Union Membership Trends by Demographics and Industry: 1973–2024

By Neeka Lucas

01 Introduction

Research question
How has union membership differed between gender, sector, industry, and education from 1973 to 2024?
Motivation
Whistleblower report reveals DOGE may have accessed and exfiltrated sensitive union data (McLaughlin, 2025)
Context and Importance
- The DOGE breach and UCSC labor strikes - Tracking trends reveals who's at risk, and why protections matter

02 Data Wrangling

Data source: Union Membership and Coverage Database from the CPS (Unionstats.com)

Wrangling challenges: merged 11 datasets into 3 by variable (education, industry, sex), converted strings to numeric, removed N/A values, extracted relevant columns

Tools: R (tidyverse: readr, tibble)

03

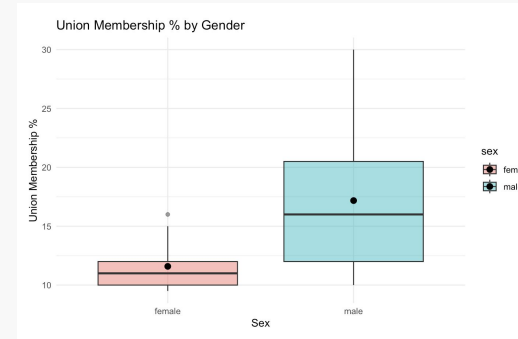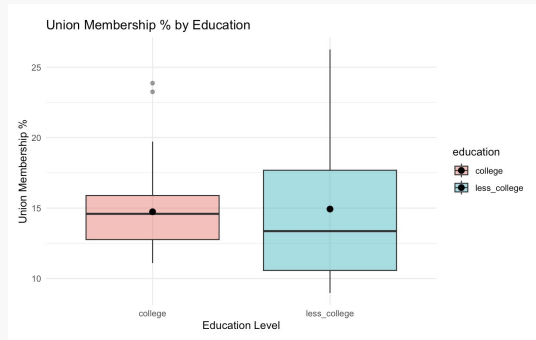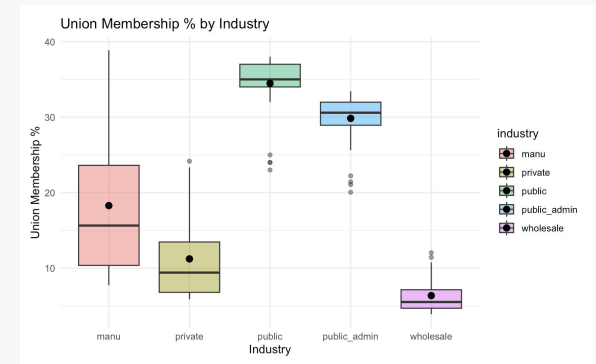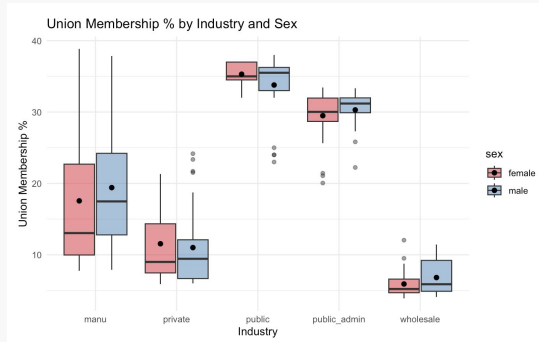# Exploratory Data Analysis (Project II)

Variables of Interest:

- year, perc_mem, per_coverage, industry, sex, education

Relationships Explored:

- trends in union membership over time
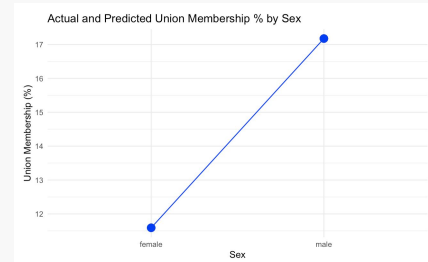- membership differences by industry, gender, and education

Exploratory Data Analysis (Project II)

04

Modeling (Project III)

Models used: logistic and linear regression

Predicted Probability of Union Membership by Industry

Actual and Predicted Union Membership % by Industry

Predicted Probability of Union Membership by Education

Actual and Predicted Union Membership % by Education

Predicted Probability of Union Membership by Sex

Actual and Predicted Union Membership % by Sex
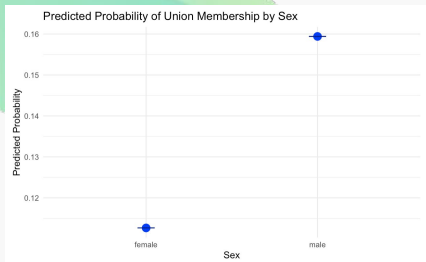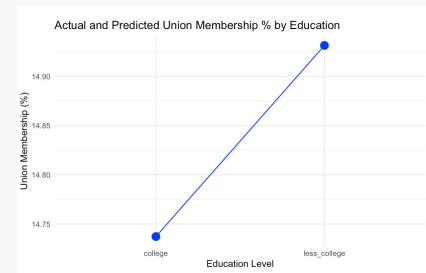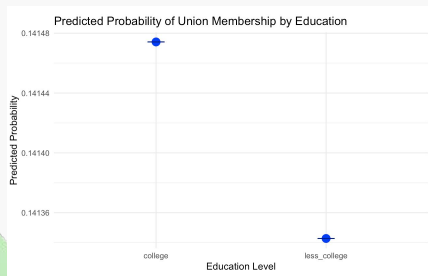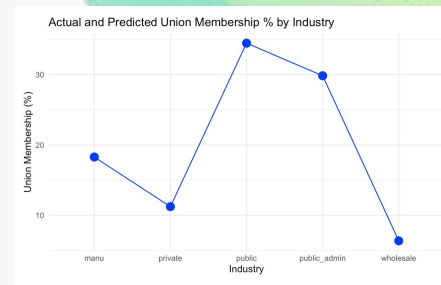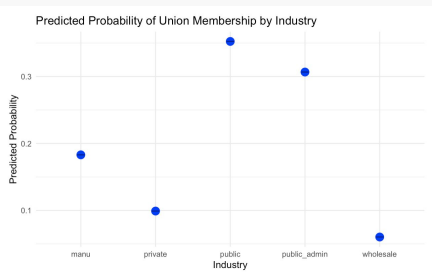
Key results and interpretation:

Logistic Regression

Education – Minimal difference, both groups had similar predicted probabilities (~14.1%).

Sex – Males had slightly higher predicted probabilities than females.

Industry – Public and public administration had the highest unionization, wholesale and private the lowest.

Linear Regression

Education – Slightly higher union membership % for less_college, difference was small.

Sex – Males showed higher average union membership %.

Industry – Public sectors had the highest %, wholesale and private the lowest.

05

Monte Carlo Simulation (Project IV)

Purpose and design:

- Simulate data from 10 distributions (gamma, beta, normal, binomial, Poisson, exponential, chi-squared, t, F, uniform) to evaluate how they affect linear and logistic regression models predicting union membership.

- Each distribution was simulated n = 100 times, using five predictors (industry, education, sex, sector, year). Models were assessed using bias of coefficients B_1, B_2, and mean squared error (MSE).

05 Monte Carlo Simulation (Project IV)

Summary of main findings:

- Linear regression performed best under normal, poisson, and uniform distributions, with low bias and MSE.
- Logistic regression was most accurate under normal and exponential distributions.

distribution <chr>	lin_bias_b1 <dbl>	lin_bias_b2 <dbl>	lin_mse <dbl>	lin_type1_error <dbl>	log_bias_b1 <dbl>	log_bias_b2 <dbl>	log_mse <dbl>	log_type1_error <dbl>
gamma	0.0105855475	−0.0191502994	0.9435197	0.04000000	0.20050971	−0.2338297	0.01137612	0.04333333
beta	0.0248115564	−0.0955204447	0.9491660	0.04666667	0.08826203	−0.3534083	0.01391430	0.05666667
normal	−0.0004480955	−0.0291448730	0.9370706	0.06000000	0.20009309	−0.2893938	0.01064469	0.05000000
binomial	−0.0102854024	0.0006369595	0.9219547	0.04333333	0.14573958	−0.1601078	0.01202230	0.04333333
poisson	−0.0016080369	0.0036393141	0.9408217	0.02666667	0.63108816	−0.7368962	0.01000439	0.06333333
exponential	0.0091110254	−0.0028943028	0.9425673	0.06000000	0.12822021	−0.3199380	0.01125477	0.04666667
chi_squared	0.0066217889	0.0044802072	0.9444686	0.03666667	1.16922703	−1.4223077	0.01040168	0.05666667
t	0.0141838883	0.0124566731	0.9559187	0.05333333	0.23472089	−0.2821159	0.01133908	0.05666667
F	0.0010898876	−0.0004642307	0.9403978	0.04666667	0.48201704	−0.7295598	0.01101808	0.03000000
uniform	0.0069140243	−0.0319284449	0.9251926	0.06666667	0.12896599	−0.1732095	0.01426070	0.04666667

06

Summary and Reflection

How has union membership differed between gender, sector, industry, and education (1973–2024)?

- Union membership was highest in the public sector and lowest in private and wholesale industries.
- Males consistently had higher membership rates than females.
- Education level had minimal impact, membership percentages were similar across levels.
- Modeling confirmed these patterns through both linear and logistic regression results.

06

Summary and Reflection

What I Learned from this full process:

- How to clean and merge complex datasets across multiple demographic variables.
- Gained skills in logistic and linear regression, and how to interpret bias, MSE, and Type I error.
- Learned to simulate data from multiple distributions to test models (Monte Carlo)

From the reproducibility test:

- the importance of clean, well-documented, and executable code so others can successfully reproduce results

06

Summary and Reflection

Ethical implications of your modeling choices - can your model results change decisions, affect people?

- Accurate models can inform policy and labor protections.
- Misinterpretation risks underrepresenting vulnerable groups.
- Transparent, reproducible methods help prevent harm.
- The DOGE breach shows why ethical data handling matters.

Thank You !

Any questions?

Work Cited

Hirsch, Barry T., David A. Macpherson, and William E. Even (2025). Union Membership, Coverage, and Earnings from the CPS. https://unionstats.com

Hirsch, B. T., Macpherson, D. A., & Even, W. E. (2025). Union Membership, Coverage, Density, and Employment: Manufacturing Workers[Data set]. Copyright by Barry T. Hirsch, David A. Macpherson, and William E. Even

Hirsch, B. T., Macpherson, D. A., & Even, W. E. (2025). Union Membership, Coverage, Density, and Employment: Wholesale and Retail Trade Workers[Data set]. Copyright by Barry T. Hirsch, David A. Macpherson, and William E. Even

Hirsch, B. T., Macpherson, D. A., & Even, W. E. (2025). Union Membership, Coverage, Density, and Employment: Public Administration[Data set]. Copyright by Barry T. Hirsch, David A. Macpherson, and William E. Even

Hirsch, B. T., Macpherson, D. A., & Even, W. E. (2025). Union Membership, Coverage, Density, and Employment: Private Sector Workers[Data set]. Copyright by Barry T. Hirsch, David A. Macpherson, and William E. Even

Hirsch, B. T., Macpherson, D. A., & Even, W. E. (2025). Union Membership, Coverage, Density, and Employment: Public Sector Workers[Data set]. Copyright by Barry T. Hirsch, David A. Macpherson, and William E. Even

Hirsch, B. T., Macpherson, D. A., & Even, W. E. (2025). Union Membership, Coverage, Density, and Employment: Workers with less than Bachelor's Degree[Data set]. Copyright by Barry T. Hirsch, David A. Macpherson, and William E. Even

Hirsch, B. T., Macpherson, D. A., & Even, W. E. (2025). Union Membership, Coverage, Density, and Employment: Workers with Bachelor's Degree or More[Data set]. Copyright by Barry T. Hirsch, David A. Macpherson, and William E. Even

Hirsch, B. T., Macpherson, D. A., & Even, W. E. (2025). Union Membership, Coverage, Density, and Employment: Male Workers [Data set]. Copyright by Barry T. Hirsch, David A. Macpherson, and William E. Even

Hirsch, B. T., Macpherson, D. A., & Even, W. E. (2025). Union Membership, Coverage, Density, and Employment: Female Workers[Data set]. Copyright by Barry T. Hirsch, David A. Macpherson, and William E. Even

Macpherson, D. A., & Hirsch, B. T. (2023). Five decades of CPS wages, methods, and union-nonunion wage gaps at unionstats.com. Industrial Relations: A Journal of Economy and Society, 62(4), 439–452. https://doi.org/10.1111/irel.12330

McLaughlin, J. (2025, April 15). A whistleblower's disclosure details how Doge may have taken sensitive labor data. NPR. https://www.npr.org/2025/04/15/nx-s1-5355896/doge-nlrb-elon-musk-spacex-security