Data Science Transfer Pathways From Associate's to Bachelor's Programs

Benjamin S. Baumer[1] Nicholas Jon Horton[2]

[1]Statistical and Data Sciences Program, Smith College, Northampton, Massachusetts, United States of America,
[2]Department of Mathematics and Statistics, Amherst College, Amherst, Massachusetts, United States of America

ABSTRACT

A substantial fraction of students who complete their college education at a public university in the United States begin their journey at one of the 935 public 2-year colleges. While the number of 4-year colleges offering bachelor's degrees in data science continues to increase, data science instruction at many 2-year colleges lags behind. A major impediment is the relative paucity of introductory data science courses that serve multiple student audiences and can easily transfer. In addition, the lack of predefined transfer pathways (or articulation agreements) for data science creates a growing disconnect that leaves students who want to study data science at a disadvantage. We describe opportunities and barriers to data science transfer pathways. Five points of curricular friction merit attention: 1) a first course in data science, 2) a second course in data science, 3) a course in scientific computing, data science workflow, and/or reproducible computing, 4) lab sciences, and 5) navigating communication, ethics, and application domain requirements in the context of general education and liberal arts course mappings. We catalog existing transfer pathways, efforts to align curricula across institutions, obstacles to overcome with minimally disruptive solutions, and approaches to foster these pathways. Improvements in these areas are critically important to ensure that a broad and diverse set of students are able to engage and succeed in undergraduate data science programs.

Keywords: articulation, associate's programs, bachelor's programs, two-year colleges, community colleges

Media Summary

The path from a 2-year college to a lucrative career in data science is beset by unnecessary curricular roadblocks, according to new research published in the *Harvard Data Science Review*.

In the United States, the job market in data science is hot, due to high salaries, expansive job growth, and comfortable working conditions. According to Glassdoor, data scientist is the third-best job in America for 2022, and has ranked among the top three every year since 2016. To prepare students to fill these positions, curricula at 4-year colleges are evolving at breakneck speeds (by their standards), but 2-year colleges are lagging behind. Because so many students who complete their college education at a public university begin their journey at one of the 935 public 2-year colleges, creating pathways from 2-year colleges to the data science job market is vitally important. The new research identifies opportunities and barriers to data science transfer pathways, calls attention to points of curricular friction, and offers minimally disruptive solutions to build these pathways.

1. Introduction

Two-year colleges (historically known as community colleges or junior colleges) play a critical role in higher education in the United States.

As of 2020, these 935 public institutions enroll more than 4.7 million students (Duffin, 2022). Two-year colleges provide associate's degrees that lead directly to employment, as well as options to transfer to bachelor's programs. In our home state of Massachusetts, the enrollment at the 15 public 2-year colleges is comparable to that of the undergraduate enrollment of the University of Massachusetts (UMass) system. While average tuition varies considerably by state, 2-year colleges are the most effective and affordable option for many students. Blumenstyk (2021) describes 2-year colleges as "the keystone for the nation's plan to help more people earn a postsecondary credential."

By all accounts, job prospects in data science are excellent, due to high salaries, expansive job growth, and comfortable working conditions. According to Glassdoor, data scientist is the third best job in America for 2022, and has ranked among the top three every year since 2016. The U.S. Bureau of Labor Statistics reports a mean annual wage of $103,930 for data scientists, and estimates that jobs will grow 22% for Computer and Information Research Scientists and 33% percent for Mathematicians and Statisticians over the next 10 years. A number of companies have reported that they cannot find sufficient skilled candidates for these positions. Due to the nature of the work, data scientists have adapted smoothly to working remotely, an increasingly relevant factor that should only improve employment prospects. The high probability of financial success for graduates in data science stands in stark contrast to the increasingly dim prospects for many master's students in other fields. Korn and Fuller (2021) conclude that 38% of master's programs at top-tier private universities in the United States do not deliver on the promise of earnings that exceed debt incurred to pay for tuition.

Providing equitable access to these desirable jobs is a challenge that is symptomatic of larger issues of class and income inequality in the United States. Several national reports (e.g., National Academies of Sciences, Engineering, and Medicine [NASEM] 2016, 2018; Rawlings-Goss et al., 2018) recognize this challenge and call for tighter partnerships between 2- and 4-year colleges. If the field of data science is serious about diversifying its workforce, then there must be paths to high-paying jobs in data science that begin at 2-year colleges, which enroll a much larger fraction of historically underserved students than 4-year colleges.

The ongoing National Science Foundation's (NSF) Data Science Corps (DSC) program focuses on creative approaches to developing a competitive and diverse workforce in data science. Through our roles as leaders of the NSF-funded DSC-WAV (Wrangle, Analyze, Visualize) program, we have had the opportunity to engage in data science projects with community organizations and to work with partners at several 2-year colleges to foster new courses and programs. This work included organizing a symposium (Data Science Symposium: Opportunities for Massachusetts Community Colleges) for academic leaders on June 13, 2022 and faculty development workshops in 2021 and 2022.

1.1. Our Contribution

The purpose of this article is to help foster connections between 2- and 4-year institutions that will lead to more transparent and flexible pathways to bachelor's degrees in data science. While we use Massachusetts as our primary example, we believe that the insights and approaches we suggest may be useful to other states. We focus exclusively on data science, including cognate disciplines of mathematics, computer science, and statistics only as they relate to data science.

We begin by briefly surveying the landscape of data science in higher education nationally (Section 2). The lack of existing transfer pathways make a bachelor's degree in data science burdensome for a 2-year college student to achieve without significant—and probably unreasonable—foresight and perseverance through administrative and bureaucratic obstacles. We use two hypothetical community college students named Alice and Bob to illustrate how these obstacles impede student progress. In Section 3, we use the bachelor's program in data science at the University of Massachusetts Dartmouth (UMassD) (which we see as representative of a curricular consensus in data science) as an example, analyze potential transfer pathways, and identify five points of friction. Our analysis leads directly to recommendations that could provide explicit pathways in data science with relatively few new courses and modest impact on existing programs (Section 4). We conclude with final thoughts in Section 5.

2. Background and Related Work

2.1. Data Science Programs in Higher Education

Since the action plan for data science articulated by Cleveland (2001), the field has continued to blossom within academia. Academic data science can be described in aspirational terms using a pyramid, with doctoral degrees rare but important for leadership and research in the field. Master's degrees are the next level, with larger numbers and considerable job opportunities. For established disciplines, bachelor's programs (offered at 4-year colleges) and associate's programs (offered at 2-year colleges), make up the third and fourth levels of the pyramid, with larger and larger numbers of students obtaining these degrees. Jobs are available at each level, with the potential for interested students to pursue more advanced degrees in order to deepen skills and expand their work opportunities. However, workforce opportunities remain opaque to too many students.

As an emerging discipline, data science has not yet matured to that extent, with master's programs leading the way, bachelor's programs on the rise, and associate's program lagging behind. Several doctoral programs in data science now exist in the United States (NASEM, 2020) and their graduates are now beginning academic and workforce careers.

Far more common are master's programs in data science and data analytics, which are offered by many universities (both online and in-person). Nationally, the growth in the number of master's degrees granted in

analytics and data science is dramatic, with more than 45,000 degrees reported in 2020 by the Institute of Advanced Analytics.

While the study of data science at the graduate level continues to evolve, its footprint is already substantial. The growth of these programs makes it possible for students at the undergraduate level to more easily identify future programs of graduate study. What undergraduate majors should best prepare a student for graduate study in data science? Computer science, statistics, and mathematics are the closest cognate disciplines, and while statistics is not always available as an undergraduate major, it is taught everywhere and can be folded into either a computer science or mathematics major, both of which are available at virtually any institution.

Historically rarer (but increasingly less so) are bachelor's degrees in data science and related fields (e.g., data analytics). These programs make up the next level of the pyramid, with larger numbers of students potentially entering the workforce (NASEM, 2018). The options—which are certain to grow in the coming years—already provide 2-year college students who are interested in data science with visible future programs of study. Gould et al. (2018) identifies six associate's degree programs in three states, including New Hampshire, Pennsylvania, and Minnesota. A number of exemplary associate's data science programs have been established in recent years (New Two-Year College Data Science, Analytics Programs on the Rise, 2022). Many others have been created across the nation.[1]

2.2. Key Concepts in Two-Year College Education

Two-year college students typically pursue associate's degrees that come in two flavors: terminal or transfer. Many associate's degrees are *terminal* (often called associate's-to-workforce), in that they are designed to prepare students for employment directly upon completion. Other associate's degrees are designed to prepare students for a smooth transfer to a 4-year institution (and even a specific bachelor's degree program at that institution) upon completion. For example, Springfield Technical Community College offers multiple degrees in computer science. The Computer Systems Engineering Tech program prepares students for various systems administration jobs after two years of study. Conversely, the Computer Science Transfer program prepares students to *transfer* to a bachelor's program in computer science, with most students presumably intending to transfer to one of the UMass campuses.

We use the term *pathway* to describe a route that a 2-year college student could take to obtain a bachelor's degree. Associate's degrees designed for transfer, as described above, are the most well-trod starting places for such pathways. But even with an associate's degree for transfer in hand, pathways are not always obvious. Many states, including Massachusetts and California, have highly visible public websites that map transfer pathways from 2-year colleges to public 4-year colleges. However, not all of these many-to-many possible pathways are mapped. For example, Bunker Hill Community College (BHCC) offers a Computer Science Transfer associate's degree, but there is no corresponding mapping to any of the UMass campuses in the system (see Section 3).

Further complicating matters are *articulation agreements*, which provide an explicit transfer pathway between one specific associate's degree program and one specific bachelor's program. These agreements may be negotiated between public or private 4-year institutions. While these one-to-one articulation agreements are helpful, they are not as visible as the many-to-many mapped pathways.

2.3. Outcomes for Undergraduates

The choice of which flavor of associate's degree to pursue has consequences for the 2-year college student. Many workforce roles for data scientists exist at the bachelor's level (De Veaux et al., 2017; NASEM, 2018), and the number is growing (Gould et al., 2018).

For those who choose further study, the bachelor's-to-master's transition is characterized by flexibility and adaptation, because graduate schools know that they will receive applications from students who attended a wide variety of undergraduate schools, and who studied highly variable subjects therein. Moreover, bachelor's programs typically involve at least 120 credit hours of study, which often provides ample flexibility for a student to deviate from any predefined curricular path. From our own experiences, we know that it is not uncommon for a traditional bachelor's student to major in say, economics, only to then decide before their senior year that they want to pursue a master's degree in data science, load up on statistics and computer science courses in their senior year, and still put together a competitive graduate school application.

It is important to remember that dramatically less flexibility is available for the associate's-to-bachelor's transition, since for 2-year college students, every credit counts. We recognize that for most 2-year college students, any credit that does not count toward their associate's degree program or their predefined transfer pathway may be considered a 'waste' of both time and money. California has been a leader in fostering smoother articulation of courses between 2-year and 4-year institutions (see https://assist.org). But while the California system provides a clear solution for existing pathways, the larger difficulties with transfer pathways are longstanding (Blumenstyk, 2021). In Massachusetts, although most students who enroll in a 2-year college program after high school intend to transfer to a bachelor's degree program, relatively few actually do so (Murnane et al., 2022).

Longer term, alternative options, including associate's-to-workforce programs (Gould et al., 2018; Rawlings-Goss et al., 2018) are desirable but outside the scope of this article. Associate's programs in cybersecurity, information technology, and web development—designed as terminal degrees—have proven effective in workforce development and the same potential exists for data science.[2]

2.4. Data Science Curricula

Undergraduate curricula in data science are now beginning to coalesce. De Veaux et al. (2017) provide curriculum guidelines for undergraduate majors in data science that are endorsed by the American Statistical Association. The "Data Science for Undergraduates: Opportunities and Options" consensus study (NASEM,

2018) provided a number of recommendations and findings relevant to undergraduate data science programs and outlined key aspects of *data acumen*. The Association for Computing Machinery (ACM) Data Science Task Force enumerated computing competencies for undergraduate data science curricula (Danyluk et al., 2021), and syllabi from example courses. Gould et al. (2018) provides curricular guidelines for 2-year college programs in data science. Comprehensive textbooks (Baumer et al., 2021; Wickham & Grolemund, 2016) and course materials (Çetinkaya-Rundel, 2020) support the teaching of a variety of different introductory data science courses. Donoho (2017) ruminates on the nature of data science as a standalone scientific discipline.

In 2019, the National Center for Education Statistics unveiled a new series of Classification of Instructional Programs (CIP) codes for data science (with the numerical designation 30.70). These new codes allow the federal government to track the growth of programs in data science and should result in an improved ability to quantify how many students are studying data science.[3]

In what might be an important stamp of legitimacy, ABET (Accreditation Board for Engineering and Technology) has begun accrediting its first undergraduate data science programs, with plans to expand to the graduate and associate's levels.

2.5. The DSC-WAV Project

While our interest in data science education is longstanding and well-documented, our specific interest in 2-year college pathways in data science is motivated by our involvement in the Data Science Corps (DSC): Wrangle, Analyze, Visualize (WAV) project (Horton et al., 2021). The first arm of the NSF-funded program links teams of undergraduate students (often data science majors) at the Five Colleges (Amherst, Hampshire, Mount Holyoke, and Smith colleges plus the University of Massachusetts-Amherst) with local, community-based organizations in the service of a real-world data science problem. Legacy et al. (2022) details how this program supports the growth of DSC-WAV student participants.

As the Data Science Corps is a workforce development initiative, the DSC-WAV project has an additional goal of growing and diversifying the data science workforce. In this fast-growing segment of the economy, highly satisfying, high-paying jobs are plentiful. After several years of working closely with our partners at Holyoke, Greenfield, and Springfield Technical community colleges on a variety of curricular- and student-focused issues, our attention is now centered on the pathway predicament. We believe that while the obstacles to transfer pathways in data science are formidable, we can overcome them with relatively nondisruptive changes. Our current focus is to help identify impediments to creating flexible and transparent transfer agreements between 2-year colleges and public universities in Massachusetts.

3. Pathways to Data Science

Explicit transfer pathways are important. The *Data Science for Undergraduates* consensus report (NASEM, 2018) recommends that: "Academic institutions should provide and evolve a range of educational pathways to

prepare students for an array of data science roles in the workplace" and that "Four-year and two-year institutions should establish a forum for dialogue across institutions on all aspects of data science education, training, and workforce development" (p. 56).

Transfer pathways for mature disciplines are well established nationally. For example, the California Assist system and the Massachusetts MassTransfer A2B Degree maps provide students in those states with easy-to-navigate, public listings of transfer pathways within their respective public higher education systems. These websites allow any student to select a 2-year college, a public university, and an intended bachelor's degree field. The system will then return a list of mapped pathways that have been preapproved by the state's Board of Higher Education for transfer. While less formal or one-to-one articulation agreements between individual programs may permit direct transfer from a 2-year college to a 4-year program, these public mapping systems are the best mechanisms for broadcasting important signals to prospective students that their academic plan is sound. Unfortunately, few pathways in data science exist.

A handful of articulation agreements between 2-year and public 4-year colleges do exist. In New Hampshire, the Public Pathways program at the University of New Hampshire at Manchester (UNH-Manchester) provides explicit transfer pathways from Great Bay, Manchester, and Nashua Community Colleges to their bachelor's degree in Analytics and Data Science. While these opportunities are advertised on New Hampshire Transfer website, they all point to UNH-Manchester (in contrast to the multiway systems in California and Massachusetts).

Fortunately, as a result of existing mapped pathways for data science–adjacent disciplines like mathematics and computer science, many courses relevant to data science exist and are easy to transfer. This includes mathematics and statistics courses (e.g., statistics, calculus, linear algebra, and discrete math) along with computer science courses (e.g., computer science I and II, data structures, and algorithms). These existing course mappings provide a solid foundation for a transfer pathway in data science—but they are not enough.

Our analysis of the gaps reveals **five points of curricular friction**:

1. A first course in data science (Data Science I)
2. A second course in data science (Data Science II)
3. A course in scientific computing, data science workflow, and/or reproducible computing
4. Lab sciences
5. Navigating communication, ethics, and application domain requirements in the context of general education and liberal arts course mappings.

Our analysis comes with recommendations for solutions. Some of these points of friction are best solved through new course offerings at the 2-year college level. Others are likely negotiable using existing courses

through careful planning, advising, and sequencing. Other sources of friction, such as institutional inertia, faculty development and retention, and technology, are no less real, but are not our focus in this article.

In the remainder of this section, we use real-world examples from our home state of Massachusetts to illustrate how these five points of friction present obstacles to transfer pathways, and offer recommendations for how they can be best overcome.

As a form of case study, we consider Alice and Bob, two hypothetical 2-year college students interested in data science. Alice attends Bunker Hill Community College (BHCC) and is pursuing the Data Analytics associate's degree. Bob attends Holyoke Community College (HCC) and is pursuing the Mathematics MassTransfer associate's degree. Both have the goal of obtaining a bachelor's degree in data science from one of the UMass campuses.

3.1. Obstacles to Transfer Pathways

For purposes of illustration, we focus on the [bachelor's program in data science at the University of Massachusetts Dartmouth](#) (UMassD). We believe that this is a helpful example because the program has existed for a number of years, conforms reasonably well to other curricular guidelines in data science, and has graduated a number of students since the program was established. However, most of our analysis and its implications are not specific to this program, and should be relevant to other programs that may be proposed in other states.[4]

Figure 1 illustrates the flow through the eight semesters of the UMassD data science major. In order for a transfer pathway to be viable, a student would have to complete the equivalent of the top half (61 credits) of the flowchart at a 2-year college. Nearly all of the courses coded as mathematics (pink), computer science (tan), lab sciences (light blue), English (gray), and university electives (green) have existing mapped equivalents at many two-year colleges in Massachusetts. However, the two data science courses (yellow) have no equivalents at the 2-year college level.

In [Section 4](#), we use Alice to illustrate why a proposed pathway from BHCC to UMassD is necessary, and how it could work.
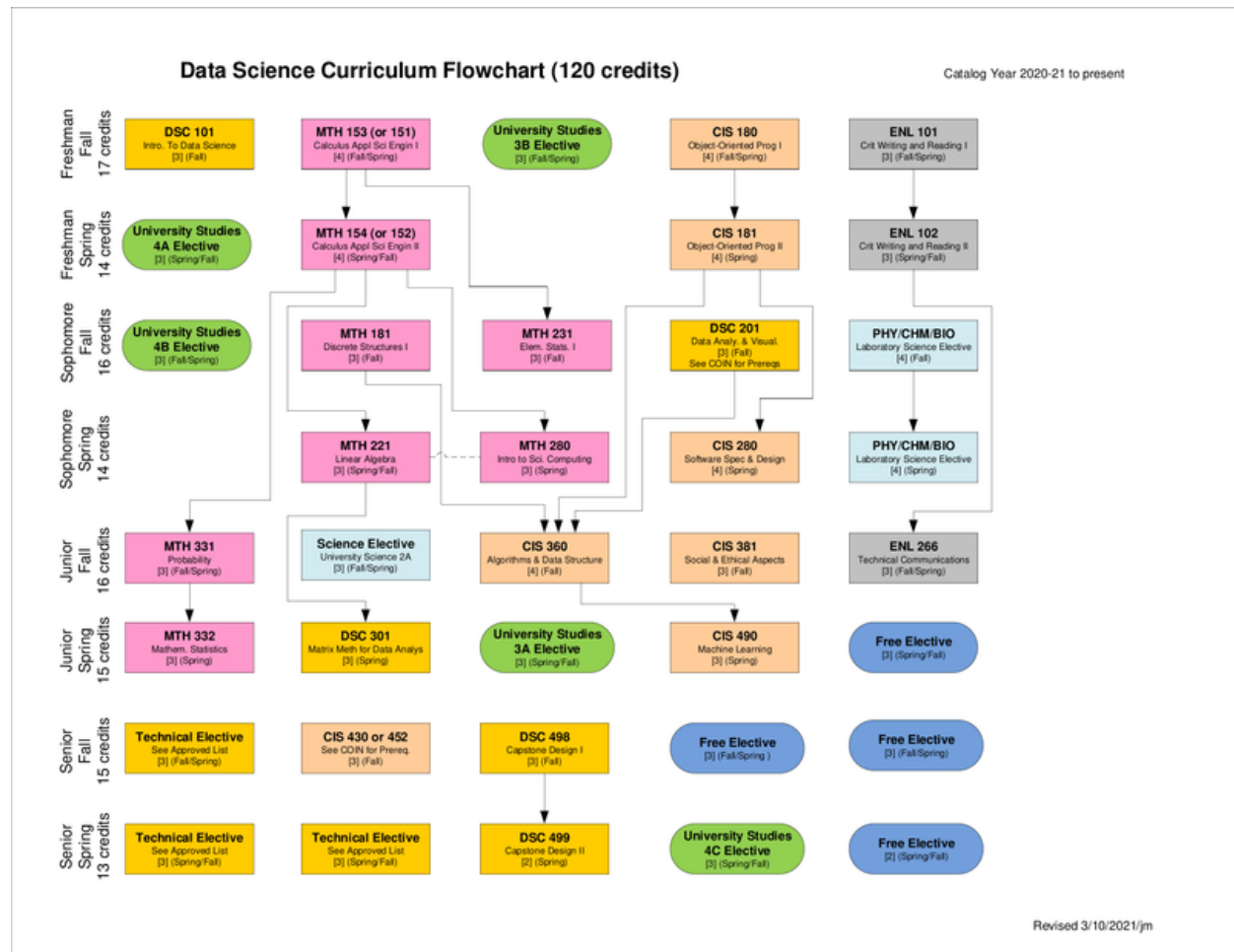
Figure 1. Flowchart illustrating progress through the data science bachelor's program at UMassD. For a transfer pathway to work, roughly the top half of the flowchart must be completed at a 2-year college. Data science courses appear in yellow rectangles.

In this section, we focus on Bob. Recall that Bob is pursuing the Mathematics MassTransfer option at HCC. This degree will set him up to seamlessly transfer to almost all of the public 4-year institutions in Massachusetts, including UMass Amherst (his first choice) and UMass Dartmouth. Bob is from the Holyoke area and wants to stay in Western Mass; UMass Amherst is nearby and many of his friends are already there. It is also the flagship campus and probably comes with the greatest prestige and employment prospects. The path of least resistance for Bob would be to transfer to UMass Amherst and major in mathematics. However, Bob took the new Introduction to Data Science course at HCC during his last semester, and now he is interested in data science, which he believes has more plentiful and lucrative employment options than mathematics. He knows his background in mathematics will serve him well, but he needs to learn more about statistics and computer programming to build out his data science skill set. As of 2022, UMass Amherst does not offer a full-blown bachelor's degree in data science, but UMass Dartmouth does. Bob faces a tough choice:

- If he transfers to UMass Amherst and pursues a bachelor's in mathematics, he probably won't have room in his schedule to flesh out his data science skills, and his employment options may suffer.
- If he transfers to UMass Amherst and then tries to change his major (perhaps to informatics with a data science concentration), he might not be able to finish his bachelor's degree in 2 years, costing him both time and money.
- If he transfers to UMass Dartmouth, he is missing several important prerequisites, so he would not be able to jump right into the junior-year data science curriculum, and risks falling even further behind while he struggles to catch up. Plus, he will be further from the support mechanisms he has closer to home.
- If he decides to stay at HCC for another semester (or another year) so that he can take computer science classes that are outside of his Mathematics MassTransfer degree, he will complicate his financial aid situation and likely have to pay out of pocket.

Bob is stuck. One reason that Bob is stuck is that he is trying to change the direction of his course of study midway through his undergraduate experience. However, it is worth emphasizing that that is much more problematic *because he started at a 2-year college*. Further, if a data science transfer pathway existed, Bob would likely have a better chance of executing his switch, or at the very least, might have recognized his interest in data science earlier in his journey. Thus, we contend that the lack of data science transfer pathways negatively impacts 2-year college students relative to their 4-year college peers.

> Visit the web version of this article to view interactive content.

3.1.1. Data Science I

Consistent with the recommendations of De Veaux et al. (2017) and Gould et al. (2018), students at UMassD take a first course in data science (Yan & Davis, 2019) in the first semester of their first year. While many other institutions nationally are now teaching introductory data science, not all students have access at their institutions. It is impossible to imagine a sensible transfer pathway in which students are not exposed to the key ideas in data science until their junior year. Irrespective of pathways to degrees, it is critically important that 2-year college students have the opportunity to develop these skills.

Note that such a course is not simply a grab bag of existing material from existing courses in statistics and computer science, but rather focuses on new components of data acumen, including the data science lifecycle, and historically underdeveloped skills like data wrangling and data visualization that support—but are not subsumed within—those existing courses. Some courses (e.g., Data 8 at Berkeley and Çetinkaya-Rundel, 2020) include elements of statistical modeling and inferential statistics, while others (e.g., SDS 192 at Smith College) do not. In either case, a first course in statistics is a separate requirement under many existing transfer pathways in mathematics.

In order for data science transfer pathways to work, 2-year colleges must offer a first course in data science. This is by far the largest obstacle to bringing these pathways online and the place where the biggest gain will be achieved in helping institutions to make data science accessible to their students. This will help to partially address the recommendation from (NASEM, 2018) that: "To prepare their graduates for this new data-driven era, academic institutions should encourage the development of a basic understanding of data science in all undergraduates." (p. 22)

Many new introductory data science courses will be developed in the coming years, and it is vital that faculty at the bachelor's and associate's levels coordinate their efforts to ensure that explicit course mappings are created that will facilitate transfer.

> **Our recommendation** is that institutions develop a flexible, shared understanding of what constitutes a first course in data science, and that any new courses developed at any institution are designed with transfer mappings in mind.

Ideally, such a first course would:

1. have minimal prerequisites;
2. dovetail in useful ways with introductory computer science and statistics courses to allow students to take these foundational courses in any order;
3. transfer to a variety of programs at the bachelor's level, and;
4. satisfy a variety of distribution requirements (in Massachusetts, this might include the R2 analytical reasoning designation at UMass Amherst).

At a high level, such a course should prepare students to demonstrate the ability to:

- use a general-purpose computational environment (e.g., Python or R) to analyze data
- scrape, process, clean, and wrangle data from various sources, including relational databases
- visualize and interpret relationships between variables in multidimensional data
- design accurate, clear, and appropriate data graphics
- communicate the results of an analysis in a correct and comprehensible manner
- collaborate within a reproducible workflow
- assess the ethical implications to society of data-based research, analyses, and technology in an informed manner.

Some courses may choose to cover statistical modeling, while others may leave that topic in the introductory statistics course. In any case, new course structures should facilitate an inclusive and engaging learning environment for students. The Dana Center (2021a) has curated a set of course design principles that we suggest be incorporated in the course development process.

Bob is lucky that he took an introductory data science course at HCC. Most 2-year colleges do not yet offer this course.

3.1.2. Data Science II

Cultivating a rich facility in data science requires repeated exposure: a single course is not sufficient for students to develop mastery. To help students along this path, bachelor's programs in data science typically include a second course in data science, often taken during the sophomore year. This course is intended to reinforce and extend fundamental skills in data wrangling, data visualization, statistical modeling, and predictive analytics. A richer treatment of data technologies and database querying in SQL may arise in such a course. The second course may be taught in a different language (e.g., Python) than the first course (e.g., R). The focus of the second course will vary from institution to institution depending on the focus of the first course (see Section 3.1.1), but we expect the general content areas to be similar to those listed above. The Data 100 course at Berkeley and the DSC 201 course at UMassD are examples of second courses in data science.

Second courses in data science obviously depend on a first course, and often build upon on other core requirements, which may include: a first course in programming, a first course in statistics, and/or linear algebra. These prerequisites have an impact on student pathways and may necessitate delaying completion of this course to the sophomore year.[5]

Given the difficulty of launching a first course in data science at 2-year colleges, it may be best, especially in the short-term, to leave the second course in data science to the universities. While not optimal, it may be feasible for transfer students to take their second course in data science during the first semester of their junior year, and while this will likely disrupt their path relative to nontransfer students, that disruption can be minimized.

To see how, note that the fifth row in Figure 1, labeled "Junior Fall" lists courses in probability, algorithms & data structures, social & ethical aspects [of computing], technical communications, and a science elective. Some 2-year colleges offer some of these courses. With appropriate planning, transfer students might be able to take at least one of these courses at their 2-year college in place of a second course in data science, which they would then take upon transferring. This exchange is possible in part because data science straddles mathematics and computer science, and data science students need not complete the entire 2-year college curriculum in *both* mathematics *and* computer science. Specifically, a computer science class taught in an appropriate language might help develop their computational foundation and may allow transfer students to be in a stronger position to excel in their subsequent courses in data science.

> **Our recommendation** is that, for the next few years, second courses in data science are left to bachelor's programs, and the credits are replaced with another course with an existing mapping. Planning should begin on course designs and frameworks for such a course to be taught at both 2- and 4-year institutions since this would support students planning to transfer as well as associate's-to-workforce programs.

Bob has not had anything like a second course in data science, so he will have to take this sophomore-level course as a junior. While this should end up being a relatively minor detour, it contributes to his feeling of not quite fitting in with his new classmates (as a 2-year college transfer student). Luckily, Bob used one of his general electives to take a programming course that prepares him for CIS 360.

3.1.3. A Course in Scientific Computing, Data Science Workflow, and Reproducible Computing

A generic bachelor's program in data science will include explicit instruction in how to advance science by computing with data in a reproducible, collaborative workflow. In some programs, this instruction will be woven into modules that permeate a series of courses. In others, there will be a standalone course that focuses on these issues. It is important that the technologies to support workflow and reproducible analysis as a component of data acumen (NASEM, 2018) should not be assumed to be known by students or left for them to learn outside of a course, lest existing disparities in background are exacerbated.

Topics in this area include version control systems (e.g., `git`), collaboration and project management tools (e.g., GitHub, Trello), software development paradigms (e.g., Agile/Scrum), document authoring software (e.g., Pandoc, variants of Markdown (Jupyter, R Markdown, Quarto), LaTeX), command line scripting (e.g., UNIX), cloud computing, as well as further exposure to R, Python, and/or SQL.

While there are existing models of such courses at 2-year colleges, they are less likely to have existing transfer mappings. Given the variety of topics in these courses and the difficulty of coordinating the content across institutions, these credits will probably have to be mapped on a one-to-one basis. One promising avenue is a course in R or Python that is outside of the main computer science sequence (which is often taught in Java or C++). An example of such a course is CSE 160 at Springfield Technical Community College.

> **Our recommendation** is that individual programs map credits where reasonably equivalent options exist, and replace them with general education or liberal arts credits where they do not.

Bob was underprepared for the mechanics of data science at the junior level. While he did use R Markdown in his data science course, he had never used GitHub, SQL, or the command line. The lack of such experience made it more difficult for him to secure an internship during his work on his associate's degree. Other frustrations after transferring occurred when he felt like he understood the material in his classes but struggled to participate in group projects because he was not as facile with the workflow tools. Two- and 4-year institutions should work to provide flexible options for students to be introduced to and deepen their understandings of reproducibility and workflow throughout their courses and programs (Horton et al., 2022).

3.1.4. Lab Sciences

Many of the existing Massachusetts transfer options in computer science (and other STEM disciplines) require two semesters of lab sciences (e.g., physics, biology, or chemistry) as a component of their general education

requirements. Requiring a student pursuing a bachelor's degree in data science to take two semesters of physics, biology, or chemistry provides an opportunity for them to learn important aspects of the scientific process as well as the collection and analysis of data. At present, many of these courses may be less germane for data science students, but there is considerable potential for them to reinforce and build basic data sciences skills for all students while building domain knowledge.

As an alternative to explore, we can imagine that a future data science–infused lab course could be developed as a way to provide more exposure to key data science topics while meeting the learning outcomes for a lab course. This is perhaps less of a 'friction-point' than an opportunity to improve data science options as well as to infuse computation and data into undergraduate STEM education (NASEM, 2022).

> **Our recommendation** is that students use existing pathways for lab sciences, choosing courses when possible that incorporate aspects of scientific data (e.g., Greenfield Community College's BIO 120 Introduction to Environmental Science)[6].

Bob completed this sequence as part of his transfer degree.

3.1.5. Communication, Ethics, and Application Domains

Bachelor's programs in data science include training in communication (how do we transfer knowledge gained from data analysis from data scientist to a broader audience? Parke, 2008) and ethics (what responsibilities do data scientists have to their users, customers, and society as a whole? Baumer et al., 2022). In addition, a domain of application is valuable (how does data science enhance our understanding of another subject?). These vital aspects of a data science curriculum cannot wait entirely until the junior year, and thus, 2-year college students must find ways to build skills in these areas before they transfer.

Most 2-year colleges offer courses in communication. If any of those courses focus on *communicating with data*, they should be taken. Courses that focus on more general writing skills are still valuable, and are already part of the general education requirements for any associate's degree. Where courses in ethics, or preferably, data ethics are available, they should be taken at the 2-year college level, as this will help to infuse ethics early in a student's education.

For those students whose application domain will intersect with the lab sciences mentioned in Section 3.1.4, that requirement might provide a helpful synergy. We imagine that this might be particularly beneficial for students interested in public health, biostatistics, or bioinformatics.

One challenge here will be ensuring that whatever these courses are, they count toward the associate's degree program.

> **Our recommendation** is that institutions think carefully and holistically about how requirements for communication, ethics, and domain application can be used to accrue credits at 2-year colleges and foster

> successful transfers.

Although Bob was exposed to ideas in data science ethics during his course at HCC, there was not time for much depth. Thus, he feels as though he is learning about these issues in depth for the first time in his junior year. He wishes he had thought to take CSI 215 (Ethical/Legal Aspects of Information Systems) at HCC, since it already transfers as CIS 381 at UMassD, which he now needs to take anyway.

3.2. Where to Situate Programs?

Unfortunately, the interdisciplinary nature of data science is in conflict with the siloing of programs within departments. The 2018 NASEM data science for undergraduates report found that many bachelor's degree programs in data science are housed in a college or school of business, a mathematics or statistics department, or a computer science department (see pages 3–5 of NASEM [2018]). A few undergraduate data science majors were described as hybrids of these three models, with joint administration/programmatic coordination. We believe that such hybrid models are better suited to ensure that students develop a deep foundation in all aspects of data acumen.

When considering where to situate associate's degree programs within departments at 2-year colleges, the compressed timeline given the 2-year nature of the degree only compounds the problem. As a result, until there are associate's degree programs in data science, even explicit transfer pathways (such as the ones we are trying to create) may force students to choose between two potentially undesirable options: obtaining an associate's degree in liberal arts studies that may not be as marketable as a degree in a more technical field, or supplementing a degree in mathematics or computer science with several additional courses. Our hope is to provide guidance about flexible pathways that could soften these rough edges that exist at present.

3.3. Alignment of Degrees With Pathways

Even without the kind of explicit pathways we are advocating for, transfer to a bachelor's in data science may still be possible. However, a student would have to forge their own pathway, which might mean taking courses at a 2-year college that were outside of the requirements of their associate's degree program, taking catch-up courses at a 4-year college once they arrive, and/or obtaining explicit transfer credit for courses that are not already mapped. **All of these obstacles add unnecessary friction, cost, and time that students and society cannot afford.**

4. Case Study: A New Transfer Pathway in Data Science

In this section, we articulate a realistic vision for a new transfer pathway from BHCC to UMassD (see Figure 1). We believe that such a pathway, if approved, would be the first of its kind[7] in Massachusetts, and one of only a handful in the nation. We hope that this proposal will serve as a model for similar potential pathways between other 2-year colleges and universities.

4.1. Bunker Hill to UMass Dartmouth

BHCC offers a data analytics option within the associate of science program. Students in this program take multiple courses in computer science, receive foundational training in statistics, linear algebra, and college writing, and are exposed to R, Python, and SQL. This is a terminal degree which is not designed for transfer. So while most of the pieces for a data science transfer pathway from BHCC to UMassD are already in place, the degree programs do not align.

Although students who attend BHCC may be more interested in staying in Boston and transferring to Northeastern via an existing articulation agreement, a public option at UMassD would be the first potential MassTransfer pathway in data science. Tables 1, 2, and 3 illustrate the current situation. Most of the course mappings in these figures are already approved by the MassTransfer system. In what follows, we provide detail about the exceptions, and discuss possibilities to reduce the number of 'wasted' (or 'stranded') credits.

Table 1. Courses that should easily transfer between BHCC's data analytics associate's degree program and UMassD's data science bachelor's degree program. These courses count toward both degrees and most mappings are already approved by the MassTransfer system. Note that 5 credits are 'lost' in the process. The 'Mapped' column indicates whether the course mapping is already approved in the MassTransfer system.

Topic	BHCC	BHCC Credit	UMassD	UMassD Credit	Mapped
Calculus I	MAT 281	4	MTH 151	4	Y
Calculus II	MAT 282	4	MTH 152	4	Y
Linear Algebra	MAT 291	4	MTH 221	3	Y
Statistics & Data Science	CIT 130 + CIT 137 + CIT 187	10	DSC 101 + DSC 201 + MTH 231	9	N
Scientific Computation	CSC 125 + CSC 225	6	MTH 280	3	N
College Writing I	ENG 111	3	ENL 101	3	Y
College Writing II	ENG 112	3	ENL 102	3	Y
Scientific Reasoning	multiple	4	NA	4	Y
Creative Work	multiple	3	NA	3	Y

Community & Cultural Contexts	multiple	3	NA	3	Y
Excel	CIT 234	3	NA	3	Y
Total	-	47	-	42	-

Table 2. Courses that are required for the data analytics associate's degree at BHCC, but do not count toward the bachelor's degree at UMassD. These credits are 'wasted' in the sense that a 2-year college student needs them to graduate, but not to transfer.

Topic	BHCC	BHCC Credit	UMassD	UMassD Credit	Mapped
Intro Statistics	MAT 181	3	MTH 147	3	Y
Problem Solving	CIT 113	3	NA	NA	N
Precalculus	MAT 197	4	MTH 150	3	Y
SQL Programming	CSC 236	3	NA	NA	N
Database Programming	CSC 240	3	NA	NA	N
Operating Systems	CIT 268	3	NA	NA	N
Total	-	19	-	6	-

Table 3. Courses that are required for the data science bachelor's program at UMassD but not for the data analytics associate's degree at BHCC. Two-year college students might have to pay extra in order to satisfy these additional requirements in order to transfer.

Topic	BHCC	BHCC Credit	UMassD	UMassD Credit	Mapped
Physics I	PHY 251	4	PHY 113	4	Y
Physics II	PHY 252	4	PHY 114	4	Y
Programming I	CSC 239	4	CIS 180	4	Y

Programming II	CSC 285	4	CIS 181	4	Y
Discrete Math	NA	NA	MTH 181	3	N
Software Design	NA	NA	CSC 280	4	N
Total	-	16	-	23	-

Alice is interested in pursuing this pathway, but because it doesn't exist, she cannot automatically enroll in the data science program at UMassD. By completing the data analytics degree program at BHCC, she will complete all of the courses listed in Tables 1 and 2. This gives her 63 credits, which is more than the 60 credits typically required for a transfer pathway. Unfortunately, only 36 of those 63 credits are transferable to UMassD, and because Alice is missing all of the courses in Table 3, she is not actually prepared for the junior-level data science curriculum at UMassD. She could pay out of pocket for the 16 transferable credits in Table 3, but that would cost her time and money, and still leave her a few requirements short. (Financial aid may also not be available to support her completion of these 'stranded' credits.) Thus, Alice's academic progress may also be impeded.

4.2. A Proposed Pathway

Tables 1, 2, and 3 lay out the building blocks from which an articulation agreement in data science between BHCC and UMassD. We argue that this articulation agreement would provide a proof-of-concept and blueprint for a more generalized *data science MassTransfer associate's degree*. The program would consist of all the courses in Tables 1 and 3, with perhaps a few tweaks.

We note first that most of the courses in Tables 1 and 3 are already mapped. The exceptions in Table 1 are the statistics, data science, and scientific computing blocks. Creating these course mappings is an active conversation among the authors and the relevant parties. We are optimistic that these mappings will remove the first, second, and third points of friction, although the student stands to lose five credits in the process.

> Visit the web version of this article to view interactive content.

Among the courses in Table 3, four are already mapped and would be part of the new transfer pathway. The physics courses remove the fourth point of friction. Two courses (discrete math and software design) have no obvious equivalent at BHCC. Discrete math is taught at many 2-year colleges, and so the lack of a discrete math option at BHCC is more an idiosyncracy of BHCC than a systemic problem. We hope that a suitable alternative course can be found, or that UMassD will accept an elective in its place. As for the software design course, BHCC offers a number of alternative courses that, while not covering the same material, seem like they

would prepare future data scientists equally well, albeit in different ways. For example, the two-course sequence in SQL Programming and Database Programming shown in Table 2 seems like it covers highly relevant material. Perhaps UMassD would accept those courses as alternatives.

In that event, suppose that we substitute the six-credit sequence in SQL programming for the discrete math and software design requirement at UMassD. Then adding these credits to those in Tables 1 and 3 results in a 69-credit curriculum of existing courses at BHCC, that would transfer as 64 credits that cover nearly all of the first 2 years of the data science curriculum at UMassD.

Using only courses that already exist, we see this as the best candidate to be the first public data science transfer pathway in Massachusetts.

5. Conclusion

At the December 2018 meeting of the National Academies Postsecondary Data Science Education Roundtable (NASEM, 2020), D.J. Patil, former Chief Data Scientist in the White House Office of Science and Technology Policy, described how his experience from a 2-year college bestowed upon him "three gifts": "a love of mathematics, an understanding of how to write in various genres, and confidence to succeed at the postsecondary level" (p. 158). He expressed that his experience at a 2-year college provided a crucial "on-ramp" to his future success in data science.

Like Patil, we see 2-year colleges as key players in developing the next generation of data science students. Our experience with the DSC-WAV project and other interactions have shown that our 2-year college system has countless committed and engaged educators and administrators working to build better futures for their students, amidst time and resource constraints. Patil testifies to the habits-of-mind and general skills that 2-year colleges are already cultivating—our goal is to align curricula so as to reduce administrative and bureaucratic obstacles.

We have focused on associate's degrees that prepare for a bachelor's degree. Other pathways, such as associate's-to-workforce are also important, and need improved flexibility and transparency.

5.1. Additional Challenges

There is considerable work needed to foster sustainable courses, structures, and programs. We acknowledge that this will require focus and attention for many years. Efforts such as the NSF-funded EDC Oceans of Data Mentoring New Data Pathways project have engaged BHCC in an effort to support new data programs.

Resource disparities at many 2-year colleges and insufficient partnerships between 2- and 4-year institutions could hamper these efforts. As but one example, due to resources and other circumstances, 2-year colleges could feel at a disadvantage and perhaps be reluctant to offer courses that are not included in guaranteed transfer systems such as MassTransfer, or courses not belonging to an already structured pathway.

Faculty development is another critical issue. At a time when data science positions are challenging for employers to fill, where will the next generation of instructors come from? This is another area where partnerships between 2- and 4-year institutions as well as industry will be critical (NASEM, 2018). An example of how this might work for data science can be seen in the strategies outlined in Enriquez et al. (2018a) for engineering transfer programs.

The changing pre-K–12 landscape raises important questions. As states are reviewing and revising their mathematics, science, and computing standards, statistics and data science are being elevated and made more explicit. We believe that this will impact the knowledge, skills, and abilities students bring to their postsecondary education. These changes may impact the future of pathways, potentially in positive ways.

We acknowledge that one of the biggest challenges to students transferring to STEM programs is that they get caught in a "mathematics maze." A failed course leads to remedial courses, which lead to more courses, which impedes progress toward completion of their program. We agree that this is an important unsolved issue (see Cafarella, 2021 for an in-depth treatment). In other disciplines (e.g., nursing), efforts such as the Dana Center's Mathematics Education for Nurses collaboration (Dana Center, 2021b) works to improve student success while helping nursing students gain the "mathematical knowledge, skills, and attitudes" (p. 1) to be successful in their career. Similar efforts would benefit future data science programs.

There are many other issues that we could address at this juncture, including aspects of associate's-to-workforce programs, challenges and opportunities of dual enrollment, and the pressing need for improved computational infrastructure. But we intentionally limit our primary focus to fostering pathways, which needs to begin by identifying barriers and resources to the widespread teaching of accessible and pedagogically sound introductory data science courses.

5.2. Bright Spots

In addition to the proof-of-concept at UNH-Manchester, there are some useful models that we can consider.

Considerable efforts toward pathways are underway in California. While the scale of the California system—which includes both the UC and Cal State constellations—provides obvious challenges, there have been encouraging developments. Models for addressing similar challenges in related disciplines (e.g., engineering) exist (Enriquez et al., 2018b).

In Ohio, 36 public institutions of higher education, 27 two-year colleges, and 9 four-year colleges approved a set of learning outcomes for a general education data science course developed by faculty from 2- and 4-year institutions (Ricardo Moena, personal communication). We see this as a necessary but not sufficient step.

In Connecticut, recent efforts have led to the establishment of associate's programs in data science with pathways to workforce or transfer (Northwestern CT Community College, 2020).

5.3. A Call to Broaden Participation

We close with some reflections on the critical role that 2-year colleges provide in terms of **affordable** options that are accessible to a **diverse** population.

The *Broadening Data Science Education* (Rawlings-Goss et al., 2018) report notes that:

> Many individuals in today's data science workforce are coming from doctoral or master's degree programs, which have seen a dramatic increase in recent years. While these advanced degrees are valuable, it is not economically feasible for all data scientists to complete four years of an undergraduate degree, then a one- or two-year master's program before they can undertake useful work. Ensuring the future growth of the workforce requires an expansion to four-year and two-year degrees. (p. 45)

At the June 2019 NASEM Roundtable meeting, Uri Treisman of the University of Texas-Austin and the Dana Center described data science programs as "powerful resources for students seeking upward mobility" (NASEM, 2020, p. 165).

Moreover, the *Broadening Data Science Education* (Rawlings-Goss et al., 2018) report suggests that: "the potential impact of the Data Divide is no less dire for our institutions of higher education" (p. 7). Such concerns lead to the finding that: "Data science would particularly benefit from broad participation by underrepresented minorities because of the many applications to problems of interest to diverse populations" (NASEM, 2018). The California Alliance for Data Science Education promotes a quotation from Jennifer Chayes (Associate Provost, Division of Computing, Data Science, and Society and Dean, School of Information, UC Berkeley) that "increasing access to data science as a career option for all students is key to making data science a more diverse and inclusive field." The *Broadening Data Science Education* (Rawlings-Goss et al., 2018) report states this even more directly:

> If we do not make diversity and inclusion a priority now, we will not have it in the future. We do not want to repeat the mistakes of the past, so we must reverse the trend for the growing divide to make and keep data science broad. Diversity will bring a lot of ideas and voices to the table, which may lead to significantly fewer models producing biased results when trained using algorithms on biased data sets. (p. 30)

We agree that 2-year colleges are the only affordable game in town and serve a key role in data science education now and in the future.

Acknowledgments

Disclosure Statement

Benjamin S. Baumer and Nicholas Jon Horton have no other financial or non-financial disclosures to share for this article.

References

Baumer, B. S. (2015). A data science course for undergraduates: Thinking with data. *The American Statistician, 69*(4), 334–342. https://doi.org/10.1080/00031305.2015.1081105

Baumer, B. S., Garcia, R. L., Kim, A. Y., Kinnaird, K. M., & Ott, M. Q. (2022). Integrating data science ethics into an undergraduate major: A case study. *Journal of Statistics and Data Science Education, 30*(1), 15–28. https://doi.org/10.1080/26939169.2022.2038041

Baumer, B. S., Kaplan, D. T., & Horton, N. J. (2021). *Modern data science with R* (2nd ed.). Chapman; Hall/CRC Press. https://mdsr-book.github.io/mdsr2e

Blumenstyk, G. (2021). The edge: The "dirty secret" that obstructs transfer. https://www.chronicle.com/newsletter/the-edge/2021-11-10

Cafarella, B. (2021). *Breaking barriers: Student success in community college mathematics*. CRC Press. https://doi.org/10.1201/9781003175803

Çetinkaya-Rundel, M. (2020). *Data science in a box*. https://datasciencebox.org/

Cleveland, W. S. (2001). Data science: An action plan for expanding the technical areas of the field of statistics. *International Statistical Review, 69*(1), 21–26. https://doi.org/10.1111/j.1751-5823.2001.tb00477.x

Dana Center. (2021a). *Data science course framework*. https://www.utdanacenter.org/sites/default/ files/2021-05/data%5C_science%5C_course%5C_framework%5C_2021%5C_final.pdf

Dana Center. (2021b). *Mathematics education for nurses*. https://www.utdanacenter.org/our-work/higher-education/collaborations/math-for-nurses

Danyluk, A., Leidig, P., Buck, S., Cassel, L., McGettrick, A., Qian, W., Servin, C., & Wang, H. (2021). *Computing competencies for undergraduate data science curricula*. Association for Computing Machinery. https://doi.org/10.1145/3453538

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*(1), 1–16. https://doi.org/10.1146/annurev-statistics-060116-053930

Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745–766. https://doi.org/10.1080/10618600.2017.1384734

Duffin, E. (2022, November 3). *Community colleges in the United States: Statistics & facts*. Statista. https://www.statista.com/topics/3468/community-colleges-in-the-united-states/#dossierContents__outerWrapper

Enriquez, A., Langhoff, N., Dunmire, E., Rebold, T., & Pong, W. (2018a). Strategies for developing, expanding, and strengthening community college engineering transfer programs. *American Society for Engineering Education*, *2018*. https://par.nsf.gov/biblio/10063235

Enriquez, A., Langhoff, N., Dunmire, E., Rebold, T., & Pong, W. (2018b). Strategies for developing, expanding, and strengthening community college engineering transfer programs. *American Society for Engineering Education*, *2018*. https://doi.org/10.18260/1-2--30995

Gould, R., Peck, R., Hanson, J., Horton, N. J., Kotz, B., Kubo, K., Malyn-Smith, J., Rudis, M., Thompson, B., Ward, M., & Wong, R. (2018). *The two-year college data science summit*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/2018TYCDS-Final-Report.pdf

Horton, N. J., Alexander, R., Piekut, A., & Rundel, C. (2022). The growing importance of reproducibility and responsible workflow in the data science and statistics curriculum. *Journal of Statistics and Data Science Education*, *30*(3), 207–208. https://doi.org/10.1080/26939169.2022.2141001

Horton, N. J., Baumer, B. S., Zieffler, A., & Barr, V. (2021). The Data Science Corps Wrangle-Analyze-Visualize program: Building data acumen for undergraduate students. *Harvard Data Science Review*, *3*(1). https://doi.org/10.1162/99608f92.8233428d

Korn, M., & Fuller, A. (2021, July 8). *"Financially hobbled for life": The elite master's degrees that don't pay off*. The Wall Street Journal. https://www.wsj.com/articles/financially-hobbled-for-life-the-elite-masters-degrees-that-dont-pay-off-11625752773

Legacy, C., Zieffler, A., Baumer, B. S., Barr, V., & Horton, N. J. (2022). Facilitating team-based data science: Lessons learned from the DSC-WAV project [Advance online publication]. *Foundations of Data Science*. https://doi.org/10.3934/fods.2022003

Murnane, R. J., Willett, J. B., Papay, J. P., Mantil, A., Mbekeani, P. P., & McDonough, A. (2022). *Building stronger community college transfer pathways: Evidence from Massachusetts*. Massachusetts Institute for a New Commonwealth. https://massinc.org/research/building-stronger-community-college-transfer-pathways/

National Academies of Sciences, Engineering, and Medicine. (2016). *Barriers and opportunities for 2-year and 4-year stem degrees: Systemic change to support students' diverse pathways*. The National Academies Press. https://doi.org/10.17226/21739

National Academies of Sciences, Engineering, and Medicine. (2018). *Data science for undergraduates: Opportunities and options*. National Academies Press. https://nas.edu/envisioningds

National Academies of Sciences, Engineering, and Medicine. (2020). *Roundtable on data science postsecondary education*. https://www.nap.edu/25804

National Academies of Sciences, Engineering, and Medicine. (2022). *Imagining the future of undergraduate STEM education*. National Academies Press. https://nap.nationalacademies.org/read/26314

New two-year college data science, analytics programs on the rise. (2022, August 1). Amstat News. https://magazine.amstat.org/blog/2022/08/01/new-two-year-programs

Northwestern CT Community College. (2020, February 11). *Northwestern CT community college launches data science degree program*. https://www.nhregister.com/news/article/Northwestern-CT-Community-College-launches-data-15074344.php

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

Rawlings-Goss, R., Cassel, L., Cragin, M., Cramer, C., Dingle, A., Friday-Stroud, S., Herron, A., Horton, N. J., R, I. T., Jordan, K., Ordonez, P., Rudis, M., Rwebangira, R., Schmitt, K., Smith, D., & Stephens, S. (2018). *Keeping data science broad: Negotiating the digital and data divide among higher education institutions*. South Big Data Hub. https://southbigdatahub.org/resources/newsblog/keeping-data-science-broad-program

Wickham, H., & Grolemund, G. (2016). *R for data science: Import, tidy, transform, visualize, and model data*. O'Reilly Media. https://r4ds.had.co.nz

Yan, D., & Davis, G. E. (2019). A first course in data science. *Journal of Statistics Education*, *27*(2), 99–109. https://doi.org/10.1080/10691898.2019.1623136

Footnotes

1. The Academic Data Science Alliance Data Science and the American Mathematical Association of Two-Year Colleges (AMATYC) have created listings that are likely incomplete. Unfortunately, no comprehensive census of programs is readily available. ↩

2. In contrast to the 45,000 master's graduates in data science referenced earlier, according to Statista (Dufflin, 2022), there were more than a million associate's degree recipients in the United States during the 2018–2019 academic year. We believe that associate's degree candidates represent an untapped data resource. ↩

3. Until recently, the new CIP codes were not classified as STEM disciplines, which had negative implications for the immigration status of international students. Efforts by the Academic Data Science Alliance and others led to reclassification of the data science CIP code. ↩

4. While data science lacks a national curriculum analogous to those in more established disciplines like computer science, mathematics, and statistics, a general framework for a bachelor's program in data science is taking shape. We have been involved in several efforts to shape such curriculum and accreditation guidelines at the national level, most notably including De Veaux et al. (2017), Gould et al. (2018), and NASEM (2018). We have also been leaders in developing courses (Baumer, 2015), undergraduate majors, and textbooks (Baumer et al., 2021) that support instruction in data science. These experiences give us some standing to anticipate what the general landscape of bachelor's programs in data science will look like over the next 10 years. The handful of bachelor's programs that currently exist will likely double at least once in that time, and one of the major goals of this article is to coordinate efforts across all parties such that the MassTransfer system is prepared to adapt to this changing landscape. ↩

5. Some 2-year college students may need to complete additional developmental math courses. Ideally, corequisite approaches (see efforts by the Dana Center [2021a]) could allow them to complete these requisites in a timely fashion without delaying their progress toward their associate's degree. ↩

6. Shodor (http://shodor.org/) has worked to incorporate data science into various STEM majors, including biology, chemistry, and physics. ↩

7. In the fall of 2021, BHCC signed an articulation agreement with Northeastern University from their associate of science in data analytics degree to Northeastern's bachelor of science in analytics degree

program. This agreement serves as a proof-of-concept that transfer pathways to *public* universities in Massachusetts can be created. ↩

References

- Baumer, B. S., Garcia, R. L., Kim, A. Y., Kinnaird, K. M., & Ott, M. Q. (2022). Integrating data science ethics into an undergraduate major: A case study. *Journal of Statistics and Data Science Education*, *30*(1), 15–28. https://doi.org/10.1080/26939169.2022.2038041 ↩

- Baumer, B. S., Kaplan, D. T., & Horton, N. J. (2021). *Modern data science with R* (2nd ed.). Chapman; Hall/CRC Press. https://mdsr-book.github.io/mdsr2e
  Baumer, B. S., Kaplan, D. T., & Horton, N. J. (2021). *Modern data science with R* (2nd). Chapman; Hall/CRC Press.

  ↩

- Blumenstyk, G. (2021). The edge: The "dirty secret" that obstructs transfer. https://www.chronicle.com/newsletter/the-edge/2021-11-10

  ↩

- Cafarella, B. (2021). *Breaking barriers: Student success in community college mathematics*. CRC Press. https://doi.org/10.1201/9781003175803 ↩

- Çetinkaya-Rundel, M. (2020). *Data science in a box*. https://datasciencebox.org/ ↩

- Cleveland, W. S. (2001). Data science: An action plan for expanding the technical areas of the field of statistics. *International Statistical Review*, *69*(1), 21–26. https://doi.org/10.1111/j.1751-5823.2001.tb00477.x ↩

- Dana Center. (2021). *Data science course framework*. https://www.utdanacenter.org/sites/default/files/2021-05/data_science_course_framework_2021_final.pdf. https://www.utdanacenter.org/sites/default/files/2021-05/data_science_course_framework_2021_final.pdf ↩

- Dana Center. (2021). *Mathematics education for nurses*. https://www.utdanacenter.org/our-work/higher-education/collaborations/math-for-nurses. https://www.utdanacenter.org/our-work/higher-education/collaborations/math-for-nurses ↩

- Danyluk, A., Leidig, P., Buck, S., Cassel, L., McGettrick, A., Qian, W., Servin, C., & Wang, H. (2021). *Computing competencies for undergraduate data science curricula*. Association for Computing Machinery. https://doi.org/10.1145/3453538 ↩

- 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*(1), 1–16. https://doi.org/10.1146/annurev-statistics-060116-053930 ↩

- Donoho, D. (2017). 50 years of data science. *Journal of Computational and Graphical Statistics*, *26*(4), 745–766. https://doi.org/10.1080/10618600.2017.1384734 ↩

- Duffin, E. (2022, November 3). *Community colleges in the United States: Statistics & facts*. Statista. https://www.statista.com/topics/3468/community-colleges-in-the-united-states/#dossierContents__outerWrapper

  ↩

- Enriquez, A., Langhoff, N., Dunmire, E., Rebold, T., & Pong, W. (2018). Strategies for developing, expanding, and strengthening community college engineering transfer programs. *American Society for Engineering Education*, *2018*. https://doi.org/10.18260/1-2--30995 ↩

- Enriquez, A., Langhoff, N., Dunmire, E., Rebold, T., & Pong, W. (2018). Strategies for developing, expanding, and strengthening community college engineering transfer programs. *American Society for Engineering Education*, *2018*. https://par.nsf.gov/biblio/10063235 ↩

- Gould, R., Peck, R., Hanson, J., Horton, N. J., Kotz, B., Kubo, K., Malyn-Smith, J., Rudis, M., Thompson, B., Ward, M., et al. (2018). *The two-year college data science summit*. American Statistical Association. https://www.amstat.org/asa/files/pdfs/2018TYCDS-Final-Report.pdf

  ↩

- Horton, N. J., Alexander, R., Piekut, A., & Rundel, C. (2022). The growing importance of reproducibility and responsible workflow in the data science and statistics curriculum. *Journal of Statistics and Data Science Education*, *30*(3), 207–208. https://doi.org/10.1080/26939169.2022.2141001 ↩

- Horton, N. J., Baumer, B. S., Zieffler, A., & Barr, V. (2021). The Data Science Corps Wrangle-Analyze-Visualize program: Building data acumen for undergraduate students. *Harvard Data Science Review*, *3*(1). https://doi.org/10.1162/99608f92.8233428d ↩

- Korn, M., & Fuller, A. (2021, July 8). *"Financially hobbled for life": The elite master's degrees that don't pay off*. The Wall Street Journal. https://www.wsj.com/articles/financially-hobbled-for-life-the-elite-masters-degrees-that-dont-pay-off-11625752773

  ↩

- Legacy, C., Zieffler, A., Baumer, B. S., Barr, V., & Horton, N. J. (2022). Facilitating team-based data science: Lessons learned from the DSC-WAV project [Advance online publication]. *Foundations of Data Science*. https://doi.org/10.3934/fods.2022003

  ↩

- Murnane, R. J., Willett, J. B., Papay, J. P., Mantil, A., Mbekeani, P. P., & McDonough, A. (2022). *Building stronger community college transfer pathways: Evidence from Massachusetts*. Massachusetts Institute for a New Commonwealth. https://massinc.org/research/building-stronger-community-college-transfer-pathways/

  ↩

- National Academies of Sciences, Engineering, and Medicine. (2016). *Barriers and opportunities for 2-year and 4-year STEM degrees: Systemic change to support students' diverse pathways*. The National Academies Press. https://doi.org/10.17226/21739 ↩

- National Academies of Sciences, Engineering, and Medicine. (2018). *Data science for undergraduates: Opportunities and options*. National Academies Press. https://nas.edu/envisioningds ↩

- National Academies of Sciences, Engineering, and Medicine. (2020). *Roundtable on data science postsecondary education*. https://www.nap.edu/25804 ↩

- National Academies of Sciences, Engineering, and Medicine. (2022). *Imagining the future of undergraduate STEM education*. National Academies Press. https://nap.nationalacademies.org/read/26314 ↩

- *New two-year college data science, analytics programs on the rise*. (2022, August 1). Amstat News. https://magazine.amstat.org/blog/2022/08/01/new-two-year-programs

  ↩

- Northwestern CT Community College. (2020, February 11). *Northwestern CT community college launches data science degree program*. https://www.nhregister.com/news/article/Northwestern-CT-Community-College-launches-data-15074344.php

  ↩

- 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 ↩

- Rawlings-Goss, R., Cassel, L., Cragin, M., Cramer, C., Dingle, A., Friday-Stroud, S., Herron, A., Horton, N. J., R, I. T., Jordan, K., Ordonez, P., Rudis, M., Rwebangira, R., Schmitt, K., Smith, D., & Stephens, S. (2018). *Keeping data science broad: Negotiating the digital and data divide among higher education institutions*. South Big Data Hub. https://southbigdatahub.org/resources/newsblog/keeping-data-science-broad-program

  ↩

- Wickham, H., & Grolemund, G. (2016). *R for data science: Import, tidy, transform, visualize, and model data*. O'Reilly Media. https://r4ds.had.co.nz ↩

- Yan, D., & Davis, G. E. (2019). A first course in data science. *Journal of Statistics Education*, *27*(2), 99–109. https://doi.org/10.1080/10691898.2019.1623136 ↩