



Gender Disparity in

2022 Aggie Hacks x Google Cloud Hackathon 🔍

Our team member



Ran Zhang

Data Modeling

Google Colab, R



Yi Huang

Data Visualization

Google Studio

Yutian Lei

Data Modeling

GCP Bigquery, R



Andrew Chao

Data Visualization

Google Studio



Table of contents

01

Problem

Gender gap exists in tech industry

02

Evidence Analysis

Gender gap in race, industry, occupations and benefits

03

Model Support

Mediation Regression

04

Recommendations

Improve gender equality in tech

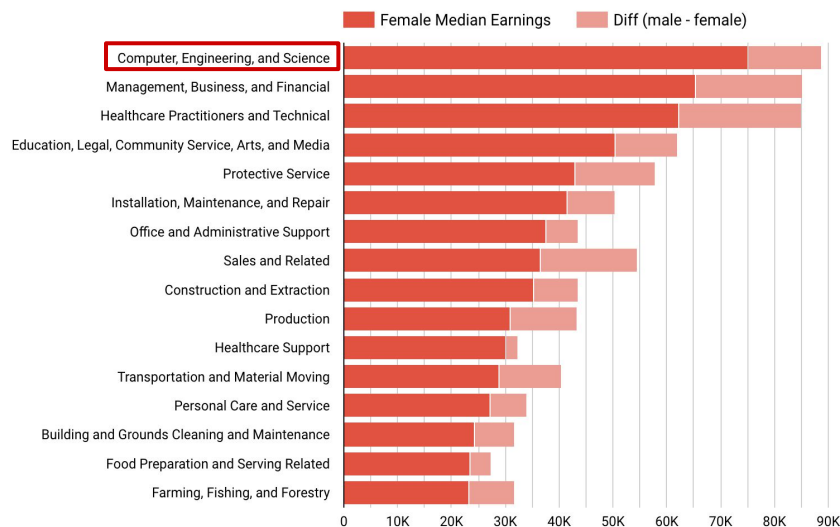
**“Employers cannot discriminate
against employees based on
gender or reproductive choices.”**

——Ruth Bader Ginsburg

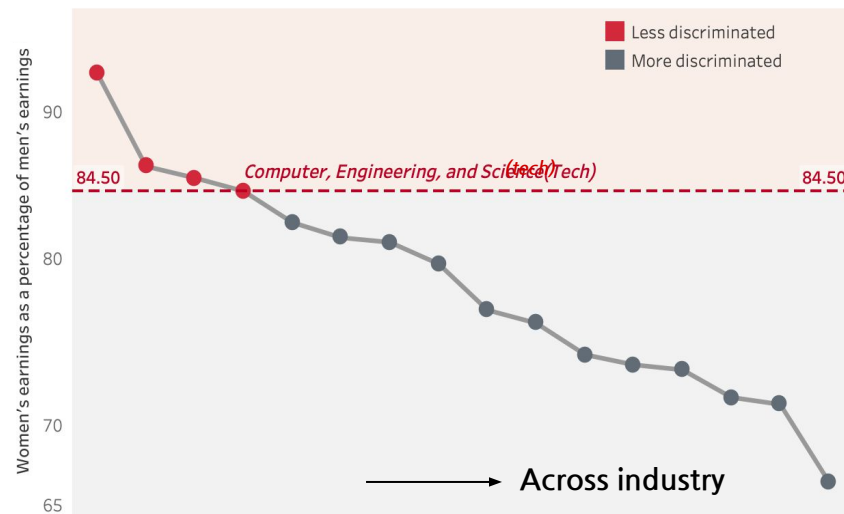


Problem: Women earn less than men in tech industry

Salary across industry



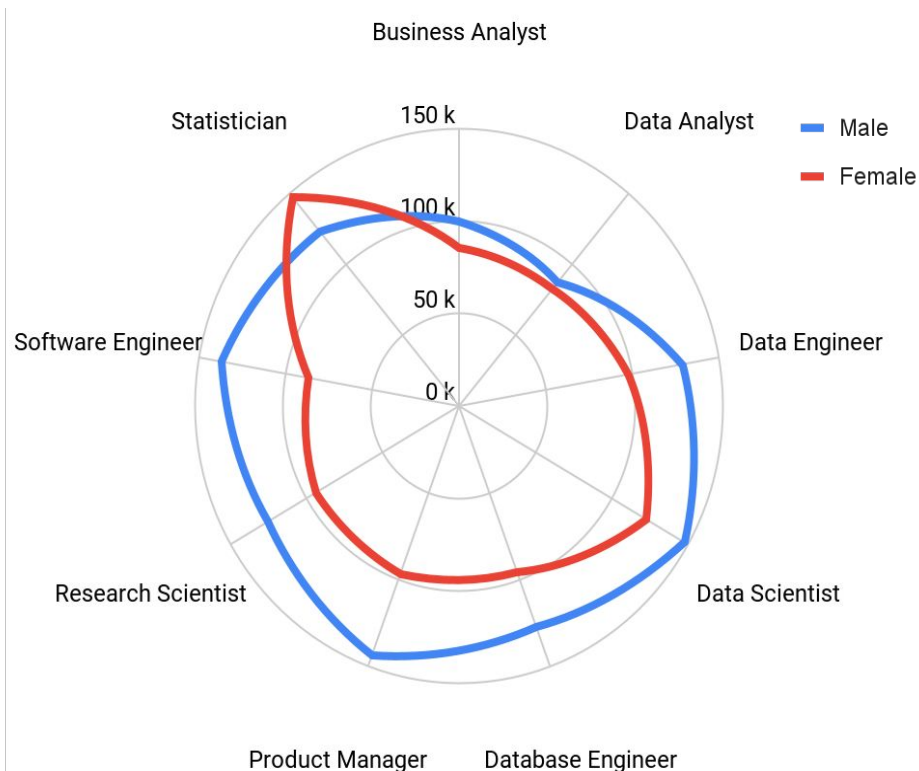
Woman Salary as percentage of Man salary



- Although women in tech industry have the highest median salary vs. other industries, there is still a ~\$13k gap between gender
- Woman in tech industry have earnings 84.5% of man's earnings, rank 4th among all industries

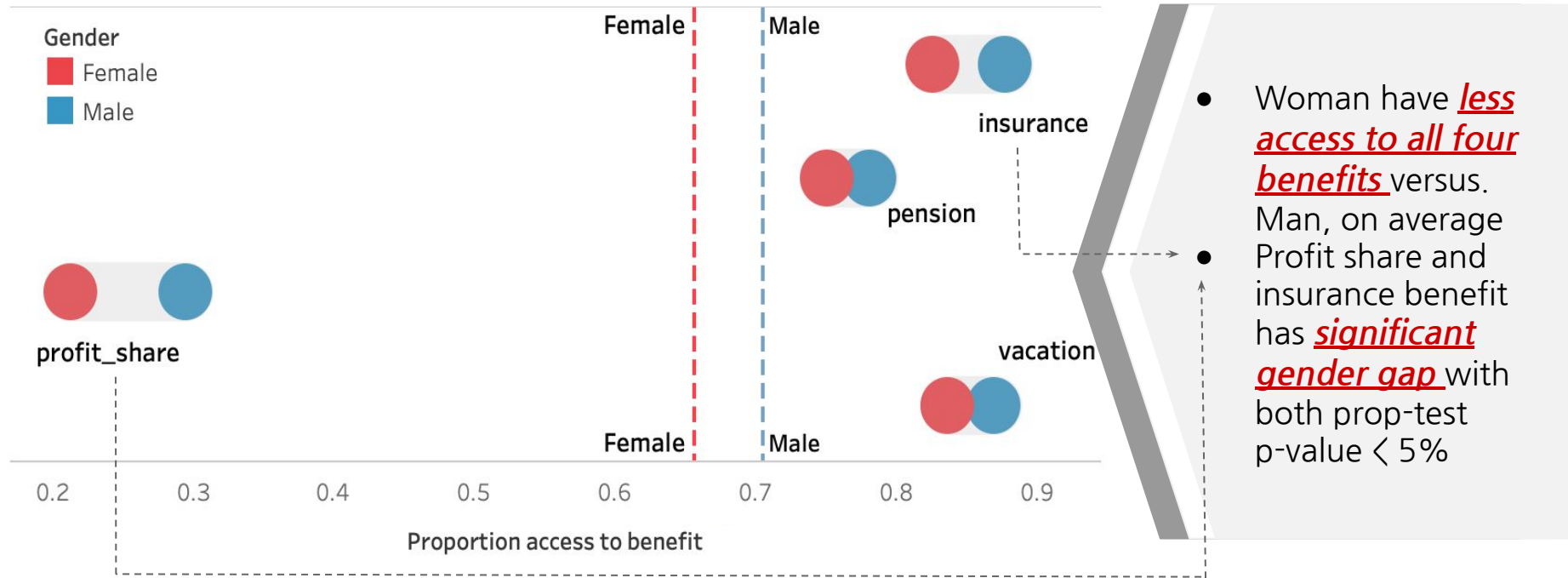
Women earn less in the same position

Average Compensation Comparison

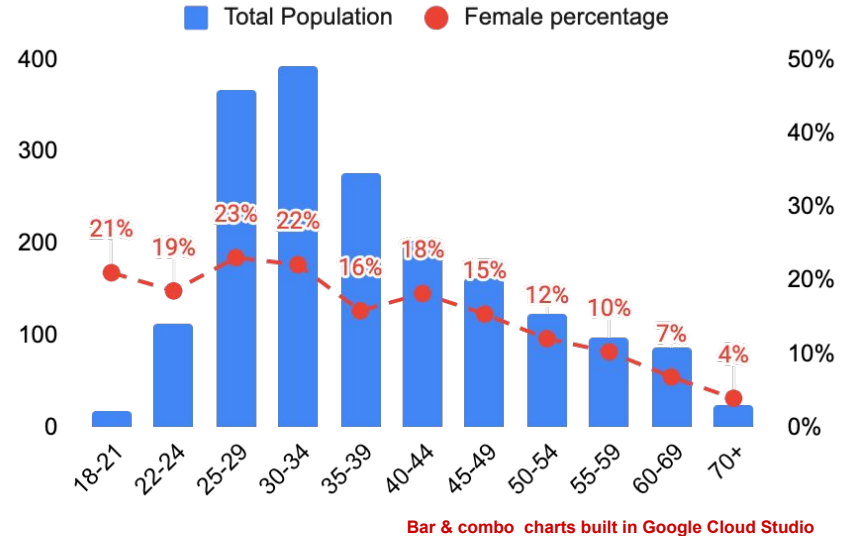
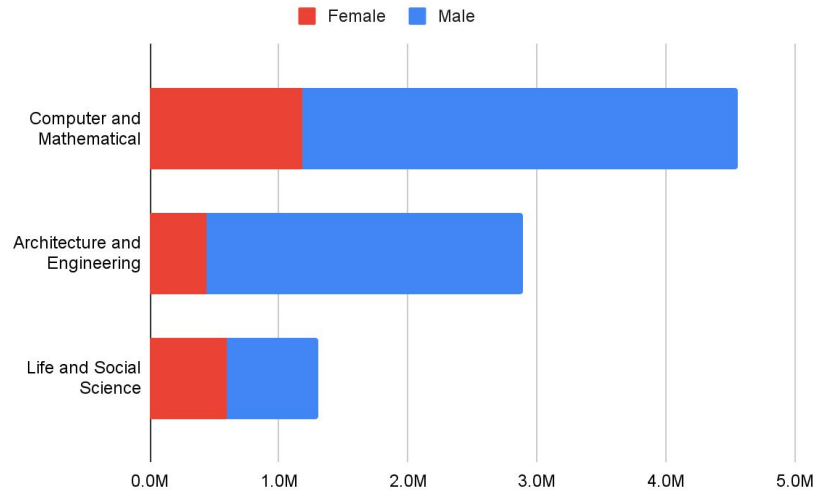


“ In the same tech position, women’s average compensation is **lower** than male. ”

Women has less access to career benefit

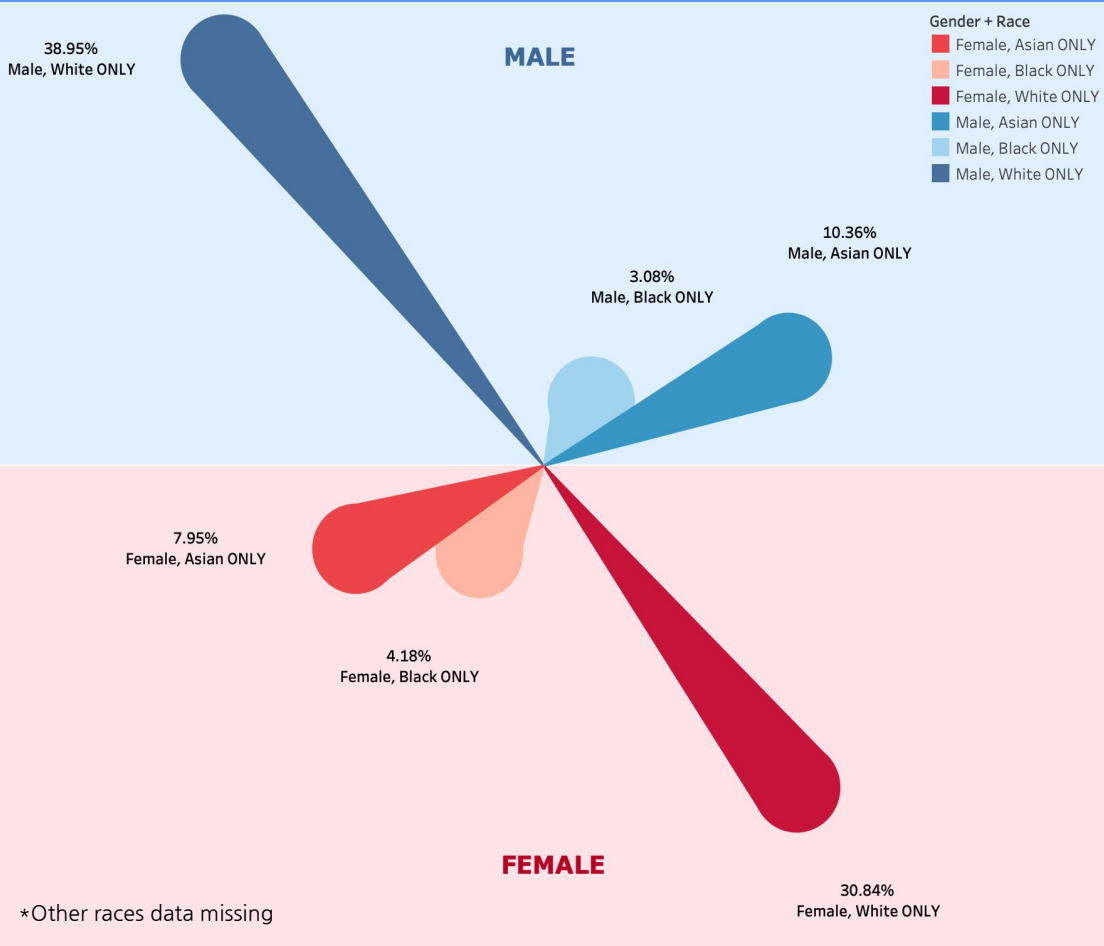


The female population is extremely lower than male population



“ Overall, in the tech industry, the female population is significant **less** than male group. As the age increased, the gender gap become bigger. ”

Gender disparity is obvious in the White and Asian group



In tech industry, female population in **White** and **Asian** group is **less** than male population

Linear Regression: Worse wage discrimination in tech industry

Race, Working experience
Current region, Age,
Annual working hours,
Year, Gender x Year

Control
Variables



Gender



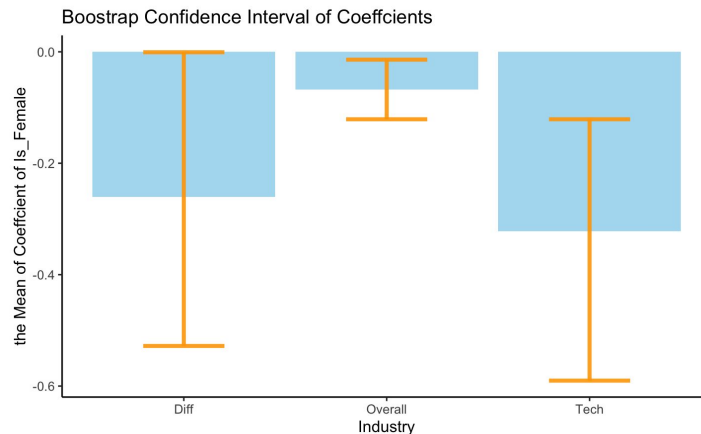
Wage

For Tech Occupations:

$$\ln(\text{Wages}) = 7.16 - 0.32 * Is_Female + \beta * X$$

For Overall Occupations:

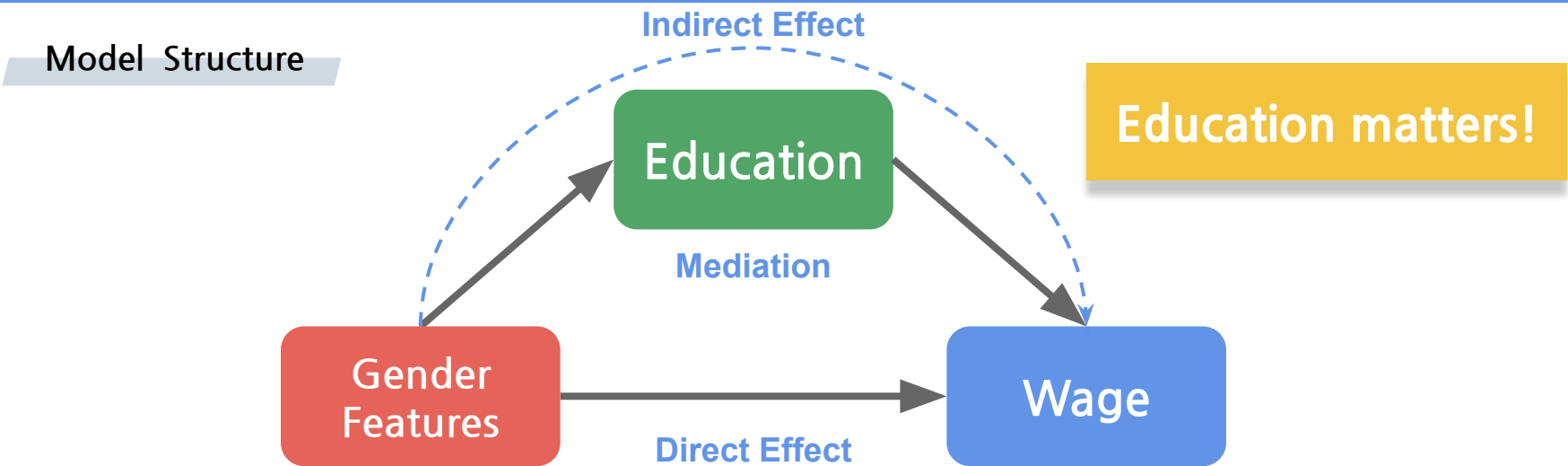
$$\ln(\text{Wages}) = 2.33 - 0.07 * Is_Female + \beta * X$$



- Wage Discrimination exists!
- Are the extents of wage discrimination significantly different?

Worse wage discrimination in tech industry at a 90% confidence interval

Mediation Regression: Can education improve women wage?



Does our research goal make sense?

Is Education effective?

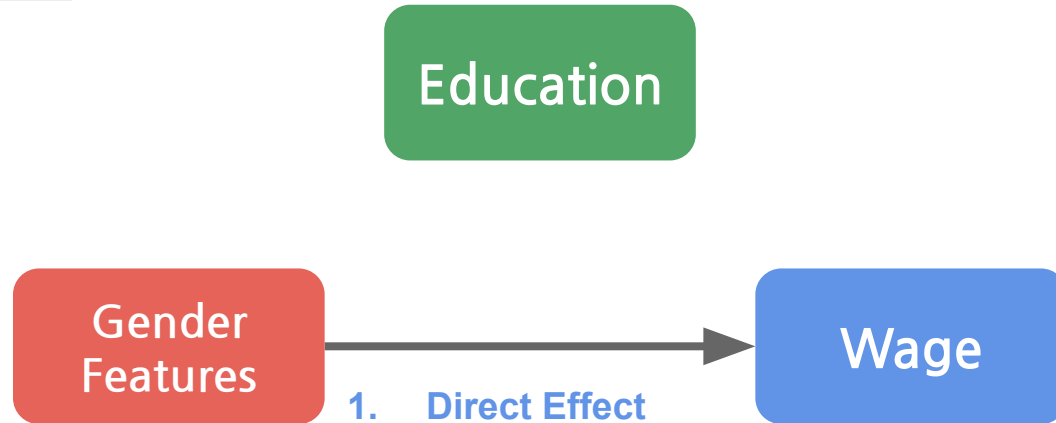
How is the indirect effect of edu?

1	Regress Wage ~ Gender Fea 1980 - 2010 (male) $R^2 = 0.16$, $p < 2.2e-16$ (female) $R^2 = 0.11$, $p < 2.2e-16$ 2017 - 2019 (male) $R^2 = 0.36$, $p < 1e-05$ (female) $R^2 = 0.16$, $p < 2.2e-16$	2	Regress Edu ~ Gender Fea 1980 - 2010 (male) $R^2 = 0.52$, $p < 2.2e-16$ (female) $R^2 = 0.67$, $p < 2.2e-16$ 2017 - 2019 (male) $R^2 = 0.38$, $p < 1.7e-16$ (female) $R^2 = 0.03$, $p < 0.002$	3	Regress Wage ~ Gender Fea + Edu 1980 - 2010 (male) $R^2 = 0.17$, $p < 2.2e-16$ (female) $R^2 = 0.11$, $p < 2.2e-16$ 2017 - 2019 (male) $R^2 = 0.37$, $p < 6.6e-6$ (female) $R^2 = 0.20$, $p < 2.2e-16$
---	--	---	---	---	---

Source: 1980-2010, 2017-2019 The Panel Study of Income Dynamics (PSID) Family-level

Mediation Regression: Can education improve women wage?

Model Structure



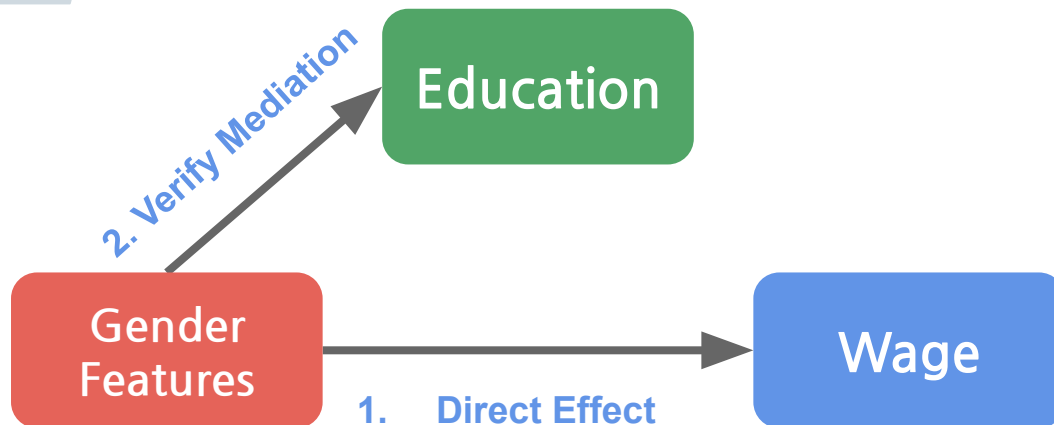
Does our research goal make sense?

1

Regress Wage ~ Gender Fea
1980 - 2010, 2017 - 2019
x
Female, Male
Gender Fea explains Wage

Mediation Regression: Can education improve women wage?

Model Structure



Does our research goal make sense?

1

Regress Wage ~ Gender Fea
1980 - 2010, 2017 - 2019
x
Female, Male
Gender Fea explains Wage

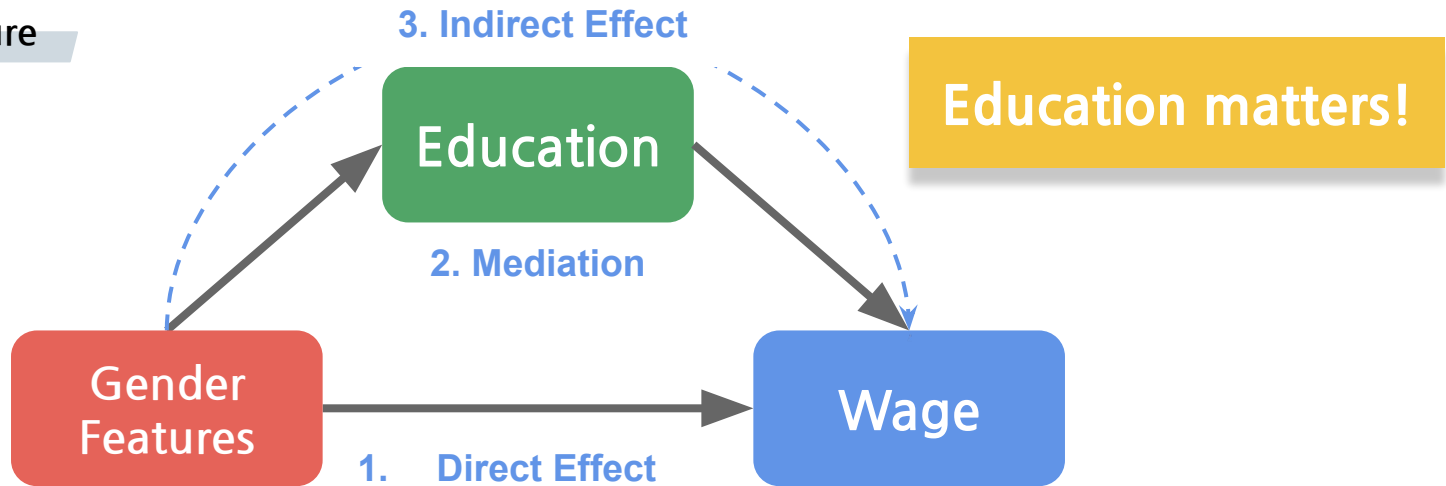
Is Education effective?

2

Regress Edu ~ Gender Fea
1980 - 2010, 2017 - 2019
x
Female, Male
Gender Fea affects Edu

Mediation Regression: Can education improve women wage?

Model Structure



Does our research goal make sense?

1

Regress Wage ~ Gender Fea
1980 - 2010, 2017 - 2019
x
Female, Male
Gender Fea explains Wage

Is Education effective?

2

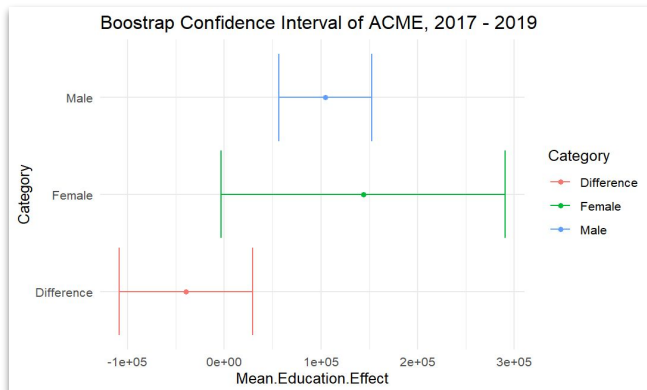
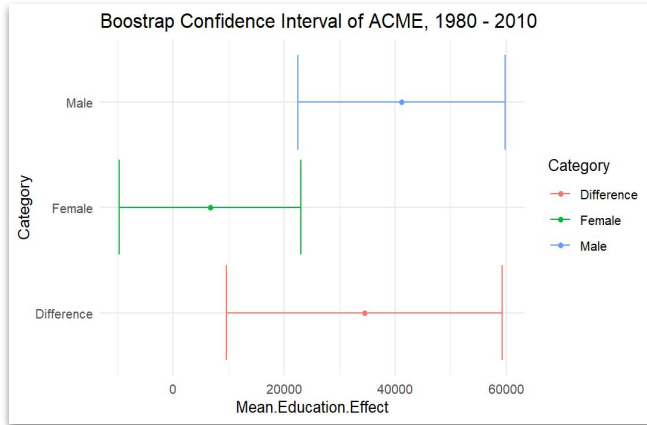
Regress Edu ~ Gender Fea
1980 - 2010, 2017 - 2019
x
Female, Male
Gender Fea affects Edu

How is the indirect effect of edu?

3

Regress Wage ~ Gender Fea + Edu
1980 - 2010, 2017 - 2019
x
Female, Male
Mediation Exists

Education for women becomes effective in recent years



Bootstrap

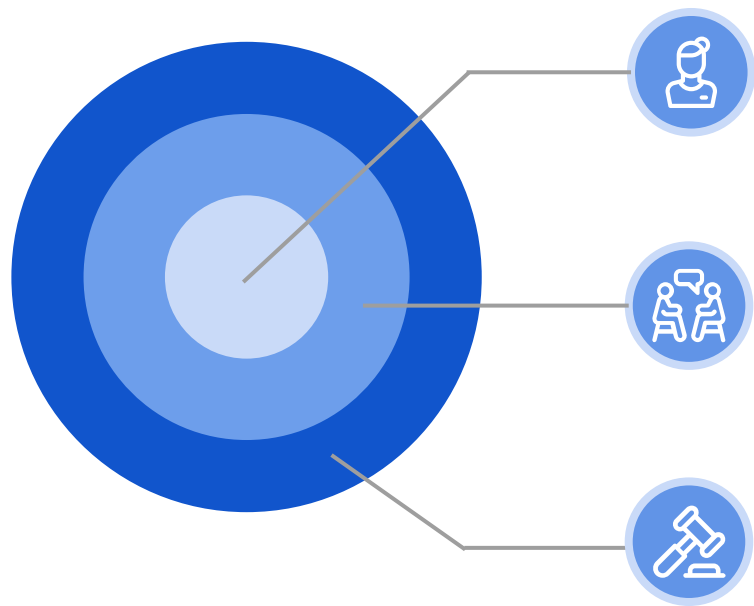
1980 - 2010

- More years in education cannot benefit female wages significantly
- More years in education can benefit male wages significantly

2017 - 2019

- More years in education significant benefit female wages
- Education effect works indifferently for man and women

Recommendations



Self Effort

- Females should be more self-motivated to get tech domain knowledge by obtaining a higher education degree, getting certificates, and attending industry conferences

Industry Effort

- build up a women friendly culture

Society Effort

- The government should invest more on improving gender equality in all race and age groups. For example, try to offer the STEM education at an earlier age to help women gain awareness of science

Q&A



- 1- 前三张：20s
- 2- 气势：20s
- 3- EDA：150s
- 4- 模型：120s
- 5- rec：60s
- 6- 过渡：30s

剩余：80s

Slides

- 1- Eva: 1-5
- 2- Andrew: 6-9
- 3- 然姐: 10-12
- 4- 甜姐：13

Limitations & Next step

1 | LIMITATIONS

- For EDA part, adding more demographic dimensions may give us new findings
- In our regression modeling dataset, we assume individuals graduating from stem-major work in the tech industry, which may bring bias into our model

2 | NEXT STEP

- Explore more EDA insights
- Given more time and data, we can also study how alternative social factors (race, age, location, etc.) would affect gender disparity in tech industries, by applying regression and machine learning models

Appendix & Reference

Datasets

1. 2019 Kaggle Machine Learning & Data Science Survey
https://www.kaggle.com/c/kaggle-survey-2019/data?select=multiple_choice_responses.csv
2. 2013-2019 National Center for Science and Engineering Statistics
<https://ncesdata.nsf.gov/builder/nscg>
3. 2019 United States Census Bureau
<https://www.census.gov/data/tables/time-series/demo/industry-occupation/median-earnings.html>
4. Salary for public sector staffs in SF, 2011- 2018
<https://www.kaggle.com/fedesoriano/gender-pay-gap-dataset>
5. 2017-2019 The Panel Study of Income Dynamics (PSID) Family-level
<https://simba.isr.umich.edu/data/data.aspx>

Appendix & Reference

Call:

```
lm(formula = WAGES ~ ., data = df_1)
```

Residuals:

Min	1Q	Median	3Q	Max
-11.6605	-0.2348	0.1588	0.5808	3.7381

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	7.160498	0.734320	9.751	< 2e-16 ***
AGE	-0.008812	0.005036	-1.750	0.0805 .
YRS.PRES.EMP	0.033051	0.006751	4.896	1.17e-06 ***
WTR.GRADUATED	-0.152699	0.176247	-0.866	0.3865
WORK.WEEKS	0.085145	0.006594	12.912	< 2e-16 ***
COMPLETED.ED	0.022544	0.034319	0.657	0.5114
SEX_2	-0.324637	0.173571	-1.870	0.0618 .
RACE_2	-0.439217	0.213781	-2.055	0.0402 *
RACE_3	-0.468961	0.786772	-0.596	0.5513
RACE_4	0.056037	0.205164	0.273	0.7848
RACE_5	0.819268	1.113097	0.736	0.4619
RACE_7	-0.861712	0.427955	-2.014	0.0444 *
CURRENT.REGION_2	-0.291249	0.177581	-1.640	0.1013
CURRENT.REGION_3	-0.097483	0.166994	-0.584	0.5595
CURRENT.REGION_4	-0.214227	0.177315	-1.208	0.2273
CURRENT.REGION_5	-0.058822	1.122990	-0.052	0.9582
CURRENT.REGION_6	-2.338277	0.556582	-4.201	2.93e-05 ***
YEAR_2019	-0.002080	0.114036	-0.018	0.9855
year_female	0.181155	0.294027	0.616	0.5380

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.564 on 873 degrees of freedom

Multiple R-squared: 0.231, Adjusted R-squared: 0.2151

F-statistic: 14.57 on 18 and 873 DF, p-value: < 2.2e-16

Call:

```
lm(formula = WAGES ~ ., data = df_3)
```

Residuals:

Min	1Q	Median	3Q	Max
-16.2253	-0.7692	-0.0275	0.7824	9.9846

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.328242	0.145085	16.047	< 2e-16 ***
AGE	-0.036052	0.001175	-30.690	< 2e-16 ***
YRS.PRES.EMP	0.060099	0.002060	29.182	< 2e-16 ***
WTR.GRADUATED	-0.103498	0.027208	-3.804	0.000143 ***
WORK.WEEKS	0.164487	0.000988	166.485	< 2e-16 ***
COMPLETED.ED	0.090740	0.006924	13.105	< 2e-16 ***
SEX_2	-0.068868	0.036233	-1.901	0.057356 .
RACE_2	-0.092317	0.051505	-1.792	0.073089 .
RACE_3	-0.033317	0.201443	-0.165	0.868639
RACE_4	-0.100846	0.134831	-0.748	0.454504
RACE_5	0.265567	0.462624	0.574	0.565944
RACE_7	0.185191	0.089404	2.071	0.038338 *
CURRENT.REGION_2	-0.195130	0.057514	-3.393	0.000694 ***
CURRENT.REGION_3	-0.128762	0.054106	-2.380	0.017333 *
CURRENT.REGION_4	-0.066592	0.060824	-1.095	0.273608
CURRENT.REGION_5	-0.652585	0.355919	-1.834	0.066743 .
CURRENT.REGION_6	-0.935387	0.227049	-4.120	3.81e-05 ***
YEAR_2019	0.018504	0.040960	0.452	0.651453
isstem	0.806376	0.081924	9.843	< 2e-16 ***
stem_female	-0.308907	0.201109	-1.536	0.124553
year_female	0.080308	0.067124	1.196	0.231555

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.116 on 16991 degrees of freedom

Multiple R-squared: 0.7861, Adjusted R-squared: 0.7859

F-statistic: 3123 on 20 and 16991 DF, p-value: < 2.2e-16

Appendix & Reference

```
> pension <- prop.test(x=c(25841, 34115), n=c(34429, 43688))  
> pension
```

2-sample test for equality of proportions with continuity correction

```
data:  c(25841, 34115) out of c(34429, 43688)  
X-squared = 99.025, df = 1, p-value < 2.2e-16  
alternative hypothesis: two.sided  
95 percent confidence interval:  
 -0.03633945 -0.02429840  
sample estimates:  
   prop 1    prop 2  
0.7505591 0.7808780
```

```
> insurance <- prop.test(x=c(28445, 38314), n=c(34429, 43688))  
> insurance
```

2-sample test for equality of proportions with continuity correction

```
data:  c(28445, 38314) out of c(34429, 43688)  
X-squared = 399.46, df = 1, p-value < 2.2e-16  
alternative hypothesis: two.sided  
95 percent confidence interval:  
 -0.05587484 -0.04572187  
sample estimates:  
   prop 1    prop 2  
0.8261930 0.8769914
```

Appendix & Reference

Mediation Regression, 1980-2010

[Code ▾](#)

Data Processing

[Hide](#)

```
old <- read.csv("psid_old_stem.csv")
psid_old <- subset(old, select = c(sex, famwgt, age, sch, white, south, LEHS, black, hisp, othrace, west, northeast, northce
ntral, annhrs, realhrwage))
psid_old <- psid_old %>% mutate(annincome = realhrwage * annhrs) # calculate annual salary/wage
psid_old <- subset(psid_old, select = -c(realhrwage, annhrs, othrace))
psid_old$sex <- 1 - psid_old$sex
```

Are X variables influencing wage?

Yes, sex(1 = Male), lower level of education(LEHS) reduces salary, while as age grows, the salary increases. These are highly significant, whereas Black people receiving less salary.

[Hide](#)

```
model.0.old.m <- lm(annincome~ . - sch, data = psid_old_male)
summary(model.0.old.m)
```

```
Call:
lm(formula = annincome ~ . - sch, data = psid_old_male)
```

```
Residuals:
```

Women in tech problems:

- **Education disparity → Tech employment disparity**

Bachelor: computer sciences (18%), engineering (20%), physical sciences (39%) and mathematics (43%)

High school: high school senior boys, 26% planned to enter STEM or biomed occupations, compared with 13% of girls

- **Retention gap → Education disparity**

Women in STEM careers are more likely to leave within the first years

- **Salary gap**

10% less than man (which we created the chart in the page 7)

- **Leadership position gap**

Women account for only 16% of senior level tech jobs and 10% of executive positions, according to the "Quantifying the Gender Gap" study by Entelo.

- **Solutions:**

- Self-development

- Keep skills relevant and up to date (taking certification and university classes, and attending industry conferences and other events)

- Tech companies

- businesses need to embed diversity and inclusion into company culture from day one
- Eliminate bias in the hiring process
- Close the pay gap

- Government

- Provide STEM education at an early age

Storyline

1. Women in tech are discriminated! And we can see from
 - a. Population disparity (including age distribution) P6
 - b. Race disparity P7
 - c. Salary disparity (from the compensation and number of women who can earn the compensation) P8
 - d. Benefit disparity P9
2. To further prove if women are discriminated in tech, we perform a wage linear regression to see if gender is significant
 - a. Proved by P11
3. So, now we proved women are discriminated in tech industry, what we can do to help women in tech?
4. We are thinking whether education can help women to improve their wage in tech industry. So, we perform the mediation regression to see if women's wage can be improved by **education**.
 - a. Proved by P12
 - b. Interesting insights -> we found the wage gap has been mitigated in the past few years P13
5. Recommendation

As we saw the discrimination exists in different kind of aspects in tech industry, here is our suggestions:
From self-improvement to industry improvement and then government can help in the last

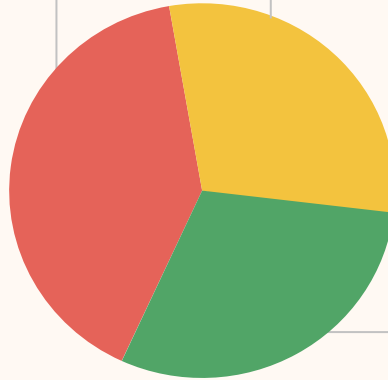
 - a. Self-development
 - i. Keep skills relevant and up to date (taking certification and university classes, and attending industry conferences and other events)
 - b. Tech companies
 - i. businesses need to embed diversity and inclusion into company culture from day one
 - ii. Eliminate bias in the hiring process
 - iii. Close the pay gap
 - c. Government
 - i. Provide STEM education at an early age

Recommendations

Industry effort

2

- Organizations like woman in tech summit should continue work on improving female status in tech industry
- Tech companies should advocate on gender equality, starting from hiring more female employees and increase female staff's access to benefits



1

Self effort

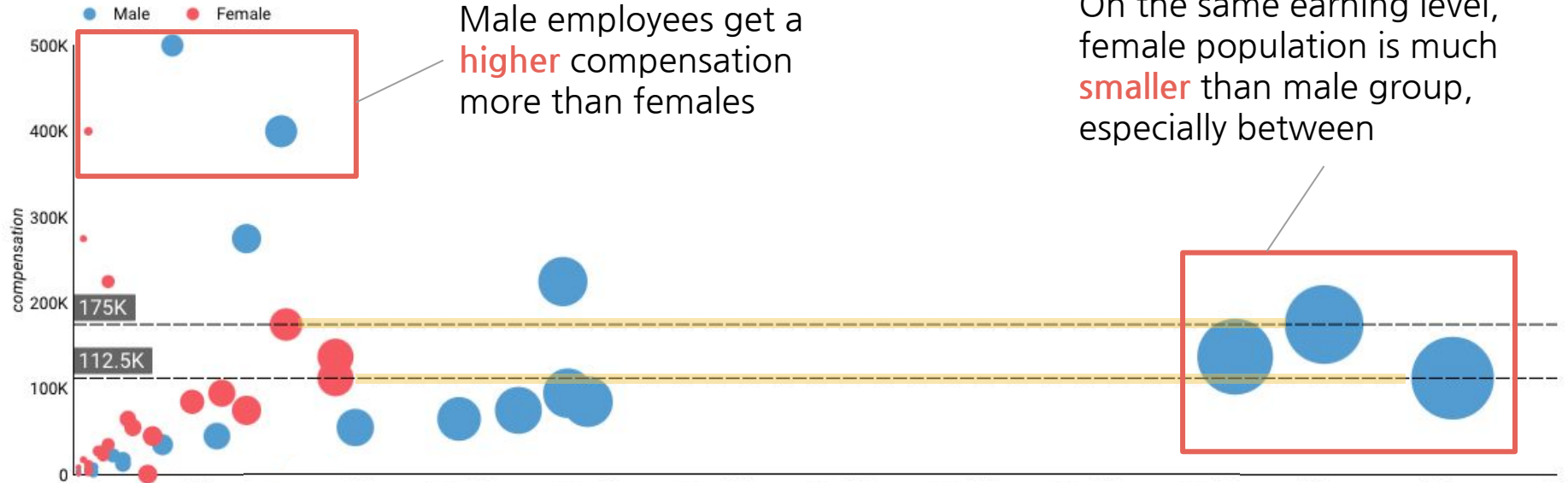
- Females should try hard to obtain a higher education degree, in order to earn an equal opportunity in tech career

3

Society effort

- The government should invest more on improving gender equality in all race and age groups

Fewer women can make the same level earning as men



Women earn less than men in tech industry

Salary across industry

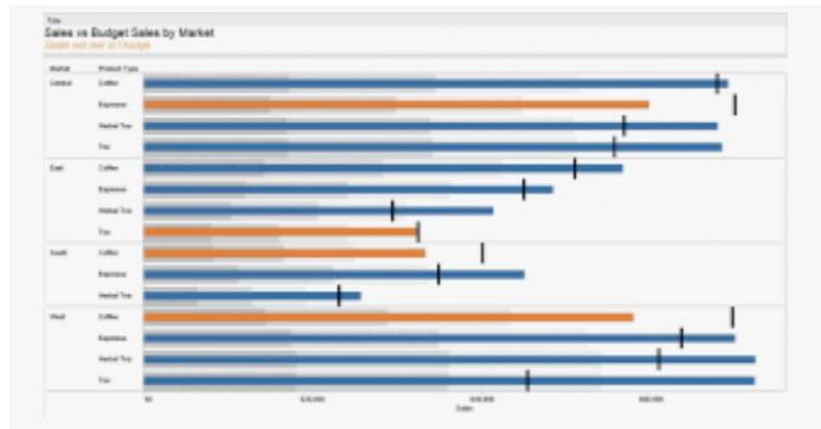
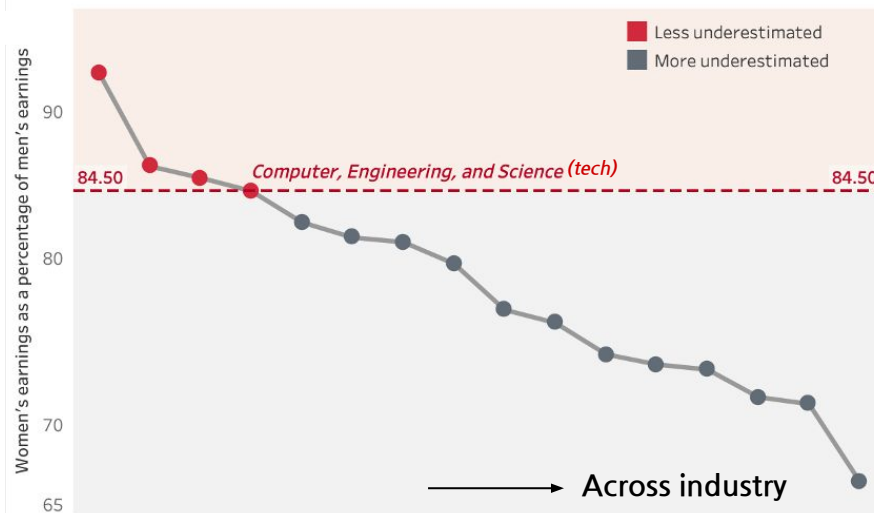


Tableau Essentials: Chart Types ...

Male Median Wage

Female Median Wage

Woman Salary as percentage of Man salary



- Although women in tech industry have the highest median salary vs. other industries, there is still a huge gap between gender.
- Woman in tech industry have earnings 84.5% of man's earnings, rank 4th among all industries.