The Gender Pay Gap in the United States Tech Industry is Largest for Women with 20 to 30 Years of Professional Coding Experience[*]

Morgaine Westin

26 April 2021

Abstract

Despite promises from prominent companies to improve representation and diversity in the workforce, representation of women and racial minorities in the tech industry continues to be a significant issue. A serious consequence of gender and racial inequalities in the workforce is the wage gap, in which women and women of colour in particular receive lower salaries than their male counterparts. Using data from the 2020 Stack Overflow Developer Survey, multiple linear regression was performed to model differences in salaries among tech professionals based on gender, while taking into consideration the effects of how gender may interact with other factors such as ethnicity and years of experience. Although there were no differences in reported salaries between men and women with less years of professional coding experience, women with 20-30 years of experience reported significantly lower salaries than men with similar levels of experience, highlighting concerns about long-term retention for women in the tech industry and the underrepresentation of women in senior-level positions.

Keywords: gender inequality, technology industry, wage gap, race, intersectionality, retention, propensity score matching, multiple linear regression

1 Introduction

In 2019, the United States tech industry had 12.1 million employees and nearly 4.6 million job postings, many of which were for new emerging tech areas (CompTIA 2020) As the tech industry continues to grow and reports indicate the median wage for tech professionals is almost double the median national wage in the United States, more people are looking to enter the field. However, despite industry growth and the number of opportunities available, the tech field has been consistently dominated by white or Asian men (Harrison 2019). Gender and racial disparities in science, technology, engineering, and medicine (STEM) fields have been well documented. Notably, women and people of colour, particularly those identifying as African American, Latinx or Hispanic, or Native American, have been disproportionately underrepresented in these fields. (Jackson, Starobin, and Laanan 2013). In response to growing calls for diversity, large tech companies such as Google, Microsoft, and Facebook have launched initiatives and pledges to help support underrepresented groups both within the workplace, and externally through funds, outreach programs, and diverse recruitment practices (Harrison 2019; Banjo and King 2020). Whether or not these initiatives are effective is up for debate, with critics calling some diversity pledges "hollow" and arguing that companies need to make more meaningful efforts before we start seeing significant diversity improvements in the data (Rooney and Khorram 2020; Twine 2018).

One contributor to the diversity problem in tech is retention, or the rate at which underrepresented groups leave the field. Retention and the "leaky pipeline" are frequently discussed in the literature surrounding representation in STEM education, where many women and racial minorities express interest in STEM fields,

---

[*]Code and data are available at: https://github.com/westinmo/2020-dev-survey-analysis.

but end up switching to different programs early on, with very few individuals in these groups going on to pursue graduate degrees or professional positions (Asai 2020). This extends to the professional world as well, where women and underrepresented racial minorities take up a reasonable portion of entry-level or new tech positions, but are more likely to leave their jobs, and less likely to hold more senior positions in the company (Tapia and Kvasny 2004). Results from a 2017 report which examined why people voluntarily left their jobs in tech found that many women and people of colour were frequently passed over for promotions, and many experienced stereotyping and discrimination in the workplace (Scott, Klein, and Onovakpuri 2017). This suggests company diversity initiatives should not only focus on diverse recruitment and hiring practices, but on creating inclusive and safe workplaces while generating more opportunities for underrepresented groups to grow within the company.

Representation of women in tech in particular has been widely discussed. Among the various STEM fields, the gender gap in computer science is one of the most apparent, where the number of women graduating with a computer science or engineering degree has been decreasing since 1983 (Singh et al. 2007). Even so, women who persist in these degrees often face more challenges when trying to enter the tech industry. A survey conducted by the organization Girls Who Code which aimed to characterize the challenges faced by college-aged women when applying for technical positions found over half of female respondents reported a negative experience, or reported knowing a woman who has had one (Girls Who Code 2019). Many of these women received dismissive, demeaning, and gender-biased remarks in the interview process, with some women experiencing sexual harassment in the form of inappropriate sexual comments and flirting (Girls Who Code 2019). Issues such as these carry on past the early hiring stages and into toxic and sexist workplaces, causing many women to leave their positions due to concerns such as sexual harassment (Scott, Klein, and Onovakpuri 2017).

Much of the research conducted on representation in STEM has focused on the experiences on white, middle class women, and has failed to take an intersectional approach to address how factors such as gender, race, and class interact to put certain groups at a greater disadvantage (Alfrey and Twine 2017). Women of colour face a "double bind" when it comes to representation in STEM fields, and are more likely to fall between the cracks of intervention programs designed to improve diversity (Malcom and others 1976). For instance, programs designed to increase representation for racialized groups are often geared towards men, while programs designed to increase female representation are often biased towards helping women in majority groups (Malcom and others 1976). Finally, individuals who are transgender or non gender conforming, such as non-binary or genderqueer are also though to be disproportionately underrepresented in the tech industry, however there is significantly less data and research done to examine the experiences of these individuals in tech compared to other underrepresented groups (Dickey 2015).

A major consequence of under-representation and gender and racial disparities in the workplace involves wage gaps. In 2019, a report from the tech job platform Hired found that 60% of men were offered higher salaries than their female counterparts for the same position at the same company, and that women were offered 3% less salary on average (Hired 2019). For many women who discovered they were being paid less than their male counterparts in the same position, the difference in salary was at least $20,000. Moreover, the wage disparities grow when examining both gender and race. While White women and Asian women earn approximately $0.97 for every dollar earned by White and Asian men, Black women and Latinx women earn $0.89 and $0.91 respectively (Hired 2019). This highlights the need to take an intersectional approach when examining wage gaps, as disparities exist even among women.

In this report, I utilized data from the 2020 Stack Overflow Developer Survey (Stack Overflow 2020) to examine representation in the tech industry, and to look for potential wage disparities based on gender and ethnicity. Multiple linear regression modelling was implemented to test the effects of gender and ethnicity on salary for tech professionals in the United States. The first section of the report provides an overview of the Developer Survey and its main results, particularly regarding the demographic characteristics of the respondents[1]. The next sections discuss the process of building and implementing the regression model used to model salary, and its results. Finally, I discuss the implications of my findings in the broader context of representation and bias in the tech industry, and the limitations of my work and how it can be extended

---

[1]Survey responses for full-time tech workers in the United States can be explored here: https://mwestin.shinyapps.io/Dev-Survey-Shiny-App/

upon. The data was prepared and analyzed in R (R Core Team 2020), primarily using the `tidyverse` (Wickham et al. 2019) package, while this report was compiled using R Markdown (Allaire et al. 2021). A full list of packages used in this analysis and report can be found in the Appendix.

2 Data

2.1 Overview of the Stack Overflow Developer Survey

Stack Overflow is a popular public platform for individuals to ask and answer questions, share knowledge, or collaborate on a wide range of topics related to computer programming (Sewak et al. 2010). The website boasts an average of 100 million user visits per month, and is used by individuals who code from a wide range of backgrounds, including professionals and enthusiasts alike (Stack Overflow 2021). In 2011, Stack Overflow launched their first Annual Developer Survey to better understand its user base. Since then, the annual survey has continued to grow alongside the website's popularity, with the most recent 2020 survey garnering close to 65,000 responses from users in over 180 countries. While there is some variation in the questions asked from year to year, the Developer Survey covers a wide range of topics related to the experiences of developers, or other people who code. This includes demographic characteristics, questions about job hunting and satisfaction, compensation, programming experience, and the different languages and libraries developers are using. Every year Stack Overflow posts a report overviewing the main findings of the Developer Survey, as well as the full anonymized dataset to download which is available under the Open Database License (ODbL).

I utilized data from the 2020 Developer Survey[2] to address my research questions regarding representation and income inequality in tech due to the popularity and usage of Stack Overflow among tech professionals and wide reach of the survey. Of course, this survey data is not necessarily representative of all tech professionals, as it was a voluntary survey and mainly limited to Stack Overflow users. Respondents were mainly sourced from onsite messaging, blog posts, email lists, Meta posts, banner ads, and social media posts, which might suggest users who were more engaged with Stack Overflow were more likely to notice the survey links and complete the survey. For the 2020 survey, Stack Overflow made additional efforts to diversify their sample by finding ways to advertise the survey outside of their own channels and target coders who may not frequent their websites, as well as promote the survey and provide outreach to underrepresented coders. According to the official report, this resulted in slightly more responses from underrepresented groups compared to previous year, however Stack Overflow acknowledged they still have work to do to increase representation in their sample.

The publicly posted dataset for the 2020 survey contains 64461 observations with 61 variables containing answers to the survey questions. Free response questions and questions with personally identifying information were not included in the dataset. A number of responses that Stack Overflow deemed "unqualified" for analysis were removed from the dataset, namely those where respondents took less than 3 minutes to complete the entire survey, which was estimated to take around 20 minutes to complete. Given the size of the dataset and range of respondents, I narrowed down the responses to focus on salaries for tech professionals living in the United States who are employed full time[3]. Of the 9765 respondents who reside in the United States and are full-time employees, 6368 also reported their income, gender, and ethnicity. Potential response bias should be considered when interpreting the results based on this data, as there may be underlying differences between respondents who reported these characteristics and those who did not. This affects both the extent to which the data is representative of tech professionals in the United States, as well as the generalizability of the model and its results.

---

[2]The overall results of the 2020 survey including more details about the survey's methodology can be found here: https://insights.stackoverflow.com/survey/2020

[3]The survey ran in February of 2020 before the World Health Organization officially declared COVID-19 to be a worldwide pandemic, so factors involving employment status and income may have changed since then

2.2 Variables

While the data from the Stack Overflow survey covers a wide range of questions about demographics and work characteristics which all provide valuable information about the tech industry, I focused my analysis on a select few variables of interest which relate to my research question. The variables I will be discussing and focusing on in my model are: salary, gender, ethnicity, and years of professional coding experience, while other variables such as developer type, age, and education level will be controlled for. When reporting salary, respondents were also asked to indicate whether the salary they input was weekly, monthly, or yearly, as well the currency. These responses were converted to annual salaries in USD by Stack Overflow based on an assumption of 12 months and 50 working weeks. Stack Overflow also trimmed the top 2% highest annual salaries in the dataset and replaced them with a threshold value equivalent to two million USD. However, to help improve model performance by reducing extreme skewness, I removed cases with annual income over $400,000[4]. In addition to this, I excluded responses where the reported annual salary was less than $10,000, as many of these responses appeared invalid and were unlikely to reflect true full time salaries[5]. After removing these responses, the overall median salary was $110,000 USD.

Respondents were asked to report their gender identity by selecting one or more of the following options: Woman, Man, Non-binary, genderqueer, or gender non-conforming, a text-entry option, and a non-response option. Out of the respondents who reported their gender identity, 87.5% identified as male, 10.4% identified as female, and 2.1% identified as non-binary, genderqueer, or gender non-conforming. As expected, female and non-binary respondents are grossly underrepresented in this sample, although there was an increase in responses from these groups compared to surveys from previous years. Reports from the National Center for Women & Information Technology (National Center for Women & Information Technology 2016) and the Pew Research Center (Pew Research Center 2021) estimate women hold around 25% of tech and computing jobs in the U.S., which suggests the numbers from the Stack Overflow survey may be an underestimate of the true number of women in the tech workforce and the results may not be entirely representative of that population. Distributions of salaries by gender and summary statistics are shown in Figure 1 and Table 1.
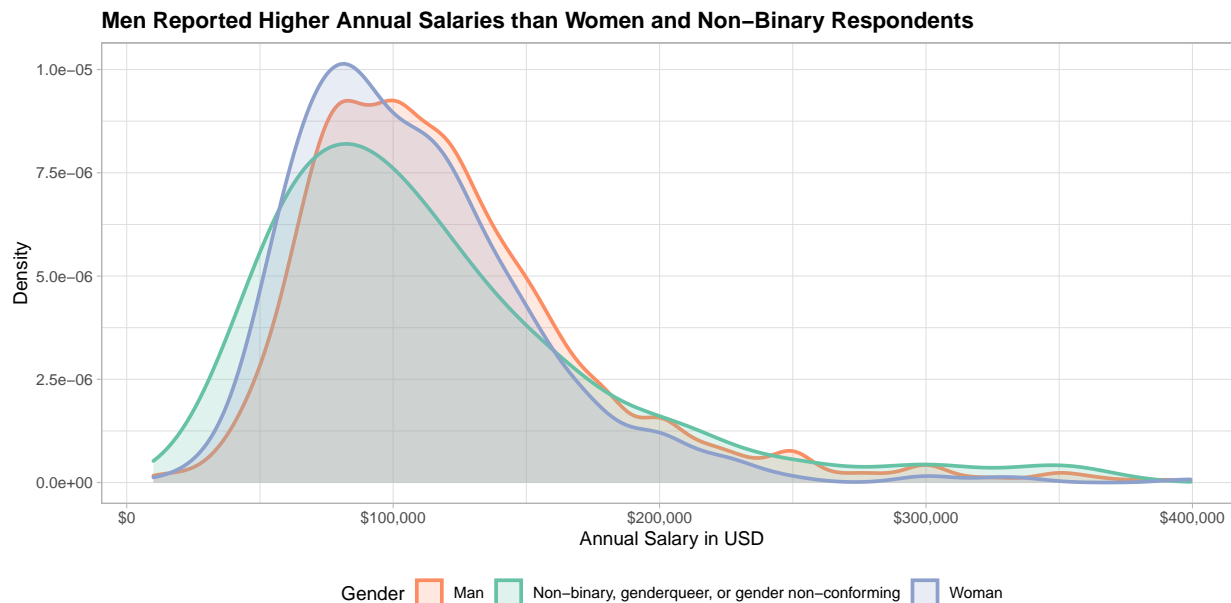

Men Reported Higher Annual Salaries than Women and Non–Binary Respondents

Figure 1: Salary Distributions for Full-Time Developers by Gender

---

[4]252 responses with annual salaries over $400,000 were removed, the majority of which were White males

[5]17 responses with annual salaries under $10,000 were removed, and the annual salaries of all except one case were under $500. Again, the majority of these respondents were White males.

Table 1: Summary Statistics for Salaries of United States Tech Workers

Gender	Respondents	Median	Mean	Standard Deviation
Man	4167	110000	119856.6	54333.32
Non-binary, genderqueer, or gender non-conforming	101	100000	116126.4	64386.71
Woman	496	100000	108263.7	46899.74

On average, median reported salaries for men were higher than they are for women and non-binary individuals, whose median reported salaries were both $100,000 USD. Non-binary respondents show the broadest range of salaries, and show higher mean salaries along with men compared to the salaries of women. This can be seen on the tail end of the density graph in Figure 1 where more individuals identifying as male or non-binary reported salaries on the higher end (upwards of $200,000 USD) compared to women. These salary distributions by gender are similar to those found in a previous analysis of the 2019 Stack Overflow Developer Survey (Silge 2019), which found the reported median salary of men was around $8,000 higher than that of women and non-binary respondents, who also reported similar salaries. While we can clearly see differences in salaries based on gender, whether or not these are meaningful differences is yet to be seen, as there appears to be a complex relationship between gender and salary in which a number of other factors need to be considered as well.

2.2.1 Ethnicity

To measure ethnicity, respondents were asked to select all options that applied to them from a predefined list of responses, along with text-entry and non-response options. To simplify analysis and avoid the overweighting of certain responses, I filtered out respondents who selected more than one ethnicity.[6] The distribution of respondents' ethnicities by gender are shown in Figure 2, with the bottom chart focusing on the ethnicities of the women in the sample.

Consistent with previous reports about racial representation in tech, an overwhelming number of respondents reported being White or of European descent. Respondents who identified as Indigenous, which included those who are Native American, Pacific Islander, or Indigenous Australian, were the least represented group in this sample. Interestingly, respondents identifying as Hispanic or Latinx were the third largest group despite reports that they are underrepresented in tech compared to other groups such as East Asians. Notably, in this subset of the sample there were no Indigenous women, and the sample sizes for non-White women were extremely small. Again, while a number of responses were removed prior to this analysis, the distribution of respondents' ethnicities is generally consistent with those in the entire dataset, as well as survey results from previous years.

Median salaries organized by ethnicity and gender are shown in Figure 3. For respondents who identified as White, South Asian, Middle Eastern, and East Asian, the median salaries for males were higher than their female counterparts. The opposite was true for respondents who identified as Southeast Asian or Hispanic/Latinx, where median salaries were higher for women. Note that there were no Indigenous women in this group, and that median salaries for Black men and women were the same. After Indigenous men, White women had the lowest median salaries compared to the other groups. However, this could be attributed to the fact that the majority of women in the sample were White and sample sizes for the other women in this group were very small, and women were overall more likely to have less professional coding experience, resulting in lower salaries.

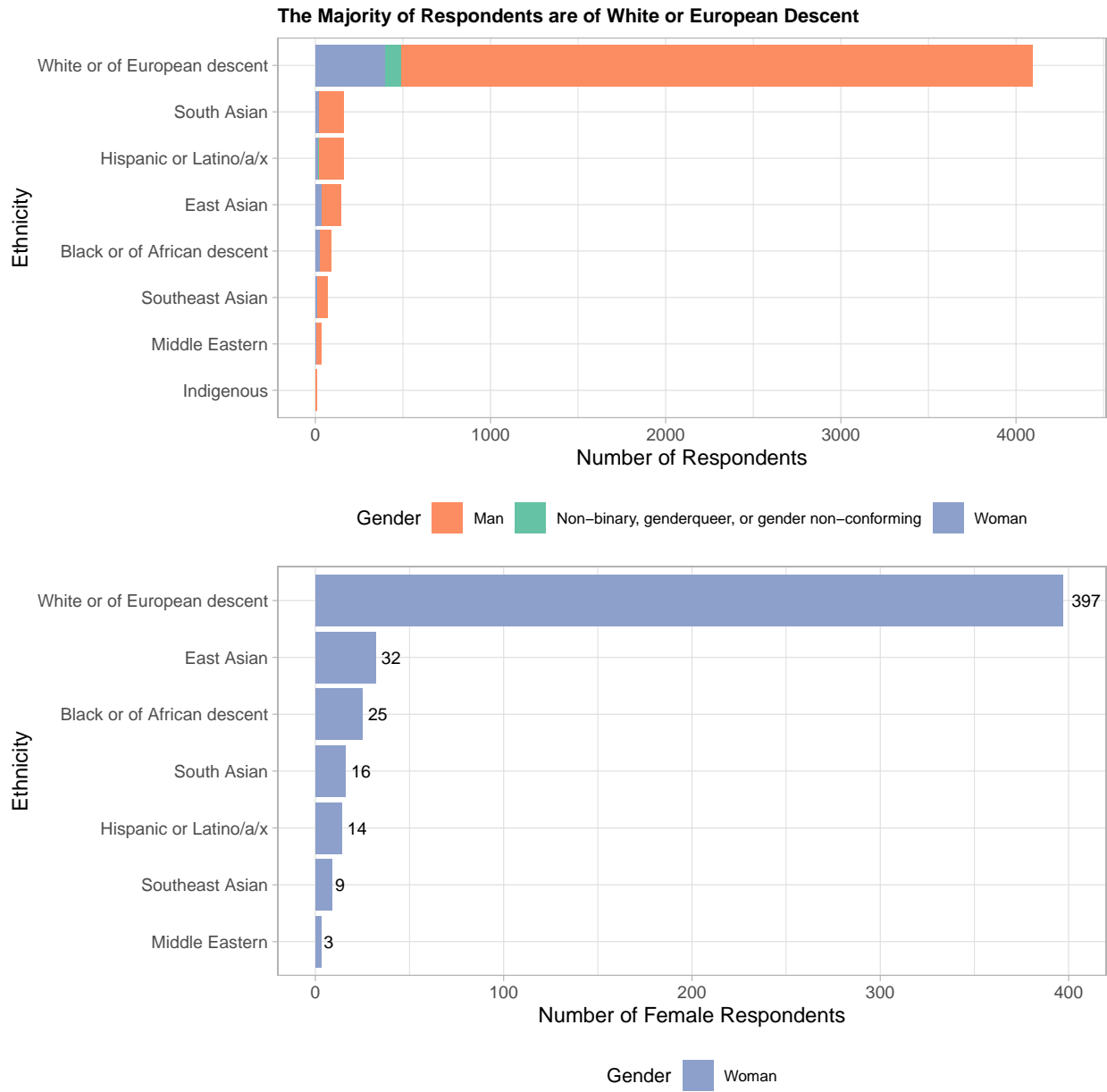[6]This resulted in 494 respondents being removed, the majority of which were male

The Majority of Respondents are of White or European Descent

Gender ■ Man ■ Non–binary, genderqueer, or gender non–conforming ■ Woman

Gender ■ Woman

Figure 2: Distribution of Respondents' Ethicities by Gender

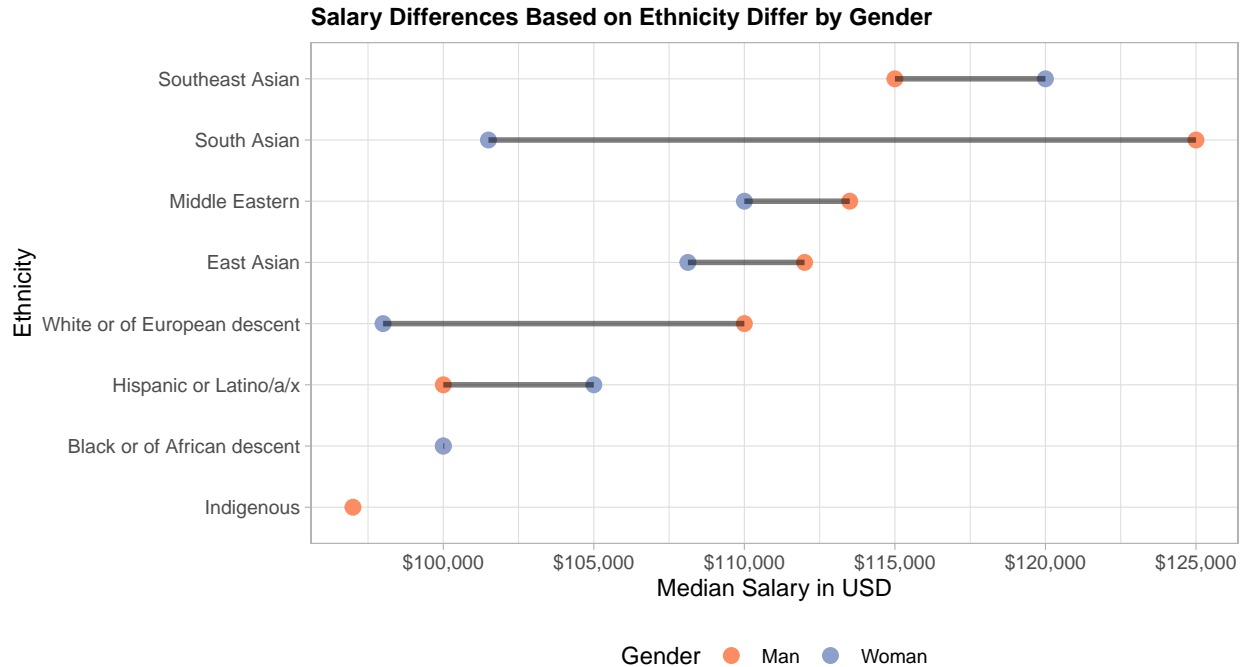**Salary Differences Based on Ethnicity Differ by Gender**

Figure 3: Differences between Median Salaries for Men and Women by Ethnicity

2.2.2 Years of Professional Coding Experience

While there were no questions on the survey which directly measured professional industry experience in years, the next closest measure assessed on the survey was a question asking how many years the respondent has been coding professionally as part of their job, not including years spent coding in school or as a hobby. As the vast majority of respondents who are employed full time reported being a developer by profession, with the rest of respondents reporting they code sometimes for their job, the number of years spent coding professionally is a good indicator of how long the respondent has been working in a tech position. Additionally, the majority of the specific job positions discussed in the next section likely require coding knowledge. However, one thing to keep in mind is that this variable does not indicate how long the respondent has been in their particular job position at a specific company, but is instead an overall estimate of how long they might have been working in the tech industry as a whole. The distribution of years of coding experience by gender is shown in Figure 4.

The majority of the respondents have been coding professionally for less than 10 years, with a large portion having less than 5 years of professional experience. This suggests many of the respondents are newer to the tech industry, and some may be holding entry-level positions. The average years of coding experience for men (mean = 9.9 years) is slightly higher than it is for women (mean = 7 years) and non-binary individuals (mean = 6.6 years)[7]. The figure indicates that larger proportions of female and non-binary respondents are newer to the industry and reported three years or less of professional coding experience compared to men. These results are similar to the Stack Overflow survey results from previous years which have shown women to be less experienced in terms of years spent coding professionally, and are more likely than men to have three years of coding experience or less (Stack Overflow 2019). As we look past the 20-25 year mark, the number of women and non-binary respondents appears to drop off, and we see men are more likely to have more years of professional experience. This is consistent with previous work on retention in the tech industry

---

[7]The mean age for men and women are 33 and 32 years respectively, and mean age for non-binary individuals is 30

which has shown women are more likely to leave tech positions than men (Scott, Klein, and Onovakpuri 2017), which would result in less women with greater years of experience.

Most Developers with Greater Years of Professional Coding Experience are Male
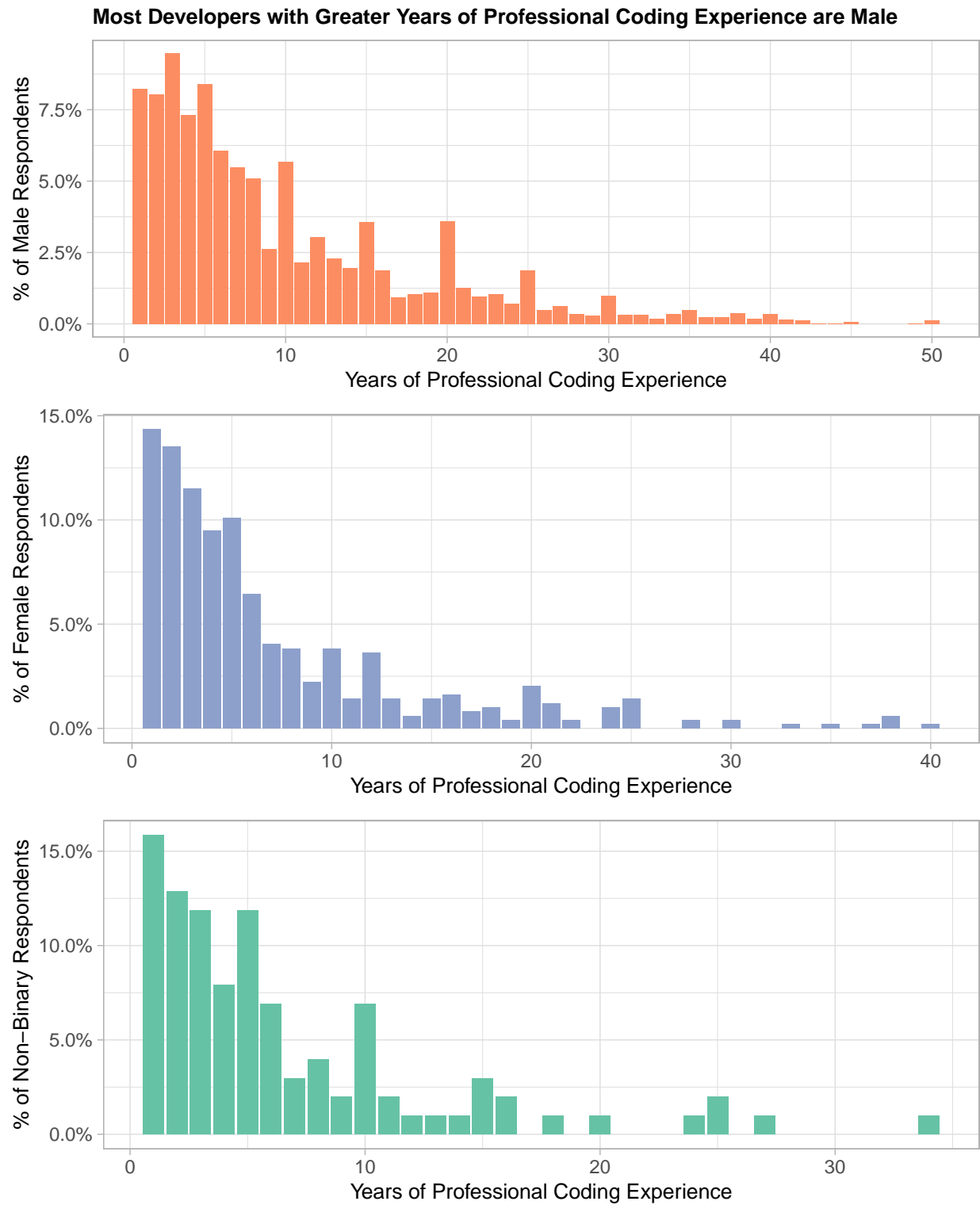

Figure 4: Proportion of Male, Female, and Non-Binary Respondents' Years of Professional Coding Experience

The relationship between years of professional coding experience and median annual salary by gender is shown in Figure 5.
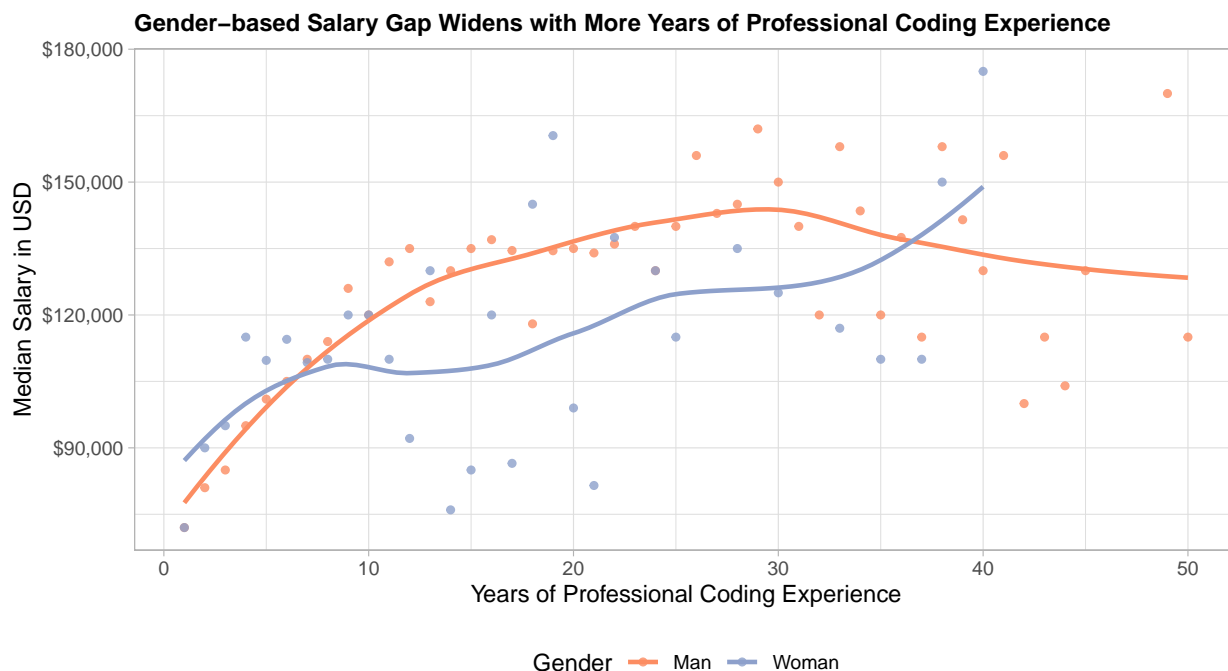

Figure 5: Relationship between Median Salary and Years of Professional Coding Experience

According to this figure, men and women with less than 5 years of experience have similar salaries, with women making slightly more than men. However, while median salaries for men continue on a relatively steady upwards trajectory as they gain more years of experience, women's salaries appear to drop, and we see significant gaps between men and women's salaries between 10 and 35 years of experience. This suggests that while men and women may start off with similar salaries, the wage gap widens over time and is more noticeable at certain levels of experience.

2.2.3 Developer Type

The Developer Type attribute in the dataset contains information about the type of job positions respondents hold. Developer type is discussed here as certain roles are compensated more than others, and gender distributions may differ depending on the role. There were a total of 23 options to select from, and respondents were asked to select all which best apply to them, resulting in a number of responses with unique combinations of these options. As a significant portion of respondents selected two or more positions, these responses were split and unnested to facilitate the analysis, however a limitation of this method is that the same respondents will now appear multiple times in the datasets which may have various implications when drawing conclusions from the analysis. To further narrow down the data and focus on individual level contributors in the tech industry, respondents holding positions such as academic researchers and scientists, as well as those in upper level management positions such as senior executives and VPs were filtered out. Figure 6 displays the salary distributions for each position type by gender and Figure 7 displays the distribution of developer positions by gender.
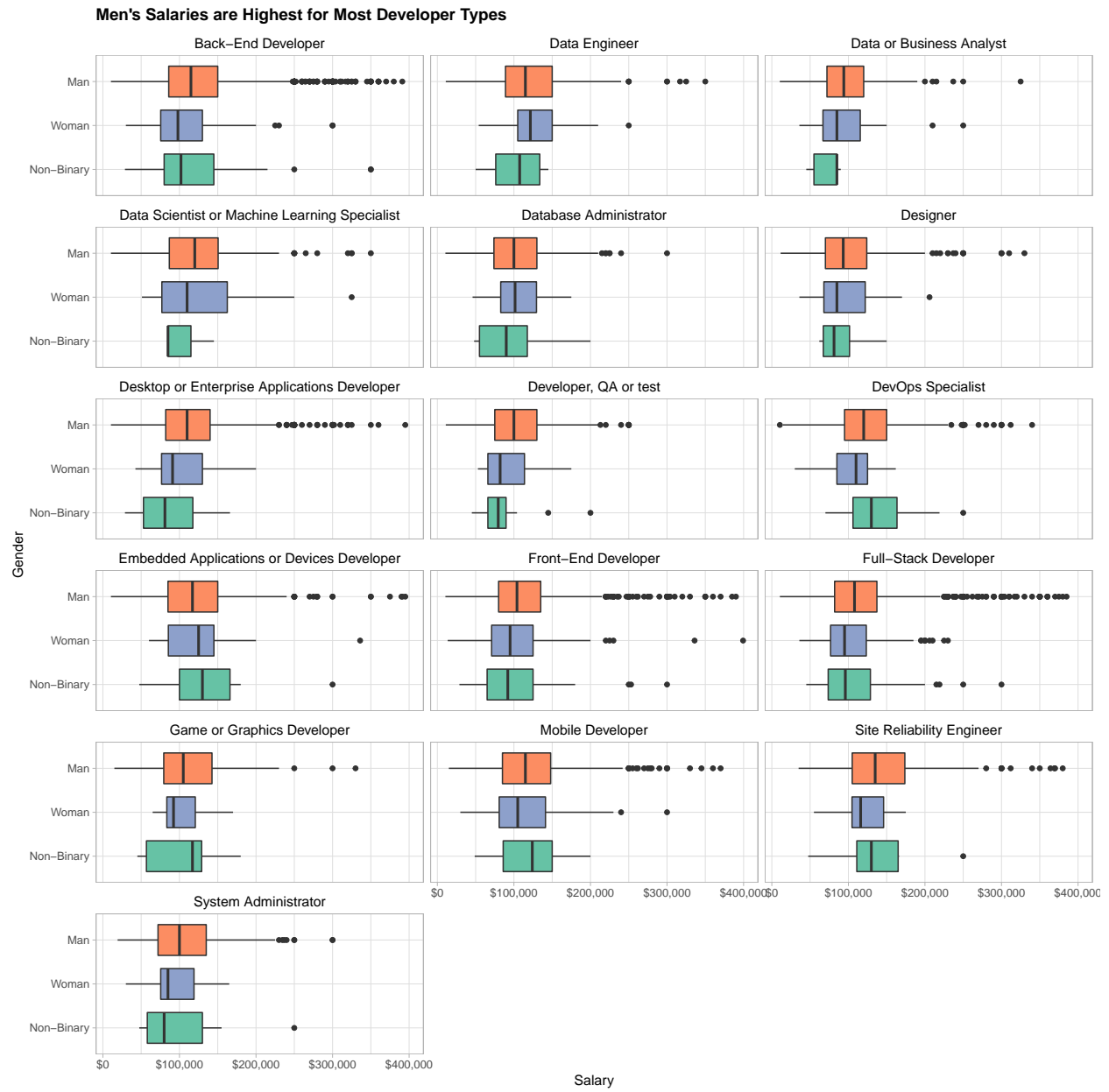
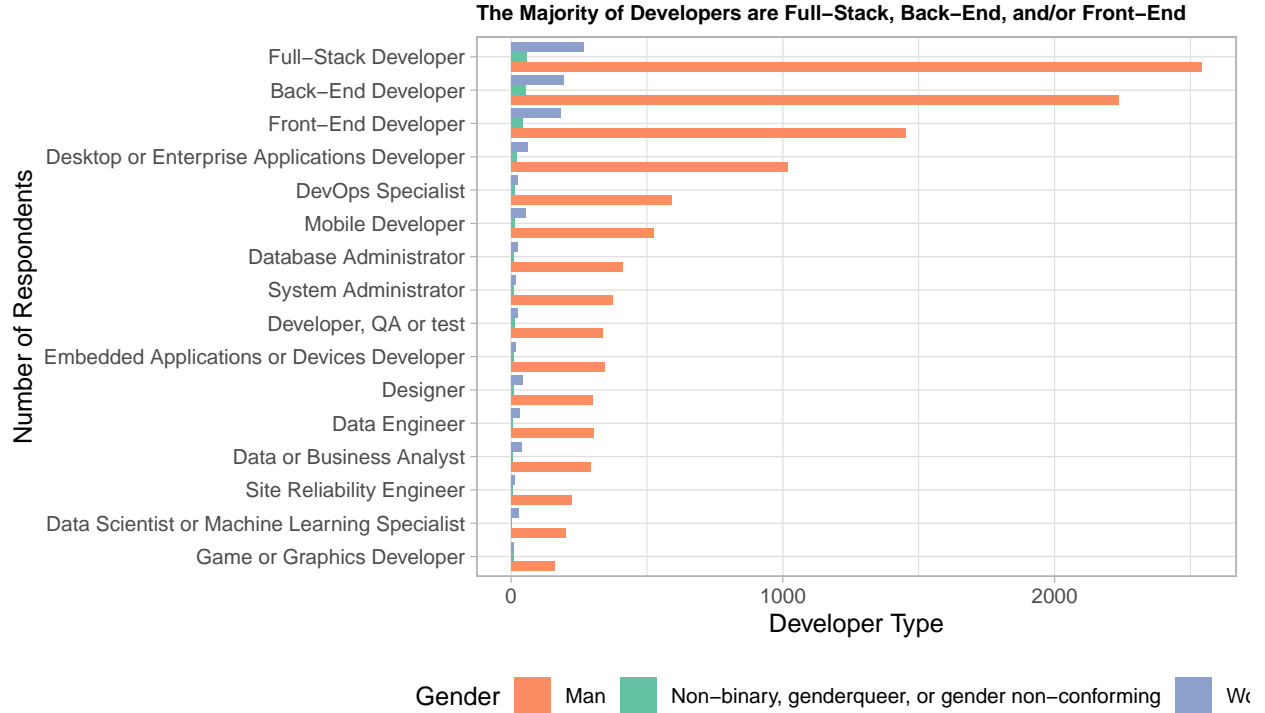Figure 6: Salary Distributions for Developer Types

Figure 7: Distribution of Developer Types

The most common positions are full-stack, back-end, and front-end developers. Before the developer type attribute was unnested, these three positions were commonly reported alongside a number of other positions. Across most job types, men's salaries are higher compared to women and non-binary individuals, however there are a few exceptions. Interestingly, non-binary individuals report higher median salaries for men for a few positions including game or graphics developers, mobile developers, and devOps specialists. However, since individuals who are non-binary take up a very small portion of these positions, it is possible that the few individuals who reported holding these positions (which may very well be the same individuals holding multiple positions) just happen to have particularly high salaries in the sample. Of the more common developer roles listed here, women appear to be underrepresented in DevOps positions, which is the developer type with the third highest average salary as reported in the official 2020 survey report (Stack Overflow 2020).

3 Model

3.1 Propensity Score Matching

Before using multiple linear regression to model the effects of gender on salary, a propensity score matching technique (PSM) was applied to the dataset to reduce the effects of confounding covariates on the main treatment variable, gender. This study uses a quasi-experimental design with observational data in which the selection of a treatment variable and subsequent allocation of respondents to treatment and control groups is a non-random process, and is instead based on subject characteristics (Austin 2011). Unlike randomized control trials where causality can be inferred after proper steps have been taken to ensure treatment and control groups share the same characteristics before being randomized to a particular intervention (Gertler et al. 2016), it is more difficult to establish causality when using observational data where baseline characteristics

may systematically differ between treatment and control groups which may confound the results (Austin 2011). PSM is a technique used to reduce the confounding effects of these baseline characteristics between groups when using observational or nonrandom data. To do this, we can compute and assign propensity scores to each respondent, which is the probability that each observation has of being in the treatment group based on the observed baseline characteristics we have (Gertler et al. 2016). Matches are then created based on these scores to ensure that aside from the treatment variable, each observation in the treated group is as similar as possible to another observation in the control group based on the selected matching characteristics.

Since I am interested in the differential effects of gender on salary, particularly if identifying as a woman has a negative effect on salary, PSM was used to match individuals who identify as men with individuals who identify as women on a number of baseline characteristics I selected from the dataset. While the experiences of non-binary, genderqueer, and gender non-conforming individuals in tech is an important area to research, from this point on my analysis of gender effects on salary will focus specifically on differences between men and women, as the prior research done to guide this study have mostly focused on representation and wage gaps for women. Since the PSM methodology used here relies on a logistic regression model to estimate propensity scores and thus requires a binary treatment variable, I made the decision to examine women as the primary treatment group compared to men, since they were the largest gender group after men and the overall sample size would be reduced based on the number of responses in the treatment group. PSM also requires a relatively large sample size, and only 2% of respondents were non-binary, which would have made it more difficult to match and fit the model and find reliable results. A logistic regression model was built to "predict" the likelihood of reporting to be a man or a woman and conduct matching based on the selected characteristics. Respondents were then matched in R (R Core Team 2020) using the `arm` package (Gelman and Su 2020).

The matching characteristics I selected were: ethnicity, education level, undergraduate major, developer type, age, age which the respondent first coded, the number of years the respondent has been coding, and the number of years the respondent has been coding professionally. These characteristics are all general demographic based attributes which were available in the dataset that I believe were relevant to match men and women on. However, one major limitation of PSM is that the matching is limited to the observed characteristics put into the model, and respondents cannot be matched on unobserved characteristics (Gertler et al. 2016). Given the complexities of gender, it is unlikely that the characteristics measured in this survey and used to perform matching sufficiently explain or cover all the underlying differences between genders. Even among the attributes measured in the survey, there may be characteristics that participants were not matched on in which gender differences may exist and therefore bias the comparisons between the two groups. Therefore, while PSM was performed to help make causal inferences about the effects of gender on salary by attempting to approximate a valid comparison group with observational data, the modeling results may still be biased as the matching of male and female observations was limited to a small selection of observable characteristics from the dataset. The final matched dataset consisted of 1804 observations, with 902 male responses and 902 female responses.

3.2 Multiple Linear Regression Model

A multiple linear regression model was employed to test the effects of gender on salary using the propensity score matched dataset. Multiple linear regression is a statistical modeling technique which aims to predict a continuous response variable based on a number of predictor variables (Olive 2017). Compared to simple linear regression which focuses on the effects of one explanatory variable on the response variable, multiple linear regression allows us to account for a number of different factors, both qualitative and quantitative, which may contribute to the response variable. This allows us to examine both the ways in which these factors individually contribute to the outcome, along with how they work together overall to explain the response variable (Marill 2004). Including multiple predictor variables in the model not only allows us to statistically control for variables whose effects may underlie the response variable, but also gives us the opportunity to examine interaction effects of variables we think may be related along with the main effects of the predictor variables. For these reasons, multiple linear regression is often used to estimate treatment effects by regressing the outcome variable on a predictor variable indicating treatment status, while accounting for

the effects of covariates in the model. If a statistically significant coefficient of the treatment variable is found in the model results, we might be able to infer the existence of a treatment effect in the data (Zanutto 2006). Of course, making such an inference also depends on steps taken in the experimental design process to ensure valid comparison groups as well as the validity of the statistical model itself.

Multiple linear regression was selected for this study in particular as I am interested in modeling the effects of various predictor variables and their interactions on salary, which is a continuous variable. While I am primarily interested in the effects of gender as a treatment variable, there are several other predictors and interactions such as ethnicity and experience that I want to explore, along with the inclusion of various demographic covariates that I would like to control for in the model. Another reason why I chose this particular method is that multiple linear regression has commonly been used to model salary in previous studies which have examined gender differences in wages and salaries (Zanutto 2006; Jena, Olenski, and Blumenthal 2016; Jagsi et al. 2013; Alkadry and Tower 2006; Rubinfeld 2000).

The following equation depicts the regression model used to predict salary.

$$log(salary_i) = \beta_0 + \beta_1 gender_i + \beta_2 ethnicity_i + \beta_3 gender_i ethnicity_i +$$
$$\beta_4 yearscodepro_i + \beta_5 gender_i yearscodepro_i + \beta_6 edlevel_i + \beta_7 devtype_i +$$
$$\beta_8 age_i + \epsilon_i$$

The output term on the left-hand of the equation represents the response variable, salary. As the salary distribution in the dataset follows a log-normal distribution, a log transformation was applied to the salary variable to normalize the data. β_0 is the model's intercept, which represents the mean salary when all predictor values are equal to 0, or categorical variables are at their reference levels. The other β terms in the equation represent coefficients the model will compute for each corresponding predictor variable (i.e. gender, ethnicity, etc.), which are the partial slopes of each predictor variable and the response variable, salary. That is, how much log(Salary) will change with each one unit increase of the predictor variable or, how much log(Salary) changes compared to the categorical variable's reference category, with all other variables held constant. Finally, ϵ_i represents the model's residual errors.

The section below discusses each predictor variable which was added into the model in more detail. The bolded variables are the main predictors I am interested in, while the rest are features that were added to the model as control variables. For categorical variables, the category with the most respondents was selected to be the reference group.

- **Gender** is the main treatment variable I am examining, and is included as a binary variable. While it was originally classified as a categorical variable with three different levels (man, woman, and non-binary), the non-binary category was removed to facilitate propensity score matching as discussed previously. Although gender is better understood as a spectrum rather than a binary construct, it was measured categorically in this survey, and will be treated in a binary manner for this analysis. The reference category is "male".
- *Ethnicity* is a categorical variable containing 7 different categories. The reference category is "White or of European descent".
- **Gender*Ethnicity** is the interaction term for the gender and ethnicity variables. This was included in the model as research on wage inequalities suggest gender and race and ethnicity often interact to affect salary (Fisher 2015).
- *YearsCodePro* is the number of years the respondent has been coding professionally. Although this was originally a continuous variable (aside from endpoints "less than 0" and "greater than 50"), I grouped the years into categories of five years each for this analysis. The relationship between years coding professionally and salary is not a straight linear one, and might be understood better in groups. For example, while earlier descriptive graphs showed a linear relationship between experience and salary for respondents with less experience, this trend diverges for respondents with greater years of experience. More importantly however, categorizing years of experience in this manner allows us to take a closer look at gender differences in salary at various levels of experience, since they follow different trends. The reference group for this category is "Less than 5 years".

- ***Gender*YearsCodePro*** Gender*YearsCodePro is the interaction term for gender and years of professional coding experience. As women often have less professional experience in the tech industry and take up a higher portion of entry level positions, there may be significant differences in the gender wage gap between men and women depending on the level of experience they have. For instance, 5 showed similar trends for women and men's salaries at less than 10 years of experience, but notable differences between genders at 10 to 30 years of experience.
- *EdLevel* is a categorical factor describing the respondents' highest level of education. There are four levels: less than Bachelor's Degree, Associate's degree, Bachelor's degree, and Graduate degree.[8]
- *DevType* was added to the model as a control variable. As discussed in the data section, this variable is a factor with 16 levels containing information about the respondents' developer role. The reference category for this variable is "Full-Stack Developer".
- *Age* was added to the model as a control variable. As I was not looking into the effects of any particular age group on salary, this variable was kept in its original form as a continuous numeric variable to simplify the model.

3.3 Assumptions and Model Checking

The multiple linear regression model was run in R (R Core Team 2020) using the `lm` function. There are several key assumptions associated with multiple linear regression analysis which need to be checked. Figure 8 is a visual check of the model's assumptions created with the `performance` package (Lüdecke et al. 2020), which contains plots checking multicollinearity, normality of residuals, homoscedasticity, and cook's distance for influential points.

The first plot visualizes multicollinearity, which exists when predictor variables are highly correlated to each other. The only variables we see high multicollinearity for are the interaction terms (i.e. ethnicity and gender*ethnicity), which makes sense since these terms include the main effects. The distribution of residuals appears relatively normal[9], although there is slight curvature on the tail ends of the Q-Q plot, the overall linear trend follows the line. Plots of the fitted values versus residuals show no discernible patterns and are equally spread around horizontal lines, suggesting assumptions of equal variance have been met. Finally, there are no notable influential observations which affect the model.

To check the model's performance as a predictor of salary, the propensity score matched dataset was split into a training and testing set using a 70:30 split ratio. After the model was trained on the testing set, it was applied to the testing portion of the data using the `predict` function. Root mean squared percentage error and median absolute percentage error were calculated by comparing the predicted values with the actual salary values of the test data, using the functions `RMSPE` and `MedianAPE` from the `MLmetrics` package (Yan 2016). Both error percentages were very small (RMSPE = 3.1% and Median APE = 1.9%), which suggests the model accurately predicted salaries within this subset of the data with minimal error. However, while these measures support the model's accuracy in this particular context, it does not necessarily mean it is generalizable to other data or effective as a "real-world" predictor of salary when taken out of the very limited constraints it was built around.

---

[8]'Less than Bachelor's Degree' was formed to include responses from individuals who reported no formal education, high-school education, some college/university education without a degree and so forth. 'Graduate degree' includes master's degrees, professional degrees, and other doctoral degrees.

[9]While the model was originally run with salaries between \$10,000 and \$400,000, after an initial performance check which showed significant skew on the left tail, the lower threshold was changed to \$30,000. The performance figure shown here and resulting model reflect this change.

Figure 8: Visual Check of Multiple Linear Regression Assumptions

4   Results

The multiple linear regression model was applied to the propensity matched dataset to predict salary. Figure 9 displays the model coefficients and their error bars indicating upper and lower estimates with a 95% confidence interval. A full numeric breakdown of the model results including p-values to indicate statistical significance is shown in Table 2.[10]

---

[10]Table created with `gtsummary` (Sjoberg et al. 2021) and `flextools` (Gohel 2021) packages.
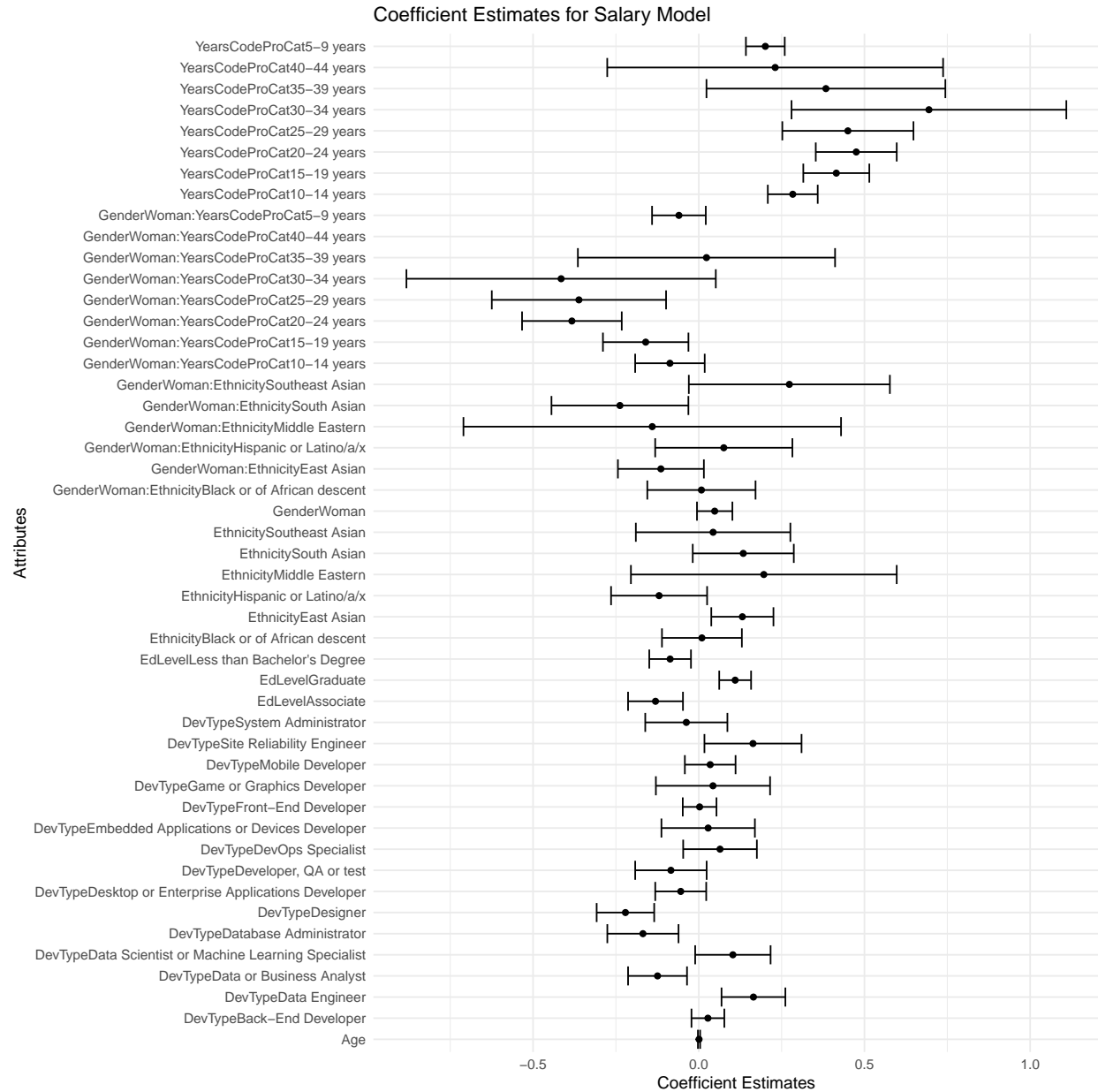
Figure 9: Coefficients for Multiple Linear Regression Model

Table 2: Multiple Linear Regression Ouput Testing the Effects of
Gender on Salary

Characteristic	Beta	95% CI[1]	p-value
(Intercept)	11.323	11.217, 11.428	**<0.001**
Gender			

[1]CI = Confidence Interval

R² = 0.193; Adjusted R-Squared = 0.172; Sigma = 0.352; Statistic = 8.96; p-value = 0.000; df = 47.0; Log-likelihood = -652; AIC = 1,402; BIC = 1,672; Deviance = 218; Residual df = 1,756; No. Obs. = 1,804

16

Characteristic	Beta	95% CI[1]	p-value
Man	—	—	
Woman	0.048	-0.005, 0.101	0.078
Ethnicity			
White or of European descent	—	—	
Black or of African descent	0.009	-0.111, 0.130	0.88
East Asian	0.132	0.038, 0.225	**0.006**
Hispanic or Latino/a/x	-0.120	-0.265, 0.025	0.11
Middle Eastern	0.196	-0.205, 0.597	0.34
South Asian	0.134	-0.018, 0.287	0.084
Southeast Asian	0.043	-0.190, 0.276	0.72
EdLevel			
Bachelor's	—	—	
Associate	-0.131	-0.213, -0.048	**0.002**
Graduate	0.110	0.062, 0.158	**<0.001**
Less than Bachelor's Degree	-0.086	-0.149, -0.024	**0.007**
Age	0.001	-0.003, 0.004	0.65
DevType			
Full-Stack Developer	—	—	
Back-End Developer	0.028	-0.022, 0.077	0.28
Data Engineer	0.165	0.069, 0.261	**<0.001**
Data or Business Analyst	-0.124	-0.213, -0.035	**0.006**
Data Scientist or Machine Learning Specialist	0.103	-0.011, 0.217	0.076
Database Administrator	-0.168	-0.276, -0.061	**0.002**
Designer	-0.221	-0.308, -0.134	**<0.001**
Desktop or Enterprise Applications Developer	-0.054	-0.131, 0.022	0.17
Developer, QA or test	-0.084	-0.192, 0.024	0.13
DevOps Specialist	0.064	-0.047, 0.175	0.26
Embedded Applications or Devices Developer	0.028	-0.113, 0.169	0.69
Front-End Developer	0.003	-0.048, 0.053	0.92
Game or Graphics Developer	0.043	-0.129, 0.215	0.63
Mobile Developer	0.035	-0.042, 0.111	0.38
Site Reliability Engineer	0.164	0.017, 0.310	0.029
System Administrator	-0.038	-0.161, 0.086	0.55

[1]CI = Confidence Interval

R^2 = 0.193; Adjusted R-Squared = 0.172; Sigma = 0.352; Statistic = 8.96; p-value = 0.000; df = 47.0; Log-likelihood = -652; AIC = 1,402; BIC = 1,672; Deviance = 218; Residual df = 1,756; No. Obs. = 1,804

Characteristic	Beta	95% CI[1]	p-value
YearsCodeProCat			
Less than 5 years	—	—	
10-14 years	0.284	0.208, 0.359	**<0.001**
15-19 years	0.415	0.315, 0.515	**<0.001**
20-24 years	0.475	0.353, 0.597	**<0.001**
25-29 years	0.450	0.252, 0.648	**<0.001**
30-34 years	0.695	0.280, 1.109	**0.001**
35-39 years	0.384	0.023, 0.744	0.037
40-44 years	0.231	-0.276, 0.737	0.37
5-9 years	0.201	0.142, 0.259	**<0.001**
Gender * Ethnicity			
Woman * Black or of African descent	0.008	-0.155, 0.171	0.92
Woman * East Asian	-0.114	-0.244, 0.015	0.084
Woman * Hispanic or Latino/a/x	0.076	-0.131, 0.283	0.47
Woman * Middle Eastern	-0.140	-0.710, 0.429	0.63
Woman * South Asian	-0.238	-0.444, -0.031	0.024
Woman * Southeast Asian	0.273	-0.030, 0.577	0.077
Gender * YearsCodeProCat			
Woman * 10-14 years	-0.087	-0.192, 0.018	0.10
Woman * 15-19 years	-0.160	-0.289, -0.031	0.015
Woman * 20-24 years	-0.383	-0.533, -0.232	**<0.001**
Woman * 25-29 years	-0.362	-0.625, -0.099	**0.007**
Woman * 30-34 years	-0.415	-0.882, 0.051	0.081
Woman * 35-39 years	0.023	-0.365, 0.411	0.91
Woman * 40-44 years			
Woman * 5-9 years	-0.060	-0.141, 0.021	0.15

[1]CI = Confidence Interval

R^2 = 0.193; Adjusted R-Squared = 0.172; Sigma = 0.352; Statistic = 8.96; p-value = 0.000; df = 47.0; Log-likelihood = -652; AIC = 1,402; BIC = 1,672; Deviance = 218; Residual df = 1,756; No. Obs. = 1,804

Figure 9 shows noticeably large error bars for the Middle Eastern ethnicity category and its interactions, as well as the years of experience categories for 30-34 years, 35-39 years, and 40-44 years. It should be noted that these groups also contain the smallest sample sizes in their respective categories and in this subsection of the dataset. In Table 2, bolded figures represent attributes in the model which were statistically significant at p < 0.01. One thing to remember when reading the model results is that a log-normal transformation was applied to the dependent variable, salary, which means the model coefficients in the output do not directly reflect dollar changes in salary associated with each variable. To interpret the coefficients in terms of percentage changes in salary, we can input the coefficients into the following equation.

$$100 * (e^{\beta_j} - 1)$$

Here, e is the exponential function and β is the estimated coefficient. The resulting number represents the percentage increase in salary associated with a 1 unit increase in the predictor variable with all other variables held constant or at reference level, or the percentage increase associated with a certain level in a categorical variable compared to the reference group with all other variables held constant (Cornell Statistical Consulting Unit 2020). The reference groups for this model were: male, White or of European descent, Bachelor's education, full-stack developer position, and less than 5 years of professional coding experience, meaning coefficients will be interpreted with these levels held constant.

4.1 Main Effects

First, the model showed no significant main effect for gender. This indicates that while holding all other predictors constant and at their reference categories, women do not earn significantly higher or lower salaries than men. Regarding ethnicity, there was a significant main effect for individuals reporting to be East Asian, who on average, earned around 14% more salary than White respondents with all other categories held at reference level. No other groups in the ethnicity category had significantly different salaries compared to White respondents. When we examine the main effect of years of professional coding experience, we see that compared to respondents at reference level with less than 5 years of experience, each category with an additional 5 years of experience is associated with significant subsequent increases in salary up (i.e. compared to those with less than 5 years of experience, respondents with 5-9 years of experience is associated with a 20% increase with salary, 20-24 years is associated with a 32% increase and so forth). This salary growth continues to increase until its peak at 30-34 years of experience, which shows a 100% salary increase compared to less than 5 years of experience. At 35-39 years of experience, we see less of an increase in salary, with the percentage growth compared to those with less than 5 years of experience being similar to that of respondents with 15-19 years of experience.

In terms of control variables, age did not appear to be a significant predictor of salary, however the respondent's highest level of completed education was significantly associated with salary. Compared to a bachelor's degree, completion of a graduate degree was associated with a 11.6% increase in salary, while completion of an associate's degree or less than a bachelor's degree were associated with significant decreases in salary. Finally, there was some variation in the impact of developer type on salary, data engineers being associated with significantly higher salaries than full-stack developers, and positions such as designers and database administrators being associated with lower salaries. Most developer types however, were not significantly associated with changes in salary compared to full-stack developers with other variables held at their reference categories.

4.2 Interaction Effects

Aside from the main effect of gender, interpretations of the coefficients discussed for the other attributes in the model have been in reference to male respondents, as they were the reference category for gender. The model also included interaction terms to examine how gender may interact with ethnicity and experience to affect salary. None of the interactions between ethnicity and gender were statistically significant at p < 0.01. Interestingly, while there was no main effect of gender on salary, the interaction between gender and years of professional coding experience was significant for women with 20-24 years of experience and women with 25-29 years of experience, in which being a woman was associated with less salary compared to men with similar years of experience at these levels. This is best illustrated in Figure 10, which plots the fitted model's interaction effects for gender and years of professional coding experience using the `interactions` package (Long 2019).
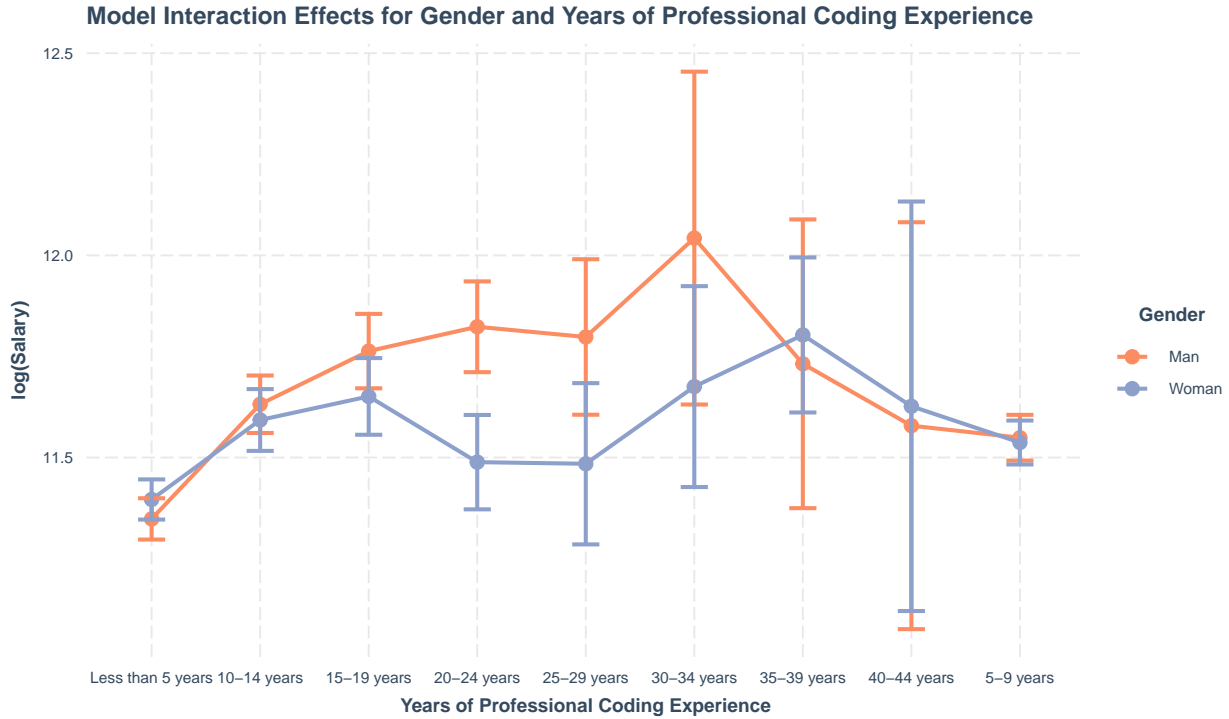
Figure 10: Intearction Plot for Gender and Experience

As discussed previously, there appears to be a main effect of experience for men where each group with an additional 5 years of experience is associated with relatively steady increases in salary until peaking at around 30-34 years of experience. Women with under 15 years of professional coding experience appear to follow this trend, as their salaries are approximately equal to men's salaries in the first three groups (less than 5 years of experience, 5-9 years of experience, and 10-14 years of experience). However, we start to see some departures from this pattern at 15-19 years of experience where women start making slightly less than men, before seeing significant differences at the 20-24 year and 25-29 year mark. While men continue to experience increases in salary, we observe noticeable drops for women who now make significantly less than men at these points, which is consistent with the significance indicators for these levels of the interaction on the model's output. Salaries appear to increase again for women with 30-34 years of experience and then converge with men's salaries onward, however these levels were not statistically significant and show very large error bars. This is likely due to the fact that the sample sizes for respondents in groups representing higher levels of professional coding experience were extremely small compared to those with less experience. While respondents with 45-49 years and 50+ years of experience existed in the original dataset, there were no women in these groups, meaning the interaction between gender and years of professional coding experience could not be modeled at these levels.

5 Discussion

5.1 Overview of Findings

In light of growing calls for diversity in the tech industry and concerns about wage inequalities based on gender and race, this paper used data from the Stack Overflow Developer Survey to examine the relationship between gender and salary for tech workers in the United States. In addition to testing for main causal effects

of gender on salary, the effects of interactions between gender and ethnicity, and gender and experience were also explored. A descriptive overview of the survey results found the vast majority of respondents were men, and most respondents were of White or European descent. After removing extremely high and low salaries, on average, men reported higher salaries than female and non-binary respondents. To test if these gender differences had a meaningful causal impact on salary, propensity score matching was used to form matching sets of male and female respondents based on several other characteristics from the dataset. A multiple linear regression model was then applied to the matched data to model salary and estimate the effects of gender, gender and ethnicity, and gender and experience. No statistically significant effects were found gender as a main treatment variable, which indicates that in general, being a woman was not associated with having a lower salary compared to men. Additionally, while ethnicity and gender did not interact to affect salary, there was a significant interaction between gender and years of professional coding experience in which women with 20-24 years of professional coding experience and women with 25-29 years reported lower salaries than men.

Although this study found no evidence for a main effect of gender on salary, this does not necessarily mean wage inequalities between men and women don't exist in the United States tech industry. However, the relationship between gender and salary is complex, and it is unlikely that the wage differences we often observe between men and women can simply be explained by the idea that women earn less solely because of their gender. Numerous other factors, including those measured in the survey and included in the model, as well as other unobserved characteristics, interact with gender in complex ways to affect salary. One of these factors is experience, as we see a large number of women with less than 5 years of experience, while individuals with more years of professional coding experience (around 20+ years) are usually men. A possible explanation for the overall wage differences between men and women can be attributed to the fact that more women in the industry are less experienced and therefore have lower salaries on average, however this places more importance on the role of experience as a predictor of salary rather than gender differences. A closer look at how gender and experience interact in both the data and modeling results revealed that men and women with equal years of professional coding experience reported similar salaries when they had under 20 years of experience, however for respondents with 20-30 years of experience, men made significantly more than women. So while men and women may start off in the industry with similar salaries and experience similar growth trajectories as they gain more experience, at 20-30 years women suddenly start reporting lower salaries while men's salaries stay roughly the same.

Given these results, what might explain the salary differences between men and women who have more years of professional coding experience? Research exploring retention problems for women in the tech industry has shown women are passed over for promotions more often than men, with more experienced women reporting they often feel "stuck" in their position and have difficulties advancing past certain stages in their careers compared to their male peers with similar levels of experience (Scott, Klein, and Onovakpuri 2017; Hewlett et al. 2008). This is commonly referred to as the glass ceiling effect, which represents the societal barriers women face in the workplace in which they often reach a point in the corporate hierarchy they are unable to advance past (Cotter et al. 2001). Although this study focused on salary differences for individual-level contributors and upper management and executive positions were removed prior to the analysis, the vast majority of individuals holding these upper level positions were still men. Therefore it is unlikely that these results can solely be explained by the fact that high ranking women with management positions and higher salaries were excluded from this analysis.

Gender-based discrimination involving key promotions or important salary negotiations may be driving the inequalities observed for respondents with more years of experience, however this data is missing a potentially critical component of the gender pay gap which has been well established in the literature, which is the effects of having children. In past years, the Stack Overflow Developer Survey asked respondents to report if they have children or other dependents that they care for, however this question was not asked in the 2020 version of the survey. According to results from the 2019 Developer Survey, around 40% of total respondents, and 60% of respondents with 10 or more years of professional coding experience reported having dependents. (Stack Overflow 2019). A previous analysis of gender and salary using data from the 2019 Developer Survey found similar interactions between gender and years of experience as this study, as well as significant interactions between gender and dependents (Silge 2019). Notably, while having dependents did not affect salary for men, it was associated with reduced salaries for women. This is consistent with previous research which

suggests motherhood is a significant contributor the gender wage gap, as the difference between men and women's earnings begins to widen after the arrival of children, with college-educated women in high-earning occupations in particular being the most affected by this (Miller 2018; Clawson and Gerstel 2014). Unequal household division of labour might result in women spending more time on family obligations and less on long working hours which may be necessary to receive a promotion, or less time invested in networking and facilitating job search activities which could lead to financially advantageous job changes (Barth, Kerr, and Olivetti 2021). At later stages in their careers, women might also be less likely to receive promotions or even seek out promotions both within and across companies due to increased need for flexibility and time for family obligations. This might explain some of the gender differences in salary observed for respondents with more years of experience. However, in the analysis of the 2019 survey data which included information about dependents, the interaction between gender and experience was still significant even after accounting for wage inequalities based on dependents, suggesting disparities still exist even for women without children.

This report also took an intersectional approach to examining gender-based wage inequalities by examining how women in underrepresented racial or ethnic groups may be disproportionately affected by the wage gap. The model results showed no significant interaction between gender and ethnicity, and most categories for ethnicity as a main effect were also insignificant compared to the salaries of White respondents. As mentioned previously, around 80% of the respondents were White, therefore the sample sizes for other non-White groups in this dataset were extremely small, and were reduced even further when the model was run after being propensity score matched. The small sample sizes for these groups may be part of the reason no significant effects or interaction effects were found in regards to ethnicity despite past literature indicating the existence of a race-gender wage gap (Hired 2019; Fisher 2015; Patten 2016).

Despite the fact that main causal effects were not found for gender or the interaction between gender and ethnicity, underrepresentation and the lived experiences of gender and race based discrimination in the United States tech industry are still serious problems. The 2020 Developer survey results indicate that women, non-binary, genderqueer, or gender non-conforming respondents, as well as non-white respondents are still vastly underrepresented in tech communities. Although it is encouraging to see a great number of women entering the industry and receiving equal compensation as men in earlier years, the fact that that there are so few women with greater years of professional coding experience, and that the women in the industry who have more experience are paid significantly less than their male peers highlights the retention problem in the industry. To ensure the women who responded to this survey with less years of professional coding experience continue to grow and advance successfully in the industry over the next 10 years, companies need to directly address the reasons which are causing women to leave the industry, which includes giving women the same opportunities to advance as men and providing fair compensation. Unfortunately, many issues underlying unfair workplace practices can be linked to structural sexism and racism, and long held ideas that tech is a "boys club" (Acker 2006; Hewlett et al. 2008). Companies need to implement more genuine and comprehensive strategies to promote inclusivity, such as auditing compensation and performance management practices at all levels of the employment cycle to mitigate bias. This also includes addressing toxic and discriminatory workplace cultures and environments which often cause women to leave their positions. Finally, regardless of this study's findings that ethnicity and gender did not interact to affect salary, it is still important for companies to take an intersectional approach when developing and implementing these strategies to provide better support and understanding of the experiences of racialized women and non-binary tech workers.

5.2   Limitations and Future Directions

There are several weaknesses and limitations with this study to address. First, as discussed in the data section, this analysis relies on self-report survey data, therefore common concerns associated with using survey data such as validity issues and response bias must be considered when interpreting the results. Non-responses for the attributes which were examined in this report were primarily removed, however there may be differences in various characteristics between those who reported certain questions and those who did not. There is also no guarantee that all responses were truthful or that all questions were interpreted in the same way. For example, an attribute such as salary could have easily been over or under reported in many

cases, which could have impacted the model results. In the survey, respondents were able to select multiple responses for questions regarding ethnicity and developer type, each which were dealt with differently in the analysis. For ethnicity, respondents who selected more than one response were filtered out to simplify the analysis. Although this was not a significant amount of respondents, it still resulted in potentially important data being lost from the analysis, especially since sample sizes for groups who were non-White were already small to begin with. Unlike ethnicity, a significant portion of respondents selected multiple responses to describe the type of developer position they have in the workplace, which makes sense given the overlap for many of these positions. To facilitate the analysis of this variable in the model, the developer types were split and unnested, resulting in a longer dataset where the same respondent appears multiple times. Of course, this has implications for both the propensity score matching and subsequent model if the same respondent (with the same salary and demographic characteristics) is being counted multiple times in the dataset as separate responses, especially for respondents with salaries which might be particularly high or low, or for respondents belong to racial or ethnic groups which were the most underrepresented in the data.

Propensity score matching was used to reduce the effects of confounding covariates which are associated with gender, and match male and female respondents based on similar characteristics. However, the PSM technique has several weaknesses which should be noted. As previously discussed, the data were only matched based on a limited number of observed characteristics, however there are a number of other confounding factors related to gender which either weren't assessed in the dataset, or were not included in the matching formula. These "hidden" differences which were unaccounted for in the matching process may have affected the modeling results in unknown ways. Another weakness of propensity score matching is that a considerable number of responses are lost when matching to a treatment group with a small sample size (Streiner and Norman 2012). While the initial dataset was quite large, since women only made up around 10% of respondents, the final matched dataset was significantly reduced to match the male responses to the female ones, meaning a lot of data was lost for the model. This was particularly noticeable when examining the effects of ethnicity, where some categories had extremely small sample sizes after the matching process. Finally, along with concerns about the process of matching in general, there have been criticisms of the PSM technique in particular which have argued that PSM often increases imbalance for covariates rather than reducing it, and increases model dependence which can generate bias (King and Nielsen 2019).

Limitations to the model itself largely involve the number of features which were not included in the model. Besides gender, ethnicity, and years of professional coding experience, the effects of several other variables such as age, education, and developer type were controlled for in the model. Given something as complex as salary, it is very unlikely that this model contains all the features which contribute to salary. Other attributes which were measured in the survey such as the size of the company the respondent works for, working hours, or the number of programming platforms and languages the respondent has worked with may also contribute to salary. There may also have been other attributes which interact with gender which were not included in the model, and may have provided more information on the overall relationship between gender and salary. For example, developer type was controlled for in this model, but there could also be a relationship between gender and developer type which could be differentially affecting salary for men and women.

Another main limitation of the model is our ability to generalize its results to the broader population, in this case the entire United States tech industry. First, the model findings are somewhat limited to the data used to build the model. The Stack Overflow Developer Survey, which is a non-random sample representing a subset of the tech industry at one point in time, is not necessarily representative of the tech industry as a whole. Second, the generalizability of the results is constrained to the attributes which were added into the model, how they were phrased and measured in the survey, and how and when respondents chose to respond to these questions. Applying this model and its results to reality is far more complex, as there are events and complex interactions in the "large world" which we could not anticipate or did not account for in the modeling process, which hinders our ability to draw meaningful conclusions about the real world based on the results found in this study (McElreath 2020). That is not to say there is no truth to these findings, however the results should be understood in the context of the data and modeling decisions made in this study, and we should be mindful of the extent to which they fit into the real world and our existing knowledge and research on this topic.

Overall, since the questions asked in the Stack Overflow Developer Survey often differ on a year to year basis,

comparisons and subsequent analyses of future and past iterations of the Developer Survey can provide new insights into how gender may interact with different variables to affect salary. Since this analysis took a very simplified approach to modeling salary and only used a limited number of hand-picked variables in the model, future work would benefit from taking a more comprehensive and systematic approach to feature selection starting from all available variables in the dataset, while also considering new interactions between gender and other variables. This paper focused on the effects of gender and salary for men and women only. Unfortunately, due to the small sample size for this group, as well as binary requirements for the PSM technique that was used, respondents who identified as non-binary, genderqueer, or gender non-conforming were not included in this analysis. Since the majority of the research on gender inequalities in the tech industry has primarily focused on women, an important area of future research would be to examine the experiences and salary implications of individuals with other gender identities who are both underrepresented and understudied in the tech industry.

Due to the weaknesses associated with propensity score matching, future work examining causal effects of gender can utilize other matching methods which may be more robust and effective at reducing imbalance and model dependence. This study also explored salary through a single model, multiple linear regression, which was chosen for its simplicity and past usage in salary modeling, however a comparison of other types of models which may also be suited to this data would also be valuable. For instance, bayesian modeling or mixed-effects models could also be applied to this data, and the results could be compared across models to gain a better understanding of how gender interacts with other variables to affect salary.

Appendix

Packages

This subsection lists the R packages used in the making of this report.

General data cleaning and analysis

- `tidyverse` (Wickham et al. 2019)
- `dplyr` (Wickham et al. 2021)
- `janitor` (Firke 2021)
- `purr` (Henry and Wickham 2020)
- `broom` (Robinson, Hayes, and Couch 2020)
- `scales` (Wickham and Seidel 2020)
- `plyr` (Wickham 2011)
- `forcats` (Wickham 2020)
- `broom` (Robinson, Hayes, and Couch 2020)
- `visdat` (Tierney 2017)

Figures

- `ggplot2` (Wickham 2016)
- `RColorBrewer` (Neuwirth 2014)
- `gridExtra` (Auguie 2017)
- `qqplotr` (Almeida, Loy, and Hofmann 2017)
- `performance` (Lüdecke et al. 2020)
- `interactions` (Long 2019)

Tables

- `kableExtra` (Zhu 2020)
- `gtsummary` (Sjoberg et al. 2021)
- `flextable` (Gohel 2021)

RMarkdown Report

- `knitr` (Xie 2021a)
- `bookdown` (Xie 2020)
- `tinytex` (Xie 2021b)

Shiny App

- `shiny` (Chang et al. 2020)
- `shinythemes` (Chang 2021)
- `plotly` (Sievert 2020)

Propensity Score Matching and Modeling

- `arm` (Gelman and Su 2020)
- `MLmetrics` (Yan 2016)

References

Acker, Joan. 2006. "Inequality Regimes: Gender, Class, and Race in Organizations." *Gender & Society* 20 (4): 441–64.

Alfrey, Lauren, and France Winddance Twine. 2017. "Gender-Fluid Geek Girls: Negotiating Inequality Regimes in the Tech Industry." *Gender & Society* 31 (1): 28–50.

Alkadry, Mohamad G, and Leslie E Tower. 2006. "Unequal Pay: The Role of Gender." *Public Administration Review* 66 (6): 888–98.

Allaire, JJ, Yihui Xie, Jonathan McPherson, Javier Luraschi, Kevin Ushey, Aron Atkins, Hadley Wickham, Joe Cheng, Winston Chang, and Richard Iannone. 2021. *Rmarkdown: Dynamic Documents for R.* https://github.com/rstudio/rmarkdown.

Almeida, Alexandre, Adam Loy, and Heike Hofmann. 2017. *Qqplotr: Quantile-Quantile Plot Extensions for 'Ggplot2'.* https://github.com/aloy/qqplotr.

Asai, David J. 2020. "Race Matters." *Cell* 181 (4): 754–57.

Auguie, Baptiste. 2017. *GridExtra: Miscellaneous Functions for "Grid" Graphics.* https://CRAN.R-project.org/package=gridExtra.

Austin, Peter C. 2011. "An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies." *Multivariate Behavioral Research* 46 (3): 399–424.

Banjo, Shelly, and Ian King. 2020. *From Apple to Facebook, Tech's New Diversity Pledges Follow Years of Failure.* Bloomberg. https://www.bloomberg.com/news/articles/2020-06-23/apple-amazon-facebook-google-microsoft-data-on-black-hiring.

Barth, Erling, Sari Pekkala Kerr, and Claudia Olivetti. 2021. "The Dynamics of Gender Earnings Differentials: Evidence from Establishment Data." *European Economic Review*, 103713.

Chang, Winston. 2021. *Shinythemes: Themes for Shiny.* https://CRAN.R-project.org/package=shinythemes.

Chang, Winston, Joe Cheng, JJ Allaire, Yihui Xie, and Jonathan McPherson. 2020. *Shiny: Web Application Framework for R.* https://CRAN.R-project.org/package=shiny.

Clawson, Dan, and Naomi Gerstel. 2014. *Unequal Time: Gender, Class, and Family in Employment Schedules.* Russell Sage Foundation.

CompTIA. 2020. *Cyberstates 2020.* https://www.cyberstates.org/pdf/CompTIA_Cyberstates_2020.pdf.

Cornell Statistical Consulting Unit. 2020. *Interpreting Regression Coefficients for Log-Transformed Variables.* https://www.cscu.cornell.edu/news/statnews/83_logv.pdf.

Cotter, David A, Joan M Hermsen, Seth Ovadia, and Reeve Vanneman. 2001. "The Glass Ceiling Effect." *Social Forces* 80 (2): 655–81.

Dickey, Megan Rose. 2015. *The Future of Trans*H4CK.* Tech Crunch. https://techcrunch.com/2015/11/29/the-future-of-transh4ck/.

Firke, Sam. 2021. *Janitor: Simple Tools for Examining and Cleaning Dirty Data.* https://CRAN.R-project.org/package=janitor.

Fisher, Milia. 2015. *Women of Color and the Gender Wage Gap.* https://www.americanprogress.org/issues/women/reports/2015/04/14/110962/women-of-color-and-the-gender-wage-gap/.

Gelman, Andrew, and Yu-Sung Su. 2020. *Arm: Data Analysis Using Regression and Multilevel/Hierarchical Models.* https://CRAN.R-project.org/package=arm.

Gertler, Paul J, Sebastian Martinez, Patrick Premand, Laura B Rawlings, and Christel MJ Vermeersch. 2016. *Impact Evaluation in Practice.* The World Bank.

Girls Who Code. 2019. *Applying for Internships as a Woman in Tech: Findings from a Survey of Gwc-Affiliated Women.* https://girlswhocode.com/wp-content/uploads/2019/08/GWC_Advocacy_InternshipApplicationExperiences_PDF_z6.pdf.

Gohel, David. 2021. *Flextable: Functions for Tabular Reporting.* https://CRAN.R-project.org/package=flextable.

Harrison, Sara. 2019. *Five Years of Tech Diversity Reports—and Little Progress.* Wired. https://www.wired.com/story/five-years-tech-diversity-reports-little-progress/.

Henry, Lionel, and Hadley Wickham. 2020. *Purrr: Functional Programming Tools.* https://CRAN.R-project.org/package=purrr.

Hewlett, Sylvia Ann, Carolyn Buck Luce, Lisa J Servon, Laura Sherbin, Peggy Shiller, Eytan Sosnovich, and Karen Sumberg. 2008. "The Athena Factor: Reversing the Brain Drain in Science, Engineering, and Technology." *Harvard Business Review Research Report* 10094: 1–100.

Hired. 2019. *The State of Wage Inequality in the Workplace.* https://hired.com/page/wage-inequality-report.

Jackson, Dimitra Lynette, Soko S Starobin, and Frankie Santos Laanan. 2013. "The Shared Experiences: Facilitating Successful Transfer of Women and Underrepresented Minorities in Stem Fields." *New Directions for Higher Education* 2013 (162): 69–76.

Jagsi, Reshma, Mr Kent A Griffith, Abigail Stewart, Ms Dana Sambuco, Ms Rochelle DeCastro, and Peter A Ubel. 2013. "Gender Differences in Salary in a Recent Cohort of Early-Career Physician-Researchers." *Academic Medicine: Journal of the Association of American Medical Colleges* 88 (11).

Jena, Anupam B, Andrew R Olenski, and Daniel M Blumenthal. 2016. "Sex Differences in Physician Salary in Us Public Medical Schools." *JAMA Internal Medicine* 176 (9): 1294–1304.

King, Gary, and Richard Alexander Nielsen. 2019. "Why Propensity Scores Should Not Be Used for Matching."

Long, Jacob A. 2019. *Interactions: Comprehensive, User-Friendly Toolkit for Probing Interactions.* https://cran.r-project.org/package=interactions.

Lüdecke, Daniel, Dominique Makowski, Philip Waggoner, and Indrajeet Patil. 2020. "Performance: Assessment of Regression Models Performance." *CRAN.* https://doi.org/10.5281/zenodo.3952174.

Malcom, Shirley Mahaley, and others. 1976. "The Double Bind: The Price of Being a Minority Woman in Science. Report of a Conference of Minority Women Scientists, Arlie House, Warrenton, Virginia."

Marill, Keith A. 2004. "Advanced Statistics: Linear Regression, Part Ii: Multiple Linear Regression." *Academic Emergency Medicine* 11 (1): 94–102.

McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan.* CRC press.

Miller, Claire Cain. 2018. "Children Hurt Women's Earnings, but Not Men's (Even in Scandinavia)." *The New York Times.*

National Center for Women & Information Technology. 2016. *Women in Tech: The Facts (2016 Update).* https://www.ncwit.org/resources/women-tech-facts-2016-update.

Neuwirth, Erich. 2014. *RColorBrewer: ColorBrewer Palettes.* https://CRAN.R-project.org/package=RColorBrewer.

Olive, David J. 2017. "Multiple Linear Regression." In *Linear Regression,* 17–83. Springer.

Patten, Eileen. 2016. *Racial, Gender Wage Gaps Persist in U.s. Despite Some Progress.* https://www.pewresearch.org/fact-tank/2021/04/14/6-facts-about-americas-stem-workforce-and-those-training-for-it/.

Pew Research Center. 2021. *Women in Tech: The Facts (2016 Update).* https://www.pewresearch.org/fact-tank/2021/04/14/6-facts-about-americas-stem-workforce-and-those-training-for-it/.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Robinson, David, Alex Hayes, and Simon Couch. 2020. *Broom: Convert Statistical Objects into Tidy Tibbles.* https://CRAN.R-project.org/package=broom.

Rooney, Kate, and Yasmin Khorram. 2020. *Tech Companies Say They Value Diversity, but Reports Show Little Change in Last Six Years.* CNBC. https://www.cnbc.com/2020/06/12/six-years-into-diversity-reports-big-tech-has-made-little-progress.html.

Rubinfeld, Daniel L. 2000. "Reference Guide on Multiple Regression." *Reference Manual on Scientific Evidence* 179: 425–69.

Scott, Allison, Freada Kapor Klein, and Uriridiakoghene Onovakpuri. 2017. *Teach Leavers Study.* Kapor Center for Social Impact. https://www.kaporcenter.org/wp-content/uploads/2017/08/TechLeavers2017.pdf.

Sewak, M, Divye Raj Khilnani, Andrew Tronson, Kari Okamoto, Shantanu Vikas Kurhekar, Lijing Zhang, Daniel Sanchez, et al. 2010. "Finding a Growth Business Model at Stack Overflow." *Inc. Stanford Case Publisher*, 1–35.

Sievert, Carson. 2020. *Interactive Web-Based Data Visualization with R, Plotly, and Shiny.* Chapman; Hall/CRC. https://plotly-r.com.

Silge, Julia. 2019. *Modeling Salary and Gender in the Tech Industry.* https://juliasilge.com/blog/salary-gender/.

Singh, Kusum, Katherine R Allen, Rebecca Scheckler, and Lisa Darlington. 2007. "Women in Computer-Related Majors: A Critical Synthesis of Research and Theory from 1994 to 2005." *Review of Educational Research* 77 (4): 500–533.

Sjoberg, Daniel D., Michael Curry, Margie Hannum, Joseph Larmarange, Karissa Whiting, and Emily C. Zabor. 2021. *Gtsummary: Presentation-Ready Data Summary and Analytic Result Tables.* https://CRAN.R-project.org/package=gtsummary.

Stack Overflow. 2019. *Developer Survey Results 2019.* https://insights.stackoverflow.com/survey/2019.

———. 2020. *Stack Overflow Annual Developer Survey 2020.* https://insights.stackoverflow.com/survey.

———. 2021. *Stack Overflow Company Page.* https://stackoverflow.com/company.

Streiner, David L, and Geoffrey R Norman. 2012. "The Pros and Cons of Propensity Scores." *Chest* 142 (6): 1380–2.

Tapia, Andrea H, and Lynette Kvasny. 2004. "Recruitment Is Never Enough: Retention of Women and Minorities in the It Workplace." In *Proceedings of the 2004 Sigmis Conference on Computer Personnel Research: Careers, Culture, and Ethics in a Networked Environment*, 84–91.

Tierney, Nicholas. 2017. "Visdat: Visualising Whole Data Frames." *JOSS* 2 (16): 355. https://doi.org/10.21105/joss.00355.

Twine, France Winddance. 2018. "Technology's Invisible Women: Black Geek Girls in Silicon Valley and the Failure of Diversity Initiatives." *International Journal of Critical Diversity Studies* 1 (1): 58–79.

Wickham, Hadley. 2011. "The Split-Apply-Combine Strategy for Data Analysis." *Journal of Statistical Software* 40 (1): 1–29. http://www.jstatsoft.org/v40/i01/.

———. 2016. *Ggplot2: Elegant Graphics for Data Analysis.* Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2020. *Forcats: Tools for Working with Categorical Variables (Factors).* https://CRAN.R-project.org/package=forcats.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation.* https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, and Dana Seidel. 2020. *Scales: Scale Functions for Visualization.* https://CRAN.R-project.org/package=scales.

Xie, Yihui. 2020. *Bookdown: Authoring Books and Technical Documents with R Markdown.* https://github.com/rstudio/bookdown.

———. 2021a. *Knitr: A General-Purpose Package for Dynamic Report Generation in R.* https://yihui.org/knitr/.

———. 2021b. *Tinytex: Helper Functions to Install and Maintain Tex Live, and Compile Latex Documents.* https://github.com/yihui/tinytex.

Yan, Yachen. 2016. *MLmetrics: Machine Learning Evaluation Metrics.* https://CRAN.R-project.org/package=MLmetrics.

Zanutto, Elaine L. 2006. "A Comparison of Propensity Score and Linear Regression Analysis of Complex Survey Data." *Journal of Data Science* 4 (1): 67–91.

Zhu, Hao. 2020. *KableExtra: Construct Complex Table with 'Kable' and Pipe Syntax.* https://CRAN.R-project.org/package=kableExtra.