Naval Research Laboratory
Washington, DC 20375-5320

NRL/5580/MR—2022/2

Graphicacy in Interactive Dashboards

MARK LIVINGSTON

Information Management Decision Architectures Branch
Information Technology Division

March 1, 2022

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
01-03-2022

2. REPORT TYPE
NRL Memorandum Report

3. DATES COVERED *(From - To)*
10-01-2018 – 02-07-2022

4. TITLE AND SUBTITLE
Graphicacy in Interactive Dashboards

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Mark Livingston

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
1L42

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Research Laboratory
4555 Overlook Avenue, SW
Washington, DC 20375-5320

8. PERFORMING ORGANIZATION REPORT NUMBER
NRL/5580/MR--2022/2

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Office of Naval Research
One Liberty Center
875 N. Randolph St, Suite 1425
Arlington, VA 22203-1995

10. SPONSOR / MONITOR'S ACRONYM(S)
ONR

11. SPONSOR / MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

DISTRIBUTION STATEMENT A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The author of a statistical graph often conveys or emphasizes a particular message through salient cues. This project aimed to characterize the relationship between the salience of features in graphs, to discover the relationship between salience and task performance with statistical graphs, and to determine how certain salient cues affect the complexity and readability of graphs.

15. SUBJECT TERMS

Graph comprehension	Visual salience	Perception
Cognition	Charts, diagrams, and plots	Human subject research

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Mark Livingston
U	U	U	U	14	19b. TELEPHONE NUMBER *(include area code)* (202) 767-0380

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

This page intentionally left blank.

Contents

This page intentionally left blank.

Executive Summary

Statistical graphs are ubiquitous in society, with heavy use in financial reports, general news reporting (especially print and online, but even on television), and in social sciences, to name just a few fields. Despite a plethora of conventional wisdom and general advice available in the popular and academic scientific literature, drawing "good" graphs still seems to be a challenge. One problem with collecting reliable data on the difficulty associated with a graph is the lack of reliable assessment methods. This was a focus of an earlier 6.1 NRL Base Program project. The next step in the process is to emphasize the message that a graph author wishes to convey. Again, the conventional wisdom and advice gives some guidance, but there is conflicting advice and a lack of empirical data. This project aimed to contribute to a solution to the latter problem.

We evaluated approaches to measuring salience of graph components; we found that traditional algorithms for salience measures and even the lone algorithm customized to graphs were lacking. However, the problem proved extraordinarily challenging, and our attempts did not advance the state of the art as much as we had hoped. Our report on this matter cited some reasons for these challenges and gives guidance to future researchers who wish to investigate this.

We next evaluated the effect of some common approaches to making selected elements of a graph more salient; we looked at color/shape cues and at text cues. To our knowledge, our work represents the first principled approach to determining what cues should be employed. This enabled us to claim some "fairness" in comparing different approaches. We found that both text cues and color/shape cues can have positive impacts on user performance (measured by error and response time), and that they did indeed attract attention to the desired areas of the graphs. We next implemented online user studies during the pandemic, studying two aspects of graphs that are believed to increase complexity and divide attention: using multiple data series and multiple dependent axes. We found that the number of points and number of data series can affect reading accuracy and time, and that axis configuration affects reading time.

This line research will hopefully continue to be a topic of investigation; we have discovered a number of fruitful avenues for scientific progress. Although this project is concluding with the close of FY21, there are further plans for data analysis sets. We expect to glean further insights from the existing and yet to be computed statistical results. Even without these finishing touches, this work has enabled us to make principled recommendations for visual representations of data that have been adopted by colleagues in the Navy.

This page intentionally left blank.

Graphicacy in Interactive Dashboards

1 Project Objective

The author of a statistical graph often conveys or emphasizes a particular message through salient cues. We need to characterize the relationship between the salience of features in bar graphs, line graphs, and similar graphs. We also need to discover the relationship between salience and task performance with statistical graphs. From these, we can formulate a theory of how users achieve situation understanding during a workflow through multiple, interactive graphs. We should validate each step using graphs that are realistic for cybersecurity scenarios.

2 Project Background

Statistical graphs are ubiquitous in society, with heavy use in financial reports, general news reporting (especially print and online, but even on television), and in social sciences, to name just a few fields. Further, there are some goals expressed in Department of Education guidance documents that relate to elements of graphicacy. One application domain that conducts a workflow through multiple interactive graphs is cybersecurity watchstanding. Cyberdefense is a highly transactional and multidimensional information domain that is a critical area of concern for the Navy and Marine Corps. While a plethora of graphs representing network events, endpoint events, system logs, configuration state and vulnerability assessments exist, there is no guidance or best practices for tailoring graphs; it merely advises developers and operators to limit parameter values to queries that will not tax the underlying system (e.g. by returning too many records to retrieve within the available time/bandwidth). Developers arbitrarily choose the format of a graph unless a particular format is requested by the user; since neither the user nor the developer are likely to be experts in visual representations, this choice is unlikely to be considered best practice for information visualization. Our objectives will enable us to discover best practices for data display in multidimensional information workflows such as in cyberdefense watchstanding.

All of these fields would benefit from an assessment of whether the main point the graph's author intends to convey is drawn in a way that the reader's attention is drawn to the important information on a graph. Many fields use information dashboards composed of multiple graphs. Our results for workflows will be a model for analysis of the quality of an information dashboard in any domain.

Manuscript approved February 25, 2022.

3 Technical Approach

We are built a library of graphs that represent cybersecurity data visualization. We adopted multiple task taxonomies from the visualization literature and adapted these to fit the tasks that are performed with graphs in cybersecurity monitoring. We began applying models of saliency that were built for natural scenes. After survey a number of cognitive and perceptually-based approaches, we found all of them to perform modestly, but none were particularly impressive. Through colleagues at the Army Research Laboratory, we discovered the Data Visualization Saliency (DVS) algorithm [13] created by researchers at Sandia National Laboratories and began a three-way collaborative project to improve its performance. Data for this work was available from the NRL Military Graphicacy project (6.1 Base Program, FY16-18), through the MASSVIS publicly available data set [2] (collected by a collaborative team of University researchers in Boston), and from a dataset collected (and held) by Sandia National Laboratory. Some graphs were taken from tests of *visualization literacy* (a term closely aligned with *graph comprehension*, our preferred term, or *graph literacy*) offered in academic papers, notably the Visualization Literacy Assessment Test [7] and the Graph Literacy Scale [4]. These tests also gave us a baseline for comparison of the graph reading skills of our users. We implemented an online test for validate our measure of graph comprehension against these measures.

We then looked to survey approaches to vary the salience of elements within those graphs, but found the literature lacking on this point. We introduced a basis for formally determining what should be the focus of a graph to make the important information salient, namely the use of GOMS modeling [3]. GOMS stands for Goals, Operators, Methods, and Selection; it was introduced as a tool for analyzing what a user needs or wishes to do with a visual interface. We then collected a data set that included eye gaze, comprehension measures, and time needed to respond. Unfortunately, this data collection was cut short by the mandatory telework policy required by the global pandemic.

We developed software to analyze the sequence of fixations; this required discovering of appropriate thresholds for how tightly clustered gaze data points must be in screen space in order to be considered a single fixation (essentially, a noise and error tolerance for the eye tracker). It also requires setting a threshold for the minimum time to be considered a fixation; this is a debate that occurs in the academic literature. We also needed to account for various regions of interest that exist in the presentation of the graph. These consist of elements of the graph: the data, the axes (including their labels), the legend, and the title, as well as the question and answer choices. NRL developed all this machinery, based on guidance from the literature, which customized the software to our data.

We lacked the ability under the pandemic situation to perform the tasks envisioned for the final year, which would have entailed another user study using the eye tracker. Therefore, we designed instead a user study to be conducted online that would study what we could study without the use of the eye tracker. This task systematically varied the number of salient elements within a graph and measured the resulting graph comprehension according to our (now-validated) measure. We designed two sets of stimuli to study variables focused on the data displayed within the graph, and another set to study variables focused on the presentation of axes within the graph. An overview of the analysis of these data sets appears in the following section.

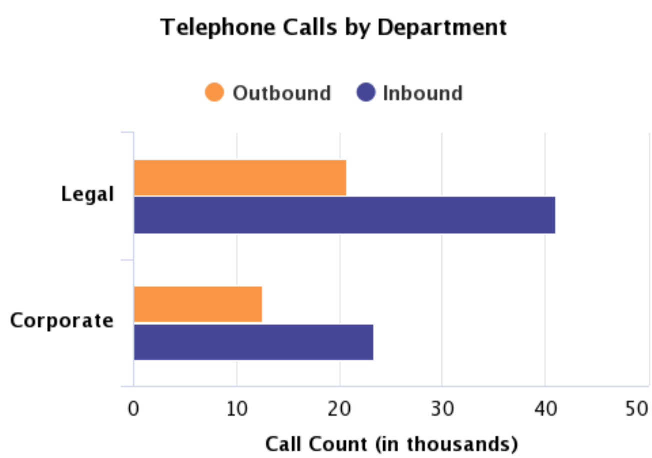
4 Results

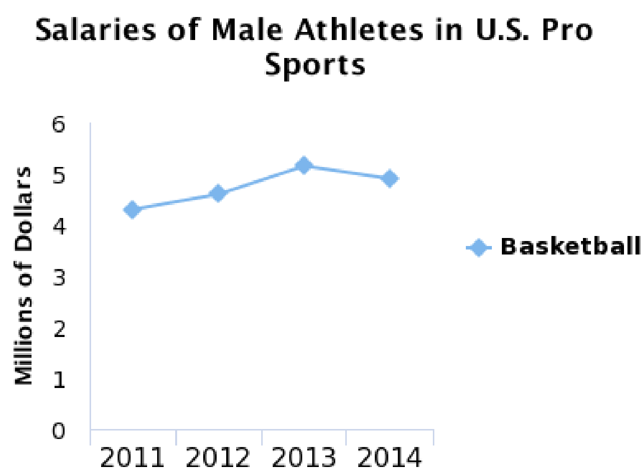
4.1 Accomplishments for FY19

The primary technical success of the first year was an investigation into the quality of various measures of salience in graphs. Quality of salience models is measured by their ability to predict eye tracking data on real graphs. The best performing salience metric was DVS. To our surprise, it uses only the location of text blocks and color-opponency model to identify salient regions. However, using a varied set of eight metrics proposed for measuring the success of prediction of eye tracking, it shows improvement ranging from 9% to 55% over the best of the other models. Our attempts to improve on these yielded small gains, but not the performance we hoped to achieve; the experience did yield some insights into the performance limitations. This work [12] was published in ACM Symposium on Applied Perception, a leading conference for work based on perceptual models.

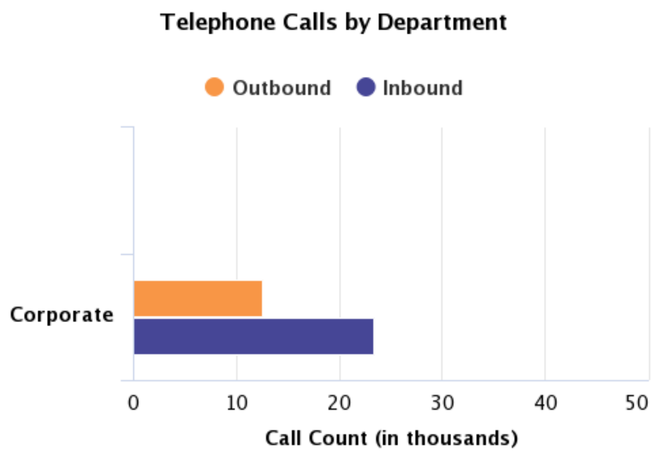
We expanded our library of graphs to include pie charts and scatterplots, tasks for the graphs, and interactions with the graphs. Finally, we began data collection in human subject research on comprehension of the graphs and formalized the process of generating queries into an algorithmic process that requires a minimum amount of author intervention (which necessarily implies author subjectivity). This work [10], which was begun under the Military Graphicacy 6.1 Base Program project, was published in the Applied Human Factors Conference and is the subject of a U.S. Patent Application (Navy Case 107230) [9].

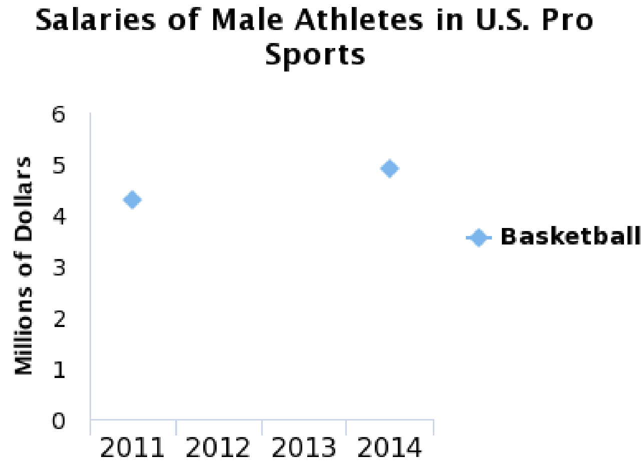
To illustrate the technique, we incorporate here two examples that were used as part of the instructions in a user study that used the algorithmic process to generate stimuli for testing graph comprehension. On the left of each pair, we present a bar graph; on the right is a line graph. The two examples apply different rules in order to create graphs that can test a reader's comprehension. The first pair are source graphs from which queries will be formed in subsequent images.



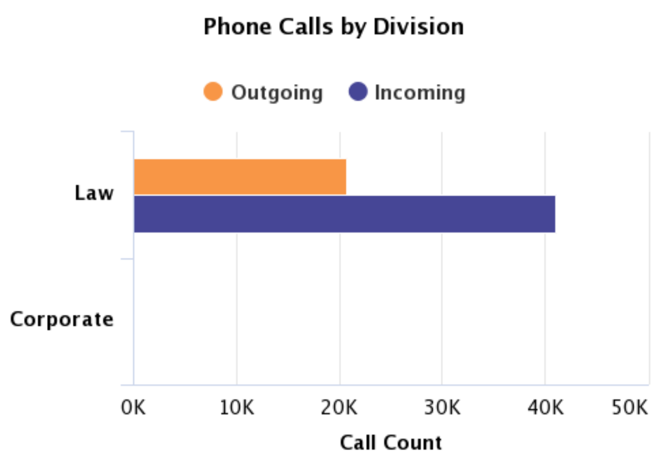


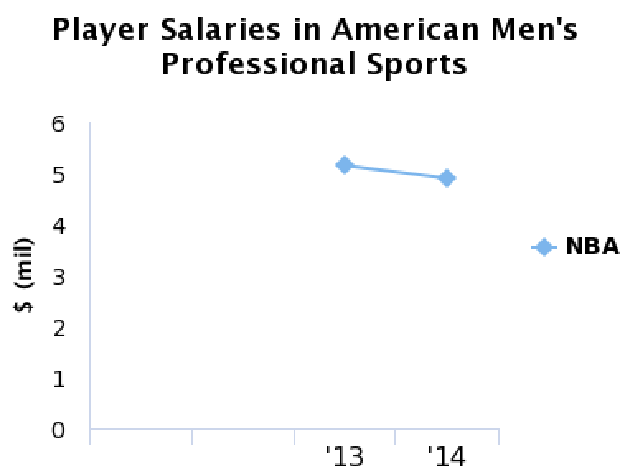
To form graphs that are the basis of queries, we take an excerpt of the graph in various ways. In the first pair, we present graphs that are analogous to verbatim quotes from prose. When we ask a reader whether each of the following graphs give information that was stated on the respective source graphs (above), the response (for each) should clearly be that it does. This is known as an *original* query type and proved to be easier than the other query types (but not trivial). Using a formal specification language for the graph made it easy for us to automate the process of extracting this graph from the source. Although not shown here, we can also engage a system of rules to change the style parameters (color, bar or line width, use of borders, marker type, use of grid lines or tick marks, etc.).



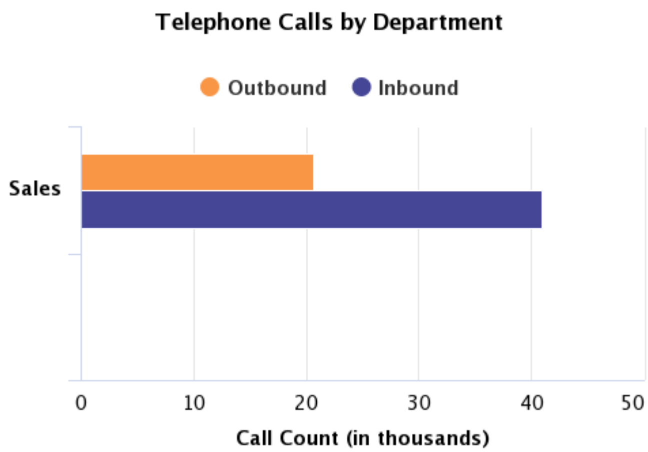


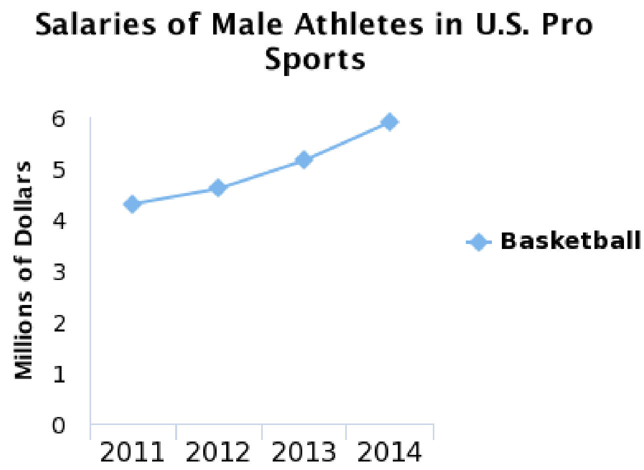
In the second type of query graph, we alter the labels using synonyms or abbreviations. This is known as a *paraphrase* query type; it proved to be much harder to recognize, for obvious reasons. Since the changes to labels do not change the gist of the meaning, and the style parameters do not affect the data, then a reader who truly comprehended the source graph should recognize this graph gives information that was stated on the source graph.



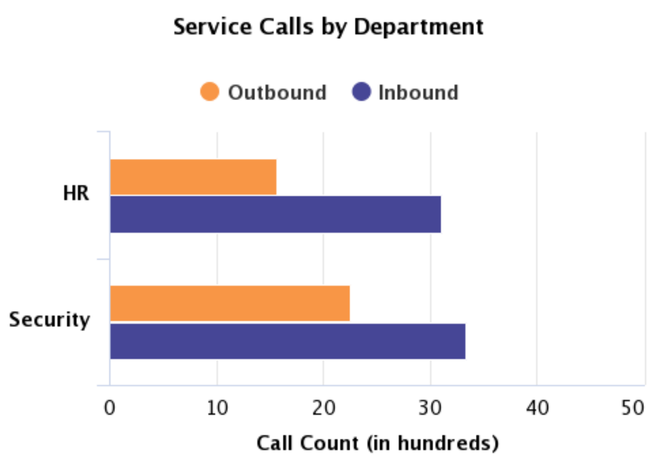


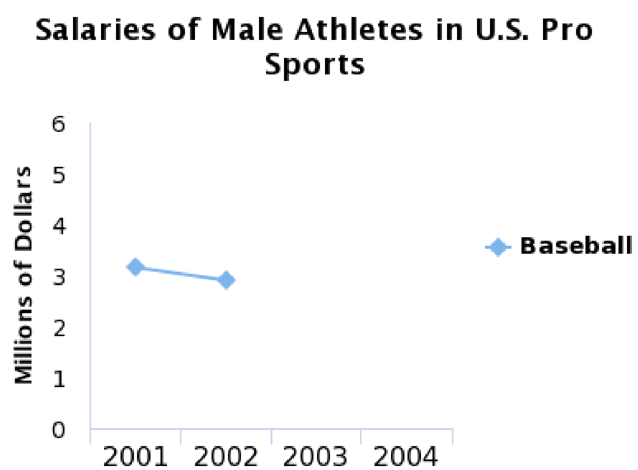
In the third type of query graph, we may move a data point; we settled on an algorithm to move a point one-third of the graph height. This guarantees that any point could move either up or down (perhaps both, in which case the software chooses randomly). We may instead choose (as done below) to change a label. Even though these changes are both simple, either should be enough (in isolation from each other, never in combination) to cause a reader to recognize that this query constitutes a *meaning change* compared to the source and gives information not stated on the source. However, because a single change can be difficult to compare with the representation in memory, this type of query can be quite challenging, and was certainly harder than the *original* or the fourth query type, discussed after the example images.





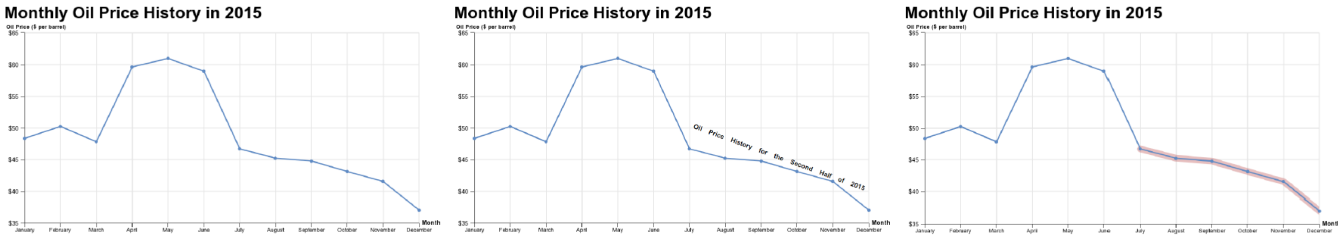
In the fourth and final type of query graph, we may move multiple data points and change multiple labels (including the title of the graph or the axes). Some combination of these changes creates a *distractor* query type. These changes should make the graph below easily recognizable as giving information that was not stated on the source graph. In fact, this query type was generally easier than *paraphrase* or *meaning change*.



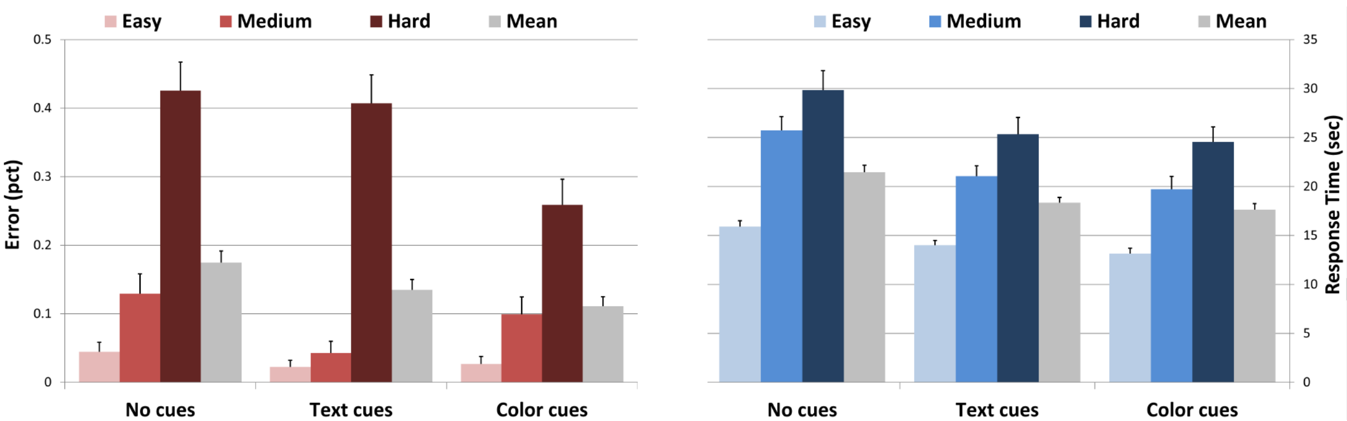


4.2 Accomplishments for FY20

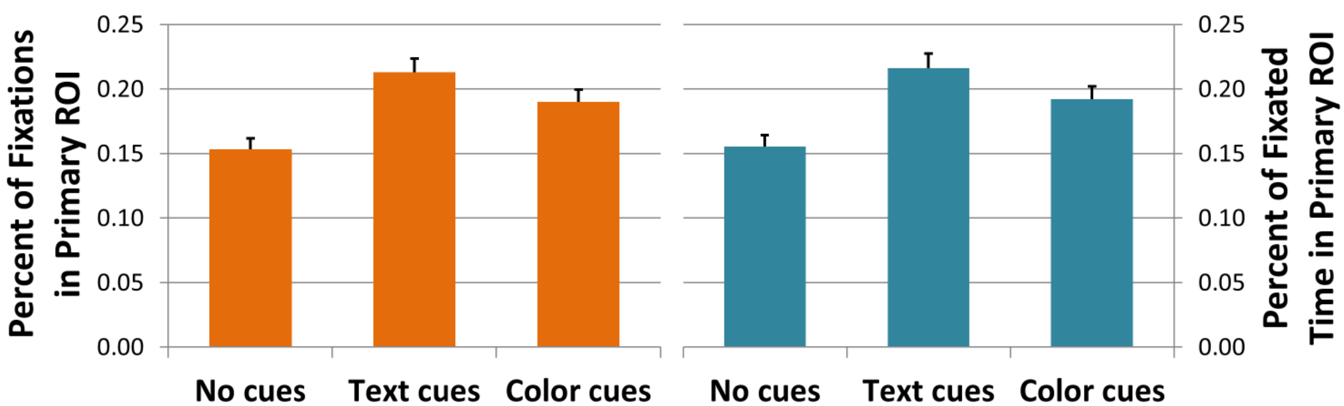
We designed, conducted, and analyzed one human subject research study on the effect of salience. We used the DVS measure to predict gaze, but found it to be of limited value. We adapted a set of tasks encapsulated in the Visualization Literacy Assessment Test (VLAT) as our tasks, built models for their performance in the GOMS paradigm, used these to identify the features and methods to increase saliency, and demonstrated performance improvement on the harder tasks when salient cues were added to the visual representation. (Error dropped from 43% to 26%.) We developed a software suite for the analysis of gaze data (including synchronization with user event data, detection of fixation and appropriate thresholds, design of regions of interest, and whether fixations were inside the various regions of interest). We showed that the salient cues attracted the gaze of study participants. Although the results from one study are preliminary, it does seem that adding salient cues could be important for difficult tasks. We also made some observations using the available literature that support these findings, although this too is based on a small number of findings. A report [11] of this study was presented at the IS&T Visualization and Data Analysis conference in January 2021, where it won the Best Paper Award. The image below gives a simple example of the non-cued, text cue, and color/shape cue version of one graph in the study.



The following two graphs give an overview of the results with respect to error (left) and response time (right).



The following two graphs give an overview of the results with respect to the regions of interest (ROIs) defined to categorize the fixation pattern of the users. The primary ROIs (result in the left graph of the following pair) generally consisted of the relevant data. The secondary ROIs (right graph) generally consisted of the corresponding labels ot that data along one or both axes and/or the legend. The exact configuration of the ROIs depended on the exact phrasing of the question and the solution approach as we modeled it. These models were developed as GOMS models [3], an approach which enabled us to introduce a formalism to applying eye tracking to this question.



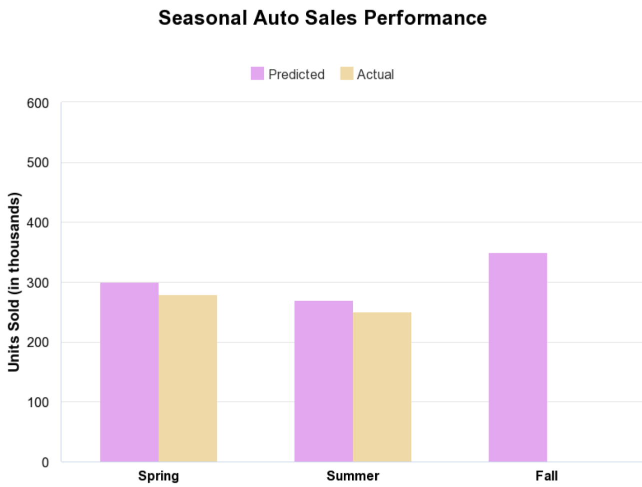
We developed a formal theory of the query generation based on a notion that we define as *visual sentences*. Drawing on the analysis of graph structure by Kosslyn [6] and the linguistics-inspired levels he used (drawn in turn from Kintsch [5]) and the Sentence Verification Technique [15, 14], we identified the minimum collection of elements to express a thought with a graph. This minimum set constitutes the simplest of visual sentences. Just as prose sentences can be simple or complex in structure and express few or many facts or thoughts, so too visual sentences can include many components and express many thoughts. We showed that visual sentences can be constructed in ways that are equivalent to statement forms of the classic taxonomy of tasks (questions) that the existing graph comprehension tests use for construction, which is due to Bertin [1]. This work resulted in a position paper [8] at the IEEE Workshop on Visual Communication, part of the IEEE VIS (Visual Analytics, Information Visualization, and Scientific Visualization) conference.

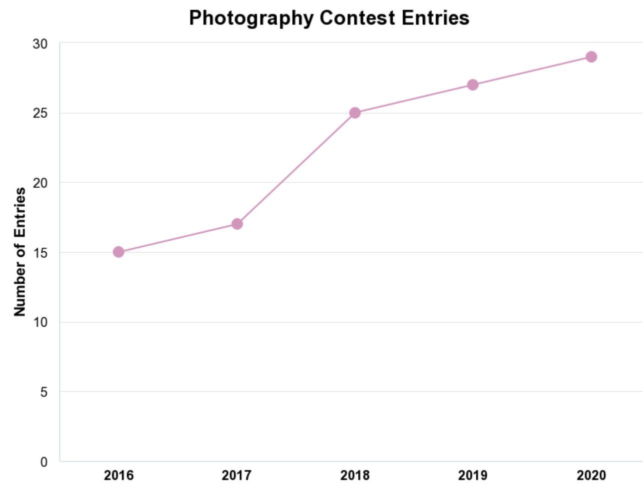
4.3 Accomplishments for FY21

We have been working to extend the analysis to include not just instances of gaze, but sequences of gaze. This required developing extensions to our software base for processing eye tracking data. We developed and used a more detailed collection of regions of interest, incorporated blink data, tried to account for users' views of the keyboard (likely when trying to enter responses with it), and the possibility that a user simply looked away or at blank space to think about a response. Analysis of the gaze is ongoing.

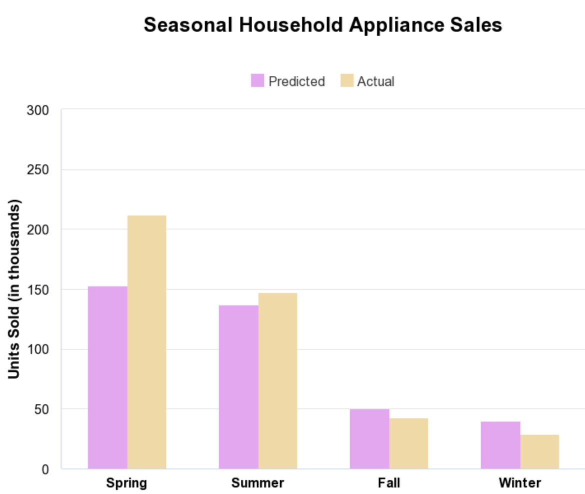
We were fortunate to have a research group at the Univ. of Tennessee at Chattanooga use our publicly-released stimuli for our online study and replicate our study. They offered the anonymized data set to us for analysis alongside our own data. While the analysis is ongoing, this has helped give us further evidence of the validity of our initial queries as measures of graph comprehension. It has shown that they are generally equivalent to the tests developed for academic literature; those are static tests that took months of development and validation for a set of questions that (in some cases, we have demonstrated) are of questionable validity (despite the statistical evidence offered by the respective authors).

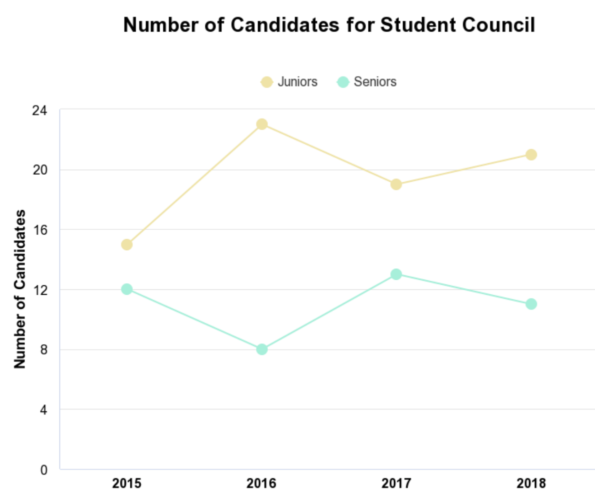
We designed, implemented, and conducted a new study that will look at the effect of the number of items on the graph, which is an aspect of the salience. Two images of graphs from the tutorial portion of the study are shown below and will help illuminate the following description of stimuli. We manipulated salience of the items by increasing the number of items present on a graph; this study was a substitute for the one originally envisioned, which would have used the eye tracker to follow gaze around a dashboard. This study will yield further insight into the effects of the salience of graph items caused by the tendency to "overload" data onto a graph. While a thorough analysis of the data is forthcoming, we can state some basic results from initial analysis. There were two independent variables of primary interest, the number of points on a graph and the number of data series into which they were organized. We drew graphs for source reading material that contained three, seven, or eleven points (five in the tutorial images below); these were either drawn as part of one data series (as in the line graph example) or divided into two series (as in the bar graph example). We found that the number of data series had a significant main effect on the response time and the time needed to study the graph (in order to feel prepared to respond to a question accurately); however, it did not have a main effect on the accuracy of the responses. The number of points, surprisingly, did not have a main effect on the accuracy. It did not have a main effect on the response time (perhaps not surprising, since all questions features exactly two points on the graph shown as part of the query). It did (also not surprisingly) have a main effect on the time respondents chose to study the source material in order to feel prepared to answer a question accurately. (Questions were shown only after a graph had been studied; the graph could not be brought back for reference.) We further observed an interaction effect between the number of data series and number of points for response time. In summary, it appears that the total number of points and whether they are part of one logical sequence or split into two affects the time needed to study a graph and have complex effects on the time needed to response to a query about the graph. But we found no evidence that it affects the accuracy of that response. We note that our volunteers overwhelmingly were experienced with graphs and of strong educational background. There are additional effects that we observed from educational background and from how points were chosen for the query graphs.

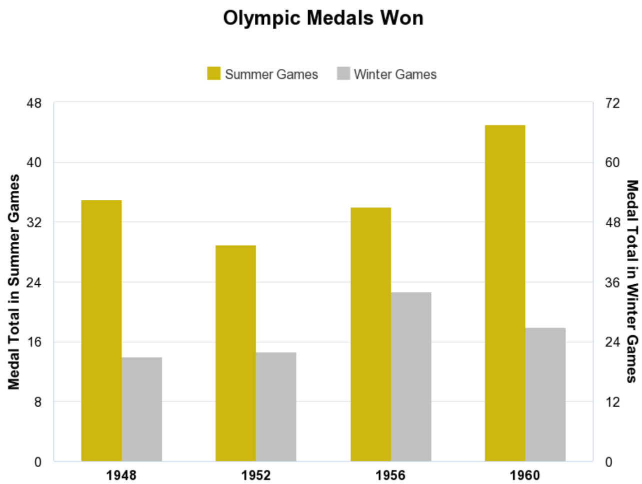


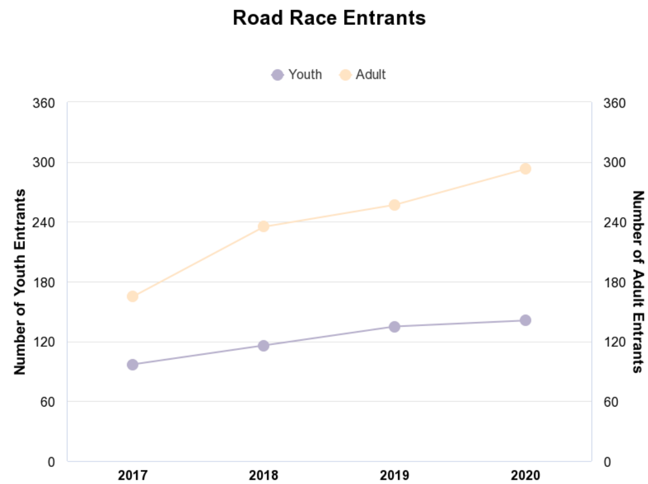


In a second study using the same framework and methodology, we studied the effect of different configurations of dependent axes. All source reading graphs in this study contained two data series, each of which contained four points (for a total of eight per graph). The source and query graphs were then configured to be drawn with one dependent axis (on the left, as in the top line of examples below), two identically drawn dependent axes (one on each side, with separate titles, as in the line graph example at the bottom right below), or two different dependent axes (with separate titles, as in the bar graph example on the bottom left below). To our surprise, the configuration of axes had no main effect on either the accuracy of, or the time needed for, the response. It did, however, have the expected effect on the study time of the graph; graph readers spent more time studying graphs with the different dependent axes than the graphs with the identically drawn axes, which in turn they spent more time studying than those with a single dependent axis. This makes sense from the standpoint of simply needing more time to process the greater number of symbols and labels present. Again in this data set, expertise with graphs (and its likely proxy, educational background) showed a main effect on the error rate; however, there was no interaction with the axis configuration.









These results from FY21 give us insight into the aspects of graph configuration that make graph readers need more time to feel that are confident that they have extracted all the information from a graph. While we have no evidence from these studies to suggest that the levels of difficulty we implemented with our source and query graphs actually affect their capabilities, we note that our pool of volunteers consisted almost entirely of people who demonstrated high competency in graph comprehension. We also note that we used only simple bar and line graphs; we cannot make any statements about other types of graphs. We plan to continue to mine the data for greater insights into these aspects of what makes graphs difficult to understand.

5 Associations and Outputs

5.1 Associated Base Program Projects

This work in many ways built on the "Military Graphicacy" Base Program project (FY16-FY18, Work Unit 1E32). That was in turn based in part on the Base Program project "Visualization of Cognitively Inspired Autonomous Decision Processes" (FY12-FY14, Work Unit 4652).

5.2 Associated External Organizations

We have worked extensively with the Army Research Laboratory and Sandia National Laboratories on the issue of how to measure salience in statistical graphs. In the course of the project, we have presented overviews to the Office of Naval Research (Basic Research Forum, 31 May 2018) and contributed ideas based on this work to white papers submitted through the DoD Human Systems Community of Interest and the DoD C4I Community of Interest. We currently have white papers under consideration at the Army Research Institute and the DoD Human Resources Activity. We also contributed a brief overview of portions of the work to the Workshop on Visualization for Scientific Discovery, Decision-Making, & Communication, sponsored by the Department of Energy's Advanced Scientific Computing Research program.

5.3 Publications and Patents

[1] Mark A. Livingston, Derek Brock, Jonathan W. Decker, Dennis J. Perzanowski, Christopher van Dolson, Joseph Mathews, and Alexander S. Lulushi. A query generation technique for measuring

comprehension of statistical graphics. In *Advances in Human Factors in Simulation and Modeling: Proceedings of Applied Human Factors and Ergonomics 2019*. Springer, 2020. Extended technical report version available.

[2] Mark A. Livingston, Derek Brock, Jonathan W. Decker, Dennis J. Perzanowski, Christopher Van Dolson, Joseph Mathews, and Alexander Lulushi. Sentence verification for statistical graphs. U.S. Patent Application, June 2020.

[3] Mark A. Livingston, Laura E. Matzen, Andre Harrison, Alex Lulushi, Mikaila Daniel, Megan Dass, Derek Brock, and Jonathan W. Decker. A study of perceptual and cognitive models applied to prediction of eye gaze within statistical graphs. In *Proceedings of ACM Symposium on Applied Perception*, September 2020.

[4] Mark A. Livingston and Derek Brock. Position: Visual sentences: Definitions and applications. In *IEEE Workshop on Visual Communication, part of IEEE VIS*, October 2020.

[5] Mark A. Livingston, Laura Matzen, Derek Brock, Andre Harrison, and Jonathan W. Decker. Testing the value of salience in statistical graphs. In *Proceedings of IS&T Visualization and Data Analysis (part of Electronic Imaging)*, January 2021.

[6] Mark A. Livingston and Derek Brock. Measuring Correct Interpretation of Graphs, Figures, and Tables. In *Workshop on Visualization for Scientific Discovery, Decision-Making, & Communication, Department of Energy Office of Advanced Scientific Computing Research*, 18-20 January 2022.

Acknowledgments

NRL team members who contributed to this work include Derek Brock (volunteer emeritus), Autumn Toney, Chistopher van Dolson, Jonathan W. Decker, and Esmé Middaugh. Numerous interns made contributions through the Science and Engineering Apprenticeship Program (Alex Lulushi, Megan Dass, Shyla Bisht), Naval Research Enterprise Internship Program (Arjun Mayur, Evan North, Claire Tector, Stacy Stoffregen, Elaine Jackson), and Historically Black Colleges and Universities / Minority Institution program (Mikaila Daniel, Nick Cummings, Jean Jerome, Jalen Davis).

References

[1] Jacques Bertin with M. Barbut et al. *Sémiologie Graphique: Les diagrammes, les réseaux, les cartes;, revised edition.* Gauthier-Villars, 1973. Translated as "Semiology of Graphics" by William J. Berg, Univ. of Wisconsin Press, 1983.

[2] Michelle Borkin, Zoya Bylinskii, Nam Wook Kim, Constance May Bainbridge, Chelsea S. Yeh, Daniel Borkin, Hanspeter Pfister, and Aude Oliva. Beyond memorability: Visualization recognition and recall. *IEEE Trans. on Visualization and Computer Graphics*, 22(1):519–528, January 2016.

[3] Stuart Card, Thomas P. Moran, and Allen Newell. *The Psychology of Human Computer Interaction.* Lawrence Erlbaum Assoc., 1983.

[4] Mirta Galesic and Rocio Garcia-Retamero. Graph literacy: A cross-cultural comparison. *Medical Decision Making*, 31(3):444–457, May/June 2011.

[5] Walter Kintsch and Praful Mangalath. The construction of meaning. *Topics in Cognitive Science*, 3(2):346–370, April 2011.

[6] Stephen M. Kosslyn. Understanding charts and graphs. *Applied Cognitive Psychology*, 3(3):185–225, July/September 1989.

[7] Sukwon Lee, Sung Hee Kim, and Bum Chul Kwon. VLAT: Development of a visualization literacy assessment test. *IEEE Trans. on Visualization and Computer Graphics*, 23(1):551–560, January 2017.

[8] Mark A. Livingston and Derek Brock. Position: Visual sentences: Definitions and applications. In *IEEE Workshop on Visual Communication, part of IEEE VIS*, October 2020.

[9] Mark A. Livingston, Derek Brock, Jonathan W. Decker, Dennis J. Perzanowski, Christopher Van Dolson, Joseph Mathews, and Alexander Lulushi. Sentence verification for statistical graphs. U.S. Patent Application, June 2020.

[10] Mark A. Livingston, Derek Brock, Jonathan W. Decker, Dennis J. Perzanowski, Christopher van Dolson, Joseph Mathews, and Alexander S. Lulushi. A query generation technique for measuring comprehension of statistical graphics. In *Advances in Human Factors in Simulation and Modeling: Proceedings of Applied Human Factors and Ergonomics 2019*. Springer, 2020.

[11] Mark A. Livingston, Laura Matzen, Derek Brock, Andre Harrison, and Jonathan W. Decker. Testing the value of salience in statistical graphs. In *Proceedings of IS&T Visualization and Data Analysis (part of Electronic Imaging)*, January 2021.

[12] Mark A. Livingston, Laura E. Matzen, Andre Harrison, Alex Lulushi, Mikaila Daniel, Megan Dass, Derek Brock, and Jonathan W. Decker. A study of perceptual and cognitive models applied to prediction of eye gaze within statistical graphs. In *Proceedings of ACM Symposium on Applied Perception*, September 2020.

[13] Laura E. Matzen, Michael J. Haass, Kristin M. Divis, Zhiyuan Wang, and Andrew T. Wilson. Data visualization saliency model: A tool for evaluating abstract data visualizations. *IEEE Trans. on Visualization and Computer Graphics*, 24(1):563–573, January 2018.

[14] James M. Royer. Developing reading and listening comprehension tests based on the sentence verification technique (SVT). *Journal of Adolescent & Adult Literacy*, 45(1):30–41, September 2001.

[15] James M. Royer, C. Nicholas Hastings, and Colin Hook. A sentence verification technique for measuring reading comprehension. *Journal of Reading Behavior*, 11(4):355–363, December 1979.