TOWARDS A FRAMEWORK FOR INTERROGATING DATA-BASED AUGMENTS PERPETUATED IN SOCIAL MEDIA

Travis Weiland
Appalachian State University, United States
weilandtj@appstate.edu

Social media platforms are becoming a dominant source for communicating and reifying data based arguments in modern society. This is a significant shift from past trends and warrants investigation. For example, in the U.S. many people now look to the president's twitter feed for the most current news and policy decisions. In the era of "fake news" it is becoming crucial for citizen's to be able to critically interrogate data-based arguments perpetuated in social media. Based on an initial analysis of a popularly perpetuated data based argument in this paper I work towards developing a framework for interrogating data based arguments from social media and discuss the implications it has for statistical literacy and the teaching of statistics. I also discuss potential future directions for investigating statistical literacy in social media and some of the methodological issues with using social media as a data source.

INTRODUCTION

A common goal that is discussed in the field of statistics education is that of statistical literacy. A statistically literate citizenry is crucial in today's data drenched societies (Steen, 2001). I too believe this is a crucial goal for us as a field, and more broadly in society, to work toward. That said I also believe that what constitutes statistical literacy is constantly changing, as it needs to be a dynamic construct as the world that we are trying to read and write with statistics is constantly changing. For example, in Gal's (2002) seminal work on adult statistical literacy he stated:

> What is "basic" knowledge cannot be discussed in absolute terms, but depends on the desired level of statistical literacy expected of citizens, on the functional demands of contexts of action (e.g., work, reading a newspaper), and on the characteristics of the larger societal context of living. (p. 9)

There are certainly core principles and practices involved in statistical literacy that draw from the core of what statistics is as a discipline such as making sense of tables and graphs and having an eye on variability (Franklin et al., 2007; Gal, 2002; Watson & Callingham, 2003). However, as Gal (2002) points out it is also important to consider literacy as a " set of cultural practices that people engage in, and hence that it is important to examine the characteristics of the texts that people may have to make sense of, and ask how these characteristics shape people's literacy practices" (p. 9). Taking this point seriously there have been significant shifts in the past decade of the characteristics of texts that people often have to make sense of, which has in turn transformed the literacy practices necessary to make sense of such texts. One of the more prevalent and powerful shifts in the texts that people consume is that of social media.

SOCIAL MEDIA AND DATA-BASED AURGUMENTS

Social media started as a way to communicate and share ideas and in a sense, a way for individuals to share their daily life with others through web-based platforms. Facebook was one of the earliest and most dominant players in social media, but many other platforms have cropped up over time such as Twitter, Pinterest, Snapchat, Tumblr, and YouTube, just to name a few. Each platform offers a slightly different mode of communication in other words each has its own characteristics of texts that people make sense of in their daily lives. These platforms all have features to their texts, which are different from the characteristics of traditional texts. For example, YouTube is a platform for sharing videos, Twitter only allows posts that are 140 characters long (at the time I am writing this they are looking at doubling this limit), and Snapchap allows users to send messages and photos that are self-deleting after viewed or a given time period.

These different mediums of communication facilitate different forms of communication. Social media sites are designed to facilitate people communicating and making connections, but often only the connections and communication that the platform predicts you want made based on

In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10, July, 2018), Kyoto, Japan.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org

your prior actions. This creates echo chambers where people are predominantly only presented with ideas and views similar to their own. That means much of the content that is data based is links to other sites, brief snippets like nine out of ten doctors say …. or data visualizations that fall under Twitter's character limit or could be used in other platforms like Pinterest. A common characteristic of social media though is that it is relatively unfiltered, subject to no specific editorial ethics, yet it used by many individuals as their main source of news and in many technologically developed nations government agencies and policy makers often use social media as a means to communicate official news and policy statements.

Unfortunately, in spite of the prevalence of social media as a mode of communication in an extensive review of the literature, I found very little research investigating social media in statistics education. In fact, in a recent article from the field of information systems the authors found there has been very little qualitative research of social media in general with much of the past research focused on quantitative analysis of frequencies of word use and the connections formed in social networks (McKenna, Myers, & Newman, 2017). Specific to statistics education, Everson and colleagues (Everson, Gundlach, & Miller, 2013) investigated the use of social media in introductory statistics courses at the graduate and undergraduate levels at multiple universities. They used social media namely Facebook and Twitter to facilitate the sharing of statistical studies and communication relevant to the courses. Everson et al. (2013) found that when it is used in certain ways social media could increase the amount of dialogue between students around the course. However, they also pointed out pitfalls to avoid. López-Zafra1 and de Paz-Cobo (2014) similarly found that Twitter could be used to help instructors communicate course material and increase students interest in statistics courses. Gundlach, Maybee, and O'Shea (2015) used a mixable system, which is meant to mimic a social media platform in use, but provides privacy protections for students and does not contain ads like public platforms. Gundlach et al. (2015) had students search for an article, video, or podcast that described statistical ideas and then share them on a mixable social media platform with their classmates who would then have to comment on the posts. These papers focused on using social media to support the instruction of statistics, but they did not investigate the use of statistics in social media, which is the focus of this paper.

Relevant to statistics education is the investigation of how social media platforms facilitate or perhaps constrain the communication of data based arguments. As citizens are faced more and more with the phenomenon of fake news, it is crucial we investigate how people interact with and make sense of data based arguments. This means that we need to rethink what practices statistical literacy should consist of in light of social media. For example, how is statistical literacy in social media settings different from in tradition paper-based media forms? How do we prepare students to make sense of data based arguments in social media? How can social media facilitate and/or constrain the communication of data? These are but a few important questions around statistical literacy and social media and far more than what can be covered in a paper of this length. In this paper, I will focus on presenting an example of prevalent data based argument in the context of the United States around the issues of gun control. I use this example to highlight several themes that have come up in my investigation of statistical literacy in social media contexts that point to some practices important to statistical literacy in such contexts that is different from statistical literacy practices in traditional texts. Since there is such a death of research in this vein, I also discuss future directions and some of the methodological issues involved in the investigation of social media in hopes others will take up such research.

METHODS

Using social media posts as data for research has its own unique challenges. To begin with qualitatively investigating social media posts is a relatively new area of research so there are few established tools or methodologies for conducting such research, though some scholars are beginning to report recommendations (McKenna et al., 2017). One of the largest issues is what are the ethical considerations of such research (McKenna et al., 2017; Townsend & Wallace, 2016). For example, the creators of social media posts are people the researcher rarely has any one-on-one interaction with nor do the creators likely know that they are participating in a study in the case that their posts are used as data for an analysis. Because of the difficulty of gaining informed consent in such situations, it is important to consider what is private and would require the permission of the

author to use, versus what is public and could be analyzed without the permission of the author. For this paper, I chose to focus my attention on Twitter because of its current prevalence in society as a news source. In the case of Twitter it is commonly agreed any post with a hashtag to a large public forum where it is clear the post was intended to be viewed publicly can be considered public (Townsend & Wallace, 2016).

The dynamic and rapidly changing nature of social media make contextualization important for making sense of the data presented. Furthermore, what you see on social media is often filtered based on what the platform determines you might want to see based on data that it has on you such as where you are accessing the site from geographically to what sites you have visited previously or searches you have carried out. To minimize the previous data that Twitter took into account I searched from the main page without being logged into any accounts. That being said I am certain some data of me as an internet user was still likely taken into account by the system when searching. The search I am presenting data from here was carried out in the southeastern United States November 6th-12th, 2017. At this time the 45th president of the US was embroiled in scandal around Russian intervention in the election that he won. Furthermore, specifically relevant to the topic I focused on for this paper a mass shooting had just occurred at a church in Texas where 26 people were killed on November 5th, also just a month prior 59 people had been killed in another mass shooting in Las Vegas. For those not familiar with the issue of gun control in the US it is a very controversial and divisive topic and was a common topic of discussion in the 2016 presidential election of which the elected president was in support of gun rights advocates and lobby groups. This is but a brief contextualization to put into context the specific data I analyzed and present in this paper.

I specifically chose the topic of gun control because of its controversial nature and because it is a common issue of discussion in the US. It is also one where many data based arguments get used, many of which are based on limited or faulty data due to the lack of data around gun deaths in the US. Furthermore, it is an ideological issue, which helps to highlight some of the issues prevalent in the characteristics of texts in social media that are different from more traditional and edited forms of media that I discuss more later in the paper.

I started by searching Twitter using a popular hashtag #guncontrol. I then scanned for tweets that discussed or presented data. An important note here is what I am considering data as in general that could mean many things. Because of my interest in investigating statistical literacy, I am referring to data as quantified data such as counts, frequencies, and measurements. I was also interested in data based arguments in general that could include the presentation of statistics and visualizations of data. A depressing observation I made very quickly was how rare it was to see any data presented in the support of arguments made for or against gun control. To try to narrow in the search to find prevalent data based arguments I then searched "#guncontrol data." From this search I then looked for data based arguments that were commonly used and repeated. Two main arguments come up often one being an article written in the New York Times entitled, "What explains U.S. mass shootings? International comparisons suggest an answer" (Fisher & Keller, 2017) and an opinion piece from the Washington Post entitled, "I used to think gun control was the answer. My research told me otherwise." (Libresco, 2017). For this paper, I chose to present my analysis of the Libresco (2017) article and its perpetuation because of the high frequency I found it to be taken up and repeated by both individuals and other media sources. To analyze the data I used open coding (Charmaz, 2014) looking for themes in the data based arguments that I present in the findings.

FINDINGS

The overall theme of the Libresco (2017) piece was that she had always been pro-gun control, but when she began to delve into the data around gun deaths while a data journalist for FiveThirtyEight she found that most gun control measures would not have prevented the majority of causes of gun deaths, namely suicide and male youth violence often gang related. Though this piece was often presented by pro-gun advocates as data based evidence that gun control is ineffective, there was surprising little actual data presented in the piece especially considering the claims made by the author. So much so that another publication took up this argument and delved into the data more specifically in an attempt to refute Libresco's claims (Lopez, 2017). My focus

here is not to critique the data based arguments made, but to investigate what practices are important for investigating such arguments. Three predominant practices I found from the analysis were the operationalization of constructs, tracing and triangulating sources, and avoiding ideological bias.

Operationalization

A common theme that appeared is operationalization. In other words, what is being counted and how is it being counted? For example, Libresco (2017) analyzed gun deaths, which could be classified as homicides, but also suicides among other classifications. On the other hand, most people who were presenting the Libresco piece were doing so as evidence that gun control would not prevent mass shootings. Mass shootings are a subset of gun deaths, one that Libresco did not spend much time discussing because they make up a small proportion of gun deaths in the US. She did point out that she found the gun bans and buyback programs in England and Australia could not be shown to reduce mass shootings there, but that was in part because of the rarity of such events in those countries to compare.

This also brings up the question of what counts as a mass shooting. After doing some investigation, I found there is no set definition, which hinders the ability to compare counts between sources as they may be counting different things. One of the more commonly used definitions comes from a report from the Congressional Research Service where they state, "for the purposes of this report, "mass shooting" is defined as a multiple homicide incident in which four or more victims are murdered with firearms, within one event, and in one or more locations in close proximity" (Krouse &Richardson, 2015, p2). The operationalization of constructs is a crucial practice in research, but I found it to also be crucial to be able to follow and critique many of the data based arguments I analyzed. I found that people often used the terms gun deaths, shootings, mass shootings, school shooting, and gun homicides interchangeably in the same tweet thread confounding what are different constructs.

Tracing and Triangulating Sources

Another important practice that became apparent during the analysis is tracing sources as they are generally mentioned by providing links to other sites (if at all) and checking multiple sources to triangulate the results and interpretations being presented. Many of the blogs and news articles I came across during the analysis would imbed hyperlinks in the texts to sources for which their claims were coming from. What was not apparent to me was how many people are aware of that or take the time to actually check such sources, which should be investigated in the future. Many Tweets however would not provide much in terms of sources other than a hyperlink to a webpage or article that supported their argument. At times, it took a significant amount of searching to trace back the sources quickly thrown out in Tweets as evidence of a point. I also never came across an instance where multiple sources were cited to support an argument. A number of citations also just went to raw data as if the interpretations of such data were self-evident.

Interestingly, in the case of the Libresco (2017) article many of the Tweets that presented it as evidence provided links to other sites that were reporting on the opinion piece rather than links directly to the piece itself. In other words, many of the sources used were peoples' interpretations of a person's interpretation of a set of data meaning that most people were being presented with interpretations at least one point if not several removed from the actual data itself. This is problematic because the results of the data have then been filtered multiple times by people and it makes it increasingly difficult to follow the sources back to the original source for verification and interrogation. Furthermore, it was a non-trivial task to find the original analysis Libresco published of her analysis of gun death data, which she was only briefly discussed in her opinion piece that was then being used by others as evidence of gun control being ineffective.

Avoiding Ideological Bias

A final important practice I found analyzing the data was that of avoiding ideological bias. For example, the Libresco (2017) article was predominantly cited by those whose views it seemed to support and was used in defense of those views. I did not see a single example of a post showing

conflicting evidence. Instead, conflicting evidence came in the form of replies from individuals with differing views. Now this is not surprising given the character limits of Tweets. However, this characteristic of the text is an important consideration when considering the practices necessary for reading the world in such a context.

The practice of avoiding ideological bias also connects to the disposition of having a critical stance that Gal (2002) discusses. However, Gal's critical stance is mostly focused on critically questioning data based articles of others for potential bias, mis-leading, or one-sided quantitative messages. In the context of the data I analyzed though I saw nothing but a back and forth of one-sided messages. Because of the unedited nature of social media content, most posts presented people's own personal views and data to support those views. In such a setting, it is important for the reader to also consider their own ideological biases as well as the author's and to reflect upon them when reading data based arguments that both support and refute their own view. The notion of having a critical self-reflective stance may be an important dispositional practice to add to statistical literacy in the context of social media. It is very easy in such an environment to simply disregard arguments that you do not agree with and merely agree with arguments that are consistent with your own view without critically examining them. When individuals are presented with data based arguments in the echo chambers of social media it is increasingly important for them to consider their own biases and to find sources and arguments outside of those echo chambers.

DISCUSSION & CONCLUSION

One of my biggest realizations from this initial investigation of social media was that in analyzing the data I was uncovering more questions than answers. This work was initially born out of my frustration with both the Brexit vote followed by the presidential election in the United States. Social media was used in both cases as a powerful form of communication in spreading arguments including data based arguments both factual and fictitious. Furthermore, in the aftermath of both votes there is more and more evidence coming to light of the purposeful use of social media to sow seeds of discontent and division by third parties in efforts to undermine the democratic process. This presents a serious threat to democracy if we do not tackle these issues head on and prepare our citizens to critically interrogate information and claims put forth and perpetuated in social media contexts. A serious consideration then is what does this mean for statistics education in the K-12 levels to prepare student to be critical citizens in the social media communities they may become a part of. Currently social media is almost completely avoided in many classroom settings other than in the case of some teachers experimenting with using it as a way of communicating course information as discussed earlier. What seems to be missing though is discussions around how to make sense of and interrogate data based arguments and information perpetuated in social media.

As I mentioned earlier, each social media platform has its own characteristics, which means that there may be different statistical literacy practices necessary for each platform. This presents a daunting challenge for a tiny field such as statistics education. A more fruitful approach however might be to look at the intersection of our field and others to consider how broader research on media literacy or information technology literacy might relate to statistical literacy and making sense of data based arguments. We might also be able to contribute to those fields, as one of the issues that came up in the findings was the general lack statistical literacy presented in the Tweets analyzed and more generally the dearth of data used to support claims and arguments in social media.

Another comment for future work I have is on investigating the practice of avoiding ideological bias, which is an area that I would argue, is under researched. More specifically, how do people's subjectivities influence their creation and interpretation of data based arguments. It has been talked about for some time how people's bias can influence there doing of statistics in the sense that when doing a statistical study we may unknowingly or knowingly alter aspects to favor our hypothesizes, this is why we have double blind studies, randomization, and peer review among other safe guards. However, what is less understood or investigated is how people's identities, ideologies, and other subjectivities influence their sense-making, uptake, and distribution of data based arguments.

There are clear limitations to this study as it only represents an initial foray into making sense of statistical literacy in the context of social media. The biggest limitation would be the contextualization of the data I discuss the analysis of here. Namely, I was analyzing posts around the issue of gun control at a point in time in the context of the United States. Much more work needs to be done to see if the findings in this context translate to others. That being said given the current dearth of research on considering statistical literacy in the context of social media it represents a starting point and part of what I have tried to do with the discussion is to present some major questions and considerations that I believe it is important for the field to take up in the future. Statistical literacy is crucial in today's modern societies and social media is becoming one of the predominant platforms for communicating information and mis-information making it a crucial context for the field of statistics education to investigate further related to how to interrogate such texts and how to prepare students to do so.

REFERENCES

Charmaz, K. (2014). *Constructing grounded theory* (2nd ed.). Thousand Oaks, CA: Sage Publications Ltd.

Everson, M., Gundlach, E., & Miller, J. (2013). Social media and the introductory statistics course. *Computers in Human Behavior*, *29*(5), A69–A81. https://doi.org/10.1016/j.chb.2012.12.033

Fisher, M., & Keller, J. (2017, November 7). What explains U.S. mass shootings? International comparisons suggest an answer. *The New York Times*. Retrieved from https://www.nytimes.com/2017/11/07/world/americas/mass-shootings-us-international.html

Franklin, C., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education (GAISE) report: a pre-K--12 curriculum framework*. Alexandria, VA: American Statistical Association.

Gal, I. (2002). Adults' statistical literacy: Meaning, components, responsibilities. *International Statistical Review*, *70*(1), 1–25.

Gundlach, E., Maybee, C., & O'Shea, K. (2015). Statistical literacy social media project for the masses. *The Journal of Faculty Development*, *29*(2), 71–80.

Libresco, L. (2017, October 3). I used to think gun control was the answer. My research told me otherwise. *Washington Post*. Retrieved from https://www.washingtonpost.com/opinions/i-used-to-think-gun-control-was-the-answer-my-research-told-me-otherwise/2017/10/03/d33edca6-a851-11e7-92d1-58c702d2d975_story.html?utm_term=.955c35e8c8c2

Lopez, G. (2017, October 4). The research is clear: gun control saves lives. *Vox*. Retrieved from https://www.vox.com/policy-and-politics/2017/10/4/16418754/gun-control-washington-post

López-Zafra, J. M., & de Paz-Cobo, S. (2014). Twitter as a learning tool in higher education. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education.* Flagstaff, Arizona, USA: Voorburg, The Netherlands: International Statistical Institute.

McKenna, B., Myers, M. D., & Newman, M. (2017). Social media in qualitative research: Challenges and recommendations. *Information and Organization*, *27*(2), 87–99. https://doi.org/10.1016/j.infoandorg.2017.03.001

Steen, L. (Ed.). (2001). *Mathematics and democracy: The case for quantitative literacy*. United States: The National Council on Education and the Disciplines.

Townsend, L., & Wallace, C. (2016). *Social media research: A guide to ethics*. Aberdeen, England: University of Aberdeen. Retrieved from http://www.dotrural.ac.uk/socialmediaresearchethics.pdf

Watson, J., & Callingham, R. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal*, *2*(2), 3–46.