Exploring the Gap Between Informal Mental and Formal Statistical Models

Hanspeter Pfister¹ Martin Wattenberg¹ Johanna Beyer¹
Carolina Nobre²

¹Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, United States of America,

²Department of Computer Science, Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts, United States of America

Published on: Jul 30, 2021

DOI: <https://doi.org/10.1162/99608f92.ba00865a>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

Hullman and Gelman argue for unifying visual exploratory data analysis (EDA) and confirmatory data analysis (CDA), with the idea that a synthesis of these two perspectives can lead people to more robust, reliable conclusions. Although EDA is sometimes viewed as a ‘model-free’ activity, Hullman and Gelman suggest we need a better understanding of the role that models play in this process. From a descriptive perspective, it seems likely that people do have a prior mental model as they approach a data set; from a normative point of view, there may be better and worse ways of using these implicit models. As a first step, they point to a need for a theory of graphical inference during EDA rooted in Bayesian inference.

We find their arguments compelling and believe that developing a new theory for EDA provides exciting avenues for future research. One significant challenge is that the informal mental models that people use during EDA may not easily translate to formal statistical models. In what follows, we discuss some of the differences between the two types of models, how studying these differences could be fruitful, and how the resulting theory might affect future visual analytics tools.

Blind Spots of Graphical Inference

One potential framework for a theory of graphical inference is Tversky’s (2002) *congruence principle*: that the perceived form of graphics should match the concepts they are meant to convey. To put this principle into practice in a statistical context requires understanding the relationship between visual perception, statistical models, and mental models.

For instance, we know that the human visual system does a reasonable job at specific basic statistical tasks: spotting outliers via attentional “popout” (Treisman & Gelade, 1980) or finding the mean of sizes of a collection of objects (Ariely, 2001). There may be more sophisticated things that we do efficiently. For example, when we look at a complex scatter plot like the Hertzsprung-Russell diagram (Spence & Garrison, 1993) (Figure 1), it seems likely that we are able to do something like clustering or kernel density estimation. But there are many statistical tasks we do less well, like guessing the kurtosis of the sizes of a collection of objects.

Figure 1. Hertzsprung-Russell diagram: 22,000 stars plotted by color vs. luminosity. What hypotheses would you make about the distribution (Wikipedia, “Hertzsprung–Russell Diagram,” 2021)?

The set of things that our eyes are best at is not the same as the set of models an expert statistician would pick for analysis. Traditionally, visualization researchers have focused on finding statistical tasks where the human visual system excels—those are the ingredients for good visualization designs. But perhaps we should pay more attention to which types of inference are hard to do graphically. These could be ripe for machine augmentation, with non-visual aspects (simulations, text descriptions, quantitative testing) that better capture the underlying models.

On the Importance of Informal Mental Models

That some statistical models are hard to represent visually is not the only issue with integrating EDA and CDA. The proposed framework by Hullman and Gelman raises the question: what are people’s actual mental models

of data, and how do they correspond to ideal statistical models? The informal models used by analysts—especially non-statisticians—may be fundamentally different from rigorous statistical models.

From one perspective, this difference is a core motivation for bringing in classic statistical techniques. The flaws of human nature in data analysis are well-documented, including cognitive biases, data collection errors, statistical mistakes, p-hacking, and many more. The point of a more rigorous, statistical approach may be precisely to address these issues.

At the same time, such an approach raises practical usability questions. Eliciting an analyst's expectations of a data set is challenging to do without relying on concepts such as uncertainty, probabilities, and general statistics. It is well documented, including in several papers by Hullman (Hullman et al., 2019; Kim et al., 2017), that the concepts of probability and uncertainty are difficult to understand for experts and non-experts alike. In addition, many users of data analysis are not educated in advanced statistics. Can we expect them to understand the concepts of Bayesian analysis and prior models?

Moreover, domain experts sometimes have functional, sophisticated mental models that remain difficult to translate into a classical statistical model. Our experience is that it can be surprisingly hard for them to make these mental models explicit. For more than a decade now, some of us have been working with neuroscientists on the reconstruction of neurons from high-resolution images of brain tissue (Kasthuri et al., 2015). Often, our collaborators will point to the reconstructed neurons and exclaim, 'I did not expect to see this.' When prompted, they cannot clearly articulate how they discovered these new insights or how exactly they differ from their expectations. These experts seem to have an implicit model of what looks 'right' in their mind that has been shaped by years of training and experience in looking at similar data. Even for these experts, it is very difficult to make those models explicit and to articulate them. This problem is exacerbated by the unprecedented and growing amounts of image data that they are able to collect and analyze (Figure 2).

Figure 2. Dense 3D neuron reconstruction at the nanometer scale from one cubic millimeter of a human cerebral cortex (zoomed-in view). Generating and confirming hypotheses in this 1.4-petabyte dataset is challenging (Shapson-Coe et al., 2021). Data available at *Explore | H01 Release* (n.d.).

One approach to this problem would be to look for ways to express prior beliefs that fit human models better. People often have mental models of data that are example- or prototype-based, what statisticians call non-parametric models (Hampton, 2006). Many common practices in visualization already reflect this. For example, it is common to compare a given company’s stock to broader sector or market indices when graphing stock prices. Implicitly, this is a comparison to a prototype as a baseline. This ‘prototype pattern’ is one example of an informal way of thinking about data that is relatively easy to translate into formal terms.

Other important mental models may be harder to capture, however. In our experience, one of the main advantages of visualizations during EDA is to detect errors in the data and to calibrate levels of trust in data sources. Users often build up subtle models of biases and other problems with different data sources. These prior beliefs are critical to analysis, but it can be highly nontrivial to make them explicit or quantitative.

(Arguably the success of the site [FiveThirtyEight](#) was based on doing exactly this in the context of political polls.) These insights about data quality issues are unglamorous, but they could be among the most important uses of visualization during the whole analysis cycle. How do visualizations to reveal data quality issues fit into a statistical modeling framework? Do we need a meta-model about how much we can rely on and trust our data?

The Role of Domain Expertise

Hullman and Gelman highlight a particularly exciting research direction, namely explicitly integrating subject matter expertise into an analysis.

An analyst's expectations of the data are influenced by their domain knowledge, biases, cognitive abilities, motivations, and even personality traits. For example, what might be unexpected and surprising to one analyst, may seem natural to an analyst deeply familiar with the domain. Successfully modeling the analysts' expectations in this scenario implies properly capturing and accounting for their domain expertise.

Indeed, visualizations encourage the viewer to do a kind of 'join' between the information they see on the screen and their own mental database. For example, for years to come, when economists see a 2020 anomaly in a time series they will say, 'Yes, naturally, the pandemic.' Or an analyst examining a US map with counties colored by sales growth may recognize an obvious correlation with demographic patterns (as pointed out in the popular xkcd comic shown in Figure 3). In each case, the user will make the connection even if the database at hand has no explicit data on events or demographics. The magic of visualizations is they let a person look for relations between tables A, B, and C—and guess that A actually depends on Z.

Figure 3. xkcd comic showing an obvious correlation between US population density and an unrelated variable. (Heatmap, n.d.)

Hullman and Gelman point to this process when they discuss the role of domain expertise in identifying and reducing model misfit. There's a natural need for confirmatory data analysis, but this often requires acquiring more data. Finding a way to allow users to quickly add in a set of standard statistics from an existing encyclopedic database could enable a fluid interplay between EDA and CDA. This would also be an opportunity for tools to help with the statistical pitfalls of multiple comparisons and data mining.

One might even imagine using machine learning techniques to give systems some kind of ‘common sense’ about data. Of course, ‘common sense’ is a broadly ambiguous term, but even partial steps in that direction could be beneficial. For instance, if every human economist knows that 2020 is an anomalous year, why shouldn’t machines know this too? One obvious challenge here is that in some domains, like astronomy, there’s likely nothing special about 2020—so any sort of common-sense statistical database will need to be scoped appropriately to a specific domain, or have some kind of semantic knowledge. However, it seems natural that a useful model of Bayesian graphical inference would take into account the fact that readers of visualizations bring a world of outside knowledge to bear on their analysis.

Making Informal Mental Models Visual—Inspiration from Visualizations for Communication

Hullman and Gelman focus on visual interfaces for exploratory and confirmatory data analysis but briefly touch on the use of visualizations for communication. They argue that stories and visual storytelling can themselves be viewed as model checks, with the ‘twist’ in a good story corresponding to a confounding of expectations.

Their insight may help us find examples of good interfaces for eliciting and checking models. Good journalism is very much about surprise, about upsetting a reader’s prior assumptions and updating their existing beliefs. Hence, we can take inspiration from journalism for examples on how we can build interactive data visualizations that encourage users to update their beliefs in presence of new data. Because journalists generally target wide audiences, these examples are particularly helpful in finding ways to help non-statisticians, users who might not be able to tell a binomial from a Gaussian distribution.

Sketching is one approach to get users to think about their biases and prior knowledge. Even if they do not have significant training in statistics, sketches encourage users to consider the shape of the data they expect before showing them the actual distribution. *The New York Times* has a powerful example of this approach, highlighting how family income affects children’s college chances (Aisch, Cox, & Quealy, 2015). This is a beautiful, simple method to elicit prior assumptions that requires no background in statistics. Sketches can also be used during an iterative process of constructing, examining, and reconstructing in a kind of ‘conversation’ with the data (Tversky & Suwa, 2009). It is interesting to consider how iterative sketching could be used to solicit statistical priors.

For data that is too complex or difficult to sketch, simulations can be a powerful alternative, allowing users to adjust parameters and see the simulated outcome. Another example from the New York Times uses a simulation approach to help readers decide if it is better to buy or rent (Bostock et al., 2014). An interesting goal for future research would be to find a way to create such simulations as easily as sketching graphs.

Making readers aware of their biases and preconceived ideas of what the data looks like in a playful way (e.g., through sketch interfaces or interactive simulations) can be a powerful way to surprise an audience and prompt them to update their existing beliefs, independent of their statistical background and training.

Conclusions

Hullman and Gelman's call for a closer relationship between EDA and CDA leads to a series of challenges and potential opportunities. We have highlighted a few here. The connecting theme in these challenges is bridging the divide between informal mental models and rigorous statistical models.

On a theoretical level, we need a better understanding of the blind spots of graphical inference. Traditional visualization research focuses on the best fits between visual perception and statistical inference. What can we say about the worst fits? We also need a deeper understanding of the mental models employed by expert practitioners. There may be an analogy with current machine learning work on understanding the results of artificial neural networks—real-life neural networks may be just as difficult to interpret!

Beyond theory, we have pointed to several areas where new tooling and interfaces may help. One of the key aspects of visual exploratory analysis is that it often suggests connections with domain expertise, including data that is not yet at hand. Finding ways to help users bring in this implicit knowledge—and even ways for machines to help make these connections—could lead to powerful new tooling. Finally, we've made the point that as we look for new interfaces to bring together EDA and CDA, we shouldn't ignore existing work, especially in journalism, that is tackling this exact problem. An interesting avenue for future work is to develop tools that allow non-experts to quickly design visualizations and interfaces for sketching and simulations.

Disclosure Statement

Hanspeter Pfister, Martin Wattenberg, Johanna Beyer, and Carolina Nobre have no financial or non-financial disclosures to share for this article.

References

- Aisch, G., Cox, A., & Quealy, K. (2015, May 28). You draw it: How family income predicts children's college chances. *The New York Times*. <https://www.nytimes.com/interactive/2015/05/28/upshot/you-draw-it-how-family-income-affects-childrens-college-chances.html>
- Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, 12(2), 157–162. <https://doi.org/10.1111/1467-9280.00327>
- Bostock, M., Carter, S., & Tse, A. (2014, May 21). Is it better to rent or buy? *The New York Times*. <https://www.nytimes.com/interactive/2014/upshot/buy-rent-calculator.html>

Explore | H01 Release. (n.d.). Retrieved July 9, 2021, from <https://h01-release.storage.googleapis.com/explore.html>

FiveThirtyEight. (n.d.). Retrieved July 9, 2021, from <https://fivethirtyeight.com/>

Hampton, J. A. (2006). Concepts as prototypes. In *Psychology of Learning and Motivation* (Vol. 46, pp. 79–113). Elsevier. [https://doi.org/10.1016/S0079-7421\(06\)46003-5](https://doi.org/10.1016/S0079-7421(06)46003-5)

Heatmap. (n.d.). Xkcd. Retrieved July 9, 2021, from <https://xkcd.com/1138/>

Hertzprung–Russell diagram. (2021). In *Wikipedia*. https://en.wikipedia.org/w/index.php?title=Hertzprung%E2%80%93Russell_diagram&oldid=1030428942

Hullman, J., Qiao, X., Correll, M., Kale, A., & Kay, M. (2019). In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1), 903–913. <https://doi.org/10.1109/TVCG.2018.2864889>

Kasthuri, N., Hayworth, K. J., Berger, D. R., Schalek, R. L., Conchello, J. A., Knowles-Barley, S., Lee, D., Vázquez-Reina, A., Kaynig, V., Jones, T. R., Roberts, M., Morgan, J. L., Tapia, J. C., Seung, H. S., Roncal, W. G., Vogelstein, J. T., Burns, R., Sussman, D. L., Priebe, C. E., ... Lichtman, J. W. (2015). Saturated reconstruction of a volume of neocortex. *Cell*, 162(3), 648–661. <https://doi.org/10.1016/j.cell.2015.06.054>

Kim, Y.-S., Reinecke, K., & Hullman, J. (2017). Explaining the gap: Visualizing one’s predictions improves recall and comprehension of data. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (pp. 1375–1386). Association for Computing Machinery. <https://doi.org/10.1145/3025453.3025592>

Shapson-Coe, A., Januszewski, M., Berger, D. R., Pope, A., Wu, Y., Blakely, T., Schalek, R. L., Li, P. H., Wang, S., Maitin-Shepard, J., Karlupia, N., Dorkenwald, S., Sjostedt, E., Leavitt, L., Lee, D., Bailey, L., Fitzmaurice, A., Kar, R., Field, B., ... Lichtman, J. W. (2021). A connectomic study of a petascale fragment of human cerebral cortex. *bioRxiv*. <https://doi.org/10.1101/2021.05.29.446289>

Spence, I., & Garrison, R. F. (1993). A remarkable scatterplot. *The American Statistician*, 47(1), 12–19. <https://doi.org/10.1080/00031305.1993.10475925>

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12(1), 97–136. [https://doi.org/10.1016/0010-0285\(80\)90005-5](https://doi.org/10.1016/0010-0285(80)90005-5)

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies*, 57(4), 247–262. <https://doi.org/10.1006/ijhc.2002.1017>

Tversky, B., & Suwa, M. (2009). Thinking with sketches. In *Tools for Innovation: The science behind the practical methods that drive new ideas* (pp. 75–84).

<https://doi.org/10.1093/acprof:oso/9780195381634.003.0004>

©2021 Hanspeter Pfister, Martin Wattenberg, Johanna Beyer, and Carolina Nobre. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](#), except where otherwise indicated with respect to particular material included in the article.