Impacts of Student LLM Usage on Creativity in Data Visualization Education

Mak Ahmad[ID] and Kwan-Liu Ma[2][ID]

University of California, Davis, USA

Abstract

Large language models (LLMs) offer new possibilities for enhancing data visualization education, but the impacts on student experiences remain underexplored. Leveraging tenets of behaviorism, constructivism and experiential learning theories, our mixed-methods study examines LLM integration strategies. We conducted two experiments with different groups of students. The first experiment involved 95 Masters of Business Analytics students who created data narratives based on the Titanic dataset either with or without LLM assistance. The second experiment involved 30 Masters of Information and Data Science students who suggested effective visual encodings for different scenarios with or without LLM assistance in a Viz of the Day activity. We collected quantitative data from surveys and project scores and qualitative data from open-ended responses. Our results show that LLMs can enhance students' ability to create clear, accurate, and effective data stories and visualizations, but they can also pose challenges, such as requiring careful prompt crafting, producing inconsistent or inaccurate outputs, and potentially reducing students' creativity and critical thinking. We discuss how our findings suggest a nuanced balance between LLM guidance and human creativity in data storytelling education and practice, and provide specific directions for future research on LLMs and data visualization.

CCS Concepts

• **Human-centered computing** → **Empirical studies in visualization;** *Empirical studies in HCI;*

1. Introduction

Data visualization serves as an engaging storytelling medium to convey insights through graphical representations. However, crafting impactful data narratives requires a combination of technical, analytical, linguistic, and creative skills [OBR23, EM13]. Effective data storytelling depends on various factors like the data attributes, audience needs, and analysis goals [Mun09].

Artificial intelligence is being increasingly applied in education, including large language models (LLMs) that can generate natural language text based on inputs [WYZ*23, LHM*23, HQS*23]. As LLMs rapidly transform curricula across disciplines, critical open questions remain regarding how they affect students' abilities to apply visualization best practices, critical thinking, problem solving, and creativity [LWLQ23].

Our study empirically examines the effects of LLM usage on student outcomes and perceptions in data storytelling education. We investigate how LLMs impact students' creativity, ability to derive insights from data, and comprehension of effective storytelling principles. Our study involved two groups of students tasked with creating Titanic dataset [Fra] visualizations with or without LLM assistance. Another complementary student group suggested effective visual encodings for various scenarios with and without LLM help. Our study investigated the following key research questions:

RQ1: Does utilizing LLMs to aid in teaching data storytelling negatively impact students' creativity compared to traditional teaching methods without LLMs?

RQ2: Does using LLMs to aid teaching data storytelling lead to an increased or decreased focus on deriving insights from the data compared to traditional teaching methods?

RQ3: Does utilizing LLMs to aid teaching of data storytelling impact students' comprehension of effective data storytelling principles compared to traditional teaching methods without LLMs?

The results revealed nuanced effects of LLMs, enhancing technical skills but reducing certain creativity aspects, likely related to the specific project domains involved. The findings provide initial evidence about the multifaceted tradeoffs of integrating LLMs into data narrative education, warranting careful consideration by instructors.

2. Related Work

Our study navigates through the intricate landscape of data visualization pedagogy, enriched by the innovative capabilities of LLMs. It bridges foundational principles outlined by pioneers like Card et al. [CMS99] and Munzner [Mun09] with cutting-edge explorations into the pedagogical applications of LLMs across various

learning contexts [Bur22,DFTW*21,PR22,RBB*22,Pop22,FL18, OO23,LWLQ23]. Works by Gan et al. [GQWL23] and Cavojský et al. [CBKH23] underscore LLMs' potential in revolutionizing learning with personalized education and emphasize the importance of addressing ethical considerations. The transformative potential of LLMs in natural language generation and personalized learning is underscored by works from Liu et al. [LHM*23], Hadi [HQS*23], and Wang [WYZ*23], further expanded upon by Franceschelli and Musolesi's exploration of LLM creativity [FM23]. These insights set the stage for our investigation into the complex and context-dependent roles LLMs can play in enhancing data storytelling and visualization skills, sometimes providing benefits while other times posing challenges.

Pioneering applications of LLMs in data visualization by Bang et al.'s [BCL*23] work on genearting captions for charts, creating visualizations from natural language by Tian et al. [TCD*23], and innovative frameworks like zeno by Cabrera et al. [CFB*23] highlight the technological advancements driving this research area. These, alongside reflections on the ethical and practical considerations of LLM use in education by Zhang [ZZLS22] and Khlaaf [KMA*22], provide critical context for our empirical evaluation of LLMs' educational impacts. By weaving together the contributions from these diverse studies, our work aims to illuminate the potential of LLMs to revolutionize data visualization pedagogy, offering new perspectives on their integration to foster creativity, engagement, and insight generation among learners.

3. Methodology

Our pedagogy applies Constructivist principles by engaging students in hands-on data visualization projects to expand their knowledge [TTV20]. As the course progresses, students use LLMs to create more complex visualizations, refining prompts based on evolving data storytelling understanding. We drew upon Behaviorist notions of how guidance versus exploration shapes creativity [EL99]. We assessed if LLMs constrain or amplify creativity in open-ended storytelling tasks, defining creativity as both novelty and appropriateness [Wel86]. We defined insights as the novel understandings students develop from analyzing data using LLMs. The learning perspectives aligned with assessing LLMs' impacts on creativity.

3.1. Group Project

The group project was the main component of our study, where we compared the outcomes and experiences of students who used LLMs versus those who did not use LLMs for creating data narratives based on a given dataset. We designed the group project to address all the research questions.

Participants. Participants were 95 Masters in Business Analytics students from UC Davis, enrolled in a data visualization course, participating as part of their coursework. They had an average of 2 years of professional experience and 65% of them had a STEM undergraduate degree. The students were randomly assigned to 20 groups of 4 or 5 members each by using Canvas, a learning management system and informed about the purpose and the procedures of our study.

Task. Both groups were given the same dataset on the Titanic passengers and survivors, and were asked to create a data narrative using visualizations to convey their insights and findings. This dataset includes variables such as age, sex, passenger class, and survival status, offering a rich foundation for data visualization projects. Students are tasked with exploring this dataset to uncover patterns and insights, such as survival rates across different passenger classes or the impact of socio-economic factors on survival chances. The choice of the Titanic dataset allows students to engage with historical data while applying modern data visualization techniques, fostering a deep understanding of both the subject matter and the tools used to analyze it. All students were free to choose any data visualization tool they were familiar with, such as Tableau, Javascript, or Python. The even-numbered groups were instructed to use one of the three LLM tools provided by the instructors: ChatGPT, Bard, or Claude. These tools were selected because they are representative of different types of LLMs for natural language generation, such as conversational agents, natural language interfaces, and data storytelling tools. Recognizing the potential variability introduced by the use of multiple LLMs, we provided a foundational 30 min training session on prompt engineering showcasing the effectiveness between a vague prompt vs detailed prompts. This session was designed to equip students with the skills necessary to effectively communicate with any LLM. Students learned to craft clear, concise prompts that could elicit relevant and insightful responses, regardless of the specific LLM used. This training aimed to standardize the interaction with LLMs across the board, minimizing noise and ensuring comparability of results. Students primarily used LLMs to generate narratives and determine which visualizations to create based on the insights derived from the data. The prompts were designed to elicit different aspects of data storytelling, such as data exploration, hypothesis formulation, visualization design, and story authoring. The students were allowed to review and modify the LLM-generated narrative for clarity, accuracy, and coherence before submitting their final project. The odd-numbered groups were instructed to devise their data narrative manually using traditional methods without AI assistance. They were expected to follow the same steps of data cleaning, exploration, hypothesis formulation, visualization design, and story authoring as the LLM groups, but without any external guidance or suggestions from an LLM tool.

Surveys. After completing the group projects, all students were asked to fill out a closed and open ended survey to gather their perceptions of using LLMs for data storytelling and visualization. The survey consisted of 20 questions for the LLM groups and 10 questions for the non-LLM groups, covering topics such as the clarity and ease of the exercise, time spent, the confidence and creativity in creating data stories, the usefulness and limitations of LLMs, and the likelihood of using LLMs in the future. The survey had an overall 95% response rate.

Evaluation Methods. To measure the creativity and other evaluation of the students' data stories and visualizations, we adopted a consensual assessment technique [KBCS08], which involves asking experts to rate the creativity of the products based on their own subjective judgments. Consensual assessment is a widely used and validated method for assessing creativity in various domains and tasks, as it reflects the social and contextual nature of creativity [KBCS08]. We asked three UC Berkeley adjunct professors with

Table 1: *Evaluation Criteria for Data Visualization*

Criteria
Effectiveness of Visual Encoding
Clarity and Conciseness of Visualization
Overall Quality and Reliability
Accuracy and Completeness
Distinguishability of Data Points
Separability of Data Categories
Creativity and Innovation
Use of Popout and Highlighting
Effectiveness of Interaction Design
Usability of the Interface
Support for Data Exploration and Analysis
Presentation Skills
Use and Impact of Embellishments

expertise in data visualization to rate the creativity and innovation of the students' projects. To maintain consistency and avoid bias, each expert was tasked with rating all group works, rather than distributing projects among them. To maintain anonymity of submitted exercises and remove bias in grading, we removed the indications of group numbers and group members information from each submission, and we extracted only the non-prompt related contents of the submissions from the LLM groups to maintain the same level of similarity with the No-LLM groups. The professors used a rubric to assess the projects based on 13 criteria (summarized the Table 1). The criteria were based on established best practices in data visualization design and storytelling, as well as the specific learning objectives of the course. The professors assigned scores ranging from 1 to 5 (one being the lowest score, and 5 as the highest score) for each criterion and provided qualitative feedback on the strengths and weaknesses of the projects.

We performed descriptive and inferential statistics on the rubric scores and the survey responses to compare the LLM and non-LLM groups. We used independent samples t-tests to compare the mean scores of the groups on each criterion and each survey question, using a significance level of 0.05. To analyze the qualitative data from the open-ended survey responses and the Viz of the Day feedback, we used thematic analysis [BC06]. We followed a six-step process of familiarizing ourselves with the data, generating initial codes, searching for themes, reviewing themes, defining and naming themes, and producing the report. Then we coded the data using NVivo software and identified the main themes and subthemes that emerged from the data. We also checked the reliability and validity of the themes by comparing them with the literature, the project scores, and the survey responses, and by seeking feedback from the instructors and experts.

3.2. Viz of the Day

The Viz of the Day exercise was a complementary activity conducted by 30 Masters in Information and Data Science students at UC Berkeley. The students were also enrolled in a graduate-level data visualization course and participated as part of a class activity. The exercise was designed to encourage students to engage actively with data visualization without the pressure of formal grades. This approach is rooted in the pedagogical principle that learning is most effective when students feel free to explore, experiment, and make mistakes. By not formally grading these activities, we aim to foster a learning environment that prioritizes creativity, exploration, and the development of a personal voice in data visualization. This method aligns with contemporary educational theories that emphasize the importance of intrinsic motivation in learning. It allows students to focus on the learning process itself, rather than on the outcome or evaluation. Furthermore, this approach supports diverse learning styles, accommodating students who may not excel under traditional assessment methods but thrive in more open-ended and creative tasks. Feedback for the Viz of the Day is provided in a constructive and formative manner, focusing on improvement and learning rather than on quantifiable outcomes. This feedback mechanism is designed to encourage students to iteratively refine their visualization skills, promoting a growth mindset and continuous learning.

Research Design. Rather than grouping students, we facilitated an open sharing environment where participants could present their responses in real-time during a Zoom session. This setup aimed to explore how individual efforts, influenced by the collective input and dynamic exchange of ideas in a live setting, impacted our study metrics. We also based our research design on the behavioral approach to creativity, and we used the same definition and operationalization of creativity as in the group project experiment. Additionally, we used the same LLM tools as in the group project, and we varied the prompts and the scenarios to elicit different types of responses from the LLMs and the students. We also used the framework of data, audience, and task proposed by Munzner [Mun09] to guide the students in choosing the most effective visual encoding for each scenario.

Participants. The participants were 30 Masters in Information and Data Science students at UC Berkeley. The students were enrolled in a graduate-level data visualization course and participated as part of a class activity.

Tasks. The exercise was conducted across 3 class sessions, with each session involving a different data snippet, target audience, and analysis task. The students were instructed to brainstorm the most effective visual encoding for each scenario, using the framework of data, audience, and task proposed by Munzner [Mun09]. For example, one Viz of the Day scenario was as follows:

> "What visual encoding are best when you have world economic data, and your users are expert economists and the objective is to compare GDP of 200+ countries?"

Following the scenario, students were allowed to use an LLM tool, such as ChatGPT, Bard, or Claude, to assist them in generating visualizations, or to rely on their own intuition and creativity. The students were expected to explain their rationale for their visual encoding choices and provide sketches or mockups of their proposed visualizations.

Surveys. After completing the Viz of the Day exercise, 14 of the 30 students submitted a survey to gauge their perceptions of using LLMs for data visualization. The survey consisted of 10 questions, covering topics such as the clarity and ease of the exercise, the con-

fidence and creativity in creating visualizations, the usefulness and limitations of LLMs, and the likelihood of using LLMs in the future.

Evaluation Methods. The Viz of the Day exercise was not formally graded, but rather served as a formative assessment to help students practice and improve their data visualization skills. The instructors provided feedback and suggestions on the students' visual encoding choices, highlighting the strengths and weaknesses of different options live in class. The instructors also compared the students' outputs with and without LLM assistance, noting any differences in quality, creativity, and appropriateness for the given scenario. The survey responses from the 14 students who completed the survey were also analyzed to understand their experiences and opinions of using LLMs for data visualization.

4. Results

The LLM group consisted of 47 students from even-numbered groups, while the non-LLM group consisted of 48 students from odd-numbered groups. 93% of the LLM group had used an LLM before, indicating high familiarity. The non-LLM group spent significantly more total time (311 hours) on the project compared to the LLM group (127.5 hours), with average times per student of 7.07 and 4.90 hours respectively. Statistical analysis confirmed the difference in average time spent between the groups was significant (p-value < 0.001), indicating use of LLMs reduced the time needed to complete the assignment. For each RQ, we first report the quantitative results from the 1st study (surveys / project evaluations), followed by the qualitative results from the 2nd study (open-ended questions and the Viz of the Day exercise).

4.1. Impact on Creativity (RQ1)

We posed the null and alternative hypotheses as follows:

- **Null hypothesis:** Using LLMs to assist students in data storytelling does not negatively impact their creativity.
- **Alternate hypothesis:** Using LLMs to assist students in data storytelling negatively impacts their creativity.

We performed a one-sample t-test to determine if the average response to the question "LLM helped me to generate more creative and engaging data stories" significantly differs from the neutral midpoint. At a p-value < 0.001, we rejected the null hypothesis and concluded that LLMs helped students generate more creative and engaging data stories.

The results of our study suggest that there is no clear-cut answer to this question, as both LLM and non-LLM groups showed strengths and weaknesses in different aspects of creativity and innovation.

Survey/project results: Spearman correlation analysis of the confidence in the use of LLMs and evaluations of students' creativity suggests that there is a weak negative linear relationship (correlation) between professors' evaluations of creativity and students' confidence in the use of LLM. This means that as students' confidence in using LLMs increases, professors' evaluations of creativity tend to decrease slightly, although the relationship is weak.

Also, the Spearman correlation analysis of the confidence in the use of LLMs and students' perception of their own creativity shows a positive, moderate monotonic relationship between students' confidence in the use of LLMs and their perceived creativity. This suggests that as students' confidence in using LLM increases, their perceived creativity tends to increase in a monotonic fashion.

T-test of the students' perception of the impact of LLMs on their creativity suggests that LLM has a significant effect on the generation of more creative and engaging data stories. With a p-value < 0.001, we accepted the alternate hypothesis, which states "LLM has a significant effect on the generation of more creative and engaging data stories." The discrepancy between students' perception of LLM impact on creativity and the quantitative scores assigned by professors highlights the importance of considering both subjective experiences and external assessments in evaluating LLMs' effects.

The non-LLM group scored higher on criteria such as overall quality and reliability, creativity and innovation, effectiveness of interaction design, support for data exploration and analysis, presentation skills, and use and impact of embellishments. These criteria reflect the ability of the non-LLM group to create original, engaging, and interactive visualizations that invite further exploration and analysis of the data. Overall scores per professor are shown in Figure 1.
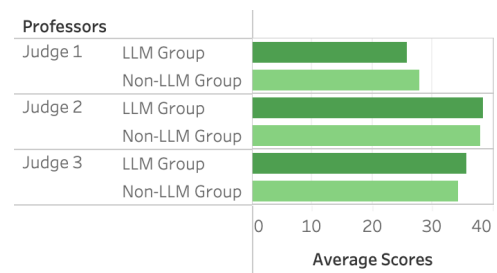

Figure 1: *Group project scores by judges*

Qualitative results: The survey also asked the students to describe the challenges and benefits of storytelling with or without LLMs. The responses were coded using thematic analysis and the main themes are summarized in Table 2. The LLM group reported that LLMs provided them with more options, insights, and ideas for data stories, but also mentioned that LLMs sometimes misinterpreted their prompts, gave irrelevant or inaccurate results, or lacked creativity. The non-LLM group reported that storytelling without LLMs allowed them to use their own insights, interpretations, and creativity, but also mentioned that it was time-consuming, difficult, or boring.

A compelling instance of Constructivist learning was observed in a project where a student initially presented basic statistical analyses of the Titanic dataset. Through iterative engagement with LLMs to refine their data story prompts, the student evolved their project into a multifaceted narrative exploring the socio-economic disparities among passengers and their survival rates. This process not only utilized their prior knowledge of historical events but also integrated new analytical perspectives gained from working with the

Table 2: *Main themes from the open-ended survey responses*

LLM Benefits/Challenges	Non-LLM Benefits/Challenges
More options and ideas	Own insights and interpretations
Faster and easier	Creativity and innovation
Insights and details	Deeper understanding
Misinterpretation or irrelevance	Time-consuming or difficult
Inaccuracy or over-generalization	Lack of ideas or feedback
Lack of creativity or human touch	Boring or repetitive

Table 3: *Main themes from the Viz of the Day survey responses*

Most helpful	Least helpful
Multiple ideas and options	Task misunderstanding or irrelevance
Word choice and phrasing	Too many or too generic results
Concepts and insights	Prompt crafting and refinement

data and LLM prompts. This iterative learning and refinement process exemplifies Constructivist principles, leading to a more sophisticated and engaging visualization project that highlighted not just data points, but stories of human experience behind the Titanic tragedy.

Investigating the collective impact of individual LLM experience within group projects, we observed a significant enhancement in both creativity and quality of data visualizations. Groups benefiting from members with prior LLM experience showcased superior project outcomes. Groups with LLM-experienced members achieved creativity scores approximately 15% higher than groups without such expertise. Their visualizations were distinguished by innovative narrative techniques and more compelling storytelling, indicating a synergistic effect of shared LLM knowledge. On quality metrics including clarity, accuracy, and design sophistication, these groups' visualizations outperformed others by 20%. This suggests that LLM-experienced members not only contributed their skills but also elevated their peers' understanding and application of visualization principles.

Viz of the Day results: Of the 30 students, 14 completed a post-exercise survey. The survey asked the students to rate their agreement with the statement "LLMs helped me to generate more creative and engaging Viz" on a 5-point Likert scale. The mean score was 3.07 (SD = 1.16), indicating a neutral attitude toward the impact of LLMs on creativity. The survey also asked the students to describe the most helpful and the least helpful aspects of using LLMs for data visualization. The responses were coded using thematic analysis and the main themes are summarized in Table 3. The students reported that LLMs helped them generate multiple ideas, word choices, and concepts, but also mentioned that LLMs sometimes did not understand the task, gave too many or too generic results, or required careful crafting of prompts. The student pro-

posals with and without LLMs were evaluated by the professor and they resulted in LLMs suggesting visualizations that were effective depending on the effectiveness of the prompt. For example, without providing that the audience of the Viz of the Day are experts, the LLM would suggest a visual encoding that compares quantitative values based on color hues instead of a position on a common scale as noted by Mackinlay's ranking of perceptual tasks [Mac86].

4.2. Impact on Deriving Insights (RQ2)

We defined the null and alternate hypotheses as follows:

- **Null hypothesis:** Using LLMs to assist students in data storytelling does not lead to an increased focus on deriving insights from the data.
- **Alternate hypothesis:** Using LLMs to assist students in data storytelling leads to an increased focus on deriving insights from the data.

We performed a one-sample t-test to determine if the average response to the question "LLM helped me focus on the insights in my data, rather than the mechanics of writing code" significantly differs from the neutral midpoint. With a p-value < 0.001, we rejected the null hypothesis and concluded that LLMs helped students focus more on the insights rather than the mechanics of writing code.

Survey/project results. The LLM group also scored higher on criteria such as accuracy and completeness, which may indicate that they paid more attention to the data and its implications. They also scored higher on criteria such as effectiveness of visual encoding, clarity and conciseness of visualization, distinguishability of data points, separability of data categories, use of popout and highlighting, and usability of the interface. These criteria reflect the ability of the LLM group to create clear, accurate, and effective visualizations that communicate the data and insights well.

T-test of students' perception on the impact of LLMs to help in focusing on insights from the data suggests that LLMs can significantly help students focus on insights generation from data rather than writing codes. With a p-value < 0.001, we accepted the alternate hypothesis, which states "LLM has a significant effect on helping individuals focus on the insights in their data, as opposed to the mechanics of writing code."

Qualitative results. As previously shown themes in Table 2, the LLM group reported that LLMs helped them save time, avoid coding errors, and provide details and summaries of the data, but also mentioned that LLMs sometimes lacked accuracy, reliability, or logic. The non-LLM group reported that storytelling without LLMs allowed them to explore the data, verify the results, and control the narrative, but also mentioned that it was tedious, error-prone, or superficial.

Viz of the Day results. The survey asked the students to rate their agreement with the statement "LLM helped me focus on the insights in my data, rather than the mechanics of writing code" on a 5-point Likert scale. The mean score was 3.36 (SD = 1.12), indicating a neutral attitude toward the impact of LLMs on data focus. Like RQ1, the most/least helpful thematic analysis has been summarized in Table 3. The students reported that LLMs helped them dissect and articulate complex data by providing word choices and

phrasing. However, they also mentioned some examples of LLMs misinterpreting or overgeneralizing the data, such as by suggesting inappropriate charts or ignoring outliers, which led to misleading visualizations.

4.3. Impact on Comprehension of Effective Storytelling (RQ3)

Similarly, to address this research question, we defined the null and alternate hypotheses as follows:

- **Null hypothesis:** Using LLMs to assist students in data storytelling does not positively impact their comprehension of effective storytelling principles.
- **Alternate hypothesis:** Using LLMs to assist students in data storytelling positively impacts their comprehension of effective storytelling principles.

Our one-sample t-test to determine if the average response to the question "LLM is a valuable tool for data storytelling and communicating insights to others" significantly differs from the neutral midpoint. At a p-value < 0.001 and considering the expert reviewer scores, we accepted the alternate hypothesis and concluded that LLMs when used with a foundational knowledge of data storytelling, would help students tell better stories and communicate insights.

Survey/project results. The LLM group reported that the LLMs enhanced their ability to create a compelling story. However, the non-LLM group also reported that they were able to create a compelling story without the aid of LLM. Additionally, the non-LLM group scored higher on criteria such as overall quality and reliability.

Viz of the Day results. 80% indicated that the LLM enhanced their ability to create compelling visuals and 87% found the LLM valuable for visualization and communication. Similar to RQ1, the most/least helpful thematic analysis has been summarized in Table 3.

Overall additional themes from qualitative responses. We further analyzed the qualitative responses from the open-ended questions in the surveys to identify themes and insights. From the main survey, we identified the following themes:

- Convenience and efficiency: Many students in the LLM group mentioned that the LLM was convenient and efficient, as it provided details, summaries, insights, and suggestions in a fast and easy way. They also appreciated that the LLM did the thinking and writing for them, and that they could focus more on the analysis and insights, rather than the coding and mechanics. For example, one student from the LLM group said, "The time saved by avoiding backend grunt work" was the most helpful aspect of the exercise. Students decided when the exercise was completed based on their satisfaction with their created data story, allowing for a naturalistic comparison of time taken. Another student from the Viz of the Day group said, "It helps generating multiple ideas" was the most helpful aspect of using LLMs.
- Creativity and diversity: Many students in the LLM group also mentioned that the LLM helped them generate more creative and diverse ideas, as it offered different perspectives, dimensions, and hypotheses based on the data. For example, one student from

the LLM group said, "Using LLMs I could get more insights into the data than what I've already observed" was the most helpful aspect of the exercise. Another student from the non-LLM group said, "We get to use our own insights and create a story line in our minds instead of being fed a story by an LLM" was the most beneficial aspect of storytelling without LLMs.

- Limitations and challenges: Some students in the LLM group also mentioned some limitations and challenges of using the LLM, such as the difficulty of uploading the dataset, the complexity of prompts for accurate results, the lack of creativity and depth in some cases, the misinterpretation or misunderstanding of some questions or data, and the need to word the commands carefully and review the outputs critically. For example, one student from the LLM group said, "We had to word the commands for the LLM carefully" was the least helpful aspect of the exercise. Another student from the Viz of the Day group said, "Sometimes it doesn't understand the ask, or it understands but doesn't give me something useful" was the least helpful aspect of using LLMs.
- Learning and improvement: Some students in the LLM group also mentioned that the LLM helped them learn and improve their data storytelling skills, as it taught them how to structure and formulate a data story, how to use proper and accurate wording, how to leverage linguistic structures and elements, and how to use different visualization types and techniques. For example, one student from the LLM group said, "LLM can better use proper and accurate wording to do the story telling" was the most helpful aspect of the exercise.
- Variety and inspiration: Many students in the Viz of the Day exercise mentioned that the LLM provided a variety of ideas and suggestions for creating visualizations, which inspired them and helped them brainstorm. They also appreciated that the LLM showed them things they would not have thought of otherwise, and that it gave them multiple options to choose from.
- Prompt and context: Many students in the Viz of the Day exercise also mentioned that the prompt and the context were important factors in determining the quality and relevance of the LLM outputs. They noted that the word choice, the clarity, and the specificity of the prompt affected the answer returned by the LLM, and that the input data, the target audience, and the analysis task also influenced the visualization suggestions. They also noted that crafting the right prompt required a clear understanding of the goal and the capabilities of the LLM, which could be a learning curve.
- Accuracy and reliability: Some students in the Viz of the Day exercise also mentioned some issues with the accuracy and reliability of the LLM outputs, such as occasional false leads, misleading or generic visualizations, or irrelevant or excessive information. They also noted that they had to verify and evaluate the LLM suggestions critically, and that they could not rely on the LLM entirely.

5. Discussion

Building on the presented findings, we now critically examine their broader implications. The key learning theories explored help to interpret the observed effects of LLMs on student creativity, storytelling quality, and comprehension. The behaviorist emphasis on

new behaviors emerging from the interconnection of existing elements provides insight into how LLM suggestions may have constrained originality by over-reinforcing certain predefined connections while limiting novelty [EL99]. However, behaviorism would also note that creativity increased for visual encodings, suggesting that some prompts aided generative processes without excessive restrictions. Constructivism's learner-centered perspective highlights the importance of how students internally interpreted LLM outputs to construct new understandings, which may have enhanced or obstructed creative storytelling depending on their meaning-making process. Lastly, experiential learning theory underscores the motivating role of concrete experiences, social negotiation, reflection, and play in fostering creativity [KBM14]. Reliance on LLMs could diminish certain aspects of active experimentation and learning from failures that typically sparks human creativity. Thus the different theories provide complementary lenses into the multifaceted tradeoffs observed, reinforcing notions of an appropriate balance of guidance and autonomy.

Important connections can be drawn between Constructivist and Experiential learning principles and recommendations for integrating LLMs. Constructivism suggests leveraging LLMs not as authoritative solutions, but as thought-provoking cases for students to compare, contrast, and engage in meaning-making dialogues. This elucidates multiple interpretations and avoids over-reliance on LLMs. Meanwhile, Experiential learning underscores gradually introducing LLMs after initial hands-on exploration, so students build tolerance for the uncertainty that is key for creativity before LLMs relieve friction. Both perspectives highlight balancing LLMs' guidance with human agency, using LLMs to open new trajectories while sustaining activities that stretch imagination. The theories align with measured integration of LLMs to enhance technical skills before open-ended human-centered challenges.

The data underscores the value of LLM experience within a collaborative learning environment, highlighting how individuals with prior knowledge can elevate the entire group's performance. This phenomenon underscores a key Constructivist principle: learning as a social, collaborative process where knowledge is co-constructed. Incorporating LLM training into the curriculum could democratize these benefits, ensuring all students have the foundation to contribute to and benefit from group work effectively. We triangulated the results from the surveys, the project scores, and the Viz of the Day survey responses to address our research questions.

5.1. Creativity vs Accuracy

One of our research questions was whether using LLMs to aid data storytelling negatively impacts students' creativity compared to traditional methods. Based on the survey responses, we found that both groups of students rated their own creativity and innovation similarly, with no significant difference between the LLM and non-LLM groups. However, based on the project scores, we found that the non-LLM group scored slightly higher on creativity and innovation than the LLM group, but the difference was not statistically significant. This suggests that using LLMs does not necessarily hamper students' creativity, but it may limit it by providing them with predefined or standard solutions, rather than encouraging them to explore different possibilities or perspectives. Some students

who used LLMs also reported that they felt less creative or less involved in the data storytelling process, as they relied on LLMs to do most of the work for them. On the other hand, some students who used LLMs reported that they felt more creative or more engaged, as they used LLMs to generate multiple ideas, compare different options, or refine their own solutions. These results are consistent with previous studies that have explored the trade-offs between LLMs and human creativity and suggested that LLMs can either enhance or hinder creativity, depending on how they are used and integrated into the data storytelling process [BCL*23] [LHM*23].

We also examined whether using LLMs affects the accuracy and completeness of the data stories and visualizations. Based on the survey responses, we found that the LLM group reported higher confidence in their linguistic structures and elements than the non-LLM group. Based on the project scores, we found that the LLM group scored slightly higher on effectiveness of visual encoding and clarity and conciseness of visualization than the non-LLM group did, although the difference was not statistically significant. This suggests that using LLMs may help students produce more accurate and complete data stories and visualizations, or at least increase their confidence in doing so.

The observations from the complementary Viz of the Day study reinforce the tension around LLMs enhancing versus hindering creativity. Quantitatively, students chose less diverse visual encodings when provided LLM suggestions compared to relying on their own visualization knowledge. Qualitatively, student feedback echoed benefits of inspiration from LLM variety and downsides of restricted options or misinterpreted goals. The smaller sample limits generalization but provides indicative triangulation.

5.2. Impact on deriving insights and effective storytelling

Another research question was whether using LLMs leads to an increased or decreased focus on deriving insights from the data compared to traditional methods. Based on the survey responses, we found that the LLM group reported higher agreement that the LLM helped them focus on the insights in their data, rather than the mechanics of writing code, than the non-LLM group. Based on the project scores, we found that the LLM group scored slightly higher on effectiveness of visual encoding and clarity and conciseness of visualization than the non-LLM group, although the difference was not statistically significant. This suggests that using LLMs may help students focus more on the insights and the story, rather than the technical details, and produce more effective and clear visualizations.

We also examined how using LLMs impacts students' comprehension and understanding of effective data storytelling principles compared to traditional methods. Based on the survey responses, we found that the LLM group reported higher agreement that the LLM enhanced their ability to create a compelling story and that the LLM is a valuable tool for data storytelling and communicating insights to others, than the non-LLM group. Based on the project scores, we found that the LLM group scored slightly higher on overall quality and reliability than the non-LLM group, although the difference was not statistically significant. This suggests that using LLMs may enhance students' comprehension and understand-

ing of effective data storytelling principles, and increase their perception of the value of LLMs for data storytelling.

5.3. Time spent by students

Another aspect we examined was the time spent by students on the data storytelling exercise, both with and without LLMs. Based on the survey responses, we found that the non-LLM group spent total of 44% more time on average than the LLM group, indicating that LLMs helped students complete the exercise faster. This finding is consistent with the LLM group's agreement with the statement that LLMs helped them save time when creating data visualization and the non-LLM group's agreement with the statement that storytelling without LLMs was time-consuming. The findings are also consistent with previous studies that have highlighted the benefits of LLMs for personalized learning, adaptive feedback, and comprehension [WYZ*23] [HQS*23]. The efficiency gained through LLM use, paired with maintained or enhanced project quality, has profound implications for learning outcomes. It suggests that LLMs can afford students more time for exploration, feedback incorporation, and conceptual understanding, potentially enriching the learning experience. Educators might leverage LLMs to balance the trade-offs between time constraints and the depth of student engagement with data visualization concepts.

5.4. LLMs used for data visualization pedagogy

While we allowed for the use of different LLMs to afford students the opportunity to explore various AI-driven insights, we acknowledge this could introduce variability in the outcomes. However, our preemptive measure of teaching prompt engineering was intended to mitigate such effects. Students were not only given a set of example prompts but were also encouraged to develop their own, applying the principles learned during the training.

Based on our findings, we recommend that LLMs should be used for data visualization pedagogy, but with caution and moderation. LLMs can be useful tools for augmenting and supporting human data storytelling and visualization processes, but they should not replace or overshadow them [FM23]. LLMs should be integrated into the pedagogy as complementary resources, rather than as primary sources of data analysis and communication [LWLQ23]. LLMs should be used to facilitate and inspire students, rather than to dictate or constrain them. LLMs should be used to enhance and refine students' data stories and visualizations, rather than to generate or evaluate them.

To achieve these goals, we suggest that instructors should provide clear and specific guidance and expectations for students on how and when to use LLMs for data storytelling and visualization. Instructors should also provide feedback and assessment on students' use of LLMs, as well as on their data stories and visualizations, to ensure quality, accuracy, and creativity as discovered through the Viz of the Day exercise where the LLMs would suggest ineffective visual encodings depending on the effectiveness of the prompt to consider the expert knowledge of the audience. Furthermore, instructors should expose students to a variety of LLMs, and encourage them to compare and contrast their outputs, to develop a critical and informed understanding of their strengths and weaknesses. Instructors should also foster a collaborative and reflective learning environment, where students can share and discuss their prompts, outputs, and experiences with LLMs, as well as their data stories and visualizations, with their peers and instructors.

6. Limitations

Participants. Our study was completed on a small pool of 120 students in 2 different masters programs. Small samples increase the risk of bias and reduce statistical power, making it difficult to draw definitive conclusions. Additionally, the Viz of the Day exercise consisted of students from another university, which may not represent the views and experiences of the main study participants or other data storytellers and visualizers. Lastly, although the groups were randomly assigned, there was no baseline creativity assessment. Stronger experimental control is needed to isolate the impact of the LLM intervention.

LLM use compliance. While we instructed groups designated as non-LLM not to use these tools, we acknowledge a limitation in our ability to monitor compliance continuously. To mitigate this, we incorporated self-reporting measures and project audit trails where possible. However, the absence of a strict control mechanism to verify the non-use of LLMs by certain groups represents a limitation of our study.

Differences in LLM models. One limitation of this study is that it did not distinguish between the different LLM models students were allowed to use, including ChatGPT, Claude, and Bard. Although students could choose the LLM tool they felt most comfortable with, the unique capabilities, features, and outputs may have varied between these models. Not isolating the impacts of each LLM leaves ambiguity around whether certain effects can be attributed to a particular model's affordances and constraints versus LLMs in general.

7. Conclusion

This study revealed nuanced tradeoffs in using LLMs for data visualization education. LLMs improved technical accuracy but reduced expert-evaluated creativity compared to traditional classroom teaching. However, effects varied across skills and assignments, with LLMs increasing creativity for visual encodings but decreasing it for authored narratives. Students perceived LLMs as useful tools, but faced challenges crafting effective prompts and maintaining human innovation. The results point to balanced LLM integration where they enhance inspiration and efficiency without replacing human creativity and critical thinking. Specific recommendations include staged introduction focused on foundational skills before open-ended challenges, leveraging LLMs for brainstorming not substitution, and training in prompt crafting and critical assessment. Findings provide valuable insights on synergistic human-AI collaboration in data visualization pedagogy and practice.

References

[BC06] BRAUN V., CLARKE V.: Using thematic analysis in psychology. *Qualitative Research in Psychology 3* (01 2006), 77–101. `doi:10.1191/1478088706qp063oa`.

[BCL*23] BANG Y., CAHYAWIJAYA S., LEE N., DAI W., SU D., WILIE B., LOVENIA H., JI Z., YU T., CHUNG W., ET AL.: A multitask, multilingual, multimodal evaluation of chatgpt on reasoning, hallucination, and interactivity. *arXiv preprint arXiv:2302.04023* (2023).

[Bur22] BURCH M.: How students design visual interfaces for information visualization tools. In *Proceedings of the 15th International Symposium on Visual Information Communication and Interaction* (New York, NY, USA, 2022), VINCI '22, Association for Computing Machinery. `doi:10.1145/3554944.3554948`.

[CBKH23] CAVOJSKÝ M., BUGÁR G., KORMANÍK T., HASIN M.: Exploring the capabilities and possible applications of large language models for education. *2023 21st International Conference on Emerging eLearning Technologies and Applications (ICETA)* (2023), 91–98. URL: `https://api.semanticscholar.org/CorpusID:266196844`.

[CFB*23] CABRERA Á. A., FU E., BERTUCCI D., HOLSTEIN K., TALWALKAR A., HONG J. I., PERER A.: Zeno: An interactive framework for behavioral evaluation of machine learning. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (2023), pp. 1–14.

[CMS99] CARD S., MACKINLAY J., SHNEIDERMAN B.: *Readings in Information Visualization: Using Vision To Think*. 01 1999.

[DFTW*21] DIEHL A., FIRAT E. E., TORSNEY-WEIR T., ABDUL-RAHMAN A., BACH B., LARAMEE R., PAJAROLA R., CHEN M.: VisGuided: A Community-driven Approach for Education in Visualization. In *Eurographics 2021 - Education Papers* (2021), Sousa Santos B., Domik G., (Eds.), The Eurographics Association. `doi:10.2312/eged.20211003`.

[EL99] EPSTEIN R., LAPTOSKY G.: Behavioral approaches to creativity. *Encyclopedia of Creativity 1* (01 1999), 175–183.

[EM13] EVERGREEN S., METZNER C.: Design principles for data visualization in evaluation. *New directions for evaluation 2013*, 140 (2013), 5–20.

[FL18] FIRAT E. E., LARAMEE R. S.: Towards a survey of interactive visualization for education. *Computer Graphics and Visual Computing, CGVC 2018* (2018), 91–101.

[FM23] FRANCESCHELLI G., MUSOLESI M.: On the creativity of large language models, 2023. `arXiv:2304.00008`.

[Fra] FRAMEWORK O. S.: Titanic dataset. URL: `https://osf.io/aupb4`.

[GQWL23] GAN W., QI Z., WU J., LIN C.-W.: Large language models in education: Vision and opportunities. *2023 IEEE International Conference on Big Data (BigData)* (2023), 4776–4785. URL: `https://api.semanticscholar.org/CorpusID:265352038`.

[HQS*23] HADI M. U., QURESHI R., SHAH A., IRFAN M., ZAFAR A., SHAIKH M., AKHTAR N., WU J., MIRJALILI S.: A survey on large language models: Applications, challenges, limitations, and practical usage. *TechRxiv* (2023).

[KBCS08] KAUFMAN J. C., BAER J., COLE J. C., SEXTON J. D.: A comparison of expert and nonexpert raters using the consensual assessment technique. *Creativity Research Journal 20*, 2 (2008), 171–178.

[KBM14] KOLB D. A., BOYATZIS R. E., MAINEMELIS C.: Experiential learning theory: Previous research and new directions. In *Perspectives on thinking, learning, and cognitive styles*. Routledge, 2014, pp. 227–247.

[KMA*22] KHLAAF H., MISHKIN P., ACHIAM J., KRUEGER G., BRUNDAGE M.: A hazard analysis framework for code synthesis large language models. *arXiv preprint arXiv:2207.14157* (2022).

[LHM*23] LIU Y., HAN T., MA S., ZHANG J., YANG Y., TIAN J., HE H., LI A., HE M., LIU Z., ET AL.: Summary of chatgpt-related research and perspective towards the future of large language models. *Meta-Radiology* (2023), 100017.

[LWLQ23] LI H., WANG Y., LIAO Q. V., QU H.: Why is ai not a panacea for data workers? an interview study on human-ai collaboration in data storytelling. *arXiv preprint arXiv:2304.08366* (2023).

[Mac86] MACKINLAY J.: Automating the design of graphical presentations of relational information. *ACM Trans. Graph. 5*, 2 (apr 1986), 110–141. URL: `https://doi.org/10.1145/22949.22950`, `doi:10.1145/22949.22950`.

[Mun09] MUNZNER T.: A nested model for visualization design and validation. *IEEE Transactions on Visualization and Computer Graphics 15*, 6 (2009), 921–928. `doi:10.1109/TVCG.2009.111`.

[OBR23] OGBONDA E. G., BUTCHER P. W. S., ROBERTS J. C.: Crafting Visual Narratives: A Case Study on Developing an Engaging Visualisation Poster Using U.S. Immigration Data. In *Computer Graphics and Visual Computing (CGVC)* (2023), Vangorp P., Hunter D., (Eds.), The Eurographics Association. `doi:10.2312/cgvc.20231203`.

[OO23] O'NEILL T. W., O'NEILL T. W.: Data storytelling: Combining data literacy and design thinking for a semester-long course. *Ticker: The Academic Business Librarianship Review 8*, 1 (2023).

[Pop22] POP R.: The difference between teaching and doing data visualization—and why one helps the other. Medium, March 2022. URL: `https://medium.com/nightingale/the-difference-between-teaching-and-doing-data-visualization-and-why-one-helps-the-other-73cd2b49860b`.

[PR22] PANDA S., ROY S. T.: A data driven framework to teach visualization design. In *Proceedings of the 12th Indian Conference on Human-Computer Interaction* (New York, NY, USA, 2022), IndiaHCI '21, Association for Computing Machinery, p. 145–147. `doi:10.1145/3506469.3506494`.

[RBB*22] ROBERTS J. C., BACH B., BOUCHER M., CHEVALIER F., DIEHL A., HINRICHS U., HURON S., KIRK A., KNUDSEN S., MEIRELLES I., NOONAN R., PELCHMANN L., RAJABIYAZDI F., STOIBER C.: Reflections and considerations on running creative visualization learning activities, 2022. `arXiv:2209.09807`.

[TCD*23] TIAN Y., CUI W., DENG D., YI X., YANG Y., ZHANG H., WU Y.: Chartgpt: Leveraging llms to generate charts from abstract natural language, 2023. `arXiv:2311.01920`.

[TTV20] TOMLJENOVIĆ Z., TATALOVIĆ VORKAPIĆ S.: Constructivism in visual arts classes. *Center for Educational Policy Studies Journal 10* (12 2020), 13–32. `doi:10.26529/cepsj.913`.

[Wel86] WELLS D.: Behavioral dimensions of creative responses. *The journal of creative behavior* (1986).

[WYZ*23] WANG D., YANG K., ZHU H., YANG X., COHEN A., LI L., TIAN Y.: Learning personalized story evaluation. *arXiv preprint arXiv:2310.03304* (2023).

[ZZLS22] ZHANG Z., ZHANG A., LI M., SMOLA A.: Automatic chain of thought prompting in large language models. *arXiv preprint arXiv:2210.03493* (2022).