Journal of Data Science 22 (2), 333–351
April 2024
DOI: 10.6339/24-JDS1129
Education in Data Science

Interdisciplinary Approaches to Teaching Communication and Ethics in Data Science: A Case Study

Scott Thatcher[1,*], K. Scott Alberts[1], and Tetyana Beregovska[1]
[1]Department of Statistics, Truman State University, Kirksville, Missouri, U.S.A.

Abstract

By its nature, data science uses ideas and methodologies from computer science and statistics, along with field-specific knowledge, to describe, learn and predict. Recently, storytelling has been highlighted as an important extension of more traditional data science skills such as coding and modeling. Three courses in our new Master in Data Science and Analytic Storytelling program were designed to include interdisciplinary modules, mainly taught by faculty in storytelling-related disciplines, such as Communication and Art & Design. These courses were PDAT 622: Narrative, Argument, and Persuasion in Data Science; PDAT 624: Principles of Design in Data Visualization; and PDAT 625: Big Data Ethics and Security.

Our first cohort serves as a natural case study, allowing us to reflectively analyze our materials and an informal student survey to explore the effects of interdisciplinarity in these novel courses. Results of the student survey show that students generally found value in these interdisciplinary course components, especially in course "signature assignments," which allow students to actively engage with course content while reinforcing technical skills from previous courses. Examples of these signature assignments are presented in this paper's supplementary materials.

Keywords *algorithmic bias; communication; design; ethics; program assessment; rhetoric; storytelling; visualization*

1 Introduction

Truman State University's 30-credit masters program in Data Science and Analytic Storytelling, launched in Fall 2021, is designed to be accessible for working professionals, using 8-week block terms, with online "low-synchronous" courses. Courses feature recorded lectures and weekly assignments, but also include "strongly recommended" weekly, hour-long instructor-led Zoom sessions where students can talk through issues in their course work and build a sense of camaraderie within their cohort.

The first block of courses, leading to a fifteen-credit Graduate Certificate in Data Science, focuses on a fairly standard set of technical topics, including an introduction to the use of tools such as R and Tidyverse, an introduction to big data management, and courses in data mining and machine learning. A second block of courses builds on that technical foundation to yield a Masters in Data Science and Analytic Storytelling, including two more electives that may concentrate on technical topics, in addition to three interdisciplinary courses whose content is the focus of this paper:

*Corresponding author. Email: thatcher@truman.edu.

Received August 1, 2023; Accepted April 4, 2024

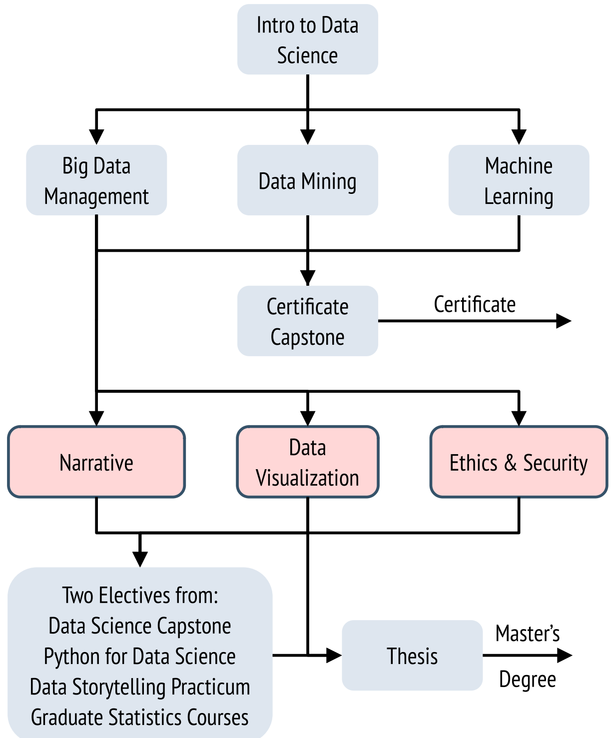

Figure 1: Flowchart of data science coursework showing the three featured interdisciplinary courses during the second half of the program.

- PDAT 622: Narrative, Argument, and Persuasion in Data Science (abbr. "Narrative"),
- PDAT 624: Principles of Design in Data Visualization (abbr. "Data Visualization"), and
- PDAT 625: Big Data Ethics and Security (abbr. "Ethics" or "Ethics & Security").

An independent thesis, which combines a technical project with storytelling components that target technical and non-technical audiences, completes the program (see Figure 1).

Two courses, Narrative and Data Visualization, were designed in collaboration with colleagues in Communication and Art & Design, respectively. Ethics & Security was designed to cover topics in ethics and data security from multiple disciplinary perspectives by faculty members within the Statistics Department who had additional training in Economics and Organizational Behavior.

These topics in communication and ethics are consistently mentioned in national curriculum guidelines and by local industry partners. Because they represent skills that continue to be refined throughout a professional's career, we sought collaboration from faculty at Truman whose expertise could give students a solid initial foundation from which to grow, in the tradition of Truman's liberal arts mission. Our purpose in this paper is to describe the collaborative process we engaged in when designing three courses that cross disciplinary boundaries, present results of an initial evaluation of student responses, and offer sample course materials that we hope may be helpful to other educators.

2 Program Background

2.1 Curriculum Guidelines

During the planning of the certificate and masters programs, the program committee consulted curriculum guidelines for statistics and data science programs, as well as published papers on industry needs. Each identified multiple areas of technical knowledge that are important for statistics and data science professionals. For example, De Veaux et al. (2017) identified broad areas in computational and statistical thinking, mathematical foundations, model-building and assessment, algorithms and software, and data curation in their data science curriculum guidelines. A data-scraping approach applied to data science job postings emphasized the importance of analytical tools and methods, implementation of big data technologies, and software development and management, among others (Radovilsky et al., 2018).

The courses in the first half of our program were designed to address many of these topics, providing an introduction to data cleaning and management, statistical thinking and traditional statistical methods, big data management, and algorithms for modeling and clustering. Later courses, such as a project-based practicum course (PDAT 626) and the thesis project, help students build skills in management, revision, and collaboration on larger coding projects, including related technologies such as Git. Because of the nature of its audience (see Section 2.2), the program does not emphasize mathematical training or software engineering as heavily as other programs might.

The three courses under consideration in this paper were motivated by calls in these guidelines for data communication skills and ethical awareness. For example,

> Graduates should be expected to write clearly, speak fluently, and construct effective visual displays and compelling written summaries.... They should be able to communicate complex statistical methods in basic terms to managers and other audiences and visualize results in an accessible manner. (American Statistical Association Undergraduate Guidelines Workgroup, 2014)

> Effective communication is a core skill of the data scientist. As members of a team, data scientists must communicate to teammates as well as to those with less intimate knowledge of the project particulars. Increasingly, data scientists communicate directly to the public via both static and interactive data visualizations. (De Veaux et al., 2017)

> Programs in data science should feature exposure to and ethical training in areas such as citation and data ownership, security and sensitivity of data, consequences and privacy concerns of data analysis, and the professionalism of transparency and reproducibility. (De Veaux et al., 2017)

Our Data Visualization course covers principles of design and accessibility as well as static and interactive visualization techniques. The Narrative course focuses on techniques of effective communication in the context of varying audiences. Also, in light of the fact that some sources have identified ethics as being important, but often relegated to only an elective area of study (Miller, 2014), we were motivated to require Ethics & Security of all students.

2.2 Program Audience and Background

Every program must choose its coverage of topic areas based not only on general guidelines in the field, but also on the needs of its student and industry audiences. Truman's Professional

Masters in Data Science and Analytic Storytelling is designed to meet the needs of undergraduates and working professionals who do not already have an extensive undergraduate background in statistics or data science. Program prerequisites include only introductory courses or work experience in statistics and computer science. Furthermore, local industry partners have expressed interest in a program of this type from which they can recruit or to which they could send their employees.

For example, a local software consultancy group has told us that they value consultants who are able to understand the needs of multiple constituencies and tailor their software solutions and communication accordingly: "What does my client need? What does my manager need?" Similarly, data scientists from a regional insurance company have told our students that perhaps only 20% of a project might involve creation of a "fancy" machine learning model, while much of the other 80% involves understanding where there is a problem to be solved and communicating with stakeholders to build consensus around a solution. Michael Terry, former Director of Data Strategy at Shelter Insurance, describes his experience as follows:

> The field of data analytics has evolved so quickly that understanding, cleaning, and analyzing data are entry-level skills. In order to advance, data analysts need to be able to talk about their work in a way that speaks to audiences at a variety of skill levels. The ability to put yourself in the place of your audience and anticipate their objections and questions is one of the hardest skills for an analyst to acquire, and the Truman students I've worked with came already trained in those skills.

For these reasons, a general design philosophy for all our courses has been to build data science skills from the ground up, emphasizing the provision of a strong conceptual framework for the material presented, as well as meaningful opportunities to apply skills to authentic data sets and problems. While we do acknowledge that our program does not take students as far in specific technical areas as would a masters program that assumes a related undergraduate major, we believe a graduate of our program should be able to do meaningful work in their field and should be able to confidently communicate and collaborate with others in the data science arena.

These considerations were one reason that we chose to emphasize interdisciplinary collaboration in Data Visualization, Narrative and Ethics. We saw these courses as opportunities to help students build a deeper conceptual foundation in these fields through the involvement of accomplished practitioners in those fields.

We believe this approach is consistent with the general approach taken by liberal arts institutions, where it has been argued in the context of the sciences that the acquisition of broad knowledge and critical thinking skills allows students to attain success in further graduate study (Cech, 1999). Furthermore, we believe our program is well situated to train professionals who plan to work in positions adjacent to data science, a need identified in sources such as Miller (2014). Table 1 shows employment data for our first graduating cohorts.

2.3 Comparison with Other Programs

To place these design decisions in context, it may be helpful to compare Truman's program to several other data science masters programs that advertise themselves as interdisciplinary in nature.

Because they serve a different audience, some programs look very different from Truman's. For example, MIT offers an interdisciplinary program in Computer Science, Economics and Data

Table 1: Employment data for graduates of our first cohorts. Most were professional students returning to update their skills, but three in this group had come to the program straight from their undergraduate degree.

Sector	Professional	Undergraduate	Total
Banking/Finance	4	0	4
Health/Pharmaceutical	3	1	4
Insurance	2	1	3
Utilities	2	0	2
Technical Writing	2	0	2
Education	1	0	1
Police Data Analyst	1	0	1
Transportation/Logistics	0	1	1

Science (MIT, 2024), which builds on an undergraduate degree of the same name. By the end of this five-year program, a student will have seen significantly more technical content than could be covered in a masters program aimed at an audience with more varied backgrounds. In this program, interdisciplinarity appears to refer to the explicit combination of topics from the three named disciplines, and data communication courses are offered as electives.

Duke, Southern Methodist and Syracuse Universities offer programs to which Truman's is perhaps more comparable (see Table 2). Program materials advertise these programs for a wide audience and stress the value of a student body with a variety of previous experience. All of these programs replace strict course-based program admission requirements with a more flexible requirement to demonstrate potential for success in some combination of quantitative, analytical and technical areas. Ability can be demonstrated through formal coursework or through other experience. Truman's requirements and philosophy of admissions are similar.

Despite these similarities, interdisciplinarity is manifested differently in each program. Duke's program materials state that students will gain both quantitative skills and experience in problem-solving, communication and team-building (Duke, 2024). Like Truman, Duke includes courses that specifically focus on ethics, communication and visualization.

Southern Methodist's program materials emphasize the interdisciplinarity inherent in the statistics, computer science and visualization aspects of data science, and offer specializations in machine learning or business analytics (SMU, 2024). "Visualization of Information" is offered as a required course, and other courses and electives appear to focus on technical topics. Courses that center on ethics and narrative are not listed.

Syracuse emphasizes "critical thinking and problem-solving skills" in its materials and offers interdisciplinarity through a wider set of tracks and electives, including AI, Business Analytics, Pipelines and Platforms, and Language Analytics (Syracuse, 2024).

Although a limited review of other programs can't speak definitively about broader patterns, it can be noted that ethics, data visualization and narrative/storytelling are recognized as important in other programs. However, it is also the case that among the programs examined here, Truman is unique in requiring all three courses in these areas. Table 2 also illustrates the trade-off inherent in requiring those three interdisciplinary courses: each other program requires at least one more course in a technical subject.

Table 2: A summary of common program topics represented by primary program course titles. Primary topics of all required courses are listed here. Electives that clearly fall within this list are also noted, but not all program electives are listed. It might be expected that topics that are not the primary focus of a course could be covered to a greater or lesser extent in other program courses. However, of the programs listed here, only Truman requires separate courses in ethics/security, visualization and narrative.

	SMU	Duke	Syracuse	Truman
Credits	33.5	42	34	30
Statistical Foundations	✓		✓	✓
Applied Statistics	✓	✓		E
Introduction to Data Science	✓	✓	✓	✓
Big Data Management	✓	✓	✓	✓
Big Data Analytics	✓		✓	
Machine Learning I	✓	✓	✓	✓
Machine Learning II	E		E	✓
Natural Language Processing	E	✓	E	
Business Analytics	E		✓	
Time Series	E			
Internship		✓		
Management Science			E	
Data Security	E		E	C
Data Ethics		✓	E	C
Data Visualization	✓	C		✓
Narrative/Storytelling		C		✓

✓: Required E: Elective C: Combined

Note: Duke teaches "soft skills" through a series of workshops and has a wider list of electives, which aren't noted here.

2.4 Interdisciplinary Course Outcomes

Learning outcomes for Truman's three interdisciplinary courses were written in collaboration with our colleagues from other departments. Table 3 lists these course-wide outcomes. Further details on each course will be given in the next section.

3 Course Design and Signature Assignments

3.1 Narrative, Argument, and Persuasion in Data Science

In PDAT 622: Narrative, Argument, and Persuasion in Data Science, faculty from Data Science and Communication collaborate to present fundamental ideas from rhetoric in a form appropriate for data-driven fields (see Table 4). Led by Communication and Data Science faculty, students read and discuss modern rhetorical literature, including works by Bitzer (1968), Fisher (1984), Rowland (1987), Toulmin (2003), and Orwell (1946). These concepts are then made specific to data science through a "signature" assignment approach, where students iteratively explore the same data set in a variety of contexts.

Table 3: Outcome statements for PDAT 622, PDAT 624, and PDAT 625.

PDAT 622: Narrative, Argument, and Persuasion in Data Science

By the end of this course, the successful student will be able to:

- respond to various rhetorical situations and audiences in order to present compelling arguments supported by data;
- adapt arguments created for specific audiences or situations to new audiences or situations;
- create persuasive visualizations that support data-driven arguments; and
- demonstrate an advanced level of proficiency in R coding using the Tidyverse collection of packages, building on skills from earlier PDAT courses.

PDAT 624: Principles of Design in Data Visualization

On successful completion of this class, students will:

- utilize visual thinking and ideation;
- critique and practice the linking between form (what is visually presented) and content (what the viewer experiences), and how these elements interact with context;
- identify and apply the elements of design (line, shape, color, scale, motion, etc.) in the display of data;
- apply the principles of design (composition, gestalt, theme and variation, etc.) in telling visual stories;
- identify and analyze the effect of historical factors, cultural milieu, political context, variously abled audience experiences, etc. and their impact on the way a message is received;
- practice moving between hand-rendered and digital formats, appreciating the strengths of each in the design process;
- develop competence in the creation of both static and dynamic multivariate information displays, ranging from individual graphs to interactive data dashboards, using tools such as
  - R and R packages such as ggplot2 for the creation of static graphs,
  - other packages, such as plotly and leaflet, for creating interactive graphs,
  - R shiny for the creation of interactive data dashboards and applications, and
  - web-based visualization tools like plot.ly and tableau; and
- identify and describe the ethical dimension of data visualization design choices.

PDAT 625: Big Data Ethics and Security

By the end of this course, a successful student will:

- identify issues important to privacy, security, and safety in the age of Big Data;
- describe algorithmic decision making and its benefits and challenges;
- identify and characterize sources of bias in real-world scenarios and enumerate approaches to reducing algorithmic bias;
- explore techniques and methods for keeping Big Data secure and private; and
- discuss data protection standards and considerations in selecting, adopting, and complying with them.

Table 4: Topics in PDAT 622: Narrative, Argument, and Persuasion in Data Science.

Module	Topic
0	Introduction to Analytical Storytelling (half-week)
1	The Rhetorical Situation
2	The Elements of a Persuasive Narrative
3	Translating Data into Logical and Persuasive Arguments
4	Adapting Arguments for Different Audiences
5	Visualizing Arguments for Different Audiences
6	Using Language to Better Illuminate Arguments
7	Case Studies and Narratives from Data
8	Evaluations and Wrap Up (half-week)

In reading Bitzer's "The Rhetorical Situation" (Bitzer, 1968), students examine the interaction of writer, subject and audience. Bitzer contends that communication requires an understanding of why the writer needs to communicate on a particular subject (exigency), what the writer hopes to achieve through communication with a particular audience (purpose), and how that particular subject can best be presented to that audience (genera).

Further readings introduce methods of argumentation, including various types of reasoning (inductive, deductive, syllogistic, and, specifically, statistical). Students first practice finding the flaws in others' arguments, and then turn attention to their own arguments.

Paralleling the introduction of these concepts, the signature assignment, titled "All Climate is Local," asks students to explore a publicly available National Oceanic and Atmospheric Administration (NOAA) climate data set for a single city of their choice, eventually communicating results to a target audience (see Table 5). In the first iteration, students are simply asked to create a statistical model, along with several data visualizations, reinforcing skills from previous courses. This assignment mirrored many statistics course assignments, where the data is taken for granted, and it is assumed that results speak for themselves. Follow-on assignments then ask students to deepen their analysis, while beginning to apply rhetorical techniques that consider exigency and audience. In the final assignment, students prepare both a 10–12 minute talk and accompanying written materials, strongly and consciously aimed at a specific audience. Possible audiences have included an auditorium of climate skeptics, a city planning board tasked with predicting climate-related changes to infrastructure costs, or even zoologists at the local zoo.

3.2 Principles of Design in Data Visualization

PDAT 624: Principles of Design in Data Visualization includes content on both static and interactive data visualization techniques, split roughly equally in the course plan (see Table 6). Collaboration with faculty in the Art Department focused on bringing students an introduction to the elements and principles of design that was informed by the methods used to teach those topics in introductory art and design courses. Lectures introduce the distinction between content and form, the importance of context, the individual elements of design, and gestalt principles of perception using examples taken from both art and statistics. Readings from the open online textbook *Introduction to Art: Design, Context, and Meaning* (Sachant et al., 2016) supplement the lectures.

Several assignments, inspired by those used in introductory design courses, encourage students to experiment, building a sense for the elements of design and the way they shape percep-

Table 5: The iterative signature assignment in PDAT 622: Narrative, Argument, and Persuasion in Data Science, including both data analysis and focus on audience and purpose.

Assignment	Topics
1	Choice of city for climate data download. Linear model and scatterplot of high and low temperatures. Student-directed exploration. **Product:** 500-word technical report summarizing findings. **Audience:** Local city planning board.
2	Exploration of changes in temperature *variation*. **Product:** 1000-word white paper with suggested course of action, or 1000-word technical paper if no change is deemed necessary. **Audience:** Local city planning board.
3	Exploration of precipitation changes. **Product:** Slides that include previous and current analysis. **Audience:** Local city planning board.
4	Refinement of climate analysis in response to feedback. **Product:** Further refinement of slide presentation. **Audience:** Choice of a different potential audience.
5	Refinement of climate analysis in response to feedback. **Product:** 10–12 minute talk to the intended audience. Written handout to accompany talk.

Table 6: Topics in PDAT 624: Principles of Design in Data Visualization.

Module	Topic
1–4	Static Visualization Elements and principles of good design (Art and Statistics) Advanced *ggplot* Color palettes/accessibility Signature Assignment: Playfair Recreation
5–8	Interactive Visualization *plotly* Geographic information and mapping *shiny* apps Signature Assignment: Shiny Visualization

tion. For example, an introductory design course might ask students to create a collection of lines and shapes, then experiment with choice of color and texture to create final products that give very different overall impressions. To emphasize that those choices are also made in data visualization, an early "deconstruction" assignment asks students to break up a finished data visualization, using its elements to create a new design that explores shape, color and form (see Figure 2).

A companion "construction" assignment gives students a tidy data set with both continuous and categorical columns, and asks them to experiment with several ways of assigning graph aes-

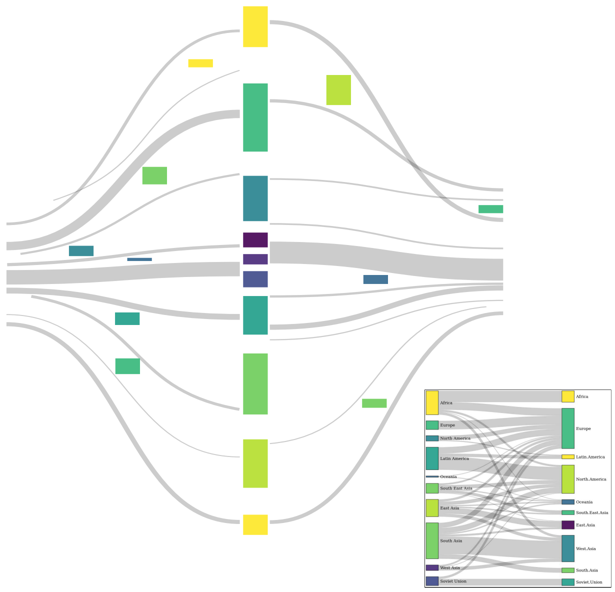

Figure 2: Example of the "deconstruction" assignment, where students are asked to experiment with design elements like color, shape and line by creating a new work that repurposes the elements of a specific data visualization outside its original data-bearing context. Student use their new design vocabulary to comment on their own work and critique examples from their peers. The original Sankey diagram (https://www.data-to-viz.com/graph/sankey.html) is shown in the inset.

thetics to the assorted variables. In both the "deconstruction" and "construction" assignments, students post their work to the class discussion board, and are asked to use the vocabulary they have learned to talk about the successful and, perhaps, less successful aspects of their designs.

Collaboration with art faculty also led to an emphasis on the process of creation and critique. At several points in the class, students are encouraged to consider the differences between creation of data visualization designs "by hand," and creation using technology. In the signature assignment for the first half of the course, students are introduced to some of the early data visualizations by William Playfair (Playfair, 1821), which were created "by hand" before well-defined patterns of data visualization had been created. Students then attempt to (1) use packages like *ggplot* to recreate Playfair's visualizations, and (2) consider new ways to express Playfair's data in effective ways (see Figure 3). One goal of that assignment is to help students think about how technology helps, but also limits, their expression. In the second large signature assignment, which involved creation of a data dashboard, students are first asked to create a pencil-and-paper design so that their design choices will not be limited to only those technological techniques they already know.

Both of these signature assignments feed into live critique sessions, which were another process-oriented tool borrowed from Art and Design. Students (meeting via Zoom in this online class) view a PDF document containing all student submissions for each assignment. Led by Art faculty, students are then asked to vote for three submissions that they feel were most effective in both design and data communication. Discussion of the highest vote-getters follows, focusing on strengths, rather than weaknesses.

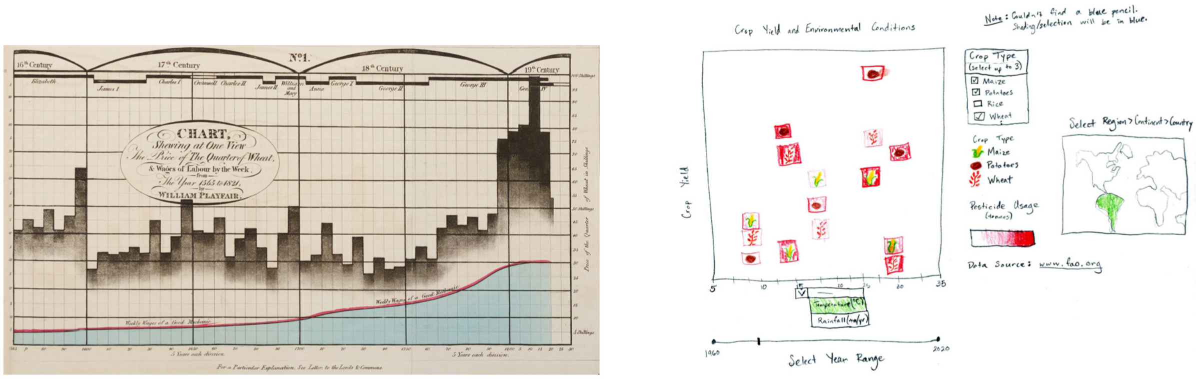

Figure 3: Williams Playfair's "Chart, shewing at one view the price of the quarter of wheat, and wages of labour by the week, the year 1565 to 1821," which is difficult to reproduce using visualization software, and an example of a student pen-and-paper dashboard mock-up.

Table 7: Topics in PDAT 625: Big Data Ethics and Security.

Module	Topic
1	The Balance between Ethics and Human Security
2	Ethics and Big Data
3	Privacy in the Big Data Age
4	Security Questions in the Current Information Landscape
5	Big Data Bias and Implications
6	Networks and Privacy
7	Legal Considerations for Ethics, Privacy and Security
8	Evaluating Your Work against Ethical Standards

3.3 Big Data Ethics and Security

Designed by a faculty member with degrees in both Economics and Statistics, in collaboration with several other Statistics faculty, PDAT 625: Big Data Ethics and Security examines those topics through philosophic, economic, technical and legal lenses. Course outcomes are summarized in Table 3, and week-by-week topics are listed in Table 7.

During the course, students explore the various dimensions of privacy, identified by computer scientist Roger Clarke (Clarke, 1997), one of the first privacy advocates: privacy of person, behavior, communication, and data. Privacy and security are then viewed through an economic lens, with application of principal-agent theory to networks components (apps, software, and hardware) and discussion of the economics of digital markets that trade data for access to services.

The course also introduces bias—statistical, human and algorithmic—and its ethical implications in data science. Students are encouraged to see that all data has context, as a product of particular human activity, and that context affects how the data should be interpreted.

One signature course assignment tackles the problem of unbalanced training data and asks students to reflect on algorithmic bias. The assignment builds on a 2016 *Pro Publica* report on algorithmic bias in predicting recidivism using commercial "black box" risk assessment software

Table 8: Pro Publica's 2016 report "Machine bias" found the following false positive and false negative rates in predicting recidivism for Black and White arrestees. A sample of student work in the algorithmic bias assignment shows a similar pattern. Both groups showed similar overall *accuracy* of 67–70%.

Source	Group	False Pos.	False Neg.	Accuracy
ProPublica	Black	44.9%	28.0%	63%
	White	23.5%	47.7%	59%
Student Work	Black	35.7%	30.2%	∼67%
	White	17.8%	51.6%	∼67%

(Angwin et al., 2016). To evaluate the commercial software (called COMPAS), *Pro Publica* used public records requests to assemble a data set of people arrested in Broward County, Florida between 2013 and 2014, along with their risk assessment scores and other arrest and demographic information. Note that this data was presumably more limited than the data available to COMPAS itself. *Pro Publica* then compared predicted two-year re-arrest based on COMPAS risk score to actual two-year re-arrest rates—the same process used by the makers of COMPAS to evaluate its performance. *Pro Publica*'s report showed that while COMPAS showed similar *accuracy* for Black and White arrestees (59–63%), it had markedly higher *false positive rates* in predicting recidivism for Black arrestees and markedly higher *false negative* rates for White arrestees (see Table 8).

The assignment asks students to first read the *Pro Publica* report to understand the context of the analysis. Then, using the data set from *Pro Publica*, students are asked to create their own predictive models for re-arrest within two years. The assignment reinforces technical skills in model building and dealing with unbalanced training data, but more importantly students are given the opportunity to reflect on the following points:

- Detailed personal information is available in public records requests (*Pro Publica* included names in its data set, but the data set provided to students had names removed).
- Students may be able to achieve better *accuracy* with limited data than the commercial COMPAS algorithm achieved with more detailed arrestee information.
- Sometimes data scientists will grapple with a data set where it is generally not possible to create an unbiased prediction.

4 Course Assessment

In a process separate from regular course evaluations, students from the initial sections of each of these three courses were asked to respond to an anonymous survey which focused on their experience of the interdisciplinary components in each course. The survey was sent out in late December, 2022 to students in the Data Visualization and Narrative courses and in early January 2023 to students in the Ethics course. All courses had ended by the time the surveys were sent out, and no compensation for responses was offered. The survey invitation was sent to 19 students from the first sections of the Data Visualization class (2 students from the spring 2021 pilot, and 17 from the fall 2022 regular section), 21 students from the first two sections of Narrative (spring and fall 2022), and 16 students from the first section of Ethics (spring 2022).

Table 9: Mean, standard deviation, five-number summary and sample size for course component ratings in each interdisciplinary course. Ratings ranged from 1: "Didn't contribute at all" to 6: "Contributed very much."

				Five-Number Summary					
Course	Component	Mean	SD	Min	Q1	Med	Q3	Max	*n*
Narrative	Signature	5.8	0.4	5	6	6	6	5	9
	Critique	5.6	0.5	5	5	6	5	5	9
	Lectures	5.0	1.1	3	4	5	4	3	9
	Rhetoric Readings	4.3	1.1	2	4	4	4	2	9
Data Vis.	Deconstruction	5.3	1.1	3	5	6	5	3	9
	Crit2	5.3	1.0	3	5	6	5	3	9
	Crit1	5.2	1.3	3	5	6	5	3	9
	Lecture	4.9	1.4	2	4	5	4	2	9
	Design Readings	4.4	1.3	2	4	5	4	2	9
Ethics	Model Bias	5.4	1.3	3	6	6	6	3	5
	Privacy Video	5.0	1.0	4	4	5	4	4	5
	Privacy Readings	5.2	0.8	4	5	5	5	4	5
	Origins	5.0	1.0	4	4	5	4	4	5
	Legal Issues Readings	5.2	0.8	4	5	5	5	4	5
	Bias Readings	5.0	1.0	4	4	5	4	4	5
	Video Critique	4.8	1.1	4	4	4	4	4	5

Faculty of each course first identified interdisciplinary course elements to include in the survey. The surveys provided a reminder of the course objectives for each course (see Table 3), then asked the students to rate the extent to which the interdisciplinary course components contributed to achieving these course objectives. Ratings ranged from 1: "Didn't Contribute at All" to 6: "Contributed Very Much." Students were also able to provide open-ended feedback on the interdisciplinary/collaborative aspects of each course.

Table 9 shows a summary of ratings from these student surveys. Sample sizes from the first course sections were small ($n = 9$, 43% resp. rate for Narrative, $n = 9$, 47% resp. rate for Data Visualization classes, and $n = 5$, 31% resp. rate for Ethics). Clearly, any conclusions should be viewed as preliminary, as it can't be guaranteed that the non-response rate did not lead to bias in the results. However, it can be seen that most ratings were positive (ranging from 4 to 6). Furthermore, in each class, the more active course elements—signature assignments and critiques—tended to have the highest mean and median ratings. Readings, on the other hand, tended to have the lowest mean scores.

Tables 10 and 11 provide a selection of open-ended responses from students. Positive comments tended to note the applicability of the skills taught in these courses to career goals, the importance of discussion and interaction, and an appreciation for the signature assignments in each course. Suggestions for improvement noted the challenging readings and a desire for more integration of interdisciplinary material in Data Visualization. One student did note that introduction of more interdisciplinary material in the coursework did reduce time available for technical coverage.

Table 10: A selection of comments on the interdisciplinary aspects of each course.

Narrative	"For my thesis, I drew on the skill [*sic*] here for finding a narrative element to make the data significant." "I liked the class overall and Dr. Alberts [*sic*] climate research made the final project even more interesting. Dr. Minch was incredibly knowledgeable and was a great source of expertise." "I really appreciated the interdisciplinary approach. It fits with Truman's stellar reputation as a liberal arts school and provides an aspect to the degree that is unique and very appreciated in various industries." "It was a lot of reading on very esoteric topics."
Data Vis.	"Collaboration and discussion are a huge part of the graduate school experience... Great opportunity to get feedback on multiple approaches that I may have chosen not to attempt." "[Deconstruction was] a favorite assignment. Key takeaways were both that context and arrangement matter and that visual elements are deliberate choices. As a storyteller, identifying and manipulating these elements are willful acts." "I felt there was a real disconnect between the art topics covered initially and what we ended up working on... In 624, there were readings from the Art dept only the first two weeks."
Ethics	**Modeling Bias Assignment** "This assignment was very challenging. It was absolutely worth the effort." "I would have liked it if the class included more assignments that let us apply what we are learning in regards to the development of predictive models." **Legal Issues** "Understanding the current legal state and any pending or future legislation will be essential for any work we do going forward. This lesson was great." **General** "The interdisciplinary nature of the course provided the perfect framework for understanding big data and how to implement within our code or our workplace. It helped me think critically about the various applications, data be taken [*sic*], and the potential security and ethical issues."

5 Discussion

Generally speaking, the authors of this paper view the introduction of interdisciplinary topics and techniques in our data science courses to be a net positive, but do acknowledge that some challenges remain, especially related to the logistics of the collaboration itself. Initial collaboration in course design required advance planning and coordination, but was made easier by the availability of time and funds to compensate all collaborators. Collaborators have continued

Table 11: A selection of comments on interdisciplinarity, more generally.

Generally Positive Student Comments
"The collaborative content in the advanced courses was outstanding and I found it immediately applicable at work. The holistic approach combining data science with argument and visualization makes this an exceptional program."
"The primary reason why I chose this program was for the analytical storytelling component. I work with data now and one of the most important aspects of my job is presenting results to stakeholders."
"The subject matter experts from other fields really helped tie things together in terms of the actual course material and real world applications… "
"I appreciated the interdisciplinary classes. In the Big Data Ethics and Security class, I appreciated our Ukrainian professor's experience with high government control of big data. Dr. Minch is a great [*sic*] at lecturing and breaking down argument pieces and laying out their importance and structure. Dr. Fine brought a lot of experience to the visualization class, his knowledge on color alone gave me a lot of think [*sic*] about and reflect on."
Student Suggestions for Improvement
"I really enjoyed the interdisciplinary aspect of the courses. I wish the non-DS professors were more interactive with the class."
"They were pretty decent. I just wished we had more examples in order to apply for more [*sic*] on our projects. The classwork was fine to understand but more time would have been helpful."
"A not insignificant drawback is the reduction of technical coverage. It's easier to pick up non-technical knowledge (definition of a rhetorical situation) than a technical skill (implementation of SVM with text classification)… I just wonder if it would be possible to implement the Communications and Art pieces in a separate, pre-matriculation course."

to interact directly with students while the courses have run, but that participation has been *gratis*, and thus not as extensive as it would have been in a truly team-taught environment. For example, students have the benefit of recorded lectures from Art faculty, but may only directly interact during critiques twice a semester.

Students have indicated that they appreciate the applicability of the interdisciplinary components in these courses to their own careers and pursuits, but as instructors, we have also noted that these adult students have brought their own experiences to the table as well. For example, discussions of ethics and data privacy were enriched by students with backgrounds in finance, insurance and law enforcement. Discussions of audience and context have benefited from the wide range of *actual* audiences that students must address in the course of their work. As we continue to evaluate and update the Data Science program, we will keep these student contributions in mind.

Although perhaps it goes without saying, we in Data Science have also appreciated the depth of experience brought by our interdisciplinary collaborators, who can speak with authority on topics we might otherwise treat superficially. For example, after hearing a lecture on color theory

Figure 4: Two statues titled *The kiss*, by Auguste Rodin (Rodin, 1882) and Constantin Brancusi (Brancusi, 1907), respectively. They share the same content ("a kiss"), but illustrate very different form. The visual example helped students articulate this distinction both inside and outside the realm of data visualization.

given by our collaborator in Art, one technically-oriented student expressed admiration for the scientific and mathematical underpinnings of the Munsell color system, which uses empirical data on human visual perception to parameterize the color space into a cylindrical coordinate system (hue, value and chroma), where each dimension is perceptually uniform and independent (Munsell, 1912; Wikipedia, 2023).

Students have also benefited from our collaborators' ability to explain concepts using examples from outside data science. For example, in Data Visualization, the distinction between form and content was illustrated using two sculptures, both titled *The Kiss* (see Figure 4). Later in the course, these visual examples provided a way to discuss the trade-offs between elegance and utility, or between complexity and simplicity. As another example, a discussion in our Narrative course used the song "Go West," as sung by the Village People in 1979, the Pet Shop Boys in 1993, and as used as a soccer anthem in 2013, to illustrate the importance of context and the "rhetorical situation" in our interpretation of the song's written lyrics. That sort of example can connect with students in a way that a discussion of data analysis cannot.

Finally, we believe that our courses benefited from the experienced use of pedagogical techniques from our partner disciplines. A prime example is the use of critique, as practiced in Art and Design. Our experience as instructors has been that peer review can sometimes be a powerful technique to aid in student learning, but it can also lead to less than ideal experiences when students are not adequately prepared or when the review itself becomes too negative in tone. The critique technique brought from Art helped alleviate some of these concerns about negativity by asking students to focus only on aspects of the work that was most effective. In that setting students were supportive and enthusiastically engaged.

To counter the concern that peer review is less effective when students don't yet have enough experience to give good feedback, our colleague from Art assured us that students are often surprised at how much agreement there is on what constitutes effective design, even in the subjective world of art. In our experience during the class, it did turn out to be the case that students were fairly good at identifying designs that promoted readability and used appropriate statistics. There was less agreement in our critique sessions on what constituted appropriate information density or appropriate use of color and other design elements. Students sometimes chose simpler data visualizations that were less multivariate in nature, while also gravitating toward designs that used more vibrant and varied color palettes than might have been recommended in lectures. Even those divergences from the "right" answers provided a good basis for class discussion, however.

Student feedback has identified at least two challenges related to our interdisciplinary course offerings: the desire for greater contact with our interdisciplinary collaborators, and the tradeoff between focus on "soft skills" and focus on greater depth of technical content. One success we have had in encouraging interdisciplinary contact has been in the recruitment of outside readers for student theses. Those readers have been pulled from diverse departments, such as Biology, Art, Accounting, English, and Physics, matching the diversity of thesis topics. While this interaction comes late in a student's career, students have valued the input of these subject-matter experts. We hope that interaction with our students will also create a pool of faculty willing to contribute to the program in the future.

We plan to address the availability of technical content through an update to our current course content and through the creation of a new on-the-ground masters program in Applied Data Science. Because data science, as a field, is constantly evolving, we are currently in the midst of our first review of program material. By identifying areas where content overlaps between courses, we have freed space in our current course list for additional content in machine learning and an introduction to generative AI. An on-the-ground masters program, currently in the planning stages, will require more credits for completion and focus on a more technical audience. The courses designed for that new program would also serve as a source of electives for students in our current program who would like to see additional content.

Regardless of these plans, we remain committed to our interdisciplinary courses in communication, visualization and ethics, which were consciously included as a way to leverage Truman's liberal arts and sciences expertise to provide students with skills that our industry partners consistently mention as important. Because technical goalposts *are* constantly shifting, we believe that time spent in the program that broadens students' perspectives on what a data scientist can be is time well spent—it has produced good results in our program, and student feedback has thus far validated the importance of including these interdisciplinary approaches.

Supplementary Material

The following files are included in the supplementary material:

1. Signature assignment in Narrative, Argument and Persuasion
2. Signature assignment on algorithmic bias in bail risk assessment
3. Algorithmic bias data set, modified from the original Pro Publica data set
4. Data visualization deconstruction assignment
5. SVG file containing components of Image 1 for the deconstruction assignment
6. SVG file containing components of Image 2 for the deconstruction assignment
7. Assignment using ggplot to recreate a historic data visualization by William Playfair
8. Data set of wheat and wages for use with the Playfair reconstruction
9. Data set of British monarchs for use with the Playfair reconstruction

Funding

The authors would like to acknowledge the Missouri Department of Higher Education and Workforce Development, though FY2020 and FY2023 MoExcels grants.

References

American Statistical Association Undergraduate Guidelines Workgroup, (2014) *Curriculum Guidelines for Undergraduate Programs in Statistical Science*. American Statistical Association, Alexandria, VA. https://www.amstat.org/docs/default-source/amstat-documents/edu-guidelines2014-11-15.pdf.

Angwin J, Larson J, Mattu S, Kirchner L (2016). Machine bias. https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing. Accessed: 2023-05-20.

Bitzer LF (1968). The rhetorical situation. *Philosophy & Rhetoric*, 1(1): 1–14. http://www.jstor.org/stable/40236733.

Brancusi C (1907). The kiss. Tylwyth Eldar. https://www.loc.gov/pictures/item/93511527/. Accessed: 2023-07-29.

Cech TR (1999). Science at liberal arts colleges: A better education? *Daedalus*, 128(1): 195–216.

Clarke R (1997). Introduction to dataveillance and information privacy, and definitions of terms. https://www.rogerclarke.com/DV/Intro.html. Accessed: 2023-07-29.

De Veaux RD, et al. (2017). Curriculum guidelines for undergraduate programs in data science. *Annual Review of Statistics and Its Application*, 4(1): 15–30. https://www.amstat.org/docs/default-source/amstat-documents/edu-datascienceguidelines.pdf. https://doi.org/10.1146/annurev-statistics-060116-053930

Duke University (2024). Master of engineering in computer science, economics, and data science (course 6-14p). https://datascience.duke.edu/academics/mids-courses/. Accessed: 2024-03-31.

Fisher W (1984). Narration as a human communication paradigm: The case for public moral argument. *Communication Monographs*, 51: 1–22. https://doi.org/10.1080/03637758409390180

Massachusetts Institute of Technology (2024). Master of engineering in computer science, economics, and data science (course 6-14p). https://catalog.mit.edu/interdisciplinary/graduate-programs/computer-science-economics-data-science/. Accessed: 2024-03-31.

Miller S (2014). Collaborative approaches needed to close the big data skills gap. *Journal of Organization Design*, 3(1): 26–30. Special Issue on Big Data and Organization Design. https://doi.org/10.7146/jod.9823

Munsell AH (1912). A pigment color system and notation. *The American Journal of Psychology*, 23(2): 236–244. https://doi.org/10.2307/1412843

Orwell G (1946). Politics and the English language. *Horizon*, 13: 252–265. https://www.orwellfoundation.com/the-orwell-foundation/orwell/essays-and-other-works/politics-and-the-english-language/

Playfair W (1821). *A Letter on Our Agricultural Distresses, Their Causes and Remedies: Accompanied with Tables and Copper-Plate Charts, Shewing and Comparing the Prices of Wheat, Bread, and Labour, from 1565 to 1821. Addressed to the Lords and Commons*. William Sams, Bookseller to His Royal Highness the Duke of York, London. https://www.google.com/books/edition/A_Letter_on_Our_Agricultural_Distresses/Ywt4gFSlL_EC.

Radovilsky Z, Hegde V, Acharya A, Uma U (2018). Skills requirements of business data analytics and data science jobs: A comparative analysis. *Journal of Supply Chain and Operations*

Management, 16(1): 82–101.

Rodin A (1882). The kiss (le baiser). https://en.wikipedia.org/wiki/File:Rodin_-_Le_Baiser_06.jpg. Accessed: 2023-07-29.

Rowland R (1987). Narrative: Mode of discourse or paradigm? *Communication Monographs*, 54(3): 264–275. https://doi.org/10.1080/03637758709390232

Sachant P, Blood P, LeMieux JA, Tekippe R (2016). *Introduction to Art: Design, Context, and Meaning*. University of North Georgia University Press, Dahlonega, Georgia. https://open.umn.edu/opentextbooks/textbooks/introduction-to-art-design-context-and-meaning.

Southern Methodist University (2024). Master of science in data science. https://datascience.smu.edu/. Accessed: 2024-03-31.

Syracuse University (2024). Ms in applied data science. https://onlinegrad.syracuse.edu/information-science/academics/. Accessed: 2024-03-31.

Toulmin S (2003). Fields of argument and modals. *Uses of Argument*. Cambridge University Press. 11–39.

Wikipedia (2023). Munsell color system. https://en.wikipedia.org/wiki/Munsell_color_system. Accessed: 2023-07-29.