SEEING DATA, FEELING NUMBERS: HOW PEOPLE INTERACT WITH DATA VISUALISATIONS

Helen Kennedy
University of Sheffield, UK

Rosemary Lucy Hill
University of Leeds, UK

Abstract

Whilst there has been considerable attention to the development of strategies to maximise effectiveness in data visualization research, only some of this has engaged actual users in the research process. Practitioners and researchers alike agree that historically, the field has lacked a user-centred knowledge base. Research which engages users often provides little information about who they are and how this might affect their engagement with visualizations. Research into user engagement with visualizations almost never considers the socio-cultural factors that affect *how* people engage and some research even attempts to bypass or 'overcome' such factors. Drawing on qualitative, empirical research with users of visualizations on *Seeing Data: are good big data visualisations possible?* (http://seeingdata.org), in this paper we identify six factors that affect engagement, which we define as socio-cultural: subject matter; source/media location; beliefs and opinions; time; emotions; and confidence and skills. We argue that to fully understand engagements with visualizations, it is important to acknowledge these factors. We conclude by reflecting on what these findings mean for how 'effectiveness' could and should be defined in relation to data visualizations. Our research suggests that the study of engagement with visualizations, to date primarily carried out within fields such as HCI or computing, can benefit from adopting qualitative approaches developed within media studies, specifically in relation to media audience research.

Context & literature

As data become more and more ubiquitous and the main way that 'ordinary' people get access to them is through visualizations, it is more important than ever to attend to questions of how people engage with data visualizations. In literature concerned with measuring the effectiveness of visualizations, there is widespread discussion of how to

optimise them, but, as some commentators have argued, more empirical research with users of visualizations is needed to ground assertions about what works in a visualization (for example Kirk 2009 and Lin et al 2009). Because some of the most influential work in the field, such as the writings of Edward Tufte (1983, 1991, 2006), draw on design instincts, a number of calls have been made for the field to pay more attention to the ordinary users of visualizations.

Since those calls were made, some research on the effectiveness of data visualizations has involved users, but we argue in this paper that this is limited in a number of ways.

1. Research in this field focuses on measures such as accuracy, consistency and speed of comprehension, as if socio-cultural factors did not play a part. For example Huang et al (2009) focused on cognitive load (that is, the amount of interpretative work the brain has to do) in their study of visualization efficiency. Borkin et al (2013) argue that being able to identify and quantify what makes a visualization memorable is important, in order to be able to design effective visualizations. Chin et al (2009) tested dynamic data visualizations to find out which visualization methods were most helpful in the quick accomplishment of tasks. Haroz and Whitney (2012) tested how layout, visual features like colour and motion and the inclusion of specific visualization elements affected users' task performance. Some research even attempts to bypass or 'overcome' social or cultural influences on engagements with visualizations, opting for techniques like electroencephalography (EEG) which go straight to the brain, an alleged biological entity unaffected by society or culture (Anderson et al 2011).

2. The authors of these papers provide little information about the participants in their studies. Huang et al tell us that their participants were IT students, while Anderson et al explain that their user group was made up of 10 men and 7 women. Borkin et al tell us more about their participants, but none of these authors consider the ways in which demographic factors might impact on visualization engagement. There is no information about who the users are in the tests carried out by Einsfeld et al (2009) other than that they were non-experts. Haroz and Whiteney tell us very little about their five participants, only that three were female and they were either graduate or postgraduate students in psychology or computer science or trained university staff. Given the small number of participants, a more precise description of each one would have been possible and desirable. Chin et al also tell us little about their users. The work of Dadzie et al (2009) is an exception. They gathered and report on demographic data about participants, used qualitative methods like observation and interviews and aimed to reproduce realistic conditions of engagements in order to elicit 'more valid feedback'. (Ziemkiewicz and Kosara (2009) and Kosara et al (2003) are also exceptions).

We argue that to fully understand how people engage with visualizations and how visualizations can be effective, it is important to acknowledge these factors. Our research suggests that the study of engagement with visualizations, to date primarily carried out within fields such as HCI or computing, can benefit from adopting qualitative approaches developed within media studies, specifically in relation to media audience research, that are attentive to social and cultural influences on engagement with

designed artefacts like visualizations. This includes the work of Stuart Hall (1973) and others in highlighting the importance of two elements in engagements with media products: the *encoding* of meaning as the media product is produced, and the *decoding* of meaning as it is used or consumed. Another important contribution from media studies (and other fields) is the recognition of the role of emotions in engagements with media artefacts.

Methods

This paper is based on qualitative, empirical research with users of visualizations on Seeing Data. We define the users with whom we worked as 'non-expert' because they were not expert in data visualization, although some had expertise in related fields, such as data, visual design, or the subject matter of some of the visualizations that we examined. The project aimed to explore factors in visualization consumption and production processes that affect engagement, and through this identify how effectiveness could and should be defined in this context. The project addressed these questions through a range of methods, including focus group research, interviews and diary-keeping.

We carried out nine, two-hour focus groups with a total of 46 participants, in four geographical locations which, given our focus on migration as a case study, we characterise in the following ways:
- Rural, high migration
- Rural, low migration
- Urban, high migration
- Urban, low migration.

We aimed to recruit participants who might be assumed to be interested in data, the visual, or migration, and so 'already-engaged' in one of the issues at the heart of the project, and others about whom we could not make these assumptions.

27 participants were female and 16 were male; ages ranged from 11 to 70, with the 30-39 age range best represented (18 participants). Employment sectors were extremely diverse, including fields like hairdressing and cleaning, local government, agricultural work, teaching, media, retail and information services. Most participants had qualifications of some kind; 19 had completed tertiary education and 11 had higher degrees. As the study took place in the UK, most participants (n=30) self-reported as British, and other nationalities included German, Indian, Lithuanian, Pakistani, Polish and Thai. Focus group participants kept diaries of their encounters with visualizations for one week before the focus group; seven participants kept longer diaries, for one month after the focus groups. The visualizations on which we based discussion in the focus groups can be seen here: http://seeingdata.org/understanding-data-visualisations/rate-these-visualisations/.

Socio-cultural factors

We argue that socio-cultural factors matter when it comes to engaging with data visualizations, and the field of visualization engagement research needs to pay attention to them. We identify six factors that affect engagement, which we define as socio-cultural:

1. **Subject matter.** Visualisations do not exist in isolation from the subject matter that they represent. When subject matter speaks to people's interest, people are engaged. As we found in our efforts to set up focus groups, it is hard to engage people around the topic of visualisations alone.

2. **Source and media location.** Participants expressed concern that the media aim to confuse, so visualisations encountered within the context of certain media were therefore suspect too. However, when media are sought out and already-trusted, the visualisations within them acquire the trustworthiness of those media.

3. **Beliefs and opinions.** As stated, participants trust visualisations that appear in trusted publications, which fit in with their worldviews. Some participants liked visualisations that confirmed their beliefs and opinions, but others liked it when data called into question existing beliefs.

4. **Time**: Engaging with visualisations can be seen as work, or as laborious, by people for whom doing so does not come easily: time is crucial in determining whether people are willing to do this 'work'. When time was made and set aside to engage with visualisations, as in our focus groups, participants found the experience enjoyable.

5. **Confidence and skills**: Some people lacked confidence in their statistical education because they did not know how to read particular chart types. Many participants doubted their 'graphicacy' skills, the combined statistical and visual capacities required to interpret graphics. Others did not have the language skills to feel confident engaging with visualisations. Some felt that critical thinking skills were also needed, such as asking what has been left out, or what point of view is being prioritised); some of the designers we interviewed highlighted this point too.

6. **Emotions**: We argue that all of these factors contributed to participants' particular emotional responses to visualisations. Strong emotions arise when considering a visualisation, which relate to the subject matter, source and location, beliefs and opinions, available time, confidence in skills and responses to the visual aspects of a visualisation. These contribute to first impressions of visualisations and also emerge from longer-term engagement with them. When we interviewed focus group participants a few weeks after their participation, none of them could remember any specific data or figures from the visualisations they looked at there. They could, however, remember the overall impressions that the visualisation made, and the way they felt.

Definitions of effectiveness

What do these findings mean for how 'effectiveness' could and should be defined in relation to data visualizations? We suggest that technical measures like memorability, speed, accuracy of recall or consistency of comprehension are not adequate for determining what users experience as an 'effective' visualization, as they fail to consider the factors that we discuss here as dimensions of effectiveness. Although challenging and not necessarily easy to implement, we propose moving away from assumptions about the existence of 'average users' or the possibility of 'universal design', which are implicit in existing definitions of effectiveness, and moving towards an acknowledgement of socio-cultural differences amongst users and the contexts in which they engage with visualizations, as this will lead to fuller understanding of how we might think about their effectiveness.

References

Anderson EW, Potter KC, Matzen LE, Shepherd JF, Preston GA and Silva CT. A user study of visualization effectiveness using EEG and cognitive load. *Computer Graphics Forum*. Wiley Online Library, 2011, p. 791-800.

Borkin MA, Vo AA, Bylinskii Z, et al. What makes a visualization memorable? *Visualization and Computer Graphics, IEEE Transactions on*. 2013; 19: 2306-15.

Chin G, Singhal M, Nakamura G, Gurumoorthi V and Freeman-Cadoret N. Visual analysis of dynamic data streams. *Information Visualization*. 2009; 8: 212-29.

Dadzie A-S, Lanfranchi V and Petrelli D. Seeing is believing: Linking data with knowledge. *Information Visualization*. 2009; 8: 197-211.

Einsfeld K, Ebert A, Kerren A and Deller M. Knowledge generation through human-centered information visualization. *Information Visualization*. 2009; 8: 180-96.

Hall S. *Encoding and decoding in the television discourse Birmingham Centre for Contemporary Cultural Studies*. [S.l.]: The University of Birmingham, 1973.

Haroz S and Whitney D. How capacity limits of attention influence information visualization effectiveness. *Visualization and Computer Graphics, IEEE Transactions on*. 2012; 18: 2402-10.

Huang W, Eades P and Hong S-H. Measuring effectiveness of graph visualizations: A cognitive load perspective. *Information Visualization*. 2009; 8: 139-52.

Kirk A. Paradox of Form vs Function in Graph Design. University of Leeds, 2009.

Kosara R, Healey CG, Interrante V, Laidlaw DH and Ware C. User studies: Why, how, and when? *IEEE Computer Graphics and Applications*. 2003: 20-5.

Lin X, Kerren A and Zhang J. Challenges in human-centered information visualization: Introduction to the special issue. *Information Visualization*. 2009; 8: 137-8.

Tufte ER and Graves-Morris P. *The visual display of quantitative information*. Graphics press Cheshire, CT, 1983.

Tufte ER. Envisioning information. *Optometry & Vision Science*. 1991; 68: 322-4.

Tufte ER. *Beautiful evidence*. Graphics Press Cheshire, CT, 2006.

Ziemkiewicz C and Kosara R. Preconceptions and individual differences in understanding visual metaphors. *Computer Graphics Forum*. Wiley Online Library, 2009, p. 911-8.