Communicating Qualitative Uncertainty in Data Visualisation

Georgia Panagiotidou[1], Andrew Vande Moere[2]

Keywords: qualitative uncertainty, critical data visualisation, digital humanities

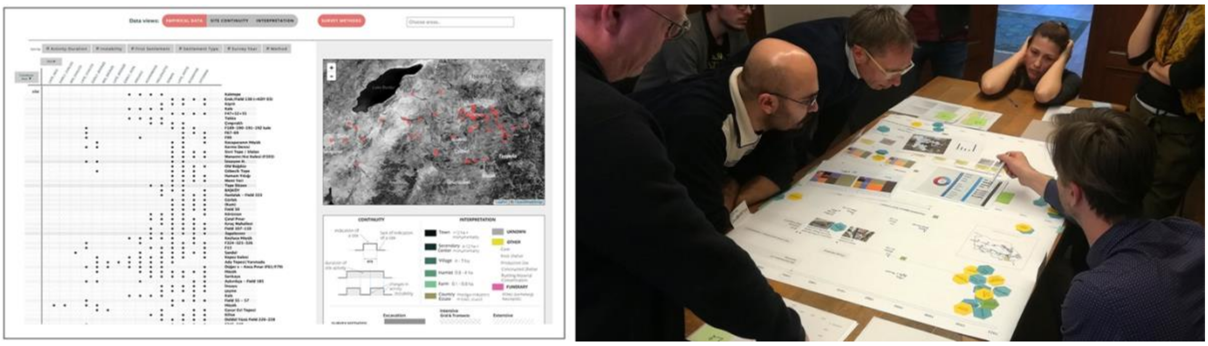

Figure 1 Snapshots from our two case studies. On the left, the interface for analysing the archaeological settlement dataset. The qualitative uncertainty regarding the data interpretation are encoded as different data views. On the right, the participatory data activity that helped reveal how human geographers approached the uncertainty of their house-hold survey dataset.

In data visualisation design and research, uncertainty is often approached in its quantifiable dimensions such as modelling errors or forecasted probabilities. This however is not the only type of uncertainty that can cloud decision-making with data. Implicit and underlying data issues such as the circumstances of collection, its storage or even assumptions and biases of its authors can jeopardize the validity of its analysis. This type of uncertainty is referred to as qualitative uncertainty [1]. Researchers using humanistic datasets are well aware of such uncertainty as they often use historical records that are partial, incomplete and the product of their historical circumstances [2]. But even beyond the humanities, critical data studies have been increasingly drawing awareness to the problems of treating data as objective, complete and devoid of human interpretation. They call for data to be examined in their situated local origins [3], to be acknowledged as products of the socio-technological settings that generate them [4] and treated as products and agents of power imbalance [5]. Once we acknowledge data as inherently incomplete and partial, we are in fact acknowledging a degree of uncertainty of working and reasoning with them. Nevertheless, being non-quantifiable, such uncertainty is rarely explicitly designed for in visualisations. We argue data visualisation research and practice should be expanded to account for non-quantifiable uncertaint*y.* Accordingly, we present here two case studies within the digital humanities in which we tried to understand and then integrate qualitative uncertainties in our visualisation designs.

Our first case discusses the visualisation of an archaeological dataset that documents sites of human settlements since Neolithic times. This spatiotemporal dataset has been progressively assembled over the period of 30 years with evolving technologies, methodologies, archaeological theory and research teams, making some of its data points only broadly comparable. As a result, the archaeologists approached our original visualisation prototypes with scepticism for their suppression of these issues. Even if the dataset was presented as complete and objective, the archaeologists were nevertheless aware of its caveats and accounted for their uncertainty in tacit ways. We engaged in a custom participatory activity and spent considerable time collaborating with archaeologists to deconstruct their scepticism in the visualisation designs and better understand their data issues. Our final design accounted for the uncertainty by depicting multiple views of the same data differing on their degree of interpretation and assumption (from the directly sampled values to the modelled ones). In this

[1] georgia.panagiotidou@kuleuven.be, KU Leuven, Belgium
[2] andrew.vandemoere@kuleuven.be, KU Leuven, Belgium

way, we used the interaction flow among these views to progressively lessen the uncertainty and help archaeologists trust their conclusions.

The second case discusses an attempt to integrate the datasets of three epistemologically distant researchers from archaeology, ecology and human geography into a single visualisation. Specifically, we tried to synthesise the previously mentioned archaeological dataset with modelled historical pollen data and land-use data collected from a contemporary door-to-door house-survey in the same region. Through a series of workshops we called data sessions, we found that qualitative uncertainties were present in all three datasets. The historical pollen dataset for example was synthesized of multiple years of fieldwork from authors using different base assumptions while the house-survey included self-reported dimensions such as hand-drawn mental maps of the residents. These uncertainties, which are epistemic in nature (i.e. originating from limited knowledge or ignorance regarding facts, numbers and science [6]), were handled differently based on the researcher's epistemological backgrounds (i.e. their disciplinary approach to knowledge creation). For example, ecologists argued in terms of alternative, hypothetical cases of the uncertain data while the human geographers minimised the uncertainty by bringing additional datasets and points of view to the topic. Our proposed visualisation design therefore used custom interfaces for each discipline to view the synthesized data, each having their own dedicated approach for handling uncertainty.

We draw from theses case studies and offer the following reflections.

First, we propose that qualitative dimensions of uncertainty can be designed for in the visualisations' interaction design as well as its visual encoding. Besides our case study, other visualisation designs examples also come close to this approach for example by communicating forecasting uncertainty through the use of animation [7], or by developing elaborate interaction patterns to navigate through the different degrees of detail of cultural collections [8].

Second, we argue that uncertainty visualisation should be understood within its situated, socio-technological context rather than only the data. One of the goals of communicating uncertainty in visualisation is to provide a feeling, experience or sensation so it can inform users' actions [9]. As we observed in our second case study, data uncertainty is not viewed or valued the same throughout researchers or projects. Even if cognitive parameters remain constant in how we 'read' the visual encodings of uncertainty, how we trust and make decisions based on that same data is not. In order to design for such data 'uncertainty experiences' therefore, we should consider them as a core part of the design study pipeline and accordingly approach them in a human-centred way.

Finally, we suggest the further development of inclusive participatory visualisation activities specifically catered for understanding uncertainty. These can be based on data visualisation activities [10] to help understand the real contexts of data use in structured ways and can be informed by critical data approaches (such as data literacy tactics [11], [12]) to reveal the more tacit aspects of qualitative uncertainty. In our case studies for instance, we deployed custom group visualisation show-and-tell activities (the data sessions), we used prototypes to unpack the sensemaking and argumentation of our collaborators and guided structured sessions documenting the data trail from collection to publication.

Given that visualisations tend to emit an 'aura of objectivity' [13], data issues are often masked thus jeopardizing the quality of the conclusions. By communicating qualitative uncertainty, visualisations can become more 'uncertain' allowing their readers to question, scrutinise and experience the underlying data and assumptions. We contribute by proposing how we can reveal uncertainty during the design process as well as how we can encode it visually, interactively and adaptively.

[1] L. M. K. Padilla, M. Powell, M. Kay, and J. Hullman, "Uncertain About Uncertainty: How Qualitative Expressions of Forecaster Confidence Impact Decision-Making With Uncertainty Visualizations," *Frontiers in Psychology*, vol. 11, no. January, 2021.

[2] F. Windhager, S. Salisu, and E. Mayr, “Exhibiting Uncertainty: Visualizing Data Quality Indicators for Cultural Collections,” *Informatics*, vol. 6, no. 3, 2019.
[3] Y. Loukissas, *All Data are Local: Thinking Critically in a Data-Driven Society*. MIT Press, 2019.
[4] R. Kitchin, *Data Lives: How Data Are Made and Shape Our World.* Policy Press, 2021.
[5] C. D’Ignazio and L. F. Klein, “Introduction: Why Data Science Needs Feminism,” in *Data Feminism*, MIT Press, 2020.
[6] A. M. Van Der Bles *et al.*, *Communicating Uncertainty about Facts, Numbers and Science*, vol. 6, no. 5. 2019.
[7] J. Hullman, P. Resnick, and E. Adar, “Hypothetical Outcome Plots Outperform Error Bars and Violin Plots for Inferences about Reliability of Variable Ordering,” *PLoS ONE*, vol. 10, no. 11, pp. 1–25, 2015.
[8] K. Glinka, C. Pietsch, M. Dörk, and Marian Dörk, “Past Visions and Reconciling Views: Visualizing Time, Texture and Themes in Cultural Collections,” *DHQ: Digital Humanities Quarterly*, vol. 11, no. 2, pp. 1–19, 2017.
[9] L. Padilla, M. Kay, and J. Hullman, “Uncertainty Visualization,” in *Handbook of Computational Statistics and Data Science*, 2020.
[10] S. Huron, B. Bach, U. Hinrichs, M. Keck, and J. Roberts, “IEEE VIS Workshop on Data Vis Activities to Facilitate Learning, Reflecting, Discussing, and Designing,” in *IEEE VIS Workshop on Data Vis Activities to Facilitate Learning, Reflecting, Discussing, and Designing.*, 2020.
[11] C. D’Ignazio, “Creative data literacy: Bridging the Gap Between the Data-haves and Data-have Nots,” *Information Design Journal*, vol. 23, no. 1, pp. 6–18, 2017.
[12] S. Braun, “Critically Engaging with Data Visualization through an Information Literacy Framework,” *DHQ: Digital Humanities Quarterly*, vol. 12, no. 4, 2018.
[13] H. Kennedy, R. L. Hill, G. Aiello, and W. Allen, “The work that visualisation conventions do,” *Information Communication and Society*, vol. 19, no. 6, pp. 715–735, 2016.