STUDENTS' UNDERSTANDING OF DATA VISUALIZATIONS

Charlotte Bolch and Tim Jacobbe
University of Florida, USA
cbolch@ufl.edu

Students are currently exposed to graphical displays and data visualizations presented to them in their courses and the media. Educators can incorporate data visualization skills into statistics courses where just basic graphical displays are usually the focus of descriptive statistics. The purpose of this study was to investigate the impact of students' experiences in learning how to read, interpret, and create data visualizations with the use of dynamic statistical software on their graph comprehension skills. A specific data visualization module addressing these skills supplemented the curriculum in an introductory statistics course. Data visualization assignments, student journals, the Survey of Attitudes Towards Statistics© and the Levels of Conceptual Understanding in Statistics assessments were used to examine students' understanding and perceptions about data visualizations and graph comprehension skills.

BACKGROUND

Most students in all educational levels are currently exposed to graphical displays and tables almost every day through the media. In the 21st century, statistics and graphical displays are increasingly being used within online/print newspapers or magazines and television as ways to add credibility to arguments or persuade consumers in advertisements (Friel, Curcio, & Bright, 2001; Watson, 2015). The ability to properly evaluate evidence and claims based on data are important skills that students need to learn during their education (Ben-Zvi & Garfield, 1997; GAISE College Report ASA Revision Committee, 2016). Helping students of all educational levels develop statistical literacy skills is a way they can understand and evaluate the quantitative information in their everyday lives to be informed citizens. Statistical literacy refers to two interrelated components: people's ability to interpret and critically evaluate statistical information; and their ability to discuss or communicate their reactions to such statistical information (Gal, 2004).

Data visualization has developed over the past ten years as a way to present complex data and information in a succinct way and perhaps discover unknown relationships and patterns (Keim, Kohlammer, Mansmann, May, & Wanner, 2010; Mirel, Kumar, Nong, Su, & Meng, 2016). Data visualization builds upon traditional graphical displays like the histogram by presenting data in a way that makes the data stand out, provides comparisons between multiple variables or subgroups, may allow the user to interact with the graphic, and may use words or pictures to create infographics (Nolan & Perrett, 2016). To understand and develop these data visualization skills, students need to first develop a foundation of statistical literacy skills.

Statistical literacy is the understanding of statistics information or research results by having the skills to organize data, construct tables, and create different data representations (Gal, 2004). A student can be taught how to *do* statistics, but they can go through an entire statistics course and *not develop the statistical thinking tools* needed to approach an issue or problem. Statistical thinking encompasses a more holistic view of learning and applying statistics than a procedural approach to "learning" statistics does (delMas, 2004; Gal, 2004). Graphical comprehension is a component of students acquiring statistical literacy skills and becoming statistically literate citizens (Ben-Zvi & Garfield, 1997; National Council of Teachers of Mathematics, 2000). Students at the kindergarten through 12th grade (4 to 6 year-olds through 17 to 19 year-olds) or college/university level may not be conducting research, but they are still exposed to data displays through graphs and tables in their everyday lives through magazines or television (Franklin et al., 2007). Therefore, the ability to understand and interpret graphical displays is an aspect of creating a well-rounded citizen. At the college/university level, instructors should cultivate statistical literacy and thinking skills, so students can have a deeper understanding of statistics. Students can use those methods to apply what they have learned to problems and issues they encounter with data, analysis, and inference in their everyday lives (Pfannkuch & Wild, 2004).

There has been a limited amount of research conducted on students' graph comprehension focused on statistics compared to research on student graph choice and construction in areas of mathematics, science, and psychology. Curcio (1987) introduced three levels of graphical comprehension questions defined as “read the data”, “read between the data”, and “read beyond the data”. In the literature, graph comprehension has been looked at from many different perspectives with some examples focusing more on cognitive ability (Lem, Onghena, Verschaffel, & Dooren, 2013; Lowrie, Diezmann, & Logan, 2012; Nicolaou, Nicolaidou, Zacharia, & Constantinou, 2007), the graph as a singular entity (Selva & Lima, 2010; Shah & Hoeffner, 2002), or a student’s ability to translate information between multiple different modes (Espinel Febles & Carrion Perez, 2006; Kosslyn, 1985).

THE PROBLEM

Educators should incorporate data visualization skills into courses such as statistics where only basic histograms, boxplots, and scatterplots are mainly taught (Friel et al., 2001; Nolan & Perrett, 2016). In traditional introductory statistics courses in the United States, the first few weeks of class focus on teaching students how to display and summarize categorical and quantitative variables using the standard graphical displays like bar charts, histograms, boxplots, or dotplots. A common pedagogical approach is to learn about these standard graphs by drawing a few by hand using small data sets. However, this should not be the only approach to teaching students data visualization skills. There has been a rapid increase in computational and graphing capabilities with statistical software, applets, and open-source software, but teaching practices have not kept up with these advancements (Nolan & Perrett, 2016).

According to the *Guidelines for Assessment and Instruction in Statistics Education (Committee) College Report 2016* in the United States, “data analysis involves much more than constructing a confidence interval or finding a p-value” (GAISE College Report ASA Revision Committee, 2016, p. 9). The guidelines recommend that one goal for introductory statistics is that students should be able to produce graphical displays and numerical summaries and interpret what graphs do and do not reveal. The Committee urges instructors to give students experience with multivariable thinking that prepares them “for challenging questions that require investigating and exploring relationships among more than two variables” (GAISE College Report ASA Revision Committee, 2016, p. 14). Also, the Committee recommends that since technology has changed the practice of statistics, there should also be a change in how statistics is taught in regards to technology and software (GAISE College Report ASA Revision Committee, 2016).

PURPOSE

The purpose of this study was to investigate the impact of students’ experiences in learning how to read, interpret, and create data visualizations with the use of technology and dynamic statistical software on their graph comprehension skills and attitudes towards statistics. A specific data visualization module addressing these skills along with teaching students how to interpret and critique data visualizations they encounter in their everyday lives was used to supplement the standard curriculum in an introductory statistics course. This module included the technicalities or issues that are present in real-world data (e.g. missing data, unstandardized variables, merging multiple data sources, etc.) and required students to apply critical thinking skills to answer research questions using this data. The module also taught students using various statistical software and technology (such as Minitab, iNZight, Google Sheets, and RAWGraphs) how to create data visualizations.

The two-week module on data visualizations included:

- Learning common vocabulary for describing data visualizations.
- Learning about data visualizations and their uses in different contexts.
- Interpretations and critiques of data visualizations in media sources.
- Deconstruction and reconstruction of a data visualization using a statistical software tool with the intention of creating a clearer and more concise visual display.
- Creating a data visualization poster using statistical software and technology that answers a research question of interest by using a real-world dataset with multiple variables, a high number of observations, and “messy” data values.

Data visualization skills were also incorporated throughout the rest of the course when appropriate during the regression, normal models, and inferential statistics units.

The learning objectives of the data visualization module within the curriculum were:

- Students should be able to read, interpret, and critique data visualizations.
- Students should be able to create data visualizations from a real-world dataset using technology that is effective and appropriate to answer their research question of interest.

This study was important because it looked at whether students' graph comprehension and attitudes towards statistics increased, stayed the same, or decreased with the data visualization module. This study was an innovative approach to teaching introductory statistics using various statistical software and open-source programs. Also, the study collected information about students' attitudes towards statistics before and after the course to see if using a data visualization module integrated with technology would change students' preconceived perceptions about statistics and the usefulness of the data visualization tools after the course.

RESEARCH QUESTIONS

- With the implementation of the data visualization module using dynamic statistical software, was there an increase in student graph comprehension?
- Did student attitudes towards statistics and the application of statistics to their everyday lives change with the implementation of the data visualization module?
- Were students able to create accurate and effective data visualizations using technology?

METHOD

The participants in the study were 6 students from a community college introductory statistics course in Fall 2017 in the southeastern part of the United States. Approval from the Institutional Review Board/Ethics Committee was received. The students in the course were approached for their participation by the researcher (who was also the instructor) during the first week of the fall semester. The course was taught in Fall 2017 in the evening with non-traditional students (adults returning to college for an associate's or bachelor's degree). This study was ongoing with the same course taught in Spring 2018, but the preliminary results provided in this paper were only from the Fall 2017 course.

Data Collection

Table 1 highlights the methods that were used to collect data for each research question. Students' graph comprehension was assessed using the beginning/intermediate assessment of the National Science Foundation funded Levels of Conceptual Understanding in Statistics (LOCUS) project (DRL-1118168). The assessment has 23 multiple-choice questions and was administered before the data visualization module began (pre-assessment) and at the end of the course (post-assessment). Student ability to create graphical displays using technology and their data visualization skills were assessed using two data visualization assignments.

Table 1. Alignment of research questions and data collection methods

Research Question	*Data Collection Methods*
Question 1: Student graph comprehension skills	-LOCUS assessment with graph comprehension questions (Pre/Post)
Question 2: Student attitudes towards statistics	-*Survey of Attitudes Towards Statistics©* (Pre/Post) -Journal entries
Question 3: Ability to create a graphical display	-Data visualization assignments

The assignments were graded using a rubric that looks at five critical tasks needed in a data visualization and assigns a competency level to each task (needs improvement, basic, or surpassed). The assignments were an assignment on how to deconstruct and reconstruct a plot using a statistical software tool and a final project using a real-world dataset where the students hypothesized a question and created a data visualization poster to answer the research question.

Students' attitudes and perceptions towards data visualizations and the application of using statistical software in their everyday lives was assessed using 1) a validated *Survey of Attitudes Toward Statistics©* (SATS-36) (Schau, 2003; Schau, Stevens, Dauphinee, & Vecchio, 1995) and 2) student journal entries. Student journals were used to understand a student's thought process, reactions, and perceptions throughout the data visualization module. Students were asked to complete a journal entry consisting of a few questions about what they learned after each class activity during the two weeks (four entries total) along with one journal entry about their reactions toward the final data visualization project. In addition, basic demographic information like age and gender were collected during the Pre-*Survey of Attitudes Towards Statistics©.*

The data visualization assignments were graded via a rubric. The level of competency for each critical task was assessed for each participant. The research method of grounded theory (Corbin & Strauss, 1990) was used to look for common themes in student journal entries about the applicability of data visualization skills and using statistical software in their everyday lives. Using the scoring guide of the SATS-36, the 36 items were grouped into six attitude components which were Affect, Cognitive Competence, Value, Difficulty, Interest, and Effort (Schau, 2017). The responses to all negatively worded items were reserved before all item responses within each attitude component were summed and divided by the number of items within each component. For each students' attitude component, the change in subscale scores between the Pre-*Survey of Attitudes Towards Statistics©* and the Post-*Survey of Attitudes Towards Statistics©* were calculated. The possible range of scores for each attitude component was between 1 and 7 with higher scores indicating more positive attitudes. The pre vs. post results for the LOCUS assessments were compared by calculating the percentage of correct responses for each student for the graph comprehension questions identified within the assessment.

RESULTS

Student Graph Comprehension

Student graph comprehension scores for the percentage of correct answers ranged from 45% to 82% on the pre-LOCUS assessment. For the post-LOCUS assessment, student graph comprehension scores ranged from 25% to 92%. Only three students (students 1, 2, and 5) demonstrated an increase in their graph comprehension scores between the pre vs. post-LOCUS assessments. See Table 2 for graph comprehension scores for all six students.

Table 2. Pre and Post-Assessment Graph Comprehension Scores

Student	*Pre-Assessment Graph Comprehension Score (% Correct)*	*Post-Assessment Graph Comprehension Score (% Correct)*
1	54%	92%
2	73%	75%
3	45%	42%
4	64%	25%
5	54%	58%
6	82%	67%

Attitudes Towards Statistics

Table 3 shows the change (increase, decrease, or no change) in each attitude component for all six students between the Pre vs. Post-*Survey of Attitudes Towards Statistics©.* Four students' feelings concerning statistics (Affect) increased over the course of the semester. Three students' attitudes about their intellectual knowledge and skills (Cognitive Competence) increased while their attitudes about statistics being less difficult (Difficulty) as an area of study increased as well (negative responses were reverse coded so higher attitude scores means less difficulty). All students' attitudes about the usefulness, relevance, and worth of statistics in their everyday lives (Value) increased during the course. Students' individual interest level in statistics (Interest) either increased or stayed the same over the semester. The amount of effort each student believed it took to learn statistics (Effort) had no change or a decrease.

From the journal entries throughout the semester, two students described that one of the activities allowed them to understand that sometimes graphical displays were created to influence people's views or deceive the reader of the graph. One student stated in their journal entry after their final project that they were able to make the direct application to what they were doing in their job analyzing surveys and writing reports for various stakeholders.

Table 3. Six attitude components of Pre vs. Post-*Survey of Attitudes Towards Statistics©*

Student	*Affect*	*Cognitive Competence*	*Value*	*Difficulty (↑ = less)*	*Interest*	*Effort (↑ = less)*
1	↓	↓	↑	↓	↑	No change
2	↓	↓	↑	↓	No change	No change
3	↑	↑	↑	↑	↑	↓
4	↑	↑	↑	↑	No change	No change
5	↑	↓	↑	No change	No change	↓
6	↑	↑	↑	↑	↑	↓

Effective and Accurate Data Visualizations

Student data visualization assignments were all rated on five different tasks (computation, analysis, synthesis, visual presentation, and written communication) on the competency levels of "needs improvement", "basic", or "surpassed". All six students showed through their data visualization assignments that their level of competency was surpassed on each task of choosing the correct graphical display, identifying key features, and communicating their findings clearly and concisely through visual presentation and written communication.

CONCLUSION

This study examined preliminary data about students' graph comprehension skills and attitudes/perceptions about statistics with the use of a two-week data visualization module in an introductory statistics course. Based on the small sample size, we could not determine whether students' graph comprehension skills improved over the course of the introductory statistics course along with their attitudes and perceptions towards statistics. The results from this study we hope can be used to inform future studies with larger samples of students to investigate if there is a relationship between the use of a data visualization module and an increase in students' graph comprehension skills. In addition, we envisioned this study with a larger sample size to inform educators about the use of a data visualization module in helping students to have a positive attitude towards statistics.

REFERENCES

Ben-Zvi, D., & Garfield, J. (1997). Statistical Literacy, Reasoning, and Thinking: Goals, Definitions, and Challenges In I. Gal & J. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 3-15). Amsterdam, Netherlands: IOS Press and the International Statistical Institute.

GAISE College Report ASA Revision Committee. (2016). *Guidelines for Assessment and Instruction in Statistics Education College Report* Retrieved from http://www.amstat.org/education/gaise

Corbin, J., & Strauss, A. (1990). Grounded Theory Research: Procedures, Canons, and Evaluative Criteria. *Qualitative Sociology, 13*(1), 3-21.

Curcio, F. R. (1987). Comprehension of Mathematical Relationships Expressed in Graphs. *Journal for Research in Mathematics Education, 18*(5), 382-393.

delMas, R. C. (2004). A Comparison of Mathematical and Statistical Reasoning. In J. Garfield & D. Ben-Zvi (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 79-95). The Netherlands: Kluwer Academic Publishers.

Espinel Febles, M. C., & Carrion Perez, J. C. (2006). *An Investigation About Translation And Interpretation Of Statistical Graphs And Tables By Students Of Primary Education.* Paper presented at the Seventh International Conference On Teaching Statistics, Salvador, Bahia, Brazil.

Franklin, C. A., Kader, G., Mewborn, D. S., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for Assessment and Instruction in Statistics Education (GAISE)*. Retrieved from Alexandria, VA: www.amstat.org/education/gaise

Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making Sense of Graphs: Critical Factors Influencing Comprehension and Instructional Implications. *Journal for Research in Mathematics Education, 32*(2), 124.

Gal, I. (2004). Statistical Literacy: Meanings, Components, Responsibilities. In J. Garfield & D. Ben-Zvi (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning, and Thinking* (pp. 47-78). The Netherlands: Kluwer Academic Publishers.

Keim, D. A., Kohlammer, J., Mansmann, F., May, T., & Wanner, F. (2010). *Mastering the Information Age: Solving Problems with Visual Analytics*. In D. A. Keim, J. Kohlhammer, & G. Ellis (Eds.). Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.363.2661

Kosslyn, S. M. (1985). Graphics and Human Information Processing: A Review of Five Books. *Journal of the American Statistical Association, 80*(391), 499-512.

Lem, S., Onghena, P., Verschaffel, L., & Dooren, W. V. (2013). External Representations For Data Distributions: In Search of Cognitive Fit. *Statistics Education Research Journal, 12*(1), 4-19.

Lowrie, T., Diezmann, C. M., & Logan, T. (2012). A Framework for Mathematics Graphical Tasks: The Influence of the Graphic Element on Student Sense Making. *Mathematics Education Research Journal, 24*(2), 169-187. doi:http://dx.doi.org/10.1007/s13394-012-0036-5

National Council of Teachers of Mathematics (2000). *Principles and Standards for School Mathematics*. Retrieved from Reston, VA: National Council of Teachers of Mathematics.

Mirel, B., Kumar, A., Nong, P., Su, G., & Meng, F. (2016). Using Interactive Data Visualizations for Exploratory Analysis in Undergraduate Genomics Coursework: Field Study Findings and Guidelines. *Journal of Science Education and Technology, 25*, 91-110. doi:10.1007/s10956-015-9579-z

Nicolaou, C. T., Nicolaidou, I. A., Zacharia, Z. C., & Constantinou, C. P. (2007). Enhancing Fourth Graders' Ability to Interpret Graphical Representations Through the Use of Microcomputer-Based Labs Implemented Within an Inquiry-Based Activity Sequence. *The Journal of Computers in Mathematics and Science Teaching, 26*(1), 75-99.

Nolan, D., & Perrett, J. (2016). Teaching and Learning Data Visualization: Ideas and Assignments. *American Statistician, 70*(3), 260-269. doi:10.1080/00031305.2015.1123651

Pfannkuch, M., & Wild, C. (2004). Towards An Understanding of Statistical Thinking. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 17-46). Dordrecht, Netherlands: Kluwer Academic Publishers.

Schau, C. (2003). *Students' Attitudes: The "Other" Important Outcome in Statistics Education*. Paper presented at the Joint Statistics Meetings, San Francisco, CA.

Schau, C. (2017). SATS Scoring. Retrieved from https://www.evaluationandstatistics.com/scoring

Schau, C., Stevens, J., Dauphinee, T. L., & Vecchio, A. D. (1995). The Development and Validation of the Survey of Attitudes Toward Statistics. *Educational and Psychological Measurement, 55*(5), 868-875. doi:http://dx.doi.org/10.1177/0013164495055005022

Selva, A., & Lima, I. (2010). *Youth And Adults Students Interpreting Bar And Line Graphs*. Paper presented at the Eighth International Conference On Teaching Statistics, Ljubljana, Slovenia.

Shah, P., & Hoeffner, J. (2002). Review of Graph Comprehension Research: Implications for Instruction. *Educational Psychology Review, 14*(1), 47-69. doi:http://dx.doi.org/10.1023/A:1013180410169

Watson, J. (2015). Statistical literacy in action. *Australian Primary Mathematics Classroom, 20*(4), 26-30.