Plotting Apples, Oranges, and Distributions: A Taxonomy of the Information in Uncertainty Visualisations

Harriet Mason* Monash University
Dianne Cook† Monash University
Sarah Goodwin‡ Monash University
Emi Tanaka§ Australian National University
Ursula Laa¶ University of Natural Resources and Life Sciences

ABSTRACT

Uncertainty visualisation is needed to communicate the risk associated with estimates and facilitate decision-making. The best visual encoding of uncertainty information relies on accurate estimation and appropriate depiction. While there is a reasonable amount of research into estimating uncertainty, effective visual representations of uncertainty are not as well researched. We believe this is primarily driven by a lack of consensus on how to assess the information depicted in a visualisation. We present a work-in-progress taxonomy that can be used to evaluate the information in a graphic.

Index Terms: Human-centered computing—Visualization—Visualization techniques—Visualization theory, concepts and paradigms; Human-centered computing—Visualization—Visualization design and evaluation methods

1 THE IMPORTANCE OF UNCERTAINTY VISUALISATION

The term "uncertainty" lacks a commonly accepted definition in the literature. Lipshitz and Strauss [13] even commented that "there are almost as many definitions of uncertainty as there are treatments of the subject". The most encompassing definition comes from Walker et al. [18] who define uncertainty as "any deviation from the unachievable ideal of completely deterministic knowledge of the relevant system". This is the definition we will adopt for this paper.

In order to make an accurate decision using an estimate of a numerical quantity, a thorough understanding of the uncertainty around that estimate is needed. There is a large amount of research suggesting that presenting uncertainty information improves decision making and increases trust in predictions, both experimentally [4, 9, 11, 17] and in practice [1].

The most effective expression of uncertainty information is through visualisation, which provides a more complete picture of risks than numerical summaries alone. Something as simple as sketching a distribution before recalling statistics or making predictions can greatly increase the accuracy of those measures [5, 7].

Unfortunately uncertainty information is often complicated and difficult to express. This means uncertainty is not conveyed as often as it should be. A survey found that only a quarter of visualisation authors included uncertainty in 50% or more of their visualisations [6]. Two of the most common explanations authors gave for omitting uncertainty were an inability to calculate it and a fear of overwhelming their audience [6]. Therefore, a set of guidelines that make it easier for authors to clearly visualise the relevant uncertainty could improve their use.

*e-mail: harriet.mason1@monash.edu
†e-mail: di.cook@monash.edu
‡e-mail: sarah.goodwin@monash.edu
§e-mail: emi.tanaka@anu.edu.au
¶e-mail: ursula.laa@boku.ac.at

2 UNCERTAINTY VISUALISATION EXPERIMENTS

A set of guidelines that highlight the information depicted in an uncertainty visualisation and the clarity with which that information is expressed would also benefit the authors of visualisation experiments.

The purpose of data visualisation is insight, but due to time limits or other constraints, most visualisation studies use multiple benchmark tests as a substitute for measuring the complicated phenomena of insight [15]. These benchmark tests commonly compare plots based on the accuracy with which users can report statistics [8, 15]. In order to compare the visual features of different plots they need to contain the same information [3]. Additionally visualisations should aim to show enough information to solve a task while avoiding irrelevant distracting information [12]. While including additional features can increase the accuracy of some conclusions, it can also bias or discount others [10].

If we are trying to test the benefits of using one plot over another, and there is a differences in the information conveyed by the two plots, then we cannot be sure if the one plot outperformed the other due to a better visual encoding of information or because of a difference in information itself. However, there is currently no reliable or standard approach to assess the information contained in a graphic. Therefore there are many experiments where one plot contains more relevant information than the other, and the conclusion of the paper is warped. Since there are numerous ways in which uncertainty information can be quantified and expressed, experiments comparing uncertainty visualisations suffer disproportionately from this issue. A taxonomy of the information contained in an uncertainty visualisation could alleviate this problem.

3 A TAXONOMY OF VISUAL UNCERTAINTY INFORMATION

3.1 Other Taxonomies

This is not the first work that tries to establish a taxonomy of visual uncertainty. Most taxonomies, such as the one developed by Walker et al. [18] focus on uncertainty rather than visual expressions of it. Potter, Rosen, and Johnson [16] developed a taxonomy that organised uncertainty visualisations based on the dimensionality of the data and the dimensionality of the uncertainty. This work focused on a simple organisation of the uncertainty visualisation methods that existed at the time. Our work differentiates itself by specifically considering the decision making goal that often accompanies an uncertainty visualisation.

3.2 Our Taxonomy

Our taxonomy was developed by reviewing the current uncertainty visualisation literature and untangling what features seem to impact the accuracy with which information can be extracted from a plot. The three elements of the information taxonomy are distribution, expression, and hierarchy & heuristics. The taxonomy is visualised in Figure 1.

Distribution is the most important aspect of the taxonomy as it highlights the importance of showing the correct variable or relationship. Very often, a single marginal or conditional distribution

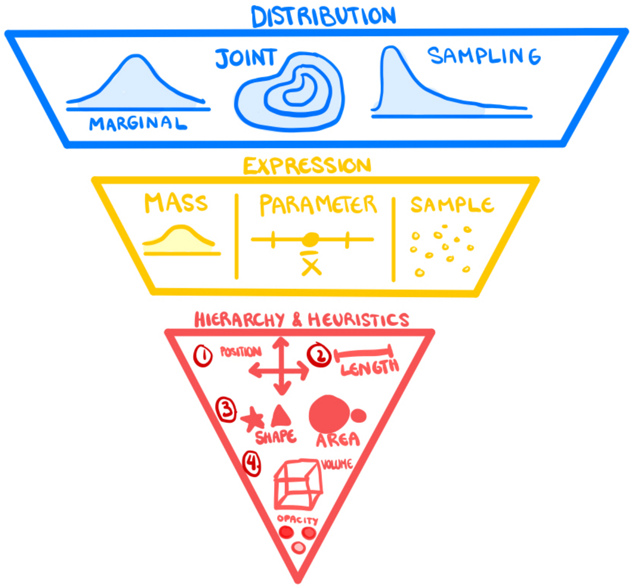

Figure 1: A visualization of the three layers that make up our taxonomy of visual information. Examples of the types of considerations that should be made at each stage of the taxonomy are provided in each section.

is expected to answer a wealth of questions related to uncertainty, even when that distribution is statistically irrelevant to the question. For example, error bars are often considered poor practice by researchers due to being unable to communicate uncertainty that it does not express [2].

Expression highlights which feature of the aforementioned distribution is visualised. While there are a large number of distributions that can be visualised, there only seems to be three ways to express a distribution. You can express a distribution using it's mass, specific parameters, or a real or simulated a sample. Different expressions of a distribution seem to be better at answering different questions. For example, a sample is most appropriate to answer questions about the frequency of an outcome.

Hierarchy & Heuristics identifies how easily a user can extract the information depicted in a graphic. Hierarchy identifies how efficiently information can be extracted using elementary perceptual tasks [3]. Heuristics leverage the existing mental connections we have between visual features and variables. For example, uncertainty can be intuitively mapped to fuzziness where more fuzzy means more uncertain [14].

4 Conclusion

This work first highlighted the importance of visualising uncertainty for decision making by disputing several misconceptions that surround the practice. We then discussed the information gap in the visualisation evaluation literature that leads researchers to frequently compare two plots that differ on several metrics. Finally we presented a taxonomy for the information contained in an uncertainty visualisation.

References

[1] J. Al-Kassab, Z. M. Ouertani, G. Schiuma, and A. Neely. Information visualization to support management decisions. *International Journal of Information Technology Decision Making*, 13(2), Mar. 2014. doi: 10.1142/S0219622014500497

[2] S. Bella, F. Fidler, J. Williams, and G. Cumming. Researchers misunderstand confidence intervals and standard error bars. *Psychological Methods*, 10(4):389–396, Dec. 2005. doi: 10.1037/1082-989X.10.4.389

[3] W. S. Cleveland and R. McGill. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387):531–554, Sept. 1984. doi: 10.1080/01621459.1984.10478080

[4] M. Fernandes, L. Walls, S. Munson, J. Hullman, and M. Kay. Uncertainty displays using quantile dotplots or cdfs improve transit decision-making. In *Proc. CHI*, p. 1–12. Association for Computing Machinery, New York, 2018. doi: 10.1145/3173574.3173718

[5] D. G. Goldstein and D. Rothschild. Lay understanding of probability distributions. *Judgment and Decision Making*, 9(1):1–14, Jan. 2014. doi: 10.1017/S1930297500004940

[6] J. Hullman. Why Authors Don't Visualize Uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):130–139, Jan. 2020. doi: 10.1109/TVCG.2019.2934287

[7] J. Hullman, M. Kay, Y. S. Kim, and S. Shrestha. Imagining Replications: Graphical Prediction Discrete Visualizations Improve Recall Estimation of Effect Uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):446–456, Jan. 2018. doi: 10.1109/TVCG.2017.2743898

[8] J. Hullman, X. Qiao, M. Correll, A. Kale, and M. Kay. In Pursuit of Error: A Survey of Uncertainty Visualization Evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):903–913, Jan. 2019. doi: 10.1109/TVCG.2018.2864889

[9] S. L. Joslyn and J. E. LeClerc. Uncertainty forecasts improve weather-related decisions and attenuate the effects of forecast error. *Journal of Experimental Psychology: Applied*, 18(1):126–140, Mar. 2012. doi: 10.1037/a0025185

[10] A. Kale, M. Kay, and J. Hullman. Visual reasoning strategies for effect size judgments and decisions. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):272–282, Feb. 2021. doi: 10.1109/TVCG.2020.3030335

[11] M. Kay, T. Kola, J. R. Hullman, and S. A. Munson. When (ish) is my bus? user-centered visualizations of uncertainty in everyday, mobile predictive systems. In *Proc. CHI*, pp. 5092–5103. Association for Computing Machinery, New York, 2016. doi: 10.1145/2858036.2858558

[12] S. M. Kosslyn. *Graph design for the eye and mind*. Oxford University Press, Aug. 2006. doi: 10.1093/acprof:oso/9780195311846.001.0001

[13] R. Lipshitz and O. Strauss. Organizational Behavior and Human Decision Processes, Volume 69, Issue 2. *Organizational Behavior and Human Decision Processes*, 69(2):149–163, Feb. 1997. doi: 10.1006/obhd.1997.2679

[14] A. M. Maceachren, R. E. Roth, J. O'Brien, B. Li, D. Swingley, and M. Gahegan. Visual semiotics uncertainty visualization: An empirical study. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2496–2505, Dec. 2012. doi: 10.1109/TVCG.2012.279

[15] C. North. Toward measuring visualization insight. *IEEE Computer Graphics and Applications*, 26(3):6–9, May 2006. doi: 10.1109/MCG.2006.70

[16] K. Potter, P. Rosen, and C. R. Johnson. From quantification to visualization: A taxonomy of uncertainty visualization approaches. *IFIP Advances in Information and Communication Technology*, 377 AICT:226–247, 2012. doi: 10.1007/978-3-642-32677-6_15

[17] S. Savelli and S. Joslyn. The advantages of predictive interval forecasts for non-expert users and the impact of visualizations. *Applied Cognitive Psychology*, 27(4):527–541, June 2013. doi: 10.1002/acp.2932

[18] W. E. Walker, P. Harremoes, J. Rotmans, J. P. Van Der Sluijs, M. B. A. Van Asselt, P. Janssen, and M. P. Krayer Von Krauss. Defining uncertainty: A conceptual basis for uncertainty management in model-based decision support. *Integrated Assessment*, 4(1):5–17, Mar. 2003. doi: 10.1076/iaij.4.1.5.16466