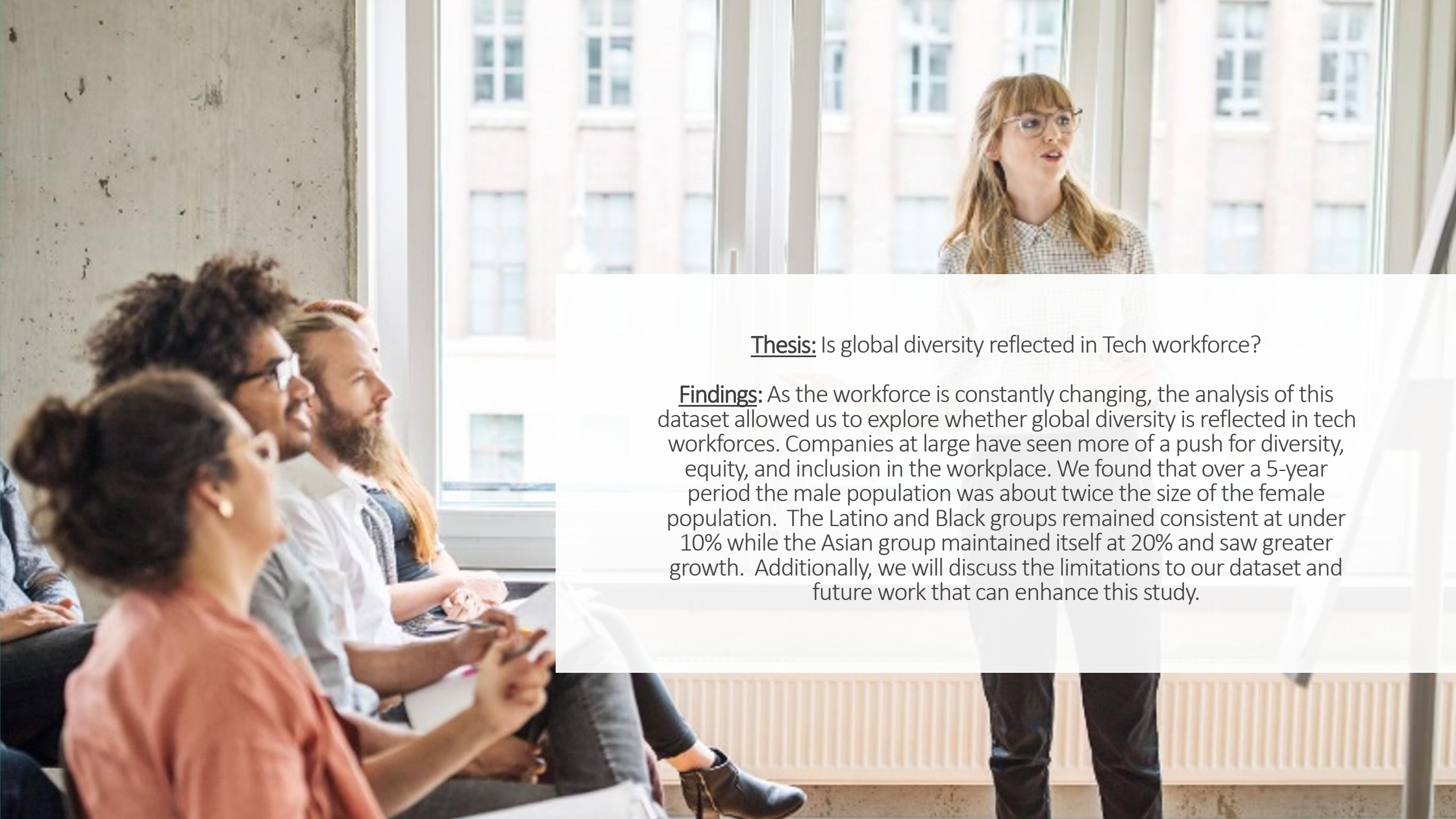
Diversity in Tech

Neyda Morales, Carlos Ruiz,
Paris Lee, Daniel Purrier

June 10, 2024

Agenda

- Thesis
- Introducing our Data
- Research Questions
- Linear & Statistical Modeling
- Bias and Limitations
- Future Work
- Call to Action
- Works Cited
- Q&A



Thesis: Is global diversity reflected in Tech workforce?

Findings: As the workforce is constantly changing, the analysis of this dataset allowed us to explore whether global diversity is reflected in tech workforces. Companies at large have seen more of a push for diversity, equity, and inclusion in the workplace. We found that over a 5-year period the male population was about twice the size of the female population. The Latino and Black groups remained consistent at under 10% while the Asian group maintained itself at 20% and saw greater growth. Additionally, we will discuss the limitations to our dataset and future work that can enhance this study.

Introducing our Data

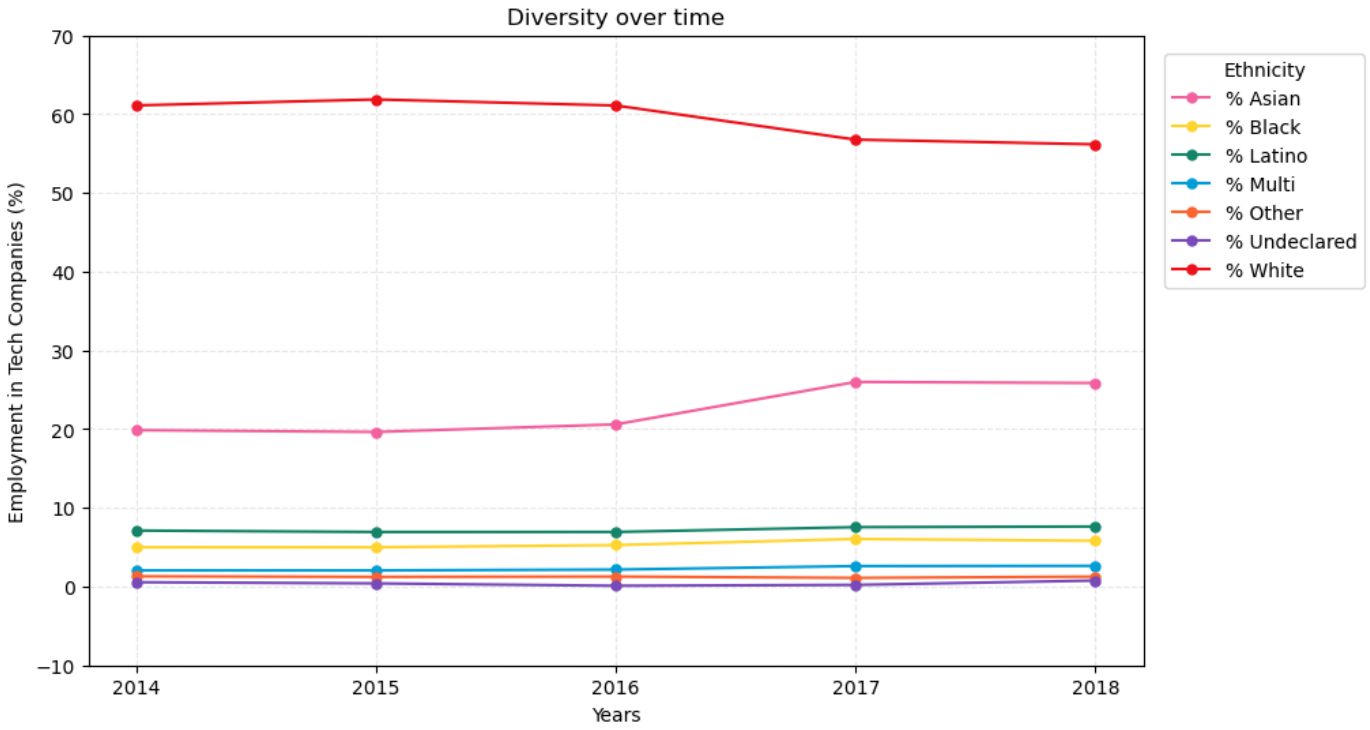
Data Cleaning Overview:

Removed duplicate of Apple, Rounded down to 0% for anything that was <1, Converted all dashes to 0%, Converted all percentages to floats

```
Index: 91 entries, 0 to 93
Data columns (total 11 columns):
 #   Column        Non-Null Count  Dtype
---  ------        --------------  -----
 0   Year          91 non-null     int64
 1   Company       91 non-null     object
 2   Female %      91 non-null     float64
 3   Male %        91 non-null     float64
 4   % White       91 non-null     float64
 5   % Asian       91 non-null     float64
 6   % Latino      91 non-null     float64
 7   % Black       91 non-null     float64
 8   % Multi       91 non-null     float64
 9   % Other       91 non-null     float64
 10  % Undeclared  91 non-null     float64
dtypes: float64(9), int64(1), object(1)
memory usage: 8.5+ KB
```

```
Year
2014                [Yahoo!, Google, Apple, Apple (excluding undeclared), Cisco, eBay
2015          [Yahoo!, Google, Apple, Apple (excluding undeclared), Cisco, eBay, HP, I
2016    [Yahoo!, Google, Apple, Apple (excluding undeclared), Cisco, eBay, HP, Indiego
2017                              [Yahoo!, Google, Apple, Cisco, eBay, HP, Indiego
2018    [Yahoo!, Google, Apple, Cisco, eBay, HP, Indiegogo, Nvidia, Dell, Ingram Micro,
```

	Year	Company	Female %	Male %	% White	% Asian	% Latino	% Black	% Multi	% Other	% Undeclared
0	2018	Yahoo!	37	63	45	44	4	2	2	3	-
1	2018	Google	31	69	53	36	4	3	4	0	-
2	2018	Apple	32	68	54	21	13	9	3	1	2
3	2018	Cisco	24	76	53	37	5	4	1	<1	-
4	2018	eBay	40	60	50	39	6	3	1	1	-
5	2018	HP	37	63	73	12	8	4	2	<1	-
6	2018	Indiegogo	50	50	58	28	7	4	-	3	-
7	2018	Nvidia	17	83	37	45	3	1	14	<1	-
8	2018	Dell	28	72	69	9	11	10	-	1	-
9	2018	Ingram Micro	31	69	52	14	19	14	1	0	-



Research Question 1

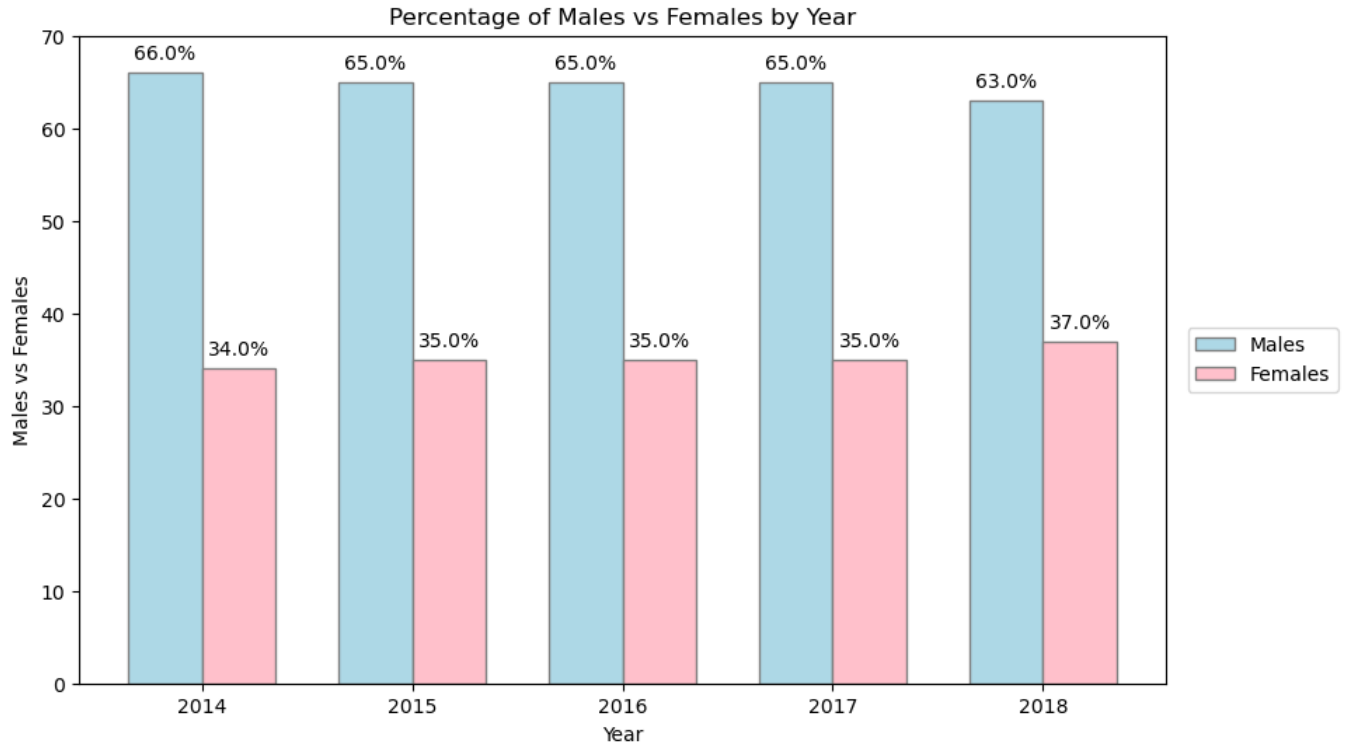
Has diversity in tech companies increased over time?

What is the percentage of Males vs Females in the tech workplace?

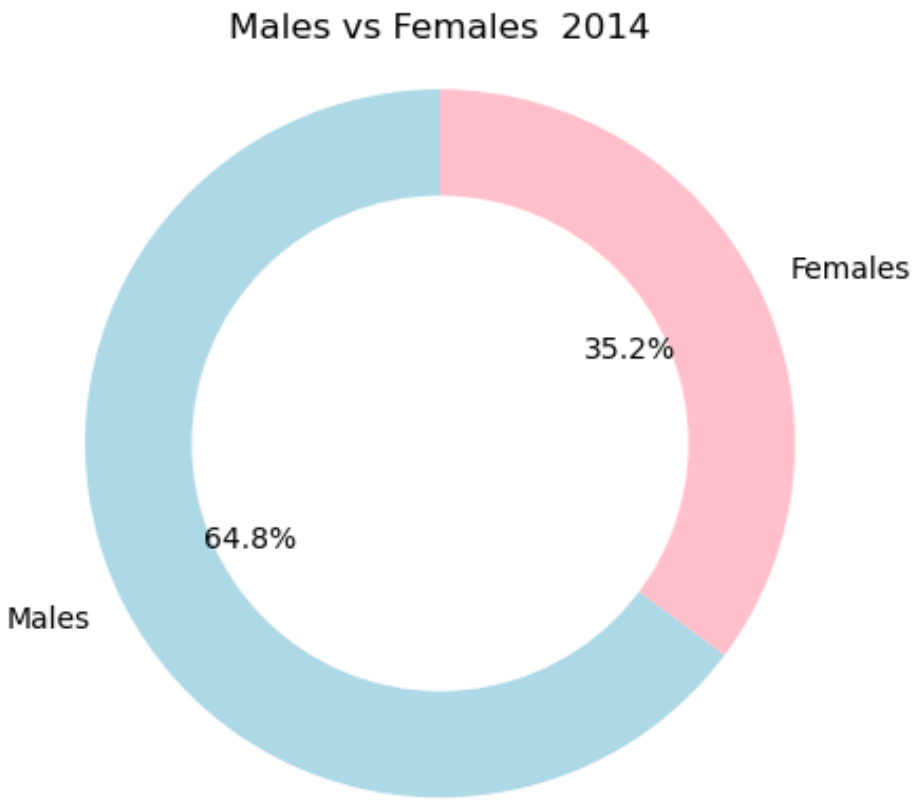
RESEARCH QUESTION 2

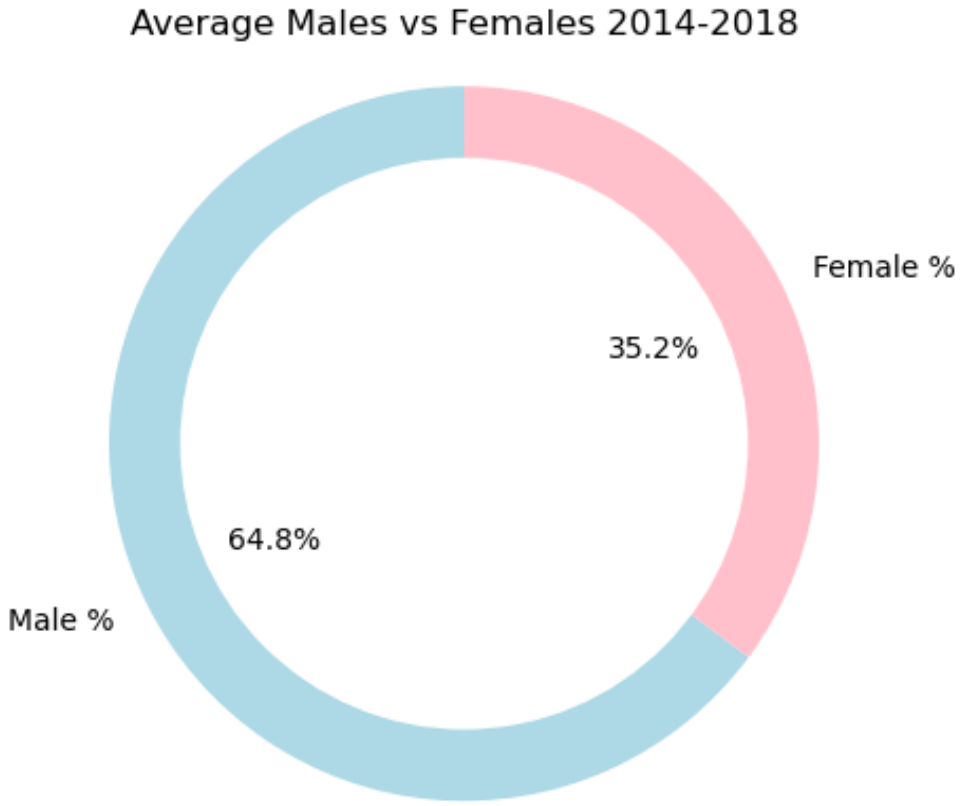
Males vs Females 2014-2018

	Year	Male %	Female %
0	2014	66.0	34.0
1	2015	65.0	35.0
2	2016	65.0	35.0
3	2017	65.0	35.0
4	2018	63.0	37.0

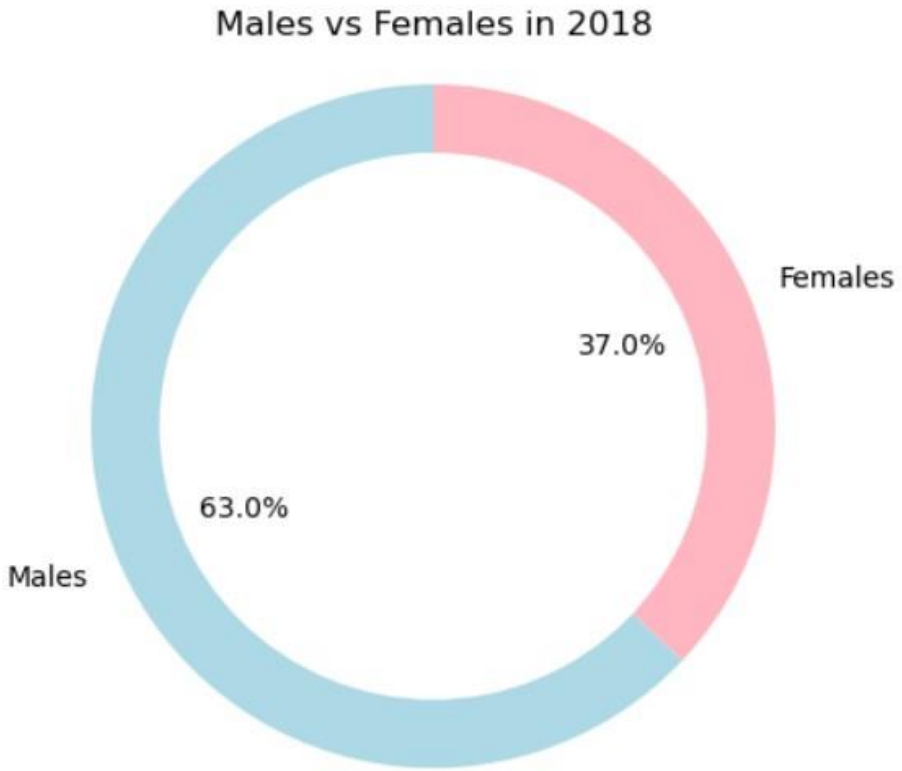


Deeper Dive


Males vs Females 2014
Females
35.2%
64.8%
Males

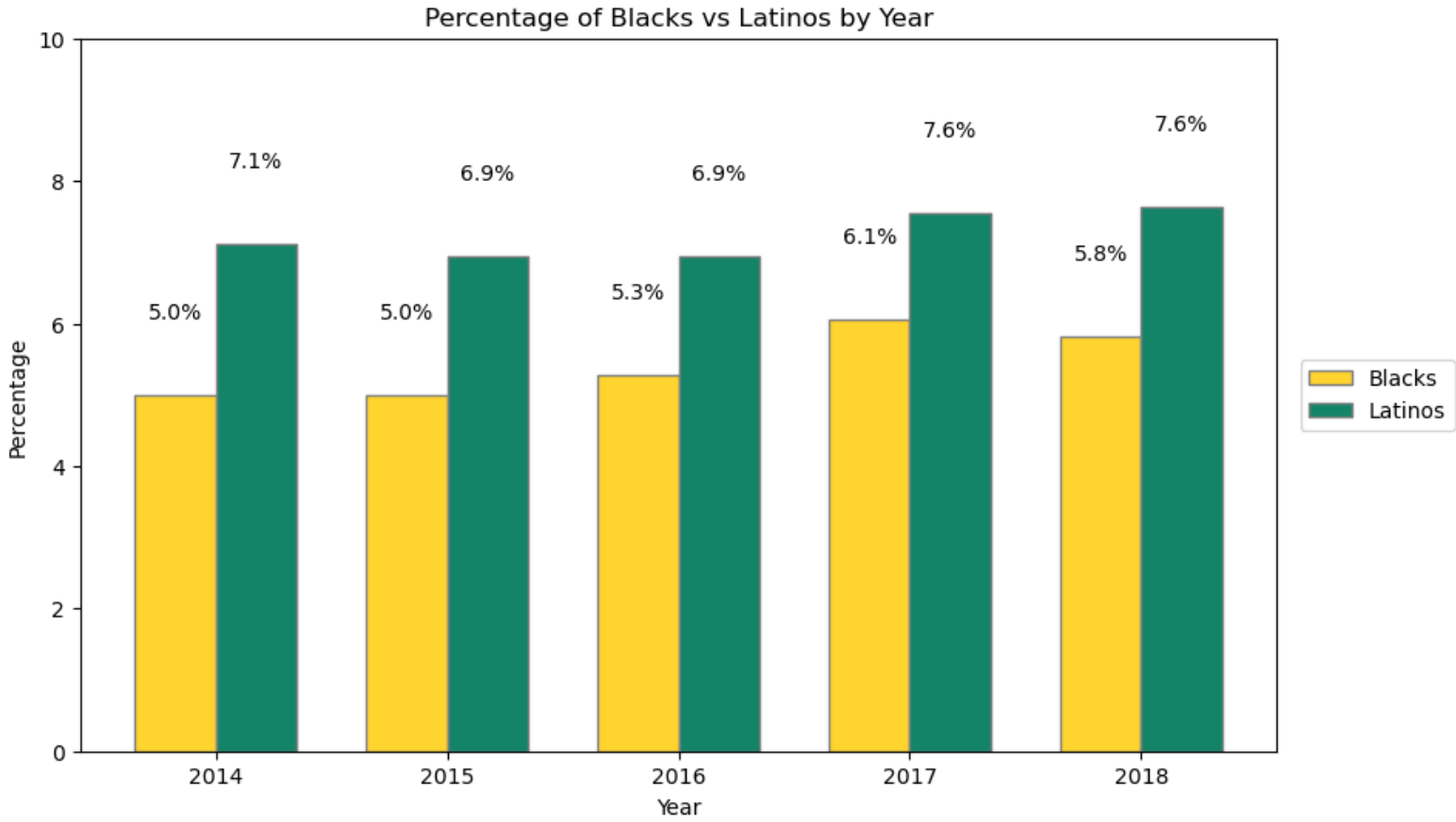

Average Males vs Females 2014-2018
Female %
35.2%
64.8%
Male %

Final Year



Research Question 3

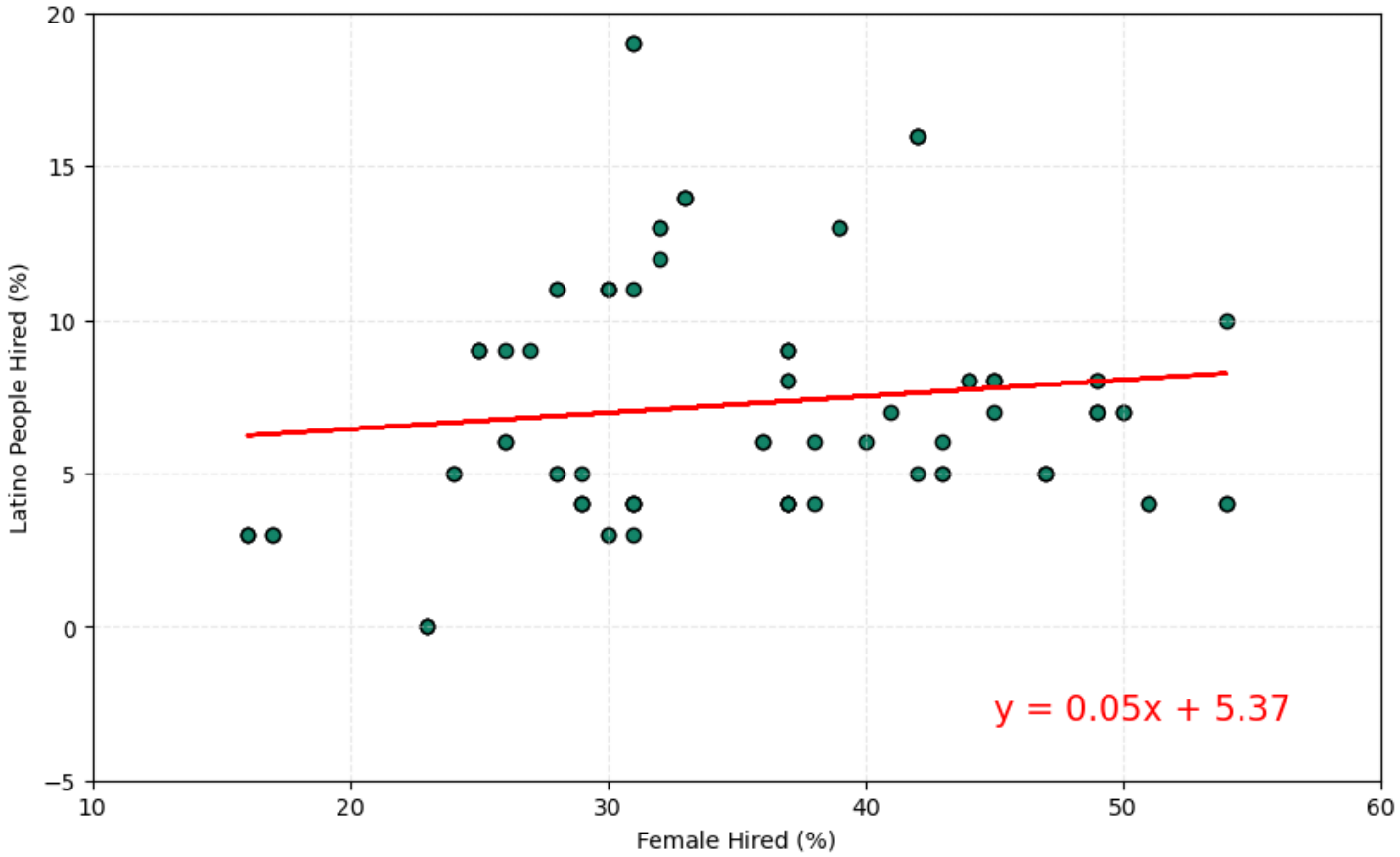
How do Black people and Latino people compare over time?



Linear & Statistical Modeling

Regression

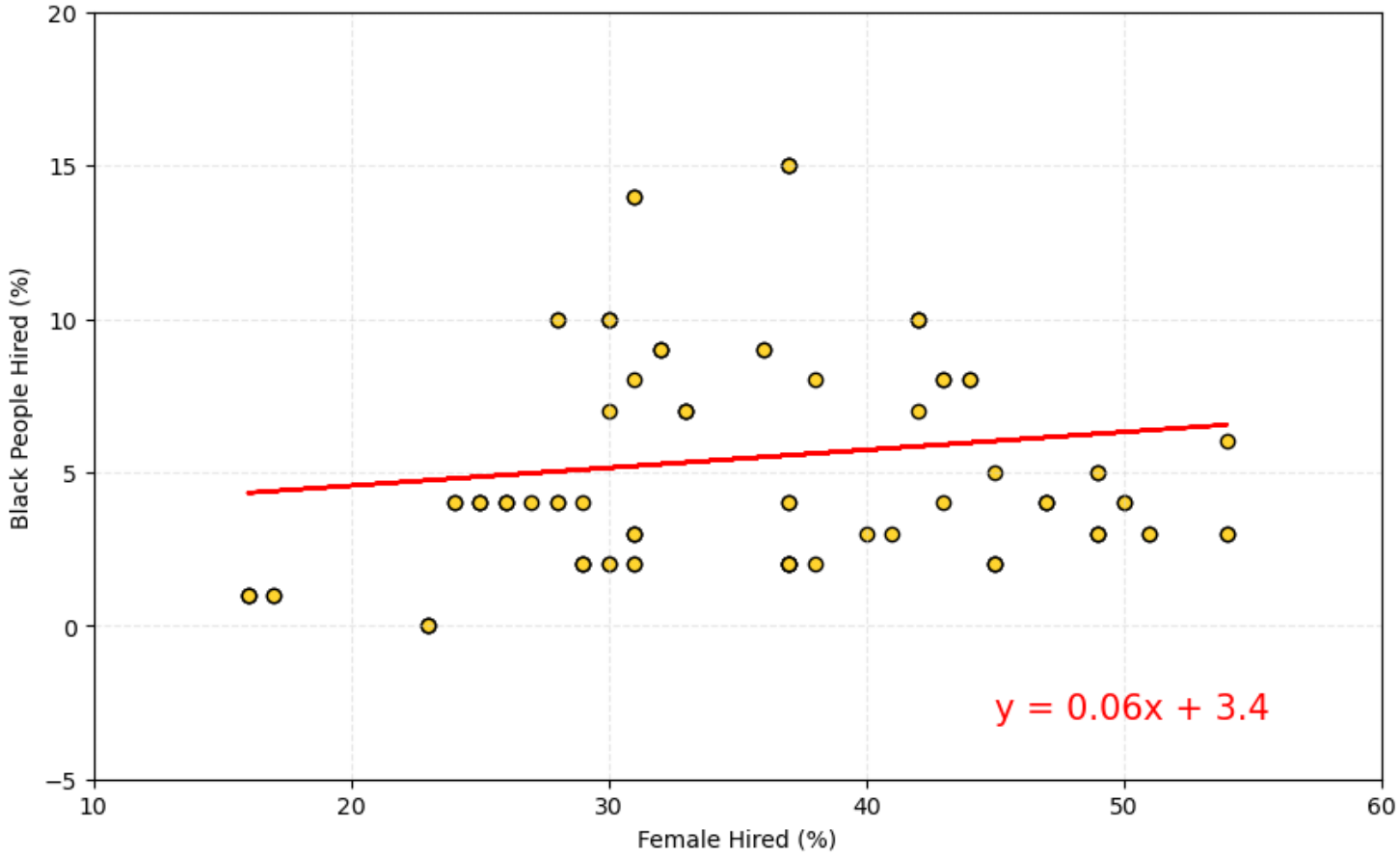
1. Are companies that hire more Women more likely to hire Latino people?



The r-value is: 0.12619415936237158

Regression

2. Are companies that hire more Women more likely to hire Black people?



The r-value is: 0.12619415936237158

Statistical Testing

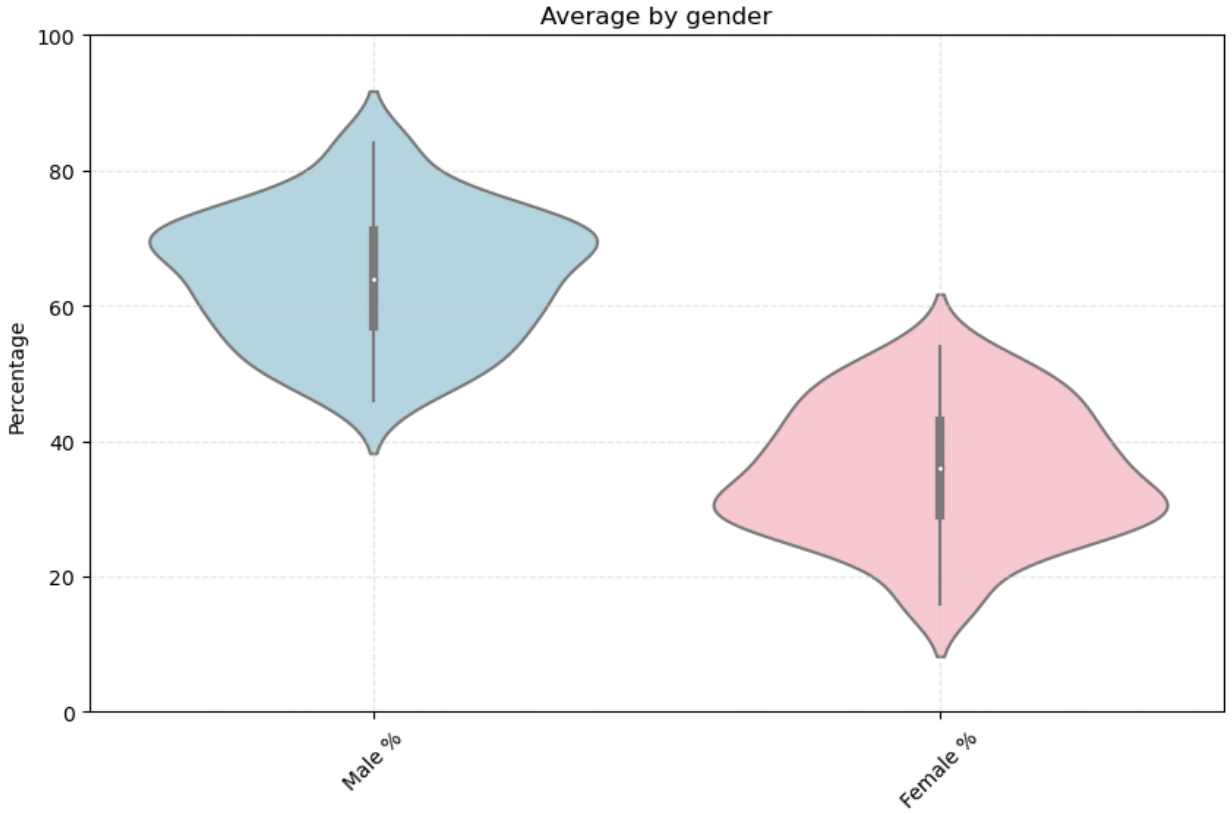
Gender - Hypothesis

- Null: There is no difference between Males vs Females working in Tech Companies.
- Alternative: There is a difference between Males vs Females working in Tech Companies.

Diversity - Hypothesis

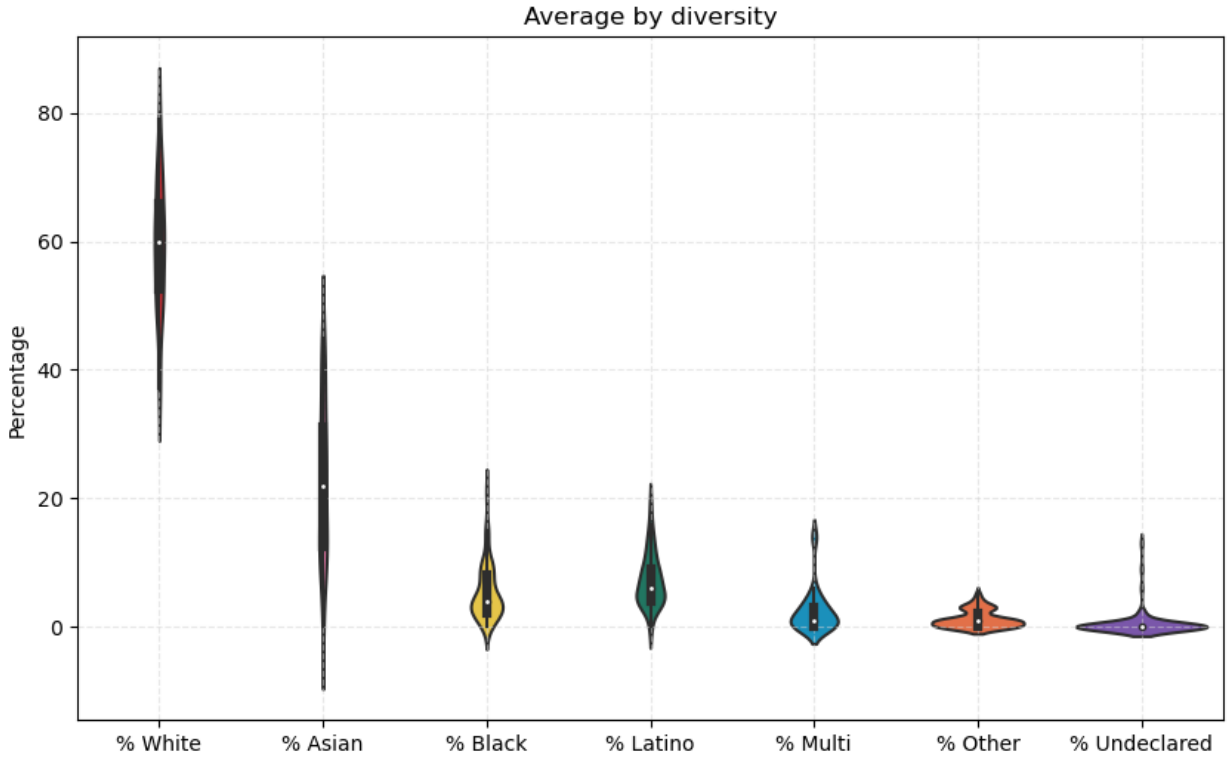
- Null: There is no difference in the averages between each diversity group.
- Alternative: There is at least one difference in the averages between each diversity group.

Gender



- We accept the alternative hypothesis – there is a significant difference between males and females across tech companies.
- TtestResult
  - (statistic=20.573574726611028, pvalue=3.78484545338512e-49, df=179.9993887209176)

Diversity



- We accept the alternative hypothesis – there is at least one diversity group with differences in the averages.
- FonewayResult
  - (statistic=985.8325475532455, pvalue=2.47133054e-316)

Bias and Limitations

SMALL LIST OF COMPANIES

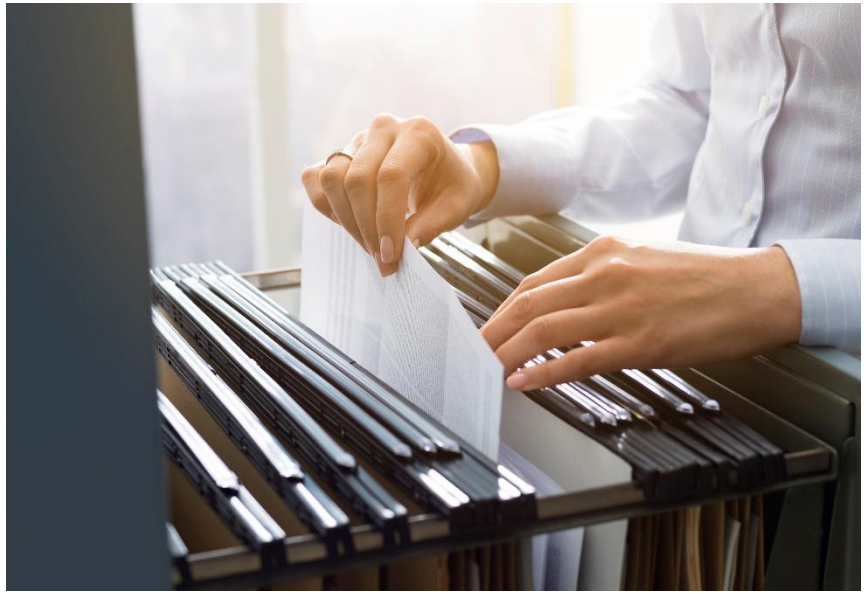
COMPANY SIZE NOT INCLUDED

RECENT YEARS NOT INCLUDED

RAW NUMBERS NOT INCLUDED (ONLY PERCENTAGES)

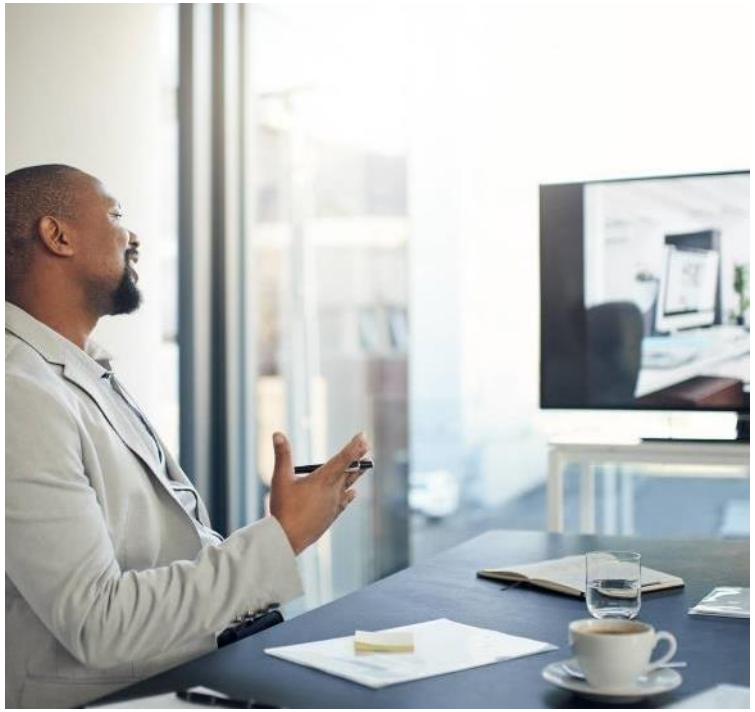
INTERSECTIONAL DATA NOT INCLUDED (EX: LATINO FEMALE, BLACK MALE)

VALUES THAT WERE UNDECLARED



Future Work

- Pull data that spans a longer period
- Incorporate company size
- Ensure that the data includes gender and race for each employee
- Obtain data that is more gender inclusive (gender diversity)
- Capture employee satisfaction data: are minority employees happy in their respective companies?



Call to Action

- Tackle unconscious bias within the workplace and on individual teams
- Introduce tech into more schools
  - workshops, camps, clubs, and after school programs centered around technology

Works Cited

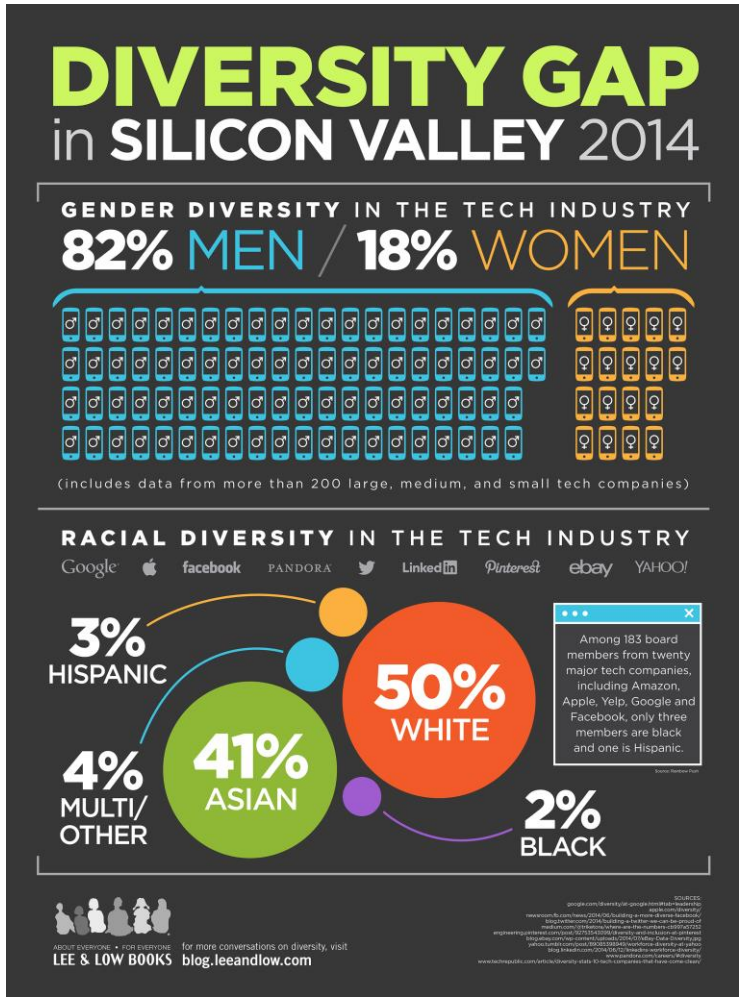
Jainaru. **"Diversity in Tech Companies."** Kaggle, www.kaggle.com/datasets/jainaru/diversity-in-tech-companies/data. Accessed 10 June 2024.

"The Diversity Gap in Silicon Valley." Lee & Low Books Blog, 12 Mar. 2015, blog.leeandlow.com/2015/03/12/the-diversity-gap-in-silicon-valley/. Accessed 10 June 2024.

"Set Color for Each Violin in Violin Plot." Stack Overflow, stackoverflow.com/questions/34058188/set-color-for-each-violin-in-violin-plot. Accessed 10 June 2024.

"The Gender Gap in Tech: Let's Talk About It." Ironhack Blog, www.ironhack.com/gb/blog/the-gender-gap-in-tech-let-s-talk-about-it. Accessed 10 June 2024.

Lardinois, Frederic. **"Black, Latino People Are Being Left Out of the Tech Workforce."** GovTech, www.govtech.com/workforce/black-latino-people-are-being-left-out-of-the-tech-workforce.html. Accessed 10 June 2024.



Findings: As the workforce is constantly changing, the analysis of this dataset allowed us to explore whether global diversity is reflected in tech workforces. Companies at large have seen more of a push for diversity, equity, and inclusion in the workplace. We found that over a 5-year period the male population was about twice the size of the female population. The Latino and Black groups remained consistent at under 10% while the Asian group maintained itself at 20% and saw greater growth. Additionally, we will discuss the limitations to our dataset and future work that can enhance this study.

Thank you

Questions?

Diversity ttest

White vs Asian: TtestResult(statistic=22.426127199064165, pvalue=3.4974324485005776e-53, df=173.78741107942048)
White vs Black: TtestResult(statistic=47.45100949423269, pvalue=6.374334017920352e-81, df=122.98495278778121)
White vs Latino: TtestResult(statistic=46.29135565132039, pvalue=4.13282193180125e-78, df=119.3735241720709)
White vs Multi: TtestResult(statistic=51.99438750290772, pvalue=4.870119400677852e-79, df=109.537654889487)
White vs Other: TtestResult(statistic=55.316345330041564, pvalue=3.910652332452989e-73, df=93.31475419432466)
White vs Undeclared: TtestResult(statistic=55.69860802269435, pvalue=7.557267383459079e-75, df=95.9104106591678)
Asian vs Black: TtestResult(statistic=12.852924987195221, pvalue=7.644253536628321e-24, df=112.92177336927801)
Asian vs Latino: TtestResult(statistic=11.581759018419703, pvalue=8.827328477772503e-21, df=110.3300092458146)
Asian vs Multi: TtestResult(statistic=15.577446894105972, pvalue=6.992841235175281e-29, df=103.40989499464588)
Asian vs Other: TtestResult(statistic=16.909733828894062, pvalue=5.166342130749162e-30, df=92.26101951932218)
Asian vs Undeclared: TtestResult(statistic=17.467632811843643, pvalue=2.8276886801936257e-31, df=94.03313383302206)
Black vs Latino: TtestResult(statistic=-2.9021422691618346, pvalue=0.004171105579260181, df=179.31680824799332)
Black vs Multi: TtestResult(statistic=5.509231046627633, pvalue=1.3317013546634914e-07, df=168.02727152668047)
Black vs Other: TtestResult(statistic=8.904425131520638, pvalue=1.4956780083145585e-14, df=107.30242991242221)
Black vs Undeclared: TtestResult(statistic=10.270102429424778, pvalue=3.7853392676111054e-18, df=120.25779691951077)
Latino vs Multi: TtestResult(statistic=9.017914867341968, pvalue=3.608912970795619e-16, df=172.49074872237733)
Latino vs Other: TtestResult(statistic=13.439520313134684, pvalue=6.477281620974978e-25, df=109.52785001293975)
Latino vs Undeclared: TtestResult(statistic=14.69482678067052, pvalue=6.301597651996616e-29, df=123.96164151053163)
Multi vs Other: TtestResult(statistic=2.92523538481525, pvalue=0.004120717814921395, df=119.35924965926462)
Multi vs Undeclared: TtestResult(statistic=4.846109485972371, pvalue=3.3103904655276187e-06, df=139.43501659992066)
Other vs Undeclared: TtestResult(statistic=3.457792920018384, pvalue=0.0006911859728477218, df=166.76559172686433)