The Promise of Portfolios: Training Modern Data Scientists

Deborah Nolan¹ Sara Stoudt^{2,3}

¹Department of Statistics, University of California Berkeley, Berkeley, California, United States of America,

²Berkeley Institute for Data Science, University of California Berkeley, Berkeley, California, United States of America,

³Statistical & Data Sciences Program, Smith College, Northampton, Massachusetts, United States of America

Published on: Jul 30, 2021

DOI: <https://doi.org/10.1162/99608f92.3c097160>

License: [Creative Commons Attribution 4.0 International License \(CC-BY 4.0\)](https://creativecommons.org/licenses/by/4.0/)

ABSTRACT

Data scientists rely on many technical skills and the ability to reason about data to solve problems. As educators grapple with how to prepare students in this new field, they are faced with identifying both what a student must know and what a student should be able to do by the end of their data science education, and also how to collect evidence of those abilities.

We present a way to unite and coordinate individual efforts toward training well-rounded data scientists: a data science portfolio that highlights strong communication. This structuring of classroom assignments provides a way to evaluate students' mastery of material in each class and also allows for a student to build a professional portfolio that remains valuable after the class is over.

Data science portfolio pieces broadly include written and visual assignments that give students practice crafting data-driven arguments and narratives for a variety of audiences. This flexible nature of the portfolio gives students a way to demonstrate their abilities in an inclusive, 'choose your own adventure' way.

As students refine and share their portfolios with others throughout a course or program, they can see their own growth and receive feedback from instructors, peers, and the broader data science community.

We provide examples of and guidance for how to implement a data science portfolio approach in single courses and wider data science programs.

Keywords: communication, data science assignments, pedagogy, portfolio, visualization

1. Introduction

Developing as a data scientist and being secure in this identity comes from practicing one's craft and gaining confidence in one's skills and ability to reason about a data-related question. As data science classes, majors, minors, and full programs emerge at the undergraduate and graduate levels, educators are faced with the challenge of how to prepare students and assess their learning. Faculty are tasked with considering what aspiring data scientists should be able to do at the end of a program, as opposed to merely what they know (Hardin et al., 2015; Bargagliotti et al., 2020), while grappling with what it means to be a data scientist. Data science takes a multifaceted approach to exploring, analyzing, and solving problems with data, and we believe that students should provide multifaceted evidence of their mastery of these skills and their ability to reason with data.

Portfolios, a collection of work that provides evidence of a person's talents and skills, have commonly been used professionally in the arts (Scolere, 2019) and have been used in a general education setting (Arter & Spandel, 2005; Burnett & Williams, 2009; Carleton College, n.d.; Crump, 2019) to collect and share student work. In industry, code portfolios (e.g., sharing projects on GitHub) have become a common add-on to a job application as evidence of various computational skills (Craig et al., 2018; Marlow & Dabbish, 2013).

We see the building of a broader data science portfolio as an opportunity for students to collect and reflect on their whole data science education, while the successful creation of a portfolio demonstrates mastery of the field more generally that can help the student synthesize what they have learned in their educational journey.

The skills a student displays in their portfolio can include the technical ones most often associated with data science: coding, statistical modeling, machine learning, and so on. However, essential skills also include those sometimes dismissed as ‘soft’ skills, such as communication, collaboration, and ethics. We consider these skills to be career success factors, and note that they are acknowledged under the umbrella of ‘data acumen’ (National Academies of Sciences Engineering Medicine [NASEM], 2018). Rather than treated as an afterthought, we argue that communication, as a crucial part of data science practice, should be explicitly taught and practiced in the data science curriculum. For this reason, we view portfolio contributions expansively to include writing for a broader audience (e.g., blogs, participation in online communities, news pieces), critically reading others' work, coding as writing, and visualizing data, in addition to common products such as formal reports and oral presentations.

Recognizing the need for alternative methods of assessment and the benefits for students both educationally and pre-professionally (Ring et al., 2017), there has been some overall guidance for using mathematical portfolios in the classroom (Koca & Lee, 1998), implementation of mathematical portfolios in particular undergraduate courses (Burks, 2010; Domínguez García et al., 2015; Sole, 2012), and an implementation branching into a statistics course at the graduate level (Keeler, 1997). These individual implementations vary in what is considered a portfolio piece from more traditional contributions like problem sets or exams to more reflective contributions.

We offer a framework to adapt the portfolio concept to the practice of modern data scientists and advocate for interweaving it throughout a single data science course as well as into the broader data science curriculum. Data science portfolios that emphasize strong communication skills such as the ability to: write precise and accurate technical reports, translate findings to a broader audience in conversation or written form without sacrificing this precision and accuracy; and connect findings to a relevant context for a variety of audiences, can unite and coordinate individual efforts toward training well-rounded data scientists and create an inclusive approach that allows students a flexible way to demonstrate their technical, reasoning, and communication abilities.

The remainder of this article is structured as follows: We first describe the strengths of a data science portfolio and its benefits for students, next we describe the details of portfolio creation based on our experiences in the classroom, then we provide structural advice for departments and programs and implementation advice for individual instructors, and we close with a discussion of related efforts and conclusions about how our approach fits in and augments the current state of data science education.

2. Data Science Portfolios and Their Strengths

Working in the ‘open’ has become a popular approach in data science (Dabbish et al., 2012) that has translated to the pedagogy of data science (Beckman et al., 2020; Glassey, 2019). Data scientists have encouraged public-facing portfolios (Kross & Guo, 2019; Laderas, n.d.; Robinson, 2017; Robinson & Nolis, 2020), and communities of practice have appeared online to build skills together and learn from one another (R for Data Science Online Learning Community, n.d.; Shrestha et al., 2021). A potentially public-facing portfolio provides an opportunity for each student to express their own identity as a data scientist, as well as reflect on their growth. However, practicing data science in the open requires going beyond storing code on GitHub. After the data munging and analysis are done, a crucial part of the process is consolidating and communicating the findings and their implications to a wide variety of audiences: data science colleagues, domain scientist collaborators, business stakeholders, and the broader public.

Although science and numbers appear objective on the surface, knowledge sharing in scientific fields, including data science, is essentially based on an argument and requires attention to elements of both style and rhetoric (Sutton, 1997; Woodard et al., 2020). A data-driven argument convinces someone of the validity of the findings, the appropriateness of the analysis, the generalizability of the conclusions, and the overall credibility of the evidence presented. A data-driven narrative is fundamentally a story, explaining how the findings fit with others' work, how it affects current views in the field, and why the problem's solution is important in the original context. Furthermore, being able to situate one's work into a broader narrative is crucial to making the work more accessible and the field more inviting to newcomers.

To practice a wider range of learning outcomes, including crafting data-driven arguments and narratives, we consider data science portfolio pieces broadly and emphasize the communication elements of each. Just as there is no one type of data scientist, there is no fixed set of requirements for a portfolio. Included work can be written, visual, or oral. Written work can include summaries of findings or thorough figure captions. Visual work can include static or interactive data visualizations. Oral work can include a recording of a presentation, a discussion among peers, or a think-aloud interview (Reinhart et al., 2020). The work, in any form, can be technical or meant for a broader audience. Products like code or a formal report are meant for others in our field while products like blog posts or contributions to Wikipedia engage a broader audience.

Despite the fact that communication has been acknowledged as an important topic in both statistics and data science education (De Veaux et al., 2017; Guidelines for Assessment and Instruction in Statistics Education

[GAISE], 2016; NASEM, 2018, 2020), it can be challenging to formalize and unify technical communication training, whether at the undergraduate or graduate level. In fact, this gap is how we got started thinking about portfolios ourselves. We participated in UC Berkeley's Art of Writing Program, which supports the creation of writing seminars in many disciplines, including those beyond the humanities (Art of Writing, n.d.), and we first implemented a portfolio approach there. Our communication-focused data science portfolio has the added benefit of providing a concrete way to interweave data-centered communication throughout the curriculum as well.

3. What Is the Benefit for Students?

Portfolio-creation is not solely about students amassing a variety of products that show their technical skills; it is also about teaching them how to communicate their work and present themselves as data scientists. Portfolio contributions require a deeper understanding of data science content. In order for students to share their work, they must craft an argument to convince their audience that what they did is defensible and what they found is believable. Students also must place their findings within a compelling narrative to make sure they have a strong response to those questioning 'so what?' Just as students can struggle with the lack of a 'right' answer in applied statistics and modeling situations, students can struggle with the absence of a single argument and narrative that effectively convinces and engages an audience.

Working through a variety of portfolio assignments provides ample practice convincing and engaging many audiences throughout a student's academic trajectory rather than waiting to produce a few shareable products toward the end of a student's academic training. By creating an expectation of sharing works in progress rather than just when the work is 'finished,' students are encouraged to break perfectionism tendencies that can hold them back. Pragmatically, students also have something to point to their abilities, beyond exam scores and grades, at any time in their academic training, for example, when they are applying for research opportunities, internships, fellowships, and jobs.

Some redundancy in styles and topics of portfolio assignments can be built into a curriculum so that skills deemed particularly important or challenging can be emphasized and revisited. Portfolios are not static snapshots of strengths and weaknesses; they evolve as students do. This continual evolution gives students the opportunity to reflect on the work they are doing rather than seeing each assignment as merely a means to a grade. As students take ownership over their portfolios and see their progress over time, they can identify and fill gaps in their understanding and track progress toward their career goals with a growth mindset. Other positive effects of reflection found in studies of portfolios in other contexts include thinking about future identities (Bennet et al., 2016) and considering professional identities that cross multiple roles such as researcher and teacher (Syvante et al., 2015). From a meta perspective, the self-reflection needed to curate a portfolio is yet another way to practice crafting an argument (how can I convince my portfolio reader that I have acquired the necessary skills?) and a narrative (how can I use my portfolio to tell someone about my identity as a data scientist?).

Our data science portfolio approach shifts a typical assignment structure of problem sets in favor of assignments that can serve two purposes: a means to evaluate a student's mastery of material and a way for a student to build a portfolio that remains valuable after the class is over.

4. What Does Building a Data Science Portfolio Look Like?

In this section we make the idea of a data science portfolio concrete by describing what the portfolio creation process looked like in two classes we have taught. The content of the classes reflects our communication-oriented motivation for experimenting with the data science portfolio. One class, which we refer to as the 'seminar,' was taught at UC Berkeley, focused on statistical writing, and had about 15 students in it. This seminar was explicitly structured around the creation of a portfolio. A second class, the 'gateway' course, taught at Smith College, focused on communication with data more broadly, and had about 60 students in it. The gateway course was taught in a way that allowed assignments to become part of a portfolio without the portfolio being the main intent of the class. We also provide a roadmap of how to integrate portfolio assignments into courses, using our classroom-tested assignments as examples.

These portfolio assignments stress the importance of argument and narrative in a data scientist's practice. Providing opportunity for frequent, low-stakes practice and placing an emphasis on refinement allows students the space to craft and the support to refine a data-driven argument. We also build in flexibility in what students work on to show that arguments are made differently in varied contexts and encourage public visibility as a data science scholar to encourage students to shape and share the personal narratives of their experience. Our implementation of these principles in practice follow.

4.1. Forming a Data-Driven Argument

Communication takes practice, and low-stakes writing helps provide that practice with less pressure to perform for a weighty grade (Elbow, 1997). In both the seminar and gateway, regular weekly assignments had students practice a variety of communication approaches at differing levels of formality. These assignments were graded on the check, check-plus, check-minus system where a good-faith effort would result in full credit (check), a piece lacking evidence of effort would result in partial credit (check-minus), and a piece of work that truly stood out would be acknowledged but not treated as extra credit (check-plus).

In both classes, students got an opportunity to take a break from creating new material to focus on refining prior work. In other contexts, students rarely get to iterate on high-stakes projects because they often come at the end of a term. However, a portfolio can include multiple drafts of a piece of work to show and emphasize progress as well as track an argument's evolution. As peer review is part of the profession for academic data scientists and practitioners alike, we incorporated it into both courses as part of the refinement process (Anders, 2020).

In the seminar, students were responsible for choosing and refining a subset of their portfolio assignments into more formal ‘products’. This refinement process required students to reflect on what subset would play to their strengths and/or interests, and receive and incorporate feedback from both a peer and the instructor. In the gateway course, iteration was built into a specific subset of assignments, requiring reflection on what was and was not effective about their work at each stage. One project had many stages, one of which was incorporating feedback into revising a single data visualization. Other assignments built on one another, a sort of forward-moving iteration. For example, one week a student would create a data visualization for a topic and then get feedback from a peer. The following week, when the student was tasked with creating a data visualization for a new topic, they also were to incorporate the past feedback into their new creation.

4.2. Forming a Data-Science Identity

Allowing students some choice in what they work on results in a diverse set of portfolio products that better reflect the individuals in the class. Practicing skills in a context relevant to students’ own interests and goals can also help motivate them (Beymer & Thomson, 2015). In the seminar, students picked from a selection of assignments related to the content of the class each week. As they explicitly had portfolios in mind, they were primed to choose activities that they felt would best represent their abilities and interests in the professional narrative they were building.

Full flexibility is not always feasible due to an instructor’s constraints for giving feedback. In the gateway course, students worked on the same set of assignments, but the assignments were designed to allow for maximal flexibility in what data students could work with or what application area they could focus on. This allowed students to tailor the material to their own interests, which was especially helpful in the context of the heterogeneous makeup of this class (students from all class years and many majors, from humanities to STEM).

In the seminar, in order to provide a way for students to reach an audience beyond the structure of the classroom, we aimed to make a printed anthology of the class’s work, but the logistics of this proved too daunting in the end. A subset of our students’ work was featured in an anthology of student work curated by the Art of Writing program. Having a tangible product is a worthy goal and may become more feasible as a team effort among faculty committing to adding portfolio creation into their classes. In an offshoot of this class, we were successful in posting student work on a class blog, and students with some technical experience could likely run their own blog with some guidance from an instructor.

In the gateway course, there were a few levels of visibility offered to broaden the students’ audience base. There was a public-facing course newsletter that students could submit work to or contribute to the production of. This went out to a list of subscribers, open to the public, a handful of times throughout the semester and included student reflections on what they had learned recently in class as well as their work samples. Acknowledging that not everyone wants their work to be so public, we also shared work in a smaller community, within the course and within subsets of the campus community. Remotely, this took the form of a

‘gallery walk’ where students were asked to reflect on what they accomplished over the course of the semester, choose a piece of work they were most proud of from the class, and display it on a slide to be browsed by those with view access to the full slide deck. In a large class that occurred remotely, this was especially helpful for students to see what other classmates had been working on.

4.3. Portfolios Throughout the Curriculum

Approaches for helping students practice honing a data-driven argument can also be implemented at the program level. As students collect portfolio contributions throughout multiple classes, they can have some choice in what ends up in their final portfolio. As they move forward in their academic career, they may choose to remove some prior work that is no longer representative of their most persuasive argumentation or their most compelling narrative. Iteration can also be practiced on a longer time scale if students are instructed to refine portfolio assignments from previous introductory courses in upper-level courses. Curation forces students to reflect on their work and how they and their arguments have grown throughout their education; it also requires critical discernment of not only what material best shows off their strengths but also what types of skills are most crucial to prove to their audience in the first place. This same eye for parsimonious yet memorable presentations will serve them well in time-constrained career settings such as job interviews.

We want students to explore pathways in data science, and not get boxed into a common portrayal of the field. A ‘choose your own adventure’ approach can also be implemented at the program level. Students may be required to have a certain number of portfolio contributions within particular categories of content or medium yet have the freedom to pick which assignments gathered from multiple classes to include. A broader audience can be cultivated by having a program-level portfolio showcase at different stages of the students' careers. This would provide a way to recruit new students to the program as well as brag about student achievements to potential future employers.

The tables of classroom-tested portfolio assignments (Tables 1 and 2) can be used to map to what level instructors are teaching and what type of communication they want students to practice and add to their portfolios in their class. An instructor of a single class may choose to focus on a column that reflects the level of their students, while a program may choose to focus on a row, or multiple rows, to spread throughout the curriculum. Reflections on the process of completing or revising any of these assignments could become their own assignments. We provide references to samples of a subset of these assignments, published online elsewhere as part of the optional, public-facing nature of our previous classes in the Appendix.

Table 1. Written portfolio assignments.

	Beginner	Intermediate	Advanced

Writing – Formal	<p>Describe the ideal data set needed to answer a question of interest.</p> <p>Summarize watching someone code, analyze, or explore data (e.g., livestreams).</p> <p>Write and revise informative captions and/or alternative text for visualizations.*</p> <p>Peer review of written work (or code).</p>	<p>Link visualizations together to tell one coherent story (e.g., storyboard).*</p> <p>Critique and/or edit a Wikipedia article on a statistics or computing topic.</p>	<p>Read, analyze, and critique a formal, data-driven article.*</p> <p>Write a personal statement about yourself as a data scientist.**</p> <p>Revise and resubmit a previous portfolio assignment.</p> <p>Read a data-related book and write a book report.</p>
Writing – Broader Audience	<p>Go to a public seminar where someone is using data to answer a question and summarize it.</p> <p>Explain your work in plain language using xkcd’s Simple Writer.</p>	<p>Write a blog post about the process of learning / exploring in data science.**</p> <p>Keep a data diary and write an accompanying blog post.</p> <p>Write a press release for a project.</p>	<p>Write a blog post explaining a data-related method.</p> <p>Find numbers in the news and assess their effectiveness and accessibility to the audience.</p>

Note. We have classroom tested a variety of *written* portfolio assignments, broken down here by level and communication type. Many beginner activities can be used at the intermediate and advanced levels (e.g., peer review). Some intermediate and advanced assignments can be modified for beginners (e.g., reading, analyzing, and critiquing a formal paper can be adapted for less technical writing, such as blog posts and news articles). Assignments marked with a single asterisk are those particularly helpful for forming a data-driven argument, and those marked with a double asterisk are those particularly helpful for developing a data science identity. Descriptions of these appear in the Appendix.

Table 2. Visual portfolio assignments.

	Beginner	Intermediate	Advanced
Visual – Formal	<p>Explore an interactive visualization.</p> <p>Collect visualizations that have a common theme and that you find effective.*</p>	<p>Participate in Tidy Tuesday, which offers a weekly data set and prompt for exploration, visualization, and sharing findings.</p>	<p>Reproduce a figure from an article (e.g., using open data from FiveThirtyEight or Our World in Data).</p>

Visual – Broader Audience	Collect data for a week and draw a Dear Data postcard.** Create a tactile data visualization. Critique user-friendliness of an interactive visualization.	Create a single visualization that quickly reveals a finding to the audience.*	Create a user-friendly interactive visualization.
----------------------------------	---	--	---

Note. We have classroom tested a variety of *visual* portfolio assignments, described and broken down here by level and communication type. Some beginner activities can be used at the intermediate and advanced levels (e.g., Dear Data). Assignments marked with a single asterisk are those particularly helpful for forming a data-driven argument, and those marked with a double asterisk are particularly helpful for developing a data science identity. Descriptions of these assignments appear in the Appendix.

Portfolio assignments can be used to supplement current course work, including formal reports and project modules. Table 3 shows how different portfolio assignments can support the writing process of a formal report, while Table 4 shows how some portfolio assignments may be elevated to a project level and further supplemented by other smaller assignments. We do not suggest that every step in the formal report process or every course project needs to be accompanied by all of these portfolio assignments; rather, we think instructors can choose a subset of steps to connect to portfolio building while potentially coordinating across multiple classes that end in a formal report or final project.

Table 3. Incorporating portfolio assignments in report writing.

Stage of Preparing and Writing a Formal Report	Relevant Portfolio Assignments
Choose a data set.	Describe the ideal data needed to answer a question of interest. Participate in Tidy Tuesday.
Exploratory data analysis (EDA).	Write a blog post about the exploration process. Write and revise captions and/or alt-text for visualizations.
Preliminary analysis.	Keep a data diary and write an accompanying blog post.
Get feedback on analysis.	Create a single visualization that reveals a finding to the audience quickly. Peer code review.

Outline report.	Link visualizations together to tell one coherent story, that is, storyboard. Read, analyze, and critique the structure of a formal, data-driven article.
Draft data description and methods.	Write a blog post about a method.
Draft results.	Write a blog post about your findings.
Draft discussion & conclusion.	Read, analyze, and critique the ending of a formal, data-driven article.
Draft introduction.	Write a press release for your paper.
Draft abstract.	Explain your work in plain language using xkcd's Simple Writer.
Get feedback on the draft.	Peer review of written work.
Revise.	Revise and resubmit.

Note. One approach to incorporating portfolio assignments into an existing class is to tie assignments to steps in the analysis and report-writing process. Descriptions of these assignments appear in the Appendix.

Table 4. Incorporating portfolio assignments in projects.

Portfolio Assignments as Course Projects	Supplementary Portfolio Assignments
Collect visualizations that have a common theme and that you find effective.	Keep a data diary and write an accompanying blog post. Reproduce a figure from an article (e.g., using open data from FiveThirtyEight or Our World in Data). Find numbers in the news and assess their effectiveness and accessibility to the audience.
Link visualizations together to tell one coherent story (e.g., storyboard).	Collect data for a week and draw a Dear Data postcard. Create a single visualization that reveals a finding to the audience quickly. Participate in Tidy Tuesday.

Create a tactile data visualization.	<p>Write and rewrite informative captions or alternative text for visualizations.</p> <p>Explain your work in plain language using xkcd’s Simple Writer.</p>
Explore an interactive visualization.	<p>Critique user-friendliness of an interactive visualization.</p> <p>Write a blog post about the process of exploring an interactive visualization.</p>

Note. Another approach to the portfolio is to choose assignments to expand into projects for a module-based course while using other assignments as supplements. Descriptions of these portfolio assignments appear in the Appendix.

5. Advice for Data Science Departments and Related Programs

We do not want to downplay the logistical challenges in developing portfolio-integrated courses and programs. Ideally, the portfolio is integrated throughout a program and faculty are well-equipped to teach and evaluate communication skills. This is likely to require program-wide assessment of student outcomes, coordinated modification to curricula, and faculty training and incentives.

We advocate that now is the time to revise our data science programs. These programs are in their infancy, and examining them holistically now is less disruptive. A program-wide approach spreads out the workload as the program takes steps toward portfolios as student products. Furthermore, during the pandemic, teaching norms were disrupted. At our institutions, faculty shifted away from traditional exam-based assessment toward projects, which are more conducive to the portfolio. Faculty developed new modes of instruction and strengthened cross-campus teaching communities. This momentum and the interdisciplinary nature of data science can be used advantageously to find faculty partners across an institution. Faculty in other STEM fields who may want to incorporate a portfolio into their curriculum and colleagues in English/Composition can be allies and partners in change.

How do instructors get training in teaching communication? The optimistic answer is that given time, as the communication of data science becomes part of data science education more broadly, additional preparation on behalf of the instructors will not be necessary. In the meantime, there is admittedly a gap to be filled in a more ad hoc way. We can speak from our own self-training. We synthesized resources from a variety of literatures, including rhetoric, science communication, the pedagogy of teaching writing, and from writers explaining their own creative processes. Time to do this was a luxury, but incentives for this kind of self-study have tremendous potential.

Another way to spur on the faculty is to take advantage of small or nontraditional pockets of money at an institution or scholarly community. Perhaps there are course release or funding opportunities for adding new

writing elements to a course or creating new courses, like a first-year seminar. These opportunities may have passed by undetected in the past because faculty were not primed to think about their role as teachers of communication skills. For example, the seminar described in Section 4 was supported by UC Berkeley’s Art of Writing program (Art of Writing, n.d.), the tactile data visualization assignment class tested in the gateway course, also described in Section 4, grew out of a curricular enhancement grant from Smith College’s Design Thinking Initiative (Design Thinking Initiative, n.d.), the Dear Data seminar references in Figure A1 was part of Berkeley’s freshman seminar series, which is accompanied by a small monetary incentive, and the Writing Data Stories course was offered as a connector to an introductory data science course (Berkeley Computing, Data Science, and Society, 2021b). These courses and seminars can serve as a testing ground for development of portfolio-related curricula that will migrate into the curriculum.

Scalability is another challenge. In our experience, giving feedback on communication-focused assignments is hard to scale, creating another potential strain. Peer review and leveraging the broader community gained by working in the open can help but does not completely replace instructor feedback. Beneficial institutional support may include training and staffing teaching assistants or graders and a commitment to balancing and valuing the teaching load of those who include portfolio assignments in their courses. This tradeoff may also be an argument for reducing class sizes, at least for courses where communication is a heavier focus.

Despite the challenges, we think the communication-forward data science portfolio approach helps students see how the material in individual classes relates to the ‘real world’ and how learning outcomes translate into tangible skills and products. It also helps faculty members evaluate students on a more well-rounded measure of competency. A letter of recommendation based on a portfolio rather than a set of grades provides a more all-inclusive narrative. From our own experience, these letters are easier to write too.

There is also an equity argument to be made. Calls for students to create portfolios as a way to get internships and jobs assume students have ‘free’ time to work on these outside of their other responsibilities. This disadvantages students who have employment or care-taking responsibilities in their outside-of-class lives. By building portfolio creation into the curriculum, everyone has a chance to leave the class not only with a grade, but also with a tangible set of products.

6. Advice for Instructors

Explicitly teaching communication skills and evaluating reasoning, especially in the form of technical writing, can be intimidating. We have ourselves felt intimidated. In our own statistics and data science training we have not received formal training on how to both craft an argument and build a narrative, let alone been given training in teaching this kind of communication style. There is an opportunity for faculty both looking to ‘level up’ their own skills but also for a broader effort to ‘teach the teachers’ as part of implementing a data science portfolio program-wide.

To get instructors started, we have provided descriptions of over 20 portfolio projects appropriate for three levels of undergraduate students, beginner, intermediate, and advanced, and spanning multiple types of communication skills (see the Appendix for more detailed descriptions). These portfolio projects can be adapted to fit into many different courses or serve as inspiration for instructors to create their own activities. We provide here some guidance for what makes an effective portfolio assignment and how to evaluate these types of assignments.

Those who want to ease into change do not need to start from scratch to help their students build a data science portfolio in their class. Many preexisting assignments may be slightly adapted to make them more portfolio friendly. Consider the following:

1. Can someone outside of the class easily digest the output without the context of the class? Consider a student presenting the work to a potential employer as the structure of the assignment is adapted.
2. Is there a baseline skill that students should practice that they can then hone in a second iteration? If yes, structuring the assignment so that it can be completed at multiple levels can help meet that learning goal.
3. Is there a way to add some flexibility or choice to a current assignment? Completely free-form assignments come with their own challenges, but curating a handful of vetted options for a topic or data set can help mitigate major ones.
4. Can variety be added to current assignments? If a syllabus is currently problem-set or code heavy, an actionable first step is to turn some assignments into writing prompts or visualization challenges.

Instructors can also make their own efforts toward creative course preparation visible to their supervisors, thanks to the philosophy of public-facing portfolios. After only one course, instructors will have products to point to as evidence of both their students' abilities but also their own pedagogical innovations.

Aside from the support-related challenges, with more creative and open-ended assignments comes the challenge of how to fairly assess student work. Students may also initially struggle with assignments and feedback that feel more subjective when they are used to more concrete tasks with clear solutions (Steen-Utheim & Hopfenbeck, 2018). This is another opportunity to learn from colleagues in the humanities who have experience assessing written arguments and narratives.

Rubrics can help standardize what to look for in each student's contribution. These can be refined as instructors see more student work and get a sense of what shows more evidence of mastery or creativity. We have had success getting advice from colleagues in other departments on how to design rubrics that avoid assigning points for specific tasks, for example, axis labels in a plot, and take a more holistic assessment strategy. We have also grouped activities into similar categories or chosen a subset of activities for students to refine further to limit the number of rubrics that need to be created for a course. We provide sample rubrics in the Appendix.

For low-stakes assignments we recommend rewarding a good-faith effort. A satisfactory–unsatisfactory binary allows an instructor to focus their time on giving more thorough written comments. Flexibility has its pros and cons. We recommend providing checkpoints within a class and within a program to ensure students do not procrastinate and, say, save multiple portfolio contributions until the end of a course. These intermediate due dates also help spread out the time needed to provide timely feedback.

7. Discussion

Perhaps the biggest opportunity that the portfolio brings to data science pedagogy is its emphasis on modern communication strategies and products. What we advocate for here is rooted in practices determined to have consistent evidence of ‘high impact’ by the Association of American Colleges and Universities: Coordinating portfolios across classes connects to learning communities in that it requires integrating concepts and skills across multiple classes, incorporating writing-intensive elements into more classes is consistent with continual writing practice throughout the curriculum, and, most directly, a communication-focused data science portfolio is consistent with an ePortfolio to reflect on and share work over time. Depending on specific implementations, a communication-focused data science portfolio can also be consistent with first-year, capstone, collaborative, and common intellectual experiences (Kuh, 2008). There are certainly efforts among the data science community that value communication and follow different approaches to teaching these skills.

One approach incorporates writing-intensive courses into the curriculum. These are in the spirit of Writing Across the Curriculum (WAC) efforts that provide a framework for developing writing courses with different motivations that are taught by faculty throughout the institution (McLeod & Soven, 2000). WAC efforts can be implemented in different ways, including the ‘writing to learn’ style that uses informal writing to help students work through content-specific concepts and the ‘writing in the disciplines’ style that prepares students to write in the forms expected of them in their future profession (McLeod & Soven, 2000). For example, this semester, UC Berkeley is experimenting with a couple of communication-focused courses: a writing-focused ‘connector’ course, classes across fields that are designed to be taken concurrently with or directly after a foundational data science course (Data 8) (Berkeley Computing, Data Science, and Society, 2021a,b), and a first-year seminar focused on visual communication and based on the Dear Data project (Lupi & Posavec, 2016, 2018). Adapting these standalone communication courses to require a portfolio as an end-product can help these courses move away from traditionally defined outputs such as formal papers or reports.

Other approaches to teaching communication spread communication skills across multiple courses in the curriculum. These are in the spirit of Writing Enhanced Curriculum (WEC) efforts that have similar goals as the WAC movement but take the approach of spreading writing over the curriculum rather than focusing on writing-intensive courses (WAC Clearinghouse, n.d.). We are aware of some programs and individuals, likely among others, that have taken advantage of these WAC and WEC resources specifically in mathematics, statistics, and data science programs (Bucknell University, n.d.; Hayden, 1989; Kinnaird, 2021; Parke, 2008; Townsend et al., 2003). This approach can work well with portfolio creation that incorporates pieces from a

variety of classes. The coordination already required to make these programs possible can be further leveraged to bring faculty together to discuss communication-focused learning outcomes for data science specifically.

There are many advantages to both of these writing program–inspired approaches to incorporating communication into the curriculum, but we see some warnings that they are not fully meeting the needs of future data scientists as is. The disadvantages of relying on a single writing-intensive course in a data science curriculum are the reliance on traditionally defined outputs, such as formal papers or reports, and the challenge of where to place this course within a major pathway so that it is not taken too late or too soon. The disadvantages of spreading communication skills across the whole curriculum include traditional course material being prioritized under time pressure, lack of coordination across different faculty members as they teach their own version of classes, and learning outcomes otherwise not being fully met.

Ideally, communication skills would be taught and practiced at multiple points throughout a student's academic trajectory; a writing-intensive course focused on data communication, offered on a regular basis, can establish a common foundation, and these skills can be explicitly reinforced in the rest of the classes within a program. An implementation of a blended WAC and WEC framework, balancing each framework's strengths and weaknesses, may be impractical and resource-intensive. Although still resource-intensive on a smaller scale, that is, at the department level, a portfolio approach can naturally support this hybrid.

A data science portfolio with an emphasis on communication can act as a layer on top of current communication and data science practice efforts, and the nature of the portfolio can mitigate some of the disadvantages to the WAC- and WEC-style approaches alone. The portfolio assignments outlined in this article broaden the idea of what the products of communication look like and give concrete, bite-sized activities that can be partitioned and spread across the curriculum.

8. Conclusion

Whether your institution is building a data science program or is considering how the rise of data science may reshape current programs, we hope you find a place for a more comprehensive and modern assessment of data science–related abilities and skills. We propose a way forward: a coordinated portfolio-building effort, with an emphasis on communication, across the curriculum.

This will take a team effort among faculty and continual reflection on what the learning goals of a program or course are. However, there is a large pool of resources and a vast literature about writing-related pedagogy to draw inspiration from. This article is a roadmap of a subset of these resources and a call to action to further engage with these resources within our own data context.

As major maps are created and revisited and the trajectory of a student through a major is examined, we hope you consider what skills you want your students to have, what types of products could provide evidence of

those skills outside of the context of a particular class, and how those products could be partitioned across the curriculum as portfolio assignments.

Acknowledgments

We acknowledge our campus colleagues, Ramona Naddaff, Rhetoric; Kathleen Donegan, English; and Evan Variano, Mechanical Engineering; and our Statistics colleagues across the country, Nick Horton and Andrew Bray, for their valuable advice in the development of this project.

Stoudt thanks Katherine Kinnaird for sharing her materials to get her started teaching the foundations course described here and for introducing her to the idea of the Writing Enhanced Curriculum.

Stoudt received inspiration for some of these activities from a storytelling workshop that was part of the “Environment and Society: Data Science for the 21st Century (DS421)” National Science Foundation Research Traineeship and took a class from Nick Howe at Smith College that used a portfolio approach.

Disclosure Statement

Nolan and Stoudt received support from The Art of Writing Program for both creation of the seminar described here and for feedback on the textbook we wrote based on the seminar (Nolan & Stoudt, 2021).

This material is based upon work supported by the National Science Foundation under Grant No. (DMS-1745640).

References

- Anders, T. (2020, February 4). *Peer review in data science courses* [Lightning talk]. Rstudio::conf 2020. <https://www.rstudio.com/resources/rstudioconf-2020/peer-review-in-data-science-courses/>
- Art of Writing. (n.d.). *Supporting a writing community at UC Berkeley*. UC Berkeley. <https://artofwriting.berkeley.edu/>
- Arter, J. A., & Spandel, V. (2005). Using portfolios of student work in instruction and assessment. *Educational Measurement: Issues and Practice*, 11(1), 36–44. <https://doi.org/10.1111/j.1745-3992.1992.tb00230.x>
- Bargagliotti, A., Binder, W., Blakesley, L., Eusufzai, Z., Fitzpatrick, B., Ford, M., Huchting, K., Larson, S., Miric, N., Rovetti, R., Seal, K., & Zachariah, T. (2020). Undergraduate learning outcomes for achieving data acumen. *Journal of Statistics Education*, 28(2), 197–211. <https://doi.org/10.1080/10691898.2020.1776653>
- Beckman, M. D., Cetinkaya-Rundel, M., Horton, N. J., Rundel, C. W., Sullivan, A. J., & Tackett, M. (2020). Implementing version control with Git and GitHub as a learning objective in statistics and data science courses.

Journal of Statistics Education, 29(Suppl. 1), S132–S144. <https://doi.org/10.1080/10691898.2020.1848485>

Bennett, D., Rowley, J., Dunbar-Hall, P., Hitchcock, M., & Blom, D. (2016). Electronic portfolios and learner identity: An ePortfolio case study in music and writing. *Journal of Further and Higher Education*, 40(1), 107–124. <http://doi.org/10.1080/0309877X.2014.895306>

Berkeley Computing, Data Science, and Society. (2021a). *Data science connector courses*. UC Berkeley. <https://data.berkeley.edu/education/connectors>

Berkeley Computing, Data Science, and Society. (2021b). *Translating numbers into words: The art of writing about data science*. UC Berkeley. <https://data.berkeley.edu/news/translating-numbers-words-art-writing-about-data-science>

Beymer, P. N., & Thomson, M. M. (2015). The effects of choice in the classroom: Is there too little or too much choice? *Support for Learning*, 30(2), 105–120. <https://doi.org/10.1111/1467-9604.12086>

Bucknell University. (n.d.). *Writing across the curriculum*. <https://www.bucknell.edu/academics/current-students/writing-across-curriculum>

Burks, R. (2010). The student mathematics portfolio: Value added to student preparation? *PRIMUS*, 20(5), 453–472. <https://doi.org/10.1080/10511970802433008>

Burnett, M. N., & Williams, J. M. (2009). Institutional uses of rubrics and e-portfolios: Spelman College and Rose-Hulman Institute. *Peer Review*, 11(1), 24–27.

Buzetto-More, N. (Ed.). (2010). *The e-portfolio paradigm: Informing, educating, assessing, and managing with e-portfolios* (pp. 109–139). Informing Science Press.

Carleton College. (n.d.). *Writing portfolio*. <https://www.carleton.edu/writing/portfolio/>

Cesal, A. (2020, July 23). Writing alt text for data visualization. *Nightingale*. <https://medium.com/nightingale/writing-alt-text-for-data-visualization-2a218ef43f81>

Craig, M., Conrad, P., Lynch, D., Lee, N., & Anthony, L. (2018). Listening to early career software developers. *Journal of Computing Sciences in Colleges*, 33(4), 138–149.

Crump, M. J. C. (2019). Portfolio and prosper. *Nature Human Behaviour*, 3(10), 1008. <https://doi.org/10.1038/s41562-019-0689-0>

Dabbish, L., Stuart, C., Tsay, J., & Herbsleb, J. (2012). Social coding in GitHub: Transparency and collaboration in an open software repository. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (pp. 1277–1286). ACM Press. <https://doi.org/10.1145/2145204.2145396>

Datawrapper. (n.d.). *Enrich your stories with charts, maps, and tables*. <https://www.datawrapper.de/>

Design Thinking Initiative. (n.d.). *Curricular enhancement and institutional capacity building*. Smith College. <https://www.smith.edu/academics/design-thinking/faculty-staff>

De Veaux, R. D., Agarwal, M., Averett, M., Baumer, B. S., Bray, A., Bressoud, T. C., Bryant, L., Cheng, L. Z., Francis, A., Gould, R., Kim, A. Y., Kretchmar, M., Lu, Q., Moskol, A., Nolan, D., Pelayo, R., Raleigh, S., Sethi, R. J., Sondjaja, M., & Ye, P. (2017). Curriculum guidelines for undergraduate programs in data science. *Annual Review of Statistics and Its Application*, 4, 15–30. <https://doi.org/10.1146/annurev-statistics-060116-053930>

Domínguez García, S., García Planas, M. I., Palau, R., & Taberna Torres, J. (2015). Modelling E-portfolio for a Linear Algebra undergraduate course. *International Journal of Education and Information Technologies*, 9, 115–121. <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.828.6140&rep=rep1&type=pdf>

Elbow, P. (1997). High stakes and low stakes in assigning and responding to writing. *New Directions for Teaching and Learning*, 1997(69), 5–13. <https://doi.org/10.1002/tl.6901>

FiveThirtyEight. (n.d.). *Data*. <https://github.com/fivethirtyeight/data>

Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report ASA Revision Committee. (2016). *Guidelines for Assessment and Instruction in Statistics Education college report*. American Statistical Association. https://www.amstat.org/docs/default-source/amstat-documents/gaisecollege_full.pdf

Gelman, A. (2015). *What to do in 2015: Your statistics diary*. Columbia University. <https://statmodeling.stat.columbia.edu/2015/01/07/2015-statistics-diary/>

Glassey, R. (2019). Adopting Git/Github within teaching: A survey of tool support. In *CompEd '19: Proceedings of the ACM Conference on Global Computing Education* (pp. 143–149). <https://doi.org/10.1145/3300115.3309518>

Hardin, J., Hoerl, R., Horton, N. J., Nolan, D., Baumer, B., Hall-Holt, O., Murrell, P., Peng, R., Roback, P., Temple Lang, D., & Ward, M. D. (2015). Data science in statistics curricula: Preparing students to “think with data.” *The American Statistician*, 69(4), 343–353. <https://doi.org/10.1080/00031305.2015.1077729>

Hayden, R. (1989). Using writing to improve student learning of statistics. *Writing Across the Curriculum*, 1(1), 3–9. <https://doi.org/10.37514/WAC-J.1997.8.1.18>

Keeler, C. M. (1997). Portfolio assessment in graduate level statistics courses. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 165–178). IOS Press.

Kinnaird, K. (2021). *Statistical & data sciences writing plan*. Smith College.

<https://www.smith.edu/sites/default/files/media/Documents/Jacobson-Center/SDS-WritingPlan-accessible.pdf>

Koca, S. A., & Lee, H. J. (1998). Portfolio assessment in mathematics education. *ERIC Digest*.

<https://www.ericdigests.org/2000-2/portfolio.htm>

Kross, S., & Guo, P. J. (2019). 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* (Paper No. 263). <https://doi.org/10.1145/3290605.3300493>

Kuh, G. D. (2008). *High-impact educational practices: What they are, who has access to them, and why they matter*. Association of American Colleges & Universities.

Laderas, T. (n.d.). *Portfolio example*. <https://github.com/laderast/portfolio-example>

Lupi, G., & Posavec, S. (2016). *Dear Data*. Princeton Architectural Press.

Lupi, G., & Posavec, S. (2018). *Observe, collect, draw!* Princeton Architectural Press.

Marlow, J., & Dabbish, L. (2013). Activity traces and signals in software developer recruitment and hiring. In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work* (pp. 145–156).

<https://doi.org/10.1145/2441776.2441794>

McLeod, S. H., & Soven, M. (Eds.). (2000). *Writing across the curriculum: A guide to developing programs*. WAC Clearinghouse.

McNamara, A. (2018, March 20). *Wikipedia in the classroom: Gender, argh* [blog post].

<https://www.amelia.mn/blog/teaching/2018/03/20/Wikipedia-in-the-classroom.html>

Mock, T. (n.d.). *Tidy Tuesday*. GitHub. <https://github.com/rfordatascience/tidytuesday>

Munroe, R. (n.d.). *Simple Writer*. Xkcd. <https://xkcd.com/simplewriter/>

National Academies of Sciences Engineering Medicine (NASEM). (2018). *Data science for undergraduates: Opportunities and options*. <https://www.nap.edu/catalog/25104/data-science-for-undergraduates-opportunities-and-options>

National Academies of Sciences Engineering Medicine. (2020). *Roundtable on Data Science Postsecondary Education*. <https://www.nationalacademies.org/our-work/roundtable-on-data-science-postsecondary-education>

Nolan, D., & Stoudt, S. (2020). Reading to write. *Significance*, 17(6), 34–37. <https://doi.org/10.1111/1740-9713.01469>

Nolan, D., & Stoudt, S. (2021). *Communicating with data: The art of writing for data science*. Oxford University Press. <https://doi.org/10.1093/oso/9780198862741.001.0001>

Our World in Data. (n.d.). *Research and data to make progress against the world's largest problems*. <https://ourworldindata.org/>

Parke, C. S. (2008). Reasoning and communicating in the language of statistics. *Journal of Statistics Education*, 16(1). <https://doi.org/10.1080/10691898.2008.11889555>

R for Data Science Online Learning Community. (n.d.). "R4DS Online Learning Community" <https://www.rfordatasci.com/>

Reinhart, A., Evans, C., Luby, A., Orellana, J., Meyer, M., Wiecezorek, J., Elliot, P., Burckhardt, P., & Nugent, R. (2020). Think-aloud interviews: A tool for exploring student statistical reasoning. *arXiv*. <https://doi.org/10.48550/arXiv.1911.00535>

Ring, G. L., Waugaman, C., & Brackett, B. (2017). The value of career ePortfolios on job applicant performance: Using data to determine effectiveness. *International Journal of ePortfolio*, 7(2), 225–236. <https://files.eric.ed.gov/fulltext/EJ1159904.pdf>

Robinson, D. (2017). *Advice to aspiring data scientists: Start a blog*. Variance Explained. <http://varianceexplained.org/r/start-blog/>

Robinson, D. (n.d.). *Tidy Tuesday live screencasts* [YouTube channel]. <https://www.youtube.com/channel/UCeiiqmVK07qhY-wyg3IZiZQ/videos>

Robinson, E., & Nolis, J. (2020). *Build a career in data science*. Manning.

Scolere, L. (2019). Brand yourself, design your future: Portfolio-building in the digital age. *New Media & Society*, 21(9), 1891–1909. <https://doi.org/10.1177/1461444819833066>

Shrestha, N., Barik, T., & Parnin, C. (2021). Remote, but connected: How #TidyTuesday provides an online community of practice for data scientists. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1), (Article 52). <https://doi.org/10.1145/3449126>

Sievert, C. (2020). *Interactive web-based data visualization with R, plotly, and Shiny*. CRC Press.

Sole, M. A. (2012). The mathematics portfolio: An alternative tool to evaluate students' progress. *Journal of Mathematics Education at Teachers College*, 3(1). <https://doi.org/10.7916/jmetc.v3i1.739>

Steen-Utheim, A., & Hopfenbeck, T. N. (2018). To do or not to do with feedback: A study of undergraduate students' engagement and use of feedback within a portfolio assignment design. *Assessment & Evaluation in*

Higher Education, 44(1), 80–96. <https://doi.org/10.1080/02602938.2018.1476669>

Stoudt, S. (2021, January 27). Data visualization beyond the screen. *Northeast Big Data Innovation Hub*. <https://nebigdatahub.org/data-visualization-beyond-the-screen/>

Sutton, B. (1997). Writing in the disciplines, first-year composition, and the research paper. *Language and Learning across the Disciplines*, 2(1), 46–57. <https://doi.org/10.37514/LLD-J.1997.2.1.04>

Svyantek, M. V., Kajfez, R. L., & McNair, L. D. (2015). Teaching vs. research: An approach to understanding graduate students' roles through ePortfolio reflection. *International Journal of ePortfolio*, 5(2), 135–154. <http://files.eric.ed.gov/fulltext/EJ1107857.pdf>

Townsend, M. A., Dobler, C. P., Jordan, J., & Miller, J. (2003). Improving statistical understanding: Using writing in the statistics classroom. *Joint Statistical Meetings*. https://apps3.cehd.umn.edu/artist/articles/JSM_Symposium.pdf

WAC Clearinghouse. (n.d.). *What designs are typical for WAC programs?* <https://wac.colostate.edu/resources/wac/intro/design/>

Woodard, V., Lee, H., & Woodard, R. (2020). Writing assignments to assess statistical thinking. *Journal of Statistics Education*, 28(1), 32–44. <https://doi.org/10.1080/10691898.2019.1696257>

Appendix

Brief Descriptions of Portfolio Assignments

Writing – Formal

Ideal data set

This assignment helps students be concrete both in picking a question of interest that has a quantifiable/measurable answer and specifying what information they need and how they would obtain it. Students pick a scientific or social question that they wonder about and describe the ideal data set needed to answer the question. This data set description will include the data collection setup, the cases and variables of interest, and an approximate order of magnitude for the desired sample size. You can also ask students what could go wrong in the data collection and what other related questions this data might be useful for.

Livestreams

This assignment lets students learn from others' thought processes and see that trial and error is part of everyone's data science experience. Students watch someone livestream (e.g., Robinson, n.d.) their coding or

data analysis and write about what they observe. Students might talk about what they learned, what they would do differently, or lingering questions they have for the coder or analyst.

Captions

This assignment helps students recognize the important role of captions as a component of a data visualization. Students can take a variety of approaches to practicing caption writing including writing captions for provided visualizations, revising preexisting captions, and creating their own visualizations and captioning them. Captions can also be practiced in the context of accessibility and alt-text (Cesal, 2020).

Peer review

This assignment gives students experience with peer review as a step in the academic process of sharing ideas and findings with a broader research community. Students read another's work, summarize it, and identify the strengths and weaknesses at both the content and presentation levels. Specific instructions for what to pay attention to can be customized by the instructor depending on the goals of the original assignment.

Storyboard

This assignment helps students find a narrative in their work and helps them break away from the tendency to talk about work in the order of their process. Students take the tables and figures they have made as part of the exploration and analysis processes, group them, and organize them to tell a story. This is an iterative process and may reveal gaps in the analysis that students must revisit. Instructors may also provide a series of tables and figures for students to arrange in a storyboard to practice the storytelling without the analysis overhead. See also [our conference presentation recording](#).

Wikipedia

This assignment helps students write about a technical topic for a nontechnical audience. Students choose to revise a data-related topic on Wikipedia to make it more clear, accessible, or comprehensive. Students may also want to write about a statistician or data scientist (McNamara, 2018).

Read to write

This assignment helps students prepare to write their own formal report by learning from another's work. Students choose a data-related paper, examine the argument, and map the organization. Further details are described in Nolan and Stoudt (2020).

Personal statement

This assignment allows students to reflect on their data-science identity. Students may model this reflection on a personal statement for graduate school, a cover letter for a data-related job, or a blog post. The statement may

include a reflection on their portfolio so far and their growth throughout their academic program.

Revise and resubmit

This assignment gives students experience with the revise and resubmit process in academic publishing. Students respond to peer or instructor feedback by not only revising their work, but also by explicitly responding to feedback with an explanation of what they changed. This can help avoid superficial revision.

Book report

This assignment allows students to read a data-related book that interests them and write about it. Students may choose to write a summary for a nontechnical audience, review the book's strengths and weaknesses, or reflect on what they learned or still wonder after reading.

General reflection

This assignment can accompany any other assignment to encourage students to think about what they have accomplished or learned, what strategies worked or did not work in the process, and how different skills or assignments fit together in relation to their career goals.

Writing – Broader Audience

Talk summary

This assignment helps students see how data is used in different fields. Students identify a talk happening on campus or online where the speaker uses data in some way and write about it. Students may choose to write a summary for a nontechnical audience, review the talk's strengths and weaknesses, or reflect on what they learned or still wonder after attending.

Simple writer

This assignment helps students replace jargon in their explanations and communicate their work in a more accessible way. Students explain their work, another's work, or a data-related topic using only common language, perhaps guided by an application that flags uncommon words (Munroe, n.d.).

Blog post

This assignment helps students write for a broader audience in a shorter, more informal style. Students may write a blog post on a range of topics, including a personal reflection on the exploration or learning process in data science or an explainer of a data-related method or finding. These should be a quick read and accessible to a reader without any data science experience.

Data diary

This assignment helps students collect data-related ideas that come to them throughout a day, week, or course (Gelman, 2015). Students informally keep track of passing thoughts or sources of inspiration. These could include new data sources, questions they may want to answer using data, or others' data work that they like and want to try to emulate. These diary entries can become fodder for other assignments such as writing a blog post or providing inspiration for a future project.

Press release

This assignment helps students share their work more broadly. Students write about their findings as if a journalist was reporting on it. This requires thinking about what the most important parts of their work are and writing about it in an engaging and approachable way that is still faithful to the data. It can be helpful to first read some press releases from the press office of their own institution.

Numbers in the news

This assignment helps students reflect on how numbers and data are talked about in news sources. Students find the use of numbers in a news article (e.g., polls, summaries of research findings, etc.) and assess how effectively they are communicated. This effectiveness discussion might include comments about the accuracy as well as the accessibility to the desired audience.

Visual – Formal

Interactive visualization exploration

This assignment helps students learn from their experience as a user of an interactive visualization to inform later creation. Students find an interactive visualization, for example, from a news source or another data scientist's portfolio, and describe the process of exploring it. This description could include what they found, how the interactivity worked for them, and what additional features they would have liked to explore.

Visualization collector

This assignment helps students identify their own visualization style and aesthetic by collecting inspiration from others' work. Students collect visualizations from a variety of outlets, including from journalists and other data scientists' portfolios. They then work to articulate why they like particular visualizations, find common themes, and make goals about how to incorporate these approaches into their own work.

Tidy Tuesday

This assignment connects students to a virtual community of practice while giving them an opportunity to practice creating visualizations and finding a story in a data set (Shrestha et al., 2021). Tidy Tuesday is a weekly project, organized and fostered by an online community. The data set is posted each week for

participants to explore, visualize, and share findings (Mock, n.d.). Participants share their process, visualizations, code, and findings through social media, blogs, and GitHub. Students can participate live with the current data set of the week or use a previous data set, all archived on the project's GitHub repository, that they find interesting. Variations on this activity may include making one publication-worthy plot, writing a blog post about their exploration process, sharing tidy and well commented code, creating a series of prototyped figures that work together to tell a story about the data set, or even suggesting a data set to be featured in the future via a GitHub issue.

Reproduce a figure

This assignment familiarizes students with open data and reproducibility principles. Students choose a visualization whose source data accompanies the source (e.g., FiveThirtyEight, n.d.; Our World in Data, n.d.) and try to recreate the figure as closely as possible. Students may also wish to iterate and try to improve the plot, discussing why they made different choices from the original creator.

Visual – Broader Audience

Dear Data

This assignment allows students to explore an artistic, hand-drawn approach to data visualization. Students collect data about themselves for a week and then sketch and refine a visualization inspired by the Dear Data postcards of Giorgia Lupi and Stefanie Posavec (Lupi & Posavec, 2016, 2018).

Tactile data visualization

This assignment helps students think about a different audience, those with visual impairments, and confront the accessibility of their data communication approach. Students are tasked with creating a visualization whose key elements are tactile rather than visual. They may use materials such as clay, fabrics with different textures, and other crafting materials. Further details are described in Stoudt (2021).

Visualization with quick finding

This assignment helps students identify their core message and create one visualization that quickly imparts that message on their audience. Students synthesize a larger project into one takeaway and workshop a visualization that quickly reveals that takeaway. This will likely take an iterative approach as students refine their message and get feedback from peers or the instructor about the effectiveness of the visualization itself.

Interactive visualization creation

This assignment allows students to explore another layer of visualization where the user has some control over their experience. Students may create wrapper functions that allow users to easily change aspects of a more

complicated plot, explore R tools like plotly or Shiny (Sievert, 2020), or take advantage of other interactive platforms such as Datawrapper (Datawrapper, n.d.).

Rubrics

Table A1. Sample blog post rubric. This sample rubric is used for assessing a blog post. Each row corresponds to a core competency that we are looking for the student to demonstrate in the blog post. The level of competency is reflected in the columns of the table (needs improvement, basic, and surpassed), and examples of how to distinguish these levels appear in the cells of the table.

Core Competency	Needs Improvement	Basic	Surpassed
Audience	unclear who the audience is	audience identified but writing is mixed in effectiveness of writing for audience	clear audience and written appropriately for that audience
Goal	unclear what the goal of the post is	goal identified but writing is mixed in effectiveness for reaching the goal	clear goal and goal reached
Engaging blog style	jargony or structured like a formal report	blog style is mixed in effectiveness, reads like formal report	blog style is effective, new spin from a typical writing about data
Pace and forward motion	too long and ‘yo-yos’ back and forth between trains of thought	reasonable length with some backtracking and/or tangents	reasonable length, continually drives the post forward toward the destination promised by the headline
General Writing	frequent technical errors in the writing	writing is technically correct but some awkward phrasing	writing is technically correct, flows well, and is free of grammar and spelling errors

Table A2. Sample visualization rubric. This sample rubric is used for assessing a visualization. Each row corresponds to a core competency that we are looking for the student to demonstrate in the visualization. The level of competency is reflected in the columns of the table (needs improvement, basic, and surpassed), and examples of how to distinguish these levels appear in the cells of the table.

Core Competency	Needs Improvement	Basic	Surpassed
-----------------	-------------------	-------	-----------

Appropriateness	incorrect chart type for data	correct chart type for data but mixed effectiveness of aesthetic choices and mappings	appropriate chart type for data and appropriate use of color, size, shape, and other aesthetic properties mapped to data
Informativeness	no or default axis labels	axis labels and annotations used but have mixed effectiveness in their readability	axis labels and other annotations are readable and informative
The Takeaway	flaws in plot overwhelm the message	aesthetically correct and appropriate plot, but story is obscured	uncovers important structure in data and is easily interpreted in the context of problem or question
Caption	no or minimal caption	caption includes description of what was plotted but doesn't take the extra step of adding context or pointing out the main takeaway	caption includes description of what has been plotted, the conclusion that the reader should take away, and details about content of plot or creation of it

Examples of Subset of Portfolio Assignments

*Dear Data**
Spring 2021

**based on the Dear Data Project by Lupi and Posavec*

Figure A1. Sample Dear Data. Nolan's sample hand-drawn visualization for her seminar based on the Dear Data project (Lupi & Posavec, 2016, 2018).

As part of the optional, public-facing nature of the gateway course described in the manuscript, the archive of the course newsletter has some examples that students approved sharing outside of the classroom environment. They can be found in posts #1–#5 at the following link: <https://buttondown.email/sstoudt/archive>. Similarly, the data “phys” project is documented at the following link: <https://nebigdatahub.org/data-visualization-beyond-the-screen/>.

As a spin-off independent study inspired by the writing seminar described above and taught by Nolan and Stoudt, students published blog posts that can be found at the following link: <https://stat198-spring18.github.io/>.

More sample student work from Nolan’s Dear Data seminar and Adam Anderson’s Writing Data Stories Connector Course is in the process of being published online (links forthcoming).

The companion website to Nolan and Stoudt’s book, found here <https://communicating-with-data.github.io/>, will also showcase student work contributed by users of the book.

©2021 Deborah Nolan and Sara Stoudt. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](https://creativecommons.org/licenses/by/4.0/), except where otherwise indicated with respect to particular material included in the article.