Gender Wage Gap Over Time

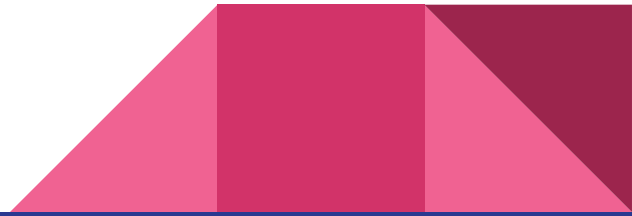
By Brian Mann

The Question

- Is there a wage gap between men and women?
  - How has it changed over time?
  - Are there any differences by education level?
  - What are the trends for the wages of men, women and overall?

The Hypothesis

- Our hypothesis is that there has been a statistically significant decline in the gap between the wages of men and women over time
- The null hypothesis is that there is no such significant difference in the wage gap between men and women over time



The Data

- The data set used to conduct this research project comes from the Economic Policy Institute's State of Working America Data Library from 2022
- This data was taken from the average hourly wage of workers separated by five different education levels, gender and race for the US population
- The hourly wages have already been converted to 2022 USD values using the CPI-U-RS index
- In this project, we will only be focusing on the data for men, women and the overall population
- Data will be analyzed using a Python Jupyter Notebook

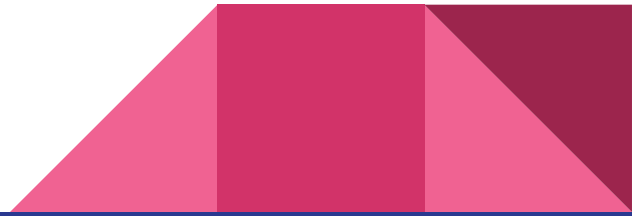
Snapshot of the Data Set

	year	less_than_hs	high_school	some_college	bachelors_degree	advanced_degree	men_less_than_hs	men_high_school	men_some_college	men_bachelor
0	1973	18.06	22.22	24.08	32.80	38.16	21.18	26.90	27.67	
1	1974	17.68	21.60	23.32	31.69	38.37	20.63	26.15	26.79	
2	1975	17.30	21.55	23.30	31.45	38.41	20.00	26.02	26.93	
3	1976	17.52	21.76	23.49	31.46	37.50	20.36	26.14	27.10	
4	1977	17.59	21.50	22.97	31.07	37.36	20.43	25.97	26.70	

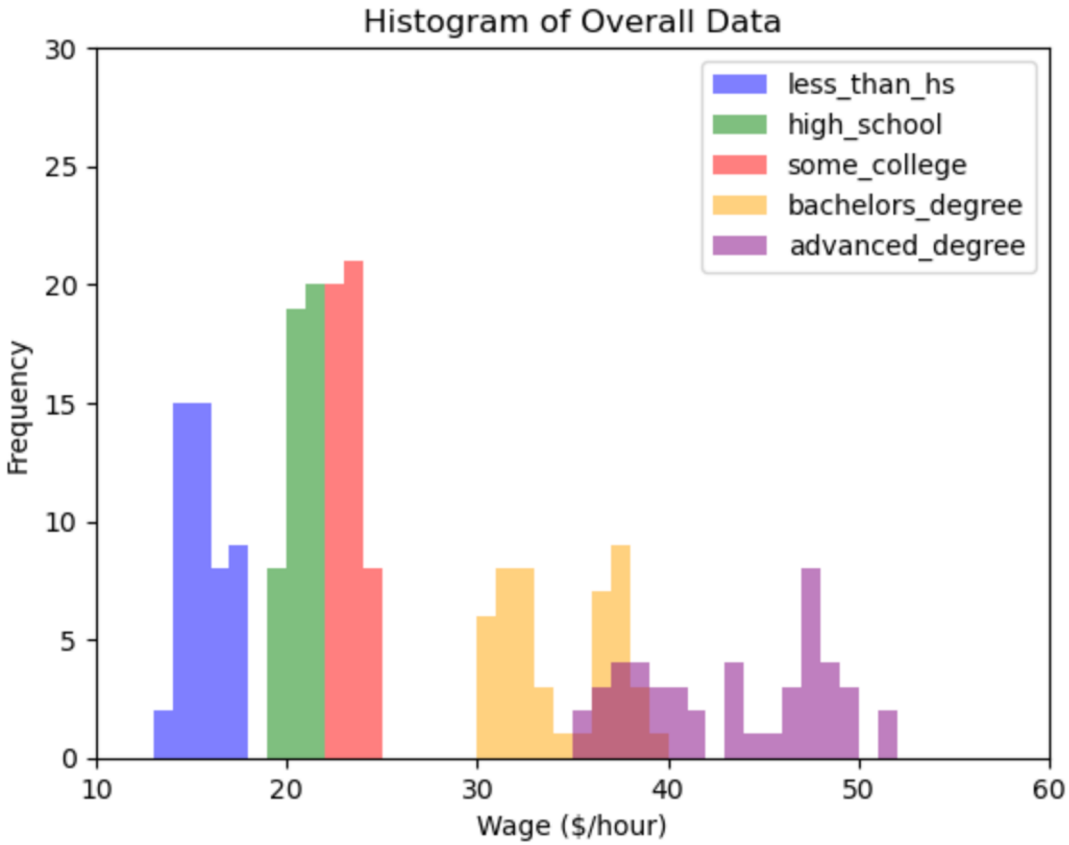
5 rows × 61 columns

The Variables

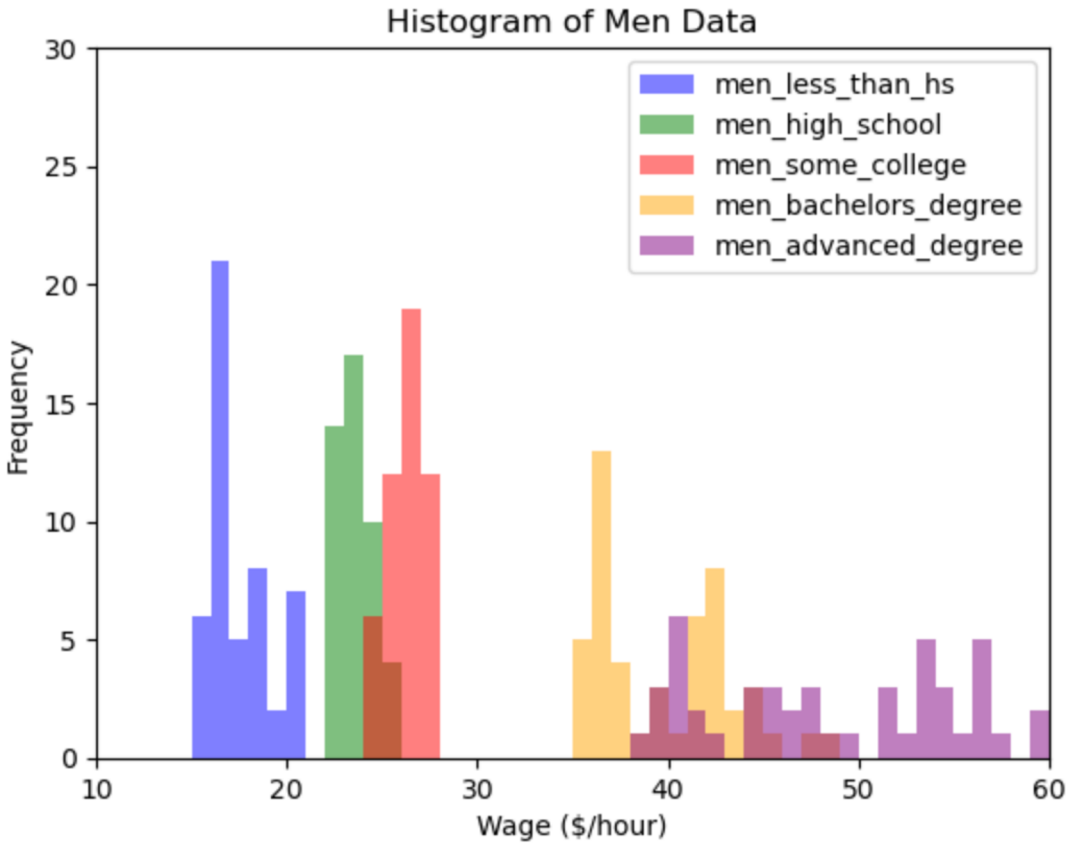
- In this analysis, we will be focusing on five variables, each representing a different level of education attained:
  - Less_then_hs: no high school diploma
  - High_school: a high school diploma
  - Some_college: at least some college education
  - Bachelors_degree: a bachelor's degree
  - Advanced_degree: a master's degree or other advanced degree
- For each of these variables, we will look at three different groupings: men, women and the overall population
- Each entry in the data set represents the average value in 2022 USD for the hourly wage of workers with the given level of education



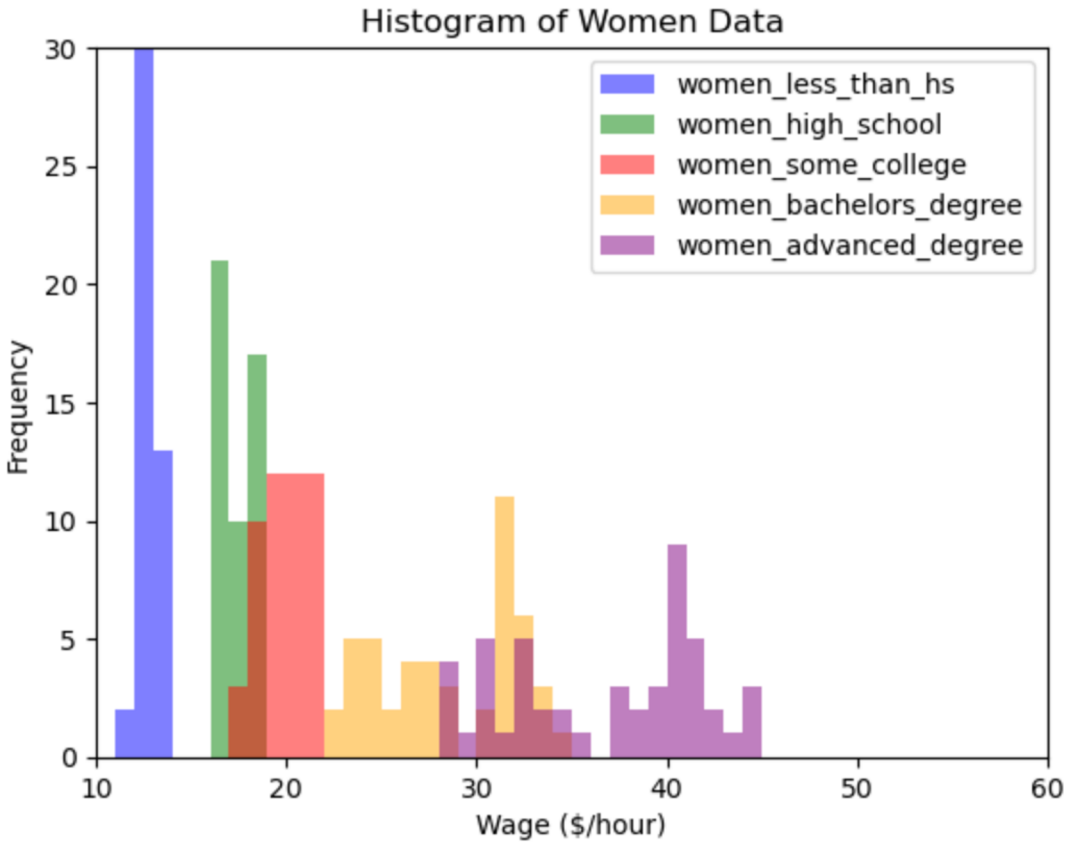
Histograms


Histogram of Overall Data
less_than_hs
high_school
some_college
bachelors_degree
advanced_degree
Frequency
Wage ($/hour)
0
5
10
15
20
25
30
10
20
30
40
50
60

Histograms


Histogram of Men Data
men_less_than_hs
men_high_school
men_some_college
men_bachelors_degree
men_advanced_degree
Frequency
Wage ($/hour)
0
5
10
15
20
25
30
10
20
30
40
50
60

Histograms


Histogram of Women Data
women_less_than_hs
women_high_school
women_some_college
women_bachelors_degree
women_advanced_degree
Frequency
Wage ($/hour)
0
5
10
15
20
25
30
10
20
30
40
50
60

Histogram Analysis

- Generally, we can see that as education increases, wages tend to increase.
  - This is the same across genders, as well as overall
- We can also see that for each level of education, the values for the male wages histogram tend to be higher than the values for the female wages histogram.
- There are essentially no outliers, as the data comes from averages over time.
  - Average wages tend to be very stable over time, especially with larger sample sizes

Descriptive Characteristics

	Mean	Median	Range	Skew
less_than_hs	15.7026	15.340	4.11	0.520736
high_school	20.8766	20.855	3.08	0.134490
some_college	23.2192	23.185	3.40	0.646484
bachelors_degree	34.7686	34.205	11.61	0.319095
advanced_degree	43.8990	44.085	18.42	0.018634

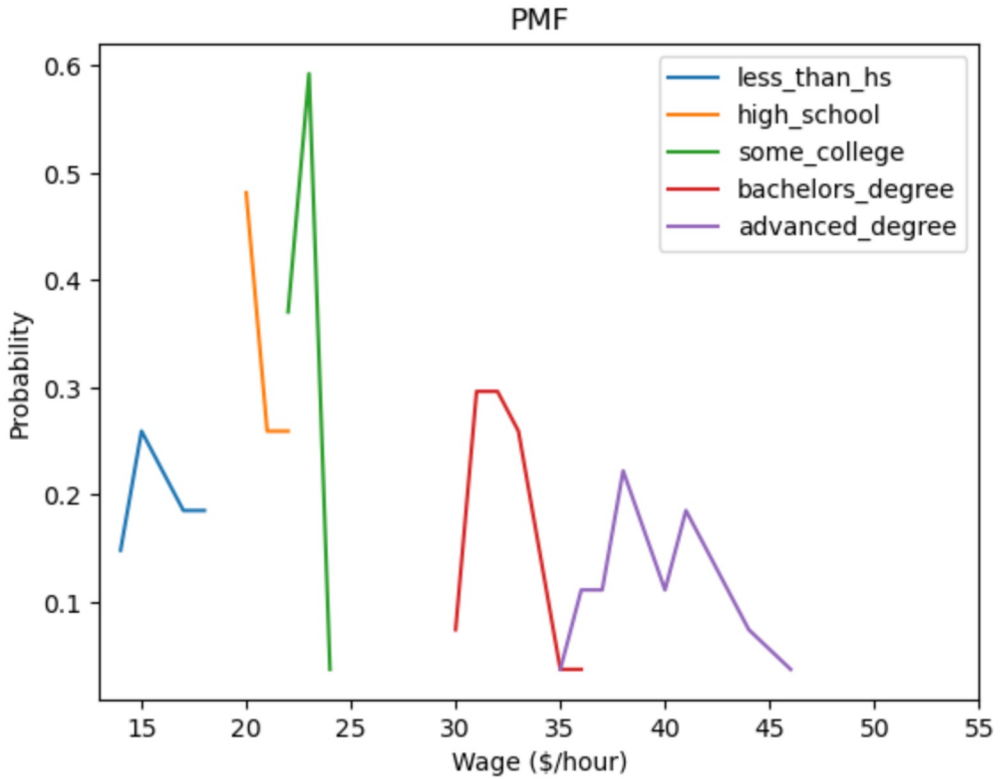
	Mean	Median	Range	Skew
men_less_than_hs	17.5652	16.905	5.79	0.674019
men_high_school	23.8326	23.695	4.79	0.697004
men_some_college	26.3338	26.365	3.77	0.004734
men_bachelors_degree	39.9884	39.485	13.85	0.564821
men_advanced_degree	49.4302	48.940	24.80	0.164481

	Mean	Median	Range	Skew
women_less_than_hs	12.8514	12.810	2.65	0.932343
women_high_school	17.5716	17.370	2.91	0.348900
women_some_college	19.9432	19.985	4.44	-0.121558
women_bachelors_degree	28.9264	29.425	12.47	-0.145332
women_advanced_degree	36.9752	38.345	17.57	-0.206381

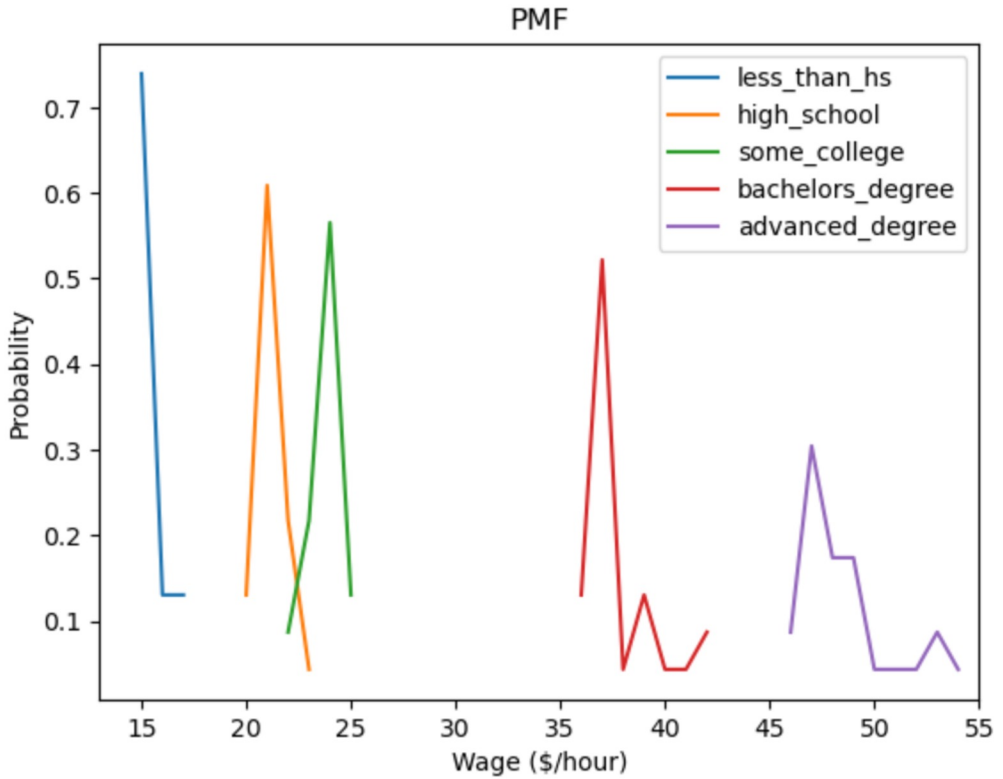
Descriptive Characteristics Analysis

- We first note that just as we suspected, for every level of education, the mean and median values are higher for men than women, indicating a clear wage gap
- Interestingly, as the level of education increases, the range of values between the highest and lowest average wage also seems to increase
- Much of the data is somewhat positively skewed, indicating that there is a higher proportion of lower average wages. This makes the mean somewhat higher than the median
- Strangely, only in the female data do we see a negative skew. Each of these occurs in higher levels of education

Scenario 1: PMF of Wages Before Pre-2000



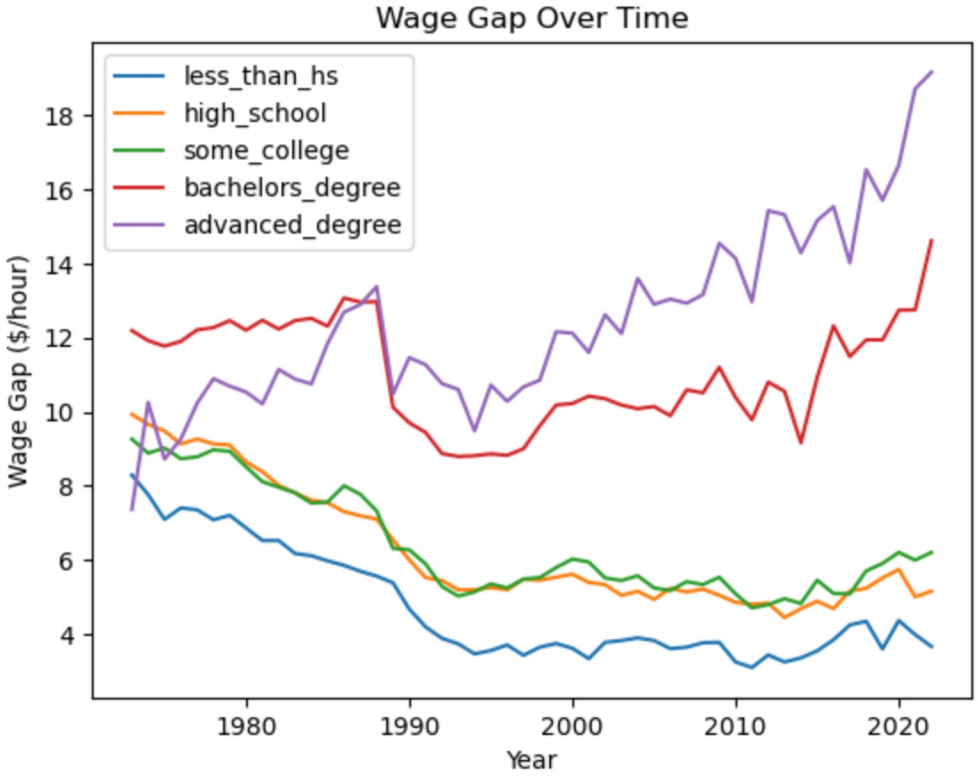
Scenario 2: PMF of Wages Before Post-2000



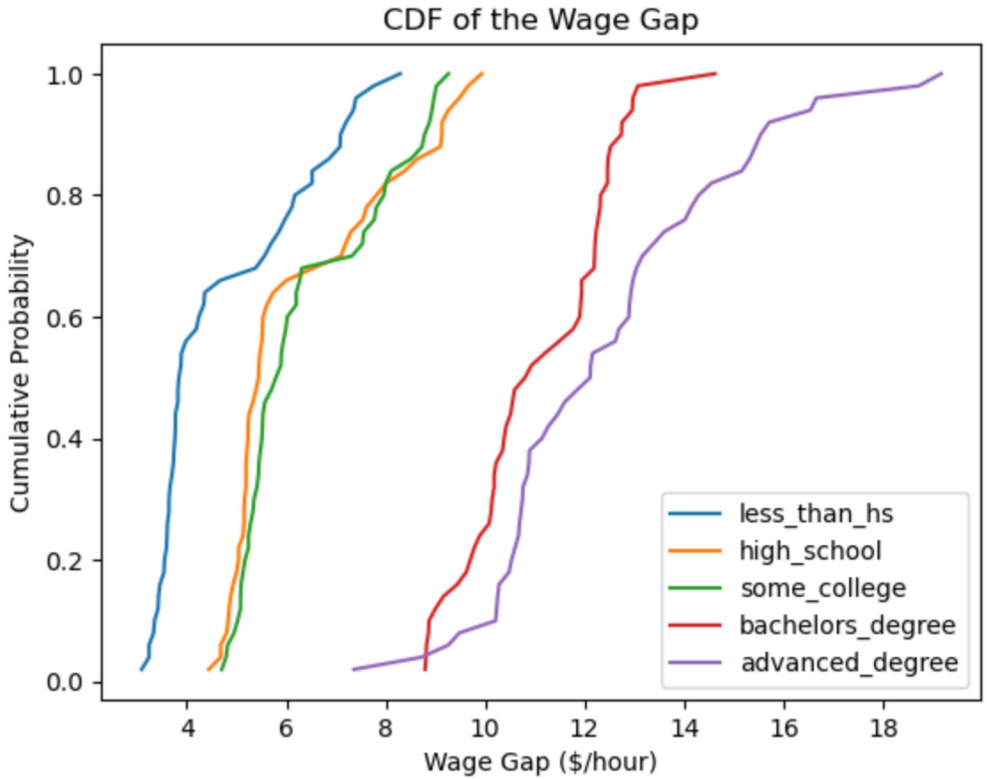
PMF Scenario Analysis

- In both the pre-2000 and post-2000 scenario, we can see that for lower education levels, there does not seem to be a significant difference
  - The chances of getting an hourly wage between $15-25 do not seem to fluctuate much
- However, for bachelor's and advanced degree holders, there does seem to be a significant increase in hourly wage from pre-2000 to post-2000

The Wage Gap Over Time



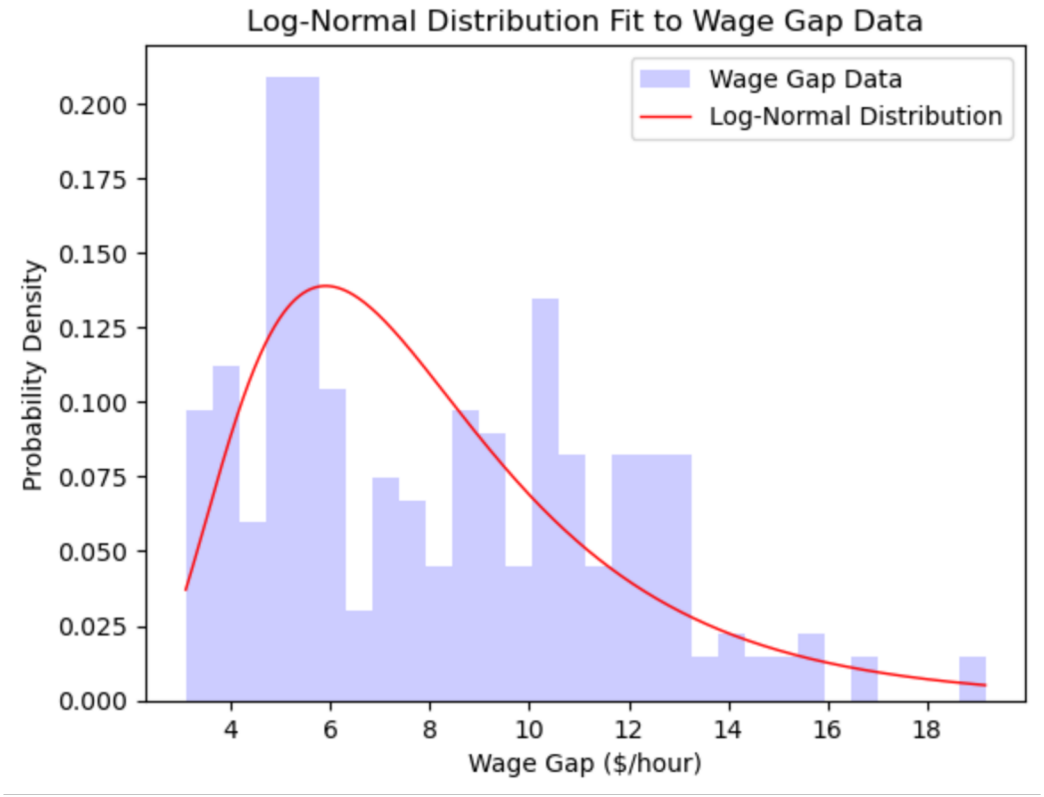
CDF of the Wage Gap



CDF Analysis

- Here, we created a new set of data that represents the difference between the average hourly wages of men and women for each level of education
- As is seen, the gap is positive (men making more than women) for every single data point
- Generally, we can see that the average wage gap between men and women tends to increase with higher levels of education
- For lower levels of education the gap is typically only $4 - $6
- The gap increases to around $10 for bachelor’s and above

Analytical Distribution

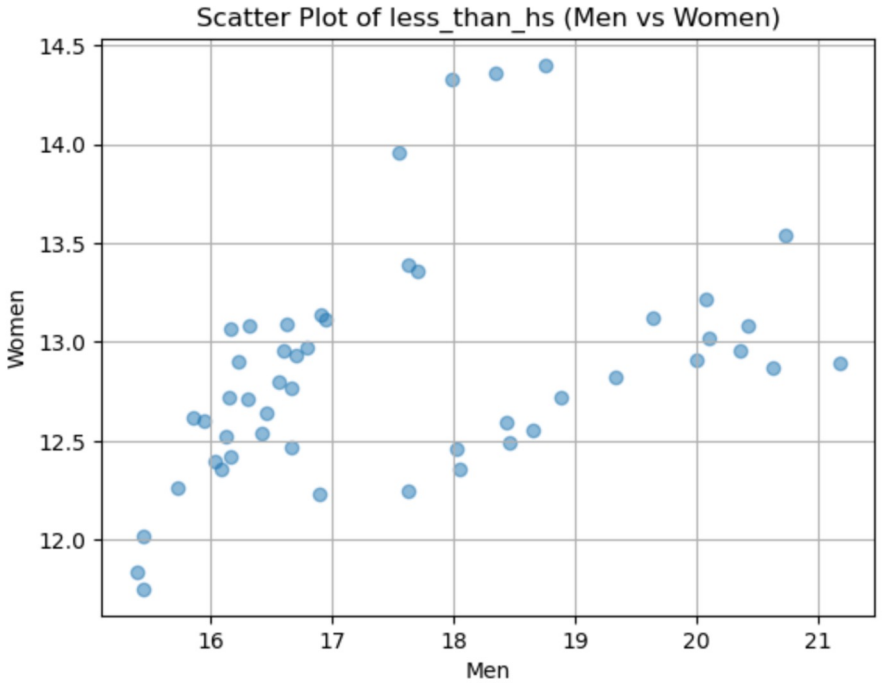


Kolmogorov-Smirnov test statistic: 0.10051534067012036
P-value: 0.011883550359003724

Analytical Distribution Analysis

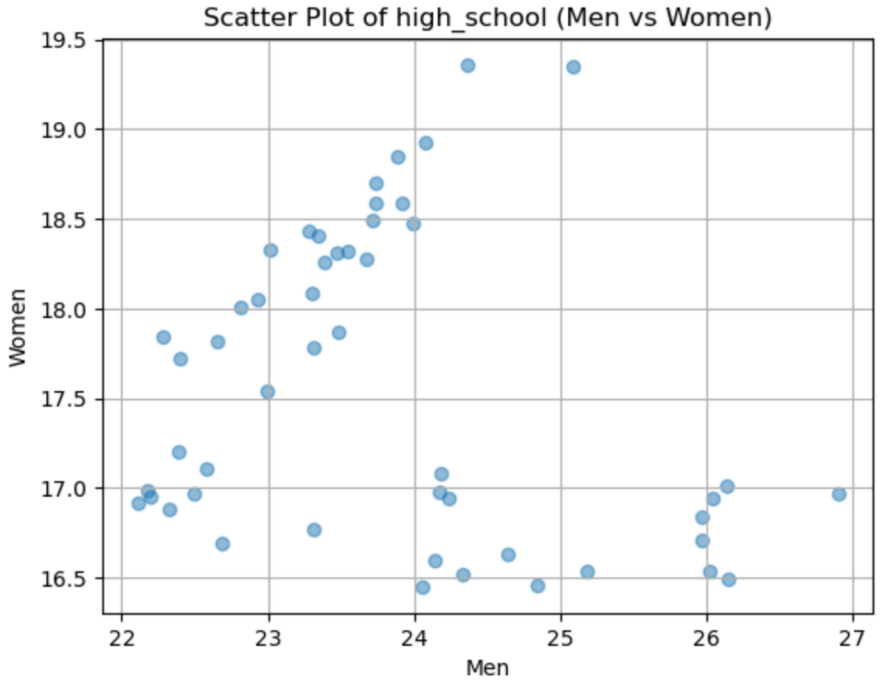
- Here, we have aggregated all the wage gap data into a single array, regardless of education level
- A log-normal distribution was chosen to model the given data, as it is usually a better choice when modeling positively skewed data
  - Income and wages are often positively skewed, as we have already seen in the descriptive statistics analysis
- Unfortunately, after testing the distribution and the sample using the Kolmogorov-Smirnov test, we arrive at the conclusion that there is a significant difference between the two
  - The test statistic is quite low (only 0.1 out of 1)
  - The p-value is also below 0.05, indicating that there is a significant difference between the distribution and the sample data
  - Visually, the two look mostly different

Scatter Plots



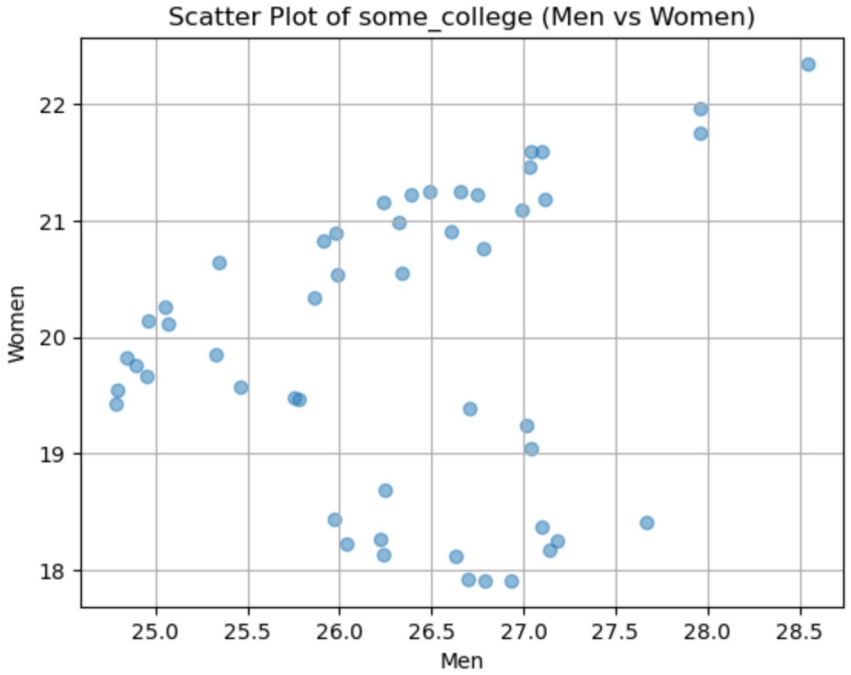
Pearson's correlation coefficient for less_than_hs: 0.40936709147072853
Covariance for less_than_hs: 0.3794027755102042

Scatter Plots



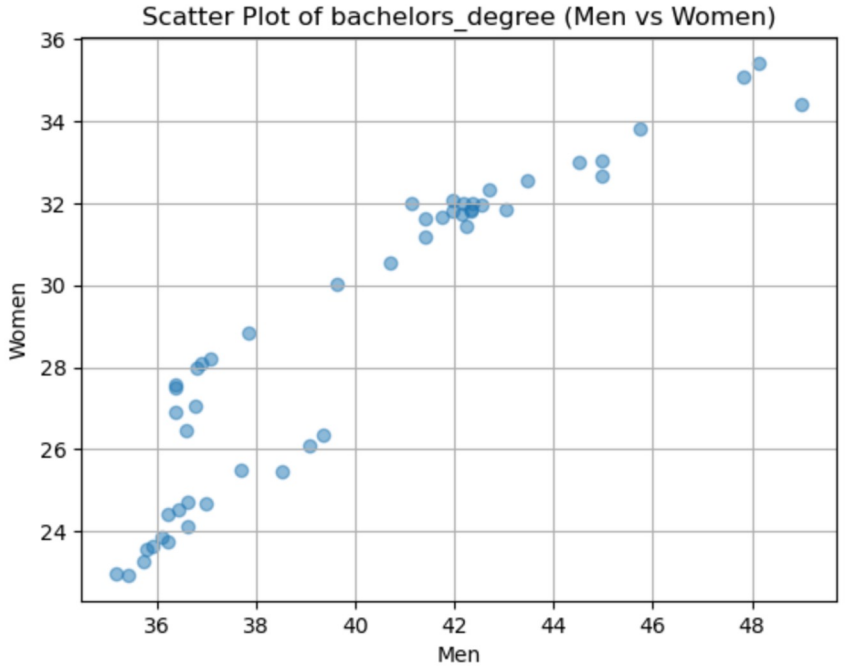
Pearson's correlation coefficient for high_school: -0.22108317658680293
Covariance for high_school: -0.23151036734693842

Scatter Plots



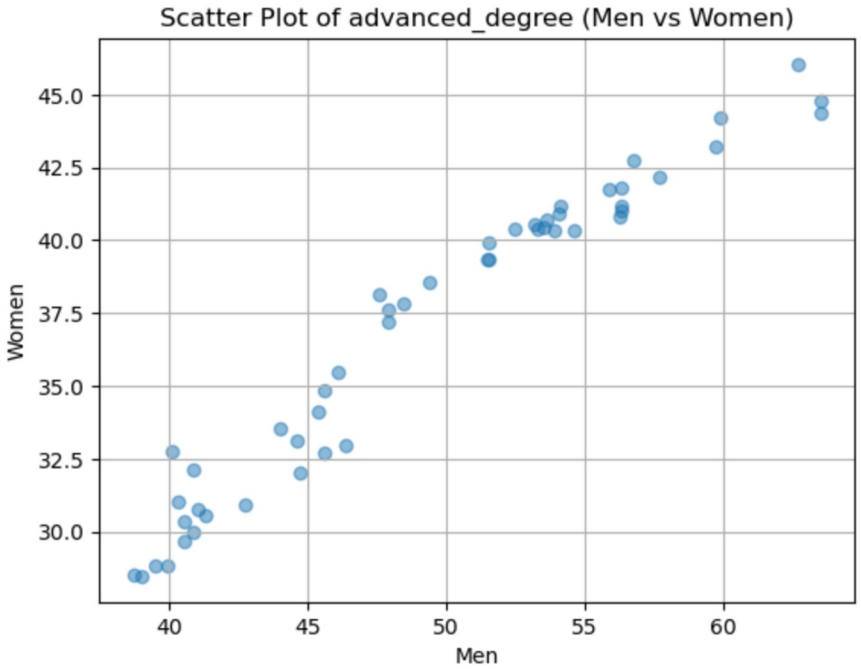

```
Pearson's correlation coefficient for some_college: 0.17219247888114797
Covariance for some_college: 0.19758555102040834
```

Scatter Plots




```
Pearson's correlation coefficient for bachelors_degree: 0.9268341535241581
Covariance for bachelors_degree: 13.017736979591835
```

Scatter Plots




```
Pearson's correlation coefficient for advanced_degree: 0.9740439154323426
Covariance for advanced_degree: 36.14953771428571
```

Scatter Plot Analysis

- Different scatter plots were generated based off of the hourly wages of men and women for each level of education
  - The x-axis represents the male wage, while the y-axis represents the female wage, with each point being the given values for one particular year
- Generally, there was not much correlation or covariance between the values of male and female wages for lower education levels
  - Those with less than a HS education or some college had a slightly positive correlation, while those with a HS diploma even had a slightly negative correlation
- However, for bachelor's and higher, the correlation and covariance were considerably high (corr > 0.9)
  - When higher-educated men had higher salaries, so did women

Hypothesis Testing Using Kendall's Tau-B

```
*** Hypothesis Test for less_than_hs ***
---
Kendall's tau-b: -0.5629086845967293
P-value: 8.224558841832918e-09
---
*** Hypothesis Test for high_school ***
---
Kendall's tau-b: -0.7247458917172676
P-value: 1.2388367167678222e-13
---
*** Hypothesis Test for some_college ***
---
Kendall's tau-b: -0.5482027973358568
P-value: 1.9860464050071154e-08
---
*** Hypothesis Test for bachelors_degree ***
---
Kendall's tau-b: 0.017966518443500928
P-value: 0.8539873239818369
---
*** Hypothesis Test for advanced_degree ***
---
Kendall's tau-b: 0.7142857142857143
P-value: 2.493596474326011e-13
---
```

Hypothesis Testing Analysis

- Here we tested the original hypothesis: There has been a significant decline in the wage gap over time
- We used Kendall's Tau-B to assess significance
  - Kendall's Tau-B is non-parametric (doesn't have to follow a normal distribution)
  - We compare Time with the Wage Gap to assess if there has been a significant decline
- For the three lowest education levels, we can infer that there has been a significant decline in the wage gap
  - For each, we have Tau-B < -0.5 and $p < 0.001$
- For bachelor's degree holders, there seems to be no significant change
  - Tau-B ~0, $p > 0.05$
- For advanced degree holders, there is actually a significant increase in the wage gap over time
  - Tau-B > 0.7, $p < 0.001$

Linear Regression

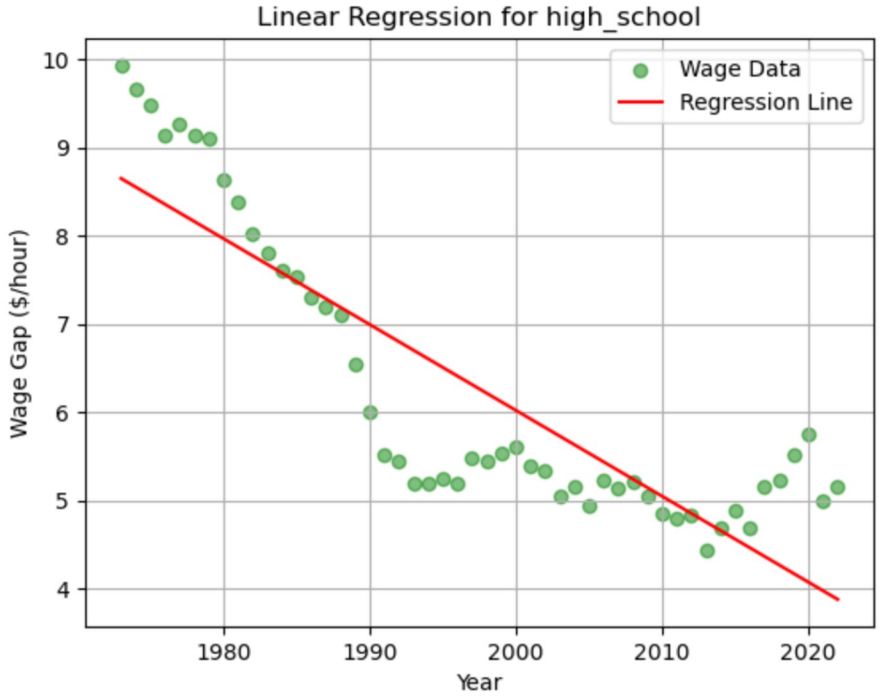
OLS Regression Results

Dep. Variable:	less_than_hs	R-squared:	0.686
Model:	OLS	Adj. R-squared:	0.680
Method:	Least Squares	F-statistic:	105.1
Date:	Fri, 01 Mar 2024	Prob (F-statistic):	1.11e-13
Time:	11:58:58	Log-Likelihood:	-61.668
No. Observations:	50	AIC:	127.3
Df Residuals:	48	BIC:	131.2
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	174.8509	16.596	10.536	0.000	141.483	208.219
years	-0.0852	0.008	-10.252	0.000	-0.102	-0.068

Omnibus:	3.267	Durbin-Watson:	0.133
Prob(Omnibus):	0.195	Jarque-Bera (JB):	1.646
Skew:	-0.048	Prob(JB):	0.439
Kurtosis:	2.116	Cond. No.	2.77e+05

Linear Regression



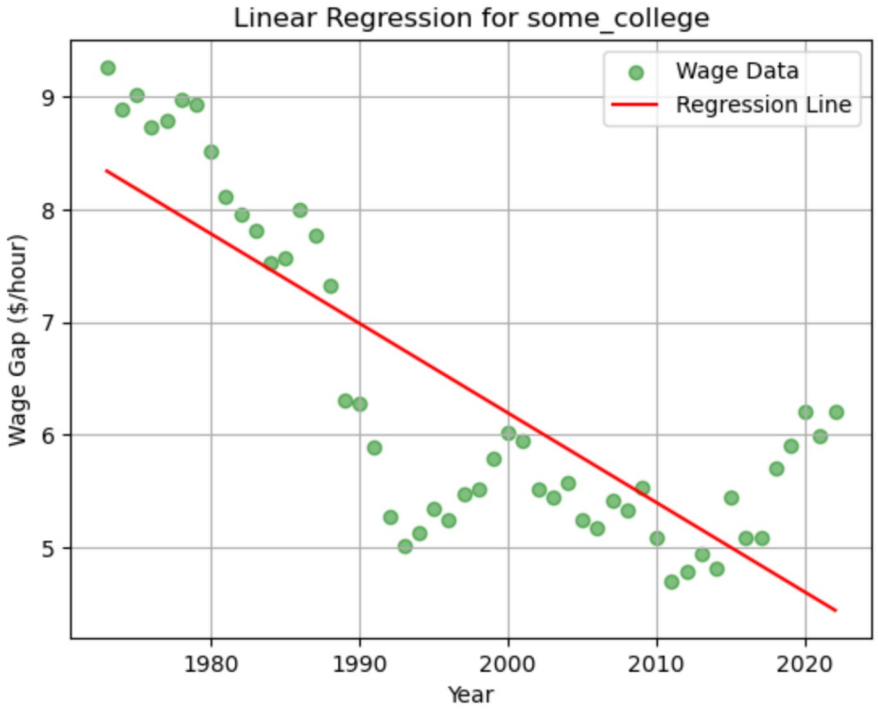
OLS Regression Results

Dep. Variable:	high_school	R-squared:	0.748
Model:	OLS	Adj. R-squared:	0.742
Method:	Least Squares	F-statistic:	142.1
Date:	Fri, 01 Mar 2024	Prob (F-statistic):	5.92e-16
Time:	11:58:58	Log-Likelihood:	-60.818
No. Observations:	50	AIC:	125.6
Df Residuals:	48	BIC:	129.5
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	200.7614	16.316	12.305	0.000	167.956	233.566
years	-0.0974	0.008	-11.921	0.000	-0.114	-0.081

Omnibus:	2.466	Durbin-Watson:	0.089
Prob(Omnibus):	0.291	Jarque-Bera (JB):	1.431
Skew:	0.062	Prob(JB):	0.489
Kurtosis:	2.181	Cond. No.	2.77e+05

Linear Regression



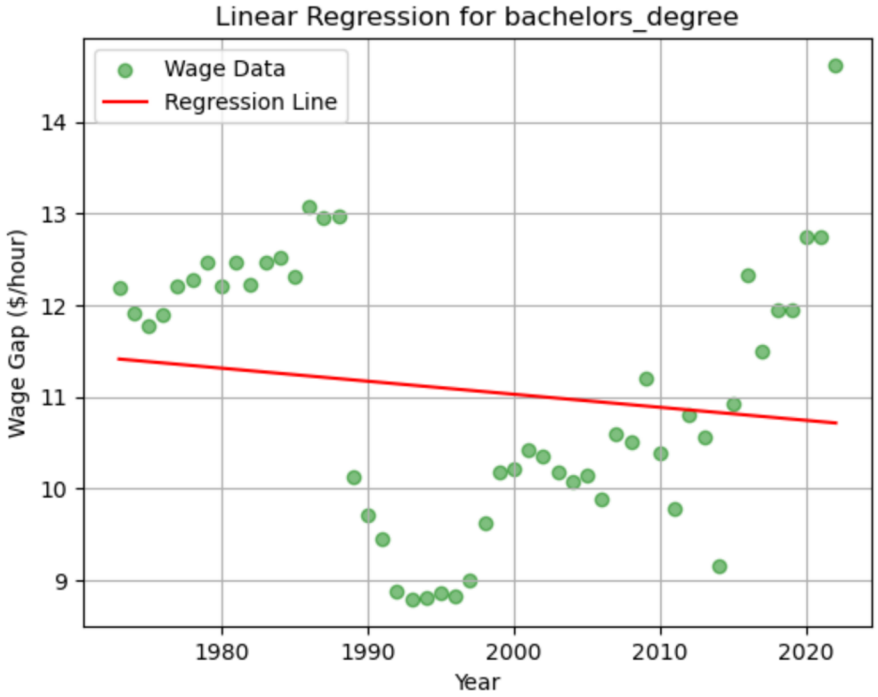
OLS Regression Results

Dep. Variable:	some_college	R-squared:	0.657
Model:	OLS	Adj. R-squared:	0.650
Method:	Least Squares	F-statistic:	92.04
Date:	Fri, 01 Mar 2024	Prob (F-statistic):	9.67e-13
Time:	11:58:58	Log-Likelihood:	-61.548
No. Observations:	50	AIC:	127.1
Df Residuals:	48	BIC:	130.9
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	165.2170	16.556	9.979	0.000	131.930	198.504
years	-0.0795	0.008	-9.594	0.000	-0.096	-0.063

Omnibus:	0.573	Durbin-Watson:	0.141
Prob(Omnibus):	0.751	Jarque-Bera (JB):	0.659
Skew:	0.005	Prob(JB):	0.719
Kurtosis:	2.438	Cond. No.	2.77e+05

Linear Regression



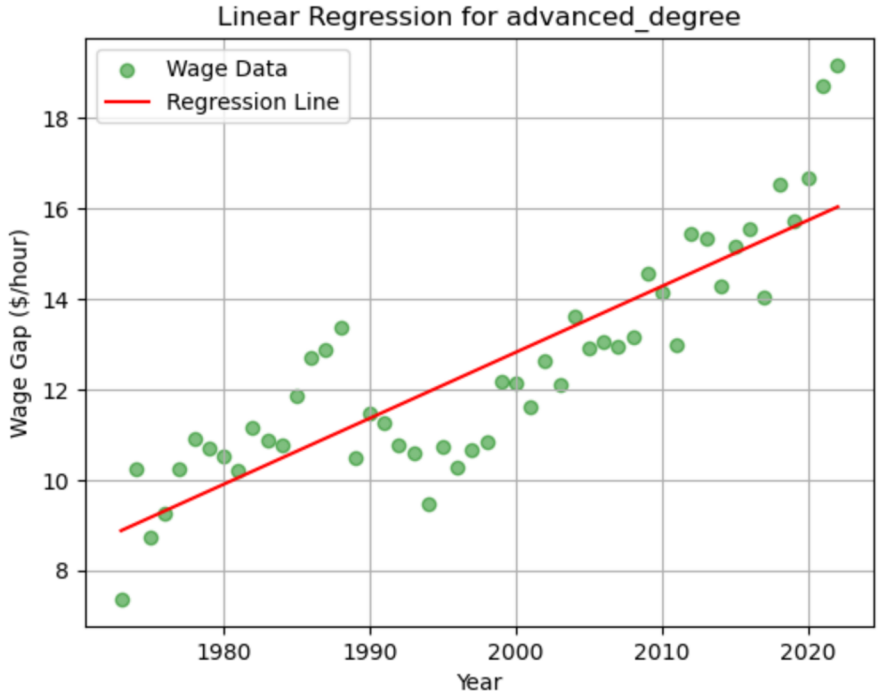
OLS Regression Results

Dep. Variable:	bachelors_degree	R-squared:	0.021
Model:	OLS	Adj. R-squared:	0.001
Method:	Least Squares	F-statistic:	1.029
Date:	Fri, 01 Mar 2024	Prob (F-statistic):	0.315
Time:	11:58:58	Log-Likelihood:	-87.930
No. Observations:	50	AIC:	179.9
Df Residuals:	48	BIC:	183.7
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	39.5306	28.061	1.409	0.165	-16.889	95.950
years	-0.0143	0.014	-1.015	0.315	-0.042	0.014

Omnibus:	0.363	Durbin-Watson:	0.263
Prob(Omnibus):	0.834	Jarque-Bera (JB):	0.528
Skew:	0.149	Prob(JB):	0.768
Kurtosis:	2.595	Cond. No.	2.77e+05

Linear Regression



OLS Regression Results

Dep. Variable:	advanced_degree	R-squared:	0.746
Model:	OLS	Adj. R-squared:	0.741
Method:	Least Squares	F-statistic:	141.1
Date:	Fri, 01 Mar 2024	Prob (F-statistic):	6.70e-16
Time:	11:58:58	Log-Likelihood:	-81.286
No. Observations:	50	AIC:	166.6
Df Residuals:	48	BIC:	170.4
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	-279.4313	24.569	-11.373	0.000	-328.831	-230.032
years	0.1461	0.012	11.881	0.000	0.121	0.171

Omnibus:	1.731	Durbin-Watson:	0.772
Prob(Omnibus):	0.421	Jarque-Bera (JB):	1.510
Skew:	0.418	Prob(JB):	0.470
Kurtosis:	2.838	Cond. No.	2.77e+05

Linear Regression Analysis

- Here we conducted a linear regression for each education level
  - Each predictor variable was the time in years
  - Each outcome variable was the hourly wage for that respective education level
  - This shows us yet another way of viewing how the wage gap has changed over time, delineated by education level
- For the three lowest education levels, the plots are all very similar, with the regression lines also sharing similar slopes
  - Interestingly, each plot seemed to show that there was a significant decline in the wage gap during the 1980's, with the gap mostly leveling off since the 1990's
- For those with bachelor's degrees, the data is quite scattered, and the slope of the regression line is mostly horizontal
  - Again, we see a large decline in the 1980's. This time, however, the gap seems to have started increasing ever since the 1990's
- For those with advanced degrees, the wage gap follows a very clear upward linear trend throughout the entire 50-year period
- The R-squared values for each model (apart from the bachelor's degree model) were all fairly high - between 0.65 - 0.75.

Overall Conclusions

- Original Hypothesis: There has been a statistically significant decline in the gap between the wages of men and women over time
- Conclusion: It depends on education level.
  - For lower lower values of education (less than high school up to some college), we concluded that there has been a statistically significant decline in the wage gap over time. However, much of this is due to a large decline during the 1980's
  - For bachelor's degree holders, no statistically significant difference was found and the null hypothesis was accepted
  - For advanced degree holders, we arrived at the opposite conclusion: there has been a significantly significant *increase* in the wage gap over time
- Overall, we can see that there still exists a significant gap in the wages of men and women, despite some progress being made for jobs that require less education
- For those with advanced degrees, the wage gap has consistently grown higher, making it an interesting trend to assess

Limitations and Future Steps

- In this research project, we looked at the average wages for a 50-year period differentiated by education level and gender
  - We did not take into account differences in job, location, race/ethnicity, and experience level
  - Each of these could be an important factor to consider when looking at differences in wage
- The data has already been adjusted for inflation using the CPI-U-RS index
  - This index is not perfect, and other factors could have been left out of the calculations that affected wages at the time such as the general buying power of the dollar
- The increase in the wage gap for advanced degrees might be heavily skewed by women in education
  - Many teaching jobs require an advanced degree despite lower wages
  - Most teachers are female
- In the future, more data should be assessed that takes into consideration differences in demographics, job and experience level

Thank You for Viewing this Presentation!