Investigating the Visual Utility of Differentially Private Scatterplots

Liudas Panavas, Tarik Crnovrsanin, Jane Lydia Adams, Jonathan Ullman, Ali Sargavad, Melanie Tory, and Cody Dunne

Abstract—Increasingly, visualization practitioners are working with, using, and studying private and sensitive data. There can be many stakeholders interested in the resulting analyses—but widespread sharing of the data can cause harm to individuals, companies, and organizations. Practitioners are increasingly turning to differential privacy to enable public sharing of data with a guaranteed amount of privacy. Differential privacy algorithms do this by aggregating data statistics with noise, and this now-private data can be released visually with differentially private scatterplots. While the private visual output is affected by the algorithm choice, privacy level, bin number, data distribution, and user task, there is little guidance on how to choose and balance the effect of these parameters. To address this gap, we had experts examine 1,200 **differentially private scatterplots** created with a variety of parameter choices and tested their ability to see aggregate patterns in the private output (i.e. the **visual utility** of the chart). We synthesized these results to provide easy-to-use guidance for visualization practitioners releasing private data through scatterplots. Our findings also provide a **ground truth for visual utility**, which we use to benchmark automated utility metrics from a variety of fields. We demonstrate how multi-scale structural similarity (MS-SSIM), the metric most strongly correlated with our study's utility results, can be used to **optimize parameter selection**. A free copy of this paper along with all supplemental materials is available at https://osf.io/wej4s/.

Index Terms—Scatterplots, differential privacy, data study, visual utility.

1 Introduction

Researchers need to analyze sensitive and personal data to answer important questions in research areas such as racial discrimination, cancer screening, or health outcomes. This data— from medical records, payment information, search queries, fitness tracking—in its unaltered state cannot be shared with the public [73], and its availability is either confined to a small group of researchers or requires lengthy information release processes [24].

Data can be more safely shared through the use of differential privacy algorithms [14], which protect individual privacy while enabling the public's ability to see aggregate patterns. As illustrated in Fig. 1, these algorithms release the data as aggregate statistics with a specified amount of noise added, guaranteeing, to a set level of privacy ϵ, that an attacker cannot know whether or not an individual's data was used in the released output. However, no such guarantees are provided for the *utility* of the output.

The privacy-protecting addition of noise can obfuscate or even alter the original patterns found in the data [46], [69]—as shown in Fig. 2 and throughout this paper. This drop in utility affects any private release of the data, whether as a statistic, table, or visualization. Given that data can contain important patterns that are much more easily understood visually [2], we are interested in investigating the possible drop in utility of visualizations showing differentially private data versus the original data.

We use the term *visual utility* to characterize how well a differentially private visualization retains its ability to let the user generate the same insights as the original visualization. There is a lack of consensus on how to best evaluate the utility of a privacy-preserving visualization [4], [57], [67], and visual utility has rarely been discussed in the differential privacy community because data is generally released as numerical values from queries on a database [13], [67].

There has been some preliminary work investigating how adding noise to protect privacy affects the resulting visualizations [46], [69]—but there is little concrete direction provided to data curators. The impact on visual utility depends on the choice of algorithm for adding the noise, privacy level ϵ, user task, data distribution, and bin size. The noise addition can prevent the data user (e.g. public health researcher) from completing their tasks and generating insights from the private scatterplots. Data curators need to balance the trade-offs between privacy and visual utility when releasing data, but they need more information about how their choices can affect visual utility [52].

The work presented here, based on an expert evaluation of 1200 differentially private scatterplots, provides evidence-based guidelines for data curators to create and evaluate their private visualizations. We focus on scatterplots because (1) they are frequently used to highlight interesting patterns and distributions between two variables [22], and (2) it is challenging for data curators using scatterplots to protect privacy because scattered points often directly correspond to individuals. Scatterplots have no aggregation like we see in many other chart idioms, so we must rely on data obfuscation approaches, such as differential privacy algorithms, to provide privacy guarantees.

Our narrow focus—only dealing with one chart idiom—

- *Liudas Panavas, Tarik Crnovrsanin, Jane Lydia Adams, Jonathan Ullman, Melanie Tory, and Cody Dunne are with Northeastern University. E-mails: [panavas.l | t.crnovrsanin | adams.jan | j.ullman | m.tory | c.dunne]@northeastern.edu*
- *Ali Sargavad is with the University of Massachusetts, Amherst. Email: asarv@cs.umass.edu.*

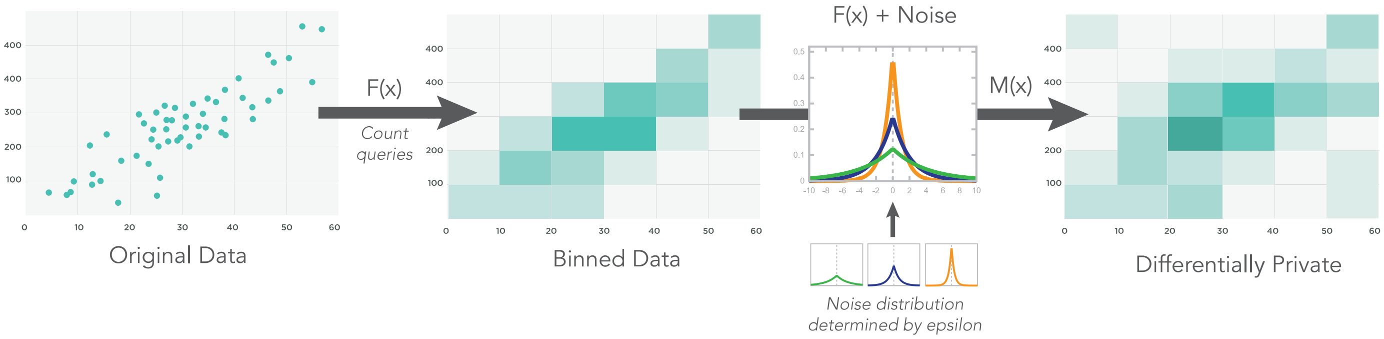


Fig. 1: Illustration of how a differentially private algorithm generates private data from the original data. The data is binned through count queries, denoted $F(x)$. Noise is added from Laplace distributions dictated by ϵ. The output is a differentially private scatterplot composed of F(x) + noise = M(x).

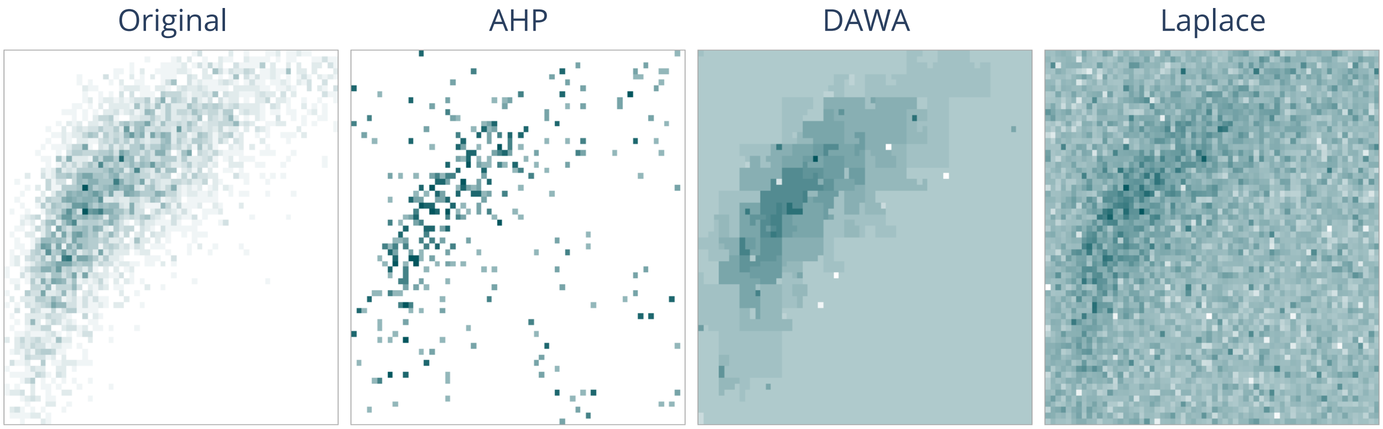


Fig. 2: A binned scatterplot (a.k.a. heatmap, leftmost plot) uses a color scale to show how many points occur in each cell. A malicious viewer could unambiguously locate a data point by knowing only the value for one of its plotted attributes—thus, determining the likely value of the point's *unknown* attribute. We can reduce the amount of private information exposed by using **differential privacy**. Here, we show the results of three of the five algorithms we tested (AHP, DAWA, Laplace) for creating differentially private scatterplots. Each scatterplot has the same theoretical *privacy guarantee*, $\epsilon = 0.2$, but each algorithm adds noise differently. **The choice of algorithm clearly affects the visual utility** of the resulting scatterplot.

lets us investigate the factors affecting visual utility thoroughly. In particular, our data study evaluated each combination of:

- **5 algorithms for adding noise:** [DAWA, AHP, AGrid, Laplace, Geometric Truncated]
- **4 privacy levels ϵ:** $[0.50, 0.10, 0.05, 0.01]$
- **2 bin sizes:** $[32 \times 32, 64 \times 64]$
- **20 data distributions:** $[0..19]$, each a canonical scatterplot from each of Pandey et al.'s categories [47]
- **3 user tasks:** [Clusters, Distributions, Correlation], filtering the tasks as appropriate for each data distribution

We created 1200 differentially private scatterplots for these combinations. For each, we had three three visualization practitioner-researchers assess its visual utility using an assessment rubric and their perceptual judgment. We analyzed the data from this expert assessment to determine how each parameter affects the visual utility of private plots.

We used our expert-generated ground truth to assess several statistical metrics and determine which best *approximates visual utility*. Currently, there are no empirically-driven metrics for quantifying the utility of differentially private visualizations [4], [57], [67]. Traditional metrics used to evaluate differentially private algorithms, such as Average Per Query Error (APQE) [29], can produce scatterplots that have the same score but produce very different visual representations [67]. For example, Fig. 4 shows two scatterplots with similar APQE (statistical evaluation metric – Fig. 3) that produce different visual results. Therefore, to assess the quality of these metrics, we compare the following evaluation metrics to our expert judgements:

- **5 metrics for visual utility:** [MS-SSIM, Average Per Query Error (APQE), Earth Mover's Distance, KSTest, Scagnostics]

Our findings fill the gap of having *accurate metrics for both privacy and utility* when generating privacy-preserving visualization. Moreover, our groundwork and methodology for comparing automated utility metrics could be applied to other comparisons of statistical metrics related to private or non-private visualizations.

In particular, the key contributions of this work are:

1) A comparison of how the combination of differential privacy algorithm, privacy level, data distribution, and bin size affects the visual utility of the resulting scatterplots for user tasks.
2) Guidance for data curators on how to adjust parameters

to create more perceptually-consistent privacy-preserving scatterplots.

3) An assessment of how well common statistical utility metrics correspond to expert ratings of perceived utility.

We hope to see future researchers adopt our methodology to understand other differentially private chart idioms, as well as to evaluate visual distribution similarities.

2 Supplemental Material

A copy of this paper, along with all supplemental materials, is available at https://osf.io/wej4s/. This includes all materials required to reproduce and replicate this study—dataset-generating code, study plot generation code, study website, collected data, data analysis code, and code to generate the figures. To avoid issues stemming from postdiction such as hindsight bias, overconfidence in post hoc explanations, and underestimating uncertainty, we preregistered our study on the Open Science Framework (OSF) before running the experiment. Our preregistered study design and analysis code is available at https://osf.io/25xhn. Please view the collected data at the accompanying website (linked).

3 Background

Our study examines how a variety of parameters affect the visual output of differentially private scatterplots and extends those results to find ways to automate utility evaluation. We first introduce the topic of differential privacy then our literature review follows the same format: an investigation of privacy-preserving visualization followed by visual utility. Research papers on privacy-preserving visualization often either help researchers select a level of privacy for specific datasets or provides general information on how privacy affects a broad range of data visualizations. In contrast, by focusing specifically on scatterplots, our research gives clear guidelines on how privacy affects scatterplots but also investigates the many other variables a data curator must decide on. We also found that the concept of visual utility is defined in many different ways, with no clear understanding of which statistical metrics best represent visual utility.

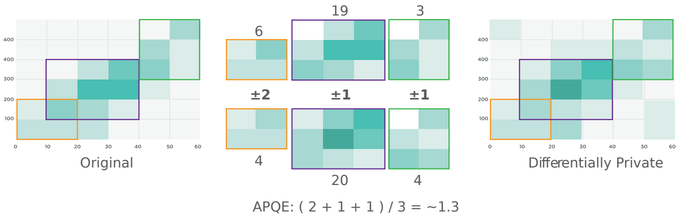


Fig. 3: Illustration of how the Average Per Query Error (APQE) automated metric is calculated across samples. The metric takes random queries and finds the difference in counts between original and private outputs dividing by total queries (middle figure). These type of automated metrics are crucial for quickly evaluating the utility of many attributes or testing new algorithms.

3.1 Differential Privacy

Differential privacy is a statistical property, or guarantee, of an algorithm. It guarantees that the query outputs on the original and private dataset are indistinguishable whether or not a specific person's data is contained in the dataset. It has been widely adopted by governments [1] and influential organizations [19] due to its mathematically-proven guarantees of privacy against "all reasonable" attacks [13]. For a more thorough non-technical explanation of how differential privacy works and protects individuals from attacks, refer to the work done by Wood et al. [64].

The basic building blocks of a differentially private algorithm, M(x), are illustrated in Fig. 1. They are: the non-private algorithm $F(x)$ and the addition of $noise$. In the case of a scatterplot, $F(x)$ is a count query over a range dictated by the bin size, and $noise$ addition is a random value chosen from a distribution specified by ϵ. ϵ has an inverse relationship to the level of injected noise, meaning smaller ϵ translates to more noise added. The more noise added, the less sure the attacker will be that their previous knowledge is being validated by the data they see. Therefore, ϵ can be thought of as a tuning knob between privacy and utility [64]. A key point to make here is that, while many parameters affect visual utility (algorithm, ϵ, bin size, data distribution, user task), ϵ is the only one that specifies the level of privacy (protection of an individual's data). By holding ϵ constant and adjusting the other parameters, we can investigate their influence on visual utility at the same theoretical privacy level. Fig. 2 demonstrates this principle by showing three private scatterplots that represent equal privacy guarantees but produce different visual outputs.

3.2 Privacy-Preserving Visualization

In a review summing up the research at the intersection of data privacy and visualization, Bhattacharjee et al. [4] discuss how visualizations have been used to empower the record owner (understanding of internet privacy policies) [9], data curator (helping them select the correct level of privacy) [8], and data recipient (maximizing the utility of

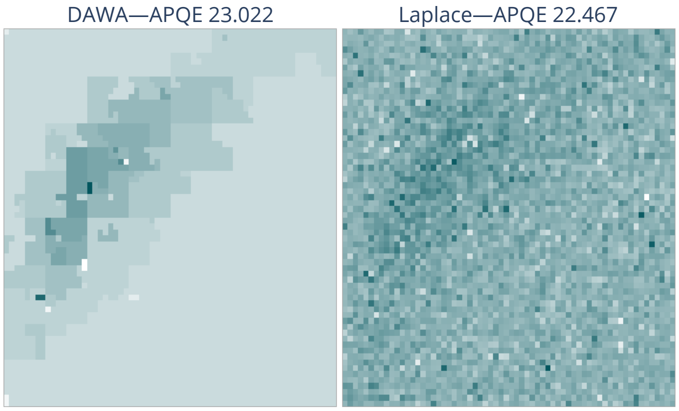


Fig. 4: DAWA & AHP have comparable Average Per Query Error (APQE) [29]—a common metric for comparing differential privacy algorithms—but produce two disparate visual results. Therefore, we cannot necessarily count on statistical metrics for our judgments of visual utility.

the disclosed data) [38]. Our work focuses on providing the data curator with a better understanding of how their data curation choices will influence the utility of the plots, which will not only make their job easier, but also benefit the data users as the key patterns and insights are clearer.

Since the data curator has to make many difficult decisions—and usually has little expertise in differential privacy—researchers have created studies and designed systems to assist them. In GraphProtector [59], Wang et al. created a user interface to help test how different algorithms hide and protect sensitive attributes in a node-link diagram. Wang et al. [60] had previously built a system that helps users understand the privacy risks associated with selecting certain privacy parameters for multi-attribute tabular data. They allow the data curator to apply various algorithms, including differential privacy, to the tabular data and and use a utility comparison view to help the user make decisions. Dobrota [10] designed a tool to help data curators select an appropriate ϵ by displaying visualizations that show the difference in values between the original and private data. Likewise, Nanayakkara et al. [44] and John et al. [34] created visualizations for selecting ϵ values. These systems guide the data curator in their parameter decision-making process for a single dataset. They do not test or provide concrete information on how the parameters affect a broad range of data and visual outputs. A data curator trying to develop a background understanding of the parameters affecting visual output would need to go through the time-consuming and tedious process of using previous systems to guess and check across a variety of datasets. In this paper, we try to quantify and analyze this information so that end users can more readily develop an intuition of what will work.

Within the narrower focus on scatterplots, several researchers have investigated the effect of applying privacy techniques. In one of the first papers on privacy-preserving visualization, Dasgupta et al. [7] demonstrated how adjusting the k value in k-anonymity will affect a scatterplot's privacy and utility based on the user's ability to encode and decode uncertainty. Zhang et al. [69] tested how different levels of privacy affect users' ability to complete a variety of tasks using scatterplots. Their study examined one algorithm (Laplace) at three levels of privacy, demonstrating that users better retain their ability to do summary tasks versus value tasks with private plots. In a similar study, Lee [39] provided some background information on how differential privacy affects visualization outputs for a variety of charts, but no concrete guidance aside from showing that visual quality quickly degrades for private scatterplots with differential privacy. From these papers the reader is left with essentially one conclusion regarding selecting parameters to create private scatterplots—that the data curator should maximize ϵ. While this is important knowledge, it only addresses one parameter the data curator must be aware of.

In the literature, there is little guidance regarding the other necessary decisions a data curator must make: algorithm, bin size, or how data distribution and user task will affect the utility of the private scatterplot. Instead of taking a broad scope and investigating many visualizations or only investigating one parameter, our study narrows the visualization type allowing us collect and analyze data on all the necessary parameter choices. By quantitatively assessing *all* the parameters necessary for the generation of differentially private scatterplots, we create concrete, easy, and accessible guidelines to increase the visual utility of privacy-preserving scatterplots.

3.3 Visual Utility

While research on differential privacy *agrees on* ϵ as a metric for privacy-preservation, there is *little consensus* about how visual utility can be qualified or quantified [4]. Recall that these are separate considerations—e.g., in Fig. 2 we fixed ϵ but the visual utility of the three private visualizations could vary markedly. Different research areas (differential privacy, data visualization, and privacy-preserving visualizations) each have come up with different definitions and metrics of utility (Table 1). This problem is compounded by the fact that two plots with the same statistical summaries can be visually distinct [4], [67] (see, e.g., Fig. 4 and [2], [41]). Therefore, a data curator expecting to retain the high visual utility of the released data may assume that an automated metric preserves both privacy and visual utility and/or revert to carefully inspecting each plot. We aim to solve the data curator's predicament by providing a better understanding of visual utility and how to automate its evaluation. We tackle this problem in three stages: collect ground truth on the visual utility of private plots, find common and effective automated metrics, and test which metrics best align with the ground truth data.

To define visual utility and evaluation criteria, we first look at data visualization literature. Utility or effectiveness is often quantitatively evaluated by measuring how accurately users can complete a task using the provided visualizations [18]. Our evaluation tasks will test if users can extract the same overall patterns from the private data since differential privacy is meant to preserve aggregate patterns without revealing individual information. Therefore, we look for metrics related to evaluating the shape similarity or similarity of two scatterplots as they correspond to pattern retention. Previous work by Matute et al. [42] and Pandey et al. [47] use human evaluators to set perceptual groupings as ground truth for plot similarity. Our work similarly uses human perception as ground truth, but we extend the definition of utility beyond overall plot similarity to be task (correlation, clusters, distribution) dependent.

The work by Zhou et al. [72] uses this definition of visual utility when designing an interactive system to modify private synthetic data generators. Parts of their interface evaluate how well the visual patterns are preserved. While their method is promising, the privacy algorithms they use and compare require an ϵ that is meaningfully higher than our maximum ϵ and larger than recommended standards to produce a visual with reasonable utility [17]. In a similar paper, Zhang et al. [69] investigate how differential privacy affects the utility of different types of visualizations and provide a basis for the work described here. They collect data on the utility of a variety of differentially private visualizations by asking users to give exact numerical answers for specific tasks. We do not use this approach as differential privacy is meant to *prevent* accurate numerical answers—participants will inherently respond erroneously due to the added noise. Instead, we ask participants to

Metric	Field	Explanation	Reasoning
MS-SSIM [61] (Wang et al.)	Image Similarity	Measures image structural similarity taking into account influences of pixels close to each other.	Our goal in retaining utility is retaining visual similarity. MS-SSIM evaluates how closely one image resembles another so therefore it will likely correspond closely with visual utility.
Average Per Query Error (APQE) (Hay et al.) [29]	Differential Privacy	Takes random 2D queries of the data and finds the difference between original data and private data.	APQE has been used to benchmark utility of a variety of algorithms across many parameters [29]. We include this metric under the assumption that smaller differences in the original and private data will lead to more similar visual outputs.
Earth Mover's Distance [60] (Wang et al.)	Privacy-preserving visualizations	Compares the probability distributions on two dimensions to evaluate similarity.	Earth Mover's Distance has been used to evaluate privacy/utility tradeoffs in multi-attribute data [60]. We include this metric because we believe the closer the distributions of two datasets, the more likely they will retain their visual patterns.
KSTest [49], [55] (Tao et al.)	Differential Privacy	Kolmogorov–Smirnov test—The probability that two samples are drawn from the same distribution.	KSTest is available in the popular differential privacy benchmarking library SDGym [49] and has been used to compare synthetic data generators [55]. We include this metric because we believe the closer the distributions of two datasets, the more likely they will retain their visual patterns.
Scagnostics [62] (Wilkinson et al.)	Data Visualization	Graph-based metrics quantifying scatterplot shape across 9 criteria.	Scagnostics has been used to evaluate scatterplot similarity [42]. We believe that if two scatterplots are similar across the 9 criteria, then they will be visually similar as well.

TABLE 1: Automated utility metrics from several domains and our reasoning for their inclusion in the study.

evaluate how well general patterns in the data are preserved. Our study borrows from this previous work to create the first large publicly-available corpus of visual utility ground truth against which automated metrics can be accurately measured.

There is a lack of guidance in the literature regarding which automated (statistical) metric one should choose to best evaluate the visual outputs of private data. We define automated as metrics that can be derived from statistics of the data and therefore require no human in the loop. Automated metrics can help data curators particularly when a dataset has many attributes or the curator has insufficient time to visually inspect the differentially private outputs to ensure they retain appropriate levels of utility for the data user. Therefore we collect common metrics from a variety of fields—described extensively in Table 1—to see which metric most effectively quantifies visual utility. All the metrics are used to evaluate utility in their respective fields, but data curators have no empirical evidence for which is best as they are never compared against one another. We implement and evaluate each of these metrics against the ground truth of visual utility set forth earlier.

To summarize, our paper is the first to collect a *large dataset of visual utility scores* and *automated metrics scores* and *compare* them against each other. We are the first to rank these metrics based on their efficacy in predicting the visual utility. This provides data curators with empirically-derived guidance for which metrics to use when they wish to automatically determine the visual utility of a private plot.

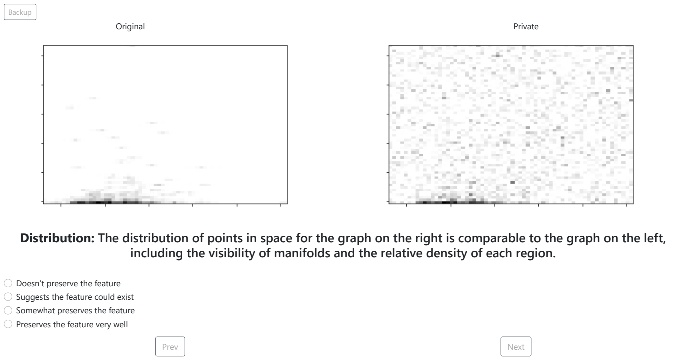
4 Experimental Design & Methods

Our goal is to provide actionable guidance to data curators for each parameter choice they must make when generating private scatterplots. Therefore, we test and evaluate how all the necessary parameter choices (differential privacy algorithm, privacy level (ϵ), bin size, data distribution, user task) will affect the visual utility of private scatterplots. Our study design process emulates the knowledge acquisition process that a data curator might go through after working with many different datasets and parameters. To do so, we conduct a *data study* [51] where many datasets are viewed by a few people rather than a few datasets being viewed by many people. Following prior work, [50], [51] we choose this design since *we anticipate that data parameters have a much bigger impact on visual utility than human perception variability*. We confirm this using an inter-rater reliability metric (see Section 4.2 for more details). This metric shows that the vast majority of the variation in the utility ratings stems from the manipulated variables rather than the reviewers' perception of the output. We also choose a small group of raters since we want our raters to have extensive data analysis experience to provide a robust measure of visual utility. This design allows us to maximize the number of variables tested.

In our study, we use three coders who discussed the coding in depth in a small pilot study to ensure there was consensus on the utility rating scale and descriptions of the task. The three coders scored a large (1,200), varied group of differentially private charts on their visual utility. The charts were generated using varying differential privacy algorithms, privacy levels, data distributions, tasks, and bin sizes (see Section 4.3 for further explanation of variables).

- **5 algorithms for adding noise:** $[\text{DAWA}, \text{AHP}, \text{AGrid}, \text{Laplace}, \text{Geometric Truncated}]$
- **4 privacy levels ϵ:** $[0.50, 0.10, 0.05, 0.01]$
- **2 bin sizes:** $[32 \times 32, 64 \times 64]$
- **20 data distributions:** $[0..19]$, each a canonical scatterplot from each of Pandey et al.'s categories [47]
- **3 user tasks:** $[\text{Clusters}, \text{Distributions}, \text{Correlation}]$, filtering the tasks as appropriate for each data distribution

The second portion of the study analyzes how automated utility metrics correspond to expert visual utility evaluations. This design expands upon the framework laid out by Matute et al. [42]. They first gather the ground truth (human visual perception) and then compare it to automated statistics (traditional utility metrics). We found common utility metrics represented in a variety of domains and found the strength of correlation with our visual utility evaluations of the different plots. The goal of this portion of the study is to help data



Distribution: The distribution of points in space for the graph on the right is comparable to the graph on the left, including the visibility of manifolds and the relative density of each region.

○ Doesn't preserve the feature
○ Suggests the feature could exist
○ Somewhat preserves the feature
○ Preserves the feature very well

Fig. 5: Screenshot of study website. Code to generate study can be found in the supplemental materials.

curators effectively automate the process of visual utility evaluation.

4.1 Procedure

To gather the visual utility evaluations for the 1200 charts, a study website was created (see Fig. 5). The website contained a page for each parameter combination with forward and back buttons. The reviewers could see the binned scatterplot with no noise added on the left and the private binned scatterplot on the right. The questions were blocked by task and randomized within tasks to help the reviewers remain cognizant of the criteria they were evaluating the private plot on. They evaluated the ability of the private plot to retain the task completion on a 4-point Likert scale, developed through discussion and consensus agreement of the coders: [**0: Doesn't** preserve the feature: I have no confidence that the feature exists. **1: Suggests** the feature could exist: It looks like the feature might exist but I have low confidence and/or the feature is shown with little clarity. **2: Somewhat** preserves the feature: I'm confident this feature exists but its fidelity is meaningfully lower than the original plot. **3: Preserves** the feature very well: I'm confident the feature exists and it is shown with high fidelity relative to the original plot.]

We chose to create four categories since the nuances between different ratings of a more fine-grained scale would be difficult to discern. Before the study was run, the three experts coalesced on the agreed-upon definitions and rankings of utility by discussing a representative sample of questions. The representative plots were created with all the same parameters used in the actual study but used data distributions that were not seen in the actual study. Having these discussions was critical to ensure there was consensus on what utility entails and resulted in a high IRR score during the data collection process.

4.2 Inter-Rater Reliability (IRR)

To validate our choice of study design and utility rating system we ensured that the variability in utility ratings stemmed from the parameter changes rather than the coders' perceptions. To do so, we ran a pilot study and calculated the inter-rater reliability (IRR) to ensure the trained coders agreed upon the evaluation of the data. We set a-priori threshold for our IRR at .6 which is classified as substantial agreement by Kevin Hallgren [27]. This allows us to attribute the differences in the scores to the data and not to the users.

To test the inter-rater reliability we follow the guidelines for choosing the correct intraclass correlation coefficient (ICC) as set out by Koo and Li [36]. Using their criteria, our ICC selection corresponds to the two-way mixed effects, absolute agreement, single rater/measurement—ICC(3,1). For our pilot study we found an IRR of .7311 with a 95% confidence interval of [0.68, 0.78]. This was an acceptable score based on Kevin Hallgren's recommendations. The IRR for the final data gathering portion was .831 with a 95% confidence interval of [0.82, 0.85]. This is well in the recommended IRR score and is classified as good by Koo and Li [36].

See our preregistration (https://osf.io/25xhn, sec. 1.5.1) for the pilot analysis & data.

4.3 Manipulated Variables

Differential privacy algorithms have varied performance based on many factors [29]. We test the performance of each combination of differential privacy algorithm (Section 4.3.1), privacy level (ϵ) (4.3.2), bin size (4.3.3), data distribution (4.3.4), and user task (4.3.5) to get a precise understanding of where they perform best.

4.3.1 5 algorithms for adding noise

This categorical variable has values: [DAWA, AHP, AGrid, Laplace, Geometric Truncated].

We selected algorithms with an open-source implementation that represent a broad range of best-in-class results. After examining papers that benchmark popular differential privacy algorithms and open-source libraries [25], [29], [55], [70], we settled upon 5 algorithms. Our differential privacy expert co-author (Prof. Ullman) verified that the selected algorithms were a relevant and representative sample.

We choose to investigate both data-independent and data-dependent algorithms, where the primary difference is how they add noise to the underlying data. Data-*independent* algorithms add noise without considering the input data, while data-*dependent* algorithms add noise based on the structure of the input [40]. Since the output of data-dependent algorithms is related to distribution, different algorithms will perform better on different shapes [29]. This results in no single algorithm outperforming all others across all types of data sizes, shapes, and privacy parameters [30], [69]. Therefore, to find the best algorithms for each set of variables, we test multiple data-independent and data-dependent algorithms.

We included **2 data-independent algorithms:** Laplace and Geometric Truncated. Both implementations can be found in the open-source library diffprivlib [32]. We choose to test the Laplace mechanism as it is (1) the original proposition for noise addition [13], (2) has proven to be a safe general choice when adding noise to visualizations [69], and (3) is a good benchmark to compare other algorithms against [29], [40], [69]. Our choice of using Geometric Truncated is motivated by Garrido et al. [25], who benchmarked algorithms available in open-source differential privacy libraries. Garrido et al. found that Geometric Truncated [26] from diffprivlib outperforms other algorithms when adding noise to count queries. As count queries are the basis of differentially private scatterplots, we consider Geometric Truncated to be our best-in-class data independent algorithm.

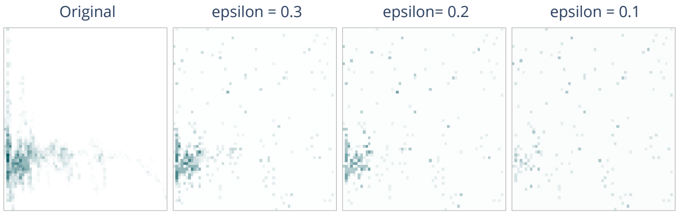


Fig. 6: Small changes in privacy level ϵ can cause large disproportional changes in graphical representation. A .1 change in ϵ from .3 to .2 and .2 to .1 changes the data privacy by nearly an equivalent amount. The dropoff in visual utility is not proportional as the plot with $\epsilon = 0.1$ retains little visual utility.

In addition to the 2 data-independent algorithms, we analyze **3 data-dependent algorithms**: DAWA [40], AGrid [48], and AHP [71]. Based on the metrics set out by Hay et al. [29], DAWA performs the best overall of all the algorithms and best overall for the small sample sizes. Perhaps more relevant for this paper, DAWA has also been found effective for maintaining the visual utility of 1D histograms [69]. AGrid was selected because it often performs well on shapes that DAWA had difficulty with [29]. Finally, while DAWA and AGrid are designed for accurately answering range queries, AHP is a general-purpose algorithm for accurately creating histograms [37]. We execute all the algorithms with the default parameters. Our target user is not a differential privacy expert so we expect them not to have the time or expertise to accurately tune the additional parameters outside of ϵ. All three of the algorithms are implemented in DPComp [11]. At https://osf.io/wej4s/ we provide instructions for using the algorithms with your own data.

4.3.2 4 privacy levels (ϵ)

This quantitative variable has values: $[0.50, 0.10, 0.05, 0.01]$.

Tuning ϵ allows the data curator to decide the tradeoff between accuracy and privacy. Since differential privacy is a relatively new concept, guidelines for setting ϵ have not yet been developed [64]. Dwork et al. state that the value of ϵ is more of a social question, so therefore we want to give data curators the ability to see multiple privacy levels [13]. While there are no strict guidelines for setting differential privacy, differential privacy experts do agree that the value of ϵ should be below one [15], [64]. Additionally, high ϵ does not allow for accurate comparisons across algorithms so we keep ϵ relatively small [15].

To find the levels of ϵ we want to test, we first ran simulations on a variety of datasets at different levels of ϵ. This practice is standard in differential privacy applications [15]. The levels of ϵ were chosen at levels that often corresponded visually with the four possible visual utility options: [doesn't, suggests, somewhat, does] retain the task. The highest level of privacy where any utility was retained at a domain size of 5,000 was $\epsilon = .01$ (Doesn't). On the other hand, at $\epsilon = .5$, the private graphs almost always retained their full utility (Does). To test the nuances of the different algorithms, we also evaluate them at $\epsilon = .1$ (somewhat) and $\epsilon = .05$ (suggests). In this way we hope to find the differences in the algorithms chosen depending on the privacy level.

Dataset Size	5,000	10,000	50,000	100,000	1,000,000
ϵ	0.1	0.05	0.01	0.005	0.0005

TABLE 2: All algorithms evaluated are scale-ϵ exchangeable [29]. For each of the dataset sizes, the corresponding ϵ will produce the same private plot if all other parameters are kept consistent. Therefore, our results can be extended to larger or smaller datasets. To find the corresponding ϵ for another dataset size use the formula: $\frac{CountOther}{CountTable} * \epsilon_{Table}$.

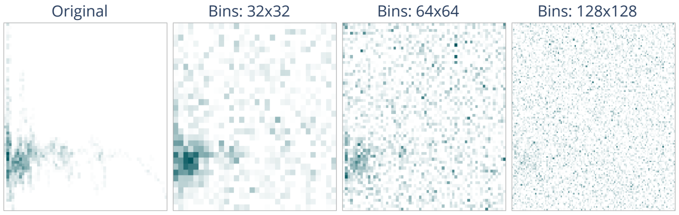


Fig. 7: At the same level of ϵ, the utility of the visualization can worsen if too fine of a bin size is selected (128x128).

One more very important point to note is that each of the algorithms chosen is scale-ϵ exchangeable [29]. What this means is that increasing the number of records or increasing the ϵ have an equivalent effect. We can think of the number of records and ϵ as being inversely proportional so two datasets, one consisting of 10,000 rows and one consisting of 1,000 rows, of the exact same distribution, will have the same error at an ϵ of .1 and 1 respectively. This is critical as it allows our results to be extended to any domain size with a simple calculation. While this is the case it is not advisable to apply differential privacy to small datasets [15]. At a certain size, the noise will either distort the visual pattern too much or the privacy guarantee will not be sufficient for any practical circumstance. Looking at Table 1 we can see how our choices of ϵ for a domain size of 5,000 can be expanded to a variety of different domain sizes.

4.3.3 2 bin sizes

This quantitative variable has values: $[32 \times 32, 64 \times 64]$

Bin size can affect the utility of the visualization by both making the patterns coarser and affecting the perturbation errors. The same amount of noise will make a smaller impact on coarser bins than finer bins. Finding the correct number of bins is a difficult problem even without privacy constraints [65]. There are a variety of approaches proposed from previous research to find the optimal binning of dataset [20], [54]. When adding noise from the differential privacy algorithms we need to additionally take into account the impact this will have on the output. A histogram with finer bins may lead to lower accuracy since the perturbations will have a larger effect on each bin [66]. Therefore, to select the appropriate number of bins we took three different approaches: previous research, algorithmic partitioning, and visual inspection.

The work by Hay et al. [29] evaluated a variety of algorithms on 2D histograms and partitioned the data into 32x32, 64,x64, 128x128, and 256x256 bins. Their datasets contained a million or more records so we focused on the lower range of bin numbers. We then ran the pilot datasets through a variety of partitioning algorithms employed by the numpy histogram_bin_edges package [28]. The different partitioning algorithms generally returned bin amounts near the 32x32 and 64x64 range. We therefore selected to study both the 32x32 and 64x64 bin amounts since they retained the nuances of the original data but the visual pattern could alter depending on the amount of noise added.

4.3.4 20 data Distributions/Scatterplot shapes

This categorical variable has values $[0..19]$, each a canonical scatterplot from each of Pandey et al.'s categories [47].

Since the shape of the data plays a crucial role in the utility of the data-dependent algorithms [29], it was important to select a representative sample of a wide variety of scatterplots. We chose to avoid synthetic datasets as those patterns are rarely seen in real-world applications [47]. Instead, our scatterplots are based on the scatterplot perceptual groupings presented by Pandey et al. [47]. In their paper, they narrow down 84,000 scatterplots to a selection of 247 representative plots which were grouped into 20 categories by human subjects based on visual similarity. Our study analyzes one scatterplot from each of the 20 categories—meeting our criteria of a varied, real, and representative sample of plots.

Many of the datasets used by Pandey et al. are small, consisting of fewer than 1000 rows. While this is common in real-world datasets, differential privacy is not designed for such small datasets [15]. We are not testing whether the algorithms work on these specific datasets, but instead, we aim to test how well the algorithms preserve the *visual patterns* these datasets represent. Therefore, each plot received additional data points until it reached 5,000 points. 5,000 rows represent the minimum amount of points where our maximum ϵ of .5 consistently produces a good visual representation. For datasets less than 5,000 points, we increase the size of the datasets preserving the bivariate distribution of the points. We added synthetic points to each dataset in a systematic manner using copulas [58], random Laplace jitter, or simply duplicating points. We chose the most appropriate technique for each dataset by visually inspecting and comparing the original and larger datasets points plotted in a scatterplot and binned scatterplot. Any plot that had more than 5,000 rows, had rows randomly removed until it also contained 5,000 rows. The code to generate the datasets and the plots generated for visual inspection can be found in the supplemental materials.

During the data study, the experts saw side-by-side the binned scatterplot with no noise and one with noise added. They were asked how well they could complete the assigned task. We set our color scale for the binned representation to the standard matplotlib library greyscale [33]. We used popular plot-generation libraries and a typical unaltered color scale to try and replicate the process of a data curator who does not have the time or expertise to go through and specifically design each graph.

4.3.5 3 user tasks

This categorical variable has values: [Clusters, Distributions, Correlation], filtering the tasks as appropriate for each data distribution.

Differential privacy changes the ability of the user to complete certain tasks more than others [69]. Therefore, we assess visual utility based on the user's ability to complete certain tasks [67], [69].

When we add differential privacy to the private scatterplots, they are meant to do just that; keep the data private. Therefore, users should not be able to extract any values about an individual (retrieve value, identify outlier) or else the whole premise of privacy dissipates. Our study uses tasks that are difficult to accurately describe numerically through aggregate statistics but can be quickly discerned by viewing a scatterplot and do not involve individual data points. This criteria results in three tasks: *identify correlation (including non-linear)*, *identify clusters*, and *characterize distribution*.

These tasks are extracted from Micallef et al.'s five most common user tasks for scatterplots [43]. The three experts went through many pilot questions and discussed each task until a succinct comprehensive definition was agreed upon:

- **Distribution:** The distribution of points in space for the non-private graph is comparable to that of the private graph, including the visibility of manifolds and the relative density of each region.
- **Correlation:** The private graph preserves the level of dependence between the two attributes—including non-linear dependence.
- **Clusters:** The clusters visible in the non-private graph—and no other clusters—are visible on the private graph and occur in the same places.

Using these agreed-upon definitions and a 4-point coding scale the experts were able to evaluate the visual utility in relation to each task.

4.4 Analysis

4.4.1 Parameter/Algorithm Comparisons

The first step in our analysis is creating one ranking from the three different raters. We use the median of the ratings. Since the data is ordinal and does not have equally defined spacing between the four ratings, we do not use the mean [53].

Next, we test how the different parameters affect the visual outputs using statistical tests and visual data inspection. The data is filtered based on the various parameters (4 privacy levels (ϵ), two bin sizes, three tasks) tested. The analysis then follows a three-step process:

1) Using the Friedman test [21], we check whether there is sufficient evidence to say that there is a difference between the five different algorithms if the Friedman test returns a p-value < .05 we continue on step two - the post-hoc analysis.
2) We conduct a post-hoc Conover [6] test to see which algorithms differ from each other. This gives us more specific insight into which algorithms are different from one another.
3) We visualize the data to examine the results from the post-hoc Conover test. The post-hoc Conover states if there is a difference but does not provide direction. Therefore we use visuals to check the practical significance of our effect sizes and determine which algorithm retained a higher utility [12].

4.4.2 Comparison to Utility Metrics

We want to measure the strength of association between the visual utility ratings generated by our coders and metrics of utility that can be generated computationally. Our coder rankings are ordinal, and the metrics are continuous. When obtaining the association of ordinal-continuous variables, it is recommended to use Kendall's coefficient of rank correlation τ_b range from -1 (perfect negative association) to 1 (perfect positive association) [35]. To test whether our different metrics produce different strengths of correlation, we will use the Fisher z-transformation to see if there is a meaningful difference between the correlation coefficients. We then create a rank order of metrics based on the highest absolute value of correlation coefficient $|\tau_b|$ that has a p-value of less than .05. The metrics and their explanations can be found in Table 1.

5 Results

We will first discuss how the different algorithms perform overall across all parameters and then break down performance based on the different levels of each parameter. As previously found in the literature [29], Ⓐ **DAWA perform best of all the algorithms** aggregated across all manipulated variables (ϵ, bin size, task, distribution, algorithm). Geometric Truncated also performs well as found in the previous literature by Zhang et al. [70]. Ⓑ **Laplace, AGrid, and AHP generally provide the lowest visual utility**.

Of all the parameters, Ⓒ **ϵ plays the largest role in determining the visual utility of the output.** At the highest $\epsilon = 0.5$, There is meaningful evidence that AHP and Geometric Truncated perform best (Post-Hoc Conover - $p < .1$ compared to other algorithms). While AHP performs well at higher information reveal (high ϵ), its utility quickly drops off at the lower ϵ's of 0.05 and 0.01. DAWA consistently performs well throughout but particularly outperforms the other algorithms at lower information reveal ($\epsilon = .05$ and $\epsilon = .01$). Ⓒ **No algorithm performs well at an $\epsilon = .01$** for a dataset of 5000 points. Only DAWA produced any charts with a utility rating that was not 0. *Implication: at higher information reveal (high ϵ) use AHP, Geometric Truncated. At low information reveal (low ϵ), DAWA is your best bet.*

The tasks also play a role in how much utility is retained when the data is privatized. Ⓓ **Identifying clusters was the task that best retained its visual utility**. This is followed by correlation and distribution for task completion difficulty. This may be because cluster tasks do not require seeing as much fine-grained detail as the other tasks. Algorithm performance also varied based on task. Ⓓ There is evidence that DAWA meaningfully outperforms the other algorithms for the correlation task, and AHP, DAWA, and Geometric Truncated take a three-way tie for the best algorithm for distribution tasks.

The bin sizes have a smaller effect on the utility than ϵ but still impact the visual utility rankings. The Ⓔ **coarser bins (32x32) better retain the utility** (Wilcoxon Signed rank test with all parameters but bins equal) It is important to keep in mind that the utility retention was made by comparing non private binned plots with private binned plots of the same bin size. For coarser bins, both DAWA or Geometric Truncated perform reasonably well but at finer bin sizes (64x64) DAWA outperforms all the algorithms (Post-Hoc Conover test $p < .1$).

Finally, our utility metrics results provide interesting insights into which automated metrics can be used to best evaluate visual utility. Ⓕ **The automated utility metric that most strongly correlates with visual utility is MS-SSIM [61]**. MS-SSIM provides data curators with the opportunity to compare any parameter combination against any other parameter combination. This allows data curators to optimize across all manipulated variables.

6 Discussion

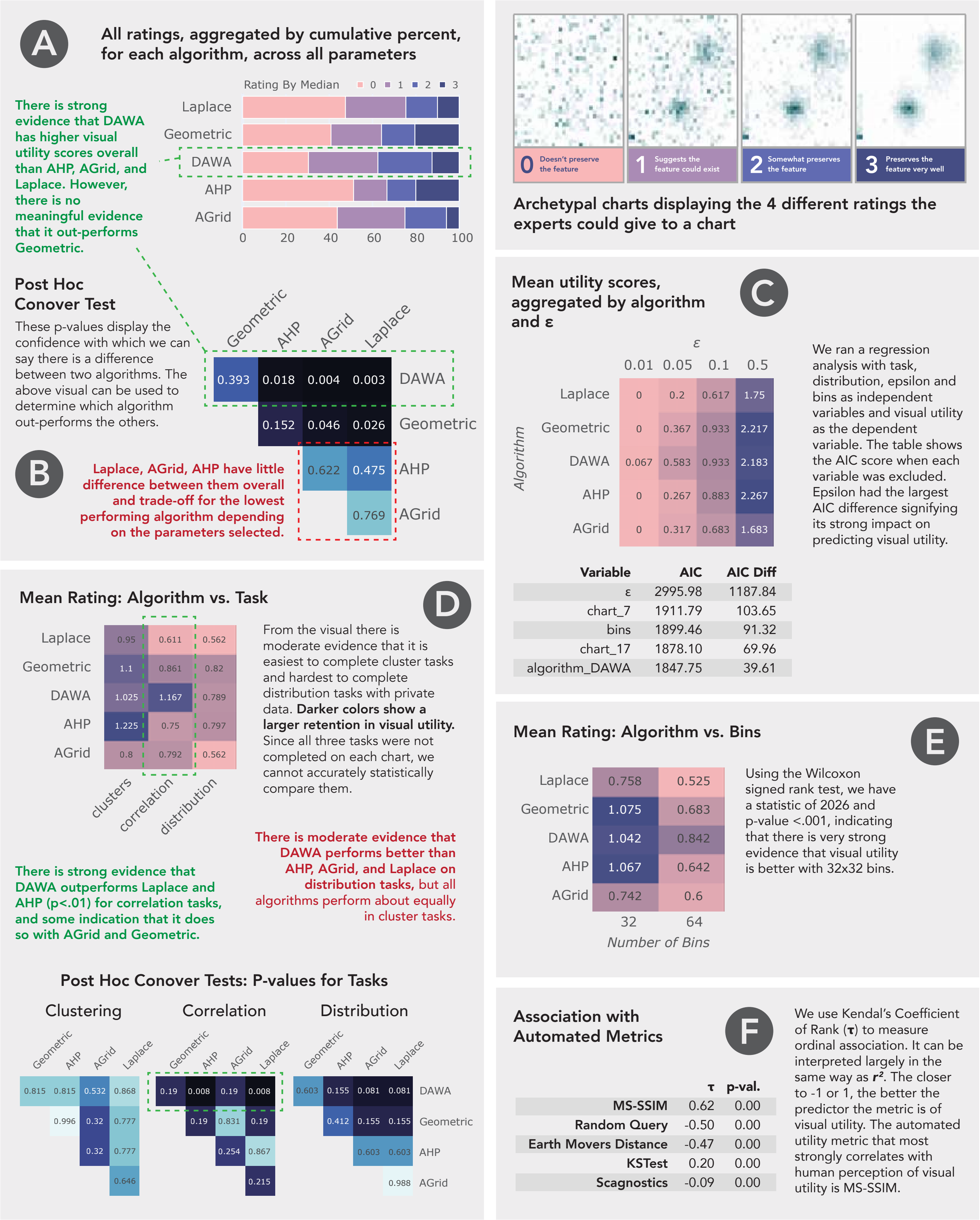
The results of our study led to many concrete guidelines but also many observations on how privacy affects the utility of private plots. In this section, we summarize observations that will help data curators maximize the visual utility of their private plots.

6.1 Signal to Noise

All the parameters that we evaluated (except the task) influence the way noise is added to the aggregated counts. This is particularly apparent between the data-independent and data-dependent algorithms. Data-dependent algorithms use various clustering strategies to group bins and make them the same count, creating large areas of uniform color (see Fig. 4 (left)). The data-independent algorithms add noise to each bin independently of the adjacent bins, often creating visualizations with the look of static TV (Fig. 2 Laplace).

By observing these variations in noise, the expert evaluators came to several conclusions about their ability to perceive the true underlying data. For low levels of privacy (small amounts of noise), the signal-to-noise ratio is often high and therefore it is easier to differentiate between what is noise and what is the underlying data. Additionally, the large areas of uniform count that data-dependent algorithms generate can be easily classified as noise and ignored. This is compounded if the user is familiar with the algorithm, as the private outputs show similar patterns of noise addition for the same algorithm. Other algorithms, such as Laplace, can create negative bin counts, which a data user can ignore since a binned scatterplot should never have a count less than 0. Finally, a prior understanding of the data has a large influence. Our experts saw the original data alongside the private data and commented that there were times they were not sure if they would have seen the patterns in the private data if they were not aware of them beforehand. Therefore a data user with a stronger prior belief in what the data shape will be can likely better confirm their beliefs and see the signal past the noise than a user completely unfamiliar with the data.

These observations may explain why DAWA and the cluster task had the highest visual utility ratings. DAWA often has large areas of uniform count (Fig. 2 Fig. 4), and as stated previously, this type of noise may help reduce the obfuscation of the real data by the noise. For clusters, the noise often does not overshadow the strong signal that represents a dense area of points. Therefore, it is easier to complete this task than (for instance) distribution, where subtle patterns are often lost.



There is strong evidence that DAWA has higher visual utility scores overall than AHP, AGrid, and Laplace. However, there is no meaningful evidence that it out-performs Geometric.

Archetypal charts displaying the 4 different ratings the experts could give to a chart

Post Hoc Conover Test

These p-values display the confidence with which we can say there is a difference between two algorithms. The above visual can be used to determine which algorithm out-performs the others.

Geometric	AHP	AGrid	Laplace	
0.393	0.018	0.004	0.003	DAWA
	0.152	0.046	0.026	Geometric
		0.622	0.475	AHP
			0.769	AGrid

Laplace, AGrid, AHP have little difference between them overall and trade-off for the lowest performing algorithm depending on the parameters selected.

Mean utility scores, aggregated by algorithm and ε

	0.01	0.05	0.1	0.5
Laplace	0	0.2	0.617	1.75
Geometric	0	0.367	0.933	2.217
DAWA	0.067	0.583	0.933	2.183
AHP	0	0.267	0.883	2.267
AGrid	0	0.317	0.683	1.683

Variable	AIC	AIC Diff
ε	2995.98	1187.84
chart_7	1911.79	103.65
bins	1899.46	91.32
chart_17	1878.10	69.96
algorithm_DAWA	1847.75	39.61

We ran a regression analysis with task, distribution, epsilon and bins as independent variables and visual utility as the dependent variable. The table shows the AIC score when each variable was excluded. Epsilon had the largest AIC difference signifying its strong impact on predicting visual utility.

Mean Rating: Algorithm vs. Task

	clusters	correlation	distribution
Laplace	0.95	0.611	0.562
Geometric	1.1	0.861	0.82
DAWA	1.025	1.167	0.789
AHP	1.225	0.75	0.797
AGrid	0.8	0.792	0.562

From the visual there is moderate evidence that it is easiest to complete cluster tasks and hardest to complete distribution tasks with private data. **Darker colors show a larger retention in visual utility.** Since all three tasks were not completed on each chart, we cannot accurately statistically compare them.

There is strong evidence that DAWA outperforms Laplace and AHP (p<.01) for correlation tasks, and some indication that it does so with AGrid and Geometric.

There is moderate evidence that DAWA performs better than AHP, AGrid, and Laplace on distribution tasks, but all algorithms perform about equally in cluster tasks.

Post Hoc Conover Tests: P-values for Tasks

Clustering

Geometric	AHP	AGrid	Laplace	
0.815	0.815	0.532	0.868	DAWA
	0.996	0.32	0.777	Geometric
		0.32	0.777	AHP
			0.646	AGrid

Correlation

Geometric	AHP	AGrid	Laplace	
0.19	0.008	0.19	0.008	DAWA
	0.19	0.831	0.19	Geometric
		0.254	0.867	AHP
			0.215	AGrid

Distribution

Geometric	AHP	AGrid	Laplace	
0.603	0.155	0.081	0.081	DAWA
	0.412	0.155	0.155	Geometric
		0.603	0.603	AHP
			0.988	AGrid

Mean Rating: Algorithm vs. Bins

	32	64
Laplace	0.758	0.525
Geometric	1.075	0.683
DAWA	1.042	0.842
AHP	1.067	0.642
AGrid	0.742	0.6

Using the Wilcoxon signed rank test, we have a statistic of 2026 and p-value <.001, indicating that there is very strong evidence that visual utility is better with 32x32 bins.

Association with Automated Metrics

	τ	p-val.
MS-SSIM	0.62	0.00
Random Query	-0.50	0.00
Earth Movers Distance	-0.47	0.00
KSTest	0.20	0.00
Scagnostics	-0.09	0.00

We use Kendal's Coefficient of Rank (**τ**) to measure ordinal association. It can be interpreted largely in the same way as ***r*²**. The closer to -1 or 1, the better the predictor the metric is of visual utility. The automated utility metric that most strongly correlates with human perception of visual utility is MS-SSIM.

Fig. 8: The figure displaying the results presented in Section 5. The figure displays A) overall algorithm performance, B) Post-hoc Conover comparison of all algorithms, C) the most impactful variable, ϵ's, affect on visual utility, D) task related performance of the algorithms, E), bin size affect on utility, and F) ranking of automated utility metrics.

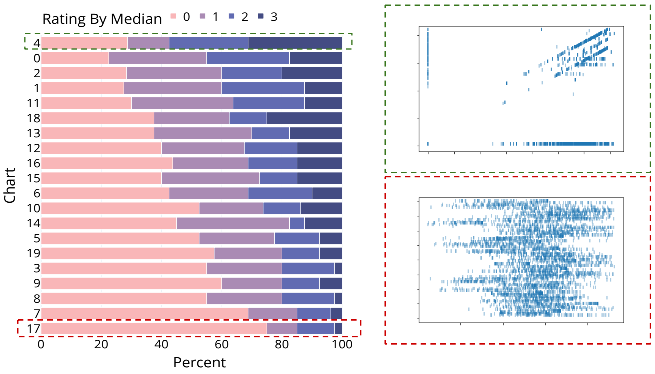


Fig. 9: The stacked bar chart consists of cumulative percents of the raters scores aggregated across all algorithms. Chart 4 (green, top) had the highest mean utility score while chart 17 (red, bottom) had the lowest mean utility scores.

6.2 Epsilon's Influence

While we tested many parameters, it is apparent from the analysis that by in large the change in utility results from changes in ϵ, with the other parameters providing small variations. One strategy that data curators can use to increase their maximum ϵ, is relaxing the ϵ constraint by using the popular method of (ϵ, δ) differential privacy outlined by Dwork et al. [16]. While "pure" differential privacy gaurantees that the maximum privacy loss is ϵ on all possible queries, the δ relaxation allows us to increase ϵ by instead guaranteeing the privacy loss does not exceed ϵ with probability at most $1 - \delta$ (pure would be a probability of 1).

Even small changes in the ϵ can make a drastic difference in the visual utility of the private plot. This is particularly true for several of the data-dependent algorithms (Fig. 6 shows the data-dependent algorithm AHP). For several algorithms, namely AHP, there is a clear dropoff point where the private plots greater than a certain ϵ retain some utility and smaller than that ϵ retain almost no utility (Fig. 8 C). On the other hand, the data-independent algorithms and DAWA have a more proportional drop in utility as privacy increases. For the algorithms that have sudden drop-offs in utility, data curators can run automated checks using the MS-SSIM metric to find the lower bounds of ϵ that retain the visualization's utility.

6.3 Data Distribution

The distribution of the data plays a role in the utility and efficacy of differential privacy and the algorithm selected. In their comparison of various algorithms on different data distributions, Hay et al. [29] discuss how certain algorithms, particularly data dependent algorithms, perform better on different distributions. They state that no algorithm performs best on all data distributions and sizes. Our findings agree with this statement but we also find that certain data distributions have lower visual utility across all algorithms (Fig. 9). In a sense, the data distribution has less of an effect within the algorithms but certain data distributions are likely to lose more of their utility across all algorithms.

We also investigated which graphs had the highest and lowest scores. Charts with large areas of uniform density performed poorly (Charts 17 and 7) while charts with narrow, "stringy" distributions (Charts 4 and 2) performed best (Fig. 9).

This finding could stem from the processes and assumptions data-dependent algorithms make. Essentially, data-dependent algorithms look for dense regions in the data. The basic premise of the data-dependent algorithm is to partition the data into groups where the signal is similar and then make those regions uniform (We can see these regions as the uniform colors in Fig. 2 DAWA). Therefore, when there is clear dense regions, the data-dependent algorithms easily group those bins together and keep them visually distinct as the counts represent opposite ends of the color scale. On the other hand, when the data distribution is more uniform, the data-dependent algorithms highlight and produce regions that are similar in count but still represent opposite ends of the color scale since the color scale range has changed. Therefore the difference in color for dense regions (Fig. 9 top) for some graphs may look the same as the difference in color of a uniform regions (Fig. 9 bottom) creating the illusion that there is greater variation between certain regions then there really is.

6.4 Visual Post Processing

In addition to the parameter optimization before the output is created, data curators can vary the visual parameters in the output to better highlight the signal amongst the noise. Researchers in the medical field have looked at this particularly with MRI images [45]. Thaker et al. [56] briefly mention the potential effectiveness of doing this when hiding the noise in private scatterplots. Previous literature corroborates this insight and has stressed the importance of selecting an appropriate color scale for the data being presented [5]. We advise the data curator to examine how different color scales and binning of the color scale may influence the visual perception of the data.

6.5 Utility Metrics

Automatic utility metrics should make it easier for data curators to decide which parameters will offer the greatest visual utility. The utility can change based on all of the parameters as evidenced by Fig. 2, Fig. 7, Fig. 10 even when ϵ remains the same. Without a way to quickly and statsitically check for visual utility retention, the creation of optimal private plots can be difficult or time consuming. Our work helps remove this barrier by benchmarking the best automated metric for visual utility. From our findings we found that MS-SSIM [61], was the best way to predict visual utility of a private plot. MS-SSIM can be used to help optimize bin size, ensure visual utility does not drop-off at certain ϵ's (Fig. 6), or to visually post process the plot (Fig. 10). While not all automated metrics are tested, we hypothesize that the area of image similarity may be a good place to look for even more accurate visual utility metrics.

Another result that may seem intuitive but has broad implications is that algorithms deliberately optimize one metric. Having an algorithm optimize utility for one metric may not correspond to optimal utility performance on a different metric (APQE does not translate well to visual utility – Fig. 4). This is evident as Geometric Truncated performed well when evaluated on visual utility but often performed the worst on the APQE metric. Since algorithms are designed to work best on a certain metric they are evaluated on, it would be interesting for future work to try and design an algorithm that is evaluated on visual pattern retention (visual utility) as the metric.

7 Future work

This work does a deep investigation of many parameters on one type of chart but we only scratch the surface of the intersection of visualization and differential privacy. Narrowing the scope to scatterplots was important to provide actionable insights and validate our methodology for finding automated utility metrics. Future work could expand upon our study in several ways. Many more parameters could be tested using our top automated utility metric MS-SSIM. The study methodology can be altered to increase the amount of plots and parameters by having the plots utility evaluated algorithmically instead of by human coders. The parameter choices could be extended and more data distributions added to better understand the nuances in how data distributions and privacy levels affect the visual utility of scatterplots. Futhermore, now that our methodology is established and verified, future work could examine other visualization types using the same process. For instance, future work could find automated utility metrics to evaluate private line charts [23] visual utility. Finally, by investigating other chart types, we can verify if our results extend to other visualization types. Since private scatterplots are 2D histograms, the results may be particularly transferable to 1D histograms. In their algorithm benchmarking study, Hay et el. [29] find similar results to ours stating that DAWA was the top algorithm for 1D histograms.

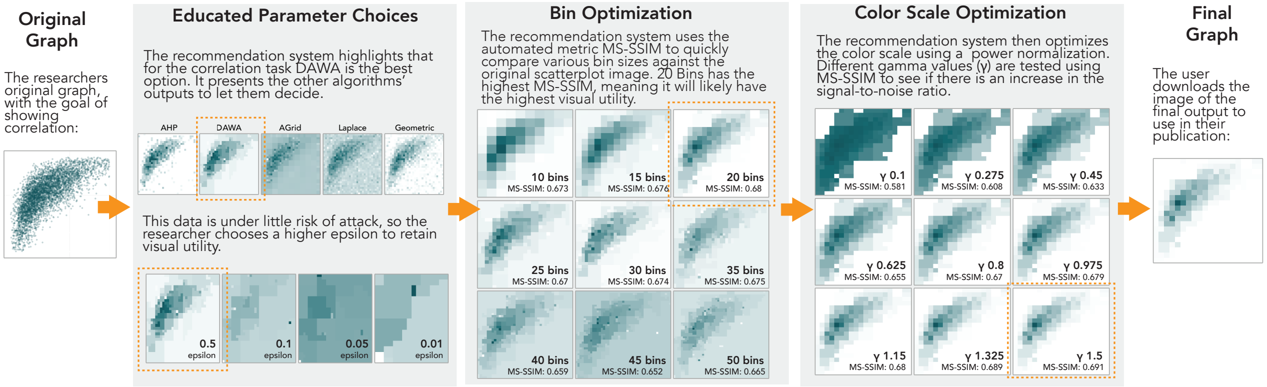
Additionally, our methodology for finding automated utility metrics for private scatterplots could be extended in several ways. First, more ground truth data gathered on a variety of data distributions would increase the validity of the correlations with the automated utility metrics. Second, a finer grain of utility scores could be generated to help differentiate subtler differences between the private plots. Third, more metrics could be run with the existing data to see if they outperform the others. Future work could also encompass creating an original computational metric or using machine learning to train an algorithm to predict whether an outputted private visualizations retains it's utility.

Other aspects of our study design could also be adjusted to extend and validate our results. First, this study could be extended to a user study with many participants. A qualitative portion would also bring in new insights into how participants and data users interact with private data visually. The structure of the questions examining utility can be adjusted to potentially better reflect the type of questions a data analyst would ask of the data. For example, instead of asking the raters to give a score of utility, study participants could be asked to examine both the original and private data and pull out key insights in an open ended format. This study could also be redone with the same parameters but with algorithm optimization. Previous work has found drastic improvements in performance if the parameters of the algorithm are optimized [29]. One algorithm in particular, AGrid, could be improved as it produced an artifact that we choose to ignore when evaluating our visual utility metrics. All these adjustments and variations demonstrate how under-explored this area is.

An extension of our work could also be creating an automated pipeline based off of the results to recommend chart parameters. As an example, a researcher could input a CSV of their data and tune the parameters of the chart to create a scatterplot that they think best displays their data (Fig. 10 Original Graph). After inputting the primary task they hope to retain the utiltiy of, the system could guide them in selecting the best algorithm and ϵ based off of our studies results (Fig. 8 C, D, Fig. 10 Educated Parameter Choices). This system could also optimize parameters using the MS-SSIM metric. For instance, as show in Fig. 10 Bin Optimization and Color Scale Optimization, parameters can quickly be optimized computationally reducing the time consuming activity of visual inspection. This kind of system could lead data curators to make clear and informed parameter decisions or help produce private plots if the dataset has many attributes.

Another area of research with many potential avenues is using visualizations to help explain differential privacy. This could overlap with many of the concepts found in explainable Artificial Intelligence (explainable AI) literature [31]. Both differential privacy and machine learning have complicated algorithms where it is hard for practitioners to understand the underlying mechanisms [1]. Input parameters can be tuned in both to create more optimal outputs [68]. Finally user trust and understanding play a pivotal role. Additionally, while we tested algorithms that take in either one or two variables and output private results, there is a growing field of research related to synthetic differentially private data generation [3] that privatizes multi-attribute datasets. These techniques employ machine learning to learn the patterns of a dataset and output synthetic data that mimics these patterns. Visualization can be of particular use in examining the outputted synthetic data to ensure the quality and information contained in the original dataset remains the same. Future work could therefore look at explainable differential privacy and see if the same methodology and systems that are effective in explaining artificial intelligence can be translated to assist and explain private data generation.

Another under-explored area is interactive private data analysis. This runs in parallel to a burgeoning area of research in differential privacy referred to as adaptive differential privacy [63]. When a user is provided access to a differentially private database, they are often given a privacy budget. The privacy budget is the total allowable ϵ to be used. In differential privacy, ϵ compounds additively with each new query on the same database (with a budget of .5, a data user could view two plots at an ϵ of .25). Therefore, user's of of an interactive visual analysis tool would have to balance their accuracy against number of queries/visualizations. Designing a system to investigate user's trust and understanding of this kind of decision making process would provide valuable insights into creating an interactive, exploratory private visual analysis system.



The researchers original graph, with the goal of showing correlation:

The recommendation system highlights that for the correlation task DAWA is the best option. It presents the other algorithms' outputs to let them decide.

This data is under little risk of attack, so the researcher chooses a higher epsilon to retain visual utility.

The recommendation system uses the automated metric MS-SSIM to quickly compare various bin sizes against the original scatterplot image. 20 Bins has the highest MS-SSIM, meaning it will likely have the highest visual utility.

The recommendation system then optimizes the color scale using a power normalization. Different gamma values (γ) are tested using MS-SSIM to see if there is an increase in the signal-to-noise ratio.

The user downloads the image of the final output to use in their publication:

Fig. 10: We manually generate charts showing how our results can be used to to create an optimal differentially private scatterplot. This process could be converted into a recommendation system in future work to help data curators generate private plots.

8 Conclusion

In our paper we take a critical and thorough look at the many parameters that a data curator would have to sift through when choosing to release sensitive data using differential privacy and scatterplots. This study provides guidance to data curators about which algorithm is best when trying to retain the visual utility of a private binned scatterplot.

We also comment on how the other parameters affect the utility. The privacy parameter, ϵ, has by far the largest impact on a private plots visual utility. Finally, we benchmark automated utility metrics against our ground truth data and demonstrate how the most strongly correlated metric, MS-SSIM, can be used to optimize certain parameters. Data curators, users, and future researchers can benefit by better understanding how parameter decisions influence differentially private scatterplots and can build on this knowledge to further explore the intersection of data privacy and visualization.

Acknowledgments

This work is supported in part by NIEHS Superfund Research Program award P42ES017198 (PROTECT). We also greatly appreciate Jane Kokernak and Laura South's editing assistance.

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] F. J. Anscombe. Graphs in statistical analysis. *The american statistician*, 27(1):17–21, 1973.

[3] C. Arnold and M. Neunhoeffer. Really useful synthetic data–a framework to evaluate the quality of differentially private synthetic data. *arXiv preprint arXiv:2004.07740*, 2020. doi: 10.48550/arXiv.2004.07740

[4] K. Bhattacharjee, M. Chen, and A. Dasgupta. Privacy-preserving data visualization: reflections on the state of the art and research opportunities. *Computer Graphics Forum*, 39(3):675–692, 2020. doi: 10.1111/cgf.14032

[5] M. Borkin, K. Gajos, A. Peters, D. Mitsouras, S. Melchionna, F. Rybicki, C. Feldman, and H. Pfister. Evaluation of artery visualizations for heart disease diagnosis. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):2479–2488, 2011. doi: 10.1109/TVCG.2011.192

[6] W. J. Conover and R. L. Iman. On multiple-comparisons procedures. *Los Alamos Sci. Lab. Tech. Rep. LA-7677-MS*, 1:14, 1979. doi: 10.1161/CIRCULATIONAHA.107.700971

[7] A. Dasgupta, M. Chen, and R. Kosara. Measuring privacy and utility in privacy-preserving visualization. *Computer Graphics Forum*, 32(8):35–47, 2013. doi: 10.1111/cgf.12142

[8] A. Dasgupta, R. Kosara, and M. Chen. Guess me if you can: A visual uncertainty model for transparent evaluation of disclosure risks in privacy-preserving data visualization. In *2019 IEEE Symposium on Visualization for Cyber Security (VizSec)*, pp. 1–10. IEEE, 2019. doi: 10.1109/VizSec48167.2019.9161608

[9] P. S. Dhotre, A. Bihani, S. Khajuria, and H. Olesen. Take it or leave it: Effective visualization of privacy policies. *Cybersecurity and Privacy: Bridging the Gap*, pp. 39–64, 2017.

[10] B. Dobrota. Measuring the quantity of data privacy and utility tradeoff for users' data: A visualization approach. Master's thesis, Utrecht University, 2021.

[11] DPComp-Org. DPComp-org/DPComp_Core, 2022.

[12] P. Dragicevic. *HCI Statistics without p-values*. PhD thesis, Inria, 2015.

[13] C. Dwork. Differential privacy: A survey of results. In *International conference on theory and applications of models of computation*, pp. 1–19. Springer, 2008. doi: 10.1007/978-3-540-79228-4_1

[14] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. Our data, ourselves: Privacy via distributed noise generation. In *Annual international conference on the theory and applications of cryptographic techniques*, pp. 486–503. Springer, 2006. doi: 10.1007/11761679_29

[15] 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

[16] C. Dwork, G. N. Rothblum, and S. Vadhan. Boosting and differential privacy. In *2010 IEEE 51st Annual Symposium on Foundations of Computer Science*, pp. 51–60. IEEE, 2010. doi: 10.1109/FOCS.2010.12

[17] C. Dwork, A. Smith, T. Steinke, and J. Ullman. Exposed! a survey of attacks on private data. *Annu. Rev. Stat. Appl*, 4(1):61–84, 2017. doi: 10.1146/annurev-statistics-060116-054123

[18] N. Elmqvist and J. S. Yi. Patterns for visualization evaluation. *Information Visualization*, 14(3):250–269, 2015. doi: 10.1145/2442576.2442588

[19] Ú. Erlingsson, V. Pihur, and A. Korolova. Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proc. 2014 ACM SIGSAC conference on computer and communications security*, pp. 1054–1067, 2014. doi: 10.1145/2660267.2660348

[20] D. Freedman and P. Diaconis. On the histogram as a density estimator: L_2 theory. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57(4):453–476, 1981. doi: 10.1007/BF01025868

[21] M. Friedman. The use of ranks to avoid the assumption of normality implicit in the analysis of variance. *Journal of the american statistical association*, 32(200):675–701, 1937.

[22] M. Friendly and D. Denis. The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences*, 41(2):103–130, 2005. doi: 10.1002/jhbs.20078

[23] L. Frigerio, A. S. d. Oliveira, L. Gomez, and P. Duverger. Differentially private generative adversarial networks for time series,

continuous, and discrete open data. In *IFIP International Conference on ICT Systems Security and Privacy Protection*, pp. 151–164. Springer, 2019. doi: 10.1007/978-3-030-22312-0_11

[24] M. Gaboardi, J. Honaker, G. King, J. Murtagh, K. Nissim, J. Ullman, and S. Vadhan. Psi ({\Psi}): a private data sharing interface. *arXiv preprint arXiv:1609.04340*, 2016. doi: 10.48550/arXiv.1609.04340

[25] G. M. Garrido, J. Near, A. Muhammad, W. He, R. Matzutt, and F. Matthes. Do i get the privacy i need? benchmarking utility in differential privacy libraries. *arXiv preprint arXiv:2109.10789*, 2021. doi: 10.48550/arXiv.2109.10789

[26] A. Ghosh, T. Roughgarden, and M. Sundararajan. Universally utility-maximizing privacy mechanisms. *SIAM Journal on Computing*, 41(6):1673–1693, 2012. doi: 10.1145/1536414.1536464

[27] K. A. Hallgren. Computing inter-rater reliability for observational data: an overview and tutorial. *Tutorials in quantitative methods for psychology*, 8(1):23, 2012. doi: 10.20982/tqmp.08.1.p023

[28] C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Cournapeau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. F. del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, Sept. 2020. doi: 10.1038/s41586-020-2649-2

[29] M. Hay, A. Machanavajjhala, G. Miklau, Y. Chen, and D. Zhang. Principled evaluation of differentially private algorithms using DPBench. *Proc. 2016 International Conference on Management of Data*, 2016. doi: 10.1145/2882903.2882931

[30] 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

[31] 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

[32] 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

[33] J. D. Hunter. Matplotlib: A 2D graphics environment. *Computing in science & engineering*, 9(03):90–95, 2007. doi: 10.1109/MCSE.2007.55

[34] 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

[35] H. Khamis. Measures of association: How to choose? *Journal of Diagnostic Medical Sonography*, 24(3):155–162, 2008. doi: 10.1177/8756479308317006

[36] T. K. Koo and M. Y. Li. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2):155–163, 2016. doi: 10.1016/j.jcm.2016.02.012

[37] I. Kotsogiannis, A. Machanavajjhala, M. Hay, and G. Miklau. Pythia: Data dependent differentially private algorithm selection. In *Proc. 2017 ACM International Conference on Management of Data*, pp. 1323–1337, 2017. doi: 10.1145/3035918.3035945

[38] S.-Y. Kung. Discriminant component analysis for privacy protection and visualization of big data. *Multimedia Tools and Applications*, 76(3):3999–4034, 2017. doi: 10.1007/s11042-015-2959-9

[39] H. B. Lee. Visualization and differential privacy. Master's thesis, University of Illinois at Urbana-Champaign, 2017.

[40] C. Li, M. Hay, G. Miklau, and Y. Wang. A data-and workload-aware algorithm for range queries under differential privacy. *arXiv preprint arXiv:1410.0265*, 2014. doi: 10.48550/arXiv.1410.0265

[41] J. Matejka and G. Fitzmaurice. Same stats, different graphs: Generating datasets with varied appearance and identical statistics through simulated annealing. In *Proc. 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 1290–1294, 2017. doi: 10.1145/3025453.3025912

[42] J. Matute, A. C. Telea, and L. Linsen. Skeleton-based scagnostics. *IEEE transactions on visualization and computer graphics*, 24(1):542–552, 2017. doi: 10.1109/TVCG.2017.2744339

[43] L. Micallef, G. Palmas, A. Oulasvirta, and T. Weinkauf. Towards perceptual optimization of the visual design of scatterplots. *IEEE transactions on visualization and computer graphics*, 23(6):1588–1599, 2017. doi: 10.1109/TVCG.2017.2674978

[44] P. Nanayakkara, J. Bater, X. He, J. Hullman, and J. Rogers. Visualizing privacy-utility trade-offs in differentially private data releases. *arXiv preprint arXiv:2201.05964*, 2022. doi: 10.48550/arXiv.2201.05964

[45] N. Nida, M. Sharif, M. U. G. Khan, M. Yasmin, and S. L. Fernandes. A framework for automatic colorization of medical imaging. *IIOAB J*, 7:202–209, 2016.

[46] L. Panavas, T. Crnovrsanin, J. L. Adams, A. Sarvghad, M. Tory, and C. Dunne. Visual utility evaluation of differentially private scatterplots. OSF Preprints, 2022.

[47] 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 *Proc. 2016 CHI Conference on Human Factors in Computing Systems*, pp. 3659–3669, 2016. doi: 10.1145/2858036.2858155

[48] W. Qardaji, W. Yang, and N. Li. Differentially private grids for geospatial data. In *2013 IEEE 29th international conference on data engineering (ICDE)*, pp. 757–768. IEEE, 2013. doi: 10.1109/ICDE.2013.6544872

[49] Sdv-Dev. SDV-Dev/SDGym: Benchmarking synthetic data generation methods., 2022.

[50] M. Sedlmair, T. Munzner, and M. Tory. Empirical guidance on scatterplot and dimension reduction technique choices. *IEEE transactions on visualization and computer graphics*, 19(12):2634–2643, 2013. doi: 10.1109/TVCG.2013.153

[51] M. Sedlmair, A. Tatu, T. Munzner, and M. Tory. A taxonomy of visual cluster separation factors. *Computer Graphics Forum*, 31(3pt4):1335–1344, 2012. doi: 10.1111/j.1467-8659.2012.03125.x

[52] J. Soria-Comas, J. Domingo-Ferrer, D. Sánchez, and D. Megías. Individual differential privacy: A utility-preserving formulation of differential privacy guarantees. *IEEE Transactions on Information Forensics and Security*, 12(6):1418–1429, 2017. doi: 10.1109/TIFS.2017.2663337

[53] L. South, D. Saffo, O. Vitek, C. Dunne, and M. A. Borkin. Effective use of Likert scales in visualization evaluations: A systematic review. *Computer Graphics Forum*, 41(3):43–55, 2022. doi: 10.1111/cgf.14521

[54] H. A. Sturges. The choice of a class interval. *Journal of the American Statistical Association*, 21(153):65–66, 1926.

[55] Y. Tao, R. McKenna, M. Hay, A. Machanavajjhala, and G. Miklau. Benchmarking differentially private synthetic data generation algorithms. *arXiv preprint arXiv:2112.09238*, 2021. doi: 10.48550/arXiv.2112.09238

[56] 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

[57] A. Torfi, E. A. Fox, and C. K. Reddy. Differentially private synthetic medical data generation using convolutional GANs. *Information Sciences*, 586:485–500, 2022. doi: 10.1016/j.ins.2021.12.018

[58] P. K. Trivedi and D. M. Zimmer. *Copula modeling: an introduction for practitioners*. Now Publishers Inc, 2007. doi: 10.1561/0800000005

[59] 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, 2018. doi: 10.1109/TVCG.2018.2865021

[60] 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

[61] Z. Wang, E. P. Simoncelli, and A. C. Bovik. Multiscale structural similarity for image quality assessment. In *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2, pp. 1398–1402. Ieee, 2003. doi: 10.1109/ACSSC.2003.1292216

[62] L. Wilkinson, A. Anand, and R. Grossman. Graph-theoretic scagnostics. In *Information Visualization, IEEE Symposium on*, pp. 21–21. IEEE Computer Society, 2005. doi: 10.1109/INFOVIS.2005.14

[63] D. Winograd-Cort, A. Haeberlen, A. Roth, and B. C. Pierce. A framework for adaptive differential privacy. *Proc. ACM on Programming Languages*, 1(ICFP):1–29, 2017. doi: 10.1145/3110254

[64] 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.

[65] Y. Xiao, L. Xiong, and C. Yuan. Differentially private data release through multidimensional partitioning. In *Workshop on Secure Data Management*, pp. 150–168. Springer, 2010. doi: 10.1007/978-3-642-15546-8_11

[66] J. Xu, Z. Zhang, X. Xiao, Y. Yang, G. Yu, and M. Winslett. Differentially private histogram publication. *The VLDB journal*, 22(6):797–822, 2013. doi: 10.1007/s00778-013-0309-y

[67] 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.

[68] D. Zhang, R. McKenna, I. Kotsogiannis, M. Hay, A. Machanavajjhala, and G. Miklau. Ektelo: A framework for defining differentially-private computations. In *Proc. 2018 International Conference on Management of Data*, pp. 115–130, 2018. doi: 10.1145/3183713.3196921

[69] 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

[70] S. Zhang, A. Hagermalm, and S. Slavnic. An evaluation of open-source tools for the provision of differential privacy. *arXiv preprint arXiv:2202.09587*, 2022. doi: 10.48550/arXiv.2202.09587

[71] X. Zhang, R. Chen, J. Xu, X. Meng, and Y. Xie. Towards accurate histogram publication under differential privacy. In *Proc. 2014 SIAM international conference on data mining*, pp. 587–595. SIAM, 2014. doi: 10.1137/1.9781611973440.68

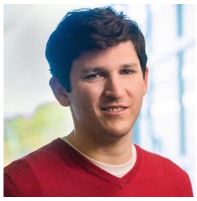
[72] J. Zhou, X. Wang, J. K. Wong, H. Wang, Z. Wang, X. Yang, X. Yan, H. Feng, H. Qu, H. Ying, et al. Dpviscreator: Incorporating pattern constraints to privacy-preserving visualizations via differential privacy. *IEEE Transactions on Visualization and Computer Graphics*, 2022. doi: 10.1109/TVCG.2022.3209391

[73] T. Zhu, G. Li, W. Zhou, and S. Y. Philip. Differentially private data publishing and analysis: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 29(8):1619–1638, 2017. doi: 10.1109/TKDE.2017.2697856

Liudas Panavas Liudas Panavas is a PhD candidate at Northeastern University. He studies visual analytics for private data and explainable artificial intelligence.

Tarik Crnovrsanin Tarik Crnovrsanin received his PhD degree from the University of Davis, California. He is a postdoctoral researcher in Visualization at Khoury College, Northeastern University. His current research interests are visualization in networks, machine learning, and explainable AI.

Jane Adams Jane Adams is a PhD student at Northeastern University in the Data Visualization Lab within Khoury College of Computer Sciences. Her research interest is information visualization for exploratory analysis of high-dimensional data, with applications in explainable A.I. and health sciences.



Jonathan Ullman Jonathan Ullman is an associate professor at Northeastern University. His research centers on the foundations of privacy for machine learning and statistics, in particular differential privacy and its surprising interplay with other topics such as statistical validity, robustness, cryptography, and fairness.

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.

Melanie Tory Melanie Tory is Director of Data Visualization Research at the Roux Institute. She earned her PhD in Computer Science from Simon Fraser University and her BSc from the University of British Columbia. She is Associate Editor of IEEE Computer Graphics and Applications, IEEE Transactions on Visualization & Computer Graphics, and Computer Graphics Forum.

Cody Dunne Cody Dunne is an assistant professor at Northeastern University whose research focuses on data visualization. He aims to help people explore and understand complex data—in particular data that includes aspects of network topology and change over time. Prof. Dunne earned his PhD at the University of Maryland.