



Vizdat: A Technology Probe to Understand the Space of Discussion Around Data Visualization on Reddit

JUMANA ALMAHMOUD, Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, USA

DAVID R. KARGER, Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, USA

Visualizations play a considerable role in explaining trends or providing evidence when consuming data online. Whether those visualizations are shared on news outlets or social networks, platforms usually allow readers to discuss their stories in comments sections. For the scope of this work, we studied the online community *r/dataisbeautiful* on Reddit. We found that chart creators were using a variety of authoring tools to share their content. Readers of these posts, *commenters*, used text mainly to discuss and critique the visual content. We noticed a need for a richer mode of communication that would show instead of telling authors what to do. Based on our findings, we introduced *Vizdat*, a lightweight tool and extension to allow users to visualize and reproduce charts in the comments sections of data stories. We used *Vizdat* as a *technology probe* with 11 Reddit users to create data visualization and discuss charts on *r/dataisbeautiful*. During the four-week field deployment period, we observed how *Vizdat* was used and interviewed the participants. We found that *commenters* saw value in accessing the visualization specifications through *Vizdat* and used those charts to structure their replies with richer modalities. As a result, visualization *authors* appreciated the feedback and less toxic discussion provided through comments embedded with modified versions of their charts. In our paper, we share these findings and other insights to understand the dynamics of forum discussion around charts.

CCS Concepts: • **Human-centered computing** → *Collaborative and social computing design and evaluation methods*.

Additional Key Words and Phrases: online discussion, technology probe, visualization, reproducibility, reddit

ACM Reference Format:

Jumana Almahmoud and David R. Karger. 2023. Vizdat: A Technology Probe to Understand the Space of Discussion Around Data Visualization on Reddit. *Proc. ACM Hum.-Comput. Interact.* 7, CSCW1, Article 110 (April 2023), 16 pages. <https://doi.org/10.1145/3579544>

1 INTRODUCTION

Data visualization has become an accessible tool for all types of users, experts, and novices in recent years. Users often share those charts to support their narratives and arguments on many outlets, including social media networks such Reddit or Facebook, as a focal point of the conversation. Users share their charts for many reasons, including support for their arguments or getting design feedback. Regardless of users' goals in sharing charts online, there is always discussion around the chart artifact [21, 24, 27, 34]. According to Heer et al. [22], commenting on data stories on online forums is a way for users to learn more about visualization by exchanging insights drawn from the

Authors' addresses: Jumana Almahmoud, jumanam@mit.edu, Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, Massachusetts, USA; David R. Karger, karger@mit.edu, Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, Massachusetts, USA.

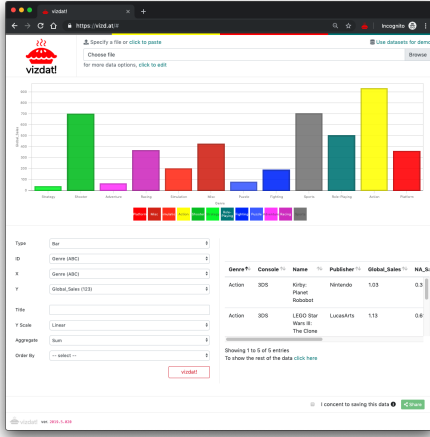


This work is licensed under a Creative Commons Attribution-NonCommercial International 4.0 License.

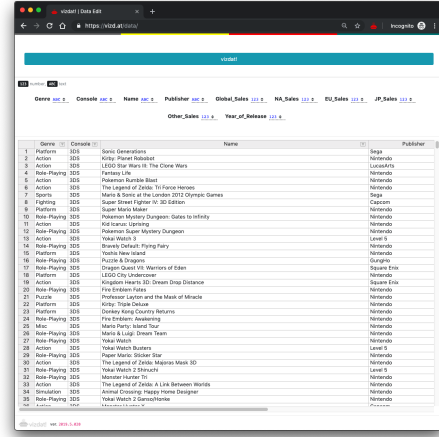
© 2023 Copyright held by the owner/author(s).

2573-0142/2023/4-ART110

<https://doi.org/10.1145/3579544>



(a) Visualization authoring interface



(b) Data set explorer to view and manipulate data

Fig. 1. Web page of the Vizdat probe

data and socializing. Previous attempts to study conversation around visualization include Many Eyes [38], which is a standalone forum that offers publishing and visualization authoring tools. We aim to study discussion where it usually occurs in existing communities rather than a new one that builds around a tool. We emphasize this as it was reported by Danis et al. [17] that Many Eyes was used as an authoring tool and most of the conversations took place on other existing online communities outside the platform, leaving no traces of those valuable discussions on the platform. Prior studies of discussion around visualizations identified key areas of challenge:

- Visualization authors who receive a large number of comments on their data stories need ways to manage the comments and incorporate valuable feedback into their design process [24].
- Users usually discuss one viewpoint of the data set that the author shared and would like more flexibility to explore other views within the discussion context [27].
- Users on social media are presented with a huge number of misleading charts, and there must be a way to identify and fix those charts [13–15, 18, 35].
- While user engagement level varies when interacting with charts[33], higher and more valuable levels of engagement are rare on discussion forums [34].
- Users prefer general and established discussion communities to share their charts. Visualization workflows should support that decentralized model [17].

We hypothesize that improving the current discussion affordances on general social media websites and forums does help in solving the challenges mentioned above. We speculate that enabling users to embed charts in their comments that could easily be modified in-situ and shared within the context of discussion will create a new space of valuable conversations. In order to explore that design space, we focused on r/dataisbeautiful on reddit. We identified key features to help improve the quality of discussion to make it more rewarding for visualization among authors, readers, and *commenters*. We then designed and deployed a technology probe [11, 25] based on the identified features to help facilitate our observations and interviews with community members.

In this paper, we make the following contributions to the CSCW community:

- A Technology probe built based on fieldwork and observations of a specific social community on the web that discusses data visualization
- Empirical evidence of social community use of our technology probe, including their reactions and interactions with each other.
- Design opportunities to support the discourse around information visualization on existing social media sites and forums.

2 RELATED WORK

2.1 Discussing Charts on the Web

One classic work that looked into the discussion around visualizations on the web is Many Eyes [38], a website and forum in which users can author, publish and discuss visualizations. Many Eyes centralized the discussion on their website by giving users a forum to share and comment on visualizations. However, on their follow-up study [17], they have found that most of the discussion around visualization happened offsite, not on their website. Users would use the platform as an authoring tool for their visualizations then share them on their existing and more established communities and blogs. A similar work is Sense.us [22], which is also centralized and uses a blog-style discussion. The main difference between Many Eyes and Sense.us is that the latter was not an authoring tool. Instead, the researchers generated a predefined set of charts and made it open to the public to discuss them.

Wattenberg et al. explored social data analysis using Bartle's taxonomy of players in online games [9] to understand the roles of users in social data analysis of NameVoyager¹. They looked into how these users interacted with each other in conversations around visualization that took place offsite. Two of the personas they discussed that resonate with what we have observed on Reddit are the *Socializers* and *Killers*. Socializers share screenshots of visualizations in their existing social communities for the sake of discussion. Killers are users who ridicule and troll other discussions and users.

More recently, Kauer et al. [27] looked at how users engage with visualization on r/dataisbeautiful and have identified reaction types that include: observation, conclusion, critique, etc., which echoes and complements existing taxonomies proposed by Heer et al. [22] and Hullman et al. [24]. Mcinnis et al. [34] extended the work done by Hullman et al. [24] and revisited discussions that happened on news websites around data on climate change issues. Particularly, they looked into comments that referred to the data, analysis, or visualizations that complement the article. They found that those comments are rare (2.4% of the comments) on those platforms but nevertheless valuable.

2.2 Visualization Authoring Tools for the Web

Looking at the spectrum of visualization tools, we find that tools such as Tableau [6], Lyra [37], Workbench [4], and Chartcultur [36] are for more experienced users. While these tools are more accessible than programming environments such as D3 [12], they still require dedicated time and effort by the user to learn how to use them. A more accessible set of tools for authoring data visualization for the web guides users in the authoring experience through a wizard interface. For example, the experience on Datawrapper [2] first prompts users to choose a data set, describe it in a form, visualize it by picking a type from a predefined list, then publish. Another example is Raw [5], a lighter tool that takes users through a similar linear approach to share images of the visualizations.

Visualization authoring tools allow users to author multiple visualization types such as line charts, maps, or dot plots. Each type of visualization supports a different goal or argument. Blount

¹<https://www.babynamewizard.com/voyager>

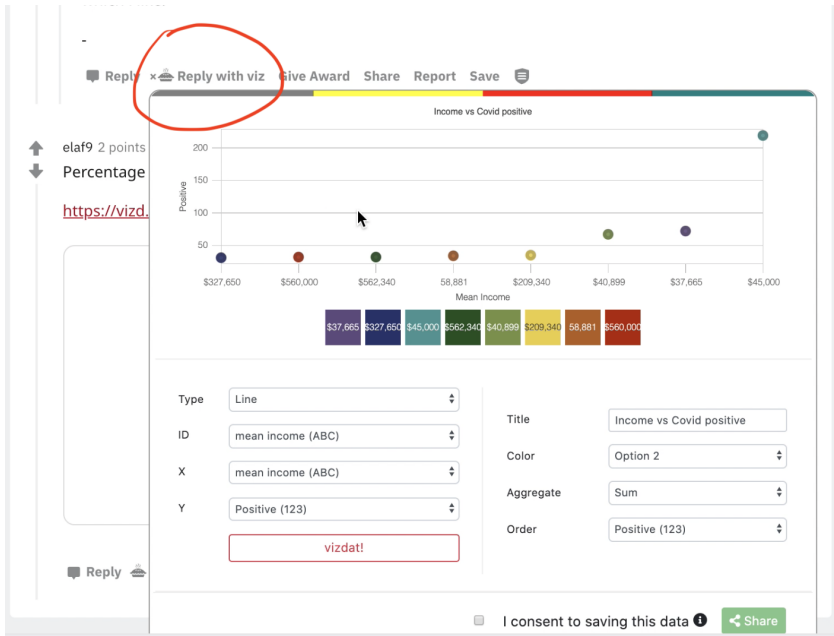


Fig. 2. Vizdat plugin, the same authoring tool as the web page without the data set explorer. Circled is the ‘Reply with viz’ button that is added to posts and comments with charts

et al., [10] observed that the most used patterns in data stories were patterns to support arguments through comparison, repetition, exploration, and animation.

2.3 Technology Probes

A probe, in general, is a tool used to discover information about the unknown with the hopes of returning with valuable information [11]. A technology probe is a special type of probe that incorporates the social science objective of gathering insights on the users and usage of the technology in the wild and the design goal of stimulating users and designers [25].

Technology probes have been used as a helpful design tool in multiple domains. For example, Haldar et al. [20] developed a technology probe to investigate how patients utilize and interpret peer-support technology during their hospital stay, as well as to identify design potential for future peer support technologies. Barbosa et al. [8] used a probe to understand how users interact in real-time with profiles generated about them for targeted ads. Another example is GAMUT by Hohman et al. [23], in which they deployed a probe to explore ways to help increase model interpretability through interactive user interfaces.

3 DESIGN

We started the design process with a formative study to explore how users discuss their shared data visualizations online [26, 29]. We focused on one sub-community on Reddit called r/dataisbeautiful, which is dedicated to sharing data visualizations and the discussion around them with more than 16 million members². We collected data and artifacts such as screenshots of comments, charts and links to resources. We also posted comments and contributed to the community with charts that

²<https://www.reddit.com/r/dataisbeautiful/>

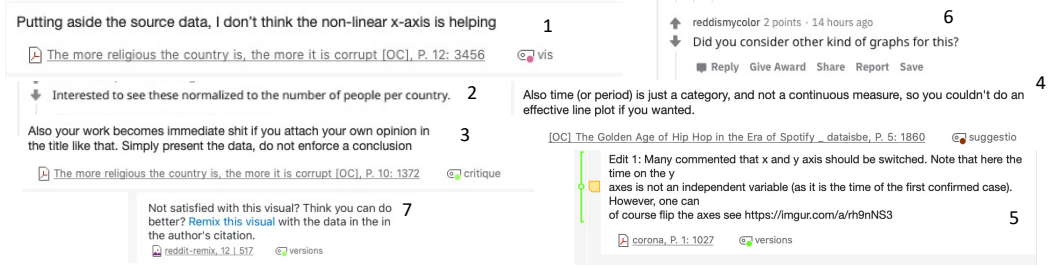


Fig. 3. Example user quotes for need finding: 1. Change Scale, 2. Aggregate, 3.Change title, 4.Change data type, 5.Manage axis 6.Change visualization type, 7.Reproduce visualization

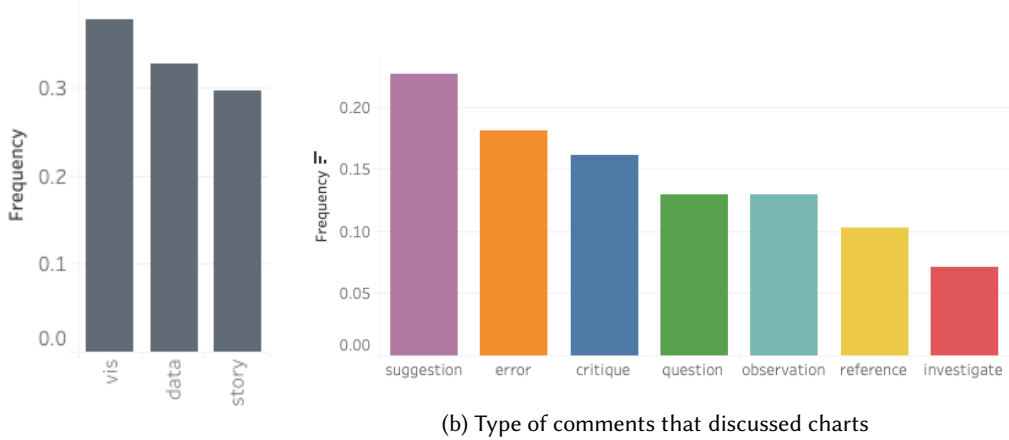
we built. We picked the most popular 27 posts to code with a total of 13181 comments. For the coding process, we started by open coding and identified the main themes through an iterative process. After agreeing on the codes, we defined the themes based on the codes. The themes were split into two categories, one category was around users' needs (Figure 3), and the second category was around the type of comments. (Figure 4). We talk more about the user needs in Section 3.1.

Most comments (90% of the 13181 comments) were irrelevant to our analysis and followed the general discussion themes found on Reddit as reported by Chandrasekharan et al. [16] such as humorous responses, comments that only express gratitude, attacks on other users, or sarcastic comments. To our surprise, the percentage of relevant comments is similar to the one reported by [34]. We were expecting to see more discussion around visualization on this subreddit since it is a specialized community for visualization, as opposed to the data journalism websites looked at in the previous study [34]. Among the 10% relevant comments, we found comments that were special to the visualization forum r/dataisbeautiful and that reflected the taxonomy proposed by Heer et al. [21, 22]. Comments on visualizations were expressing observations, questions about the chart, data, or tools used, criticizing the design, method, or analysis [24]. The general themes were also consistent with what [24] found in their work on data journalism websites. They consisted of comments that addressed the visualization itself, the data, or the story (Figure 4 shows the frequency of codes in our data).

In addition to the trends in comments, we found different types of content creators (authors). Some authors are highly invested; they answer and ask questions in their reply to their audience and explain their process and design decisions. For example, one author replied on his post: *"On purpose, I used a different color palette here since it is not eras but subgenres. Thanks for the feedback - what colors would you suggest?"*. On the other side, some authors were passive and silent, rarely interacting with the reactions of the commenters.

We observed workflows on r/dataisbeatiful that could be improved. Most noticeably, we noticed comments and posts that attempted to *reproduce* charts. For example, an author created a new post to share an updated version of their previously shared chart. The author noted in the comments section of his post: *"I actually posted this here a while ago. I received a lot of suggestions on how to make this interactive dashboard better. Here are the few things I have added ..."* We believe that creating a new post takes the updated version out of context as the discussion and critique happened on another post. Another insightful example is a comment by a user suggesting edits on a visualization post that could be easily demonstrated in-situ instead of being told: *"Could be interesting to see this same chart with the dots sized for GDP per capita."*

Another comment provided a more extensive list of suggestions that could possibly be demonstrated using fewer words: *"Since the title is 'Religion -> Corruption', I think it may be somewhat*



(a) The general themes of discussion around charts

(b) Type of comments that discussed charts

Fig. 4. Frequency of codes

misleading... A better approach might be to plot % non-atheist vs. the reciprocal of the corruption score ("corruption")... Alternatively, you could change the title to say something like "The more atheist a country is, the less corrupt it is"... I would also add the links to your data/code right under the plot". The user suggested edits, including changing the title, the axis, the aggregation of the data and labels, which are simple things that could be easily demonstrated if users had access to the specifications of the visualization.

3.1 Probe Features

We noticed that users needed ways to version and reuse the charts they were discussing. However, we know that Reddit users are there for leisure and socializing most of the time [16] so our probe needed to be convenient. We decided to deploy a probe in the comments section without interrupting users' regular workflows. Given the critical nature of r/dataisbeautiful, we concluded that users needed a way to access, modify and re-share visualization specifications and data when composing a comment on a post with a chart. Based on our need finding we built a probe that allows users to modify the scale, aggregate, manage axis, change viz type, change the data type, change the title, control colors, reproduce visualizations, manage data, add interactivity, and change labels. According to previous work [30], making the data transparent and available in addition to the visualization gives users some agency in co-designing the other perspectives of the data story in the comments section. The probe was designed to work with most platforms with a comments section. It provides a way for users to share links to visualization and render those links as interactive visualizations. For the scope of this study, we optimized the experience to blend in seamlessly in the comments section on Reddit.

4 VIZDAT

Vizdat consists of two main components: Web App (Figure 1)³ and Chrome Extension (Figure 2)⁴. Users create visualizations by uploading their data to the extension or website. The engine then translates the data into a suggested visualization in which users can modify or change by changing the options on the GUI. When users upload a dataset, *Vizdat* detects the type of data and suggests a visualization accordingly. The goal is to help novice users to understand their data by exploring the suggested chart [31, 32]. For the suggested visualizations, we configured Datalib [1] and Jiminy [3].

Through the *Vizdat* Chrome extension (Figure 6), users get to experience more of the in-situ visualization. As in the web app, users can build visualizations using the tool. In addition, the extension detects links to visualizations that were built using the tool, renders them as interactive visualizations by embedding iframes, and attaches a “reply with viz” button to the comment or the post that shared that chart. When users click on the “reply with viz” button for a post or comment, the visualization specs will appear in which users can modify and include in their comments. We store the data along with the visualization configuration to allow users to reproduce visualizations. To make the process even more lightweight, we allow users to send a URL with the data, and the web app will simply return an interactive visualization with a link that could be shared.

4.1 Probe Deployment

We deployed the probe on r/dataisbeautiful by creating posts online using the tool. In addition, we approached users through the comments section by replying to their posts with a *Vizdat* version of their charts and suggested the use of the tool. In all of our interactions, we emphasized that we are part of a research to improve discussion around information visualization. The authors were given the flair “Viz Researcher”⁵ to show their role. We approached an admin within the community and made sure we followed the proper norms of that subreddit.

5 INTERVIEWS

While we have made the probe functional on all discussion platforms with a comments section, for the scope of this paper, we focused our interviews on r/dataisbeautiful. The type of users that we have interviewed on r/dataisbeautiful are content Creators or *authors*, *Commenters* and an admin. Interviews were conducted either through google hangout (video and voice calls) or through the chat service on Reddit. Five of the six commenters we interviewed preferred to communicate through the chat service, and we accommodated their request. Commenters were more private but open when we offered to communicate with them through the textual chat service on Reddit. The interviews took around 60 to 80 minutes through either medium, with more delays when communicating through chat. Participants who were more comfortable with the textual chat offered us descriptive replies to our questions through text, which did not affect the quality or the richness of the data we collected. We rewarded participants with Reddit coins as an appreciation for their time. Each participant agreed on a consent form before participating in the interview and this work was deemed exempt by the local Institutional Review Board.

We interviewed 11 users on Reddit—six commenters, four content creators and one admin. We refer to who actively created posts with their original charts on any subreddit as content creator or *authors*. Users who actively commented on charts on any subreddit without creating their posts are considered *commenters*. The admin was a user who actively posted and commented but was part of the admin team of the subreddit r/dataisbeautiful. The admin was the most experienced leader in

³<https://vizd.at/>

⁴<https://chrome.google.com/webstore/detail/vizdat/oblpmgjlmainfnjjmjfonbjkghknf>

⁵<https://mods.reddithelp.com/hc/en-us/articles/360010541651-User-Flair>

Table 1. Interview Participants. (DiB stands for DataisBeautiful)

User	Years on DiB	Background	Tools used
A1	5	Decent programming and data experience. Student interested improving visualization skills.	R and ggplot
A2	3	No programming or data experience some design experience. Learning skills in data storytelling.	Data Studio, Floursih
A3	8	No Programming experience. some data and design experience. Student and hobbyist.	MS Excel, Photoshop.
A4	2	Advanced programming and data experience. Decent design experience. Data Scientist learning to improve storytelling and visualization skills.	ggplot2
C1	8	Decent programming and data experience.	NA
C2	2	Advanced programming and data experience.	NA
C3	8	Advanced data experience. No programming experience.	NA
C4	7	No programming or data experience. Decent Design experience.	NA
C5	8	No experience or interest in posting, just enjoys the community.	NA
C6	7	No experience or interest in posting, just enjoys the community.	NA
Admin	8	Advance programming and data experience	Python and Pandas

the community and has been a moderator in the subreddit for more than eight years. In addition to his experience in leading the community, he is a visualization researcher and practitioner. We started our interviews with him with an open-ended approach as we wanted to learn more about the community and get directions from the moderator. Given the advanced experience and position this admin holds in this community, he was able to grant us approval to conduct our work in the community and point us to the right resources we needed. With the rich information he provided, the data we collected from him was rich, and we did not need to consult another admin, as our main targeted groups were commenters and authors. We spent the rest of our time interviewing community members who are more relevant to our user group. Those type of users are not mutually exclusive, as some content creators do comment sometimes and so on. Table 1 lists the users who participated on our interviews.

We approached our participants through different ways:

- We posted original content on Reddit of data visualizations created using our probe Vizdat. We approached the *commenters* of those posts.
- We commented on posts that discussed data stories with our version of the chart using Vizdat. We approached the author of that post or users who interacted with our comments.

We sent direct messages to the users who interacted with posts or comments built using vizdat. After getting a response, we asked users for their consent and if they were comfortable with being part of our interviews. We asked participants about their motivation to share charts on Reddit or participate in discussions around charts. We then asked them to reflect on their usage of *Vizdat* and whether they had suggestions to make their discussion around charts more rewarding. Participants

have used vizdat at least one time before the interview. For authors, at least attempt to generate one chart. For commenters, at least trying to recreate one chart. All data is self-reported, and no private data was collected as part of this study. The authors of this study always declared their identity as researchers conducting a study when communicating with users.

6 RESULTS

In this section, we share the results of our interviews with the 11 Reddit users. Using a grounded theory approach, the authors performed a qualitative analysis of the interview transcripts starting with open coding the excerpts through an iterative process. After agreeing on the codes, the authors grouped together the codes to form the following themes:

6.1 Trends and Challenges With Reddit

6.1.1 Sharing visualizations as images or videos is limiting. Visualization authors use many tools to create their charts and export them as images or videos. This freedom of choice creates a challenge when trying to introduce tools or standards to reproduce those charts, as each tool has its own specifications [19]. As the moderator mentioned when discussing the idea of using a version of Vizdat to unify the chart authoring activity:

God, it would be an amazing feature to do that for just general data is beautiful, because people create their visualizations in so many different ways from so many different data sources. - (Admin)

In our interviews with multiple authors, we could see that some are passionate about their workflows and image in the community. Such authors are hesitant to use authoring tools that do not give them more control over the aesthetics and branding of their charts. Those users used the tool in ways that we did not anticipate. They generated charts through Vizdat of their data set and shared them in the source comments, where authors usually provide a link to the data sources and tools used. They provided other users with the visualization specifications and data in a reusable and reproducible format. Still, they maintained their sense of style in the original post (Figures 5 and 6).

While most users authored their charts using applications, some used code to generate their contributions. For those users, workflows such as exposing the code in the comments and rendering in situ work better than a template-based experience. As an experienced *commenter* shared with us: *“Most useful would probably be something like a **Jupyter Notebook that is integrated with Reddit to easily tweak the source code in the comment field and post the plot as the comment**”* - (C4)

6.1.2 The effort spent to author and share visualizations and the effort spent to discuss them. From our conversations with authors and *commenters*, we noticed that the level of effort spent by commenters is small in general. To elaborate, users on Reddit are presented with hundreds of posts from multiple subreddits. When they interact with a post, they rarely look for a deeper kind of interaction in general. For example, commenter (C1) reported that he would use the tool if many constructive comments relied on the reproducible chart in the discussion. However, to them, that kind of conversation was too rare [34] to give up their current practices and workflows: *“So while that seems like it would be the most helpful/useful tool to me [Vizdat], I think I would not use it for a lot my activity on /r/dataisbeautiful since I would guess if you think of it in terms of the 90-9-1% rule that type of commenting would be in the 1% bucket.”*

Commenters suggested more direct ways to comment on visualizations. For example (C6) suggested: *“the most helpful tool for explaining would be having an **in-browser paint program so that someone could doodle over the plot without having to download the image, edit it, and re upload** for people not familiar with whatever language was used to make it.”* Another commenter

(C1) suggested something similar: “Doodle by finger on the existing visual or sketch an alternative visual.”

6.1.3 Users participate to learn from each other. All of the participants we have interviewed mentioned that they engage in r/dataisbeautiful to learn something, whether it is to learn about information visualization, data manipulation, or stories. For example, *commenter* (C3) mentioned: “My comments have always been about layout. It really is my weakness. And I’m always looking for a way to make things look better.” They appreciate a workflow integrated with the discussion that provides them with the flexibility to explore visualization layouts and read other users’ feedback.

As for visualization authors, they also find focused discussions around their data stories valuable in improving their skills: “many of them [comments] are valuable in terms of improvement of the design and wording, the choice of the chart type and what people perceive as “good” and “bad” and what they struggle with when being confronted with a chart in an unfiltered way”- (A4) Focused discussions around charts include expressing other viewpoints, constructive criticism, and observations.

6.2 Affordances of Vizdat

6.2.1 The technique “Show, don’t tell” applies nicely to discussions around information visualizations. Content creators appreciated feedback, especially comments that showed how to improve their charts. Most content creators emphasized that regardless of the toxic nature of feedback on Reddit, they still appreciated the few valuable insights that they gained out of the discussion and wished it was more structured and filtered. The moderator explained to us that they always try to push for more constructive conversations anchored around the visualizations. However, it is hard to implement as the platform expands with its current design: “When dataisbeautiful started taking off and it started becoming less of a niche community and more of a broadly popular community. We tried to push off the expert discussion over the visualization. I guess I could call myself a visualization expert and I still like having that discussion. **I try to push that discussion over to visuals, to the visualization.**” He then explained that a workflow such as the one provided by Vizdat is ideal for supporting their vision of the kind of conversations they wish to see more of on r/dataisbeautiful:

*“I think that’s [Vizdat] a great learning experience. If you explain why a pie chart is bad or especially a racing bar chart is bad for this particular data, and you **see that and learn** from that, even if they’re not experts or even highly competent in the subject, it’s a great learning experience.” - (Admin)*

Authors also appreciated the toned down criticism achieved through showing the changes and sharing them as part of the conversations: “Of course I would like such a tool [vizdat]. Constructive feedback is always welcome, even if sometimes it comes off as a bit rude”- (A3)

6.2.2 The process of generating charts matters in the discussion around them. Users appreciated sharing the process of generating visualization posts. For authors, sharing the process added more substance and validity to their visualizations and analysis [7] which invites more constructive conversations. *Commenters* also appreciated shared processes to learn and discuss the design decisions. For example, one user shared a line graph and had one comment questioning his decision: “Why not a stacked area graph? I find that much visually intuitive especially when you show a 100% sum” - (A4). **The author then replied to that comment with a link to how the visualization looked like as a stacked area chart**, arguing that lines are more intuitive than stacked graphs. The conversation continued with that link as a reference point. In our conversation with the moderator, he expressed appreciation for discussions that revolved around the process with actionable insights that all members in the community could view and learn from: “**Ideal case is where all posts were like that; everything is constructive...** then someone who is really knowledgeable coming in and

saying, hey, here's how you can improve your visualization. And then even better, the original creator comes back and says, OK, I tried that. Here's how it looks. That would be like top notch dataisbeautiful discussion, in my opinion"- (Admin).

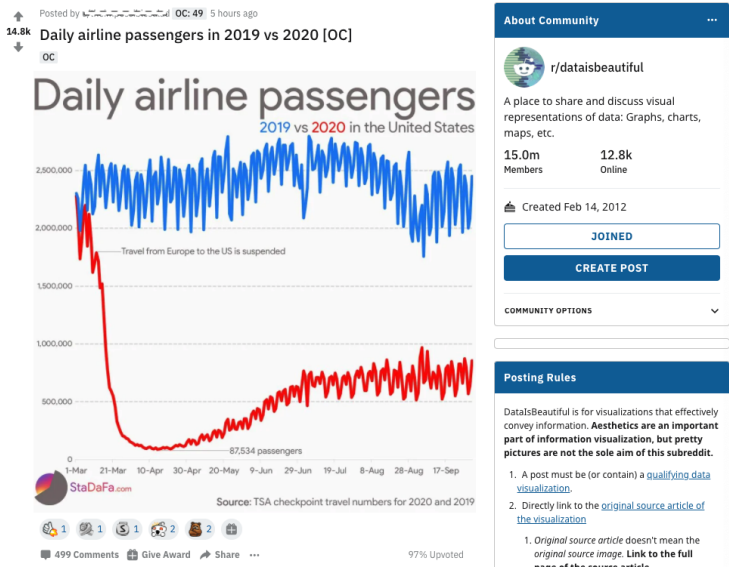


Fig. 5. An example of a chart on r/dataisbeautiful. Author X used Excel and Photoshop

6.2.3 One discussion, one data set, and multiple stories. Dedicated visualization authors have different views and data that they would like to share. They usually add a comment to their post with links to blog posts or images. With *Vizdat*, users attached different presentations of their data set to their comments. For example the author of Figure 6 (A3) explained:

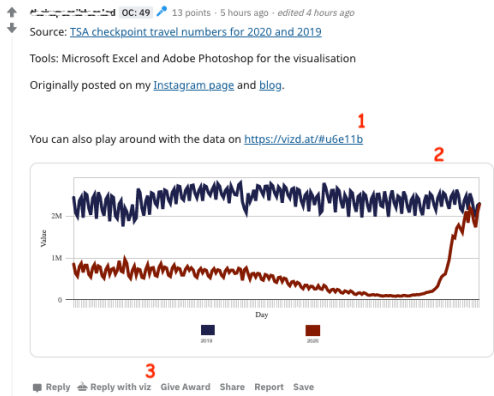


Fig. 6. Author X embedded Vizdat chart to his source comment. 1.Vizdat link to the chart, 2.Using the plugin users can see and interact with the chart, 3.'Reply with viz' button to reply with a modified version of the chart

The reason I am using the tool [Vizdat], I just believe it's worthwhile to provide people with the raw data so they can play around and create their own visualizations.

Currently, users are limited by the submission process on Reddit. Users can only submit an image or video and add comments with additional images or links. In terms of interactions, users can quote and reply. However, users need more efficient ways in the comments section to modify the quoted chart and share that modified chart as part of the conversation. After becoming accustomed to the probe, the moderator saw value in integrating the probe with users' daily experience on r/dataisbeautiful: *"I have not seen anything on Reddit for that [reproducing charts]. It would be amazing if we could create a label that said automatically mix this in."*

He then followed up by explaining the challenges on the platform as discussing visualizations requires more specific support than what already exists on online forums: *"Reddit is by no means perfect. It was designed for a very specific thing. And now we're trying to hack all kinds of stuff on top of it. So that's why we really try to stick to open-source things and things built into Reddit wherever possible. You know, a lot of a lot of people use Chrome and Firefox extensions on top of Reddit. I think that's good."*

We know from our conversations with authors and commenters, that while Reddit has its limitations, they appreciated the large existing community and would rather be where people are than move to dedicated platforms for their visualizations. For example, commenter (C4) mentioned that they enjoy the rich conversation on Reddit in general and appreciate the posts from r/dataisbeautiful. However, he rarely participated because he did not have much to add but enjoyed the idea of a probe that renders more information than text to support the discussion. He explained that he is interested in reading comments that show *"either confusing presentation or [where] data implies some further results that could be more interesting than the OP [original poster]."*

6.3 Suggestions

6.3.1 Sharing reproducible visualization workflows could potentially facilitate a more rewarding discussion around them. In addition to sharing the process of authoring charts, sharing workflows that are reproducible by community members provided a more prosperous environment for constructive conversations. The workflows included the data sources formatted, visualization specifications, and source files. After using the probe, commenter (C1) suggested sharing code instead of a template-based approach: *"Alternatively, having an in-browser IDE where someone familiar with the source language could make edits and have a neat final copy would be nice."* In addition to that, the moderator emphasized the importance of workflows that allow users to remix and share charts by asking authors to add a comment to their posts (source comments- authors' supplementary comments on their posts) with information about their charts:

*if someone sees something cool and they're like that's a great idea. How do you make this? I have an idea how to improve it... **We strongly encourage remakes and reboots of things.** That's why we want to make sure that those **source comments are of high quality** so people can attempt to reproduce a work -(Admin)*

However, there are challenges to this ideal world. According to the moderator, the quality of user-generated source comments determines the possibility of reproducibility of that chart: *"those source comments vary tremendously in quality. Some people spend hours typing up the whole essay describing how they did it and why. Some people are just, source Google, tool Excel. **Practically useless as a source comment.** So we're trying to find good ways to encourage OC [original content] creators to do them and maybe some minimum rules or something like that [Vizdat]."*

6.3.2 Workflows to support authors in managing a large number of comments on their posts. Authors would like to categorize comments in order to reduce redundancy when navigating through

discussion and feedback on their charts. Authors hope that by categorizing comments, users who are commenting become aware of what's been said to say something else. Engaged authors often find themselves repeating their replies to those comments as they do not have a way to reply once to all those comments that ask the same question. Authors would like a way to automate that process and provide them with a filtered view of the comments tagged by type of comment (critique, observation, appreciation etc.) or by the type of critique in the comment (critiquing the presentation vs. the data vs. the analysis vs. the conclusions). Author (A1) suggested: *"Maybe different categories for comments so people can see if someone else has already commented what they're about to comment. I get a lot of the same comments over and over"*. Another author (A4) explained that in addition to managing the comments, they need a way to update their posts accordingly without creating a new post: *"A bit more sorting - often the same topic pops up several times (e.g. "why is rapper xy not included" or "the gray is hard to see". And most importantly for me: **an option to update the chart or pin the updated version below!**"*

6.3.3 Workflows to support commenters when reading and composing comments. Currently, users share their charts in the form of images or videos. The social media platform is not aware of the *content* of those images or videos. *Content* includes data points in the chart, specs of the visualization, including chart type, semantics, conclusions, or trends. Users enjoyed the idea of sharing charts with more details that were provided through the probe. More engaged *commenters* such as learners and expert users suggested views on the forum that provide them with contextualized and synthesized views of the interactions in a particular post: *"A section where it would compare in detail how it's done in the post vs how it can be improved. A section where it would give examples of other posts that deal with the same or a similar subject but are visually better"* - (C4)

To add more meaning to the content shared on the platform, the moderator shared that on r/dataisbeautiful, users can type certain keywords to trigger bots that fill in the comments with pre-defined feedback and suggestions to minimize the effort of *commenters* and share common knowledge of visualization: *"one of our moderators did created a little tool. And not many people know about this. You can type in the comments an exclamation mark then colorblind and an auto moderator will pick up on that and respond to that comment with informative resources around colorblindness. And there's a whole bunch of them that we have."*

7 DISCUSSION

Using *Vizdat* as a technology probe helped us actively engage with the social community that discusses visualization on r/dataisbeautiful. We were able to create and reply to posts and watch how users interacted with them in the wild through the probe. We also learned from community members who used the probe about their current practices and what kinds of solutions they would like for improvement. We designed one probe to elicit information from two types of experiences: the content creation experience and the commenting experience. We believe there is value in sharing the visualization specification in the context of the discussion and not away through a link to another page or tool. The probe worked well in describing the vision to our participants as they were able to see a new button in the comments section that prompts them to "Reply with vis" on posts created using the probe. They were also able to see the visualizations rendered and made interactive on the platform, which is something not available on Reddit.

While the probe was limited compared to other visualization tools, it was functional with multiple features to elicit usage and discussion [25]. Authors understood and appreciated the idea of making their charts reproducible in the platform, so they **used *Vizdat* to generate charts for their source comments (Authors' supplementary comments on their posts)**. On the other hand,

commenters form a larger community. Some commented on the design and wanted more control over the aesthetics to continue using the probe. Others cared more about the underlying data analysis and wanted more features to support their workflows. Regardless, tools for commenters should be close to their usual work flows, requiring the fewest amount of steps and effort from users. Instead of using a button in the comments section, NLP could support the process of commenting by reading and recognizing what the user is saying about the visualizations and translating that text to a modified version of the original visualization.

Visualizations on r/dataisbeautiful are like black boxes. While users see the final outcome, most of the analysis and process are hidden. Koesten et al. [28] discussed practices and channels used to support data sense-making. In addition to the practices they mentioned (code, emails, repositories, slides, etc.), we found that the first comment in any post on r/dataisbeautiful serves as a form of documentation for the visualizations shared. The structure and quality of documentation depended on the original creator of the post, which creates an inconsistent experience. From our conversations with users, we found that there is great value in exposing the details of these visualizations to make them reproducible. Authoring tools such as Tableau do give users the ability to share their workbooks online. **However, the modifications on that chart have to be performed in a different context away from the discussion.** Authoring tools should allow for sharing the code of these charts to enable platforms to parse that content and render it properly to fit into the discussion content, instead of sharing these charts as videos or images.

Finally, the discussions around charts could be in the form of annotations on these charts instead of discussion trees. Doing so helps contextualize and localize the conversations. What often happens in the comments section is that one comment will be upvoted and have most of the interactions, reducing the visibility of other comments about the visuals. The annotations could be in the form of modified versions of the chart addressing the annotated area, creating a network of versions for others to go through while preserving the provenance of these data stories.

8 CONCLUSION AND LIMITATIONS

In this paper, we shared our observations and probe deployment story on r/dataisbeautiful. We found that chart creators were using various authoring tools to share their content, making it challenging to create one tool to parse and reuse those charts. Readers of those posts, the *commenters*, used text mainly to discuss and critique the visual content. We noticed a need for richer modes of communication that show instead of telling authors what to do. Based on our findings, we introduced *Vizdat*, a web tool and extension to allow users to author and reproduce charts in the comments section of data stories. We used *Vizdat* as a technology probe with 11 Reddit users to create data visualization and reply with charts on r/dataisbeautiful. We observed the probe usage in the community and interviewed the participants. We found that *commenters* found value in accessing the visualization specifications to structure their replies with richer modalities by including a modified version of the original chart. *Authors* appreciated the feedback and the healthier discussion provided through comments embedded with modified versions of their charts.

One of the challenges we faced during the probe deployment was recruiting users to interview, given the anonymous nature of Reddit. Some users were willing to talk to us outside the platform through video and audio calls. In contrast, others were hesitant and were more approachable through text chat on the forum, which affected the consistency of the collected data. Another challenge was the authors' adoption. We noticed that most content creators maintain a reputation and branding, and a simple probing tool was not enough. However, the tool helped probe them to

think about the missing parts of such tools and reflect on their current practices. *Vizdat* is a plugin that can work on any platform that has a comments section. We have tested it on blogs, The Upshot, Facebook, and Twitter). However, to scope our contribution, all the analysis and insights were from *r/dataisbeautiful*. We hope to expand our studies to include other communities on Facebook, Twitter, and news outlets to help make our results more general and inclusive.

ACKNOWLEDGMENTS

To the amazing content creators and users on Reddit for their help and useful insights

REFERENCES

- [1] 2019. Datalib: JavaScript Data Utilities. <https://vega.github.io/datalib/>
- [2] 2019. Datawrapper. <https://www.datawrapper.de/index.html>
- [3] 2019. Library for inferring which type of data visualization can be rendered from a JSON dataset, and with which data field(s): Vizzuality/jiminy. <https://github.com/Vizzuality/jiminy> original-date: 2016-03-02T10:22:07Z.
- [4] 2019. Workbench – Data workspaces made simple. <http://workbenchdata.com/>
- [5] 2020. RAWGraphs. <https://rawgraphs.io/>
- [6] 2021. Tableau: Business Intelligence and Analytics Software. <https://www.tableau.com/>
- [7] Jumana Almahmoud, Robert DeLine, and Steven M. Drucker. 2021. How Teams Communicate about the Quality of ML Models: A Case Study at an International Technology Company. *Proceedings of the ACM on Human-Computer Interaction* 5, GROUP (July 2021), 222:1–222:24. <https://doi.org/10.1145/3463934>
- [8] Natã M. Barbosa, Gang Wang, Blase Ur, and Yang Wang. 2021. Who Am I?: A Design Probe Exploring Real-Time Transparency about Online and Offline User Profiling Underlying Targeted Ads. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 5, 3 (Sept. 2021), 1–32. <https://doi.org/10.1145/3478122>
- [9] Richard Bartle. 1996. Hearts, clubs, diamonds, spades: Players who suit MUDs. vol. 1, no. 1 (June 1996). <http://www.mud.co.uk/richard/hcds.htm>
- [10] Tom Blount, Laura Koesten, Yuchen Zhao, and Elena Simperl. 2020. Understanding the Use of Narrative Patterns by Novice Data Storytellers. In *Proceedings of the 4th International Conference on Computer-Human Interaction Research and Applications*. SCITEPRESS - Science and Technology Publications, Budapest, Hungary, 128–138. <https://doi.org/10.5220/0010121601280138>
- [11] Kirsten Boehner, Janet Vertesi, Phoebe Sengers, and Paul Dourish. 2007. How HCI interprets the probes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery, New York, NY, USA, 1077–1086. <https://doi.org/10.1145/1240624.1240789>
- [12] Mike Bostock. 2011. D3.js - Data-Driven Documents. <https://d3js.org/>
- [13] Katy Börner, Adam Maltese, Russell Nelson Balliet, and Joe Heimlich. 2016. Investigating aspects of data visualization literacy using 20 information visualizations and 273 science museum visitors. *Information Visualization* 15, 3 (July 2016), 198–213. <https://doi.org/10.1177/1473871615594652> Publisher: SAGE Publications.
- [14] Alberto Cairo. 2019. *How Charts Lie: Getting Smarter about Visual Information*. W. W. Norton & Company. Google-Books-ID: qP2KdWAAQBAJ.
- [15] André Calero Valdez, Martina Ziefle, and Michael Sedlmair. 2018. Studying Biases in Visualization Research: Framework and Methods. In *Cognitive Biases in Visualizations*, Geoffrey Ellis (Ed.). Springer International Publishing, Cham, 13–27. https://doi.org/10.1007/978-3-319-95831-6_2
- [16] Eshwar Chandrasekharan, Mattia Samory, Shagun Jhaver, Hunter Charvat, Amy Bruckman, Cliff Lampe, Jacob Eisenstein, and Eric Gilbert. 2018. The Internet’s Hidden Rules: An Empirical Study of Reddit Norm Violations at Micro, Meso, and Macro Scales. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (Nov. 2018), 32:1–32:25. <https://doi.org/10.1145/3274301>
- [17] Catalina M. Danis, Fernanda B. Viegas, Martin Wattenberg, and Jesse Kriss. 2008. Your place or mine?: visualization as a community component. In *Proceeding of the twenty-sixth annual CHI conference on Human factors in computing systems - CHI '08*. ACM Press, Florence, Italy, 275. <https://doi.org/10.1145/1357054.1357102>
- [18] Kirstin Fontichiaro and Jo Angela Oehrli. 2016. Why Data Literacy Matters. *Knowledge Quest* 44, 5 (2016), 21–27. <https://eric.ed.gov/?id=EJ1099487> Publisher: American Association of School Librarians.
- [19] Arpita Ghosh and Jon Kleinberg. 2013. Incentivizing participation in online forums for education. In *Proceedings of the fourteenth ACM conference on Electronic commerce (EC '13)*. Association for Computing Machinery, New York, NY, USA, 525–542. <https://doi.org/10.1145/2482540.2482587>
- [20] Shefali Haldar, Yoojung Kim, Sonali R. Mishra, Andrea L. Hartzler, Ari H. Pollack, and Wanda Pratt. 2020. The Patient Advice System: A Technology Probe Study to Enable Peer Support in the Hospital. *Proceedings of the ACM on*

- Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 112:1–112:23. <https://doi.org/10.1145/3415183>
- [21] Jeffrey Heer. 2006. Socializing Visualization. *Social Visualization Workshop, CHI* (2006), 10.
 - [22] Jeffrey Heer, Fernanda B. Viégas, and Martin Wattenberg. 2007. Voyagers and voyeurs: supporting asynchronous collaborative information visualization. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*. ACM Press, San Jose, California, USA, 1029. <https://doi.org/10.1145/1240624.1240781>
 - [23] Fred Hohman, Andrew Head, Rich Caruana, Robert DeLine, and Steven M. Drucker. 2019. Gamut: A Design Probe to Understand How Data Scientists Understand Machine Learning Models. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (CHI '19)*. ACM, New York, NY, USA, 579:1–579:13. <https://doi.org/10.1145/3290605.3300809> event-place: Glasgow, Scotland Uk.
 - [24] Jessica Hullman, Nicholas Diakopoulos, Elaheh Momeni, and Eytan Adar. 2015. Content, Context, and Critique: Commenting on a Data Visualization Blog. In *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing - CSCW '15*. ACM Press, Vancouver, BC, Canada, 1170–1175. <https://doi.org/10.1145/2675133.2675207>
 - [25] Hilary Hutchinson, Wendy Mackay, Bo Westerlund, Benjamin B. Bederson, Allison Druin, Catherine Plaisant, Michel Beaudouin-Lafon, Stéphane Conversy, Helen Evans, Heiko Hansen, Nicolas Roussel, and Björn Eiderbäck. 2003. Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. Association for Computing Machinery, New York, NY, USA, 17–24. <https://doi.org/10.1145/642611.642616>
 - [26] Danny L. Jorgensen. 2015. Participant Observation. In *Emerging Trends in the Social and Behavioral Sciences*. American Cancer Society, 1–15. <https://doi.org/10.1002/9781118900772.etrds0247> _eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/9781118900772.etrds0247>.
 - [27] Tobias Kauer, Arran Ridley, Marian Dörk, and Benjamin Bach. 2021. The Public Life of Data: Investigating Reactions to Visualizations on Reddit. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (May 2021), 1–12. <https://doi.org/10.1145/3411764.3445720> arXiv: 2103.08525.
 - [28] Laura Koesten, Kathleen Gregory, Paul Groth, and Elena Simperl. 2021. Talking datasets – Understanding data sensemaking behaviours. *International Journal of Human-Computer Studies* 146 (Feb. 2021), 102562. <https://doi.org/10.1016/j.ijhcs.2020.102562>
 - [29] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2010. *Research Methods in Human-Computer Interaction*. Wiley Publishing.
 - [30] David Lloyd and Jason Dykes. 2011. Human-Centered Approaches in Geovisualization Design: Investigating Multiple Methods Through a Long-Term Case Study. *IEEE Transactions on Visualization and Computer Graphics* 17, 12 (Dec. 2011), 2498–2507. <https://doi.org/10.1109/TVCG.2011.209> Conference Name: IEEE Transactions on Visualization and Computer Graphics.
 - [31] Jock Mackinlay. 1986. Automating the design of graphical presentations of relational information. *ACM Transactions on Graphics* 5, 2 (April 1986), 110–141. <https://doi.org/10.1145/22949.22950>
 - [32] Jock Mackinlay, Pat Hanrahan, and Chris Stolte. 2007. Show Me: Automatic Presentation for Visual Analysis. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (Nov. 2007), 1137–1144. <https://doi.org/10.1109/TVCG.2007.70594>
 - [33] Narges Mahyar, Sung-Hee Kim, and Bum Chul Kwon. 2018. Towards a Taxonomy for Evaluating User Engagement in Information Visualization. *Workshop on Personal Visualization: Exploring Everyday Life* Vol. 3. 2. (2018), 4.
 - [34] Brian James McInnis, Lu Sun, Jungwon Shin, and Steven P. Dow. 2020. Rare, but Valuable: Understanding Data-centered Talk in News Website Comment Sections. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (Oct. 2020), 174:1–174:27. <https://doi.org/10.1145/3415245>
 - [35] Paul Parsons. 2018. Promoting Representational Fluency for Cognitive Bias Mitigation in Information Visualization. In *Cognitive Biases in Visualizations*, Geoffrey Ellis (Ed.). Springer International Publishing, Cham, 137–147. https://doi.org/10.1007/978-3-319-95831-6_10
 - [36] Donghao Ren, Bongshin Lee, and Matthew Brehmer. 2019. Charticulator: Interactive Construction of Bespoke Chart Layouts. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (Jan. 2019), 789–799. <https://doi.org/10.1109/TVCG.2018.2865158>
 - [37] Arvind Satyanarayan and Jeffrey Heer. 2014. Lyra: An Interactive Visualization Design Environment: Lyra: An Interactive Visualization Design Environment. *Computer Graphics Forum* 33, 3 (June 2014), 351–360. <https://doi.org/10.1111/cgf.12391>
 - [38] Fernanda B. Viegas, Martin Wattenberg, Frank van Ham, Jesse Kriss, and Matt McKeon. 2007. ManyEyes: a Site for Visualization at Internet Scale. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (Nov. 2007), 1121–1128. <https://doi.org/10.1109/TVCG.2007.70577>

Received January 2022; revised July 2022; accepted November 2022