Coding Bootcamps and Gender Equity

John Vandivier[1]

Abstract

Insights from this study empower policymakers and employers to combat the gendered debt crisis in the United States and improve industrial gender diversity. This paper provides a systematic metastudy of the literature on hirability plus novel data and analysis to provide an explanation of the higher levels of gender diversity in coding bootcamps relative to computer science programs. Results show that women display equal preference for a programming career compared to men. Contrastive explanation from people and things personality orientation does not explain desire for a programming career, although it does explain aversion to computer science college enrollment. Evidence indicates that women prefer coding bootcamps in part due to their shorter duration and content. As an industry, information technology can encourage female participation by supporting part-time work from home and hiring individuals that do not hold a computer science degree.

Keywords: labor economics, alternative education, coding bootcamps, gender equity, debt crisis, student debt, education economics

2010 MSC: TODO

Email address: vandivier_john@yahoo.com (John Vandivier)

1. Introduction

This paper provides a review of current knowledge and original empirical evidence to solve for the present gendered debt crisis. Given that sixty percent of college students are female[1], it is no surprise that about sixty percent of student debt is held by women in the United States[2]. The problem of debt payoff is a harder task for the average woman. In 2020, the average female worker in the United States made eighty-four cents on the dollar compared to a male. The female labor force participation rate is historically lower than the rate for men, and further, the recent pandemic has disproportionately impacted women, pushing labor force participation to a 33-year low in 2021[4].

Two naive solutions to the gendered debt crisis include increasing earnings for women and reducing the college debt obtained by women. This paper will argue that these naive solutions are exactly appropriate. I will argue that digital bootcamps are an ideal tool for equitable wage normalization, and that reductions in educational debt subsidy solve the problem of excess debt retention. It turns out that non-females benefit from the same activities.

Glassdoor is a leading job search platform. Glassdoor reports that in the United States and abroad, occupational sorting alone explains the majority of the wage gap[9]. The present paper provides original research on the degree to which this sorting is voluntary. If sorting is involuntary or switching costs are high, we have identified the root cause of the issue. If sorting is voluntary and switching costs are low, disparities in wages are not a problem to be solved.

Burning Glass is another job market analytics platform, similar to Glassdoor. Burning Glass reports that in 2013, the average entry-level STEM job paid 66,123 dollars, with 52,299 dollars for the average non-STEM job. This difference represents a twenty-six percent premium for a STEM job[5]. Some institutions including NASA have begun to support initiatives related to STEAM, or STEM plus art[6]. Artistic disciplines are lower-earning than ordinary STEM roles. Institutional support of STEAM over STEM is a move in the wrong direction. It is a move that enables perpetuation or of existing wage gaps.

As the acronym grows, the salary advantage is diluted, and the reverse is also true. Top-paying degree fields include computer-related, engineering, and mathematical fields[3]. Computer-related jobs are at the top in pay but at the bottom in female program enrollment and industry employment. Pew notes that women make up half of those employed in STEM fields, but only twenty-five percent of computer jobs and fifteen percent of engineering jobs[7].

Software engineering and data engineering are at the intersection of computer-related jobs and engineering. A degree in computer science is the most common degree among software engineers[8]. Figure [FIG 1] provides historic data by gender on computer science degree enrollment, software engineering as a profession, and female graduation from coding bootcamps.

Software engineers are also referred to as software developers and a number of other job titles on the market. In 2013, the average salary for software developers was 92,820 dollars[10]. This is a forty percent premium compared to the average STEM job using the earlier Burning Glass data. In short, increasing female labor in software development is a potent strategy to increase wages. Coding bootcamps are a known path to such increased participation, and the utilization of coding bootcamps involves a double-benefit because the cost of attendence is generally lower.

If coding bootcamps are so effective, why are they not already widely used? The present metastudy answers this question in three ways. First, I document that coding bootcamps are already widely used and usage is increasing. Second, I show that usage is associated with hirability. This provides evidence for specific predictive models that are reused in this paper for an updated future outlook. Third, I review ways in which the positive rate of adoption is currently constrained and can be strategically improved.

2. Methodology

The method of this paper follows a five-step process. First, I systematically review claims in the existing literature on hirability and alternative credentials.

Second, I collect novel data for the purpose of replicating and testing claims in the existing literature. Third, I test some new hypotheses on the novel data set. Fourth, I perform a meta-analysis across current and prior findings.

The papers in this systematic meta-analysis use a technical definition of hirability that first appears in [11]. The seminal paper uses a public dataset that is hosted by Mendeley Data[12]. The first group of papers considered for inclusion in this meta-analysis are papers that cite the seminal work or the associated dataset. In addition, all papers from any author in the first group are also considered. Papers that do not provide new data nor make any new predictions or claims related to hirability are excluded.

Searches were conducted on Google Scholar, Google Dataset Search, Mendeley Data, and the Social Science Research Network (SSRN). Ten papers, including the present paper, were considered for inclusion. Five candidate papers are included in the analysis and five were excluded[1].

Each of the five papers in the meta-analysis follow a questionnaire design. Appendex TODO contains a copy of the questionnaire used to collect the novel data analyzed in this paper. This questionnaire was administered in October of 2021. Respondents are United States citizens at or over the age of eighteen. Qualified respondents participated in the survey through the Amazon Mechanical Turk platform.

Responses are investigated using an exploratory data analysis (EDA), descriptive statistics, robust feature selection, and regression analysis. Descriptive statistics focus on identifying differences in questionnaire response by gender. Feature selection and regression analysis focus on identifying whether gender or gender-interacted variables explain differences in evaluation of alternative credentials or demand for a career in programming. Novel results are subsequently compared with other findings from the systematic review.

EDA uncovered an interesting four-way interaction between the industry of

[1] Three papers are excluded because they are unrelated to a study of hirability. Two other papers are excluded because they lack new data, predictions, and claims.

occupation, gender, risk preference, and naive preference for a programming career. The subsequent section provides a detailed description of the whole dataset, but this singular finding resulted in a situation where the number of independent variables exceeded the number of samples. This fact of the dataset informs the methods of feature selection that are used.

EDA replicates a finding in prior research that skill gaps without absolute overqualification effects fit the data better than skill gaps that allow for absolute overqualification. Absolute overqualification occurs when a job candidate is perceived to have a higher level of skill compared to an ideal applicant.

I extend the concept of non-overqualification into comparative skill gaps and find the extension increases intuition without an important loss of explanatory power. A skill gap is the difference between the ideal skill level and the observed skill level. The skill gap for a recent college graduate minus the skill gap for an alternatively educated non-college graduate (ACNG) is a comparative skill gap. A positive comparative skill gap is counterintuitive because it may or may not indicate that a respondent views a recent college graduate as having relatively ideal skills compared to an ACNG.

Ignoring relative overqualification improves intuition. Relative overqualification occurs when an ACNG is perceived with higher skill than a recent college graduate. In the current data set, relative overqualification effects are small enough that ignoring them does not importantly change model explanatory power. By ignoring both absolute and relative overqualification, a comparative skill gap can be intuitively interpreted as a pure economic bad.

I refer to these effects as comparative skill gaps in the remainder of the paper. A positive comparative skill gap indicates that an ACNG is perceived to have less skill than a recent college graduate and also less than an ideal level of that skill.

Robust feature selection is accomplished by conducting multiple feature selection algorithms. These algorithms include the elastic net, univariate selection, and sequential forward-selection (SFS). Leave-one-out cross-validation is used to tune the elastic net and terminate the SFS process[2]. The elastic net is an ideal feature selection tool because it combines the benefits of ridge regression and the lasso regression while mitigating their respective drawbacks. These benefits include stable feature selection and accounting for collinearity. Elastic net can be used with linear and logistic regression, with categorical variables or otherwise, and it is a recommended variable selection tool when the number of predictors (k) is larger than the number of observations (n)[13].

Hirability is used as the main dependent variable for regression analysis, but two other dependent variables were also studied to ensure completeness and robustness of findings. Univariate selection is a trivial process that correlates features one-by-one to the dependent variable, or hirability in the case of this study. Univariate selection is not preferred to elastic net selection in large part because it does not account for collinearity. Still, features that are selected by univariate selection are relatively strong, so employing the method adds an interesting nuance to results.

Several prior studies in the meta-analysis used recursive feature elimination (RFE). Recursive feature elimination involves including a large set of variables concurrently in a model and removing them one at a time until some condition is reached. Because $k > n$ in the current analysis, it is not possible to run a long regression of this type. RFE is relatively stable and generally preferred to SFS, but SFS is otherwise similar in nature to RFE. As a result, it is a useful tool for the purpose of replication and also for checking the robustness of findings produced by the elastic net. SFS is conducted using linear regression and also using scaled vector regression, which is somewhat orthogonal to the way the elastic net is computed. This means that SFS could potentially uncover interesting nonlinear relations that the elastic net may not consider.

Finally, regression analysis is used to examine the magnitudes of the relatively small number of interesting independent variables identified through

[2]Specifically, I use the ElasticNetCV module from the Scikit-learn library in Python[21]. The analytical code and novel data used in this study are available at TODO ANONYMIZE.

robust feature selection. Ordinary least squares (OLS) regression is conducted with two different dependent variables. Logistic regression is conducted with a third dependent variable.

3. Description of Data

This paper leverages a novel set of online questionnaire responses ($n = 114$)[3]. The questionnaire includes ninety-four questions. The questionnaire is organized into ten sections. The last section is administrative in nature. The other nine sections correspond to groups of related factors. Personality information includes continuous measures of the Big Five personality traits and for grit. Other than those factors, all other variables are categorical or Likert-type responses. The majority of Likert-type questions use a 10-point scale, and such variables are often analyzed as continuous data [4].

Appendix TODO contains summary statistics for each variable. The main dependent variable of interest is hirability. Hirability is a non-price indicator of willingness to hire on the basis of an alternative credential. Two other variables that are investigated as both dependent and independent variables are employment in the information technology industry and naive preference for a programming career.

4. Results

The novel questionnaire administered for the present gender analysis resulted in one hundred and fourteen responses. The mean hirability response is 8.1 out of ten, and the median is 8. Mean hirability among males is not significantly

[3]The data is available here TODO ANONYMIZE

[4]It is an accepted practice to treat Likert-type responses as either categorical or continuous for regression analysis. Jaccard and Wan provide support for continuous analysis of Likert-type data. They note that severe departures from the assumptions on cardinality "do not seem to affect Type I and Type II errors dramatically," particularly when the Likert scale is five or more points[**?**].

different from mean hirability among females ($p < 0.51$). Eight responses were dropped for three reasons and the remaining responses include some partial responses. In the cleaned dataset ($n = 106$), mean hirability is 8.2 out of ten, the median remains at 8, and hirability remains invariant by gender ($p < 0.68$). Invariance of hirability by gender is consistent with the insignificance of gender in three of four other papers in the meta-analysis.

The purpose of the current study is to detect gender effects, but only two observations of "Other" gender were obtained. These responses were dropped because they are insufficient to yield meaningful statistical analysis. Three more responses were dropped for lack of variation in skill question responses. Three last responses were dropped as negative outliers in hirability response. These three negative outliers responded with a hirability of less than five. The remaining responses for hirability are approximately normal in distribution.

Table TODO shows dates of observation, sample sizes, and mean hirability for the five papers in the meta-analysis. The collective data indicates a positive trend over time. Two explanations of a positive trend to hirability exist prior in the literature. The earliest theory is the concept of employer-led favorability[14]. The second explanation is a positive exposure effect that results from coronavirus-induced remote activities[15]. These theories are compatible with each other and they indiciate that hirability is expected to remain high to the extent that social activity remains remote, particularly with respect to industry norms.

Risk preference did not differ significantly by gender ($n = 99, p < 0.57$) and naive preference for a programming career did not differ by gender ($n = 99, p < 0.74$). Preference for working with a peer that is a college graduate is significantly higher among men ($n = 105, p < 0.05$). When asked to agree or disagree that, "I prefer to hire or work with a person that has a college degree rather [sic] a person that holds a reputable certification or non-college credential," seventy-seven percent of men agreed while only fifty-eight percent of women agreed. High female enrollment into coding bootcamps relative to accredited computer science programs is not simply explained by differential interest in programming as a career, risk preference, or perceived value of alternative credentials.

Professional psychology suggests that men tend to be things-oriented and therefore will prefer a career in software development. The current study presents conflicting evidence in that women and men do not differ in preference for a career in programming. The conflict can be reconciled by realizing that modern software development is largely collaborative and social. The social nature of modern programming is consistent with high value for soft skills as emphasized by one paper in the meta-study[16].

A survey of software developers from Stack Overflow adds nuance to the relationship between programming and things-oriented work[17]. Programmers may specialize in a number of areas. Graphic designers, web developers, and mobile application developers are more frequently female. Systems and database administrators are more often male. Notice that those specialties that tend female are associated with user interfaces and graphics. This indicates that an explanation from orientation towards things or people works as a partial explanation of specialty selection, but the explanation fails at the aggregate level of the modern software development job family. Contrasting holistic and typical software development practices with the curriculum of a typical computer science program indicates an emphasis on things that may discourage female interest in the field of study, even while interest in the programming occupation persists.

The absence of differences in means is noteworthy, but a better analysis accounts for interactions and collinearity between gender and other variables. As previously noted, risk aversion did not differ significantly by gender. This is surprising, because substantial external literature demonstrates a link between gender and risk aversion[18, 19]. Further research invariant to gender shows a link between risk aversion and occupational preference[20]. These prior findings motivated exploration of a three-way interaction that turned out to significantly explain hirability. Further exploration showed that naive preference for a programming career varies importantly with respect to the three-way interaction.

When categorical response options are treated as distinct dummy variables, the ninety-four question survey produces one hundred and fourty-three variables. When the four-way interaction plus related lower-level interactions are counted, there are two hundred and twenty-four variables.

These variable pools were seperately subjected to elastic net regularization with leave-one-out cross-validation. Male identification is the only gender variable in the pool of one hundred and fourty-three variables. When these factors are ranked for importance by their contribution to out-of-sample explanatory power, gender identification is retained in the top fourty percent of variables, and it is retained in model that maximizes out-of-sample R^2.

When the variable pool that includes interactions is reduced, many gender-interacted variables are retained in the model that maximizes out-of-sample R^2 ($\lambda_1/\lambda_2 = 0.09, k = 48, R^2 = 0.71$). From a pool of two hundred and twenty-four variables, fourty-eight variables are retained in this model. This approximately represents the top twenty percent variables in the pool by importance. Eleven of these important variables are gender-related. This confirms the idea that interacted gender effects are more important than gender alone. Interestingly, the single most important gender effect using this approach is gender interacted with risk preference, preference for a programming career, and occupation in the information technology industry. This aligns with the special relationship that has been observed between alternative education and the information technology industry.

In addition to the elastic net, four kinds of sequential factor selection (SFS) are tested to identify potentially important features. The point of this exercise is to identify features in a way that is robust to multiple analytical approaches and multiple concepts of importance. Two of the four SFS procedures use univariate selection. One approach was to use the Benjamini-Hochberg procedure to estimate the false discovery rate. Features were included up to an allowed false discovery rate of fifty percent. The secon univariate approach selects combinations of features that minimize the family-wise error rate. Features were included up to a family-wise error rate of fifty percent.

The other two SFS procedures use forward selection of variables in a multiple regression. The first approach is a multiple support vector regression (SVR) that minimizes mean squared error. The second approach is an ordinary least squares (OLS) multiple regression that optimizes for median absolute error. OLS optimization against mean squared error was avoided to preclude the linear analysis from becoming an extraneous subset of the vector analysis. In both cases, forward selection continued until fifty-six variables were selected. This is equivalent to taking the top twenty-five percent of variables.

One hundred and twenty-six features are identified as important by at least one of the six feature selection approaches. The next important result relates to the direction of the coefficients of these variables. Before those coefficients are identified, however, additional steps must be taken to eliminate overfitting. This is because several of the feature selection approaches do not account for collinearity. Included variable bias is also an important concern to mitigate.

To address these concerns in an inferrential study, backward feature elimination (BFE) optimizing on the p-value of each feature is used. When $p > 0.5$ it becomes likely that a coefficient is equal to zero. Such variables should be eliminated to avoid included variable bias and overfit. The sample size of the current study precludes multiple regression of one hundred and twenty-six features. As a second-best approach, I reduce the variable pool using supervised and theory-driven stepwise regression.

I begin by applying BFE to variables identified by the elastic net, exluding gender interactions. Sixty-seven features survive this reduction step. I then add features identified by the elastic net inclusive of interacted gender. I apply BFE again, then I add SFS-selected variables, and apply BFE again. In the end, the original pool of one hundred and twenty-six features is reduced to a pool of seventy-six features. Each feature in this model has $p < 0.5$, the model fully explains the observed data ($R^2 = 0.997$). The Akaike information criterion (AIC) for the reduced model is substantially better than the AIC for a model built on the results of the elastic net alone. Table TODO shows the coefficients for this model.

In the preferred model, gender is independently important, as well as various gender interactions. Male identification is negatively associated with hirability. Gender interacted with naive preference for a programming career is positive with a large magnitude. The four-way interaction with gender and the information technology industry is significant and negative, but small. Summing across relevant effects, the total effect of male identification within the information technology industry is positive on hirability[5]. This result indicates that men in the tech industry are actively encouraging the use of alternative credentials. This advocacy has a positive impact on women, because coding bootcamps disproportionately enroll women compared to college computer science programs.

Two of five papers in the meta-analysis identify gender as important[6]. All three papers that did not identify gender as important used OLS estimation and selected a preferred model by maximizing the adjusted r-squared value. The current paper replicates that a simple measure of gender is not included in the OLS model that maximizes adjusted r-squared. This paper argues that such a model is underfit, however, because it omits factors where $p < 0.5$ and therefore ignores likely true effects. In addition, the preferred model in this paper has a lower AIC. In addition, in an underfit model that maximizes adjusted r-squared, simple gender is not an included factor but many gender interactions are retained. Of note, the four-way interaction that includes gender and the information technology industry is retained in such a model. Given these results, the claim in the current paper that gender has an important effect should not be taken as a flat disagreement with any other paper in the meta-study.

Further robust validation of the preferred model, associated factors, and theory is accomplished by using the same model to explain two other independent variables. In theory, an ideal economic system empowers individuals with economic mobility that is not restricted to income mobility. If an individual in such

[5] Among male respondents in the information technology industry, the average level of risk seeking is 6.4 and the average naive preference for a programming career is 7.6. Within the information technology industry, then, the total male identification effect is: $0.6 = \beta_m + \beta_p(X_p) + \beta_r(X_r) + \beta_{pr}(X_{pr}) = -0.765 + 0.634(7.6) - 0.325(6.4) - 0.028(7.6 * 6.4)$

[6] This paper and [11].

a system desired to become a programmer, a strong correlation with actually becoming a programmer is to be expected. It is troubling to some small extent that current discrepencies by gender in the labor force exist. Gender-related exit effects and historical effects could theoretically explain discrepencies in current labor without implying a current problem of gender-related economic mobility. The greater problem is that women desire to program as much as men, but fewer females in the software engineering labor pipeline are observed.

Two other dependent factors that were analyzed include present employment in the information technology industry and naive preference for a programming career. Both dependent factors are substantially explained with minimal modification to the preferred model of hirability. Both of these factors are independent factors in the preferred model of hirability, so dropping them as independent factors is one part of the required minimal modifications. After the change in specification, some variables obtained a large p-value, so BFE to a p-value of 0.5 is applied to improve AIC and avoid included variable bias. Naive preference for a programming career is modeled using OLS and employment. Employment in the information technology industry is a binary outcome, so it is investigated using logit and probit.

Table TODO shows the direction of effect for factors in the preferred models of hirability, naive preference for a programming career, and current employment in the information technology industry. A general expectation of a consistent direction of effect is based on the theoretical positive correlation between these variables. A case in which variables obtain signs in opposite directions is no threat to any particular theory or finding, but it should be taken into account when strategies for creating change are formed. Specifically, factors of naive preference for a programming career might inspire a strategy to increase female preference for a programming career, but such a strategy would probably want to be developed in a way that avoids reducing hirability or current employment in the information technology industry.

The logit and probit analysis show that logit is generally a better fit. Many factors are cut from these models to avoid a singular matrix or excessive quasi-seperation. Gathering additional samples would likely allow singular matrix and quasi-seperation to be avoided without needing to cut as many factors. In short, there are reasons to think that the logit model is underfit, so one potentially important finding should be noted with skepticism. The finding is that gender is an important factor in the preferred logit model, but naive preference for a programming career is excluded due to a lack of significance.

Logistical modeling prior to BFE identifies a large positive coefficient on naive preference for a programming career ($\beta = 0.62, 0.5 < p < 0.51$). Forcing the factor into the preferred logit model identifies a small positive coefficient ($\beta = 0.02, p < 0.96$). An average of these effect estimates provides a magnitude comparable to the effect for gender[7]. Because a positive coefficient is robustly identified and consistent with theoretical expectation, it is difficult to accept the null hypothesis of a coefficient of zero, despite the high p-value of this factor. It seems more accurate to claim that the logistic analysis lacks the power to precisely estimate the magnitude of the coefficient, but a best estimate of the effect puts it on par with gender.

The logistic analysis does establish that gender better explains current employment in the information technology industry compared to naive preference for a programming career. This result is consistent with a narrative of sexism in the information technology industry or the hiring process, but I argue that this is an inappropriate interpretation of the result. This interpretation is incorrect because having a programming job is importantly determined by factors other than the ability to obtain such a job. Specifically, exit effects and historical ef-

[7]The average coefficient estimate for naive preference for a programming career is $0.32 = (0.62 + 0.02)/2$. This factor is measured on a scale from one to ten, so the effect can be multiplied by five for a rough comparison with gender. The resulting dummified coefficient of 1.6 is not meaningfully different from the preferred effect estimate of the gender dummy at 1.65. Also compare the estimated total average effects. The mean response for naive preference for a programming career is 7.54, so the total estimated effect is: $2.4 = 0.32(7.54)$. The estimated effect for male identification is: $1.1 = 0.65(1.65)$. The effect for gender is identified more confidently, but it is not clear that it is larger in magnitude.

fects are large and important. Census data from 1970 shows that females made up about 22.5 percent of programmers[?]. [FIG 1] showed that the proportion of men in programming has trended sideways since 2017 at about ninety percent. These historical facts allow for male identification to correlate with holding a programming job, even if men and women are hired in proportion to their application rate. Further, if men and women are hired proportionally, but women leave disproportionately, men will eventually dominate the labor pool.

Women exit industry at a higher rate than men across industries. The need to care for children is an important factor of the decision to stay at home for American women[22]. Wealthy women decide to stay at home more frequently than do either lower-income or middle-income women. Because programming is a lucrative field, and Americans currently engage in a pattern of assortive mating, female programmers will tend to have higher personal and household earnings. This might lead to high industrial exit. As a result, even with equitable hiring practices, a high share of male labor could be explained by historical and exit effects.

The model that explains naive preference for a programming career establishes little when assessed in isolation, but some results are interesting when considered alongside prior findings. Previous resarch found that alternative learning provider exposure effects are positive on hirability. In the present model we can see that exposure is also associated with positive naive preference for a programming career.

Previous research on skill gaps suggested curricula changes that would enable candidates to increase hirability in the eyes of employers. Specifically, an increased focus on soft skills like teamwork and communication were advised. Current results show that a perceived skill gap in teamwork for alternatively educated individuals is negatively associated with naive preference for a programming career. An opposite effect is found for communicaiton skills. This could indicate that shrinking skill gaps for teamwork are more important than shrinking skill gaps for communication from the perspective of alternative education consumers. A variety of other communication and marketing strategies are suggested to programming education providers on the basis of age, location, and other effects.

Table TODO shows explanatory power by factor group on hirability. Independent factors for ethnicity, industry of employment, and state of residence have been consistently identified as important in other papers, so these factor groups provide a meaningful standard for identifying other factors as important. Table TODO (the same table) also includes explanatory power of the preferred linear model, the model without gender-derivative features, and a model of the gender-derivative features in isolation.

5. Conclusions

This paper systematically investigated the research on hirability to understand the relationship between gender and coding bootcamps. This analysis contributes to the development of a solution to the gendered student debt crisis in the United States. The results of this paper may also be used by employers seeking to improve employee gender diversity, by learning providers, consumers of education, policymakers, and many others. A number of results in the literature were replicated, refined, and extended. Novel data and analysis were also presented.

A simple model of the labor composition of an industry must account for historical labor composition, entry, and exit. Ensmenger gives a thoughtful treatment to the history of the subject[**?**]. A simplistic summary would be to say United States government statistics indicate that men have made up roughly eighty percent of professional computer programmers from 1970 through 2018. Government statistics consistently show a ten percent difference compared to Stack Overflow developer survey results, but the trend is similarly horizontal through 2021[23].

NPR provides a popular explanation from culture for the small share of female programmers[25]. In this story, male-targeted adverstisements led to a reduction in the share of female computer science enrollment. First, this misses important gender-related enrollment effects that have to do with post-war male educational enrollment. Second, the share of female computer science enrollment does not seem to be strongly related to the share of professional female programmers.

Ada Lovelace is widely considered to be the first programmer despite lacking a computer science degree[26]. In 1980 and prior, women were a general minority in college rather than uniquely a minority in computer science programs[27]. In 1970 the share of professional female computer programmers was already 22.5 percent. That number exceeds the share of female computer science students by over sixty percent[28]. Professional female programming appears to historically precede and often exceed female computer science program enrollment, rather than to lag it. From 1970 to today, the share of female computer programmers has hardly changed despite large variation in the share of female computer science enrollment. Third, and most directly, the current study shows that men and women do not differ in their naive preference for a programming career. The gender-specific cultural bias for programming does not appear to exist at the individual level.

Psychological literature indicates that women are significantly more people-oriented than men, but the present paper notes that modern programming is not a decidedly things-oriented industry. Indeed, employers see a deficit in soft skills among college and bootcamp graduates alike. This paper improved such analysis by directly investigating naive preference for a programming career. An interesting result is there there was not a difference in mean response by gender. This seems to present a problem because the industry is predominantly male, but it could be that historical effects and exit effects explain this descrepency without maligning the hiring process, so further research is needed.

The current analysis showed that gender effects are independently important after holding constant this kind of brute preference for work and also while holding constant other psychological variables including variables for Big Five personality traits.

Individual desire to stay at home is an important factor that could be explored in future research. This measure works as an opportunity cost of employment that may disproportionately impact women. If men and women both have a high naive preference for a programming career, but women have a higher relative desire to stay at home, then the true preference for a programming career net of opportunity cost is effectively higher for men. Computer programming is already a remote-friendly job family, but facilitating part-time employment might boost female participation among women that would generally like to stay home but perhaps take on an occassional high-paying project.

A strategy of incentivizing women that stay at home to pick up part-time remote work is given additional empirical merit by a study from the Bureau of Labor Statistics (BLS). The BLS found that married white women were the most likely group to choose part-time work[24]. As earlier discussed, wealthy women are more likely than other women to choose to stay at home. Given that married white women are a large component of wealthy women, enabling part-time programming work from home is a strategy with the potential to substantially increase the share of female programmers.

Larger sample sizes are suggested as an improvement for future research that intends to explain individual-level employment in the information technology industry. Causal analysis and dynamic analysis would be useful improvements as well. A model of naive preference for a programming career in the present study disclosed many correlated variables, but experimental or causal analysis would better inform strategic application of results. For instance, a large negative effect on hirability for the state of New York is identified in the current study. Does it follow that marketers should avoid geotargeting ads to New York? Precisely the opposite might be true. Exposure to digital ads might improve public opinion in New York at a greater rate per dollar spent compared to purchasing digital ads in a geography where opinion is already high.

An explanation from things orientation does not substantially explain naive preference for a programming career, but it does seem to help explain the low enrollment of women into computer science programs. Table TODO shows the percentage for various college majors among respondents that write code professionally and studied at the university level[8]. This table shows the percentage of male programmers that studied with a given major. Notice that men comprise a larger share of computer science and information technology programs. Female participation can be encouraged by broader use of web development and design programs. Coding bootcamps often emphasize web development, which partially explains their relatively high female participation. Gender diversity is also encouraged by hiring individuals with a degree unrelated to programming.

Women have earned the majority of doctoral degrees in the United States beginning with the Fall cohort of 2008[https://www.aei.org/carpe-diem/women-earned-majority-of-doctoral-degrees-in-2019-for-11th-straight-year-and-outnumber-men-in-grad-school-141-to-100]. In 2017, women earned fifty-three percent of doctoral degrees. In the Fall of 2017, women made up 54.8 percent of 4-year enrollments, 60.0 percent of two-year enrollments, and 65.8 percent of less-than-two-year enrollments. Women made up a larger share of part-time enrollments in each category. Part-time students that were enrolled in a less-than-2-year program were 66.8 percent female. [9] The trend indicates that shorter programs trend disproportionately female. This partially explains the relatively high enrollment of women into coding bootcamps compared to computer science programs.

`TODO: format citations`

[1]https://www.wsj.com/articles/college-university-fall-higher-education-men-women-enrollment-admissions-back-to-school-11630948233

[2]https://educationdata.org/student-loan-debt-by-gender

[3]https://fortune.com/2021/09/01/highest-paying-college-majors-starting-salaries/

[4]https://nwlc.org/resources/january-jobs-day-2021/

[5]https://www.burning-glass.com/research-project/stem/

[8]This data is derived from the 2020 Stack Overflow Developer Survey[30].

[9][https://nces.ed.gov/pubs2018/2018151.pdf] 4-year: 54.8 = 869,804/(716,453+869,804) 2-year: 60.0 = 11,515/(7,692+11,515) less-than-2-year: 65.8 = 2,053/(1,068+2,053) part time, less-than-2-year: 66.8 = 377/(187+377)

[6]https://www.nasa.gov/press-release/nasa-announces-virtual-webb-steam-day-event-for-students-educators

[7]https://www.pewresearch.org/science/2021/04/01/stem-jobs-see-uneven-progress-in-increasing-gender-racial-and-ethnic-diversity/

[8]https://www.careerexplorer.com/careers/software-engineer/education/

[9]https://www.glassdoor.com/research/gender-pay-gap-2019/

[10]https://codesubmit.io/blog/the-evolution-of-developer-salaries/#from-2013-to-2019

[11]https://www.tandfonline.com/doi/abs/10.1080/13504851.2020.1767279?journalCode=rael20

[12]https://data.mendeley.com/datasets/75t88s6sdt/1

[13]https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1467-9868.2005.00503.x

[14]Preliminary Attitudinal Trends in Alternative Postsecondary Learning

[15]COVID-19 and Alternative Postsecondary Learning

[16]Conformity and Soft Skills as Determinants of Alternatively Credentialed Non-College Graduate Hirability

[17]https://insights.stackoverflow.com/survey/2017

[18]https://academic.oup.com/jeea/article-abstract/7/2-3/649/2296043

[19]https://psycnet.apa.org/buy/1999-13573-004

[20]https://www.sciencedirect.com/science/article/pii/S105353571000154X

[21]https://scikit-learn.org/stable/about.html#citing-scikit-learn

[22]https://www.deseret.com/2019/1/24/20664077/rich-and-poor-married-moms-are-more-likely-to-stay-at-home-with-the-kids-but-for-entirely-different

[23]https://insights.stackoverflow.com/survey

[24]https://www.bls.gov/opub/mlr/2018/article/pdf/who-chooses-part-time-work-and-why.pdf

[25]https://www.npr.org/sections/money/2014/10/21/357629765/when-women-stopped-coding

[26]https://www.scientificamerican.com/article/in-celebration-of-ada-lovelace-the-first-computer-programmer/

[27]https://nces.ed.gov/programs/digest/d17/tables/dt17_303.70.asp

[28]https://www.gcu.edu/blog/gcu-experience/analysis-women-computing

[29]https://medium.com/@uw_data_scientist/girls-are-not-more-likely-to-drop-out-of-cs-and-engineering-70e03fdbfca

[30]https://insights.stackoverflow.com/survey/2020#developer-profile-undergraduate-major-professional-developers2