Data Visualization Use in JASIST: An Exploratory Look Through Years 2001- 2021

Jeffcoat, S. O. SUNY Albany, USA | sjeffcoat@albany.edu
Rogers, Monica SUNY Buffalo, USA | rogers36@buffalo.edu

ABSTRACT
Broadly, the information science field has grown to encompass a more extensive set of data science competencies as the role of data, data literacy, and data visualization has grown. Schwabish explains, "now more than ever, content must be visual if it is to travel far" (2021). Through coding selected years across 20 years of JASIST journal, from the years 2001, 2006, 2011, 2016, and 2021; we were able to not only inventory the proportion of images that are data visualizations but also to identify trends and changes over time in the data visualizations contained in JASIST publications. Findings suggested that overall, the importance and complexity of data visualizations in publications increased over the date range studied. Additionally, as data visualizations were used more often, they were observed to become more complex. Still, once distilled into their simplest forms, a limited variety of data visualizations was used across the dataset. Therefore, there is a clear need for increased data literacy targeted toward complex data visualizations. Furthermore, opportunities exist to improve the types of data visualizations used in academic writing.

KEYWORDS
Data Visualization; Data Literacy; Data Comprehension

INTRODUCTION
Data visualization, which transforms abstract data into physical visions (for example, length, position, shape, color, and so on), is a powerful means to present compelling stories of data to humans. The uses of data visualizations vary. One idea states, "the fundamental goal of visualizations is to generate a particular insight or to execute a specific task by emphasizing distinct features of the underlying dataset" (Lurie and Mason 2007; Anderson et al. 2011). These insights provided by data visualizations can lead to researchers being able to generate or investigate hypotheses. Additionally, Brehmer and Munzner explain, data visualizations can act as a, "presentation of information to a particular audience by telling a persuasive and data-supported story for decision-making purposes" (Brehmer and Munzner, 2013). Numerous studies underline the emerging need for data competencies, data services, and visualization within the information science field. Since 2015, the amount of research publications on data science topics within information science journals has been steadily increasing (Virkus & Garoufallou, 2019). Broadly, the information science field has grown to encompass a more extensive set of data science competencies as the role of data, data literacy, and data visualization has grown. Schwabish explains, "now more than ever, content must be visual if it is to travel far" (2021). Readers are overwhelmed with information in the forms of data, news, and text. Visuals make it easier for readers to recognize and recall information. Yet, many researchers were never taught how to present their work visually (Schwabish, 2021).

The Journal of the Association for Information Science and Technology
The Journal of the Association for Information Science and Technology (JASIST) was established in 1950 as an American Documentation Quarterly. The new journal was a publication of the American Documentation Institute (ADI), which had formed in 1937 around a group of researchers and practitioners interested in the emerging technology of microfilm as a medium for the preservation and dissemination of documents and knowledge (Tate, 1950). With an impact factor of 2.678, JASIST has played a "vital role" in the dissemination of scholarly articles in libraries and information science since 1950" (Mukherjee, 2006). JASIST has committed to providing, "intellectual leadership by publishing original research that focuses on the production, discovery, recording, storage, representation, retrieval, presentation, manipulation, dissemination, use, and evaluation of information and on the tools and techniques associated with these processes" (About the Journal, 2022). This commitment makes the publication a representative selection for evaluating data visualization trends in the information science community.

RESEARCH QUESTION AND SIGNIFICANCE
Question 1: In selected years of JASIST articles between 2001- 2021, what proportion of images are data visualizations?

Three categories of images were established: Data visualizations, Information visualizations, and Visual information. Additionally, uncodeable (those which did not fit into the established definitions) data visualizations were coded as such. The "Defining Visualization Types" section below presents the definitions of each of these categories.

Question 2: What are the trends and changes in the data visualizations in the articles?

Mid-Year Conference of the Association for Information Science & Technology | April 11-13, 2023, Virtual Event. Author(s) retain copyright, but ASIS&T receives an exclusive publication license.

Our research takes an inductive approach- measuring the use of data visualizations in JASIST, a top information science journal. Using a comparative study of the articles published between 2001 and 2021, we were able to make a series of conclusions about the use of data visualizations in JASIST, which could be interpreted as representative of data visualization within the broader information science discipline. This research is essential in that it adds to the literature in that it considers, analyzes, and evaluates how or the frequency with which data visualizations are used in academic journals. Additionally, this contribution to the literature is the first known study of the use of data visualizations in JASIST spanning the years 2001- 2021.

LIMITATIONS

Limitations existed, whereas the corpus contained only selected years across 20 years of articles from 2001, 2006, 2011, 2016, and 2021. Each author's determination of "table" or "figure" designations for submissions to JASIST is made independently. As JASIST does not provide a formal naming structure, tables and figures are more likely misnamed or interchangeably named by authors. These inconsistencies presented a coding challenge as the visual interpretation occasionally does not match the title images used to discern "tables" from "figures." Researcher bias was addressed by implementing an inter-coder reliability process and using an outside source for the visual dictionary/code book of identifiable data visualizations.

DEFINING VISUALIZATION TYPES

Data visualization

Oxford's Lexicon dictionary defines data visualization as, "the representation of information in the form of a chart, diagram, picture, etc." and secondarily, "a chart, diagram, picture, etc. that is created as a visual representation of information" (Data Visualization English Definition and Meaning | Lexico.Com, n.d.). In this study, the term is best understood to describe a graph, table, or chart created to explain a scientific phenomenon and which employs data collected during the research process described in the publication.

Information visualization

The term "information visualization" does not have a specific definition and instead includes other suggested search terms such as "visualization, "informatization,' 'information,' 'pre-visualization' and, 'data visualization' (Information Visualization English Definition and Meaning | Lexico.Com, n.d.). Spence offers an explanation of what information visualization is not—scientific visualization. Scientific visualizations, "relate to and represent visually, something 'physical'. Thus the flow of water in a pipe, the nature of the weather in a mountainous area and the stresses in a girder are usually—and usefully—displayed directly superimposed on or at least close to, a realistic representation of the physical thing. By contrast, information visualization deals with abstract quantities such as baseball scores, connections between known criminals, fluctuating exchange rates, and electrical voltages and is used as a tool to enhance a human being's acquisition of insight into abstract ideas" (Spence, 2006). In this study, information visualization describes a figure that displays information visualized to explain a process. An example is a flow chart demonstrating a theoretical or physical function.

Visual information

The term "visual information" includes other suggested search terms such as "information," 'informational' "informatization,' 'non- information," and lastly, 'informatisation' (Visual Information English Definition and Meaning | Lexico.com, n.d.). Together with text and audio information, visual information is considered a form of unstructured data (Sudhir, 2016). Rizkallah (2017) explains that unstructured data includes documents, messages, social media posts, pictures, videos, and audio. In this study, the term is best understood to describe an observation documented as a visual example. This image has not been altered to include data collected through the research related to the published article.

METHODS

Data collection was conducted by identifying member-accessible JASIST journal publications via the JASIST journal repository. Using a structured approach, the JASIST database was searched for 2001 to 2021. Journals published every five years between this date range were selected. JASIST publications from 2001, 2006, 2011,

2016, and 2021 were selected. A total of 925 articles were included in our initial corpus. There were no exclusion criteria applied. Therefore non-article journal submissions were also analyzed, including book reviews, letters from the editor, and erratum, which contained no data or visualizations. This inclusive analysis resulted in fewer selection biases and a complete understanding of the data and research question. A review of the results of each article led to the creation of a codebook. We began by using the Visualization Literacy Assessment Test (VLAT) as a validated data visualization tool upon which to base the data visualization codebook developed. While the VLAT was sufficient in the early stages of data analysis, an alternative data visualization reference tool was added as the complexity of the data visualizations observed surpassed the 12 data visualization types included in the VLAT test. Ultimately, the Financial Times' Visual Vocabulary was utilized to code the data visualizations in the analyzed articles. The Visual Vocabulary is a tool developed in 2001 to act as a starting point for visualizing data imperative to the story being told by the research (Smith et al., 2021). The story aspect of each data visualization was necessary as the coding process sometimes required determining the author's intent to distinguish one data visualization type from another.

We established and confirmed the coding rules during a series of meetings over a one year period. Coding practices were primarily determined based on coding conventions proven successful in previous studies. After establishing coding rules, a codebook was developed, and The authors decided to code all articles from 2001, 2006, 2011, 2016, and 2021. All figures and tables were counted and coded for composition. Tables were calculated and categorized as "data" or "text." Figures were estimated and coded as being composed of data visualization(s), visual information, or information visualization(s) or determined to be "uncodeable." We then identified data visualizations using the Visual Vocabulary. Coding procedures were modeled after the social media analysis of Lovejoy and Saxton (2012). Both authors began coding the articles in the first and last publication years (2001 and 2021). As noted by Lovejoy and Saxton (2012), "discrepancies between codings were discussed and coding rules refined until 100% agreement was reached." Using the increasingly refined rules, the last three publication years were coded (2006, 2011, 2016). Discrepancies were again discussed and agreed upon before moving forward. Finally, a neutral third reviewer was asked to code a subset of the articles. An 83.6% inter-coder agreement and a Holsti's Method score of .836 were reached, indicating a high level of intercoder reliability.

EMPIRICAL DATA

For all years analyzed (See Table 1), combined there were a total of 4779 figures identified, with 961 (20%) being information visualizations, 653 (14%) were visual information, 3047 (645) were data visualizations, and only 166 (3%) were uncodeable. One notable trend was that the percentage of data visualization figures started lower in 2001 at only 49%, increasing to 73% by 2011 and fluctuating to 65% in 2016 but back up to 68% in 2021.

Year	# of figures	# of info viz	% info viz	# visual info	% viz info	# uncodeable	% uncodeable	# of data viz	% data viz	# unique data viz types used
2001	496	120	24%	120	24%	16	3%	244	49%	16
2006	877	226	26%	176	20%	25	3%	475	54%	23
2011	1106	187	17%	125	11%	45	4%	805	73%	32
2016	1461	288	20%	162	11%	79	5%	956	65%	37
2021	839	140	17%	70	8%	1	0%	567	68%	29
Total	4779	961	20%	653	14%	166	3%	3047	64%	

Table 1. Classifications of Figures Coded

We identified forty-seven unique data visualizations from the codebook as being used in the papers analyzed (See Table 2). Only ten data visualization types were used in each of the five years of documents included in this analysis. They were: bar, bubble, connected scatterplot, histogram, line, network, pie, scatterplot, stacked bar, and Venn. Six data visualizations were used throughout four publication years, 12 were used throughout three publication years, six were used throughout two publication years, and 13 were used throughout only one year.

Unique data visualization type		Number of years used
line	stacked bar	5
scatterplot	bar	
connected scatterplot	pie	
network	venn	
histogram	bubble	
column	dot strip plot	4
paired column	cumulative curve	
boxplot	diverging bar	
frequency polygons	slope	3
seismogram	dot density	
heat map	chord	
dot plot	marimekko	
radar	column line timeline	
beeswarm	xy heatmap	
ordered column	candlestick	2
gridplot	flow map	
bump	stacked area	
sankey	basic choropleth	1
area	density plot	
calendar heatmap	lollipop	
contour map	ordered proportional symbol	
spine	streamgraph	
violin plot	waterfall	
vertical timeline		

Table 2. Unique Data Visualizations Recorded

Only 21 of the 47 identified data visualizations were used frequently enough to constitute at least 1% or more of the 4779 visualizations used. The least used 25 visualizations were used a combined 131 times. The visualizations used frequently enough to constitute at least 1% of all visualizations are presented in Table 3.

Data viz Type	Number of years appeared in	total of viz type across all years	as percent of all data viz's
line	5	658	23%
scatterplot	5	372	13%
connected scatterplot	5	312	11%
column	4	221	8%
paired column	4	212	7%
network	5	128	5%
histogram	5	112	4%
stacked bar	5	110	4%
bar	5	96	3%
pie	5	92	3%
boxplot	4	81	3%
frequency polygons	3	45	2%
venn	5	42	1%
bubble	5	37	1%
seismogram	3	33	1%
heat map	3	29	1%
dot strip plot	4	27	1%
cumulative curve	4	22	1%
dot plot	3	20	1%
radar	3	18	1%
beeswarm	3	17	1%

Table 3. The Most Frequently Used Visualizations

CONCLUSIONS

In 2001 there were only 16 unique data visualization types used. This number increased as years passed, with a peak of 37 unique data visualization types used in 2016 and settling at 29 unique data visualization types used in 2021. These finds suggest that, overall, data visualizations' importance and complexity were increasing in publications. Additionally, data visualization literacy for the authors and the readers (in this case, both reviewers and readers) increased. These findings align with the idea that "visualization literacy,' is an emerging literacy and is gaining importance within the library and information science field (Wright et al., 2012; Carlson et al., 2015; Federer, 2018)

As authors used data visualizations more often, they became more complex. Although it became apparent that researchers began to utilize standardized templates provided by popular software such as Microsoft Excel to generate data visualizations, advances in technology have allowed for the automated generation of data visualizations that far surpass what was capable in 2001. Previous research further suggests that charts are no longer simplistic, making it increasingly challenging for viewers to, "effectively perceive the displayed data" (Reitsma, 2019). An increased understanding by viewers of deciphering data visualizations is imperative in advancing transparency in the sciences.

Surprisingly, the authors employed a limited variety of data visualizations across the dataset. After distilling the data visualizations coded into their elemental compositions based on use, several patterns were easily identifiable. Line charts were used the most often, followed by scatterplots and connected scatterplots. Variety determinations were based on similarity. Connected scatterplots and line charts are functionally similar. Column and paired columns charts are similar and functionally the same as bar and paired bar charts. Histograms are similar in use to column and bar charts. With this in mind, this analysis shows that only four main visualizations were used most frequently: scatterplots, line charts, column charts, and networks.

The findings suggest the need for a more directed approach to increasing the types of data visualizations used in academic writing. Utilization of a more diverse range of data visualizations will increase the reader's capacity to comprehend the presented data. For instance, while a column chart may technically show the varied ages of a population, data visualizations such as an isotype (pictogram) or grouped symbol chart better depict the story of the generations represented in the data. Future research will be directed towards extracting further insight regarding data visualizations seen in 2021 articles. Despite COVID- 19 related research, on average, authors utilized fewer data visualizations in papers in 2020 than in 2016. This phenomenon will need to be investigated with a more profound analysis. Subsequent research will also explore trends found among the uncodeable figures and propose suggestions for expanding the data visualizations used by researchers.

REFERENCES

About the Journal. (2022). The Journal of the Association for Information Science and Technology (JASIST). https://asistdl.onlinelibrary.wiley.com/journal/23301643

Anderson, E. W., Potter, K. C., Matzen, L. E., Shepherd, J. F., Preston, G. A., & Silva, C. T. (2011). A User Study of Visualization Effectiveness Using EEG and Cognitive Load. Computer Graphics Forum, 30(3), 791–800. https://doi.org/10.1111/j.1467-8659.2011.01928.x

Bevan, N., Kirakowski, J., & Maissel, J. (1991). What is Usability? Proceedings of the 4th International Conference on HCI. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.630.1555&rep=rep1&type=pdf

Brehmer, M., & Munzner, T. (2013). A Multi-Level Typology of Abstract Visualization Tasks. IEEE Transactions on Visualization and Computer Graphics, 19(12), 2376–2385. https://doi.org/10.1109/TVCG.2013.124

Carlson, J., Research D., Nelson, M. S., Associate, Johnston, L. R., Research D., & Koshoffer, A., Science In. (2015). Developing Data Literacy Programs: Working with Faculty, Graduate Students and Undergraduates. Bulletin of the Association for Information Science and Technology, 41(6), 14–17

Cawthon, N., & Moere, A. V. (2007). The Effect of Aesthetic on the Usability of Data Visualization. 2007 11th International Conference Information Visualization (IV '07), 637–648. https://doi.org/10.1109/IV.2007.147 Chart English Definition and Meaning | Lexico.com. (n.d.). Lexico Dictionaries | English. Retrieved June 14, 2022, from https://www.lexico.com/search?utf8=%E2%9C%93&filter=noad&dictionary=en&query=information+visualization

Data Visualization English Definition and Meaning | Lexico.com. (n.d.). Lexico Dictionaries | English. Retrieved June 14, 2022, from https://www.lexico.com/en/definition/data_visualization

DATA VISUALIZATION English Definition and Meaning | Lexico.com. (n.d.). Lexico Dictionaries | English. Retrieved May 3, 2022, from https://www.lexico.com/en/definition/data_visualization

D'Ignazio, C., & Klein, L. F. (2020). Data feminism. The MIT Press.

Du Bois, W. E. B., Battle-Baptiste, W., & Rusert, B. (2018). W.E.B Du Bois's data portraits: visualizing Black America (First edition). The W.E.B. Du Bois Center At the University of Massachusetts Amherst; Princeton Architectural Press. https://ebookcentral.proquest.com/lib/albanyedu-ebooks/reader.action?docID=5515147

Duffy, T., & Waller, R. (1985). Designing Usable Texts. Orlando: Academic Press.

FIGURE English Definition and Meaning | Lexico.com. (n.d.). Lexico Dictionaries | English. Retrieved May 3, 2022, from https://www.lexico.com/en/definition/figure

Federer, L. (2018). Defining data librarianship: a survey of competencies, skills, and training. Journal of the Medical Library Association: JMLA, 106(3), 294–303.

Goldberg, K. L. (2020). Adopting analytics: A framework for nonprofit organizations. University of Maryland University College.

Graph English Definition and Meaning | Lexico.com. (n.d.). Lexico Dictionaries | English. Retrieved June 14, 2022, from https://www.lexico.com/search?utf8=%E2%9C%93&filter=noad&dictionary=en&query=information+visualization

Gu, Z. (2022). Complex heatmap visualization. IMeta, 1(3). https://doi.org/10.1002/imt2.43 Hemalata Iyer. (n.d.). Core Competencies for Visual Resources Management. Institute of Museum and Library Services (IMLS), 8. http://www.vraweb.org/wp-content/uploads/2016/09/iyer_core_competencies.pdf

Hemalata Iyer. (2009a). A Profession in Transition: Towards Development and Implementation of Standards for Visual Resources Management. Part A - The Organization's Perspective. Information Research, 14(3). http://www.informationr.net/ir/14-3/paper412.html

Hemalata Iyer. (2009b). A profession in transition: towards development and implementation of standards for visual resources management. Part B - the professional's perspective and beyond. Information Research, 14(4). http://www.informationr.net/ir/14-4/paper413.html

Information Visualization English Definition and Meaning | Lexico.com. (n.d.). Lexico Dictionaries | English. Retrieved June 14, 2022, from https://www.lexico.com/search?utf8=%E2%9C%93&filter=noad&dictionary=en&query=information+visualization

JOURNAL English Definition and Meaning | Lexico.com. (n.d.). Lexico Dictionaries | English. Retrieved May 3, 2022, from https://www.lexico.com/en/definition/journal

Lee, B., Brown, D., Lee, B., Hurter, C., Drucker, S., & Dwyer, T. (2020). Data Visceralization: Enabling Deeper Understanding of Data Using Virtual Reality. https://doi.org/10.48550/ARXIV.2009.00059

Lee, T. (David), Lee-Geiller, S., & Lee, B.-K. (2020). Are pictures worth a thousand words? The effect of information presentation type on citizen perceptions of government websites. Government Information Quarterly, 37(3), 101482. https://doi.org/10.1016/j.giq.2020.101482

Lovejoy, K., & Saxton, G. D. (2012). Information, Community, and Action: How Nonprofit Organizations Use Social Media*. Journal of Computer-Mediated Communication, 17(3), 337–353. https://doi.org/10.1111/j.1083-6101.2012.01576.x

Mukherjee, B. (2009). Journal of the American Society for Information Science and Technology (2000—2007): a bibliometric study. IFLA Journal, 35(4), 341–358. https://doi.org/10.1177/0340035209352429

Nicholas H. Lurie, & Charlotte H. Mason. (2007). Visual Representation: Implications for Decision Making. Journal of Marketing, 71(1), 160–177. https://www.researchgate.net/profile/NicholasLurie/publication/276941370_Visual_Representation_Implications_for_Decision_Making/links/5aa7bed2a6fdc c9576d5969f/Visual-Representation-Implications-for-Decision-Making.pdf

O'Brien, H. L., & Toms, E. G. (2010). The development and evaluation of a survey to measure user engagement. Journal of the American Society for Information Science and Technology, 61(1), 50–69. https://doi.org/10.1002/asi.21229

Qin, X., Luo, Y., Tang, N., & Li, G. (2020). Making data visualization more efficient and effective: a survey. The VLDB Journal, 29(1), 93–117. https://doi.org/10.1007/s00778-019-00588-3

Reitsma, R. M. (2019). The Future of Data: Too Much Visualization, Too Little Understanding? Dialectic, 2(2). https://doi.org/https://doi.org/10.3998/dialectic.14932326.0002.207

Rizkallah, J. (n.d.). Council Post: The Big (Unstructured) Data Problem. Forbes. Retrieved June 14, 2022, from https://www.forbes.com/sites/forbestechcouncil/2017/06/05/the-big-unstructured-data-problem/

Schwabish, J. A. (2021). Better data visualizations: a guide for scholars, researchers, and wonks. Columbia University Press.

Smillie, R. (1986). Design Strategies for Job Performance Aids (p. 39) [Technical]. reprinted by the Navy Personnel Research and Development Center. https://apps.dtic.mil/dtic/tr/fulltext/u2/a174358.pdf

Smith, A., Campbell, C., Bott, I., Faunce, L., Parrish, G., Ehrenberg-Shannon, B., McCallum, P., & Stabe, M. (2021, March 7). Charts that work: FT visual vocabulary guide. Financial Times. https://www.ft.com/content/c7bb24c9-964d-479f-ba24-03a2b2df6e85

Sudhir, K. (2016). Editorial—The Exploration-Exploitation Tradeoff and Efficiency in Knowledge Production. Marketing Science, 35(1), 1–9. https://doi.org/10.1287/mksc.2015.0974

Table English Definition and Meaning | Lexico.com. (n.d.). Lexico Dictionaries | English. Retrieved May 3, 2022, from https://www.lexico.com/en/definition/table

Tate, V. D. (1950). Special notice to former subscribers of the journal of documentary reproduction. American Documentation, 1(1), 1–2. https://doi.org/10.1002/asi.5090010101

Virkus, S., & Garoufallou, E. (2019). Data science from a library and information science perspective. Data Technologies and Applications, 53(4), 422–441. https://doi.org/10.1108/DTA-05-2019-0076

Wright, S. et. al. (2012) A Multi-Institutional Project to Develop Discipline-Specific Data Literacy Instruction for Graduate Students. Libraries Faculty and Staff Presentations. Paper 10. http://docs.lib. purdue.edu/lib_fspres/10