Harvard Data Science Review • Issue 3.3, Summer 2021

Visualization and Inferences

Jean-Daniel Fekete¹

¹Aviz, French National Centre for Scientific Research, French National Institute for Research in Computer Science and Automation, Université Paris-Saclay, Paris, France

Published on: Jul 30, 2021

DOI: <https://doi.org/10.1162/99608f92.049ac5b4>

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

Introduction

In their article, [Hullman and Gelman \(2021\)](#) raise several issues related to the possible misuse of visualization as supported by popular visualization systems. Starting from the seminal work of ([Tukey, 1977](#)), they explain that the separation between *exploratory data analysis* (EDA) and *confirmatory data analysis* (CDA) is not as clear as stated originally. The flexibility and power of analysis environments such as RStudio or Jupyter lead to mixing EDA and CDA, with too little control over the inferences made using visualizations and possibly leading to wrong decisions in the end.

[Hullman and Gelman \(2021\)](#) are trying to clarify the process by which analysts use EDA and CDA to reach sound generalizable conclusions and this clarification is much needed. In particular, because the visualization community and the statistics community have different views on what the process should be and it is clear that a discussion between the two communities (and more) can help align the views and hopefully agree on problems and possible solutions. [Hullman and Gelman \(2021\)](#)'s article provides a very comprehensive review of EDA vs. CDA and lists most of the problems and controversies from the authors' perspective.

Coming from an HCI and visualization perspective, I would agree on the fact that the modern data analysis process, fast and flexible, should come with safeguards. However, I will try to explain why some of the solutions proposed by [Hullman and Gelman \(2021\)](#) related to misusing statistics may be overlooking the role and process of visual exploration, and that their proposed safeguards could lead to under-utilizing the human visual capabilities. My three main points are that:

- EDA should remain separate from CDA as much as possible,
- visualization is about perception and patterns, not just about inferences, and
- improving analysis systems should be done in a layered way and not in a monolithic way.

The goal of this comment is to contribute to the dialogue on this very important problem.

An initial important comment is that one meaning of EDA used in [Hullman and Gelman \(2021\)](#) is inspired from ([Gelman, 2003](#)): “We view model checking as the comparison of data to replicated data under the model. This includes ‘exploratory data analysis’ and ‘confirmatory data analysis’ as special cases: EDA is the graphical comparison and CDA is the p-value, but they are based on the same hypothesis test”; and from [Gelman, 2004](#)): “exploratory data analysis is the search for unanticipated areas of model misfit.” Both definitions restrict EDA to model checking when the original intent of the EDA book is: “to expose its readers and users to a considerable variety of techniques for looking more effectively at one’s data” ([Tukey, 1977](#)), a much wider definition. When [Hullman and Gelman \(2021\)](#) discuss the problems of EDA, their critiques seem to apply to the wider definition, whereas when they discuss the solutions, they seem to apply to the more restricted definitions. My arguments here are structured around the wider definition.

The Roles of Visualization in the Hypothetico-Deductive Method

The hypothetico-deductive method (HDM) has been the cornerstone of the scientific process at least since the beginning of the 20th century. With this method, a scientific inquiry starts with a hypothesis that can be falsified or proven depending on the epistemological system and proceeds with rigorous deductive methods to check the hypothesis and reach a verifiable conclusion.

Hypothesis Generation

Yet the HDM says nothing about where the hypothesis comes from. Visualization is a powerful mechanism that helps humans generate hypotheses from data by relying on their visual system. Once the hypothesis is generated, it can enter the standard HDM process without raising any problem (almost) since the HDM is agnostic about the origin of the hypothesis.

With modern, fast, and flexible tools, interesting hypotheses (often called ‘insights’ in the visualization community) can be generated quickly. When trying to validate a hypothesis obtained from visualizing data, the most important safeguard is to avoid reusing the exploration data for confirmation. This problem is well known in statistics. It raises the issue of ‘data-analysis literacy’ caused by the ease of use of the new tools that lower the usability barrier but does not enforce literacy.

Visualization helps generate hypotheses because it relies on the human perception of patterns. ([Boy et al., 2014](#)) define *visualization literacy* as: “The ability to confidently use a given data visualization to translate questions specified in the data domain into visual queries in the visual domain, as well as interpreting visual patterns in the visual domain as properties in the data domain.”

The hypothesis-generation aspect of visualization comes from the second part of the definition: when a human sees a visual pattern in a chart, it implies that there is a related property in the data. That property needs to be retrieved through a translation process and may turn out to be trivial or surprising once translated. Yet, the pattern detection is triggered by the perceptual system and therefore not or marginally subject to prejudices (though it is sensitive to biases). For example, a trivial pattern that the visual system recognizes at a glance is the linear alignment of points in a scatter plot or any 2D chart. When spotted in a histogram, this pattern means that the histogram values follow a linear progression. When spotted in the Ulam spiral ([Gardner, 1964](#)), the pattern means that there is some structure in prime numbers but mathematicians have not been able to formalize it yet. The pattern detection mechanism is extremely fast and efficient for several kinds of patterns, and therefore humans can generate hypotheses at high speed. This is why visualization systems are typically optimized for showing patterns. Most of the time, the patterns will be uninteresting once translated. By quickly iterating through the visual representations, more patterns can be exposed, more properties inferred, most of them leading to uninteresting hypotheses. The remaining interesting ones can be fed later to the HDM for confirmation or invalidation. In addition to insights and model checks, visualization will also reveal weird patterns, unsuitable to analysis and potentially requiring further investigations, as explained by [Wilkinson](#).

(1999). It is easy to overlook this translation process from pattern to property when our visualization literacy increases.

When [Hullman and Gelman \(2021\)](#) write, “Any exploratory graph should be interpretable as a model check, a comparison to ‘the expected,’” they seem to merge the pattern spotting stage, the translation stage, and model selection. The first stage is not a model check, and the translation can generate multiple models or just surprise and no model, as for the Ulam spiral. The model check can only happen after one or more models have been hypothesized from the pattern.

Hypothesis Testing

The first part of the visualization literacy definition “confidently use a given data visualization to translate questions specified in the data domain into visual queries in the visual domain” ([Boy et al., 2014](#)) allows quickly testing hypotheses and is useful with graphical inference testing. Yet, it raises several issues, some of them discussed in [Hullman and Gelman \(2021\)](#). The translation from data domain to visual domain has limitations: some data-domain properties cannot be efficiently translated into visual queries, and those that can, do so with different levels of accuracy.

Is using visualization more efficient than using a symbolic representation for solving a problem and running a series of computations and tests? [Larkin and Simon](#) try to formalize what ‘better’ representation of a problem means when comparing two data representations of the resolution, one solving the problem using symbolic inferences only (equivalent to function calls using a specified language and some data structures), and the other using a mix of symbolic and graphical inferences. They compare the resolution of a pulley problem and a geometrical problem with or without the support of a diagram. They translate the inference process required to solve each problem into low-level inference steps. In their translation, the inference steps are identical with or without diagrams but some steps are more efficiently resolved using perception than using a symbolic data structure. In summary, [Larkin and Simon](#) write: “We believe the right assumption is that diagrams and the human visual system provide, at essentially zero cost, all the inferences we have called ‘perceptual.’”

Can the process described by [Larkin and Simon](#) be used to solve a statistical problem in a similar way, with and without visualization? I believe that it would lead to a realistic description of a statistical inference process, highlighting the steps lending themselves to graphical inferences. Using Bayesian reasoning, as argued by [Hullman and Gelman \(2021\)](#), the usefulness and effectiveness of visualization would become more apparent when the high-level reasoning process is translated into low-level inference steps, akin to the process described by [Larkin and Simon](#). With a concrete decomposition of a reasoning process into steps, some of the steps can only be resolved at the cognitive level (e.g., logical deductions), others using computation (computations of p-values), and a few with graphical inferences.

Yet, graphical inferences in visualization are different from geometry. With visualization, some visual queries are accurate (two lines cross) and others are heuristics (estimating the average value of a histogram to compare

two histograms) and can lead to errors. Using a visual check to answer low-level inference steps would come with some level of confidence and accuracy depending on the heuristics used. Deciding if that level is good enough is a trade-off between time and accuracy. Using our visual system, we can perform tens to hundreds of low-level estimations per minute, but if we want to be accurate, we often need to perform computations using a computer language and each of the computations can take minutes to type and run.

This view of inferences from [Larkin and Simon](#) differs from what [Hullman and Gelman \(2021\)](#) discuss, but I would argue that some of the questions the latter ask are not related to visualization but to the inference process. Is visualization ‘model-free’? The patterns are produced by a visualization technique and data so if the data is transformed to follow a model, the patterns will relate to that model. For example, HCI research frequently uses one-way ANOVA to compare several interactive techniques under varying conditions, typically measuring completion time and the number of errors. Applying an ANOVA on experimental data requires it to be *normal or close to normal* ([Scheffé, 1959](#)). It is accepted in HCI to show the actual distribution to claim that it looks normal instead of using a normality test because most of the (frequentist) tests are too sensitive and the ANOVA is robust to slight violations of normality. Visualization is good enough for that case. Additionally, if the experimental data is skewed (as in Figure 4 of [Hullman and Gelman \(2021\)](#)), it is common to transform it (e.g., log-transform completion time in HCI experiments) so the transformed data looks normal and can be used for the ANOVA. A first examination of the data will allow spotting the well-known skewness of completion time, and a second examination of the log-transformed data will suffice to recognize a close-to-normal distribution that can be handled by the ANOVA. The visualization is therefore not ‘model-free’ but expects a model, helps viewers spot it quickly, and can help spot other distributions as well. This process shows that visual estimation can be sufficient to replace a rigorous normality statistical analysis. However, the process relies on a certain level of statistical literacy about ANOVA, normality tests, the shape of a quasi-normal distribution, and the shape of a skewed distribution fixable by a log-transform. The inference process to perform the analysis produces a sequence of computations and tests, and a few of the tests can be done visually, leading to new computations.

When [Hullman and Gelman \(2021\)](#) mention problems with novice data analysts using visualization to answer complex questions, as in their Figure 3, my interpretation would be that, without proper visualization literacy and statistical literacy, the inference steps performed using visual perception could easily be wrong.

According to [Larkin and Simon](#), graphical inferences are steps in a sequential inferential process that can be performed visually and efficiently. The model proposed by [Larkin and Simon](#) is simplistic in many ways; a real decomposition is certainly much more complex but currently out of reach, hidden in the human brain. In particular, with experience (increased visualization literacy), the translation of high-level questions to low-level inferences could use more complex strategies: e.g., rely on multiple fast visual checks instead of longer computational checks. [Hullman and Gelman \(2021\)](#), as well as the two initial articles on the topic ([Gelman, 2003](#); [Gelman, 2004](#)), discuss a truncated view of the whole resolution problem, mainly focusing on visually

checking models at an advanced stage of statistical analysis when a few models are already hypothesized and explored. In that particular case, the low-level inferences they describe are visual at some point but have required a long series of choices to find the right visual representation for fast and accurate decisions.

Additionally, the visual checks are not always conclusive. Heuristic checks remain a problem in this framework and can be hard to control cognitively. A rigorous statistical process using computed tests would be more accurate but at the expense of time. Using heuristics, humans are sometimes very good at making fast and accurate decisions and sometimes very bad ([Gladwell, 2006](#); [Gigerenzer, 2008](#)). The Bayesian reasoning described in [Hullman and Gelman \(2021\)](#) remains for now cognitively complex. It does not explain how the decomposition into low-level operations is done to allow graphical inferences, and in particular how to make sure that the graphical inferences are accurate enough. More research is needed in visualization to better understand and control our perceptual abilities and decide when graphical inference is good enough for particular checks. Currently, graphical inferences are performed using ad-hoc representations that we know well enough to trust them, not always grounded by empirical studies.

Safeguards and Interference

[Hullman and Gelman \(2021\)](#) mention safeguards regarding the multiple comparison problem amplified by the speed of graphical inferences. Their goal is to encourage more robust inferences by promoting the visualization of uncertainty and avoiding aggregation.

As explained above, the level of robustness of the inferences is also a matter of time and resources, and therefore I would consider it a problem of enlightened choice. Analysts should be able to choose their trade-offs between robustness and time/resources, in an accurate and confident way. Therefore, as stressed by [Hullman and Gelman \(2021\)](#), controlling the uncertainty is essential.

However, instead of addressing the problem in a ‘depth-first’ fashion, where each step from pattern perception to model checking is validated or checked on the spot, I believe that a ‘breadth-first’ process is preferable. Allowing rapid visual scanning of visualizations will generate many hypotheses at first, some of them being invalidated immediately or later during an insight collection session. Then, all the insights should be assessed and examined in a validation process, using a mix of visualization, computations, and sometimes heavier experiments. One important issue with the early phases of the exploration is to make sure the visualizations remain simple and effective at quickly checking hypotheses. The reason for staging the process by layers of robustness lies in the cost of switching between visual perception (for spotting patterns) and cognitive activities (for checking higher-level statistical properties). Staying in ‘visual perception mode’ to collect all the possible patterns/insights is more efficient than switching back and forth from perception to ‘statistical inference mode.’

Additionally, showing uncertainty and detailed information results in more complex visualizations that are harder to read and less efficient to parse visually, requiring a higher visualization literacy. In the worst case, additional marks used to express the uncertainty interfere with the low-level perception process and limit the

efficiency of visualization. More often, they are difficult to interpret. More research is needed to understand the trade-offs, when fast uncertainty-unaware perception is more useful than slower uncertainty-aware perception.

I agree with [Hullman and Gelman \(2021\)](#) that current visualization systems are designed for fast pattern detection and low robustness. However, I believe that the transition to more robustness should be done through layering, maybe switching the visual exploration system altogether from Tableau to RStudio and others, making sure the hypotheses gathered in the first systems, and the visualizations that triggered them, are transmitted to the next. Figure 8 of [Hullman and Gelman \(2021\)](#) shows well multiple levels of uncertainty and complexity that highlight the trade-offs between simplicity and speed vs. complexity and precision, the latter requiring a high level of visualization literacy to extract information confidently from the charts.

Using this layered scenario, an exploratory process could start with fast, simple, and effective visualizations to gather initial insights, followed by more complex and uncertainty aware ones to decimate the spurious insights/hypotheses, followed by more complex model-based analyses to validate or invalidate the remaining hypotheses using various levels of sophistication: frequentist, Bayesian, multiverse, and beyond.

Conclusion

EDA goes beyond model checking; it has a generative aspect on its own based on the capability of the visual system to spot patterns without a priori meaning. A visualization is good because it emphasizes important patterns. Interaction allows quickly switching between different visualizations and visual configurations to be exposed to more patterns. Changing the parameters of visualizations allows more patterns to appear (e.g., using a log scale rather than a linear one for revealing new alignments of points). This capability offered by visualization is new in the scientific toolbox ([Fekete et al., 2008](#)) and needs to be advertised more broadly.

Graphical inferences are an efficient method (most of the time) to check hypotheses and often they, themselves, rely on models (visualizations are not always model-free). The high-level sense-making loop ([Pirulli & Card, 2005](#)) supported by interactive data analysis tools relies heavily on graphical inferences but does not control them well. These inferences can be wrong due to aggregation hiding important information such as uncertainty, variability, raw size, etc. Unfortunately, exposing all the information also reduces the efficiency of the visualizations and requires higher degrees of literacy from the analyst. Visualization research need to investigate how to best distill the right level of information in a layered model.

I share [Hullman and Gelman \(2021\)](#)'s belief that we need to improve our data analysis tools from EDA to CDA, but these improvements should be introduced with caution and in a layered fashion to avoid interfering with the human perceptual capabilities. Still, more research is needed to survey all the graphical inference methods, their effectiveness, their limits, and their accuracy to inform the data analysts. I am not sure we will find a theory of graphical inferences but we certainly need more empirical research. Designing consistent and provenance-aware analysis systems to first generate and then confirm hypotheses, with varying levels of confidence, is becoming essential. It would mitigate the proliferation of incompatible tools, and turn them into

complementary tools for addressing data analysis supporting several levels of confidence and different levels of skills.

Acknowledgments

Thanks to Anastasia Bezerianos for her feedback, and to the [Aviz Team](#) at Inria and Univ. Paris-Saclay for insightful discussions on [Hullman and Gelman \(2021\)](#).

Disclosure Statement

Jean-Daniel Fekete has no financial or non-financial disclosures to share for this article.

References

- Boy, J., Rensink, R. A., Bertini, E., & Fekete, J.-D. (2014). A Principled Way of Assessing Visualization Literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12), 1963–1972. <https://doi.org/10.1109/TVCG.2014.2346984>
- Fekete, J.-D., Van Wijk, J., Stasko, J. T., & North, C. (2008). The value of information visualization. In A. Kerren, J. T. Stasko, J.-D. Fekete, & C. North (Eds.), *Information Visualization: Human-Centered Issues and Perspectives* (pp. 1–18). Springer. https://doi.org/10.1007/978-3-540-70956-5_1
- Gardner, M. (1964). Mathematical games: The remarkable lore of the prime number. *Scientific American*, 210, 120–128. <https://doi.org/10.1038/scientificamerican0364-120>
- Gelman, A. (2003). A Bayesian formulation of exploratory data analysis and goodness-of-fit testing. *International Statistical Review*, 71(2), 369–382. <https://doi.org/https://doi.org/10.1111/j.1751-5823.2003.tb00203.x>
- Gelman, A. (2004). Exploratory data analysis for complex models. *Journal of Computational and Graphical Statistics*, 13(4), 755–779. <https://doi.org/10.1198/106186004X11435>
- Gigerenzer, G. (2008). *Gut feelings: Short cuts to better decision making*. Penguin Books Limited.
- Gladwell, M. (2006). *Blink: The power of thinking without thinking*. Penguin Books Limited.
- Hullman, J., & Gelman, A. (2021). Designing for interactive exploratory data analysis requires theories of graphical inference. *Harvard Data Science Review*, 3(3). <https://doi.org/10.1162/99608f92.3ab8a587>
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1), 65–100. <https://doi.org/10.1111/j.1551-6708.1987.tb00863.x>

Pirolli, P., & Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. *Proceedings of International Conference on Intelligence Analysis*, 5, 2–4.

Scheffé, H. (1959). *The analysis of variance* (p. 477). Wiley.

Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley. <http://opac.inria.fr/record=b1080310>

Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54(8), 594–604. <https://doi.org/10.1037/0003-066X.54.8.594>

©2021 Jean-Daniel Fekete. 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.

References

- Boy, J., Rensink, R. A., Bertini, E., & Fekete, J.-D. (2014). A Principled Way of Assessing Visualization Literacy. *IEEE Trans. Vis. Comput. Graphics*, 20(12), 10. <https://doi.org/10.1109/TVCG.2014.2346984> ↵
- Fekete, J.-D., Van Wijk, J., Stasko, J. T., & North, C. (2008). The Value of Information Visualization. In Springer (Ed.), *Information Visualization: Human-Centered Issues and Perspectives* (pp. 1–18). Springer. <https://hal.inria.fr/hal-00701741> ↵
- Gardner, M. (1964). Mathematical games: The remarkable lore of the prime number. *Scientific American*, 210, 120–128. <https://doi.org/10.1038/scientificamerican0364-120> ↵
- Gelman, A. (2003). A bayesian formulation of exploratory data analysis and goodness-of-fit testing*. *International Statistical Review*, 71(2), 369–382. <https://doi.org/https://doi.org/10.1111/j.1751-5823.2003.tb00203.x> ↵
- Gelman, A. (2004). Exploratory data analysis for complex models. *Journal of Computational and Graphical Statistics*, 13(4), 755–779. <https://doi.org/10.1198/106186004X11435> ↵
- Gigerenzer, G. (2008). *Gut feelings: Short cuts to better decision making*. Penguin Books Limited. <https://books.google.fr/books?id=1nS6TuRcceIC> ↵
- Gladwell, M. (2006). *Blink: The power of thinking without thinking*. Penguin Books Limited. https://books.google.fr/books?id=FiPLHNvT%7B%5Ctextbackslash%7D_NkC ↵
- Hullman, J., & Gelman, A. (2021). Designing for interactive exploratory data analysis requires theories of graphical inference. *Harvard Data Science Review*, 0(0). ↵
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cogn. Sci.*, 11(1), 65–100. <https://doi.org/10.1111/j.1551-6708.1987.tb00863.x> ↵
- Pirolli, P., & Card, S. (2005). The sensemaking process and leverage points for analyst technology as identified through cognitive task analysis. *Proceedings of International Conference on Intelligence Analysis*, 5, 2–4. ↵

- Scheffé, H. (1959). *The analysis of variance* (p. 477). Wiley. ↵
- Tukey, J. W. (1977). *Exploratory data analysis*. Addison-Wesley. <http://opac.inria.fr/record=b1080310> ↵
- Wilkinson, L. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist*, 54, 594–604. %20http://www.apastyle.org/manual/related/wilkinson-1999.pdf%20 ↵