

Cross-Disciplinary Faculty Development in Data Science Principles for Classroom Integration

Brandeis Marshall Computer and Information Sciences Spelman College Atlanta, Georgia USA brandeis.marshall@spelman.edu Susan Geier Educational Psychology Lafayette, Indiana USA srwgeier@gmail.com

ABSTRACT

Data science in practice leverages the expertise in computer science, mathematics and statistics with applications in any field using data. The formalization of data science educational and pedagogical strategic remain in their infancy. College faculty from various disciplines are tasked with designing and delivering data science instruction without the formal knowledge of how data science principles are executed in practice. We call this the data science instruction gap. Also, these faculties are implementing their discipline's standard pedagogical strategies to their understanding of data science. In this paper, we present our crossdisciplinary instructional program model designed to narrow the data science instruction gap for faculty. It is designed to scaffold college faculties' data science learning to support their disciplinespecific data science instruction. We provide individualized and group-based support structures to instill data science principles and transition them from learners to educators in data science. Lastly, we share our model's impact on and value to faculty as well as make recommendations for model adoption.

KEYWORDS

data science, faculty development, curricula design

ACM Reference format:

Brandeis Marshall and Susan Geier. 2020. Cross-Disciplinary Faculty Development in Data Science Principles for Classroom Integration. In Proceedings of 2020 ACM Conference Special Interest Group on Computer Science Education (SIGCSE'20), March 11-14, 2020, Portland, OR. ACM, NY, NY, USA, 7 pages. https://doi.org/10.1145/3328778.3366801

1 Introduction

In their 2016 McKinsey Global Institute report, the data scientist shortfall estimates have increased to 250,000 alongside not to mention data-literate "translators" in business, medicine and other

SIGCSE '20, March 11-14, 2020, Portland, OR, USA.

© 2020 Association of Computing Machinery.

ACM ISBN 978-1-4503-6793-6/20/03...\$15.00

https://doi.org/10.1145/3328778.3366801

fields [12]. Given the increasing need for skilled data practitioners at all levels, the need for data science instruction in different fields has become paramount. The lens of data understanding and use has varying emphases and applications depending on the field. Each discipline brings context to what is coded computationally we build these technologies to help us improve mental, physical, and spiritual health, financial products, and services. We depend on reliable data as input into specific algorithms producing relevant knowledge with accurate interpretation to inform our decision-making process. The data science field is designed to address these complex and interdependent relationships. As shown in Figure 1a, data science [7] is perceived as an overlapping blend of computer science, mathematics, and some other domain discipline. The degree and depth of this interdisciplinary overlap is not specified. Data, on the other hand, are a commodity that is at the center of all aspects coded.

The role and contributions of faculty in the data science ecosystem have yet to be well-researched. The role of faculty in student academic and professional development, on the other hand, is well-documented [21]. Faculty are the first point-of-contacts, academic advisors, mentors, and role models. At small liberal arts colleges, this impact is even greater with lower student-faculty ratios [16]. Institutions are therefore tasked with developing and delivering data science instruction as part of their STEM offerings.

The challenge we face is what we call the data science instruction gap. The data science instruction gap is two-fold. First, there is a lack of a common blueprint as to the scope and shape of the data science field. Second, many faculty lack formal or informal training in data science principles thus making it challenging to translate that knowledge to their classrooms. However, faculty do bring their domain expertise as well as the standard pedagogical practices associated with their fields. The impending bachelor's recipient production of data science aware graduates hinges on faculty embedding data science in a multitude of ways.

In this work, we consider foundational aspects of data science instruction in providing clarity around defining data science (as many interpretations exist), its application to other fields and most importantly, building faculty capacity in domain-specific data science instruction. We make the following contributions:

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.



Figure 1. (a) Visual representation of disciplines that comprise of data science (b) A visual representation of the data science workflow, shown in [2]

- We present our cross-disciplinary intervention program model designed to narrow the data science instruction gap for faculty.
- We provide scaffolding techniques to assist college faculties in their data science learning and support their disciplinespecific data science instruction.
- We share our model's impact on and value to faculty as well as make recommendations for model adoption.

In Section 2, the related literature is discussed. Section 3 describes our cross-disciplinary data science intervention program for faculty. In Section 4, the impact of the learner and educator phases are shared as well as faculty perceptions of the program. Lastly, we conclude in Section 5.

2 Related Work

With data science typically associated under the STEM umbrella, there is a large existing body of scholarship on effective STEM education and evidence-based pedagogical practices [4, 6, 8, 10, 13, 19, 23]. The impact of active learning in supporting learner's understanding of concepts and the ability to demonstrate new skills are well-researched. Problem-based learning and projectbased learning are notable key learner-centered approaches where learners apply new concepts to real-world examples. In addition, the active learning paradigm can aid learners in finding and leveraging connections amongst concepts. In computing, pair programming, live coding, and whiteboard coding are active learning approached intended to build a learner's self-efficacy, competence, and confidence in computational problem-solving. Unfortunately much of the instruction reverts to powerpoint slides delivered by the instructor. Devising mechanisms to help faculty evolve their teaching practices are an element of this work. [11].

The adoption of effective educational innovations, particularly in data science, is heavily influenced by its domain specification. Programs in data science are labeled business analytics, business intelligence, data analytics, applied statistics, data mining, and information management and/or part of computer science,

statistics, business or mathematics academic units. The dominance of one field/discipline further impacts which concepts are being taught and to what depth. We present our model program designed to narrow the data science instruction gap for faculty. We describe the program design and components implemented to transition faculty from learners to educators.

Due to the newness of data science as a field, there is little core and general data science curriculum available in the literature. The National Science Foundation's Computer and Information Sciences Directorate constructed a working group that outlines a national strategic trajectory of data science with a computing focus [3]. The need for a national data science education and training agenda was listed as one of three target areas. The National Academies of Science, Engineering and Mathematics [17] brought together the STEM disciplines to begin articulating the scope and shape of data science at the undergraduate level and shared several degree program exemplars. In either circumstance, the entry pre-requisite knowledge for data science tends to starts at the upper-level undergraduate classification, e.g., databases and linear algebra. Just-in-time theoretical foundations are therefore coupled with these advanced concepts creating technical gaps [4, 6, 10, 13, 23].

Data science in practice has evolved faster than the construction of evidence-based data science instruction and pedagogy strategies. In response, data science education opportunities range in structure from 1 to 3 day short courses (trainings or workshops) to online certificates consisting of 4 or 5 15-week courses to oncampus 5 to 6-year doctoral degree programs. For example, Data Carpentry provides prepared lessons and conducts 1-2 day workshops with the intent to build data acumen and self-efficacy of researchers in data science [20]. They are filling a data skills gap for those who have at least intermediate technical science skills. At University of California at Berkeley, their Foundations of Data Science 15-week course (data8.org) introduces coding, statistics and machine learning concepts to learners (primarily early undergraduates) who have at least intermediate mathematics preparedness, according to the institution's admission criteria.

The alternative options fall within the immersive programs (bootcamps) and MOOC paradigm, which have low completion rates and rigor of instruction varies widely [1, 9, 18, 22]. The

attainment of transferrable data science skills remains elusive. Our efforts to address the data science instruction gap enlist educators in higher education due to their domain expertise, trajectory of impact in the classroom environment and potential regulation of instructional rigor through college accreditation processes.

3 Our Cross-Disciplinary Data Science Intervention Program Model

The existing data science instruction environment consists of selfcontained prepared short lessons and, on the opposite extreme, specialized long course sequences. Prepared short lessons, e.g., Data Carpentry, presupposes instructor content knowledge, data acumen, and technology familiarity. Lecture time is a scarce resource in higher education. Executing even a single prepared data science lesson can exceed the allotted class time given the slate of course topics. Also, prepared lessons have specific learning objectives and outcomes that may be a mismatch with an existing college-level course. While these lessons can serve as introductory exposure to data skills, they are stand-alone workshops as a cocurricular activity not designed for classroom integration.

Existing specialized long course sequences offered by MIT, Harvard University, Stanford University, Carnegie Mellon University, Duke University and other online-based immersive programs available through coursera, deeplearning.ai, and fast.ai, require significant prerequisite knowledge where a small portion of learners thrive. Again, instructor content knowledge, data acumen, and technology familiarity are assumed as well as course fit within the faculty's home academic unit and institution. Thus, college-level full course adoption or adaption could be difficult to navigate and sustain requiring additional levels of institutional support.

To design and develop data science instructional materials, many educators rely on reference books since there are no course textbooks with comprehensive terminology, problems, exercises, case studies, and projects. Learners are uncertain of assessing their data science understanding and skills attainment. There is substantial curricula design and development work to be accomplished by educators for comprehensive, connected and cohesive content delivery.

We present our cross-disciplinary data science intervention program model designed to narrow the data science instructional gap for faculty. In the learner phase of the program, faculty engage in an intensive instructional program during a two-week retreat. We aim to provide and support faculty in developing their data science awareness, broadening their data science understanding and exposing them to data science technologies. In the educator phase, we intend to support faculty in customizing their curricula for incorporating data science principles. Our objective is to capture how data science may manifest in different disciplines and courses. Table 1 displays the faculty participants 'general demographic information.

3.1 Learner Phase: Characterizing Data Science

Data science at the college level is positioned as an interdisciplinary discipline, typically housed within a business, computer science, mathematics, or statistics department.

Гable	1. Faculty	Participant	Characteristics
-------	------------	-------------	-----------------

Departments Included	Biochemistry, Biology, Business, Computer Science, Economics, Environmental Science, Health Science, Mathematics, Physics, Political Science, Psychology, and Art and Visual Culture			
Gender	10 Women; 10 Men			
Position Titles	Full Professor (1) Associate Professor (12); Assistant Professor (4); Instructor/Lecturer (3)			
Average Number of Years Teaching	13 years (Min 2, Max 34)			
Average Class Size	24 students (Min 10, Max 40)			
Statistics	Eight participants have a basic understanding of statistics; 11 have a more advanced understanding, and 1 participant has no experience with statistics.			
Computer Science (CS)/Programming (P)	7 participants are proficient in CS/P; 12 have minimal experience and 1 participant has no experience with CS/P.			
Python, R	7 participants have varying levels of experience with Python; 6 have experience with R; 9 participants have no experience with Python or R			

Each discipline offers a distinct interpretation of data science given its newness as a recognized field of study. The varied framing of 'what is data science?' necessitates the establishment of a unifying definition that serves as a consistent point of reference for faculty in their learning. We define data science as "an ecosystem dedicated to the systematic collection, management, analysis, visualization, explanation, and preservation of both structured and unstructured data. Through scientific methods and processes, the field of data science intends to extract impactful knowledge and insights to better the human condition" [15].

In Figure 1b, we make data science more concrete and tangible to our faculty participants via a visual workflow [2]. While presenting the workflow, we explained each phase in greater detail. The lively discussion amongst faculty participants led to the enhanced version of the data science workflow examples later provided to the faculty for use in their courses. An overview of the data science workflow elements is presented in Figure 1b. There are five elements: 1. Data acquisition and cleanup focus on data sourcing, collection, and (re)formatting. 2. Storage and management consider how to organize data collected for effective handling of complex requests. 3. For data analysis, algorithms and processes are leveraged to transition raw data into actionable insights or information. 4. Data visualization centers on presenting findings in a visual form to help data practitioners determine if their analyses make sense. 5. Data communication and storytelling address the data practitioner's journaling of these phases alongside the dissemination of insights in an accessible manner for non-data professionals. The output of one phase feeds as input into the next phase in hopes of describing a comprehensive understanding of the data. With each phase, there is a suite of technologies and common practices intended to support comprehension.

Several publicly-accessible data science projects are shared with the faculty participants including the Iris Dataset (https://github.com/venky14/iris-dataset) and course-specific data science modules (https://ds-modules.github.io/DS-Modules/). Seeing data science in practice helps to solidify the workflow. However, faculty participants desire experience in constructing a data science project from scratch incorporating elements from the data science workflow. As an example, we use a public dataset of the New York State's Take 5 drawing covering the past few years. Although multiple data formats are available, we select the CSV file format. Before conducting exploratory data analysis (EDA), we first establish our dataset investigation objective, which is evaluating the randomness of these drawings. Then, we must "clean" the data such that the computer can mimic the processing of the numbers as a human. Table 2 shows the data in its original form where the computer would read the middle column as one large number rather than five distinct ones. Table 3 displays the reformatted data after some wrangling using common Python pandas methods where the winning numbers are split into the five distinct items with labeled columns.

Table 2. Snapshot of NYS's Take 5	Winning Drawing (raw data)
-----------------------------------	----------------------------

Draw Date	Winning Numbers	Bonus #	
5/2/18	02 22 26 32 36		
5/4/18	01 05 12 16 31	35	

Table 3. Snapshot of NYS's Take 5 Winning Drawing (cleaned data)						
Draw Date	WN1	WN2	WN3	WN4	WN5	Bonus
05-02- 2018	2	22	26	32	36	Null
05-04- 2018	1	5	12	16	31	35

Then, the descriptive statistics and data analyses occur using other Python pandas methods. Visualizing data with matplotlib and seaborn libraries are cornerstones of EDA. The Jupyter Notebook provides an interactive web application environment to write and execute any coding elements as well as integrate documentation and interpretations of outcomes for both the technical and nontechnical audiences. The final product becomes a report detailing a comprehensive data science workflow.

Our data science definition, workflow visual aid, and Jupyter Notebook examples help to anchor the faculty participants in a newly framed set of concepts. A data science starter slide deck, about 20 slides, covered this material in three sections: defining data science, data science careers, and data science workflow indepth. The slide deck serves as a valuable introduction to data science for the faculty participants. Faculty began to forge connections between each element of the data science workflow and their selected discipline-specific, data-centric course. With classroom implementations as the main deliverable, faculty are encouraged to gauge their competencies and confidence related to data science principles in the context of their domain expertise.

3.2 Educator Phase: Scaffolding in Data Science Principles

During the educator phase, faculty participants gather every month to ideate and share curricula structure, lesson plans, and assessment mechanisms [5]. Faculty participants distill similarities and differences in how data collection, cleaning, analysis, visualization, and storytelling interlace across and within their disciplines and courses. Key similarities across disciplines encompass degree of familiarity and competency of statistics and computing. Key differences focus on particular discipline conventions, e.g. research method processes and practices, and standard technology tools common to a particular discipline.

Through continued discussions, the predominant roles of the faculty's domain and discipline's standard technology infrastructure result in a method to support their data science curricula development. While integrating data science concepts into their courses, faculty refine the scope of data science to align with specific technologies commonly used in their domains. For instance, an elective physics course may consider Python as the preferred programming language; however, this would depend on both faculty and learners having the requisite algorithmic thinking and design skills in that language.

We use the prepared short lesson structure to provide a universal framing for data science and couple it with the specialized course structure to administer the contextual data science principles for each discipline. Each faculty anchor their data science instruction by first delivering a prepared 20-slide lecture. Additionally, we provide data science learning support structures, e.g., individualized consultations, seminars, and exercises. During the individualized consultations with the lead program instructor, the status of the faculty participant classroom integration is reviewed. The lead program instructor offers potential resources and suggests ways to integrate data science principles that would not adversely impact the delivery of other course topics. The monthly seminars and exercises guide the data science curricula design and development work overtime. The intent is to showcase the most relevant data science workflow elements in the data science curricula unit with a referential context and understanding of the data science workflow.

Through the learner and educator phases, faculty participants create and deliver customized domain-specific data science instruction. One business faculty member thought the emphasis for their students should be on the data analysis, visualization, and communication workflow elements. In their opinion, Storage should be part of the computer science disciplines since it requires more technical expertise. Data acquisition and cleanup would be addressed at a higher level in the research realm. As for the best way to introduce data science to business undergraduates, a hands-on activity that gives students common datasets and then teaches them how to look within those datasets for interesting correlations may resonate.

4 Impact of Learner and Educator Phases

Based on their classroom integrations, there are several key observations: (1) each discipline's interpretation and communication with respect to data vary, (2) prepared lessons do not influence classroom adoption or adaption, (3) time investment to create curriculum could be significant and (4) a faculty's mindset about data science, whether growth or fixed, directly impacts the depth of their intervention. In general, the level of interpretation that we envisioned was more substantial. We recommend domain-specific data science courses for each department to capture the specific discipline and technology dependencies.

4.1 Faculty Perceptions

4.1.1 In anticipation of participating in the program. We asked participants to state their curricular and scholarly objectives related to participating in the program and their motivations and concerns related to participating in the project. Regarding curricular objectives, all responses focused on expanding their students' knowledge and skills associated with data science by:

- Relating data science to decision-making
- Learning the data science components and transferring that knowledge to their students
- Presenting real-world applications using mathematics and data science
- Introducing Python programming to their students
- Providing students with modern tools and resources to address modern problems

Their scholarly objectives centered on expanding their abilities to manage and analyze data sets related to their research interests and current projects. The majority of participants were motivated to participate in the program because they recognized the growing influence data science is having in the marketplace. They also wanted to expose their students to using data science in research. Other reasons for participating: making existing courses more interesting and relevant to their students and learning techniques that would enhance their own research processes.

We asked participants to share concerns they might have about learning data science concepts and developing a data science module for their courses. The majority were concerned about the time constraints, a few about their own and their students' lack of prior programming experience, and a few participants were concerned that their students would not have access to the software/resources needed. Time constraints would become the most significant obstacle for introducing data science concepts in their courses.

Representative comments made by faculty before attending the program follow:

"I would like to integrate more data science discussion in our required methods courses to raise awareness of data science among students. ... Learning more about data science would allow many more of our students to imagine themselves as data scientists and to recognize how well they can do in data science careers."

"I want to learn more about data science, its connections to other disciplines, what skills are necessary to be successful and how my mathematical background is useful. I have recently had a growing interest in data analytics, data mining and information management and would like to use this opportunity to get valuable information and training in these areas."

"A few years ago, I was on the job market and noticed data science as an evolving field. Thus, I have increasing become very interested in data science. I want to expose students to this evolving career."

4.1.2 After participating in the two-week retreat. Participants commented about the importance and value of engaging with retreat colleagues to gain a better understanding of developing the data science module. Many commented about how their perceptions of data science had changed from their retreat experience. It was clear from the observations, discussions and survey responses that the faculty were better prepared to develop and share data science curriculum with their students. Faculty participants also acknowledged the need for the continued support they would receive during individual meetings and virtual Q & A. They particularly valued the retreat and looked forward to the monthly seminars with their cohorts where they could share successes and challenges and receive feedback and suggestions from experts and knowledgeable colleagues. Therefore, the value for a year-long program rather than a one- or two-week event with no follow-up was confirmed. Faculty comments follow: "Attending the retreat helped me to formalize what data science is and to see more clearly the importance of it as a discipline and its potential value to my students and me. It was great hearing and seeing various perspectives on it from the point of view of colleagues in different disciplines ... "

"I was very unsure of what all data science entailed. I was familiar with some skills that were necessary, but did not have a full understanding as I do now. I think these two weeks have been invaluable. I especially enjoyed constructed time set aside for us to collaborate with our colleagues, form bonds, and build networks between us while learning and considering implementation of Data Science"

4.1.3 Curriculum Development Phase. The curriculum development phase began in week two of the summer retreat and continued during the academic year via individual meetings, monthly seminars, and, targeted assignments designed to facilitate data science instruction in a step-wise process. Faculty aligned their data science curriculum with one or more of the workflow elements. The STEM and business faculty emphasized the data analysis element and typically represented all elements in their instruction. The non-STEM faculty typically emphasized the visualization and communication elements. Faculty often remarked about the value of the instructional supports built into



Figure 2. Student post-instruction survey outcomes

the program. Many considered this scaffolding to be essential as they incorporated data science into their courses.

"The individual meeting allowed me to discuss what I planned to do and if it sufficiently met the requirements of the program. Oftentimes, you have an idea in your head but are unsure if it will work. The individual meetings allowed me to articulate that plan and move forward with its implementation. I can comment that I should have used my resources more frequently and effectively."

"The monthly seminars were useful in that it was nice to see the other participants and to see where they were in their module development as well as share success and challenges. These seminars were difficult to work into my schedule, but I feel like they were crucial to the success of the [model program]"

"I found the Data Science Module Guidelines, introductory slide deck, individual meetings and monthly seminars very useful to be progressing in the work step by step. It was of particular convenience to have all the content in a central repository in Piazza so we could always go back and check."

4.1.4 *Curriculum Development Phase.* Faculty participants had several discussions with the lead instructor and each other as they developed their data science infusions for their courses. Representative comments follow:

"The [learner phase] helped a lot. I don't think I would have been able to do my module without the retreat. The retreat is what made me feel comfortable enough to be able to talk about Data Science with my students. Also, the retreat is where I had the time to start developing my module. Once the semester started, it was difficult to find time to work on the module. So, the retreat afforded me the time to think about and develop the module."

4.1.5 Classroom Implementation Phase. As we gathered faculty perceptions of our model, we collected their perceptions of student engagement of their lessons. In general, data science lesson was effective in terms of introducing students to data science, connecting data science to their disciplines, and exploring data science career information. Some faculty remained unsure about their effectiveness in delivering domain-specific data science instruction. Nevertheless, the faculty indicated that students were generally interested, some students wanted additional/more

advanced information about data science, the career information was interesting for most students.

Nearly 250 students were exposed to our data science intervention model. We collected students' attitudes toward data science postinstruction and display outcomes of their impressions in Figure. 2. Two impactful Likert-scale questions posed were: 1) data science is relevant to my undergraduate degree, and 2) I would consider taking data science courses if they were offered. Over 80% of the students who completed the post-instruction survey agreed that data science was relevant in their undergraduate matriculation plus they were willing to further their data science instruction. Additionally, we asked 162 students if their interest and understanding of data science had increased due to the data science instruction they received in their course. Seventy-six percent (76%) of student respondents agreed that their interest had increased. Moreover, 97% of the students agreed that their understanding of data science had increased.

4.1.6 *Challenges*. Finding time was the most challenging aspect of the program: finding time to develop the module during the academic year, finding time during class to cover the material, and finding time to learn the data science material to assist students.

"The main constraint in the development process was time. I'd wish I had more time to dedicate to this area as I find it fascinating. Learning curve is steep so it requires good amount of time for me to learn the content before being able to communicate well to students. I know I'll feel much more comfortable in future semesters. I'm planning to continue including Data Science as part of [my course].

5 Conclusions

Prepared lessons may provide some flexibility; however the structured program described provides the scaffolding needed to effectively prepare the participants to develop their disciplinerelated data science curriculum. Faculty participants benefited from the collaborative engagements with colleagues during the retreat, the seminars and after they implemented their data science content in their courses. We suggest taking the data science instruction one step further. It would be beneficial to extend the data science instruction beyond one or two topics offered in a course to an entire course designed to highlight data science within the specific discipline of the major course of study.

ACKNOWLEDGMENTS

This work has been supported by the National Science Foundation under Grant No. 1623362.

REFERENCES

- [1] C. A. Arbeit, A. Bentz, E. F. Cataldi, and H. Sanders, "Alternative and Independent: The Universe of Technology-Related Bootcamps", RTI Press Publication No. RR-0033-1902. Research Triangle Park, NC: RTI Press. https://doi.org/10.3768/rtipress.2018.rr.0033.1902, 2019.
- [2] A. Azam. "The First Rule of Data Science", Berkeley Science Review. 2014. Online: http://berkeleysciencereview.com/article/first-rule-data-science/
- [3] F. Berman (co-chair), R. Rutenbar (co-chair), H. Christensen, S. Davidson, D. Estrin, M. Franklin, B. Hailpern, M. Martonosi, P. Raghavan, V. Stodden, and A. Szalay.. "Realizing the Potential of Data Science: Final Report from the National Science Foundation Computer and Information Science and Engineering Advisory Committee Data Science Working Group, "National Science Foundation Computer and Information Science and Engineering Advisory Committee Report, December 2016. Online: https://www.nsf.gov/cise/ac-data-science-report/CISEACDataScienceReport1.19.17.pdf

- [4] B. Baumer. A data science course for undergraduates: Thinking with data. The American Statistician 69, 4, 334-342, 2015.
- [5] N. C.C. Brown, M. Kölling, "A Tale of Three Sites: Resource and Knowledge Sharing Amongst Computer Science Educators", Proceedings of the ACM Conference on International Computing Education Research. pp. 27-34, 2013.
- [6] J. Brunner, Robert & J. Kim, Edward. Teaching Data Science. Procedia Computer Science. 80. pp. 1947-1956, 2016. 10.1016/j.procs.2016.05.513.
- [7] L. Cao. Data Science: A Comprehensive Overview, ACM Computing Surveys (CSUR), v.50 n.3, p.1-42, 2017.
- [8] J. Daugherty. Infusing engineering concepts: Teaching engineering design. National Center for Engineering and Technology Education, 2012. http://files.eric.ed.gov/fulltext/ED537384.pdf
- [9] D.F. Feldon, S. Jeong, J. Peugh, J. Roksa, C. Maahs-Fladung, A. Shenoy, M. Oliva, "Null effects of boot camps and short-format training for PhD students in life sciences", Proceeding of the National Academies of Sciences, 114(37):9854–9858, 2017. doi: 10.1073/pnas.1705783114.
- [10] J. Hardin, R. Hoerl, N J. Horton, D. Nolan, B. Baumer, O. Hall-Holt, P. Murrell, R. Peng, P. Roback, D. Temple Lang & M. D. Ward. Data Science in Statistics Curricula: Preparing Students to "Think with Data", The American Statistician, 69:4, pp. 343-353, 2015. DOI: 10.1080/00031305.2015.1077729
- [11] C. Henderson and M.H. Dancy, "Physics Faculty and Educational Researchers: Divergent Expectations as Barriers to the Diffusion of Innovations," American Journal of Physics, vol. 76, Jan. 2008, pp. 79-91.
- [12] N. Henke, J. Bughin, M. Chui, J. Manyika, T. Saleh, B. Wiseman, G. Sethupathy, "The Age of Analytics: Competing in a Data-Driven World", McKinsey Global Institute, pp. 1-136, 2016. Online: https://www.mckinsey.com/businessfunctions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-adata-driven-world
- [13] Stephanie C. Hicks and Rafael A. Irizarry. A Guide to Teaching Data Science, The American Statistician, 72:4, 382-391, 2018. DOI: 10.1080/00031305.2017.1356747;

https://amstat.tandfonline.com/doi/abs/10.1080/00031305.2017.1356747#.XQGsot NKg1g

- [14] D. Kolb. Experiential Learning as the Science of Learning and Development. Englewood Cliffs, NJ: Prentice Hall, 1984.
- [15] B. Marshall and S. Geier. "Targeted Curricular Innovations in Data Science", Proceedings of the IEEE Frontiers in Education Conference, 2019.
- [16] J. Monks and R. Schmidt. The impact of class size and number of students on outcomes in higher education, 2010. [Electronic version]. Retrieved August 29, 2019 from Cornell University, School of Industrial and Labor Relations site: http://digitalcommons.ilr.cornell.edu/workingpapers/114/
- [17] National Academies of Sciences, Engineering, and Medicine, "Data Science for Undergraduates: Opportunities and Options," Washington, DC: The National Academies Press, 2018. Online: https://doi.org/10.17226/25104.
- [18] J. Reich and J. Ruipérez-Valiente, "The MOOC pivot", Science. 363. 130-131. 10.1126/science.aav7958, 2019.
- [19] C. Taylor, J. Spacco, D. Bunde, T. Zeume, Z. Butler, M. Barnas, H. Bort, F. Maiorana, C. Hovey, "Promoting the Adoption of Educational Innovations. 368-368, 2018. 10.1145/3197091.3205847.
- [20] T. K. Teal, K. A. Cranston, H. Lapp, E. White, G. Wilson, K. Ram, A. Pawlik, "Data Carpentry: Workshops to Increase Data Literacy for Researchers", International Journal of Digital Curation 10 (1), 135-143, 2015.
- [21] P. D. Umbach and M. R. Wawrzynski. Faculty Do Matter: The Role of College Faculty in Student Learning and Engagement, Research in Higher Education 46 (2):153-184, 2005.
- [22] K. Word, "When Do Workshops Work? A Response to the 'Null Effects' paper from Feldon et al.", https://software-carpentry.org/blog/2017/12/response-nulleffects.html.
- [23] C. Wymbs. Managing the innovation process: Infusing data analytics into the undergraduate business curriculum (lessons learned and next steps). *Journal of Information Systems Education*, 27(1), 61–74, 2016.