

Challenges and Opportunities of Teaching Data Visualization Together with Data Science

Shri Harini Ramesh*
Carleton University

Fateme Rajabiyazdi†
Bruyere Research Institute
Carleton University

ABSTRACT

With the increasing amount of data globally, analyzing and visualizing data are becoming essential skills across various professions. It is important to equip university students with these essential data skills. To learn, design, and develop data visualization, students need knowledge of programming and data science topics. Many university programs lack dedicated data science courses for undergraduate students, making it important to introduce these concepts through integrated courses. However, combining data science and data visualization into one course can be challenging due to the time constraints and the heavy load of learning. In this paper, we discuss the development of teaching data science and data visualization together in one course and share the results of the post-course evaluation survey. From the survey's results, we identified four challenges, including difficulty in learning multiple tools and diverse data science topics, varying proficiency levels with tools and libraries, and selecting and cleaning datasets. We also distilled five opportunities for developing a successful data science and visualization course. These opportunities include clarifying the course structure, emphasizing visualization literacy early in the course, updating the course content according to student needs, using large real-world datasets, learning from industry professionals, and promoting collaboration among students.

Index Terms: Data Visualization, Data Science, Education, Challenges, Opportunities

1 INTRODUCTION

In today's data-driven world, data visualization skills are highly valued and in demand across numerous industries, including finance, healthcare, marketing, technology, and academia [10, 18]. It is important to equip university students with essential data visualization skills for their future careers [22].

Students need a prior background in programming and data science to learn, design, and develop data visualization. However, students often lack strong programming and/or data science backgrounds. Previous studies have explored various methods of teaching data visualization to students without a programming background and have offered solutions such as using GUI-based visualization tools, including Tableau [20] or providing hands-on training for developing visualizations through projects [15], and activities [9, 16].

Moreover, some university programs lack a dedicated data science course, particularly at the undergraduate level, leaving students without knowledge of data science. Offering courses that combine data science and data visualization for these students can be instrumental. Such courses need to focus on providing students with the skills of the entire data science pipeline, from data collection and cleaning to analysis, interpretation, as well as data visualization. However,

combining data science and data visualization into one course can be challenging due to time constraints and the heavy load of learning both concepts. So, it is important to introduce and evaluate the effectiveness of such courses and share the findings with the visualization community to continuously improve the curriculum [1].

Thus, we developed and evaluated the effectiveness of teaching a course that combines data science and data visualization. We conducted a survey of the enrolled students and analyzed the responses. From the survey results, we identified four challenges of teaching data visualization together with data science, including [CH1] difficulty in learning and using multiple tools in a single term, [CH2] heterogeneous proficiency with tools and libraries, [CH3] difficulty learning data science-related topics (i.e., artificial intelligence (AI), machine learning (ML), and data visualization) in one course, and [CH4] challenges in selecting and cleaning a dataset. We also distilled five opportunities for enhancing future courses focused on teaching data visualization with data science, including [OP1] making the course structure clear and updating the content according to student needs, [OP2] introducing large real-world datasets for course assignments, [OP3] learning from industry professionals, [OP4] enhancing learning through collaboration, and [OP5] emphasizing visualization literacy and making interactive visualizations early in the course. We believe that the challenges and opportunities identified in our study will provide a foundation for educators when teaching data visualization together with data science.

2 COURSE BACKGROUND

We introduced and developed a new course, "Introduction to Data Science and Data Visualization," teaching both data science and data visualization to undergraduate students at an engineering department. This was a comprehensive 14-week course that included instructor-led lectures, class activities, guest lectures, and datathons (hands-on tutorial lab sessions). We held in-person lectures twice a week, each lasting 80 minutes; 20 minutes of each lecture was dedicated to class activities involving small tasks related to the lecture content, often using specific software or tools. We also invited guest speakers from academia, government, and industry—experts in their fields—to share their knowledge during lectures. Furthermore, we organized a total of 11 tutorial lab sessions, known as 'Datathons,' held weekly. Each session lasted 180 minutes and took place in person. The prerequisites for this course were knowledge of statistics and programming fundamentals to ensure students had proficiency in software development in at least one language.

The learning outcomes for students in this course included:

1. Explain different properties of data
2. Collect, clean, and process data
3. Analyze data using various techniques and tools
4. Assess existing visualizations based on visualization theory and principles
5. Design effective visual representations of data for exploration and communication

*e-mail: shrihariniramesh@cmail.carleton.ca

†e-mail: fatemerajabiyazdi@cunet.carleton.ca

We assessed students through five components: datathons, an individual project, a midterm, class activities, and guest lectures. Datathons comprised 50% of the final grade. Students had to complete at least 10 datathons. The evaluation was based on participation, presentation of the results, and the ability to answer questions posed in the datathons. The individual project was worth 25% of the final grade. Students had to select their own dataset, analyze the dataset, develop a visualization, and find insights from the data. It had three deliverables: a 2-page preliminary report, a class demo, and a detailed final report. The midterm, contributing 20% of the final grade, required students to achieve at least 50% to pass. The midterm had seven questions totaling 100 marks, with an additional bonus question worth 10 marks. Class activities and participation in guest lectures made up the remaining 5% of the final grade.

2.1 Course Syllabus

The course had six modules: Data Collection and Storage, Data Cleaning and Integration, Data Visualization, Correlation and Regression, Clustering and Classification, and Text Analysis and Text Visualization. Each module was comprised of a lecture, a class activity, and a datathon.

2.1.1 Module 1: Data Collection and Storage

Lectures: Module 1 spanned two weeks. In the first week, we introduced the roles and responsibilities of a data scientist, the essential skills required for the profession, the interconnectedness of data science and data visualization, and an overview of the data science lifecycle—from data collection to visualization. In the second week, we focused on data storage concepts and Structured Query Language (SQL) fundamentals. We covered SQL basics (creating tables, managing data types, and writing data queries) and introduced a database engine, SQL Lite (SQLite). Additionally, we introduced modern computing infrastructures, Google Colab, and cloud Tensor Processing Units (TPUs).

Class Activities: In the first week, we asked students to use tools such as the Application Programming Interface (API) Connector to extract data. In the second week, we asked students to write SQL queries for a given sample dataset, such as identifying products from specific countries and calculating prices.

Datathons: In the first week, we asked students to design survey questions using tools such as Google Forms and conduct the survey with their fellow students. To evaluate, we asked them to present the survey questions and responses, explain the reasoning behind the survey questions, discuss any challenges they faced during the survey conduction, and describe how they stored the survey data. In the second week, we asked students to write queries (e.g., create tables, delete entries) using SQLite and integrate SQLite with Python. They executed all these tasks using Python notebook in Google Colab. Then, to evaluate, we asked them to present their notebook and discuss the challenges they faced during this process.

2.1.2 Module 2: Data Cleaning and Integration

Lectures: Module 2 spanned two weeks. In the first week, we had a guest panel of data scientists from the city of Ottawa who discussed the city's open data and the process of making data open to the public. We also lectured about the importance of data cleaning, sources of error in data, methods to prevent errors, and ways to identify errors. In the second week, we discussed data wrangling, data integration, tools for data wrangling, and data integration.

Class Activities: In the first week, we conducted an activity where students had to measure and edit the distance between two string sequences. In the second week, we had activities for using OpenRefine to clean a dataset and using Tableau to integrate and join datasets.

Datathons: In the first week, we asked students to work on cleaning data using a Python notebook in Google Colab. We provided a

dataset and asked them to identify missing values and handle them using various data-filling methods. We also asked them to deal with duplicates and find outliers using visualization and methods such as z-score and interquartile range. To evaluate, we asked them to present the identified outliers in the data through visualization and discuss the potential impact of these outliers on the data analysis. In the second week, we asked students to connect various data sources in Tableau and create visualizations to conduct a preliminary analysis of the data. Students later presented their findings and conclusions based on the insights extracted from the visualizations.

2.1.3 Module 3: Data Visualization

Lectures: Module 3 spanned two weeks. In the first week, we discussed the importance of data visualization, effective and ineffective visual representations, and designing data visualizations. In the second week, we covered cognitive bias, visualization mirages, equity awareness in visualization, and an introduction to statistics.

Class Activities: In the first week, we conducted an activity in which students visualized two quantities to learn different methods for visualizing data [8]. In the second week, we used a public dataset available in Tableau and asked the students to visualize it to analyze the data and identify any mirages in the visualization.

Datathons: In the first week, we asked students to create a visual storyboard in Tableau using the provided dataset to gain insights and address questions presented in the datathon about the dataset. In the second week, we provided a different dataset and asked students to create visualizations using Power BI. They had to create visualizations for the questions posed in the datathon and an additional analysis question of their own that presented interesting insights about the data. For both of the week's evaluations, we asked them to present their findings and conclusions based on the insights extracted from the visualizations.

2.1.4 Module 4: Correlation and Regression

Lectures: Module 4 spanned three weeks. In the first week, we covered topics including the relationship between statistics and data science, P-hacking, power analysis, and exploratory data analysis. In the second week, we focused on correlation, regression analysis, and regression models. In the third week, we covered dimensionality reduction, principal component analysis (PCA), and T-distributed stochastic neighbor embedding (T-SNE).

Class Activities: In the first week, we asked students to critique a visualization they found online. In the second week, we asked students to perform regression analysis and visualize the results using Tableau Pulse, an AI-driven, intuitive tool for exploring datasets and automating insights discovery. In the third week, we had two activities. For the first activity, we provided students with a dataset and asked them to conduct regression models and visualize the results in Rstudio. For the second activity, we played a game with color names, teaching the Stroop effect [25].

Datathons: In the first week, we asked students to explore frequency distributions, visualize categorical data, and answer statistical questions (e.g., mean, standard deviation, normal distributions) using Python notebooks in Google Colab. In the second week, students implemented linear and polynomial regression and performed exploratory data analysis. In the third week, we asked students to implement multiple linear regression and create interactive visualizations for PCA representation. For all three weeks, we evaluated students by the quality of insights they shared when answering the data analysis questions for the provided dataset.

2.1.5 Module 5: Clustering and Classification

Lectures: Module 5 spanned two weeks. In the first week, we discussed clustering, its applications, and implementation. In the second week, we covered topics including classification, clustering, and their differences. We also had a guest speaker, a technologist

aiming to make technology understandable for the next generation. They spoke about the AI hype cycle in the industry and the ethical considerations that are often overlooked in the rush to be the biggest, the best, and the first.

Class Activities: In the first week, we invited two guest speakers: a chief executive officer from a data analytics company and a representative from Career Services. The motivation behind these guest lectures was to enhance student’s understanding of various data scientist roles, their day-to-day responsibilities, available jobs and roles in the industry and academia for data science, and the required skills. Students were encouraged to participate in the discussions and interact with the speakers. In the second week, we introduced students to a visualization tool called EduClust [11], a visual education platform for clustering algorithms. We asked students to experiment with different clustering algorithms and parameters using EduClust tool.

Datathons: In the first week, students implemented k-means clustering k-nearest neighbors, addressed overfitting and underfitting, and used decision trees, random forests, and naive Bayes classifiers using Python notebooks in Google Colab. To evaluate, we asked several questions, which students needed to answer by analyzing the provided dataset and then sharing the insights in class. In the second week, we didn’t have a datathon because it was a holiday.

2.1.6 Module 6: Text Analysis and Text Visualization

Lectures: Module 6 spanned one week. we introduced text data, text data processing, named entity recognition, document-level metrics, term frequency-inverse document frequency, and text visualizations.

Class Activities: In this week, we conducted three activities. For the first activity, we asked students to choose one text visualization from online resources and critique it for the class. For the second activity, we asked students to upload a document into Voyant ¹, a text analysis tool, visualize the text data using this tool and share their findings with the class. Then, for the third activity, we asked students to work with Julia to visualize text data.

Datathons: In this week, we students worked with text data in Tableau, where they cleaned, processed, and visualized the frequency and sentiment in the text data. To evaluate, we asked them to share the insights gained by visualizing the text data.

3 SURVEY STUDY

We conducted an anonymous survey with students enrolled in the 14-week long “Introduction to Data Science and Data Visualization” course in Winter 2024. Our aim was to assess the effectiveness of the course, especially in equipping undergraduate students with both data science and data visualization skills in a single course. We used Qualtrics to develop the survey, which included a mix of twelve Likert agreement scales ranging from strongly disagree to strongly agree, seven open-ended questions, and five multiple-answer questions. We began with demographic questions, including gender, age, academic program, and year of study. We then asked questions about each part of the course, including the datathons, the individual project, the tools learned, and the overall course structure. During the last lecture of the course, we shared our study recruitment poster with the students containing details about the study and a QR code linking to the survey. Interested students could scan the QR code, which directed them to the survey after they signed a consent form. Data collection took place in April 2024, and participants received a \$10 CAD e-gift card as compensation for their participation in the survey. Once we collected the data, we cleaned and performed statistical analysis for the Likert scale and multiple-choice questions in Qualtrics. We also used open coding [12] and affinity diagram technique [13] to analyze the open-ended questions.

¹<https://github.com/voyanttools/Voyant>

4 RESULTS

Out of 26 students enrolled in the class, 21 participated in the survey. The participants’ demographics breakdown is provided in Table 1. The students had basic knowledge in fundamental programming, object-oriented programming, programming with Python and Java, algorithms, data structures, and statistical analysis from their previous years’ coursework in their undergraduate studies.

Table 1: Demographic breakdown of survey participants.

Category	Details	Responses
Age	18-24 years	20
	25-34 years	1
Gender	Male	15
	Female	6
Program	Software Engineering	15
	Computer Systems Engineering	5
	Electrical Engineering	1
Year of Study	4th Year	17
	5th Year	4

We present the results as follows: (i) Course structure and content, (ii) Datathons, (iii) Individual project, (iv) Skills and tools, (v) Career aspirations, and (vi) Student confidence. A summary of the results for the Likert scale questions is provided in Table 2.

(i) Course structure and content: Teaching data science and data visualization in one course required us to cover many topics within a term. Thus, one of our aims when developing the course was to provide a clear and easy-to-follow structure to ensure students could keep up with the topics without feeling overwhelmed. This was reflected in the feedback, with the majority of students, 62% agreed, 29% strongly agreed with the statement “Overall organization and structure of the course were clear and easy to follow”.

(ii) Datathons: Datathons were developed to complement lectures and provide hands-on experience with the concepts and tools learned in class. When asked if “Lectures and datathons complement each other”, 57% of students agreed, and 14% strongly agreed. However, 5% disagreed, indicating that there is room for improvement in integrating these components. Furthermore, datathons aimed to foster collaboration among students. In our survey, more than 80% of the students (43% agreed and 43% strongly agreed) reported that “Datathons encouraged collaboration and discussion among students”. In another survey question, 15% of the students disagreed with the statement, “Group participation in datathons enhanced the learning experience and made problem-solving easier.” This indicates that although most students found group work beneficial, some students encountered challenges with collaboration. This could be because, in datathons, a few students took control of creating and presenting visualizations to ensure high grades, which limited opportunities for others to showcase their skills.

(iii) Individual project: The individual project component was developed for students to independently apply the skills and knowledge acquired throughout the course and develop a visualization system for their chosen dataset. When asked to respond to the statement, “The lectures and datathons helped me to develop the skills necessary to undertake individual projects independently,” 72% of the students affirmed that the components developed the requisite skills for dataset analysis and visualization creation. More than 90% of the students (48% agreed and 43% strongly agreed) reported that the “The individual project provided a valuable practical experience that reinforced the understanding of course concepts.”. Furthermore, all students agreed to the statement, “The individual project allowed me to explore topics of personal interest within the scope of the course.” Some students even created their own datasets by collecting

Table 2: Response of the students to the Likert agreement scale questions.

Topic	Statement	Strongly Disagree	Disagree	Neutral	Agree	Strongly Agree
(i) Course structure	Overall organization and structure of the course were clear and easy to follow	0%	0%	10%	62%	29%
(ii) Datathons	Lectures and datathons complement each other	0%	5%	24%	57%	14%
	Datathons encouraged collaboration and discussion among students	0%	0%	14%	43%	43%
	Group participation in datathons enhanced my learning experience and made problem-solving easier	5%	10%	10%	52%	24%
(iii) Individual project	The lectures and datathons helped me to develop the skills necessary to undertake individual projects independently	0%	0%	29%	48%	24%
	The individual project provided a valuable practical experience that reinforced my understanding of course concepts	0%	0%	10%	48%	43%
	The individual project allowed me to explore topics of personal interest within the scope of the course	0%	0%	0%	52%	48%
(iv) Skills and tools	The course equipped me with the skills and knowledge required to analyze various datasets	0%	0%	5%	71%	24%
(v) Career aspirations	This course inspired me to pursue a career in data science and data visualization	0%	14%	48%	33%	5%
(vi) Student confidence	I am confident in my ability to use Python for analysis and visualization in the future	0%	0%	14%	24%	71%
	I am confident in my ability to use Tableau/Power BI for analysis and visualization in the future	0%	0%	29%	62%	10%

Table 3: Responses to the multiple choice question “Which specific modules of the course do you feel most confident in after completing the course?” with choices corresponding to the six course modules.

Modules	Responses
Data collection and storage	43%
Data cleaning and integration	71%
Data visualization	90%
Correlation and regression	19%
Clustering and classification	19%
Text analysis and text visualization	5%

data from various online sources and literature, demonstrating their ability to independently gather and manage data.

(iv) Skills and tools: The course aimed to provide students with essential data science and data visualization skills and tools to analyze and visualize data. In the survey, 95% of the students agreed with the statement, “The course equipped me with the skills and knowledge required to analyze various datasets.”

(v) Career aspirations: The course was developed for fourth and fifth-year students who were close to graduating within 1-2 terms and pursuing a career. When asked to respond to the statement, “This course inspired me to pursue a career in data science and data visualization”. Responses were mixed, with 33% agreeing and 5% strongly agreeing that the course inspired them to pursue a career in these fields, while 14% disagreed, and almost half of the students remained neutral. While the course successfully motivated a substantial portion of participants, it’s understandable that not all students would envision these fields as their future careers after a single course. Career decisions often evolve over time and are influenced by various factors beyond academic exposure. Encouraging even a fraction of students to consider these career paths can be seen as a positive outcome.

(vi) Student confidence: Throughout the course, students were taught various tools for analysis and visualization, including Python, Tableau, and PowerBI. When asked to respond to the statements, “I am confident in my ability to use Python for analysis and visualization in the future” and “I am confident in my ability to use Tableau/PowerBI for analysis and visualization in the future,” stu-

dents expressed greater confidence in their ability to use Python (96%) compared to Tableau or PowerBI (72%). This could also be due to the fact that a few students had prior knowledge and hands-on experience with Python, whereas Tableau and PowerBI were new to most of them. In addition to these questions, we also asked a multiple-choice question, “Which specific modules of the course do you feel most confident in after completing the course?” (Table 3). Students expressed that they are most confident in data visualization, with 90% feeling assured. However, confidence was low (19%) in correlation and regression, as well as clustering and classification modules. The lowest confidence was in text analysis and visualization, with just 5% feeling confident. The lower confidence in these topics can be attributed to the complex statistical and machine-learning concepts involved. Moreover, covering a broad range of topics within a term by combining data science with data visualization may have impacted the depth of coverage and practical experience gained in these advanced areas.

5 DISCUSSION

We analyzed the results of our survey and identified 4 challenges and 5 opportunities in teaching data visualization together with data science in one course (Table 4).

5.1 Challenges

CH1: Difficulty in learning and using multiple tools in a single term- To provide students with essential data science and data visualization skills, it was necessary to teach a wide range of tools. The objective of this introductory course was to give students a basic knowledge of data visualization tools to inspire them to explore further if they were interested. However, due to the short timeframe, students found it challenging to learn the tools and effectively utilize them. For example, one student mentioned, “I think personally grasping how to use the tableau was difficult. Lots of usefulness in the program, but maybe it was only for me, but trying to take the visualization ideas in my head to tableau was hard” (P21). Students also encountered difficulties when using these tools for their individual projects. As one student stated, “Challenges with the individual project involved using Tableau/PowerBI effectively to visualize the data” (P10). While the goal of the project was to give them an

Table 4: Challenges and opportunities in teaching data visualization together with data science.

Tag	Challenges and Opportunities
CH1	Difficulty in learning and using multiple tools in a single term
CH2	Heterogeneous proficiency with tools and libraries
CH3	Difficulty learning data science-related topics (i.e., AI, ML, and data visualization) in one course
CH4	Challenges in selecting and cleaning a dataset
OP1	Making the course structure clear and updating the content according to student needs
OP2	Introducing large real-world datasets for course assignments
OP3	Learning from industry professionals
OP4	Enhancing learning through collaboration
OP5	Emphasizing visualization literacy and making interactive visualizations early in the course

overview of the entire data analysis pipeline, from data collection to visualization, it was not meant to evaluate their ability to create advanced visualizations. However, students became interested in delving deeper into the tools as they worked on their projects, wanting to create more advanced visualizations. This posed a challenge for them within the given time, “[challenges I encountered while working on my individual project was] mainly figuring out how to do more complex operations within tableau to create the visualizations that I envisioned” (P14). These experiences illustrate the tension between the introductory nature of the course and the students’ desires to explore further. Echoing the results of previous research [20], our survey results also highlight the struggle in learning and using multiple data analysis and data visualization tools in one term. Grouping students into different groups where they can choose their tools and learn about them could be a solution. This approach has also been offered in previous studies [14].

CH2: Heterogeneous proficiency with tools and libraries- When developing the course, one of our goals was to ensure that every student could learn the various tools and libraries, regardless of their prior knowledge. The datathons were structured to guide students through tasks that ranged from operations, such as uploading data to Google Colab, to more complex processes involving statistical analysis and visualization. For some students, the tasks seemed too simple and basic, as they were already familiar with these operations. For instance, they commented that “Many datathons were fairly easy” (P18), and “Some of the datathons seemed repetitive” (P4). On the other hand, for students who were new to Python and data visualization, the same tasks posed a challenge. They felt overwhelmed at times and suggested: “. . . making the datathons slightly shorter would make it easier” (P10). Despite receiving detailed explanations for advanced concepts, some students felt that they needed more time to understand them effectively. This reflects the challenge of varying levels of complexity each student faced with the same tools and libraries. This challenge echoes previous research, which emphasized that methods effective for one group of learners may not be as effective for another [1] and that students often show varying levels of proficiency with visualization tools [23]. Offering free-form courses where students can choose their projects, libraries, and tools depending on their proficiency could be one solution [19].

CH3: Difficulty learning data science-related topics (i.e., AI, ML, and data visualization) in one course- Our course covered a wide array of topics in data science, ranging from foundational statistics such as mean, median, and mode to advanced topics such as ML and AI, along with data visualization. Many students struggled with the advanced topics. e.g., “Some of the machine learning stuff was a bit challenging” (P9), reflecting the difficulty they encountered when delving into ML algorithms and techniques. Moreover, certain ML and AI topics require knowledge of statistics. Although

students typically learn foundational courses such as statistics and mathematics in their first year, they often lose familiarity over time. Similarly, previous studies showed students from non-computer science backgrounds without a strong mathematical foundation found some data science topics difficult to understand [27]. In our survey, one student mentioned, “Some of the correlation and regression concepts were initially hard to grasp without a refresh on basic statistics” (P16). Students expressed interest in refreshing their understanding of some statistics and mathematical concepts within the course. Even though basic statistics were covered in the course, it’s not feasible to cover all the foundational math and statistics in the given timeframe. Doing so would have left even less time for other topics. This reflects the challenge of teaching and learning a wide range of data science topics, especially the advanced ones, along with data visualization. Data visualization itself is a complex and advanced topic, and learning it with other essential data science topics could add extra difficulties. While it might be impractical to require students to take a data science course before enrolling in a data visualization course, dedicating a few sessions at the beginning of the term to assess students’ proficiency in foundational topics and refresh those topics as necessary could support students in learning advanced topics throughout the term.

CH4: Challenges in selecting and cleaning a dataset- An important step for learning how to design and develop a data visualization is to create a data visualization from scratch, find a dataset, or collect data and clean the data. Despite covering topics such as data selection and cleaning in the course, students encountered challenges in finding appropriate datasets for their individual projects. The abundance of available datasets often leads to choice paralysis, where students find it difficult to pinpoint one that aligns with their interests. As one student noted, “finding what dataset interested/that I wanted to visualize was hard. With lots of datasets available, it was hard for me to differentiate what would be best for me to select” (P21). Moreover, even after selecting a dataset of interest, students frequently faced issues concerning its quality and size. Some datasets were too small to derive meaningful analysis from, e.g., “The dataset I chose was not large enough to extract more than 6 or so insights, I was not able to overcome this issue” (P6). Additionally, some students found their chosen datasets to be excessively complex and filled with inaccuracies, requiring substantial time for thorough cleaning and organization before they could commence their analysis and visualization tasks. One student expressed their challenge, “How to handle mistakes in real datasets?” (P8). Despite their prior experience with partially cleaned datasets in datathons and class activities, students encountered difficulties when dealing with messy real-world data they collected. This aligns with previous research emphasizing the complexities of data acquisition and preparation in project-based data visualization education [15]. This underscores the challenge of working with curated/partially cleaned datasets in tutorials and handling the intricacies of raw data sourced from real-world scenarios. It is important for the community to curate a list of recently collected datasets that could serve as practice for students in data science and data visualization courses. These datasets should have an appropriate level of complexity for learning purposes, not to overwhelm students, and be exciting at the same time.

5.2 Opportunities

OP1: Making the course structure clear and updating the content according to student needs- To equip students with essential data science and data visualization skills, it is crucial to cover a broad range of topics in both fields. However, the need to complete the course within one term introduced time constraints. Therefore, efforts were taken to provide a clear and easy-to-follow structure, ensuring students could keep up with the topics without feeling overwhelmed. Students appreciated this course structure, noting, “The course was well structured and covered a wide range of relevant

topics related to the field” (P16). Another student expressed, “*It was honestly the most enjoyable course I’ve taken in my university career. For the first time, I didn’t dread having to attend classes*” (P14). This highlights the importance of making the course structure clear to the students early in the course to set the right expectations for students. During the term, we got informal feedback from students on different aspects of the course and made changes to the ways we presented the content as necessary. One student highlighted, “*I really appreciate feedback on the go and updating the course and datathons based on what students would like to learn*” (P19). This underscores the importance of flexibility and adapting the content of the course based on student’s skills and needs. Future courses should prioritize making the course structure clear to the students while also being adaptable to change the content to student feedback, as highlighted in our survey with students and also in previous research in visualization education [1,26] With students coming from various prior knowledge levels, it’s essential to gather feedback continuously and update the course content based on students’ needs.

OP2: Introducing large real-world datasets for course assignments- Students highly valued the practical experience gained from datathons, as they provided real-world applications beyond theoretical knowledge. One student noted, “*Datathons proved a ‘real-world’ use case beyond the theoretical. We had the chance to start from a messy dataset and develop it into a clean visualization*” (P11). While datathons aimed to simulate real-world scenarios, they often utilized small datasets to help them with the learning curve. This approach was made to ensure that students new to data science and data visualization could learn foundational concepts without being overwhelmed by the complexity of handling vast amounts of data. However, students expressed a growing interest in working with larger datasets to gain more real-world hands-on experience. One student suggested, “*Maybe introduce larger datasets ... Maybe also explore some pipeline aspects (e.g., pulling data from AWS, Azure)*” (P14). Future data science and visualization courses should consider utilizing large real-world datasets, especially with students who have a background in engineering and programming. Integrating larger datasets into future courses not only better prepares students for real-world scenarios but also familiarizes them with the complexities involved, such as scalability, readability, and the substantial computing power often required [7,24]. For students from non-engineering or programming backgrounds, caution should be exercised when introducing large datasets.

OP3: Learning from industry professionals- Guest lectures were incorporated into the course to provide students with diverse perspectives from industry, government, and academia. In regular lectures, students gain knowledge of specific topics and theoretical foundations for data science and visualization. However, guest lectures bring an added dimension to the concepts they’ve learned in class [17]. Guest lectures from industry professionals are particularly beneficial in data science and data visualization courses as these skills are highly demanded in various industries. Firsthand insights into industry practices can bridge the gap between theory and practice, highlighting the evolving demands and applications in the field. One student highlighted, “*I also really enjoyed the guest lectures, having people in the industry talk to us from both the private and city of Ottawa was cool*” (P21). While the guest lectures were generally well-received, some feedback indicated variability in their effectiveness e.g., “*Some guest lecturers were better than others. For example, the individuals from [data analytics company] were very good, but the one about making presentations was somewhat repetitive to things we have already been taught in previous courses*” (P18). The feedback highlights the importance of guest lectures and students’ interest in learning directly from industry experiences rather than solely focusing on acquiring specific skills related to student success. This aligns with previous research indicating that students have shown interest in introducing industry guest lectures

into the mass communication curriculum [21]. Thus, future courses should prioritize guest lectures from professionals who are working with real-world data, perhaps in industrial settings. We also acknowledge that it could be challenging to bring industry people into the classroom, and alternative methods, such as virtual guest lectures or industry collaboration site visits, can be explored.

OP4: Enhancing learning through collaboration- Collaboration among students is helpful in exchanging ideas and approaches, enhancing their learning experience [2]. In this course, all the datathons and many class activities were conducted in groups to emphasize the importance of collaboration and to make problem-solving easier. One student appreciated this aspect, saying, “*Working in groups allows us to compare ideas and approaches; which visualizations work and which do not*” (P11). Collaborative learning in datathons helped students peer review each other’s work, provide comments, and inspire one another. They also supported each other during their individual projects. When one student was stuck, another could offer assistance, fostering an environment of collaborative learning. For instance, one student shared, “*...find[ing] good insights for the project [was challenging]... I overcame it by getting my peers’ opinions on graphs I formed and explained*” (P5). However, some students expressed a desire to work alone at times while still having access to peer support. One student remarked, “*A lot of the datathons were more beneficial to complete individually (the group aspect was beneficial to ask questions if we got stuck, but we always each did it individually)*” (P13). Although research in data visualization education supports and encourages group work for active learning and exposure to diverse student work [3,6], feedback from our students highlights the importance of providing both group work and individual tasks. For a successful data science and data visualization project, a variety of skills, such as design, development, and engineering, are necessary, making collaboration useful. However, it can be challenging to divide tasks properly and fairly and to ensure all group members learn everything. Future courses should incorporate both collaborative activities and opportunities for students to work independently. This approach will provide a comprehensive understanding, allowing students to benefit from peer insights while also developing their individual skills and confidence in tackling data science and data visualization challenges.

OP5: Emphasizing visualization literacy and making interactive visualizations early in the course- It’s becoming increasingly important to teach visualization literacy, which involves the ability to read, interpret, and create data visualizations [4]. It goes beyond simply understanding data representations and includes the capacity to critically evaluate, interact effectively, and thoughtfully design and develop data visualizations [5]. Although the course was structured to cover data visualization topics in the third and fourth weeks, students expressed a strong desire to learn these skills earlier. One student emphasized, “*Have all the ‘How to make good data visualization’ classes at the beginning.*” This feedback highlights the need for a knowledge of visualization literacy from the start. During the course, students showed an interest in learning about good practices in creating ‘right’ data visualizations. Due to student interest, we added a lecture on “Good practices in creating ‘right’ data visualizations” midway through the course, during the datathon, whenever time allowed. Additionally, students expressed a desire to learn how to create more aesthetically pleasing and dynamic visuals, with one student noting, “*Teach us how to make more pretty and dynamic visuals*” (P1). These insights highlight the strong interest students have in learning how to make interactive data visualizations. Previous research has also emphasized the importance of supporting the development of skills and knowledge in interactive exploration [1]. Future courses should aim to cover these topics more thoroughly and emphasize them early on, ensuring that students are well-equipped to create effective visualizations from the start.

6 CONCLUSION

When teaching data visualization to university-level students who lack prior knowledge of data science, it is essential to incorporate data science concepts into the visualization curriculum. Thus, our aim was to develop and evaluate the effectiveness of teaching a course that combines data science and data visualization. We conducted a survey with the students enrolled in the course. In this paper, we described the development of a course that combines data science and data visualization and shared the survey results. Based on our results, we identified four challenges in teaching data visualization together with data science. We also distilled five opportunities for enhancing future courses focused on teaching data visualization together with data science. We hope that the challenges and opportunities highlighted in our study will provide valuable insights for educators and researchers in the visualization community. However, we acknowledge that these challenges and opportunities are specific to the context of teaching data visualization together with data science in a single course and may not be generalizable to other courses.

ACKNOWLEDGEMENTS

We sincerely appreciate the valuable input of all survey participants. Additionally, we acknowledge the generous support provided by Carleton University to conduct this research.

REFERENCES

- [1] B. Bach, M. Keck, F. Rajabiyazdi, T. Losev, I. Meirelles, J. Dykes, R. S. Laramée, M. AlKadi, C. Stoiber, S. Huron, C. Perin, L. Morais, W. Aigner, D. Kosminsky, M. Boucher, S. Knudsen, A. Manataki, J. Aerts, U. Hinrichs, J. C. Roberts, and S. Carpendale. Challenges and opportunities in data visualization education: A call to action. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):649–660, 2024. doi: 10.1109/TVCG.2023.3327378
- [2] Z. Beasley, A. Friedman, L. Pieg, and P. Rosen. Leveraging peer feedback to improve visualization education. In *2020 IEEE Pacific Visualization Symposium (PacificVis)*, pp. 146–155, 2020. doi: 10.1109/PacificVis48177.2020.1261
- [3] Z. J. Beasley, A. Friedman, and P. Rosen. Through the looking glass: Insights into visualization pedagogy through sentiment analysis of peer review text. *IEEE Computer Graphics and Applications*, 41(6):59–70, 2021. doi: 10.1109/MCG.2021.3115387
- [4] K. Börner, A. Bueckle, and M. Ginda. Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences*, 116:1857 – 1864, 2019.
- [5] J. Boy, R. A. Rensink, E. Bertini, and J.-D. Fekete. A Principled Way of Assessing Visualization Literacy. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):10, Nov. 2014. doi: 10.1109/TVCG.2014.2346984
- [6] M. Burch and E. Melby. What more than a hundred project groups reveal about teaching visualization. *J. Vis.*, 23(5):895–911, oct 2020. doi: 10.1007/s12650-020-00659-6
- [7] D. Burlinson, M. Mehedint, C. Grafer, K. Subramanian, J. Payton, P. Goolkasian, M. Youngblood, and R. Kosara. Bridges: A system to enable creation of engaging data structures assignments with real-world data and visualizations. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education, SIGCSE '16*, p. 18–23. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2839509.2844635
- [8] R. Content. 45 ways to communicate two quantities. <https://rockcontent.com/blog/45-ways-to-communicate-two-quantities/>, 2012.
- [9] A. Diehl, E. E. Firat, T. Torsney-Weir, A. Abdul-Rahman, B. Bach, R. Laramée, R. Pajarola, and M. Chen. VisGuided: A Community-driven Approach for Education in Visualization. In B. Sousa Santos and G. Domik, eds., *Eurographics 2021 - Education Papers*. The Eurographics Association, 2021. doi: 10.2312/eged.20211003
- [10] A. Eltweri, A. Faccia, and N. Sawan. The importance of big data visualisations for auditors’ decisions. In *Proceedings of the 2021 5th International Conference on Software and E-Business, ICSEB '21*, p. 7–12. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3507485.3507487
- [11] J. Fuchs, P. Isenberg, A. Bezerianos, M. Miller, and D. Keim. EduClust - A Visualization Application for Teaching Clustering Algorithms. In M. Tarini and E. Galin, eds., *Eurographics 2019 - Education Papers*. The Eurographics Association, 2019. doi: 10.2312/eged.20191023
- [12] L. A. Goodman. Snowball sampling. *The Annals of Mathematical Statistics*, 32(1):148–170, 1961. doi: 10.1214/aoms/1177705148
- [13] R. Hartson and P. Pyla. *The UX Book: Process and Guidelines for Ensuring a Quality User Experience*. Morgan Kaufmann Publishers Inc., Elsevier, San Francisco, CA, USA, 2nd ed., 2018. doi: 10.1016/C2013-0-19285-9
- [14] M. Hedayati and M. Kay. “choose-your-own” d3 labs for learning to adapt online code. In *2023 IEEE VIS Workshop on Visualization Education, Literacy, and Activities (EduVis)*, pp. 49–54, 2023. doi: 10.1109/EduVis60792.2023.00014
- [15] D. Kammer, E. Stoll, and A. Urban. Experience of teaching data visualization using project-based learning. *CoRR*, abs/2111.04428, 2021.
- [16] S. Knudsen, M. B. Bennetsen, T. K. Høj, C. Jensen, R. L. N. Jørgensen, and C. S. Loft. Reflections on designing and running visualization design and programming activities in courses with many students. In *2023 IEEE VIS Workshop on Visualization Education, Literacy, and Activities (EduVis)*, pp. 55–64, 2023. doi: 10.1109/EduVis60792.2023.00015
- [17] B. R. Krogstie and J. Krogstie. Guest lectures in it education-recommendations based on an empirical study. 2018.
- [18] S. Kross and P. J. Guo. Practitioners teaching data science in industry and academia: Expectations, workflows, and challenges. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, p. 1–14. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3290605.3300493
- [19] B. Langstone, A. Ouskine, C. Haberl, and F. Rajabiyazdi. Critical reflections on teaching data visualization free form, 2022. Medium.
- [20] L. Y.-H. Lo, Y. Ming, and H. Qu. Learning vis tools: Teaching data visualization tutorials. In *2019 IEEE Visualization Conference (VIS)*, pp. 11–15, 2019. doi: 10.1109/VISUAL.2019.8933751
- [21] P. F. Merle and C. Craig. Be my guest: A survey of mass communication students’ perception of guest speakers. *College Teaching*, 65(2):41–49, 2017. doi: 10.1080/87567555.2016.1232691
- [22] K. Mike, B. Kimelfeld, and O. Hazzan. The Birth of a New Discipline: Data Science Education. *Harvard Data Science Review*, 5(4), oct 27 2023. <https://hdr.mitpress.mit.edu/pub/by22efhx>.
- [23] J. Seo, S. O’Modhrain, Y. Xia, S. Kamath, B. Lee, and J. M. Coughlan. Designing Born-Accessible Courses in Data Science and Visualization: Challenges and Opportunities of a Remote Curriculum Taught by Blind Instructors to Blind Students. In E. E. Firat, R. S. Laramée, and N. S. Andersen, eds., *EuroVis 2024 - Education Papers*. The Eurographics Association, 2024. doi: 10.2312/eved.20241053
- [24] J. Strahler, M. McQuaigue, A. Goncharow, D. Burlinson, K. Subramanian, E. Saule, and J. Payton. Real-world assignments at scale to reinforce the importance of algorithms and complexity. *J. Comput. Sci. Coll.*, 35(8):166–175, apr 2020.
- [25] J. R. Stroop. Studies of interference in serial verbal reactions. *Journal of experimental psychology*, 18(6):643, 1935.
- [26] U. H. Syeda, L. South, J. Raynor, L. Panavas, D. Saffo, T. Morriss, C. Dunne, and M. A. Borkin. Vis Repligogy: Towards a Culture of Facilitating Replication Studies in Visualization Pedagogy and Research. In E. E. Firat, R. S. Laramée, and N. S. Andersen, eds., *EuroVis 2024 - Education Papers*. The Eurographics Association, 2024. doi: 10.2312/eved.20241054
- [27] Y. Velaj, D. Dolezal, R. Ambros, C. Plant, and R. Motschnig. Teaching data science to non-computer science students: A learner-centered approach. In *2023 IEEE Frontiers in Education Conference (FIE)*, pp. 1–9, 2023. doi: 10.1109/FIE58773.2023.10343463