City University of New York (CUNY)
CUNY Academic Works

Dissertations, Theses, and Capstone Projects | CUNY Graduate Center

6-2021

Data Analysis and Visualization to Dismantle Gender Discrimination in the Field of Technology

Quinn Bolewicki
The Graduate Center, City University of New York

How does access to this work benefit you? Let us know!

More information about this work at: https://academicworks.cuny.edu/gc_etds/4418
Discover additional works at: https://academicworks.cuny.edu

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

DATA ANALYSIS AND VISUALIZATION TO DISMANTLE GENDER DISCRIMINATION IN THE FIELD OF TECHNOLOGY

by

QUINN BOLEWICKI

A capstone project submitted to the Graduate Faculty in Data Analysis and Visualization

in partial fulfillment of the requirements of the Master of Science,

The City University of New York

2021

© 2021

QUINN BOLEWICKI

All Rights Reserved

Data Analysis and Visualization to Dismantle Gender Discrimination in the Field of Technology

by

Quinn Bolewicki

This manuscript has been read and accepted for the Graduate Faculty in Data Analysis and Visualization in satisfaction of the capstone project requirement for the degree of Master of Science

May 9, 2021	Aucher Serr
Date	Capstone Project Advisor
May 13, 2021	Matthew K. Gold
Date	Program Director

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Data Analysis and Visualization to Dismantle Gender Discrimination in the Field of Technology

by

Quinn Bolewicki

Advisor: Aucher Serr

In the United States, a significant population is facing an uphill battle trying to thrive in an industry that has seen exponential growth in recent years. Women, who account for approximately 50.8% of the U.S. population[1] are statistically underpaid and underrepresented in science, technology, engineering, and mathematics (STEM).[2] Despite women-led technology teams establishing a 21% greater return on investment than teams who don't, and young women largely outperforming men in math according to a 2015 study, there are only three fortune 500 companies led by women, and they comprise only 10% of internet entrepreneurs.[3] Research generates hundreds of articles, infographics, and reports showing that women are capable of, yet not thriving in technology occupations. *Brent Ozar Unlimited* and *PayScale* have provided data from the past seven years on top tech companies demographics and salaries. I have analyzed this

[1] Bureau, US Census. *Census.gov*, www.census.gov/.

[2] Hill, Catherine, et al. *Why so Few: Women in Science, Technology, Engineering, and Mathematics*. AAUW, 2010.

[3] "Infographic: How Does Access to Technology Lead to Gender Equality?" *Ignite*, 3 Dec. 2014, ignite.globalfundforwomen.org/gallery/infographic-how-does-access-technology-lead-gender-equality.

data using various statistical methods and created Envision Equality. Using this platform, I've visualized and displayed data from these sources in the form of charts and infographics. The intention of this project is to start a dialogue surrounding the issue of gender discrimination in technology occupations, and begin to address it.

TABLE OF CONTENTS

LIST OF FIGURES

DIGITAL MANIFEST

I. Capstone Project Whitepaper (PDF)

II. WARC Files

a. Project Website

Archived version of https://qbolewicki.github.io/Capstone_site/

III. Code

a. Zip file containing the contents of GitHub repository at the time of deposit.

(https://github.com/Qbolewicki/Capstone_site)

TECHNICAL SPECIFICATIONS

Website

Envision Equality is a webpage built entirely over the course of the Spring 2021 semester. The goal of this website is to raise awareness around the issue of gender inequality in the field of technology. The website by design is clean and simple to interact with. It was built using HTML and CSS, and features data visualizations created with JavaScript and Adobe Illustrator. This project is composed of 23 elements and over 300 hours of planning and building. At its foundation are seven HTML files that call upon two JavaScript documents, three JavaScript libraries, four PNG's, two SVG's, two CSV's, and a CSS style sheet. Theses files then return a structure for the entire website.

The JavaScript provides the logic for creating the "Gender Population" and "Hours Worked" visualizations. It is important to note that if the data in the attached CSV's change, the visualizations will update upon refreshing the browser to reflect these changes. The third visualization, "Pay Gap", is an SVG which was created using crowdsourced data from *Brent Ozar Unlimited*. With the average pay gap between men and women in data occupations calculated, Adobe Illustrator was used to measure and reflect this. By scaling the SVG to 500 pixels and removing 6.37% of the illustration, the gender pay gap was visually represented. Finally, the SVG was embedded in an HTML spreadsheet which now displays a static image on the website. The code for latest version of this project can be viewed at https://github.com/Qbolewicki/Capstone_site. The final version of the website can be viewed published at https://Qbolewicki.github.io/Capstone_site/.

Datasets

Research for this project began in January 2021, but the decision was made to source existing data in March. There is a problem though, regarding salary and gender data in the United States. "Even in the case of publicly traded companies, salaries are only typically disclosed for top executives. Privately held companies are not compelled to disclose salary information to the public, because the general public can't invest in private companies."[4] For this reason, as well as potentially implicating themselves in breaking the law, it is difficult to find data from companies themselves regarding gender inequality. For existing data on database occupations, I consulted *Brent Ozar Unlimited* on April 9th, 2021. This dataset titled "Data Professional Salary Survey Result*s"*[5] was crowdsourced. It contains 6,894 survey responses from the years 2017 to 2019 in Microsoft Excel format. The data needed to be cleaned to create "Pay Gap". As with any crowdsourced material, there were inevitably misunderstandings and misleading responses. An example of this can be observed in the "Gender" field where participants chose to respond with "Dragon" or "Attack Helicopter". Another example was the participant who claimed to work 200 hours a week. There are 160 hours in a week, so that survey response was omitted for the analysis.

Another data source for this project was from *PayScale*. This is American compensation software and data company that helps employers manage employee compensation, and

---

[4] Rowley, Jason D. *Visa Data Shows Billion-Dollar Tech Startups Pay Premium Salaries for Tech Talent*. 25 Apr. 2019, news.crunchbase.com/news/visa-data-shows-billion-dollar-tech-startups-pay-premium-salaries-for-tech-talent/.

[5] Ozar, Brent. "The 2019 Data Professional Salary Survey Results." *Brent Ozar Unlimited®,* 6 Jan. 2019, www.brentozar.com/archive/2019/01/the-2019-data-professional-salary-survey-results/.

employees understand their worth in the job market.[6] I chose *PayScale* for their extensive databases on crowd sourced and company sourced information. Additionally, they have published original research on the gender pay gap.[7] I first accessed the data used for "Gender Population" and "Hours Worked" on March 25th, 2021. The data titled, "By the Numbers: Comparing Tech Employee Salary, Age, Stress and More" [8] was in the form of a table on *PayScale's* website. The table was reasonably small, so I was able to manually transcribe it into a Microsoft Excel spreadsheet.

Analysis

For the data used in "Pay Gap" I utilized the *Brent Ozar Unlimited* dataset. This data contained two very important fields for this visualization: "Gender" and "SalaryUSD". To see if there was a gender pay gap, I first had to separate rows into two different sheets by "Male" and "Female" indicated in the "Gender" column. Within these sheets in Microsoft Excel, I used a formula to calculate the average salaries which was $91,739 for women and $97,580 for men. Lastly, I quantified the percentage of that gap, which for this dataset equated to 6.37%. I created an SVG in Adobe Illustrator, which was scaled to 500 pixels wide. I then calculated 6.37% of 500 which is 31.85. At that point I removed 32 pixels from the SVG. We can now view a 6.37% gap in the illustration of the gender symbol. This missing section illustrates the gender pay gap

[6] Bieber, Christy. "Pros and Cons of an Open Salary Policy." *The Motley Fool*, 28 June 2018, www.fool.com/careers/2018/06/28/pros-and-cons-of-an-open-salary-policy.aspx.

[7] "Racial and Gender Pay Gap Statistics for 2021." *PayScale*, 2021, www.payscale.com/data/gender-pay-gap.

[8] "Highest Tech Salaries & Other Job Stats." *PayScale*, 2015, www.payscale.com/data-packages/top-tech-companies-compared/tech-salaries.

observed in the *Brent Ozar Unlimited* dataset. This illustration, although created manually, can be recreated using any SVG in Adobe Illustrator.

For the "Weekly Hours" visualization, I chose to analyze average number of hours worked per week according to gender. This visualization was also sourced from the *Brent Ozar Unlimited* dataset. It required two different Excel sheets separated by "Male" and "Female" indicated in the "Gender" column. Within these sheets in Excel, I used a formula to calculate average number of hours worked per week in the "HoursWorkedPerWeek" column. On average, women worked 43.24 hours per week and men worked 43.08 hours per week. I transcribed the data into a CSV to be embedded into the JavaScript for "Hours Worked".

Lastly, I transcribed the data for our radial chart, "Gender Distribution". This data was sourced and examined by *PayScale* and did not require further analysis in Microsoft Excel. I converted two columns from the Excel spreadsheet into a CSV to be embedded in JavaScript for this visualization. The columns that I chose to focus on from the original data are titled "Employer Name" and "% Female". These titles were changed to "Company" and "Female" in the CSV to remove special characters and spaces.

There were several key takeaways from these analyses. According to the *Brent Ozar Unlimited* dataset, female data professionals on average work more hours per week than males and fulfill roles that are paid 6.37% less. These findings are demonstrated in the "Pay Gap" illustration, and "Weekly Hours" bar chart. In "Pay Gap", the viewer can observe a 6.37% gap in the illustration of the gender symbol. This gap symbolizes the gender pay gap in data professional occupations. In the "Weekly Hours" bar chart, it is presented that women in data professional occupations work slightly more hours per week than their male peers. This visualization could be used to nullify any argument that more women choose to work part-time,

which would account for a discrepancy in average salary between genders. Finally, we have the takeaways from the *PayScale* data, which indicates that many familiar technology companies are male dominated. This is demonstrated using a radial graph that is distinctly unevenly distributed.

Although unfortunate, the data and analysis results confirmed my suspicion that there are existing biases surrounding gender in various technology occupations. The pay gap is extreme, as well as the number of men versus women employed. The analysis for "Hours Worked" did give me a moment of pause. I was sure that some of the discrepancy in annual salary could be accounted for by part time employees. Upon analysis, this hypothesis was disproven. I look forward to exploring more arguments surrounding this topic through data analysis and visualization.

LIST OF ABBREVIATIONS

Abbreviation	Explanation
CSS	Cascading Style Sheets
CSV	Comma-separated values
HTML	Hypertext Markup Language
PDF	Portable Document Format
STEM	Science, Technology, Engineering, and Math
SVG	Scalable Vector Graphics
WARC	Web ARChive
ZIP	Archive file format

NARRATIVE

Capstone Project Description

Women as a community are underrepresented in science, technology, engineering, and mathematics (STEM).[9] 11 years ago, I began my own journey as a woman working in the field of technology. During that time, I have become exceedingly aware of a gender imbalance in the workplace. In February 2021, I plunged into research regarding gender inequality in STEM occupations. Over the course of my graduate career, I discovered my passion for making socio-cultural progress using data visualizations. During this time, I found myself focusing on challenges specific to women. For this semester-long project I used data and technology to retaliate against the biases of their own industries. I gathered, cleaned, and analyzed data regarding gender and technology occupations. Additionally, I created a platform in the form of a website where I could visualize these data findings and build upon the project in the future. This would prove that not only are women capable of exceeding in STEM, but the data supports it. To attempt this operation, I met with my advisor, Professor Serr. My advisor guided me though developing this idea into a practical and cohesive capstone project, that I now refer to as *Envision Equality*.

Throughout this project, design decisions have been a major component. On *Envision Equality,* everything from colors to fonts were carefully considered factors. It was important that the final output of this project felt bold, groundbreaking, and sophisticated. A good deal of inspiration was taken from observing how high-end tech product pages looked and functioned.

[9] Hill, Catherine, et al. *Why so Few: Women in Science, Technology, Engineering, and Mathematics*. AAUW, 2010.

Blues, golds, and grays were used to contrast the dark background and create cohesion throughout the various webpages. I chose to avoid sharp corners in the navigation and stock images to give the site a modern and relevant tone. Additionally, the same colors were assigned to represent men and women to increase legibility, and small caps were used in visualizations and throughout the website to highlight important text. The resulting website and data visualizations are featured on https://Qbolewicki.github.io/Capstone_site/.

Focus Area and Previous Course of Study

This capstone project is built upon my experience in technology, hours of research, and my time as a graduate student in the CUNY Graduate Center Data Analysis and Visualization program. My first data visualization course, Lev Manovich's "Data Visualization", taught me the fundamentals of statistical data analysis. I learned how to translate data analysis into an effective visualization by any means necessary. This course built the foundation I needed to take my existing skillset and use it to successfully produce an infographic. With this in mind, I created "Pay Gap" using data analyzed from Microsoft Excel and presented it using Adobe Illustrator.

"Hours Worked" and "Gender Distribution" visualizations were profoundly influenced by "Interactive Data Visualization", instructed by Ellie Frymire and Aucher Serr. This course taught the fundamentals of building data visualizations using JavaScript; specifically using a library called D3. Using JavaScript and databases to build visualizations gives them dynamic transitions when prompted. For "Gender Distribution" this transition not only captures the attention of the viewer but alludes that there is an inequality in the categories being presented.

In addition to these highly influential courses, I took several other classes and workshops throughout my time at CUNY Graduate Center that impacted the way I approached my capstone

project. In 2018, I took a workshop called "Understanding (&) Building on the Web with HTML/CSS" instructed by CUNY Graduate Center Digital Fellows, which taught the basics of HTML and CSS. It was there that I launched my first local server and composed a website. This workshop gave me the confidence to pursue building a website for my capstone project. Michelle McSweeney's "Introduction to Data Visualization and Design: Fundamentals" was an in-depth investigation into creating a data visualization project that looked cohesive and professional. It was in that course that I began to explore my personal aesthetic and see it take shape in the form of big data analysis and visualization. In the final semester before perusing my capstone project, I enrolled in "Data Analysis Methods" instructed by Howard Everson. The benefit from taking this class was unquantifiable; I completed the course with a confidence in statistics that I have never previously been able to achieve. Every course that I have enrolled in at the CUNY Graduate Center has had some impact on the outcome of my capstone project. Although influenced by my life experience, these courses gave me the tools I needed to bring this project to fruition.

Evaluation

My original scope for the capstone project included creating data based on survey results from people working in technology, statistically analyzing that data, and creating live visualizations on a website. Although the project pivoted along the way, I was not deterred from the main goal of empowering women in technology though awareness and validation. In this section of the white paper, I will address aspects of the project that were not realized and give the solutions that were implemented.

I. The Survey

A major component of this project was the way in which data was to be collected. The original scope included sending a survey with six questions to those working in STEM fields. The survey was created and can be viewed at https://forms.gle/4GkcMEuWmu8z4Au96. Additionally, various means of outreach were compiled, but time became a considerable factor when the Human Research Protection Program declared the project would in fact need approval. There were numerous steps that needed to be taken to ensure the protection of the potential human subjects. Upon receiving this statement with approximately six weeks left to complete my capstone project, I chose to research existing crowdsourced data.

II. Live Visualizations

Formerly, I wanted to have a living database. This meant gathering live data from surveys from across the United States and displaying visualizations which evolved over time. Once the Human Research Protection Program was concerned, I chose to focus on data collected between 2014 and 2019 to ensure quality and reliability. Although this did not mean that the visualizations had to be stagnant, it did mean that they weren't going to be updated with live data on an ongoing basis. As a consequence, we would lose a key component of the project. With this in mind I decided to create visualizations that transitioned dynamically upon refreshing the webpage. This would give the visualizations an interactive component, while being fed existing data that did not change.

Over the course of completing this project, there were numerous successes and failures. One accomplishment was the adaptability presented in the face of many challenges. Time restraints were a factor that although carefully considered, ended up being the greatest hindrance.Despite these challenges, the project maintained its core values and was successful in its goal of raising awareness around gender inequality in STEM occupations.

Continuation of the Project

The project was successfully published as a webpage within the parameters outlined this semester. With that in mind, visualizations will be added, and I will reach out to various newsletters and publications to maximize outreach. In the future, I plan to create an additional three visualizations. These will demonstrate ways in which this field could improve upon gender equality practices. The visualizations would indicate job satisfaction, examine the ways in which men as opposed to women leave their positions and whether it was voluntary, and finally present a time series to analyze whether the gender gap is closing or expanding further.

Additionally, there are some aspects of this capstone project that I would like to improve upon as I expand my knowledge in JavaScript, HTML, and CSS. Two of those aspects are user experience and optimal knowledge gained. An example of improved user experience would be making the visualizations have scroll-based transitions rather than entering upon refreshing the webpage. This way the user would actively load the visualizations by scrolling down the website, creating a direct connection between the user and data. The goal of this interaction would be to create an investment on the end of the user; they must scroll to view the transition, and consequently the next visualization.

My goals were to create a website featuring data visualizations by a woman, for everyone. I want to have an impact on the lives of women in technology, and possibly the fieldas a whole. Ultimately, I hope these visualizations will start a dialogue surrounding the inequality in STEM occupations, and solutions to this problem. That discourse would make thiscapstone project a great success.

FIGURES AND DIAGRAMS

I. Pay Gap

a. File Type:

SVG

b. File Name:

gender_gap.svg

This illustration displays my gender pay gap analysis from the *Brent Ozar Unlimited* dataset. I chose to combine the gender symbols into one graphic to represent unity. I then removed 6.37% of the image to simulate the 6.37% gender pay gap in professional data occupations.

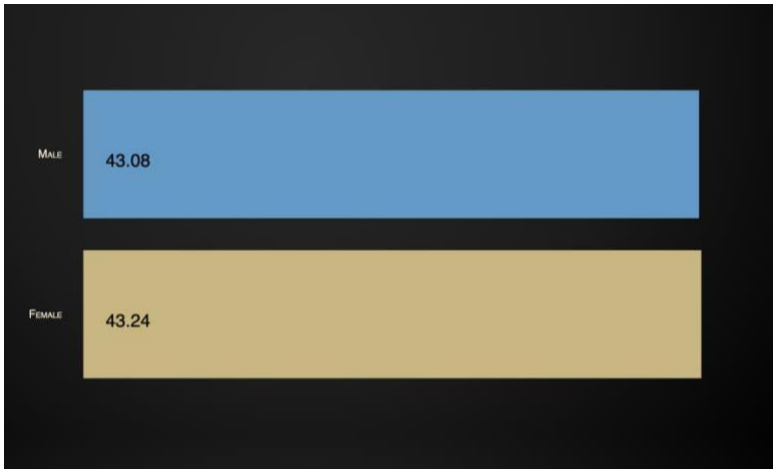
II. Hours Worked

a. File Type:

JavaScript

b. File Name:

barchart.js



"Hours Worked" is a horizontal bar graph displaying data from *Brent Ozar Unlimited*, regarding number of hours reported as worked weekly by men versus women. A bar graph is a simple and effective way to show that there is no major difference in number of hours reported according to gender.

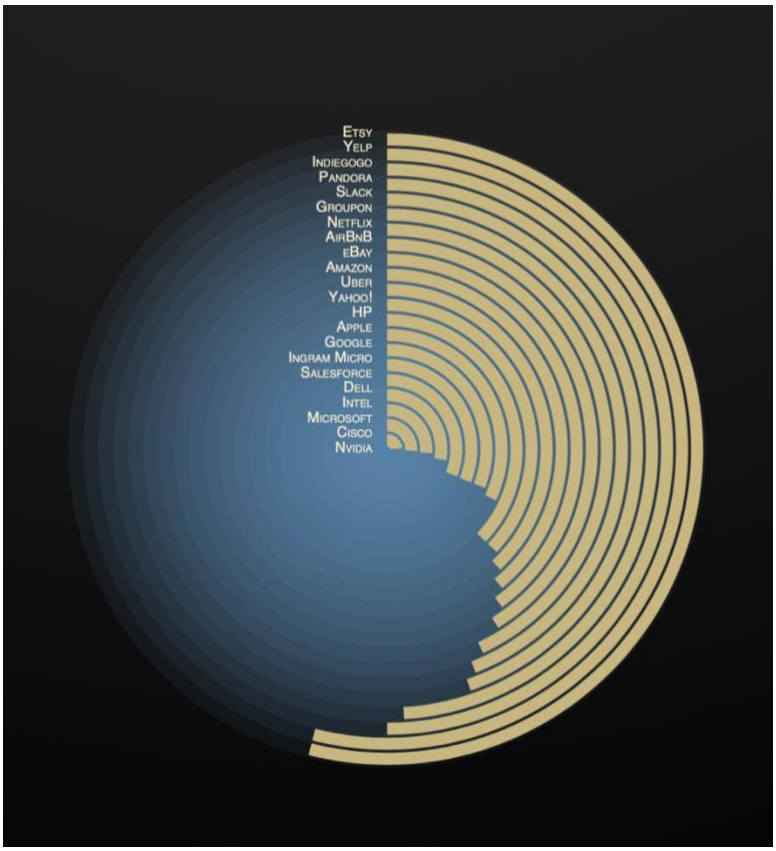
III. Gender Distribution

a. File Type:

JavaScript

b. File Name:

circlegraph.js



Here, a radial graph represents and female population in a series of technology companies. This container can easily be recognized as less than 50% full when the data from *PayScale* transitions in. Additionally, using bars allows the viewer to easily identify the most and the least diverse companies in the dataset.

List of Variables

hours_worked.csv

Title	Description
Gender	Identifying gender
Hours	Number of hours reported as worked per week

diversity_2018.csv

Title	Description
Company	Top technology companies in 2015
Female	Percent of the company identifying as female

Glossary of Functions

JavaScript

arcTween(d, i): start count of bars

define(runtime, observer): defines runtime and observer libraries

getInnerRadius(index): radius of chart

getOuterRadius(index): space between bars

rad2deg(angle): start angle of chart

Microsoft Excel

AVERAGE: calculate average of total in selected cells

Bibliography

"About the Coalition." *Reboot Representation*, https://www.rebootrepresentation.org/about-the-coalition/. Accessed 10 Feb. 2021.

Agarwal, Dr Pragya. "Gender Bias In STEM: Women in Tech Still Facing Discrimination." *Forbes*, https://www.forbes.com/sites/pragyaagarwaleurope/2020/03/04/gender-bias-in-stem-women-in-tech-report-facing-discrimination/. Accessed 10 Feb. 2021.

Ahuja, M. K. "Women in the Information Technology Profession: A Literature Review, Synthesis and Research Agenda." *European Journal of Information Systems*, vol. 11, no. 1, Taylor & Francis, Mar. 2002, pp. 20–34. *Orsociety.tandfonline.com (Atypon)*, doi:10.1057/palgrave.ejis.3000417.

Bostock, Mike. *D3.Js – Data-Driven Documents*. https://d3js.org/. Accessed 10 Feb. 2021.

Breslin, Jennifer and USA. "We Must Act Now for Science and Technology to Drive Gender Equality." *Ignite – Global Fund for Women*, 18 Feb. 2015, http://ignite.globalfundforwomen.org/gallery/we-must-act-now-science-and-technology-drive-gender-equality.

Closing the Gender Gap for Women in Technology | McKinsey. https://www.mckinsey.com/industries/technology-media-and-telecommunications/our-insights/closing-the-tech-gender-gap-through-philanthropy-and-corporate-social-responsibility#. Accessed 10 Feb. 2021.

Etzkowitz, Henry, et al. *Athena Unbound: The Advancement of Women in Science and Technology*. Cambridge University Press, 2000.

Hill, Catherine, et al. "Why So Few? Women in Science, Technology, Engineering, and Mathematics." *American Association of University Women*, American Association of University Women, 2010. *ERIC*, https://eric.ed.gov/?id=ED509653.

"Home." *Reboot Representation*, https://www.rebootrepresentation.org/. Accessed 10 Feb. 2021.

"Infographic: How Does Access to Technology Lead to Gender Equality?" *Ignite – Global Fund for Women*, 6 Nov. 2014, http://ignite.globalfundforwomen.org/gallery/infographic-how-does-access-technology-lead-gender-equality.

"ITU, UN Women Announce New Global Awards for Outstanding Work in Technology for Gender Equality." *UN Women*, https://www.unwomen.org/en/news/stories/2014/7/itu-awards-for-gender-equality. Accessed 10 Feb. 2021.

Kolhatkar, Sheelah. "The Disrupters." *The New Yorker*, https://www.newyorker.com/magazine/2017/11/20/the-tech-industrys-gender-discrimination-problem. Accessed 10 Feb. 2021.

Lemons, Mary A., and Monica Parzinger. "Gender Schemas: A Cognitive Explanation of Discrimination of Women in Technology." *Journal of Business and Psychology*, vol. 22, no. 1, Sept. 2007, pp. 91–98. *Springer Link*, doi:10.1007/s10869-007-9050-0.

NW, 1615 L. St, et al. "How Bad Is Gender Discrimination in Tech? Men, Women Disagree." *Pew Research Center*, https://www.pewresearch.org/fact-tank/2017/10/10/women-are-more-concerned-than-men-about-gender-discrimination-in-tech-industry/. Accessed 10 Feb. 2021.

"Obstacles to Female Leadership." *Pew Research Center's Social & Demographic Trends Project*, 14 Jan. 2015, https://www.pewresearch.org/social-trends/2015/01/14/chapter-3-obstacles-to-female-leadership/.

Ozkazanc-Pan, Banu, and Susan Clark Muntean. "Networking towards (in)Equality: Women Entrepreneurs in Technology." *Gender, Work & Organization*, vol. 25, no. 4, 2018, pp. 379–400.*Wiley Online Library*, doi:https://doi.org/10.1111/gwao.12225.