**Megan Ozeran – Ethics of Data Visualization**

**1. Project background**

The ethics of data visualization have not yet been formally defined, and it is crucial that we do so.

Current technologies make data collection, data analysis, and data visualization easier than ever before, even as the amount and variety of data continue to expand. Computer-mediated data analysis is no longer simply in the domain of computer scientists, as more tools and software are made available to researchers with a variety of technical expertise.

This leads to two intersecting issues. The first is that newer technological tools and methods continue to spark ethical quandaries, and the fields of computer ethics, information ethics, and related applied ethics are still in flux. The second issue is that increasingly more people are able to use these new tools and methods, but they are not always taught to consider related ethical questions.

To help inform this project, I have read extensively: articles, books, blog posts, and essays from both researchers and professionals. In addition, my colleague Dr. John Gallagher and I surveyed data visualization experts nationally, and Illinois faculty who do or teach data visualization, in order to discover what values are already practiced and taught. By synthesizing the numerous voices and ideas expressed in the literature and by individuals, I have begun to tease out some principles that we can include in library visualization instruction. The initial list follows, along with guiding questions or explanation and a few examples.

**2. Ethical principles for data visualization**

**Consider the selection, emphasis and framing**What data have you chosen to use? What are you focusing on? What are you leaving out? Why? How can you convey these decisions to your audience? This is also a place to rethink binaries, or consider categorization in general. How are things categorized in the data? How do these categories inherently misrepresent the world?

For instance, the map on the following page is one of many that cartographer Denis Wood created for a specific neighborhood in North Carolina. Here he chose to map power lines (“squirrel highways”), which allow squirrels to traverse the neighborhood, as a way of demonstrating the subjectivity of maps. Street maps for humans are not objective, they’re just subjective in a way that we’re used to seeing. In the data and the map Wood created, he made very specific decisions about *selection* and *emphasis* that pointedly leave out the human experience of the neighborhood. The *frame* for all the maps in this collection is, of course, the boundaries of the neighborhood – which requires making decisions around what constitutes those boundaries.



“Squirrel Highways”

by Denis Wood

*Everything Sings: Maps for a Narrative Atlas*. Los Angeles, CA: Siglio Press, 2010.

**Consider the aesthetics**Nothing is neutral. Even attempts toward minimalism, simplicity, and/or “clarity” are not neutral – they promote a specific way of seeing and thinking about the data. Aesthetics influence the tone of the visualization, how its content is perceived, and what people feel about it.

**Be transparently truthful**Be as truthful as you can, and also explain to your audience how this truth was obtained, both in the data analysis and in the visualizing of the data. Don’t mislead with your visuals. Be careful with choices that skew, such as excluding data you deem as outliers, or making data transformations or abstractions that lose detail. (Accuracy is one of the most commonly identified values in data visualization – perhaps the only one everyone already agrees upon. But accuracy is a slippery thing, because “all models are wrong.” [https://en.wikipedia.org/wiki/All\_models\_are\_wrong])

For instance, the map below attempts to represent the full complexity of white supremacy mob violence in the United States. The specific view copied below is a view upon clicking on an individual lynching record in Richmond County, Georgia. Notice at the bottom of the page is a link titled, “Should I believe this…?”, which leads to extensive details about the data and how it is all represented on the map. It also points out important caveats for how the map can or cannot be used to make certain conclusions.



“Map of White Supremacy Mob Violence” (detail), by auut studio, 2016, on Monroe Work Today
http://www.monroeworktoday.org/explore/

**Make labor visible**Who were all the people behind the data, its collection, its analysis, and so forth? How did their labor make this visualization possible? How can you show this labor in you visualization?

**Consider impact on human welfare**Who will be using the visualization? To what end? Will decisions be made that can affect human welfare, whether mental, physical, financial, etc.? Will sharing this visualization violate anyone’s privacy, or right to self-determination? How might the visualization be used or misused for political purposes?

It may be useful to refer again to cartographer Denis Wood. In the book *Everything Sings*, he wrote that in the 1960s, “Mapmaking was understood as a trade or as a fee-for-service profession: the neutral, unbiased, value-free provision of maps for employers or clients who wanted to bomb the land, mine it, drill it for oil, run roads across it, plant suburban divisions on it, promote it as a tourist destination, or buy and sell it” (p. 14). He also pointed out that university-trained mapmakers were helping the U.S. military bomb Vietnam. We need to seriously think about how our visualizations will be used.

**Promote empathy**This can be considered in at least two ways. First, promote empathy for the people whom the data represents. For example, if you are representing death, a minimalist, “objective” visualization may not be the best way to respect those people. Second, promote empathy for your audience. Make your visualization accessible, usable, and understandable. Provide ways for them to learn more based on questions that may arise from viewing your visualization.

For instance, in the map above about white supremacy mob violence, providing the name of each victim (and pointing out when they are unnamed) reminds the audience that these are all human lives behind the data.

**3. Future Work**

There are two components to this project: defining ethics of data visualization, and teaching those ethics. Both of these need to be improved upon. The initial ethical principles I outlined above will need to be reviewed by the larger community, as they will need to reflect what the community at large truly believes the ethics of data visualization are (or should be). As these ethics are better defined, instruction will need to find new ways to incorporate the principles that are most important for students to learn. It will require a cyclical process of teaching, assessment, and improvement of instructional modules to determine the best way to teach the ethics of data visualization.