

title\*

Sub

Morgaine Westin

29 March 2021

### Abstract

wip

## 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, including the tech industry (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).

A major contributor to the diversity problem in tech is retention, or the rate at which underrepresented groups leave the field. Retention is frequently discussed in the literature surrounding representation in STEM education, where many women and racial minorities express interest in STEM fields and begin studying them at high rates, 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 is commonly referred to as the leaky pipeline, which extends from early on in STEM education (such as in highschool), to graduate and professional positions in STEM (Clark Blickenstaff\* 2005). As we look further down the pipeline, we see less high-level positions being held by underrepresented groups. This idea 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.

---

\*Code and data are available at: <https://github.com/westinmo/2020-dev-survey-analysis>.

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). These concerns are often worse for women of colour, particularly black women, Latina women, and Native American women. 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 thought 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 underrepresentation 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 Hispanic 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 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, as well as the interaction between the two, on income 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. The next sections discuss the process of building and implementing the regression model used to model income, 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 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).

## 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<sup>1</sup> 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 wanted to narrow down and focus on the effects of gender and ethnicity on income specifically for individuals living in the United States with a full-time employment status<sup>2</sup>. Of the 9765 respondents who reside in the United States and are full-time employees, 6860 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.

## 2.2 Variables

## 3 Model

## 4 Results

## 5 Discussion

---

<sup>1</sup>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>

<sup>2</sup>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

# Appendix

## References

- Alfrey, Lauren, and France Winddance Twine. 2017. “Gender-Fluid Geek Girls: Negotiating Inequality Regimes in the Tech Industry.” *Gender & Society* 31 (1): 28–50.
- 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>.
- Asai, David J. 2020. “Race Matters.” *Cell* 181 (4): 754–57.
- 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>.
- Clark Blickenstaff\*, Jacob. 2005. “Women and Science Careers: Leaky Pipeline or Gender Filter?” *Gender and Education* 17 (4): 369–86.
- CompTIA. 2020. *Cyberstates 2020*. [https://www.cyberstates.org/pdf/CompTIA\\_Cyberstates\\_2020.pdf](https://www.cyberstates.org/pdf/CompTIA_Cyberstates_2020.pdf).
- Dickey, Megan Rose. 2015. *The Future of Trans\*H4CK*. Tech Crunch. <https://techcrunch.com/2015/11/29/the-future-of-trans4ck/>.
- 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](https://girlswhocode.com/wp-content/uploads/2019/08/GWC_Advocacy_InternshipApplicationExperiences_PDF_z6.pdf).
- 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/>.
- 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.
- 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.”
- R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- 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>.
- Scott, Allison, Freeda 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.
- 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.
- Stack Overflow. 2021. *Stack Overflow Company Page*. <https://www.bloomberg.com/news/articles/2020-06-23/apple-amazon-facebook-google-microsoft-data-on-black-hiring>.

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.

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, 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>.