Illuminating the Landscape of Differential Privacy: An Interview Study on the Use of Visualization in Real-World Deployments

Liudas Panavas*, Amit Sarker*, Sara Di Bartolomeo, Ali Sarvghad, Cody Dunne, and Narges Mahyar

Abstract—As Differential Privacy (DP) transitions from theory to practice, visualization has surfaced as a catalyst in promoting acceptance and usage. Despite the potential of visualization tools to support differential privacy implementation, their development is limited by a lack of understanding of the overall deployment process, practitioner challenges, and the role of visual tools in real-world deployments. To narrow this gap, we interviewed 18 professionals from various backgrounds who regularly engage with differential privacy in their work. Our objectives were to understand the differential privacy implementation process and associated challenges; explore the actors (individuals involved in differential privacy implementation), how they use or struggle to use visualization; and identify the benefits and challenges of using visualization in the implementation process. Our results delineate the differential privacy implementation process into five distinct stages and highlight the main actors alongside the diverse visualization applications and shortcomings. We find that visualizations can be used to build foundational differential privacy knowledge, describe implementation parameters, and evaluate private outputs. However, the visualization strategies described often fail to address the diverse technical backgrounds and varied privacy and accuracy concerns of users, hindering effective communication between the different actors involved in the implementation process. From our findings, we propose three research directions: visualizations for setting and evaluating noise addition, evaluation of uncertainty visualization related to trust in differential privacy, and research focused on pedagogical visualizations for complex data science topics. A free copy of this paper and all supplemental materials are available at https://osf.io/qhyzt/?view_only=1a5c7d7553c840ab9f125d88bc13946f

Index Terms—Differential Privacy, Data Visualization, Human-Subjects Qualitative Studies, Uncertainty Visualization.

1 Introduction

With sensitive data being collected at scale and driving scientific advancements in medicine, social science, and artificial intelligence, implementing effective data anonymization techniques has become essential [64]. Differential privacy has become the gold standard of data anonymization techniques by adding calibrated noise to statistical queries to protect individuals' privacy [22]. Despite being an established academic research area with solid foundations and notable existing literature [28], [36], adoption of differential privacy by industry has been rare and faced substantial challenges [18], [27], [47]. Existing work has extensively investigated theoretical aspects of differential privacy but largely overlooked the challenges faced by practitioners, including lack of expertise and trust among adopters [15], [18], [23], [28]. These operational challenges often arise due to the complex nature of differential privacy and must be addressed for it to become usable in real-world deployments [15].

Industry practitioners have identified visualization as a possible solution to address these operational challenges, given its ability to convey and communicate complex data topics [23], [28], [57], [66], [76], [83]. Prior research has also demonstrated the efficacy of visualization in enhancing the comprehension and communication of differential privacy to end users and privacy practitioners [23], [28], [52], [67], [86], [90]. However, despite visualizations' potential to directly address these functional challenges of differential privacy, the extent to which visualization can support differential privacy implementations remains unclear. While acknowledging the effectiveness of visualization, these prior works have only briefly touched upon its potential application without thoroughly exploring its role in real-world deployment processes [23], [28].

The limited use of visualization in differential privacy deployments may stem from challenges, including a lack of understanding about the overall deployment process, the needs of practitioners, and visualization's potential benefits [28], [67]. Moreover, with only a cursory examination of visualization, it is difficult to establish where visualization researchers should focus future efforts. To better understand how visualization can help bridge the gap between theory and industry practice, our study focuses on describing the differential privacy deployment pipeline, the actors involved, and the factors contributing to the limited adoption of visualization in industry practices.

To help determine what challenges limit differential privacy implementation—and what remedies may be possible through visualization use—we conducted an in-depth interview study with 18 industry practitioners who have been a part of real-world differential privacy deployments (Table 1). These interviews aimed to gain a nuanced understanding of the differential privacy pipeline, the actors involved and identify key communication challenges between them. The interview questions evaluate how visualization can aid differential privacy implementations by exploring the benefits and challenges of using visualization. This study employed a semi-structured interview approach to gathering data from

- *The first and second authors made equal contributions.
- *Liudas Panavas, Sara Di Bartolomeo, and Cody Dunne are with Northeastern University. E-mails: [panavas.l | dibartolomeo.s | c.dunne]@northeastern.edu*
- *Amit Sarker, Ali Sargavad, and Narges Mahyar are with the University of Massachusetts, Amherst. E-mails: asarker@cics.umass.edu, [asarv | nmahyar]@cs.umass.edu.*

participants who worked in differential privacy deployments, held a range of job responsibilities, and had different levels of experience with visualization. The data was analyzed using a deductive thematic analysis approach, with themes derived from the research objectives and formulated as interrogative questions [12]. Our research lays the groundwork for examining the practical challenges of implementing differential privacy in real-world contexts, focusing on understanding the role of visualization in facilitating this process.

Our analysis delineates the differential privacy pipeline, pinpointing the responsibilities of actors at each phase. We consistently found that successful implementations require the collaborative input of individuals with diverse expertise. Furthermore, visualization proves to be an important instrument for facilitating communication, enhancing education, and enabling the effective application of differential privacy. It is used to convey differential privacy theory, analyze implementation parameters, support decision-making, and evaluate the quality of private data. However, some key obstacles, such as technical intricacies of depicting abstract privacy models, achieving clarity for a diverse audience, and illustrating error metrics, pose challenges in leveraging visualizations effectively for differential privacy implementation. After analyzing the results, we conducted a brief follow-up study with five additional practitioners (Table 3) to evaluate our implementation pipeline (Fig. 1) and provide additional visualization examples. They agreed that the implementation pipeline was captured accurately; however, they expressed additional nuances in terms of complexity and effort for the various stages based on the data release strategy.

Our study builds upon prior work by identifying five common implementation stages, providing a more granular categorization of actors in differential privacy, identifying common communication channels among them, and examining the roadblocks stemming from communication issues. Additionally, we surface the challenges and opportunities of using visualizations in differential privacy implementation, highlighting the need for further visualization research to advance practical application. In particular, we contribute: (1) a characterization of differential privacy deployment pipelines and their associated actors; (2) an examination of the role of visualization and its use in alleviating implementation challenges; and (3) three collaborative research directions focusing on visualizations for evaluating noise, trust through uncertainty visualization, and educational visualizations for differential privacy.

2 Background and Related Work

Below we provide an overview of differential privacy, including an examination of industry adoption thus far, a discussion of important implementations, and a summary of relevant literature on those implementations. We then map the current literature at the intersection of differential privacy and visualization. This intersection covers visualizations that convey theoretical knowledge about differential privacy, explore the privacy-utility trade-off, and visualize differentially private data outputs. We motivate our work based on the limitations of the relevant literature and the challenges of differential privacy adoption.

2.1 Differential Privacy Background

Consider a scenario where the U.S. Department of Housing and Urban Development is formulating a fair, affordable housing policy. They utilize Census data processed with differential privacy to identify patterns such as the unique housing challenges faced by same-sex unmarried partners [26]. By doing so, they can introduce policy provisions like stricter penalties for landlords found guilty of discrimination. However, it is also crucial to note potential risks. If individual-level data was disclosed without differential privacy, it could lead to targeted discrimination or harassment, particularly in areas where same-sex couples might face hostility. This highlights the dual purpose differential privacy serves—it enables the creation of well-informed, equitable policies while safeguarding the privacy of vulnerable individuals.

The U.S. Census Bureau's application of differential privacy in the 2020 Census is a compelling illustration of differential privacy's broader potential [1]. It introduces carefully calibrated noise (bounded by a parameter ε) to the data, obscuring individual information while preserving overall population statistics. This process inherently alters the original data values and distribution shape, preventing unwanted identification of individuals. However, this alteration can also impact the data's utility [60], [89]. Yet, a distinctive advantage of differential privacy over other data anonymization techniques (e.g., k-anonymity [72], l-diversity [48], t-closeness [46]) is its ability to quantify the induced error. This key feature allows stakeholders to incorporate this error when making critical decisions, providing a clear advantage over other methods that offer less predictability and control [6]. Thus, the Census Bureau's use of differential privacy exemplifies its critical role in data privacy: providing a robust mathematical framework for balancing data utility with privacy protection.

Despite its distinct advantages and important applications, the widespread adoption of differential privacy in organizational workflows is still in its infancy [23], [28], [47]. While some success stories exist, the factors contributing to effective implementations remain under-explored. Existing literature often focuses on final technical decisions, overlooking the implementation process and its challenges [3], [32], [33], [73]. This gap hinders researchers' ability to address practical issues effectively [16].

Recently organizations have begun to fill this need for shared information by providing insights into their implementation processes. The publication associated with Wikimedia's differential privacy data release presents a six-stage implementation process figure and briefly discusses the associated challenges [87]. Urban Institute [77] and Tumult Labs [88] provide implementation workflow diagrams but lack any further detail. Our implementation process description encompasses these previously delineated stages and extends them by describing the process of learning and introducing differential privacy and the the challenges faced after deployment.

Additionally, while the challenge of having many stakeholders is often discussed [10], [16], only Cummings et al. provide a detailed stakeholder analysis and characterization [15]. They identify six groups categorized into end users and implementation stakeholders. This work provides a basis for the actors used in our research and highlights the importance and difficulty of negotiation and compromise among these groups [16], [50].

Our research expands the documentation of implementation challenges and stakeholder negotiations by providing a detailed description of the differential privacy implementation workflow, supported by insights from various organizations and individuals. We confirm the breakdown of roles found through our interviews and Cummings et al.'s [15] work through a follow up interview. We then extend their work by introducing two new axes—DP knowledge and technical proficiency—that help explain why certain communication challenges occur. Additionally, we outline the primary communication channels between various actors

throughout the implementation pipeline highlighting areas where visualizations can aid in the implementation process. This work builds a more comprehensive picture by combining implementation stages alongside stakeholder negotiations into one cohesive picture.

2.2 Differential Privacy and Visualization in Practice

Differential privacy has progressed from a theoretical concept to being deployed in various industries [22]. Major tech giants such as Apple [74], Google [3], Microsoft [21], Meta [44], Uber [40], and LinkedIn [43] have embraced the power of differential privacy to analyze and release user data. They have also contributed to the development of open-source systems and frameworks like OpenDP [68], OpenMined [55], and Diffprivlib [35] to democratize accessibility to differential privacy tools. Additionally, new companies like Tumult Labs [7] have emerged to offer differential privacy as a service. Lastly, government agencies, such as the U.S. Census [1], have also adopted differential privacy for their data releases.

Despite its growing popularity, differential privacy is still not widely adopted or accepted [28], [47]. Damien Desfontaines purports to list all major adoptions of differential privacy as of May 2024 by industry in one short webpage [19]. In the few pieces of literature examining differential privacy deployments [23], [28], there are many reasons listed for the slow uptake, including a lack of trust in the private data's accuracy, difficulties in integrating differential privacy into data analysis workflows, and challenges in setting parameters and standards.

While there are many challenges, visualization is pointed to as a potential aid in deploying differential privacy and increasing its accessibility [23], [28]. Garrido et al. [28] conducted interviews with industry practitioners not experienced with differential privacy to understand the barriers to adopting differential privacy. Their findings indicate a desire for visual dashboards and tools to inspect data quality and the impact of differential privacy measures. Meanwhile, Dwork et al. [23] interviewed experts in differential privacy from various industries and found that simulations and examples helped promote understanding and buy-in.

Despite these promising indications, visualization remains largely underutilized in differential privacy deployments. Work by Garrido et al. [28] indicates that open-source frameworks like OpenMined [55], Tumult Labs [7], and DiffPrivLib [35] rarely incorporate visualizations into their software. Visualization is used sparingly in documentation to illustrate the concept of differential privacy, the impact of parameter choices, and to demonstrate the results of anonymizing data. An exception to this trend is when tech companies collaborate with non-profit organizations, as seen in Facebook Mobility reports [32], Google Covid Mobility reports [3], or Microsoft IOM reports [24] and the DPCreator tool generated by OpenDP [67]. Visualizations are often the main way data is presented in these cases. Similarly, in cases where data is released for public use, such as in the Census [1], accompanying visualizations and educational resources on differential privacy are typically provided.

The use of visualizations in differential privacy is becoming more prevalent, underscoring the need for comprehensive research into its role. Yet, academic attention to visualizations may be limited due to their frequent internal use and lack of documentation [41]. Research is needed to see how it is used and whether it has the potential to address some of the challenges and misconceptions surrounding differential privacy.

2.3 Communicating Differential Privacy Theory Visually

Due to its widespread applicability, people of all backgrounds are required to interact and understand differential privacy. Since it is a complex topic, many users may not fully comprehend the theory, risks, or implications of working with differentially private systems [41]. Research has investigated how visualizations can effectively communicate risks and foundational knowledge of differential privacy [11], [41], [70], [90]. Studies have explored various visualizations, such as visual metaphors [41], spinner visuals [11], heatmaps [90], hypothetical outcome plots [70], and frequency-framed visualizations [53] to understand their impact on users' decision-making when it comes to sharing personal data.

The findings of these studies are varied. Some show that visualizations can increase users' knowledge of differential privacy by demonstrating concepts such as 'the noise is increasing privacy' [41]. Other times visualization led to surprising implications, such as when participants chose to be 'more honest' by choosing a lower level of anonymization [11] or were overconfident in their self-assessed comprehension of differential privacy [70]. The varying interpretations of differential privacy theory and parameters can lead to issues and barriers that may impact decision-making in real-world deployments. To understand how different levels of comprehension and interpretation of differential privacy theory and parameters can cause implementation barriers, it is vital to investigate how practitioners successfully develop an understanding of differential privacy and the consequences when they do not.

2.4 Visualization for Implementation Parameters

Research has explored visualization techniques aimed at supporting data custodians in their decision-making process for the release of differentially private data. Visualizations are often embedded in software that helps users set implementation parameters. Nanayakkara et al. [52] employ quantile dot plots and hypothetical outcome plots to assist in setting and splitting epsilon. John et al. [39] use a 3D line chart to reduce the complexity of privacy parameter setting. Additionally, there is work aimed at helping data custodians publish their data with a differential privacy guarantee, using visualizations at various stages of the privacy preservation pipeline to aid in privacy-utility tradeoff decisions. DPVisCreator [94] provides histograms, scatterplots, and statistical metrics to help data custodians evaluate if visual patterns are retained at varying privacy levels. Using node-link diagrams, Wang et al. [84] allow for the manipulation of privacy models and assess their impact on data utility. Overlook [75] employs histogram comparisons to aid in setting epsilon for different data columns. These visual tools can help users understand and evaluate implementation parameters (such as the privacy parameter epsilon).

While these tools are useful, they often overlook specific challenges some practitioners face (e.g., Cummings et al.'s discussion [15]). For example, practitioners deal with more than just the epsilon setting; they also struggle with choosing parameters such as delta, data bounds, and interpreting accuracy metrics [15], [67]. Cummings et al.'s work points out some unaddressed concerns and demonstrates that there could be additional unacknowledged practitioner needs. Through our interviews, we see how or if these challenges are addressed and gain insights into the visualization tools that are favored, dismissed, or ignored in differential privacy workflows. These understandings can guide the development of more intuitive, user-friendly tools tailored for implementing differential privacy.

2.5 Visualizations for Differential Privacy Releases

Once the data has gone through the differential privacy implementation process, the resulting noisy data is released to the public. Due to the size of the datasets involved, the data is often presented using visualizations. Researchers have investigated the accuracy and trust of users when working with visualizations involving differentially private data. Hay et al. [31] and Zhang et al. [92] have explored the challenges of visualizing data with differential privacy and the impact that differential privacy can have on the utility of visualizations. These studies have raised concerns about the effects of differential privacy on visual pattern retention. To address this, other researchers have conducted empirical studies to quantify the effects of differential privacy on visual utility using a limited set of visualizations [60], [93].

Despite these efforts, it is still uncertain if laboratory experiments extend to real-world applications. The current literature examines a narrow subset of users and does not ask them to complete more than very basic analysis tasks with differentially private data. Previous reviews of real deployments [10] indicate that the challenges can be far more than just the ability to pull out reliable information from the data. Individuals may not trust the data enough in the first place or need additional proof that the resulting analysis would remain the same. Additionally, in real deployments, certain releases such as the U.S. Census [1] display uncertainty but do not incorporate it in their visualizations, while others such as Google Mobility reports [3] discuss the uncertainty in technical documentation but do not display it publicly. These challenges parallel the issues uncertainty visualization research has worked to solve [38]. Relevant research for differential privacy can be found in how uncertainty visualization affects trust [58], decision quality [25], confidence in data analysis [13].

While this work is a good foundation, the uncertainty stemming from differential privacy is deliberate and, therefore, may change the results of previous studies. It is currently unknown what the motivations and feedback on data release strategies have been for large-scale deployments of differentially private data releases. Through interviews with differential privacy practitioners, we can gain a valuable understanding of their motivations for choosing how to present the idea of noise addition, as well as users' responses to these deployments. This can guide practitioners on how to effectively communicate differential privacy to end-users in various use cases and ensure the successful deployment of differentially private data.

3 Interview Study Method

We conducted an interview study with 18 industry professionals (Data Custodian, Differential Privacy Expert, Domain Expert, and Management). Our study aimed to achieve three objectives: (**O1**) understand the differential privacy implementation pipeline and the associated challenges and opportunities; (**O2**) explore the different actors involved in the pipeline, their interactions, and use of visualizations; (**O3**) identify the benefits and challenges of using visualization in the implementation process.

This section outlines the interview procedures and data analysis process employed in the study. We adopted a semi-structured interview approach [20] to accomplish the study objectives, gain a deep understanding of various actors' perspectives, experiences, and challenges, and enable follow-up questions. While our study focused on understanding the role of visualization within the differential privacy pipeline, emphasizing its potential benefits in this context, we also paid close attention to instances where the interviewees highlighted the challenges of utilizing visualization or suggested alternative methodologies. We recognize that the field of differential privacy is still maturing, and the challenges it presents are complex. While most existing work concentrated on technical solutions of differential privacy, our approach diverts from the mainstream to explore how visualizations could potentially address some of these challenges. Our interview questions are in our supplemental materials [4].

3.1 Participants

We recruited 18 participants with expertise in differential privacy from industrial settings and holding various job responsibilities in organizations. The participants were knowledgeable about the latest developments and trends in the field of differential privacy, which was crucial for identifying potential research opportunities and challenges. Utilizing the snowball sampling strategy [45] from our personal and professional networks, we targeted individuals working with differential privacy.

We sought to recruit a diverse group of participants by expanding our search to online platforms, such as relevant Slack channels and forums dedicated to differential privacy—e.g., OpenDP [68] and OpenMined [55]. Although all the participants were based in the U.S., they represented diverse positions within their organizations, which helped to provide a well-rounded understanding of differential privacy and visualization practices within the country. Furthermore, the participants held a variety of job titles, including senior scientists, senior research engineers, privacy engineers, and federal employees.

Correspondence with participants occurred primarily through direct email, informing them of the study's objectives. Participation in the study was voluntary, and participants were provided with a $50 Amazon gift card at the conclusion of the interview, except for seven federal employees who declined the gift card due to institutional regulations. Table 1 provides information on the interview participants, including their job responsibilities in their organization, organization type, and their background with differential privacy. Additional background or demographic information cannot be provided due to the small sample of individuals working in this domain.

3.2 Procedure

After obtaining approval from our Institutional Review Board (IRB), we emailed prospective participants and dispatched the approved consent form to them. Upon receiving consent, we scheduled semi-structured interviews, which were conducted online via Zoom and ranged in duration from 30 to 45 minutes. We began each interview with a brief introduction outlining the study's primary objectives and the interview process. We asked participants questions concerning their job responsibilities in their organization, the current practices and barriers implementing differential privacy, and how visualizations intersect with their work in differential privacy. Due to the semi-structured nature of our interviews, we asked follow-up questions enabling a more thorough and nuanced exploration of the relevant topics and resulting in a richer and more insightful data collection process.

It is important to acknowledge that visualizations may not be the solution in all stages of the differential privacy pipeline. Yet, prior work has proven the effectiveness of visualization in enriching the understanding of differential privacy processes [23], [28], [52],

TABLE 1: We conducted 18 interviews with differential privacy practitioners. This table shows their high-level job responsibilities, organization type, and backgrounds with differential privacy.

ID	Responsibilities	Organization	Differential Privacy Background
P1	Data Custodian	Non-Profit	Navigates the tension between data transparency & privacy
P2	Data Custodian	Government	Involved in several DP releases, worked on managing and protecting sensitive data
P3	Data Custodian	Government	Worked with the DP experts and stakeholders to find a balance between data accuracy and privacy
P4	Data Custodian	Government	Developed privacy methods for releasing DP data, involved in decision-making for privacy budget choices
P5	Data Custodian	Government	Ensure the accuracy of data, Worked closely with external stakeholders to help them understand the data
P6	DP Expert	Industry	Involved in contributing to the theory of differential privacy, conducting research, and writing research papers
P7	DP Expert	Government	Internal developer of privacy algorithms, scientific lead on DP projects, explain privacy algorithm
P8	DP Expert	Industry	Involved in designing and developing differential private methods
P9	DP Expert	Industry	Algorithm designer for data privacy, communicates DP to practitioners
P10	DP Expert	Industry	Communicates DP to technical and non-technical audiences, approves anonymization strategies
P11	DP Expert	Industry	Developed an interactive interface to aid in selecting privacy budgets while balancing the accuracy of data
P12	DP Expert	Industry	Taught DP to non-technical individuals, and developed privacy-preserving algorithms
P13	Domain Expert	Non-Profit	Involved in consulting on DP research projects, providing quality checks on proposals
P14	Domain Expert	Industry	Explained differential privacy to different stakeholders
P15	Domain Expert	Industry	Involved in conveying data insights to experts and management on several DP projects
P16	Management	Government	Involved in explaining DP to non-technical individuals, supervised the development of DP implementation
P17	Management	Industry	Led several projects on differential privacy, and made decisions on privacy parameters
P18	Management	Government	Supervised the implementation of differential privacy and enhanced privacy measures for all data

[67], [86], [90]. Thus, the emphasis on the role of visualization is the main goal of this paper and a perspective we have striven to incorporate into our study. It is notable that significant attention has been given to technical aspects of differential privacy, such as mathematics and statistics. However, such a focus tends to limit exploration in alternative areas that could potentially offer novel insights or solutions. Therefore, by making our interest in visualization explicit to our study participants, we placed the spotlight on visualization, enabling us to unearth new, visualization-centric approaches that have demonstrated promise in the field of differential privacy. For each interviewee, we asked six main questions found in our supplemental materials [4]. Two members of our research team (first and second authors) conducted interviews in tandem to facilitate note-taking and ensure asking appropriate follow-up questions. We documented each interview through detailed notes and audio recordings.

3.3 Data Gathering and Analysis

We collected approximately 700 minutes of audio recordings from 18 interviews and transcribed each one. We formatted the transcriptions using spreadsheets with a separate sheet for each participant. Each sheet contained separate cells for participant ID, notes, questions, follow-ups, and answers. The gathered data were analyzed in two phases. During the initial phase, two independent coders (first and second authors) analyzed six randomly selected interviews (P1, P4, P7, P10, P12, P17). The remaining 12 interviews are analyzed in the second phase of our analysis. They created another spreadsheet with a separate sheet corresponding to one of the eight themes (Table 2). We derived the themes by mapping the research objectives to eight interrogative questions mentioned in Table 2, which then guided the analysis of the transcriptions. For example, one of our research objectives was to explore the actors involved in the differential privacy pipeline, so we developed an interrogative question of '**Who** is using visualization with differential privacy?'. Refer to Table 2 for more information on the mapping process. This method of interrogative questions has been employed to categorize relevant literature in other domains, demonstrating the effectiveness of this approach in uncovering crucial insights [34].

TABLE 2: This table shows the mapping process of our study objectives (see Section 3 for objective descriptions) with the eight interrogative questions. The bolded question words (**When**, **Who**, **Why**, . . .) are the derived themes.

Objectives	Themes
O1	• **When** is visualization used in the DP pipeline?
O2	• **Who** is using visualization with DP?
O3	Benefits: • **Why** is visualization used with DP? • **What** is visualized in DP? • **How** is visualization used with DP? Challenges: • **Why** is visualization **not** used with DP? • **What** is **not** visualized in DP? • **How** is visualization **not** used with DP?

We analyzed the data using a deductive thematic analysis [12] utilizing the derived themes that reflect our study objectives. First, the coders refined the definitions of the derived themes. They analyzed three interviews and extracted quotes, adding them to a shared sheet. They removed redundant quotes and collaboratively reached a consensus on the quote dataset for each interview. Afterward, the two coders independently added the quotes to the corresponding theme sheets. They placed any unclear quote in a separate sheet with an additional column explaining the dilemma. After analyzing each interview, the coders convened to review and discuss the unclear quotes they had identified. They also examined each other's coding for consistency and accuracy. In the case of

TABLE 3: We conducted five follow-up interviews with differential privacy practitioners. This table shows their high-level job responsibilities, organization type, and backgrounds with differential privacy.

ID	Responsibilities	Organization	Differential Privacy Background
P19	DP Expert	Academia	Designed improved methods to communicate the statistical nature of differential privacy
P20	DP Expert	Industry	Developed tools and methodologies to release differentially private statistics
P21	Data Custodian	Government	Assisted in modernizing PET's by implementing differential privacy
P22	Data Custodian	Industry	Researched solutions to implementing differentially private server
P23	DP Expert	Industry	Assisted various organizations in implementing differential privacy strategies

disagreements on particular quotes, they reviewed the relevant data, including revisiting the transcription and listening to the audio recording, to better understand each other's perspectives and reach a consensus. After analyzing the third interview and this iterative process, the coders agreed that they fully understood the themes, and the definitions had reached their final form (supplemental materials: [4]). To ensure agreement on the proper data categorization in the themes, they calculated the inter-rater reliability on the remaining three interviews, obtaining a Cohen's kappa score of 0.82, which signifies a strong agreement between the two coders [29]. From this point on, in the second phase of the analysis, the coders analyzed the remaining 12 interviews (6 by coder 1 and 6 by coder 2). Once all the transcriptions were analyzed, they discerned the patterns in the data in each theme to answer the research objectives.

3.4 Follow-up Interviews

Building on our initial study, we conducted follow-up interviews with five additional differential privacy practitioners (information in Table 3). In these interviews, we asked participants to examine and discuss the stages and actors of our proposed pipeline by showing them Fig. 1. Our goal was to ensure with a new independent group of practitioners that we had correctly characterized the implementation process. Furthermore, to help delineate the differences between differential privacy and data science workflows, we asked them about the unique aspects of working with differential privacy they had not experienced before in their professional careers. Finally, participants were to show visualizations they had created or encountered and assign them to the appropriate implementation stage. The interview questions for the follow-up interviews can be found in the supplemental materials [4].

4 Results

In this section, we first introduce the actors involved in differential privacy implementation, their activities, and the challenges they face. We then propose a pipeline that captures the high-level stages of differential privacy implementation, drawing on the findings of our interview study (Fig. 1). Finally, we use this pipeline to structure our investigation of visualization's role (in terms of benefits and challenges) in differential privacy implementation.

4.1 Differential Privacy Implementation Pipeline

4.1.1 Primary Actors

The individuals involved in the differential privacy pipeline come from various backgrounds and technical levels. This diversity of backgrounds is one of the key challenges in implementing and creating tools for differential privacy implementation [28]. All of our interviewees agreed that it is important to consider the audience when communicating differential privacy concepts such as privacy guarantees or implementation parameters. Technical terms may be suitable for mathematicians or statisticians, but a more general audience may need other forms of aid, such as visual representations, to understand the concepts better.

We found two common characteristics among our interviewees involved in the differential privacy pipeline. First is their level of understanding of differential privacy. Most studies divide actors into two groups: those with no differential privacy experience [28] vs. differential privacy experts [23]. Our interviews, however, revealed that this dichotomous division does not fully capture actors with varying levels of differential privacy understanding. Hence, based on our qualitative data analysis, we have categorized the understanding level of actors involved in the differential privacy pipeline into five distinct groups: Awareness, Familiarity, Comprehension, Expertise, and Mastery. We provide more information on this understanding level in our supplemental materials [4]. Problems often arise due to varying levels of understanding among the actors involved, leading to issues with communication and roadblocks along the implementation pipeline. For instance, actors with less differential privacy understanding or experience often must make important technical decisions such as decisions related to the use of differential privacy, benefits, and ethical considerations [16]. One participant noted:

> Technical people are the only ones who really understand it [differential privacy], and policy people are the least equipped to actually make decisions, even though they are ultimately the ones who need to make them. (P13)

The second important characteristic is the actor's background as either non-technical or technical. Broadly, we can characterize these two groups as follows: *Technical people*: Experts in math, statistics, computer science, data analysis, and programming who are proficient with complex concepts and algorithms. *Non-technical people*: People knowledgeable in areas like law, policy, or ethics, without strong technical backgrounds but aware of differential privacy implications and involved in decision-making. Technical individuals with mathematical expertise may be able to understand and handle complex explanations and theories. However, non-technical individuals may benefit from more straightforward explanations and visuals to facilitate their understanding.

> We need to ensure that everyone involved understands the uncertainty and can account for it in their analysis. If they are technical people, such as statisticians and scientists, they are smart enough to handle this complexity. But when communicating with government officials, it may be necessary to provide simpler explanations of uncertainty, maybe using some graphs. (P10)

Individuals can apply differential privacy guarantees on their own data and serve all the actors' responsibilities, however, all of our interviews showed that implementation in practice requires collaboration by people with varying expertise. This parallels other

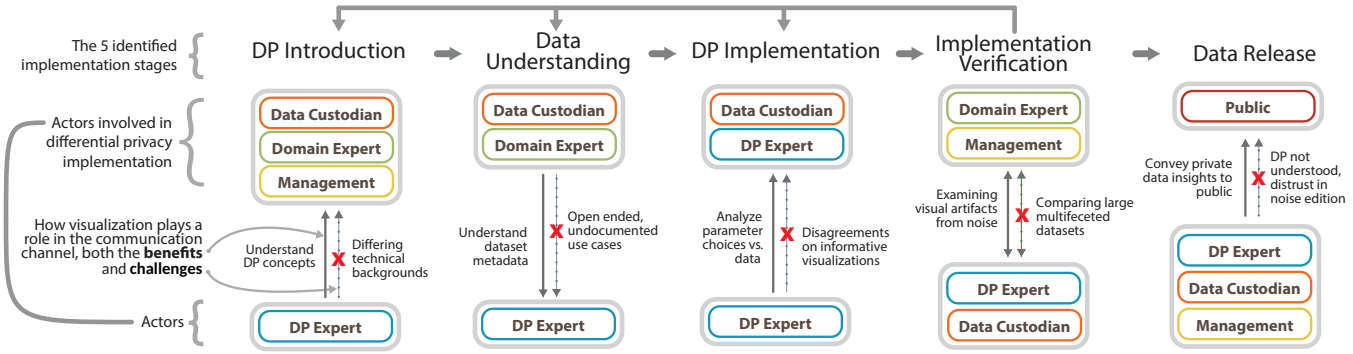

Figure 1: The figure depicts the five stages of the differential privacy (DP) pipeline, highlighting the actors involved in each stage. Arrows connecting the actor blocks signify their communication. Solid arrows represent instances where visualization aids in communication (Benefits), and in contrast, crossed arrows indicate difficulties (Challenges) in visualization use. The arrows between the pipeline stages represent the flow from one stage to another. Problems in a particular stage can cause a return to previous stages. The terms 'Benefits' and 'Challenges' refer to the advantages and challenges of using visualization within the pipeline, respectively, as discussed in Section 4.2.

complex multi-stakeholder workflows, such as the data science pipeline, where collaborative work between actors with diverse expertise is necessary for successful implementation [91]. Based on our careful data analysis, we identify this diversity of actors and broadly characterize them as ***Data Custodians***, ***Differential Privacy Experts***, ***Domain Experts***, ***Management***, and the ***Public***. These actors were identified organically by asking participants about their job responsibilities in the field of differential privacy. We analyzed and grouped similar job descriptions, eventually identifying these five distinct actors. It is worth noting that some participants played multiple job responsibilities within this context.

Within the differential privacy pipeline, the engagements of the data custodian and differential privacy expert are vital. Being responsible for safeguarding individuals' identity and the privacy of their sensitive data, the data custodian is integral to every differential privacy implementation. On the other hand, the differential privacy experts, with their specialized knowledge and experience, are critical in translating various actors' privacy requirements and concerns into differential privacy implementation procedures. This involves verifying and overseeing the implementation process. Despite the availability of open-source libraries, tools, and tutorials for implementing differential privacy, the consequences of incorrect implementation—such as reputational damage for organizations and potential legal issues—make the involvement of a differential privacy expert crucial. The need for domain experts, management, and the public depends on the specific project's requirements. Everyone except for the public, and in some cases, the domain experts, has access to the raw data. Table 4 describes all the actors involved in a differential privacy pipeline and the major tasks they perform in the pipeline.

TABLE 4: This table shows the actors involved in the Differential Privacy (DP) pipeline and the major tasks they perform to make the pipeline operational. The data in this table is based on the results of the study.

Actors	Tasks
Data Custodian	• Facilitators of the differential privacy pipeline • Describe the benefits of DP to all the stakeholders • Convince the stakeholders why DP is the right choice • Ensure smooth communication among different actors
DP Expert	• Serve as a reference point for all things related to DP • Responsible for implementing DP guarantees on raw data • Understanding the data for DP implementation • Work closely with other actors specifically data custodians • Ensure privacy is maintained while maintaining utility
Domain Expert	• They are the data and downstream analysis task experts • Communicate information regarding data understanding • Verify and ensure that data can be used for future analysis • Documenting the various use cases for the data • Assist in making decisions on data preprocessing
Management	• Make strategic decisions on privacy-utility trade-offs • Responsible for creating or approving DP policies • Make decisions regarding the choice of differential privacy • Allocate the necessary resources for DP implementation
Public	• Individuals who have access to the private data release • They are the end users of the DP implementation • Responsible for comprehending the DP uncertainty

4.1.2 Pipeline Stages

Based on our qualitative data analysis, we propose that the differential privacy landscape can often be characterized as a pipeline with five distinct stages: ***Differential Privacy Introduction***, ***Data Understanding***, ***Differential Privacy Implementation***, ***Implementation Verification***, and ***Data Release***. Each stage typically involves specific key activities carried out by different actors in the project. By discussing each stage in detail, we aim to provide a clear and practical framework for differential privacy implementation that will be useful for practitioners and researchers alike. To delve deeper into the implementation complexities, we will consider the case of the U.S. Census Bureau's application of differential privacy in the 2020 Census as an example in each stage.

The first stage, **Differential Privacy Introduction**, involves introducing differential privacy concepts to all the stakeholders. This is typically performed by a differential privacy expert, who helps the stakeholders understand the benefits and workings of differential privacy. Tools used at this stage often include educational resources such as online tutorials and academic papers, presentations, and interactive visualization tools to help explain the concepts [86], [90]. In the context of the U.S. Census Bureau's application of differential privacy, this step would have involved the Bureau's

TABLE 5: This table presents the objectives, challenges, and opportunities for visualization research associated with each stage of the Differential Privacy (DP) pipeline. The data in this table is based on the results of the study.

Stages	Objectives	Challenges	Visualization Opportunities
DP Introduction Section 4.2.1	• Comprehending DP theory • Advocating DP adoption • Establishing trust in DP	• Diverse technical backgrounds • Insufficient DP explanatory visuals • Wide variety of DP applications	• Creating visuals for DP education to explain complex concepts, tailored to user's expertise and familiarity • Designing visuals to illustrate privacy risk and the benefits of DP
Data Understanding Section 4.2.2	• Understanding data properties • Understanding use cases • Generating dataset metadata • Identification of data artifacts	• Uncertainty in use cases • Difficulties in describing metadata • Data quality assurance	• Highlighting data artifacts (outliers, skewed data) in visual formats, making it easier to identify and address issues
DP Implementation Section 4.2.3	• Adding DP noise • Privacy parameters tuning • Generating DP release	• Data misinterpretations • Lack of visuals for interpreting DP parameters • Lack of implementation literature	• Developing interactive visualizations that allow for experimentation with DP parameters • Demonstrating hyperparameter effects through visuals, such as the effect of varying epsilon values • Visualizations to aid in privacy negotiations related to tradeoffs in privacy, accuracy, bias, and representation
Implementation Verification Section 4.2.4	• Validation of noise addition • Ensuring privacy guarantees • Checking utility preservation	• Ensuring appropriate data utility • Presenting digestible information • Informed decision-making	• Visualizing noise addition to ensure the DP process does not result in data that is unrealistic or misleading • Illustrating data utility preservation visually, ensuring that data remains useful post-DP application
DP Data Release Section 4.2.5	• Data release to the end-user • Ensuring privacy compliance	• Misconceptions about DP • Distrust in DP noise • Managing privacy regulations	• Visualizing privacy compliance in released data, showing how data aligns with privacy regulations • Developing user-friendly visualizations to enhance the public's understanding and trust in DP-protected data

privacy experts explaining the importance of privacy preservation and the mechanisms of differential privacy to stakeholders such as government officials and lawmakers.

In the **Data Understanding** stage, data custodians and differential privacy experts often work together to understand the specifics of the data to be analyzed. They identify and discuss key characteristics of the data (e.g. metadata and distribution) and potential vulnerabilities and privacy risks. Tools used at this stage include data visualization and analysis tools to explore and understand the data [94]. For instance, the 2020 Census data includes vast amounts of sensitive data such as an individual's age, sex, race, and income. Hence, the Bureau's data custodians and differential privacy experts would collaborate to discern the most relevant and beneficial data statistics to release to serve the Bureau's information dissemination agenda and goals while preserving individuals' privacy.

During the **Differential Privacy Implementation** stage, carefully calculated noise is introduced to the data or the query results. The differential privacy expert typically manages this task and uses libraries and software specifically designed to apply differential privacy in data analysis tasks [52]. In the Census example, this stage would have involved adding noise to the collected data. This step is crucial, as the noise level needs to be carefully tailored to balance the competing needs of data utility and privacy protection. For example, the Bureau might have to determine the appropriate noise levels to apply to population counts for various geographic levels—from the national level down to city blocks. These counts are vital for many purposes, including redistricting [42] and allocation of federal funds [63]. Too much noise may result in distorted population counts that, for instance, lead to incorrect redistricting decisions. Too little noise, on the other hand, could potentially compromise privacy.

Once the implementation is complete, the **Implementation Verification** stage begins. Typically, data custodians and differential privacy experts inspect the differential privacy settings, parameters, and outcomes to ensure the accuracy of the information and the correct execution of the differential privacy process. The inspection often involves statistical checks and sometimes the use of tools (e.g., Google's Differential Privacy Library [30], IBM's Diffprivlib [35]) for privacy auditing [28]. Differential privacy experts and data custodians communicate several times with domain experts and management, ensuring that the verification process and outcomes meet their requirements and provisions. For instance, in the case of the 2020 census, the demographic analysis team, a group of domain experts that work closely with census data to study population characteristics, would check and verify if they could carry out their analyses effectively with the differential private noise added to the data. The pipeline would move forward if the results were consistent with their expectations and the noise was within acceptable limits. Otherwise, the demographics team would provide feedback to the differential privacy experts, and the privacy parameters would be adjusted accordingly, triggering another round of checks.

The final stage, **Data Release**, involves releasing the privacy-preserving results to the relevant stakeholders and audience. This process is a collaborative effort between the differential privacy experts, data custodians, and project management. The output could be in the form of a report, an interactive data analysis dashboard, or a simple CSV or TSV file, depending on the specifics of the project. In all cases, ensuring that the released data maintains the intended privacy guarantees while still being useful for the stakeholders' specific needs is crucial. For the 2020 Census, this meant releasing the privacy-preserving data in a form that met the end users' essential requirements while adequately protecting the privacy of the individuals represented in the data. The Bureau also took necessary provisions to increase the public understanding and trust of the differentially private information released. This was executed via conducting public seminars, holding workshops, and providing tutorials [81], [82]. These educational sessions focused

on explaining the differential privacy technique, its importance, and the implications for data accuracy. They also emphasized that introducing noise was a necessary trade-off for maintaining privacy. Moreover, in collaboration with differential privacy experts, the Bureau developed a series of publicly accessible materials, including blog posts [79], infographics [78], and videos [80]. These resources clearly communicated the benefits of differential privacy, explaining that despite the introduction of noise, the data remained valid and reliable for essential government functions and public use.

Table 5 describes the primary objectives, major challenges to overcome to keep the pipeline operational, and opportunities for the visualization researchers to contribute in each stage of the pipeline. It is important to note that a typical implementation of differential privacy often involves all stages. However, the emphasis and time spent on a stage can vary significantly based on the project's context, goals, and specific requirements. Moreover, the structure we outline here represents the high-level stages of the pipeline based on our data analysis. Different scenarios, data types, and goals may necessitate adjustments and deviations from this pipeline.

4.2 The Role of Visualization

This section examines the communication between the actors involved and the extent to which visualization is employed within each stage in the differential privacy pipeline. For each stage, we first show the communication channel among the actors. We then report how visualization aids in accomplishing the stage's objectives, and discuss the challenges arising from its use or absence, respectively organized under **Benefits** and **Challenges**. We use the themes outlined in Table 2 under **(O3)** to describe our findings from the interview data. Fig. 1 shows the Benefits and Challenges of using visualization broken down by differential privacy implementation stages.

4.2.1 Stage 1: Differential Privacy Introduction

DP Expert → Data Custodian, Management, Domain Expert

Benefits: In this stage, differential privacy experts must convince data custodians, management, and domain experts that differential privacy is the right choice and help the data custodians become familiar with differential privacy principles. A crucial task in this stage involves educating non-technical actors about the essential concepts of differential privacy. Establishing this common understanding is necessary to advance the implementation forward and facilitate discussions about complex differential privacy implementation criteria. According to one participant, "*visualization is seen as crucial in the early stage in making DP concepts accessible and understandable - P10*", emphasizing the vital role visualization plays in this stage to translate complex differential privacy principles into digestible visual formats. Another interviewee explains privacy transformations by simply turning a *green block* (P10), representing confidential data, into a circle, or using *"little stick figures to [...] visually tell how differential privacy worked - P10"*. These visualizations stand out not just for their simplicity but also for their originality.

Based on our findings, the focus in this stage is on using visualization for communicating fundamental privacy concepts such as the explanation of data sensitivity, the impact of anonymity, the implications of privacy budgets, and the effects of noise addition. For instance, one participant mentioned using a "*presentation ... to explain the difference between differential privacy and anonymity with concrete explanation - P6*". Moreover, differential privacy experts frequently utilize visual representations like bar charts to illustrate the difference between true counts and those altered by differential privacy noise for a count query. An interviewee described this approach:

> Visualization is used to show differences in the data distribution before and after applying DP noise to emphasize that the overall shape of the distribution remains largely the same. (P3)

Our interview data shows that differential privacy experts adopt various visualization methodologies tailored to the audience's expertise and familiarity with differential privacy concepts. One participant described:

> Every single time I start a new project with a team that I haven't already worked with, I will send them the high-level flowchart of the process first and then some graphs that describe both the low-level concepts and the context in which it makes sense to use this [differential privacy]. (P7)

This approach underscores the strategic use of visualization to aid understanding, starting from broad overviews to more detailed, concept-specific illustrations. The introduction of privacy loss vs. accuracy graphs is another method a differential privacy expert mentioned, designed to visualize the trade-off between privacy protection and data utility: "*So one of the initial things we tried was making these kinds of privacy loss vs. accuracy graphs. Showing kind of privacy loss frontier - P8*". This visualization technique effectively communicates the critical decision-making aspect of selecting privacy loss budgets, a cornerstone in the application of differential privacy.

Challenges: While the introduction stage is crucial for the success of the differential privacy implementation, people often misinterpret the parameters or are left unconvinced that differential privacy is appropriate and needed. Our data analysis reveals that a diverse range of technical backgrounds and varying familiarity with the underlying statistical concepts of the actors make designing effective and easily understandable visualizations challenging. Adding further complexity is the wide variety of applications within differential privacy, requiring different visualizations for each scenario. As one interviewee highlighted, "*the diversity of applications and measurements presents challenges in creating effective visualizations - P10*".

Moreover, "limited educational resources" and the "lack of established visualization frameworks" for differential privacy concepts can inhibit the use of visual aids during the introduction stage, as mentioned by several interviewees (P1, P4, P7). Several high-level concepts, such as attack models, privacy budget, individual contribution, and data bounds, pose challenges when attempting to visualize them. This difficulty arises because they are inherently abstract and complex. Consequently, "*these concepts are seldom represented in well-documented visualizations - P6"*. Furthermore, common visualizations such as bimodal distribution curves to illustrate privacy guarantees or individual contribution may not effectively communicate the concept to people with less technical expertise. An interviewee noted:

> I think some visual communication, for example, an interactive web app or even a static visualization, could help show how limiting a person's contribution works. It just does not seem to be effectively explained right now. (P1)

4.2.2 Stage 2: Data Understanding

DP Expert ⟵ Domain Expert, Data Custodian

Benefits: The information flow in this stage is reversed, where the communication is now less focused on differential privacy but instead on the exploration of data characteristics, error distributions,

and the identification of potential artifacts. Our analysis reveals that data custodians and domain experts must build consensus on the use cases and error metrics to ensure data utility. Subsequently, they convey this information to differential privacy experts, providing them with an understanding of the dataset metadata so "*the noise does not create undesirable outcomes - P4*".

Based on our interview findings, the rationale for utilizing visualization in the data understanding stage is multifaceted. Primarily, it aids in exploring and comprehending the underlying data structure. As one interviewee noted, the aim was to "*build approaches to visually identify signatures of undesirable artifacts in the data - P4*", highlighting the need for visualizations that help users detect anomalies that could potentially compromise privacy or data utility. Another participant shared that the visualization of raw data helped clarify how the data should be presented to the differential privacy experts.

> We plan to release a ranked bar chart showing the most viewed pages from specific countries, leading to a natural question about how rankings might shift between actual and privacy-protected ('fake') data. (P5)

Moreover, our findings suggest that data custodians can more easily identify and explore potential quality issues within the dataset by visualizing the raw data. For instance, missing values or inconsistent data entries become more apparent in a graphical representation, which helps clarify the data quality metrics. Furthermore, one participant mentioned using visuals to gain a deeper understanding of data relationships, describing "visualization" as a *more straightforward* method for understanding the data.

> For joint distributions, we used heat maps and scatter plots or binned scatter plots to visualize the data. This is particularly relevant for some of the tabular and geographical data we've dealt with. Visualizations made it more straightforward. (P9)

Challenges: During the data understanding stage, the varied nature of datasets poses significant challenges. With a wide array of possible applications and interpretations of the data, it becomes difficult for data custodians and domain experts to document and prioritize the statistics that need to be preserved. According to one participant, these variations in use cases lead to issues related to differential privacy implementation and reduced visualization usage. Because it is not possible to "*systematically sit down and write down quantitatively what are all the ways in which people can use the data - P7*". Moreover, each use case might require a unique implementation strategy and the appropriate error metric may vary accordingly. This complexity can discourage the consistent usage of visualizations as these need to be tailored to the specific context. One of our interviewees described this issue:

> We were interested in having a visualization expert try to build approaches to visually identify signatures of undesirable artifacts in the data. This was a very broad mandate for the person working on this. In fact, it might have been too broad, as, in my opinion, this ultimately did not really pan out that well. (P4)

4.2.3 Stage 3: Differential Privacy Implementation

DP Expert $\longrightarrow$ DP Expert, Data Custodian

Benefits: The differential privacy implementation is where the differential privacy experts apply noise to make the output meet differential privacy guarantees. It is important to note that noise can also be applied using software libraries [7], [35], [56], [68]. Based on our findings, differential privacy experts use visual representations in this stage to interpret and communicate specific aspects derived from the differential privacy algorithms to themselves or data custodians. These aspects include privacy-utility trade-offs, how the noise addition impacts the statistical properties of the data, and different privacy parameter choices. As one participant noted, visualization is instrumental in "*helping people design, for example, choose parameters of the DP algorithm to optimize utility-privacy trade-offs and make the right decision for the people who will then use the data - P10*".

The most common visualization practice is visualizing how various hyperparameter choices (clamping, truncation, noise addition) affect the outputted data. For instance, a tool developed by Google [30] allows users to adjust hyperparameters and immediately see the impact on data accuracy. One of our interviewees stated that this tool uses comparative histograms of private vs. non-private data to represent the difference visually, effectively demonstrating "*what part of the inaccuracy of the data comes from noise vs. clamping vs. truncation - P11*".

According to our findings, differential privacy experts often use line charts to plot different epsilon values against corresponding accuracy metrics. An interviewee stated that finding the optimal balance between privacy and utility is still challenging. With the current visualization tools, they pick the approximate value.

> So we might have multiple line graphs. We check how many counties pass a single quality threshold by checking single or multiple privacy thresholds and what happens if we change the quality threshold. We look at those two graphs and kind of get a sense this is approximately the quality and privacy threshold we should pick. (P17)

Moreover, differential privacy experts sometimes use the Receiving Operating Characteristics (ROC) [61] curves to determine the approximate balance between privacy and accuracy. This approach enables decision-makers to identify the inflection point where the trade-off is optimized.

> Our primary data visualization instrument is constructing a receiving operating characteristics (ROC) curve, where the horizontal axis is the epsilon level, and the vertical axis is one minus bias or mean squared error. So, as you move toward zero, it curves up and asymptotes toward zero. Finally, you want to find the inflection point in the curve. (P16)

Challenges: While error metrics and privacy-utility trade-offs were talked about often, few visualizations were developed or publicly available for those parameters in this stage. One interviewee highlighted that large tech companies involved in differential privacy have tailored their "*progress towards visualization and experimentation exclusively for internal developers, rather than the general public - P1*". Another participant stated that "*early disagreements on the kinds of error measures and a lack of familiarity with differential privacy affected the production and use of visualizations - P7*". Moreover, when differential privacy experts needed to see errors, their expertise in analyzing tables and quantiles enabled them to scan and identify potential issues.

> We knew what we were looking for, as we were deeply familiar with the most common patterns due to repeatedly examining the tables and quantiles. [...] The familiarity with the tables was just easier than creating and using any kind of visual format. (P7)

Additionally, different actors have different preferences for visualization tools and techniques, which can cause confusion and inconsistencies in the implementation stage as described by one participant. "*There are so many potential graphs and this is overwhelming for people who want to use visualization - P3*".

4.2.4 Stage 4: Implementation Verification

DP Expert, Data Custodian $\longleftrightarrow$ Domain Expert, Management

Benefits: The implementation verification stage aims to rigorously test and evaluate the differential privacy algorithm and noise

addition process to ensure the data meets appropriate quality standards. According to our data analysis, visualization plays a significant role in implementation verification. It enables differential privacy experts and data custodians to detect potential privacy vulnerabilities and improve the differential privacy algorithm's performance. For example, one participant noted, "*heat maps are used to visually represent the distribution of geographical data before and after the application of differential privacy - P3*". These heat maps highlight areas where the noise addition might be causing significant data distortion, thereby indicating potential areas for algorithmic refinement.

Our interview data reveal that data custodians and differential privacy experts communicate back and forth to identify and resolve any undesirable artifacts, outliers, marginals, data leakage, or inadequate randomization of the noise addition process. The data custodians then communicate the findings with domain experts and management for final approval. To aid in this communication, interactive dashboards allow management to explore the data and its transformations through differential privacy by themselves.

> Visualizations play a vital role when it comes to conversations with management, who often only get brief glimpses of the process outputs. For them, well-packaged visualizations like dashboards are extremely useful. I would like to reach a point where the pipelines for producing these visualizations are more systematic as part of error and privacy analysis. (P16)

Challenges: Despite the important roles of visualization in this verification stage of the differential privacy pipeline, some challenges prevent data custodians and differential privacy experts from using visualizations effectively. One of our interviewees (P7) mentioned that the primary issue is the lack of time, as creating effective visualizations can be a time-consuming task. Moreover, data custodians and differential privacy experts have to deal with large and complex datasets, which further increases the time and effort required for visualization.

> I think lack of time is the biggest hindrance to visualizations being used. [...] We need to spend a lot of time analyzing the data, we really don't have the time to think and create good visuals. (P7)

Another issue is the difficulty in plotting specific error metrics that are crucial to evaluating the differential privacy algorithm's performance. Specific error metrics are customized to evaluate the performance or impact of differential privacy in particular scenarios, i.e., Precision and Recall in Privacy-Preserving Record Linkage [62], F1-Score for Privacy-Preserving Classification [69]. These error metrics often depend on specific contextual information about the data, making them harder to represent abstractly compared to general metrics (e.g., mean squared error, mean absolute error). This complexity, in turn, makes the visualization process non-trivial. One of our interviewees noted: "*general metrics are easy to plot but not interpretable or useful while specific metrics are hard to plot, but are useful - P3*".

4.2.5 Stage 5: Data Release

DP Expert, Data Custodian, Management ⟶ Public

Benefits: Data release is a critical step in the differential privacy pipeline where data custodians communicate with differential privacy experts and management to build the differential privacy release and finally communicate the data with the end users. Our interviewees (P1, P8, P10, P12) mentioned that visualization at this stage is not merely about presenting data; it is about communicating complex concepts like uncertainty and the effects of differential privacy in a manner that is accessible to a broad audience.

> In terms of representing the results, I think the community mobility reports by Google in the COVID-19 pandemic are actually really nice. This report involved PDFs with graphs showing changes in mobility, with considerations on how to represent uncertainty. (P10)

Moreover, another interviewee shared, "*I often use error bars to communicate the uncertainty and a bar chart visualization inspired by the Washington Post's election results coverage, where they used color gradients to indicate the 95% confidence interval - P1*". This approach helps in quantifying the uncertainty, making it visually understandable for the audience. One participant also mentioned using visualizations to compare differential privacy processed data and the original data to highlight the effects of differential privacy. "*A specific example involves the census applying a new differential privacy algorithm to 2010 data and then showing some histograms to people for comparison - P8*".

Another participant described their approach of using visualization in data release stating that "*we've developed a visualization to walk people through the process [...] how the data come into the algorithm, how they flow through it, and finally get the private output - P12*". This step-by-step visualization aids in understanding how differential privacy works, making the complex process more accessible. Additionally, organizations often provide some resources to explain the changes the data has undergone. They use videos with visualizations [51] to explain the differential privacy process to the end users. Data custodians and differential privacy experts can communicate the potential risks and benefits to the end users by visualizing the uncertainty in the data and explaining the benefits of differential privacy. One participant described this:

> The concept of releasing data that may be flawed or off by a certain amount can be counterintuitive and may cause distrust. To preempt this, it is essential to say, "We know the data is flawed, but in a measurable way, and we'll show that using visualizations". The upside is that this approach lets us protect privacy while releasing more granular data. (P1)

Challenges: While visualization is utilized in certain aspects of the data release stage, challenges persist in effectively communicating the concept within data releases. This can be attributed to several reasons, such as a lack of understanding of the randomization mechanism used in differential privacy or difficulties associated with interpreting noisy data. The randomization mechanism can be complex, and "*end users of the private data often do not care about the details of anonymization. All they want to know is it's anonymous, and that's good enough for them - P17*".

Additionally, the diversity in public expectations complicates this task further, as each individual may seek different assurances from the data. This makes it impractical to satisfy all through a single or even a set of visualizations.

> I haven't seen really great visualizations for showing to the public, ones that would convince people that the data is okay. I think we've struggled to get those kinds of visualizations out because everyone wants to see something different. But I can't put up every possible graph. So you kind of throw your hands up and say, 'Well, I don't know what to do.' (P4)

4.3 Follow-up Results

All participants concurred with the overarching structure of our pipeline and the actors involved, yet highlighted potential subtleties. They noted that while our pipeline reflects the workflow typical of large-scale private data releases (i.e. Census), adopting an approach that facilitates differentially private SQL-like interactions could significantly alter the importance of several phases. Specifically, the implementation phase might shift from developing an algorithm to simply selecting the privacy parameter ε.

The additional visualizations participants showed did not reveal any novel findings as they were primarily line charts focusing on privacy-accuracy trade-offs. While not providing any additional visualizations, each participant continued to express the difficulties in understanding and learning differential privacy as a challenge they had not experienced previously in their careers. They underscored the critical role of visualizations in these educational contexts. This included showing that visuals not only helped explain differential privacy concepts to others but also improved their own understanding of the trade-offs involved through visual exploration. Participants P19 and P20 shared insights on how visualizations featured in a textbook by Near and Abuah helped solidify their understanding of differential privacy fundamentals [54].

5 Discussion

Visualization researchers can make a direct, immediate impact in the growing field of differential privacy. The challenges practitioners face reflect broader questions that visualization research can aid in and learn from. How can visualization assist in multi-faceted decision-making tasks? Which uncertainty visualization strategy can increase trust? How can visualizations help teach a growing number of complex data science topics? As differential privacy transitions from theory to practice, visualization researchers can step in and work collaboratively with differential privacy practitioners. Using our pipeline and breakdown of actors, they can have a common vocabulary to discuss pain points and visualization solutions. As a starting point for this research collaboration, we offer avenues of research for visualization researchers to assist differential privacy practitioners and, in turn, help push visualization research forward.

5.1 Deliberate Uncertainty

One of the largest differences in differential privacy deployments stems from the change in how uncertainty comes into the data. In visualization research, almost every study on uncertainty is in relation to uncertainty *inherent* in the data, whether it be from sampling or modeling [59]. On the other hand, differential privacy introduces *deliberate* uncertainty. This uncertainty is pulled from distributions defined by the practitioner and intentionally obfuscates the preciseness of the data. This pivotal difference opens up new research questions for differential privacy and visualization researchers to tackle together.

One avenue of research should examine the effects of uncertainty visualization strategies in regard to setting privacy parameters. In uncertainty visualization creation and evaluation, research most often asks what visualization will maximize the performance of the end user [38]. Differential privacy uncertainty evaluation would change that by asking what visualization would best help someone decide the level of uncertainty. This raises questions that are partially addressed in the literature. Do different uncertainty visualization designs affect the level of privacy (uncertainty) when people release data? Do different presentations of uncertainty influence practitioners' understanding and interpretation of the privacy risks and utility of their data [53]? Research has highlighted that certain visualizations better illustrate uncertainty and lead to more accurate statistical judgments [8]. We lack similar evaluations on how users reason about and make decisions regarding deliberate uncertainty.

This challenge of appropriate visualization choices is compounded by practitioners having to evaluate the accuracy, fairness, and privacy of thousands or millions of statistics at once [2].

> My impression is that they [data experts] frequently felt overwhelmed by the number of things that might change. And because of that, determining acceptable visualizations became crucial to concretely see what kinds of patterns did and did not emerge. (P7)

As differential privacy matures, this type of visualization knowledge will be important to allow practitioners and researchers to balance privacy and accuracy appropriately. Just as suboptimal uncertainty visualization design can lead to incorrect assumptions by decision-makers, the lack of evaluations on various visualization strategies may result in individuals inadvertently compromising their data protection or releasing unusable data.

5.2 Trust in Differential Privacy

Another avenue of research with immediate implications is examining uncertainty visualizations' effect on users' trust when working with differentially private data. Theoretically, displaying the *inherent* uncertainty typically enhances users' trust in the dataset, as reported in previous literature [65]. On the other hand, as documented by the U.S. Census and the subsequent lawsuits filed against it, the reaction towards noise addition from differential privacy can be negative [10].

The implications of deliberate uncertainty may cause this change in reaction. For example, applying differential privacy could theoretically return a negative value for a count query. This is, as one participant put, *"a counter-intuitive notion and may cause distrust - P1"* but is a perfectly reasonable outcome of adding noise [17]. Displaying uncertainty visualizations (confidence bars, distribution curves) over top of the private negative counts clashes with fundamental data expectations, making standard uncertainty visualizations ineffective in such contexts. While differential privacy implementation, verification, and post-processing steps can mitigate large errors, there are chances the data contradicts users preconceived notions in a similar fashion. Differential privacy experts need help communicating this deliberate uncertainty to not just prevent distrust but also demonstrate the benefits of differential privacy:

> So to preempt that, say, "Hey, we know that this data is flawed, and it's flawed in a measurable way. We're going to communicate that flaw and also highlight the benefit that comes with it." (P1)

We suggest that visualization researchers work on understanding how a general audience interprets and interacts with data released through differential privacy, given its unique element of deliberate noise addition. One way to do this is to leverage Hullman et al.'s uncertainty evaluation methodology [37], focusing on "confidence" as the desired effect. Does knowing that the noise is deliberate rather than inherent cause different reactions to uncertainty visualizations, particularly related to trust? What uncertainty visualization design best supports user confidence and understanding of the differentially private data? Studies employing this approach should also consider the larger context of differential privacy releases, assessing both the trust in data validity and comprehension of differential privacy's purpose [37]. Helping pinpoint the most effective strategies can have immediate practical effects as organizations continue to release their private data. This can reduce the friction stemming from miscommunication about deliberate noise addition.

5.3 Pedagogy for Diverse Backgrounds

Visualization researchers can draw important lessons from the practices of differential privacy experts in using visualizations to simplify and teach complex data science concepts to a variety of

individuals. Unlike data science pipelines, visualizations in the context of differential privacy deployments are heavily focused on pedagogical tasks (DP Introduction) [14]. Almost every interviewee mentioned creating or seeing visualizations explaining the mathematical concepts underlying differential privacy. This reflects a broader trend toward the increased use of pedagogical visualizations in data science [14].

> So, I think what we really need is more visualization to kind of bridge that gap from the complex math behind differential privacy to something more digestible for lay data users. It's about pushing for more education, but in a way that's approachable for people who aren't computer scientists or mathematicians, to really improve collaboration and understanding across the board. (P5)

This task of simplifying complex differential privacy concepts has inspired practitioners to innovate with visual solutions that are creative and effective (see Section 4.2.1). Our interviews revealed a variety of straightforward yet creative techniques for teaching through visualization [9]. Additionally, visualizations were often built up to explain a concept similar to data storytelling: *"the more complicated the concept is, the more like little paving stones, you have to bridge that gap between the user and the expert - P10"* [49]. Furthermore, adapting visualizations to suit different audiences is considered essential, as another practitioner emphasized the need to *"develop different communications visualizations based on specific audiences - P13"* [5].

Working with differential privacy practitioners to explore this rich collection of visualizations and practices could advance our understanding of how to use visualizations to make complex data science topics accessible and engaging to various stakeholders. We can learn from this foundation to build out better visuals. Several differential privacy experts specifically mentioned wanting assistance in crafting these educational visualizations.

> It'd be nice if the DP community had a set of visuals that they've tested out with focus groups with various levels of mathematical ability, who can say this visualization does something for us. Trying to whip up these visuals on your own is tough. If they don't click with your crowd, you might as well toss them out the window. (P4)

This concept of pedagogical visualization is growing in areas such as AI, but there is still much to learn [71], [85]. What visualization design strategies can enhance pedagogical understanding of data science topics? How do these designs change based on the technical background of the audience? How can we adapt data storytelling to teach data science topics? Studying this invites us into an exciting exploration of how visual tools can bridge the gap between complex concepts and broader comprehension.

6 Limitations of the Study

One of the limitations of our study is using the snowball tactic for participant recruitment. This approach may have led to a sample not representative of the broader population working in differential privacy, especially those working on local differential privacy deployments or smaller deployments. It is crucial to acknowledge that the insights obtained from this interview study may be influenced by the interviewees' varying depth and breadth of knowledge on the subject matter. Finally, since all of our participants were based in the US, our study may not fully capture the experiences and perspectives of individuals from different cultural backgrounds or regions, which could limit the generalizability of our findings to a broader population. Future research should aim to address this limitation by utilizing a more diverse sample, including domain experts and participants from a range of deployment sizes, including smaller organizations and individuals applying differential privacy to their personal data.

7 Conclusion

Our study illuminates the critical role of visualization in implementing differential privacy. Through interviews with 18 practitioners, we identified the key stages of the differential privacy pipeline, the responsibilities of the actors involved, and the diverse applications and challenges of visualization at each stage. We found that visualizations are increasingly important in differential privacy implementations and visualization research can have an immediate tangible impact. Visualizations are already being used to assist individuals in introducing differential privacy, selecting implementation parameters, and evaluating private data. Furthermore, the unique challenges in differential privacy related to decision-making, trust, and education open new avenues for collaborative research that mutually benefit privacy practitioners and visualization researchers. Moving forward, we hope this research establishes a common understanding and vocabulary for researchers poised to work in this understudied yet exciting cross-disciplinary area of research.

Acknowledgements

The authors appreciate the comments of each of the anonymous reviewers and would like to thank Melanie Tory and Evanthia Dimara for their feedback. This work was supported by the National Science Foundation SATC grant # 1954814.

References

[1] J. M. Abowd. The US Census Bureau adopts differential privacy. In *Proc. 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 2867–2867, 2018. doi: 10.1145/3219819.3226070

[2] T. Adeleye, S. Berghel, D. Desfontaines, M. Hay, I. Johnson, C. Lemoisson, A. Machanavajjhala, T. Magerlein, G. Modena, D. Pujol, et al. Publishing Wikipedia usage data with strong privacy guarantees. *arXiv preprint arXiv:2308.16298*, 2023. doi: 10.48550/arXiv.2308.16298

[3] A. Aktay, S. Bavadekar, G. Cossoul, J. Davis, D. Desfontaines, A. Fabrikant, E. Gabrilovich, K. Gadepalli, B. Gipson, M. Guevara, et al. Google COVID-19 community mobility reports: anonymization process description (version 1.1). *arXiv preprint arXiv:2004.04145*, 2020. doi: 10.48550/arXiv.2004.04145

[4] Anonymous for peer review. Illuminating the landscape of differential privacy: An interview study on the use of visualization in real-world deployments—online supplement, 2024. Available at osf.io/qhyzt/?view_only=1a5c7d7553c840ab9f125d88bc13946f.

[5] E. P. Baumer, M. Jasim, A. Sarvghad, and N. Mahyar. Of course it's political! a critical inquiry into underemphasized dimensions in civic text visualization. In *Computer Graphics Forum*, vol. 41, pp. 1–14. Wiley Online Library, 2022. doi: 10.1111/cgf.14518

[6] S. H. Begum and F. Nausheen. A comparative analysis of differential privacy vs other privacy mechanisms for big data. In *2018 2nd International Conference on Inventive Systems and Control (ICISC)*, pp. 512–516. IEEE, 2018. doi: 10.1109/ICISC.2018.8399125

[7] S. Berghel, P. Bohannon, D. Desfontaines, C. Estes, S. Haney, L. Hartman, M. Hay, A. Machanavajjhala, T. Magerlein, G. Miklau, et al. Tumult Analytics: a robust, easy-to-use, scalable, and expressive framework for differential privacy. *arXiv preprint arXiv:2212.04133*, 2022. doi: 10.48550/arXiv.2212.04133

[8] G.-P. Bonneau, H.-C. Hege, C. R. Johnson, M. M. Oliveira, K. Potter, P. Rheingans, and T. Schultz. Overview and state-of-the-art of uncertainty visualization. *Scientific visualization: Uncertainty, multifield, biomedical, and scalable visualization*, pp. 3–27, 2014. doi: 10.1007/978-1-4471-6497-5_1

[9] C. M. Bowen. The art of data privacy. *Significance*, 19(1):14–19, 2022. doi: 10.1111/1740-9713.01608

[10] D. Boyd and J. Sarathy. Differential perspectives: Epistemic disconnects surrounding the US Census Bureau's use of differential privacy. *Harvard Data Science Review (Forthcoming)*, 2022. doi: 10.1162/99608f92.66882f0e

[11] B. Bullek, S. Garboski, D. J. Mir, and E. M. Peck. Towards understanding differential privacy: When do people trust randomized response technique? In *Proc. 2017 CHI Conference on Human Factors in Computing Systems*, pp. 3833–3837, 2017. doi: 10.1145/3025453.3025698

[12] A. Castleberry and A. Nolen. Thematic analysis of qualitative research data: Is it as easy as it sounds? *Currents in pharmacy teaching and learning*, 10(6):807–815, 2018. doi: 10.1016/j.cptl.2018.03.019

[13] M. Correll and M. Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2142–2151, 2014. doi: 10.1109/TVCG.2014.2346298

[14] A. Crisan, B. Fiore-Gartland, and M. Tory. Passing the data baton: A retrospective analysis on data science work and workers. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1860–1870, 2020. doi: 10.1109/TVCG.2020.3030340

[15] R. Cummings, D. Desfontaines, D. Evans, R. Geambasu, M. Jagielski, Y. Huang, P. Kairouz, G. Kamath, S. Oh, O. Ohrimenko, et al. Challenges towards the next frontier in privacy. *arXiv preprint arXiv:2304.06929*, 2023. doi: 10.48550/arXiv.2304.06929

[16] R. Cummings and J. Sarathy. Centering policy and practice: Research gaps around usable differential privacy. In *2023 5th IEEE International Conference on Trust, Privacy and Security in Intelligent Systems and Applications (TPS-ISA)*, pp. 122–135. IEEE Computer Society, 2023. doi: 10.1109/TPS-ISA58951.2023.00024

[17] D. Desfontaines. Differential privacy in practice (easy version). https://desfontain.es/privacy/differential-privacy-in-practice.html, 11 2018. Accessed on March 1, 2024. Ted is writing things (personal blog).

[18] D. Desfontaines. *Lowering the cost of anonymization*. PhD thesis, ETH Zurich, 2020. doi: 10.3929/ethz-b-000508570

[19] D. Desfontaines. A list of real-world uses of differential privacy. https://desfontain.es/privacy/real-world-differential-privacy.html, 10 2021. Ted is writing things (personal blog).

[20] B. DiCicco-Bloom and B. F. Crabtree. The qualitative research interview. *Medical education*, 40(4):314–321, 2006. doi: 10.1111/j.1365-2929.2006.02418.x

[21] B. Ding, J. Kulkarni, and S. Yekhanin. Collecting telemetry data privately. *Advances in Neural Information Processing Systems*, 30, 2017. doi: 10.48550/arXiv.1712.01524

[22] C. Dwork. Differential privacy. In *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II 33*, pp. 1–12. Springer, 2006. doi: 10.1007/11787006_1

[23] C. Dwork, N. Kohli, and D. Mulligan. Differential privacy in practice: Expose your epsilons! *Journal of Privacy and Confidentiality*, 9(2), 2019. doi: 10.29012/jpc.689

[24] D. Edge, W. Yang, K. Lytvynets, H. Cook, C. Galez-Davis, H. Darnton, and C. M. White. Design of a privacy-preserving data platform for collaboration against human trafficking. *arXiv preprint arXiv:2005.05688*, 2020. doi: 10.48550/arXiv.2005.05688

[25] M. Fernandes, L. Walls, S. Munson, J. Hullman, and M. Kay. Uncertainty displays using quantile dotplots or cdfs improve transit decision-making. In *Proc. 2018 CHI Conference on Human Factors in Computing Systems*, pp. 1–12, 2018. doi: 10.1145/3173574.3173718

[26] S. Friedman, A. Reynolds, S. Scovill, F. Brassier, R. Campbell, and M. Ballou. An estimate of housing discrimination against same-sex couples. *Available at SSRN 2284243*, 2013. doi: 10.2139/ssrn.2284243

[27] S. L. Garfinkel, J. M. Abowd, and S. Powazek. Issues encountered deploying differential privacy. In *Proc. 2018 Workshop on Privacy in the Electronic Society*, pp. 133–137, 2018. doi: 10.1145/3267323.3268949

[28] G. M. Garrido, X. Liu, F. Matthes, and D. Song. Lessons learned: Surveying the practicality of differential privacy in the industry. *arXiv preprint arXiv:2211.03898*, 2022. doi: 10.48550/arXiv.2211.03898

[29] N. Gisev, J. S. Bell, and T. F. Chen. Interrater agreement and interrater reliability: key concepts, approaches, and applications. *Research in Social and Administrative Pharmacy*, 9(3):330–338, 2013. doi: 10.1016/j.sapharm.2012.04.004

[30] Google. Differential privacy by Google. https://github.com/google/differential-privacy.

[31] M. Hay, A. Machanavajjhala, G. Miklau, Y. Chen, D. Zhang, and G. Bissias. Exploring privacy-accuracy tradeoffs using DPComp. In *Proc. 2016 International Conference on Management of Data*, pp. 2101–2104, 2016. doi: 10.1145/2882903.2899387

[32] A. Herdağdelen and A. Dow. Protecting privacy in facebook mobility data during the COVID-19 response (2020). https://research.fb.com/blog/2020/06/protecting-privacy-in-facebook-mobility-data-during-the-covid-19-response, 2021.

[33] S. Hod and R. Canetti. Differentially private release of Israel's national registry of live births. *arXiv preprint arXiv:2405.00267*, 2024.

[34] F. Hohman, M. Kahng, R. Pienta, and D. H. Chau. Visual analytics in deep learning: An interrogative survey for the next frontiers. *IEEE Transactions on Visualization and Computer Graphics*, 25(8):2674–2693, 2018. doi: 10.1109/TVCG.2018.2843369

[35] N. Holohan, S. Braghin, P. Mac Aonghusa, and K. Levacher. Diffprivlib: the IBM differential privacy library. *arXiv preprint arXiv:1907.02444*, 2019. doi: 10.48550/arXiv.1907.02444

[36] X. Hu, M. Yuan, J. Yao, Y. Deng, L. Chen, Q. Yang, H. Guan, and J. Zeng. Differential privacy in telco big data platform. *Proc. VLDB Endowment*, 8(12):1692–1703, 2015. doi: 10.14778/2824032.2824067

[37] J. Hullman. Why authors don't visualize uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):130–139, 2019. doi: 10.1109/TVCG.2019.2934287

[38] J. Hullman, X. Qiao, M. Correll, A. Kale, and M. Kay. In pursuit of error: A survey of uncertainty visualization evaluation. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):903–913, 2018. doi: 10.1109/TVCG.2018.2864889

[39] M. F. S. John, G. Denker, P. Laud, K. Martiny, A. Pankova, and D. Pavlovic. Decision support for sharing data using differential privacy. In *2021 IEEE Symposium on Visualization for Cyber Security (VizSec)*, pp. 26–35. IEEE, 2021. doi: 10.1109/VizSec53666.2021.00008

[40] N. Johnson, J. P. Near, and D. Song. Towards practical differential privacy for SQL queries. *Proc. VLDB Endowment*, 11(5):526–539, 2018. doi: 10.1145/3187009.3177733

[41] F. Karegar, A. S. Alaqra, and S. Fischer-Hübner. Exploring user-suitable metaphors for differentially private data analyses. In *18th Symposium on Usable Privacy and Security*, pp. 175–193, 2022.

[42] C. T. Kenny, S. Kuriwaki, C. McCartan, E. T. Rosenman, T. Simko, and K. Imai. The use of differential privacy for census data and its impact on redistricting: The case of the 2020 U.S. Census. *Science advances*, 7(41):eabk3283, 2021. doi: 10.1126/sciadv.abk3283

[43] K. Kenthapadi and T. T. Tran. PriPeARL: A framework for privacy-preserving analytics and reporting at Linkedin. In *Proc. 27th ACM International Conference on Information and Knowledge Management*, pp. 2183–2191, 2018. doi: 10.1145/3269206.3272031

[44] D. Kifer, S. Messing, A. Roth, A. Thakurta, and D. Zhang. Guidelines for implementing and auditing differentially private systems. *arXiv preprint arXiv:2002.04049*, 2020. doi: 10.48550/arXiv.2002.04049

[45] S. Lewis. Qualitative inquiry and research design: Choosing among five approaches. *Health promotion practice*, 16(4):473–475, 2015. doi: 10.1177/1524839915580941

[46] N. Li, T. Li, and S. Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *2007 IEEE 23rd international conference on data engineering*, pp. 106–115. IEEE, 2006. doi: 10.1109/ICDE.2007.367856

[47] A. Machanavajjhala, X. He, and M. Hay. Differential privacy in the wild: A tutorial on current practices & open challenges. In *Proc. 2017 ACM International Conference on Management of Data*, pp. 1727–1730, 2017. doi: 10.1145/3035918.3054779

[48] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3–es, 2007. doi: 10.1145/1217299.1217302

[49] S. A. Matei and L. Hunter. Data storytelling is not storytelling with data: A framework for storytelling in science communication and data journalism. *The Information Society*, 37(5):312–322, 2021. doi: 10.1080/01972243.2021.1951415

[50] G. Miklau. Negotiating privacy/utility trade-offs under differential privacy. In *Privacy Engineering Practice and Respect (PEPR '22)*. USENIX Association, June 2022.

[51] minutephysics. Protecting privacy with MATH (collab with the Census). https://www.youtube.com/watch?v=pT19VwBAqKA, Sep 2019.

[52] P. Nanayakkara, J. Bater, X. He, J. Hullman, and J. Rogers. Visualizing privacy-utility trade-offs in differentially private data releases. *Proceedings on Privacy Enhancing Technologies*, 2022(2):601–618, 2022. doi: 10.2478/popets-2022-0058

[53] P. Nanayakkara, M. A. Smart, R. Cummings, G. Kaptchuk, and E. Redmiles. What are the chances? explaining the epsilon parameter in differential privacy. *arXiv preprint arXiv:2303.00738*, 2023. doi: 10.48550/arXiv.2303.00738

[54] J. P. Near and C. Abuah. *Programming Differential Privacy*, vol. 1. Joseph P. Near and Chiké Abuah, 2021.

[55] OpenMined. OpenMined/PIPELINEDP: PipelineDP is a python framework for applying differentially private aggregations to large datasets using batch processing systems such as Apache Spark, Apache Beam, and more. https://github.com/OpenMined/PipelineDP.

[56] OpenMined. PyDP: The python differential privacy library. https://github.com/OpenMined/PyDP.

[57] J. J. Otten, K. Cheng, and A. Drewnowski. Infographics and public policy: using data visualization to convey complex information. *Health Affairs*, 34(11):1901–1907, 2015. doi: 10.1377/hlthaff.2015.0642

[58] L. Padilla, R. Fygenson, S. C. Castro, and E. Bertini. Multiple forecast visualizations: Trade-offs in trust and performance in multiple COVID-19 forecast visualizations. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):12–22, 2022. doi: 10.1109/TVCG.2022.3209457

[59] L. Padilla, M. Kay, and J. Hullman. Uncertainty visualization, 2020. doi: 10.1002/9781118445112.stat08296

[60] L. Panavas, T. Crnovrsanin, J. L. Adams, J. Ullman, A. Sargavad, M. Tory, and C. Dunne. Investigating the visual utility of differentially private scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–16, 2023. Preprint & supplemental material: https://osf.io/b5zvn/. doi: 10.1109/TVCG.2023.3292391

[61] M. S. Pepe. Receiver operating characteristic methodology. *Journal of the American Statistical Association*, 95(449):308–311, 2000. doi: 10.1080/01621459.2000.10473930

[62] S. M. Randall, A. M. Ferrante, J. H. Boyd, J. K. Bauer, and J. B. Semmens. Privacy-preserving record linkage on large real world datasets. *Journal of biomedical informatics*, 50:205–212, 2014. doi: 10.1016/j.jbi.2013.12.003

[63] A. Reamer. Counting for dollars 2020: The role of the decennial Census in the geographic distribution of federal funds. Technical report, The George Washington Institute of Public Policy, 2018. Initial Analysis: 16 Large Census-guided Financial Assistance Programs.

[64] L. Rocher, J. M. Hendrickx, and Y.-A. De Montjoye. Estimating the success of re-identifications in incomplete datasets using generative models. *Nature communications*, 10(1):1–9, 2019. doi: 10.1038/s41467-019-10933-3

[65] D. Sacha, H. Senaratne, B. C. Kwon, G. Ellis, and D. A. Keim. The role of uncertainty, awareness, and trust in visual analytics. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):240–249, 2015. doi: 10.1109/TVCG.2015.2467591

[66] W. Samek, T. Wiegand, and K.-R. Müller. Explainable artificial intelligence: Understanding, visualizing and interpreting deep learning models. *arXiv preprint arXiv:1708.08296*, 2017. doi: 10.48550/arXiv.1708.08296

[67] J. Sarathy, S. Song, A. Haque, T. Schlatter, and S. Vadhan. Don't look at the data! How differential privacy reconfigures the practices of data science. *arXiv preprint arXiv:2302.11775*, 2023. doi: 10.48550/arXiv.2302.11775

[68] M. Shoemate, A. Vyrros, C. McCallum, R. Prasad, P. Durbin, S. Casacuberta Puig, E. Cowan, V. Xu, Z. Ratliff, N. Berrios, A. Whitworth, M. Eliot, C. Lebeda, O. Renard, and C. McKay Bowen. OpenDP library. https://github.com/opendp/opendp.

[69] A. K. Singh and R. Gupta. A privacy-preserving model based on differential approach for sensitive data in cloud environment. *Multimedia Tools and Applications*, 81(23):33127–33150, 2022. doi: 10.1007/s11042-021-11751-w

[70] M. A. Smart, D. Sood, and K. Vaccaro. Understanding risks of privacy theater with differential privacy. *Proc. ACM on Human-Computer Interaction*, 6(CSCW2), nov 2022. doi: 10.1145/3555762

[71] D. Smilkov, S. Carter, D. Sculley, F. B. Viégas, and M. Wattenberg. Direct-manipulation visualization of deep networks. *arXiv preprint arXiv:1708.03788*, 2017. doi: 10.48550/arXiv.1708.03788

[72] L. Sweeney. k-anonymity: A model for protecting privacy. *International journal of uncertainty, fuzziness and knowledge-based systems*, 10(05):557–570, 2002. doi: 10.1142/S0218488502001648

[73] J. Tang, A. Korolova, X. Bai, X. Wang, and X. Wang. Privacy loss in Apple's implementation of differential privacy on macOS 10.12. *arXiv preprint arXiv:1709.02753*, 2017. doi: 10.48550/arXiv.1709.02753

[74] D. P. Team. Learning with privacy at scale. https://machinelearning.apple.com/research/learning-with-privacy-at-scale.

[75] P. Thaker, M. Budiu, P. Gopalan, U. Wieder, and M. Zaharia. Overlook: Differentially private exploratory visualization for big data. *arXiv preprint arXiv:2006.12018*, 2020. doi: 10.48550/arXiv.2006.12018

[76] A. Unwin. Why is data visualization important? What is important in data visualization? *Harvard Data Science Review*, 2(1):1, 2020. doi: 10.1162/99608f92.8ae4d525

[77] Urban Institute. Considerations for consistent terminology when teaching data privacy methods. Urban Institute: Considerations for Consistent Terminology, June 2024. Accessed: 2024-06-21.

[78] U.S. Census Bureau. 2020 decennial Census visualizations and infographics. https://www.census.gov/programs-surveys/decennial-census/decade/2020/2020-visualizations.html, 2020–2023. Accessed on March 1, 2024.

[79] U.S. Census Bureau. 2020 Census blog posts. https://www.census.gov/programs-surveys/decennial-census/decade/2020/news/blog-posts.html, 2021-2023. Accessed on March 1, 2024.

[80] U.S. Census Bureau. 2020 Census: What is the Census? https://www.census.gov/library/video/2019/2020-census-what-is-the-census.html, Aug 2022. Accessed on March 1, 2024.

[81] U.S. Census Bureau. 2020 Census resources. https://www.census.gov/data/academy/topics/2020-census.html, Sep 2023. Accessed on March 1, 2024.

[82] U.S. Census Bureau. Recorded webinars. https://www.census.gov/data/academy/webinars.html, May 2023. Accessed on March 1, 2024. Website: Census.gov.

[83] A. Vellido. The importance of interpretability and visualization in machine learning for applications in medicine and health care. *Neural computing and applications*, 32(24):18069–18083, 2020. doi: 10.1007/s00521-019-04051-w

[84] X. Wang, J.-K. Chou, W. Chen, H. Guan, W. Chen, T. Lao, and K.-L. Ma. A utility-aware visual approach for anonymizing multi-attribute tabular data. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):351–360, 2017. doi: 10.1109/TVCG.2017.2745139

[85] Z. J. Wang, R. Turko, O. Shaikh, H. Park, N. Das, F. Hohman, M. Kahng, and D. H. P. Chau. CNN Explainer: learning convolutional neural networks with interactive visualization. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1396–1406, 2020. doi: 10.1109/TVCG.2020.3030418

[86] Z. A. Wen, J. Jia, H. Yan, Y. Yao, Z. Liu, and C. Dong. The influence of explanation designs on user understanding differential privacy and making data-sharing decision. *Information Sciences*, 642:118799, 2023. doi: 10.1016/j.ins.2023.03.024

[87] Pageviews Analysis: Comparison of pageviews across multiple pages. https://pageviews.toolforge.org/?project=en.wikipedia.org.

[88] WireWheel. Differential privacy lessons for enterprises. https://wirewheel.io/blog/differential-privacy-lessons-for-enterprises/, June 2024. Accessed: 2024-06-21.

[89] A. Wood, M. Altman, A. Bembenek, M. Bun, M. Gaboardi, J. Honaker, K. Nissim, D. R. O'Brien, T. Steinke, and S. Vadhan. Differential privacy: A primer for a non-technical audience. *Vanderbilt Journal of Entertainment and Technology Law*, 21:209, 2018.

[90] A. Xiong, C. Wu, T. Wang, R. W. Proctor, J. Blocki, N. Li, and S. Jha. Using illustrations to communicate differential privacy trust models: An investigation of users' comprehension, perception, and data sharing decision. *arXiv preprint arXiv:2202.10014*, 2022. doi: 10.48550/arXiv.2202.10014

[91] A. X. Zhang, M. Muller, and D. Wang. How do data science workers collaborate? Roles, workflows, and tools. *Proc. ACM on Human-Computer Interaction*, 4(CSCW1):1–23, 2020. doi: 10.1145/3392826

[92] D. Zhang, M. Hay, G. Miklau, and B. O'Connor. Challenges of visualizing differentially private data. *Theory and Practice of Differential Privacy*, 2016:1–3, 2016.

[93] D. Zhang, A. Sarvghad, and G. Miklau. Investigating visual analysis of differentially private data. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1786–1796, 2020. doi: 10.1109/TVCG.2020.3030369

[94] J. Zhou, X. Wang, J. K. Wong, H. Wang, Z. Wang, X. Yang, X. Yan, H. Feng, H. Qu, H. Ying, et al. DPVisCreator: Incorporating pattern constraints to privacy-preserving visualizations via differential privacy. *IEEE Transactions on Visualization and Computer Graphics*, 29(1):809–819, 2022. doi: 10.1109/TVCG.2022.3209391

Liudas Panavas Liudas Panavas is a PhD candidate at Northeastern University. His research encompasses decision making for complex trade-offs with a primary focus on comparing ML models and usable differential privacy.

Amit Sarker Amit Sarker is a PhD student at the University of Massachusetts Amherst. His current research focuses on the design and implementation of visual data exploration platforms tailored for differentially private data and the fairness of Differential Privacy algorithms.

Sara Di Bartolomeo Sara Di Bartolomeo is currently a postdoctoral researcher at the University of Konstanz in Germany. She earned her PhD in Computer Science from Northeastern University in Boston. Her research interests encompass various areas of data visualization, with a primary focus on graph layout algorithms, particularly layered and dynamic graphs.

Ali Sargavad Ali Sarvghad is a Research Assistant Professor in the Manning College of Information and Computer Sciences at the University of Massachusetts Amherst. He has PhD in Computer science from the University of Victoria, Canada. His research interests are inclusive and accessible data visualization, privacy-preserving visual analytics, and Virtual reality.

Cody Dunne Cody Dunne is an Associate Professor at Northeastern University whose research focuses on data visualization. He aims to create readable visualizations of network relationships that help users understand complex data. His research leverages the algorithmic aspects of computer science as well as human-computer interaction (HCI) methodologies.

Narges Mahyar Narges Mahyar is an Associate Professor at the Manning College of Information and Computer Sciences, University of Massachusetts Amherst. She designs, develops, and researches innovative visualization techniques and social computing tools to assist both experts and non-experts in making data-driven decisions. Narges holds a PhD in Computer Science from the University of Victoria, Canada.