Reflecting on the Design Criteria for Explanatory Visualizations

Aritra Dasgupta
New Jersey Institute of Technology

Enrico Bertini
New York University

Reflecting on the Design Criteria for Explanatory Visualizations

Robert Kosara
Tableau Software
rkosara@tableau.com

Aritra Dasgupta
PNNL
aritra.dasgupta@pnnl.gov

Enrico Bertini
New York University
enrico.bertini@nyu.edu

ABSTRACT

The visualization field has developed a good set of design criteria, metrics, and methods to assess visualization techniques and systems. These are all focused on analytical and exploratory uses, however. A large class of visualizations are created to present and communicate data and issues, however, and are seen by millions of people. We do not currently have a good grasp of what criteria should be used to systematically design and compare them, and how to do that. The aim of this paper is to raise the issue, describe different uses of visualizations, and propose criteria that should be considered while designing and critiquing them.

1. INTRODUCTION

Visualizations are not only used to explore and analyze data, but also to present and communicate findings and insights. Little research has looked into this category of explanatory visualizations so far, despite this being the much more common use case: millions of people see visualizations in news media every day.

Most existing work focuses on exploration and analysis, with the tacit assumption that the same techniques also work well for presentation or communication. When techniques are discussed, this is done based on criteria that are useful for assessing their potential for these tasks, but not necessarily for explanatory purposes. This leads to misunderstandings and to the dismissal of techniques that work well for presentation or communication even if they may not be very useful for analysis [12].

Traditionally, the purpose of visualization has been to enable users to find insights in data [16] themselves. But in many cases, data is presented to get a particular point across, whether that is an insight, an observation or awareness, or even a call to action or make a decision. For example, when a journalist creates a visualization for reporting on the current weather situation, the goal there is to mainly present the key trends and create awareness among the general public. When climate scientists create visualizations for communicating their results to policy makers on climate change, they are mainly calling for actions.

In this paper, we describe the different purposes visualization is used for today, and how they differ from the traditional analytical visualization. We then discuss the different design priorities for each of them, and finally propose a number of design criteria and challenges specific to explanatory visualizations.

C4PGV '16 Baltimore, MD USA

ACM ISBN –. . . –
DOI:

2. USES AND USERS OF VISUALIZATION

There is a large range of types of users and different uses by those users of visualization. Any particular user, or group of users, will have a number of different ways it employs visualization. We group these into *visualization ecosystems*, and discuss the three broad types of usages scenarios below.

2.1 Visualization Ecosystems

Creation of a visualization is driven by the intent of the designer, keeping the user or the audience in mind. The intents can be different, based on the audience even when the creator is the same. For example, a climate scientist can design a visualization of historical temperature data for both disseminating scientific knowledge within the climate science community (Figure 1), or communicating to policy-makers about key indicators and impacts. In the first case, the goal is exploratory analysis, while in the second case, the goal is presentation of results. Owing to these differences in who creates a visualization (whether it is the data producer or a visualization expert) and who the audience is, the scenarios can involve very different approaches and techniques. Hence, we argue, it is necessary to define the role of the different players in the visualization ecosystem and their interactions in the usage scenarios. We believe, by following the connections in this ecosystem, we can better reflect upon the design priorities of visualizations and evaluate them better.

Visualization ecosystems (Table 1) are comprised of the data producer, the data consumer or the visualization audience, and the designer; and the means adopted to provide value out of the data. Each row in Table 1 is a possible usage scenario (we do not claim that this is an exhaustive list). In some cases, the producers themselves can be the consumers. For example, we have collaborated with climate scientists, who design visualizations of the data they generate through simulation experiments, for exploring the patterns for further analysis. They also create visualizations for policy-makers, like, government committee on climate change, for communicating the key trends so that actions or

Data Producer	Data Consumer	Visualization Designer	Means	Output
Scientist	Self	Self	Analysis	Insight
	Science community	Self	Presentation	Insight
	Science community	Visualization Expert	Guided Analytics	Insight, Decisions
	Policy Makers	Self	Presentation	Decisions
Open Source	Public	Journalist	Presentation	Observation, Awareness
	Policy Makers	Journalist	Presentation	Insight, Awareness
Organizations	Public	Visualization Expert	Presentation	Observation, Awareness
	Policy Makers	Visualization Expert	Guided Analytics	Decisions

Table 1: Different usage scenarios that form visualization ecosystems involving the data producers and consumers. These lead to different means being adopted to derive value out of data. This can be an analysis process to derive insight, presentation to communicate findings and lead to awareness, or guided analytics to lead to decisions or action points. Design priorities of visualizations should adapt to these different scenarios.

decisions can be taken. Journalists do not generally produce data, but they do design visualizations for presenting news stories involving generally open source data. They extract the salient patterns that suit their story points and create visualizations for creating awareness or letting the public make observations about current affairs.

Visualization experts are generally involved in providing visual analytics solutions to both scientists and policy-makers. In these cases, the consumers may or may not know what they are looking for in the data. The visualization provides important visual cues that can lead to key insights that can be iterated upon, or immediate decisions. For example, visualization experts designing an urban visualization system can help city officials make key decisions on making the traffic management system more efficient.

2.2 Analysis

In analysis, the user treats the visualization as externalized memory in which he or she works. Since it is usually necessary to try out many different approaches and ask many different questions, remembering the details is not a priority; only key insights need to be remembered, and visualization tools often provide at least rudimentary support to bookmark or otherwise record key steps. The visualization techniques can also be entirely generic and minimalist, with more generic tools actually reducing cognitive load due to familiarity. The purpose of analysis is almost always to derive insight.

In this scenario the consumers can themselves be the producer (Table 1), or at least, the consumer know their data well and generally have an idea about what they are trying to find. The analysis process is iterative in nature, where the user directly interacts with the visualization. Often, multiple iterations are needed to derive *insight*. Analysis is usually the means for deriving insight, when domain experts themselves design a visualization (for example, a biologist looking at heat map of gene expression data) for their own analysis.

2.3 Guided Analytics

In addition to pure exploration/analysis and presentation, there is also a hybrid, commonly called *guided analytics*. This approach can be either guided by data analysis that points the user to potentially interesting features in the data [17] or by a predefined sequence of actions and views that is likely to lead the user to a conclusion. A typical example of this little-researched type of analysis is the selection of a consumer product like a smartphone or camera based on criteria; the user is led through a series of steps to pick criteria of interest, with the ability to explore what other options exist. In this scenario, the consumer might be familiar with the domain, but not necessarily a data producer. The consumer also might have very open-ended questions about the data, like a climate scientists wanting to know about spatio-temporal similarity of climate models or an administrator wanting to know about patterns of traffic congestion in a city.

When visualization designers create visualizations for domain experts or policy-makers, it is usually with the goal of providing seed points in the visualization on which the analytics can be based upon. For example, a visualization created for similarity analysis of climate models, can provide an initial overview of which models are similar and dissimilar, which can act as the seed points for the scientists to further drill down into why those models are similar or dissimilar. In addition to insight, the eventual output of guided analytics serve as incentives for decision-making.

2.4 Presentation and Communication

Presentation has very different goals and thus requires different techniques and strategies. The goal of a presentation is to communicate a set of key points, and for the audience to remember them. A large number of similar charts is not likely to be remembered in any detail [1]. Rather, a presentation needs to provide information in a way that is memorable and that might even contain a call to action [8]. This may include certain styling and formatting choices, and even embellishments such as images, to help create context.

The purpose of presentation, in addition to providing insight, can be to spread social or political awareness among the public. The typical scenario is when journalists design visualizations for the general public and policy-makers. The design has to tell a compelling story about the findings of the scientists to non-technical stakeholders. This use of visualization is slowly being recognized, but is still not receiving nearly as much attention as analysis [13].

In the course of our collaboration with climate scientists [7], we found that they the same design for data analysis and publications or presentation of their results within the community (Figure 1). This often leads to design problems, because although in this case the audience might be familiar with the data, but they might not always know what to look for in the data.

3. DESIGN PRIORITIES

The goals and criteria for exploration and analysis are very similar, and differ considerably from the needs of presentation and communication.

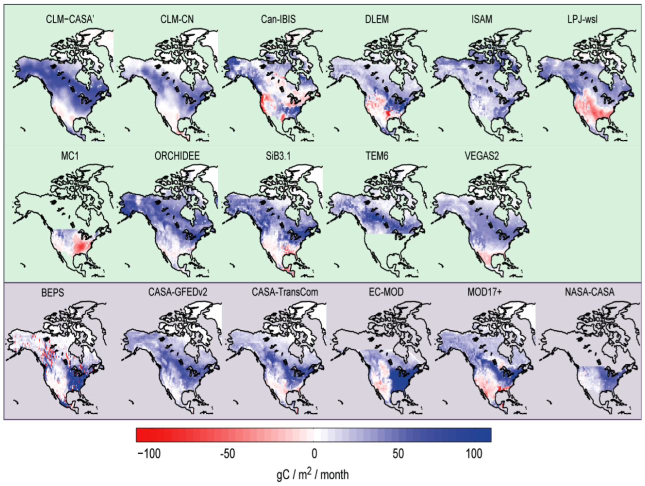

Figure 1: Maps representing climate models, designed by climate scientists for presenting their findings on model similarity [11]. The design is suboptimal for presentation purposes, as the maps are ordered randomly without any emphasis on communicating the degree of similaroty among the models.

3.1 Analysis

There is a continuum between exploration and analysis in visualization; the two have much in common. Exploration tends to be less based on questions and more on data patterns, while analysis is more goal-directed and based on knowledge about the data.

Generality. Exploration and analysis tools need to be as general as possible, since by its definition exploration is done on unknown data. General tools provide the means to work with a wide range of data and remain useful even when there are outliers or unexpected patterns.

The visualization literature is almost exclusively concerned with general tools, so the above might seem self-evident. However, in the context of presentation, very specific techniques (like the Connected Scatterplot [10]) are sometimes used that can be the right choice in that context.

Quick Iteration. Visualization tools are generally designed to allow for quick iteration to ask many different questions and look at the data in many different ways. This is a huge advantage visualization has, and also requires general tools. But this is not of relevance for presentation, where the presentation piece is constructed once, with great care, and then consumed many times.

Dead Ends. Similarly, running into dead ends is a natural part of the exploration process and even a good sign that one is covering a large space of possible hypotheses. A narrative is linear, however, and each step has to have a purpose. Adding many tangential or irrelevant threads to a story only serves to confuse the viewer by making it harder to follow the main idea.

Trust. When data producers want to analyze their own data, like scientists producing data through simulations, it is very important for them to trust what is being shown in the visualization. If there are too many transformations or abstractions that are used and not communicated within the representation or the technique, they may not believe the trends that the visualization shows. Especially when visualization experts design systems for guided analytics, these systems have to be self-contained to explain clearly, what the patterns mean and how they were derived.

Fidelity Related to trust, an important issue when designing visualizations for data producers, is to preserve the fidelity of the data as much as possible. For example, domain scientists across many disciplines spend a considerable amount of time generating their own data. When they are the data consumers through a visualization medium, it is imperative that information loss is minimized so that there is no bias in their conclusions.

3.2 Guided Analytics

While many of the design priorities in case of analysis and guided analytics overlap due to their similar goals, in this section we highlight those features that are specific to guided analytics.

Information Scent. In case of data with large cardinality and dimensionality, it is essential to provide users with some information scent about the salient relationships and trends. This can be done by creating layers of abstraction for providing overview, which can be used to drill down into why those relationships are present, e.g., with a meta view and data views.

System Feedback. One of the under-researched areas of visualization is to how the system can provide the user with feedback about her actions. Since guided analytics is mostly concerned with providing action points to the user, this is critical in prompting the user if say for example, the resolution chosen creates too much information loss.

Provenance. The ability to trace results and reproduce them is critical, especially when the data has been prepared by somebody else. Knowing not just the source of the data, but also how it has been processed before being shown, can change the interpretation of the visualization significantly.

3.3 Presentation and Communication

The goals in presentation and communication are quite different from the above. There is a much larger emphasis on guiding the user, but also reducing the amount of data shown to just the crucial parts. News media examples (Figure 2) provide a good illustration of the points we make below.The key difference between analysis, whether free or guided, and presentation is that in the latter case, the point to be made has been defined by an author, while analysis is open-ended by design. Having a pre-defined message may seem limiting and biased when viewed from a pure analysis perspective, however it is a requirement when the goal is communication of an issue. Since we assume the underlying data of the presentation to be available, it is always possible to perform further analysis on it using exploration and analysis tools.

Guidance. Presenting data means guiding the viewer through some kind of sequence. This is in contrast to exploration tools, which do not try to guide so as not to bias the exploration. But when presenting, guidance is a key element, otherwise the viewer is left to discover the information himself.

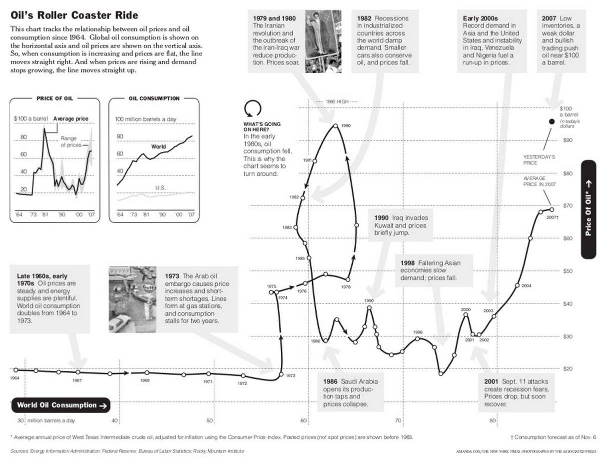

Figure 2: A newspaper graphic needs to draw attention, explain how it is read, and guide the reader. These are not typically priorities in analytical visualization. Graphic by Amanda Cox, *The New York Times* [5].

Specificity/Focus. Presentation tools are often specific not just to the type of data, but the specific data values being presented. This is not a shortcoming, but can be a strength: specific, tailored presentation is more likely to be remembered [1, 2]. General techniques like bar charts are effective, but also not distinctive or exciting.

Semantics. In addition to the numbers, visualization in the communication context often needs to also explain its context and what it even is about. This is not an issue when a subject matter expert is dealing with his or her own data, but when somebody is flipping through a newspaper or a website, context needs to be established and communicated clearly.

Efficiency. The attention span of data consumers, such as the general public, in case of presentation or communication, is limited. The design has to be efficient enough to enable quick interpretation of the intended message. This category encompasses the principles of effectiveness [15], use of pre-attentive features [4] and visual variables for efficient search for patterns [3].

Emphasis. Visualization techniques are typically designed not to bias or emphasize particular data, but to leave that to the user. However, when trying to make a point, it is often useful and necessary to emphasize particular elements of the data for clarity, or at least to give viewers a starting point. Merely showing the data does not usually provide a clear enough message for people to understand. Charts should be self-contained through the use of proper labelling, grids and annotations if necessary, which help emphasize the intended message. Improper use of these auxiliary information however can clutter charts and make the decoding process inefficient, by increasing the non-data ink [9].

Expressiveness. The expressiveness [15] criterion dictates whether the visual representation is well-matched with the properties of the data attributes. A lack of expressiveness would mean that the elements of the visualization design are unable to convey the intended message. This can often be the problem when visualization experts are not the visualization designer. For example, for showing temporal trends of temperature for different regions, for hundred years of data, if all data points are plotted in a line chart, it would be too cluttered with a lot a of jaggedness. For visualizations to be expressive, it is often necessary for designers to choose an appropriate level of abstraction, which in this case can be average over certain time periods, thus maximizing the ability of users to observe or be aware of the salient trends.

Low visual complexity. General visualization techniques can be visually complex. But in many cases, decision makers cannot afford the time for sifting through very complex displays. Visualizations designed for presentation rather than analysis need to be simplified, use more familiar techniques, and provide hints as to how to read them.

4. EVALUATING EXPLANATORY VISUALIZATIONS

Based on the descriptions of the different types of visualization described above, we propose a number of criteria to be used in evaluating explanatory visualization. We are also very aware of the challenges involved in this type of evaluation, some of which are also discussed in this section.

4.1 Criteria

The criteria for evaluating visualizations designed for communication and presentation are different than the very performance centric ones usually used for analysis (accuracy, error rate, task completion time).

Draw interest. Presenting information often means competing for attention. This is true on a website just as it is in a more formal presentation (where attendees might be distracted by email, etc.). Drawing attention to the visualization is therefore a key issue for presentation-oriented visualization that is not typically a concern for analytical tools.

Engagement. Once the viewers' attention has been captured, the question is how long they will stay with the view to explore, etc., before moving on. Generally, longer engagement should be better. However, that metric alone could be misleading because it might also measure time people spend confused over what they are shown, trying to decipher a needlessly complex visualization. On the other hand, they might also move on quickly if they cannot understand what is shown, and how.

Willingness to explore. Many interactive visualizations on the web are never actually interacted with by the vast majority of users who see them. Giving users clear instructions, affordances, and a reason to interact will lead to more engagement and, presumably, more information transfer.

Memorability. The point of a presentation is for the audience to take something away. At least some of the facts therefore need to be memorable enough to 'stick,' at least for a short time. A critical question in this context is, what should stick? Individual data values? Trends? The overall message? The unusual visual design used? Each of these might be helpful, but we do not currently know what is needed in which case.

Communicate the encoding (in addition to the data). For the audience to even understand what is being shown, the visualization often needs to communicate how it represents the data. This is especially critical in the case where no presenter is present to explain. A visualization's effectiveness in communicating its own design is not usually tested in user studies.

Persuade. Do people change their opinions after looking at visualization, as opposed to a paragraph of text? While visualization has traditionally been looked upon as a tool for informing, with the growing acceptance of visualization by mainstream journalist, there needs to be more research on how explanatory visualizations can be more persuasive [14].

Inspire action. Ultimately, the goal of many presentations is to inspire action or lead to a decision. How effective a technique or particular design is in doing this is difficult to assess (and there are usually many other factors), but is also perhaps the most important question to ask.

Some of these points can be tested with existing methods, at least to an extent. Using dwell time as a proxy for engagement risks also measuring people's confusion, at least until we can establish a way for people to signal the difference. Some of the criteria above can only be measured on a large scale, such as how much a given design draws interest.

4.2 Challenges

Some of the criteria above, and the overall question of evaluation, involve some unique challenges (and certainly more than are in this list).

Measurability. In analysis, many aspects of visualization can be measured relatively easily, such as information loss and uncertainty [6], and we have developed a good set of criteria and ways of measuring them. In presentation or communication, we need to first establish what we need to measure, and how we can go about doing that.

Heuristics. Given that many of the criteria listed above cannot be measured easily or at all (at least not directly), it is going to be necessary to develop heuristics and proxies that are reliable, accurate, and understood well.

Audience Size and Sample. Just as with analytical visualization, the types of users that are targeted by explanatory visualization vary widely. Given the less specific and harder to measure criteria to use in this area, it is also likely going to be necessary to conduct much larger studies. How to recruit users for studies, and how to make sure that they are representative of the intended target audience, are open questions at this point.

5. CONCLUSION

When it comes to using visualization for presentation and communication, we are currently either misapplying the wrong criteria when evaluating techniques, or we are unable to evaluate at all. A clearer understanding of the goals and tasks in explanatory visualization is needed, which will require rethinking some of the user models we keep using in visualization. We have tried to argue for the need for these different criteria, listed some of them. We hope that this will pave the way for more work into this very exciting and important area.

6. REFERENCES

[1] S. Bateman, R. Mandryk, C. Gutwin, A. Genest, D. McDine, and C. Brooks. Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts. In *ACM Conference on Human Factors in Computing Systems*, pages 2573–2582, 2010.

[2] R. Borgo, A. Abdul-Rahman, F. Mohamed, P. W. Grant, I. Reppa, and L. Floridi. An Empirical Study on Using Visual Embellishments in Visualization. *IEEE TVCG*, 18(12):2759–2768, Dec. 2012.

[3] W. Cleveland and R. McGill. Graphical perception: The visual decoding of quantitative information on graphical displays of data. *Journal of the Royal Statistical Society.*, 150(3):192–229, 1987.

[4] W. Colin. *Information visualization: perception for design*, volume 5. 2000.

[5] A. Cox. Oil's Roller Coaster Ride. http://nyti.ms/yfZB8X, 2007.

[6] A. Dasgupta and R. Kosara. The importance of tracing data through the visualization pipeline. In *Proceedings of the BELIV Workshop*, page 9. ACM, 2012.

[7] A. Dasgupta, J. Poco, Y. Wei, R. Cook, E. Bertini, and C. T. Silva. Bridging theory with practice: An exploratory study of visualization use and design for climate model comparison. *IEEE TVCG*, 21(9):996–1014, 2015.

[8] N. Duarte. *Resonate: Present Visual Stories that Transform Audiences*. Wiley, 2010.

[9] S. Few. Show me the numbers. *Designing Tables and Graphs to Enlighten*, 2012.

[10] S. Haroz, R. Kosara, and S. Franconeri. The Connected Scatterplot for Presenting Paired Time Series. *IEEE Transactions on Visualization and Computer Graphics*, 22(9):2174–2186, 2015.

[11] D. Huntzinger, W. M. Post, Y. Wei, A. Michalak, et al. North american carbon program (nacp) regional interim synthesis: Terrestrial biospheric model intercomparison. *Ecological Modelling*, 232:144–157, 2012.

[12] R. Kosara. Presentation-Oriented Visualization Techniques. *IEEE Computer Graphics and Applications*, 36(1):80–85, 2016.

[13] R. Kosara and J. Mackinlay. Storytelling: The next step for visualization. *Computer*, 46(5):44–50, 2013.

[14] H. Lam, E. Bertini, P. Isenberg, C. Plaisant, and S. Carpendale. Empirical studies in information visualization: Seven scenarios. *TVCG*, 18(9):1520–1536, 2012.

[15] J. Mackinlay. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics (TOG)*, 5(2):110–141, 1986.

[16] C. North. Toward Measuring Visualization Insight. *Computer Graphics and Applications (CG\&A), Visualization Viewpoints*, 26:6–9, 2006.

[17] L. Wilkinson, A. Anand, and R. Grossman. High-Dimensional Visual Analytics: Interactive Exploration Guided by Pairwise Views of Point Distributions. *Transactions on Visualization and Computer Graphics*, 12(6):1363–1372, 2006.