

THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

The Public Life of Data: Investigating Reactions to Visualizations on Reddit

Citation for published version:
Kauer, T, Ridley, A, Dörk, M & Bach, B 2021, The Public Life of Data: Investigating Reactions to Visualizations on Reddit. in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21).*, 612, ACM Association for Computing Machinery, The ACM CHI Conference on Human Factors in Computing Systems 2021, Virtual Conference, Japan, 8/05/21.
https://doi.org/10.1145/3411764.3445720

Digital Object Identifier (DOI):
10.1145/3411764.3445720

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

OPEN ACCESS

Download date: 29. Mar. 2024

The Public Life of Data: Investigating Reactions to Visualizations on Reddit

Tobias Kauer
Design Informatics,
University of Edinburgh, UK
Urban Complexity Lab,
University of Applied Sciences Potsdam, DE
t.kauer@ed.ac.uk

Arran Ridley
Urban Complexity Lab,
University of Applied Sciences Potsdam, DE
arran.ridley@fh-potsdam.de

Marian Dörk
Urban Complexity Lab,
University of Applied Sciences Potsdam, DE
marian.doerk@fh-potsdam.de

Benjamin Bach
Design Informatics,
University of Edinburgh, UK
bbach@ed.ac.uk

ABSTRACT

This research investigates how people engage with data visualizations when commenting on the social platform Reddit. There has been considerable research on collaborative sensemaking with visualizations and the personal relation of people with data. Yet, little is known about how public audiences without specific expertise and shared incentives openly express their thoughts, feelings, and insights in response to data visualizations. Motivated by the extensive social exchange around visualizations in online communities, this research examines characteristics and motivations of people's reactions to posts featuring visualizations. Following a Grounded Theory approach, we study 475 reactions from the `/r/dataisbeautiful` community, identify ten distinguishable reaction types, and consider their contribution to the discourse. A follow-up survey with 168 Reddit users clarified their intentions to react. Our results help understand the role of personal perspectives on data and inform future interfaces that integrate audience reactions into visualizations to foster a public discourse about data.

KEYWORDS

data visualization, reddit, personal data

ACM Reference Format:
Tobias Kauer, Arran Ridley, Marian Dörk, and Benjamin Bach. 2021. The Public Life of Data: Investigating Reactions to Visualizations on Reddit. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3411764.3445720

1 INTRODUCTION

The growing familiarity with data visualization and the emergence of easily accessible visualization authoring tools is fundamentally changing the ways in which visualizations are created, shared, and used. Today, visualizations are published through a wide range of channels, including news articles [28], social media [29], discussion websites, blogs, reports, government websites,[1] and museums [13]. With this development, visualizations reach an ever widening and increasingly diverse range of audiences with their individual opinions, perspectives, and understandings. Learning how these audience engage with visualizations is key to understand how visualizations are used, how they can cause harm, and how they can contribute to productive discussions and informed decision-making.

Social data-analysis [39] describes the process of collaborative sensemaking with visualizations in online spaces. Many tools and platforms have been created for social data analysis, allowing for annotations, comments, and discussion threads [e.g., 11, 36, 37, 39, 40]. Consequently, some investigations [8, 11, 14, 39] have started reporting on peoples' comments and engagement with these platforms, reporting on how user comments refer to context, express criticism, and create a shared space for insights. While collaborative sensemaking is often associated with domain expertise and a commonly shared goal to explore data and generate insights [e.g., 11, 37, 40], much engagement with visualization is happening without specialized or shared domain expertise or common goal.

Here, we are interested in how these 'general audiences' engage with visualizations online. As our study suggests, engagement of these non-expert audiences can be very different and include motivations beyond (collaborative) sensemaking such as entertainment [30]. There is much more to learn about how people engage with visualization in the broadest sense, including aspects such as critique [14], emotions [19], personal experience [26] and expertise, and potentially visualization literacy. A better understanding of these social practices can inform the design of a future generation of annotation and collaboration tools as well as methods to allow larger and more diverse audiences to express their viewpoints and engage in collaborative exchanges and decision making with visualization.

To widen our understanding of how people engage with data visualizations in the wild, our paper analyzes user comments on the Reddit forum `/r/dataisbeautiful`. We study community members' expressions and introduce the term *reaction* to describe these

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
CHI '21, May 8–13, 2021, Yokohama, Japan
© 2021 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445720

[1] https://www.ons.gov.uk

expressions; we define a reaction as *any verbal or textual expression upon viewing, reading, and/or interacting with a data visualization.* Through iterative, multi-person coding of 475 comments, based on principles of Grounded Theory, we obtained a classification of 10 distinct reactions types: *observations, hypotheses, opinions, conclusions, clarifications, proposals, critiques, additional information, testimonies*, and *jokes* (Section 4). Reactions were found to point to four main *scopes* around these visualizations: *visual representation, data and data analysis, topic*, and *insights*.

These types and scopes provide us with a structured framework to analyze user's engagement described by our reactions. Our findings highlight a broad and diverse range of engagement types (reactions) and their respective subflavors, complementing prior research [14, 39]. An online survey with 168 posters of comments from our collection delivers additional information about the audience's demographic and individual motivations (Section 5). We found motivations to include both personal (intrinsic) goals and extrinsic goals to actively contribute to a public discourse around data. Finally, we discuss how our findings can inform the design of future generations of tools for collaborative and public discourses based on data visualizations (Section 6).

2 BACKGROUND

This research builds on prior studies of engagement with data visualization and collaborative sensemaking as well as recent research on personal data practices and user annotations.

2.1 Engagement with Data Visualization

The different aspects of the relationship between a visualization and its audience are often discussed in visualization as *engagement*, a term that however lacks a firm definition. For example, Haroz et al. [10] refer to engagement as a measurement of effectiveness when interacting with visualization, alongside memorability or performance. Other sources describe engagement as a prerequisite for learning, ranging from *not-viewing* over *responding* to *presenting* [25]. Similarly, Mahyar et al. describe engagement by the means of Bloom's taxonomy of learning goals, ranging from *low* to *high* [23]; *low* has been attributed to viewing and simple interaction, while *high* has been used to describe higher-level cognitive tasks such as hypothesis testing and decision making. Similar notions to describe engagement refer to *"the [whole] processes of looking, reading, interpreting and thinking that take place when people cast their eyes on data visualizations and try to make sense of them."* [20].

From another perspective, engagement has been used to describe how visualizations can get citizens informed about and involved in social issues [5]. In that sense, engagement is a process that may be triggered by a visualization, but eventually happens outside of the visualization, e.g., in the form of social change and advocacy, rather than sensemaking and interaction. Often such engagement will include other persons. Against the background of co-located and in-person conversations about visualizations, Walny et al. [37] provide a crucial differentiation between *direct engagement*, which is *"characterized by a visitor's observable interactions with the space, artifacts, and people"*, and *data engagement*, defined *"by the extent to which a visitor connects with [...] the data behind the visualizations"* [37].

Developing a framework of different reaction types can Some user engagement can be measured more easily, e.g,. from interaction logs, questionnaires, and observations [3, 15, 32, 37]. For example, prior studies [32, 33, 38] report on observations on how general audiences engage with visualizations in co-located and real-world settings. However, rather than aiming to collecting a wide range of reactions, these studies focus on understanding positive and negative impressions towards a designed artefact such as an interactive table top application, a data comic, or a polling display. Data engagement and engagement beyond the artifact, on the other side, is hard to measure and assess. By relying on comments that users provide on the web, we explore a specific resource to understand engagement, while at the same time being objective and neutral about the type of engagement. Our reaction types are meant to broaden our understanding of engagement and can be seen as a specific form of *visualization engagement* with data and visualization, highlighting forms of *data engagement*.

2.2 Collaborative Sensemaking

To support engagement with visualization beyond viewing and exploration, a range of tools and interfaces have been built, supporting *social data analysis* [39] and *collaborative visualization* [17]. These tools provide the audience with annotation capabilities ranging from free form annotation (e.g., text, pen-annotations) to specific pointers or recommending annotations [8]. For example, CommentSpace [40] introduced a fixed set of tags (e.g., *question, hypothesis, to-do, evidence-for, evidence-against, etc.*) for authors to classify their own comments.

Comments can also be linked to specific data points, providing evidence-for or evidence-against the given claims. Building on the idea of *anchored conversations* [2], the sense.us platform by Heer et al. [11] demonstrates a mechanism for linking comments with specific parts and states of a visualization, an idea recently extended by PixelClipper [37]. The visualization community platform ManyEyes [36] allowed users to upload data, create visualizations, share the visualizations online, and exchange comments with other community members. Finally, ChartAccent provides annotation functionality to support storytelling, available only to the authors but not to collaborators and audiences [27]. In general, most of these tools, except ManyEyes, target specific groups of people, with specific tasks, and a common goal of sensemaking but little is known what people think beyond the features of these tools. To not introduce any functional bias into our research, we look only at textual comments while seeking to study an audience as large and diverse as possible.

2.3 Studying User Annotations

User annotations and comments have been studied in controlled user setups with both, business intelligence experts [8] and non-expert audiences [34, 35]. An analysis of 300 annotations sourced though a lab study [34, 35], identified six dimensions of comments (*insight on data, multiple observations, data units, level of interpretation, co-reference, detected patterns*). However, these studies investigated annotation features with a goal of collaborative sensemaking or evaluated specific commenting features of the designed tools. Moreover, as a controlled laboratory setup, these study results do not

capture intrinsic motivations and reactions, such as personal traits, or expressions of opinions or feelings.

Wattenberg [39] reports on user comments made across a variety of social media while studying reactions to his BabyName Voyager. He points out the benefit of being able to point to evidence in a visualization when sharing one's comment and notes that a personal perspective can act as a simple entry point for people to engage with a visualization.

To the best of our knowledge there are only two systematic studies investigating reactions to visualization in online communities. Analyzing comments on a visual news website, Hullman et al. [14] found that *"comments drew on personal knowledge, opinions, experiences, or narratives"*. Analyzing comments on their Sense.us visualization commenting platform, Heer et al.[11] present a classification of annotations. However, annotations were created in controlled studies as well as intranets, while our work considers a vast public participation. Their classification eventually focuses on sensemaking categories such as *observations, hypotheses, data integrity*, and social feedback (e.g., *socializing*, *affirmation*). In contrast, our work identified additional types of reactions such as critique and personal relations to data. Thus, we expand on these observations and devise a clear typology of reactions alongside a deeper analysis of personal drivers for engagement and public discourse.

2.4 Personal Relations with Data

A person's particular situation, knowledge, and experience is highly influential when engaging with a data visualization [19, 20]. People's cultural and social context as well as the emotional condition play a considerable role when making sense of data. Any type of engagement can be related to current affairs and highly personal as reported by Peck et al.[26]; for example, well-established visualization tasks (e.g., information extraction, hypothesis-building) seem to be overshadowed by biographical or personal factors; people would not trust data that is sourced from institutions with a political identity different from their own—no matter the topic or insight. In other cases, people would deem charts more useful if they depicted a topic they had first-hand experience with—independent of the extracted information. In this paper we investigate the notion that *data is personal* and want to understand how personal relations to data are expressed in public settings.

3 DATA & METHODOLOGY

For our study, we chose to investigate Reddit for its abundance and variety of data visualization posts, the lively discussions in response to them, and the ready availability of the data for qualitative analysis. Reddit has 430 million monthly active users (*redditors*)[2]—for comparison, Twitter has 336 million[3]—who share media, links or self-created content in groups (*subreddits*) that focus on a broad range of cultural, political, or social topics. It has been conceptualized as a public sphere in which people engage collaboratively to express and form opinions [6]. Contrary to other social platforms like Twitter or Facebook, Reddit users remain semi-anonymous, do not form networks [21], and do not have "an a priori built social network" [24]. Posts are generally public and can be commented on in hierarchical threads by anybody. Posts and reactions to them receive a score by the audience (*karma*), making reddit "the internet's largest social voting community" [9].

`blockbuster`	Blockbuster Video US store locations between 1986 and 2019
`breaking-bad`	The absolute quality of Breaking Bad.
`bushfire`	Area of land burnt in Australia and area of smoke coverage shown as equivalent area over Europe
`cases`	COVID-19 spread from January 23 through March 14th. (Multiple people independently told me to post this here)
`daylight`	Hours of daylight as a function of day of the year and latitude
`deaths`	Coronavirus Deaths vs Other Epidemics From Day of First Death (Since 2000)
`donations`	Top 10 Highest Covid-19 donations with the percentage of their net worth
`employment`	To show just how insane this week's unemployment numbers are, I animated initial unemployment insurance claims from 1967 until now. These numbers are just astonishing.
`garbage`	The Great Pacific Garbage Patch
`golf`	Hardest Golfing President
`google`	Trending Google Searches by State Between 2018 and 2020
`IMDb`	Worst Episode Ever? The Most Commonly Rated Shows on IMDb and Their Lowest Rated Episodes
`minecraft`	Where is each ore found in a minecraft world?
`net-worth`	2020 Presidential Candidates by Net Worth
`oil`	US oil price
`ramsey`	What percent Gordon Ramsay's Kitchen Nightmares are still open?
`rates`	This chart comparing infection rates between Italy and the US
`stimulus`	Where the money goes in the US Senate's $2T coronavirus stimulus bill
`tuition`	US College Tuition & Fees vs. Overall Inflation
`tulsa`	Attendance at Donald Trump's rally in Tulsa, compared to the number of tickets Trump claimed were requested.

Table 1: Original titles of the visualization posts from which the reactions were obtained. We use short tags to identify the visualizations throughout this paper.

Our study investigates visualization reactions expressed on the visualization-focused subreddit `/r/dataisbeautiful`. What kind of content is allowed in this community is specified by a set of rules, stating that all posts must contain a data visualization that has at least one computer generated element. The rules also enforce a standard by sanctioning sensationalized headlines (*"Clickbait posts will be removed"*) and encouraging the transparent display of the data sources. An additional set of rules is aimed at the audience to establish a code-of-conduct, inviting constructive feedback to the presented visualization and prohibiting intentionally rude or hateful comments. To extract data and content from Reddit, we use the official API through the Python library PRAW.[4]

3.1 Data Selection

To find a suitable subset of comments for this study, we followed the following five-step procedure:

(1) **Pilot studies:** We examined single visualization posts and their reactions and tested preliminary coding schemes on small numbers of them. In these pilots, we observed that some posts, e.g. visualization of quantified-self data and humorous posts, attract many inappropriate or off-topic reactions.
(2) **Visualization selection:** We included the 26 most popular visualization posts according to their *karma* score over the span of one year (Oct-2019 - Sept-2020) in our study. From these 26 posts, we manually removed six posts that mostly attracted off-topic reactions and did not consider them for data collection.

[2] https://archive.is/w04Nx
[3] https://www.statista.com/statistics/282087/number-of-monthly-active-twitter-users/
[4] https://praw.readthedocs.io/en/latest/index.html

(3) **Data Collection:** We used the reddit API to collect 68.000 reactions from the remaining 20 visualization posts.
(4) **Filtering:** We filtered these reactions based on the following four criteria and ended up with a set of 792 reactions.
   - As we intended to look only at direct reactions to the visualizations, we only included first-level comments but no replies to any comments.
   - We ignored comments by the visualization author and by automated bots, identifiable through their username.
   - We ignored comments that were deleted by the reaction authors or removed by moderators.
   - We set a threshold of five upvotes for each comment to be considered, to remove comments off topic, illegible, or otherwise irrelevant.
(5) **Sample Sizing:** Since the number of reactions differed along the 20 posts, we randomly selected 25 reactions per visualization. Since some visualizations had less than 25 reactions, our final dataset contains 475 reactions.

Generally speaking, the visualizations we considered use common and well-known techniques such as barcharts, linecharts, or heatmaps. Nine visualizations are animated, the other eleven are static images. While the range of depicted topics is broad, we found several visualizations about CoViD-19, the 2020 presidential election or pop-cultural data like IMDb-ratings. Table 1 shows the list of visualizations included in our study, with titles and clickable links.

3.2 Open Coding

To obtain a high-level understanding of reactions, we engaged in open and close coding following Grounded Theory [31]. Our goal was to find independent and non-overlapping concepts in the data, able to describe the range of users' reactions.

Our coding scheme developed over eight major iterations. Each iteration involved between two to four of the authors and refined both the set of comments and visualizations as well as our coding scheme. We started our open coding by independently coding a random set of comments. We then build a first scheme with four dimensions: *i)* types of reactions, e.g., observation, critique, clarification, etc, *ii)* a user's stance (negative, neutral, positive), *iii)* their conviction (weak, neutral, strong), and *iv)* any presented evidence (e.g., external, anecdotal).

Throughout the early iterations, we observed that many of these dimensions depended very much on the subjective interpretations of the coders. Through several iterations of close coding, we removed dimensions and categories that led to misleading coding. Each new iteration was tested with new set of randomly selected comments to prevent us from fitting our scheme to a specific set of comments. In the final coding scheme, 20% of the data was coded by all four authors and discussed to reach consensus. The four authors independently used this scheme to categorize comments and test, whether the assumed categories could exhaustively and without overlap characterize the actual data. We merged categories that were hard to discern from our data (e.g., observations and interpretations) and added new categories where comments did not fit any of the existing categories or to clarify important differences among comments within a category.

Throughout coding, we observed that many comments included notions about the commenter's personal perspectives. While we started out with an individual reaction type 'personal reference' many of these comments overlapped with other reaction types. We thus introduced an optional *personal* flag to tease out the role of personal qualifications in the reactions. Many of these personal comments included subtle expressions as first-person pronouns (I, me, my) or indications of experience (*as someone who [...]*), to articulate personal positions (*in my opinion [...]*).

3.3 Final Scheme

After our 8 major iterations, we converged towards a coding scheme with two dimensions: Ten *reaction types* (Table 2) and four *reaction scopes* (Figure 2). While we started out with the intention to only allow one type per reaction, we accepted that comments are often more nuanced and ambiguous to be classified as one type only. We attribute this to the very nature of human thought and expression. While we tried to minimize overlap between types of reactions as much as possible, we accepted some comments to contain two reaction types, rather than eliminating overlap.

3.4 Online Survey

We conducted a survey and contacted the authors of all 475 comments, 168 participants (35%) of which completed it. In the survey questionnaire, we asked for basic demographic information and compared them to the overall average of the Reddit user base, reported on.[5] 10% of the participants in our survey were not cis-male (overall on Reddit: 33%). Our two most prominent age groups were 19-29 year old (survey: 45%, overall: 50%) and 30-39 year olds (survey: 45%; overall: 9%) The participants lived most of their lives in North America (58%) or Europe (28%). 39% had a Bachelor's degree, 25% had a Master's degree, 24% had a high school diploma. 45% stated that they had some formal training with data, statistics or visualization. We asked the participants about their motivation for their reaction and report back on the results in Sections 4 and 5.

4 VISUALIZATION REACTIONS

This section presents ten reaction types and four scopes. The reaction types describe the kind of contributions that people make when responding to a visualization. The scopes describe which aspect of a visualization the comment refers to. In the following, we start by explaining scopes and then present the reaction types. We include verbatim examples of reactions and show what visualization they refer to in a `[tag]`. We complement the results of our analysis with the motivations the posters shared in the survey. Direct quotes from the survey are indicated through a *background color*.

4.1 Scopes

Akin to Hullman et al.'s [14] targets, scopes denote the part of a visualization post a commenter refers to. Since posts on `/r/dataisbeautiful` are structurally different from comments on a news blog and do not come with an accompanying article, we encountered four slightly different scopes:

[5] https://www.pewresearch.org/internet/2013/07/03/6-of-online-adults-are-reddit-users



Figure 1: Examples of the 20 popular visualizations we obtained from `/r/dataisbeautiful`. In total we considered 475 reactions. Visualizations varied in their topics, aesthetic appeal, visualization techniques, and the use of animations.

Data—This scope describes reactions about the dataset. Comments referred to data quality and the ability to inform (*"What is the composition of the large corporation other? Loans? Grants? Tax credits?"* [STIMULUS]), data collection methods (*"do you mind discussing how you parsed the world file?"* [MINECRAFT]), or the accuracy of data (*"The swine flu killed at least 18000 people in the end. I find this quite misleading"* [DEATHS]).

Visualization—Comments with this scope refer to the visual representation and its creation. Comments referred to representation aesthetics (*"Great Graphic"* [STIMULUS]), visual variables (*"Any meaning to the different colours?"* [CASES]), or other design decisions (*"This map is using the Mercator projection, which can be a bit deceptive"* [BUSHFIRE]).

Insight—Whenever a comment referred to information that can be extracted from the visualization, we call this an 'insight'. Comments referred to observed data points or trends (*"It really revved up in the last 4 days."* [CASES], raised questions about a finding (*"We can see where it goes.... but... where does it come from?"* [STIMULUS]), or any other contemplation about the presented information (*"Why is there a seasonal trend to this? What month does that happen and any idea why?"* [EMPLOYMENT]).

Topic—Comments with this scope referred to the topic illustrated by the visualization, rather than any specific insight. Generally, these comments were found to be independent from the specific information contained in a visualization: *"We had so much time to start quarantining people travelling from China. We all blew it."* [CASES].

The majority of the 475 comments referred to either an insight (188) or to the topic (187) (Figure 2). This supports Kennedy et al.'s [20] findings that a visualization's subject matter is an important driver for engagement—people engage with topics they are interested in. 117 reactions mainly referred to the data source, and 82 talked about the visual representation of that data.

4.2 Reaction Types

We now report on the ten reaction types that we identified. Each type indicates a distinct contribution that a comment makes in response to a data visualization. We provide a brief description, the frequency, and example reactions for each. We also discuss potential subgroups that are characterized by similar phrases or lines of argumentation. Table 2 provides an overview of the reaction types.

Observations point out insights that were made directly in the visualization or interpretations derived from its data: *"wow, House episodes rating are incredible consistent"* [IMDB]. These reactions consist of individual, often simple observations of outliers (*"May 24th, when Minnesota became the center of the world for a little bit."* [GOOGLE]) and trends (*"Wow that's crazy I didn't think a company could expand that quickly. And then to see it disappear as quickly is crazy"* [BLOCKBUSTER]). Often, observations are a necessary prerequisite for other reactions, such as opinions and conclusions. Their scopes are predominantly insights, yet they are not merely about expressing what a person took away from a chart. Observations are often expressed in a humorous way and combine specific data points or trends with a joke: *"That line dips any lower it'll hit oil"* [OIL]. We also found a number of observational reactions that include subjective or personal expressions, such as *"I like"*, *"I'm surprised"*, *"I'm particularly amused by"*, *"Fascinating that"*, and *"Wow, that is [...]"*. These expressions reveal genuine moments of surprise associated with finding something interesting and relevant. Moreover, they show people's personal stance with respect to that observation. We flagged these observations as *personal* for subsequent analysis.

Conclusions take a step back and reflect on the observed information and its consequences. One characteristic group of conclusive reactions simply summarize the presented information and refer to

Type	Description	#
OBSERVATION	Stating the obvious; point out specific data points or trends; express insights and interpretations	62
CONCLUSION	Taking a step back to provide an overview over the presented information and draw conclusions or call to action	29
HYPOTHESIS	Providing a speculative explanation to explain the presented information. This can range from tentatively asked questions to bold claims that introduce additional data to argue for an explanation	56
CLARIFICATION	Asking for information to better understand the data, a visualization or its insight	50
PROPOSAL	Propose future work and possible adaptations of the data or visualization	43
CRITIQUE	Feedback to the visualization author, ranging from expressing constructive feedback and pointing out improvements to disagreement with the visualization and contestation of data.	90
ADDITIONAL INFORMATION	Introduce additional information to allow people to better contextualize or relate to the presented data. Includes background information, comparisons with similar data, trivia, and links to external sources with additional data.	58
TESTIMONY	Providing personal information by speaking from first-hand experiences and sharing anecdotes or memories	41
OPINION	Providing a personal perspective by sharing opinions, but also feelings and emotional reactions.	92
OTHER	Unfitting in other categories, off-topic or not understandable reactions	19

Table 2: Reaction types, their meaning, and frequency of occurrence (#). The sum of occurrence is larger than number of reactions due to some double encodings

insights the reaction author had: *"Season 5 was flawless, particularly those last 3 episodes [...]"* [BREAKING-BAD]. Beyond that, we found a number of comments that put individual observations into perspective: *"When billions of dollars can be depicted using a thin, one pixel line, it truly shows just how much two trillion dollars is"* [STIMULUS]. Other reactions draw conclusions from the data to point out necessary next steps or make a call to action: *"This is exactly why any push for free college MUST tackle the irresponsible spending and ballooning administrative costs by colleges"* [TUITION]. Conclusions can include a personal stance and can be motivated by expressing how this stance is not in line with the take-away message of the visualization: *"I wanted to summarise the data in a different way to how the [author] did"*.

Hypotheses provide speculative explanations for the presented information. While some hypotheses are phrased as short, tentative questions (*"Is it possible the only driver for cost of college is not inflation but rather demand?"* [TUITION], the majority of them introduces additional information and argues for a specific explanation of the presented data: *"[...] This perfectly tracks with the drive to transfer our publicly shared national treasure into the private accounts of the exceptionally wealthy, starting with Reagan in the 80's. This is wealth inequality come home to roost. "* [TUITION]. Some of these claims introduce bold theories that are generally not externally verified: *"This is all incentivized by the government [...]"* [TUITION].

Clarifications are questions addressed to the visualization author or the public audience. They are generally posed to better understand the presented information. We found a wide variety of clarification questions: Some people needed help with correctly interpreting the information (*"How is it negative, I know how these things work"*, oil) or encoding the visualization (*"What does the wideness of the color mean? [...]"* [MINECRAFT]). Some intend to get more details about specific data points (*"What the hell happened in season 4 of the UK version?"* [RAMSEY]) or background information about the topic (*"Why is there a seasonal trend to this? What month does that happen and any idea why?"* [BUSHFIRE]). One recurring clarification seeks more information about the used datasets and how they are generated: *"Why is 'nation security' part of the 'large corperations branch?"* [STIMULUS]. These clarifications are generally phrased in a more critical tone, as they question the design decisions that were made to produce the used data. This critical stance is also observable in some skeptical questions that question the correctness of single data points: *"[...] your chart is showing that both poles go from 0 hours of daylight to 24 hours of daylight in what appears to be a single day. See day 82 for example. You sure your math is right?"* [DAYLIGHT]. Such questions often co-occur with the reaction type critique and are a frequent case of double-encoding.

Proposals suggest future work and possible amendments to the dataset and adaptations of the visualizations. While we observed different kinds of proposals in the reactions, a majority of proposals suggested to introduce additional data to the existing visualization to provide more context: *"Can we add the common flu to this and start on January 1st?"* [DEATHS]. Others were interested in using the same visualization to display another related dataset: *"Can I get one of these for every spending bill? [...]"* [STIMULUS]. On the other hand, reactions proposed to improve or change the visualization: *"Kinda wish the opening and closing were more distinctively different"* [BLOCKBUSTER], sometimes with the intention to make the presentation more relatable: *"Please do British Columbia Canada! My wife doesn't get how bad and big it is"* [BUSHFIRE]. One participant described the motivation to react with a proposal with the *"Desire for more than one perspective on a dataset"*.

Critique gathers reactions that express a form of resistance. In contrast to proposals, they identify specific issues with the current form of visualization or data—often in a harsh or ironic speech. The most predominant kinds of critique are contestations that scrutinize the validity of the used data for the visualization: The scope of 52 (out of 113, 46% XX) critical reactions is on the data. These data contestations are threefold: Firstly, people point out missing data that they deem relevant to get a whole picture of the presented information: *"Should have included cancer and AIDS to humble down those COVID stats"* [DEATHS]. Secondly, people indicate that the presented data does not accurately represent the topic and should be generated in another way: *"It seems that this graphic uses worldwide statistics for Covid19, but USA only stats for some of the other outbreaks, in order to give an overly dramatic impression"* [DEATHS]. Thirdly, people contest the correctness of single data points, sometimes by introducing additional sources: *"This is wrong. Based on the Deutsche Bank info that came out last week. Trump is only worth about 600-700 M. [...]"* [NET-WORTH]. This last point illustrates how contestations often co-occur with the provision of additional information. The motivation behind this kind of critique often is *"to combat misleading or unhelpful data visualizations"*. Beyond these contestations of data we observe 36 (32% XX) criticisms of the visualization. The reactions point out hard to understand or possibly misleading representations (*"This chart succeeds at illustrating the wealth of Bloomberg, but does a very poor job of displaying/contrasting most of the candidates. Consider a log plot"*

	Observation	Conclusion	Hypothesis	Clarification	Proposal	Critique	Addtitional	Testimony	Opinion	Joke	Other	**Total**	Personal %
Topic	2	10	37	18	8	12	37	29	49	29	15	**246**	29%
Vis	3	3	1	8	12	37	6	1	21	5	2	**99**	34%
Data	6	2	10	19	31	53	20	3	12	9	0	**165**	22%
Insight	56	26	27	25	2	11	13	10	50	37	5	**262**	30%
Total	**67**	**41**	**75**	**70**	**53**	**113**	**76**	**43**	**132**	**80**	**22**		
Personal %	31%	24%	24%	27%	36%	23%	14%	91%	39%	23%	14%		

Figure 2: A heatmap indicating the absolute distribution of types across scopes. Values that stick out are e.g., how critical reactions most often relate to the data, whereas the scope of observations most often is an insight. Yellow cells indicate the relative distribution of personal flags. By design, testimonies and opinions most often have a personal notion.

[NET-WORTH] and often co-occur with suggestions for improvement. Surprisingly many reactions address the aesthetic appeal of visualizations and we often observed harsh criticism for a perceived lack of beauty or the use of animation in charts.

Additional information is provided in order to contextualize the visualization. This can be background knowledge that helps to understand the subject (*"[...] The garbage patch is not a big island of floating plastic bottles, etc. It's a big sparse conglomeration of micro plastics[...]"* [GARBAGE] or to put the presented data into perspective (*"Instead we had 13 million people watch the rally"* [TULSA]). Since additional information is used to explain existing or introduce new knowledge, we observe that these reactions are among the longest in our study (>500 characters on average), with them being double as long as the overall average. People's motivation to provide additional information is to contribute to sensemaking by adding what they deem relevant: *"I had information pertaining to the topic at hand and just wanted to point out that the data was more significant."* Others also have the discussion that can unfold from this kind of reaction in mind: *"[...] i decided to post it to add the idea to the conversation"*.

Testimonies are reports of first-hand experiences with the presented information, usually with a strong focus on the topic (70%). Testimonies express forms of local knowledge and anecdotal evidences of people's everyday lives, oftentimes disclosing biographical or private information. The largest group of testimonies talk about a personal spatial relation to the presented information: *"I live along the east coast of Australia in an area not directly affected by the fire but heavily impacted by the smoke [...]"* [BUSHFIRE]; *"We don't have daylight savings here in India. Now I know why that's a phenomenon in the West. [...]"* [DAYLIGHT]. Others relate to the topic through memories and experiences they have made in the past: *"I remember being a kid wanting to work at block buster one day. They were gone by the time I was old enough to work"* [BLOCKBUSTER]. A third group talks about the impact the presented data has on their everyday lives and disclose information about their jobs or families: *"[...] I'm a little pissed I'm an essential worker throughout this and will be getting less pay than being unemployed [...]"* [STIMULUS]. The last example shows that testimonies frequently co-occur with opinions.

Opinions provide a personal perspective by sharing an individual judgement or point-of-view. This can be simple feedback to the visualization author (*"I really enjoy this graph. It really bothers me when people conflate millionaires and billionaires"* [NET-WORTH]; *"This data is most certainly not beautiful"* [RATES]). Often, opinions carry a notion of personal account, yet they do not disclose the same level of personal information as testimonies: *"I think I will ever get or understand the appeal to play golf. Played once and was bored out of my mind"* [GOLF]. Visualizations that depict everyday topics many people can connect with—such as pop-cultural data about TV shows—show a high number of opinionated reactions. On the one hand, people have opinions about the topic, people who watched the depicted TV show express how they liked it (*"Glad to see Person of Interest get some love. Always felt that was an underrated show. "* [IMDB]. However, we also found some contestations when people relate their own experience to the data and express an opinionated contestation: *"How they gunna hate on Towlie? one of my favorite episodes!"* [IMDB]. This is backed up by the motivation to create these kinds of posts: *"I liked the show. I agree with the visualization"*. This tension between an insight on the data and the commenter's own stance on that matter is a recurring theme we found in many opinionated reactions: *"Might as well call this 'Why I hate people: An animated explanation' "* [GOOGLE].

How many redditors does it take to change a light bulb? Nine to express all prior types and one to make **jokes** about it. Across all visualizations, we observe many reactions that are primarily intended to be funny, including sarcasm (*"ah yes, the two presidents of the united states"* [GOLF]), pop-cultural references (*"Game of Thrones' ending is why I have trust issues."* [CASES]), observational humour (*"Looks like every Facebook party ever organized"* [TULSA]), and puns (*"This chart points out the ultimate irony... It could not break into Bad..."* [BREAKING-BAD]). 19 reactions we collected would not fit into any of the presented categories. This includes verbal insults, automated responses from bots or things we could not read or understand, e.g., *"Meanwhile a mysterious diplomat traverses his homeland searching unsuccessfully for a single hand to shake"* [BUSHFIRES].

5 THE PERSONAL AND THE PUBLIC: WHY PEOPLE ENGAGE WITH VISUALIZATIONS ONLINE

While reaction types inform us about *what* type of engagement people shared, this section looks into the *why*, i.e., asking for the motivations behind people's comments. Based on the 168 survey replies we obtained from from comments, we report on findings about personal (intrinsic) motivations and motivations to support the public discourse (extrinsic).

5.1 Personal Perspectives and Experiences

A major theme that we observed when coding the reactions was that personal perspectives were expressed in many comments and throughout all types of reactions: testimonies, opinions, emotions (e.g., in observations), adding personal knowledge (e.g., in additional information), proposals, and critiques. Building on prior work

on personal perspectives in data visualization [e.g., 14, 19, 26] we discuss four drivers for expressing personal experiences and perspectives and seek to understand what they add to the discourse around data visualization.

Expressing opinions—We observed that reactions expressing opinions, i.e., points of view or judgements about a topic, are an inherent part of the sensemaking process, as they act as an intermediary between the presented information and a viewer's self-perception. This self-perception can be shaped by cultural values or political views, (e.g., *"[This] is a perfect encapsulation of Trump's opportunism, his hypocrisy and lack of self-awareness"* [GOLF]) but also relate to personal preferences, e.g., how people react to the presented ratings of a TV show they enjoyed. Expressions of opinions vary in their expressivity including typical phrases like "I think / love / hate [...]" or "in my opinion [...]". Paired with a scope of topic or insight—specifically when talking about their own observations—they further indicate that people refocus on their own position and juxtaposing it to what the data suggests when making sense of it. Commenters are well aware of this dynamic that prefers personal relation over information: *"I believe my comment was upvoted because others agreed with it, not because it offered any insight or analysis."*

Matching knowledges—Other reactions demonstrate how people compare extracted information against their own existing knowledge. Oftentimes, this simply results in people stating what they learned through observation and interpretation. In case of a mismatch between presented and existing knowledge, we observe two possible reactions. We found several reactions that imply how the new information challenges previous assumptions, often accompanied by an expression of astonishment (*"Bernie Sanders is 78 years old with a NW of only 2.5 million? That's honestly pretty surprising to me [...]"* [NET-WORTH]. If however, the viewer is not convinced by the presented data, they express their own knowledge in the form of critiques, motivated by not letting something they perceive as wrong remain uncommented: *"I have a compulsion for calling out bullshit by using my experience and education"*.

Sharing emotions—Redditors post emotional reactions that are evoked by the visualization. While many reactions contain emotions, they do not typically carry the central message of the reaction, but often set the underlying tone of a reaction. People prelude their reactions with how a visualization makes them *"feel"*, or what their *"gut feeling"* [GARBAGE] about the data is. Some reactions however have feelings or emotions as a first class citizen: *"I'm simultaneously horrified yet excited to see how this looks [...]"* [CASES]. In the survey, one participant points out how expressing emotions is the main driver of her reaction: *"I can't really explain my motivation other than wanting to share how this phenomenon 'feels', even if feelings [do not equal] science."* Another participant states: *"This was about the hard feelings I still have [...] and less about data as a motivation to post."* This emphasizes the role of emotional statements as a counter-weight against the typical impression of data and visualization to be rational and objective [7].

Claiming interest—We found some expressions that were shaped by a viewer's personal interest, which is clearly stated in their motivation: *"I appreciate data visualization as someone with a background in the sciences, and also have an interest in politics, and so I felt motivated to comment [...]"*. When formulating proposals, people often asked for an adaptation that they can better relate to personally, for example, in order to *"satisfy a curious itch"*. This might be a change of the visualization in order to better understand how the presented information affects them personally (*"Can we get another chart that outlines the reduction of state and federal funds dedicated to higher education? I know my state only contributes to less than 10% of the state's higher education system"* [TUITION]), or the adaptation of new data, for example, to display the ratings of a TV Show they watched and want to learn more about. In the survey, a participant described his motivation to post a clarification with a *"genuine interest to get an apolitical answer [...]"* in order to learn more about particular aspects that were not further described in the visualization. People also ask for clarifications to express their *"Desire for more than one perspective on a dataset"*, which can be met by adding anecdotal evidence.

Adding anecdotal evidence—When considering a viewer's personal involvement in the visualizations, testimonies get to the heart of what the previous drivers already indicate. They are expressions of inherently personal, sometimes even intimate relations that complement the visualization with first-hand experiences. Against the background that visualizations can depict data that may be specific to a single country, time-period or demographic group, these reactions offer subjective perspectives that can help the audience to better relate to the data. Several participants confirm this motivation: *"I was looking to provide a human story behind the data. The bushfires in Australia earlier this year were devastating and my comment was intended to describe how the fires being discussed in the data is beautiful post was affecting millions of people"* or *"I am a regular reader of Reddit, and golf a lot. The data was interesting to me and I wanted to add some context to it for those that may not golf and do not quite understand if the rounds of golf represented are a lot or not very many. If an activity is foreign to someone a number matters little to them. Context or comparison to better understand the number can be helpful."*

5.2 Informing the Public Discourse

We found that many personal reactions, though motivated by internal reasons, were meant to contribute to the public discourse and collaborative sensemaking among community members. For example, while some reactions only speak to the visualization author (*"To compliment the and give constructive feedback"*, *"encouragement [...] and asking for more information"*). Other comments aimed to add to a public discourse more broadly: *"posting my opinion to a wide audience"*. Below, we characterize four motivations of people to express themselves in public. Considering that reactions are also motivated by personal expression it is important to note that these drivers are not mutually exclusive with the personal ones reported in section 5.1.

Participate in public sensemaking—Several participants explained their motivation to share their existing knowledge or gained insights with others as constructive and objective conclusion to share facts, findings, and implications from the visualization with others (*"[...] understand and share the implications of the data"*; *"share an interesting conclusion [...]"*; *"share an observation/trend"*).

Adding one's own story—Personal experiences can constitute a crucial quality of the reactions to data visualizations we studied.

The prospect of providing complementary human perspectives onto the data is particularly relevant for the discourse to humanize data, complement findings, *"providing local or 'inside' information"*, or to *"contextualize the data for others[...]. I wanted people remember or know what it was like."*.

Rectifying what's wrong—Correcting wrong information goes beyond the previously described sensemaking. Similar to previous work [14], we observed how people express critique about the topic, data, or visual representation: *"Data is often used to justify one side of things. Unfortunately data can be manipulated to make concepts, ideas, etc. seem better or worse [...]"*. While correcting old or introducing new information is an important motivation, we only see few cases that use and link to verifiable sources. In some cases, the source is named but not linked—or different authors provide different "correct" numbers fixing the data. The majority of reactions claiming that some of the presented data is wrong and suggest other data remain unverified and are intertwined with personal perspectives. Eventually, critique on the visual design is specifically important and points to increasing awareness of slants and biases in visualization: *"A lot of information is presented in a biased way, where the numbers are accurate but are portrayed in biased light. I try to rectify that"*.

Sparking discussion—Reactions can be an invitation for others to participate. Participants express reactions in order to further *"engage with the community"* and trigger more follow-up reactions: they want *"to bring attention to a particularly interesting data event and provide a starting point for further discussion"* and feel that the statements they make *"furthered the discussion around the topic"*. People are curious what other's make of their observations and opinions: *"I thought it'd be interesting to see how people would react to it [and] if others felt the same [...]"*.

Besides personal and public motivations, a third group of motivation can be summarized *"acquir[ing] upvotes"*, or having fun by contributing jokes: *"sparkles! I love sparkles"* [CASES].

6 DISCUSSION

In this section, we summarize and discuss our findings, reflect on our methodology, discuss implications for interface design to support engagement with visualizations, and relate our findings to other findings in the literature.

6.1 Main Findings

We found a great **diversity of reactions** in the visualization-focused online community `/r/dataisbeautiful`. Previous classifications [34, 35] and interfaces [11, 40] primarily focused on annotations—a different, yet arguably comparable kind of audience-authored reaction—as tools for collaborative sensemaking. In contrast, reactions on Reddit serve a wider variety of purposes and express not only analytical insight, but also personal perspectives, critical questions, and additional information. Furthermore, the comments themselves constitute an integral part of the visualization community on Reddit and as such they also exhibit humor, sarcasm, and banter. The dynamic between the visualization and the people that comment on it is oftentimes harmonic and sociable, for example, when expressing observations or evoking memories, but can also be tense and antagonistic, especially when critical comments and opinions are expressed.

We were interested in understanding how an open and potentially heterogenous community reacts to visualizations, differing from from lab studies and specific communities (such as university students and employers of a tech company) and a wide range of different types visualizations. Our results confirm findings by Heer et al. [11] and reveal additional categories. Our taxonomy is meant to capture 'exclusive' categories in order to provide a better picture of individual types of reactions. Finally, we are most interested in how people report on their personal relation and their motivations in contributing to commenting, something entirely missing in the previous studies.

For example, we found various **levels of sensemaking** across almost all reaction types. While observations can be associated with low engagement in Mahyar et al.'s taxonomy [23], reactions such as hypotheses, clarifications, and conclusions indicate a higher level of engagement, often based on personal experience, wider contextual knowledge, or profound interest. A specific type of comments added comparisons and other information to help translating the insights of a visualization from one context to another. Eventually, critiques can be seen as a distinct form of sensemaking, given that they imply some sort of personal cognitive effort that includes understanding an issue, comparing it with previous observations or guidelines, finding a point of disagreement, and expressing this disagreement intelligibly. In many cases, the critiques we found were highly constructive, suggesting the means to overcome a particular weakness.

The **role of personal perspective and experience** is prevalent across all types of reactions and ranges from expressing opinions and emotions to contributing personal experiences (testimonies) and adding or providing evidence. We also found that critique is not necessarily expressed from a neutral or analytical perspective: more than one third of critical reactions have a personal notion to them. This supports many of the findings by Peck et al. [26] and also shows that personality does not only shape perception, but also the sensemaking and communication after viewing a visualization. Confirming Kennedy et al.'s [19] observations about the role of emotions in engagements with data, we found multiple instances that expressed the relevance of people's *feeling of numbers*. Eventually, we identified several **drivers for engaging with visualization**, including personal motivations and contributions to the public discourse.

6.2 Reflecting on Methodology and Data

Our findings are characteristic for Reddit, specifically the `/r/dataisbeautiful` community. Reddit is featuring non-interactive content, other than e.g., many contemporary news outlets and their visual stories. Besides the ensuing discussions, Reddit content comes without additional information, e.g., such as in Hullman et al.'s study [14]. While related studies on personal perception [26] and feelings [19] have been carried out in co-located, synchronous workshops, or interviews, our study focused on an online setting. For the most part, we were able to support existing results and deepen our knowledge of the audience behavior when reacting to visualizations and their motivation behind it.

We deliberately **ignored discussion threads**, i.e., the nested replies to comments, in order to focus on the primary reactions to the visualizations. However, we often resorted to replies in order to understand and agree on the meaning of single comment as these replies often clarified the initial intention of a comment. Future work should consider these discussions to better understand the audience's perception of reactions, for example, to find out how successful testimonies are in humanizing the data or determine whether people change their mind based on critiques. The reaction types provide a structured basis for such an endeavour and can be used to further analyze how discussions evolve.

To account for the complex nature of the expressions we encountered, we **iteratively adjusted the coding scheme**. Comments contained traces of different reaction types and this ambiguity required intense discussion among the authors. Correctly and consistently interpreting a reaction depends on inside knowledge about the topic, the specific online community, and the used jargon and abbreviations. Some of this is due to the varying familiarity with Reddit and the `/r/dataisbeautiful` community among the authors. Still, some comments could not be classified due to potential irony, missing context, or where the reaction would not fit in any of the ten categories.

More research is necessary to **generalize our findings** and determine whether they are specific to one online community or whether they apply to other platforms, communities, and even co-located (uncontrolled) settings. Our framework can serve as a reference point to understand how reaction types, their quality, and quantity change with context (e.g., online, situated, synchronous, asynchronous [17]), audience, media (e.g., pen and paper, data physicalization, interactive), and goals of engagement (e.g., open, entertainment, collaborative sensemaking, design crits). Evidently, more research is necessary to understand these specific parameters that influence human engagement with data visualizations.

Our reactions are **based on textual comments**. Other sources of textual comments could include forums [4]. Other means of data capturing, for example through audio recordings, interaction, pointing, drawing, or specific functions in existing annotations interfaces, can provide complementary information while at the same time requiring specific analysis methods.

6.3 Design Implications and Future Work

Beyond the often discussed cases for engagement exploration and communication, our work informs mechanisms of questioning and contestation (CRITIQUES, CLARIFICATIONS), user contribution (TESTIMONIES, ADDITIONAL INFORMATION, PROPOSALS, OPINIONS), and more sensemaking tasks (OBSERVATIONS, HYPOTHESIS, CONCLUSION). Informed by these reactions (4) and drivers (5) for personal contributions,this section discusses design implications for user interfaces, complementing previous implications by Heer et al. [11].

G1—Allow for **type-specific annotations** to capture the richness of observers reactions and to support participation and discourse around data. Such rich annotations can entail audience-driven labeling of elements and regions of a visualization, highlighting relations between elements, linking reactions to data points, or drawing on the visualization. While some techniques are present in existing systems (e.g. Lyra, CommentSpace, Sense.us), our typology can be used to better tailor annotation features to suit the specific needs of individual reaction types. For example, interfaces could provide features to collect and categorize OPINIONS such as done using Twitter for general discussions [16] and link them to the data. We can imagine visual polling interfaces in which user reactions drive visual properties of elements that are put to the vote, e.g., the roughness of a visual shape or a highly controversial opinion. Interfaces for data contestation could allow people to *rectify whats wrong* and provide features to express a CRITIQUE of data points or the quality or provenance of the data set, and add details on the annotator's confidence or the provided ADDITIONAL INFORMATION as evidence. CLARIFICATIONS and PROPOSALS aim to improve or re-create visualizations as proposed by Hullman at al., [14], raising questions about data access and visualization creation tools.

G2—Supporting **personalized stories and visualizations** can capture an audience's TESTIMONIES and display them within the context of a given visualization. Driven by people's willingness to *provide anecdotal evidence* and *add their own story*, this allows for a juxtaposition of the data's bird's-eye perspective on a topic and a local, situated reaction to it and thus provides a more convincing narrative as proposed by one study participant: *"visualization is powerful it's even more powerful when supplemented by real experiences"*. We imagine interfaces to provide personal testimonies in a more participatory and empowering way than 'just' annotating and commenting on a given visualization. Challenges are posed by allowing for easy authoring of such stories as well as storage and navigating these stories. Eventually, such stories can include not only text, but other media such as photos, oral histories [22], customized visualizations, or data comics [1, 41]. This can help to include external information and adapt a given visualization to different contexts (e.g., using alternative geographic maps).

G3—**Visualizing the discourse** around a given visualization can help analysts and an audience understand controversial parts and aspects of a visualization and the data displayed. For example, similar to hotspots in a document [12], a visualization could display a heatmap to highlight parts with many reactions, broken down to different reaction types, e.g. OPINIONS, CONTESTATIONS, CLARIFICATIONS, CRITIQUES, OBSERVATIONS or, as suggested by Heer et al. [11], keyword tagging. Complementary analytical charts could show specific insights and trends about reactions such as reaction type, frequency over time, types of OPINIONS, or terms mentioned in CRITIQUE posts using techniques from NLP and sentiment analysis. Showing which and how elements of a data set or visualization are reacted to could inform both further audience engagement as well as a detailed analysis into the content of the discourse. An eminent challenge here is again to design easy-to-use interfaces to submit reactions and to navigate an existing discourse and its reactions.

G4—Open discourse and annotation requires **guidance and moderation**. **Guidance** should help participants express reactions and participate a discourse. For example, besides explanatory messages, visualizations of discourse (G3) can provide cues and affordances to an ongoing discourse [12] and provide for *social navigation cues* [11]. Scoring systems could allow audiences interacting with the discussion itself [18]. **Moderation** must prevent vandalism and misinformation as well as prevent discussions from going off-topic.

Also, moderation can help finding agreement and resolving conflicts for CONTESTATIONS and CRITIQUES, eventually updating a given visualization or data alongside respective notifications to inform the audience of such updates.

G5—Gratify people's participation. While one goal of any (visualization-based) discourse should be to enrich an audiences' understanding of the data (see Sec 5, *matching knowledge, add anecdotal knowledge, participate in public sensemaking, sparking discussion, rectifying what's wrong*), we found many other motivations for people to engage with visualizations (express opinion, claim interest, share emotions, express humor) which could be leveraged by social media platforms in to attract attention to a topic and give the audience an enriched channel of expression and participation. More simple motivations can include attention to one's comments or posting humoristic comments. One challenge here is how to get people engaged in intellectually demanding reactions such as CRITIQUES and CONTESTATIONS, or suggestions on visualization redesign.

7 CONCLUSION

We see a huge research potential in providing more expressive and inclusive interfaces for engaging with and discussing content in data visualizations; this includes both online as well as offline media. To inform these interfaces, this paper has investigated reactions to data visualizations on the social platform Reddit, seeking to better understand different types of reactions and the drivers that motivate people to express them in a public forum. We coded comments to popular posts containing data visualizations and developed a coding scheme of ten types of reactions. We extended prior classification systems and added personal relations to data and a wide variety of critiques on the presented visualizations and their data sets. In a subsequent online survey, we identified nine drivers of people to engage with a visualization and express themselves. We differentiated between drivers that share a rather personal motivation from those that primarily aim to inform the public discourse. Our results show a variety of ways for people to personally, analytically, or critically connect with data when they are visualised in online media. Our work demonstrates people's willingness to engage and put effort into presenting personal perspectives to a public forum. From our findings, we deduced five design implications to inform future interfaces for participatory data visualization systems that support a range of different reaction types.

ACKNOWLEDGEMENTS

We would like to thank the reviewers for their valuable feedback and especially the program subcommittee member who helped shepherding our paper.

REFERENCES

[1] Benjamin Bach, Nathalie Henry Riche, Sheelagh Carpendale, and Hanspeter Pfister. 2017. The emerging genre of data comics. *IEEE computer graphics and applications* 37, 3 (2017), 6–13.

[2] Elizabeth F Churchill, Jonathan Trevor, Sara Bly, Les Nelson, and Davor Cubranic. 2000. Anchored conversations: chatting in the context of a document. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 454–461.

[3] Matthew Conlen, Alex Kale, and Jeffrey Heer. 2019. Capture & analysis of active reading behaviors for interactive articles on the web. In *Computer Graphics Forum*, Vol. 38. Wiley Online Library, 687–698.

[4] Alexandra Diehl, Alfie Abdul-Rahman, Mennatallah El-Assady, Benjamin Bach, Daniel A Keim, and Min Chen. 2018. VisGuides: A Forum for Discussing Visualization Guidelines.. In *EuroVis (Short Papers)*. 61–65.

[5] Marian Dörk, Patrick Feng, Christopher Collins, and Sheelagh Carpendale. 2013. Critical InfoVis: exploring the politics of visualization. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 2189–2198.

[6] Bryan Dosono, Bryan Semaan, and Jeff Hemsley. 2017. Exploring AAPI identity online: Political ideology as a factor affecting identity work on Reddit. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2528–2535.

[7] Catherine D'Ignazio and Lauren F Klein. 2016. Feminist data visualization. In *Workshop on Visualization for the Digital Humanities (VIS4DH), Baltimore. IEEE*.

[8] Micheline Elias and Anastasia Bezerianos. 2012. Annotating BI visualization dashboards: Needs & challenges. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1641–1650.

[9] Eric Gilbert. 2013. Widespread underprovision on Reddit. In *Proceedings of the 2013 conference on Computer supported cooperative work*. 803–808.

[10] Steve Haroz, Robert Kosara, and Steven L Franconeri. 2015. Isotype visualization: Working memory, performance, and engagement with pictographs. In *Proceedings of the 33rd annual ACM conference on human factors in computing systems*. 1191–1200.

[11] 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*. 1029–1038.

[12] William C Hill, James D Hollan, Dave Wroblewski, and Tim McCandless. 1992. Edit wear and read wear. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 3–9.

[13] Uta Hinrichs, Holly Schmidt, and Sheelagh Carpendale. 2008. EMDialog: Bringing information visualization into the museum. *IEEE transactions on visualization and computer graphics* 14, 6 (2008), 1181–1188.

[14] 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*. 1170–1175.

[15] Ya-Hsin Hung and Paul Parsons. 2017. Assessing user engagement in information visualization. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 1708–1717.

[16] Samuel Huron, Petra Isenberg, and Jean Daniel Fekete. 2013. PolemicTweet: Video annotation and analysis through tagged tweets. In *IFIP Conference on Human-Computer Interaction*. Springer, 135–152.

[17] Petra Isenberg, Niklas Elmqvist, Jean Scholtz, Daniel Cernea, Kwan-Liu Ma, and Hans Hagen. 2011. Collaborative visualization: Definition, challenges, and research agenda. *Information Visualization* 10, 4 (2011), 310–326.

[18] Tobias Kauer, Benjamin Bach, and Marian Dörk. 2019. Participatory Deep Maps: Towards Discursive User Engagement With Data Visualizations. In *CityVis: 2nd Workshop on Urban Data at IEEE VIS*.

[19] Helen Kennedy and Rosemary Lucy Hill. 2018. The feeling of numbers: Emotions in everyday engagements with data and their visualisation. *Sociology* 52, 4 (2018), 830–848.

[20] Helen Kennedy, Rosemary Lucy Hill, William Allen, and Andy Kirk. 2016. Engaging with (big) data visualizations: Factors that affect engagement and resulting new definitions of effectiveness. *First Monday* 21, 11 (2016).

[21] Danielle K Kilgo, Joseph J Yoo, Vinicio Sinta, Stephanie Geise, Melissa Suran, and Thomas J Johnson. 2016. Led it on Reddit: An exploratory study examining opinion leadership on Reddit. *First Monday* (2016).

[22] Manissa M Maharawal and Erin McElroy. 2018. The Anti-Eviction Mapping Project: Counter Mapping and Oral History toward Bay Area Housing Justice. *Annals of the American Association of Geographers* 108, 2 (2018), 380–389.

[23] Narges Mahyar, Sung-Hee Kim, and Bum Chul Kwon. 2015. Towards a taxonomy for evaluating user engagement in information visualization. In *Workshop on Personal Visualization: Exploring Everyday Life*, Vol. 3. 2.

[24] Alexey N Medvedev, Renaud Lambiotte, and Jean-Charles Delvenne. 2017. The anatomy of Reddit: An overview of academic research. In *Dynamics on and of Complex Networks*. Springer, 183–204.

[25] Thomas L Naps, Guido Rößling, Vicki Almstrum, Wanda Dann, Rudolf Fleischer, Chris Hundhausen, Ari Korhonen, Lauri Malmi, Myles McNally, Susan Rodger, et al. 2002. Exploring the role of visualization and engagement in computer science education. In *Working group reports from ITiCSE on Innovation and technology in computer science education*. 131–152.

[26] Evan M Peck, Sofia E Ayuso, and Omar El-Etr. 2019. Data is personal: Attitudes and perceptions of data visualization in rural pennsylvania. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[27] Donghao Ren, Matthew Brehmer, Bongshin Lee, Tobias Höllerer, and Eun Kyoung Choe. 2017. Chartaccent: Annotation for data-driven storytelling. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*. Ieee, 230–239.

[28] Edward Segel and Jeffrey Heer. 2010. Narrative visualization: Telling stories with data. *IEEE transactions on visualization and computer graphics* 16, 6 (2010),

1139–1148.

[29] Xinhuan Shu, Aoyu Wu, Junxiu Tang, Benjamin Bach, Yingcai Wu, and Huamin Qu. 2020. What Makes a Data-GIF Understandable? *arXiv preprint arXiv:2008.07227* (2020).

[30] David Sprague and Melanie Tory. 2012. Exploring how and why people use visualizations in casual contexts: Modeling user goals and regulated motivations. *Information Visualization* 11, 2 (2012), 106–123.

[31] Anselm Strauss and Juliet Corbin. 1994. Grounded theory methodology. *Handbook of qualitative research* 17, 1 (1994), 273–285.

[32] Alice Thudt, Uta Hinrichs, and Sheelagh Carpendale. 2012. The bohemian bookshelf: supporting serendipitous book discoveries through information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1461–1470.

[33] Nina Valkanova, Robert Walter, Andrew Vande Moere, and Jörg Müller. 2014. MyPosition: sparking civic discourse by a public interactive poll visualization. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. 1323–1332.

[34] Pierre Vanhulst, Florian Evequoz, Raphael Tuor, and Denis Lalanne. 2018. A descriptive attribute-based framework for annotations in data visualization. In *International Joint Conference on Computer Vision, Imaging and Computer Graphics*. Springer, 143–166.

[35] Pierre Vanhulst, Florian Évéquoz, Raphaël Tuor, and Denis Lalanne. 2018. Designing a Classification for User-authored Annotations in Data Visualization.. In *VISIGRAPP (3: IVAPP)*. 85–96.

[36] 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 (2007), 1121–1128.

[37] Jagoda Walny, Sarah Storteboom, Richard Pusch, Steven Munsu Hwang, Søren Knudsen, Sheelagh Carpendale, and Wesley Willett. 2020. PixelClipper: Supporting Public Engagement and Conversation About Visualizations. *IEEE Computer Graphics and Applications* 40, 2 (2020), 57–70.

[38] Zezhong Wang, Shunming Wang, Matteo Farinella, Dave Murray-Rust, Nathalie Henry Riche, and Benjamin Bach. 2019. Comparing Effectiveness and Engagement of Data Comics and Infographics. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–12.

[39] Martin Wattenberg. 2005. Baby names, visualization, and social data analysis. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.* IEEE, 1–7.

[40] Wesley Willett, Jeffrey Heer, Joseph Hellerstein, and Maneesh Agrawala. 2011. CommentSpace: structured support for collaborative visual analysis. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 3131–3140.

[41] Zhenpeng Zhao, Rachael Marr, and Niklas Elmqvist. 2015. Data comics: Sequential art for data-driven storytelling. *tech. report* (2015).