Unions Are Good!

Joseph Thiers

Contents

1 Introduction **4**
- 1.1 What This Analysis Is All About 4
- 1.2 Where the Data Came From 4
- 1.3 Why This Matters 5

2 Data Preparation **6**
- 2.1 Getting the Data Ready 6
- 2.2 Quick Data Overview 6
- 2.3 Filtering to Focus on the Present 7
- 2.4 Merging the Data 7
- 2.5 Making Data Manipulation Easier 7

3 Analysis and Aggregations **9**
- 3.1 Calculating Key Metrics 9
- 3.2 Rescaling Data for Better Visualization 10

4 Visualization of Results **12**
- 4.1 Average Wage Trends Over Time 12
- 4.2 Union Wages Comparison by Year 14
- 4.3 Employment Trends by State, Including Union Membership 15
- 4.4 Union vs Nonunion Wage Trends Over Time 17
- 4.5 Union Wage Premium Over Time 17
- 4.6 Union Membership Trends Over Time 20

4.7 Employment in Different Sectors Over Time 21

5 Discussion and Summary 23
5.1 Wage Growth and Union Influence 23
5.2 Decline in Union Membership 24
5.3 Shifts in Employment by Sector 24
5.4 Implications for the Future 24

6 Appendix 26
6.1 Additional Code 26
6.2 Data Source Links 27
6.3 Tools and Packages Used 28
6.4 Additional Reading and Resources 28

Chapter 1

Introduction

1.1 What This Analysis Is All About

In this zine, we're diving into data on wages, demographics, and state info to see how union membership stacks up against nonunion wages. We want to understand what makes a difference in pay and job security across different states and communities. Spoiler alert: unions have a big impact, and we're here to show you how!

1.2 Where the Data Came From

We used datasets from R for Data Science Online Learning Community [2024] that cover wage and union membership trends across the U.S. from 1973 to 2022. These datasets come from government surveys and are packed with info on union membership, wages, and employment stats. We use three main `datasets-demographics.csv`, `wages.csv`, and `states.csv` which let us track how unionization has affected wages over time. The data tells a compelling story about how unions have shaped workers' lives for decades, and we're here to shine a light on that.

1.3 Why This Matters

We want to understand how unions impact wages and employment patterns. Are union members earning more? Are they better off than nonunion workers? Which communities benefit the most from union membership? Our goal is to highlight the real power of unions in making workers' lives better. We hope to give you a better idea of how collective bargaining makes a difference, both for individuals and for entire communities.

Chapter 2

Data Preparation

2.1 Getting the Data Ready

We loaded three datasets—demographics, wages, and states using some simple code in R. These datasets have key info: `demographics.csv` tells us about employment by different demographic groups, `wages.csv` gives us union vs. nonunion wages, and `states.csv` tells us about state-level union activity and employment. Here's how we got started:

```
setwd("~/Documents/School/RaceIncomeCalifornia")
demographics <- readr::read_csv('demographics.csv')
wages <- readr::read_csv('wages.csv')
states <- readr::read_csv('states.csv')
```

2.2 Quick Data Overview

We took a close look at what's inside each dataset. The `demographics.csv` shows union membership and employment data broken down by things like gender, race, and age. The `wages.csv` lets us compare wages for union and nonunion workers, while `states.csv` gives a bigger picture at the state level. Understanding what's in each dataset helps us figure out how to connect it all together to tell a story about unions and wages.

2.3 Filtering to Focus on the Present

I decided to focus on data from 2000 onwards. Why? Because we want to see what's happening now, or at least in the recent past. Older data might not be as relevant to the present-day challenges and victories of workers, so we filtered it down to what's most useful.

```
filtered_wages <- wages[wages$year >= 2000, ]
filtered_states <- states[states$year >= 2000, ]
filtered_demographics <- demographics[demographics$year >= 2000, ]
```

2.4 Merging the Data

We merged the wages and states datasets to make it easier to look at everything together. This helps us connect wages (union and nonunion) with state-level data, like union membership rates. By combining this data, we can get a fuller picture of how unions are making a difference for workers in each state.

```
merged_data <- merge(states, wages, by = "year", all.x = TRUE)
```

2.5 Making Data Manipulation Easier

I converted the filtered datasets into a faster, more efficient format using data.table. This makes it easier to manipulate big datasets without a lot of waiting around, which is especially helpful when plotting graphs.

```r
if (!requireNamespace("data.table", quietly = TRUE)) {
  install.packages("data.table")
}
library(data.table)

setDT(filtered_wages)
setDT(filtered_states)
setDT(filtered_demographics)

setDT(filtered_wages)
setDT(filtered_states)
setDT(filtered_demographics)
```

Chapter 3

Analysis and Aggregations

3.1 Calculating Key Metrics

In this chapter, I calculated some key metrics to understand wage trends and union impact. I wanted to get a clear sense of how union and nonunion wages compare and how union membership has changed over time.

- **Mean Wages**: First, we calculated the average wages for union and nonunion workers over the years. This helped us see the difference in pay between the two groups.

```
wage_trends <- filtered_wages[,
      list(mean_union_wage = mean(union_wage,
                                  na.rm = TRUE),
      mean_nonunion_wage = mean(nonunion_wage,
                                na.rm = TRUE)), by = list(year)]
```

This code calculates the average union and nonunion wages for each year from 2000 onwards, allowing us to visualize and compare wage trends between unionized and nonunionized workers. These calculations provide an overview of how wages for both categories have evolved over the years.

- **Union Wage Premium**: I also calculated the "union wage premium," which is basically the extra amount union workers make compared to nonunion workers.

```
union_wage_premium <- filtered_wages[,
      list(raw_premium = mean(union_wage_premium_raw, na.rm = TRUE),
      adjusted_premium = mean(union_wage_premium_adjusted,
                              na.rm = TRUE)), by = year]
```

The union wage premium calculation helps us quantify the economic benefits of union membership. The adjusted premium, in particular, sheds light on whether the observed wage differential is purely due to union membership or influenced by other worker characteristics.

- **Union Membership Trends**: I looked at how union membership has changed over time. Understanding how union density (the percentage of workers in a union) has shifted helps us connect the dots between membership levels and wage benefits.

```
membership_trends <- filtered_demographics[,
        list(average_p_members = mean(p_members, na.rm = TRUE)), by = year]
```

By calculating the average percentage of union members for each year, we can identify any notable increases or decreases in union membership. These trends help inform our understanding of how unions have gained or lost influence over the past few decades.

3.2 Rescaling Data for Better Visualization

In addition to the core aggregations described above, further calculations were performed to deepen our understanding of the dataset and prepare for visualizations.

- **Average Wage per Year**: I also calculated an overall average wage (for both union and nonunion workers) for each year to provide a broader view of wage growth over time.

```
average_wage_by_year <- filtered_wages[,
      list(average_wage = (mean(union_wage, na.rm = TRUE) +
      mean(nonunion_wage, na.rm = TRUE)) / 2),
        by = list(year)]
```

This calculation helps in understanding whether wages have increased consistently over the years and if union membership has had an impact on overall wage trends. By averaging the wages of union and nonunion workers, we provide a generalized view of wage progression in the broader labor market.

- **Members Rescaling**: I rescaled the number of union members so we could better visualize membership levels alongside other employment metrics. This helps us create clearer, more impactful graphs.

```
merged_data$members_scaled <- merged_data$members * 100
```

The rescaling of union membership data helps align the values with other metrics, making visualizations more intuitive and interpretable. It ensures that union membership figures can be easily compared alongside other variables such as employment or wages in different graphical plots.

With these metrics calculated, we're ready to dive into visualizing the power of unions and how they impact workers' lives.

Chapter 4

Visualization of Results

4.1 Average Wage Trends Over Time

I used ggplot to create a graph that shows the average wages from 2000 to 2020. It's easy to see that wages have generally gone up—but union wages are leading the way.

```
library(ggplot2)
ggplot(average_wage_by_year, aes(x = year, y = average_wage)) +
  geom_line() +
  labs(x = "Year", y = "Average Wage") +
  theme_minimal() +
  scale_color_brewer(palette = "Set1")
```

The 4.1 visualization displays the trend of average wages over the two-decade span starting from the year 2000 up to 2020. This plot shows a steady increase in wages, highlighting an upward trajectory that suggests a positive economic trend. The x-axis represents the years, while the y-axis denotes the average wage figures. The line graph is particu-

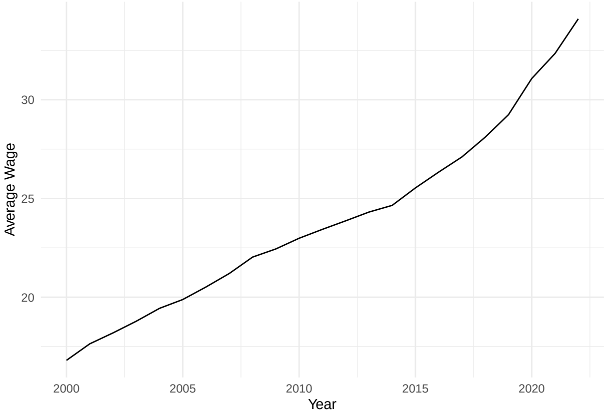

Figure 4.1: Average Wage

larly useful for tracking changes over time, making it easier to visualize the rate of growth or any periods of stagnation.

4.2 Union Wages Comparison by Year

Next, we used a boxplot to compare union wages over different years. This gives us a look at how wages have varied for union workers.

```
ggplot(filtered_wages, aes(x = factor(year), y = union_wage)) +
  geom_boxplot() +
  labs(x = "Year", y = "Union Wage") +
  theme_minimal()
```

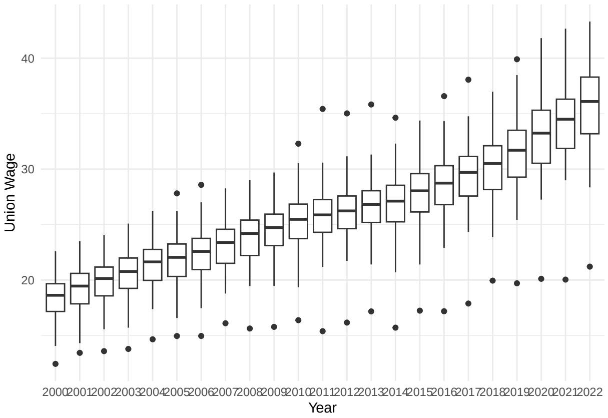

Figure 4.2: Union Wages by Year

This boxplot 4.2 illustrates union wages over different years, capturing the range of wages as well as median values. By comparing the median line in each box, we can see if union wages have experienced upward or downward trends. The presence of outliers in some years indicates that there were wage values that significantly deviated from the rest of the dataset, providing insights into wage inequality within union jobs. Additionally, the increasing height of the boxes over time may suggest growing wage variability among union workers.

4.3 Employment Trends by State, Including Union Membership

I wanted to show how employment and union membership vary across different states. For this, we used a combination of bar charts and line graphs. The idea here is to illustrate not just the number of employed people in each state, but also how many of them are union members. This is important because higher union membership often means better wages and working conditions for more people.

```
ggplot() +
  geom_bar(data = states, aes(x = state, y = employment),
           stat = "identity", fill = "steelblue", alpha = 0.7) +
  geom_line(data = merged_data,
            aes(x = state, y = members_scaled,
                group = state), color = "red") +
  labs(x = "State",
       y = "Employment/union members", y.sec = "Members (scaled)") +
  theme_minimal() +
  scale_fill_brewer(palette = "Pastel1") +
  coord_flip() +
  scale_y_continuous(sec.axis = sec_axis(~ . / 100, name = "Members"))
```

In this graph 4.3, the blue bars show how many people are working in each state, while the red line shows how many of them are in unions. The states with the highest union membership—like California, New York, and Pennsylvania—are leading the way in worker rights and fair treatment. It's pretty simple: more unions = better lives for workers.

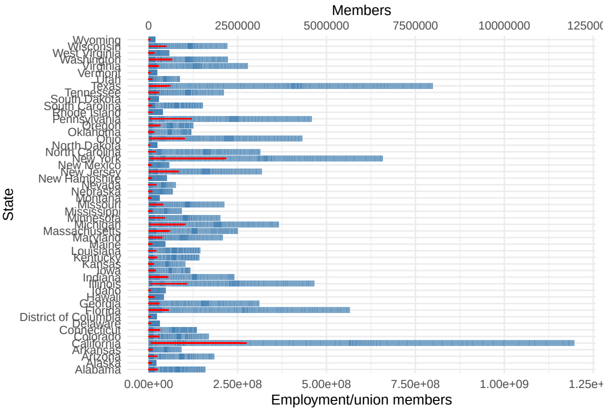

Figure 4.3: Employement Trends by State

4.4 Union vs Nonunion Wage Trends Over Time

We wanted to make it crystal clear how much better union wages are compared to nonunion wages. This graph shows just that—union vs. nonunion wages over time.

```
ggplot(wages, aes(x = year)) +
  geom_smooth(aes(y = union_wage, color = "Union Wage"),
              se = FALSE, method = "loess") +
  geom_smooth(aes(y = nonunion_wage, color = "Nonunion Wage"),
              se = FALSE, method = "loess") +
  labs(x = "Year", y = "Wage",
       color = "Type of Wage") +
  theme_minimal(base_size = 14) +
  theme(legend.position = "bottom") +
  scale_color_manual(values = c("Union Wage" = "blue", "Nonunion Wage" = "green"))
  guides(color = guide_legend(title = "Wage Type"))
```

```
## `geom_smooth()` using formula = 'y ~ x'
## `geom_smooth()` using formula = 'y ~ x'
```

Union wages (in blue) are consistently higher than nonunion wages (in green). It's proof that collective bargaining works. When workers join together, they get better pay—it's as simple as that.

4.5 Union Wage Premium Over Time

I also took a look at the "union wage premium," which is just a fancy way of saying how much extra money union workers make compared to nonunion workers. This graph shows that premium over time.

```
ggplot() +
  geom_point(data = wages, aes(x = year, y = union_wage_premium_raw),
             color = "orange") +
  geom_smooth(data = wages,
  aes(x = year,
```

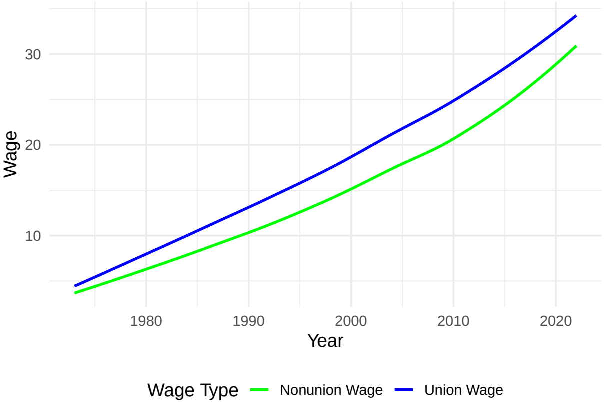

Figure 4.4: Union Vs. Union Wage Over Time

```
    y = union_wage_premium_adjusted), method = "loess", color = "purple") +
  labs(x = "Year", y = "Wage Premium") +
  theme_minimal() +
  scale_color_brewer(palette = "Dark2")
```

```
## `geom_smooth()` using formula = 'y ~ x'
```

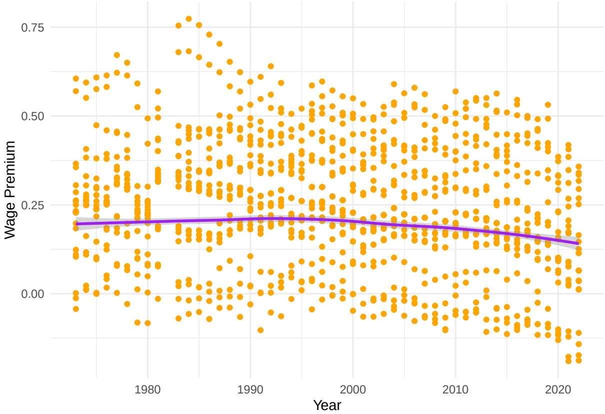

Figure 4.5: Union Wage Premium Over Time

The orange points show the raw premium, while the purple line is the adjusted premium. What this graph tells us is that unions are getting workers paid more, even when you take other factors into account. That's real value for workers. But we see that over time this is trending down.

4.6 Union Membership Trends Over Time

We also looked at how union membership has changed over time with this area chart. It's important because union membership has been on the decline, and we need to change that.

```
ggplot(demographics, aes(x = year, y = p_members)) +
  geom_area(fill = "lightblue", alpha = 0.5) +
  labs(x = "Year", y = "Percentage of Union Members") +
  theme_minimal() +
  scale_fill_brewer(palette = "Blues")
```

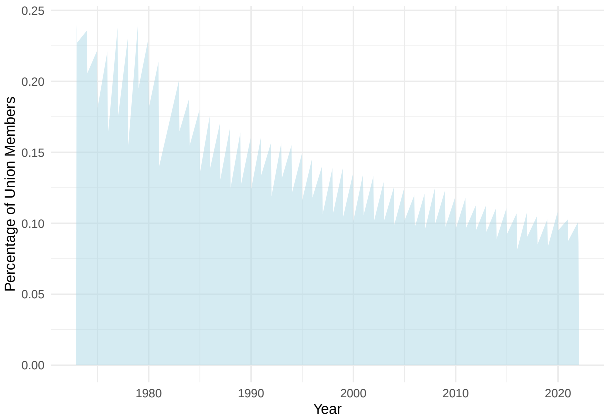

Figure 4.6: Uniont Membership Trends

The drop in union membership is clear. Fewer union members mean less bargaining power, and that's why wages haven't risen as fast as they should. We need to support union growth to bring back fair wages and worker protections.

4.7 Employment in Different Sectors Over Time

Lastly, we looked at employment trends across different sectors. This stacked bar graph shows how employment has changed from manufacturing to more service-focused jobs over time.

```
ggplot(states, aes(x = year, y = employment, fill = sector)) +
  geom_bar(stat = "identity", position = "stack") +
  labs(x = "Year", y = "Employment") +
  theme_minimal() +
  scale_fill_brewer(palette = "Spectral")
```

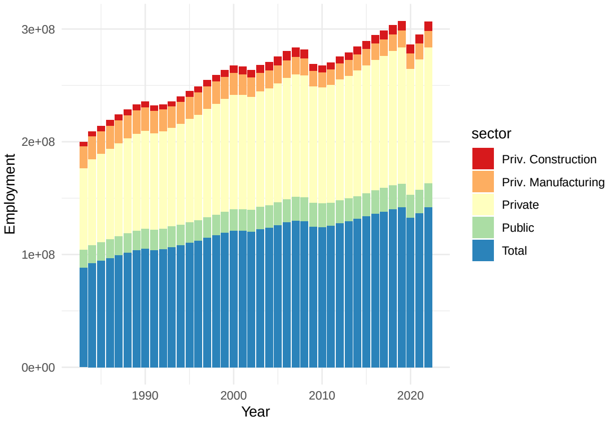

Figure 4.7: Setor Employement

The private sector—especially services—has grown a lot, while manufacturing has shrunk. We need unions in all these sectors to keep fighting for fair pay and worker rights, no matter what kind of work people are doing.

Chapter 5

Discussion and Summary

5.1 Wage Growth and Union Influence

Une thing is crystal clear: union workers make more money. The graphs and numbers show that if you're in a union, you're getting paid better compared to folks who aren't. We're talking about both the raw numbers (the difference in pay) and the adjusted numbers (where we consider factors like age, job type, etc.). And guess what? Even when you adjust for those things, union workers still come out on top.

When we looked at the trends over time, union wages have consistently been higher. Sure, there's been some narrowing of the gap lately, but unions are still the best bet for fair wages. If unions start losing power, everyone's pay could start sliding back. We need to keep unions strong if we want to keep fighting for better wages for all workers.

5.2 Decline in Union Membership

One of the more worrying things we saw is how union membership has been dropping for the past couple of decades. There are a bunch of reasons for this. The economy has shifted—manufacturing, where unions were strong, has shrunk, and we've moved towards more service jobs, which are harder to organize. Plus, with globalization, a lot of jobs went overseas or became automated, and that took away a lot of union power.

And let's not forget the gig economy. More and more people are working gig jobs or part-time, which makes it tough for them to organize and join unions. Fewer unions mean less power for workers to negotiate fair pay and benefits, and that's not good for anyone. If we want to turn this around, we need to make it easier for workers in all kinds of jobs to join unions and fight for their rights.

5.3 Shifts in Employment by Sector

Another big trend we saw is that employment has shifted from industries like manufacturing and construction into the service sector. The service industry has been booming, but it hasn't been easy for unions to gain a foothold there. That means a lot of workers in service jobs are missing out on the protections and pay boosts that unions can provide.

The public sector is one of the few places where unions are still doing okay, but even there, unions are under pressure from budget cuts and changes to policies that aim to cut spending. If we want to improve job quality for everyone, unions need to find ways to grow in these new industries. The service sector needs unions more than ever to help protect workers' rights and improve their pay.

5.4 Implications for the Future

So, what does all this mean for the future? Well, we've got some challenges, but also opportunities. Unions are still crucial for securing better wages, but the decline in membership is a problem we need to solve. The answer might be new organizing strategies—unions need to adapt to reach workers in tech, service, and gig industries. If unions can innovate, they can still be a powerful force for worker rights.

If union membership keeps shrinking, we're going to see more income inequality. That means the gap between the rich and the poor will get wider, and that's bad news for everyone. Unions help level the playing field, so we need to support them, whether through better labor laws or by organizing in new industries.

The shift to more service-based jobs could be an opportunity if unions can step in and start organizing those workers. The future of the labor market is going to depend a lot on how well unions, workers, and policymakers can adapt to these changes. We need to work together to make sure everyone gets fair wages and decent working conditions, no matter where they work.

Unions have always been about making things fair for workers, and that mission hasn't changed. We need to keep building union power to make sure everyone gets a fair shake in this rapidly changing economy.

UNIONS MATTER!

Your take away from this should be that the more of us who join unions, the better all of our wages will be!

Chapter 6

Appendix

6.1 Additional Code

In this appendix, we provide additional R code that was used throughout the analysis but not directly referenced in earlier chapters. This section includes helper functions, data transformations, and any other relevant scripts that may be helpful for those wishing to replicate or extend the analysis.

These additional scripts include: - Code to handle missing data. - Scripts for creating specific subsets of the data that were not covered in the main analysis but may provide additional insights. - Custom ggplot themes that were used for aesthetic consistency across all visualizations.

6.1.1 Handling Missing Data

In some cases, the datasets included missing values that needed to be handled to avoid skewing the results. Below is the R code used for addressing these missing values by imputing where appropriate or removing incomplete cases:

```
filtered_wages <-
  filtered_wages[!is.na(filtered_wages$union_wage) &
                 !is.na(filtered_wages$nonunion_wage), ]

filtered_demographics[is.na(filtered_demographics$p_members),
                      p_members := mean(p_members, na.rm = TRUE)]
```

These steps ensured that the datasets used for visualization and analysis were complete and representative, reducing potential biases caused by missing information.

6.1.2 Creating Data Subsets

To further explore wage differences across different demographic groups, subsets of the data were created based on gender and race. These subsets allowed for a more detailed examination of how union membership impacted specific groups differently.

```
female_workers <- filtered_demographics[facet == "female"]

black_workers <- filtered_demographics[facet == "black"]
```

These subsets were instrumental in exploring intersectional wage disparities and understanding how union membership benefits might vary among different demographic groups.

6.2 Data Source Links

For transparency and reproducibility, below are the direct links to the original datasets used in the analysis. Researchers and readers are encouraged to access these datasets to validate the findings or to conduct further analyses.

- Demographics Dataset: This dataset includes information on union membership, employment, and demographic segmentation from 1973 to 2022.
- Wages Dataset: This dataset provides data on union and nonunion wages, wage caps, and wage premiums.
- States Dataset: This dataset offers a detailed state-level breakdown of union membership and employment metrics.

Providing these links ensures that the analysis is open and accessible, encouraging collaboration and further research. These datasets are rich in historical and demographic details, making them valuable resources for anyone interested in labor economics or union studies.

6.3 Tools and Packages Used

This analysis was conducted using the R programming language, leveraging several packages that were critical for data manipulation, analysis, and visualization:

- `data.table`: Used for efficient data manipulation and aggregation. Its fast operations are particularly useful for working with large datasets.
- `ggplot2`: A versatile package for data visualization that was used to create all the plots in this analysis. Its ability to produce highly customizable plots made it ideal for exploring trends in wages and union membership.
- `dplyr`: Used for data wrangling tasks such as filtering, summarizing, and joining datasets. The intuitive syntax of `dplyr` helped streamline these tasks, making the analysis more readable.

Each of these tools played an important role in the analysis workflow, contributing to the accuracy and clarity of the results.

6.4 Additional Reading and Resources

For readers interested in learning more about the topics covered in this analysis, here are some additional resources:

- **"The State of Working America" by Economic Policy Institute**: A comprehensive resource for understanding labor market trends in the U.S., including union membership and wage disparities.
- **"Labor Economics" by George J. Borjas**: This book provides a detailed introduction to labor economics, including the impact of unions on wages and employment.
- **TidyTuesday Project**: The datasets used in this analysis were sourced from TidyTuesday, a weekly data project aimed at promoting data literacy and practice within the R community. More information can be found at TidyTuesday GitHub.

These resources are recommended for readers who wish to deepen their understanding of labor economics, union influence, and data analysis techniques.

Bibliography

R for Data Science Online Learning Community. Tidytuesday: A weekly data project aimed at improving data wrangling and visualization skills, 2024. URL https://github.com/rfordatascience/tidytuesday. Accessed: 2024-11-17.