Datathons: An Experience Report of Data Hackathons for Data Science Education

Craig Anslow
Dept. of Computer Science
Middlesex University
London, UK
c.anslow@mdx.ac.uk

John Brosz
Libraries & Cultural Resources
University of Calgary
Calgary, Alberta, Canada
jdlbrosz@ucalgary.ca

Frank Maurer[1]
Mike Boyes[2]
[1]Dept. of Computer Science
[2]Dept. of Psychology
University of Calgary
Calgary, Alberta, Canada
{fmaurer,boyes}@ucalgary.ca

ABSTRACT

Large amounts of data are becoming increasingly available through open data repositories as well as companies and governments collecting data to improve decision making and efficiencies. Consequently there is a need to increase the data literacy of computer science students. Data science is a relatively new area within computer science and the curriculum is rapidly evolving along with the tools required to perform analytics which students need to learn how to effectively use. To address the needs of students learning key data science and analytics skills we propose augmenting existing data science curriculums with hackathon events that focus on data also known as *datathons*. In this paper we present our experience at hosting and running four datathons that involved students and members from the community coming together to solve challenging problems with data from not-for-profit social good organizations and publicly open data. Our reported experience from our datathons will help inform other academics and community groups who also wish to host datathons to help facilitate their students and members to learn key data science and analytics skills.

CCS Concepts

•**Social and professional topics** → *Computer science education; Information science education; Computing literacy;*

Keywords

Analytics, Community Engagement, Datathon, Data Science, Hackathon, Open Data

1. INTRODUCTION

Due to the ever increasing availability of data over recent years, acquiring key data science and analytics skills have become important areas of interest. The aim of data science is the extraction of knowledge from large volumes of data that are structured or unstructured [6]. There are many techniques of data science that require mastering such as analysis, capture, data curation, search, sharing, storage, transfer, visualization, and information privacy.

Many types of companies commonly use data science and analytics techniques to better understand the behaviour of their customers. Governments use analytics to better understand the impact of the policies they put in place. As companies, organizations, and governments begin to rely more and more on data for decision making the demand for data scientists will grow rapidly. A report by McKinsey Company stated by 2018, the US alone could face a shortage of up to 190,000 people with analytical skills as well as 1.5 million managers and analysts with the know-how to make effective data driven decisions [11]. Some columnists have even called data scientists "the sexiest job of the 21st century" [5].

Universities and schools are rapidly introducing courses and degrees on data science, however, they are still in their infancy and as of yet few students have degrees in this subject [15]. Some universities are establishing new interdisciplinary research institutes focusing on data science, for example in 2015 the Alan Turing Institute[1] was created which involves a number of leading UK universities. To address the challenge outside of academia many community based organizations are running events and focused groups to help people up skill and acquire key data science and analytics skills to complement their existing jobs.

We propose augmenting these practices and computer science curricula with hackathon events that focus on data also known as *datathons*. In this paper we present our experience at hosting four datathons over a two year period where students and members from the community came together to solve challenging problems with data from two not-for-profit social good organizations (SGOs) and publicly open data. The purpose of the datathons was for people to learn and put into practice new data science and analytics skills which they can then utilize for their studies and jobs. Another purpose lies in providing our partner SGOs that lack the expertise and resources to make use of data science and analytics, with data science knowledge and skills from within the community to better understand their data. This experience report about our datathons will help inform other academics and community groups who also wish to host datathons to help facilitate their students and members to learn key data science and analytics skills.

SIGCSE '16, March 02-05, 2016, Memphis, TN, USA
© 2016 ACM. ISBN 978-1-4503-3685-7/16/03…$15.00
DOI: http://dx.doi.org/10.1145/2839509.2844568

[1] https://en.wikipedia.org/wiki/Alan_Turing_Institute

2. RELATED WORK

Cleveland first coined the term data science in 2001 [4]. The purpose for doing so was to help expand the technical areas of the field of statistics focusing on the data analyst. These areas included multidisciplinary investigations, models and methods for structuring data, computing with data, pedagogy, tool evaluation, and theory. Our work focuses on multidisciplinary investigations and pedagogy.

Data science has since become an active field within many parts of academia. To meet the demands for data scientists by prospective employers in businesses and governments universities are rapidly establishing data science and analytics courses and degrees to address these demands [5].

Anderson et al. [2] report on an undergraduate data science degree at the College of Charleston started in 2003. The program was initially called Discovery Informatics but later changed to Data Science in 2012 to address the shift in focus within academia. Some of the lessons they learned include that the name change helped improve program awareness, the benefits of allowing students to take courses from multiple disciplines, and that introductory courses in data science helped with retention. They also get students to participate in an online data science competition. Others have explored computer science courses with a focus on information security and analytics [9], big data analytics [10], healthcare informatics [20], and data science for software engineers [12].

Howe et al. [8] question the role the database community has within data science. They propose that they should be one of the main groups leading the conversation about data science especially given their close relationship with database users. They pose a number of questions to consider focusing on what data science is, what level to teach data science in the computer science (CS) curriculum, which areas of CS and elsewhere should be engaged in forming data science, and how to get industry involved in the classroom.

With regard to data science curriculum a number of universities have explored creating non-CS courses that have a data science focus. Sullivan [17] created a course for non-majors that provides a data-centric introduction to computer science. Varvel et al. [19] investigated iSchools and Library and Information Science to see how current programs and courses supported data science initiatives. They identified 475 courses in 158 programs at 55 schools that concentrated on data topics in courses and categorized the courses into data centric, data inclusive, digital, and traditional. They found a relatively small number of schools offering programs specifically designed to educate data professionals, but expect that to grow rapidly.

Gil [7] proposes topics for a course on data science for non-programmers and non-CS majors. The course outlines lessons on parallel and distributed computing where students learn the principles and benefits of these areas by practicing with workflows to understand key concepts.

Plaue and Cook [14] report on lessons learned from an interdisciplinary service course on data journalism which was a collaboration between the CS department and the college of journalism at their institution. The course was well received by students from a variety of backgrounds.

Topi [18] argues that data science involves multiple disciplines not just computer science, mathematics, and statistics, but should include others such as information systems. Topi also notes that the ACM Education Council has a group exploring the feasibility of launching an initiative to create data science curricular and degrees.

3. DATATHONS

We propose augmenting the data science curriculum with practical experience by having students participate in data hackathons also known as *datathons* to learn key data science and analytics skills. A datathon is an event similar to a hackathon where people come together over a certain time period, commonly a weekend, to work on problems with a specific dataset. There are a number of steps involved in a datathon including preparing, planning, recruiting participants, arranging an appropriate venue, preparing and documenting the data, logistics, and hosting the event. We describe the benefits of datathons and outline our experience at running four datathons. Datathons provide unique opportunities for computer science students to acquire experience in working with data, participate in interdisciplinary teams, communicate their ideas effectively, apply (or gain) project management skills, and engage with community organizations to work on real-world data problems.

Data science skills. Data science and analytics skills are important throughout computer science, whether in working directly with data, handling the data associated with software development such as testing results, as well as in communicating data to others. Additionally, datathons expose students to real-world data and its inherent problems such as reliability, lack of documentation, unused data fields, privacy, anonymization, incorrect data, and unstructured data.

Interdisciplinary teams. Many areas of computer science practice rely on teams working with non-computer scientists. The most prevalent example is that of requirements gathering with clients for development projects. In other areas, such as in game development or tools development for scientists, computer scientists work in partnership with professionals from other disciplines. Exposure to such environments will help computer scientists learn to communicate their capabilities, learn processes for working with these other professionals, and will deepen their understanding of their profession's strengths and weaknesses.

Communication skills. Beyond the communication inherent in working with other disciplines, participants also gain experience in communicating the goals or benefits of their projects when demonstrating results.

Project management skills. As with any project, developing a deliverable application or data analysis report within a short period of time will provide a venue to apply the planning, documentation, communication, resource management, and related skills that students have been taught (or will be formally taught) as part of their CS program.

Community engagement. Another important aspect of datathons is that they provide a point of engagement for departments, faculty, and students to work with community organizations on real-world data problems. Such events draw the interest of local media as well as businesses looking to hire graduates or engage in research.

4. CASE STUDIES

In our case studies we discuss two types of datathons where we report on a total of four datathon events (two of each type) held over a two year period. The first case study is a very data-focused approach where participants work with an outside organization's data, providing expertise in working with the data, and helping the organization answer questions that they lack the resources and skills to answer themselves. The second case study is a slight variation on a hackathon where the goal is to form teams and then create an app that makes use of specific, often public, datasets.

4.1 Data for Good

Calgary Data for Good (DFG)[2] is a local organization inspired by DataKind.org and is working for positive social action through "data in the service of humanity." DFG brings together data scientists with social good organizations (SGOs) through a collaborative approach that leads to shared insights, greater understanding, and positive action with data. DFG creates a network of data scientists from the community and universities to inspire new applications of data science skills and analytics in meeting the needs of SGOs.

DFG organizes regular meetups and datathon events for its members. These datathons are somewhat different than a typical hackathon. Rather than creating a specific application, the goal instead is to help the participating SGO make better use of their data. Due to back and forth with the SGO these datathons are much less of a self-contained event than hackathons, requiring a great deal of collaboration with the organization, preparation of their datasets, and in developing reasonably scoped objectives that can be accomplished by the group over the course of the datathon. Our datathons featured 30-50 volunteer data scientists working with datasets over a weekend.

While requiring much more organizational effort the involvement of the SGO in the datathons brings many benefits. It exposes participants to real, messy data, as well as the back-and-forth partnership effort needed to work with a client. The SGO also brings their mission and values, encouraging participants who might not be motivated by other factors (e.g., prizes, experience).

Recruiting an SGO is not necessarily as difficult as one might imagine. A common scenario among SGOs (particularly not-for-profit organizations) is possessing large amounts of data but lacking the resources or expertise to make use of it. These organizations gather a great deal of data to support grant applications, reporting requirements, and in tracking operations. However, typical non-profit funding usually includes very little or no support for administrative overhead, such as data management and analysis. This problem noted by Daniel Perron, IT Analyst for the 2015 SGO, *"We don't have the resources in house to break down all of this data, we have the reports, we have the raw data, but that doesn't really tell us what the trends look like"* [13]. Consequently these organizations make minimal use of their data but are often keen to see the data used to improve their efforts and organizational practices.

Due to the differing backgrounds of the participants in the datathons it is not possible to provide all of the software they might make use of. However participants, being interested and active in data science, in most cases had their own copies of the software tools they preferred. These tools and languages included C#, Excel, Python, R, Ruby, Splunk, SPSS, and Tableau. Most participants brought their own laptops and other equipment.

A key element at the datathons is making data accessible. As participants make use of a variety of tools, having the data converted to as many different formats as possible is helpful. In our experience CSV, Excel, and MS Access formats were the most used. It was also important to ensure data is available through physical transfer (external drives: flash, HDD) in case of network or server problems.

Another organizational point is separating the participants into teams to work on problems for the SGO. Naturally this separation is neither hard nor fixed. Our approach has been to do this based on tasks/objectives.

In 2014 the participating SGO was Sustainable Alberta. They organize a yearly national event tracking commuters and were interested in improving the experience of participating individuals and teams. The tasks created to address the SGO's needs were: improving information and statistics about groups that have participated for several years; create dashboards for participating individuals and teams; and examining ten years of historic data for trends in participation as well as analyzing behavior by community and geography. Groups varied from 3-15 members and a selection of findings was published by Anslow et al. [3].

The 2015 SGO was the Calgary Distress Centre, a 24-hour support, free crisis counseling and resource referral service. The SGO's objectives were expressed as six questions and participants were instructed to join whichever group they wished. Volunteer participants from the SGO attended and spread out with at least one representative per group. While this may not have been necessary, it was very helpful in obtaining on-the-spot context for the data. The questions were: does call content depend upon the time of day, what issues do Distress Centre personnel most successfully handle, do calls related to particular concerns change over time; are there monthly or seasonal patterns, are there correlations between call data and public datasets such as suicide rates or weather records, does interactive mapping show patterns in 211 informational calls (a separate local information service), and what is the difference in interactions between online chats and SMS texts? A summary of the findings has been compiled by DFG [1].

The creation of these tasks/objectives and preparing the data occurred through pre-event preparation. The organizers met several times with the SGO to ensure data was available, sufficiently anonymized, and documented. A larger meeting occurred two weeks in advance with a group of ten experienced participants. The data and objectives were described to this group and then were revised before the datathon based on the group's feedback. This feedback was essential in ensuring objectives were appropriately scoped and potentially answerable as well as that the dataset was sufficiently structured and documented.

The weekend of the years' events started Friday night with an introduction both in terms of motivation as well as a technical introduction to working with the dataset from SGO personnel (see Figure 1(a)). The remainder of the evening was social time for people to meet one another, ask questions, and start planning for the next day. Saturday started

[2] http://www.meetup.com/Data-for-Good-Calgary/

(a) Introductory social event. (b) A group exploring a dataset. (c) A group brainstorming together.

Figure 1: Datathons - illustrating different aspects of the datathons.

early; objectives were introduced, data access was provided, and groups were formed. People worked throughout the day (Figure 1(b)). Sunday consisted of a little more work; mostly time to gather results and assemble a presentation. Then each group presented their findings to the entire groups as well as the SGO attendees.

Feedback from both the SGOs and participants has been positive. The 2015 SGO noted that they have obtained several results: data is now shaped and formatted in a way that it can be more easily used and analyzed, a number of strong but preliminary results have been found that need to be validated; their organization can better perceive the value and power of data; they were introduced to a variety of useful tools; and they were connected to a group of knowledgeable volunteers [1].

Participants appreciated the experience, interfacing well with people from different disciplines: *"On Saturday morning, our team began putting that game plan into action, which proved a lot more difficult than anticipated. Different backgrounds meant we had different approaches to the data, so the first few hours were spent trying to get everyone's ideas heard, while simultaneously deciding how to modify the data into an appropriate format for the analyses. After lunch, everyone settled into a groove, talking quietly in nuclear groups and concentrating on their assignments*" [16]. Summarizing the experience Sabine Syeffarth also noted: *"I would highly recommend that anyone who has an interest in working as an analyst or data scientist take part in events like this. Apart from the great feeling you get by making a difference, it is also great to see how your own skills measure up against professionals. The benefit of the datathon is you can connect with people who have similar interests, learn new tools, and all in all, have a fun - albeit stressful - weekend!"* [16].

4.2 Canadian Open Data Experience

Data-focused hackathons are a popular means that local, regional, and federal governments are actively sponsoring to gather interest in open government data initiatives. These events aim to create apps that make use of recently released datasets. The events generated interest among developers and technology innovators in government datasets with the hope of creating useful and popular applications that will draw public interest and attention to these datasets.

Our involvement has been with the Canadian Government's Canadian Open Data Experience (CODE)[3] event. This event encourages developers to create applications using the Canadian government's open data portal[4]. While self-organized teams were free to explore the data ahead of the event; the event itself consisted of a 48-hour coding sprint at the start of which the desired application theme(s) were announced, and then the apps were developed.

We organized two hubs as part of CODE in 2014 and 2015. The 2014 theme for apps was "solving problems and increasing productivity through the use of open data." In 2015 there was the option of working towards one of three themes: youth employment, business opportunities, or healthy living. After the event a shortlist of fifteen finalist teams from across the country was created from which prize winners were selected: one for each application theme, a fan favourite, and a best student team. Participants had the option of working entirely on their own or at organized hubs where multiple groups worked in the same space.

In 2014[5] sixteen participants created four teams (Figure 1(c)). One team's app was selected as one of the 15 national finalists and presented their app for judges at the finals in Toronto. In 2015 fourteen participants formed five teams.

Organizationally these events are straightforward to organize as the data organization, event goal, and prizes were provided by the government event organizers. For our institution the key organizational elements were to locally publicize the event, recruit participants, and secure an appropriate space. The group structure and technologies were dependent on CODE's specifications. As this event was a competition we found it useful to provide programming both before and after to assist in group formation and to encourage post-event discussions on what they worked on.

The apps the teams built during CODE had to utilize the open data portal and many of them utilized visualization techniques and made web apps. One team built an app that explored which city is best in Canada based on economy, housing, weather, and crime data. One app looked at where crashes happen on roads in Canada and the best time to travel to avoid bad spots. Another app explored how not-for-profit organizations spend their funds. Finally, one app explored election data in the province of Quebec. All the students reported that it was a great activity to participate in, and one team stated: *"The CODE competition was in the core of what we believe in. It feeds into our skills as computer scientists in the sense that using data for the benefit of Canadians is the group's philosophy. It was a very good learning experience and it was a true measure of character, and we really pushed ourselves."*

[3] https://www.canadianopendataexperience.ca/

[4] http://open.canada.ca/

[5] https://www.youtube.com/watch?v=egw39gz9-tc

5. DISCUSSION

There are a number of aspects that we would like to draw upon based on our datathon hosting experience.

Interdisciplinary Teams. It is important to recruit a diverse set of participants in the datathons as working on data science teams requires interdisciplinary skills. It is critical that each group has access to people with different skill sets in order to solve the questions that are raised using different tools. Recruiting people beyond universities from community groups and other technical and user groups is important as it brings in fresh ideas, processes, and tools.

When recruiting it is important to emphasize that these events are learning environments and that prospective participants need not be experts, that whatever data skills they possess will be useful. To provide an anecdote from the 2015 datathon, one participant was worried about attending as they lacked statistics and visualization knowledge; however within their group this participant was able to perform that low-level parsing of the data that had the statistics and visualization experts in the group stumped. It is also important to ensure a core cadre of participants who have strong skills in certain areas as other participants can learn from those people with the specific skill sets such as statistics, data visualization, or data mining techniques. In the datathons we tailored several of the objectives to align with the known expertise of the groups.

Group Work. When forming groups it is important that participants are self-organized and that participants feel free to join and work in any group. Since these events are voluntary people are giving up their time to work on problems they are interested in. Hence if they are forced to do something they are unlikely to be engaged and go home early. However, it is possible to guide participants towards groups at the beginning by outlining the data science skills required to answer the questions associated with each task.

Data. A critical element for running a datathon is the data itself. The data that is made available during the datathon must be easily accessible for the participants. We found providing data through an internally hosted server on a database to be successful. Using cloud services such as DropBox allowed snapshots of the data to be available and made it easy for participants to share their findings. In order to make best use of the data it needs to be well structured, cleaned and anonymized. We found working with SGOs to accomplish this task requires significant time up-front as they were not used to doing these kinds of tasks with their data. Access and permission to the data must be made clear at the beginning of the datathon so participants know exactly what they can and cannot do with the data during and after the datathon. When working with open data as part of the datathons it was important that the students understand how to make best use of the data and what steps were involved in processing the data. In the events with open data there were fewer concerns as to what the participants did with the data post-datathon.

Goals. We found that for the datathons to be effective it was imperative that the SGO had specific goals up front at the beginning of the event. The goals that were defined were mainly geared towards the needs of the SGO, but it was also important to have realizable goals for the participants so they were able to accomplish significant results within the limited time of the event. The tasks that were selected helped the SGO meet their initial analytical enquiry goals. We found that both of the SGOs that participated were very happy with the results achieved.

Tools. We experienced that participants used a variety of tools to solve the problems as part of the datathons. Some used statistics focused tools such as R, SPSS, and MATLAB. Others used some visualization tools such as Tableau and Excel while participants who were developers used toolkits like D3 to develop visualizations and JavaScript UI frameworks to build dashboards. As part of the datathons we helped participants to learn some of these tools. To aid with tool learning we had break off groups run some mini tutorials to help others get up to speed with the tools.

Community Engagement. It was critical for the success of our datathons to partner with not-for-profit SGOs. Without these organizations participants would not have experienced real-world data problems when working with large and complex data sets. By working with these SGOs we had members from these organizations who were onsite during the datathon to help participants with all the specific questions they had about the data. This was also a profound effect for the members of these organizations who got new insights into how data scientists work with their data and the kinds of techniques they use in order to address the answers to the questions they seek.

Logistics and Planning. Particularly when working with SGOs, we found it very helpful a small group of data experts vet the data organization and documentation as well as the planned questions and tasks for the groups to ensure everything was on track. We began the majority of advertising and recruiting notices a month in advance through newsletters, posters, mailing lists, and so forth. Where possible we posted a reminder the week of the event. We found the best way to encourage participation was to have a recruitment event a couple of weeks prior to introduce the datathon. Having an introductory social evening the night before the actual event helped to set the stage of the datathons. During this introductory event the problems and objectives of the datathon were laid out to all participants which gave a useful overview on what needed accomplishing. This was a time for networking, allowing participants to start learning about one another and begin considering what they will do during the datathon. These events also provided participants the opportunity to ask all the questions they wanted to know before the datathon actually happened and for them to see what role they could play in making the datathon a success. The actual day of the event started with a short invigorating session to get the participants up to speed and then immediately working in the groups. We had checkins at the end of the day where teams reported what they had been working on. On the final morning we had all teams come back and give a formal presentation to the rest of the participants to demonstrate what outcomes they had achieved, and providing a cap-stone to the weekend. During the event it is important to have breaks for food. Given that people were volunteering their time, we found it more fitting to find a sponsor willing to donate some money towards food rather than having a user pay system. Participants felt they were more valued if there was someone else providing food.

Venue. We hosted our datathons in two locations; one in the department of computer science and the other in the library. From our feedback having the right venue makes the event more effective. If the room is not appealing or supportive of working in groups participants are less motivated to work. We found that it is useful to have a main room where everyone is welcome to hang out, network, and work together. This room was used for the introductory and presentation sessions. The room was configurable with chairs, tables, and several whiteboards for note-taking and planning. It was important that there sufficient power outlets and extension cables for participants to plug their equipment into. Given the nature of the datathon we found that groups needed space to have discussions, so we set up a number of break out rooms where smaller groups of people could work more closely together. We decided to use an adjoining room that had a kitchen for the food room where participants were freely available to come and go as they wanted. This allowed for a smooth transition from working on the datathon to some down time to take a break.

Media. We had a number of media agencies interested in the datathons which helped raise the profiles of some of the SGOs. It is important to get a representative of the group to speak and liaise with the media to make sure the right messages are being delivered about the importance of learning data science and analytical skills.

Things to Avoid. We that felt unlike many hackathons that it was best not to go all night; by making these events more compatible with travel times and family commitments we believe we were able to attract more diverse and experienced participants. We found that it was best to not group all technical people into one team and instead spread specialists out so that others can learn from their expertise. Rather than prescribe what participants worked on during the datathon we found it more effective for participants to have the freedom to work on what they wanted to. Lastly, for one of the early datathons we had a problem with all the data being on one server and experienced a single point of failure problem. Hence we recommend providing different aspects of the data from different servers as well as using cloud services such as DropBox when possible.

6. CONCLUSIONS

As data continues to grow within our lives, businesses, organizations, and governments there is a greater need to make sense of all the data. With this increase in data there is a need for people to learn and up-skill themselves with key data science and analytics techniques and tools to make better informed decisions. Universities are beginning to address this challenge by setting up new courses and degrees to support the new area of data science and analytics. We agree with universities diversifying their programs, and we encourage students and those who want to up-skill themselves to participate in data hackathons to get hands on experience with real-world data problems. In this paper we reported our experience at hosting four datathons that involved students and members from the community coming together to solve challenging problems with publicly open data and data from not-for-profit organizations. Our reported experience will help other academics and community groups who also wish to host datathons to promote data science.

Acknowledgments

Thanks to Data For Good Calgary (and Geoff Zakaib) for co-organizing the datathon. Many thanks to Sustainable Alberta Association and the Calgary Distress Center for their participation. This research was partially funded by Mitacs and supported by Project VALCRI, European Commission Grant FP7-IP-608142.

7. REFERENCES

[1] T. Al-towaitee. Distress center datadata technical summary. Data for Good Calgary Meetup Website, August 2015. http://goo.gl/7ciZnH.

[2] P. Anderson, J. Bowring, R. McCauley, G. Pothering, and C. Starr. An undergraduate degree in data science: Curriculum and a decade of implementation experience. In *SIGCSE*, pages 145–150. ACM, 2014.

[3] C. Anslow, B. Jackel, K. Mehmood, P. Fairie, A. D'Souza, M. Underwood, and K. Teh. CommuterVis: Towards understanding commuter behaviour. In *VIS Workshop on Business Visualization*. IEEE, 2014.

[4] W. Cleveland. Data science: An action plan for expanding the technical areas of the field of statistics. *ISI Review*, 69:21–26, 2001.

[5] T. Davenport and D. Patil. Data scientist: The sexiest job of the 21st century. Harvard Business Review, October 2012. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century/.

[6] V. Dhar. Data science and prediction. *Commun. ACM*, 56(12):64–73, Dec. 2013.

[7] Y. Gil. Teaching parallelism without programming: A data science curriculum for non-cs students. In *Workshop on Education for High-Performance Computing (EduHPC)*, pages 42–48. IEEE, 2014.

[8] B. Howe, M. Franklin, J. Freire, J. Frew, T. Kraska, and R. Ramakrishnan. Should we all be teaching "intro to data science" instead of "intro to databases"? In *SIGMOD*, pages 917–918. ACM, 2014.

[9] S. Kumar. Designing a graduate program in information security and analytics: Masters program in information security and analytics (MISA). In *SIGITE*, pages 141–146. ACM, 2014.

[10] A. Mahadev and K. Wurst. Developing concentrations in big data analytics and software development at a small liberal arts university. *J. Comp. Sci. Coll.*, 30(3):92–98, Jan. 2015.

[11] J. Manyika, M. Chui, B. Brown, J. Bughin, R. Dobbs, C. Roxburgh, and A. Hung Byers. Big data: The next frontier for innovation, competition, and productivity. McKinsey Global Institute, May 2011. http://www.mckinsey.com/insights/business_technology/big_data_the_next_frontier_for_innovation.

[12] T. Menzies, E. Kocaguneli, F. Peters, B. Turhan, and L. Minku. Data science for software engineering. In *ICSE*, pages 1484–1486. IEEE/ACM, 2013.

[13] M. Modjeski. Data for Good and Calgary Distress Centre to turn raw information into efficiencies. Metro Calgary, May 2015. http://goo.gl/OhoY23.

[14] C. Plaue and L. Cook. Data journalism: Lessons learned while designing an interdisciplinary service course. In *SIGCSE*, pages 126–131. ACM, 2015.

[15] R. Schutt and C. O'Neil. *Doing Data Science: Straight Talk from the Frontline*. O'Reilly Media, Inc., 2013.

[16] S. Seyffarth. Behind the scenes of a datathon! Cybera Tech Radar Blog, June 2015. http://goo.gl/m3QcMn.

[17] D. Sullivan. A data-centric introduction to computer science for non-majors. In *SIGCSE*, pages 71–76. ACM, 2013.

[18] H. Topi. Data science and information systems: Relationship of love or hate? *ACM Inroads*, 6(1):26–27, Feb. 2015.

[19] V. Varvel, E. Bammerlin, and C. Palmer. Education for data professionals: A study of current courses and programs. In *iConference*, pages 527–529. ACM, 2012.

[20] G. Zheeng, C. Zhang, and L. Li. Bringing business intelligence to healthcare informatics curriculum: A preliminary investigation. In *SIGCSE*, pages 205–210. ACM, 2014.