Comparing Interactions Between Collective Bargaining, Wages and Income, and House of Representative Elections

Ramiro Arevalo* Melanie Klein† Michael Taffe‡

2024-01-12

Abstract

For our project, our goal is to examine the relationships between collective bargaining agreements, U.S. house election results, and incomes/wages of Americans. The motivation for this project is the rise in wealth inequality and populism in America and the recent surge in union activity. In our project, we find that all three metrics are related. Using regression testing we show that for unions and political leanings, this relationship has surged and deepened in the last four decades from 1984 to 2019

1 Introduction

Unions play a crucial role in negotiating wages and benefits for workers through collective bargaining agreements. The political landscape, particularly the House of Representatives, holds significant legislative authority in influencing workers' rights and labor policies. We examine the relationship between unions and collective bargaining agreements, U.S. House of Representatives election results, and economic variables including the incomes and wages of Americans. The recent surge in union activity, increased political polarization, and a rise in wealth inequality motivate these project topics.

In this paper, we use three research questions to guide our analysis:

1. Is there any relationship between collective bargaining rates among workers to house delegation makeup?
2. Is there any relationship between collective bargaining rates among workers and wages?
3. Do collective bargaining rates effect wealth inequality?

Our research questions guide our two hypotheses:

1. There is a significant correlation between collective bargaining rates and higher amounts of Democrat representation in the House of Representatives.
2. Unions lead to higher wages.

*American University
†American University
‡American University

We use five data sets that are joined into two data frames to explore our research questions. We create a data frame that stores information at the national level about income and union wages, and a data frame that stores information at the state level about minimum wages, union employment, and House election results.

When conducting a regression analysis, it demonstrates that the correlation between collective bargaining and Democratic representation in the House has gotten stronger in the last four decades from 1984 to 2019. Additionally, when conducting t-tests among union and non union wages, there is a statistically significant difference that demonstrates a decreasing wage differential over the same time period.

2 Historical Reference and Overlap Points of Union Wages and Effects of Collective Bargaining Agreements

Analyzing union data sets along with U.S. House of Representative data allows us to examine and search for correlation between wages and the increase/decrease of union participation in national regions. We investigated decreases in wages and union participation, along with widening gaps in income between non union, and union members, and the overall increases in wages. Our study's findings are in line with Western and Rosenfeld (2011). *Unions, Norms and the Rise in U.S. Wage Inequality* found "The decline of organized labor in the United States coincided with a large increase in wage inequality" from 1973 to 2007.

Our graphs convey, and affirm Wilmers (2017), reference point that "even in a period of labor weakness (2000-2014), union activism increases the wages of union members. Utilizing the House of Representatives data set of both Democrats and Republicans, we also find a correlation between the downward slope in union wages and enrollment during a Republican presidency which also came with a large Republican majority in the House during the Bush administration. Particularly since the early 2000s, when the Bush administration's NLRB appointments led to a series of decisions that were especially hostile to organized labor Milkman and Luce (2017).

3 Data and Methods

The data being analyzed is from a collection of five data sets. The first two data sets, `union_wages` and `union_states`, we collect from the source Tidy Tuesday, although the original data comes from the Union Membership and Coverage Database at Georgia State University Johnthegeek (2023a). The tables relevant to this paper detail union membership counts by state and year, as well as union and non union wages nationally. The third data set, `house`, we also collect from the source Tidy Tuesday, although the original data comes from the MIT Election Data and Science Lab Johnthegeek (2023b). The data includes every person who ran for a seat in the U.S. House of Representatives, but we only look at the election winner. The fourth data set, `income_distribution`, we also collect from the source Tidy Tuesday, although the original data comes from the U.S. Census Johnthegeek (2021). The data tracks the mean and median income nationally in the United States. The fifth data set, `minimum_wage`, we collect from the U.S. Department of Labor and tracks state minimum wage and the federal minimum wage LisleJoem (2021).

All of the data sets undergo basic cleaning so that the columns are correctly formatted, variable names are consistent, there are no NA values, and relevant variables are selected. For the House election data, the table is manipulated so that only rows of candidates who won their election remain. A binary variable is created in which we code the outcome as 1 if the winner was a

Democrat and code the outcome as 0 if the winner was a Republican. We create an additional variable to show the proportion of the House delegation that is Democratic for every combination of years and states. For the union data, we filter it to not include union sectors, such as construction unions, and instead look at all unions. For the income distribution data, we filter it to not include race because it is outside of the scope of this study.

We combine the five data sets into two data sets: one that stores data at the state level and another that stores data at the national level. `full_state_data` joins the `union_states`, `house`, and `minimum_wage` data sets. There are `8730` observations and each row represents a year and a state. `full_ntl_data` joins the `union_wages` and `income_distribution` data sets. There are `36` observations and each row represents a distinct year. Variables regarding union employment are not provided at the national level, only at the state level, so we summarize and add them to the national data set. Economic variables in the national data are not originally adjusted to 2020 values, so we calculate the rate from the state data that contains adjusted and non-adjusted values for federal minimum wage, and then apply that rate to the national data. This allows all of the economic data to be standardized and account for inflation. Among the original five data sets, there is data recorded from 1967 to 2022. Not all of the data sets cover that time frame, so once joined into two data sets, we filter the data to only include the years 1984 to 2019.

3.1 Union Data

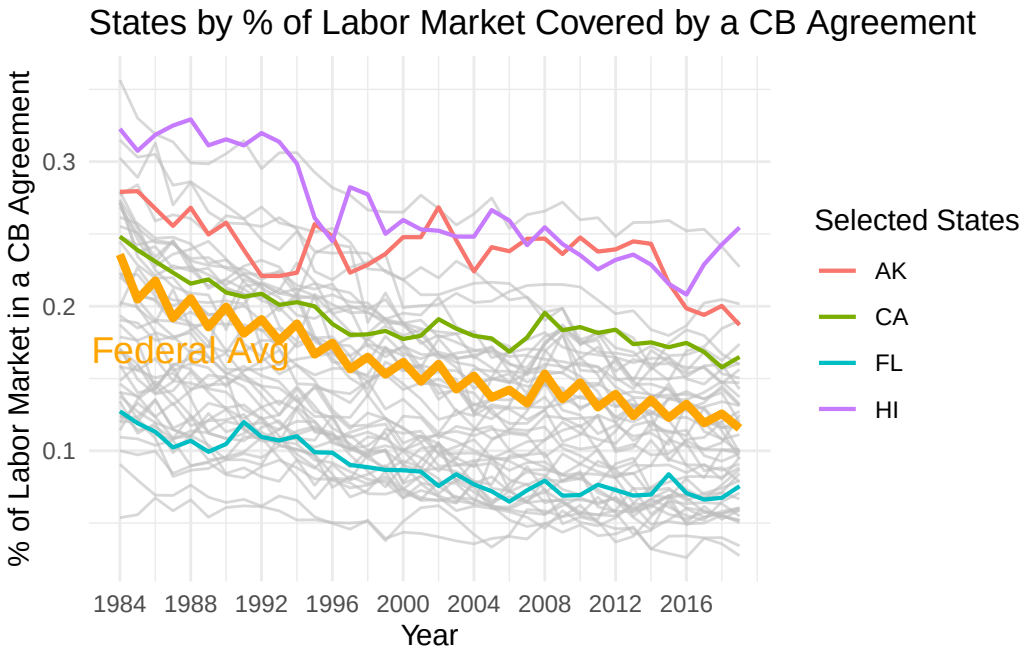

Figure 1: Union Data

Figure 1 shows what is commonly known about collective bargaining agreements in the United States: they have been on the decline since at least the 1980s. Unfortunately our data does not extend to the last four years where some reports indicate union membership is on the rise. Early on, we see that southern and more Republican leaning states have lower collective bargaining rates than more Democrat and northern / western states. There is a very small bump in membership in 2008, which we hypothesize happened in relation to the 2008 financial crisis.

3.2 Income Data

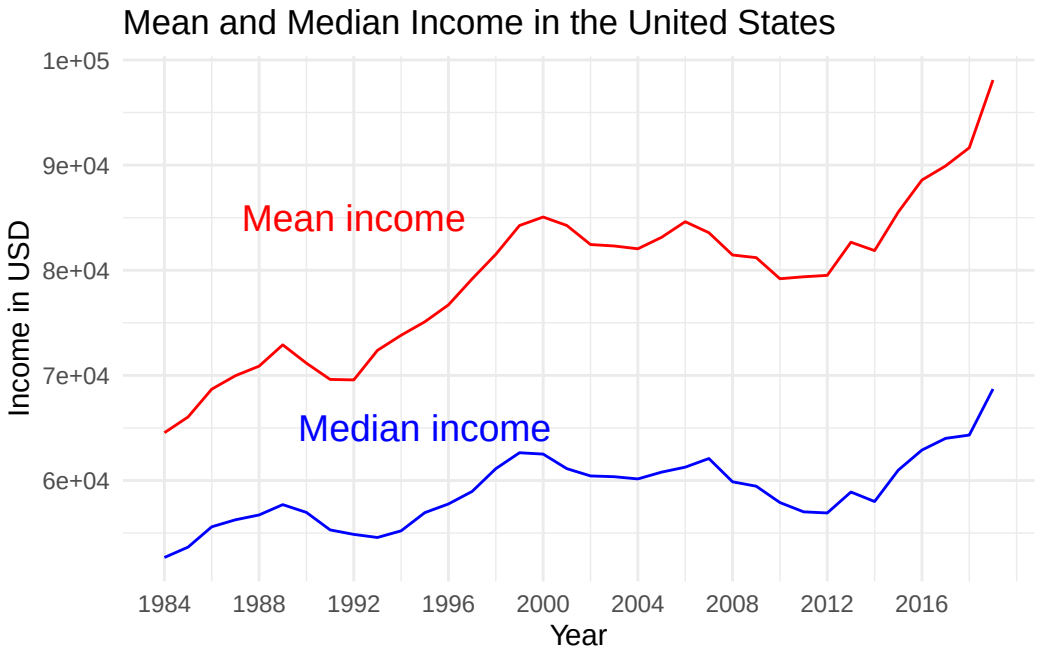

Figure 2: Income Data

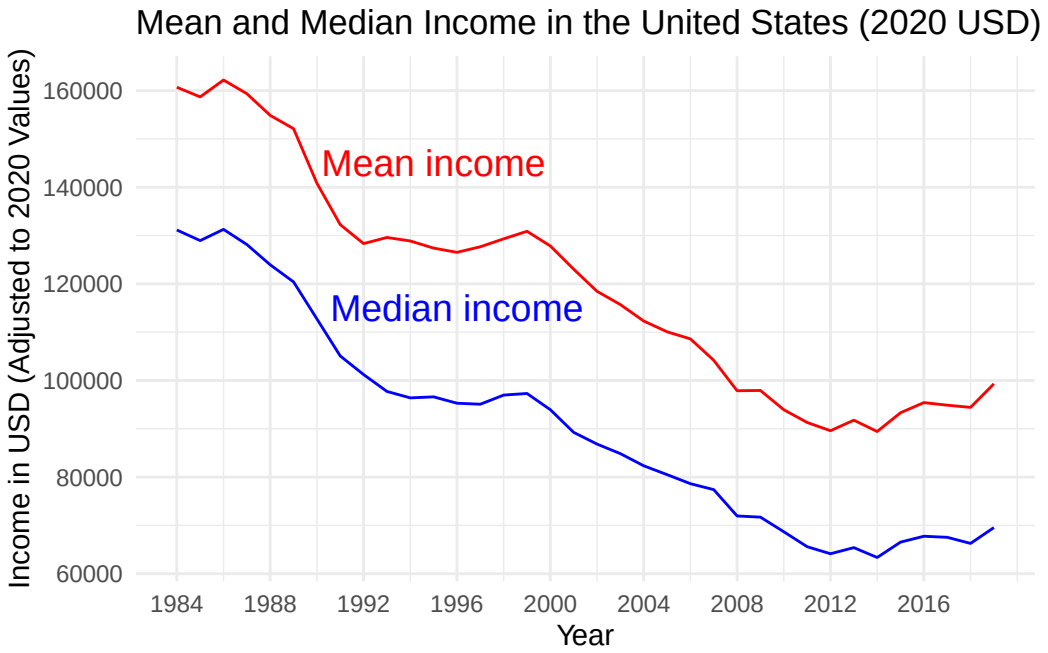

Figure 3: Income Data (adjusted)

Figure 2 and Figure 3 show that unadjusted income increases, but when accounting for inflation, Americans are earning less per year now than they did 4 decades ago. As we will see later on when examining wage data at a national level, there are numerous external reasons for this. Most notably, an aging population that is no longer working is likely driving the average down. It is possible that the average worker is earning more but we do not have the data to verify this.

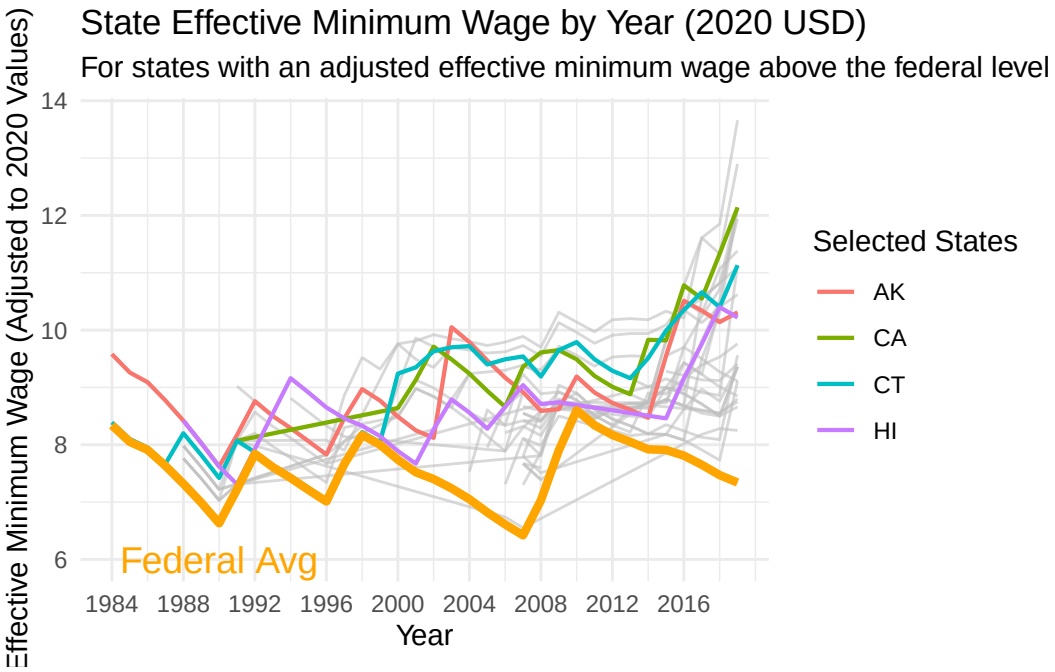

Figure 4: Wage Data

Figure 4 removes all data points where the state effective minimum wage falls below the federal amount. From this we see that more states are taking it into their own hands to protect workers rights. We see a large jump in 2007 which may be related to the financial crisis. There is another spike around 2014.

3.3 House of Representatives Data

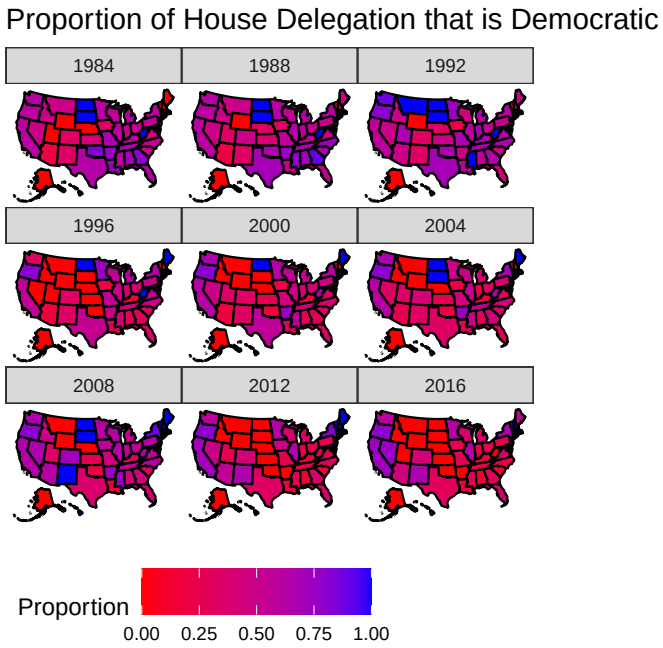

Figure 5: House Data

In Figure 5 we see that, especially in the northern-central states, the country has grown to be more Republican in the last 4 decades. However, in the last 10 years this has begun to shift again. During this whole time, it appears that New England has become more Democratic. It is important to note that the states in the north-central region typically have only one to three representatives which can make changes appear much more drastic than they actually are.

3.4 Collective Bargaining and Wage Data

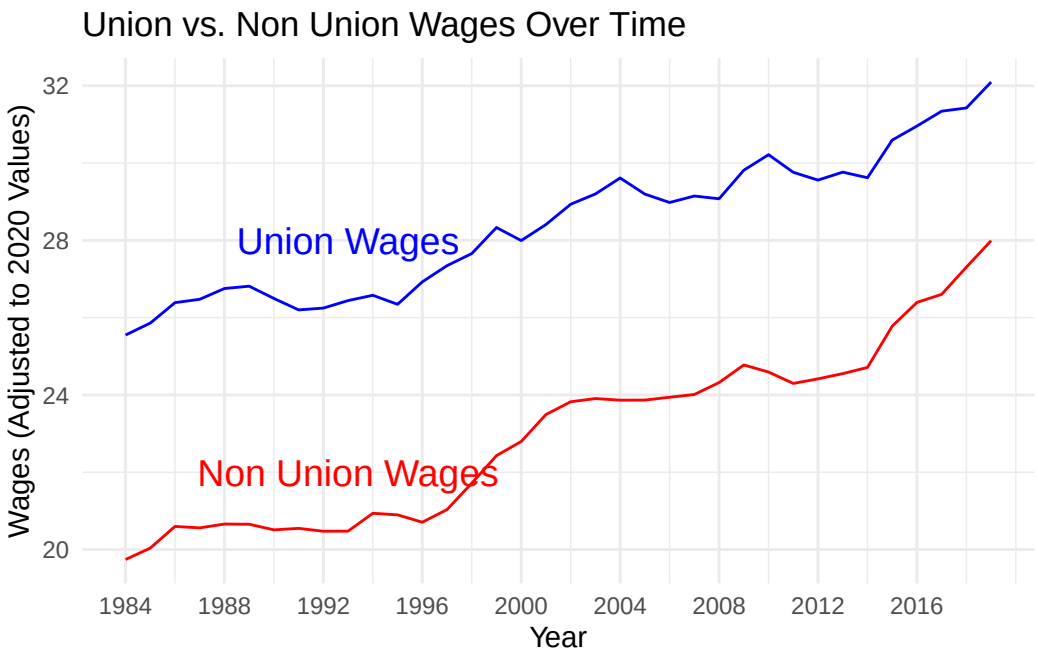

Figure 6: Union Wage Data

Figure 6 is rather unintuitive, as we would expect adjusted wages to be decreasing in this time, based on Figure 3. There are a number of potential causes for this. First, Americans work less hours today than they did in 1980. Second, and more likely, our income variables include those who do not work. As the U.S. experiences a "graying of society", this could be driving down mean incomes.

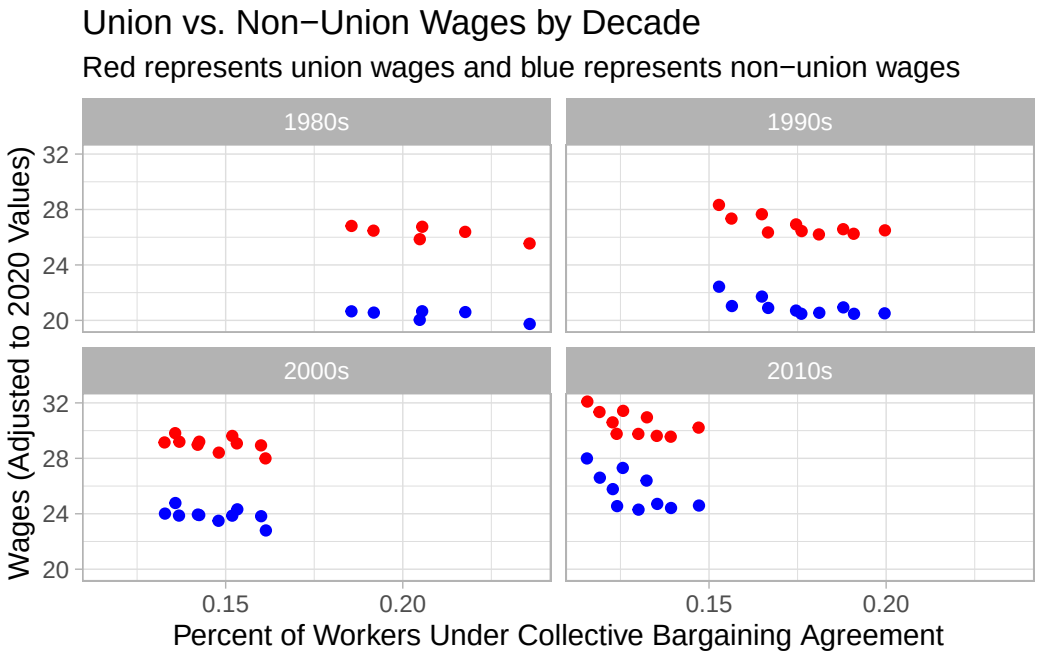

Figure 7: Collective Bargain Data

Figure 7 is incredibly interesting. First, a quick view of this graph shows the exact opposite of our hypothesis. Looking at each decade, we see collective bargaining rates decreasing and wage increasing. Then, in each decade that relationship continues. Still, we see that in every decade, a collective bargaining agreement is associated with higher wages than for a worker over non-collective-bargaining-jobs. However, by the 2010s, it appears that the distance between union and non union wages are shrinking. Still, this finding should be taken in the context that the graph does not account for the number of workers in each category and other confounding variables.

3.5 Collective Bargaining and House of Representatives Data

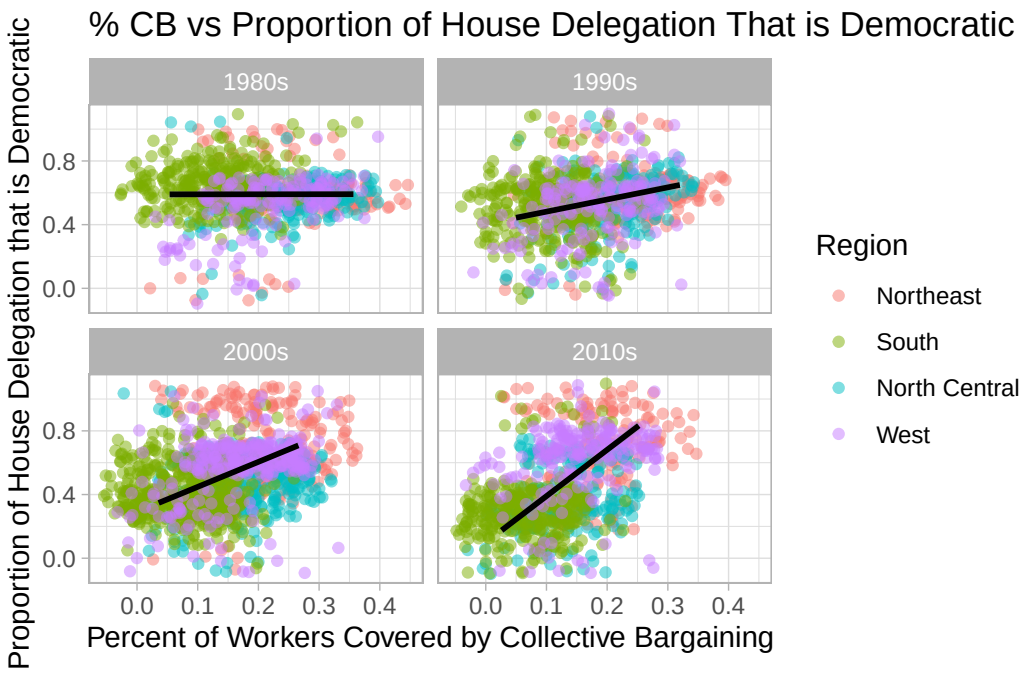

Figure 8: Collective Bargain Data by Decade

Figure 8 demonstrates the political association with collective bargaining agreements over the last 4 decades In the 1980s, there was almost no relationship between collective bargaining agreements and party affiliation, as seen by the horizontal regression line. However, by the 2010s, there is a significant increase in regression slope.

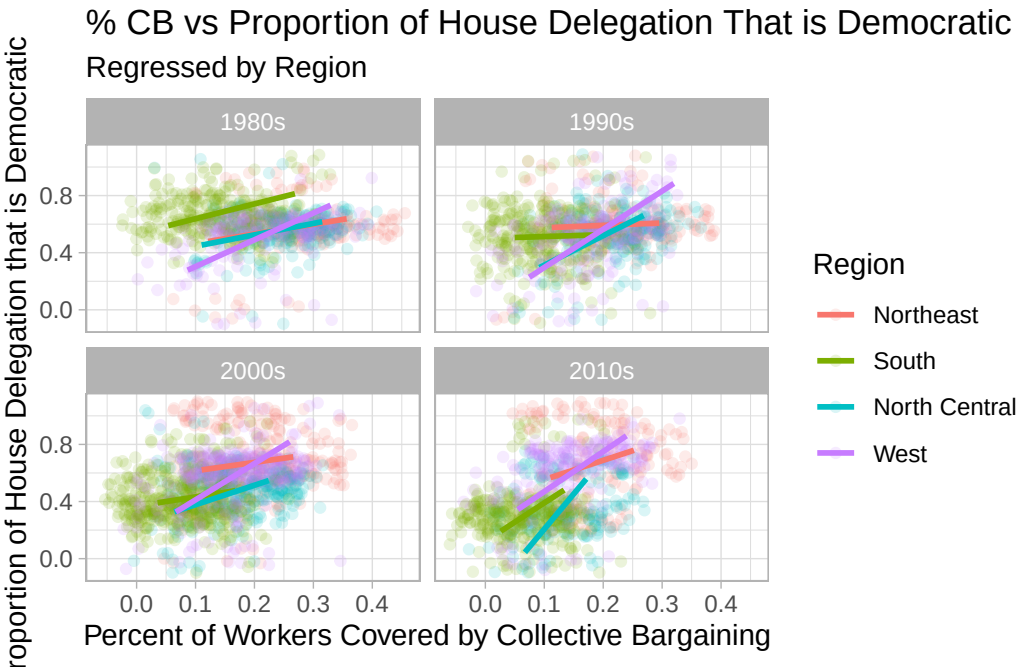

Figure 9: Collective Bargain Region by Decade and Region

In Figure 9 we can see that the North Central and South regions are the main drivers behind this, seen in the steepening of the regression slope.

4 Exploring Correlations: Linear Regression Analyses and T-Tests on Collective Bargaining and Wage Disparities

Based on our exploratory data analysis, we conduct linear regression analyses and t-tests. The linear regressions use data from `full_state_data` and explore how collective bargaining agreements are correlated with Democratic representatives being elected to the House in all four regions. The results are used to examine if the North Central region and Southern region are driving this stronger correlation over time. Additionally, the t-test uses data from `full_ntl_data` to understand if the observed variations in wages between unionized and non-unionized workers are statistically meaningful. By looking at different decades, we aim to analyze potential shifts in wage consistencies or differences over time. This provides insight into the dynamics of the labor market.

4.1 Predicting Democratic House Delegations with Percentage of Workers in Collective Bargaining Argreements

Building on the results in Figure 8 and Figure 9, Table 1 performs a linear regression between the percent of workers in a collective bargaining agreement and the proportion of the House delegation that is Democratic for the 1980s and 2010s in each region, as well as for all of the regions combined

Table 1: Linear Regression

	Reggression_Run	Decade	Region	Estimate	Std. Error	t value	P_Value
1	percent_collective_bargain	1980s	all	-0.03	0.06	-0.44	0.66
2	percent_collective_bargain	1980s	Northeast	0.58	0.19	3.12	0.00
3	percent_collective_bargain	1980s	West	1.81	0.14	13.25	0.00
4	percent_collective_bargain	1980s	North Central	0.98	0.15	6.45	0.00
5	percent_collective_bargain	1980s	South	0.97	0.13	7.50	0.00
6	percent_collective_bargain	2010s	Northeast	1.41	0.25	5.65	0.00
7	percent_collective_bargain	2010s	West	2.61	0.18	14.85	0.00
8	percent_collective_bargain	2010s	North Central	4.10	0.25	16.31	0.00
9	percent_collective_bargain	2010s	South	2.28	0.24	9.44	0.00
10	percent_collective_bargain	2010s	all	2.86	0.07	38.37	0.00

For all linear regressions, $p < 0.05$, except for in the 1980s when examining `all` regions, $p = 0.659$.

Figure 10 plots the `Estimate` values calculated in the linear regressions. The regression analysis shows that the South and North Central regions in the United States are the main regions that have seen union membership rates and political leanings show increased correlation. The South and North Central regions have the greatest increase in their estimate values from the 1980s to the 2010s among the four regions.

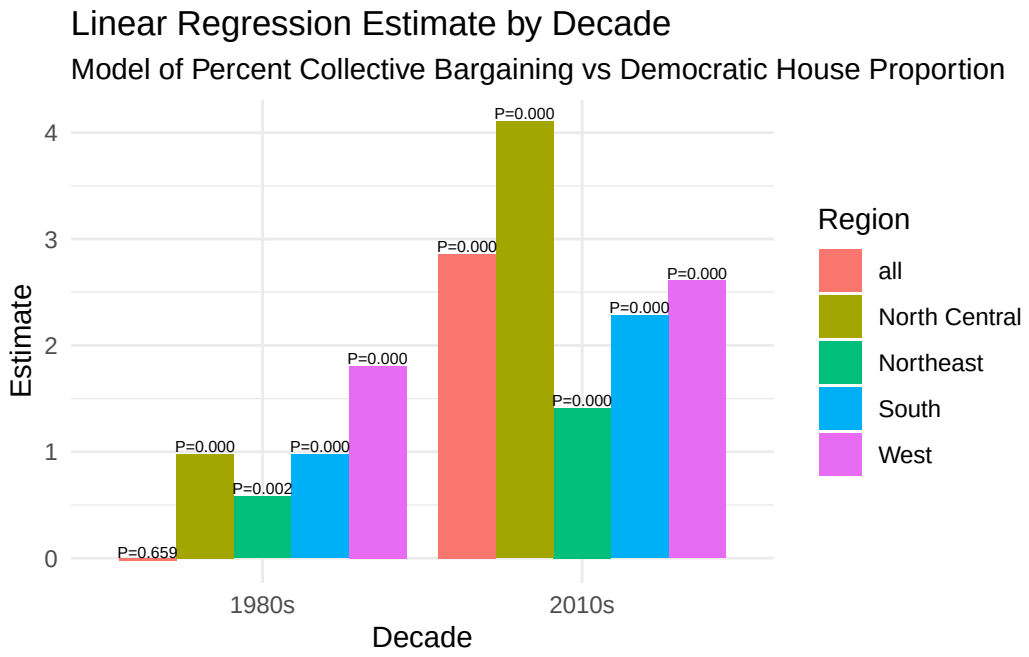

Figure 10: Linear Regression Estimates

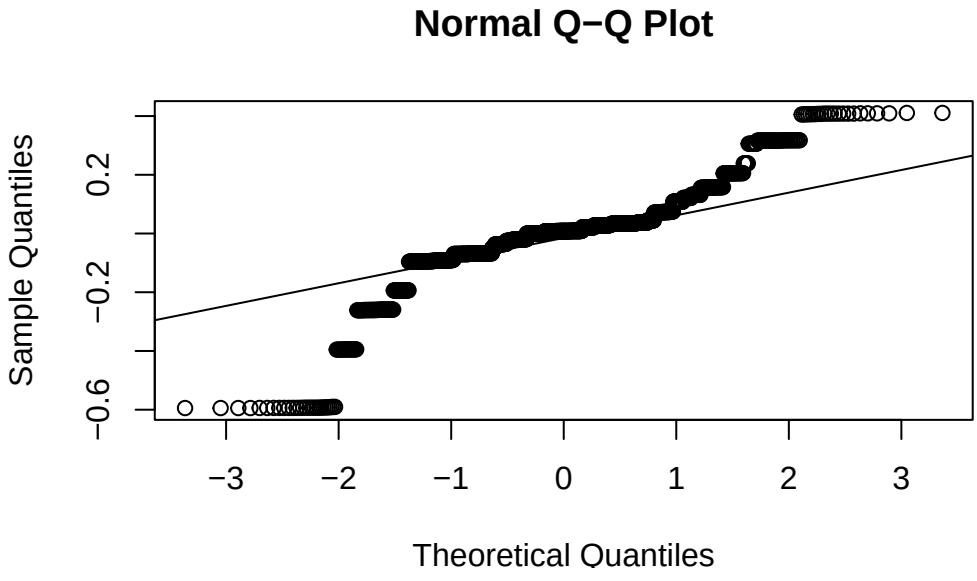

Figure 11: QQ Plot 1980s

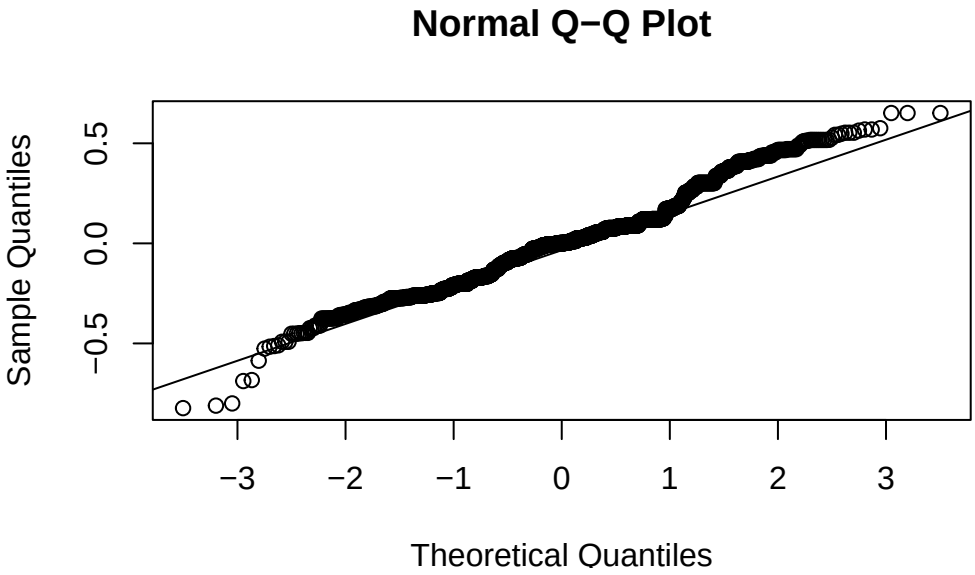

Figure 12: QQ Plot 2010s

The QQ plot in Figure 11 from the 1980s shows a very bimodal distribution, while the QQ plot in Figure 12 from the 2010s regression model shows a more normal, slightly light-tailed distribution.

4.2 Comparing Wages between Union and Non Union Workers

Building on the results in Figure 7, our results seen in Table 2 for these t-tests determine if there is a statistically significant difference between union wages and non union wages in the 1980s and the 2010s.

The difference between union wages and non union wages means in the 1980s is 5.9313876, with a 95% confidence interval that covers $(5.35, 6.51)$. The difference between union wages and non union wages means in the 2010s is 4.8695495, with a 95% confidence interval that covers $(3.79, 5.95)$. The 95% confidence interval for the 1980s has a smaller range than the 95% confidence interval for the 2010s.

Table 2: Wages by Union and Non Union Workers

metric	t_out_1980s	t_out_2010s
t	-22.85	-9.55
df	9.41	15.63
p-value	1.46e-09	6.39e-08
95%-CI-LB	-6.51	-5.95
95%-CI-UB	-5.35	-3.79
Mean_of_X	20.38	25.66
Mean_of_Y	26.31	30.53

5 Discussion

From our data analysis, it can be inferred that collective bargaining agreements have become a very political concept. The charts in our exploratory data analysis and regression models show this. This partially supports our hypothesis. While the correlation is a newer phenomenon, the relationship appears to be strong today. The relationship between unions and wages/income in the United States is much harder to draw inferences. While our t-tests and some charts show that unions can help workers earn more, other graphs show that unions have a negative relationship on incomes. Our second hypothesis thus remains without strong support.

There are multiple limitations to our project that we found during our research, that we hope to address in future research. First, our regression models can be improved by adding a time series element to the model. Furthermore, we can use a multiple linear regression model instead of multiple SLR models. For instance, in an MLR model with a time series element, we can take into account the success of unions in a given time period and state. We can also include other socioeconomic factors at the state level such as housing prices and general costs of living. This can provide a more accurate image of what trends are occurring in the data. Next, data past 2019 can provide useful insights into the union work done by the Biden administration. Finally, with more variables we may be able to expose more nuanced relationships. Two example of this would be union wages and other income indicators at the state level and election data for the U.S. Senate and individual state legislatures.

Overall, in our project, we are able to provide merit to our first hypothesis, showing a growing relationship between collective bargaining agreement rates and House delegation makeup. We were not able to substantiate our second hypothesis as some of our graphs showed the opposite of our predictions while others and our t-test did indicate our predictions were correct.

References

Johnthegeek. 2021. "Wealth and Income over Time." https://github.com/rfordatascience/tidytuesday/blob/master/data/2021/2021-02-09/readme.md.

———. 2023a. "Union Membership in the United States." https://github.com/rfordatascience/tidytuesday/blob/master/data/2023/2023-09-05/readme.md.

———. 2023b. "US House Election Results." https://github.com/rfordatascience/tidytuesday/blob/master/data/2023/2023-11-07/readme.md.

LisleJoem. 2021. "US Minimum Wage by State from 1968 to 2020 and 2020 Equivalent Dollars." https://www.kaggle.com/datasets/lislejoem/us-minimum-wage-by-state-from-1968-to-2017.

Milkman, Ruth, and Stephanie Luce. 2017. "Labor Unions and the Great Recession." *RSF: The Russell Sage Foundation Journal of the Social Sciences* 3 (3): 145–65.

Western, Bruce, and Jake Rosenfeld. 2011. "Unions, Norms, and the Rise in US Wage Inequality." *American Sociological Review* 76 (4): 513–37.

Wilmers, Nathan. 2017. "Labor Unions as Activist Organizations: A Union Power Approach to Estimating Union Wage Effects." *Social Forces* 95 (4): 1451–78.