How Gender, Ethnicity, and Public Presentation Shape Coding Perseverance after Hackathons

Emilia Gan
efgan@uw.edu
University of Washington
Seattle, Washington, USA

Tyler Menezes
tylermenezes@codeday.org
CodeDay
Seattle, Washington, USA

Benjamin Mako Hill
makohill@uw.edu
University of Washington
Seattle, Washington, USA

ABSTRACT

Hackathon-style coding events are a popular and promising approach to broadening participation in computer science and engineering. We present a quantitative analysis of self-reported perseverance in coding after hackathon-style events for 4,703 hackathon participants run by the nonprofit organization CodeDay. Drawing from previous work on broadening participation in computing, we test hypotheses that seek to answer three questions about whether and how hackathon-style coding events support continued engagement in computing among young people: (1) Are participants from underrepresented groups as likely to continue to engage in coding after attending a hackathon-style event? (2) Are participants more likely to continue to code after hackathon-style events if they attend events with demographically similar peers? (3) Are participants more likely to continue to code after a hackathon-style event if they present their work? In line with many studies of broadening participation, we find that members of underrepresented groups are less likely to report continuing to engage in programming 10 weeks after hackathon-style events. However, we find that these participants are more likely to report continuing to code when a larger proportion of attendees at their event share their gender or ethnicity. We also find that membership in underrepresented groups is associated with a greater likelihood of continued engagement when participants present their work to others at the end of events. Our work contributes to the literature on both education and broadening participation in computing by outlining several conditions under which hackathon-style events may be effective in promoting continued engagement among underrepresented young people.

CCS CONCEPTS

• **Applied Computing → Education**; • **User Characteristics →** *Gender*; *Race and Ethnicity*.

KEYWORDS

Gender and Diversity, Outreach, Belonging, Inclusiveness, Hackathons, Coding Interest

1 INTRODUCTION

Computing technology underpins today's society and it would be difficult to identify a single domain that has not been affected by computers. It is critical, therefore, that young people have experiences with computing that go beyond simply being users of information technology. Unfortunately, there is persistent demographic inequality at all levels of computing, both in industry and academia, with both women and people of color vastly underrepresented. While the root causes of this underrepresentation are varied and complex, there is widespread recognition that broadening participation in computing must involve building interest in programming among young people who are currently underrepresented.

An increasingly popular approach to building interest and confidence in computing is hackathon-style events that seek to give participants hands-on experience with coding in a social environment. Hackathons are still a relatively recent phenomenon. The first hackathon was held by OpenBSD in 1999 for professional developers [23], and the first college hackathon occurred 10 years later [25]. Hackathons are now commonly used as educational events, with more than 200 events sponsored annually by Major League Hacking, an organization formed to support student hackathons [21]. Increasingly, these events are open to pre-college students. Hack Club, an organization that curates a list of hackathons for high school students, reports 57 events planned for 2022 before the end of July and over 100 events each year for the previous 4 years [11].

Although the popularity of hackathons has exploded, it is not clear that they have been effective in promoting increased diversity in computing. In 2019, Major League Hacking published data on the 215 events they had supported over the previous year, reporting that 73% of the attendees identified as male, 24% as female, and 3% as other [31]. Ethnic diversity was similarly lacking—84% of the attendees identified themselves as white/Caucasian or Asian/Pacific Islander [31].

CodeDay is a non-profit organization with a mission to make computing more accessible to underserved communities by organizing hackathon-like events for high school students and organizing mentored internships for high school and college students [6]. CodeDay events typically occur in multiple cities over the same weekend. Although CodeDay locations are still predominantly within the United States, there are some international events. This study uses

data collected by CodeDay from people attending more than 100 events to test hypotheses about the relationship between membership in unrepresented groups, demographics of other learners at events (which may or may not convey a sense of belonging for certain participants), public presentation of completed work, and continued participation in programming.

Our study uses data from 7,207 attendees of CodeDay events. Because our sample includes 2,074 individuals who identify as Black, Latinx, Native American, or other, we are able to gain valuable insights about the interactions between gender and ethnicity which are often difficult to measure due to the low number of underrepresented participants at individual events. This work contributes to research on informal learning and broadening participation in computing by using unique survey data to evaluate hackathon-style events in terms of their ability to encourage longer-term engagement among those with little or no interest in computing. Moreover, our work shows how event-level characteristics like overall demographics and the incorporation of public presentation are related to longer-term participation in computing—especially by attendees from underrepresented groups.

2 BACKGROUND

While many programs have been implemented to address inequalities in computing, gaps in access, participation, and opportunities persist. The U.S. Equal Employment Opportunity Commission's report on "Diversity in High Tech" reveals that computing remains a primarily White and male domain, especially in the highest paid executive-level positions, and that women, Hispanics, and African Americans are underrepresented at all levels of the computing industry [7]. The lack of gender and ethnic diversity in the computing workforce is reflective of a long sequence of inequities and imbalances, beginning from early childhood experiences and continuing throughout adulthood [19, 35].

2.1 Hackathons

A portmanteau of the terms "hacking" and "marathon," "hackathons" are time-bounded events where participants collaborate intensively on programming tasks in a social, usually co-located, environment, [2]. Hackathons often culminate in a public presentation of projects, often accompanied by the awarding of prizes. The value of these prizes can be considerable, ranging from cash, to computer hardware, to job interviews, and to internship offers [29]. Inequality in participation at early career stages can have far-reaching career implications on the demographics of computing [27, 29]. In recent years, hackathons have become an important platform for gaining skills and recognition in the computer industry, especially in individuals' early career stages [9, 22, 38]. Hackathons have also left an indelible mark on the culture of work in computing and on workplace norms [38]. Not surprisingly, most hackathons are attended predominantly by white and Asian men [31].

The male-dominated and, at times, "macho" culture of hackathons can be off-putting to women. For example, in a study by Warner and Guo, 16% of the hackathon attendees identifying as women surveyed cited "hacker culture" as a negative aspect of the event, compared to only 2% of the attendees surveyed who identified as men [36]. In their work on hackathon culture, Decker *et al.* point

out that efforts to increase participation by women, such as special preregistration periods or "women only" events, can emphasize and increase the sense of separation and nonbelonging that they are trying to counteract [8]. They go on to suggest that consciously designing events to be more collaborative and less focused on competition can result in more welcoming events for all participants [8]. Although persistence in coding during and after hackathons has not been extensively studied, some researchers have reported significant dropout rates, with up to 30% to 50% of participants not participating in the presentation stage of events [12, 38]. The benefits of organizing events with various options to participate, collaborate, and define goals have been reinforced by Kos' work with female-focused hackathon participants, showing that women often attend with the goal of exploring new topics and learning new skills and find that the competitive nature of these events works against these goals [15].

Several hackathons have been specifically designed to broaden participation. For example, some hackathons, such as Womxn/Hacks, are dedicated to improving equity in computing for female-identifying and non-gender-conforming people [37], while others, such as Hack the Gap, organize hackathons and other programs to promote inclusivity more broadly [3]. As in other hackathons, participants in these events work on projects of their own choosing and may present their work to others at the end of the event. CodeDay events are organized to increase interest in computing among high school students who might not be exposed to computing and make an effort to recruit computing novices [20].

2.2 Underrepresentation and Coding Persistence

Research has provided insight into some of the factors that contribute to the lack of diversity in computing. For example, when forming opinions on career options, it is important for young people to be able to see themselves fitting in [33]. Newcomers to coding must overcome a variety of obstacles before reaching the point where they see themselves as programmers. Computing careers are still viewed, perhaps for good reasons, as unwelcoming domains for those who do not fit the stereotypical image of a programmer. Thus, as children, girls and members of underrepresented minority groups may not know of anyone working in computing that they can identify with, making it harder for them to see themselves in computing careers.

Previous research by Margolis *et al.* [18], Jepson and Perl [14], and Resnick and Rusk [28], among many others, has shown that early exposure to computing can help counteract the negative effects of the stereotypes commonly associated with computing and the lack of representation in the computing workforce. Popular initiatives, such as Scratch and Code.org, have shown that computing can have wide-reaching appeal, as long as participants are given the opportunity to feel valued and empowered. Despite the popularity of such programs, many young people enter college without having had such positive experiences.

Our first hypothesis aims to provide insight into this issue. The participants in this study are primarily high school students. Many of these students are likely in the process of figuring out what they might want to do after they finish school. Given historical trends,

we expect that a significant proportion of these students already believe that a career in computing is not for them. Consequently, we hypothesize that *(H1) participants from groups who are underrepresented in computing are less likely to continue coding than participants in well represented groups*. Clearly, there are multiple demographic dimensions along which an individual might be underrepresented. Although gender inequality in computing is widely studied [18], there is also a lack of ethnic diversity [1]. As a result, we split our first hypothesis into two subparts. We hypothesize both that *(a) participants who do not identify as male will be less likely to continue coding* and *(b) participants who do not identify as White will be less likely to continue coding*.

2.3 Demographic Peers and Coding Persistence

Students who come from groups underrepresented in computing often face significant barriers to participation. These include a lack of access to technology and a lack of role models and mentors [16, 35]. As a result, these students are less likely to self-identify as "coders" or see computing as a field open to them. Research by Cheryan and Plaut [4] and Tellhed *et al.* [30] has shown that the ability to identify with others in a group learning to code is key to developing interest in belonging to that group. Our sense is that the demographics of typical hackathons could discourage participants from underrepresented groups by suggesting, perhaps implicitly, that they might not belong. Walton and Cohen have shown that for college students, if individuals feel that they have few friends in a given domain, their sense of belonging and the ability to perform to their full potential are negatively affected [34]. Given that social connection is a basic human need, hackathons that promote a feeling of belonging may be especially effective in helping beginners overcome initial challenges to continue participating in programming activities. For students from demographic groups that have not typically been well represented in computing, attending diverse events may be an important first step in creating more of a sense of connection. For our second hypothesis, we look at how individual participants are affected by attending an event with others of their same gender or ethnicity. We hypothesize that *(H2) participants who are more demographically similar to others at their event are more likely to continue coding*. Once again, we break this down into two subhypotheses based on gender and ethnicity: *(a) participants who attend with others of their gender will be more likely to continue coding* and *(b) participants who attend with others of their ethnicity will be more likely to continue coding*.

2.4 Presenting One's Work and Coding Persistence

There is reason to believe that the presentation of work by participants at the end of hackathon-style events can be beneficial. Kos includes both encouraging participants to present and providing opportunities to demo projects in noncompetitive settings in a list of design recommendations for creating inclusive hackathons [15]. A study of participants in the Scratch online community showed that public sharing of user-created artifacts is associated with higher levels of subsequent participation, but also revealed that groups less likely to participate are also less likely to share at the earliest stages of their participation [10]. A proposed explanation is that sharing work can initiate a virtuous cycle in which a sense of belonging leads to increased sharing, which, in turn, contributes to an increased sense of belonging [10]. Furthermore, connecting with peers and feeling accepted by them is important for the development of one's social identity and sense of belonging [32]. This dynamic can be elusive for members of underrepresented groups.

Often, hackathons are set up such that end-of-event presentations are a celebration of participants' accomplishments. We suspect that among students who do not already feel connected to coding, presenting their work is more likely to be viewed as a risk with the potential to expose their lack of expertise and label them as outsiders. Successful presentation of work to a supportive audience of peers can therefore be especially significant to these participants as it validates that this work is something they can do successfully and be recognized for. Consequently, we hypothesize that *(H3) participants who present their work are more likely to continue coding*. For the third time, we split this into two subhypotheses based on gender and ethnicity: *(a) participants who do not identify as male and who present their work will be more likely to continue coding*, and *(b) participants who do not identify as White and who present their work will be more likely to continue coding*.

3 EMPIRICAL SETTING

We test our hypotheses using a unique dataset from the nonprofit organization CodeDay. Since its founding in 2009, more than 58,000 people have attended CodeDay events, and more than 70% of CodeDay attendees are from groups recognized as underrepresented in computer science [6]. Although events were held virtually during the COVID-19 pandemic beginning in March 2020, events described in this dataset were all held in-person at various locations in the United States over a weekend (including overnight on Saturday) before COVID. Since November 2021, the organization has returned to running in-person events. Events are organized to promote attendance by young people local to the areas where the events are held.

Each CodeDay event follows a similar format. Participants first spend an hour presenting ideas for self-directed projects and self-select into teams of 1-6. At this time, an optional 1-hour beginner coding workshop is offered. For the next 19-20 hours, teams work to create projects using their chosen technology. Industry mentors and more experienced students provide guidance as needed.

CodeDay structures the final two hours of each session using one of four methods: (1) participants are asked to present to their peers and a panel of judges; (2) participants are asked to meet with judges in a science fair-style exhibition format and a subset are randomly selected to present to their peers; (3) participants are asked to meet with judges in a science fair-style exhibition and winners are asked to present to their peers; (4) participants are asked to meet with judges in a science fair-style exhibition format with no presentations to peers.

When included, presentations are comprised of a 2-minute group demonstration of the finished project with narration, with no audience questions or feedback. The focus of these presentations is

Demographics of All Participants Providing Gender & Ethnicity Information (%)
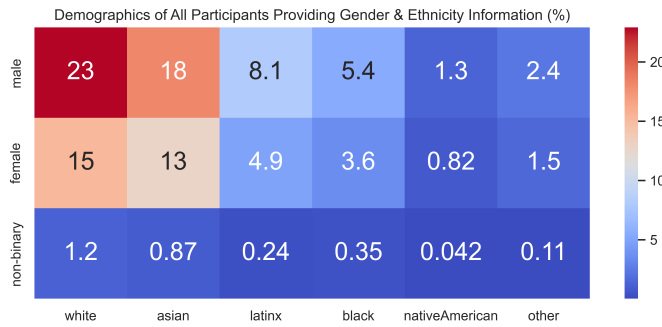

Figure 1: Heatmap showing demographics of CodeDay attendees shown as percentage of attendees (considering only attendees for whom data are available).

on the finished work and participants are asked not to create slides or a script. Although participants can technically opt out of presentations, CodeDay staff make an effort to encourage all students to present [20] and 12,729 of the 14,057 student in our full dataset (more than 90%) are listed as having presented.

Similarly, meetings with judges focus on demonstrations of the work without slides or a script. Judges are encouraged to ask questions as needed to determine scores in the field of effort, creativity, and polish. The average duration of a meeting with judges is three minutes, and the average number of judges present is three [20].

4 METHODOLOGY

4.1 Data

This study uses fully anonymized data from 141 unique CodeDay sessions on 7 CodeDay weekends and 28 different locations held between November 8, 2014 and November 12, 2016. Table 1 provides some numerical overviews of the dataset used in this project. We see that the events ranged in size from 13 to 351 attendees. Most events had fewer than 150 attendees. Survey data were collected from 14,057 hackathon attendees by CodeDay. Of these individuals, 7,207 provided demographic data. All participants who indicated that they had "low" or "no" interest in coding prior to attending an event were sent a 10-week postevent follow-up survey via email.

In Figure 1, we present a breakdown of attendees grouped by gender and ethnicity for all participants for whom we had such data (we describe how we operationalize each in the following subsection). A demographic breakdown by interest (i.e., attendees expressing little prior interest in coding, attendees expressing no prior interest in coding, and attendees who received, but did not respond to, the follow-up survey) did not differ greatly from the heatmap shown. Additional cross tabulations are provided in the online resources accompanying this paper (see Appendix A).

Of a total of 3,451 postevent surveys sent out to students reporting "low" interest, 2,411 responses were received (a 70% response rate). Of the 1,252 individuals who reported "no" interest in coding pre-event, 779 responded (a 62% response rate). These response rates are quite high by social scientific standards, especially since respondents were not paid or reimbursed in any way [13]. The core dataset of 3,190 participants used in our analyses includes all

participants for whom we have complete demographic data and who responded to both the initial and follow-up surveys.

4.2 Measures

We test our hypotheses using a set of dichotomous independent measures that we construct from the dataset shared with us by CodeDay. Our dependent variable *Continued* is a dichotomous measure of continued participation in coding that is set to 1 if a user reported that they were still engaging in coding activities in the 10-week postevent survey and 0 if the user reported that they were not.

Our first key independent variable is *NonMale* and is assigned 0 for participants who reported their gender as Male and 1 for participants who self-identified as Female, nonbinary, or other. While we recognize that it is more common to code biological sex as male/female/etc and gender identity as man/boy, woman/girl, and nonbinary, we code "gender" as *NonMale* in our analyses for two reasons: the survey asked for gender in terms of male, female, or nonbinary, and we are interested in the experience of underrepresented groups in general. We describe this as gender and not sex because CodeDay staff explained to us that the survey was about gender identity, not biological sex. We use a binary gender specification for statistical inference purposes, not because we believe that gender identity can be adequately represented as a binary. While we chose to use the terms "Male" and NotMale" in our analysis, rather than "Boy" and "NotBoy," it is important to recognize that the subjects of our study are children. Our second key independent variable is *Non-White*, with participants who identified as White coded as 0 and all others coded as 1.

In addition to the overall demographic breakdown of our dataset, we were also interested in the details of the demographic composition of events from the personal perspectives of individual participants. Accordingly, for each individual attendee, we separately calculated the proportion of other participants attending each attendee's CodeDay event who shared the focal individual's reported gender (*Gender Prop.*) and ethnicity (*Ethnicity Prop.*) and used these proportions to test Hypothesis 2. Although these proportions can theoretically range between 0 and 1, when we look only at participants who provide demographics data, the large majority of the data was between 0.1 and 0.3 for ethnicity and between 0.1 and 0.4 for gender.

To test Hypothesis 3, we constructed a measure, *Presented*, where participants who presented were assigned 1 and those who did not present were assigned 0. It is important to note that whether participants presented was typically determined at the event level. The large majority of participants in our dataset attended events where all groups presented in some form (131 of 141 events) and each attendee was recorded as having presented. That said, there was no requirement for each member of the group to actively participate in the presentation in these cases. At a further four events, none of the participants presented, leaving only six events where whether to present or not may have been offered as an option to attendees.

Coding Interest, used as a control variable, was also a dichotomous measure. It was coded 1 when participants reported low prior interest in coding and 0 when no prior interest was reported. Because we only have outcome data on participants who reported

Table 1: Descriptive statistics for dataset from CodeDay used in this analysis

Parameter	Value
Number of events	141
Size of smallest event	13
Size of largest event	351
Mean event size	99.7
Standard deviation (event size)	70.3
Median event size	81
Total number of attendees	14057
Total number of attendees with complete demographics data	7207
Number of attendees who were sent a postevent survey	4703
Number of attendees who completed postevent survey	3190
Number of attendees with complete 'Presented' data and postevent survey	3107

low or no interest, this specification captures all variation in coding interest in our dataset.

4.3 Analytic Plan

We first use descriptive statistics to understand the gender and ethnicity distributions of the participants in CodeDay events. The Python code we used to explore and visualize our data is available in the online material accompanying this paper (link provided in Appendix A). We then test all our hypotheses using multilevel logistic regression. Pampel [24] explains that while linear regression is well suited to outcomes that are continuous, logistic regression can be used to model differences in probability as log odds and is well suited for dichotomous dependent variables, such as whether or not an event takes place. Prior level of expressed coding interest of the participants is used as a control variable across all our analyses and is expected to be strongly associated with continued participation. Because participants are clustered within events in a way that might threaten the statistical assumption of independence behind logistic regression, we fit generalized linear mixed-effects regression models with a random intercept term (u) associated with each event. The R code that we used to evaluate our models is available in the online material accompanying our paper.

4.3.1 Hypothesis 1: Demographics and persistence. Hypothesis 1 looks at the association between gender (*NonMale*) or ethnicity (*NonWhite*), separately, on continued participation in coding activities:

For **Hypothesis 1a**, we used the following model to assess the effect of being nonmale:

$$\log\left(\frac{\hat{p}(continued)}{\hat{p}(1-continued)}\right) = \hat{\beta}_0 + \hat{\beta}_1 NonMale + \hat{\beta}_2 Coding\ Interest + u$$

For **Hypothesis 1b**, we used the model below to assess the effect of being non-White:

$$\log\left(\frac{\hat{p}(continued)}{\hat{p}(1-continued)}\right) = \hat{\beta}_0 + \hat{\beta}_1 NonWhite + \hat{\beta}_2 Coding\ Interest + u$$

4.3.2 Hypothesis 2: Peer demographics and persistence. Hypothesis 2 uses each individual as a reference point and looks at the

association between the proportion of attendees who share similar demographics (*Gender Prop.* and *Ethnicity Prop.*) and continued participation in coding activities. Because we are particularly interested in the ability of these events to broaden participation, we present the tests of Hypothesis 2 both in general and among groups stratified by *NonMale* and *NonWhite*. This allows us to see the effect of demographic similarity between attendees among members of underrepresented groups.

For **Hypothesis 2a**, we used the following model to assess the effect of attending events with others who share the same self-identified gender:

$$\log\left(\frac{\hat{p}(continued)}{\hat{p}(1-continued)}\right) = \hat{\beta}_0 + \hat{\beta}_1 Gender\ Prop. + \\ \hat{\beta}_2 Coding\ Interest + u$$

For **Hypothesis 2b**, we used the following model to assess the effect of attending events with others who share the same self-identified ethnicity:

$$\log\left(\frac{\hat{p}(continued)}{\hat{p}(1-continued)}\right) = \hat{\beta}_0 + \hat{\beta}_1 Ethnicity\ Prop. + \\ \hat{\beta}_2 Coding\ Interest + u$$

4.3.3 Hypothesis 3: Presentation and persistence. In Hypothesis 3, we look at the association between presenting one's work publicly (*Presented*) and continued participation, combined with demographic measures.

In **Hypothesis 3a**, we look at the association of presentation and gender with continued activity in coding. We use two different measures to look at the effect of gender. The first is a binary measure that indicates whether the attendee is of an underrepresented gender group (*NonMale*), as in Hypothesis 1a. The second is our measure of the proportion of attendees at the same event who share the gender of the attendee (*Gender Prop.*), as in Hypothesis 2a. We also consider the interaction between gender and presentation. Our

model for **Hypothesis 3a** is thus:

$$\log\left(\frac{\hat{p}(continued)}{\hat{p}(1-continued)}\right) = \hat{\beta}_0 + \hat{\beta}_1 Presented + \hat{\beta}_2 Coding\ Interest +$$
$$\hat{\beta}_3 NonMale + \hat{\beta}_4 Gender\ Prop. +$$
$$\hat{\beta}_5 NonMale * Presented + u$$

Similarly, with **Hypothesis 3b**, we look at the association of presenting and ethnicity using both a binary measure of ethnicity (*NonWhite*) and a measure of how many others attending the event share the attendee's self-identified ethnicity (*Ethnicity Prop*). Therefore, our model for **Hypothesis 3b** is:

$$\log\left(\frac{\hat{p}(continued)}{\hat{p}(1-continued)}\right) = \hat{\beta}_0 + \hat{\beta}_1 Presented + \hat{\beta}_2 Coding\ Interest +$$
$$\hat{\beta}_3 NonWhite + \hat{\beta}_4 Ethnicity\ Prop. +$$
$$\hat{\beta}_5 NonWhite * Presented + u$$

5 RESULTS

Tables 2 and 3 present a set of descriptive bivariate statistics from the follow-up surveys sent to these attendees that capture the basic patterns in our data. In these tables, we include data for nonrespondents to the survey (the category "unknown" in the tables) and calculate persistence percentages while both including and excluding these unknowns. The combined category in each table is the same and is provided for ease of comparison with the groupings that are divided by gender (nonmale/male) or ethnicity (non-White/White). When unknowns are left out, it should be noted that the proportions of respondents in each subgroup that persisted in coding activities after the event fall within a relatively narrow range, with a low of 62% (for the nonmale subgroup) and a high of 67.4% (for the White subgroup).

As described in our Analytical Plan Section, coding interest was included as a control variable in all our statistical analyses. Not surprisingly, a higher level of prior coding interest—that is, "little" interest, as opposed to "no" interest—was strongly associated with a higher likelihood of continuing participation. Furthermore, as can be seen in Tables 4 through 9, the coefficients associated with coding interest fall within the range of 0.58–0.74, and within the even narrower range of 0.64–0.69 for most analyses.

H1a looks at the relationship between gender and continued participation. As shown in Table 4, we see that nonmale participants are less likely to continue coding after the event than male participants. Table 5 shows the results of the analysis carried out to test H1b, focusing on the relationship between ethnicity and continued participation. Here, we see that non-White participants are less likely to continue coding after the event than White participants. Both results are as hypothesized.

In Tables 6 and 7, we see results of the analyses testing H2a and H2b, which explore the effect of having others of the same gender and ethnicity, respectively, at an event. Our results indicate that attending events with others of the same gender is not significantly associated with continued participation for either male or nonmale
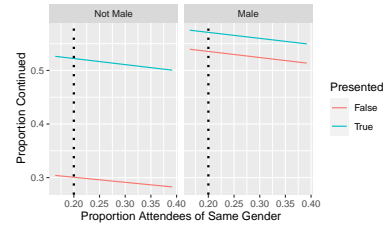

Figure 2: Line plot showing predicted probability of continued participation given hypothetical proportions of similar-gender attendees and presentation status (H3a). Keep in mind that negative slope of the lines and the difference between boys who did and not present are not statistically significant relationships in our model.

participants. However, when considering ethnicity, having others of the same ethnicity at the event was significantly associated with increased likelihood of continued participation for non-White, but not for White, attendees.

Table 8 presents the results of the analysis performed for H3a focusing on gender. We see a negative association between *NonMale* and continued participation that is consistent with our findings for H1a. Although the effect of presenting on its own was not found to be significant in the analysis for H3a, the result for the interaction term indicates that presenting is significantly associated with a higher likelihood of continued participation for nonmale attendees who present their projects.

Table 9 presents the results of the analysis performed for H3a focusing on ethnicity. We see support for the argument of H2b that the presence of others of one's own ethnicity is significantly associated with continued participation. On the other hand, presenting one's project at the event appears to be associated with a slightly lower likelihood of continuing to participate in coding after the event. However, looking at the interaction term, we see that the situation is not so straightforward. For attendees who are non-White, presenting was associated with an increased likelihood of continued participation. Although interesting, neither of these last two findings is statistically significant.

Our findings for H3a (gender focus) are illustrated in Figure 2. Our findings for H3b (ethnicity focus) are illustrated in Figure 3. Hypothesis 3 explores the association between presenting at the event and continued participation in coding. However, since the models we use for H3 contain our ethnicity or gender variables as well, this allows us to plug in different values for these variables and plot the resulting predictions. The steps followed to produce these plots are described in detail in the online materials accompanying this paper. Both figures show model-predicted levels of continued participation (our y axis) drawn from our models for prototypical individuals with coding interest held at our sample mean and with the proportion of attendees with similar gender or ethnicity varying between the 10th and 90th percentile in our sample (our x axis).

Figure 2 shows that, as we predicted in H1a, being nonmale is associated with a decreased likelihood of continuing to participate in coding after the event. Also in line with our prediction for H3a,

Table 2: Descriptive data on the persistence of coding by gender. The table shows results obtained from the follow-up surveys. In each "% resp." column, we remove unknowns and only consider data from respondents to follow-up surveys.

Continued?	NonMale	%	% resp.	Male	%	% resp.	Combined	%	% resp.
Unknown	621	0.312	−	892	0.328	−	1513	0.322	−
Yes	846	0.426	0.620	1210	0.445	0.663	2056	0.437	0.645
No	519	0.261	0.380	615	0.226	0.337	1134	0.241	0.355
Total	1986			2717			4703		

Table 3: Descriptive data on the persistence of coding by ethnicity. The table shows results obtained from the follow-up surveys. In each "% resp." column, we remove unknowns and only consider data from respondents to follow-up surveys.

Continued?	NonWhite	%	% resp.	White	%	% resp.	Combined	%	% resp.
Unknown	979	0.322	−	534	0.320	−	1513	0.322	−
Yes	1292	0.426	0.628	764	0.458	0.674	2056	0.437	0.645
No	765	0.252	0.372	369	0.221	0.326	1134	0.241	0.355
Total	3036			1667			4703		

Table 4: H1a: Association between gender and continued participation in postevent coding. We see that the variable *NonMale* is associated with a decreased probability of continued participation.

H1a (Gender)	coefficient (std. error)
(Intercept)	$0.35\ (0.10)^{***}$
NonMale	$-0.18\ (0.08)^{*}$
Coding Interest	$0.65\ (0.09)^{***}$
Num. obs.	3190

$^{***}p < 0.001;\ ^{**}p < 0.01;\ ^{*}p < 0.05$

Table 5: H1b: Association between ethnicity and continued participation in postevent coding. We see that the variable *NonWhite* is associated with a decreased likelihood of continued participation.

H1b (Ethnicity)	coefficient (std. error)
(Intercept)	$0.39\ (0.11)^{***}$
Non White	$-0.19\ (0.08)^{*}$
Coding Interest	$0.65\ (0.09)^{***}$
Num. obs.	3190

$^{***}p < 0.001;\ ^{**}p < 0.01;\ ^{*}p < 0.05$

Table 6: H2a: Association between gender proportion and continued participation in coding. We find no statistically significant associations, other than the expected association with expressed *Coding Interest*.

H2a (Gender)	Combined	NonMale	Male
(Intercept)	0.10 (0.16)	0.28 (0.25)	0.47 (0.33)
Gender Prop.	0.64 (0.49)	−1.04 (1.02)	−0.29 (1.00)
Coding Interest	$0.65\ (0.09)^{***}$	$0.74\ (0.14)^{***}$	$0.58\ (0.12)^{***}$
Num. obs.	3190	1365	1825

$^{***}p < 0.001;\ ^{**}p < 0.01;\ ^{*}p < 0.05$

Table 7: H2b: Association between ethnicity proportion and continued participation in coding. We see significant associations for *Ethnicity Prop.* for all attendees combined and for *Non White*, but not White attendees, analyzed separately.

H2b (Ethnicity)	Combined	Non White	White
(Intercept)	−0.04 (0.13)	−0.01 (0.16)	−0.02 (0.31)
Ethnicity Prop.	$1.97\ (0.60)^{**}$	$1.71\ (0.87)^{*}$	1.78 (1.37)
Coding Interest	$0.65\ (0.09)^{***}$	$0.65\ (0.11)^{***}$	$0.64\ (0.15)^{***}$
Num. obs.	3190	2057	1133

$^{***}p < 0.001;\ ^{**}p < 0.01;\ ^{*}p < 0.05$

Table 8: H3a: Association of presentation and gender with continued participation in coding. Although *Presented*, alone, was not shown to be significant, the interaction between having presented and not identifying as male (*NonMale * Presented*) was shown to be significant.

H3a (Gender)	coefficient (std. error)
(Intercept)	0.36 (0.44)
Presented	0.14 (0.39)
Coding Interest	$0.69\ (0.09)^{***}$
NonMale	$-0.99\ (0.37)^{**}$
Gender Prop.	−0.45 (0.69)
NonMale × Presented	$0.79\ (0.37)^{*}$
Num. obs.	3107

$^{***}p < 0.001;\ ^{**}p < 0.01;\ ^{*}p < 0.05$

presentation is associated with an increased likelihood of continuing coding for attendees, regardless of gender. Furthermore, while presenting is associated with higher levels of perseverance for all attendees, the plot clearly displays that this effect is much more significant for nonmale attendees. The downward slope evident as the gender proportion increases is contrary to our prediction for H2a, that attending events with others of similar gender would

Table 9: H3b: Association of presentation and ethnicity with continued participation in coding. Here, the only association of interest found is for *Ethnicity Prop.* – as was seen earlier, attending events with peers of similar ethnicity is associated with continued participation.

H3b (Ethnicity)	coefficient (std. error)
(Intercept)	0.25 (0.53)
Presented	−0.24 (0.53)
Coding Interest	0.69 (0.09)***
Non-White	−0.89 (0.48)
Ethnicity Prop.	1.96 (0.74)**
Non-White × Presented	0.87 (0.49)
Num. obs.	3107

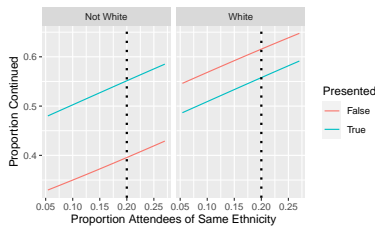***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

Figure 3: Line plot showing predicted likelihood of continued participation given hypothetical proportions of similar-ethnicity attendees and presentation status (H3b). Keep in mind that the differences in terms of presentation are not statistically significant.

be associated with increased persistence with coding. Although this effect is interesting, it does not reflect a statistically significant relationship.

Consider, for example, a prototypical nonmale participant who attends an event where 20% of participants are also nonmale (indicated by the dashed vertical line in the plot). If this hypothetical person were to present, they would be predicted to have a probability of continuing to code of 52.2%. However, if this participant did not present, the predicted probability of continued coding drops to just 30%. For a boy attending an event where 20% of the other attendees are also boys, presenting would be associated with a predicted probability of continued coding of 57.1%—not significantly higher than the predicted probability of 53.5% had he not presented.

Figure 3 shows that, for participants who do not present (red lines in Figure 3), White participants are consistently predicted to continue coding at higher rates than non-White participants. This is consistent with our prediction for H1b. However, this effect essentially disappears when we compare White and non-White attendees who presented their work. The predicted values for these attendees create almost identical lines (blue lines in Figure 3). Looking still at the effect of presenting, we see that while our prediction for H3b held for non-White attendees, presenting among White attendees is associated with a slightly decreased likelihood of continued coding

in this group of attendees, contrary to our prediction. This relationship, although not statistically significant, is interesting, as it suggests that presenting might eliminate any advantage that membership in a well-represented group provides to attendees. The plot also clearly shows the positive association between attendees of similar ethnicity and continued coding, supporting our prediction for H2b.

Consider, for example, a prototypical non-White participant attending an event where 20% of the other attendees are also non-White. If this hypothetical person were to present, presenting would be associated with a predicted probability of continued coding of 55.2%—higher than the predicted likelihood of 39.6%, had the participant not presented. For a prototypical White participant who attends an event where 20% of the participants are also White, presenting is associated with a predicted likelihood of continued coding of 55.8%, essentially the same probability as found for the prototypical non-White participant. However, if this hypothetical White participant had not presented, the predicted probability would have been 61.6 %.

6 DISCUSSION

Since its first 24-hour event in Seattle in 2011, close to 60,000 young people have participated in nearly 500 CodeDay events [6]. Although CodeDay is the only organization that we know of currently hosting hackathons aimed at broadening participation among high school age participants at this scale, we believe that this demonstrates the potential of these hackathon-style events. But how well does CodeDay actually meet its goal of engaging more underrepresented students in computing long-term? And what might the organization do to be even more effective? We believe that our analysis offers insights into the characteristics of CodeDay events that contribute to long-term participation in coding. We believe these insights can be used in the organization of future events aimed at broadening participation by many organizations.

6.1 Hypothesis 1: Demographics and persistence

In our test of H1, we find support for our predictions for both gender and ethnicity: being nonmale or non-White is associated with a lower likelihood of continued engagement. This is also apparent in Tables 2 and 3 and is not surprising. Margolis and Fisher comprehensively described factors that contribute to gender inequality in computing 20 years ago [17]. Margolis, with others, continued to look at how education, race, and computing are intertwined, making clear how long-standing societal inequality was reflected in schools and limited access to the opportunities offered by computing [19]. Their work both provides context and emphasizes the relevance of current work around inequity in computing. Adolescence is a critical period in identity formation—young people make decisions about who they are and where they fit in [17, 19]. Such deliberations directly impact such decisions as college major. The messages society sends to girls and members of ethnic minority groups, starting in early childhood, have not changed much over the years. Numerous programs have been established to introduce young children to computing, including many designed specifically for girls. Despite these efforts to counteract them, gender gaps are stubbornly persistent.

CodeDay seems to be seeing some promising results in their efforts to actively recruit participants from underrepresented populations to many of their events and programs. Our demographic data show that events are bringing in students from diverse backgrounds including students who, for whatever reason, do not currently feel connected to computing or see opportunities for themselves there. We view the results related to event demographics illustrated in Figure 1 as a hopeful sign. Furthermore, as seen in Tables 2 and 3, although attendees who identify as nonmale continue at lower rates than those who identify as male, and those who identify as non-White continue at lower rates than those who identify as White, the rates are not vastly different.

6.2 Hypothesis 2: Event demographics and persistence

Looking at H2, we find partial support for our prediction that members of a well-represented group might feel just as out of place as members of a traditionally underrepresented group if attending an event where few others shared their gender or ethnicity, and that this would be reflected in lower persistence in coding after the event. Our prediction seemed to be supported for both gender and ethnicity. Looking closer at the effect for gender, however, we see that the sign of the coefficient for *Gender Prop.* changes when attendees are grouped as either nonmale or male (Table 6, last two columns). This suggests that for both, when considered separately, the likelihood of continued engagement decreases as the proportion of other attendees who share the same gender identification increases. This is contrary to our prediction.

As Figure 2 illustrates, nonmales continue at lower rates than males. In this lower end of the range, their lower base levels for continued engagement dominate. As we move to higher values for *Gender Prop.*, boys are overrepresented so their higher base level of continued engagement dominates. This transition shows how our combined model might mask the slight negative association between *Gender Prop.* and continued engagement revealed by our regression analysis. Although not statistically significant, the possibility of this relationship is still thought provoking and fits the observation made by Decker *et al.* that segregated events can increase the sense of separation and exclusion from mainstream computing culture [8]. We imagine that there might be some "sweet spot," where the gender distribution of an event benefits all attendees.

When looking at ethnicity, we find that attending events with others of one's own ethnicity is associated with a greater likelihood of continued participation for both non-White and White attendees. However, the result for White attendees was not statistically significant. We speculate that this could be because White attendees are not subject to the same level of self-doubt about whether or not they belong in computing, given that the computing workforce is predominantly White. Another contributing factor could be that White attendees are not subject to the daily stress of surviving in a society that is not structured to include them and do not have to deal with microaggressions attendees from other ethnic backgrounds may have to contend with. Consequently, White attendees at an event where many others look like them might barely notice this fact. For attendees who are not used to thinking of computing as a welcoming space for people like them, seeing others from similar ethnic backgrounds at CodeDay events could be both surprising and positively motivating.

6.3 Hypothesis 3: Presentation and persistence

Finally, for H3, we find partial support for our predictions. When looking at presentation in combination with ethnicity (Table 9), we did not see a significant impact of presentation alone. However, it is interesting to note that the coefficient for presentation is negative, suggesting a slightly decreased likelihood of continued participation, while the interaction term of ethnicity and presentation is positive. This suggests that presenting could be associated with a slightly lower likelihood of continued participation for White attendees who present and a somewhat higher likelihood of continued participation for non-White attendees who do so. In this model, increasing the proportion of others of the same ethnicity was found to be significantly associated with increased participation, regardless of presentation (as hypothesized in H2b).

6.4 Future Directions

As is often the case, our findings have raised new questions that warrant further exploration. The influences that determine whether or not someone decides to engage in coding are dynamic, interacting, and complex to elucidate. Our current study has provided a partial view of these influences, but much remains to be studied and understood. After completing this analysis, we hope researchers will investigate the following questions: 1) What motivates participants expressing low or no prior interest in coding to attend these events? 2) What structural features of these events do participants find most helpful and influential in how they perceive coding and computing? 3) What does long-term participation look like? and 4) What impact does intersectionality have in how participants experience the event and in their long-term coding persistence?

The dataset we used provides a historical record of CodeDay's success at reaching young people from diverse backgrounds. Our quantitative analysis strongly suggests that the diversity of events contributes to the success of these events as measured by long-term persistence in coding. Of course, our quantitative analysis cannot reveal rich insights into the motivations or experiences of individual attendees. Cheryan *et al.* have demonstrated the importance of environment in creating a sense of belonging [5]. Although CodeDay staff have worked hard to create an environment that will be perceived as welcoming and safe by all participants, to date they have not been able to assess what elements of events are most critical to achieving this goal. To explore the subjective experiences of attendees, we hope to conduct a mixed methods study of this population in the future. In addition to collecting survey data at different time points, we hope to also conduct observations of events and interviews of participants. This approach has the potential to provide valuable information on how young people view coding and their place in computing at a critical decision-making point in their lives.

7 LIMITATIONS

Because event attendees were not required to provide demographic data, no data was available for 48.7% of the attendees. It is possible

that had we been able to include demographics for these participants, our results would be different. We believe our *Gender Prop.* and *Ethnicity Prop.* measures are most likely to be affected by not having complete demographics data because some individual events had large numbers of attendees who did not provide demographics data. We are encouraged that in spite of this, we are still able to see correlations between event demographics and persistence. We plan to conduct qualitative studies in the future to assess the large-scale trends identified in this work and to understand individual participant motivations.

Another possible threat to the validity of our findings stems from the fact that no data on the age of the attendees was collected. Therefore, we cannot evaluate the effect age has on presenting and/or continuing coding and possible interactions between age and other variables in the analysis. It is possible that the event format might affect different age groups to different degrees, especially as older attendees may have had more experience with coding. Although age data are not available, we know that most of CodeDay's events were publicized to high school students and that most of the attendees came from this age group. Additionally, we expect that college students who may have attended events tended to be more interested in coding because they would have needed to specifically seek out the events. In these cases, they would not have been included in our analysis because follow-up surveys were only sent to those attendees who expressed little or no prior interest in coding. Consequently, we would not expect the analysis to change significantly if age data were available and feel comfortable that our analysis adequately represents the age group of interest (high schoolers).

Considering the importance of role models and mentors, we should also acknowledge that mentor demographics could influence attendees' motivation to continue coding and the development of a sense of belonging in computing. Although approximately 60% of the students in the dataset (for whom demographics were available) were members of a population underrepresented in computing, it is estimated that only about a quarter of CodeDay's mentors are [20]. Because CodeDay mentors are typically software engineers recruited from local companies, the diversity of mentors often reflects the low diversity in the computer industry. No demographic data was collected from events about mentors.

A final threat concerns the external validity of our results and the degree to which our results generalize to other hackathons. As hackathons have become more popular, efforts have been made to make events more welcoming and diverse [26]. We believe that our results will generalize better to such hackathons than to more traditional hackathon events. Our data are also specific to the United States, a country with its own unique issues around diversity and inclusion, which may limit the generalizability of our findings.

8 CONCLUSION

Given the historically low number of women and people of color in computing, great efforts have been made to broaden participation. For designers of hackathon-style events seeking to address this inequality, our findings indicate that working to promote diversity in participant ethnicity might be as important and at least as successful. In the context of this study, attending events with others of the same gender did not have the same impact on persistence in coding as attending events with others of the same ethnicity. Our findings should be of interest to educators and others seeking to promote inclusivity and diversity.

REFERENCES

[1] Sharla Alegria. 2020. What do we mean by broadening participation? Race, inequality, and diversity in tech work. *Sociology Compass* 14 (03 2020). https://doi.org/10.1111/soc4.12793

[2] Gerard Briscoe. 2014. Digital Innovation: The Hackathon Phenomenon. http://www.creativeworkslondon.org.uk/wp-content/uploads/2013/11/Digital-Innovation-The-Hackathon-Phenomenon.pdf Accessed: 2022-03-08.

[3] Yahsmene Butler. 2021. 2020 Wrapped. https://www.hackthegap.com/blog/overcoming-2020 Accessed: 2021-08-12.

[4] Sapna Cheryan and Victoria C Plaut. 2010. Explaining Underrepresentation: A Theory of Precluded Interest. *Sex roles* 63, 7 (2010), 475–488.

[5] Sapna Cheryan, Victoria C. Plaut, Paul G. Davies, and Claude M. Steele. 2009. Ambient belonging: how stereotypical cues impact gender participation in computer science. *Journal of personality and social psychology* 97 6 (2009), 1045–60.

[6] CodeDay. 2021. CodeDay Homepage. https://www.codeday.org/ Accessed: 2021-07-29.

[7] U.S. Equal Employment Opportunity Commission. 2016. Special Report: Diversity in High Tech. https://www.eeoc.gov/special-report/diversity-high-tech

[8] Adrienne Decker, Kurt Eiselt, and Kimberly Voll. 2015. Understanding and improving the culture of hackathons: Think global hack local. In *2015 IEEE Frontiers in Education Conference (FIE)*. 1–8. https://doi.org/10.1109/FIE.2015.7344211

[9] Joanna F. DeFranco, Christopher S. Eagle, James Bret Michael, John Viega, and Jeffrey Voas. 2021. Hackathons 101. *Computer* 54, 5 (2021), 65–69. https://doi.org/10.1109/MC.2021.3064435

[10] Emilia F. Gan, Benjamin Mako Hill, and Sayamindu Dasgupta. 2018. Gender, Feedback, and Learners' Decisions to Share Their Creative Computing Projects. *Proc. ACM Hum.-Comput. Interact.* 2, CSCW, Article 54 (Nov. 2018), 23 pages. https://doi.org/10.1145/3274323

[11] HackClub. 2022. High School Hackathons in 2022. https://hackathons.hackclub.com/ Accessed: 2022-02-24.

[12] Caroline D Hardin. 2021. "Learning from Mistakes is the Best Thing": Risk and Persistence at Hackathons. In *21st Koli Calling International Conference on Computing Education Research* (Joensuu, Finland) *(Koli Calling '21)*. Association for Computing Machinery, New York, NY, USA, Article 20, 10 pages. https://doi.org/10.1145/3488042.3490175

[13] Brooks Holtom, Yehuda Baruch, Herman Aguinis, and Gary A Ballinger. 2022. Survey response rates: Trends and a validity assessment framework. *Human Relations* 75, 8 (2022), 1560–1584. https://doi.org/10.1177/00187267211070769 arXiv:https://doi.org/10.1177/00187267211070769

[14] Andrea Jepson and Teri Perl. 2002. Priming the Pipeline. *SIGCSE Bull.* 34, 2 (June 2002), 36–39. https://doi.org/10.1145/543812.543826

[15] Brittany Ann Kos. 2019. Understanding Female-Focused Hackathon Participants' Collaboration Styles and Event Goals. In *Proceedings of the International Conference on Game Jams, Hackathons and Game Creation Events 2019* (San Francisco, CA, USA) *(ICGJ 2019)*. Association for Computing Machinery, New York, NY, USA, Article 5, 4 pages. https://doi.org/10.1145/3316287.3316292

[16] Michael J. Lee. 2019. Increasing Minority Youths' Participation in Computing through near-Peer Mentorship. *J. Comput. Sci. Coll.* 35, 3 (Oct. 2019), 47–56.

[17] Jane Margolis and Allan Fisher. 2002. *Unlocking the clubhouse : women in computing.* MIT Press, Cambridge, Mass.

[18] J. Margolis, A. Fisher, and F. Miller. 1999. Caring about connections: gender and computing. *IEEE Technology and Society Magazine* 18, 4 (1999), 13–20. https://doi.org/10.1109/44.808844

[19] Jane Margolis, Joanna Goode, Rachel Estrella, Kimberly Nao, and Jennifer Jellison Holme. 2019. *Stuck in the Shallow End - Education, Race, and Computing.* The MIT Press.

[20] Tyler Menezes. 2021. personal communication.

[21] MLH. 2022. What is Major League Hacking? https://mlh.io/about Accessed: 2022-02-24.

[22] Arnab Nandi and Meris Mandernach. 2016. Hackathons as an Informal Learning Platform. In *Proceedings of the 47th ACM Technical Symposium on Computing*

Science Education (Memphis, Tennessee, USA) *(SIGCSE '16)*. Association for Computing Machinery, New York, NY, USA, 346–351. https://doi.org/10.1145/2839509.2844590

[23] OpenBSD. 2022. OpenBSD hackathons. https://openbsd.org/hackathons.html Accessed: 2022-02-24.

[24] F. C. Pampel. 2000. Logistic Regression. https://doi.org/10.4135/9781412984805

[25] PennApps. 2022. PennApps The Original College Hackathon. https://pennapps.com Accessed: 2022-02-24.

[26] Emily Porter, Chris Bopp, Elizabeth Gerber, and Amy Voida. 2017. *Reappropriating Hackathons: The Production Work of the CHI4Good Day of Service.* Association for Computing Machinery, New York, NY, USA, 810–814. https://doi.org/10.1145/3025453.3025637

[27] Kjetil Raaen and Per Lauvås. 2018. How companies find and evaluate graduate computer programmers. In *NIK.* https://ojs.bibsys.no/index.php/NIK/issue/view/41 Accessed: 2022-03-08.

[28] M. Resnick and N. Rusk. 1996. The Computer Clubhouse: Preparing for life in a digital world. *IBM Systems Journal* 35, 3.4 (1996), 431–439. https://doi.org/10.1147/sj.353.0431

[29] Amy Ru and Foaad Khosmood. 2020. Hackathons for Workforce Development: A Case Study. In *International Conference on Game Jams, Hackathons and Game Creation Events 2020* (Osaka, Japan) *(ICGJ20)*. Association for Computing Machinery, New York, NY, USA, 30–33. https://doi.org/10.1145/3409456.3409462

[30] Una Tellhed, Martin Bäckström, and Fredrik Björklund. 2017. Will I Fit in and Do Well? The Importance of Social Belongingness and Self-Efficacy for Explaining Gender Differences in Interest in STEM and HEED Majors. *Sex roles* 77, 1 (2017), 86–96.

[31] James Tetazoo. 2021. Who's going to MLH hackathons? https://news.mlh.io/mlh-hackathon-demographics-09-03-2019 Accessed: 2021-08-11.

[32] Roli Varma. 2006. Making Computer Science Minority-Friendly. *Commun. ACM* 49, 2 (Feb. 2006), 129–134. https://doi.org/10.1145/1113034.1113041

[33] G. Walton, Geoffrey L. Cohen, David Cwir, and S. Spencer. 2012. Mere belonging: the power of social connections. *Journal of personality and social psychology* 102 3 (2012), 513–32.

[34] Gregory M Walton and Geoffrey L Cohen. 2007. A Question of Belonging: Race, Social Fit, and Achievement. *Journal of personality and social psychology* 92, 1 (2007), 82–96.

[35] Jennifer Wang and Sepehr Hejazi Moghadam. 2017. Diversity Barriers in K-12 Computer Science Education: Structural and Social. In *Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education* (Seattle, Washington, USA) *(SIGCSE '17)*. Association for Computing Machinery, New York, NY, USA, 615–620. https://doi.org/10.1145/3017680.3017734

[36] Jeremy Warner and Philip J. Guo. 2017. Hack.Edu: Examining How College Hackathons Are Perceived By Student Attendees and Non-Attendees. In *Proceedings of the 2017 ACM Conference on International Computing Education Research* (Tacoma, Washington, USA) *(ICER '17)*. Association for Computing Machinery, New York, NY, USA, 254–262. https://doi.org/10.1145/3105726.3106174

[37] Womxn/Hacks. 2021. Womxn/Hacks Homepage. https://www.womxnhacks.com/ Accessed: 2021-08-12.

[38] Sharon Zukin and Max Papadantonakis. 2017. Hackathons as Co-optation Ritual: Socializing Workers and Institutionalizing Innovation in the "New" Economy. *Precarious Work* 31 (2017), 157–181.

A ONLINE RESOURCES

Supplemental materials to accompany this paper are available on the Harvard Dataverse at the following link: https://doi.org/10.7910/DVN/ORNV88. These materials include the following items:

- A description of the dataset used in this project, along with a description of the dataset formatting.
- R and Python code used in this project to conduct statistical analyses and produce data visualizations.
- A more detailed description of the demographic breakdown of the dataset.
- A detailed explanation of how the predictions were generated for the prototypical event attendees visualized in our results sections.