Can Graphical Tools Help Analysts Implement Differential Privacy?

Onyinye Dibia
Onyinye.Dibia@uvm.edu
University of Vermont

Mengyi Lu
Mengyi.Lu@uvm.edu
University of Vermont

Prianka Bhattacharjee
Prianka.Bhattacharjee@uvm.edu
University of Vermont

Chuck McCallum
cmccallum@g.harvard.edu
Harvard

Joseph P. Near
jnear@uvm.edu
University of Vermont

Yuanyuan Feng
Yuanyuan.Feng@uvm.edu
University of Vermont

1 Introduction

Differential privacy (DP) has become a widely adopted framework for privacy-preserving data analysis. However, translating DP from theory into practice remains difficult. In practice, these challenges often surface through the *tools* that practitioners rely on. Many commonly used DP systems are delivered as application programming interfaces (APIs) or libraries. However, usability studies with data practitioners indicate that usability issues with API tools may be a bottleneck for broader DP adoption [14, 19].

One avenue for addressing these challenges is the use of *guided tools with graphical user interfaces (GUIs)*. Outside the DP context, GUIs and interactive systems have long been shown to support data analysis by reducing syntactic burden, externalizing reasoning, and helping users incrementally specify and refine analysis goals [10, 20]. Within DP, however, GUI tools have not been as well explored as API-based tools. Prior GUI tools include PSI (Ψ) [8, 13] and DP Creator [18]; more recently, the OpenDP project introduced DP Wizard [16]. All of these tools provide a graphical wizard-style interface to guide non-experts through the process of performing a DP data release. DP Wizard provides a novel addition compared to previous tools: instead of outputting DP data directly, it produces a Python notebook implemented using the OpenDP API, such that executing the notebook produces the desired DP data. This structure is designed to add flexibility to the tool (by allowing the user to modify the notebook after it is generated) and to aid users in learning to use the OpenDP API (by providing a starting point for API use tailored to the user's task).

We conduct a usability study of DP Wizard with 18 subjects to assess whether GUI tools can reduce the barriers to adopting DP and improve the ability of non-experts to correctly perform DP data releases. Our study combines a task-completion component with a semi-structured interview, and seeks to answer three research questions. **RQ1**: *To what extent does a guided graphical tool support users to configure and run DP analyses?* **RQ2**: *Does using a guided graphical tool lower the barrier to using an API-based tool for DP implementation?* **RQ3**: *What are the usability benefits and drawbacks associated with a guided graphical tool for DP, compared to API-based DP tooling?*

Our results, presented in Section 4, suggest that GUI-based tools can help analysts correctly perform DP data releases, but that challenges remain (especially in setting parameters); that GUI-based tools can make analysts more successful in transitioning to API-based tools; and that GUI-based tools address some—but not all—of the usability barriers associated with API-based tools.

2 Related Work

This work builds on prior research examining challenges in understanding differential privacy (DP), the usability of DP tools, and the role of graphical or guided systems in supporting complex data analysis workflows. A substantial body of research has shown that DP is difficult to understand, even for users with technical or data science backgrounds. Prior studies have documented widespread confusion around core DP concepts [4, 5]; for end users, explanations fail to increase understanding [4, 12, 21]. Collectively, this work highlights a persistent gap between formal DP guarantees and how users reason about privacy in applied contexts. Most tools for DP are API-based and are designed to support practitioners in implementing DP data releases. Usability studies have shown, however, that these tools remain difficult to use correctly in practice [5, 14, 19]. LLM-assisted workflows help, but still do not produce consistently correct results [2]. Outside of DP, GUI-based tools can improve usability [9, 11]. Within DP, only a few tools have explored this approach [13, 16, 18]. The goal of this work is to gain a better understanding of the usability benefits of GUI-based tools for DP data release.

3 Methods and Study Design

We used usability testing [6, 15] as the primary method to study how data practitioners work with the OpenDP DP Wizard. Following prior work on DP tool usability [3, 14], our study combined task-based evaluation with surveys, semi-structured interviews, and a think-aloud protocol.

DP Wizard. DP Wizard [16] is a GUI-based tool that guides users through configuring and running DP analyses. As output, it generates executable notebook-based implementations of the desired analysis implemented using the OpenDP library. DP Wizard is designed to lower DP implementation barriers for users with less technical expertise, and also to support users in using programmable DP workflows using OpenDP API library.

3.1 Study Procedure

3.1.1 Recruitment and Screening. This study was approved by our university's Institutional Review Board (IRB). We conducted a pilot study with two graduate students to refine task instructions and survey/interview materials.

We recruited 18 participants (6 DP experts, 12 non-experts), aiming for a sample size consistent with prior DP usability studies and a higher proportion of users with limited DP experience. Non-experts were primarily computer science and data science graduate students, many early-stage PhD students, along with two data practitioners with minimal DP experience.

We distributed a recruitment advertisement with a link to an eligibility survey via the OpenDP mailing list and Slack channel, the Friends of the CMU Usable Privacy and Security Laboratory mailing list, our university's Computer Science mailing list, and professional networks. The survey (Appendix C) obtained informed consent and assessed technical background through self-reported experience and knowledge checks (three Python questions, four DP questions). Participants were eligible if they reported sufficient data science experience and answered at least one Python question correctly. We classified participants who answered at least three of four DP questions correctly as DP experts; all others were classified as non-experts.

Of the 47 respondents who began the survey, 32 completed it and 29 met eligibility criteria; all eligible respondents were invited. Participants were randomly assigned to one of two conditions (private dataset only, or private + synthetic dataset). Sessions were conducted remotely via Microsoft Teams using a standardized protocol and lasted approximately one hour. Participants received a $40 electronic gift card. All participants were at least 18 years old; 33.3% identified as female and 66.7% as male.

3.1.2 Pre-task Procedures. Before starting the study tasks, we verified that participants were able to share their screens and were familiar with the think-aloud protocol. To provide participants with the necessary background for the study tasks, we required participants to complete a tutorial for their assigned study condition, which included executable sample DP tasks using the DP Wizard (see Appendix D), and we provided a brief handout introducing the fundamentals of differential privacy. During the study, we allowed participants to refer back to the tutorial and handout, as well as consult the DP Wizard and OpenDP documentation as needed.

3.1.3 Usability Testing Tasks. We designed four usability testing tasks focused on differentially private data analysis. The tasks used two datasets: a private dataset and a synthetic dataset representing restaurant visits across a week (see Appendix A). Participants completed the tasks within a one-hour study session. Tasks 1 and 2 were performed using the DP Wizard's graphical interface, while Tasks 3 and 4 required participants to work with the Jupyter notebook generated by the Wizard. We assigned a total privacy budget of $\varepsilon = 1$ for all tasks. Participants could configure all other privacy-related parameters themselves. Throughout the session, we instructed participants to follow a think-aloud protocol, and we recorded both their verbalizations and on-screen interactions. Participants performed the following data releases and tasks:

(1) Calculate the average time spent by visitors per day.
(2) Calculate the median time spent by visitors per day.
(3) Change Epsilon ε from 1 to 0.5 and re-run the analysis.
(4) Calculate the total time (sum) spent by visitors per day.

3.1.4 Post-task Procedures. After completing the study tasks, participants first took part in a semi-structured post-task interview and then completed a post-task survey (see Appendices E and F). The interview focused on participants' experiences with the DP Wizard interface and the generated Jupyter notebook, including aspects they found intuitive or difficult to understand or modify, as well as suggestions for improving the tool. The post-task survey assessed participants' understanding of differential privacy after the study by repeating selected questions from the eligibility survey and collected additional feedback about their overall study experience.

3.2 Usability Measurements and Data Analysis

RQ1 (DP Implementation Support). We measured task performance for Wizard-based tasks by recording completion status (complete, partial, failed) and correctness relative to reference solutions. We also coded interface interactions and points of confusion observed in screen recordings and think-aloud data.

RQ2 (From GUI to Code). We assessed participants' ability to work with the generated Jupyter notebook by recording whether they could locate, understand, and modify relevant code to complete Tasks 3 and 4. We coded modification success, errors, and interruptions, and analyzed think-aloud and interview data to capture reasoning and confidence.

RQ3 (Benefits and Drawbacks). We analyzed qualitative data from think-aloud protocols, interviews, and surveys to identify perceived strengths, usability barriers, and areas for improvement.

Data Analysis The measurement items for RQ1–RQ3 were operationalized as structured codes in a shared codebook that grounded both quantitative summaries and qualitative analysis (see Appendix G). We developed the codebook after study completion by mapping research questions to observable behaviors and task outcomes, informed by an initial review of transcripts, screen recordings, and generated notebooks.

Organized around the three research questions, the codebook captured task performance, configuration choices, reasoning about DP parameters, interface interactions, and notebook use. Three researchers independently coded all artifacts and resolved disagreements through discussion. We derived quantitative summaries by aggregating coded observations and synthesized qualitative themes from associated excerpts. We additionally conducted a hybrid thematic analysis [1, 7, 17], combining deductive DP-informed codes with inductive identification of emergent usability themes. This process linked observed behaviors, quantitative outcomes, and participant reasoning to the results presented below.

4 Results

4.1 RQ1: DP Implementation Support

4.1.1 Task completion and correctness. Overall, DP Wizard enabled most participants to complete the assigned analyses but their solution correctness and conceptual understanding varied. All participants (18/18) were able to configure and complete the required analyses at least once using the tool. See Appendix B for full results.

Task-Level Time Distribution Participants showed a clear pattern in time allocation. Task 1 took the longest on average (7.74 minutes), reflecting the effort required to configure the analysis from scratch and understand the interface. Task 2 was much shorter (2.21 minutes), suggesting that once participants formed a mental model of the workflow, switching from mean to median required minimal additional effort. Tasks 3 and 4 (about 6.3 minutes each) required modifying the generated notebook, which introduced additional complexity. Overall, about one-third of time was spent in the DP Wizard GUI and two-thirds within the notebook, indicating that code-level interaction demanded more effort.

Per-Participant Task Time Distribution All 18 participants completed the tasks, but total time varied widely (12–42 minutes), reflecting differences in cognitive effort. Task 1 generally functioned as a learning phase, while Task 2 was consistently brief. Some participants progressed smoothly, whereas others slowed substantially during notebook modification, especially in Task 4. Tutorial time also varied considerably but showed no clear relationship with overall performance. Overall, the GUI supported efficient initial setup, while notebook-based edits introduced greater variability in time and effort.

4.1.2 Configuration Correctness. We evaluated five behaviors to assess correct DP configuration. Figure 1 shows how well participants achieved these behaviors in the study. All 18 participants correctly selected columns, aggregation functions, and grouping operations, demonstrating that the graphical interface effectively supports structural configuration. Seventeen participants exported and reviewed results, indicating strong engagement with outputs rather than blind execution.

In contrast, only 5 participants correctly configured contribution bounds, making this the weakest area. Although some participants conceptually considered contribution limits, most failed to translate that reasoning into correct parameter values. Fourteen participants correctly understood the distinction between public and private datasets, suggesting this concept was more intuitive. Only 5 participants inspected the private dataset directly, potentially violating DP assumptions. Overall, structural tasks were well supported, but deeper privacy parameterization (especially contribution bounds) remained challenging.

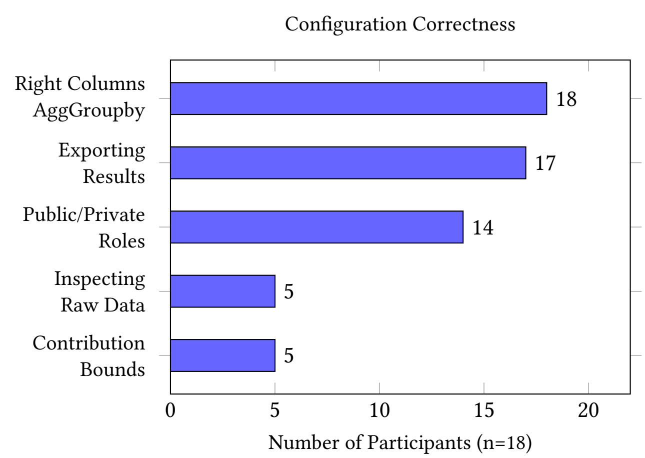

Figure 1: Participant Success Rates Across Configuration Dimensions.

4.1.3 Conceptual Understanding. We assessed seven indicators of conceptual understanding; overall results appear in Figure 2. Twelve participants understood what results should generally look like and reflected on whether outputs were reasonable, suggesting that expectation formation and result evaluation were closely linked. However, only ten participants successfully fixed issues when results appeared incorrect. Engagement with privacy parameters varied. All 18 participants considered epsilon, making it the most salient concept. Twelve reflected on clipping bounds, and nine considered contribution bounds. In contrast, only two participants thought about multiple releases, indicating that cumulative privacy loss was largely overlooked. Overall, participants demonstrated solid procedural competence and basic output evaluation, but reasoning about more advanced privacy concepts was uneven.

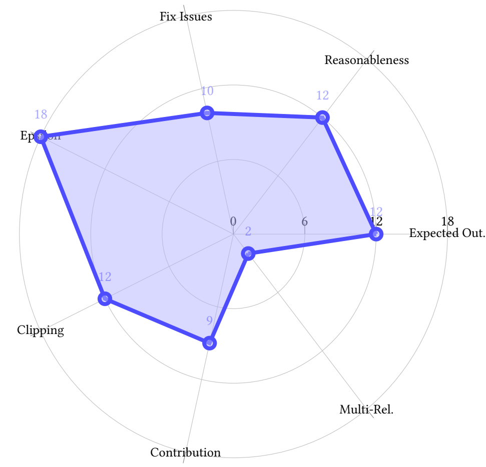

Figure 2: Participant conceptual understanding across seven dimensions (N=18).

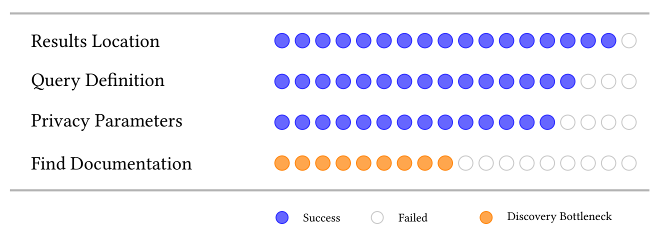

Figure 3: Iconic representation of notebook navigation success ($N = 18$). The horizontal borders emphasize the functional grouping of navigational tasks.

4.2 RQ2: From GUI to Code

4.2.1 Notebook Modification (Behavioral Data). Figure 3 shows participants' ability to navigate the Python notebook generated by DP Wizard, and Figure 4 shows participants' ability to modify the notebook as required by Tasks 3 and 4. All 18 participants were able to navigate the notebook and make basic code changes, indicating that the notebook was accessible and that DP Wizard reduced syntactic barriers to interacting with OpenDP. Fourteen participants also reported confidence in modifying the code.

However, only 4 participants were able to verify whether their modified analyses were conceptually correct. This gap suggests that while structural manipulation was manageable, reasoning about DP parameters (especially bounding and sensitivity) remained difficult.

17 participants reused the generated notebook as a template for later tasks, indicating that it functioned as a supportive scaffold where users could make localized edits rather than start from scratch.

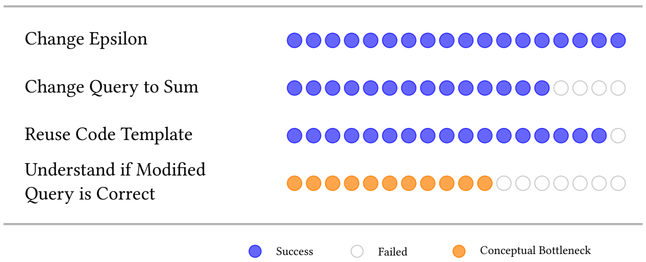

Figure 4: User performance in code modification ($N = 18$).

4.2.2 Subjective Experience with the Generated Notebook. Think-aloud and interview data show that most participants viewed the notebook positively. Fifteen described it as helpful, three partially agreed, and only one found it unhelpful. Participants did not treat it as a black-box output; instead, they experienced it as an explanatory artifact. Its structure and comments appeared to ease the transition from GUI configuration to code, reducing intimidation and supporting confidence in interacting with the generated analysis.

4.3 RQ3: Usability Benefits and Drawbacks

4.3.1 API Barriers. Figure 5 summarizes reported usability challenges across notebook comprehension, documentation access, and understanding API behavior. Eleven of 18 participants struggled to understand parts of the notebook, indicating that while it was navigable, DP-related code was not always clear. Only four participants reported difficulty finding documentation, suggesting access was not the main issue. The most significant challenge was understanding how the API works, particularly privacy budgeting. Thirteen participants reported difficulty reasoning about privacy parameters and budget tracking; conceptual understanding of API behavior and DP mechanisms remains a barrier.

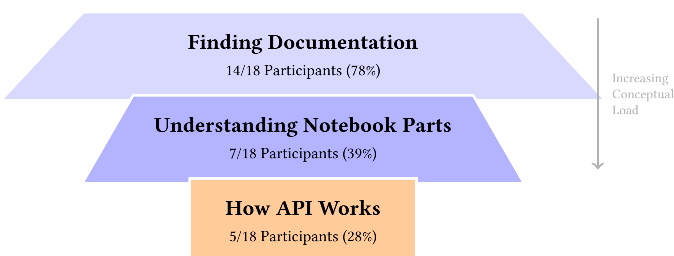

Figure 5: API Barrier Funnel: Participant success rates across different cognitive stages.

4.3.2 Design Implications. Figure 6 contrasts components perceived as useful versus missing. Tutorials (15/18) and UI elements (14/18) were seen as strong features, along with notebook comments (13/18) and structure (12/18). These findings suggest the tool effectively supports procedural workflows.

However, deeper instructional scaffolding was perceived as lacking. Many participants reported missing or insufficient parameter

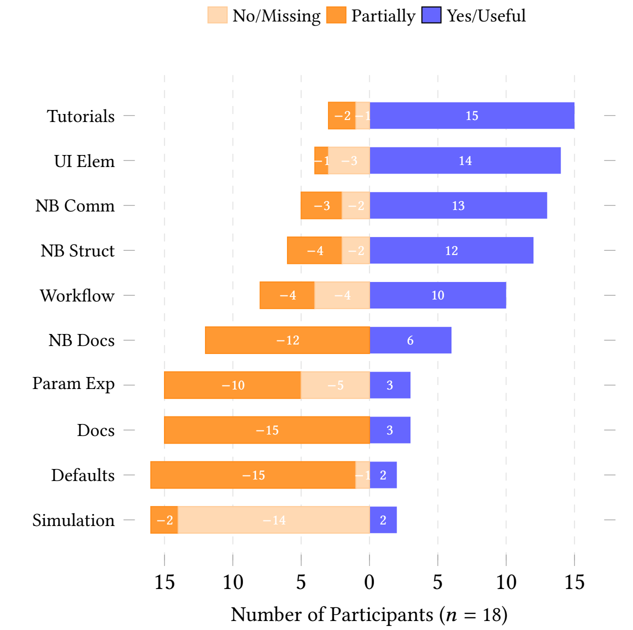

Figure 6: Participant feedback on tool components ($n = 18$).

explanations (10/18), default settings (15/18), documentation (15/18), and notebook documentation (12/18). This indicates that while the interface reduces mechanical friction, conceptual guidance—especially around parameter meaning and implications—remains underdeveloped.

5 Discussion & Recommendations

RQ1: DP Implementation Support. Our results suggest that graphical tools like DP Wizard *do* reduce barriers associated with API tools to correctly implementing DP data releases. All participants in our study correctly completed the study tasks, in contrast to findings of previous studies on API tools [14, 19]. However, graphical tools *do not* eliminate conceptual challenges associated with setting parameters, which are common to both graphical and API tools; additional work is needed to provide support to users in setting parameters.

RQ2: From GUI to Code. Our results suggest that graphical tools like DP Wizard *do* help users transition into API tools via generated notebooks. The notebooks provide a starting point for users' modifications, and serve as a form of customized documentation for the API. This result is encouraging, given the usability challenges of API tools reported in prior work [14, 19].

RQ3: Usability Benefits and Drawbacks. Our results suggest several remaining barriers: users struggled to understand how to set parameters and did not understand the defaults (highlighting the need from RQ1 for work to support users in setting parameters), and API understanding remained a challenge despite support from the generated notebook.

References

[1] Virginia Braun and Victoria Clarke. Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2):77–101, 2006.

[2] Shao-Yu Chu, Yuhe Tian, Yu-Xiang Wang, and Haojian Jin. Dpcheatsheet: Using worked and erroneous llm-usage examples to scaffold differential privacy implementation. *arXiv preprint arXiv:2509.12590*, 2025.

[3] Lynne Cooke. Assessing concurrent think-aloud protocol as a usability test method: A technical communication approach. *IEEE Transactions on Professional Communication*, 53(3):202–215, 2010.

[4] Rachel Cummings, Gabriel Kaptchuk, and Elissa M Redmiles. " i need a better description": An investigation into user expectations for differential privacy. In *Proceedings of the 2021 ACM SIGSAC Conference on Computer and Communications Security*, pages 3037–3052, 2021.

[5] Onyinye Dibia, Prianka Bhattacharjee, Brad Stenger, Steven Baldasty, Mako Bates, Ivoline C Ngong, Yuanyuan Feng, and Joseph P Near. Sok: Usability studies in differential privacy. *arXiv preprint arXiv:2412.16825*, 2024.

[6] Joseph F Dumas and Janice C Redish. *A practical guide to usability testing*. Greenwood Publishing Group Inc., 1993.

[7] Jennifer Fereday and Eimear Muir-Cochrane. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International journal of qualitative methods*, 5(1):80–92, 2006.

[8] Marco Gaboardi, James Honaker, Gary King, Jack Murtagh, Kobbi Nissim, Jonathan Ullman, and Salil Vadhan. Psi ({\Psi}): a private data sharing interface. *arXiv preprint arXiv:1609.04340*, 2016.

[9] Jeffrey Heer, Michael Bostock, and Vadim Ogievetsky. A tour through the visualization zoo. *Communications of the ACM*, 53(6):59–67, 2010.

[10] Jessica Hullman and Andrew Gelman. Designing for interactive exploratory data analysis requires theories of graphical inference. *Harvard Data Science Review*, 3(3):10–1162, 2021.

[11] Daniel Keim, Gennady Andrienko, Jean-Daniel Fekete, Carsten Görg, Jörn Kohlhammer, and Guy Melançon. Visual analytics: Definition, process, and challenges. In *Information visualization: Human-centered issues and perspectives*, pages 154–175. Springer, 2008.

[12] Patrick Kühtreiber, Viktoriya Pak, and Delphine Reinhardt. Replication: the effect of differential privacy communication on german users' comprehension and data sharing attitudes. In *Eighteenth Symposium on Usable Privacy and Security (SOUPS 2022)*, pages 117–134, 2022.

[13] Jack Murtagh, Kathryn Taylor, George Kellaris, and Salil Vadhan. Usable differential privacy: A case study with psi. *arXiv preprint arXiv:1809.04103*, 2018.

[14] Ivoline C Ngong, Brad Stenger, Joseph P Near, and Yuanyuan Feng. Evaluating the usability of differential privacy tools with data practitioners. In *Twentieth Symposium on Usable Privacy and Security (SOUPS 2024)*, pages 21–40, 2024.

[15] Jakob Nielsen. *Usability engineering*. Morgan Kaufmann, 1994.

[16] OpenDP Team. DP Wizard. https://github.com/opendp/dp-wizard, August 2025.

[17] Kevin Proudfoot. Inductive/deductive hybrid thematic analysis in mixed methods research. *Journal of mixed methods research*, 17(3):308–326, 2023.

[18] Jayshree Sarathy, Sophia Song, Audrey Haque, Tania Schlatter, and Salil Vadhan. Don't look at the data! how differential privacy reconfigures the practices of data science. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2023.

[19] Patrick Song, Jayshree Sarathy, Michael Shoemate, and Salil Vadhan. " i inherently just trust that it works": Investigating mental models of open-source libraries for differential privacy. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW2):1–39, 2024.

[20] April Yi Wang, Anant Mittal, Christopher Brooks, and Steve Oney. How data scientists use computational notebooks for real-time collaboration. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–30, 2019.

[21] Aiping Xiong, Tianhao Wang, Ninghui Li, and Somesh Jha. Towards effective differential privacy communication for users' data sharing decision and comprehension. In *2020 IEEE Symposium on Security and Privacy (SP)*, pages 392–410. IEEE, 2020.

Appendix A: Dataset Details

We used two datasets in this study: a private dataset and a synthetic public dataset. Both datasets were provided to participants as CSV files and share the same schema.

A.1 Schema

Each dataset contains the following columns:

- **VisitorId**: Unique identifier for a visitor.
- **Time spent (minutes)**: Duration of the visit in minutes.
- **Money spent (euros)**: Amount spent during the visit.
- **Day**: Integer representing the day of the week (1–7).

Each row represents a single restaurant visit. Visitors may appear multiple times per day and across multiple days.

A.2 Private Dataset

The private dataset (`week_data.csv`) was used for all Wizard-based tasks. It contains records of restaurant visits over seven days. The dataset includes repeated visits by the same visitor, enabling evaluation of contribution bounding and aggregation behavior under differential privacy.

An example row is shown below:

```
VisitorId, Time spent (minutes), Money spent (euros), Day
580, 29, 17, 1
1215, 45, 18, 1
```

A.3 Synthetic Dataset

The synthetic dataset (`week_data_synthetic.csv`) mirrors the schema of the private dataset but contains artificially generated values. It was provided in one study condition to allow participants to compare behavior across private and synthetic data.

Example rows include:

```
VisitorId, Time spent (minutes), Money spent (euros), Day
970, 38, 22, 4
534, 38, 16.4, 6
```

A.4 Purpose

The datasets were designed to:

- Require grouping by day
- Require aggregation (mean, median, sum)
- Include repeated visitor contributions
- Enable evaluation of privacy parameters and contribution limits

Both datasets are available in the supplementary materials.

Appendix B: Tasks and Solutions

Participants completed a sequence of tasks across two environments: the DP Wizard (GUI-based interface) and a Jupyter Notebook using the OpenDP library.

The tasks were framed around a dataset representing restaurant visits over a 7-day period. Each record corresponded to a single visit and included a visitor ID. Visitors could appear multiple times per day and across multiple days. The dataset described in Appendix 5 was uploaded into the tool.

B.1 DP Wizard Tasks

Participants first completed the following tasks using the DP Wizard interface.

(1) **Task 1 (Mean):** Compute the differentially private average time spent in the restaurant by visitors per day.
(2) **Task 2 (Median):** Compute the differentially private median time spent by visitors per day.

These tasks required participants to:

- Select the appropriate aggregation function (mean or median)
- Configure grouping by day
- Set or confirm the privacy parameter ϵ
- Specify necessary bounds and contribution limits

B.2 Notebook-Based Tasks

After completing the Wizard tasks, participants transitioned to a Jupyter Notebook environment preconfigured with OpenDP.

(3) **Task 3 (Epsilon Adjustment):**
  - Change ϵ from 1.0 to 0.5.
  - Re-run the analysis.
  - In a Markdown cell, briefly describe how the output changed.

(4) **Task 4 (Sum):** Compute the differentially private total time (sum) spent by visitors per day.

Participants were informed that they could increase ϵ if additional accuracy was needed.

B.3 Reference Solutions

We developed reference solutions for each task to verify correctness. Participant solutions were evaluated based on:

- Correct selection of aggregation mechanism
- Proper use of grouping by day
- Appropriate privacy parameter configuration
- Successful execution without runtime errors

The reference implementations are available in the supplementary materials.

Appendix C: Eligibility Survey

For questions that test participants' understanding, we highlight the correct answer in **bold**. The eligibility survey was administered via Qualtrics prior to scheduling study sessions.

Consent and Basic Eligibility

Participants were first presented with an IRB-approved consent form. After reviewing the form, they were asked:

- I have read and understood the information above. (No / Yes)
- I want to proceed to complete the eligibility survey for this research study. (No / Yes)

Participants then answered the following screening questions:

- Are you at least 18 years old? (No / Yes)
- Do you currently reside in the United States? (No / Yes)
- Have you performed statistical data analysis in Python? (No / Yes)

- Have you used Jupyter Notebook before? (No / Yes)
- Are you willing to participate in a study that involves using a data science interface—either a graphical tool or a coding-based tool (Python in Jupyter Notebook)—to complete data analysis tasks? (No / Yes)
- Are you willing to participate in a 1.5-hour remote usability study via Microsoft Teams? (No / Yes)

Experience Questions

Participants reported their experience with Python and Jupyter Notebook:

(1) How many years have you been coding in Python?
  - 0–1
  - 2–3
  - 3+

(2) How many years have you been using Jupyter Notebook?
  - 0–1
  - 2–3
  - 3+

(3) Which of the following best describes how you use Python and Jupyter Notebook for statistical analysis?
  - They are my preferred language/tool
  - I am comfortable using them but prefer other languages/tools (e.g., R)
  - I can work with them but often need to resort to documentation
  - I rarely use them and need additional time to get familiar with them

Python Knowledge Questions

(4) Use “set” instead of “list” as a Python data structure for a sequence of elements when:
  - elements will be appended to increase the size of the sequence
  - the order of items is important
  - **it is important to know if the sequence contains a specific item**
  - it is important to know the item with maximum value in the sequence
  - I don’t know

(5) When you create a multi-index object with the “groupby” operator in pandas, how many levels of labels are required to retrieve a value?
  - 1
  - 2
  - 3
  - **One for every index**
  - I don’t know

(6) What is the output of the following code?

```
str1 = "DataScience is fun!"
print(str1[4:12])
```

  - **Science**
  - Data Sci
  - aScience
  - Error
  - I don’t know

Differential Privacy Knowledge Questions

(7) Have you heard of the term differential privacy (DP) before?
  - No
  - Yes

(8) Have you ever written code to implement differential privacy (DP) in any capacity?
  - No
  - Yes

(9) In differential privacy, which value of the privacy parameter ϵ provides stronger privacy?
  - $\boldsymbol{\epsilon = 0.1}$
  - $\epsilon = 1.0$
  - I don’t know

(10) Releasing two differentially private statistics, one with $\epsilon_1 = 0.1$ and the other with $\epsilon_2 = 0.5$, results in a total privacy loss of:
  - $\epsilon = 0.1$
  - $\epsilon = 0.5$
  - $\boldsymbol{\epsilon = 0.6}$
  - $\epsilon = 0.05$
  - I don’t know

(11) Which of the following is an advantage of using Differential Privacy?
  - It guarantees complete anonymity of the data subjects
  - It ensures that the data is completely accurate
  - **It provides a tradeoff between privacy and utility of the data**
  - It is a computationally simple method for preserving privacy in large datasets
  - I don’t know

(12) If the mechanism M returns a number and satisfies differential privacy with $\epsilon = 0.1$, does abs$(M(x))$ satisfy differential privacy, where abs is the absolute value function?
  - No, not necessarily
  - **Yes, for** $\epsilon = 0.1$
  - Yes, for some $\epsilon > 0.1$
  - I don’t know

Demographics

Participants were also asked to report:

- Age
- Gender
- Undergraduate or graduate student status

Contact Information

At the end of the survey, participants were asked to provide an email address for scheduling purposes.

Appendix D: The Handouts and Tutorial Materials

Participants were provided with structured instructional materials to support task completion. These materials consisted of (1) a study task handout, (2) a differential privacy (DP) concept handout, and (3) a DP Wizard tutorial guide.

D.1 Study Task Handout

The study task handout described:

- The study scenario (restaurant visits dataset across seven days)
- A description of dataset columns (VisitorID, time spent, money spent, day)
- The sequence of tasks to be completed in the DP Wizard and Jupyter Notebook
- Instructions for transitioning between the Wizard and the Notebook environments

The handout specified task goals but did not provide solutions or step-by-step instructions for configuring privacy parameters.

D.2 Differential Privacy (DP) Concept Handout

Participants were also provided with a brief DP handout introducing core concepts necessary for the study tasks. The handout included:

- A high-level explanation of differential privacy
- The role of the privacy parameter ϵ
- The privacy-accuracy tradeoff
- A brief explanation of contribution bounding
- An overview of aggregation mechanisms (mean, median, sum)

The DP handout was intended to provide conceptual grounding without teaching specific implementation details in OpenDP.

D.3 DP Wizard Tutorial

Participants received a tutorial describing how to use the DP Wizard interface, including:

- Uploading datasets
- Selecting aggregation functions
- Grouping by columns
- Setting privacy parameters
- Configuring bounds and contribution limits
- Exporting and running the generated Jupyter Notebook

The tutorial focused on tool navigation and workflow rather than differential privacy theory.

D.4 External Documentation

Participants were permitted to consult the official OpenDP documentation:

https://docs.opendp.org/en/stable/index.html

D.5 Availability of Materials

All handouts, tutorial materials, task descriptions, and reference solutions are available in the supplementary repository accompanying this paper.

Appendix E: Post Task Interview

Thank you for completing (or attempting to complete) the study tasks using the DP Wizard and the generated Jupyter notebook. We now ask a few questions to understand your experience and gather feedback about the tool and workflow.

(1) What part of the DP Wizard interface was easiest to understand?
(2) What part of the DP Wizard workflow was confusing or difficult to modify?
(3) What part of the generated Jupyter notebook was easiest to understand?
(4) In the notebook, what aspects were confusing or difficult to modify?
(5) What could have helped make the process easier when working with both the Wizard and the generated notebook?
(6) Do you have any suggestions for improving the DP Wizard?
(7) Do you have any additional comments or feedback about your overall experience?

Appendix F: Post Task Survey

Participants completed the following post-task survey after finishing the usability tasks.

Please rate the following statements about your experience with the DP Wizard using a 5-point Likert scale (Strongly Disagree, Somewhat Disagree, Neutral, Somewhat Agree, Strongly Agree):

(1) I think that I would like to use DP Wizard frequently.
(2) I found DP Wizard unnecessarily complex.
(3) I thought DP Wizard was easy to use.
(4) I found DP Wizard very cumbersome to use.
(5) I felt very confident using DP Wizard.
(6) I would imagine that most people would learn to use DP Wizard very quickly.
(7) I needed to learn a lot of things before I could get going with DP Wizard.
(8) I found the various features in DP Wizard well integrated.
(9) I thought there was too much inconsistency in DP Wizard.
(10) I found that DP Wizard introduced DP concepts appropriately for me to perform the tasks.
(11) I felt I had to learn too many technical details (e.g., classes, dependencies) to complete the tasks.
(12) Once I completed the tutorial, it was easier to perform the remaining tasks.
(13) I found the documentation provided in DP Wizard simple and easy to read.

Appendix G: Additional Tables

A Participant Details

ID	Condition	DP Expertise	DP Answers Correct	Non-Male
P01	One	Expert	4/4	
P02	One	Novice†	4/4	
P03	Two	Novice	2/4	
P04	Two	Expert	4/4	
P05	One	Novice†	2/4	
P06	One	Novice†	3/4	x
P07	One	Novice	1/4	x
P08	Two	Expert	3/4	x
P09	Two	Novice	2/4	x
P10	Two	Novice	1/4	
P11	Two	Expert	4/4	
P12	Two	Expert	4/4	
P13	One	Novice	2/4	
P14	One	Novice	1/4	x
P15	One	Novice	1/4	
P16	One	Expert	4/4	
P17	One	Novice†	2/4	x
P18	Two	Novice	0/4	

Table 1: Summary of the 18 study participants. † indicates participants whose DP expertise categorization was updated after the study session.

B Additional Results

Task	Avg (min)	Percentage	Task Modality
Task 1	7.74	34.40%	Wizard
Task 2	2.21	9.80%	Wizard
Task 3	6.25	27.80%	Notebook
Task 4	6.33	28.10%	Notebook

Table 2: Average time spent and percentage distribution per task.

ID	Tut.	T1	T2	T3	T4	Total
P1	10.17	10.23	1.08	3.45	3.73	18.50
P2	4.35	9.25	—	13.87	2.35	25.47
P3	2.70	13.53	1.08	5.97	15.40	35.98
P4	17.58	6.72	2.05	9.35	0.47	18.58
P5	10.20	9.12	1.12	4.38	3.88	18.50
P6	4.37	3.72	3.32	4.20	0.90	12.14
P7	5.45	6.67	11.82	7.11	16.55	42.15
P8	21.83	5.82	—	3.45	9.92	19.19
P9	8.30	3.18	0.37	6.23	11.40	21.18
P10	16.33	9.50	1.00	5.42	2.22	18.14
P11	5.65	2.95	4.03	11.30	7.17	25.45
P12	5.68	11.95	1.07	5.50	3.67	22.19
P13	9.10	6.67	1.97	7.17	3.00	18.81
P14	4.11	10.87	0.70	4.37	8.87	24.81
P15	7.32	5.38	—	3.67	11.23	20.28
P16	15.15	3.67	1.67	4.33	3.88	13.55
P17	5.05	10.35	7.50	7.42	7.07	32.34
P18	14.73	9.90	1.00	5.42	2.22	18.54
Mean	9.34	**7.74**	**2.21**	**6.25**	**6.33**	**23.11**
Percentage	—	**34.4%**	**9.8%**	**27.8%**	**28.1%**	-

Table 3: Task Performance Metrics ($N = 18$)