Institute of
Higher Education
UNIVERSITY OF GEORGIA

IHE Research Projects Series

IHE Research Projects Series 2019-003
Submitted to series: April 17, 2019

Limitations in Data Analytics: Considerations Related to Ethics, Security, and Possible Misrepresentation in Data Reports and Visualizations

Karen L. Webber
Institute of Higher Education, University of Georgia, kwebber@uga.edu
&
Jillian Morn
Emory University School of Law

Webber, K.L. & Morn, J. (2019). Limitations in data analytics: Considerations related to ethics, security, and possible misrepresentation in data reports and visualizations. *IHE Research Projects Series, 2019-003*. Retrieved from: https://ihe.uga.edu/rps/2019_003.

Citation: Webber, K.L. & Morn, J. (2019, in progress). Limitations in data analytics: Considerations related to ethics, security, and possible misrepresentation in data reports and visualizations. in K.L. Webber, & H. Zheng. *Data Analytics in Higher Education*. Baltimore, MD; Johns Hopkins University Press.

Limitations in Data Analytics: Considerations Related to Ethics, Security, and Possible Misrepresentation in Data Reports and Visualizations

Chapter 4

Karen Webber and Jillian Morn

(April 2019 draft)

Introduction

Rapid increases in technology have led to an escalating demand for and use of data and advanced analytics to inform decision making in higher education. Data analytics related to students and their experiences in admissions, enrollment management, financial aid funds dispersement, and activities related to academic learning and performance has evolved quite rapidly (Lester et al. 2017) and continue to evolve each day as the field of data science captures more attention across the education, government, and business sectors. The growing ubiquity and granularity of data offer new opportunities for research and decision making. There are significant benefits, but also risks associated with the increased availability of data as the higher education industry shifts toward a greater reliance on big data analytics and open access (Mathies, 2018). In particular, many decision makers in higher education lack formal training and/or expertise with educational data (Horta et al. 2017) resulting in a data knowledge gap within senior institutional management (Ransbotham et al. 2015). In addition, there are an increasing number of data analyses or visualizations that, when presented without the proper context, result in inaccurate conclusions. Context is critical, as there are many questions around the legitimacy, intentionality, and even the ideology of data use within higher education (Calderon, 2015). Inaccurate conclusions can lead to incorrect discussions that can affect student performance, possible program development, and/or policy changes.

There are a number of important considerations that should be addressed related to the growing use of data and analytics in higher education, and some of these topics are covered in other chapters of this book (for example, data governance, appropriate use of predictive analytics, and institutional policies on data distribution). In this chapter, we examine another important topic -- the potential for misuse or misunderstanding of data reports or visualizations. The ease with which data can now be collected and stored tempts institution officials to gather more data, often before decision are made for how the data will be used. Senior leaders in higher education may lack formal training and/or expertise with educational data (Horta, Bouwma-Gearhart, & Park, 2017) and may be less familiar with details of the data elements or advanced statistical treatments of data. The confluence of these issues certainly prompts the need and value for data experts, often those in Institutional Research, who can provide context-based information presented in ways that can be meaningfully interpreted and used.

Even though large volumes of data can be stored easily and often, at low cost, data analysts should be mindful of some important limitations that can plague analytics and the effective presentation of data for decision making. Following an overview of the uses and misuses of data, we examine five limitations related to the use of data and analytics in higher education. First, we examine some principles of cognitive science that remind us of practices to ensure accurate presentation and/or interpretation of data and visuals. Secondly, we discuss how misunderstandings and/or misinterpretations occur through the use of misuse of data definitions, through mathematical calculations, through data visualizations, and by intentional manipulation or falsification of data. A third and final point discussed in this chapter focuses on the ethical use of data in higher education.

Overview of the Uses and Misuses of Data within Higher Education

Individuals in higher education institutions (HEIs) use data in a variety of ways from financial and resource management to learning more about the activities of community members (students and staff). Parnell et al. (2018) report that student information systems (admissions, financial aid, and academic course data) are the most frequently used data systems in higher education. However, the fastest growing segment of data use in higher education more recently relates to student behaviours and their experiences, particularly on academic performance (i.e., student success and institutional engagement, Lester et al. 2017). The evolution of data use within HEIs is arguably no different from what is occurring in the rest of society; the movement into a digital era where technologies are collecting, processing, and presenting information at high speeds (GAO 2016; Shepherd 2004).

In higher education, there are strong arguments for using data for evidence-based decision-making (Calderon, 2015; Webber, 2018). However, data has been misused /or misinterpreted and on a number of occasions, often publicized in national news outlets. Recent examples highlight, in most cases, a lack of data protocols or poor administrator judgement lead to data misuse (recent examples include misreporting of business school scores at Temple University and misreporting of undergraduate student admissions scores at Claremont College). While these past data misuses cause concern, there are other impending data challenges that require our attention as well. Mathies (2018) examines two recent disruptions in particular, the layering of coding within software systems (Somers 2017) and platform capitalism (Robertson 2018; Srnicek, 2017) that will likely lead to future data misuses in higher education. These are discussed below, and again in the concluding chapter 12 in this book.

Cognitive Science and the Use of Data and Visualizations

Along with considerations given to what data is collected and how it is used, attention must be given to principles of cognitive science and how an individual may benefit or be hindered in designing or interpreting reports and data displays. Whether poor data visuals were intentional or unintentional, both essentially lead to misinterpretation or misunderstandings. Well-designed visualizations are the goal for accurate interpretation. They can convey complex data or information and can minimize data misuse (Drake, Pytlarz, & Patel 2018). With the increasing use of data and analytics software in higher education, understanding how data are visualised and what they show is becoming increasingly important. This is particularly relevant related to third-party analytics tools, as their core elements are proprietary and often not shared with clients (Alamuddin et al., 2016).

Data Visualizations

Data visualizations consist of the tools, technologies, techniques, and methodologies to display, explore, and communicate data and information in a graphical form (Drake, Pytlarz,

& Patel 2018). Reviewing research on cognitive science and humans' understanding of visuals are helpful in developing better data visualizations in higher education.

Although visual-spatial displays are ubiquitous in human communication (Hegarty 2011), recent advances in computer graphics and human–computer interaction techniques, dynamic and interactive displays have become commonplace and changed the way we think about and display data. Cognitive scientists argue that representations that are informationally equivalent (contain the same information) are not necessarily computationally equivalent (Larkin & Simon 1987). In the case of visual-spatial displays, Hegarty (2011) reports that there is substantial evidence that task performance can be dramatically different based on varying visual displays of the same information.

A visual system helps the viewer senses the features of an image or display such as color and shape. The system helps encodes these features to the viewer's brain constructing an internal (cognitive) representation of the image or display. Exactly which of these features are encoded depends on the viewer's attention, one's goals and expectations, and the salience of the display (Hegarty, 2001). For example, one difficulty in display comprehension is viewer distraction by highly salient but task-irrelevant information; a viewer will fail to encode the critical information although it is present. In addition to basic perceptual, attentional, and encoding processes, which construct a representation of the external display, the user of a display typically applies knowledge to construct a representation of its referent (Hegarty, 2001). This includes knowledge of the display conventions such as the meaning of the x and y axes in a graph, types of information typically included and which are omitted, and which aspects that can be taken literally (such as the relative length and configuration of roads on a road map) and which are not (such as the color and width).

According to Tversky et al., (2002), an individual's intuitions about the effectiveness of a display do not always conform to the display's actual effectiveness. Intuitively, animations of physical processes, such as the workings of machines, and biological mechanisms should be a good means of communicating how processes work. In animations of physical processes, the shapes, locations, and movements of parts of the representation should correspond directly to the shapes, locations, and movements of their referents. However, in a review of several papers comparing animated to static displays, Tversky et al. (2002) indicated that there were no advantages to animations over static displays, pointing out that animations are often ineffective because they are too fast or too complex for the brain to process well.

The Application of Cognitive Science in Data Visualization

A number of scholars offer important information on the how the human brain processes information, including Edward Tufte, Daniel Kahneman, and Stephen Kosslyn. Other scholars, including Alberto Cairo and Stephen Few add their insights on effective and misguided visual presentations of data. As a leading scholar in visual displays, Tufte (1984; 2001) believes that statistical graphics, just as statistical calculations, are only as good as they are developed. A graphic (or calculation), no matter how clever or fancy, cannot rescue an ill-specified (preposterous) model or limited sample size dataset (2001). Further, Tufte (2001) asserts that graphics, when done well, can reveal data and be more precise and informative than conventional statistical computations. He offers a number of important points related to data displays, among these points, graphics should:

- Induce the viewer to think about the substance rather than about methodology, graphic design, or graphic production;
- Make large data seem coherent;

- Encourage the eye to compare different pieces of data;
- Be closely integrated with the statistical and verbal descriptions of a data set; and
- Show the data.

Graphic excellence is important. According to Tufte (2001, p. 92) graphic excellence has the following characteristics:

- Substance, statistics, and design – presentation of data must be well-designed and interesting
- Clarity, precision, and efficiency – Complex ideas are communicated clearly, with little extraneous or unnecessary information
- Parsimonious – the graphic gives the viewer the greatest number of ideas in the shortest time with the least in smallest spaces.
- Multivariate – most graphics will review more than one variable at a time, and
- Truthful – the graph tells the truth about the data.

Tufte (2001) is also concerned with how accurately visual elements represent the data at hand. *Graphical integrity* may seem straight forward, but because data is representable in a variety of ways, there is a tendency to scale data disproportionately in order to make it fit in the space allowed. This can lead to false representations of the data and incorrect conclusions. Because of the variations in the ways data is presentable and often, the limits in space, Tufte (2001) offers six principles of graphical integrity:

- Representation of numbers should match the true proportions;
- Labelling should be clear and detailed;
- Design should not vary; show only data variation;
- To represent money, use well-known units;
- The number of dimensions represented should be the same as the number of dimensions in the data; and
- Context is essential; graphics must not quote data out of context.

Another scholar who has studied judgement and decision making, Daniel Kahneman (2014), argues that individuals too often use ‘gut instincts’ to reach speedy conclusions when reading graphics. The fallacy of this ‘fast thinking’ is that the survival process only looks at the information at hand and does not actively look for additional information. Fast thinking accounts for different understandings by how vividly data points appear in one’s mind and not according to their objective importance. Kahneman (2014) believes there is a premium on presenting a coherent and logically-structured narrative; we are very good at remembering them, as logic and association tie different components together.

Although people often make and remember associations, Tversky and Kahneman (1981, 1986) believe that individuals systematically violate the requirements of consistency and coherence when making decisions. This can lead to inaccurate decisions or misunderstood information. Tversky and Kahneman (1986) trace these violations to the psychological principles that govern the perception of decision-making and the evaluation of options. Decision-making in this way aligns with expected utility. They argue the value function is S-shaped and when faced with a choice, “a rational decision maker will prefer the prospect that offers the highest expected utility” (p. 453).

Along with other scholars mentioned above, Alberto Cairo (2013, 2016) also addresses principles of good data presentation. Cairo (date) believes that mistakes can lead to misinterpretation/bad design; those mistakes include truncated axes, key variables omitted, oversimplification, using the wrong data visualization form for the data, and lacking key annotations. According to Cairo (2013), an infographic seeks to tell the story that its designer seeks to explain, but a data visualization lets people build their own insights based on the evidence provided. Cairo believes that interactive graphics highlight relevant facts first (for example the relationship between income and life expectancy across countries) and then allows the viewer to explore the dataset underlying those facts (by presenting a set of numbers). As such, interactive graphics simultaneously are infographics *and* visualizations of data with at least two layers: a presentation one, and an exploration one.

Organizing the works of other relevant scholars, Hegarty (2011) summarizes a number of principles for effective graphics. Through five principles that include relevant points from other major experts such as Tufte (2001) and Kosslyn (2006), and Tversky (2002), Hegarty astutely points to a number of important considerations, reinforcing the need to think about the needs and capabilities of the audience. Principles related to expressiveness of displays, for example, speak to the importance of presenting reports or visuals that display only the relevant data or information so that the viewer can focus on the intended numbers or image (Hergarty, 2011). Similar to any good writing, the principles of pragmatics urge the analyst or visual designer to make the most important information easy to see and understand. Being mindful of scale as well as use of consistent colors, symbols, and other conventions help the reader/user to understand the meanings intended.

The above discussion points to how and why some presentations of data can fall prey to misuse or misinterpretation. While this can happen via basic numeric tables, the advances in data analytics and infographics has facilitated the increased likelihood of misuse or misinterpretation. While Mathies' (2018) organizes into four groups, Webber (2018) reasons that data in higher education has the potential to be misused or misunderstood in six ways. Misinterpretations or use can occur when the following are carried out incorrectly: 1) statistical treatment of data; 2) use of data definitions; 3) use of color, scale, and size or proportion; 4) type of chart or graph; 5) representation of context; and 6) multiple challenges that exist from the previous five categories. Sometimes the data user may lack knowledge of data definitions or may lack access to the data. Flawed data governance may allow data use as not intended or when accessible by outside entities (data breaches). These instances often violate the privacy of individuals as well as institution itself. In many cases of data misuse within these three categories, it is often a definitional, mismanagement or misapplication issue (Mathies, 2018). While some good examples exist (e.g., Northeastern University, see: (K will get from Rana) or The University of Georgia, see https://datamanagement.uga.edu/management-and-governance), many HEIs do not have organizational structures or policies adequately addressing the organizational-wide perspective on data governance and its distribution. The lack of policies, protocols, and structures can be remedied by instituting new procedures and guidelines that address data misuse. Mathies (2018) suggests that institutions develop and implement a data sharing mandate that can help build broader groups of data users on a campus. Indeed, such a mandate might be helpful to reduce or eliminate data silos, but we agree with Mathies (2018) that well written policies and strong data governance will be required.

Charts or Graphics That Can Lead to Misinterpretation

As data visualizations become more frequent in higher education, so too is the possibility of visuals that lead to misinterpretation. Below are two examples of how data visuals mislead consumers within the context of other important information on higher education policy, process, and practice. When the proper context for data are used, there is legitimacy and greater accuracy of conclusions drawn (Calderon, 2015). Due to limited space, this section only includes just two structured examples of how data, when reported or visualised, have been misused. Additional examples of poorly designed visuals for quantitative information by Stephen Few are available at http://perceptualedge.com/examples.php.

1. Misinterpretation Based On Different Calculations Used.

To interpret Table 4.1 correctly, it is particularly important to know how a mean or average score were derived. The figures below show the Mean Debt of ALL students versus only BORROWERS. These are not significantly different groups as the BORROWERS are a subset of the ALL group. In short, this is not a comparison between two distinct groups. Perhaps a better graphic would be to include data on the NON-BORROWERS group allowing for comparisons between two distinct groups as well as to the whole group mean.

(Insert Table 4.1 about here)

2. **Misinterpretation Based on Y-axis Scale**.

Figures 4.1 through 4.4 show two bar and two line charts with different Y-axis scales that can lead to different interpretations of the same data. In the first chart, the Y-axis is much smaller (300 points) while the second chart the Y-axis is much bigger (1200 points). The Y-axis in Figure 4.1 makes the change over time appear more dramatic because the amount of change accounted for (roughly 125 points) allows for a 42 percent change graphically (out of the 300 points). The Y axis in Figure 4.2 makes the change over time appear smaller due to the larger possible change available (1200 points) and thus accounts for only about a 10 percent change graphically. The line charts in Figures 4.3 and 4.4 emphasize the same point on perception based on scale. Beginning the Y-axis scale at zero is often advised, but can lead consumers to infer a smaller change. The difference in comprehension could lead to dramatically different policy interventions.

(Insert Figures 4.1 through 4.4 here)

Dashboards and infographics can be especially challenging to complete clearly. Dashboard designers often seek to include a number of different metrics or charts about multiple subgroups on one page. Dashboards and infographics that attempt to provide information on multiple points require the viewer to see and understand multiple items, often leaving the reader to do too much mental work, likely not taking in all of the information.

Readers who are interested in examples of infographics with design issues may wish to follow principles of good graphics design expressed by Taie (2019) at Visme.co. Additionally, readers may wish to review Chibana's (2018) or French's (2018a) version of mistakes commonly made with infographics. Chibana (2018) purports that interactive graphics require readers to do too much mental work. For example, she believes a common mistake is to separate the legend from the main data, forcing the reader to look back and forth between the central visualization and the meaning of each icon or colour (see more at http://blog.visme.co/bad-infographics/#2UFZW4vCLrJADu0p.99). Mistakes include: showing too many visual points that limit the brain's ability to focus (visual data junk); using too many colours and/or too little white space; using inaccurate scales; and having no hierarchy of data or placement of information. To see French's (2018a) full list of mistakes and ways to fix, see: https://www.columnfivemedia.com/how-to-fix-the-15-common-infographic-design-mistakes.

Data Ethics

Finally, we turn our attention to the ethical implications of data misuse in higher education. Similar tp the ethics of human subjects in research, higher education officials must wrestle with the balance of the benefit that can be achieved from more data in light of respect and beneficence of the individual involved. Recent events such as the Facebook/Cambridge Analytics scandal are helping users become aware of the hidden dangers that are deeply embedded in internet sites and has prompted social media companies to be more transparent with the ways in which they collect data to enhance user experience. Authors such as Foer (2018) urge extreme caution, or at least user acknowledgement, that companies such as Google, Facebook, and Amazon are very strategic and effective at guiding individuals' access to information, futher arguing that these companies' algorithms suggest items for purchase or information to read on that website are modifying the way individuals think and behave. In the age of big data and predictive analytics, so too must higher education institutions consider the ethics of how student and employee data is collected, stored, and secured, and how the data is used and the effects of that data on individuals.

The technological advancements that have enabled greater data collection and analysis of higher education activities and outcomes have opened the doors for new ethical dilemmas. To some degree, more data points can provide a more robust picture of a person or an activity, but it seems likely that there is a point of diminishing returns that higher education officials should seek to find. In addition to the need for greater institutional resources (including advanced technical instruments, policies to ensure proposer use and storage of data, and staff members to complete analyses), officials should consider the value-added benefits as well the possible infringements on privacy and the potential for algorithms that may heighten unintended bias.

Collecting and using large volumes of data on students, indeed, is a double-edged sword. Accurate data on today's college students can contribute to informed decision making. It can help officials consider ways to ensure student success from application and entry through the degree or certificate experience, and into post-institution career success. However, while good data can assist in making better, informed decisions, the excitement about or enthusiasm for the ease of data collection and possibilities to increase student success may cloud one's thinking on issues of privacy. The collection of data from learning management systems as well as 'swipe card' data from building entry is now being used to examine the relationships between academic study (time in class or our-of-class study) and time spent on extracurricular activities (e.g., use of athletic facilities or student unions). Knowing when and how to report on data collection is important. Entrance into a building from swipe card data or logging in to a course management system (CMS) lacks detail on what is happening when the student is inside the building or the CMS and may prompt the use of many erroneous assumptions. User knowledge of how and when data is being used is important. In the K-12 sector, Reidenberg, Russel, and Kovnot (2013) reported that schools failed to inform parents about the use of data in the school's cloud services including learning analytics programs. These authors argue that the lack of effective transparency undermines accountability to school communities and to the public (p. 7).

Ethical and Privacy Considerations
Considering ethics and privacy in data use and data governance is incredibly important. There are a number of reasons why data is used for decision support and it is necessary to consider the implications for its use for both individual students and staff as well as for institutional management. Ekowo & Palmer (2017) remind us that "using data ethically is complex, and no magic formula exists" (p.3). Adopting an ethical framework to engage with data is important as the continued fervor for data-informed decisions pushes for collection and storage of more personal data.

The emerging ethical considerations for higher education administrators and researchers center on data privacy, complacency with algorithms and vendor products (including e-learning vendor products), and the ways in which the results of data analysis are used. With greater data collection, higher education institutions and researchers may be empowered to address questions and issues, uncovers patterns of behavior, and help students, staff, and organizations thrive. However, with more data collection efforts comes a higher risk of unintentional disclosures or leaks. Administrators need to consider the ways in which data is stored in the long and short term, how it will be managed, accessed, and protected.

Privacy of student information is important for education due to the possible adverse impact that inappropriate use or disclosures may have on student learning and social development as well as the possible impact on students' well-being by amplifying performance-related stress (Reidenberg & Schaub 2018). Similar to notions of stereotype threat (Steele & Aronson 1995), early awareness systems could prematurely label a student in a failing category that may prompt a student to be less motivated or be hesitant to seek tutoring due to fear of being labeled a failure. Learning analytics programs may gather large amounts of data on student performance in a variety of contexts, from student performance on specific exam questions to overall course assessments to large-scale measures at the institution or school system level. The variation in curriculum, instructor, pedagogy, classroom conditions, and/or mix of student demographics make it very difficult to compare student performance. This variation may call into question the validity of analytic methods that have not accounted for the variance inherent in these and other moderators.

HEI officials should consider how the data they collect may be subject to compliance/alignment with regional, national, and supranational policies. Under FERPA, students have a right to request access to view their academic record. The use of online course click-through data, GIS data on card swipes, and the results of early awareness systems rest in a fuzzy area of interpretation and it seems likely that institutional or federal guidelines on student protections will be revisited in the near future. The European Union's recent General Data Protection Regulation (GDPR) affects US institutions enrolling and employing EU citizens, and should be added to HEI discussions on student privacy. In addition to codifying the right to erasure, EU GDPR requires institutions and companies to document consent or a stated business purpose for processing data on EU citizens.

While respect for privacy is important, Prinsloo and Slade (2015) urge us to consider the implications of students' 'opting out' of data collection. These authors proffer than higher education officials must determine if and when students should have the option to opt out of institutional surveillance. In light of an institution's role to ensure appropriate support and guidance to students, these authors ask important questions such as how opting out may impact the institution's fiduciary role.

Prinsloo and Slade (2015) found that some institutions are creating specific policies on ethical uses of analytics, and while they believe these policies are helpful in creating transparency and 'boundary concerns' (p. 84), further issues remain related to issues of informed consent and how and when students can opt out of data collection. They recommend that HEI officials learn from the privacy issues of social media data and, despite a traditional paternalistic approach, engage more proactively with students to inform and more directly involve them in ways in which both individual and aggregated data can and should be used. Greater engagement with students, they believe, will gain greater trust and cooperation from students.

In *Weapons of Math Destruction*, O'Neil (2016; 2018) provides chilling detail on the ways in which large volumes of data and intentions to identify student success can go wrong. Algorithms are conceived as fair, objective, and efficient forms of analysis (O'Neil, 2016). Predictive analytics even be a way to have more data informed decision making in higher education as opposed to opinion and hunch-based strategy (Green, 2018). However, algorithms are constructed by people, and therefore reflect the judgements and priorities of their creators (O'Neil, 2016). Algorithms can encode human prejudice, misunderstanding, and bias into systems that determine access and outcomes and when they are not continuously updated and/or rely on inaccurate proxy measures can lead to the destructive outcomes (O'Neil, 2016). Like Harel Ben Shahar (2017), O'Neil (2016) argues that destructive effects abound in ill-conceived algorithms, many of which discriminate or further stratify individuals, particularly students in education.

We agree with these scholars who recommend that when constructing algorithms, variable selection should be based on direct measures (not proxies) whenever possible. Administrators selecting vendor products should be mindful of data sharing, privacy, and intellectual property, not providing vendor collection of data unless all forms of use are specified. End users of these products should know what is going to the measurements and be able to evaluate if these are suitable measures. Further, products should be continuously updated based on new data.

In administration, data is more often used as a weapon than as a resource (Green, 2018). Too often, says Green (2018), officials use data to point fingers at failing departments and not for identifying the strategies for success. A data-informed decision can certainly benefit from data, but should also include the nuances and contexts before decisions are made.

Additionally, learning analytics programs gather large amounts of data on student performance in a variety of contexts, from student performance on specific exam questions to overall course assessments to large-scale measures at the institution or school system level. The variation in curriculum, instructor pedagogy, classroom conditions, and mix of student demographics make it very difficult to compare. This variation may call into question the validity of analytic methods that have not accounted for the variance inherent in these and other moderators.

Recommendations for Ethical Use of Data

Following O'Neil's (2018) proposal to implement at Data Bill of Rights, and similar to Calo's (2013) recommendation for a 'consumer subject review board,' Mathies (2018b) urges higher education officials to consider a data-sharing mandate that would allow institutional data to be more accessible to campus colleagues but done so within a data governance plan. In such a data bill of rights institution officials would be required to develop a plan that respects and protects individuals' data, requires programming language that limits coding failures, and includes a data ethics board to review and ensure good data practices (for more details, see Mathies, 2018b).

As higher education officials consider the value and potential consequences of vendor products (e.g., enrollment management, student learning assessment, and advising), we recommend that higher education officials evaluate the analyses that underlie predictive models, use caution in implementing early awareness systems, and strategize on how to react to the results of analysis. In addition to good data governance. Institutional officials should be firm in fully understanding how the vendor product completes calculations (when unknown referred to as the 'black box') and how one vendor product (for example, one used for admissions or enrollment management) relates to data being used in another vendor product (for example, used for student advising). Ekowo and Palmer's (2017) guidelines for ethical uses of predictive analytics in higher education may be helpful for general structures and may be especially helpful to provide support to minimize inaccurate model use.

Conclusion

Evidence-based decision-making requires access to, proper use of, and accurate interpretation of data. The volumes of data available today tempt us to use and share information frequently, and in a variety of ways. The complexities of HEIs and today's students can benefit by the use and analysis of many data points, but users must be mindful to use data in ethical ways using statistically correct analyses and proper visuals. Large volumes of data and new technologies have allowed greater insight into student and institutional behaviors, but proper data management and policies on the use and release of data are critical. Data governance policies require regular review and modification, as processes and protocols become dated. This is particularly true as pre-platform governance frameworks often lack the conceptual tools to envision data technology progress (Weatherby, 2018). Both past and likely future data misuses highlight the need for foresight as well as oversight, from a technical as well a contextual perspective. This is where planning and institutional research (IR) professionals can play an important role. Most senior institutional leaders rely on a small group of senior associates to provide context-based information to manage an institution; if IR or planning professionals are not in this select group it compounds data issues with possible misinterpretations and/or data nuances, quirks, and policy implications that might go unnoticed or ignored (Webber 2018).

As discussed above, cognitive psychologists and other scholars remind us of important principles related to how one 'sees' data and how the brain interprets that data into information. Colour, spacing, scale, and the calculation of numeric data are important to consider when designing graphs and charts. Today's new infographics, while intuitively appealing, may offer challenges to read and understand fully. Tufte's (2001) word still ring true: *above all else, show the data*. Echoing Tufte (2001) and other cognitive scientists, tips to improve data visualizations include removing superfluous information that do not support the study, be mindful of placement (i.e., spacing) of images to facilitate comparisons, using callouts wisely, ensuring sufficient contrast between colors and patterns, including a zero baseline if possible, and designing for comprehension (French, 2018b). Following these suggestions can reduce a great deal of misunderstanding of data visualizations. Tufte's (2001) principle on clarity, precision, and efficiency is particularly important however, it can be challenging to communicate complex ideas clearly and with little extraneous or unnecessary information added. Additionally, parsimonious graphics that provide the greatest number of ideas in the shortest time and in smallest spaces (Tufte, 2001) aligns with Kahneman's (2014) fast-thinking process; encouraging individuals to looks only at the information at hand and not actively look for additional information.

Higher education officials have a duty, often a legal one, to protect student and staff data. They also have a duty to provide the higher education community with the most accurate data and information in a timely manner. Appropriate data governance procedures are essential, and senior personnel who are responsible for institutional research tasks should be integral in data governance policies on campus. Along with data policies that allow access to institutional data, good visual presentations are increasingly popular; when done well, they are a good way to share information to the internal as well as the external community. When data visualizations do not follow principles of good design, they can cause confusion or provide misinformation. The examples above illustrate how data and visualizations are misused or misinterpreted. Seeing these examples and understanding how they have been ineffective or misused can hopefully lead to better data visualizations in the future.

Implications for the Future

The use of data and desire for continued larger volumes of data for informed decision making will only increase in the years ahead in higher education. While there are significant benefits that can serve to improve one or more aspects of higher education, there are also risks. When misused /or misinterpreted, senior leaders may make the wrong decisions, ineffective or unproductive policies may be enacted, and/or students may not be as successful.

To minimize or eliminate misinterpretations and misunderstandings, institution officials should ensure that data dictionaries or other documents that include definitions for institutional data should be readily available to users. One good way to ensure the use of data dictionaries is by embedding them in interactive dashboards or other reports. Analysts who create reports and visuals must be ever-mindful of good principles of graphic design and the need to minimize cognitive strain.

Once reports or visualizations are completed, data analysts or unit leaders should also strive to engage with users to ensure a full understanding of data definitions, the context in which the data resides, and any nuances that affect the data. Most often, a fact-to-face meeting to discuss a new data table, visual, or infographic can be a highly effective way to ensure that the users understand the information presented and offers a great opportunity to answer questions, to answers questions that may arise, or to engage in discussions that may relate to new policies and/or programs.

While data analytics in higher education have great potential to help students and institutions succeed, they also present the need to be aware of and strategic in ensuring ethical and secure use of data. Reidenberg and Shaub (2018) suggest that transparency, accountability, and security should be integral aspects of learning analytics technology rather than afterthoughts, and that higher education officials and commercial learning vendors establish appropriate safeguards to govern appropriate access and use of learning analytics data. They further suggest that legal safeguards for education privacy should reflect the reality of data-driven education by expanding privacy protections to clearly cover learning analytics. These are important points that can help move the use of data analytics in higher education to a stronger and effective position in data informed decision making.

References

Alamuddin, R., Brown, J., & Kurzweil, M. (2016). Student data in the digital era: An overview of current practices. Resource document. Ithaka S+R. http://www.sr.ithaka.org/wp-content/uploads/2016/09/SR_Report_Student_ Data_Digital_Era-090616.pdf. Accessed 27 Apr 2017.

Breslow, L. A., Trafton, G. J., & Ratwani, R. M. (2009). A perceptual process approach to selecting color scales for complex visualizations. *Journal of Experimental Psychology: Applied, 15*, 25–34.

Cairo, A. (2013). *The functional art: An introduction to information graphics and visualizations*. Berkeley, CA: New Riders.

Cairo, A. (2016). *The truthful art: Data, charts, and maps for communication*. Berkeley, CA: New Riders.

Calderon, A. (2015). In light of globalization, massification, and marketization: Some considerations on the uses of data in higher education. In K. Webber and A. Calderon (Eds.), *Institutional research and planning in higher education: Global contexts and themes* (pp.288-306). New York: Routledge Press.

Calo, R. (2013) Consumer subject review boards: A thought experiment. *Stanford Law Review Online* 66: 97. Available at: https://www.stanfordlawreview.org/online/privacy-and-big-data -consumer-subject-review-boards/

Chibani, N. (2018). 11 Common Infographics Mistakes. Accessed 10 June, 2018 at: http://blog.visme.co/bad-infographics/

Drake, B., Pytlarz, I., & Patel, M. (2018) Let me paint you a picture: Utilizing visualizations to make data more accessible. In K. Webber (Ed.) *Building capacity in institutional research and decision support in higher education*. Knowledge Studies in Higher Education, 4. Cham: Springer. 81-93.

Few, S. (2012). *Show me the numbers: Designing tables and graphics to enlighten* (2nd edition). Oakland, CA: Analytics Press.

French, K. (2018a). How to Fix the 15 Most Common Infographic Design Mistakes. Retrieved from: https://www.columnfivemedia.com/how-to-fix-the-15-common-infographic-design-mistakes

French, K. (2018b). 25 Tips to Instantly Improve Your Data Visualization Design. Retrieved from: https://www.columnfivemedia.com/25-tips-to-upgrade-your-data-visualization-design

Gattis, M., & Holyoak, K. J. (1996). Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology: Learning, Memory & Cognition, 22*, 231–239.

Harel Ben Shahar T (2017) Educational justice and big data. *Theory and Research in Education* 15(3): 306–320.

Hegarty, M., Canham, M., & Fabrikant, S. I. (2010). Thinking about the weather: How display salience and knowledge affect performance in a graphic inference task. *Journal of Experimental Psychology: Learning, Memory and Cognition, 36*, 37–53.

Horta, M., Bouwma-Gearhart, J. & Park. H. (2017). Data driven decision making in the era of accountability: Fostering faculty data cultures for learning. *The Review of Higher Education, 40*(3), 391-426.

GAO. (2016). *Data and analytics: Emerging opportunities and challenges*. https://www.gao.gov/assets/690/680265.pdf. Accessed February 2018.

Hegarty, M. (2011). The cognitive science of visual-spatial displays: Implications for design. *Topics in Cognitive Science, 3*, 446-474.

Jaschik, S. (2018, July 10). Lies, damned lies and rankings. *Inside higher Education*. https://www.insidehighered.com/news/2018/07/10/temple-ousts-business-dean-after-report-finds-online-mba-program-years-submitted. Accessed 12 July 2018.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Strauss & Giroux.

Kosslyn, S. (1994). *Elements of graphic design*. New York: W.H. Freeman & Co.

Lester, J., Klein, C., Rangwala, H. & Johri, A. (2017). Learning analytics in higher education. *ASHE higher education report*. 43 (5), 9-135.

Mathies, C. (2018). Uses and misuses of data. In K. Webber (Ed.) *Building capacity in institutional research and decision support in higher education*. Knowledge Studies in Higher Education, 4. Cham: Springer. 95-111.

Mathies, C. (*in press*). Ethical use of data. In C. Mathies & C. Ferland (Eds.) IR in the digital era. *New directions for institutional research*. 178. Boston: Wiley.

Novick, L. R., & Hurley, S. M. (2001). To matrix, network, or hierarchy: That is the question. *Cognitive Psychology, 42,* 158–216.

Palmer, S. E. (1978). Fundamental aspects of cognitive representations. In E. Rosch & B. B. Lloyd (Eds.), Cognition and categorization (pp. 259–303). Mahwah, NJ: Lawrence Erlbaum Associates.

Parnell, A., Jones, D. Wesaw, A., & Brooks, C. (2018). *Institution's use of data analytics for student success*. National Association of Student Personnel Administrators, Inc.; The Association for Institutional Research; and EDUCAUSE. https://www.naspa.org/rpi/reports/data-and-analytics-for-student-success. Accessed April 2018.

Peebles, D., & Cheng, P. C.-H. (2003). Modeling the effect of task and graphical representation on response latency in a graph-reading task. *Human Factors, 45,* 28–46.

Ransbotham, S., Kiron, D., and Prentice, P.K. (2015). Minding the analytics gap. *MIT Sloan management review*. Spring, 63-68.

Reidenberg, J., & Schaub, F. (2018). Achieving big data privacy in education. *Theory and Research in Education, 00*, 1-7, DOI:10.1177/147787851885308

Reidenberg JR, Russell NC, Kovnot J, et al. (2013) Privacy and cloud computing in public schools. Report, Center on Law and Information Policy, Fordham University School of Law. Available at: http://ir.lawnet.fordham.edu/clip/2/

Robertson, S. (2018). Platform Capitalism and the new value economy in the academy. *Culture, Politics, and Global justice research cluster working paper*. Cambridge University. https://cpgj.files.wordpress.com/2018/04/cpgj-working-paper-12.pdf. Accessed February 2018.

Royal Society. (2010). *The scientific century: Securing our future prosperity*. Policy document 02/10. London: Author.

Schwartz, D. L. (1995). Reasoning about the referent of a picture versus reasoning about the picture as the referent: An effect of visual realism. *Memory & Cognition, 23,* 709–722.

Shah, P., & Carpenter, P. (1995). Conceptual limitations in comprehending line graphs. *Journal of Experimental Psychology: General, 124,* 337–370.

Sheperd, J. (2004). What is the digital era? In G. Doukidis, N. Mylonopoulos, and N. Pouloudi (Eds.) Social and economic transformation in the digital era. 1-18. http://www.igi-global.com/chapter/digital-era/29024. Accessed September 2017.

Srnicek, N. (2016). *Platform Capitalism*. Cambridge, UK: Polity.

Strauss, V. (2012, August 23). Elite university lies to college rankers for years. *Washington Post*. https://www.washingtonpost.com/blogs/answer-sheet/post/does-it-matter-that-an-elite-university-lied-to-college-rankers-for-years/2012/08/22/e89b39d2-ec68-11e1-aca7-272630dfd152_blog.html?utm_term=.48cabcd2a554. Accessed 17 Mar 2017.

Steele, C., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans, Journal of Personality and Social Psychology, 69(5), 797-811.

Somers, J. (2017). The coming software apocalypse. *The Atlantic*. September 2017. https://www.theatlantic.com/technology/archive/2017/09/saving-the-world-from-code/540393/. Accessed February 2018.

Taei, P. (2019). A beginner's guide to shareable infographics. Retrieved at: https://www.visme.co/wp-content/uploads/2017/03/How%20to%20Make%20an%20Infographic%20-%20A%20Visual%20Guide%20for%20Beginners%20By%20Visme.pdf

Tufte, E. (2001). *The visual display of quantitative data*. Cheshire, CT: Graphics Press.

Tversky, A. (2001). Spatial schemas in depictions. In M. Gattis (Ed.) *Spatial schemas and abstract thought.* (pp. 79-112). Cambridge, MA: MIT Press.

Tversky, A. & Kahneman, D. (1974). Judgement under uncertainty: Heuristics and biases. *Science, 185*(4157), 1124-1131.

Tversky, A. & Kahneman, D. (1986). Rational choice and framing of decisions. *The Journal of Business 59*(4). Part 2: The Behavioral Foundations of Economic Theory (Oct. 1986), 5251-5278.

Weatherby, L. (2018). Delete your account: on the theory of platform capitalism. *Los Angeles Review of Books*. https://lareviewofbooks.org/article/delete-your-account-on-the-theory-of-platform-capitalism/#! Accessed April 2018.

Webber, K.L. (2018). The future of IR and decision support: Ensuring a seat at the table. In K. Webber (Ed). *Building capacity in institutional research and decision support in higher education*. Cham, Switzerland: Springer. 261-276.

Webber, K.L. & Burns, R.A. (2018). Graduate student debt 2000 to 2012: Trends and implications. Submitted for consideration to peer-reviewed journal, April 2018.

Yeh, M., & Wickens, C. D. (2001). Attentional filtering in the design of electronic map displays: A comparison of color coding, intensity coding, and decluttering techniques. *Human Factors, 43,* 543–562.

Table 4.1

Cumulative Amount Borrowed for Graduate Education, US Institutions, 2012

Graduate Degree Recipients	2011-2012		
Graduate Degree	% Borrowing	Mean Debt (All)	Mean Debt (Borrowers)
Total	63.85	34,412.80	53,894.33
Degree Type			
Master's degree (all)	62.95	25,075.66	39,837.16
Professional degree (all)	84.76	109,385.40	129,053.20
Doctoral degree (all)	49.72	36,587.93	73,582.36
Gender			
Male	62.97	34,778.65	55,234.19
Female	64.45	34,167.62	53,017.00

Source: NPSAS 2012 data from Webber & Burns (2018)

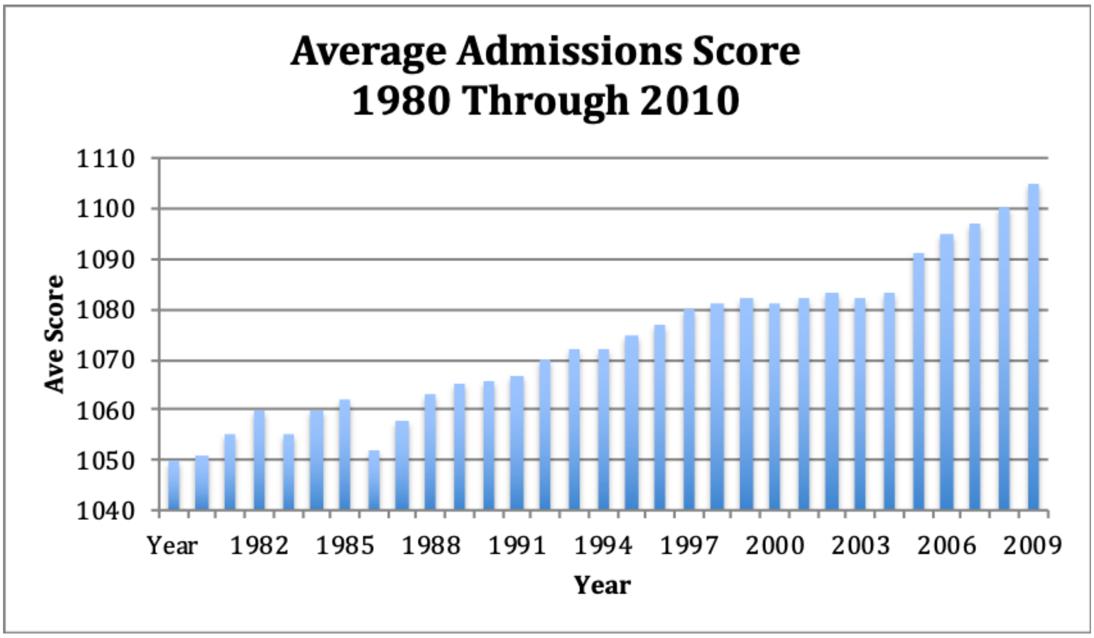

Figure 4.1 Bar Chart of Fictitious Admissions Score Data with Y axis scale from 1040 to 1120

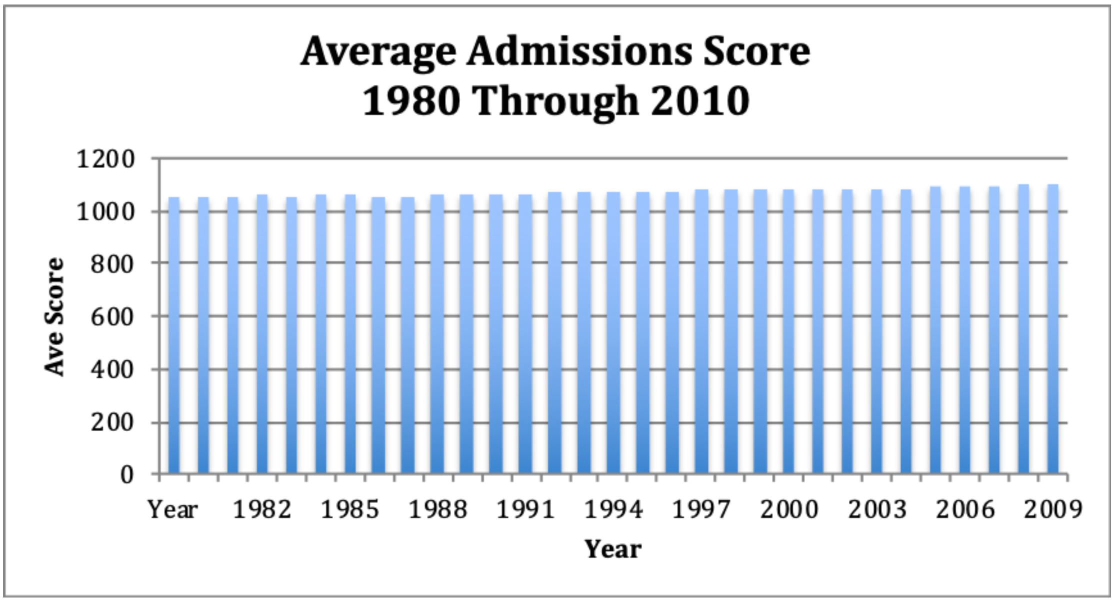

Figure 4.2 Bar Chart of Fictitious Admissions Score Data with Y Axis scale from 0 to 1500

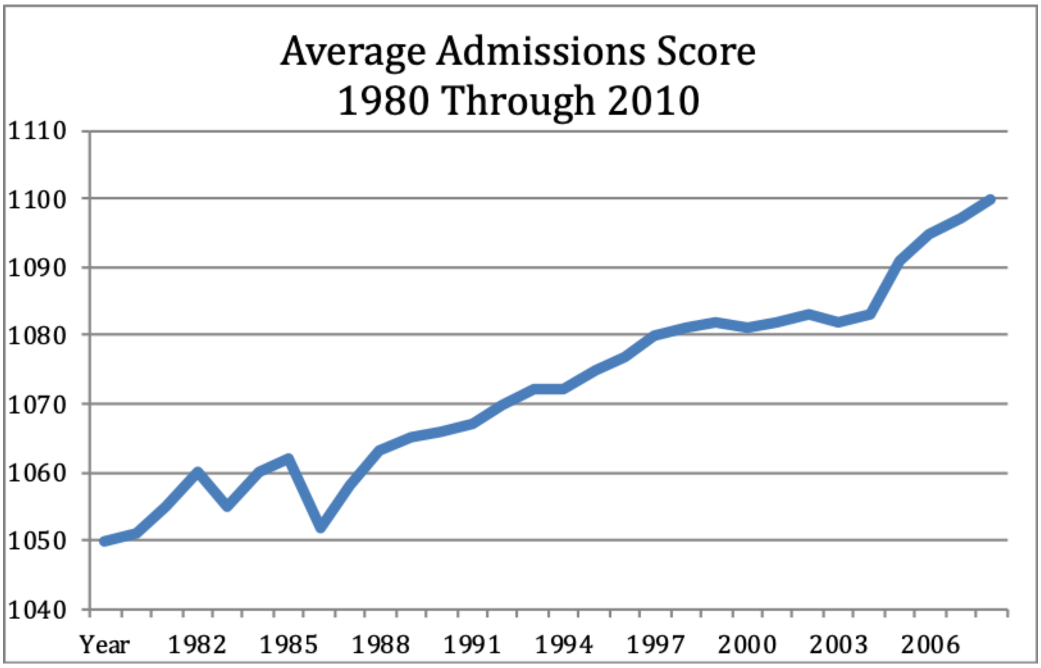

Figure 4.3 Line Chart of Fictitious Admissions Score Data with Y Axis scale from 1040 to 1120

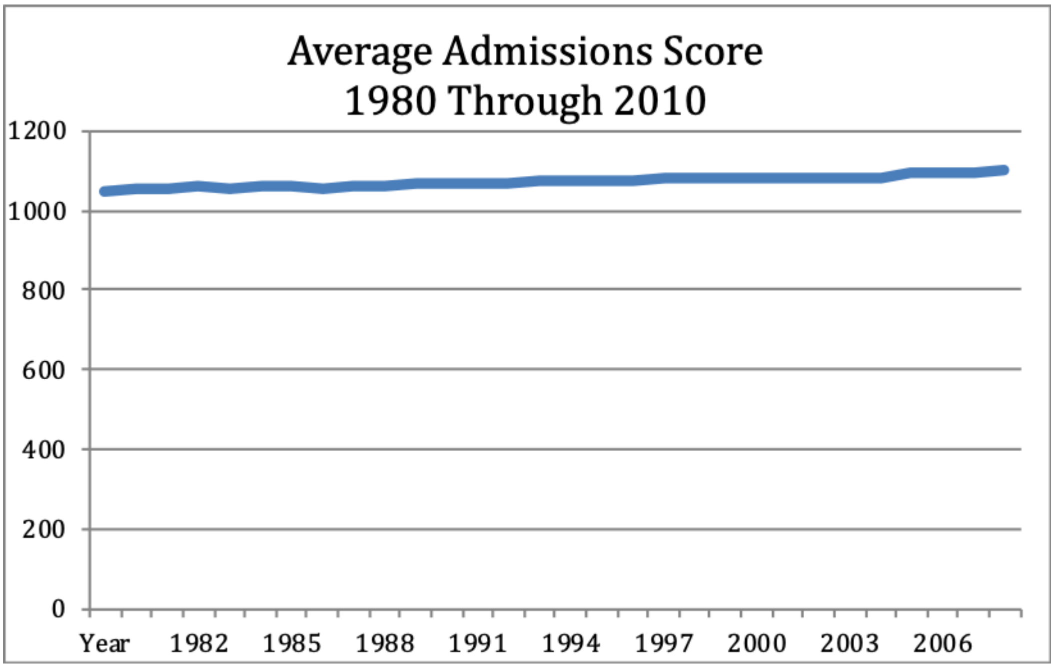

Figure 4.4 Line Chart of Fictitious Admissions Score Data with Y Axis scale from 0 to 1500