with labeling of the year beside the data point. This caused the display to be visually cluttered, and difficult to tell trends across year (see Fig. 3).

It is interesting to see the correspondence between the participant's comprehension questionnaire scores and their ease-of-use ratings (see Figs. 5 and 6). It shows that participant tends to have higher level of comprehend when the visualization tool is easier to use. In particular, it seems that the simple individual graphs clearly labeled

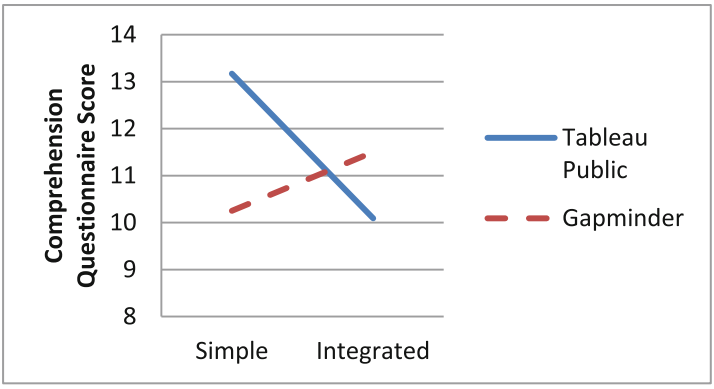

Fig. 5. Observed interaction effect for comprehension questionnaire scores

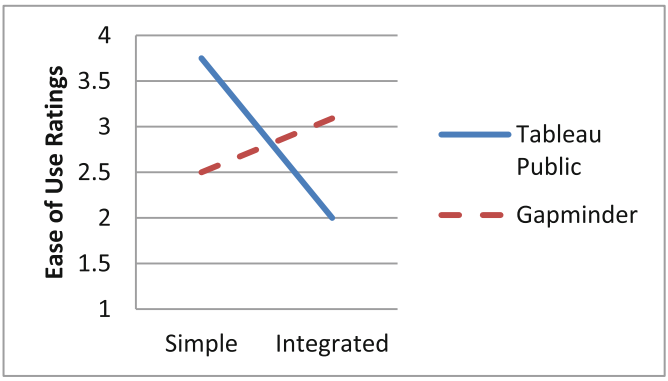

Fig. 6. Observed interaction effect for ease of use ratings

with country names in Tableau/Simple displays helped users to comprehend the data trend. On the contrary, when information about all three variables was shown simultaneously on the same graph with different coding as used in Tableau/Integrated display, it caused difficulty in user's comprehension. The use of animation in the Gapminder/Integrated was helpful for demonstrating data trend, resulting in higher comprehension score and ease-of-use rating than Gapminder/Simple display. Although Gapminder/Simple also displayed each variable with single graphs, the way for displaying country information was not as straightforward, and required prior geographical knowledge of the user. This may explain the worse user performance with the Gapminder/Simple displays.

5 Conclusion

This study demonstrates the importance of design of visualization tools in user's comprehension of data trends. Tableau/Simple display led to the higher level of comprehension and ease-of-use rating because of its use of familiar graphical features to most users-the simplicity of the continuous line to represent trends and the direct labeling of the countries. On the other hand, Gapminder/Simple caused difficulties for some users to correctly associate data line with corresponding countries because the program assumed geographic knowledge of the users. This conclusion contradicts our original hypothesis, but it does make sense that use of familiar visual features has a positive effect on comprehension. Also, the visualization tool should not assume prior knowledge of user and always provide redundant ways to assist users.

Another interesting finding was the correspondence between the comprehension performance and the ease-of-use ratings of users. Users are able to better comprehend data trends when the visualization tool is easier to use. Our results highlight importance of ease-of-use for visualization tool design. When designing visualization tools, we need to consider the ease-of-use factor for the tool to assist user's understandings of data trends.

This experiment collected data from college students with minimum training in data visualization tools. Further study could explore how training in use of visualization tool may change the comprehension scores. Further investigation can also be performed on other visualization techniques used in several different programs to determine the design effectiveness.

References

Battista, V., Cheng, E.: Motion charts: telling stories with statistics. In: Joint Statistical Meetings, Miami Beach, Florida (2011)

Heer, J., Mackinlay, J., Stolte, C., Agrawala, M.: Graphical histories for visualization: supporting analysis, communication, and evaluation. IEEE Trans. Visual. Comput. Graph. **14**(6), 1189–1196 (2008)

Heer, J., Shneiderman, B.: Interactive dynamics for visual analysis. Commun. ACM **55**(4), 45–54 (2012)

Kienle, H.M.; Muller, H.A.: Requirements of software visualization tools: a literature survey. In: 4th IEEE International Workshop on Visualizing Software for Understanding and Analysis, 2007, VISSOFT 2007 (2007)

O'Hare, D.D., Stenhouse, N.N.: Redesigning a graphic weather display for pilots. Ergon. Des. **16**(4), 11–15 (2008)

Rosling, H.: Visual technology unveils the beauty of statistics and swaps policy from dissemination to access. Stat. J. IAOS **24**(1/2), 103–104 (2007)