Chapter 12
Real Clients, Real Problems, Real Data: Client-Driven Statistics Education

Talithia D. Williams and Susan E. Martonosi

Abstract In this chapter we describe two client-focused educational experiences at Harvey Mudd College that offer students the opportunity to work on real problems for real clients using real data. The first is the Harvey Mudd College Clinic capstone program, in which teams of students spend an academic year working on a project for an external sponsor. The second is a course project in an upper level statistics elective in which the students analyze data provided by a campus partner. For both of these, we describe their structure, recent projects, as well as student and client feedback. We also offer our reflections on how providing these educational experiences has influenced us personally and professionally.

Keywords Course projects • Capstone • Statistics education • Client projects

12.1 Introduction

Demand has recently surged for data-savvy individuals in organizations ranging from government agencies to start-up businesses to nonprofit organizations. The rise of data science, an interdisciplinary field that combines computer science, statistics, and mathematics to gain insights from large data sets, has led to a surplus of jobs for data scientists, but the supply of workers who are equipped to solve data-driven problems is limited. The skills needed to be a successful data scientist increasingly depend on a combination of technical expertise, effective communication, teamwork, and attention to the client's needs. While the phrase "data-driven

---

MSC Code
97K80

T.D. Williams (✉) • S.E. Martonosi
Department of Mathematics, Harvey Mudd College,
301 Platt Blvd., Claremont, CA 91711, USA
e-mail: twilliams@hmc.edu; martonosi@g.hmc.edu

J. Dewar et al. (eds.), *Mathematics Education*, Association for Women in Mathematics Series 7, DOI 10.1007/978-3-319-44950-0_12

statistics education" might seem redundant, the reality is that statistics is often taught without requiring the analysis of real data and seldom taught with considerations of the needs of a client. At Harvey Mudd College (HMC), we have created client-focused data science experiences both in the classroom and in the senior level capstone course, known as "Clinic."

HMC is an undergraduate liberal arts college of science, engineering and mathematics that is part of the Claremont Consortium consisting of four other undergraduate colleges and two graduate institutions. All HMC students complete a common core curriculum in the technical disciplines represented on campus (mathematics, computer science, chemistry, biology, physics, and engineering) and an extensive sequence in the humanities, social sciences, and the arts. Therefore, we have a student body that is proficient in the STEM fields and for whom data analysis is likely to become an integral part of their future careers.

The purpose of this chapter is twofold. First, we illustrate two types of data analysis course experiences: a client-sponsored field capstone program in Sect. 12.2, and a project run within a statistical linear models elective in Sect. 12.3. We provide the reader with implementation models that they could adapt to their own institution. Second, we reflect on our experiences teaching these courses and describe some of the challenges and opportunities they have yielded. Williams, the first author, has taught the statistical linear models course using client-based projects, so in Sect. 12.3 the narration changes to the first person. Sect. 12.4 concludes the chapter by looking to the future of statistics education. Throughout this chapter, we use the term *deliverable*, commonly used in industry, to denote a report, a presentation, an algorithm, or other work product that is delivered to a client in the course of a project.

12.2 The Harvey Mudd College Clinic Program

The Harvey Mudd College Clinic program started in the HMC engineering department in 1963. Drawing its name from the training of medical students, in which they practice their skills during clinical rotations on real patients under the watchful eye of an experienced physician, the HMC Clinic was developed to give students the opportunity to practice their skills on a real engineering problem sponsored by an external client under the supervision of a faculty member. This experience helps students synthesize their classroom knowledge and bridge the gap between the theory of the classroom and the real world.

In 1973, the HMC mathematics department adopted Clinic in its curriculum, and since then, the departments of physics and computer science have followed suit (Borrelli 2010). Participation in Clinic is a graduation requirement for engineering, computer science, and computer science/mathematics joint majors, while mathematics and physics majors can choose either Clinic or a traditional thesis to fulfill their capstone requirement. (Biology and chemistry majors typically complete a traditional thesis, although some have participated in Clinic when an appropriate project was available.) Our departments work collaboratively to recruit interesting Clinic projects. When a project requires interdisciplinary skills as is increasingly the case, we create cross-departmental Clinic teams.

Coming from applications-driven disciplines (statistics and operations research), both authors were naturally drawn to Clinic. In our disciplines, the word "research" does not always imply proof-based theoretical advances, and often refers to the creative application of existing methodologies in new ways to solve real problems. This is the heart and soul of Clinic. Given our skills and interests, we have been involved in the program as faculty advisors, director, or associate director for six and ten years, respectively.

We often hear that HMC's unique focus on science, engineering and mathematics make it difficult for instructors at other institutions to implement programs like Clinic. However, as surveyed by Martonosi (2012), and by Martonosi and Williams (2016), other institutions have taken notice of the HMC Clinic program and have successfully emulated it. The Olin College website describes their Senior Capstone Program in Engineering, which was modeled very closely after HMC's Clinic (Olin College 2016). Moreover, in recognition of the broad impact the program has had on engineering education worldwide, the professors who cofounded the program in 1963 were awarded the 2012 Bernard M. Gordon Prize for Innovation in Engineering and Technology Education by the National Academy of Engineering (Harvey Mudd College 2012). Our hope is that readers can adapt some of the characteristics described in this chapter to create a Clinic-like program at their own institution.

We outline the structure of the program in Sect. 12.2.1, describe some of the projects that have been completed in Sect. 12.2.2, provide excerpts of student and liaison feedback in Sects. 12.2.3 and 12.2.4, and reflect on our personal experiences in the program in Sect. 12.2.5.

12.2.1 Structure of the Mathematics Clinic Program

We start by listing the cast of characters involved in the Mathematics Clinic: the team, the team's project manager, the team's faculty advisor, the sponsoring organization of the project, the sponsor's liaison, and the director and associate director of the Mathematics Clinic program. We describe each of these in more detail in the subsequent paragraphs.

In Clinic, teams of four to six students, mostly seniors, work for a full academic year on a problem posed by an external sponsoring organization. The students are responsible for determining the appropriate methodology for solving the problem, conducting relevant literature review, managing the project timeline, and preparing all deliverables in a professional manner. The teams are formed by the Clinic director and faculty advisors to balance student project preferences (as determined by a survey), background in skills needed for the project, grade point averages, student interest in serving as a team's project manager, and to avoid known personality conflicts. Clinic strengthens students' technical skills by exposing them to a complex, real problem. Additionally, Clinic builds professional skills by requiring students to communicate clearly both orally and in writing, manage a large project, and work effectively in a team.

One student on the team is appointed to serve as the team's project manager, acting as the primary point of communication between the team and the sponsor's liaison. The project manager also ensures that the team is making steady progress towards deliverables and intermediate deadlines. He or she does this in addition to technical contributions to the project alongside the other team members.

The teams are advised by a faculty member and by a liaison from the sponsoring organization. The faculty member serves primarily as a coach, mentoring the students on their research habits, team dynamics and communication. Because we feel that lessons learned from mistakes often have more impact than those learned from successes, the faculty advisor intervenes in the specifics of a project only when the team is heading far off course. For this reason, the faculty advisor need not have expertise in the mathematical methodology used in the project. The role of the sponsor's liaison is to provide domain expertise and context to the students throughout the year. The team meets weekly with the liaison by teleconference or Skype to share intermediate results and to ensure that the team's direction aligns with the project's goals. The liaison must also provide the team with data and background information in a timely fashion.

The Mathematics Clinic director and, in some years, an associate director oversee the three to five Clinic projects being run in the department each year. Their primary responsibility is to recruit projects of sufficient quantity and quality, working together with the other departmental Clinic directors and HMC's Director of Corporate Relations. The directors leverage the HMC alumni network and make site visits to companies across the west coast and, occasionally, other parts of the country. We charge a substantial fixed fee per project, which covers administrative costs, travel (for recruiting trips and for the teams to visit the sponsors), computing equipment and software, and other supplies. Because of the fee, we seek sponsors who are invested in the outcome of the project and, accordingly, we assign them the intellectual property rights to the completed work. We advise potential sponsors to propose projects whose results are not critically needed in the short-term, but the outcomes of which will be very useful to them in a few years' time. This mitigates some of the risk associated with relying on a team of undergraduates to complete a project that the company cares about. We also require the sponsor to identify a liaison who will be able to dedicate sufficient time and energy to the team to ensure that the team will have consistent access to contextual information needed to produce a useful product.

Clinic counts as a regular three-unit course in each semester, and we expect the students to devote approximately 10 h per week on the project. Of these 10 h, one hour is spent in a weekly meeting with the faculty advisor, one is spent in a weekly teleconference with the liaison, and one is spent in a weekly classroom session. The remaining 7 h are spent working on the project, and we strongly encourage the teams to schedule those 7 h as a team. We have found that the team can work more productively and resolve obstacles more quickly if everyone is in the same room working at the same time. Moreover, we have found fewer issues of students relying on their teammates to do most of the work when they are held accountable to the rest of the team for their work hours.

The weekly classroom sessions involve professional development workshops that we created as director and associate director and which continue to be used. The topics are effective team dynamics, teleconference and site visit etiquette, project management tools, and effective oral presentations. To keep the students engaged and help them retain the information, we structure these workshops as comical, improvisational skits. In these skits, faculty advisors portray students having traits and behaviors we have observed over the years as being detrimental to a team's success, such as interrupting teammates, arriving late to meetings, or not allowing every member of the team to contribute. The class discusses what they saw and what the characters should have done differently. The students enjoy the humor in a full panoply of bad behaviors parodied in a series of two-minute skits. At the start of the year, it is hard for them to realize how true-to-life these skits are; however, over the year, most of these behaviors do indeed arise on some of the teams, and the memory of the skits gives the students a starting point for addressing them. During the spring semester, Clinic teams from all departments (approximately 45 teams per year) assemble together for the weekly classroom sessions, which are dedicated to team presentations. Each team makes one 15–20 min presentation over the semester, allowing the students to review each other's work and practice giving oral presentations.

Imposing intermediate milestones and deliverables helps students manage the large, open-ended project:

- Clinic and Liaison Orientation: In early fall, liaisons are invited to campus to debrief the team on the context of the problem.
- Fall and Spring Site Visits: Each team travels to the sponsor's site early in the fall semester and again late in the spring semester. The teams gain insight into the context of their project and make presentations to the sponsoring organization.
- Marathon Push: In the first month of the project, teams devote time to the project beyond the required ten hours per week to immerse themselves in the project and write the team's Statement of Work (SOW).
- SOW: This document summarizes the team's understanding of the problem, literature review, proposed methodology, and timeline for the year. The liaison reads and approves the SOW in writing.
- Fall and Spring Presentations: In November, each team gives a presentation to the full group of Mathematics Clinic students about the progress they have made on their project. Throughout the spring semester, each Clinic team on campus makes a presentation to all Clinic students, faculty advisors and special visitors.
- Projects Day: In early May, all liaisons are invited to campus for our annual celebration of Clinic. Each team makes several presentations of their final project and presents a poster of their work in a general poster session. Receptions and a celebratory dinner conclude the event.
- Reports and Deliverables: At the end of the fall semester, the team submits a midyear report to the liaison. At the end of the academic year, the team delivers to the sponsor a final report, along with all software, computer code, data and other intellectual property of the project.

12.2.2 Recent Projects

The range of mathematical disciplines represented in the completed Mathematics Clinic projects over the past 40 years is breathtaking, including modeling pollutant transport in the atmosphere, fraud detection, and optimal control of satellite motion. Abstracts for all past Mathematics Clinic projects can be viewed on the Mathematics Clinic website (Harvey Mudd College Mathematics Clinic 2016). Although our Mathematics Clinic program does not recruit solely statistics projects, we have seen a rise in data-focused Clinic projects in recent years. For example, in the past 5 years (2010–2015), exactly half (11 out of 22) of our Mathematics Clinic projects (or those run jointly between mathematics and another department) fall in the category of data analytics, while in the preceding 5 years, that percentage was closer to 20 % (Harvey Mudd College Mathematics Clinic 2016), (Harvey Mudd College Computer Science Clinic 2016).

This shift towards data projects poses some challenges to our department. First, there is a sense of loss for the more traditional applied mathematics projects in the areas of fluid mechanics, differential equations, and linear algebra. While these areas of mathematics are still relevant, current sponsors are more interested in areas pertaining to analytics. This is especially the case for federal sponsors such as national laboratories, where funding that can be used to pay the Clinic fee is more plentiful in cutting-edge research areas like data science. A second problem is that some of our faculty feel ill-equipped to advise data-focused projects. Although the faculty advisor is not responsible for completing the project, many advisors prefer a project in their general research area. In the past, it was easier to match faculty advisors to projects in their areas of expertise. Third, for the authors, who serve as two of only a handful of statistical experts at HMC, the rise in data-related Clinics across campus results in a lot of ad hoc advising of Clinic teams on matters of experimental design and data analysis. This can take up a great deal of time on top of our regular teaching responsibilities, particularly during the spring semester when Clinic teams are completing their analyses.

12.2.3 Student Experience and Feedback

The Clinic experience is often a seminal one for students, motivating them and boosting their confidence before they venture into the "real world." At the end of each semester, students complete peer- and self-evaluations in which they reflect on the Clinic experience. In addition to showing appreciation for the technical knowledge gained in the experience, the evaluations invariably emphasize the professional skills gained and the students' satisfaction in having worked on a real problem for a real client:

> I have learned a lot about machine learning and software development, but also about documentation, working in a team, and research in industry. ... It has been fascinating to study mathematical ideas that are applied so readily to an industrial problem.

> I enjoyed in particular the aspects of heuristic design and adaptation, as well as the practice doing things like teleconferences, presentations, and reports on our progress, all clearly reflecting the sorts of work we can expect to do in industry.

The evaluations also commonly refer to the challenge of Clinic, and the satisfaction that arises from meeting that challenge:

> We've certainly run into our share of issues along the way, but I don't think the clinic experience would have been nearly as valuable if we didn't have those challenges to overcome.
>
> Our advisor held us to a very high standard, which was sometimes stressful but I think it encouraged us to work very hard and resulted in a final product which we are proud of.

Most often, student complaints about the Clinic experience stem from the difficulties in working with a team:

> I feel like our clinic team may have had a fair bit of trouble keeping on top of deadlines this year, but that we have produced a good result. I'm not sure why we were so often behind, but if I had to hazard [a guess], I would say communication troubles.

Despite the stress of Clinic, it is an experience that shapes a student's professional trajectory. Moreover, because our Clinic program is well-known, we often hear prospective students speak of Clinic as one of the deciding factors in choosing to attend HMC.

12.2.4 Client Feedback

It is not only the students who find the experience rewarding. Each year, we survey the liaisons, asking them how well the team met the project's goals, managed the project, and communicated with the liaison, and how they rate their overall satisfaction with Clinic. On a five-point scale, five being the best, the College scores higher than four, on average, on these questions.

Additionally, the Clinic Advisory Council, a committee of approximately 20 representatives from industry who have engaged with the Clinic program in the past, conducts phone interviews of all liaisons to discuss their experience with the program. Some comments from recent Mathematics Clinic liaisons are:

> I was very impressed by the level of enthusiasm and knowledge of the students. Very well done.
>
> [The liaison] realizes the quality of the "scarce resource" of [HMC students] and is intent upon identifying and hiring the best fits for his company, using the Clinic Program as a way to get to know the team members.

Of course, not every project is completed successfully. In some instances, the liaison and the team realize that the original proposal is infeasible or no longer in the sponsor's interests, so they work collaboratively to redefine the project scope. In other instances, however, the team simply does not meet the project's goals, usually because of poor team communication or project management. Fortunately, truly unsuccessful instances comprise only 5–10 % of all projects. One reason for our high success rate is that Clinic directors thoroughly vet projects in advance to ensure they are of an appropriate scope for the team. Another reason is that teams are encouraged to distinguish in their SOW between achievable baseline goals and "stretch goals" that will be completed if time permits. A third reason is that teams communicate regularly with the liaison so that expectations can be effectively managed if the project scope needs to be adjusted during the year.

12.2.5 Our Experiences

We have both served as faculty advisors of Clinic teams several times. Additionally, the second author served as Clinic director for 5 years, and the first author served as associate Clinic director for 1 year. In this section, we describe some of the personal impacts these roles have had on us.

As faculty advisors, we have found two primary challenges. The first is knowing when to be the "good guy" and when to be the "bad guy" in our team interactions. Generally, we try to observe from a distance to allow the students to direct the flow of the project. At times, however, when a team repeatedly fails to follow our suggestions, misses deadlines, or engages in unproductive habits or behaviors that are disrespectful to the liaison's time, we step in more assertively. The second challenge has been keeping the team and the liaison focused on the project scope. We have occasionally encountered liaisons that push students beyond what is expected from a three-unit course and have had to intervene on the students' behalf.

Being a Clinic faculty advisor sometimes forces us to step out of our comfort zones when the project ventures into an area of mathematical sciences we are less familiar with, but it is rewarding to broaden our knowledge. The opportunity to work directly with many of the companies that hire our students has provided us insight into the future needs and directions for mathematical sciences in industry. This allows us to develop new course materials that are relevant and cutting-edge for our students.

During our time administering Clinic as director and associate director, we developed valuable leadership and administrative skills, balancing the needs of students, faculty advisors, liaisons, and college administrators. The development of these skills was accompanied by a similar development of confidence. Directing the Clinic dramatically reduced time for scholarly activities such as research, but it was enjoyable work, from learning about the fascinating work done at prospective sponsoring organizations, to mentoring the student teams as they developed project management and communication skills.

12.3 Client Projects in Statistics Courses

Clinic is one model for offering statistics students client-driven projects at a departmental scale. On the scale of an individual instructor, final projects in beginning and upper level statistics courses can be used to gauge students' understanding and mastery of the material. One such course that began from this frame of reference was the HMC statistical linear models course (Math 158). I (the first author) based the course on the Guidelines for Assessment and Instruction in Statistics Education (GAISE) (American Statistical Association 2005), whose six broad recommendations include:

1. Emphasize statistical literacy and develop statistical thinking.
2. Use real data.
3. Stress conceptual understanding, rather than mere knowledge of procedures.
4. Foster active learning in the classroom.
5. Use technology for developing concepts and analyzing data.
6. Use assessments to improve and evaluate student learning.

When I first taught the course, I assigned a project based on an interesting dataset I could find and tested students on their ability to correctly analyze the data and present their results before their classmates. While this model proved sufficient in meeting course objectives, the GAISE guidelines, and student assessment, it lacked a connection to a data-driven client experience that would challenge students by placing them in unfamiliar territory. Lazar et al. (2011) have documented the improvement in the statistical analysis and quality of work of students who engage in a consulting-like experience. Much of the noted improvement is a result of incorporating real, messy data into projects and emphasizing the responsibility that students have to clients.

To give the students a more authentic data analysis experience, I restructured the course project to involve a local client with a real problem needing data analytics. In the remainder of this section, I describe that project. Section 12.3.1 describes its structure, Sect. 12.3.2 gives some examples of recent problem statements, Sects. 12.3.3 and 12.3.4 describe the student and client experience, and Sect. 12.3.5 provides some of my personal reflections and advice from having run such a project.

12.3.1 Structure of Community Client Engagement

12.3.1.1 Team Structure

In a typical semester of statistical linear models, students come from several of the Claremont Colleges with varying backgrounds. While an introductory statistics course is a prerequisite, the style of the course depends on the institution in which it was taken. Some courses are full semester and use the open-source statistical programming language R. Others are half semester and teach statistical analysis using Minitab®. This poses a challenge when constructing teams that must coalesce to effectively tackle a project and produce tangible results.

A solution I have found is to intentionally place students in teams based on their strengths. I give a survey to students at the beginning of the semester that asks for their previous computing experience, past statistics courses, and perceived strengths when working in teams. I use this information to create teams that have at least one strong computing person, a strong statistical person, and an effective oral and written communicator. While all students experience all aspects of the project, this distribution helps teams manage more independently by providing them in-house experts in areas of critical need. Teams are typically made up of three to five students, depending on the size of the course.

Once the teams are in place, they are responsible for scheduling additional time outside of class to work on the project together. All students are required to evaluate their teammates' contribution and I take the overall evaluation into account when determining the final project grade. I also try to be especially sensitive to women and underrepresented students by not placing them in groups where they are the only minority. Instead I place them on teams with at least two women, at least two minorities, or some combination of the two.

12.3.1.2 In-Class Lab Experience

I typically reserve two classes per month for an in-class lab where I present a new topic or type of data analysis that needs to be performed on their project dataset. Students bring their personal laptop to class and I reserve additional laptops for students that don't have one. The in-class lab allows me to observe how the teams are working together, gauge their progress so far and answer questions that often arise during their external meetings. I'm also able to observe and intervene if students are having difficulties or becoming disengaged. If the client is local, I often invite them to class to be available to answer questions and provide direction, especially during the first few lab sessions.

The structure of the in-class lab usually begins with stated goals for the session, for instance, to write an R program that will perform an exploratory data analysis and produce various plots. While the analysis of the data occurs over the entire semester, by setting up-front, measurable goals for the lab sessions, students understand that I expect them to have a deliverable by the end of class. To help novice R users get up to speed, it might sound counterintuitive that I tend to pair them together rather than pair a novice with a strong R programmer. However, except in cases where the strong R programmer is also a good mentor and tutor, pairing the novices together often leads to better collaboration.

12.3.1.3 Final Presentation Experience

Final presentations in semester-long statistics courses are commonly used to evaluate students' understanding of the material and their ability to communicate results effectively (Khachatryan 2015). What makes the final presentation experience particularly relevant in a client consulting environment is that students have to explain potentially complicated statistical analysis to people who are not experts in the field. Students learn that they must omit theoretical details and instead deliver results that are clear, concise, and visually appealing. The rich details of the analysis are included in a final written report that each team submits both to me and to the client.

The program and location for the final presentations can be used to create a meaningful closing experience for both the students and the client. Each team, dressed in business casual attire, presents its results and fields any questions that arise. For the past four years, we have had an on-site meeting at the end of the semester where the client invited all staff to the presentation and provided light refreshments.

12.3.1.4 Building Client Partnerships

A key aspect of the success of the course project has been the connection to the community-based client. In response to my solicitation to various constituencies around the Consortium for large-scale data analysis, Sam Kome, the Director of Strategic Initiatives and Information Technology at the Claremont Colleges Library contacted me. The library had large amounts of data and was seeking someone to analyze it.

The steps taken to build the community client relationship and provide a rich experience for students can easily be replicated in other data-driven courses:

1. Send an email to faculty and staff at your campus soliciting data.
   For example, an introductory statistics class could work with the cafeteria staff members to visualize demand by hour of the day, day of the week, or menu items.
2. Involve the community client in developing the project with the students.
   I found that staff members were excited to talk to the class about their data and project goals.
3. Be proactive in getting data at the beginning of the course.
   Students can immediately begin exploratory analysis and apply more advanced techniques as the semester progresses.
4. Create in-class lab days where students work in teams on the final project.
   While this required removing some lecture topics from the syllabus, the hands-on project time was an opportunity for the community client to attend class and answer individual teams' questions.
5. Encourage your client to remain available throughout the semester.
   Our client encouraged the students to contact him by email to review initial results or answer additional questions.
6. Set up a meaningful presentation experience.
   Presentations normally occur at the client facility or in a nice room on campus. Ample time is set aside for informal conversations, and light refreshments are provided.

Although the dedication of class time to the project reduces the number of topics we can treat in the course, the quality of the final project and presentation never ceases to amaze me. Giving students an unknown, unrefined dataset forces them to become researchers, ask their own questions, and go in multiple directions.

12.3.2 Recent Projects

In a recent project, the client provided patrons' library Wi-Fi usage data, and each team independently determined the type of analysis required to address the client's needs. Once teams formulated a project direction, they shared among themselves to prevent overlap, and with the client to verify that the analysis would meet their needs. Three approaches to this problem are summarized below as examples:

Team 1: This team established a heuristic for deciding between active and passive connection and defined a metric for wireless usage based on the data transferred and amount of time used. With this metric, the team determined how various factors, such as roles (student, faculty, staff), radio types, and signal quality impacted wireless use.

Team 2: This team performed a comparative analysis specific to Apple products, since they comprise the majority of products connecting to the Wi-Fi. They examined how Macs, iPods, iPads, and iPhones successfully connected to the network and the subsequent signal strength. They compared the performance activity common to each of these device types in terms of data usage and connection time.

Team 3: This team examined measures of connectivity success, including time spent connected, megabytes of data used, number of attempts made to connect, and signal strength, to determine the quality of the Wi-Fi connection.

Through these different approaches, the teams were able to visualize the ways in which various devices connected to the library Wi-Fi and better understand the duration of library visits by campus, day of the week and time of day.

12.3.3 Student Experience and Feedback

Through conversations with students and the written course evaluations I have received, students:

- Appreciated the active learning aspect of the in-class labs;
- Felt a sense of purpose by working with real data for clients who valued their solutions;
- Were genuinely surprised to see the excited reactions of the staff to their results; and
- Realized that there are many types of analyses that can be done on a single dataset.

One student sent an email the following semester stating:

> I just wanted to send you a quick note saying thanks for the statistical skills you helped hone in Math 158 last semester. I am doing an experiment for a cognitive science class and I've relied heavily on the techniques and tools you showed us throughout the class in order to analyze the results. It is very useful to be comfortable analyzing data, so thank you for imparting that ability throughout the course.

However, some students were frustrated by the open ended approach, the time it took to clean the dataset, and having to rely on team members who were less dependable.

12.3.4 Client Feedback

We have worked with the Claremont Colleges Library for the past five years and they have been very pleased with the partnership. Below is an excerpt from a letter that Sam Kome sent to the Harvey Mudd College Dean of the Faculty highlighting the experience from the Library's perspective:

> The wireless analysis identified and described a significant authentication issue that actually affected all the Claremont campuses. The library was disproportionately affected, and was able to use the analysis to adjust wireless provision and nearly eliminate complaints. We very much look forward to continuing this fruitful collaboration. ... We have found Professor Williams students' work to be thorough and thoughtful, and each presentation yields novel and directly useful information. We also appreciate that the students become deeply engaged with the data and they frequently express gaining a greater understanding of the complexity of today's academic library.

12.3.5 The Instructor's Experience

Creating an authentic data-driven experience required me to rethink my role in the classroom and allow my students to take ownership of the experience. I vividly recall the moment that students took charge of the project with the Claremont Colleges Library. I had invited the staff members to class to talk about the data. After the staff did a very brief presentation, there was an awkward silence as students looked to me for direction. I told them, "I'm not analyzing their data, so I don't have any questions for them. But in three months, you get to stand up and present your results in front of the entire staff. This is your time. We can take as much or as little of it as you need." This was the moment that the students took ownership of their project.

One closing thought on this experience would be the following:

> *Make the mundane meaningful.*

Although necessary, the process of cleaning real-world big data is often tedious and frustrating. Students frequently spend hours writing code to parse the data into a readable format to finally begin performing the statistical analysis. During one particular in-class lab day, two recent HMC alumni, Kyle and Russell, were visiting campus for a recruiting trip and asked to observe the class. The students were cleaning data that day and although I constantly reminded them that this is a necessary process of data analysis, they were obviously frustrated and disappointed with the task.

I asked Kyle and Russell if they would like to look at the data, and the two of them began cleaning the data as I continued helping students. After a short time, Kyle and Russell had written a script that read in the data, parsed it, corrected the formatting issues, and produced a multiple time series plot. I asked them to display their code on the screen and walk the students through their process. Russell told the class, "I wish I would have had a class like this back when I was at Mudd. In a typical five-day work week, we spend four days just cleaning the data. The stats is pretty easy after that." The mood of the class changed following their presentation.

As a professor, this moment was significant because it provided immediate purpose to a typically mundane classroom lab session by merging it with a real-life career experience. In every lecture, lab and presentation experience, students are learning statistical methodologies to prepare them for their careers. I now intentionally invite former students back to allow them to provide that rich perspective.

12.4 Looking Ahead

What does the future hold for statistics education? The need for computational and analytics skills to mine large data sets is expected to grow (McKinsey Global Institute 2011). Although HMC is unique amongst liberal arts colleges in that its only offered majors are in science, engineering and mathematics, we believe fervently that statistics is itself a liberal art (Moore 1998). It is no longer reserved for specialists, or even just for scientists and engineers. Rather, in the Information Age, it is imperative that all students learn to understand and critically interpret data put before them. Reflecting this trend, data science courses offered at HMC and other Claremont Colleges are routinely filled to capacity with students from many majors, and we are seeking ways to offer more data science electives while still maintaining our existing mathematics curriculum. Client-focused experiences such as Clinic and projects in upper-level statistics electives are one avenue to help prepare students to work effectively with data. We encourage our readers to consider how these types of experiences can be incorporated in their own curricula.

Acknowledgements The authors thank several anonymous reviewers for their thoughtful and detailed feedback that greatly improved this manuscript.

References

American Statistical Association. (2005). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report*. Resource document. American Statistical Association. Retrieved March 31, 2016, from http://www.amstat.org/education/gaise/GaiseCollege_Full.pdf.

Borrelli, R. (2010). The doctor is in. *Notices of the American Mathematical Society, 57*(9), 1127–1130.

Harvey Mudd College. (2012). *Harvey Mudd College faculty members share prestigious 2012 Bernard M. Gordon Prize*. Retrieved March 31, 2016, from https://www.hmc.edu/clinic/gordon-prize/.

Harvey Mudd College Computer Science Clinic. (2016). *Clinic projects*. Retrieved March 31, 2016, from https://www.cs.hmc.edu/clinic/projects/.

Harvey Mudd College Mathematics Clinic. (2016). *Projects indexed by year*. Retrieved March 31, 2016, from https://www.math.hmc.edu/clinic/projects/years/.

Khachatryan, D. (2015). Incorporating statistical consulting case studies in introductory time series courses. *The American Statistician, 69*(4), 387–396. doi:10.1080/00031305.2015.1026611.

Lazar, N. A., Reeves, J., & Franklin, C. (2011). A capstone course for undergraduate statistics majors. *The American Statistician, 65*(3), 183–189. doi:10.1198/tast.2011.10240.

Martonosi, S. E. (2012). Project-based ORMS education. *Wiley Encyclopedia of Operations Research and Management Science* (pp. 1–15). doi:10.1002/9780470400531.eorms1059.

Martonosi, S. E., Williams, T. D. (2016). *A survey of statistical capstone projects*. Manuscript submitted for publication.

McKinsey Global Institute (2011). *Big data: the next frontier for innovation, competition, and productivity*. Retrieved May 12, 2016, from http://www.mckinsey.com/business-functions/business-technology/our-insights/big-data-the-next-frontier-for-innovation.

Moore, D. S. (1998). Statistics among the liberal arts. *Journal of the American Statistical Association, 93*(444), 1253–1259. doi:10.1080/01621459.1998.10473786.

Olin College. (2016). *Senior capstone program in engineering (SCOPE)*. Retrieved March 31, 2016, from http://www.olin.edu/collaborate/scope/.