Report of the Working Group

Data Science and the Digital Humanities

Working Group Members:

Jeannie Albrecht, Department of Computer Science
Jon Bakija, Department of Economics
Phoebe Cohen, Department of Geosciences
Richard De Veaux, Department of Mathematics and Statistics
Beth Fischer, Williams College Museum of Art
Barron Koralesky, Office for Information Technology
Tiku Majumder, Department of Physics
Laura Muller, Quantitative Skills Programs and Peer Tutoring
Jason J. Storm, Department of Religion
Chad Topaz, Dept. of Mathematics and Statistics
Chad Weinard, Williams College Museum of Art
Safa Zaki, Dept. of Psychology

June 2020

Table of Contents

Executive summary

The intersections and connections among science, technology, and society have helped shape history. These connections are at the heart of issues that are critical to human existence, such as climate change, medical ethics, economic growth, and inequality and social justice. At the same time, society is grappling with the role of data and digital information in all aspects of our lives, including politics, the arts, health care, law enforcement, and more. As data have increasingly become the medium for the collection, storage and transmission of information, all liberal arts students will need to have basic literacy in data science and the digital humanities.

Data science is an interdisciplinary approach through which inquiry across the liberal arts can be framed and carried out by leveraging the study of data. A data science approach integrates knowledge from computing and from the mathematical/statistical sciences with critical expertise in the domain of inquiry. Data science exists alongside a related but distinct field, *digital humanities.* Digital humanities involves the use of digital tools, methods, and approaches to extend the human capacity to explore questions relating to people, cultures, and communities.

Williams has the foundation of a strong data science and digital humanities program in the form of student interest, selected course offerings, faculty scholarship, and various offices, programs, and individuals. Strategic planning presents the opportunity to build on these existing strengths and create a signature program in data science and digital humanities. Such a program would teach data-driven approaches across the arts, humanities, social sciences, and natural sciences. It would support student and faculty research addressing pressing contemporary challenges through the use and critique of digital methods. And it would promote the coherent, clear, and credible use of data and digital technology in scholarship and communication, as well as the ethical use of technology and data.

Vision and goals

The relationship between the liberal arts and technology is at a critical juncture. At the present moment, humanity needs a greater understanding of the interconnections between science, technology, and society. These interconnections are at the heart of crucial issues including climate change, medical ethics, economic growth, and social justice. In addition, our society is grappling with the role of data and information in all aspects of our lives, from our political systems, to health care, to social media, and even our kitchen appliances.

What is data science?

Many areas of modern society including politics, the arts, health, science, and commerce, increasingly rely on the large quantities of data enabled by a revolution in computing over the past 30 years. As a result of this revolution, the field of *data science* was born. Data science develops and applies tools from mathematics, statistics, and computing to study applications in a wide range of domains, even ones that did not historically adopt data-driven approaches. It is crucial to understand two characteristics of data science. First, the effective practice of data science cannot occur without *domain expertise*, that is, expertise in the domain to which data science tools are being applied. The skills of data science live, then, at the intersection of three areas: quantitative tools, computing, and domain expertise. Second, data science is much more than this aggregated set of skills. Data science complements knowledge in traditional disciplines through its attention to best practices for using data and the social-ethical implications and challenges of doing so. Because of the multiple skill areas from which data science draws, the wide range of domains of applicability, and the engagement of new questions related to ethics and society, data science is a fundamentally interdisciplinary field that supports critical thinking.

Data science education is at the core of the liberal arts, contributing substantially to fields that span all divisions at the college. Examples include using data science to: uncover trends related to gender in 20th-century literature; understand patterns of crime in urban areas; predict the spread of epidemic disease; detect fraudulent artworks; help cities predict health code violations in restaurants; determine which colleges and universities do the best and worst jobs of promoting intergenerational income mobility; attribute authorship of Beatles songs and *The Federalist Papers*; learn what public policies work best to fight poverty around the globe; analyze large volumes of 911 emergency call data to shed light on how to conserve police resources and spare citizens contact with the criminal justice system that is avoidable; and predict the patterns of activity in the brain associated with the meanings of particular nouns. As data has increasingly become the medium for the collection, storage and transmission of information, all liberal arts students will need to have basic literacy in data science.

What are the digital humanities?

Data science does not exist in a vacuum. During a time of unparalleled access to data and computational tools, it is crucial to communicate knowledge effectively, to be mindful of how knowledge is produced, and to be aware of the history, ethics, and social impact of digital work. *Digital humanities* refers to the use of digital tools, methods, or approaches to extend the human capacity to explore questions (and pose new kinds of questions) relating to people, cultures, or communities. Critically, digital humanities also includes using methods or critiques developed in the humanities to consider how technology and computing intersect with social, cultural, and historical factors. Data science skills are often a key component of digital humanities work, along with tools and approaches for digitization and research, analysis and interpretation, presentation and dissemination—all working together with domain expertise. Digital humanities thrives on interdisciplinary connections, collaboration, and open scholarship.

Examples of digital approaches to humanities topics include creating virtual reconstructions of places or artifacts to connect isolated research results or test proposed ideas; mapping or networking ideas, people, and events to determine patterns; creating online archives or exhibitions that allow novel/greater access to resources; the use of digital tools and visualization methods to research artifacts; and the creation of written and/or visual art using digital tools or digital artifacts. Often the approach of the digital humanities is exploratory, and is meant to challenge existing patterns of thought or prompt new ones.

Digital humanities at Williams will draw on the college's unique strengths and opportunities. At Williams, the digital humanities would bring digital tools and approaches to the study of the humanities, and bring humanities questions, tools and approaches to the study of the "digital" in disciplines like computer science, math and statistics. Further, digital humanities at Williams would connect the fine arts to the sciences and humanities through a shared interest in critical making. Perhaps most importantly, the digital humanities at Williams would focus on pedagogy, developing project-based approaches to teaching and learning marked by collaboration and interdisciplinary connections.

Managing rapid technological change for the good of all will require technologists and scientists expert in ethics and mindful of history, social context and impacts. We will need historians and scholars conversant in data and technology. Creative ideas will come from artists trained in critical thinking and fluent in digital culture. More broadly, citizens will need to be equipped with the skills and knowledge to make choices and make their voices heard. This is the opportunity for digital humanities at Williams.

Vision, mission, and values

This strategic initiative will empower the Williams community to think critically, inquire, and create in a world transformed by data and digital technology.

The vision will be realized through a four-pronged mission that will:

- Equip students, staff, and faculty across the arts, humanities, social sciences, and natural sciences with digital and data-driven approaches.
- Enhance digital and data science approaches with domain knowledge and humanities-driven analysis.
- Address pressing contemporary challenges through the use and critique of digital and data-driven approaches.
- Facilitate the coherent, clear, and credible use of data and digital technology in scholarship and communication as well as the ethical use of technology and data.

Our initiative incorporates several other high-level goals:

- Integrate data science and digital humanities in both the curriculum and co-curriculum
- Support the use and understanding of data on the part of students, faculty, and staff
- Connect the sciences, social sciences, and humanities

Certain key values have informed all aspects of this initiative:

- Collaboration across disciplinary boundaries
- An emphasis on evidence-based reasoning
- A focus on inclusion, diversity, and equity
- Social responsibility
- A culture of assessment

Description and appraisal

Data science

Williams has the seeds of a strong data science initiative in the form of student interest, selected course offerings, faculty scholarship, and various offices, programs, and individuals. However, these efforts lack a centralized home, a specific mission, and a defined curricular pathway. This diffusion makes it difficult to assess the breadth of data science work being done on campus by faculty, staff, and students. Nonetheless, we discuss a few key points below.

Curricular structures

Students can theoretically pursue majors related to data science through the contract major option. While we have not audited the details of contract majors in recent years, we have looked at the total number of annual contract majors in each division over the past 12 academic years (graduating classes of 2008–2019). The median number in each division is between zero and two per year, so it is unlikely that the contract major program is contributing significantly to data science education on campus.

We have also compiled the average number of double majors per year over the past twelve years that include a major from mathematics, statistics, or computer science. This figure is 11 students per year for Division 1 majors, 31 students per year for Division 2 majors, and 16 students per year for natural science majors in Division III (astronomy, astrophysics, biology, chemistry, geosciences, and physics). Thus, there are potentially 58 double majors per year (on average) that combine a data science tool area with a potential domain area of application. Additionally, an average of 6 students per year double major in two of mathematics, statistics, and computer science and therefore combine two data science tool areas in their studies.

Finally, for 2010–2019, we identified an average of 148 students per year who graduate with a single major in Division 1 or 2 and take at least two courses in computer science, mathematics, or statistics. We therefore have a substantial number of graduates who (single) major in the arts, languages, humanities, and social sciences, and who would be well positioned to integrate computational and/or quantitative approaches into their work.

Courses

Courses at Williams that contribute to the teaching of data science tools include the following:

- CSCI: 134 (Introduction to Computer Science), 136 (Data Structures and Advanced Programming), 237 (Computer Organization), 256 (Algorithm Design and Analysis), 315 (Computational Biology), 358 (Applied Algorithms), 361 (Theory of Computation), 374 (Machine Learning)
- ECON: 255 (Econometrics), 371 (Time Series Econometrics and Empirical Methods for Macro), 471 (Topics in Advanced Econometrics)
- POEC: 253 (Empirical Methods in Political Economy)
- GEOS: 214 (Mastering Geographic Information Systems)
- MATH: 150/151 (Multivariable Calculus), 250 (Linear Algebra), 307 (Computational Linear Algebra)

- PSYC: 201 (Experimentation and Statistics)
- STAT: 101 (Elementary Statistics and Data Analysis), 161 (Introductory Statistics for Social Science), 201 (Statistics and Data Analysis), 202 (Introduction to Statistical Modeling), 346 (Regression and Forecasting), 360 (Statistical Inference), 442 (Statistical Learning and Data Mining)

Scholarship

Scholars in many parts of campus are already engaging tools from data science. A complete accounting is beyond the scope of this document, but we mention a few examples here.

The recent Data Science Boot Camp for faculty involved nine projects in data science, all of which will continue throughout 2020. Carolina Melgarejo-Torres (Spanish) studies large corpora of Spanish text spoken by native speakers in order to understand the roles of and restrictions of certain words. Ben Snyder (Sociology) is analyzing data related to spatiotemporal variation of student emotions on the Williams campus. Derek Dean (Biology) is designing data-intensive laboratory activities to teach students about gut development in sea urchins. Tomas Adalsteinsson (Golf) is aggregating a large, first-of-its-kind database of golf statistics in order to uncover predictors of strong performance and to develop new strategies for his team. David Gürçay-Morris (theater) is interested in the critical environments and the creative process as manifest through the development and review of video games. Prisca Gayles (Africana Studies) is investigating the representation of Afrodescendant Argentinians within the country's census data. Ezra Feldman (English) has launched a study of the Williams College Libraries' acquisitions data. Sara Dubow (History) is mining massive amounts of Congressional data to understand the role of conscientious objection and religious accommodation in federal law. Paul Karabinos (Geosciences) is using zircon chemistry data to determine where certain igneous rocks were formed.

The vast majority of economics faculty and research students engage primarily in empirical work involving statistical analysis of data sets. This work requires programming, data curation, data management, and causal inference. Data analysis and/or machine learning play important roles in faculty research in many other disciplines across the college, including for example psychology, biology, physics, geosciences, statistics, computer science, and political science.

Within physics, Daniel Aalberts uses data mining of proteomics and genomics data to discover which sequence features influence protein expression and messenger RNA stability. Those model inferences allow him and his colleagues to redesign synonymous genes to produce much more protein than the native gene of interest. Graham Giovenetti, the most recent hire in the physics department, is involved in experimental searches for evidence of dark matter particles, which involves collecting and processing vast amounts of data over long periods of time. In general, these modern particle astrophysics experiments record several megabytes of data for every particle interaction that occurs within a detector, collecting petabyte-scale datasets over the lifetime of an experiment. His group increasingly relies on supercomputers and machine-learning techniques to process these large, high-dimensionality datasets and extract interesting physical parameters, benefiting from, and in many cases contributing to, the field of data science.

This small subset of examples gives a sense of the breadth of inquiry on campus. Later in this document, we propose curricular initiatives in data science. Beyond the direct benefit to the students themselves, this training would increase the pool of students who are qualified to work as research assistants on projects involving data manipulation and analysis, increasing faculty's collaboration, and potentially co-authorship, with students.

Existing resources for data science

Beyond the individuals who undertake activities related to data science, and beyond the physical spaces allocated to them, resources for data science are situated in a wide range of offices and programs at the college, providing a foundation on which to build a future initiative. We mention here a few key resources. Overall, Williams has some excellent resources, and at the same time, would benefit from coordination among these.

Libraries

The Libraries support data science in three broad areas: managing access to data, assisting faculty and staff in their use and management of data, and preserving and making accessible the products of data science. Librarians license access to datasets that support the curriculum and faculty and student research. As a Federal Government Depository, the Libraries provide access to federal government data. Librarians also facilitate access and use of collections as data in cases like the Hathi Trust Research Center. Working through the College's Data Governance Policy, librarians are able to supply data from their own operations. Librarians teach users how to search for data and primary sources, which involves discussing who produces information and the limits to access, and teach the critical assessment of data and sources, including the politics of information and data (*e.g.,* sessions on biases in Google and in library subject headings). The digital archivist advises faculty on best practices in data storage and management, on providing access to datasets in our institutional repository, and on the long term preservation of datasets, often with an eye towards compliance with grantor requirements.

In all these areas, the librarians recognize a need to increase capacity to support data sciences at Williams. More specifically, they would like to: expand their knowledge of data fields relevant to their roles as subject liaisons; expand their knowledge of data structures and data organization; and develop a broad basic understanding of core data science tools and how they relate to data they assist users in finding. These advances would enable a move from informing others about resources to actually training and teaching them.

Office for Information Technology

Office for Information Technology (OIT) staff support teaching, learning, and scholarship through the application of technology. They consult with faculty, students and staff in a number of ways. Starting by helping to refine pedagogical and research goals, they can facilitate the development of workflows for gathering data and other materials; research and recommend tools for aggregation, analysis, visualization and display; help with the development of assignments, rubrics, assessments and teaching materials; develop workshops, teach class sessions and provide consultations for individuals; and develop and implement required systems, infrastructure and integrations. They provide classroom, media development and computation environments, develop and oversee student workers for projects, and consult with and support students on their original projects and scholarship. OIT realizes that they will need to increase the depth and scale of this support if the college moves forward with this initiative. This would entail professional development for staff, as well as additional investments to support a constantly changing technology environment.

Quantitative Skills Program

The Office of Quantitative Skills Programs and Peer Support (QSPPS) supports the development and use of quantitative skills across the Williams curriculum. This office is specifically charged with developing a more coherent infrastructure for STEM+ pedagogy and programming to support both Williams' students and faculty. It also assists with dissemination and assessment of new initiatives around engaged learning in quantitative skills and quantitative reasoning. As the college's efforts around data science and digital humanities grow, Quantitative Skills Programs and Peer Support anticipates spending more time working with faculty across the curriculum to think about scaffolding data science skill sets, including those faculty who have less experience in thinking about quantitative reasoning as an aspect of their teaching. QSPPS also oversees the Peer Academic Support Network of tutors in academic areas. In recent years, there has been increased demand for tutoring around data analysis using the software package R in all areas of Division 3. The data science initiative would likely increase the demand for statistical and programming support for students.

Geographic Information Systems

Geographic Information Systems (GIS) comprise a core set of tools used for spatial analysis in data science. GIS are used across many disciplines at the College, from biology to political science to classics. Currently, these methods are formally taught in just one course, in the geosciences. If we move forward with the initiative, we may want to find additional opportunities to teach and support GIS methods.

Data Science Boot Camp for Faculty

Cultivating interest and investment in the data sciences from faculty spanning all academic divisions will be critical in the successful growth of a data science program. As a first step, the college ran a three week data science training course, or *boot camp,* for nine faculty during the 2020 Winter Study period. The initiative was supported by a grant from the Sherman Fairchild Foundation and included faculty from across the college (see Data Science Scholarship) who had little to no experience with data science but wanted to learn more about the field. During the boot camp, participants:

- Learned about the field of data science and how it could become part of the curriculum by hearing from guest speakers who use data science in their research and teaching;
- Completed data science education modules designed to teach coding, data acquisition, data cleaning, and data visualization at an elementary level;
- Began work on a project in their field that involves a data science approach, and that will continue throughout 2020.

Institutional comparisons

Data science is a relatively new and rapidly evolving field of study. Professional academic associations including professors of statistics, computer science, and mathematics from many different schools have recently produced reports offering guidelines for structuring a data science curriculum. These reports include the 2016 Park City Math Institute's "Curriculum Guidelines for Undergraduate Programs in Data Science," and the ACM Data Science Task Force's "Computing Competencies for Undergraduate Data Science Curricula" (Draft 2, 2019). Our own college is home to some of the leaders of these efforts. Statistics Professor Dick De Veaux was the lead author on the first report, and Computer Science Professor Andrea Danyluk was the co-chair of the committee that produced the second report. Colleges and universities have been experimenting with the design of data science curricula, including coursework, majors, and minors within data science itself, as well as related coursework and data science tracks within other disciplines and majors. To explore ideas for how to design a data science curriculum, we read the two reports noted above, investigated data science curricula at fifty top liberal arts colleges in the US,[1] and also looked into examples of data science curricula at a few selected large research universities.

In Appendix 3, we provide more detailed information on data science curricula at other schools, including: examples of introductory-level courses in data science; requirements for majors or minors in data science at selected schools including Berkeley, Denison, Macalester, Mount Holyoke, Smith, and Wesleyan; and the major curriculum proposed by the Park City Math Institute. Appendix 3 provides a fairly representative picture of the range of different approaches to designing a data science curriculum, and highlights some particularly innovative ideas. In the rest of this section and in Appendix 4, we summarize a few major themes and lessons uncovered by our review of data science curricula at these other schools.

Which schools offer a major in data science?

We identified eight colleges that currently offer a major in data science, sometimes under a different name. Claremont-McKenna, Mount Holyoke, and Wellesley all have majors called "Data Science," while the equivalent major at Smith is called "Statistical and Data Sciences," and the equivalent major at the US Military Academy is called "Applied Statistics and Data Science." Denison calls its equivalent major "Data Analytics." According to one of the founders of the Denison program (Havill, 2019, p. 12), "we intentionally left the word 'science' out of the name of the program, opting for 'Data Analytics' over 'Data Science,' to signal that this course of study is not geared exclusively to science-oriented students." The other two of the existing eight major programs at top liberal arts colleges are rather different from the first six. Bowdoin calls its major "Digital and Computational Studies," and this major puts relatively less emphasis on intermediate and upper-level statistics and computer science, and more emphasis on interdisciplinary courses combining data science with concerns from across the liberal arts. Bucknell's major is called "Business Analytics," and combines many common elements of a data science major with elements of an undergraduate business degree. One other top liberal arts college that we know of, Dickinson, is currently considering a proposal for a new major in "Data

[1] To study data science curricula at top liberal arts colleges, we visited the websites of each of the 50 highest-ranked schools in the 2020 *US News* list of National Liberal Arts Colleges, searched for "Data Science," looked over the lists of majors, minors, and areas of study, and read descriptions of the curricula that we found.

Analytics." While our review of data science programs at larger research universities was less systematic, some examples that we looked into include the Data Science majors at the University of California at Berkeley and the University of Michigan, and the Statistics and Data Science major at Yale University.

Which Schools offer a minor in data science?

Twelve of the top fifty liberal arts colleges that we reviewed currently offer a minor, concentration, or equivalent in data science or something similar, including Bowdoin, Claremont-McKenna, Colby, Connecticut College, Davidson, Lafayette, Macalester, Mount Holyoke, Union College, Wesleyan, and Whitman. The minor is called "Data Analytics" at Union, "Data Analysis" at Wesleyan, "Data, Information, and Society" at Connecticut College, "Digital and Computational Studies" at Bowdoin, and "Data Science" everywhere else. Wesleyan also offers an eight-course "Applied Data Science Certificate." In addition, the Middlebury faculty recently proposed the creation of a new minor in Data Science. Examples at research universities that we examined include a "Certificate in Data Fluency" at Brown University and the Data Science minor at Stanford University.

Data science curriculum without a major or minor

Many of the other liberal arts colleges that we investigated offered some elements of a data science curriculum, without yet offering a major or minor. For instance, Amherst, Grinnell, Middlebury, and Oberlin all offer introductory courses in data science. Bates has a program in Digital and Computation Studies that offers numerous courses but does not yet offer a major or minor. Similarly, Grinnell has an "interdisciplinary initiative" in data science that offers several courses but not yet a major or minor. And of course, virtually all of the schools we looked at offered at least some coursework in statistics and computer science.

Data science tracks in other majors, and other methods of promoting interdisciplinarity

One recent innovation has been the development of data science "tracks" within other majors -- new versions of majors (outside of statistics, data science, computer science, and math) that combine some of the traditional content of that major with various data science requirements. For example, the economics department at the Massachusetts Institute of Technology created a new major in "Computer Science, Economics, and Data Science," which is now significantly more popular than the traditional economics major. The University of Chicago now offers a "BA in Economics with Specialization in Data Science." Both of those majors make some of the traditional core requirements of an economics major (such as intermediate macroeconomics) optional, and require a certain amount of coursework in computer science and data science instead.

Another development has been the creation of data science majors or minors that explicitly require significant coursework in some particular domain of knowledge, and include some measures to integrate the study of data science with the study of that domain. Many of the data science majors and minors mentioned earlier are designed to complement study in other fields, and some explicitly require a set of courses in some other complementary discipline, along with some kind of integrative exercise. For example, the Data Analytics major at Denison requires students to take three or four classes from a menu of options in a particular domain, such as Anthropology and Sociology, Biology, Economics, Philosophy, Physics, Political Science, or

Psychology, and to conduct a capstone independent data analysis project on a topic in that domain (see Table 3.4 for further details).

In addition to what we have summarized above, our examination of data science initiatives includes an analysis of different pathways into data science (and, critically, their relationship to issues of equity, diversity and inclusion), different manifestations of upper level curricula, and more; see Appendix 4.

Digital humanities

Digital humanities activities are on the rise across the Williams campus. Even without a formal curriculum or campus center, we see students ambitiously pursuing digital humanities in contract majors, double majors, independent studies and honors theses. Faculty are integrating digital humanities into their classes and research. Though faculty and students rarely label their approach as "digital humanities," the computational tools and methods they seek mark a clear need. Expert staff – many of whom have promoted digital humanities at Williams for years – are eager to find ways to provide additional support that is coordinated, creative, and strategic.

Curricular structures

Williams has no formal curricular pathways for students wishing to pursue digital humanities, and no specific classes providing an introduction to the field. Two existing pathways may provide some options: contract majors and double majors. However, both options have limitations. For a contract major, since there is no set pathway, each student must find their own combination of courses, and provide justification for it. This results in too much separation between tools/methods and domain, too many extraneous courses, and too much uncoordinated advising and faculty work. A similar set of challenges affects double majors: there is no mechanism that guarantees a well-founded approach to digital humanities.

Contract majors and double majors in digital humanities lack coordinated support. There are no community-building opportunities for students, nor any co-curricular options that simultaneously are focused on digital humanities and are guaranteed to reach the appropriate audience of students. Both groups will find that expert staff support is essential, not only for learning how to use technical tools, but also for learning how those tools fit within critical inquiry for a given field. However, there is neither accounting of nor recognition of the workshops and on-to-one tutorials where support happens. Most often, students wishing to develop a course of study in digital humanities must work on their own, without coherent organizational support, to connect digital approaches and technologies to the specific framing and content of their disciplines.

Courses

At Williams, most student training in digital humanities is provided through insertions into existing classes on an ad hoc basis, often in the form of student projects. These additions can add engaging, dynamic new elements to a syllabus, but there are shortcomings. The digital humanities components are often added late in the development of a course, often in a way that is "illustrative" of digital humanities for specific content, and frequently non-critically, as a way to make an "interactive" project or as an alternative to a traditional research paper. Such courses often rely on Williams staff engagement at a level that is not sustainable as more courses add such projects. Recent student projects include examples involving GIS and text mining—applications that show student excitement and ambition, but also leaned heavily on last-minute support from staff. Because these tools are not built into the curriculum or the academic support structure, students have difficulty integrating such tools and approaches into their studies, and critically assessing their strengths and weaknesses. Digital humanities elements are too often inserted without proper skill building, appropriate methodology, or critical context, and students are not taught to evaluate the quality of digital projects in their fields.

Independent study classes are another place where digital humanities tools and approaches are taught and learned. These courses offer flexibility to marshal resources (faculty, staff, technical know-how and domain knowledge) for student learning around a given topic, much like a contract major on a small scale. Recent independent study projects by computer science students have visualized the Williams College Museum of Art's (WCMA's) collection in new ways. In the process, students learned about the texture and historical nuance of humanities data, the idiosyncrasies of digitized art, and the curatorial opportunities afforded by new perspectives on collections, as well as new web frameworks and interfaces. As they are project-based, tailored, focused, resourced, independent studies are a good model for teaching and learning digital humanities tools and approaches. The challenge is making them scale.

Honors theses also show a lively interest in digital humanities across departments. A preliminary assessment of the last two years of student theses, however, suggests that many students are working without needed training and support in digital approaches, and many are not aware of the tools and approaches that might deepen their research or help them use digital tools in methodologically appropriate ways.

Scholarship

While there is no mechanism by which we can know all scholarly efforts at Williams that relate to digital humanities, we mention a few examples of such work here. In the digital humanities, scholarship includes not only publication, but data curation, tool development, exhibitions, program design, and more.

Chad Topaz and collaborators from the Department of Mathematics, along with Ondine Chavoya from Art History and Kevin Murphy from WCMA, conducted the first large-scale analysis of the gender and ethnic diversity of artists whose works are held in major U.S. art museums. This work, covered in the national and internal press, catalyzed conversation across the museum sector and led to Williams faculty and one student participating in an invitation-only data hackathon at the National Gallery of Art. WCMA curator Horace Ballard published an article for the International Journal for Digital Art History that recounts how WCMA Digital provides not just new data and tools to engage the collection, but also suggests a new intellectual framework for curatorial practice. Professor of Anthropology Antonia Foias curated *The Seeds of Divinity*, an exhibition that used digital tools and approaches to develop and present new research on Mesoamerican objects in WCMA's collection; see interview from the Society of Cultural Anthropology. Chad Weinard, director of WCMA Digital, describes the intellectual impact of *Pink Art*, the 2017 WCMA exhibition developed in collaboration with Computer Science, in the title chapter for the anthology *Humanizing the Digital*.

Existing resources for digital humanities

Beyond the individuals who undertake teaching and scholarship related to digital humanities, and beyond the physical spaces allocated to them, support for digital humanities efforts comes from many sources on campus. These sources include the Libraries, OIT, the Visual Resources Center, and WCMA. Many Williams staff have skills in digital humanities topics and tools and are eager to integrate critical digital humanities approaches into their work with students and faculty. However, as with data science, in digital humanities, these resources are uncoordinated and largely offered on an ad hoc basis, without scaffolding and context. Seen from the perspective of a digital humanities lifecycle, the current resource structure supports the beginning (content collection/domain knowledge) and the end (presentation/production), but there is a need for structures in between that connect content to tools and approaches, and connect different parts of the digital humanities process.

Libraries

Williams Libraries offers support for digital humanities by facilitating access to digitized collections and resources (especially via department-specific liaisons), promoting open digital scholarship through initiatives like Unbound (a collaboration with OIT), and through outreach programs. Librarians support digital humanities in many of the ways outlined above for data science.

Visual Resources Center

The Visual Resources Center supports digital image collections, presentation, and tools, with particular expertise in art and art history.

Office for Information Technology

OIT staff support digital humanities through the application of technology in the same ways as mentioned earlier for data science.

Williams College Museum of Art

WCMA supports digital humanities activities on campus through WCMA Digital, a grant-funded initiative that seeks to spark new kinds of engagement with the collection, and with the humanities more broadly, through digital tools and approaches. WCMA Digital currently supports a full-time director, a part-time developer, and a contract designer (through spring 2020). A WCMA-funded digital humanities postdoctoral fellow, tasked with faculty support, is on staff for a two-year term (until June 2021). WCMA Digital focuses on three areas of practice:

- *Data.* The project develops, maintains and supports the use of WCMA's collection data and images. WCMA's collections provide for a living, evolving data set. The museum also offers a unique, complete set of digital images. Both resources open new conversations and collaborations, connecting departments like computer science, math, statistics and biology to art and humanities.
- *Support for teaching, learning and research.* WCMA excels at helping faculty integrate art into their courses; WCMA Digital adds a digital dimension to this effort. A Digital Humanities Postdoctoral Fellow is tasked with helping faculty (and students) make use of WCMA's digital collection, and integrate digital tools and approaches more broadly.

- *Tools.* The project designs, develops and supports digital tools that help students, faculty, curators and museum staff visualize and use the collection. Technical expertise, making and building and important parts of digital practice.

As part of the art museum, WCMA Digital benefits from WCMA's position on campus as a connector, convener and collaborator. The project has been able to reach a broader public by generating exhibitions (*Accession Number*, *Pink Art, All At Once*) that explore art, data, and digital culture. Proximity to the arts (and artists/makers), along with a focus on teaching and learning, make WCMA Digital unique among digital humanities programs. WCMA Digital also enjoys support from WCMA's grants and development team and communications staff. These are key elements to consider for the success of future initiatives. This program is grant-funded and term-limited, ending in June 2020.

Institutional comparisons

While digital humanities centers came to prominence at large research universities in the 1990s, programs at small liberal arts colleges are more recent and are distinctive in their focus on pedagogy, ethics and social context, and creative making. The Digital Scholarship Lab at the University of Richmond focuses on building and maintaining digital tools (much like a university digital humanities center), but those tools are meant to engage undergraduate students. The Digital Humanities Initiative at Hamilton College and the Center for Digital Liberal Arts at Occidental College have an explicit pedagogical focus in their missions. At Occidental, the program began as part of a larger vision to transform the college for the twenty-first century.

Other programs include the Grinnell College Digital Liberal Arts Collaborative, the Grinnell College Data Analysis and Social Inquiry Lab, the Five College Digital Humanities Program, and programs entitled "Digital Humanities" or "Digital Liberal Arts" at the Five Colleges, Carleton, Middlebury, Macalester, Mount Holyoke, and Kenyon. Spelman College is another institution to watch, given their initiative for a Center for Innovation and the Arts, which draws together art, art history, technology and innovation.

For additional context, see "A Survey of Digital Humanities Centers in the United States" and "Should Liberal Arts Campuses Do Digital Humanities? Process and Products in the Small College World".

Strategies

Data science

Curriculum

The cornerstone of Williams' efforts in data science should be a forward-looking curriculum that serves students well. In the following discussion, we keep in mind three groups of students as defined in the Harvard Data Science Review:

Group I: "Students who do not aim to acquire deep expertise but need to gain a basic understanding of data science, whether they recognize it or not, and for whom traditional introductory statistics courses are an incomplete solution… Students may be motivated to [learn] out of simple curiosity about this important part of their world, a desire to interpret information better, or as career preparation."

Group II: "Students who are intrigued by the power of data science as a tool that can lead to advances in their field of primary interest, which might be physics, chemistry, medicine, public health, climate science, political science, sociology, economics, or many others, including the humanities. Theirs is often a bridging role."

Group III: "Students who intend to become full-fledged experts by specializing in data science, with ambitions to contribute to advances in the field through their teaching or research, whether in universities or within companies, government, or nonprofit organizations."

While there are many different ways one could design a data science curriculum, with significant pros and cons to each (see Appendix 3 and Appendix 4), here is one possible way forward.

Introduction to data science sequence

Group I could be served by a two-course sequence, Introduction to Data Science, which could be offered every year. This sequence would be targeted at students with little or no prior experience with computing or statistics. The course would expose students to different data types and sources, and to the process of transforming data to a format suitable for analysis. It would also introduce students to elementary notions of visualization, estimation, modeling, prediction and inference. We envision that this class would be taught in part through case studies, involving a progressive series of more to less-manicured data designed to enhance students' computational and analytical abilities. A student completing the two-course sequence in Introductory Data Science would be able to:

- Discuss the digitization of society from historical, philosophical, and other perspectives
- Acquire, wrangle, and explore data coming from a range of source types
- Evaluate the legitimacy of data
- Analyze how a data collection process influences the scope of inference
- Summarize, visualize, and analyze data
- Perform elementary statistical estimation, prediction, inference, and modeling
- Understand elementary algorithms and computer coding concepts

- Apply frameworks from data ethics to case studies and projects
- Communicate results in graphical, oral, and written formats

Besides providing an introduction to data science (including its relation to society), this sequence would provide an integrated introduction to computer science and statistics with no prerequisites. After completing it, a student would be able to take CSCI 136 (Data Structures) and either STAT 202 (Introduction to Statistical Modeling) or ECON 255 (Econometrics). The existence of this course sequence may have the effect of reducing enrollment pressures on introductory courses in CSCI and STAT.

Data science concentration

To provide a suitable option for students in Group II, we envision a concentration in Data Science that would bring together the three pillars of the field. A possible pathway might include:

- Introduction to Data Science I and II (see above)
- One of STAT 202 (Introduction to Statistical Modeling), STAT 302 (Applied Statistical Modeling), STAT 346 (Regression and Forecasting), or ECON 255 (Econometrics)
- Data Structures and Advanced Programming (CSCI 136)
- CSCI 374 (Machine Learning) or a (new) 300-level STAT course in machine learning that requires fewer prerequisites
- One additional mathematics, statistics, computer science, or GIS class at the 200 or 300 level that is relevant to data science, as approved by the director of the concentration
- Capstone experience: a project in the domain of the student's major (or one of their majors)

This set of courses is smaller than that needed for a major, and would provide a coherent experience to students wanting to focus on applications of data science tools to domain areas. While we cannot know *a priori*, it is possible that the availability of a concentration will have the effect of reducing the number of students who double major, and these students would in turn benefit from the freedom to further explore the Williams curriculum.

Data science major

For students in Group III, a fully-fledged interdisciplinary major in Data Science may be appropriate. This major would be focused on the methods of data science rather than on applications. Students in Group III may currently be majors or double majors in CSCI, MATH, or STAT. Our proposed curriculum would streamline their studies. Note, however, that due to the double counting rule, double majoring in Data Science and CSCI/MATH/STAT would be challenging. A possible structure of such a major would be:

- Core: Introduction to Data Science I and II
- Mathematical Skills: Multivariable Calculus (MATH 150/151), Discrete Mathematics (MATH 200), and Linear Algebra (MATH 250)
- Statistical Modeling: One of STAT 202, STAT 302, STAT 346 or ECON 255
- Data Structures and Advanced Programming (CSCI 136)
- Algorithm Design and Analysis (CSCI 256)

- Capstone: Machine Learning (CSCI 374) or Statistical Learning and Data Mining (STAT 442) or another pre-approved elective in CSCI or STAT
- Elective in a domain area of application (pre-approved)

Co-curriculum and institutional structure

For data science to flourish at Williams, we recommend that the college investigate co-curricular and supporting structures that would increase capacity, build community, and maintain quality.

To support students, we recommend the following:

- Development and maintenance of recommended computer configurations for students pursuing data science, and continued involvement of Financial Aid to ensure that all students can afford to own a computer;
- Cultivation of in-house expertise on data science in the 68 Center for Career Exploration;
- Identification of off-campus pro bono data science activities for students through the Center for Learning in Action;
- Dissemination of opportunities for participation in data hackathons and similar events;
- Creation of a curated, centralized online listing of student opportunities related to data science including on-campus research, off-campus internships, jobs, and volunteer work, special events, courses, and more.

To develop staff and faculty capacity, we recommend the following:

- Deployment of a data science boot camp for staff and social sciences/humanities boot camp to bridge gaps between divisions; these could be modeled in part after the faculty data science boot camp taught by Chad Topaz during January, 2020;
- Establishment of opportunities for staff and faculty to apply for other professional development related to data science.

For community building, we recommend the following:

- A regularly-scheduled data science speaker series;
- A program of visiting scholars/practitioners in data science who would interact with students, staff, and faculty;
- A weekly data science lunch or tea hour for informal exchange and networking;
- Space on campus that serves as a home for data science.

To ensure a healthy and viable data science initiative, we recommend the following:

- Establishment of a tightly-structured steering committee that includes key stakeholders from on campus as well as selected scholars and practitioners from off campus.

Resource considerations

While the college already has a strong foundation in data science and digital humanities, it will not be possible to create a vibrant program in these areas without some additional resources.

While the complete resource needs will have to be more fully explored during the operational phase of planning, we do have some thoughts about some likely areas of focus.

Faculty and staff human resources

One of the challenges in building a program in data science and digital humanities is that we are already experiencing high demand pressures in some of the fields that would contribute most directly to the core classes. We thought creatively about ways to support a new curriculum through a reconfiguration of existing faculty, but we were unable to identify a pathway that did not involve at least some additional hiring in cognate disciplines. The most immediate needs in this regard would be the Introduction to Data Science sequence and electives for an eventual concentration and a major (the creation of which would require faculty votes). Further needs could be met through strategic hiring in domain areas and GIS, which could happen as part of the ordinary course of hiring in those fields.

While the faculty resources would be essential to the creation of a successful program, there may be additional staff needs as well. One area of future support is in the area of data ethics. We want to make sure that our students, faculty, and staff approach their data work ethically and practice good data governance. While some of this work is already being done by our institutional research office, we would like data ethics to be infused into all aspects of the data science/digital humanities curriculum and program. This could be achieved through a dedicated director of data ethics, or perhaps through a more distributed model of faculty and staff support. We can also imagine increased needs for practical support in the areas of coding and statistics, professional development in OIT and the libraries, as well as lab instruction for GIS. Finally, there may also be opportunities to increase the capacity of the '68 Center for Career Exploration to support students seeking data science careers.

Most successful programs also have a dedicated faculty leader, as well as staff support. While there are many models that could work at Williams, we strongly recommend that the college think carefully about a leadership structure that will support the program in its early phases and provide administrative continuity and support throughout its existence.

We recognize that faculty and staff are the most precious resources the college has and that it must be prudent in considering any increase in college FTE. Nevertheless, we believe that a successful data science program will require at least some additional faculty and staff resources. The environment for hiring staff and faculty in areas related to data science is incredibly challenging. We recommend that the college adopt a creative and flexible approach to hiring that would give us the greatest chance of attracting talented and diverse candidates from a wide range of backgrounds.

Infrastructure

The college may want to invest in additional classroom space (through renovation or new construction) that would support learning environments that are highly computational and highly collaborative. Few appropriate spaces exist on campus at the present time. Further, to the extent that the program attracts additional faculty and staff resources, there may be needs for additional office and lab space. Finally, there would almost surely be needs for computational resources, including access to new computing environments and software, enhancement of computational horsepower, web server space, web storage space for big data, data licensing, and more.

Other programmatic budget

Finally, we have recommended numerous programs and sub-initiatives, which would come with funding requirements of their own. Examples include speaker series, weekly lunches/teas, visiting scholars/practitioners, and professional development opportunities for staff and faculty.

Digital humanities

Curriculum

A digital humanities curriculum at Williams would allow humanities students to apply a wide range of digital tools and approaches to their chosen discipline. Humanities students would gain data literacy, technical capability, and skills to critique the utility, appropriateness and opportunities afforded by digital tools and approaches for their field.

A digital humanities concentration would allow students to add a new digital dimension to their course of study, without requiring a second major far outside their focus. An introductory course (or two) will give students a foundation in data literacy, computational analysis, and other tools; such knowledge and skills can be applied to a domain over the course of several electives; and is brought to fruition through a capstone project. (For background, see "Digital Humanities Pedagogy as Essential Liberal Education: A Framework for Curriculum Development.")

A possible structure for this concentration is:

- Introductory course(s) teaching data literacy, computational analysis, digital literacy (creating digital scholarship and media), information literacy (finding, using and evaluating digital primary sources), presentation/interpretation (modeling, reconstruction, simulation type topics), ethics, and social context;
- Electives courses in
  - Domain areas that are adapted or created to use digital tools and approaches as primary modes of inquiry
  - Computer Science that are adapted or created to emphasize arts and humanities approaches to data, images, creative processes
  - Ethics, history, and social science that provide social context for digital tools, approaches, and culture
- Capstone experience in which students carry out a substantial digital humanities project within a given domain, developed with faculty/staff advisors

The introductory course/s is/are an opportunity to provide an accessible foundation for digital humanities study across disciplines. Ideally, the course(s) would have no prerequisites and would be welcoming to students with a wide range of backgrounds. It would develop quantitative skills through data literacy, using real-world data sets and project-based work. The course would introduce computational thinking; students would not learn programming per se, but would be able to tell what code does, and adapt open-source tools that use Javascript, Python, R. In digital humanities, many beginner-level tools are opaque in function; students need to learn to make critical assessments of what the tools do, when to use them, and what their limits are. Critically, students will understand the role of visualization or modeling processes and tools as interpretation—creative, biased, powerful—not just as direct illustrations of reality. Though the course can only offer an introduction, it can set the stage for new kinds of deep engagement with the liberal arts. We imagine this course could be, in fact, the same course as the Introduction to Data Science Course (perhaps cross-listed, or with an alternative name) described previously.

The electives and capstone experience would provide opportunities to put new digital tools and approaches to work in the context of domain knowledge. Some classes at Williams could already serve as digital humanities electives. In many cases, however, developing digital humanities electives within various humanities disciplines will require additional efforts in faculty development and staff support. Such efforts would also be crucial to making capstone experiences successful, as advisors and mentors would need to be equipped and their efforts coordinated. The work of advising, supporting, and collaborating is already in demand for independent study projects and contract majors related to digital humanities.

Co-curriculum and institutional structure

While curriculum is important, the cornerstone of digital humanities lies outside curricular structures, in dynamic interdisciplinary programs, community-building, and collaborative partnerships that support teaching, learning and research.

We propose a digital humanities center that adds a digital dimension to Williams' vision of the liberal arts. Data literacy would be a foundation, which would add rigor and social context to the gathering, handling and analysis of data. Presentation tools would get critical attention as well, as visualization is a form of interpretation. Most importantly, we recommend a holistic approach, where content would be connected to presentation through thoughtful analysis and critical thinking. The center would directly support faculty, students and staff as they explore new approaches to teaching, learning and research. Digital humanities is marked by project-based collaboration, inclusive interdisciplinarity, and a fruitful pairing of making and critical thinking.

More specifically, a digital humanities center should:

Build community. Convene groups by hosting lectures, symposia, artist projects and exhibitions that bring new ideas and people to campus to explore digital humanities, and technology at the intersection of arts and culture. Facilitate open-access scholarship and public impact by getting Williams' digital humanities projects and ideas out into the world. Connect units and individuals by hosting meet-ups for staff (and faculty) with technical skills to share in-progress work and research interests. Coordinate campus resources by serving as a hub for resources, people, projects, and events.

Support teaching, learning, and research. Support faculty through group workshops, individual coaching, and other means in order to provide skills and approaches to teach with digital tools and advise students. Support students through workshops, individual coaching, and guidance with class projects and independent study projects, as well as by connecting students interested in digital humanities with people, resources, and pathways that fit their interests. Support staff who utilize digital humanities skills and tools through individual coaching and professional development.

Make and maintain. Synergize the “maker” impulse that underlies much digital humanities work with the pervasive influence of the arts on the Williams campus via initiatives in creative coding, immersive installations, and so forth. Act as project partners, prototyping and/or managing digital projects, writing technical and design briefs for grants, and developing and maintaining data resources. Provide a robust outward-facing presence by developing grant programs or other funding opportunities, and generate publicity as Williams’ impact on digital culture grows.

We suggest that WCMA, the libraries, and OIT are the institutional hub for digital humanities at Williams. At most colleges, a digital humanities center might be in the library or in the information technology department. For Williams, adding the art museum is critical. WCMA is an established connector, convener and collaborator on campus. It is a site of public interchange, with exhibitions and access to publicity and grants apparatus. The museum excels at partnering with faculty to enhance teaching and learning with art and technology. The museum also has its own digital needs and opportunities, providing an ongoing source of digital projects (including a data set, image set and visualization tools) for a technical team and for collaboration with faculty and students.. Most importantly, WCMA Digital has already provided a pilot for a digital humanities center. Programs, projects, staff, partners, and potential funders are in place, poised for ongoing institutional commitment.

Resource considerations

To put the recommendations made above into practice, there will inevitably be some resource implications.

Faculty and staff human resources

To build on the work of WCMA Digital, a digital humanities center would benefit from:

- A full-time director and a postdoctoral fellow;
- A developer (part-time or contractor)
- Faculty resources to teach digital humanities introductory courses (potentially jointly with introductory data science courses);
- A formal collaborative working group with Libraries, OIT, and faculty advisors.

We are aware that not all of these needs can be met in the earliest years of the program, but we wanted to convey a sense of what the steady state of the program might look like. The most immediate need would be a full-time director (with experience as a developer/designer) who could perform many of the critical functions listed above in the early years of the program.

Infrastructure

The digital humanities center need not be a physical center, but there would still be a need for offices, studio spaces for meeting and making, and space for community gatherings. Existing spaces may be appropriate; proposed new spaces like a new WCMA building or a repurposed Lawrence Hall offer opportunities for the future.

Other programmatic considerations

Finally, digital humanities resources should include support for programs such as speaker series, weekly lunches/teas, visiting scholars/practitioners, and professional development opportunities for staff and faculty.

Open questions

The ideas above sketch out a possible future for data science at Williams, but many open questions remain, and must be carefully thought through in the next phase of planning.

- Are “data science” and “digital humanities” good choices for the names of these initiatives? These terms have the advantage of being recognizable and conveying commonly-understood meaning outside of our institution. At the same time, within our institution, they may be misconstrued as referring only to the natural sciences (for the former) and only to the humanities (for the latter). If these are not good naming conventions, what should the names of the initiatives be instead?
- How should data science efforts be organized, administratively speaking? For instance, should they be centralized in a department or interdisciplinary institute?
- Similarly, how should these efforts be organized spatially? We have recommended dedicated space, but neither the amount of space nor its distribution on campus have been specified.
- Is the particular approach to designing a data science introductory sequence, concentration, and major outlined above a good one that satisfies the needs and constraints of the college?
- How will new curricular offerings in data science articulate with current offerings in mathematics, statistics, and computer science?
- How should data science and digital humanities efforts be assessed?
- What are the potential synergies between data science and digital humanities efforts, including at the curricular level?
- More specifically, are the goals of the introductory data science courses and introductory digital humanities courses close enough that they could be the same courses?
- What are the potential synergies between this strategic initiative and the strategic initiative in Science, Technology, and Society?

Appendix 1: Working group charge

Many areas of modern society including politics, the arts, health, science, and commerce, increasingly rely on the large quantities of data enabled by a revolution in computing over the past 30 years. The field needed to support the analysis of and critical thinking about such data is called data science. Data science, however, does not exist in a vacuum. During a time of unparalleled access to data and computational tools, it is crucial to communicate knowledge effectively, to be mindful of how knowledge is produced, and to be aware of the history, ethics, and social impact of digital work. The working group on data science, digital scholarship, and science and technology studies will focus on strategies to:

- Equip students, staff, and faculty across the arts, humanities, social sciences, and natural sciences with digital and data-driven approaches
- Enhance digital and data science approaches with domain knowledge and humanities-driven analysis
- Address pressing contemporary challenges through the use and critique of digital and data-driven approaches
- Facilitate the coherent, clear, and credible use of data and digital technology in scholarship and communication
- Promote the ethical use of technology and data

This working group will consider the following questions:

- How can we create a research environment that supports faculty and student work involving data science and digital scholarship?
- What are the appropriate curricular structures to train students in these areas?
- What can Williams learn from other institutions that have undertaken initiatives in data science and digital scholarship?
- What resources are necessary to undertake new initiatives?
- How will the college assess its level of success in the areas of data science and digital scholarship?

Appendix 2: Methods and outreach

Members of our working group:

- Met 14 times from June, 2019 through January, 2020
- Conducted research on data science and digital humanities programs across the country
- Conducted a focus group with students in early September, 2019
- Participated in the campus-wide Strategic Planning Day in October, 2019
- Spoke with the President's Administrative Group in December, 2019
- Hosted a visit from external data science expert, Prof. Brandeis Marshall (Spelman College) in January, 2020

Appendix 3: Data science curricula at other institutions

Table 3.1 -- Examples of introductory data science courses at other schools

School	Course Title	Notes
Amherst	Data Science	Introductory statistics and introductory computer science are prerequisites
Berkeley	Data 8: Foundations of Data Science	No prerequisites; substitutes for introductory statistics in requirements for data science major
Bowdoin	Introduction to Digital and Computational Studies or Data Driven Societies	No prerequisites; both cover some statistics and computer programming; either can serve as the introductory course for Digital and Computational Studies Major, which does not require any other statistics or computer science
Brown	Data Science Fluency	No prerequisites. Required for "Certificate in Data Fluency" along with intro computer science, a course in "Data, Ethics, and Society", one elective, and one experiential learning component
Claremont-McKenna	Foundations of Data Science	Introductory computer programming course is a prerequisite; complement to introductory statistics; required for Data Science major, along with intro statistics and intro computer science
Connecticut College	Thematic Inquiry	Course on data analysis, data ethics, and data communication. No prerequisites; complement to introductory statistics; required for Data Information and Society "pathway," along with introductory statistics
Davidson	Data Science and Society	No prerequisites; not open to students with any prior statistics or computer science coursework; complement to introductory statistics; can serve as an elective for data science minor, whereas intro computer science and intro statistics are required for the minor
Denison	Introduction to Data Analytics	No prerequisites; complement to introductory statistics; required for Data Analytics major along with intro stats and intro computer science
Grinnell	Introduction to Data Science	Introductory statistics is a prerequisite, prior computer science course is recommended

Table 3.1 (continued) -- Examples of introductory data science courses at other schools

School	Course Title	Notes
Lafayette	Principles of Data Science	Both introductory statistics and introductory computer science are prerequisites; required for data science minor
Macalester	Introduction to Data Science	No prerequisite; complement to introductory statistics; not required for data science minor (which does require both intro stats and intro computer science), but can optionally substitute for intermediate computer science requirement
Middlebury	Introduction to Data Science	Introductory statistics is a prerequisite
Oberlin	Introduction to Data Science	No prerequisites
Smith	Introduction to Data Science	No prerequisites; complement to introductory statistics; required for Data Science major along with intro stats and intro computer science
Stanford	Data Science	No prerequisites; complement to introductory statistics; required for Data Science minor along with intro stats and intro computer science
Union	Data Analytics	Introductory computer science is a prerequisite; complement to introductory statistics; required for Data Analytics minor along with intro stats and intro computer science
Wesleyan	Digging the Digital Era: A Data Science Primer	No prerequisites; substitute for introductory statistics; this is one among many ways to satisfy intro stats requirement for Data Science minor or certificate, which also requires intro computer science
Yale	YData: An Introduction to Data Science	No prerequisites; can substitute for introductory statistics in requirements for Data Science major, but is one among multiple options; based on Berkeley's Data 8

Table 3.2 -- Examples of interesting topics courses with a data science emphasis, requiring little or no prior background

School	Course	Notes
Harvard	ECON 1152 – Using Big Data to Solve Economic and Social Problems	No prerequisites. See Vox article on this course here.
Grinnell	HIS 100-04 - Introduction to Historical Inquiry: Digital Histories	
Berkeley	LEGALST 88 - Crime and Punishment: Taking the Measure of the US Justice System	Examples of "connector courses" to be taken concurrently with Berkeley's Data 8 – Foundations of Data Science
Berkeley	L&S 88 - Child Development Around the World: Analyzing Household Data Sets	
Berkeley	ESPM 88B – Data Sciences in Ecology and the Environment	
Berkeley	L&S 88 - Rediscovering Texts as Data	
Berkeley	HIST 88 – How Does History Count? Exploring Japanese-American Internment through Digital Sources	
Berkeley	L&S 88 – Data Science for Cognitive Neuroscience	

Table 3.3 -- Berkeley Data Science Major

Introductory statistics / data science requirement	Data 8: Foundations of Data Science (has no prerequisites, satisfies statistics prerequisite for other majors)	Students can take "connector courses" concurrently with Data 8 (which is facilitated by Berkeley's flexible credit hour system). These courses apply data science tools to particular domains. Examples include "Economic Models, "Exploring Geospatial Data," "Data Science for Cognitive Neuroscience," "Literature and Data," etc.
Introductory computer science requirement	One "Program Structures" Course, such as CS 61A: Structure and Interpretation of Computer Programs or CS 88: Computational Structures in Data Science	
Intermediate statistics and data science requirement	Data 100: Principles and Techniques of Data Science	
Intermediate computer science requirement	CS 61B: Data Structures	
Math requirements	Calculus 1 and 2, Linear Algebra, Probability	
Upper-level data science requirements	1 advanced data science course (e.g. machine learning)	
Electives	2 electives in "Computational and Inferential Depth" (e.g., "Efficient Algorithms and Intractable Problems," Econometrics, etc.)	
Domain course requirement	3 courses from a particular domain	
Human context and ethics requirement	Take one course from a list of options (e.g. Philosophy course on "Moral Questions of Data Science," or course on "Theory and Method in the Digital Humanities")	

Table 3.4 -- Dension Data Analytics Major

Introductory data science requirement	DA 101 - Introduction to Data Analytics (has no prerequisites)
Introductory computer science requirement	CS 111 – Discovering Computer Science: Scientific Data and Dynamics; or CS 112 Discovering Computer Science: Markets, Polls, and Social Networks
Introductory statistics requirements	DA/MATH 220 – Applied Statistics
Intermediate computer science requirement	DA 210/CS 181 – Data Systems
Math requirements	MATH 135 – Single Variable Calculus; or MATH 145 Multi-variable Calculus
Practicum requirement	DA 301 Practicum in Data Analytics (involves a semester-long group data analysis project fall of junior year)
Upper-level data science requirements	DA 350 Advanced Methods for Data Analytics (includes, for example, machine learning)
Summer experience requirement	"…students must complete a DA summer experience (internship or research project). This experience must be approved by the Data Analytics Program Committee, and is normally undertaken during the summer before the senior year."
Colloquium requirement	DA 200 - Data Analytics Colloquium (required once as a sophomore and once as a junior or senior) – partial credit course that involves attending talks by guest speakers
Capstone	DA 401 - Seminar in Data Analytics (involves an individual independent data analysis research project, "applied to a problem in their chosen domain" – see below for examples of domains)
Domain course requirement: students must take the designated courses in one of the following domains:	Anthropology and Sociology (3 courses) • ANSO 100 - People, Culture and Society • ANSO 343 - Demography of Africa; or ANSO 347 Power in Society • ANSO 351 - Survey Research Methods

Table 3.4 (continued) -- Dension Data Analytics Major

Domain course requirement (continued): students must take the designated courses in <u>one</u> of the following domains:	<u>Biology (4 courses)</u> ● BIOL 210 - Molecular Biology and Unicellular Life ● BIOL 220 - Multicellular Life ● BIOL 230 - Ecology and Evolution and one of the following: ○ BIOL 309 - Computational Biology ○ BIOL 345 - Eukaryotic Cell Biology ○ BIOL 356 - Special Topics (Biostatistics) <u>Economics (4 courses)</u> ● ECON 101 - Introductory Macroeconomics ● ECON 102 - Introductory Microeconomics ● ECON 302 - Intermediate Microeconomic Analysis ● ECON 467 - Econometrics II <u>Philosophy (3 courses)</u> ● PHIL 121 - Ethics: Philosophical Considerations of Morality; or PHIL 126 - Social and Political Philosophy ● PHIL 205 – Logic ● PHIL 210 - Philosophy of Science <u>Physics (3 courses)</u> ● Either PHYS 121 – General Physics I & PHYS 122 General Physics II; or PHYS 125 - Principles of Physics I: Quarks to Cosmos & PHYS 126 - Principles of Physics II & PHYS 127 - Principles of Physics III ● PHYS 312 - Experimental Physics <u>Political Science (3 courses)</u> ● POSC 201 - Analyzing Politics and any two of the following: ○ POSC 213 - Doing Political Science: American Political Behavior ○ POSC 307 - The Politics of Congress ○ POSC 309 - Campaigns and Elections ○ POSC 311 - Political Organizations in the U.S <u>Psychology (3 courses)</u> ● PSYC 100 - Introduction to Psychology ● PSYC 200 - Research Methods and Statistics ● PSYC 2XX/3XX - Psychology elective (except research courses, 370, 410, 361-364, 451-452)

For further information on the Denison Data Analytics program, see: Jessen Havill. 2019. “Embracing the Liberal Arts in an Interdisciplinary Data Analytics Program.” *SIGCSE '19: Proceedings of the 50th ACM Technical Symposium on Computer Science Education*. February, pp. 9-14. <https://doi.org/10.1145/3287324.3287436>.

Table 3.5 -- Macalester Data Science Minor

Introductory statistics requirement	STAT 155 - Introduction to Statistical Modeling
Introductory computer science requirement	CSCI 123 – Core Concepts in Computer Science
Introductory data science or intermediate / advanced computer science requirement	One of: STAT 112 – Introduction to Data Science (no prerequisites); or CSCI 127 – Object Oriented Programming and Abstractions; or CSCI 128 – Data Structures; or CSCI 302 – Introduction to Database Management Systems; or CSCI 365 – Computational Linear Algebra; or CSCI 440 – Collective Intelligence; or CSCI 484 – Introduction to Artificial Intelligence. Some of these courses have other prerequisites, e.g., CSCI 484 requires CSCI 221 – Algorithm Design and Analysis
Intermediate statistics requirement	One of: STAT 253 – Statistical Machine Learning; or STAT 453 - Survival Analysis; or STAT 454 – Bayesian Statistics
Math requirements	Minor can be completed without any additional math courses, but some of the courses above have additional math prerequisites (e.g., linear algebra)
Domain course requirement	"Two courses in a single domain meeting the following criteria: the courses should be focused around a theme (e.g., bioinformatics) and not a broad discipline AND that theme should [be] one that allows the possibility of data science-related activities, either within the courses themselves or in the student's future career or further education." Examples include Econometrics, Geosciences course on GIS, etc., list available here.

Table 3.6 -- Mount Holyoke Data Science Major

Introductory statistics requirement	STAT-140 Introduction to the Ideas and Applications of Statistics

Introductory computer science requirement	COMSC-151 Introduction to Computational Problem Solving
Intermediate statistics requirements	STAT-242 Intermediate Statistics AND STAT-340 Applied Regression Methods
Intermediate computer science requirement	COMSC-205 Data Structures
Math requirements	Calculus 1, Calculus 2, Linear Algebra
Upper-level data science requirement	COMSC-335 Machine Learning
Electives	None
Domain course requirement	“Two courses at the 200 level or above within a single domain area; A domain area -- chosen in consultation with the student's Data Science advisor -- is defined as any College-defined major excluding mathematics, statistics, and computer science. Course selection must be approved by the student's Data Science advisor.”
Ethics or social implications requirement	“The study of ethics in relation to data science is integrated throughout the curriculum and emphasized in this integrative capstone course.”
Capstone	DATA-390 Data Science Capstone

Table 3.7 -- Mount Holyoke Nexus in Data Science

Statistics requirement	One statistics course at 200 level or higher from approved list of courses
Computer science requirement	One computer science course at the 200 level or higher from approved list of courses
Domain course requirement	One course from an application area (e.g., biology, economics, English, psychology, sociology) at the 200 level or higher from approved list of courses
Electives	One "elective course that demonstrates an interest in data science…"
Internship	
Required course related to internship	COLL-211 Reflecting Back: Connecting Internship and Research to Your Liberal Arts Education

At Mount Holyoke, a "Nexus" is an interdisciplinary concentration or minor, with an internship component.

Table 3.8 -- Smith College Statistical and Data Science Major

Introductory statistics requirement	SDS 201 - Statistical Methods for Undergraduate Research or SDS 220 - Introduction to Probability and Statistics
Introductory computer science requirement	CSC 111 - Introduction to Computer Science Through Programming
Introductory data science requirement	SDS 192: Introduction to Data Science (no prerequisites)
Intermediate statistics requirement	MTH/SDS 291 – Multiple Regression
Math requirement	MTH 211 – Linear Algebra
Computing depth requirement	Choose one of the following: CSC 151: Programming Languages CSC 212: Data Structures CSC 220: Advanced Programming Techniques CSC/SDS 235: Visual Analytics—must take programming intensive track CSC 294: Computational Machine Learning CSC/SDS 352: Parallel & Distributed Computing
Statistics depth requirement	Choose one of the following: MTH/SDS 290: Research Design and Analysis SDS 293: Machine Learning MTH/SDS 320: Mathematical Statistics SDS 390: Topics in SDS (Categorical Data Analysis, Structural Equation Modeling, or Statistical Analysis of Social Networks)
Communication requirement	Choose one of the following: CSC/SDS 109: Communicating with Data CSC/SDS 235: Visual Analytics SDS 236: Data Journalism

Table 3.8 (continued) -- Smith College Statistical and Data Science Major

Domain course requirement	"Every student is required to take a course that allows them to conduct a substantial data analysis project evaluated by an expert in a specific domain of application." Examples: An honors thesis (not in statistics or comp. sci.) SDS 300: Applications of Statistical & Data Sciences PSY 301: Research Design and Analysis PSY 358: Research Seminar in Clinical Psychology ECO 311: Seminar: Topics in Economic Development ECO 363: Seminar: Inequality EGR 389: Data Mining BIO 334: Bioinformatics & Comparative Mol Bio NSC 318: Neurobiology
Capstone	SDS 410: Capstone

The Smith College Statistical and Data Science Minor requires the following six courses from the list of major requirements above: introductory statistics; Introductory computer science; introductory data science; intermediate statistics; computing depth; and communication.

Table 3.9 -- Wesleyan Data Analysis Minor

One introductory statistics or data science requirement	One of MATH132 - Elementary Statistics; or PHYS/QAC221 - Modeling and Data Analysis: From Molecules to Markets; or PSYC200 Statistics: An Activity-Based Approach; or QAC201 - Applied Data Analysis; or QAC211 Digging the Digital Era: A Data Science Primer; or QAC250 - An Introduction to Data Journalism
Mathematical, statistical, and computing foundation courses: choose **two** courses from the following list, each from a different group	Group 1: Mathematical Foundations MATH221 - Vectors and Matrices MATH223 - Linear Algebra MATH228 - Discrete Mathematics MATH274 - Graph Theory Group 2: Statistical Foundations ECON300 - Quantitative Methods in Economics GOVT367/QAC302 - Political Science by the Numbers MATH231 - An Introduction to Probability MATH232 - Mathematical Statistics Group 3 – Computing Foundations BIOL265 - Bioinformatics Programming CSCI112 - Introduction to Programming CSCI115 - How to Design Programs CSCI211 - Computer Science I CSCI212 - Computer Science II
Two applied electives	Choose two courses from an extensive list, including courses such as "Introduction to GIS," "Economics of Big Data," etc. Complete list available here.

Table 3.10 -- Wesleyan Applied Data Science Certificate
Same requirements as the Data Analysis Minor shown in table 2.9 above, plus the following requirements:

Two applied data science courses selected from the following list of three options:	QAC305 - Exploratory Data Analysis and Pattern Discovery QAC385 - Applications of Machine Learning in Data Analysis QAC386 - Quantitative Textual Analysis: Introduction to Text Mining
One capstone "practicum"	"The capstone Data Analysis Practicum that includes an ethics and epistemology seminar discussion as well as completing an independent data science project."

Table 3.11 -- Park City Math Institute 2016 Proposed Curriculum for Data Science Major

Requirement	Courses	Topics covered
Introduction to data science (2 courses)	Introduction to Data Science I Introduction to Data Science II	Introduction to high-level language; exploring and manipulating data; functions and basic coding; introduction to modeling, both deterministic and stochastic; concepts of projects and code management; databases; introduction to data collection and statistical inference
Mathematical foundations (2 courses)	Mathematics for Data Science I Mathematics for Data Science II	Elements of calculus, linear algebra, and probability most relevant to data science
Computational thinking (2 courses)	Algorithms and Software Foundations Data Curation – Databases and Data Management	Algorithm design; programming concepts and data structures; tools and environments; scaling for big data; principles of data management
Statistical thinking (2 courses)	Introduction to Statistical Models Statistical and Machine Learning	Standard topics from introductory and intermediate statistics; machine learning
Course in an outside discipline (1 course)	One course in some other domain that is relevant to data science	
Capstone course	Data in Context – Capstone Experience	A capstone experience in which students consider scientific questions, collect and analyze data, and communicate the results
Source: Richard D. De Veaux, *et al.* 2017. "Curriculum Guidelines for Undergraduate Programs in Data Science." *Annual Review of Statistics and Its Application*. Vol. 4 (March), pp. 15-30. <https://doi.org/10.1146/annurev-statistics-060116-053930>.		

Appendix 4: Further description and appraisal of data science at other institutions

Pathways into data science

One of the big questions in the design of data science curricula is how best to construct pathways into the study of data science. The approach that requires the least change relative to existing course offerings at many schools would be to start the study of data science with a standard introductory course in statistics (such as STAT 161 or STAT 201 at Williams) and a standard introductory course in computer science (such as CSCI 134 at Williams). However, some leading experts on the development of data science curricula, including De Veaux *et al* (2017), Danyluk *et al* (2019), and Lue (2019), have argued that this is not the optimal approach. Rather, these authors argue that schools ought to develop dedicated introductory-level data science gateway courses that convey what is distinctive about data science (relative to statistics and computer science), and include elements of data science that are often missing or given little emphasis in introductory statistics and computer science courses, such as data curation and management, data ethics, working on original data-based projects in many different domains of knowledge that are of interest to a variety of students, and so forth.

Developing a new data science curriculum offers a great opportunity to attract a more diverse set of students into STEM fields, and also to get a more diverse set of students using data science tools in other fields of study in the social sciences, humanities, and arts. As Lue (2019, p. 2) puts it:

> "The widening application of data science methods to nearly every field imaginable in the natural sciences, social sciences, and humanities opens up avenues for engagement based on what students care about and the challenges they are most interested in tackling. Data science therefore provides an opportunity to build an inclusive STEM curriculum from the ground up that connects with multiple disciplines as well as the personal passions of students."

The degree to which the data science curriculum actually succeeds at achieving goals of diversity and inclusion will depend heavily on whether attractive and appropriate pathways into the study of data science are offered. Thus, we give the issue of pathways into the study of data science some special attention here.

Table 3.1 provides links to and information about introductory data science courses that we identified in our review of data science curricula at top liberal arts colleges and selected research universities. We found that the most common approach is to offer a single-semester introductory course in data science, which complements but does not replace introductory statistics or introductory computer science. The following schools offered an introductory data science course, and required all three of introductory data science, introductory statistics, and introductory computer science for the data science major or minor: Claremont-McKenna, Denison, Lafayette, Smith, Stanford, and Union College. A slight variation on this approach is to require both introductory statistics and introductory computer science for the data science major or minor, but to treat the one-semester introductory data science course as an optional elective; schools taking this approach include Davidson and Macalester. Connecticut College offers a

"Data Information and Society Pathway," similar to a minor or concentration, which requires both an introductory data science course and introductory statistics, but does not require any separate computer science courses. Several schools that do not yet have a data science major or minor also offer introductory data science courses that are viewed as a complement to, rather than a substitute for, introductory statistics and introductory computer science, including Amherst, Grinnell, Middlebury, and Oberlin.

Another less common approach is to offer a single-semester introduction to data science course that can serve as a *substitute* for introductory statistics, presumably because it covers much of the content that would traditionally be included in an introductory statistics course; this approach is used at Berkeley, Brown, Wesleyan, and Yale. Bowdoin's Digital and Computational Studies major requires students to choose one of two introductory level courses in the subject that cover some statistics and computer programming, and then does not require any other statistics or computer science courses for the major.

Most of the introductory data science courses listed above have no prerequisites, including those at Berkeley, Bowdoin, Brown, Connecticut College, Davidson, Denison, Macalester, Oberlin, Smith, Stanford, Wesleyan, and Yale. Introductory statistics is the only prerequisite to introductory data science at Grinnell and Middlebury. Introductory computer science is the only prerequisite to introductory data science at Claremont-McKenna and Union College. Introductory data science has both introductory statistics and introductory computer science as prerequisites at Amherst and Lafayette. There are some trade-offs here. An introductory data science course with no prerequisites might do a better job of showcasing what is interesting about data science to a wider audience of students, and might serve the needs of students who want to use data science tools in certain domain areas (*e.g.*, digital humanities and digital arts), whereas statistics might not always be so necessary. On the other hand, if all students entering introductory data science already had a background in introductory statistics, then there could be more room for other content and less redundancy with introductory statistics. Students working on independent or group data analysis research projects in the introductory data science course could get farther and do more sophisticated and rewarding work. Over time, it might also become increasingly common for students to learn statistics in high school, and a statistics prerequisite would enable those students to build on and reinforce what they've learned about statistics in their introductory data science course.

A few of the schools that we reviewed, including Colby, Mount Holyoke, the US Military Academy, the University of Michigan, and Wellesley, offer data science majors or minors, but do not have an introductory data science course at all. Rather, they just require introductory statistics and introductory computer science as the gateway courses for the study of data science.

The Park City Math Institute report on "Curriculum Guidelines for Undergraduate Programs in Data Science" advocates a two-semester introductory sequence in data science that includes the elements of introductory statistics and introductory computer science that are most pertinent to the study of data science, along with other topics that are typically included in the types of one-semester introductory data science courses discussed above. A potential advantage of this approach is that it could better integrate the study of introductory statistics and computer science with applications and data-based research projects on domain topics that a diverse array of students care about. This might work better at inspiring a broader array of students to pursue further academic work in this area, and might promote the success of students with a wider array of backgrounds and prior preparation. Another potential advantage would be efficiency in terms of reducing the total number of courses that a student would need to take in order to complete a data science concentration or major. None of the schools that we reviewed currently offers such a two-semester introductory data science sequence that is designed to substitute for both introductory statistics and computer science. That might reflect the fact that all of these data science programs are relatively new, and starting out with something closer to existing statistics and computer science curriculum makes for an easier transition as a new data science curriculum is created. Thus, some of these schools might gradually transition to something more like what is recommended in the Park City report over time, and such an approach is worth considering at Williams.

With that said, the Park City approach also raises a number of questions, such as: what to do about students with AP credit in statistics or computer science; what to do about students who start in this data science sequence but then decide they want to pursue a major in statistics or computer science, or start in statistics or computer science and then decide they want to do data science; and what are the implications for other majors such as economics that currently require introductory statistics and are concerned about increasing the total number of courses required for their majors (already effectively 11 courses in economics once one considers the existing requirements of introductory statistics and single-variable calculus). The approach commonly used at other schools of having a one-semester intro to data science course that complements rather than substitutes for introductory statistics and introductory computer science arguably handles those questions more smoothly. Ultimately, it may turn out that the best approach is to offer a variety of different possible pathways into the study of data science, so students with greatly varying interests, needs, and degrees of preparation will be able to find a pathway that is a good fit.

It is also important to consider the possibility of offering courses in many different disciplines that are accessible to first- and second-year students that can help them see why data science is relevant to issues they care about, and can serve as pathways to further study of data science and its applications in various fields. Table 3.2 offers some examples. At Berkeley, "connector courses" in various different disciplines are designed to be taken concurrently with their introductory data science course, Data 8: Foundations of Data Science. These courses apply the tools learned in introductory data science to questions in other domains. Perusing that list of courses provides numerous excellent examples of how data science can be applied to questions across the liberal arts. Some of the many examples include courses such as: Crime and Punishment: Taking the Measure of the US Justice System; Child Development Around the World: Analyzing Household Data Sets; Data Sciences in Ecology and the Environment; Rediscovering Texts as Data; How Does History Count? Exploring Japanese-American Internment through Digital Sources; and Data Science for Cognitive Neuroscience. It would be difficult to transplant this approach exactly to Williams without major changes to our course-credit system, as Berkeley's relatively flexible credit hour system makes it work well there. But one can imagine courses such as these eventually proliferating throughout the curriculum at Williams, perhaps with one or two semesters of introductory data science as a prerequisite. Another great example is a new introductory-level economics course offered by Raj Chetty at

Harvard, "Using Big Data to Solve Economic and Social Problems." This course illustrates how tools of data science and economics can be usefully applied to shed light on questions relating to income inequality, socioeconomic mobility, education reform, racial disparities in economic opportunity, and the environment, among other pressing topics. The course has reportedly had some success in increasing the diversity of students choosing to study economics (Matthews, 2019). At Grinnell, the History Department offers an introductory-level course on Introduction to Historical Inquiry: Digital Histories which provides another nice example.

Design of intermediate and upper-level curricula at other schools

In addition to the introductory sequences discussed above, it is common for data science majors and minors to include some or all of the following elements: some intermediate coursework in statistics and computer science; an advanced data science course covering techniques such as machine learning; some math (often including calculus and/or linear algebra); coursework on the ethical or social implications of data science; one or more courses in another domain; and a capstone course, often involving an original data analysis project. A challenge is that this can add up to a very large number of courses, which is particularly concerning at a place like Williams where the norm is just nine courses for a major (or up to eleven courses for an interdisciplinary major). This is part of what motivates efforts to create and offer dedicated data science courses that include just the elements of math, statistics and computer science that are most pertinent to data science. On the other hand, this latter approach raises a number of questions about how the data science curriculum would interact with the rest of the existing curriculum. For example, what should the institution do about students who take a two-semester "math for data science" sequence that includes elements of several different conventional math courses, but then discover that to get where they want to go academically (*e.g.*, graduate school or another major), they need to learn the elements of those conventional math courses that were left out?

In relation to existing statistics, computer science, and mathematics curricula, data science curricula are on a spectrum. At one end of the spectrum would be data science programs where all of the required courses are conventional statistics, computer science, and math courses. The data science majors at Mount Holyoke (Table 3.6) and Wellesley are good examples of this approach. At the other end of the spectrum would be the major proposed in the Park City Math Institute report (Table 3.11), which extracts selected topics from a much broader array of math, statistics, and computer science courses and compresses them into a significantly smaller number of dedicated data science courses. No school that we reviewed offers a major or minor that follows the Park City approach, but Berkeley (Table 3.3) and Denison (Table 3.4) offer interesting and innovative examples of data science programs that are closer to that end of the spectrum. The Statistical and Data Science Major at Smith (Table 3.8) and the data science minors at Macalester (Table 3.5) and Wesleyan (Table 3.9) provide good examples of data science programs that are somewhere near the middle of that spectrum.

Data science programs vary greatly in terms of how much coursework in other domains is required, and how well integrated the study of data science is with those other domains. Berkeley (Table 3.3) and Denison (Table 3.4) seem to do a particularly good job of integration, but at the expense of requiring a very large number of courses to complete the data science major.

There is substantial variation across schools in terms of how ethical and social questions about data science are integrated into the data science curriculum. Certainly these questions should be highlighted in any data science curriculum. Concerns related to data ethics have recently captured the public's attention. For example, the book *Weapons of Math Destruction* by Cathy O'Neil discusses how the careless use of big data and algorithms can exacerbate inequalities and injustices across economic, social, and racial dimensions. An approach used in some data science programs is to require a course relating to the ethical, humanistic, or social implications or applications of data science. Berkeley's data science major has a "Human context and ethics requirement," where students are required to take one course from a list of options, such as a Philosophy course on "Moral Questions of Data Science," or a course on "Theory and Method in the Digital Humanities." By contrast, the Smith College Statistics and Data Science Major does not require a separate course on data ethics, but rather attempts to "incorporate data ethics in many classes so that it is fully integrated throughout the curriculum" (Ott, 2019). An advantage of the latter approach is that students can see how ethical and social questions come up in many different areas of data science in an integrated way. But a potential drawback is that data science courses are generally taught by statisticians or computer scientists who have no formal training in ethics and moral philosophy, and so may be unprepared to provide the students with a systematic theoretical ethical framework for thinking about these issues. As a result, it could be that a combination of the approaches is best, but this also contributes to a potentially very large number of required courses for a major or concentration.

Other examples of data science across the curriculum

Courses collected under the Digital and Computational Studies rubric at Bowdoin offer some effective examples of how data science methods and social and humanistic implications of technology and data can be built into the curriculum across the social sciences, humanities, and the arts. Course titles include "How to Read a Million Books," "The Nature of Data: Introduction to Environmental Analysis," "GIS and Remote Sensing: Understanding Place," "Social and Economic Networks," "Digital Text Analysis," "Art, Technology, and Social Change," "Cognition in Analog and Digital Environments," and "Technology and the Common Good."