

INFORMATION VISUALISATION IN EDUCATION: A REVIEW OF CURRENT TOOLS AND PRACTICES

Ajay Sivanand and Dr. Brian Frank
Queen's University
a.sivanand@queensu.ca, brian.frank@queensu.ca

Abstract –Faculties at post-secondary institutions have employed systems to gather vast amounts of assessment data with the ultimate goal of using the data to improve the student learning experience. Unfortunately, the large amount of data and complex relationships make it difficult for instructors and administrators to interpret and enact program and policy changes based on it. Worse yet, this is still a nascent problem and there are few supports upon which they may lean. Information Visualization (IV) is one technique that has shown promise in facilitating the extraction of meaningful information, and a tool that can support a faculty's extraction and application processes would be highly beneficial.

This paper looks at the current work of IV in education. We look at what applications these visualizations are being used in, what considerations went into designing them, and how they were evaluated. This information will hopefully provide insight into how to build visualizations for a variety of other educational contexts.

Keywords: information visualization, education data mining, learning outcomes assessment, assessment data

1. INTRODUCTION

In recent years, post-secondary faculties across a variety of disciplines have spent considerable effort building systems to collect assessment data with the goal of helping improve the student experience [4]. However, currently, using this data to guide program improvement decisions at any level is inhibited by a lack of familiarity among program coordinators and assessment practitioners to extract actionable information effectively [2]. Data sets of this magnitude and complexity are a new challenge that require an entirely new skill set; resources for which are still scarce.

Education Data Mining is an entire research field that is dedicated to this pursuit of resolving research issues stemming from large sets of data originating in an education context [16]. There are many techniques in development that are trying to solve this problem

including ones taking advantage of inferential statistics, machine-learning, and data-mining algorithms. Information Visualization (IV) is one approach that looks especially promising for its ability to “take advantage of the human eye’s broad bandwidth pathway into the mind to allow users to see, explore, and understand large amounts” of data (p. 604). As one review of the field summarizes: “it is recognized that data mining algorithms alone are not enough. Without expressive visualizations and interfaces it is hard to achieve the necessary flexibility to understand results, generate new hypotheses, and test them on the fly in a natural interactive environment” [6].

IV is a growing field in education. To gain a sense of the state of the art currently, this paper surveys the research work done to create tools and education designs that utilize IV. Note that, visualizations can often be used to help discover patterns and illustrate findings in many research situations, as we commonly see with line graphs and scatter plots in publications. This review however is specifically looking at publications that are primarily focused on visualizations in education, rather than the larger body of work that may have utilized one or more visualizations as part of their broader research interests. This particular selection of papers provides insights into visualizations that are larger in scope than an educational research’s specific data-mining goals.

Since there has yet to be a comprehensive review of the literature on the subject, this paper looks to identify areas of education that IV is being utilized, as well as successes and shortcomings that can be built on and explored, respectively with future research. We have looked at 12 papers, spanning 7 different tools and visualization sets. The next section will discuss the visualizations’ purposes and intended end users, followed by a section detailing the IV theories and design considerations that were taken into account creating the visualizations. Finally, we discuss the ways in which visualizations have been evaluated and ways to iterate positively.

2. USES AND USERS

There are a wide variety of uses for visualizations in education. For the purposes of this paper, we are going to group them by where the data they are visualizing originate. These groups are Course management systems (CMS), intelligent tutoring systems and interactive learning environments, and finally program curricula and learning outcomes.

2.1. Course Management Systems

Course management systems like WebCT, Moodle, and Desire2Learn are now almost ubiquitous among post-secondary institutions to disseminate course material, provide assessments, and foster class discussion. This is especially important for courses offered through distance learning. CourseVis is a tool built as an add-in for WebCT to counter the high drop-out rates, students' feeling of isolation, lack of motivation, and difficulties with finding support from tutors and peers found in distance learning [12]. The tool provides a number of visualizations that allow for monitoring of CMS discussion, student progress, CMS materials access frequency, and general student behaviour on the CMS. Here, the primary intended user is the instructors of the course. The visualizations would help them provide real time feedback about where to intervene as well as summative feedback once the course is finished. A follow up to the tool, GISMO [11], was built with much of same functionality along with improvements. It was instead built for the open source CMS, Moodle, allowing for greater accessibility among users. This can still be downloaded from the project's page.

The Student Activity Visualization Tool groups students by the activities they perform on online learning platforms [7]. It uses k-means clustering, a common data-mining algorithm that lends itself nicely to visualization. The tool applies the technique to 6 different activity types such including messaging activity and testing activity. The tool's visualizations can provide insight for instructors running a course through the platform as well as staff members managing the platform; giving them information about what features of the platform are being utilized and by whom.

SNAPP is a tool that uses data from online discussion forums, often found within CMSs, to create a sociogram [1]. This can provide instructors with real-time interaction patterns between students like identifying isolated and highly interactive participants; indicating an instructor's role as central or peripheral to the network; monitoring network evolution and discussion continuity; and more. Identification of these patterns can be used as a way to evaluate the impact of an intervention on student engagement. An additional benefit of the tool is that it

works on the user's client and is independent of where or on what platform the discussions take place.

2.2. Intelligent Tutoring Systems

Intelligent tutoring systems are an increasingly popular software tutors with which students interact. What makes them 'intelligent' is their ability to personalize the tutoring technique of the course material based on the students' prior performance. They do this by storing a representation of the student, referred to as a student model, that they base their tutoring decisions on [22]. ViSMod (Visualization of Bayesian Student Model) is a tool that externalizes this student model in the form of an interactive visualization [21]. The paper suggests that by letting students understand, explore, inspect, and modify their Bayesian models, it empowers them to take more control of their learning. Additionally, teachers can compare student models to see where students are struggling and excelling.

There are more intricate interactive environments that are specialized to teach such subjects as first-order logic and physics; where students can interact with a multi-level problem inputting their answer through many steps. The tool InVis, capitalizes on this extra information by creating interactive visualizations of the pathways that students take to a correct or incorrect solution [8]. The visualization uses the aggregate student solutions to provide insight as to what steps might be causing the most trouble in the class as well as pointing out where struggling students are getting stuck. In this case, the tool is meant primarily for the instructor of the course.

One study proposed three novel visualizations as a way of externalizing group activity and interaction [9]. The visualizations rely on log data that track each member's contribution to the project as well as team correspondence and delegation. From the resulting visualizations, the authors were able to gain insight into a number of group dynamics. With the visualizations, through just one snapshot, students inside and instructors outside the group are able to determine if a team is functioning in an effective manner. Mirroring this information either back to the team or a supervisor allows them to take action accordingly.

2.3. Program Curricula and Learning Outcomes

Not all data has to be acquired from entirely digital mediums. With the increasing use of learning outcomes assessment among institutions, the sheer volume of details to manage a degree program has become extremely unwieldy and complex. Program coordinators can find themselves lost trying to gain a sense of how their courses build to their program outcomes [20]. Vaitis, Nilsson, and Nary [19] created visualizations to map the teaching methods of a course in a medical school degree program

to its expected learning outcomes. They use a subset of IV called Visual Analytics, that combines the visualisations with data analysis to give more detailed summary of the data. The authors also created a visualisation that used student data from a particular exam in conjunction with the curriculum data to give a more granular picture of which learning outcomes for a course were being achieved and which were not. In this case, the authors were program coordinators themselves and were primarily trying to create visualizations for their own purposes due to the lack of appropriate tools currently available.

One aspect that does become apparent when looking at these papers together, is that the majority of them are only useful once you, as an instructor or program coordinator, have already committed to the system it is built into or the educational theory that it is based on. If you either use a different LMS/CMS or map your teaching in a different way, the available resources have little immediate value.

3. VISUALIZATION DESIGN PROCESS

Once a researcher has a clear idea of what data-mining problems they want to resolve, they then have to create visualizations that will provide adequate enough insight with which a user can make informed decisions for improvement in learning. Given the now wide range of visualization options from the standard line and scatterplots to tree graphs and heat maps and even completely novel ones, the question does not so much become *how* to visualize data, but *how best* to visualize your data. There are a number of authors in the IV field that provide advice on how to choose or create a visualization to best suit the goal's needs. For example, Börner and Polly [5] operationalized the process of visualization creation in their version of a workflow to build one for almost any scenario, starting with the stakeholder needs and making decisions based on the type of data (ie. categorical, ordinal, numerical). In this section we look at how each of the visualization projects, of the ones that reported it, arrived at their design.

InVis [8] designed their visualizations by heavily relying on the *visual information seeking mantra* from Schneiderman [17]. The main message of the mantra is boiled into three statements: overview first, zoom and filter, details-on-demand. Schneiderman also provides advice on how to handle different data types based on his taxonomy. The InVis team also discusses in detail how, where, and why they put the mantra to work on their visualizations in good detail, which can be useful for others looking to do similar work.

The team behind CourseVis [12] and GISMO [11] do not specify a guideline or workflow that they followed, but based on their descriptions, their visualizations seemed to originate in the data that is the pillar of the data-mining requirements. Their goal seems to be to

display as much information in a single view as possible, even using multiple stacked data using multiple y-axis for the same x-axis in a temporal plot. This maximizes the information available in an overview snapshot. The tool still does allow for filtering and zooming, but the overview seems to be where they think there is the most value since it what they provide as demonstration.

While Vaitsis et al. [20] unfortunately do not provide much detail about their design decisions regarding their final visualizations, they do describe why they chose to use visual analytics. Largely, it boiled down to the data being so complex that the extra information provided by the data analysis overlaid on the visualizations were required to adequately fit their needs.

Additionally, one theme that was common among many of the projects was interactivity. The teams behind CourseVis [12], InVis [8], ViSMod [22] all mention the interactivity as a key requirement and/or benefit of their visualizations to be successful at their respective tasks. This added feature allows for visualizations to be dense with information while remaining clean and uncluttered. Interactivity allows users to zoom in and out of rich areas, clicking on aspects of the visualization which highlight features or provide more information.

Each paper does detail their own technical configurations, in which there are a variety of visualization software tools and packages that are used by each of the authors. However, there lacked enough commonality to seem worth reporting on what software was used. Vaitsis et al. even reported that there was no clear choice for software and so had to conduct their own survey of options to find which worked best for their specific needs [20].

4. EVALUATION AND EXCELLENCE

An obviously important aspect when building these visualizations is evaluating their effectiveness in doing what they were intended to: providing insight on the learning from the data. Few of the papers that we were looked at discuss how the visualizations were evaluated once they were created. Of the ones that do, the visualizations are evaluated in two major categories of metrics: their usability and their usefulness. This distinction however isn't mentioned in most of the papers and only becomes apparent when looking at them together. The usability refers to the design aspects of the visualization and how much information can be interpreted from the visualization. While usefulness refers to how much that extracted information can actually be used to towards its intended purpose and how accurate that is to the real world. Unfortunately, only half of the studies we looked at published any material about conducting evaluations on their visualizations.

4.1 Usability

To promote usability, there is a substantial wealth of literature in information visualization that provide design principles, guidelines, and workflows to follow in order to maximize the amount of data extracted. Although, recent advances in computational power have increased the complexity of visualizations, the theory for the field has been in development for much longer. Bertin [3] showed that there were three main aims behind how one chose to represent data, moving from high to low-level acts of graphical interpretation. Each aim provided a guide as to which data and how much data can be presented effectively at once. Tufte [18] discusses the key aspects that one must worry about to promote 'graphical excellence'. MacKinlay [10] provides rankings for the effectiveness of different visual variables based on the kind of data being visualized. Additionally, there are even tests that have been created [15] to measure these technical aspects of a visualization. As such, there is a lot of foundational work with which to build visualizations on to promote its usability.

With CourseVis, the team conducted a study with 6 instructors, giving them the tool with data from a mock course and asking them to see whether they can gain any insight into the classroom [12]. The InVis team performed a similar study, providing 15 participants with mock visualizations and asking them to answer a set of questions which asked them discern certain details about the class from the visualizations [8]. Both studies included a qualitative portion that let the users provide more open feedback. The ViSMod team used this approach as well and asked participants to interact with their visualizations, change the settings, and provide feedback about what they thought worked best [22]. All studies used the feedback to iterate the designs of their visualization tools to make them more readable and user-friendly.

4.2 Usefulness

Usefulness is harder concept to promote and evaluate. No standard practices exist to ensure that your visualization is going to be useful. This is an intrinsic problem to the field since the measure of usefulness is going to be entirely unique to the data-mining goals of the visualization end users. Additionally, the visualizations can only show what information that can be extracted from the data. Understanding the root of the problem, or steps to fix it would involve inference and assumptions on the part of the end user. One way to partially deal with this is by designing your visualizations around the end user needs. By first scoping out what problems the data is being used to solve, the visualizations can be designed such that solutions to those problems would arise from the

end result. That way, evaluations can be run with these participants after to test whether the visualizations meet the participant's originally imagined standards.

For CourseVis, the CMS tool, the team conducted a survey with distance learning instructors to find out what information the instructors would need to conduct their online learning activities [13]. After creating the tool, they then ran a focus group and semi-structured interviews to see whether instructors found any value in the visualizations that the tool provided [12], [14]. The team behind VisMod performed similar studies to understand their requirements, although provide less detail about their process or results [22]. By contrast, in the case with some visualizations, the measure of usefulness does not require a separate study. Through their efforts to visually map the learning outcomes of their medical program to course assignments, Vaitsis et al. discovered there were gaps in their expected learning outcomes [20]. As such, sometimes visualizations can provide immediate insight into your system, without much need for inference or assumption.

Additionally, the team behind VisMod shows another interesting approach to gaining an understanding of the usefulness of a visualization. Using an exploratory approach, they tested their software with a fairly large sample of students (N = 110) [22]. They were able to find novel ways in which mirroring their student model visualization to the student had an effect on their learning. The generalizability of the findings might be limited, but conducting such studies gives the team an opportunity to find the strengths in their approach and iteratively capitalize on them to create a more robust tool.

When trying to use visualizations to monitor student behaviour, an important aspect with looking at usefulness is that the inferences made from the visualizations are an accurate portrayal of what is happening with the students. This process to test the accuracy of inferences is likely a much more intensive task, which might be the reason that it is an aspect that we did not see reported almost at all. Kay et al. [9], in their study of externalizing group behaviours, give us some insight into what that process might look like. In addition to the visualizations they created and tested, the authors also had the students fill out surveys about their thoughts on how the group functioned. These responses were compared against the information inferred from the visualizations by the instructor and researchers. Kay does caution that this kind of information should not be used to compare students to a normative scale, but should instead be used primarily to identify possible problem areas that would prompt a possibly more involved intervention.

5. CONCLUSION

Only a few of these visualization efforts have so far completed and documented the full process of acquiring

requirements, systematically designing visualizations, and then evaluating their merit. These gaps in the literature leaves the researchers looking to take advantage of these studies with missing puzzle pieces of how to proceed with their own IV endeavours. A large number of the papers that were reviewed in this paper were either reports or from journal proceedings, and as such it is possible that the authors of the respective papers addressed some of these shortfalls in their study but failed to mention it in the reporting.

Still, the variety in applications for information visualization and successes so far make it a promising field to look to for support in an educational world where data is becoming increasingly available and brimming with underutilized potential.

Acknowledgements

The authors acknowledge the support of the Higher Education Council of Ontario for funding this project.

References

- [1] A. Bakharia and S. Dawson, "Snapp," *Proc. 1st Int. Conf. Learn. Anal. Knowl. - LAK '11*, p. 168, 2011.
- [2] T. W. Banta and C. Blaich, "Closing the Assessment Loop," *Chang. Mag. High. Learn.*, vol. 43, no. 1, pp. 22–27, 2010.
- [3] J. Bertin, *Graphics and graphic information processing*. Walter de Gruyter, 1981.
- [4] C. Blaich and K. Wise, "From gathering to using assessment results: Lessons from the Wabash National Study. (NILOA Occasional Paper No. 8)," *Wabash Natl. Study - Impact Teach. Pract. Institutional Cond. Student Growth*, pp. 1–46, 2011.
- [5] K. Börner and D. E. Polley, *Visual insights: A practical guide to making sense of data*. MIT Press, 2014.
- [6] D. Brodbeck, R. Mazza, and D. Lalanne, "Interactive visualization - A survey," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5440 LNCS, pp. 27–46, 2009.
- [7] M. S. Chiritoiu, M. C. Mihaescu, and D. D. Burdescu, "Students Activity Visualization Tool," *Int. Educ. Data Min. Soc. IEDMS*, 2013.
- [8] M. W. M. Johnson, M. Eagle, and T. Barnes, "Invis: An interactive visualization tool for exploring interaction networks," *Proc. EDM*, vol. 2013, p. 65, 2013.
- [9] J. Kay, N. Maisonneuve, K. Yacef, and P. Reimann, "The Big Five and Visualisations of Team Work Activity," *Lect. Notes Comput. Sci.*, vol. 4053, no. 2006, pp. 197–206, 2006.
- [10] J. Mackinlay, "Automating the design of graphical presentations of relational information," *ACM Trans. Graph.*, vol. 5, no. 2, pp. 110–141, 1986.
- [11] R. Mazza and C. Milani, "Exploring usage analysis in learning systems: Gaining insights from visualisations," *AIED 2005 Work. Usage Anal. Learn. Syst.*, pp. 65–72, 2005.
- [12] R. Mazza and V. Dimitrova, "Generation of graphical representations of student tracking data in course management systems," *Proc. Int. Conf. Inf. Vis.*, vol. 2005, pp. 253–258, 2005.
- [13] R. Mazza and V. Dimitrova, "Informing the design of a course data visualisator: an empirical study," *5th Int. Conf. New ...*, no. ICNEE, pp. 215–220, 2003.
- [14] R. Mazza, "Evaluating Information Visualization Applications with Focus Groups: the CourseVis experience," *Focus (Madison)*, p. 1, 2006.
- [15] C. Plaisant, "The challenge of information visualization evaluation," *Proc. Work. Conf. Adv. Vis. interfaces - AVI '04*, pp. 109–116, 2004.
- [16] C. Romero and S. Ventura, "Data mining in education," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 3, no. 1, pp. 12–27, 2013.
- [17] B. Shneiderman, "The eyes have it: a task by data type taxonomy for information visualizations," *Proc. 1996 IEEE Symp. Vis. Lang.*, pp. 336–343, 1996.
- [18] E. R. Tufte and P. R. Graves-Morris, *The visual display of quantitative information*, vol. 2, no. 9. Graphics press Cheshire, CT, 1983.
- [19] C. Vaitsis, G. Nilsson, and N. Zary, "Big Data in Medical Informatics: Improving Education through Visual Analytics," *Stud. Heal. Technol.*, vol. 205, no. August, pp. 1163–1167, 2014.
- [20] C. Vaitsis, G. Nilsson, and N. Zary, "Visual analytics in healthcare education: exploring novel ways to analyze and represent big data in undergraduate medical education," *PeerJ*, vol. 2, p. e683, 2014.
- [21] J. D. Zapata-Rivera and J. E. Greer, "Inspecting and Visualizing Distributed Bayesian Student Models," *Intell. Tutoring Syst.*, vol. Lecture No, pp. 544–553, 2000.
- [22] J. D. Zapata-Rivera and J. E. Greer, "Interacting with Inspectable Bayesian Student Models," *Int. J. Artif. Intell. Educ.*, vol. 14, pp. 127–163, 2004.