Data Visualization in Online Educational Research

Xue Wen
Louisiana State University

Xuan Wang
The University of Texas Rio Grande Valley

Chapter 12
Data Visualization in Online Educational Research

Xue Wen
https://orcid.org/0000-0002-7529-9436
Louisiana State University, USA

Xuan Wang
https://orcid.org/0000-0001-9183-3080
The University of Texas Rio Grande Valley, USA

ABSTRACT

This chapter presents a general and practical guideline that is intended to introduce the traditional visualization methods (word clouds), and the advanced visualization methods including interactive visualization (heatmap matrix) and dynamic visualization (dashboard), which can be applied in quantitative, qualitative, and mixed-methods research. This chapter also presents the potentials of each visualization method for assisting researchers in choosing the most appropriate one in the web-based research study. Graduate students, educational researchers, and practitioners can contribute to take strengths from each visual analytical method to enhance the reach of significant research findings into the public sphere. By leveraging the novel visualization techniques used in the web-based research study, while staying true to the analytical methods of research design, graduate students, educational researchers, and practitioners will gain a broader understanding of big data and analytics for data use and representation in the field of education.

INTRODUCTION

Advances in technology have created numerous opportunities for research in online education. With the vast needs of resources for a diversity of learners, various online learning platforms have developed free online courses for online learners to acquire new skills, advance their careers, and deliver quality educational experiences at scale (Chen, Chen, Liu, Shi, Wu, & Qu, 2016). Massive open online courses (MOOCs), a popular online learning platform in recent years, have emerged and attracted a remarkable

DOI: 10.4018/978-1-7998-1173-2.ch012

amount of public attention (Hollands & Kazi, 2019). The term MOOCs—massive open online courses, was first used in the educational community in 2008 by Stephen Downes and George Siemens (MAUT-McGill University, 2018). Educators intended to explore the possibility for interactions between a wide variety of participants made possible by online tools that provide a more productive learning environment than traditional tools would allow (Chen et al., 2016). Learners from all over the world can enroll in more than 1,000 courses, and the number of registrants has reached 10 million (Rollins, 2018). Because of its volume and complexity, MOOCs often generate large, heterogeneous datasets comprising clickstream data, contributions to discussion forums, and various performance metrics (Vieira, Parsons, & Byrd, 2018). Thus, to study MOOCs or related online learning platforms, educational researchers would be required to master visual learning analytics, educational data mining, and visual analytics to capture rich data about students and their online learning behaviors.

From this perspective, Vieira et al. (2018) define a new term called *visual learning analytics*, which can be defined as an integration of learning analytics, educational data mining, and visual analytics, to illustrate how designers and researchers can employ data visualization approaches for analyzing educational data. To help readers systematically review the field of data visualization in online educational research, the authors present a brief agenda for the field of visual learning analytics in an educational context in this chapter. Also, based on the unique characteristics of MOOCs' environment, this chapter focuses on illustrating three data visualization techniques that help readers understand which types of educational data can be customized and how to visualize the educational data in MOOC scenarios. Visual learning analytics is a discipline that shows significant promise in helping users gain insight into data visualizations (Vieira et al., 2018). This term integrates data analysis, visual representations, and user interactions to leverage the strengths of technology and humans. In the context of the web-based environment, visual learning analytics use computational tools and methods for understanding educational phenomena, such as students' learning paths, the effectiveness of learning materials, and different approaches that students use for a given task through visualization discourse.

While there is a large body of work providing innovative visualization skills in data representation, educational researchers, instructors, and school administrators typically have a difficult time processing and interpreting big data about online education. This is because they have a limited understanding of necessary data mining and processing techniques. Challenges for educators mainly arise when analyzing MOOC data. First, MOOC data is often large, complex, and heterogeneous. Second, the users of the analytics system in MOOC are often course instructors, educational researchers, and students, who usually have little to no knowledge of data analytical techniques (Qu & Chen, 2015). Given this information, the purpose of this chapter is to illustrate three scenarios in online learning platforms along with the possible solution of data visualization techniques. This chapter aims to enhance the understandings of researchers, instructors, stakeholders, and students of necessary data processing and interpretation in online educational research. Therefore, this chapter focuses on the discussions of data visualization approaches by incorporating the basic concepts of visual learning analytics. This chapter looks at *what* types of data in online educational research can be approached using visualization techniques and *how* they can be approached. On one hand, the "*what*" focuses on characterizing the data source, data analytical approach, tools, and purposes. On the other hand, the "*how*" focuses on the possible solutions in visualization that can specifically help readers to understand big data and data analytics in online educational research. Therefore, the objectives of the chapter are: (1) to identify the primary uses of visualizations in online educational research, (2) to customize data visualization techniques for online

educational research, and (3) to create visualization tools that can help advance the readers' understandings in the field of data visualizations. The research questions for this chapter are:

1. What data sources, approaches, and purposes in the use of data visualization can be applied to online educational research?
2. How does each visualization technique address the type of data in online educational research?

The remainder of this chapter is structured as follows: (1) Section two briefly describes the concept of visual learning analytics, which is useful for introducing readers to information regarding approaches to increase the effectiveness of the visualizations contained in the online learning platform MOOCs. In parallel to the visual learning analytics, the authors emphasize the use of visualization in the educational context. In order to gain a better understanding of the whole course or specific tasks in online environment (Verbert, Duval, Klerkx, Govaerts, & Santos, 2013), the authors address practical influences of the data visualizations in MOOC to look closer at how instructors, course designers, or researchers take advantage of each of these visual approaches. Additionally, students could use visualizations to reflect on aspects of their learning behavior because the visualizations could potentially assist them, for example, in managing time effectively, accessing essential learning resources, or gaining a rich picture of their progress in the online learning context (Bodily & Verbert, 2017; Reimers & Neovesky, 2015); (2) section three provides the examples/scenarios from online educational research using appropriate visualization techniques; and (3) section four discusses the direction of future development/research for the technique, summarizes the core outcomes of this chapter, and addresses the limitations/weaknesses of the visualization examples.

BACKGROUND

Visual Learning Analytics

Visual learning analytics applied to educational contexts, especially the online learning platforms, provides students and teachers with a set of tools and spaces to learn (Conde-Gonzales, Garcia-Penalvo, Aguilar, & Theron, 2014). The term *visual learning analytics* recently emerged in the field of educational research and has gained considerable attention to represent current research on the intersection of data analysis and visualization in web-based studies (Dawson, 2010; Echeverria, Martinez-Maldonaldo, Shum, Chiluiza, Granda, & Conati, 2018; Ritsos & Roberts, 2014; Verbert, Manouselis, Drachsler, & Duval, 2012; Vieira et al., 2018). Vieira et al. (2018) think that data visualization belongs to the emerging discipline of visual learning analytics, which integrates three specialized subfields: learning analytics, educational data mining, and visual analytics (Siemens, 2013; Thomas & Cook, 2005; Vieira et al., 2018). Learning analytics was originally defined as "the measurement, collection, analysis, and reporting of data about learners and their contexts, for the purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens, 2013, p.1383). By contrast, educational data mining focuses on developing methods for "exploring the unique types of data that come from educational settings" (Siemens, 2013, p.1383). Furthermore, visual analytics was defined as "the science of analytical reasoning facilitated by interactive visual interfaces" (Thomas & Cook, 2005, p.4). By combining the three concepts, Vieira et

al. (2018) defines visual learning analytics as "the use of computational tools and methods for understanding educational phenomena through interactive visualization techniques" (p.120).

Previous literature also mentioned many challenges that educators face in the use of visual learning analytics tools in physical classroom settings (Verbert et al., 2013) and suggested that it can be challenging to bring emerging technologies into real classroom environments since the social sciences tend to adopt technologies at a slow pace relative to technological innovation. Ritsos and Roberts (2014) indicated that visualization is often challenging to interpret, although the purpose of visualization tools is to enrich the learning process for both students and instructors. Due to these limitations, visualizations are often used in blended, virtual, and/or MOOC dashboards, which are very often basic techniques they use (e,g., bar plots, scatter plots); however, these techniques offer limited interactivity for the user and are not informed by pedagogical practices (Dawson, 2010; Ritsos & Roberts, 2014). Therefore, Vieira et al. (2018) propose that if educators had access to compelling visualizations of educational data, they could potentially use them to provide formative feedback and to improve instructional quality. If students were provided abundant resources for learning data visualizations, it could promote their metacognitive skills and enable them to choose their learning path (U.S Department of Education, 2017). If researchers became familiar with these visual learning analytics tools, it could help them to gain a thorough understanding of extensive, complex data derived from online studies (Qu & Chen, 2015).

The Use of Visualization in Education

Visualization emerged as a separate discipline in 1980 (Telea, 2014) as a reaction to the increasing amount of data generated by computer calculations (Olshannikova, Ometov, Koucheryavy, & Olsson, 2015). In short, visualization is the graphical display of information. The purpose of any visualization (e.g., textbook, article, movie, TV show, computer program, etc.) used in an educational context is to facilitate the process of turning information into knowledge (e.g., idea, concept, fact, algorithm, relationship, etc.) (National Forum on Educational Statistics, 2016). To achieve this, a visualization must create connections between the knowledge the learner already has and the knowledge being taught. Hence, to design effective visualizations, it is necessary to understand what the learner knows (Segenchuk, 1997). This is important in the context of education (Segenchuk, 1997). The underlying theory of the visualization that is relevant to what the learner knows is about the representation of knowledge – internal representations and external representations (Gilbert, 2008). A representation is a likeness or simulation of some idea, concept, or object (Gilbert, 2005). In a discussion on this topic, Rapp and Kurby (2008) state:

External representation is one that is available in the environment, like the aforementioned skyline, flag, or blueprint. These representations often stand for or correspond to additional concepts or notions, such as a flag both being an object itself and a symbol of some geographical region, group of people or socio-cultural perspective (p. 32).

Maps and graphs are examples of external representations because they organize data into presentations that are easier to interpret and understand than their original forms (e.g., the collected numbers that comprise a bar graph) by summarizing concepts in remarkable, systematic ways. On the contrary, the internal representation is available in the environment but is held in the viewer's or learner's mind. One example of internal representation is a learner's mental activity of processing abstract concepts, solving problems and making decisions (Rapp & Kurby, 2008). The dichotomy between external and internal

representations is both theoretically and practically important because these two types of representations necessarily interact throughout students' learning experiences in a variety of ways. Moreover, it has been situated in the setting of blended, virtual, and/or MOOC for understanding learning behaviors, teaching progress, as well as assessment and feedback (Atapattu, Falkner, & Tarmazdi, 2016; Chen et al., 2016; Munoz-Merino, Ruiperez-Valiente, Alario-Hoyos, Perez-Sanagustin, & Kloos, 2015).

Regarding the purpose of using visualization strategies in online context, relevant research using MOOC data focus on instructional design (Chen et al., 2016), exploring usage pattern (Munoz-Merino et al., 2015), understanding discussion forums (Atapattu et al., 2016), and retention (Chen, Chen, Zhao, Boyer, Veeramachaneni, & Qu, 2016). For instance, Chen et al. (2016) introduced a comprehensive visualization system to enable course instructors and education experts to analyze the peaks or the video segments that generate numerous clickstreams. The purpose of these visualizations was to inform the instructors that students were having an issue with a video (e.g., the concept was not clearly explained), as indicated by the visualized peaks. Munoz et al. (2015) showed how different visualizations within and between MOOC courses help teachers make quick and informed decisions in their case study. The results from the case study revealed that visualizations enable the whole comparison of a large number of students at a glance, that visualized metrics can help teachers identify students' effectiveness with the exercises, and that these metrics can help teachers understand the ways the students used the video lectures. Besides, commonly found in the context of MOOCs, the research explored the phenomenon of retention, which is a challenge for the MOOC learning environment (Chen et al., 2016). For instance, Chen et al. (2016) introduced a visual analytics system DropoutSeer, which not only aids instructors and education experts to catch the reasons for dropout but also enable researchers to identify key features which can further enhance the performance of the models. Additionally, it is interesting to see that discussion forum activity is an indispensable component of MOOCs. Discussion forums are difficult to manage and to understand when a large number of students are discussing ideas (Vieira et al., 2018). Atapattu et al. (2016) utilized an open-source topic visualization dashboard for topic analysis of discussion contents from three different MOOC courses. This dashboard allows users to explore the analysis by manipulating different variables, such as votes, views, instructor replies, and time-series analysis.

Visual learning analytics can support the creation of data visualizations throughout the online educational research, and help educators rethink this interdisciplinary approach to adhere more strictly to their educational goals. Therefore, the authors' motivation for developing this chapter is to present three effective visualization techniques by user-friendly web tools to analyze available educational data, informed by visual learning analytics and visualization theory in education. The demonstrations of the three examples in the following sections can help readers to gain a systematic review of data visualizations in online research from the perspective of discussion content, students' online behavior, and students' usage patterns.

METHODOLOGY

The research questions of this chapter are: (1) What are existing data sources, approaches, audiences, and purposes in the use of visualizations for online educational research? and (2) How does each visualization technique address the type of data in online educational research? To answer the two research questions, the authors primarily followed a quantitative approach and applied the process of three quantitative analytical techniques: word clouds, correlational analysis, and descriptive statistics. Word clouds

can help educational researchers to research both qualitative and quantitative scenarios. For instance, word clouds produce the spoken (interview data) and written responses (observation field notes) of informants by visualizing the data in an artistic way (McNaught & Lam, 2010); word clouds can also quantify qualitative data by turning the data from words or images into numbers. However, McNaught and Lam (2010) emphasize that word clouds can be a useful tool for preliminary analysis and validation of previous findings, but not a stand-alone research tool comparable to traditional qualitative analysis methods. For correlational analysis and descriptive statistics, the authors of this study introduce two innovative web-based analytical software: *Python Jupyter Notebook* and *Tableau* to produce a heatmap matrix and for analyzing MOOC data, respectively. Authors retrieved online learning platforms for discussion activity from a hybrid graduate-level course and MOOC courses. Additionally, the authors used information from Harvard University and Massachusetts Institute of Technology (MIT) to provide insight into data visualization for readers (educator, researcher, and students). Of note, two different data sources were retrieved to demonstrate the visualization examples in the following section. Although the provided data sources may limit the scope of this chapter, the authors' purpose is to motivate readers to think about how a large dataset from online learning platforms can be utilized and customized based on the selected analytical approaches. The authors present the details of data collection and data analysis in the next section.

Research Context: MOOCs

Major MOOC platforms, such as Coursera and EdX, can provide raw data to course instructors and their collaboration partners. The raw data includes the essential information in learner profiles (i.e., student demographics, learning performances, learning behaviors), video viewing histories, clickstreams of course video, and activities in course forums. MOOC data is often large, complex, and heterogeneous, since it contains structured, unstructured (such as text in course forums), spatial, and temporal information. For instance, considering the impossibility for instructors and teaching assistants to grade tens of thousands of student quizzes, peer grading is widely used in MOOCs. It becomes challenging to extract useful information if the peer grading data is parse (i.e., each student only grades a few questions) (Figure 1).

Data Sources

In the first example, the data was retrieved from a module of an online discussion forum in a hybrid (50% in class and 50% online) introductory master-level course. This course consisted of seven modules, and each module is two weeks in length. The first week was conducted in class physically, and the second week was self-paced online learning via Moodle. In the second week of each module, students were assigned discussion tasks related to the module topic that was provided by the instructor. During that week, these students were required to actively participate in the discussion forums. The objective of this example was to introduce the word clouds analytical technique, which can be useful in analyzing the written responses of informants in online discussions. Below is a sample of an extract of the postings from student A (anonymous).

Figure 1. A screenshot of an open online course from EdX platform

Student A

Instructor: What different kinds of thinking are required of educational researchers?

A: A curious mind is essential for an educational researcher to form research questions. Curiosity is similar to skepticism, which is referred by the textbook as one of the attitudes expected from scientists. Without curiosity and skepticism, there would not be groundbreaking or new research, as theories are not being challenged but accepted. Another vital component of thinking in educational research is objectivity. Researchers should never infuse any personal opinion into a study. The purpose of research is to collect a wide array of data and provide evidence to explain certain phenomena. Any biased data would affect the validity and reliability of the research.

Below is a sample of an extract of the feedback from classmates for student A's posting:

B respond to A: I, too, agree with your point on skepticism. I read in the text that researchers must remain skeptical. This is a very important mindset to possess as skepticism leads to a new discovery!

In the second and third example, the authors used an existing MOOCs dataset from Kaggle metadata website (Kaggle Inc, 2019). In 2012, a non-profit learning platform, co-founded by MIT and Harvard University, launched open online courses on EdX. This dataset includes 290 Harvard and MIT online courses, 250 thousand certifications, 4.5 million participants, and 28 million participant hours on the EdX platform since 2012. The objective of this example is to explore the possible relationship among the factors that could affect each other in order to understand which factors are important in the online classes provided by the two elite universities.

Data Analysis Tools

Wordle (http://www.wordle.net/) (Figure 2) is a free visual presentation software application that allows researchers to produce word-cloud analyses of the spoken and written responses of informants in research projects. Recently, *Wordle* has been implemented in the academic environment, emerging as a potential strategy to analyze teaching and learning experiences in an online setting (synchronous or asynchronous format) (Williams, Parkes, & Davies, 2013). *Wordle* offers an initial data analysis tool, that is relatively easily applied, pointing researchers to areas of potential interest for further analysis or research. However, *Wordle* cannot replace more detailed and, potentially more rigorous analytical qualitative or quantitative techniques in research (McNaught & Lam, 2010).

Figure 2. A screenshot of Wordle interface

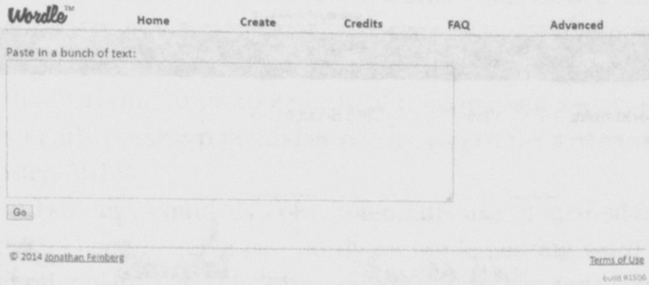
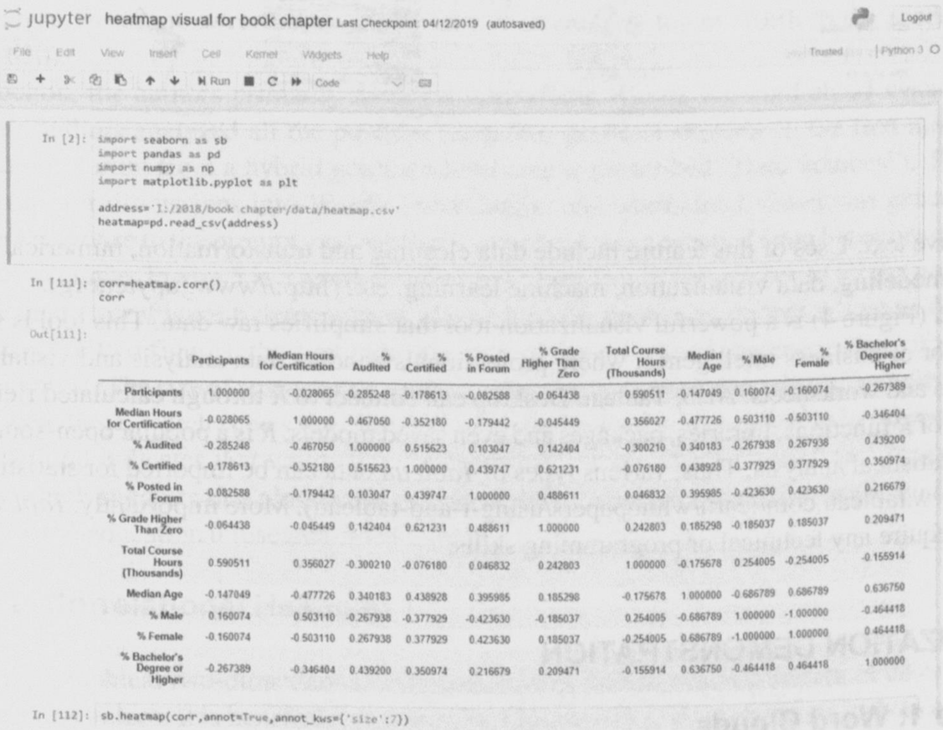
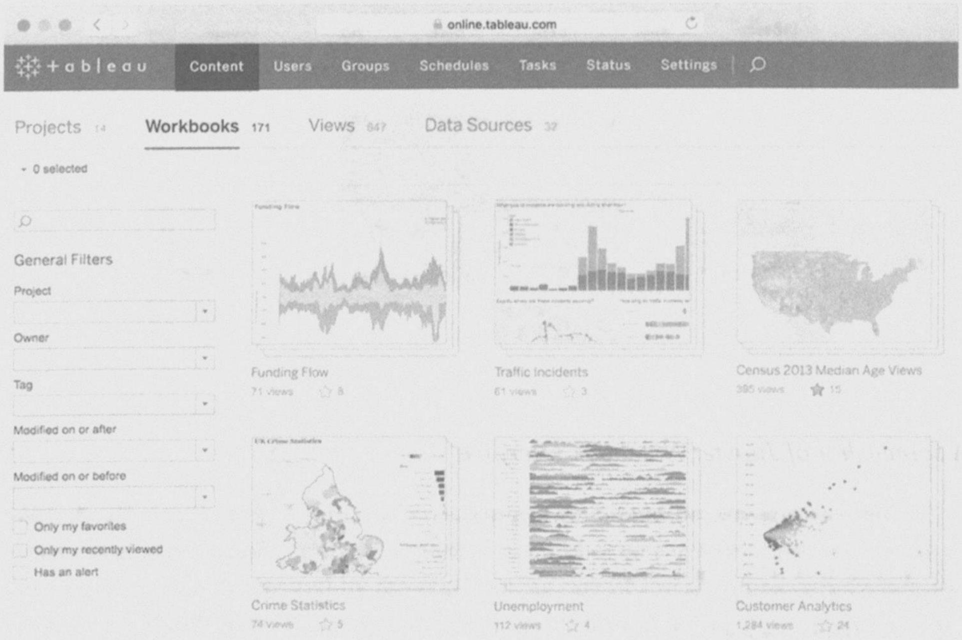


Figure 3. A screenshot of Jupyter Notebook interface



Python is an interpreted, object-oriented, high-level programming language with dynamic semantics (https://www.python.org/). *Python* is easy to learn and use. The number of features in the language itself is modest, requiring relatively little investment of time or effort to produce one's first programs. The syntax is designed to be readable and straightforward. This simplicity makes Python an ideal teaching language, and it lets newcomers pick it up quickly *(*http://www.inforworld.com*)*. One of the functions in the *Python* application is *The Jupyter Notebook* (Figure 3). *The Jupyter Notebook* is an open-source web application that allows one to create and share documents that contain live code, equations, visualizations,

Figure 4. A screenshot of Tableau interface



and narrative text. Uses of this feature include data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, etc. *(*http://www.jupyter.org*)*.

Tableau (Figure 4) is a powerful visualization tool that simplifies raw data. This tool is widely used in the sector of business intelligence, where professionals conduct data analysis and visualizations via dashboards and worksheets. Also, Tableau Desktop can connect to *R* through calculated fields and take advantage of *R* functions, libraries, packages and even saved models. *R* is a popular open-source environment for statistical analysis. Thus, various types of *Tableau* data can be imported for statistical analysis (https://www.tableau.com/learn/whitepapers/using-r-and-tableau). More importantly, *Tableau* software does not require any technical or programming skills.

VISUALIZATION DEMONSTRATION

Example 1: Word Clouds

A word cloud, which is one of the useful data visualization techniques, is commonly used in web-based research within the context of online discussions. A word cloud takes the most frequently used words in a particular text and randomly displays them by size, based on their frequencies (deNoyelles & Reyes-Foster, 2015). Word clouds also show variation in color, typography, and composition, offering an aesthetically pleasing presentation of words. Because of this, word clouds are used in a myriad of creative, playful ways by a broad array of users (McNaught & Lam, 2010).

The use of word clouds has risen in popularity over the last decades. With the fast development of big data and visualization technologies throughout the world, word clouds have become efficient and effective for representing a quantitative element of a corpus in a straightforward manner (deNoyelles & Reyes-Foster, 2015). Recently, word clouds have quickly come into prominence within academic environments. For educators striving to promote critical thinking and engagement in the online classroom, word clouds consequently possess potential as pedagogical tools within asynchronous discussions (deNoyelles & Reyes-Foster, 2015).

Besides, the word-cloud strategy could also be a potentially useful method for qualitative analysis of the text. Williams et al. (2011) conducted a case study on implementing word cloud visuals to quickly and effectively highlight both positive and negative areas in the student experience at the induction stage of a management education program course. Through the holistic review of literature on word clouds in web-based research, the authors of this chapter identified that synchronous discussion forums, online classrooms, or web-based research projects have widely adopted *Wordle* as a research tool of qualitative analysis or preliminary analysis. *Wordle* has been described by its developer, Jonathan Feinberg, as a "toy" (Feinberg, 2009). McNaught and Lam (2010) were attracted to this description of the application, and they analyze *Wordle* to ascertain its utility as "... a scholarly toy of worth to the academic community" (p. 630).

In this section, the authors primarily focus on introducing *Wordle* for word cloud visualization of text data. The authors analyzed all the postings from four graduate students in the first module of an online discussion forum from a hybrid graduate-level course (described "Data Sources"). The authors fed the text of their discussions into *Wordle* individually: one word-cloud visual was generated from each student's response to the prompt, and another word-cloud was generated from his or her classmates' feedback. For example, Figure 5 is a word-cloud visual of student A's responses to all the prompts provided by the instructor. Figure 6 demonstrates classmates' comments on student A's post. With even a cursory look at the two figures, the first indicates that student A valued the importance of research in educational context. The size of some words, such as "skepticism", "confidentiality", "validity", "rights" and "data", were large, which shows that those words were frequently mentioned in her discussion. The second word-cloud indicates that words like "educational", "researcher", "agree" and "enjoyed" were outstanding, which suggests that classmates provided certain amount of positive words on student A's perspective on the educational research-related topic.

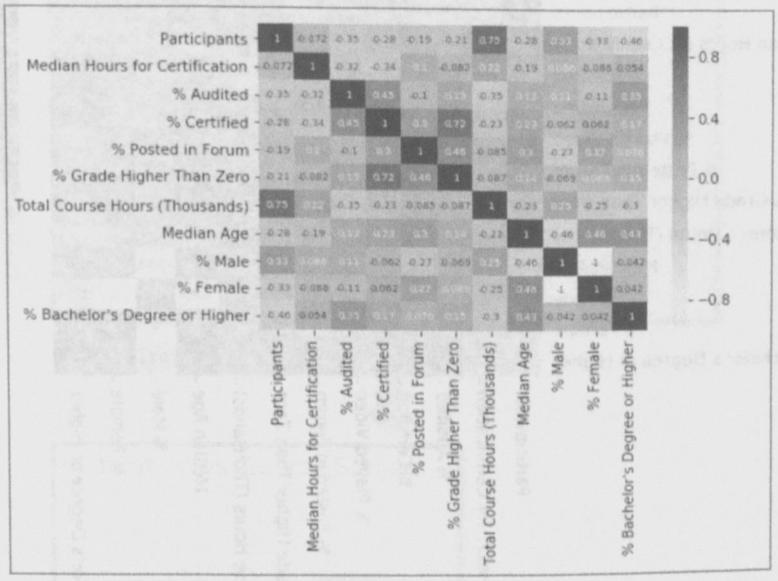
Example 2: Correlational Heatmap

A heatmap is a graphical two-dimensional representation of data that uses a system of color-coding to represent different values. It is beneficial to researchers because they allow them not only to extract specific data points, but also to display a general view of numerical data. More elaborated heat maps allow researchers to understand the underlying mechanism of the complex data sets. There are multiple ways to display heatmaps, but all share one property in common – they use color to communicate the relationship between data values that may be difficult to understand if presented numerically in a spreadsheet. With the rise of big data analytics of data from MOOCs, educational researchers tend to prefer clear visual aids to lower the learning curve and help to analyze the data more intuitively and efficiently compared with traditional methods (Atapattu et al., 2016). For example, educational researchers use heatmaps to demonstrate correlational statistics for the end-users of the analytics systems, including course instructors, education researchers, and students who have little to no knowledge of data mining techniques.

Figure 5. Word cloud visual of student A's discussing posts in an online course

Figure 6. Word cloud visual of classmates' feedback on student A's discussing posts in an online course

Figure 7. Heatmap visual on learners' characteristics and learning behavior from Harvardx online platform



In this example, the authors employed the python programming on *Jupyter Notebook* to produce two traditional heatmap visuals with different colors to understand how one factor would affect another in a MOOCs dataset (described in "Data Sources"). Figure 7 demonstrates the correlational heatmap for Harvardx. For instance, the number of participants was positively correlated with the proportion of male participation (r= 0.33) but negatively correlated with the proportion of learners with higher grades than zero (r=-0.21) or the proportion of certifiers (r=-0.28). Readers may infer that an increasing number of male online learners were enrolled in the Harvard open online courses; and that while more learners enrolled, less obtained certifications. Figure 8 demonstrates the correlational heatmap for MITx. For instance, the participation rate was positively correlated with the proportion of learners who liked to play video in the open online class and positively correlated with the proportion of learners who earned a higher grade than zero. The codes for the heatmap analysis are provided in Appendix A.

Example 3: Dynamic Dashboard

With the rise of educational technologies, such as Learning Management Systems (LMS) and MOOCs, a new term called *Learning Dashboard* is being used in online educational platforms (Schwendimann et al., 2017). Learning Dashboards typically apprehend and visualize traces of learning activities to facilitate awareness, reflection, and sense-making, and to enable learners to identify goals and track progress towards these goals (Mazza & Dimitrova, 2004; Schwendimann et al., 2017). The Learning Dashboard builds on research in the field of information visualization, learning analytics, and educational data mining. Besides, it transforms data into visualizations, such as graphs, gauges, dials, and maps. Several synonymous terms for learning dashboards are currently in use, including *Educational Dashboard*, *Dashboard for Learning Analytics*, *Data Dashboard*, and *Web Dashboard* (Schwendimann et al., 2017).

Figure 8. Heatmap visual on learners' characteristics and learning behavior from MITx online platform

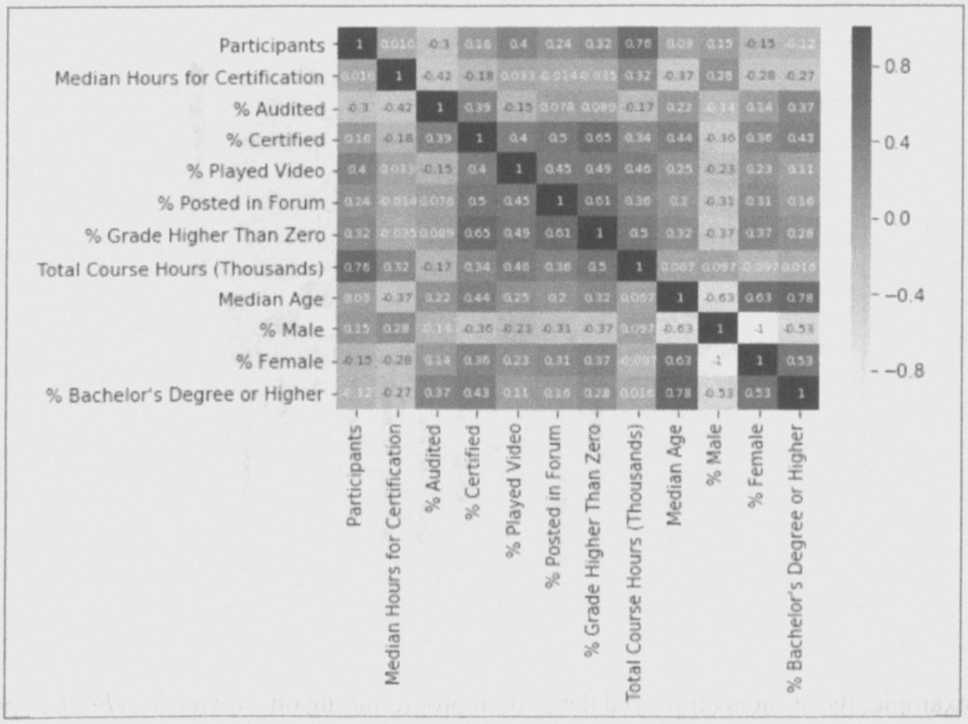


Figure 9a (right), 9b (left). Two major panes in Tableau

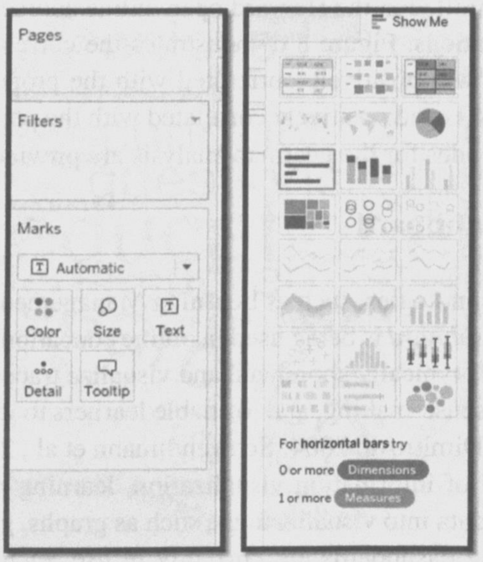
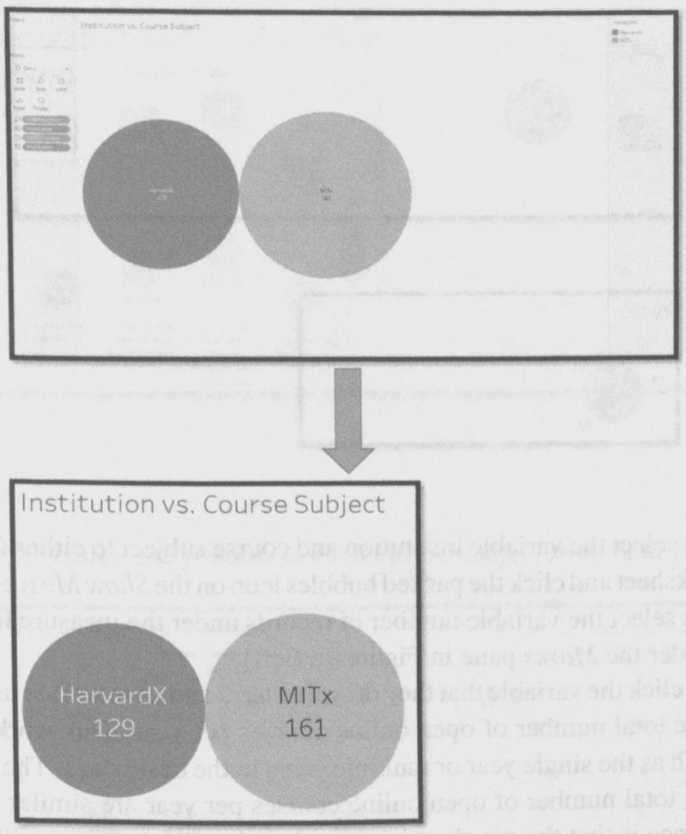


Figure 10. Institution vs. course subject

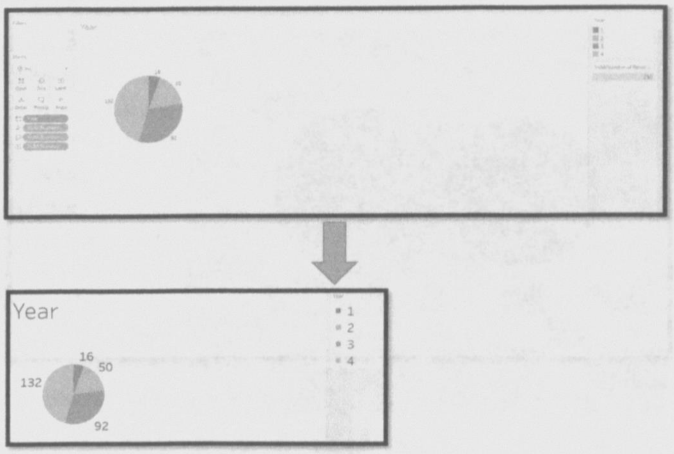


In this section, the authors systematically introduce each visualized worksheet and an integrated interactive dashboard in *Tableau* using the same MOOCs data (described in "Data Souces section") as used in previous sections. Figure 9a illustrates the graph types and the function that provides a vertical pane to select the type for variables. Researchers adjust the other properties for each graph by the functions in Figure 9b. A filter is used for selecting the data range in the current worksheet. Furthermore, researchers use various properties under the *Marks* pane in Figure 9a to navigate a visual by changing its appearances, such as color, font size, graph size, and other functions in the current worksheet.

The *Columns* and *Rows* pane at the top of the worksheet is important for navigating the selection or removal of variables. Variables are separated as *Dimensions* and *Measures* after uploading the data in Tableau. The *Dimensions* represents the categorical variables, and the *Measures* denotes the continuous variables. The option of *Dimensions* and *Measures* is provided to change the data type. Below is an example to illustrate the process of building a dynamic dashboard with the same MOOCs data. Authors demonstrate descriptive statistics of the data and explore the differences among online courses in each institution. Each worksheet was built before the dashboard. Therefore, the comparisons and essential factors can be examined in one dynamic dashboard based on aggregated worksheets. Figure 10 illustrates the first worksheet: institution vs. course subject and the total number of course subjects in each institution.

The following steps show the process of constructing a visual presented in Figure 10 by *Tableau*:

Figure 11. Worksheet for the number of courses per year from both Harvardx and MITx



Step 1: The authors select the variable institution and course subject to either *Columns* shelf or *Rows* shelf in the current worksheet and click the packed bubbles icon on the *Show Me* menu in Figure 9a (right).

Step 2: The authors select the variable number of records under the measure for listing the variables and drag it to *Label* under the *Marks* pane in Figure 9b (left).

Step 3: The authors click the variable that they drags in step 2 and choose *Sum* under the pop-up menu.

Figure 11 shows the total number of open online courses per year. This worksheet is used to filter the specific period, such as the single year or multiple years in the dashboard. The steps to complete this worksheet for showing total number of open online courses per year are similar to the previous worksheet. The only difference is that the pie chart icon needs to be selected on the *Show Me* menu. Figure 12 compares the average certified percentage of each course subject to the average certified at least 50% course content accessed of each course subject. In this worksheet, the average percentage of different levels of certifiers in each subject can be presented in one worksheet. With a closer look at the overall graph, a higher proportion of certifiers had accessed to over 50% of content in each course subject. In other words, the majority of learners who received certifications were those who have access to more than half of the content for each course subject.

The following steps show the process of constructing a visual presented in Figure 11 by *Tableau*:

Step 1: The authors select the variable "course subject" to columns and variable "% certified" to rows.

Step 2: The authors drag the variable "% certified of >50% course content accessed" onto the view to the opposite axis. They drop the field with the presence of the black dashed line.

Step 3: To make the scales on the axes equal, the authors right-click the axis of the second measure ">50% course content accessed" that was added in the graph and choose *Synchronize Axis*.

Step 4: The authors customize the mark type for each distinct measure. For this example, they choose the market type square and circle.

Figure 13 compares the percentage of female and male in each course subject. The steps for this worksheet include the function of using measure values and measure names in a view. The measure values and measure names are *Tableau*-generated fields that serve as containers for more than one measure. In this example, the measure names contain female and male, the side-by-side bar graph can be used to reduce the complexity of the process. The side-by-side bar graph is available under the *Show Me* menu.

Figure 12. Worksheet for certifiers in each course subject from both Harvardx and MITx

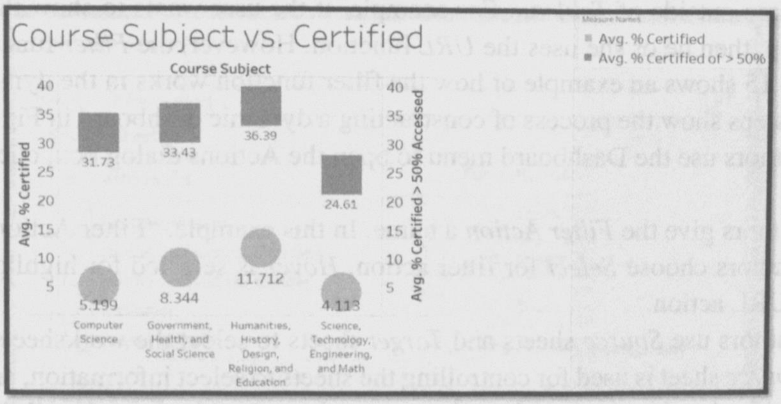
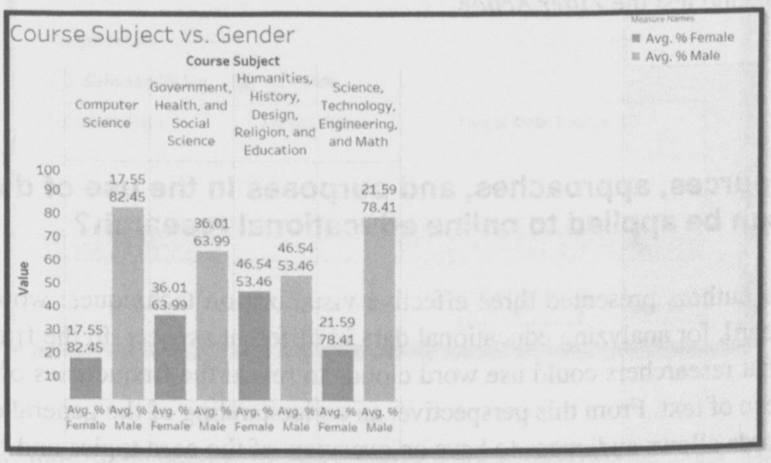


Figure 13. Worksheet for males and females in each course subject from both Harvardx and MITx



The following steps show the process of constructing a visual presented in Figure 13 by *Tableau*:

Step 1: The authors select the variables "course subject", "%female", and "%male" and move them to *Columns* or *Rows* in this worksheet.

Step 2: The authors select side by sidebar under the *Show Me* menu to main the comparison in each column in the graph.

Step 3: The authors drag the variable "%female" and "% male" to the text to have the percentage number at the top of each histogram.

Once the user has created all the worksheets for the dashboard, he or she should click the dashboard button to put them together. The order position, and size are decided by preferences. The most critical process of building a dynamic dashboard is adding actions in the dashboard. There are three different types of add-in functions in the table, including *Filter*, *Highlight*, and *URL* (Figure 14). The *Filter* function is used for capturing information between worksheets, the *Highlight* function is used for calling attention

to marks of interest by coloring select marks, and the *URL* function has used a hyperlink that spots to a web-based resource outside of *Tableau*. For example, if the user wants to show the location of a zip code in google map, then he or she uses the *URL* function. However, the *Filter* function is most useful in *Tableau*. Figure 15 shows an example of how the filter function works in the dynamic dashboard.

The following steps show the process of constructing a dynamic dashboard in Figure 15 by *Tableau*:

Step 1: The authors use the Dashboard menu to open the Actions dialog box, click Add *Action*, and select *Filter*.

Step 2: The authors give the *Filter Action* a name. In this example, "Filter Action" is the name.

Step 3: The authors choose *Select* for filter action. *Hover* is selected for highlight action, and the *Menu* is used for URL action.

Step 4: The authors use *Source* sheets and *Target* sheets to select the worksheets in the interactive dashboard. The *Source* sheet is used for controlling the sheets to select information, and the *Target* sheet is used for measuring the changes based on the source sheet. In this example, all the changes in each worksheet need to be selected after the information is selected, so all the worksheets are selected under *Source* sheets and *Target* sheets.

Step 5: Choose *Show All Values* to include all the unselected information, and the program will make the unselect values shaded, rather than disappear.

Step 6: Click Ok and test the *Filter Action*.

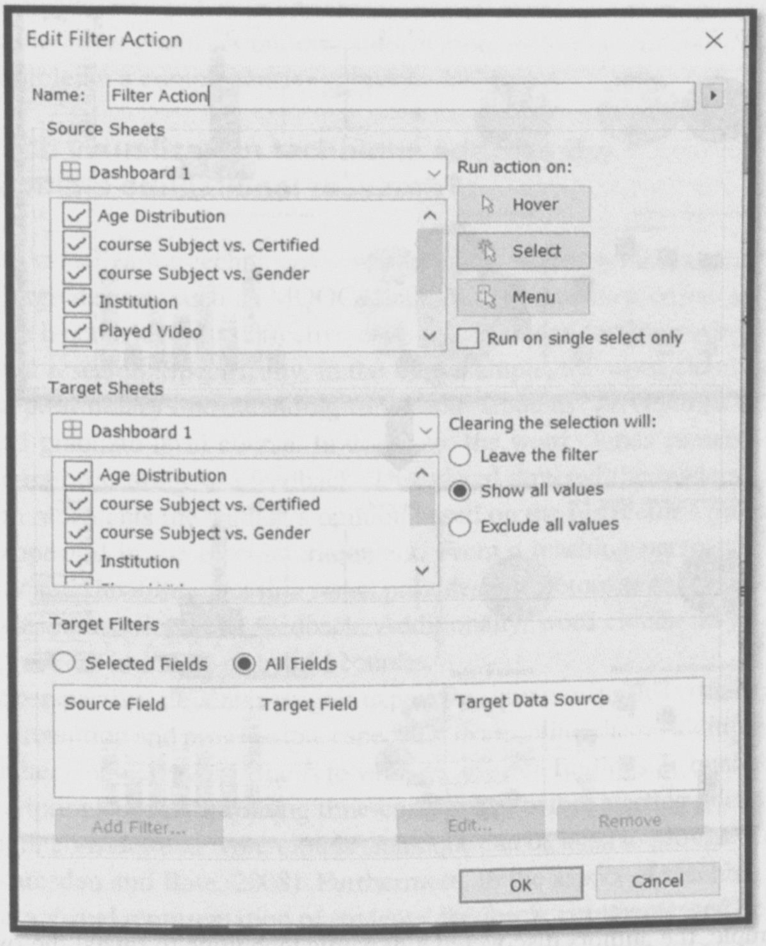
DISCUSSION

1. What data sources, approaches, and purposes in the use of data visualization can be applied to online educational research?

In this chapter, the authors presented three effective visualization techniques: word clouds, heatmap matrix, and dashboard, for analyzing educational data in different aspects. In the first example, the authors discovered that researchers could use word clouds to reveal the frequencies of the diverse words that appear in a piece of text. From this perspective, an understanding of the general composition of the frequently used words allows audiences to have an overview of the core topics and the main themes in a text. However, previous research has also confirmed some of the practical issues in the use of word clouds (Ramsden & Bate, 2008). The critical issue of this technique is that frequency sometimes does not determine the importance even though that the visual presentation often leads the observer to this conclusion (Ramsden & Bate, 2008 - cited in Williams et al., 2013). For any experienced qualitative researcher, it is often a single and contrary response by the person being interviewed that is most illuminating (Williams et al., 2013). This indicates that researchers need to triangulate the data by exploring data in-depth and from different perspectives factors that the word clouds technique has brought into the researcher's consciousness. Despite the weakness, the visual representation of the textual data provides potential avenues for exploration in the text analysis.

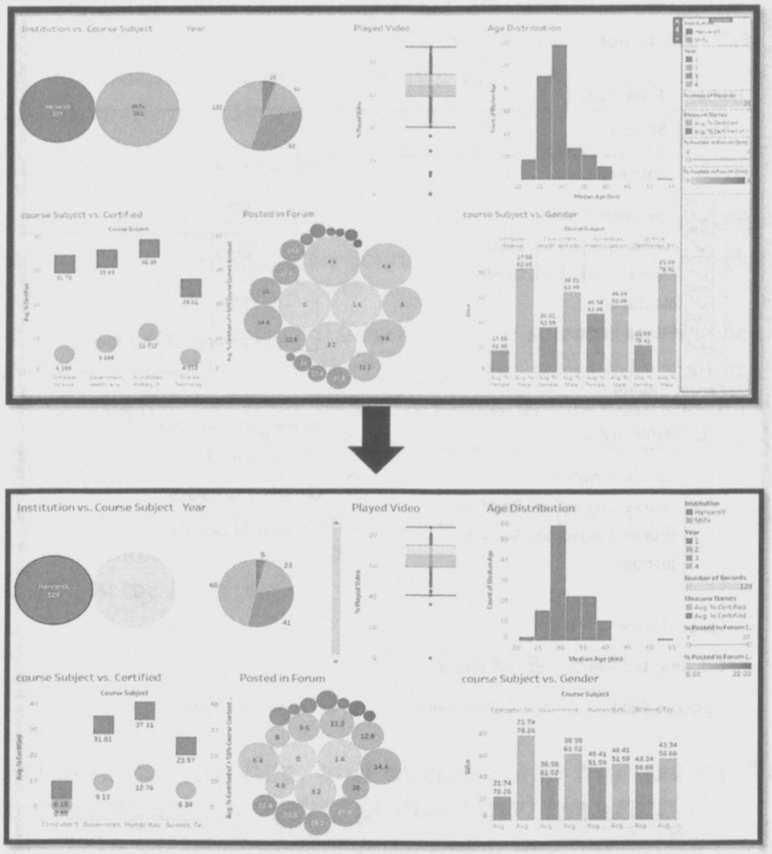
In the second example, the authors demonstrated that the heatmap matrix could be used to reveal the relationships among student demographic characteristics and course-taking patterns. The style of the heatmap matrix used color – specifically, saturation to encode values within cells. From this perspective, representing the data with a matrix can be beneficial, as it facilitates the extraction of exact numerical values or single bits of information, even with large sets of information. Besides, using color

Figure 14. A screenshot of Filter Action box in Tableau



code saturation can help with quickly reading patterns on a quantitative scale. Hence, the heatmap matrix can be used in quantitative research, especially the correlational research, which investigators attempt to understand or explore the possible association among important numerical factors. In this example, one heatmap matrix helps the authors to identify which two factors are correlated by showing the Pearson *R* statistic in both a positive and negative way. The larger the statistic is, the stronger the relationship between the two factors will be. Another significant advantage of heatmap matrix is that it helps investigators quickly identify which important quantitative variables can contribute to statistical models (e.g., regression models) if they are interested in building sophisticated predictive models. For instance, when researchers have a large dataset, such as MOOCs data, with numerous variables /factors listed, investigators often find it difficult to find an optimal analytical approach to make the best use of the data. In this situation, correlational statistics are often chosen at the exploratory data stage. When there is an efficient way to visualize the correlational statistics, such as heatmap matrix, investigators can quickly identify the key factors and feel confident in accessing advanced predictive models which are considered as appropriate quantitative methods for the context of the study.

Figure 15. An example of a comprehensive dynamic dashboard after Filter Action



In the third example, the authors discovered a dynamic technique to reveal the overall explanations with all the information from the same MOOCs data. The dynamic dashboard was illustrated by applying the popular visualization tool *Tableau*. In this example, the authors arranged the insights of numerical data quantitatively so that readers who see them can understand their implications and how to act on them clearly in online educational research. *Tableau* not only shows different visuals, such as histograms, bubbles, maps, and pie charts, but also introduces the process of transforming data from static to dynamic. Each stage also mentions the substance of the single worksheet and its function in the dashboard. The authors also used the example of capturing the significant changes in the dynamic dashboard in the end. Importantly, the dashboard is comprised of multiple practical implications of quantitative analyses to any kind of audiences (students, educators, administrators, researchers, and stakeholders). The appeal of this technique is obvious. For those who have a difficult time processing big data, *Tableau* dashboard provides an alternative to transform it into a straightforward format. In addition to quantitative analyses, this dashboard can be applied in content analysis (e.g., studying communicative activities by turning quantitative data into qualitative data), qualitative analysis (e.g., surveys on topics such as students' perceptions about online course), and mixed methods analysis (e.g., survey results such as students' satisfaction are rich with qualitative data, but analysis sometimes requires quantitative comparison, ag-gregations, etc.). To transform data into a format that audiences can comprehend from a distance, the

authors turned to a data visualization known as *Tableau* worksheet. For example, the authors aggregated course selections by institutions and course preferences by gender. In interacting with readers, the authors demonstrated a clear logic for gradually building information, including students' MOOC learning trends and comparisons visible, in a comprehensive dynamic dashboard.

2. How does each visualization technique address the type of data in online educational research?

In general, all three visualization techniques – word clouds, heatmap matrix, and dashboard – can be used to analyze a large dataset, such as MOOCs data. In the context of online teaching and learning, these techniques can be considered as effective data visualizations for addressing big or complex data in online educational research. Specifically, in the first example, the word clouds effectively provided readers with a fast, preliminary understanding of the four students' perceptions regarding educational research in a hybrid graduate-level course. In doing so, the word clouds presented a visual of words based on each discussion posting and feedback. This visual directed the readers' attention to the size of the words, which represents the student's opinion based on the instructor's prompt question and the frequency that it appeared in the discussion content. From a teaching perspective, educators cannot ignore the creativity and flexibility that this visual provides instructors as an alternative way of examining students' discussion activities and feedback. Additionally, word clouds may prove useful in their understanding of practices in online or hybrid courses.

From a learning perspective, students may gain experience in exploring this visual representation, which can help to improve retention and progression, especially in an online class. Additionally, from a research perspective, researchers may use word clouds to validate research findings. In qualitative research, when studying the transcriptions in full becoming time-consuming, before starting a long analytical process, the researchers might consider that word clouds technique can be used to provide a quick outline of the qualitative data (Ramsden and Bate, 2008). Furthermore, in the aspect of teaching and learning, word clouds can serve as a visual representation of students' feedback, comments, and perspectives in online courses' modules such as discussion forums. Therefore, instructors or educational researchers will get a quick and precise visualization of the prominence of students' standpoints and context of students' learning experiences (Williams et al., 2013). When word clouds become an instructional tool in online class, they can potentially promote students' critical thinking, quality of learning, and peer interaction. For instance, DeNoyelles and Reyes-Foster (2015) conducted a study that required students in their online class to analyze two famous speeches in the form of word clouds. Students were then given a discussion prompt. Students who were exposed to the word clouds performed significantly better in responding to the discussion prompt than students who were not. These evidences support the word cloud technique's function as an innovation in deciphering unstructured qualitative data, garnering feedback, positive or negative comments, as well as enhancing online teaching and learning environment.

Correlation statistics is considered as an effective analytical method for showing the extent of the relationship among numeric variables. In this chapter, the authors presented the heatmap matrix using a popular, free web-based interactive tool called *Python Jupyter Notebook*. The results from the second example indicated that between 2012 and 2016, online learners in general were more likely to choose MIT open online courses (MITx) compared with Harvard open online course enrollment. Those who chose MIT courses were more likely to complete the course and achieved a better learning outcome

compared to those who chose Harvard open online courses. This is just a simple scenario for showing readers how to use the basics in correlational visualization.

A visualization like a heatmap matrix can be used to reveal online students' usage pattern and learning behavior in the context of web-based or online. For instance, Qu and Chen (2015) revealed their findings that most students watched online course video for no more than 6 minutes (meaning the videos could only hold their attention for this amount of time), and courses of longer videos usually had a higher dropout rate. They showed these pieces of evidence by demonstrating the correlational visualization: scatterplot. Therefore, they proposed that course video design guidelines should include creating shorter videos and avoiding abrupt visual transitions. However, the advantage of heatmap matrix over scatterplot is that the former is more suitable for showing pattern among multiple variables. This is especially important when investigators want to detect any significant relationships using extensive complex information with multiple attribute columns in the dataset. Also, correlational visualizations can be used in the exploratory analysis before the development of a predictive model in online research. For instance, when investigators want to build a dropout prediction model based on the MOOCs data (Chen, Chen, Zhao et al., 2016), they may choose to use heatmap matrix to color highlight the essential variables (students' demographics), so the dropout rate could be potentially contributed to important predictors (e.g., instructors are usually interested in whether a leaner will finish a course by reviewing if he or she receives a certificate or pass the final exam).

Dynamic dashboard in *Tableau* is powerful to drive visuals and data analysis. From simplicity to complexity, *Tableau*'s visual analytics gradually bring audiences to a flexible front-end for data exploration with the necessary analytical depth in online educational research. The results from the third example showed that MITx held more open online course than Havardx in general between 2012 and 2016. Among these course subjects, Humanities, History, Design, Religion, and Education (HHRDE) typically had the most certifiers and explorers, followed by Government, Health, and Social Science (GHS). However, participants had the greatest difficulty in earning certifications from Science, Technology, Engineering, and Mathematics (STEM). This was perhaps because the specialization of the STEM courses was at a very high level compared to the more general interest nature of HHRDE courses. In comparing participation by gender, female participation was quite low in computer science and STEM courses. This is an example for showing audiences how to visualize learning traces for learners and teachers, to facilitate awareness, reflection, and sense-making (Schwendimann et al., 2017; Verbert et al., 2013). For online research used data from a MOOC platform (e.g., EdX), the dashboard systematically analyzed the demographics and behavior of participants in each course subject provided by the MOOC platform. The dashboard also comprehensively established an analytical framework for understanding the nascent online learning context of MOOC. For those who choose this technique, they should to fully understand the research question and optimal goals for building the dynamic dashboard to maintain useful and sufficient information in every worksheet. Thanks to the features of *Tableau* dashboard, the authors encourage readers to attempt this tool to tackle the challenges in producing creative, timely, and accurate visuals when using big or complex data in online educational research.

FUTURE RESEARCH DIRECTIONS

These examples can be analyzed under the lens of existing online/web-based research to identify future directions. First, bringing more data visualization tools into online classrooms may enable instructors to provide personalized, formative, and summative feedback. This approach not only enables educators to better understand the learning process beyond a final product, but also promotes students' metacognitive development. Future studies can explore the effect of using visual analytics on students' metacognitive development and self-regulation (Azevedo & Hadwin, 2005 - cited in Vieira et al., 2018).

Visual learning analytics provides the merits of dealing with large dataset (Vieira et al., 2018). Moreover, including students' demographics and historical data can help researchers to better understand their online learning process. Besides, more innovative and interactive visualization tools may help deal with the complexity of the heterogeneous data (Dawson, 2010). Future research should also explore how to better incorporate information visualization into educational research fields. There is a substantial body of knowledge in educational literature explaining how students learn, how students interact with each other, how instructors can support student learning (e.g., Bransford et al., 1999; Schwartz, Tsang, & Blair, 2016), and how students interact in collaborative data visualization environments (Byrd & Vieira, 2017). For example, the design of word clouds for students can be informed by educational studies on critical thinking and engagement (DeNoyelles & Reyes-Foster, 2015). Finally, researchers tend to advocate for more sophisticated visualization (Dawson, 2010). The authors define the term *sophisticated visualizations* in the context of this chapter as: (1) being interactive; (2) being dynamic; (3) being creative. These three elements may also help educators, educational researchers, and stakeholders to gain additional insight from visual analytics tools.

CONCLUSION

The field of data visualization has started to emerge as an opportunity to provide insights and inform instructional decisions using a large amount of heterogeneous data. This chapter introduced a few existing visualization tools and approaches in the context of online learning platforms. The authors purposefully selected three types of visuals: word clouds, correlational heatmap, and dynamic dashboard, to represent the state of the art of the current visualization methods. The goals of this chapter were: (1) to identify the primary uses of visualizations in online educational research, and (2) to customize data visualization techniques for online educational research; and (3) to create visualization tools that can help advance the readers' understandings in the field of data visualization.

The main limitation of this chapter is one inherent to any related data visualization presentations. The authors explored a limited set of venues, and this selection can bring bias to the results of the examples provided. However, the selected literature and data sources are relevant and well-established in the areas of big data and data visualization. The authors hope that this chapter will spur more interdisciplinary work at the intersection of visualization and education, leading to the development of new tools, techniques, methods, theories, and frameworks for the emerging field of visual analytics and big data in online educational research.

REFERENCES

Atapattu, T., Falkner, K. E., & Tarmazdi, H. (2016). *Topic-wise classification of MOOC discussions: A visual analytics approach*. Paper presented at the International Conference on Educational Data Mining (EDM). Retrieved from https://eric.ed.gov/?id=ED592727

Azevedo, R., & Hadwin, A. F. (2005). Scaffolding self-regulated learning and metacognition – implications for the design of computer-based scaffolds. *Instructional Science*, *33*(5-6), 367–379. doi:10.100711251-005-1272-9

Bodily, R., & Verbert, K. (2017). Trends and issues in student-facing learning analytics reporting systems research. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference* (pp. 309-318). ACM. 10.1145/3027385.3027403

Bransford, J. D., Brown, A., & Cocking, R. (1999). *How people learn: Mind, brain, experience, and school*. Washington, DC: National Research Council.

Byrd, V., & Vieira, C. (2017). *Visualization: A conduit for collaborative undergraduate research experiences*. Paper presented at 2017 ASEE Annual Conference & Exposition, Columbus, OH. 10.18260/1-2--29108

Chen, Q., Chen, Y., Liu, D., Shi, C., Wu, Y., & Qu, H. (2016). Peakvizor: Visual analytics of peaks in video clickstreams from massive open online courses. In Proceedings of Transactions on Visualization and Computer Graphics (vol. 22, pp. 2315–2330). Baltimore, MD: IEEE.

Chen, Y., Chen, Q., Zhao, M., Boyer, S., Veeramachaneni, K., & Qu, H. (2016). DropoutSeer: Visualizing learning patterns in massive open online courses for dropout reasoning and prediction. In *Proceeding of Conference on Visual Analytics Science and Technology (VAST)* (pp. 111–120). Baltimore, MD: IEEE. 10.1109/VAST.2016.7883517

Conde, M. A., Garcia-Penalvo, F. J., Gomez-Aguilar, D. A., & Theron, R. (2014). Visual learning analytics techniques applied in software engineering subjects. In *Proceedings of Frontiers in Education Conference (FIE)* (pp. 1-9). Madrid, Spain: IEEE. 10.1109/FIE.2014.7044486

Dawson, S. (2010). 'Seeing' the learning community: An exploration of the development of a resource for monitoring online student networking. *British Journal of Educational Technology*, *41*(5), 736–752. doi:10.1111/j.1467-8535.2009.00970.x

DeNoyelles, A., & Reyes-Foster, B. (2015). Using word clouds in online discussions to support critical thinking and engagement. *Online Learning*, *19*(4). Retrieved from https://eric.ed.gov/?id=EJ1079614

Echeverria, V., Martinez-Maldonaldo, R., Buckingham Shum, S., Chiluiza, K., Granda, R., & Conati, C. (2018). Exploratory versus explanatory visual learning analytics: Driving teachers' attention through educational data storytelling. *Journal of Learning Analytics*, *5*(3), 72–97. doi:10.18608/jla.2018.53.6

Gaebel, M. (2014). MOOCs: Massive open online courses. EUA occasional papers. *European University Association*. Retrieved from https://eric.ed.gov/?id=ED571272

Gilbert, J. K. (2005). *Visualization in science education*. Springer. doi:10.1007/1-4020-3613-2

Gilbert, J. K. (2008). Visualization: An emergent field of practice and enquiry in science education. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Eds.), *Visualization: Theory and practice in science education* (pp. 3–24). Springer. doi:10.1007/978-1-4020-5267-5_1

Hollands, F., & Kazi, A. (2019). *MOOC-based alternative credentials: What's the value for the learner?* Retrieved from https://er.educause.edu/articles/2019/6/mooc-based-alternative-credentials-whats-the-value-for-the-learner

Holt, A. J. (2018, June 6). *Generating correlational heat maps in seaborn.* Retrieved from https://ajh1143.github.io/Corr/

Kaggle Inc. (2019). *Online courses from Harvard and MIT: What subjects or courses are the most popular on edX?* Retrieved from https://www.kaggle.com/edx/course-study

MAUT- McGill University. (2018). *A brief history of MOOCs.* Retrieved from https://www.mcgill.ca/maut/current-issues/moocs/history

Mazza, R., & Dimitrova, V. (2004). Visualizing student tracking data to support instructors in web-based distance education. In *Proceedings of the 13th international World Wide Web Conference on Alternate Track Papers & Posters* (pp. 154-161). New York, NY: ACM. 10.1145/1013367.1013393

McNaught, C., & Lam, P. (2010). Using Wordle as a supplementary research tool. *Qualitative Report, 15*(3), 630–643.

National Forum on Education Statistics. (2016). *Forum guide to data visualization: A resource for education agencies* (NFES 2017-016). Washington, DC: National Center for Education Statistics.

Olshannikova, E., Ometov, A., Koucheryavy, Y., & Olsson, T. (2015). Visualizing big data with augmented and virtual reality: Challenges and research agenda. *Journal of Big Data, 2*(1), 22. doi:10.118640537-015-0031-2

Qu, H., & Chen, Q. (2015). Visual analytics for MOOC data. *IEEE Computer Graphics and Applications, 35*(6), 69–75. doi:10.1109/MCG.2015.137 PMID:26594957

Rapp, D. N., & Kurby, C. A. (2008). The "ins" and "outs" of learning: Internal representations and external visualizations. In J. K. Gilbert, M. Reiner, & M. Nakhleh (Eds.), *Visualization: Theory and practice in science education* (pp. 29–52). Springer. doi:10.1007/978-1-4020-5267-5_2

Reimers, G., & Neovesky, A. (2015). Student focused dashboards - an analysis of current student dashboards and what students really want. In *Proceedings of the 7th International Conference on Computer Supported Education* (pp. 399-404). Lisbon, Portugal: CSEDU. 10.5220/0005475103990404

Ritsos, P. D., & Roberts, J. C. (2014). Towards more visual analytics in learning analytics. In *Proceedings of the 5th EuroVis Workshop on Visual Analytics* (pp. 61–65). Swansea, UK: Eurographics Association.

Rollins, A. (2018). *What's A MOOC? History, principals, and characteristics.* Retrieved from https://elearningindustry.com/whats-a-mooc-history-principles-characteristics

Schwartz, D. L., Tsang, J. M., & Blair, K. P. (2016). *The ABCs of how we learn: 26 scientifically proven approaches, how they work, and when to use them.* New York, NY: W.W. Norton & Company.

Segenchuk, S. (1997). *The role of visualization in education*. Retrieved from https://web.cs.wpi.edu/~matt/courses/cs563/talks/education/IEindex.html

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *The American Behavioral Scientist*, *57*(10), 1380–1400. doi:10.1177/0002764213498851

Telea, A. C. (2014). *Data visualization: Principles and practice*. Wellesley, MA: AK Peters. doi:10.1201/b17217

Thomas, J. J., & Cook, K. A. (2006). A visual analytics agenda. *IEEE Computer Graphics and Applications*, *26*(1), 10-13. doi:10.1109/mcg.2006.5

U. S. Department of Education. (2017). Reimagining the role of technology on education. *2017 National Education Technology Plan Update*. Retrieved from https://tech.ed.gov/files/2017/01/NETP17.pdf

Verbert, K., Manouselis, N., Drachsler, H., & Duval, E. (2012). Dataset-driven research to support learning and knowledge analytics. *Journal of Educational Technology & Society*, *15*(3), 133–148. Retrieved from https://www.jstor.org/stable/pdf/jeductechsoci.15.3.133.pdf

Vieira, C., Parsons, P., & Byrd, V. (2018). Visual learning analytics of educational data: A systematic literature review and research agenda. *Computers & Education*, *122*, 119–135. doi:10.1016/j.compedu.2018.03.018

Williams, W., Parkes, E. L., & Davies, P. (2013). Wordle: A method for analyzing MBA student induction experience. *International Journal of Management Education*, *11*(1), 44–53. doi:10.1016/j.ijme.2012.10.002

ADDITIONAL READING

Freund, R., & Wilson, W. (2010). *Statistical methods*. New York, NY: Elsevier.

Trochim, M. K. (2006). *Research methods knowledge base*. Retrieved from http://www.socialresearchmethods.net/kb/index.php

KEY TERMS AND DEFINITIONS

Big Data: A large volume of data that may be analyzed computationally and systematically to reveal characteristics, patterns, trends, and associations in various fields.

Correlational Heatmap: A visual data visualization tool that demonstrates correlational statistics by applying color-coding to represent the relationship between data values. It is advantageous to explore the two-dimensional data.

Dashboard: A web-based program that apprehends and visualizes traces of learning activities, to facilitate awareness, reflection, and sense-making, and to enable learners to identify goals and track progress towards these goals.

MOOCs: Also knowns as Massive open online courses, MOOC was first used in 2008 in the educational community by Stephen Downes and George Siemens.

Python Jupyter Notebook: A data analysis program that produces a standard correlational heatmap visual for demonstrating the potential relationship among the factors by applying Python Programming.

Tableau: A powerful and fast-growing data visualization tool that is used in the industry of business intelligence. This tool helps to simplify raw data into a straightforward, understandable format.

Visualization: A data analysis technique that emphasizes the external representation of abstract or concrete ideas to help people understand the meaning of the expressed information, such as images, diagrams, and animations.

Word Clouds: A visualization of text in which the more frequently used words are adequately emphasized by occupying more prominence in the representation.

Wordle: A web-based platform that demonstrates a fast and rich visual to enable researchers to obtain a basic understanding of the data (textual information). This platform produces word-cloud analyses of the spoken and written responses of informants in research projects.

APPENDIX

Python Programming Code in Example 2

```
import seaborn as sb
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
address='I:/2018/book chapter/data/heatmap.csv'
heatmap=pd.read_csv(address)
corr=heatmap.corr()
corr
sb.heatmap(corr,annot=True,annot_kws={'size':7})
#harvard heatmap
sb.heatmap(corr_h,annot=True,annot_kws={'size':7},cmap="Greens")
#MIT heatmap
sb.heatmap(corr_m,annot=True,annot_kws={'size':7},cmap="BuPu")
```