Harvard Data Science Review • Issue 2.4, Fall 2020

The Opportunities of Two-Year College Data Science

Brian Kotz¹

¹Department of Mathematics Statistics and Data Science, Montgomery College, Germantown, Maryland, United States of America

Published on: Oct 29, 2020

DOI: <https://doi.org/10.1162/99608f92.75aed58b>

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

ABSTRACT

Two-year colleges are poised to play a substantial and possibly transformative role in data science and undergraduate data science education. Current two-year college data science programs provide affordable and rigorous certificate and degree experiences that instill data acumen in students who do not seek or do not fit within the traditional four-year college paradigm. Additional two-year college data science certificate and degree programs are certain to develop as conversations continue with four-year colleges regarding matters of transferability, student achievement, and program evolution. Developing a two-year college data science program is not an easy task, but deep discussions of postsecondary data science education will be incomplete if they fail to consider the opportunities that two-year colleges provide. Two-year college data science educators will therefore need to continue to be active participants in the discussions of both the field and its educational practices going forward. In addition to supporting students in their data science programs with a comprehensive approach respectful of both student diversity and local needs, two-year colleges may also have an opportunity—or even an obligation—to effectively instill principles of general data literacy in their broader undergraduate populations. Additional resources, continued professional development, and effective leadership will be required. These ideas are discussed both generally and within the framework of one two-year college program.

Keywords: data science education, undergraduate education, two-year college, community college, data acumen

1. Thinking Broadly: A Current Role for Two-Year Colleges

In June of 2019, when the National Academies of Sciences, Engineering, and Medicine’s (NASEM) Roundtable on Data Science Postsecondary Education hosted their 11th meeting, they devoted the entirety of their agenda to data science at two-year colleges. In addition to facilitating some new collaborations and showcasing many of the efforts that have already occurred in two-year colleges, the Roundtable [meeting](#) served as a natural complement to the Academies’ recent *Data Science for Undergraduates: Opportunities and Options* [report](#) (sponsored by the National Science Foundation) where two-year college data science was highlighted and discussed as a viable undergraduate modality.

Many two-year colleges offer two distinct paths of education for their communities. One is what we refer to at our two-year college as Workforce Development and Continuing Education, that is, noncredit classes for community education, updating skills, English skills for adult speakers of other languages, professional licensure in trades, and so on. In a data science setting, this skill-based coursework may include gaining familiarity with data structures, programming languages, and so on. The need to provide affordable and flexible learning opportunities in these skills is great and cannot be overlooked. On the other hand, as expertise

in skills is only one component of data science, many two-year colleges have developed programs within their other distinct educational path—a path of credit-bearing, college-level classes such as those that transfer to four-year colleges and universities or are required for an Associate of Arts or Associate of Sciences degree. Montgomery College’s Data Science Certificate program, which launched in Fall 2017, resides in this path.

In this credit-bearing model, two-year colleges provide undergraduate curricula to students who do not seek a traditional four-year college experience. As such, two-year colleges can reach motivated students who may not be able to consider (or may not need to consider) the four-year college model for their goals. Although the size of this population for data science is not known, interest and enrollments are growing, at least in the greater Baltimore-Washington area. Given the balance of flexibility and rigor that two-year colleges can provide to their students, two-year college data science programs are poised to play a vital role in addressing the calls for both a more data literate workforce and more data literate transferees to four-year college undergraduate programs. As Laura Haas, Co-Chair of the Committee on Envisioning the Data Science Discipline: The Undergraduate Perspective, pointed out in her remarks in *the Harvard Data Science Review (HDSR)*, “We found that many different forms of data science education and multiple pathways will help students from a variety of backgrounds—educational and demographic—to succeed at levels ranging from basic to expert” (Haas et al., 2019). Two-year colleges are definitely a part of that mix. In order for the data science education community to reach and to support these students, two-year colleges will need to play an even larger role in the discussions of data science education going forward.

2. Professional Support for and Interest in Two-Year College Data Science

One theme that became clear throughout the NASEM Roundtable meeting was that there is great support and interest regarding the development of two-year college data science programs. For example, the American Mathematical Association of Two-Year Colleges (AMATYC) now has an established formal community of individuals within its ranks dedicated to supporting and improving data science in two-year colleges. While that community includes representatives from established two-year college programs, it also includes (and welcomes) representatives and academic leaders from other colleges and stakeholder organizations. Also of note, the National Science Foundation (NSF) has sponsored multiple projects in support of two-year college data science education and professional development, including the 2018 Two-Year College Data Science Summit (TYCDSS) hosted by the American Statistical Association (ASA). The [final report](#) from that summit is publicly available and contains many recommendations and resources for both existing and future two-year college programs.

As several organizations have provided funding and expertise in two-year college data science efforts, these contributions from a broad community have nurtured a growth and camaraderie that has elevated data science in two-year colleges quite rapidly. Whether it is through attempts to integrate data literacy into coursework, to develop full programs, or to better define the “levels ranging from basic to expert” mentioned earlier, the

commitment of these foundations and organizations and the entrepreneurial, pioneering spirit of these educators, scientists, industry representatives, and public officials suggest a bright future for two-year colleges in the field.

One additional appeal of two-year college programs is a perceived barrier reduction in educational accessibility. While I think that a great deal more research needs to be done here, it could be that many traditional access barriers in other fields might not be present in data science, and two-year colleges are excellent places to start that investigation. Two-year colleges by their very nature provide first-year and second-year academic experiences for their students; and there could be advantages both to the students and to the discipline in having students experience data science at an early point in their academic careers. One might argue that this early accessibility could also potentially invite a more diverse group of students to the field. Also keep in mind that the open nature of many data science resources and tools (not to mention the data!) acknowledges the field as one of accessibility and portability. As society continues to recognize the importance of data science, how it affects everyone's lives, how it resides in social sciences, business, history, technology, and other domains, I suspect there will be a greater call for the undergraduate educators of the field to organize and to collaborate such that students in any particular demographic group have the same access and opportunity as students in any other group. Two-year colleges have historically made accessibility a high priority, and their data science curricula will be no exception.

3. Acting Locally: Planning for Our College's Program, Mindful of the Needs of Our Discipline and Community

In late 2014, current and potential students were asking our statistics faculty if our college planned to offer 'more' beyond our introductory level statistics courses. Some students specifically inquired about data science as they had heard the term a great deal in the media, in their work, and in other disciplines. Other students wanted to know if we would be considering courses that presented analysis techniques and tools for larger data sets as the term 'big data' was also becoming more common. In each case, the students were unsure of which department or program to turn to: mathematics, computer science, computer applications, business, and so on. Concurrently, our faculty's ASA members were noting that the updated *Curriculum Guidelines for Undergraduate Programs in Statistical Science* listed "Increased importance of data science" as a key point, saying "Working with data requires extensive computing skills. To be prepared for statistics and data science careers, students need facility with professional statistical analysis software, the ability to access and wrangle data in various ways, and the ability to perform algorithmic problem solving" (ASA, 2014)—all of which often went beyond the scope and requirements of our introductory offerings at the time. Additionally, the interest in two-year college data science was growing more palpable both from our own community and nationally, and organizations were beginning to [partner](#) with two-year colleges to consider programs and explore pathways to meet regional labor demand.

As we started proposing the idea of a formal Data Science Certificate to our STEM administrators in late Spring 2015, a serendipitous event occurred just days before our meeting with the provost. Earlier in 2015, Dr. DJ Patil had been appointed by President Obama as the first Chief Data Scientist of the United States, and in May of that year, on his White House blog, Dr. Patil [posted](#) “I’m the US Chief Data Scientist—and I got my start in community college” (Holst, 2015). He went on to talk about how his two-year college experience was transformative and how it did more than just start him on his math/data pathway. Not only was this a very timely item to mention to the provost, but it reinforced many of the reasons why we were considering developing our program. Whether it would be a recent high school graduate wanting to learn more about the field, an established member of the workforce seeking to improve their understanding of new analytical capabilities, a local college student recognizing the future of data science in their field, or any other member of the community, we wanted to provide the education they were seeking, mindful of the potentially transformative experiences and opportunities that are available to our students each day.

With our administration’s encouragement to further pursue the idea of a program, we were soon able to gain insight from and establish stronger relationships with local businesses, national academic and professional organizations, our county government, and the federal government (see Appendix A). While these organizations and potential partners provided specific advice and recommendations regarding software, analysis tools, data mining techniques, and so on, more importantly, they also helped us to develop the larger picture of what our program might look like and how a certificate curriculum may fit within our existing structure. Five items became clear to us during this period:

- Students would need to have affordable access to both data and tools such that they could experience the ‘messiness’ of real data rather than the often sanitized, smaller data sets to which they may be accustomed.
- There was an opportunity to expose students to key components of data acumen and to engage students in numerous fields/areas of interest; and in some cases, this engagement could be of community benefit.
- It was often both a challenge and a priority for employers to find data analysts with experience in presenting, communicating, and collaborating effectively, and developing these skills in students in a data science environment could improve both the student’s marketability and the impact of their analyses.
- Individual component courses could (and should) be made available and accessible to all qualified students, not just those pursuing a certificate, and
- Students who were seeking a certificate should have an authentic capstone experience that not only involved as many aspects of the data life cycle (Wing, 2019) as possible but also culminated with an external presentation.

4. Building Our Local Model and Connecting With Statistics

The program that we were envisioning would need to meet several criteria in order to eventually be approved by our Curriculum Committee and the Maryland Higher Education Commission (MHEC). For one thing, it would be inappropriate for the new program to draw too many resources away from our primary objectives of

student success; yet at the same time, it was becoming obvious that a modest Data Science Certificate program could enhance our community and fulfill many aspects of our college's mission and vision [statements](#), particularly in that we would be a college “characterized by agility and relevance as it meets the dynamic challenges facing our students and community.” Additionally, we obviously wanted to have a program that respected our conversations with and the letters of support we gathered from the ASA, Dr. Patil, local partners, and so on. In Fall 2016, we built the framework of our curriculum mindful of a hypothetical student's progression toward a certificate on a course-by-course basis: 1) start with statistics, 2) introduce data science, 3) provide a space to focus on presentation and visualization, 4) learn analysis and data management techniques beyond those of the introductory courses, and lastly, 5) engage in a comprehensive, culminating experience that was impactful to a program partner. These last four items became our DATA courses of Introduction to Data Science, Data Visualization and Communication, Statistical Methods in Data Science, and the Capstone Experience in Data Science, respectively. Formal course descriptions are available through our program [website](#) and course catalog.

Having our general introductory statistics course (Elements of Statistics) serve as our first course in the certificate progression was both logistically pragmatic and philosophically deliberate. This course often has a semester enrollment of close to 2,000 students with many of these students experiencing the field of statistics for the first time. Although the overwhelming majority of these students do not initially intend to pursue further work in the field, by the course's end, some students are interested in additional opportunities (if their schedule allows) to perform analyses of larger data sets and at a more robust level. Likewise, students from our introductory-level Biostatistics and Statistics for Business and Economics courses sometimes complete their classes with a more defined interest in learning further statistical techniques or performing larger scale analyses, particularly in their specific fields. We therefore determined that any one of these three courses (or their equivalent from another institution) could serve as a prerequisite to our Introduction to Data Science course (DATA 101).

This decision also facilitated a framework whereby students who had an exposure to introductory statistics but who were not necessarily seeking a data science certificate could still engage in an introductory data science experience; and we felt that this was a viable and more accessible pathway for students to engage in as opposed to a pathway with prerequisites in programming or calculus. But on a more pedagogical level, and acknowledging that my background is in statistics and statistics education, in many ways, we also felt that the skills that a student needed for developing greater data acumen dovetail the foundational skills that a student needs in order to reason statistically—skills that could also serve as a data science foundation prior to experiences with computing, advanced prediction models, working with nontraditional data types, and so on. As more and more high schools and colleges conduct their introductory statistics courses in a holistic way that encourages communication and interpretation and deemphasizes formulas and rote computations, students begin to see a more comprehensive picture of the journey of data in its collection and generation, its analysis, and the presentation and communication of that analysis. Looking more broadly for one moment, I believe that

it will be incumbent upon data science educators to continue addressing the fundamental question of how will providing a curriculum for helping all students develop data and statistical literacy differ from a data science curriculum (if it will differ at all) going forward.

5. Implementing Our Program With an Eye to the Broader Data Science Education Landscape

Our college's program is very much in its infancy. In February 2017, the new courses were approved by the college and cataloged; and MHEC approved our Data Science Certificate program in August of that year. Our first classes were offered in the Fall 2017 semester, and with the original idea that most students would engage in one course per semester, we conferred our first certificates on schedule in May 2019 to three students. Our initial program goals were quite modest, mindful that like most two-year colleges, we often had greater institutional urgencies and priorities regarding student completion and retention. Montgomery College's program was developed mindful of our resources, local needs, and partner feedback; and that allowed us to provide a certain level of customizability in response to our community's characteristics and needs, something at the very heart of most two-year colleges' missions. Our administration also encouraged us to implement the courses with fully open resources of no cost to students. We also very much wanted our program to support the full range of our student population: students of various backgrounds, interest levels, facility with mathematics, workforce experience, and so on. Bear in mind that as a two-year college, our students vary greatly in terms of age, educational background, citizenship, and so on; and we wanted a data science program that reflected our whole college population (not just a certain demographic) and included those groups who may be traditionally underrepresented in traditional STEM fields.

In both of our 100-level courses of Introduction to Data Science (DATA 101) and Data Visualization and Communication (DATA 110), our faculty are mindful of rapidly changing technology, our program's commitment to a zero-cost experience in terms of resources, and a focus on knowledge rather than software. As a result, each semester may require a thoughtful and at times lengthy revision of materials for the course. Additionally, with students arriving with different levels of experience in programming, there can be coding, knitting, and publishing frustrations. Rather than seeing this as an impediment to student success, the courses embrace this variability as a catalyst for collaboration and teamwork. While students gain experience individually through coding, web scraping, data mining, and creating portfolios in both GitHub and RPubS, they also discover the importance of transparency, code-sharing, and replication through their interactions with peers. With all assignments and projects shared and posted, students are encouraged to highlight new discoveries both in their own work and in the work of others. Along with other benefits, this helps students to consider meaningful interactivity in their visualizations that may improve the presentation of their analyses. Interestingly, it is in some of the more collaborative exercises with feedback from others in the class that the students develop their greatest appreciation for the benefits of clear communication, both among peers and to a more general audience.

Mindful of this last point, just as our students have varying levels of comfort with the technical aspects of the course, they also have varying levels of experience (and comfort) with different presentation and writing skills. Writing assignments that consider diverse communication styles are interwoven throughout the DATA 110 course, and students present to each other regularly. They learn to make an elevator pitch, to discuss ethical practices, and to carefully consider the aesthetics of their work. At the same time, the students must consider that their future audience may not be a group of peers with similar interests or with a similar level of experience in data science. Consider, for example, that the annual ASA DataFest competitions (in which our students have participated) require a team to distill their many, many hours of wrangling, analysis, and discovery into a presentation of just five minutes and using only two or three slides. Communicating both the results and the applicability of the research in a concise, meaningful, and nontechnical manner is an important tenet in much of the work of our program. This is not just providing outsourced employer training; rather, we are providing an experience that is germane to the full data science experience and was requested and recommended by our community partners.

At the 200 level, it is in our Statistical Methods in Data Science class (DATA 201), and the Capstone Experience in Data Science (DATA 205) that students take the final steps toward our certificate and a deeper dive into the discipline. While our 100-level courses provide thoughtful, practical, and challenging introductory data science experiences to students, it is at the 200 level that students develop a greater exposure to more advanced analytical methods and software and become more involved with the complete data life cycle in an integrated way. For the DATA 205 Capstone Experience, we first partnered with our county government through its Department of Technology Services and their open data initiative, dataMontgomery. Through arrangements made between their department's Data Services management and our program coordinator, our capstone students have the invaluable opportunity to work closely with data professionals and to engage in meaningful, civic-minded research using large, content-rich data. The course culminates with [presentations](#) to members of our local county government and staff, all leading to discussions with the data owners and ideas for future work together. While these specific experiences are important milestones for our recently launched program, in a larger sense, they exemplify much of what both the two-year college and data science communities aspire to with their undergraduate experiences. Our students benefit, the county benefits, and both the college and the county government are enthused to maintain this arrangement for the foreseeable future.

For continued program guidance and in keeping with best practices shared from other programs and organizations, we developed an Advisory Board consisting of local stakeholders and partners including professional data scientists, academic association representatives, and potential employers. In addition to what we continue to learn from our own experiences and the broader data science education community, the Board provides an invaluable resource in terms of discussing curricular development, assessing student achievement, and recommending general program direction. Additionally, during the first two years of our program, the NASEM Roundtable on Data Science Postsecondary Education concurrently and indirectly provided us with great insight via their Roundtable meetings as we were able to hear data science leaders validate many of our

guiding principles and objectives. This was particularly true regarding the importance of providing students with opportunities to strengthen business and leadership skills (including communication, storytelling, collaboration, giving and receiving feedback, and tailoring presentations to different audiences) and the potential for data science to better engage women and minorities (see Appendix B).

Launching a two-year college program is challenging, and there are many constraints in terms of program sustainability, accreditation, approval, and so on. Regardless of a student's background or enthusiasm, in many instances two-year college students might not partake in advanced coursework in data science (or any other discipline) if the coursework is not transferable, eligible for financial aid, and so on. Aware of the challenges that interested students might face when pursuing a certificate, our administration is confident that when four-year degrees are developed to which our data science students can transfer, an associate's degree will likewise be developed by our college, thereby addressing many of these challenges. Another issue is attracting and recruiting students to a new program; and despite frequent mention of data science in employment listings and certain media, a great deal of outreach is often needed. Our faculty go to great lengths to avail themselves to students, to promote the benefits of the field to both students and colleagues, and to reach out to our local business, research, and academic communities. This is very important as in each of these groups, members vary in their understanding and optimism as to what a two-year college data science program graduate can contribute. Recruiting and developing data science faculty are also essential, and we were quite fortunate that there were data scientists within our part-time faculty as well as professional development support by our administration for both part-time and full-time faculty. As with any new two-year or four-year college program, innovative thinking, support, and effective internal and external partnerships will be crucial going forward.

At the launch of our program in Fall 2017, we had 30 enrollments for our DATA courses, all of which were in DATA 101. Two-and-a-half-years later, our total DATA enrollments had nearly tripled with 89 students in Spring 2020 spanning all of our courses. We experienced a slight drop in Fall 2020 enrollments; but as the portability of the instruction to a fully online environment was very manageable when mandated last March due to COVID-19, we have continued to provide sections of each DATA course in Fall 2020, and we fully expect the curriculum to remain in demand in 2021 and beyond. As for the program certificates, as mentioned earlier, our first certificates were conferred to three students in May 2019, and 12 more students earned their certificates this past academic year. Although these numbers may seem small to readers at large universities with full degree programs, our administration was thrilled that we had meaningfully involved the college in this field and had provided a benefit to our students and our community in an impactful way, all with minimal redistribution of college resources.

6. Back to the Broader Picture: Future Challenges and Concerns for Two-Year Colleges

While two-year college data science programs may present unique and rewarding opportunities, they also share many of the same concerns of their four-year college counterparts. As mentioned in the *Opportunities and*

Options report, “these are early days for undergraduate data science education, academic institutions should be prepared to evolve programs over time” (NASEM, 2018)—but what exactly will that evolution look like, and to where will it lead? Who will be involved in the discussions that move undergraduate data science forward, and will two-year colleges be an important part of those conversations?

Another item of concern is assessment of curriculum and program effectiveness at a larger scale. We all want to make sure that the education that we are providing is preparing students for larger goals than just the next job. How will we measure success, and how will our discipline change from those discussions? Any science, and particularly an emerging one, needs to have some flexibility for growth, new perspectives, and new definition. The ambitious central mission that Xiao-Li Meng introduced in the inaugural *HDSR*—namely, “to help define and shape what DS (Data Science) is or should be” (Meng, 2019)—will require communication both within and between our institutions, our professional organizations, industry, government, and other key stakeholders—two-year colleges included.

Managing expectations has been key in our college’s development of good relationships, and just as our college does not deem a graduate of our introductory statistics class to be a statistician, or a graduate of our introductory physics class to be a physicist, so too we do not claim that a graduate of our college’s introductory data science course or our certificate program is necessarily a data scientist. However, as in these other disciplines, we have the opportunity to inspire and to provide a meaningful and affordable first step for our students. At a minimum, by offering courses in data science, two-year colleges assist in creating a more data literate and data aware community, and it is there in which I believe our next great opportunities (and next worries) may lie.

7. Moving Forward: Next Opportunities and Greatest Needs

Years ago (and maybe it is still the case in some university systems), the ‘go to’ general education mathematics elective for undergraduates was something like College Algebra. In time, courses that included applications such as financial mathematics and exploratory data analysis were encouraged. As mentioned before, at Montgomery College, the largest general education elective in our mathematics department is our Elements of Statistics course with an enrollment of over 4,000 students annually. This is in part due to the realization by many transferring institutions and specific degree programs that an exposure to basic statistical thinking is an outcome that higher education wants for its undergraduates. If we as a discipline want introductory data science to become such a widely accepted course in our academic systems (both two-year and four-year), we will need to demonstrate even more about the field’s applications to society, to continue to infuse discussions regarding the appropriate use of data in our curricula, to stress the importance of communication skills, and a great deal more. While some colleges and universities (listed in section 3 of the *Opportunities and Options* report) have developed successful models where introductory data science is accessible and relevant to the general undergraduate population, this may still not be the norm at most institutions.

One of the recommendations of the *Opportunities and Options* report that is relevant here is as follows: “Recommendation 2.3 - 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” (NASEM, 2018)—and note that it says “*all*.” As Alfred Hero, Co-Chair of the Committee on Envisioning the Data Science Discipline: The Undergraduate Perspective, remarked, “in our discussions the committee did support incorporating data literacy as a general requirement for all undergraduate students, supplementing writing, communication, and numeracy requirements that are common today.” (Haas et al., 2019). Defining the components and boundaries of such data literacy is a challenge, but many undergraduate programs, two-year colleges included, have begun to address this question.

The larger issue to me is one of *where* the students will find this data literacy. It is not clear if this general data literacy will be distinct from (or a subset of) the “data acumen” for aspiring data scientists that was discussed in the *Opportunities and Options* report—but it is a good bet that the data science educators responsible for instilling data acumen in their program’s students will also be the educators at the forefront of developing (and possibly executing) courses of general data literacy to their larger undergraduate populations. The opportunities are intriguing, the benefits are many, but the scale may be overwhelming. Another key point is that many undergraduate students do not necessarily engage in their introductory coursework at four-year institutions. Preliminary findings from the 2015 Conference Board of Mathematical Sciences (CBMS) *Survey Report* (the most recent report available) listed an estimated 280,000 introductory probability and statistics enrollments in two-year colleges in Fall 2015 accounting for over 44% of such enrollments in U.S. higher education (CBMS, 2018). Keep in mind also that over 220,000 Advanced Placement Statistics exams and nearly 448,000 Advanced Placement Calculus exams were taken by high school students in 2018 (College Board, 2018). With the many advances in teacher preparation in statistics education at the K-12 level in recent decades, more high school graduates are arriving at our college with a firm background in introductory statistics and an eagerness to work with data on a larger scale and in more depth.

If ‘Data Science 101’ becomes a staple for undergraduates at four-year colleges, I believe two-year colleges will develop and offer equivalent (or identical) courses in the name of student achievement and institutional mission success. If executed well (and preferably not at the expense of other college programs), this general education platform may end up being the next great opportunity for two-year colleges, for data science, and for our graduates. But this also means that two-year colleges and four-year colleges will continue to share one major concern that we are already grappling with, namely, the need for qualified data science instructors to meet demands. Whether it is for their specific data science programs or for the general education of their undergraduates, I believe that all colleges will continue to recognize the merits of data science education and thus only add to any scarcity that may presently exist. Professional development and teacher education in data science must become an even higher priority for the field, particularly at the undergraduate level. As shown in the Roundtable Meeting summary and the TYCDSS report, successful initiatives exist that are addressing this concern, but more support and more initiatives are needed. Additionally, if we continue to advocate that

statistics and data science become an essential part of the culture at the K-12 level, it will be incumbent upon postsecondary educators to support the efforts of K-12 teachers seeking preparation and continuing education in those fields.

8. Conclusion

Given that our discipline seeks greater data literacy in our workforce, in our undergraduate population, and in our general society, we must consider the opportunities of two-year colleges. These colleges represent a resource that can greatly assist in achieving these goals. With a history of providing accessibility and transformative experiences for their students and for developing programs to help meet the needs of society, two-year colleges will be key contributors to the future of data science and undergraduate data science education. The need for more data science educators is great, and many opportunities may not be realized without addressing this need.

Acknowledgments

The author would like to thank John Hamman, Margaret Latimer, Kathryn Linehan, and Rachel Saidi for their comments on previous versions of this article.

Disclosure Statement

Brian Kotz has no financial or non-financial disclosures to share for this article.

References

American Statistical Association. (2014). *Curriculum guidelines for undergraduate programs in statistical science*. <http://www.amstat.org/education/pdfs/guidelines2014-11-15.pdf>

Blair, R., Kirkman, E., Maxwell, J.W. *CBMS 2015: Statistical abstract of undergraduate programs in the mathematical sciences in the United States, Fall 2015 CBMS Survey*. American Mathematical Society. <http://www.ams.org/profession/data/cbms-survey/cbms2015>

College Board. (2018). AP exam volume changes (2008–2018). <https://secure-media.collegeboard.org/digitalServices/pdf/research/2018/2018-Exam-Volume-Change.pdf>

Gould, R., Peck, R., Hanson, J., Horton, N., Kotz, B., Kubo, K., Malyn-Smith, J., Rudis, M., Thompson, B., Ward, M.D., and Wong, R. (2018). *The two-year college data science summit: A report on NSF DUE-1735199*. <https://www.amstat.org/asa/files/pdfs/2018TYCDSFinal-Report.pdf>

Haas, L., Hero, A., & Lue, R. (2019). Highlights of the National Academies Report on "Undergraduate Data Science: Opportunities and Options." *Harvard Data Science Review*, 1(1).

<https://doi.org/10.1162/99608f92.38f16b68>

Holst, L. (2015). Email from DJ Patil: "How I Became Chief Data Scientist." *Obama White House Archives*.
<https://obamawhitehouse.archives.gov/blog/2015/05/06/email-dj-patil-how-i-became-chief-data-scientist>

Meng, X.-L. (2019). Data science: An artificial ecosystem. *Harvard Data Science Review*, 1(1).
<https://doi.org/10.1162/99608f92.ba20f892>

National Academies of Sciences, Engineering, and Medicine. (2018). *Data science for undergraduates: Opportunities and options*. The National Academies Press. <https://doi.org/10.17226/25104>

National Academies of Sciences, Engineering, and Medicine. (2020). *Roundtable on data science postsecondary Education: A compilation of meeting highlights*. The National Academies Press.
<https://doi.org/10.17226/25804>

Wing, J. M. (2019). The data life cycle. *Harvard Data Science Review*, 1(1).
<https://doi.org/10.1162/99608f92.e26845b4>

Appendix

Appendix A: Early Key Partners and Advocates

Key partners and advocates supporting our early program development included:

- The Washington, DC, area's ASA DataFest competition and its hosts, Summit Consulting and the University of Maryland's Joint Program in Survey Methodology,
- The Montgomery County (MD) Government's Department of Technology Services and its [dataMontgomery](#) open data and open government initiatives,
- The ASA's Section on Statistics Education (as it was known back then, now the "Section on Statistics and Data Science Education"),
- The National Academies of Sciences, Engineering, and Medicine (NASEM) and its Roundtable members who enthusiastically supported our program development and welcomed our participation in their data science education initiatives, and
- Dr. DJ Patil who invited our college's data science Program Development Team to a private meeting at The White House complex in February of 2016.

Appendix B: Related Comments and Insights from Roundtable Meetings

Although the 11th meeting of the NASEM Roundtable on Data Science Postsecondary Education (June 2019) focused on two-year college data science education, comments and insights from earlier meetings of the

Roundtable also addressed many of the principles of our program and the general mission of two-year colleges. From the *Roundtable on Data Science Postsecondary Education: A Compilation of Meeting Highlights*:

Meeting #3: Data Science Education in the Workplace – May 1, 2017

(Emily) Plachy (IBM) suggested it would be helpful if stakeholders created a public education system for data science where organizations could share ideas for workplace training.

(Ashley) Lanier and (Ashley) Campana (Booz Allen Hamilton) noted that ... employees with data science specialties need to learn “consulting skills” such as communicating, storytelling, working with clients, working in a team, understanding an audience, and choosing the right approaches.

(Nicholas) Horton (Amherst) noted that community colleges could play a role in training because of their low-cost, flexible offerings.

Patrick Riley, Google, responded that while people working in traditional technical fields talk predominantly with other technical people, data scientists need to be able to explain difficult concepts to non-technical audiences. (Claudia) Perlich (New York University) agreed that students have to learn to frame problems clearly for non-technical audiences. Riley suggested that students would benefit from practice exercises in which they have to present summaries of analyses to varied audiences.

Meeting #4: Alternative Mechanisms for Data Science Education – October 20, 2017

(Al) Hero (University of Michigan) said that he has witnessed a decline in both data and visual literacy among high school students; he wondered how to reverse these trends and how to engage more students in science.

Deborah Nolan, University of California, Berkeley... emphasized the need for undergraduate faculty to continue to focus on teaching fundamentals instead of emerging technologies.

Meeting #8: Challenges and Opportunities to Better Engage Women and Minorities in Data Science Education – September 17, 2018

Eric Kolaczyk, Boston University... suggested that the emergence of data science, with its focus on new paradigms, has the potential to create a watershed moment to better engage women and minorities in STEM fields and beyond.

Meeting #9: Motivating Data Science Education Through Social Good – December 10, 2018

(DJ) Patil championed the role of 2-year institutions in offering introductions to data science for social good.

(Rahul) Bhargava teaches a cross-disciplinary course, hosted by the MIT Humanities department, called Data Storytelling Studio, in which students “consider the emotional, aesthetic, and practical effects of different [data] presentation methods.”

Meeting #10: Improving Coordination Between Academia and Industry – March 29, 2019

(Mary Ellen) Sullivan said that MassMutual emphasizes skills that are essential for business but rarely developed at the undergraduate level, such as leading, giving and getting feedback, and tailoring presentations to different audiences.

(Peter) Norvig (Google) said that one of Google’s most significant responsibilities is to help grow the field of data science, starting at the K-12 level by developing curriculum and educating teachers.

Daniel Marcu, Amazon ... noted that members of industry and academia alike should be making efforts to enhance their communication and collaboration.

©2020 Brian Kotz. This article is licensed under a Creative Commons Attribution (CC BY 4.0) [International license](#), except where otherwise indicated with respect to particular material included in the article.

Preview image for this article courtesy of Montgomery College.