Professional Development for High School Computer Science Teachers on Data Science through a Feminist Lens

Anna Baynes
Department of Computer Science
California State University, Sacramento
shaverdian@csus.edu

Aaminah Norris
Department of Teaching Credentials
California State University, Sacramento
aaminah.norris@csus.edu

***Abstract*—This Research-to-Practice Work in Progress paper presents a recent professional development for high school computer science teachers. The court system increasingly relies on algorithms to rate a criminal defendant's likelihood of recidivism. These algorithms utilize machine learning models tuned on biased datasets, resulting in false assessments. There are several other alerting examples of how a software engineer's personal bias affects their design and development of technology. This bias leads to poor usability of technology for underserved groups. Data science applies machine learning on massive datasets to invent new technology to improve everyday life from healthcare to consumer needs. Designing and developing data science solutions for a diverse population requires an unbiased approach. Data feminism, which emerges from the intersecting fields of data science and feminist theory, presents a methodology to solve data science problems given three principles: 1) invent new ways to represent data unknowns, 2) invent new ways to reference the material economy behind the data, 3) make dissent possible. In this paper, we present a case study on a week-long online professional development for computer science high school teachers on data science using a data feminist lens. We discuss the data science tools and technology taught during the weeklong event. We share examples of the data feminism principles and class exercises. The computer science teachers also participated in community building activities. The culminating weeklong challenge was to analyze an earthquake dataset, which included themes of socioeconomic disparity within a city. The teachers worked in teams to explore this dataset with the new data science techniques and recommend resource management solutions for a city's earthquake response. In this paper, we explore the question: how the participating computer science high school teachers incorporate data feminism into their data science pipeline? We analyze the audio transcripts collected during the week to create visual analytics on our findings. We consider the impacts of culturally responsive approaches to computer science K-12 curriculum and future professional development in this space.**

I. Introduction

With the expansion of computer science K-12 education requirements, the need to further train and support teachers in subject matter material increases. Teachers often enroll in professional development courses during their free time as a way to learn new computer science material. Faculty from the departments of Computer Science and College of Education at Sacramento State University partnered with Elk Grove Unified school district to secure a broadening impact project which offers professional development to computer science teachers as a means to explore recruitment and retention of girls in their classes. The participating teachers were all high school computer science teachers with various backgrounds in computer science. We designed and developed a weeklong summer intensive fully online professional development for the participating teachers on data science. Data science is an integrates skills from various fields in computer science, statistics, design, and more. During our professional development, we taught the teachers an introductory background on data science including the analytical pipeline of data collection processing, analysis, and data visualization. We trained the teachers to use a new data wrangling tool to analyze data sets. Beyond the computational aspects of data science, our goal was to understand this topic through a data feminist viewpoint.

We presented the teachers with the data science feminist principles proposed in [15]. The authors in [15] describe how data feminism aims to alter the traditional approach by following the three tenants summarized: 1) invent new ways to represent data unknowns, 2) invent new ways to reference the material economy behind the data, 3) make dissent possible. Data science is riddled with uncertainty in every stage of the analytical pipeline. The first tenant warns the data scientist to be cautious of these uncertainties and their effects. The second tenant notes that there are various stakeholders and the important to understand each voice in an application surrounding the dataset. The third tenant challenges the data science to present interactive visualizations beyond filters and sliders so that the user can question the facts and view different realities.

We designed the professional development week to teach data science by applying these three data feminist principles. We do not propose new data feminist practices, instead we incorporate the principles in [15] into our lesson plan. Along with the tools and techniques taught, participating teachers read and discussed data feminist approaches to solving problems. On the third day, we presented a data challenge based on the IEEE VAST 2019 Challenge scenario and dataset. The scenario includes a fictional scenario of using social and

seismic data sets of over 80,000 rows to help town officials to determine the extent of a recent earthquake's damage and dispatch the appropriate resources. The IEEE VAST Challenge included several metadata describing the town, neighborhoods, and pre-earthquake issues in town. The challenge is to look beyond the numbers and to understand the socioeconomic diversity in the different neighborhoods in town to produce a data visualization to guide officials on how to help the town.

We reviewed audio zoom recordings of the daily meetings to see how participants apply data feminism principles in their discussions of solving the data science challenge. We incorporated these transcriptions into a timeline data visualization to see the progression of the participants using data feminism. We analyzed how the participants use the three data feminism principles in their discussions. In this paper, we present related work, the daily professional development agenda, the problem statement proposed, and the data visualization designed to study our problem statement.

II. Related Work

The professional development week weaved in topics of data science education, data feminist viewpoints, culturally responsive learning communities, and collaboration on goals to recruit girls into computer science courses. In this section, we share background related to these related fields. There have been several supportive studies on the importance of professional development offered to computer science teachers to learn new lesson plans [1], [5], [4] and [2]. Professional development offered to teachers directly impacts their students. For example, [3], [23], and in [8] summarizes several computer science professional development programs intended to increase access to curriculum material and lesson plans. These professional development programs are becoming an extremely important need as computer science becomes a fundamental core subject in k-12 education. In our professional development week, we brainstormed several community building activities to learn and share the participants' cultures. Integrating culturally relevant pedagogy with computer science is an emerging area to improve computer science education for underrepresented students. In [6] and [7] authors explore case studies of practical approaches to culturally relevant pedagogy professional development for high school computer science teachers.

Integrating culturally relevant applications in class is challenging when computer science teachers are already burdened with busy schedules in teaching introductory programming, data structures, and algorithms. Data science is a field which applies computer science solutions to real world issues. Next, we review relevant curriculum and lesson plans in data science. Exploring Computer Science (ECS) [9] is a widely used K-12 computer science education curriculum which engages students in computational thinking. There is an increased effort to educate K-12 students on computer science topics [15], [24]. Another example of a data science workshop for students is [11], where students work on data science examples. Given the data-driven world, data science skills are in demand.

Data feminism [22] includes several principles on how identity, power, and knowledge is created and the effects of gender, race, and class [19]. Feminist theory and increasing female participation in computer science are orthogonal topics. We reviewed recent works which studied these topics. For example, [18] presents a computer science camp for middle school girls with female role models, which saw an increased participation in computer science courses. In this paper, we do not study the results of the teachers returning to the classroom after the professional development. In [17] authors discuss the importance of user centric design, but admit the pitfalls of the current industry where within organizations, computer scientists and engineers control technology design decision. In the professional development, we trained the teachers to ask about who collected the data, why the data was collected, and who were the major stakeholders prior to examining the actual data [19]. Data feminism encourages pluralism and listening to multiple voices instead of one loud voice [14]. This approach leads to data scientists investigating the sources of uncertainty in a dataset. The ethical impact of computing is a core computer science topic taught in high school programs. The data feminist principles of uncertainty, bias, and dissent parallel computer science ethics subjects, and are explored in several related works [25], [26], [27], [28].

III. Daily Agenda and Activities

In this section we layout the daily two-hour agenda of activities and lessons. Prior to each meeting, the research team planned activities ranging from data science, community building, and discussions. The participants included a five-person research team from Sacramento State University, two administrators, and the eight high school computer science teacher participants from Elk Grove Unified school district.

A. Day 1

On the first day we included a community building activity around leveraging oxytocin through collaboration in this Research Practice Partnership (RPP). We discussed how to humanize the RPP to develop effective methods for cultivating a caring, supportive, culturally responsive, and rigorous professional learning community. We discussed techniques to bring community guidelines to the classroom including using gender-conscious language, communicating needs, setting norms, and being responsible. We focused on community building the first day to alleviate the stress and pressure of an intensive professional development week. Next, we spent 60 minutes on an introduction to Data Science. Teachers were introduced to the definition, motivation for the field, and learned about the Data Science Pipeline of collecting, processing, analyzing and representing a data set.

B. Day 2

Day two's lesson was on data cleaning [21] and discussing the data feminism principles from [15]. Data is usually collected in ways that causes uncertainty, missing values, unknown stakeholders, and more. We introduced the teacher

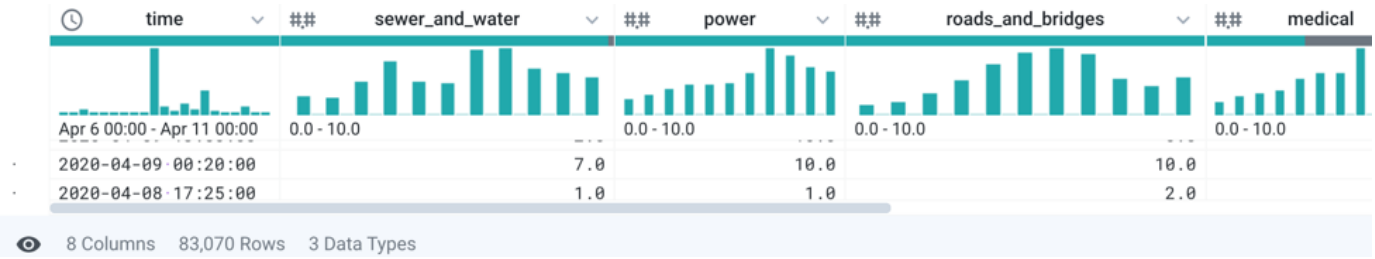

Fig. 1. Trifacta visualizing challenge dataset

participants to a data wrangling tool called Trifacta Wrangler [20], which uses a familiar spreadsheet style combined with visualizations and smart suggestions of data cleaning operations. The teachers used a sample small weather data set to learn about some basic data cleaning operations. Next, we discussed the data feminism principles and integrated it into the data cleaning lesson to help the teachers understand the principles.

C. Day 3

On day three, after a community building exercise, we presented on the value of data visualization. We showed examples of different data visualization in recording information (e.g. blueprints, seismographs), analyzing data to support reasoning (e.g. finding patterns and discovering errors in data), and to communicating information to others (e.g. share and persuade). Then, we discussed the IEEE VAST Challenge and its fictional scenario. We opened the data set in Microsoft Excel and quickly realized that the data set of more than 80,000 rows is unmanageable. Therefore, we used Trifacta show visualizations of the summary statistics and helped the teacher participants to understand the data set. We discussed the meta data and then discussed the challenge question: Identify the top neighborhood the officials should help and why. For the rest of the day, we split the teams into four groups which included a member of the research team with the groups of the practitioners (e.g. teacher participants and administrators). Each group spent time using Trifacta Wrangler to solve the data challenge and begin to create visualizations to support their decision. We asked each team to particularly focus on incorporating the data feminism principles from [15] into their challenge solution. Figure 1 shows the screen of a teacher participant using Trifacta Wrangler to analyze the challenge dataset. One of the first data uncertainties noticed in the dataset was that in the Shake Intensity column there are missing values. A participating teacher asked if a missing value meant zero or if it had a different meaning. We connected this question to the first data feminist principle on data uncertainty.

D. Day 4

On day four, after a quick community building activity, the group split into their teams from the previous day to continue to solve their data challenge. There was an ambiance of a hackathon to build a solution in teams. The research team visited each team and helped facilitate the discussion and answer roadblock questions. The teams then presented their findings. Some teams created a compelling story about the different neighborhoods and some teams followed an analytical approach with spreadsheets and graphs. Overall the final decision on which neighborhood to help was unanimous.

E. Day 5

Day five was reserved for discussions on data feminism. We shared ideas on how teaching with a data feminist approach is different from the traditional way computer science is taught. We discussed how this related to what the participating teachers currently teach in their computer science courses.

IV. Problem Statement and Data Analysis

The intensive weeklong professional development generated many insightful discussions. We recorded each zoom session and reviewed and edited the audio transcripts. The dataset consisted of only textual data of the recorded transcripts, where each row of data is tuple of the speaker and their transcribed words. We used this dataset to create the timeline shown in Figure 2. The timeline shows the daily progression of how participants applied data feminism to solve the data science challenge with the following problem statement: How did the participants utilize the three data feminism principles in their discussions?

We transformed the dataset to build the timeline visualization. First, we reviewed the transcripts of each day and categorized key phrases on data science related discussions into the three principles of data feminism: 1) invent new ways to represent data unknowns, 2) invent new ways to reference the material economy behind the data, 3) make dissent possible. We also annotated each phrase based on who said it (i.e. research team, administers, or participating teachers). We combined the dataset into a timeline visualization to see which principles are covered, by which group of individuals, and how the week progresses. Figure 2 shows the visualization summarizing the weekly progression.

We included only day three through day four when we began working on the data challenge. The visualization shows that in the beginning of the professional development, the research team mostly applied data feminism in their discussions because we were in the initial process of the training and introducing the data science concepts. We see on day four the

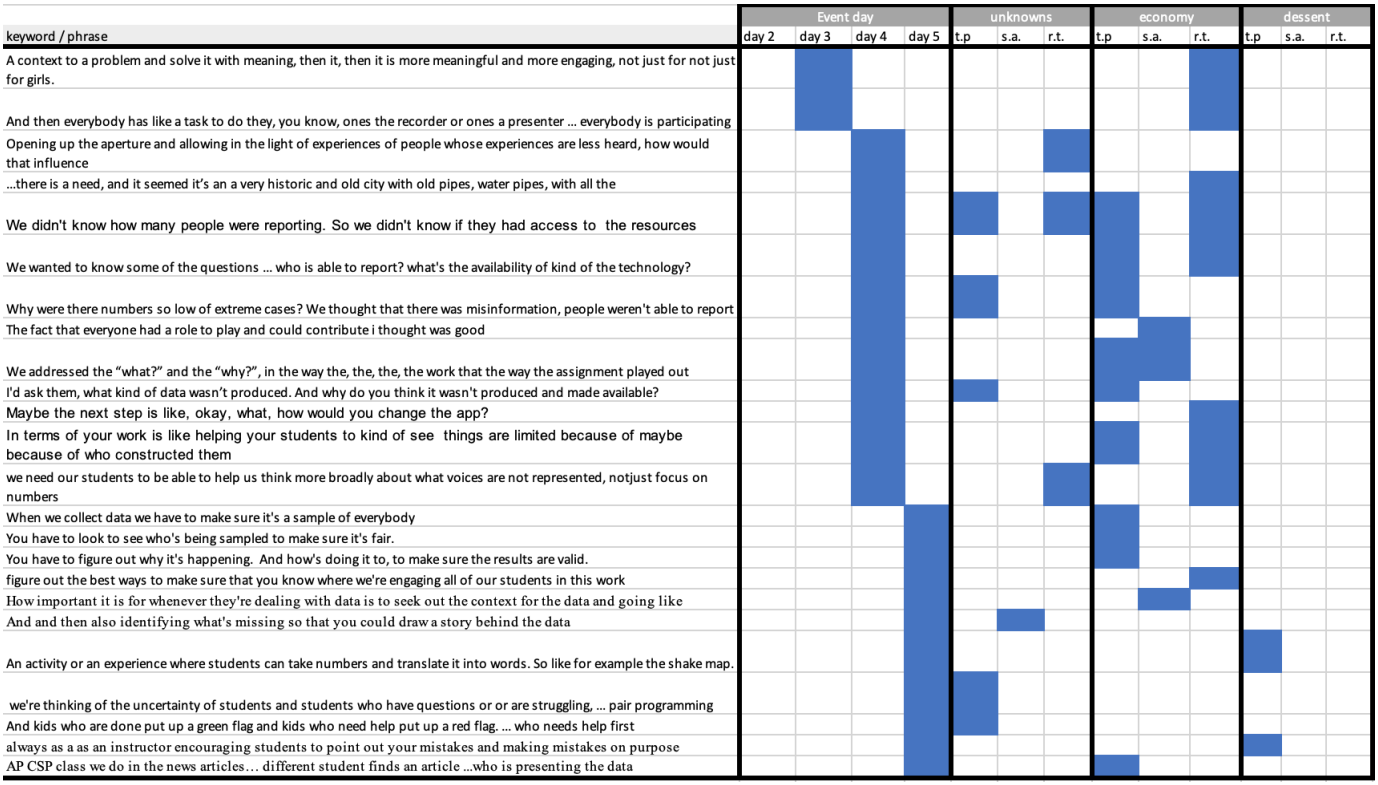

Fig. 2. Timeline visualization of categorized text. T.P., S.A., R.T., represent teacher participants, school administrators, and research team respectively.

participating teachers use the data feminism principles to describe their data challenge solution. For example, "Why were there numbers so low of extreme cases? We thought that there was misinformation, people weren't able to report," shows an application of the first principle, representing unknowns. We noticed that the least applied data feminism principle was the third principle on making dissent possible. The reason the third principle may not have been applied as much as the other two is because we did not have enough time in the week to re-evaluate our analysis and think about ways to deconstruct our thought process to think about other solutions. Overall the results are positive that the participating teachers were able to understand the data feminism principles and apply it to a new data challenge during the intensive five-day professional development.

V. Conclusion

In this paper, we present our findings after an intensive professional development week designed to teach Data Science to high school computer science teachers. In the lessons, we included readings and discussions on data feminism approach to data analysis. We present a data visualization of the transcriptions to see the progression of the participants using data feminism. We analyzed how the participants use the three data feminism principles in their discussions. We see that the participants began to see the importance of these principles. In future work, we hope to study correlations between different approaches to studying data science and increasing the inclusion of females and underrepresented groups in the computer science classroom.

Acknowledgment

We gratefully acknowledge our participants for engaging in this professional development. This work was supported by a grant from the National Science Foundation (1923602).

References

[1] Hur, Jung Won and Andrzejewski, Carey E and Marghitu, Daniela, "Girls and computer science: experiences, perceptions, and career aspirations," Taylor & Francis, vol. 27, pp. 100–120, 2017.

[2] Ravitz, Jason and Stephenson, Chris and Parker, Karen and Blazevski, Juliane, Early lessons from evaluation of computer science teacher professional development in Google's CS4HS program, ACM Transactions on Computing Education (TOCE), vol. 17, pp. 1–6, 2017.

[3] Ladner, Richard E and Stefik, Andreas, "AccessCSforall: making computer science accessible to K-12 students in the United States," ACM SIGACCESS Accessibility and Computing, ACM New York, NY, USA, pp. 3–8, 2017.

[4] Vakil, Sepehr. "Ethics, identity, and political vision: Toward a justice-centered approach to equity in computer science education." Harvard Educational Review 88.1 (2018): 26-52.

[5] Morais, Isabel, and Mayra S. Bachrach. "Analyzing the impact of computer science workshops on middle school teachers." 2019 IEEE Integrated STEM Education Conference (ISEC). IEEE, 2019.

[6] Madkins, Tia C., et al. "Culturally relevant computer science pedagogy: From theory to practice." 2019 Research on Equity and Sustained Participation in Engineering, Computing, and Technology (RESPECT). IEEE, 2019.

[7] Codding, Diane, et al. "Culturally responsive and equity-focused computer science professional development." Society for Information Technology and Teacher Education International Conference. Association for the Advancement of Computing in Education (AACE), 2019.

[8] Sheridan, Scott, Chrystalla Mouza, and Lori Pollock. "Professional development for computer science education: design and outcomes from a case study teacher." Society for Information Technology and Teacher Education International Conference. Association for the Advancement of Computing in Education (AACE), 2019.

[9] Joanna Goode, Gail Chapman, and Jane Margolis. 2012. Beyond curriculum: the exploring computer science program. ACM Inroads 3, 2 (2012), 47–53.

[10] Yong Zheng. 2019. A Course on Applied AI and Data Science: Recommender Systems. (2019), 43–48.

[11] Shashank Srikant and Varun Aggarwal. 2017. Introducing data science to school kids. (2017), 561–566.

[12] Steven D Essinger and Gail L Rosen. 2011. An introduction to machine learning for students in secondary education. (2011), 243–248.

[13] Jai W Kang, Qi Yu, Erik Golen, and Edward P Holden. 2018. IT curriculum: Coping with technology trends and industry demands. (2018), 44–49.

[14] Hill, Rosemary Lucy. "What Is at Stake in Data Visualization?: A Feminist Critique of the Rhetorical Power of Data Visualizations in the Media." Data Visualization in Society, edited by Martin Engebretsen and Helen Kennedy, Amsterdam University Press, Amsterdam, 2020, pp. 391–406.

[15] D'Ignazio, Catherine, and Rahul Bhargava. "Data Visualization Literacy: A Feminist Starting Point." Data Visualization in Society, edited by Martin Engebretsen and Helen Kennedy, Amsterdam University Press, Amsterdam, 2020, pp. 207–222.

[16] D'Ignazio, Catherine, and Lauren F. Klein. "Feminist data visualization." Workshop on Visualization for the Digital Humanities (VIS4DH), Baltimore. IEEE. 2016.

[17] Cukier, Wendy, Denise Shortt, and Irene Devine. "Gender and information technology: Implications of definitions." ACM SIGCSE Bulletin 34.4 (2002): 142-148.

[18] Burge, Janet E., et al. "Girls on the go: a CS summer camp to attract and inspire female high school students." Proceeding of the 44th ACM technical symposium on Computer science education. 2013.

[19] Wajcman, Judy. "From women and technology to gendered technoscience." Information, Community and Society 10.3 (2007): 287-298.

[20] Hellerstein, Joseph M., Jeffrey Heer, and Sean Kandel. "Self-Service Data Preparation: Research to Practice." IEEE Data Eng. Bull. 41.2 (2018): 23-34

[21] Rattenbury, Tye, et al. Principles of data wrangling: Practical techniques for data preparation. " O'Reilly Media, Inc.", 2017.

[22] Wyer, Mary, et al., eds. Women, science, and technology: A reader in feminist science studies. Routledge, 2013.

[23] Simpson, Amari T., et al. "Impacting Teacher and Counselor Practices as They Support Traditionally Underrepresented Students to Pursue STEM Majors and Careers." 2020 IEEE Frontiers in Education Conference. IEEE, 2020.

[24] Mithun, Shamima, and Xiao Luo. "Design and Evaluate the Factors for Flipped Classrooms for Data Management Courses." 2020 IEEE Frontiers in Education Conference. IEEE, 2020.

[25] Smith, Sally, et al. "Computing degree apprenticeships: An opportunity to address gender imbalance in the IT sector?." 2020 IEEE Frontiers in Education Conference. IEEE, 2020.

[26] Peters, Anne-Kathrin, et al. "Care ethics to develop computing and engineering education for sustainability." 2020 IEEE Frontiers in Education Conference. IEEE, 2020.

[27] Scott, Andrew, and Scott Barlowe. "How software works: Computational thinking and ethics before CS1." 2016 IEEE Frontiers in Education Conference. IEEE, 2016.

[28] Thoroughman, Kurt A., and Joseph A. O'Sullivan. "High impact practices toward personal and professional identity in introductory and advanced engineering seminar courses." 2016 IEEE Frontiers in Education Conference. IEEE, 2016.