Investigating Visual Analysis of Differentially Private Data

Dan Zhang, Ali Sarvghad, and Gerome Miklau

Abstract—Differential Privacy is an emerging privacy model with increasing popularity in many domains. It functions by adding carefully calibrated noise to data that blurs information about individuals while preserving overall statistics about the population. Theoretically, it is possible to produce robust privacy-preserving visualizations by plotting differentially private data. However, noise-induced data perturbations can alter visual patterns and impact the utility of a private visualization. We still know little about the challenges and opportunities for visual data exploration and analysis using private visualizations. As a first step towards filling this gap, we conducted a crowdsourced experiment, measuring participants' performance under three levels of privacy (high, low, non-private) for combinations of eight analysis tasks and four visualization types (bar chart, pie chart, line chart, scatter plot). Our findings show that for participants' accuracy for summary tasks (e.g., find clusters in data) was higher that value tasks (e.g., retrieve a certain value). We also found that under DP, pie chart and line chart offer similar or better accuracy than bar chart. In this work, we contribute the results of our empirical study, investigating the task-based effectiveness of basic private visualizations, a dichotomous model for defining and measuring user success in performing visual analysis tasks under DP, and a set of distribution metrics for tuning the injection to improve the utility of private visualizations.

Index Terms—Differential privacy, information visualization

1 Introduction

In this paper, we present the results of an empirical study that investigates challenges and opportunities for visual data analysis under *Differential Privacy (DP)*, an emerging model for protecting sensitive data from leakage [15]. Due to its strong guarantee of preserving individuals' privacy, Differential Privacy has recently been adopted by many industry leaders such as Google [16] and Apple [18] and government institutions such as the U.S. Census Bureau [1, 19]. This rising popularity makes the investigation of DP in the context of visual data analysis a relevant and timely problem. Additionally, in the face of the current COVID-19 pandemic and possible similar future crises, it is critical to investigate privacy-preserving data sharing and analysis methods to enable us to synergize global efforts to combat these crises.

Differential privacy typically functions by adding *carefully calibrated noise* to data that blurs information about individuals while preserving overall statistics about the population. A higher level of noise translates to a higher level of privacy protection. However, the injection of noise results in the alteration of data values and distribution shape.

Depending on the magnitude of data perturbation, moderate to extreme visual discrepancies can happen between a private visualization and its non-private counterpart. Fig. 1 shows an example of such effects. This figure also shows the difference in the notion of "success" between private and non-private visualizations. A user's success performing a visual analysis task on histogram A depends on correct perception and decoding of visually encoded information. In this case, the measurement of success is binary with values of pass/fail. However, a user's success performing a similar task on histogram B depends not only on perceptual accuracy but also on the magnitude of data perturbation. Even when a user achieves perceptual accuracy and correctly identifies the visual artifact of interest (e.g., the smallest bar in the histogram), any readings from the target will still be different from the non-noisy data. Depending on the distance between noisy and non-noisy values, the user can achieve degrees of success. In this case, accuracy is a fuzzy variable with degrees of a pass or fail. This phenomenon imposes a grand challenge and raises a critical question: is it possible to perform visual data analysis on differentially private visualizations and trust the outcomes? For instance, do patterns in a private line chart indicating improvements in patients' conditions on specific treatment match similar patterns in the non-private line chart?

Existing work in the confluence of privacy and visual data analysis has been mainly focused on the use of syntactic privacy models such as k-anonymity and l-diversity (e.g., [7,8,45]) and we know little about the challenges and opportunities of supporting visual data analysis under differential privacy. To fill this gap, we investigated the following two research questions:

- (RQ1) What is the relationship between the noise-injection level, visualization type, data analysis task, and users' performance (accuracy and time to complete tasks)?
- (RQ2) Is it possible to tune noise injection to improve the utility of private visualizations?

To investigate RQ1, we performed a crowd-sourced user study and examined the effects of three privacy levels (high, low, non-private) for combinations of eight analysis tasks and four visualization types (bar chart, pie chart, line chart, scatter plot). A central challenge in this phase was the assessment of a user's accuracy and task success. Injection of noise and consequent perturbations of data and visual patterns can result in erroneous findings, even if a user's answer to a task is correct based on the private visualization presented to the user. We set forward a dichotomous assessment method that measures accuracy and task success based on the notions of perceptual and perturbation accuracy. A detailed description of this method is presented in Section 4.8. In our study, we only considered univariate visualization. The main rationale behind this decision was to eliminate the possible interactions between two sets of noisy variables and their possible effects on participants' performance. We measured the task success rate and response time for 204 participants. We found that the rate of user success dropped for all tasks as the noise level increased. However, the rate of decline was not consistent across all tasks. In particular, "summary tasks" (e.g., Characterize Distribution) seem to be less sensitive to the injection of noise in comparison to "value tasks" (e.g., Compute Derived Value).

There has been prior research on designing differentially private algorithms [15, 20, 29], and on the comparison of algorithm performance [21] for answering range queries. However, it is unclear how noise injection influences downstream visual tasks. Given a certain privacy protection level, various differentially private noise injection

- *Dan Zhang is with University of Massachusetts, Amherst. E-mail: dzhang@cs.umass.edu.*
- *Ali Sarvghad is with University of Massachusetts, Amherst. E-mail: asarv@cs.umass.edu.*
- *Gerome Miklau is with University of Massachusetts, Amherst. E-mail: miklau@cs.umass.edu.*

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

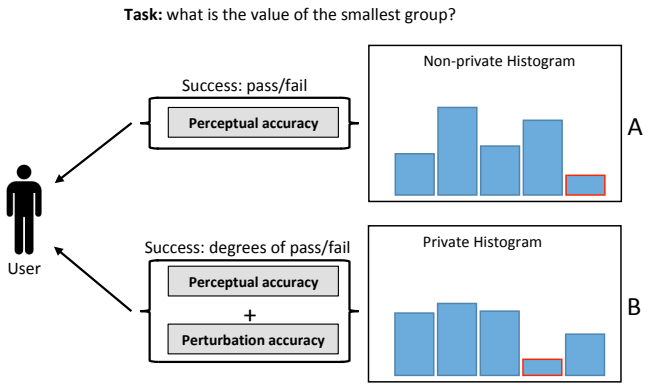


Fig. 1: On the right-hand side, histogram B (bottom) is the private counterpart of the histogram A (top). Data perturbation caused by the injection of noise has resulted in noticeable alterations in visual patterns and data values. Consequently, the utility of histogram B for supporting several visual analysis tasks is compromised. For instance, for the task, "identify the group with smallest value" using histogram B, even if the user correctly identifies the bar with the red border as the smallest, still, his finding is erroneous based on the non-private histogram, A.

algorithms could produce completely different outputs. When visualized, some outputs could contain visual artifacts that make visual tasks substantially harder for users. We investigated the possibility of tuning noise injection through wisely choosing algorithms to improve perceptual accuracy for specific visual tasks. We introduced three basic *distribution metrics* to quantify the shape of noisy algorithm outputs, measuring to what extent they preserve prominent visual features essential for a specific task. *Peakedness Score*, *Anomaly Score*, and *Clusteredness Score* respectively quantify to what extent there exists a single peak, an anomaly data point, and clear cluster boundaries. Then we performed several rounds of simulations using three popular differentially private algorithms Laplace [15], DAWA [29], and MWEM [20] and compared the perceptual distribution metrics of their noisy output. The results of these simulations indicate that the Laplace mechanism works better across different tasks compared with the more complex DAWA and MWEM algorithms.

Our work is the first on differential privacy and visual data analysis that provides an understanding of task-based effectiveness of private visualizations and creates the possibility of managing noise injection based on analysis tasks and visualization. The main contributions of this work are:

- The results of an empirical study, investigating the task-based effectiveness of basic private visualizations.
- A dichotomous model for defining and measuring user success in performing visual analysis tasks under DP.
- A set of distribution metrics for tuning the injection to improve the utility of private visualizations.

In the rest of this paper, we first provide a more detailed explanation of differential privacy followed by prior research related to our work. Next, we present detailed descriptions of work performed to investigate RQ1 and RQ2. Finally, we present a comprehensive discussion of findings and a set of empirical guidelines. We conclude the paper with the limitations of our work and plans for future investigation.

2 Differential Privacy

Differential privacy is classically defined for input data represented as a single table, in which each row contains information about an individual. The formal definition is as follows [15]:

A randomized algorithm $\mathcal{A}$ satisfies ε-differential privacy if for all databases D and D' that differ on one record, and for any subset of outputs $S \subseteq Range(\mathcal{A})$,

$$Pr(\mathcal{A}(D) \in S) \leq e^{\varepsilon} \times Pr(\mathcal{A}(D') \in S)$$

Since differentially private algorithms are randomized, we can think of the algorithm's output, on a given input, as a probability distribution over possible outputs. The definition requires that, if the algorithm runs on any two databases that differ on the details of any individual's record, the output distributions will be "close", where close is formalized by the e^{ε} term. Consequently, seeing the output of the algorithm cannot reveal much about any single contributor's data. The privacy parameter ε, therefore, controls the level of privacy protection: smaller ε means a stricter limit on privacy loss and, in general, this means the algorithm output will be able to communicate less information about the underlying data.

The data owner must choose an ε value to grant to a user of sensitive data. This parameter can be thought of as a privacy loss "budget". Granting a higher ε to a user means they can receive more accurate results from the private algorithm. In the research literature, a common value for ε is 0.1, in which case $e^{\varepsilon} \approx 1.1$ and the likelihood of any output cannot differ by more than about 10% on inputs that differ by one record. In practical deployments, higher ε values have been used and may still provide reasonable privacy protection. In this paper, we explore a range of ε values because it is one factor that impacts the effectiveness of private visualizations.

Laplace Mechanism A standard method for achieving differential privacy is the Laplace Mechanism (although there are many other mechanisms). When the Laplace Mechanism is used to answer a query over a sensitive database (e.g., how many people have the marital status of "divorced") the true query result is computed on the data and then carefully calibrated noise is added to the answer before it is returned to the user. In particular, a sample is drawn from the Laplace distribution with mean zero and a specified scale factor determined by the ε parameter and a property of the query called its sensitivity. The process is ε-differentially private, and the resulting "noisy" query answer may be shared with the user. The formal definition of the Laplace mechanism is as follows [15]:

Let $f(D)$ denote a function on D that outputs a vector in $\mathbb{R}^d$. The Laplace Mechanism is:

$$\mathcal{A}_{LM}(D) = f(D) + (Z_1, ..., Z_d)$$

where Z_i are i.i.d random variables from Laplace$(\Delta f/\varepsilon)$.

Above, Laplace(b) denotes the Laplace probability distribution centered at 0 with scale b, and Δf is called the *sensitivity of* f and is the maximum difference in f between any two databases D and D' that differ only by a single record: $\Delta f = \max_{D,D'} \|f(D) - f(D')\|_1$.

The Laplace Mechanism is commonly used when f is a histogram-generating function, which counts the number of records in a set of disjoint ranges or categories. In this case, adding or removing a single record to the input will only affect the counts in one of the histogram bins by precisely 1. Thus the sensitivity of f is 1, and the Laplace Mechanism adds random noise sampled from Laplace$(1/\varepsilon)$ to each histogram bins and releases the noisy histogram.

Example 1 *Consider the Census data introduced above and the vector-valued histogram function that returns the count of people with each of seven possible marital statuses. If the true counts are* $\langle 100, 51, 9, 45, 134, 123, 12\rangle$ *the noisy counts returned by the Laplace Mechanism using* $\varepsilon = 1.0$ *might be* $\langle 102.1, 49.9, 9.6, 44.4, 134.9, 121.5, 11.0\rangle$*, the result of adding independent samples to each count from the distribution Laplace*(1)*. With 95% confidence, samples from this distribution will fall within* -3 *and* 3*, which is modest noise and, for example, would allow us to reliably identify the fifth status as the most frequent. Using* $\varepsilon = .1$*, the*

confidence interval is $[-30, 30]$*, and such random noise could distort conclusions on the most frequent status.*

Using a smaller privacy parameter produces noisier output and, therefore, less utility. However, the relationship between privacy and utility could depend on the design of the private algorithm, the task being performed, and the underlying input data. One major goal of this work is to better understand this tradeoff for private visualization.

The differential privacy research community has worked actively on designing algorithms that offer the highest accuracy for a given degree of privacy protection. Much of this work is task and domain-specific. For example, there are algorithms for privately releasing sets of aggregate statistics (e.g., [2, 10, 22, 29, 49, 50]), algorithms for private learning classifiers (e.g., [17, 25, 37]), and algorithms for privately analyzing motifs in graphs (e.g., [26]). While differentially private algorithm design has been a subject of intense research effort in recent years, the task of presenting or exploring data through visualization, under the constraint of privacy, has received far too little attention given its obvious importance for users trying to gain insights from data.

Fig. 2 provides an overview of the private visualization pipeline. Sensitive input data is privatized, for a given ε, using some differentially private algorithm, to get $\bar{D}$. This noisy estimate of the true data is used as input to a visualization tool, which produces a visualization as output. Because $\bar{D}$ is differentially private, the resulting visualization enjoys this property as well. Users then use the visualization for analysis and decision making. We observe that two types of error can impact the user's success in performing an analysis task: 1) errors caused by the injection of noise and resulting alterations of data values and visual patterns, and 2) perceptual and cognitive errors. Part of our investigation aims at understanding the two uncertainty sources and how they interact.

3 Related Work

3.1 Privacy-preserving visualizations

The goal of a privacy-preserving visualization is to protect individuals' identities and sensitive information from exposure while still allowing users to make sense and gain knowledge from it. Some prior work in this area has investigated the use of visual uncertainty for preserving privacy. Dasgupta and Kosara [13] introduced a pixel-based clustering technique for parallel coordinates called "Screen-Space Sanitization" that combines pixels to increase visual uncertainty in areas of visualization where privacy could be breached. Using a similar approach, Archambault et al. [4] and Oksanen et al. [34] suggest the aggregation of visual components for building privacy-preserving histograms and heatmaps. Deliberate reduction of visual accuracy increases visualization uncertainty and reduces the possibility of guessing the exact values, but does not satisfy a rigorous definition of privacy that can provably resist attacks.

Data sanitization techniques (e.g., k-anonymity, l-diversity, and t-closeness) have also been investigated for building private visualizations. GraphProtector [45] supports building privacy-preserving graphs of social networks. Users can combine multiple privacy protection schemes as a hybrid approach to fine-tune privacy protection. Chou et al. utilize data sanitization to build private Sankey and Iceplot visualizations for representing temporal event sequence data [8] and constructing private network visualizations [7].

Bhattacharjee et al. [5] provided a thorough and systematic analysis of state-of-the-art approaches, methods, and techniques used in privacy-preserving data visualization, and reflected on a wide range of challenges and research opportunities. Prior work mainly assumes that the data owner designs and deploys privacy mechanisms specific to the domain and visualization types, and the end-user consumes the private visualization product. This approach enables building effective private visualizations for specific tasks, visualizations types, and data domains. However, it may not work in exploratory data analysis where a user's questions and tasks are not known in advance. Currently, we lack a domain-agnostic understating of the relationship between tasks, visualizations, and privacy. The majority of prior work has also been focusing on using syntactic privacy models based on anonymization, which have fallen prey to a range of attacks [14, 32, 33].

To fill these gaps, we investigate the use of differential privacy in exploratory data analysis. Wang et al. [46] developed a visualization technique that helps users to dynamically gauge the loss of utility for a single task by anonymizing multi-attribute tabular data through building matrix-based and tree-based models for utility and privacy. In this work, we aim to better understand the relationships between noise level, task, visualization, and participants' performance. Instead of proposing an approach to modify a specific visualization technique that meets the privacy guarantee, we investigate a more general pipeline of private visualization where we can easily swap in different private algorithms or visualization techniques.

3.2 Privacy-Utility trade-off

There is an inevitable privacy-utility trade-off associated with all privacy-preserving mechanisms. Typically, a stricter guarantee of privacy results in more loss of information and hence lowers the accuracy and reduces analysis utility. For example, in the differential privacy model, using a smaller ε provides a stronger privacy guarantee, but reduces the accuracy of data analysis due to larger perturbation of data.

Deploying privacy mechanisms while balancing the privacy-utility tradeoff is a non-trivial task. It requires proper measurements and an understanding of both privacy and utility. The data mining community has proposed several metrics for evaluating the utility and quality of data after anonymization/privacy-preservation methods are applied. For example, DPBENCH [21] is a principled framework for evaluating differential privacy algorithms for answering 1-D and 2-D range queries. Dasgupta et al. [13] consider discernibility as a utility metric that measures the number of records that cannot be distinguished from one another. In the visualization field, Dasgupta et al. [12] state that utility can be regarded as a function of visual uncertainty. They introduce metrics for quantifying the visual uncertainty in cluster-based visualizations such as scatterplot and parallel coordinates.

An investigation by Zhang et al. [51] shows that the utility of performing visual tasks does not line up well with accuracy measures commonly used for algorithms answering range queries. In fact, the authors of [6, 8, 30] argue that the evaluation of utility should depend on user analytical tasks and how accurately they can be performed. Prior research (e.g., [7, 8, 12] has only investigated the impact of privacy on utility for a specific task, data, and visualization type. Taking a domain- and data-agnostic approach, in this work, we investigate the privacy-utility tension for basic data exploration tasks and visualization types.

Outside the realm of privacy, Saket et al. [39] conducted a study to investigate the effectiveness of five basic visualization types in relation to ten common data exploration tasks. The effectiveness of visualization consists of three main metrics: the success rate of performing the task, the performance time, and user preference. Results show that the effectiveness of visualization varies significantly across tasks. To understand the effectiveness of private visualizations and how conclusions might change under privacy, we use a similar experimental design like the set of tasks and visualization.

3.3 Communicating uncertainty under DP

Differentially private data, like data generated from many other privacy-preserving techniques, is inherently uncertain. DP mechanisms achieve the privacy guarantee by adding noise to computations on the sensitive data, making any single output a random instance of some underlying distribution. Therefore, dealing with uncertainty is an essential aspect of differentially private data visualization.

There has been a rich line of work on visualizing under uncertainty. One common solution is to present the level of uncertainty explicitly. The uncertainty statistic can be represented as an overlaying layer encoded in the formats of error bars or summary plots [38]. Sanyal et al. [40] investigated the effectiveness of different glyphs and markers in conveying uncertainty. However, error bars can be hard to read in multi-dimensional visualizations. An alternative is to color-code uncertainty information. Maceachren et al. [31] used hue and saturation to indicate

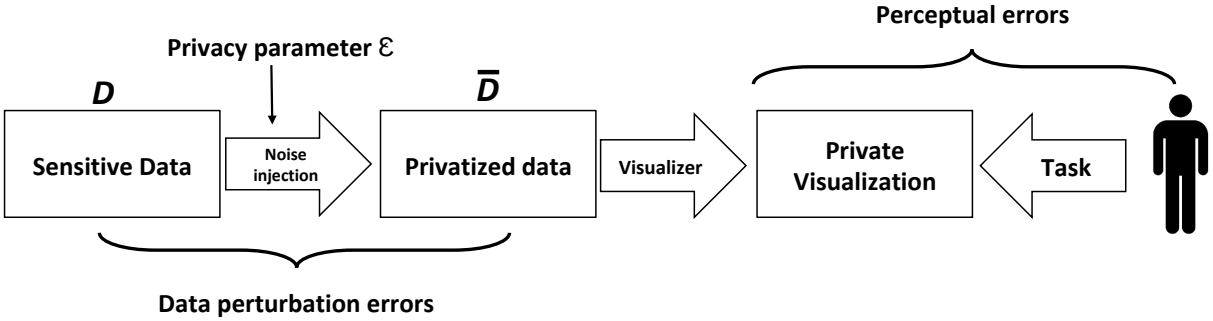


Fig. 2: The pipeline of generating private visualizations. First, sensitive data $\boldsymbol{D}$ is privatized to get $\bar{\boldsymbol{D}}$ by adding carefully calibrated noise using a differentially private algorithm. Next, privatized data is visualized using a visualization engine. Finally, a user performs data analysis tasks on private visualizations. Two major sources of error can impact the user's success in performing the task: 1) data perturbation errors caused by the injection of noise and alteration of data values and visual patterns, and 2) perceptual errors. The second type of error is common between private and non-private visualizations.

uncertainty levels, and Hengl et al. [23] proposed mixing white pixels to represent high uncertainty. A critical first step towards communicating uncertainty is selecting metrics (e.g., standard deviation) to quantify uncertainty. This can be challenging for differentially private output. For simple algorithms like the Laplace mechanism [15], we could easily derive scale and variability of the noise from the privacy parameters. However, the computation of such uncertainty for complex algorithms can be hard. For example, the DAWA [29] algorithm first applies grouping based on the characteristics of the input data distribution to achieve better accuracy. Thus each data group has different uncertainty levels and carefully designed approaches are needed to calculate these local uncertainty statistics privately.

In this work, we do not attempt to display uncertainty information to eliminate possible interference with users' task performance. Investigating methods and effects of communicating uncertainty for private visualizations a goal for our future research.

4 RQ1: Investigating the utility of private visualizations

To investigate RQ1, we conducted a crowd-sourced empirical study on Amazon Mechanical Turk[1]. The rest of this section provides detailed information about the design of the study, the data analysis process, and our findings.

4.1 Dataset

IPUMS-CPS [41] is a collection of datasets that harmonizes microdata from the monthly U.S. labor force survey and the Census Current Population Survey (CPS), covering the period 1962 to the present. We used a data extract which is a subset of the Annual Social and Economic (ASEC) data from 2010. We chose this population survey data for two main reasons. First, it is similar to the data that could soon be protected (by the U.S. Census Bureau) using differential privacy. Second, it contains data attributes with which many study participants will be familiar, hence reducing the chance of failed tasks due to a user's unfamiliarity with the data semantics.

The data extract contains personal survey information on 159,277 individuals, each contributing one row to the dataset. For our experiment, we selected a subset of numerical and categorical attributes: Age, Sex, Race, Marital status, Education status, and Total income. Selected data were organized in a tabular format where each row represented information about an individual.

4.2 Private algorithm

We selected the Laplace Mechanism as the privacy algorithm in our study. Although there are several other algorithms (i.e., [20, 29]) for generating private histograms, Laplace is relatively simple, fast, and competitive for the task of generating a single private histogram [21]. It offers a good compromise in terms of speed and utility and acts as a building block for many more complex algorithms.

In a post-processing step, we modified the output of the Laplace mechanism and replaced all the negative counts with zeros. The rationale behind this decision was to eliminate the change of producing private histograms with negative values for some of the bins which are clearly invalid. As introduced in Section 2, the differential privacy guarantee will not be compromised by this simple post-processing step. We consider the non-negative histogram to be the final privatized data and use it in later visualization steps. Since all differentially private algorithms are randomized, any single output is only a random sample from a distribution. For each experimental setting, we choose five random seeds that define randomized trials. Visualizations and tasks are judged on the average performance over these random trials.

4.3 Privacy Parameter Setting

In this study, we considered three privacy levels of privacy parameter: 1) $\varepsilon = \infty$, 2) $\varepsilon = 0.01$, and 3) $\varepsilon = 0.001$. As mentioned in Section 2, a smaller privacy parameter enforces stronger privacy. An $\varepsilon = \infty$ results in a non-private visualization which offers no privacy guarantee but also implies no noise. This non-private setting provided a baseline against which we assessed the utility of private visualizations. The value $\varepsilon = 0.001$ offers a stronger privacy guarantee and requires more distortion of the data compared to the value $\varepsilon = 0.01$. Both values might be considered low ε in practice. However, we are studying the release of a single visualization while in practice many releases would be made and a global ε bound needs to govern all interactions. In addition, the impact of noise is dependent on the size of the dataset thus the choice of ε is also dependent on the dataset. The selection of these ε thresholds was based on experimentation and preliminary testing with the study dataset. In particular, we paid careful attention that selected ε values offer strong privacy but are not too strict that the noise added leads to completely useless private visualizations.

Our goal in this work is to understand the regime in which noise from the privacy mechanism impacts the utility of private visualizations. As such, these carefully engineered ε values suit our purpose.

4.4 Participants

We recruited a total of 204 subjects based in the U.S., with an approval rate greater or equal to 95%. Each subject was allowed to participate in the study only once.

4.5 Tasks

Following prior research on task-based effectiveness of basic visualizations [39], we used the Amar et al. [3] taxonomy of low-level data analysis tasks for selection of study tasks. Those tasks are real-world

[1] https://www.mturk.com

tasks users came up with while exploring datasets with different visualizations tools and have been used in different studies for visual effectiveness evaluation. We excluded two tasks Find Correlation and Order which involve two variables and could not be performed on univariate visualizations. The 8 selected low-level tasks act as building blocks of more complex tasks. The following is a list of selected tasks. For each task, we also provide a concise explanation of how they were used, along with an example. We use the term "visualization feature" in the following descriptions to refer to element(s) in visualization such as: a bar in a bar chart, a group of points in a scatter plot, or a peak in a line chart:

Retrieve Value We asked participants to retrieve the value of a certain visualization feature. For example, *what is the number of people in the Age Range 15-20?*

Filter Given a range, we asked participants to identify visualization features in that range. For example, *which Age Ranges have Group Size between 25,000 and 35,000?*

Compute Derived Value We asked participants to derive a new value using the visualization. For example, *what is the sum of Group Sizes for Marital Status Single and Divorced?*

Find Maximum We asked participants to find the visualization feature with the largest value. For example, *which Income Range has the largest Group Size?*

Determine Range For a set of visualization features, we asked participants to identify the range of their values. For example, *what is the range of Group Sizes for all Age Ranges? Select the correct pair of minimum and maximum Group Size.*

Characterize Distribution Given a condition, we asked participants to identify the distribution of values based on the condition. For example, *what is the percentage of Income Ranges that have Group Size larger than 25K?*

Find Anomalies We asked participants to find visualization features with abnormal values. We manually modified the visualizations to include easy-to-detect anomalies like zero counts and extremely large counts. For example, *which Age Range has abnormal Group Size?*

Cluster We asked participants to put visualization features of similar values into the same cluster and report the number of clusters. For example, *what is the number of clusters based on the Group Sizes of Income groups?*

4.6 Visualization types

In this study, following prior empirical work [39], we examined four visualization techniques that are commonly incorporated in various visualization dashboards [28]: bar chart, pie chart, line chart, and scatterplot. To generate the visualizations, we took the private output of the Laplace mechanism and used the Matplotlib [24] visualization library to plot the privatized data. To maintain visual consistency, we fixed the chart size to be 500 by 350 pixels and the font size 12. We used Matplotlib's default color palette to generate pie charts and the same blue color for visual elements in other plots. The participants only viewed the final visualization and were not aware of the existence (or lack thereof) of noise in the encoded data.

4.7 Experimental procedure

Instructions and warm-up tasks At the beginning of each study session, a participant was given a brief written description of the purpose of the study, the data that would be collected, and their rights, together with a consent form. Upon consenting to participate, the participant was given information about the workflow of the study and a quick optional tutorial explaining the visualization techniques used in the study. Next, the participant performed a short warm-up session answering one sample question for each task. Warm-up questions were similar to the actual study questions but performed on a manually synthesized dataset. During the warm-up practice, the participant received feedback on whether their answer was correct along with corresponding explanations. After the successful completion of the warm-up session, the participant moved on to the actual study tasks.

Main questions For each combination of task, visualization, and privacy level, we sampled 3 univariate histograms from 3 different attributes. As introduced in Section 2, differentially private algorithms are randomized, making every output a single sample of a probability distribution. Thus, for each experimental configuration, we generated five differentially private outputs using different random seeds. This results in a total of 180 different questions (4 *Visualizations* $\times$ 3 *Histograms* $\times$ 3 *Privacy Levels* $\times$ 5 *Random Seeds*) for each task.

In the main experiment, each participant answered a sequence of 48 multiple-choice questions, organized as 6 randomly sampled questions for each of the 8 tasks. For each question, we showed the user a single private visualization together with a brief description of the task and data. Given the visualization, we asked the participant to answer a single question associated with one of the eight tasks outlined above. The participant would move on to the next question after submitting their answer to the current question. The average completion time for the study was around 15 minutes, and we paid each user two dollars.

Validation Low-quality responses are not rare in online crowdsourcing studies. To have a better sense of the overall quality of a user's responses, we introduced validation questions as recommended in [35]. For each worker, we randomly injected four validation questions, which are exact replicates of the warm-up questions seen before in the instructions. We considered the responses provided by a worker valid only if they answered three out of the four validation questions correctly. Otherwise, all responses from the worker would be discarded since they either failed to understand the task or perhaps made selections randomly.

Data collection Throughout the study, we collect worker responses to the multiple-choice questions and record the time they used to answer each question. At the end of the study, we ask the worker to fill out a simple demographic questionnaire asking about their gender and approximate age. Finally, we ask them to rate the overall difficulty of the question on a ten-point scale and provide short text feedback if they choose.

4.8 Data analysis

We collected responses from a total of 204 MTurk workers and filtered out low-quality data according to the validation criteria described in Section 4.7. Among these workers, 176 finished the study and provided valid responses (105 male, 69 female, 2 other). 73% workers are in the age range from 18 to 40 years old; the rest are above the age of 40. On a scale of 1-10 (where 10 is the most difficult), the workers reported an average difficulty score of 4.25 (out of a maximum of 10). This shows our questions have a reasonable level of difficulty and that most people did not find the tasks confusingly difficult or trivially easy.

We analyzed collected data to quantify participants' performance in terms of time and accuracy under different experimental conditions. While the analysis of performance time was a straightforward process, the assessment of accuracy was challenging. Typically, a user's success in performing a task on a non-private visualization is assessed by determining whether the user 1) correctly identifies the visual artifact(s) of interest as requested by the task, and 2) correctly decodes the visualization to retrieve or derive values and draw conclusions. However, a similar assessment of user success cannot be used under private visualization. Due to the injection of noise and consequent perturbation of data, evaluating only the two aspects cannot guarantee user success in visual tasks with respect to the underlying sensitive data. In this work, we evaluate task success in terms of *perceptual accuracy* and *perturbation accuracy*.

4.8.1 Dichotomous assessment of task success under DP

Perceptual accuracy To capture the information loss in human perception and cognition while performing visual tasks, we measure the

Table 1: Perceptual accuracy and response time comparison for each task showing visualization types that are significantly better or worse than others in terms of both accuracy and response time.

Task	Perceptual accuracy			Response time		
	Better	Worse	ANOVA	Faster	Slower	ANOVA
Retrieve Value	pie chart	-	$F_{3,44} = 2.61, p < 0.05$	bar chart	line chart	$F_{3,44} = 4.71, p < 0.01$
Filter	scatterplot	-	$F_{3,44} = 5.35, p < 0.01$	scatterplot	-	$F_{3,44} = 6.81, p < 0.01$
Compute Derived Value	-	-	$F_{3,44} = 1.19, p > 0.05$	-	pie chart	$F_{3,44} = 4.64, p < 0.01$
Find Maximum	-	-	$F_{3,44} = 0.90, p > 0.05$	-	pie chart	$F_{3,44} = 3.47, p < 0.05$
Determine Range	-	-	$F_{3,44} = 0.28, p > 0.05$	-	pie chart	$F_{3,44} = 4.50, p < 0.01$
Characterize Distribution	-	line chart	$F_{3,44} = 12.14, p < 0.01$	scatterplot, barchart	pie chart, line chart	$F_{3,44} = 15.38, p < 0.01$
Find Anomalies	-	pie chart	$F_{3,44} = 8.86, p < 0.01$	-	pie chart	$F_{3,44} = 35.57, p < 0.01$
Cluster	scatterplot	-	$F_{3,44} = 3.74, p < 0.01$	scatterplot	pie chart	$F_{3,44} = 22.84, p < 0.01$

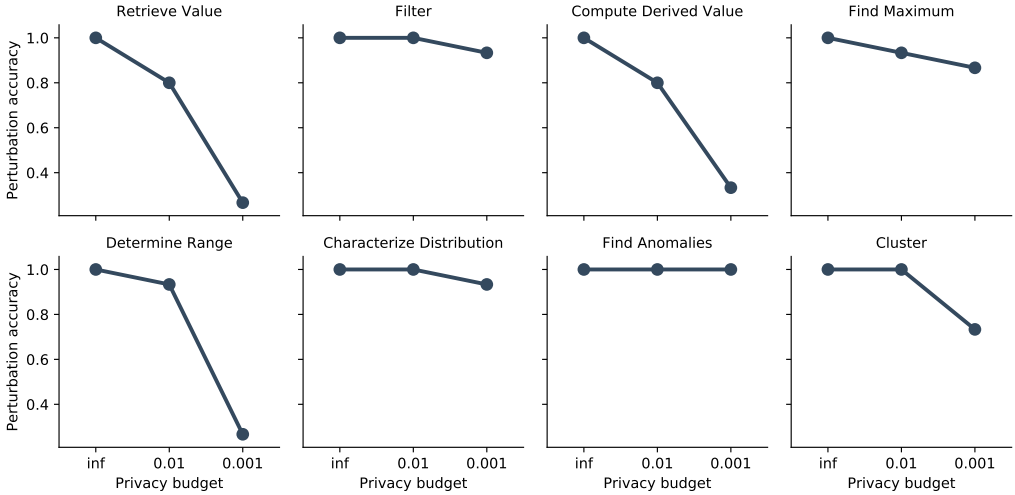


Fig. 3: Perturbation accuracy of different tasks for the non-private case (ε=inf), a low privacy level (privacy parameter $\varepsilon = 0.01$) and a high privacy level (privacy parameter $\varepsilon = 0.001$). Accuracy decreases as we spend less privacy budget. But the accuracy drop is more severe for tasks involving numerical value retrieval or estimation (e.g. Retrieve Value, Compute Derived Value)

perceptual accuracy defined as the rate of "perceptual successes". Here we compare the participants response with the task answer based on the encoded data, whether or not noise has been added to the visualized data. The response is considered a perceptual success if it matches the encoded answer, although it could be different from the answer based on the true data. For instance, if a user task was to "find the group with largest value" using a private bar chart, we considered the user perceptually accurate only if they successfully found the highest bar in the noisy bar chart. Perceptual accuracy describes a users ability to perform visual tasks, consistent with the assessment of visual effectiveness in prior work [27, 39].

Perturbation accuracy Continuing with the example task of finding the group with the maximum count, there are other potential sources of task failure. Even if the user had a perceptual success, there are chances that the highest bar in the private visualization is different from the one in the non-private visualization due to noise injection. In such a case, the user still failed to gain accurate information from the non-private data source. To isolate and measure the information loss from noise injection of the private algorithms for data exploration tasks, we introduce the notion of a perturbation failure. We compare the results of performing a task on the non-private sensitive data and its privatized counterpart after noise injection. If there is a mismatch, then it is considered a perturbation failure. *Perturbation accuracy* for a task is defined to be the rate of perturbation success after noise injection.

Perturbation for privacy may dramatically change the overall patterns of data, causing failures for summary tasks. It may also lead to changes in individual values, making tasks related to value retrieval fail. Due to the injection of noise, exact retrieval of data value is not feasible. Taking a heuristic approach, we considered an *error tolerance range* to decide if the distance between the non-private and private answers was acceptable. More specifically, we considered a range of $\pm$ 1K in which estimation errors were tolerated. For instance, for the non-noisy value of 10K, we accepted any answer in the range of [9K-11K] as acceptable. This error tolerance range was based on careful experimentation with our dataset and consideration of the range and distribution of data values. The histograms used in the study have average values of around 10K to 20K, so the tolerance range is approximately 5% to 10% of the data values.

We consider a visual task successful if both the perturbation and perceptual conditions are satisfied, meaning the information needed to perform the specific task is preserved after the perturbation and the user successfully performs the task on the noisy data. We evaluate all the tasks in relation to perceptual accuracy and perturbation accuracy using this dichotomous model.

In the analysis of perceptual effectiveness, we conducted a one-way repeated measures analysis of variance (ANOVA) for each task to test the differences of effectiveness across different visualizations and tasks. The performance time data was not normally distributed, so it was log-transformed to meet the normality assumption. In the analysis of perturbation error, to avoid the influence of perceptual uncertainty, we

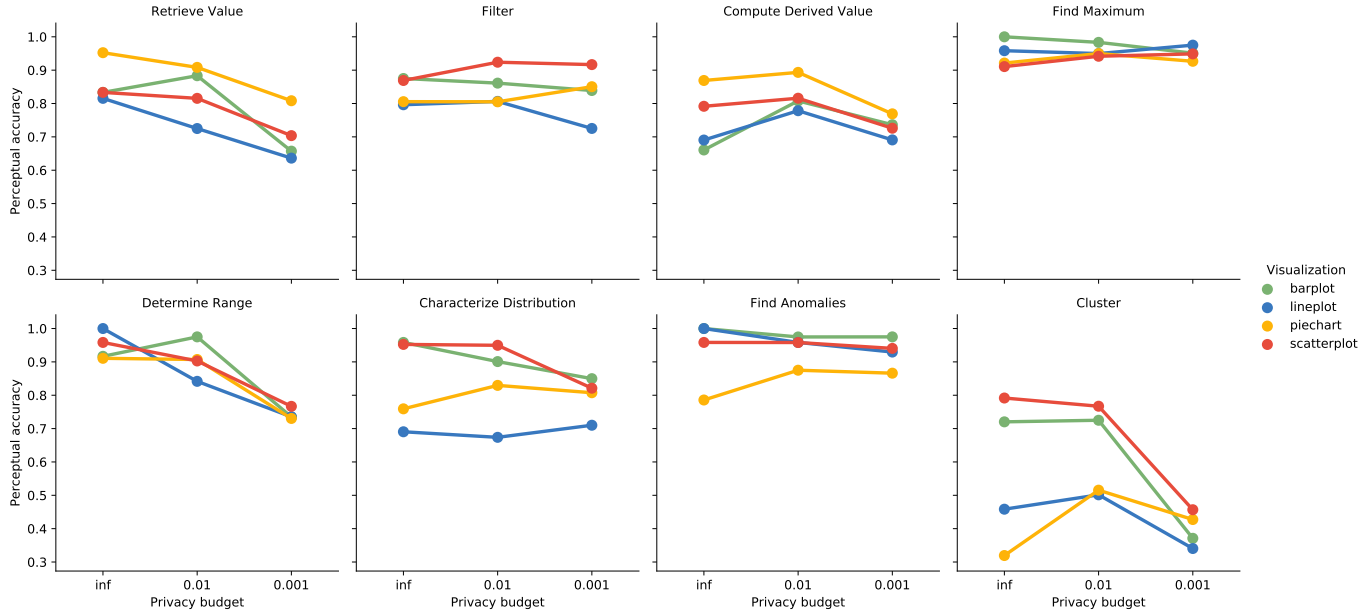


Fig. 4: Perceptual accuracy for different analysis tasks, visualization types using different privacy budget. As the privacy level gets stricter (less privacy budget), the perceptual accuracy changes for some configurations. The noise added for privacy protection needs influences people's ability to perform visual tasks.

conduct the task using only the data values without any visualization. The calculation of perturbation accuracy is done off-line and involves no users.

4.9 Findings

First, we investigate the perceptual effectiveness of the four visualization types and compare the success rate and (log-transformed) response time. Consistent with effectiveness comparisons conducted in an earlier study [39], we found that bar chart and scatterplot are the visualization types that are most accurate and have the best response times, while pie chart has the worst response time but higher accuracy than line chart. Table 1 shows detailed comparisons with the results of ANOVA tests.

Next, Fig. 3 shows the perturbation accuracy trends with varying privacy levels across analysis tasks. For all the tasks, task success rates on perturbed data drop as we move to higher noise levels (i.e., smaller ε). This was an expected outcome since the privacy-utility tension is a known phenomenon under differential privacy. While the rates of accuracy decline seem to be almost consistent within the tasks, there are noticeable differences between them. In particular, we found perceptible differences between *summary tasks*, including Filter, Characterize Distribution, Find Anomalies, Find Maximum, and *value tasks* including Retrieve value, Compute Derived Value, and Determine Range. At a similar level of noise, the rates of accuracy loss for value tasks are higher than those of summary tasks. This finding suggests that different tasks might have varying degrees of *noise tolerance*. The task Cluster showed a mixed pattern where the success rate was highly preserved for a lower level of noise but then sharply plummeted as the noise increased.

Furthermore, we want to see if noise injection influences the users' ability to perform visual tasks. Small multiples in Fig. 4 show the further breakdown of perceptual accuracy over different privacy levels. We omit the result for performance time because there is no significant difference in participant response time across different noise levels. It is interesting to see that the average perceptual accuracy drops slightly as we move to stricter privacy levels (i.e., smaller privacy budgets) for many tasks and visualizations. For task Characterize Distribution, the privacy budget has significant impact on the visual accuracy of scatterplot ($F_{2,14} = 2.99, p < 0.1$). For task Determine Range, the privacy budget used influences visual accuracy of bar chart ($F_{2,14} = 4.94, p < 0.1$). For task Cluster, visual perception of both scatterplot ($F_{2,14} = 5.64, p < 0.1$) and bar chart ($F_{2,14} = 7.13, p < 0.1$) are influenced by the noise level. This shows that visual tasks, especially the more complex summary tasks, can become harder for people when the underlying data is noisy. In other words, the information loss from data perturbation could lead to a higher level of uncertainty in visual perception. However, unlike perturbation accuracy, the influence of noise injection isn't always monotonic. Users' ability to accurately perceive information from visualizations could increase or decrease as more noise is added. To better understand this effect, we conduct a second phase of our study in the next section.

5 RQ2: Tailoring Noise Injection to Analysis Task

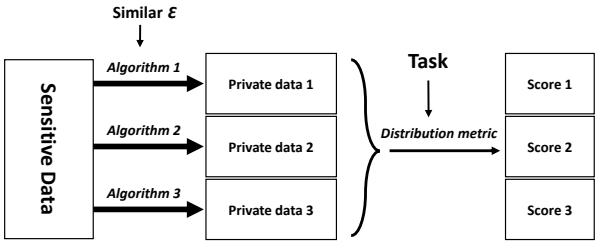
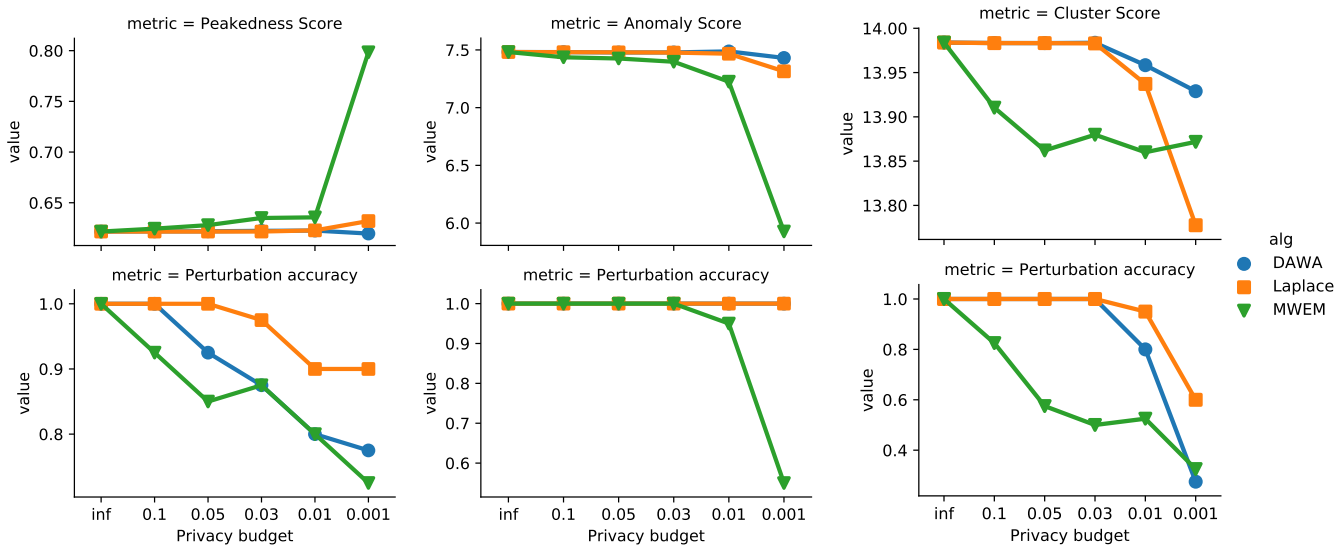


Fig. 5: This figure shows our proposed model for tuning noise injection to tasks using our suggested distribution metrics. First, sensitive data is privatized using alternative DP algorithms (e.g. Algorithms 1, 2 & 3). All the privatized data meet a certain required level of privacy (similar ε). Next, based on the task at hand (e.g., Find the group with maximum value), the related distribution metric (e.g., Peakedness Score) is utilized to calculate a score for each set of privatized data. The privatized dataset with the highest score offers a data distribution shape that will better support the task.

Prior work in data visualization has shown that characteristics of underlying data such as distribution shape impact visualization in perceptible ways [27, 36, 42, 43]. As a simple example, it is easier to find the bin with maximum value in a unimodal histogram with values $\langle 11, 10, 30, 12 \rangle$ than a flat histogram with values $\langle 11, 10, 13, 12 \rangle$. The same level of privacy protection can be achieved by various DP algo-



(a) Task Find Maximum (b) Task Find Anomalies (c) Task Cluster

Fig. 6: For the three summary tasks considered, the upper row shows distribution metrics for different algorithms and the lower row shows task success rate from data perturbation at the corresponding privacy level.

rithms (e.g., [20,29]). However, depending on the specific mechanism chosen, the injection of noise can lead to entirely different data distributions, which are consequently reflected in private visualizations. Drawing on the findings from both fields, we investigated the possibility of tuning the noise injection to result in a private data distribution that will facilitate performing a certain task assuming we have the advanced knowledge of the task at hand and privacy level required.

In this work, we suggest a way of tuning the noise injection for summary tasks. Fig. 5 provides a schematic of our approach. At the core of our model lie "distribution metrics" that quantify the distribution shape of private data. This enables us to compare the privatized output of several algorithms and select one that would best support a task. In the following section, we provide details of our suggested metrics:

5.1 Distribution Metrics

Inspired by prior work on Scagnostics (e.g., [11,47,48]), we suggest three *distribution metrics* which quantify the shape of the data distribution. Each metric is designed and corresponds to a specific summary task. The reason for focusing on summary tasks was that the success of these tasks mainly relies on the user's perceptual accuracy which in turn is related to the shape of the data distribution [43,44]. While for value tasks, there are no consistent relationships between peoples ability to read a single data point and the overall data distribution.

Peakedness Score: this metric is designed for the Find Maximum task. For $\vec{x}$, it is calculated as:

$$P = \frac{m_1}{\sum_{i=1}^{n} x_i} + 1 - \frac{m_2}{m_1}$$

where m_1, m_2 are the largest and second largest value of the distribution $\vec{x}$. For higher Peakedness Scores, the max point stands out more and it is easier to perform the task Find Maximum accurately.

Anomaly Score: this metric is designed for the Find Anomaly task. It is calculated as:

$$A = \frac{\max(|x_i - \text{mean}(\vec{x})|)}{\text{std}(\vec{x})}$$

it finds the furthest point from the sample mean and normalizes the distance by the standard deviation of the distribution. The higher the score is, the more likely the point is an outlier and it is easier for a user to detect it visually.

Cluster Score: this metric is designed for the task Cluster. It is calculated using weights from soft clustering which yield soft assignments of data points to clusters. We generate the soft clustering weights from existing clustering results. In this work, we use the Mean-shift [9] clustering algorithm which iteratively moves data points towards the mode. Unlike many other popular cluster algorithms, it does not require a pre-specified number of clusters. More specifically, given data $\vec{x}$, each data point x_i is assigned to a cluster $Clu(j)$ using the hard mean-shift clustering. The soft assignment weight is the likelihood that a data point i belongs to cluster j:

$$L_{i,j} = \frac{1}{\sum_{k=1}^{c} (\frac{x_i - c_j}{x_i - c_k})^2}$$

where c_j refers to the center of the jth cluster.

The Cluster Score of a distribution is the sum of likelihoods for all data points in the assignment, $L = \sum_{i=1}^{c} L_{i,Clu(i)}$ where $Clu(i)$ is the assigned cluster index for each data point i. Similarly, the higher the clustering score, the more likely there will be coherent clusters with clear boundaries.

5.2 Preliminary Evaluation

To assess the feasibility of using our suggested model to tune the injection of noise for summary tasks, we investigate the influence of privacy mechanisms on the data distribution. More specifically, we chose three representative differentially private algorithms and empirically compare their output at the same privacy levels over multiple runs.

With metrics to measure the visual difficulty of data distributions, we still need to understand how noise injection influences these metrics and later visual perception. Next, we compare three widely used differentially private algorithms in terms of their impact on different distribution metrics. Besides the Laplace algorithm, we also consider the MWEM [20] algorithm which iteratively updates the histogram using noisy data estimations and the DAWA [29] algorithm which forms carefully chosen groups before noise injection.

We run these algorithms over 4 input histograms from the census data and measure the metrics of the noisy histogram. The upper row in Fig. 6 shows average distribution metric scores over all the input

histograms and 10 random trials for each configuration. The lower row shows their perturbation accuracy for the corresponding task.

A good private algorithm should push the distribution towards the easy side of the spectrum as much as possible while at the same time preserving the underlying task answer. In other words, we want algorithms that produce outputs with good distribution metrics which lead to good perceptual accuracy. But we do not want the algorithm to be exaggerating the visual pattern too much and causing a perturbation failure. For example, for task Find Maximum, the Peakedness Score is high using the privacy parameter 0.001. While at the same time, the perturbation accuracy is low, showing that the noise shifted the distribution too much and created an easy-to-perceive but incorrect peak. Thus, in this case, MWEM is not a good algorithm choice.

5.3 Findings

For the three summary tasks considered, the MWEM algorithm tends to provide the lowest perturbation accuracy and only provides better perceptual metrics for one task: Find Maximum. For the task Find Anomalies, DAWA and Laplace have comparable perturbation accuracy, and DAWA tends to produce higher perceptual metrics with a lower privacy budget. So it is best to use DAWA as the DP mechanism for this task to gain good end-to-end visual utility, outperforming Laplace by a small margin. For task Find Maximum and Cluster, Laplace has noticeably higher perturbation accuracy than the other two algorithms, making it a wise choice. In real-world data exploration, it is not always clear what the downstream tasks are. Our findings show that despite its simplicity, the Laplace mechanism is a safe general choice for DP algorithms.

6 Discussion

In this section, we reflect on our findings and discuss strategies to better support visual data analysis under differential privacy.

The findings of our study show that the rate of perturbation accuracy loss (for the same level of noise) differs between summary and value tasks. In particular, summary tasks seem to be more tolerant of the injection of noise while value tasks are noticeably more sensitive. This finding has an important implication: it enables users to more efficiently manage their assigned privacy budget while analyzing data. For instance, based on the knowledge that the summary task Find Anomalies is highly resistant to high levels of noise, they can spend a smaller fraction of their budget on this task. Considering that the amount of privacy budget assigned to a user has a limit, efficient budget management is very important. Similarly, understanding the higher sensitivity of value tasks to the injection of noise and their higher rate perturbation accuracy loss can benefit users. In this case, such knowledge can inhibit performing queries that would result in futile outcomes. For example, consider the value task Determine Range, with a severe loss of accuracy under high-levels of noise, the user can decide to increase the privacy budget spent on the task to get more reliable answers or avoid the task altogether. In the setting of our study, performing a task like Filter, Find Maximum or Find Anomalies with privacy parameter 0.001 only introduces less than 10% additional error compared to using privacy parameter 0.01 but saves 90% of the privacy budget. However, for a value task such as Retrieve value, the accuracy drop can be as significant as 60%, and it is worth spending more privacy budget to get acceptable accuracy.

Although various differentially private algorithms have been designed to achieve higher query accuracy, it remains unclear which algorithms or noise injection mechanism will better facilitate downstream visual analysis. Newer and complex algorithms like [20, 29] apply specialized techniques to increase the accuracy of target workload queries. However, our initial investigation shows that these techniques can produce visual artifacts that make the data "harder" when used in visual analysis tasks. For example, the MWEM algorithms can produce plateau-shaped distribution since it updates queries in groups identified by workload queries. Our findings show that, without a specific task, the simplest Laplace mechanism is the safe choice for private visualization.

Our work is the first in the confluence of differential privacy and visual data analysis that enables managing the privacy budget based on analysis tasks and visualization. This might be even more important for exploratory visual data analysis (EVDA). EVDA revolves around the continued formulation and evaluation of questions and hypotheses by the user. Many times, an analysis avenue does not result in any interesting insights and knowledge, and users move on to investigating other aspects of data. Under such conditions, the effective management of the privacy budget is even more critical. In practical data exploration, we do not always know the exact sequence of operations to perform ahead of time. So it can be hard to generate an optimal global allocation of privacy budget. However, with the knowledge about noise tolerance of tasks, we could go with a greedy approach trying to avoid spending too much of the privacy budget at each step in the iterative process.

7 Limitations and Future Work

In designing this study, we paid careful attention to meet a high-level of internal and external validity. However, similar to any other empirical study, our findings, their implications, and derived guidelines and suggestions are subject to some limitations.

To investigate RQ1, we only considered the Laplace algorithm due to its relative ease of implementation, acceptable performance, and insensitivity to the dataset size. In the second part of this work (RQ2), we included two other algorithms MWEM and DAWA. However, there are several existing DP algorithms that could be utilized for generating private visualizations. One of our future directions is to extend the evaluation of RQ1 by including a more extensive range of DP algorithms.

In this study, we only investigate univariate visualizations. While univariate visualizations such as histograms are widely used, it remains important to extend our investigation to multivariate private visualizations. Multivariate private visualizations introduce several challenges such as: a wider variety of potential visualization types, more complex privacy algorithms to cope with sparser data, and, typically, greater noise relative to the magnitude of plotted statistics.

Since private data has inherent uncertainty, as future work, we also plan to investigate visualization techniques that use additional visual channels like error bars, summary plots [38], hue, and saturation of colormap [23, 31] to encode the uncertainty.

We used simulations as the first step towards the assessment of our proposed distribution metrics. These preliminary evaluations provide initial evidence of the utility of these metrics. However, further user studies are required for assessing their practical usefulness. As part of our future work, we intend to evaluate these metrics while participants are performing open-ended visual data exploration. We aim to understand if the findings of this study will enable them to manage their privacy budget more efficiently, and also select DP settings that will improve their performance.

8 Conclusion

In this work, we present the results of a crowd-sourced empirical study that investigates visual data analysis under Differential Privacy. We examine the effects of privacy levels (high, low, non-private), eight different analysis tasks, and four visualization types on participants' performance. Our findings show that summary tasks are more tolerant of the higher levels of noise than the value tasks. One of the main implications of understanding the relationship between task, level of noise, and accuracy is that it enables analysts to more efficiently manage and allocate their privacy budget. In this work, we also introduce a set of metrics that can be used to analyze the shape of data distribution before the injection of noise. The outcome of this analysis can then be used for tuning the DP model, such as the selection of a DP algorithm that provides better utility by reducing perceptual errors.

Acknowledgments

We appreciate the comments of each of the anonymous reviewers. This material is based upon work supported by the National Science Foundation under Grant Nos. 1409143, 1741254.

REFERENCES

[1] https://onthemap.ces.census.gov/, 2010.

[2] G. Acs, C. Castelluccia, and R. Chen. Differentially private histogram publishing through lossy compression. In *2012 IEEE 12th International Conference on Data Mining*, pp. 1–10. IEEE, 2012.

[3] R. Amar, J. Eagan, and J. Stasko. Low-level components of analytic activity in information visualization. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pp. 111–117. IEEE, 2005.

[4] D. Archambault and N. Hurley. Visualization of trends in subscriber attributes of communities on mobile telecommunications networks. *Social Network Analysis and Mining*, 4(1):205, 2014.

[5] K. Bhattacharjee, M. Chen, and A. Dasgupta. Privacy-preserving data visualization: Reflections on the state of the art and research opportunities. *STAR*, 39(3), 2020.

[6] J. Brickell and V. Shmatikov. The cost of privacy: destruction of data-mining utility in anonymized data publishing. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 70–78. ACM, 2008.

[7] J.-K. Chou, C. Bryan, and K.-L. Ma. Privacy preserving visualization for social network data with ontology information. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 11–20. IEEE, 2017.

[8] J.-K. Chou, Y. Wang, and K.-L. Ma. Privacy preserving visualization: A study on event sequence data. In *Computer Graphics Forum*. Wiley Online Library, 2018.

[9] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on pattern analysis and machine intelligence*, 24(5):603–619, 2002.

[10] G. Cormode, C. Procopiuc, D. Srivastava, E. Shen, and T. Yu. Differentially private spatial decompositions. In *2012 IEEE 28th International Conference on Data Engineering*, pp. 20–31. IEEE, 2012.

[11] T. N. Dang, A. Anand, and L. Wilkinson. Timeseer: Scagnostics for high-dimensional time series. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):470–483, 2012.

[12] A. Dasgupta, M. Chen, and R. Kosara. Measuring privacy and utility in privacy-preserving visualization. In *Computer Graphics Forum*, vol. 32, pp. 35–47. Wiley Online Library, 2013.

[13] A. Dasgupta and R. Kosara. Adaptive privacy-preserving visualization using parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2241–2248, 2011.

[14] I. Dinur and K. Nissim. Revealing information while preserving privacy. In *PODS*, pp. 202–210, 2003.

[15] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography conference*, pp. 265–284. Springer, 2006.

[16] Ú. Erlingsson, V. Pihur, and A. Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC conference on computer and communications security*, pp. 1054–1067. ACM, 2014.

[17] A. Friedman and A. Schuster. Data mining with differential privacy. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 493–502. ACM, 2010.

[18] A. Greenberg. Apple's 'differential privacy' is about collecting your data—but not *your* data. *Wired*, Jun 13 2016.

[19] S. Haney, A. Machanavajjhala, J. M. Abowd, M. Graham, M. Kutzbach, and L. Vilhuber. Utility cost of formal privacy for releasing national employer-employee statistics. In *Proceedings of the 2017 ACM International Conference on Management of Data*, pp. 1339–1354, 2017.

[20] M. Hardt, K. Ligett, and F. McSherry. A simple and practical algorithm for differentially private data release. In *Advances in Neural Information Processing Systems*, pp. 2339–2347, 2012.

[21] M. Hay, A. Machanavajjhala, G. Miklau, Y. Chen, and D. Zhang. Principled evaluation of differentially private algorithms using dpbench. In *Proceedings of the 2016 International Conference on Management of Data*, pp. 139–154. ACM, 2016.

[22] M. Hay, V. Rastogi, G. Miklau, and D. Suciu. Boosting the accuracy of differentially private histograms through consistency. *PVLDB*, 2010.

[23] T. Hengl and N. Toomanian. Maps are not what they seem: representing uncertainty in soil-property maps. In *Proc. Accuracy*, pp. 805–813, 2006.

[24] J. D. Hunter. Matplotlib: A 2d graphics environment. *Computing in Science & Engineering*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55

[25] G. Jagannathan, K. Pillaipakkamnatt, and R. N. Wright. A practical differentially private random decision tree classifier. In *2009 IEEE International Conference on Data Mining Workshops*, pp. 114–121. IEEE, 2009.

[26] V. Karwa, S. Raskhodnikova, A. Smith, and G. Yaroslavtsev. Private analysis of graph structure. *Proceedings of the VLDB Endowment*, 4(11):1146–1157, 2011.

[27] Y. Kim and J. Heer. Assessing effects of task and data distribution on the effectiveness of visual encodings. In *Computer Graphics Forum*, vol. 37, pp. 157–167. Wiley Online Library, 2018.

[28] S. Lee, S.-H. Kim, and B. C. Kwon. Vlat: Development of a visualization literacy assessment test. *IEEE transactions on visualization and computer graphics*, 23(1):551–560, 2016.

[29] C. Li, M. Hay, and G. Miklau. A data- and workload-aware algorithm for range queries under differential privacy. *PVLDB*, 2014.

[30] T. Li and N. Li. On the tradeoff between privacy and utility in data publishing. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 517–526. ACM, 2009.

[31] A. M. MacEachren, A. Robinson, S. Hopper, S. Gardner, R. Murray, M. Gahegan, and E. Hetzler. Visualizing geospatial information uncertainty: What we know and what we need to know. *Cartography and Geographic Information Science*, 32(3):139–160, 2005.

[32] A. Narayanan and V. Shmatikov. Robust de-anonymization of large sparse datasets. In *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pp. 111–125, 2008. doi: 10.1109/SP.2008.33

[33] A. Narayanan and V. Shmatikov. De-anonymizing social networks. In *Security and Privacy*, 2009.

[34] J. Oksanen, C. Bergman, J. Sainio, and J. Westerholm. Methods for deriving and calibrating privacy-preserving heat maps from mobile sports tracking application data. *Journal of Transport Geography*, 48:135–144, 2015.

[35] J. S. Olson and W. A. Kellogg. *Ways of Knowing in HCI*, vol. 2. Springer, 2014.

[36] A. V. Pandey, J. Krause, C. Felix, J. Boy, and E. Bertini. Towards understanding human similarity perception in the analysis of large sets of scatter plots. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pp. 3659–3669, 2016.

[37] N. Papernot, M. Abadi, U. Erlingsson, I. Goodfellow, and K. Talwar. Semi-supervised knowledge transfer for deep learning from private training data. *arXiv preprint arXiv:1610.05755*, 2016.

[38] K. Potter, J. Kniss, R. Riesenfeld, and C. R. Johnson. Visualizing summary statistics and uncertainty. In *Computer Graphics Forum*, vol. 29, pp. 823–832. Wiley Online Library, 2010.

[39] B. Saket, A. Endert, and C. Demiralp. Task-based effectiveness of basic visualizations. *IEEE transactions on visualization and computer graphics*, 2018.

[40] J. Sanyal, S. Zhang, G. Bhattacharya, P. Amburn, and R. Moorhead. A user study to compare four uncertainty visualization methods for 1d and 2d datasets. *IEEE transactions on visualization and computer graphics*, 15(6):1209–1218, 2009.

[41] R. R. S. R. Sarah Flood, Miriam King and J. R. Warren. Integrated public use microdata series, current population survey: Version 6.0 [dataset]. https://doi.org/10.18128/D030.V6.0, 2018.

[42] A. Sarikaya and M. Gleicher. Scatterplots: Tasks, data, and designs. *IEEE transactions on visualization and computer graphics*, 24(1):402–412, 2017.

[43] R. Veras and C. Collins. Discriminability tests for visualization effectiveness and scalability. *IEEE transactions on visualization and computer graphics*, 26(1):749–758, 2019.

[44] L. Visengeriyeva. Advancing data curation with metadata and statistical relational learning. 2020.

[45] X. Wang, W. Chen, J.-K. Chou, C. Bryan, H. Guan, W. Chen, R. Pan, and K.-L. Ma. Graphprotector: A visual interface for employing and assessing multiple privacy preserving graph algorithms. *IEEE transactions on visualization and computer graphics*, 25(1):193–203, 2019.

[46] 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, 2018.

[47] L. Wilkinson, A. Anand, and R. Grossman. Graph-theoretic scagnostics. In *IEEE Symposium on Information Visualization, 2005. INFOVIS 2005.*, pp. 157–164. IEEE, 2005.

[48] L. Wilkinson and G. Wills. Scagnostics distributions. *Journal of Computational and Graphical Statistics*, 17(2):473–491, 2008.

[49] X. Xiao, G. Wang, and J. Gehrke. Differential privacy via wavelet transforms. In *ICDE*, pp. 225–236, 2010.

[50] Y. Xiao, L. Xiong, L. Fan, S. Goryczka, and H. Li. DPCube: Differentially private histogram release through multidimensional partitioning. *Transactions of Data Privacy*, 7(3), 2014.

[51] D. Zhang, M. Hay, G. Miklau, and B. O'Connor. Challenges of visualizing differentially private data. 2016.