PRACTICE

A Student Research Course on Data Analytics Problems from Industry: PIC Math

Michael Dorff, *Brigham Young University*
Suzanne Weekes, *Worcester Polytechnic Institute*

Abstract

The PIC Math program works with mathematical sciences faculty at US institutions to help them prepare students for careers in today's workforce. PIC Math faculty engage their students to solve data science and other research problems that come directly from industry. To accomplish the program's objectives, the authors conduct a three-day summer training workshop for faculty, provide resources and training so faculty can successfully teach a semester-long course in which students are mentored in solving a research problem, and organize a student summer recognition conference. In the first three years, 107 faculty members from 101 institutions, including 16 historically black colleges and universities and Hispanic-serving institutions, have participated in PIC Math. Faculty mentored 1,397 undergraduate students in research that came from more than 170 business and government agencies.

Keywords: *business, data analytics, industry, PIC Math, real-world problems, undergraduate research*

doi: 10.18833/spur/2/4/2

Imagine a semester course in which the faculty member does not lecture but instead the students work in small groups to solve data analytics problems that come from business, industry, or government (BIG). In this course, the students' main assignments are to solve a semester-long industry problem, write a final 12-page paper about their research results, and give a 12-minute oral presentation. Also, each research problem originates from an external partner who provides the problem, answers questions from students, and receives periodic updates from students. The students learn how to solve data analytics problems, how to work in groups to solve a research problem, and how to communicate effectively with their peers in the group and with their industry contact. Such a course is the heart of the Preparation for Industrial Careers in Mathematical Sciences (PIC Math) program. PIC Math is a program funded by the National Science Foundation (NSF) and the National Security Agency (NSA) with a goal of preparing mathematics and statistics students for industrial careers by engaging them in research problems from BIG. To date, more than 1,300 undergraduates at more than 100 US universities and colleges have taken PIC Math courses.

Examples of data analytics problems solved by students include the following:

- Although the population of Youngstown, Ohio, had undergone a dramatic decline and a shift in the location of residents over the past 40 years, the police department was still using police beats based on a division of the city created decades ago. Students in the PIC Math course at Youngstown State University received 2014 crime data from the police department, analyzed the data, and proposed several new models for more equitable divisions of the city into police beats (Haigler et al. 2016). After the students presented their results to the police department and the mayor, the police department adopted one of the models proposed.
- The Field Museum in Chicago implemented a crowdsourcing project in which the public could help classify microscopic plants from a large sample collection. The museum received hundreds of thousands of pieces of crowdsourced data, most of which was usable. However, some was not. Students at Roosevelt University in

Chicago were asked to devise a criteria for determining which of the crowdsourced data was usable and which should be rejected.

- Online players at Kongregate, a browser-based video game website owned by GameStop Corp, rate the video games on the site. However, some users were submitting fake ratings to improve the reputation of their game and obtain benefits. Equipped with a large amount of data from Kongregate on the users who rate games (userID, country, IP address, age of account, number of points, and so forth), a group of students at Winthrop University in South Carolina responded to Kongregate's invitation and developed an algorithm using data-mining strategies to detect fraudulent ratings.

Background

There are growing opportunities for students with expertise in data analytics. In 2011, the McKinsey Global Institute report on data sciences or data analytics (Manyika et al. 2011, 3) projected that, by 2018, "the United States alone faces a shortage of 140,000 to 190,000 people with deep analytical skills as well as 1.5 million managers and analysts to analyze big data and make decisions based on their findings."

To help increase the number of students prepared for data analytics, a better connection between the work of government, business, industry, and nonprofits and that of educational institutions is needed. The National Academy of Sciences produced a report (National Research Council 2012) presenting 10 breakthrough actions vital to the prosperity and security of the United States. It recommended that businesses and universities should work closely together to develop new programs that "address strategic workforce gaps for science-based employers" (National Research Council 2012, 92) and that BIG should be involved in university programs "to provide internships, student projects, advice on curriculum design, and real-time information on employment opportunities" (National Research Council 2012, 143). Also, the Society for Industrial and Applied Mathematics (SIAM) released a report on mathematical sciences in industry (SIAM 2012) containing a list of suggestions and strategies based on surveys, interviews, and focus sessions with mathematical and computational scientists from industry. A major theme among these suggestions is to promote more interactions between academia and industry, and to expose students to the type of work done in industry.

A successful way to prepare students for data science careers is to provide them with opportunities to work on significant, real problems from BIG. In other words, provide them with research opportunities consisting of problems from business, industry, and government instead of problems from academia. In addition, care must be taken not to simplify the problems for the students so much that they start to resemble standard classroom problems. In addition, the benefits of undergraduate research are numerous. Studies have shown that students who are engaged in research are more successful during and after college in terms of problem solving, critical thinking, independent thinking, creativity, intellectual curiosity, disciplinary excitement, and communication skills (Bauer and Bennett 2003; Lei and Chuang 2009; Mabrouk 2009; Seymour et al. 2004). Many of these are skills desired by employers of data scientists for their employees. Also, studies have shown that students from underrepresented groups have improvements in grades, retention rates, and motivation to succeed in graduate school (Barlow and Villarejo 2004; Ishiyama 2001; Ishiyama and Hopkins 2002). Combining these two components—research problems coming from industry and students being mentored in undergraduate research—can be a successful approach to increase the number of students who graduate from college prepared for a career in data science.

The PIC Math Program

The PIC Math program, founded and directed by the authors, prepares mathematical sciences students for data science and other careers in industry by engaging them in research problems that come directly from industry. Specifically, it aims to increase awareness among students and faculty in the mathematical sciences about career options in industry, provide research experience working on real-world problems from BIG, and prepare students for industrial careers. The PIC Math program includes the following:

- a three-day faculty summer workshop that provides faculty with information and training to prepare students for industrial careers,
- a spring-semester course in which students work on a research problem from industry and communicate their work while learning skills to prepare for a career in industry,
- a student recognition conference.

Faculty applying to be considered for the program must submit a one-page statement of interest, provide a letter of support from their supervising chair or dean, and give basic information about themselves and their prospective PIC Math students.

Faculty Summer Training Workshop

A three-day faculty summer workshop is held on the campus of Brigham Young University to train faculty members to teach the course and mentor students in research. Topics for the workshop include the following:

- information on nonacademic careers and internships for undergraduate students,
- guidance on developing business and industry connections and partnerships,

- exposure to data science and other problems that arise in industry,
- training on mentoring students to work on research and on helping students develop skills that are valued by employers (e.g., speaking and writing, working in groups, working on open-ended unsolved research problems, and coding or programming), and
- preparation for teaching a PIC Math course.

Typically, the attendees include about 35 first-time PIC Math instructors, three returning PIC Math professors who share their experiences from participating in the program during a previous year, three mathematical scientists who work in industry, and the PIC Math co-directors. Some self-funded, non-PIC Math faculty often attend because of their interest in the program. Feedback about the workshop from the faculty participants has been extremely positive.

A Spring-Semester Research Course Using Problems from Industry

PIC Math faculty members teach a credit-bearing course in the spring semester at their institution focused on solving data science or other industry problems. The course targets undergraduate students majoring in mathematics and statistics but is open to students of all majors. Class sizes typically range from 5 to 15 students. Teams of 3 to 5 students are formed, and each team spends the semester working on a research problem received from a BIG partner. This helps the students develop their problem-solving, teamwork, and communication skills. Occasionally, two teams in a class may elect to work on the same industrial problem, each taking a different approach or focusing on a different aspect of the larger problem.

The PIC Math program also has produced the PIC Math Industrial Case Study videos (see MAA n.d.a). These are four, 2-video sequences highlighting some research problems encountered by mathematicians and statisticians outside of academia. The first video in each sequence features a mathematical scientist in industry talking about his or her career and the type of research problems in his or her workplace. The second video in the sequence features a faculty member presenting some technical background and an approach that can be used to make progress on the industrial research problem. For example, the video sequence pair "Improving Marketing Strategies" presents a problem in which data analytics is used to help an online company distinguish between its business customers and its private consumers from messages on gift cards. Featured on this pair of videos are Jonathan Adler Nolis, a consultant, and Talithia Williams, an associate professor of mathematics at Harvey Mudd College. At the start of the semester, it is a helpful activity for the students to watch the first of one of the video pairs in class, then discuss how to begin to solve the problem presented in the video. The next step is for students to view the companion video that explains a way to approach the problem.

Faculty members usually take a few days at the start of the PIC Math course presenting background material that will be helpful for the research project. This material could include some background on coding and using software to help deal with large data sets, or a discussion on how to clean data (e.g., how are errors or omitted data in the data set handled). However, faculty members are encouraged to allocate most of the class time for student work in their groups to solve the research problem. For this, students could discuss potential approaches to solving the problem, break the problem into smaller and more manageable pieces, work on solving those smaller pieces of the problem, learn new concepts they might need to solve a piece of the problem, interact with the company liaison, present the current status of their solution to the class, and keep a written record of what they have tried so far. Each group in the course finishes with a solution to its research problem, creates a video recording of a 12-minute oral presentation of its solution, and writes a 12-page paper on its research problem and solution. These are shared with the team's industrial liaison who proposed the problem and who worked with the students during the class.

Finding appropriate research problems is important. The program directors could just give the faculty members a research problem that they have obtained from contacts in industry. However, it is better to train the faculty members on how to develop contacts in BIG and on how to obtain a suitable research problem for their students to work on in the PIC Math course. By training faculty members to develop BIG contacts and supporting them in this progress, the likelihood is increased that faculty members will continue to teach a PIC Math-type course after the end of the NSF grant and thus sustain the overall effort. All PIC Math instructors make some contacts in industry, and a majority of the instructors have obtained suitable research problems from these contacts for their students to work on during the spring-semester course. The PIC Math program directors, however, also find industrial liaisons and research projects so that if a faculty member is unable to secure an appropriate project in time for a spring-semester course, they can work with one of these reserve problems.

About 90 percent of the problems obtained from industry are data science problems such as those mentioned at the beginning of this article. Many companies may have a large amount of data but may not have the time or the expertise to analyze it. These companies are willing to share this data with faculty members and students to obtain answers to their questions and to participate in the process of training and developing the future labor market. Also, all of the problems have a contact person from industry, which provides the students with experience communicating and interacting with people in industry.

Faculty members teaching the PIC Math course and others can consult the PIC Math course resource website (see MAA n.d.b) for sample course syllabi and schedules, sample research problems from industry, sample video presentations by students on their research, and other resources. The site is not limited to just PIC Math participants and has been used as a resource by some non-PIC Math faculty to teach a PIC Math-type course.

At the end of the PIC Math course, the faculty member selects one research group in the course as the "official PIC Math group." That group submits its 12-page written paper and the recording of its 12-minute oral presentation to the PIC Math directors. Each portfolio is sent to two mathematical scientists from industry or university faculty who have extensive experience working on research problems from industry for review. Reviewer feedback is sent to the PIC Math faculty.

Student Recognition Event

The PIC Math program holds a student recognition session at a national summer mathematics conference with funding for students from the "official PIC Math group" to attend. During this session, the students present their solution to the research problem and attend presentations from two industrial mathematicians about their professional experiences.

Results

During the first three years of the program, 107 faculty members from 101 higher education institutions participated in the PIC Math program. These institutions are located in 32 US states and Washington, DC (see Figure 1), and include 10 Hispanic-serving institutions (HSIs) and six historically black colleges and universities (HBCUs). In the mathematics departments at these 101 institutions, 14 offer a PhD degree, 23 offer a MS/MA degree, 63 offer BS/BA degrees, and one offers an AS degree as its highest degree. In total, 1,397 undergraduate students participated in the PIC Math program during the first three years—40 percent were female students, and 21 percent were students from underrepresented ethnic groups. These statistics are summarized in Table 1. Altogether, participating undergraduate students have coauthored 147 research reports and papers, and have delivered more than 150 conference presentations. There have been 166 industrial partners.

Findings from an October 2016 survey of PIC Math faculty and students indicate that the main goals of the PIC Math program are being met. Program faculty were asked about their opinions about the PIC Math summer workshop, as well as their experiences working with industry partners and in teaching the course. Students were asked for their opinions about the course, and how it affected their thinking about careers in industry and their interest in mathematics careers in industry. Both faculty and students reported an increased awareness of nonacademic career options and felt the real-world problems addressed in the PIC Math courses were challenging and led to increased understanding of the math involved. They also reported that the program greatly contributed to the outcomes for the course (see Table 2).

Stories from students participating in the program show that students were offered jobs in industry as a result of

FIGURE 1. Locations of Participating Universities/Colleges

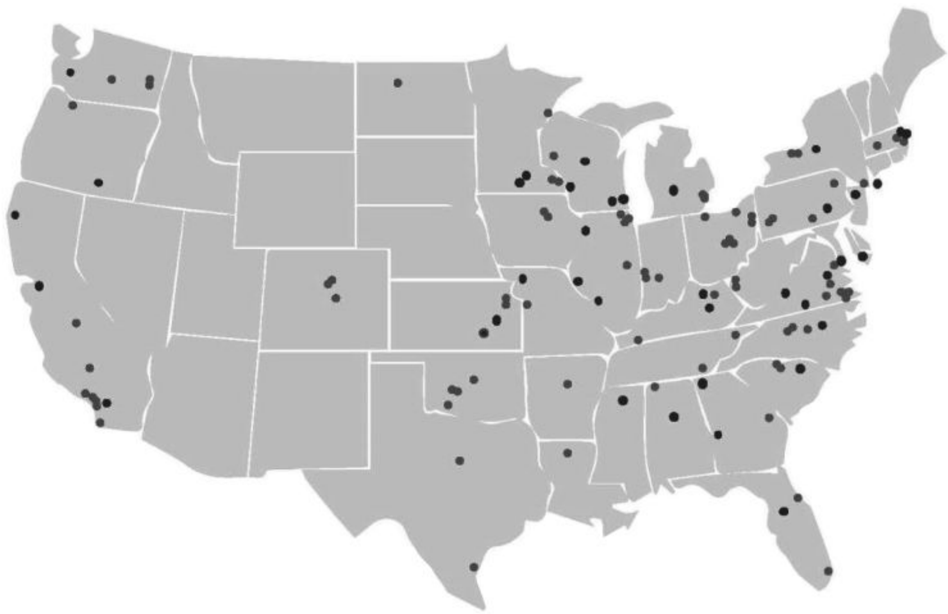

TABLE 1. Summary of PIC Math Participation Data

Number of participating undergraduate students	1,397
Number (%) of female students	558 (40%)
Number (%) of students from underrepresented groups	295 (21%)
Number of participating faculty	107
Number of institutions	101
Number of historically black institutions/Hispanic-serving institutions	16

TABLE 2. Percentage of Respondents Reporting Course Resulted in Considerable or Quite a Bit of Change in Course Learning Outcomes

	% students	% faculty
Awareness of career options in industry	70	77
Confidence in ability to work research problems from industry	91	86
Preparations for career in industry	73	84
Ability to work in groups	87	83
Ability to work with industrial sponsor	76	71
Ability to communicate mathematics through writing	78	81
Ability to communicate mathematics through oral presentations	81	81

Note: 47 faculty responses, 58 student responses

their experience in the PIC Math course. A female student at Virginia State University (an HBCU) commented that "[t]he PIC Math helped me get my first job. The experience of successfully working in groups, and problem solving were key components in my interview. When asked about leading, and how I worked well with others, I used the PIC Math as a prime example." A male student at Winona State University in Minnesota wrote, "In the PIC course we struggled a lot just getting the data to a useable status which has taught me to never expect anything to be nice and pretty. In the PIC course we worked as a team while you [sic] set deadlines that may or may not have been realistic. That is exactly the same scenario in my job now.'' Finally, a male student at the University of Pittsburgh stated that "I wholeheartedly believe that Math 1103 [PIC Math] better prepared me for my career in engineering than any other single course I took at Pitt."

Establishment of a Similar Course at an Institution

The PIC Math directors are available to communicate with other faculty members who are interested in teaching a PIC Math-type data science course at their institution. Data from a private resource website that contains sample syllabi, course schedules, letters to industry partners requesting research problems, data science problems from industry, and grading rubrics can be shared. Faculty interested in participating in the summer workshop can visit the PIC Math Workshop on Data Science webpage (see MAA n.d.c) for information and should contact the authors.

Acknowledgments

PIC Math is a program of the Mathematical Association of America (MAA) and is currently funded by NSF under grant DMS-1722275 and formerly under DMS-1345499 with SIAM. Support also is provided by NSA.

References

Barlow, Amy E., and Merna Villarejo. 2004. "Making a Difference for Minorities: Evaluation of an Educational Enrichment Program." *Journal of Research in Science Teaching* 41: 861–881. doi: 10.1002/tea.20029

Bauer, Karen W., and Joan S. Bennett. 2003. "Alumni's Perceptions Used to Assess Undergraduate Research Experience." *Journal of Higher Education* 74: 210–230. doi: 10.1080/00221546.2003.11777197

Haigler, Sebastian, Ashley Orr, Eric Shehadi, Jenna Wise, and Kristi Yazvac. 2016. "Redistricting Youngstown Police Beats." *SIAM Undergraduate Research Online* 9: 404–421. Accessed May 4, 2019. doi: 10.1137/15S013880

Ishiyama, John. 2001. "Undergraduate Research and the Success of First-Generation, Low-Income College Students." *CUR Quarterly* 21(1): 36–41.

Ishiyama, John, and Valerie M. Hopkins. 2002. "Assessing the Impact of a Graduate-School Preparation Program on First-Generation, Low-Income College Students at a Public Liberal Arts University." *Journal of College Student Retention* 4: 393–405. doi: 10.2190/cb17-uenx-n064-u1hq

Lei, Simon A., and Ning-Kuang Chuang. 2009. "Undergraduate Research Assistantship: A Comparison of Benefits and Costs from Faculty and Students' Perspectives." *Education* 130: 232–240.

Mabrouk, Patricia A. 2009. "Survey Study Investigating the Significance of Conference Participation to Undergraduate Research Students." *Journal of Chemical Education* 86: 1335–1340. doi: 10.1021/ed086p1335

Manyika, James, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, and Angela Hung Byers. 2011. *Big Data: The Next Frontier for Innovation, Competition, and Productivity.* Washington, DC: McKinsey Global Institute.

Mathematical Association of America (MAA). n.d.a. "PIC Math Industrial Case Study Videos." Accessed May 4, 2019. https://www.maa.org/programs/faculty-and-departments/pic-math/solving-real-world-problems

Mathematical Association of America (MAA). n.d.b. "Course Resources: PIC Math." Accessed May 4, 2019. https://www.maa.org/programs/students/undergraduate-research/pic-math/course-resources

Mathematical Association of America (MAA). n.d.c. "PIC Math Workshop on Data Science." Accessed May 4, 2019. https://www.maa.org/picmath

National Research Council. 2012. *Research Universities and the Future of America: Ten Breakthrough Actions Vital to Our Nation's Prosperity and Security*. Washington, DC: National Academies Press. doi: 10.17226/13396

Seymour, Elaine, Anne-Barrie Hunter, Sandra L. Laursen, and Tracee DeAntoni. 2004. "Establishing the Benefits of Research Experiences for Undergraduates: First Findings from a Three-Year Study." *Science Education* 88: 493–534. doi: 10.1002/sce.10131

Society for Industrial and Applied Mathematics (SIAM). 2012. *Mathematics in Industry*. Philadelphia: Society for Industrial and Applied Mathematics.

Michael Dorff
Brigham Young University, mdorff@math.byu.edu

Michael Dorff is department chair and professor of mathematics at Brigham Young University. He earned his PhD from the University of Kentucky in 1997 and was the founder and director of the NSF-funded national Center for Undergraduate Research in Mathematics (CURM). He is a Fellow of the American Mathematical Society and has received numerous teaching awards. In 2019, he began a two-year term as president of the Mathematical Association of America.

Suzanne Weekes is a professor in the Department of Mathematical Sciences at Worcester Polytechnic Institute (WPI). She earned her PhD from the University of Michigan in 1995. Her research interests are in numerical methods for differential equations. She has led the Center for Industrial Mathematics and Statistics at WPI and the WPI REU Program in Industrial Mathematics and Statistics. She is a founding co-director of the Mathematical Sciences Research Institute Undergraduate Program (MSRI-UP).