Multifactorial Analysis of Demographic, Economic, and Educational Influences on U.S. Voting Patterns, 2016-2020

Xiang Fu
Faculty of Computing and Data Sciences, Boston University
xfu@bu.edu

CAS PO 399 / 599 Data Science for Politics, Summer 2024 **Boston, MA**

Research Question

Research Question

How did the relationship between county-level demographic factors, education levels, and unemployment rates and the Democratic vote share change between the 2016 and 2020 presidential elections?

Literature Review

Literature Review

Racial composition has been a strong predictor of voting behavior, with changing demographics linked to **shifts in party allegiance** (Abrajano & Hajnal, 2015; Frey, 2018)

Economic indicators, such as unemployment rates and median household income, have historically **influenced political preferences** (Wright, 2012; Gelman et al., 2010).

Educational attainment has become increasingly **predictive of voting behavior**, giving rise to a notable "diploma divide" in recent elections (Sides et al., 2018; Tyson \& Maniam, 2016).

Interaction between race and education has gained prominence in **predicting voting behavior**, while economic factors interact with demographic characteristics to influence political preferences (Schaffner et al., 2018; Bartels, 2016).

Hypotheses

Hypotheses

{H1} Counties with increasing racial diversity will show a positive change in Democratic vote share from 2016 to 2020.

{H2} The impact of education levels on Democratic vote share will vary across income groups, with a stronger positive relationship in higher-income counties.

{H3} Counties experiencing higher unemployment rates in 2016 will show a negative change in Democratic vote share in 2020, reflecting dissatisfaction with economic conditions.

{H4} There will be significant interaction effects between education levels and racial demographics, with the impact of education on Democratic vote share differing across racial groups.

Datasets

Datasets

{demographic_data} Racial composition and median household income

{education_data} Educational attainment for individuals 25 and older

{unemployment_2012, unemployment_2016} County-level unemployment rates for 2012 and 2016

{election_data} Democratic vote share in the 2016 and 2020 presidential elections

Results and Analysis

Hypothesis 1

Results

Impact of Racial Diversity on Democratic Vote Share Change

Diversity index shows a very slight positive relationship with changes in Democratic vote share

Relationship is **not statistically significant** ($p > 0.05$)

Confidence interval crosses zero, cannot rule out the possibility of no effect or even a slight negative effect

Do not have strong evidence to support H1

Racial diversity **alone** may **not be a reliable predictor** of changes in Democratic vote share at the county level

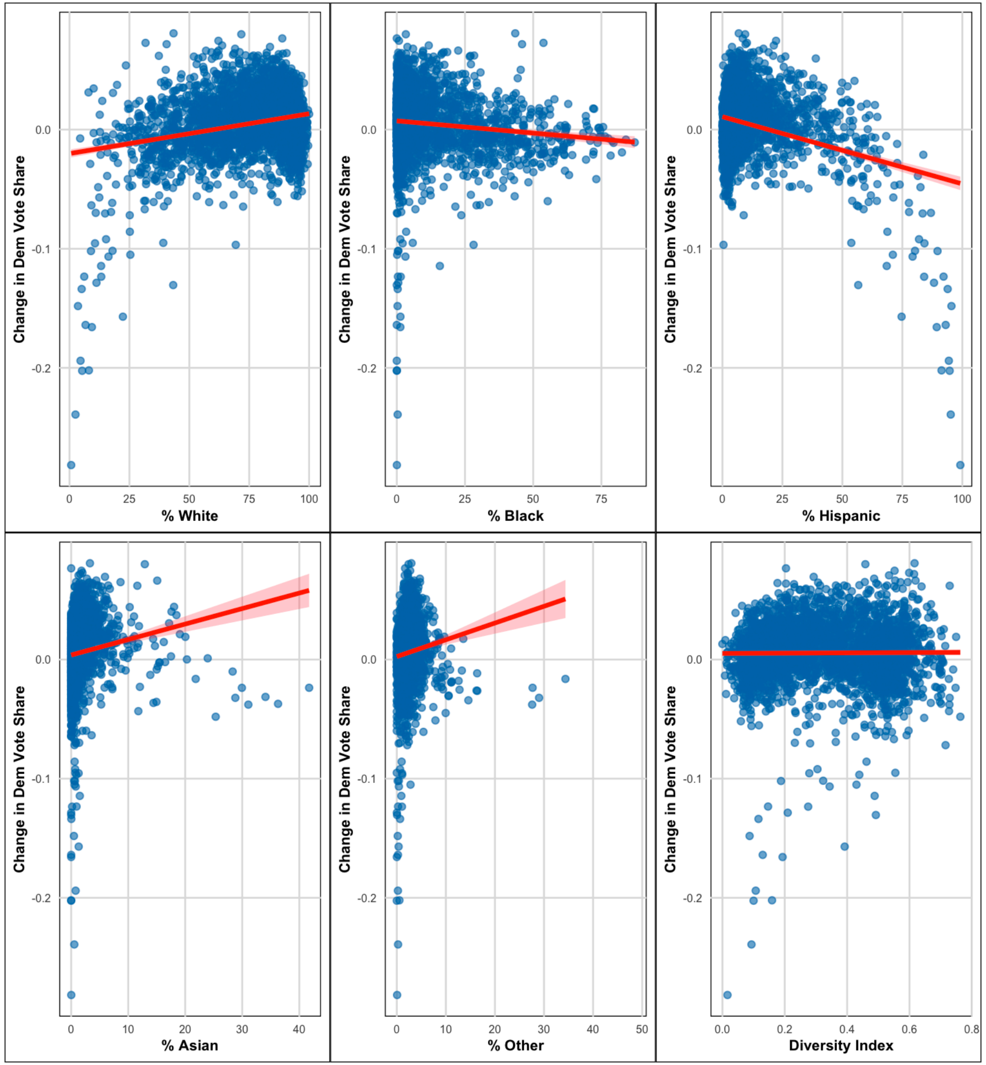
Coefficient	Estimate	Std. Error	CI Lower	CI Upper	P-Value
Estimate	0.001 199 228	0.002 536 833	−0.003 774 809	0.006 173 265	0.636

Results

Impact of Racial Diversity on Democratic Vote Share Change

Nearly flat line for the diversity index plot (bottom right)

While the overall diversity index shows little effect, the individual racial group plots reveal interesting patterns (e.g., positive trend for % White, negative for % Hispanic)



Results

Impact of Racial Diversity on Democratic Vote Share Change

X-axis: Percentage with Bachelor's Degree

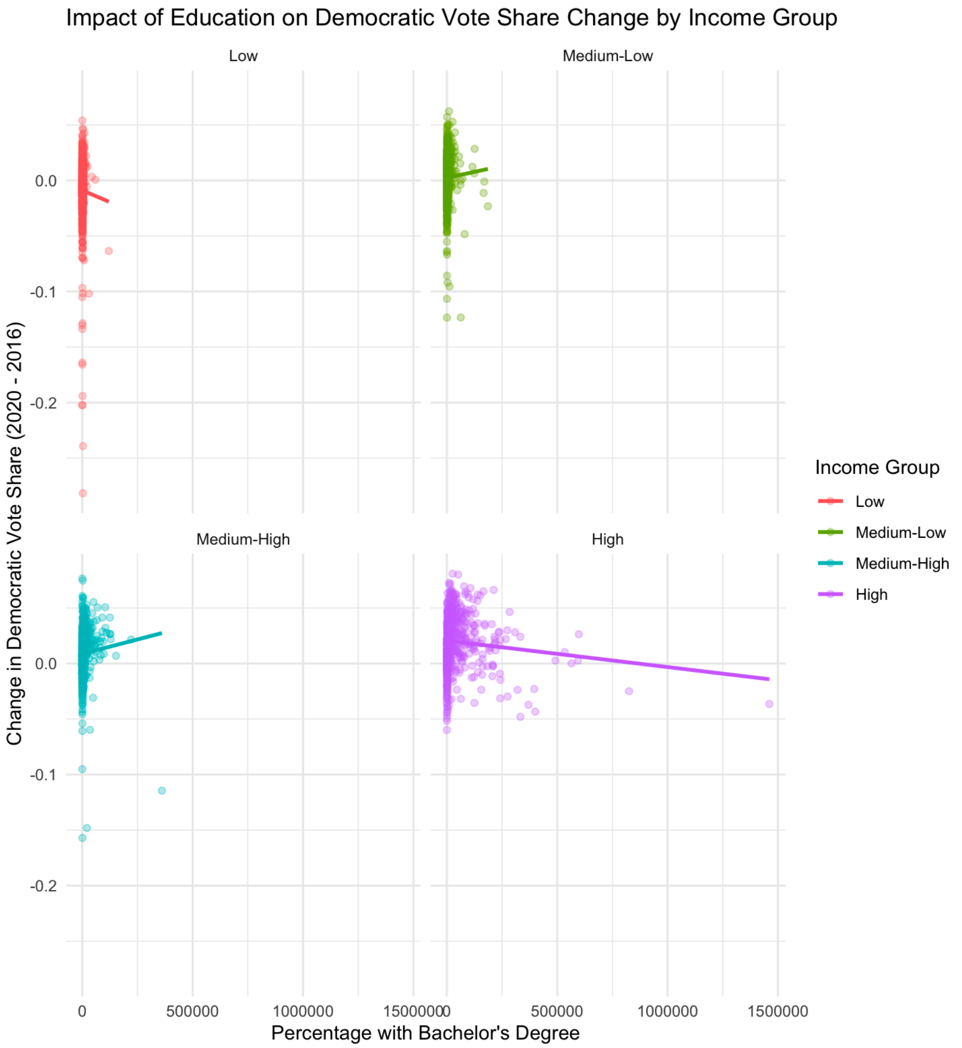
Y-axis: Change in Democratic Vote Share (2020 - 2016)

Varying relationships observed across income groups

Positive trend for Medium-Low income group

Negative trend for High income group

Mixed results for Low and Medium-High income groups



Hypothesis 2

Results

Interactions: Diversity, Education, and Income

Significant interactions between diversity and both education and income

Model explains 19.7% of variance in Democratic vote share change (R-squared: 0.197)

All predictors and interactions are statistically significant ($p < 0.001$)

Negative interaction terms suggest:

Diversity effect decreases as education/income increase
Education/income effects decrease as diversity increases

Variable	**Estimate**	**Std. Error**	**t value**	**p-value**
Intercept	-0.0467	0.0038	-12.363	$<2 \times 10^{-16}$
Diversity Index	0.0370	0.0088	4.217	2.54×10^{-5}
Bachelor's Degree	4.211×10^{-7}	5.590×10^{-8}	7.532	6.49×10^{-14}
Median Household Income	9.479×10^{-7}	7.075×10^{-8}	13.398	$<2 \times 10^{-16}$
Diversity Index:Bachelor's Degree	-6.996×10^{-7}	8.680×10^{-8}	-8.060	1.08×10^{-15}
Diversity Index:Median Household Income	-6.721×10^{-7}	1.616×10^{-7}	-4.160	3.27×10^{-5}

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

Residual standard error: 0.0231 on 3106 degrees of freedom

Multiple $R^2 = 0.197$, Adjusted $R^2 = 0.1957$

F-statistic: 152.4 on 5 and 3106 DF, p-value: $< 2.2 \times 10^{-16}$

Results

Impact of Education on Democratic Vote Share Change Across Income Groups

Low and High income groups show slight negative slopes

Medium-Low and Medium-High groups show slight positive slopes

Only the High income group shows a statistically significant relationship ($p < 0.01$)

Income Group	Intercept	Slope	p-value
Low	$-0.008\,87$	-8.54×10^{-8}	0.648
Medium-Low	$0.002\,34$	4.31×10^{-8}	0.377
Medium-High	$0.008\,73$	5.17×10^{-8}	0.108
High	0.0206	-2.38×10^{-8}	$0.007\,89$

Results

Impact of Education on Democratic Vote Share Change Across Income Groups

Interaction Model Summary

Significant differences in intercepts across income groups

Interaction terms (bachelor_degree:income_group) are not statistically significant

Income group itself has a significant effect on Democratic vote share change

Term	Estimate	Std. Error	t value	p-value
Intercept	-8.866×10^{-3}	8.948×10^{-4}	-9.908	$<2 \times 10^{-16}$
bachelor_degree	-8.545×10^{-8}	1.537×10^{-7}	-0.556	0.578
income_groupMedium-Low	1.120×10^{-2}	1.267×10^{-3}	8.841	$<2 \times 10^{-16}$
income_groupMedium-High	1.759×10^{-2}	1.275×10^{-3}	13.800	$<2 \times 10^{-16}$
income_groupHigh	2.946×10^{-2}	1.288×10^{-3}	22.872	$<2 \times 10^{-16}$
bachelor_degree:income_groupMedium-Low	1.285×10^{-7}	1.633×10^{-7}	0.787	0.431
bachelor_degree:income_groupMedium-High	1.372×10^{-7}	1.579×10^{-7}	0.869	0.385
bachelor_degree:income_groupHigh	6.166×10^{-8}	1.539×10^{-7}	0.401	0.689

Results

Impact of Education on Democratic Vote Share Change Across Income Groups

ANOVA Results

Both bachelor_degree and income_group are significant predictors

The interaction between bachelor_degree and income_group is not significant (p = 0.1398)

Income group explains more variance than education level

Term	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bachelor_degree	1	0.010 62	0.010 616	19.1081	1.276×10^{-5}
income_group	3	0.327 68	0.109 228	196.6098	$<2 \times 10^{-16}$
bachelor_degree:income_group	3	0.003 05	0.001 016	1.8283	0.1398
Residuals	3104	1.724 45	0.000 556		

Results

Impact of Education on Democratic Vote Share Change Across Income Groups

Correlation: Income and Education's Impact on Vote Share

Weak positive correlation between income and education's impact on Democratic vote share

Statistically significant, but small effect size

Suggests a slight tendency for education to have a stronger positive effect in higher-income areas

$$\text{Correlation} = 0.1716,\ p < 2.2 \times 10^{-16},\ 95\%\ \text{CI: } [0.137, 0.206]$$

Hypothesis 3

Results

Impact of Unemployment on Democratic Vote Share Change

2016 Unemployment Rate vs. Democratic Vote Share Change

Significant negative relationship ($p < 0.001$)

For every 1% increase in 2016 unemployment, Dem share decreased by 0.44%

R-squared: 0.09926 (9.93% of variance explained)

Highly significant F-statistic (p-value: 2.2e-16)

Variable	Estimate	Std. Error	t value	$Pr(>\|t\|)$
Intercept	0.0284	0.0013	21.65	$<2 \times 10^{-16}$
Unemployment Rate 2016	−0.0044	0.0002	−18.51	$<2 \times 10^{-16}$

R-squared: 0.09926, Adjusted R-squared: 0.09897

F-statistic: 342.7 on 1 and 3110 DF, p-value: ¡ 2.2e-16

Results

Impact of Unemployment on Democratic Vote Share Change

Change in Unemployment Rate (2012-2016) vs. Democratic Vote Share Change

Weak negative relationship ($p < 0.05$)

For every 1% increase in unemployment change, Dem share decreased by 0.05%

Model explains only 0.13% of the variance in Dem vote share change

Variable	Estimate	Std. Error	t value	$Pr(>\|t\|)$
Intercept	0.0042	0.0008	5.172	2.47×10^{-7}
Change in Unemployment Rate	−0.0005	0.0003	−2.036	0.0418

R-squared: 0.001331, Adjusted R-squared: 0.00101

F-statistic: 4.146 on 1 and 3110 DF, p-value: 0.04182

Results

Impact of Unemployment on Democratic Vote Share Change

Correlation: 2016 Unemployment and Democratic Vote Share Change

Correlation coefficient: -0.3150572

Moderate negative correlation

Highly significant ($p < 2.2e-16$)

95% CI: [-0.3463595, -0.2830541]

Statistic	Value
Correlation coefficient	-0.3151
95% CI lower bound	-0.3464
95% CI upper bound	-0.2831
p-value	¡ 2.2e-16

Results

Unemployment Rate Changes by County (2012-2016)

County-level map of unemployment rate changes

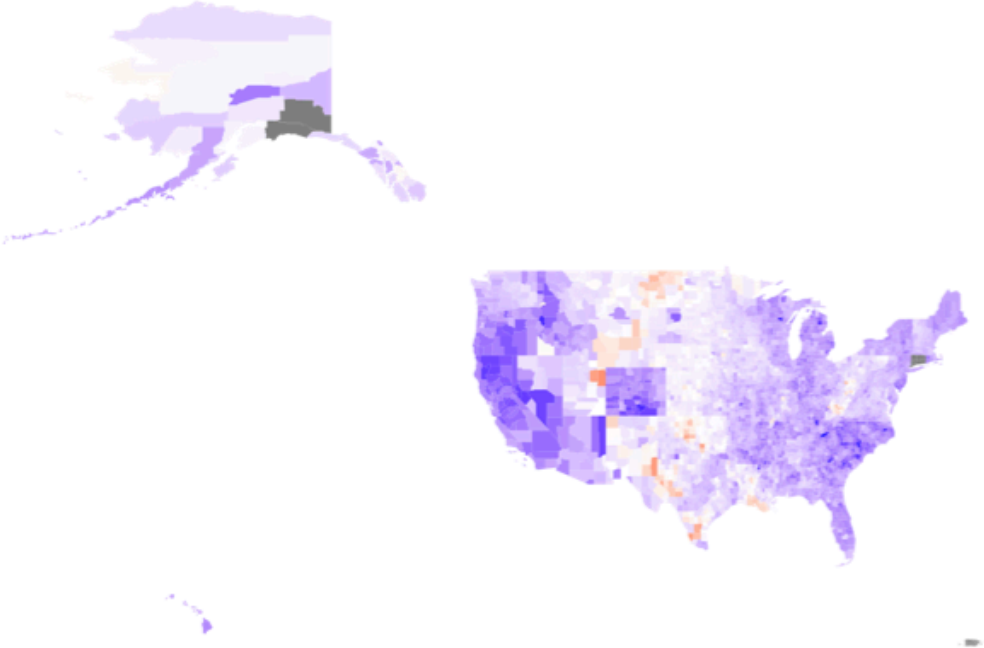
Blue areas: Decreased unemployment

Red areas: Increased unemployment

Significant regional variations evident

Midwest shows improvement

Energy-producing regions faced challenges



Results

Impact of Unemployment on Democratic Vote Share Change

Regional Variations in Unemployment Impact

Significant regional differences in unemployment effect

Strongest negative effect in the Northeast

Weakest effect in the West

Model explains 16.28% of the variance in Dem vote share change

Variable	**Estimate**	**Std. Error**	**t value**	**Pr(> |t|)**
Intercept	0.0188	0.0023	8.336	$<2 \times 10^{-16}$
Unemployment Rate 2016	−0.0029	0.0005	−6.019	1.96×10^{-9}
Region Northeast	0.0357	0.0070	5.076	4.09×10^{-7}
Region South	0.0127	0.0031	4.149	3.43×10^{-5}
Region West	0.0132	0.0037	3.615	0.000 305
Unemployment: Northeast	−0.0040	0.0014	−2.896	0.003 807
Unemployment: South	−0.0027	0.0006	−4.590	4.60×10^{-6}
Unemployment: West	−0.0002	0.0007	−0.317	0.750 940

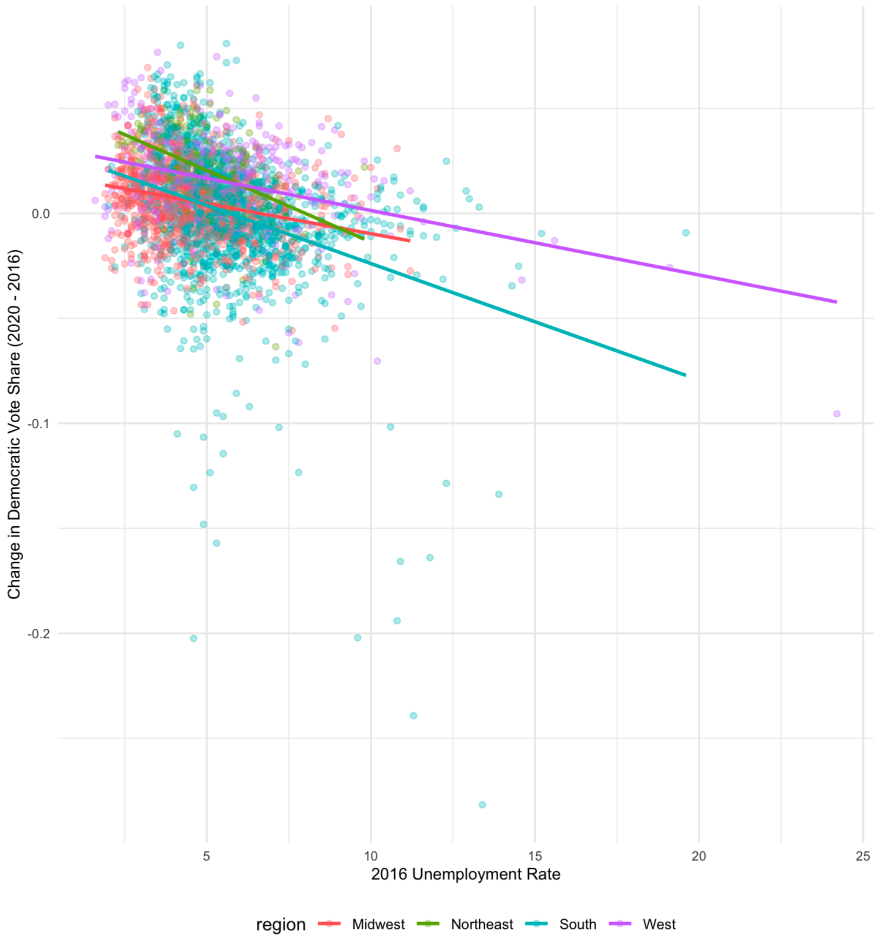
R-squared: 0.1628, Adjusted R-squared: 0.1609

F-statistic: 86.25 on 7 and 3104 DF, p-value: ¡ 2.2e-16

Results

Impact of Unemployment on Democratic Vote Share Change

Unemployment Rate vs. Democratic Vote Share Change by Region



Hypothesis 4

Results

Main Interaction Model Results

All variables and interactions are statistically significant ($p < 0.05$)

Negative main effects for bachelor's degree and all racial/ethnic percentages

Positive interaction effects between bachelor's degree and racial/ethnic percentages

Strongest negative main effect: Hispanic percentage (-8.944 x 10^-4)

Strongest positive interaction: Bachelor's degree and Hispanic percentage (6.905 x 10^-9)

Intercept is positive and significant (3.870 x 10^-2)

Variable	**Estimate**	**Std. Error**	**t value**	**Pr(> |t|)**
Intercept	3.870×10^{-2}	5.417×10^{-3}	7.144	1.12×10^{-12}
bachelor_degree	-7.639×10^{-7}	9.095×10^{-8}	−8.399	$<2 \times 10^{-16}$
pct_white	-2.965×10^{-4}	5.679×10^{-5}	−5.220	1.90×10^{-7}
pct_black	-5.419×10^{-4}	6.149×10^{-5}	−8.813	$<2 \times 10^{-16}$
pct_hispanic	-8.944×10^{-4}	6.447×10^{-5}	−13.874	$<2 \times 10^{-16}$
bachelor_degree:pct_white	1.304×10^{-8}	1.085×10^{-9}	12.016	$<2 \times 10^{-16}$
bachelor_degree:pct_black	5.005×10^{-9}	1.146×10^{-9}	4.366	1.31×10^{-5}
bachelor_degree:pct_hispanic	6.905×10^{-9}	1.195×10^{-9}	5.779	8.27×10^{-9}

Results
Robust Standard Errors

All variables remain statistically significant ($p < 0.05$)

Main effects still negative for bachelor's degree and racial/ethnic percentages

Interaction effects remain positive

Strongest negative main effect: Hispanic percentage (-8.944 x 10^-4)
Strongest positive interaction:

Bachelor's degree and White percentage (1.304 x 10^-8)

Some changes in significance levels, but overall conclusions hold

Variable	**Estimate**		**Std. Error**		**t value**	**Pr(> |t|)**
Intercept	3.870	$\times 10^{-2}$	6.710	$\times 10^{-3}$	5.767	8.852×10^{-9}
bachelor_degree	−7.639	$\times 10^{-7}$	1.310	$\times 10^{-7}$	−5.831	6.061×10^{-9}
pct_white	−2.965	$\times 10^{-4}$	6.868	$\times 10^{-5}$	−4.316	1.636×10^{-5}
pct_black	−5.419	$\times 10^{-4}$	7.037	$\times 10^{-5}$	−7.701	1.806×10^{-14}
pct_hispanic	−8.944	$\times 10^{-4}$	1.146	$\times 10^{-4}$	−7.807	7.971×10^{-15}
bachelor_degree:pct_white	1.304	$\times 10^{-8}$	1.560	$\times 10^{-9}$	8.358	$<2 \times 10^{-16}$
bachelor_degree:pct_black	5.005	$\times 10^{-9}$	1.842	$\times 10^{-9}$	2.717	0.006 621
bachelor_degree:pct_hispanic	6.905	$\times 10^{-9}$	1.914	$\times 10^{-9}$	3.608	0.000 313

Results

ANOVA Results - Model Comparison

Bachelor's degree: Significant ($F = 19.1081$, $p < 0.0001$)

Income group: Highly significant ($F = 196.6098$, $p < 2e-16$)

Interaction (bachelor's degree:income group): Not significant ($F = 1.8283$, $p = 0.1398$)

Income group explains the most variance (Sum Sq = 0.32768)

Bachelor's degree contributes less (Sum Sq = 0.01062)

Interaction term adds minimal explanatory power

Term	Df	Sum Sq	Mean Sq	F value	Pr(>F)
bachelor_degree	1	0.010 62	0.010 616	19.1081	1.276×10^{-5}
income_group	3	0.327 68	0.109 228	196.6098	$<2 \times 10^{-16}$
bachelor_degree:income_group	3	0.003 05	0.001 016	1.8283	0.1398
Residuals	3104	1.724 45	0.000 556		

Results
Standardized Coefficients Model

All variables statistically significant ($p < 0.05$)

Bachelor's degree: Strong positive main effect (0.017140)

Racial/ethnic percentages: Negative main effects

Strongest: Hispanic (-0.011071)
Followed by: Black (-0.006815), White (-0.002331)

Positive interaction effects between bachelor's degree and racial/ethnic percentages

Strongest: White (0.013283)
Followed by: Hispanic (0.004824), Black (0.003639)

Variable	Estimate	Std. Error	t value	$\Pr(>\lvert t \rvert)$
Intercept	0.007 559 2	0.000 429 4	17.604	$<2 \times 10^{-16}$
bachelor_degree	0.017 140 0	0.000 946 4	18.111	$<2 \times 10^{-16}$
pct_white	−0.002 331 1	0.001 123 2	−2.075	0.038
pct_black	−0.006 814 9	0.000 861 0	−7.915	3.40×10^{-15}
pct_hispanic	−0.011 070 8	0.000 863 0	−12.828	$<2 \times 10^{-16}$
bachelor_degree:pct_white	0.013 283 1	0.001 105 5	12.016	$<2 \times 10^{-16}$
bachelor_degree:pct_black	0.003 638 8	0.000 833 4	4.366	1.31×10^{-5}
bachelor_degree:pct_hispanic	0.004 823 9	0.000 834 8	5.779	8.27×10^{-9}

Results

Models for Majority White/ Black/Hispanic Counties

Majority White Counties:

Positive intercept
Positive effect of bachelor's degree
Both highly significant

Majority Black Counties:

Negative intercept
Positive effect of bachelor's degree
Both significant

Majority Hispanic Counties:

Negative intercept
Positive effect of bachelor's degree
Intercept significant, bachelor's degree not significant

Variable	**Estimate**	**Std. Error**	**t value**	**Pr(> |t|)**
Majority White Counties				
Intercept	6.260×10^{-3}	4.097×10^{-4}	15.28	$<2 \times 10^{-16}$
bachelor_degree	1.916×10^{-7}	1.308×10^{-8}	14.65	$<2 \times 10^{-16}$
Majority Black Counties				
Intercept	-9.135×10^{-3}	1.822×10^{-3}	−5.013	2.51×10^{-6}
bachelor_degree	1.999×10^{-7}	7.618×10^{-8}	2.624	0.0101
Majority Hispanic Counties				
Intercept	-4.956×10^{-2}	6.476×10^{-3}	−7.653	1.32×10^{-11}
bachelor_degree	3.497×10^{-8}	1.285×10^{-7}	0.272	0.786

Results
Interaction Effects of Education and Racial Demographics

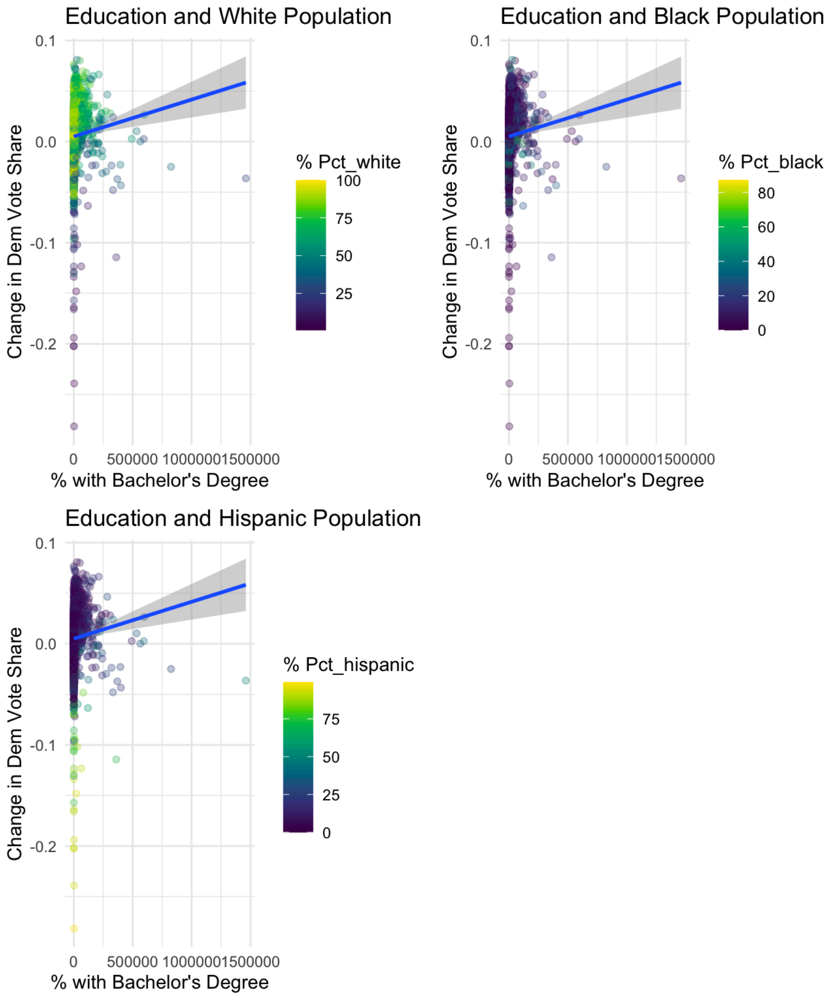
Positive relationship between education and Democratic vote share across all racial groups

Education and White population Hispanic population shows steeper positive slope than Black population

Higher variability in vote share changes for counties with lower education levels

Counties with high % of White population and low education show largest negative changes

Most positive changes occur in highly educated counties across all racial groups



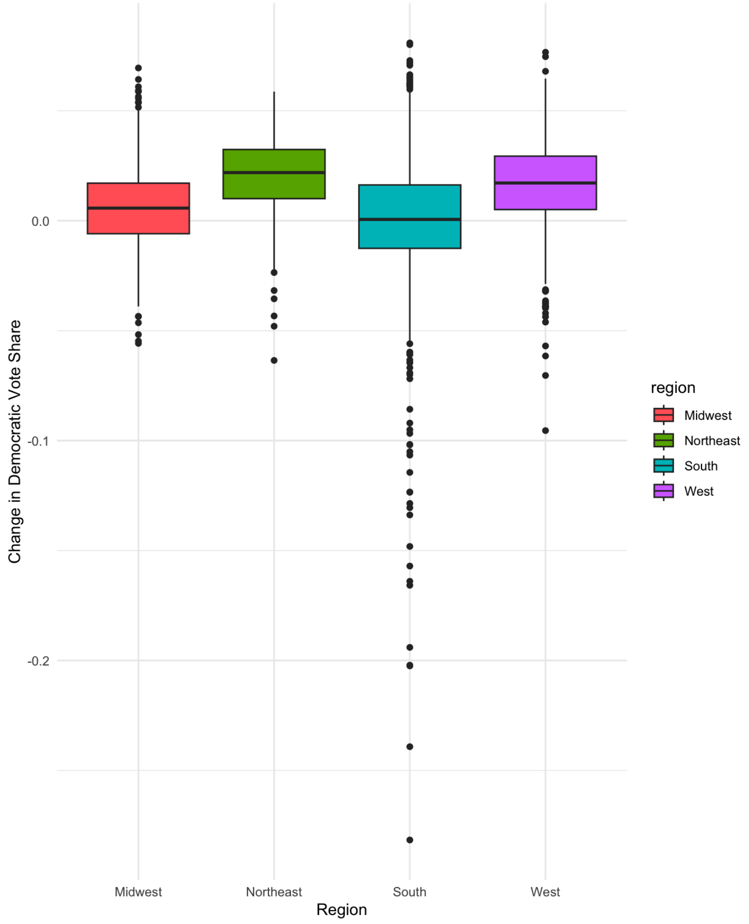
Results
Change in Democratic Vote Share by Region

West shows the largest spread in vote share changes

Northeast has the highest median increase in Democratic vote share

South shows the smallest median change

Midwest appears to have a slight negative median change



Results
Regional Differences in Key Variable Effects

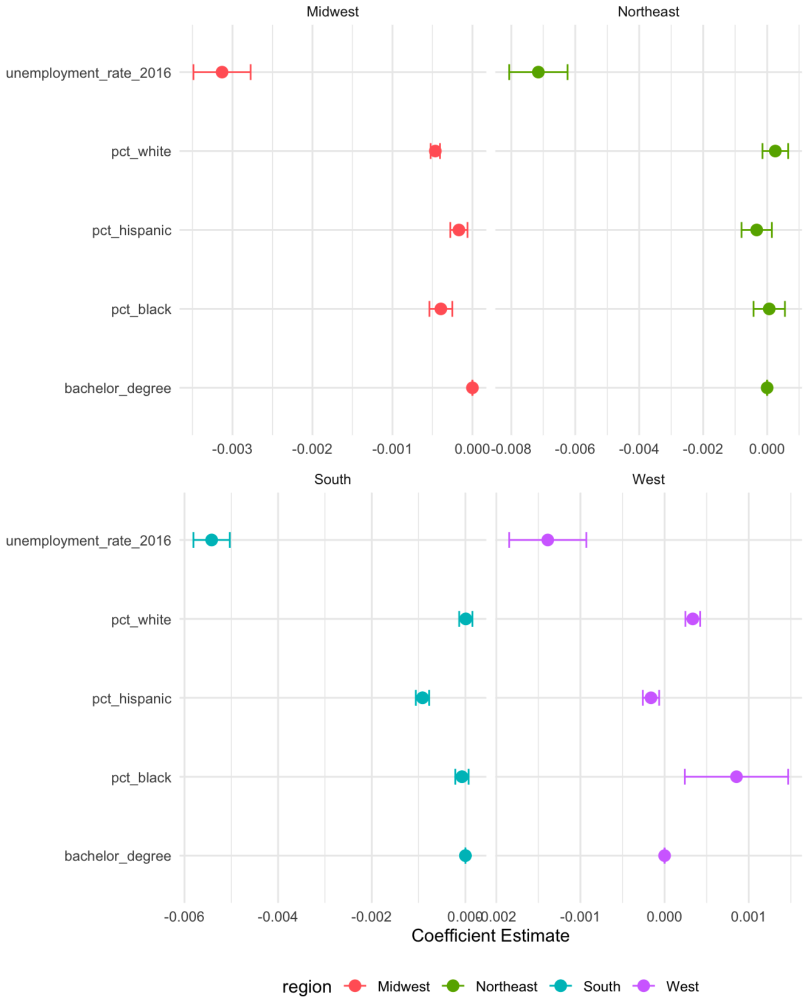
Unemployment rate consistently negative across all regions

Education (bachelor_degree) effect varies by region: Positive in Midwest and South, Negative in Northeast and West

Racial demographics show varied effects: White percentage: negative in Midwest, positive in West

Black percentage: negative in Midwest, positive in West

Hispanic percentage: negative across all regions, strongest in South



Results

Average Change in Democratic Vote Share by State

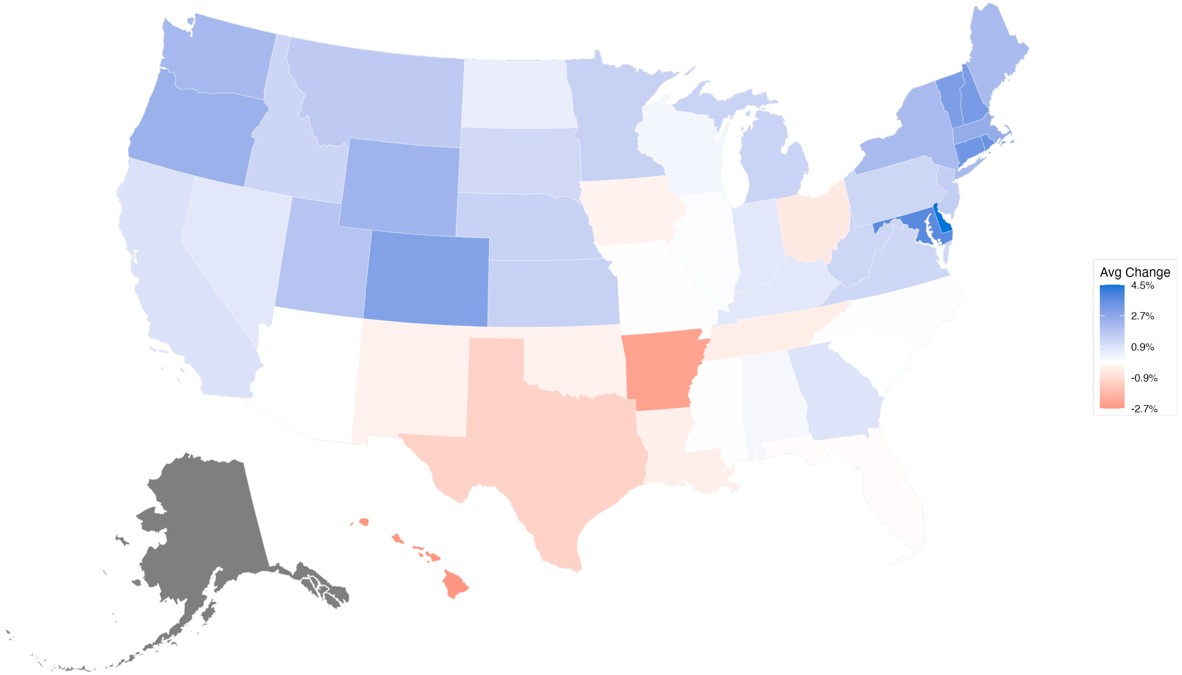
Northeastern states show mostly positive changes

Many Southern states show negative changes

Western states show a mix of positive and negative changes

Midwest states generally show small changes

Range of changes: -2.75% to +4.54%



Results
Regional Model Results

Significant regional variations in effects of variables

Unemployment consistently negative across all regions

Education (bachelor's degree) shows mixed effects

Racial demographics have varying impacts by region

Region	Variable	Estimate	Std. Error	P-value
Midwest	pct_white	$-0.000\,464$	$0.000\,058\,5$	5.37×10^{-15}
	pct_black	$-0.000\,395$	$0.000\,143$	5.93×10^{-3}
	pct_hispanic	$-0.000\,169$	$0.000\,107$	1.17×10^{-1}
	bachelor_degree	5.22	1.76	2.98×10^{-3}
	unemployment_rate_2016	$-0.003\,13$	$0.000\,357$	7.89×10^{-18}
Northeast	pct_white	$0.000\,255$	$0.000\,402$	5.26×10^{-1}
	pct_black	$0.000\,065\,0$	$0.000\,491$	8.95×10^{-1}
	pct_hispanic	$-0.000\,326$	$0.000\,473$	4.91×10^{-1}
	bachelor_degree	-4.27	2.50	8.95×10^{-2}
	unemployment_rate_2016	$-0.007\,15$	$0.000\,912$	2.10×10^{-13}
South	pct_white	8.81	$0.000\,141$	9.50×10^{-1}
	pct_black	-7.29	$0.000\,142$	6.08×10^{-1}
	pct_hispanic	$-0.000\,916$	$0.000\,143$	1.81×10^{-10}
	bachelor_degree	1.66	1.99	1.49×10^{-16}
	unemployment_rate_2016	$-0.005\,42$	$0.000\,388$	1.33×10^{-41}
West	pct_white	$0.000\,335$	$0.000\,087\,7$	1.56×10^{-4}
	pct_black	$0.000\,854$	$0.000\,614$	1.65×10^{-1}
	pct_hispanic	$-0.000\,161$	$0.000\,097\,1$	9.72×10^{-2}
	bachelor_degree	-1.89	1.12	9.24×10^{-2}
	unemployment_rate_2016	$-0.001\,39$	$0.000\,458$	2.60×10^{-3}

Results
Interaction Effects between Variables and Regions

Key variables: race/ethnicity, education, unemployment, geographic region

Statistically significant interactions found for several variable-region pairs

Largest effects seen for racial composition and unemployment across regions

Education (bachelor's degree) shows varying impact by region

Some interactions not statistically significant (e.g. Hispanic population in Northeast/West)

Variable	**Estimate**	**Std. Error**	**P-value**
Intercept	6.224×10^{-2}	7.528×10^{-3}	$<2 \times 10^{-16}$
pct_white	-4.640×10^{-4}	7.254×10^{-5}	1.82×10^{-10}
pct_black	-3.947×10^{-4}	1.775×10^{-4}	0.026 28
pct_hispanic	-1.688×10^{-4}	1.333×10^{-4}	0.205 52
bachelor_degree	5.224×10^{-8}	2.177×10^{-8}	0.016 46
unemployment_rate_2016	-3.131×10^{-3}	4.434×10^{-4}	2.04×10^{-12}
regionNortheast	-2.369×10^{-2}	5.298×10^{-2}	0.654 87
regionSouth	-2.278×10^{-2}	1.371×10^{-2}	0.096 64
regionWest	-6.075×10^{-2}	1.219×10^{-2}	6.57×10^{-7}
pct_white:regionNortheast	7.192×10^{-4}	5.574×10^{-4}	0.197 06
pct_white:regionSouth	4.729×10^{-4}	1.406×10^{-4}	0.000 78
pct_white:regionWest	7.990×10^{-4}	1.201×10^{-4}	3.38×10^{-11}
pct_black:regionNortheast	4.596×10^{-4}	6.976×10^{-4}	0.510 01
pct_black:regionSouth	3.218×10^{-4}	2.149×10^{-4}	0.134 43
pct_black:regionWest	1.249×10^{-3}	6.929×10^{-4}	0.071 52
pct_hispanic:regionNortheast	-1.576×10^{-4}	6.643×10^{-4}	0.812 47
pct_hispanic:regionSouth	-7.472×10^{-4}	1.804×10^{-4}	3.54×10^{-5}
pct_hispanic:regionWest	7.330×10^{-6}	1.703×10^{-4}	0.965 66
bachelor_degree:regionNortheast	-9.491×10^{-8}	4.070×10^{-8}	0.019 76
bachelor_degree:regionSouth	1.138×10^{-7}	2.758×10^{-8}	3.78×10^{-5}
bachelor_degree:regionWest	-7.113×10^{-8}	2.496×10^{-8}	0.004 41
unemployment_rate_2016:regionNortheast	-4.023×10^{-3}	1.329×10^{-3}	0.002 49
unemployment_rate_2016:regionSouth	-2.290×10^{-3}	5.533×10^{-4}	3.60×10^{-5}
unemployment_rate_2016:regionWest	1.742×10^{-3}	6.681×10^{-4}	0.009 15

Robustness Checks

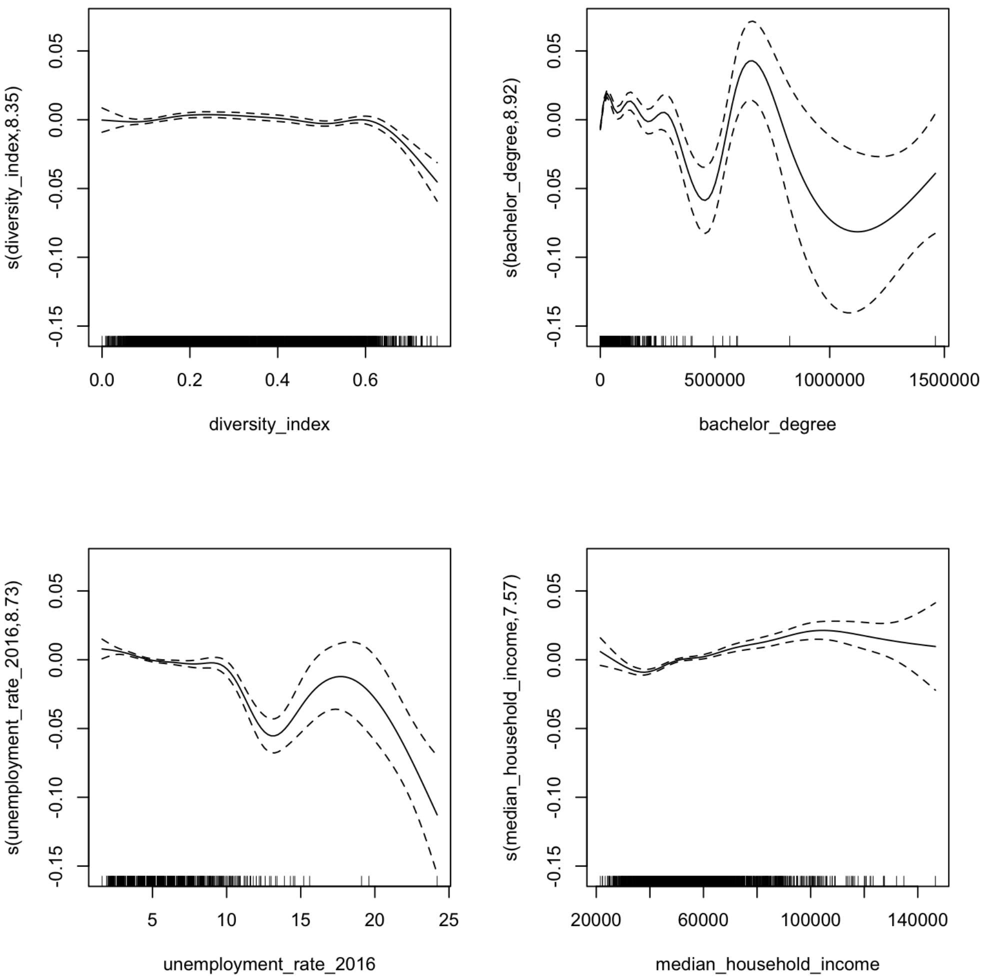
Results
Robustness Checks

Non-linear Effects on Change in Democratic Vote Share

All variables show **significant non-linear relationships** ($p < 0.001$)
Model explains 30.6% of deviance in Democratic vote share change
Adjusted R-squared: 0.299

Variable Effects:
Bachelor's Degree (edf: 8.920)
Unemployment Rate 2016 (edf: 8.728)
Diversity Index (edf: 8.350)
Median Household Income (edf: 7.572)

Simple linear models may not capture the full complexity of voting behavior



Random Forest & XGBoost Models

Results
Machine Learning Model Performance

Both Random Forest and XGBoost models show similar performance

RMSE of 0.0175 indicates average prediction error of 1.75 percentage points

R^2 of approximately 0.626 suggests models explain about 62.6% of variance

Random Forest slightly better in MAE, XGBoost marginally better in R^2

Model	RMSE	MAE	R^2
Random Forest	0.0175	0.0129	0.6261
XGBoost	0.0175	0.0131	0.6269

Results

Feature Importance Overview

Top 3 features by average rank:

Hispanic population % (1.5)
Bachelor's degree attainment (1.5)
Median household income (3.0)

Demographic factors (Hispanic, Black, Asian %) **highly influential**

Education (bachelor's degree) consistently important across models

Economic factor (median household income) ranks 3rd

Unemployment rate and diversity index less important overall

Feature	RF Importance	XGB Importance	RF Rank	XGB Rank	Avg Rank
pct_hispanic	34.900 03	0.286 488 46	2	1	1.5
bachelor_degree	39.879 16	0.140 505 87	1	2	1.5
median_household_income	30.999 48	0.130 975 93	3	3	3.0
pct_black	25.938 24	0.065 567 26	4	7	5.5
pct_asian	22.433 41	0.081 653 33	7	4	5.5
median_age	23.421 84	0.075 616 63	6	6	6.0
pct_white	19.909 63	0.079 338 65	8	5	6.5
unemployment_rate_2016	25.129 17	0.048 576 20	5	8	6.5
pct_native_american	13.532 72	0.048 078 73	10	9	9.5
diversity_index	15.462 51	0.043 198 93	9	10	9.5

Results

Partial Dependence Analysis: Diversity Index

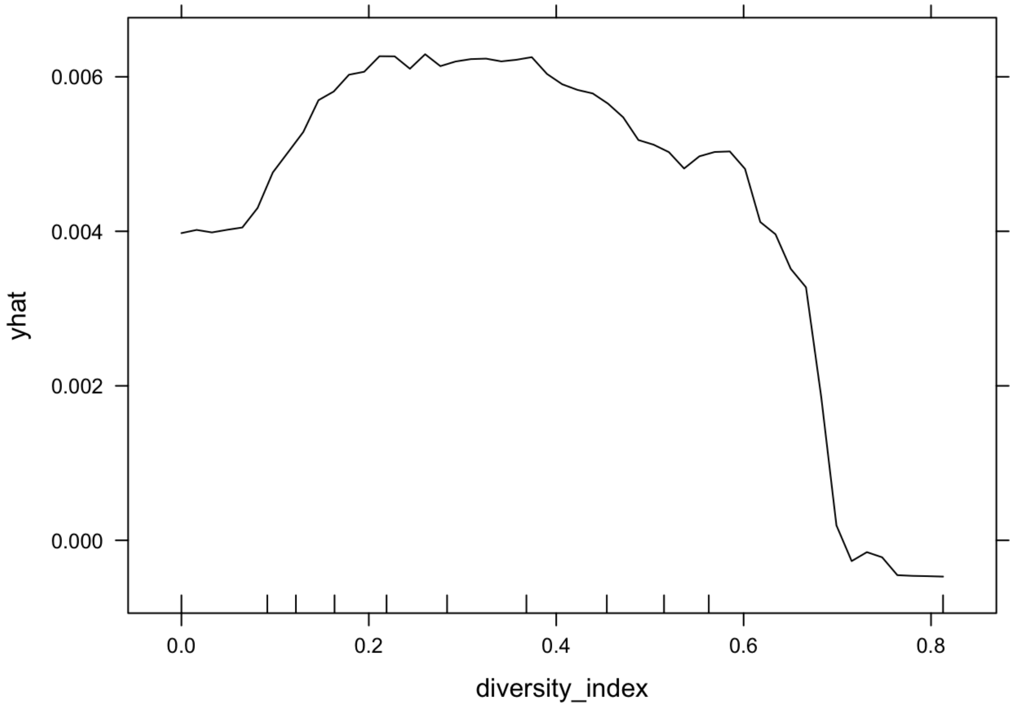
Non-linear, resembling an inverted U-shape

Initial increase from 0 to ~0.2 diversity index
Peak around 0.2-0.4 diversity index
Gradual decline from ~0.4 to 0.6
Sharp drop after 0.6 diversity index

Y-axis ('yhat') ranges from 0 to ~0.006

Largest positive effect at peak (0.006)

Effect becomes near-zero or slightly negative at high diversity indices (>0.7)

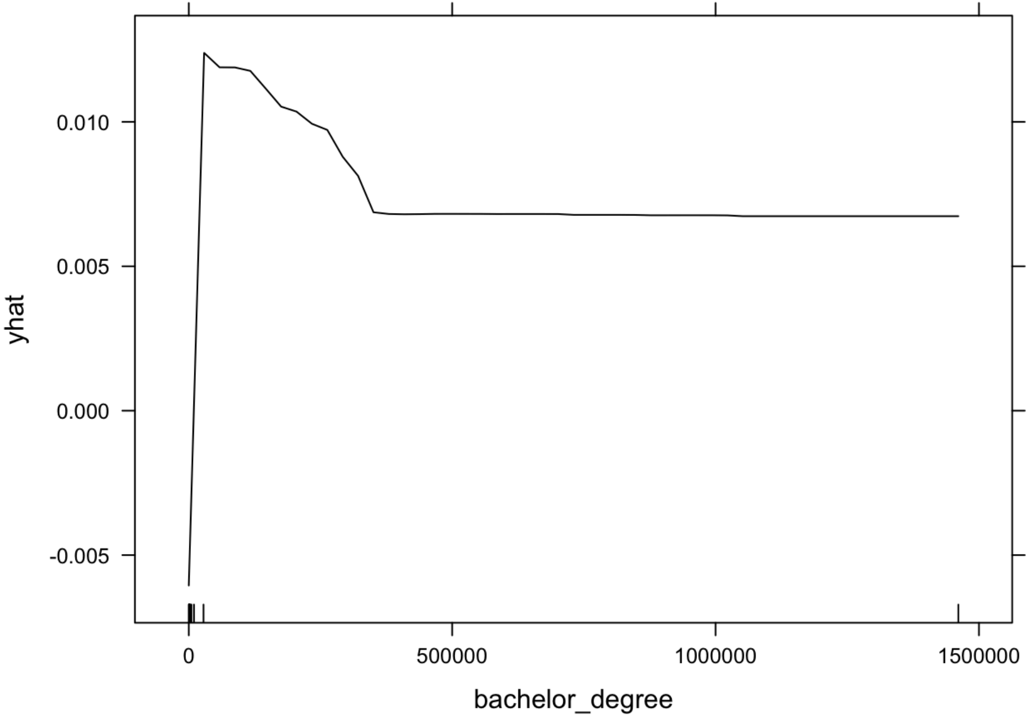


Results
Partial Dependence Analysis: Education

Sharp initial increase from negative to positive values
Followed by a gradual decline
Levels off to a constant value for higher bachelor degree numbers

Highly non-linear relationship
Saturation point occurs around 500,000 bachelor degrees
After this point, the effect remains constant regardless of further increases

Maximum 'yhat' value is about 0.012, higher than the peak for diversity index (0.006)
Range of effect is larger: from about -0.006 to 0.012, compared to 0 to 0.006 for diversity

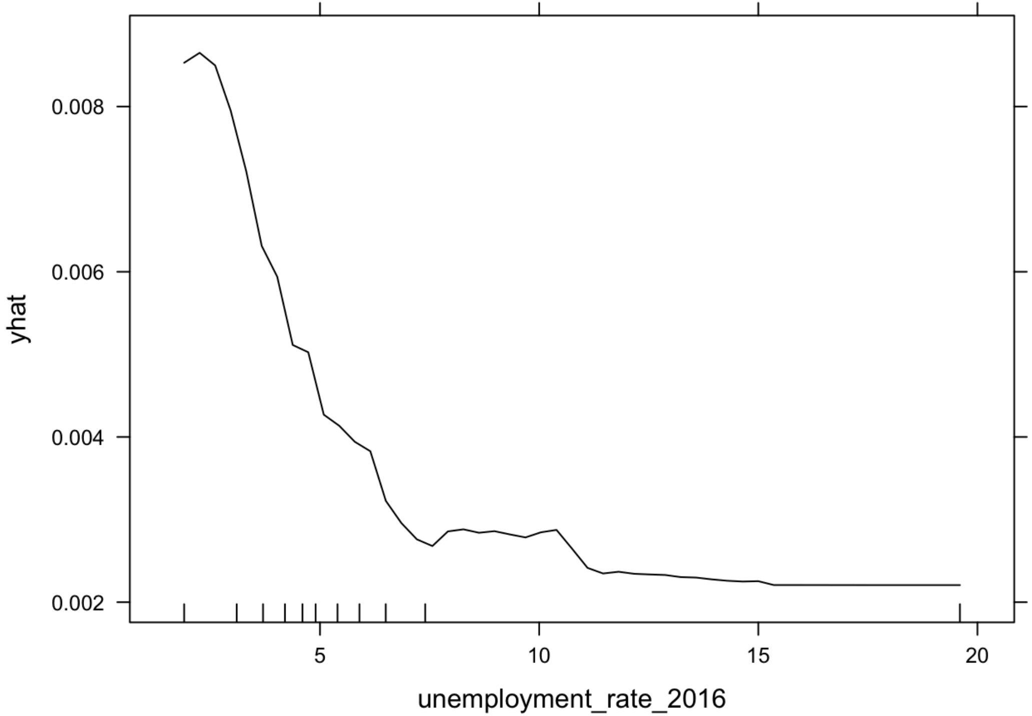


Results

Partial Dependence Analysis: Unemployment

Overall negative relationship
Non-linear, with a steep initial decline followed by a more gradual decrease
Approaches an asymptote for higher unemployment rates

Sharp decline from 0% to about 5% unemployment
Transition point around 5-7% where the slope becomes less steep
Minimal change in effect beyond ~10% unemployment
Maximum 'yhat' value around 0.009, lower than bachelor degree effect (0.012) but higher than diversity index peak (0.006)
Range of effect is approximately 0.002 to 0.009, smaller than



Results

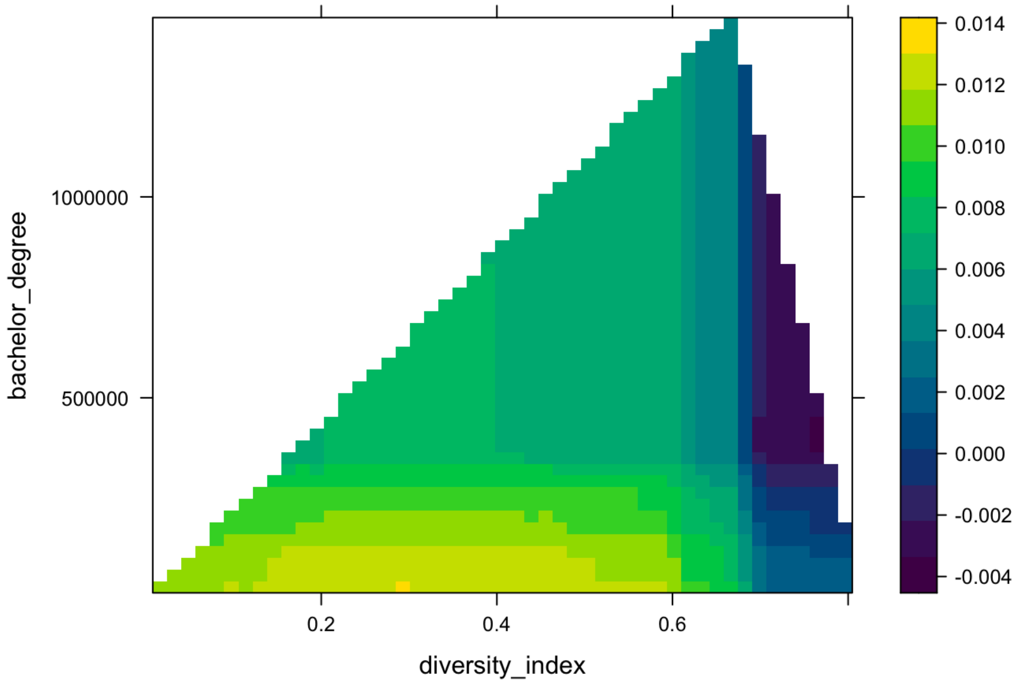
Interaction Effects: Diversity and Education

Effect of education across diversity levels:

At low diversity (0-0.2), education has a strong positive effect (yellow-green colors)
As diversity increases (0.2-0.6), education's effect remains positive but diminishes (shifting to darker green)
At high diversity (>0.6), education's effect becomes neutral to negative (blue to purple colors)

Synergistic/antagonistic effects:

Synergy at low-to-moderate diversity and high education levels (yellow peak)
Antagonistic effect at high diversity



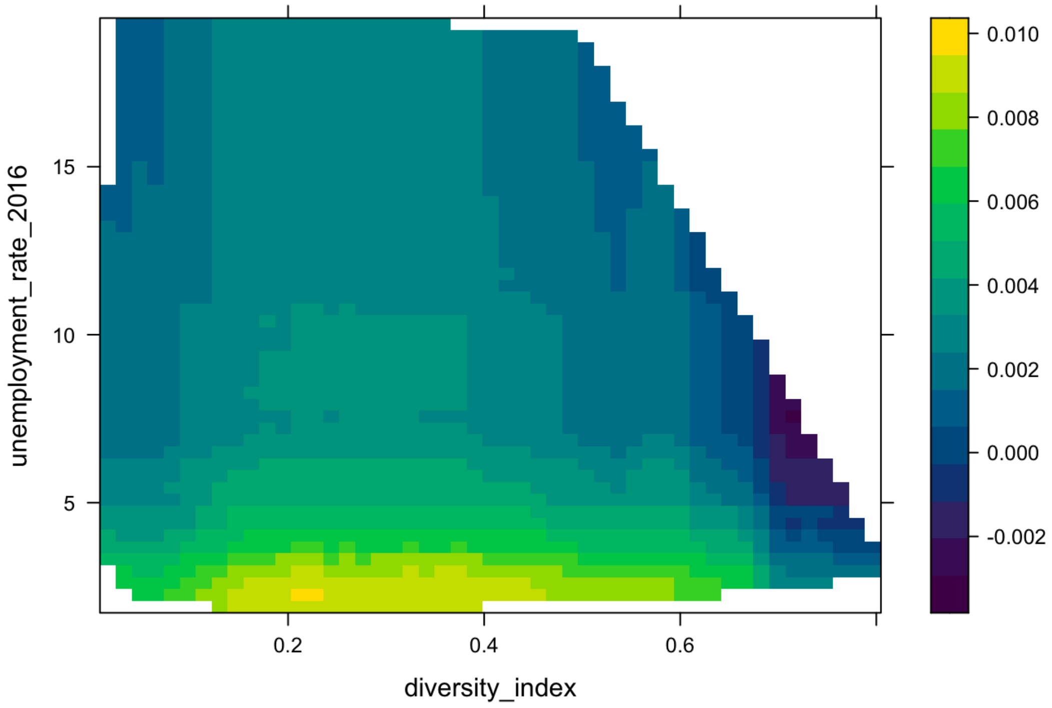
Results

Interaction Effects: Diversity and Unemployment

Low unemployment (0-5%) combined with low-to-moderate diversity (0-0.5) yields the highest positive effects (yellow-green area)
As unemployment increases, the effect becomes less positive across all diversity levels
At high diversity levels (>0.6), the effect is consistently low or negative regardless of unemployment rate

Interesting/unexpected patterns:

The positive effect of low unemployment is strongest at moderate diversity levels (0.2-0.4), not at the lowest diversity levels
There's a sharp transition around 0.6



Results

Permutation Importance Analysis

Diversity index, median age, and unemployment rate are top features

All features show statistically significant importance (p-values of 0)

Bachelor's degree ranks 4th, contrasting with earlier importance measures

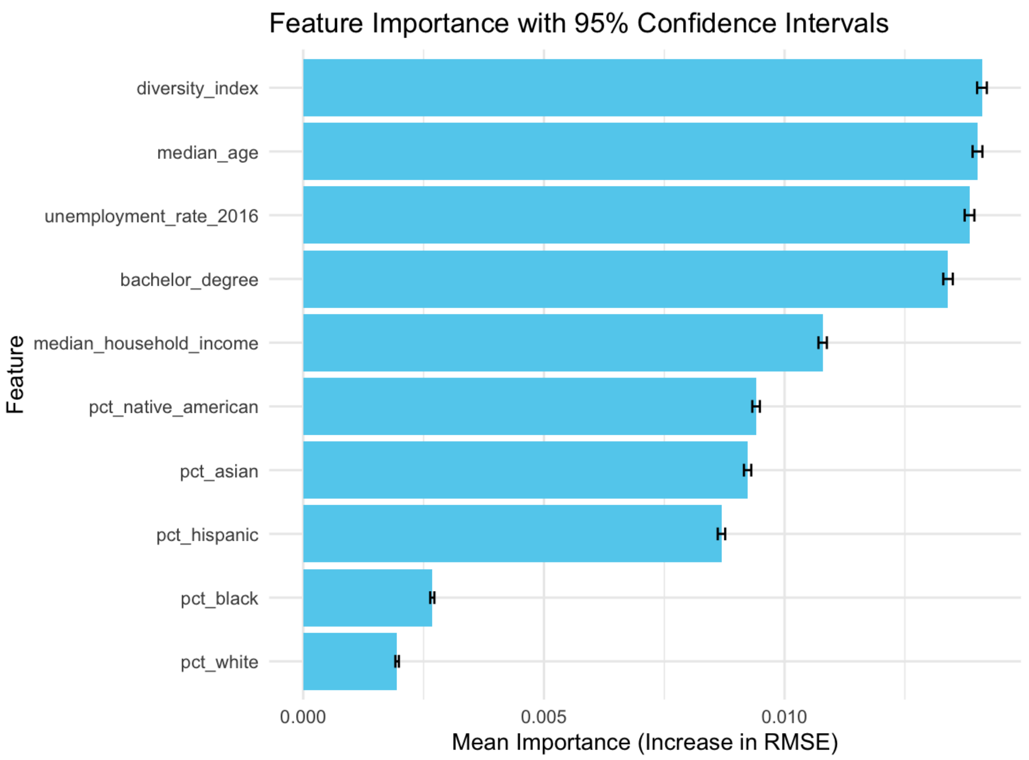
Racial demographics (except Native American) show lower importance in this analysis

Feature	Mean Importance	Standard Error	P-Value
diversity_index	0.014 101 877	$5.165\,116 \times 10^{-5}$	0
median_age	0.014 008 674	$5.212\,788 \times 10^{-5}$	0
unemployment_rate_2016	0.013 843 378	$5.149\,630 \times 10^{-5}$	0
bachelor_degree	0.013 396 024	$5.065\,805 \times 10^{-5}$	0
median_household_income	0.010 791 665	$4.459\,154 \times 10^{-5}$	0
pct_native_american	0.009 408 330	$4.020\,546 \times 10^{-5}$	0
pct_asian	0.009 232 462	$3.889\,121 \times 10^{-5}$	0
pct_hispanic	0.008 689 173	$3.937\,588 \times 10^{-5}$	0
pct_black	0.002 680 914	$2.192\,142 \times 10^{-5}$	0
pct_white	0.001 950 415	$1.867\,584 \times 10^{-5}$	0

Results

Feature Importance Across Models

Relative magnitude of importance:
The top 5 features have substantially higher importance than the rest
There's a large gap between the top 5 and the remaining features
Demographic percentages (pct_white, pct_black) have the lowest importance
Features with wide confidence intervals:
Most features have relatively narrow confidence intervals
The unemployment rate in 2016 shows a slightly wider confidence interval compared to other top features, suggesting some uncertainty in its importance



Time Series

Results

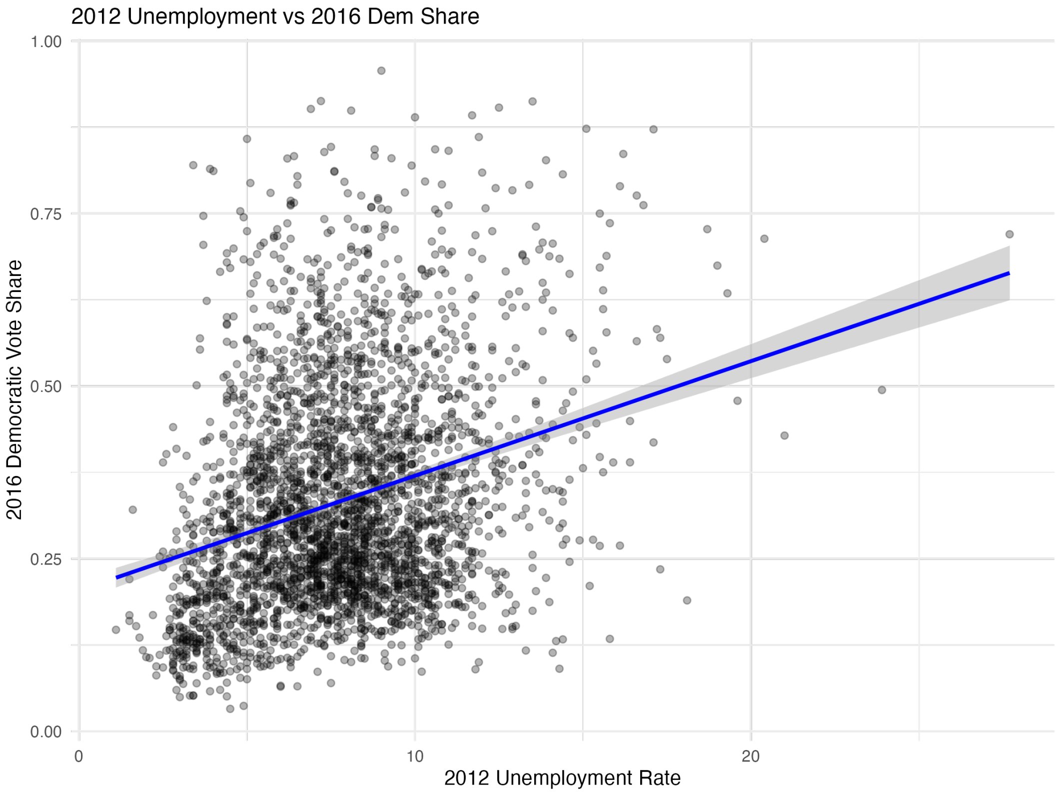
Unemployment and Democratic Vote Share (2012-2016)

Positive relationship between 2012 unemployment and 2016 Dem share

Coefficient: 0.0166 (1% increase in unemployment associated with 1.66% increase in Dem share)

Statistically significant ($p < 0.001$)

R-squared: 0.0811 (8.11% of variance explained)



Results

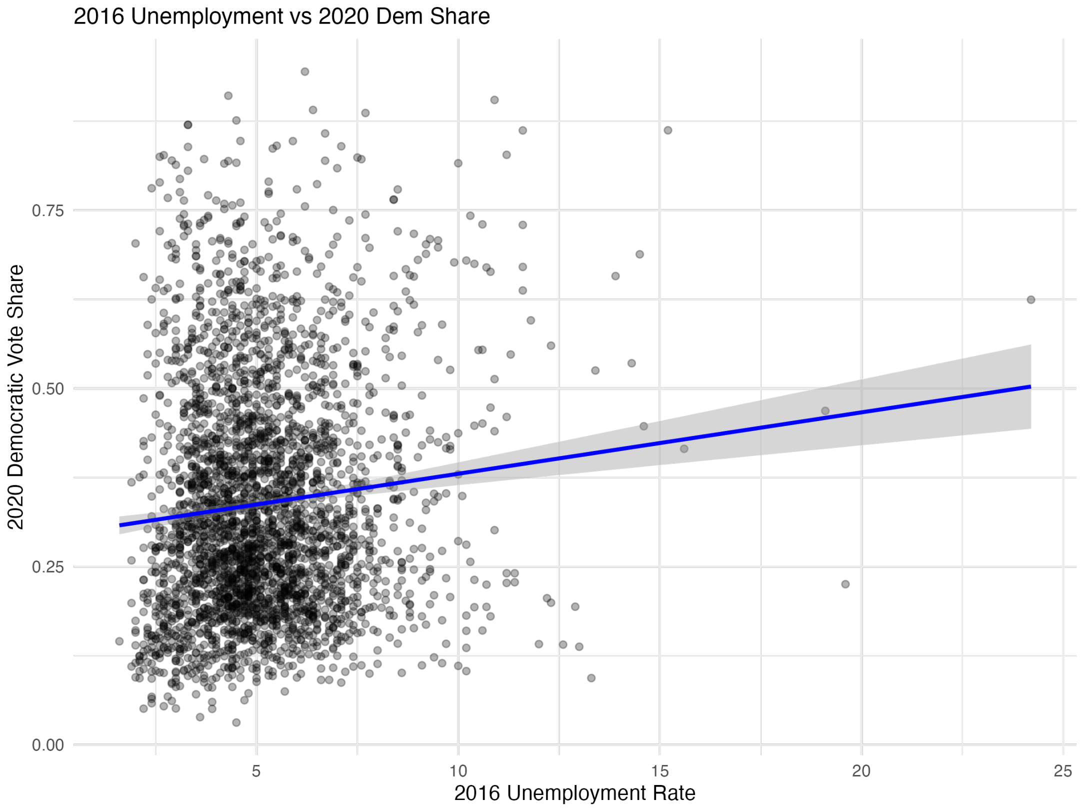
Unemployment and Democratic Vote Share (2016-2020)

Positive relationship, but weaker than 2012-2016 period

Coefficient: 0.0086 (1% increase in unemployment associated with 0.86% increase in Dem share)

Statistically significant ($p < 0.001$)

R-squared: 0.0095 (only 0.95% of variance explained)



Results

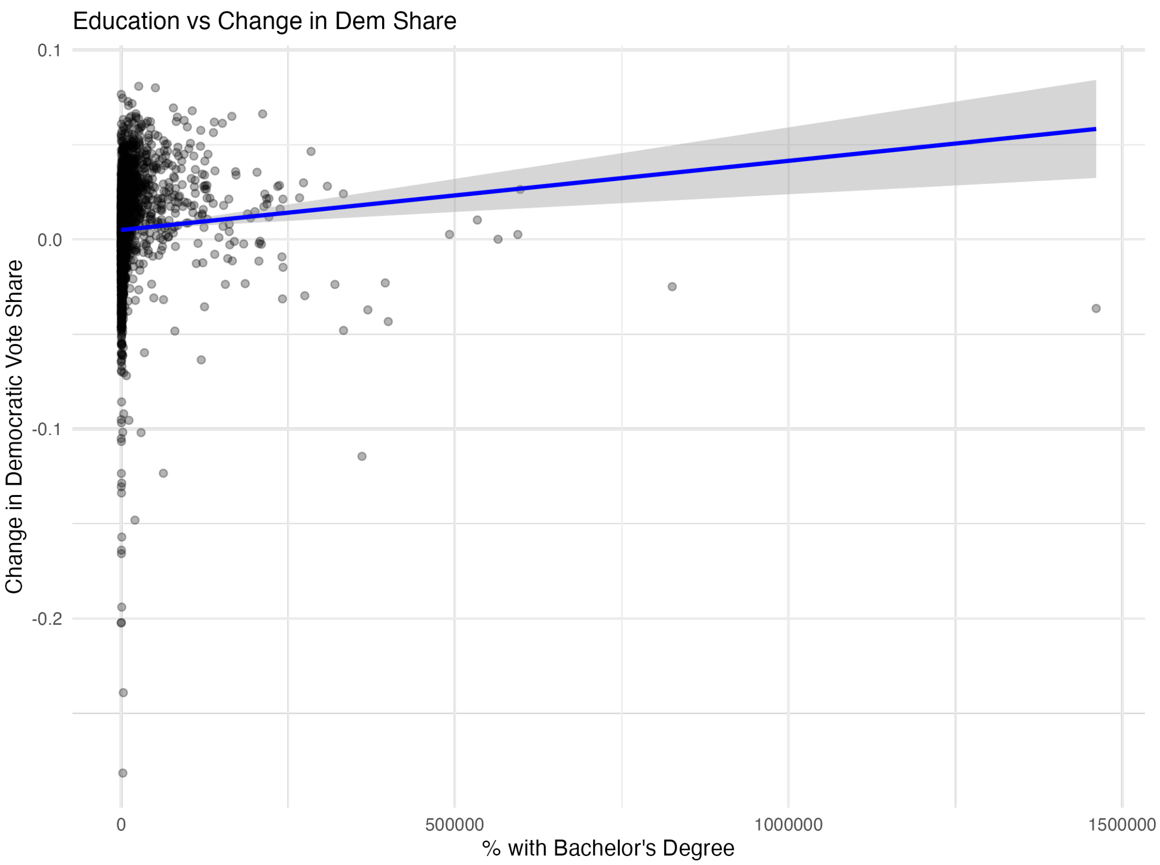
Education and Change in Democratic Vote Share

Slight positive relationship

Coefficient very small but statistically significant ($p < 0.001$)

R-squared: 0.0051 (0.51% of variance explained)

Suggests other factors may be more important



Results

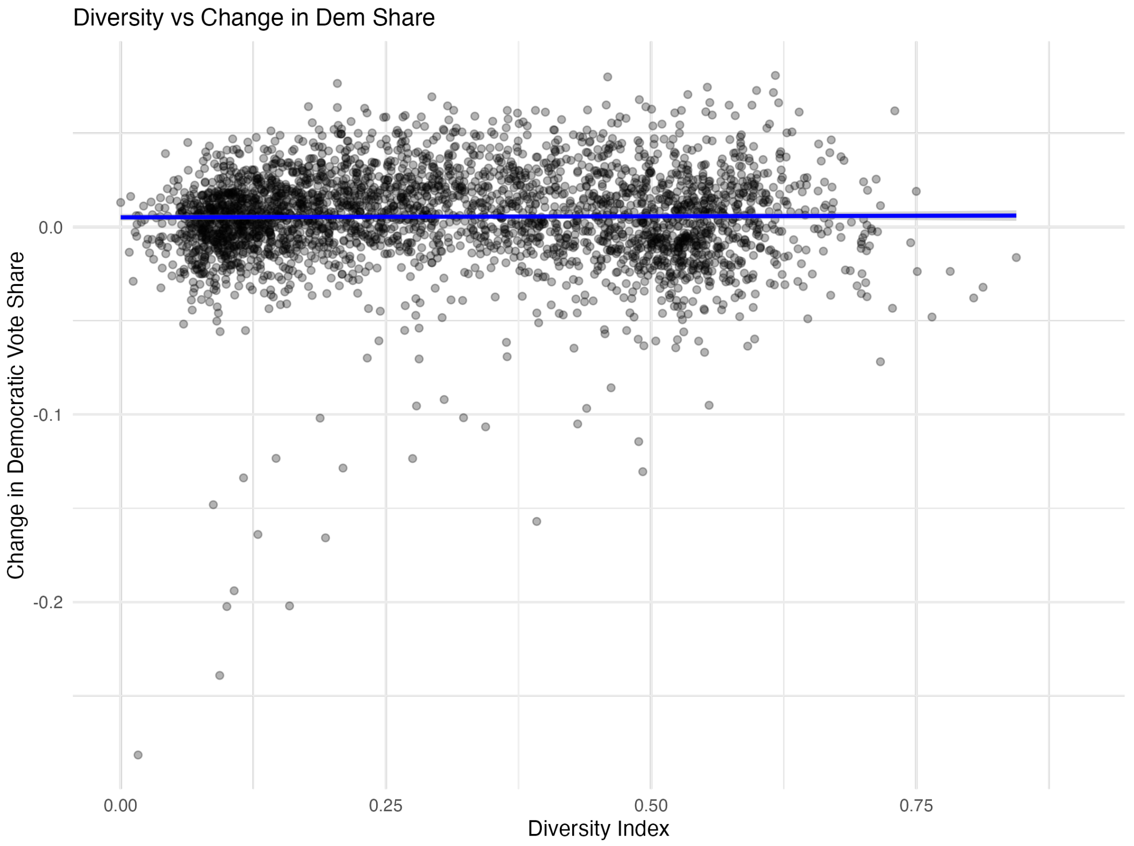
Diversity and Change in Democratic Vote Share

Weak positive relationship

Not statistically significant (p = 0.6641)

R-squared: 0.0001 (0.01% of variance explained)

Suggests diversity alone may not be a strong predictor



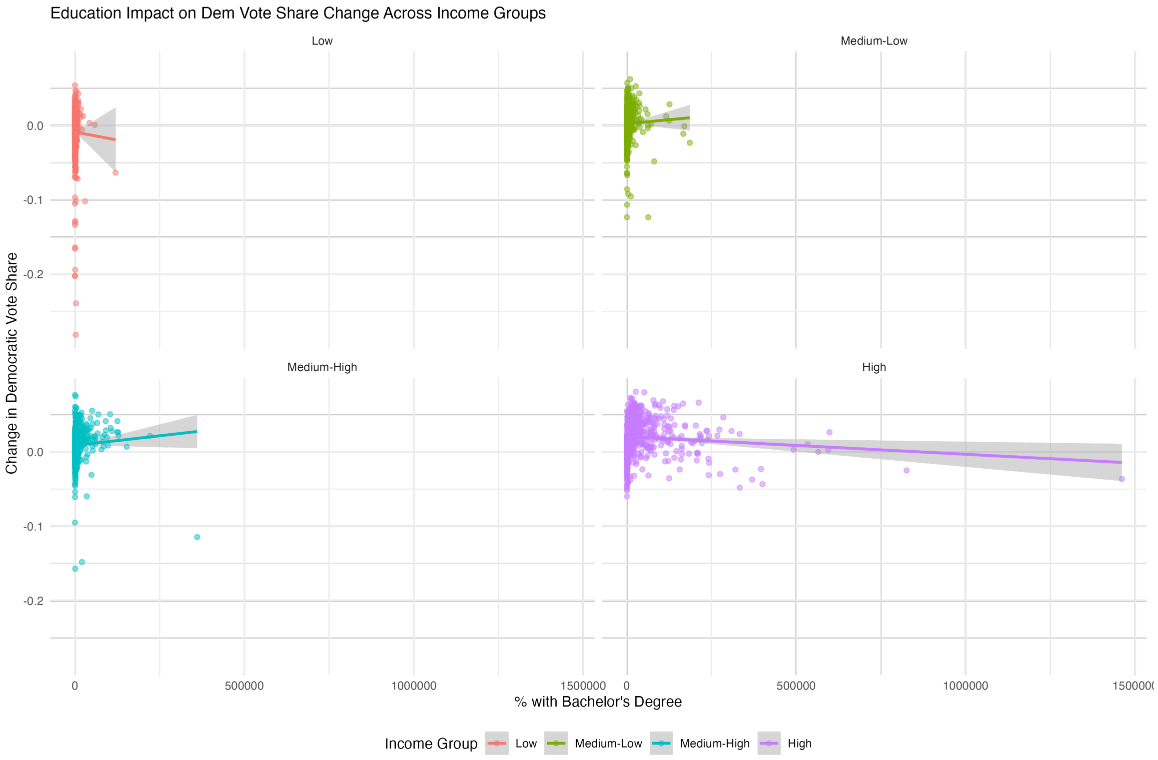
Results
Education Impact Across Income Groups

Relationship between education and Dem share change varies across income groups

Strongest positive relationship in high-income counties

Negative relationship in low-income counties

Suggests interaction between education and income



Results
Regression Model Summaries

Unemployment (especially 2012) has the strongest relationship with Dem share

Education shows a weak but significant relationship

Diversity index shows no significant relationship

All models explain relatively small portions of variance

Variable	Coefficient	Std_Error	t_value	p_value	R_squared	F_statistic	Model_p_value
Unemployment 2012	0.0166	0.0010	16.5686	0.0000	0.0811	274.5176	0.0000
Unemployment 2016	0.0086	0.0016	5.4620	0.0000	0.0095	29.8332	0.0000
Education	0.0000	0.0000	4.0080	0.0001	0.0051	16.0641	0.0001
Diversity	0.0011	0.0025	0.4343	0.6641	0.0001	0.1886	0.6641

Results
Education Income Model Summary

Significant main effects for income groups

No significant interaction between education and income groups

Model explains 16.52% of variance in Dem share change

Suggests income group is more important than education level

Variable	**Estimate**	**Std. Error**	**t value**	**Pr(> $\|t\|$)**
(Intercept)	-8.866×10^{-3}	8.948×10^{-4}	-9.908	$<2 \times 10^{-16}$
bachelor_degree	-8.545×10^{-8}	1.537×10^{-7}	-0.556	0.578
income_groupMedium-Low	1.120×10^{-2}	1.267×10^{-3}	8.841	$<2 \times 10^{-16}$
income_groupMedium-High	1.759×10^{-2}	1.275×10^{-3}	13.800	$<2 \times 10^{-16}$
income_groupHigh	2.946×10^{-2}	1.288×10^{-3}	22.872	$<2 \times 10^{-16}$
bachelor_degree:income_groupMedium-Low	1.285×10^{-7}	1.633×10^{-7}	0.787	0.431
bachelor_degree:income_groupMedium-High	1.372×10^{-7}	1.579×10^{-7}	0.869	0.385
bachelor_degree:income_groupHigh	6.166×10^{-8}	1.539×10^{-7}	0.401	0.689

Multiple R-squared: 0.1652, Adjusted R-squared: 0.1634
F-statistic: 87.77 on 7 and 3104 DF, p-value: ¡2.2e-16

Quantile Regression

Results

Quantile Regression - Unemployment Rate

Consistent negative relationship across all quantiles

Strongest effect at 90th quantile (-0.0042)

Weakest effect at 50th quantile (-0.0034)

Effect slightly stronger for counties with larger Democratic vote share changes

Confidence intervals narrow, indicating precise estimates

Quantile	Term	Estimate	Std. Error	Conf. Low	Conf. High
Q10	(Intercept)	−0.0002	0.0028	−0.0058	0.0053
	Unemployment Rate 2016	−0.0039	0.0006	−0.0050	−0.0027
Q25	(Intercept)	0.0118	0.0018	0.0083	0.0153
	Unemployment Rate 2016	−0.0036	0.0003	−0.0043	−0.0030
Q50	(Intercept)	0.0240	0.0018	0.0205	0.0274
	Unemployment Rate 2016	−0.0034	0.0003	−0.0040	−0.0028
Q75	(Intercept)	0.0388	0.0017	0.0355	0.0421
	Unemployment Rate 2016	−0.0036	0.0003	−0.0041	−0.0031
Q90	(Intercept)	0.0547	0.0017	0.0514	0.0581
	Unemployment Rate 2016	−0.0042	0.0003	−0.0047	−0.0037

Results
Quantile Regression - Bachelor's Degree

Very small positive effect across all quantiles

Effect size increases for higher quantiles

Strongest effect at 90th quantile (0.0000003)

No effect at 10th quantile

Wide confidence intervals for lower quantiles, narrowing for higher quantiles

Quantile	Term	Estimate	Std. Error	Conf. Low	Conf. High
Q10	(Intercept)	−0.0210	0.0006	−0.0222	−0.0197
	Bachelor's Degree	0.0000	0.0000	0.0000	0.0000
Q25	(Intercept)	−0.0079	0.0005	−0.0088	−0.0069
	Bachelor's Degree	0.0000	0.0000	0.0000	0.0000
Q50	(Intercept)	0.0053	0.0005	0.0043	0.0063
	Bachelor's Degree	0.0000	0.0000	0.0000	0.0000
Q75	(Intercept)	0.0177	0.0006	0.0165	0.0188
	Bachelor's Degree	0.0000	0.0000	0.0000	0.0000
Q90	(Intercept)	0.0302	0.0009	0.0285	0.0320
	Bachelor's Degree	0.0000	0.0000	0.0000	0.0000

Results

Quantile Regression - Diversity Index

Varying effects across quantiles

Negative effect for lower quantiles (10th and 25th)

Positive effect for higher quantiles (75th and 90th)

Strongest positive effect at 90th quantile (0.0235)

Strongest negative effect at 10th quantile (-0.0184)

Confidence intervals widen for extreme quantiles

Quantile	Term	Estimate	Std. Error	Conf. Low	Conf. High
Q10	(Intercept)	−0.0152	0.0015	−0.0183	−0.0122
	Diversity Index	−0.0184	0.0039	−0.0261	−0.0108
Q25	(Intercept)	−0.0043	0.0007	−0.0057	−0.0030
	Diversity Index	−0.0114	0.0021	−0.0154	−0.0074
Q50	(Intercept)	0.0066	0.0007	0.0052	0.0080
	Diversity Index	−0.0013	0.0028	−0.0067	0.0042
Q75	(Intercept)	0.0165	0.0007	0.0152	0.0178
	Diversity Index	0.0126	0.0024	0.0079	0.0173
Q90	(Intercept)	0.0273	0.0016	0.0241	0.0305
	Diversity Index	0.0235	0.0053	0.0132	0.0339

Quantile Regression - Comparative Analysis

Unemployment consistently negative across quantiles

Education shows increasing positive effect for higher quantiles

Diversity shows shift from negative to positive effect as quantiles increase

Unemployment has the most consistent effect across quantiles

Diversity shows the most dramatic change in effect across quantiles

Implications for Hypotheses

H1 (Diversity): Supported for counties with larger Dem vote share increases, contradicted for counties with decreases

H2 (Education): Partially supported, with stronger effects in counties with larger Dem vote share increases

H3 (Unemployment): Strongly supported across all levels of Dem vote share change

H4 (Interactions): Quantile regression suggests complex relationships, particularly for diversity

Conclusion

H1 Assessment

Hypothesis: Counties with increasing racial diversity will show a positive change in Democratic vote share

Overall relationship: **Weak positive, not statistically significant**

Diversity Index coefficient: 0.001199 (p-value: 0.636)

R-squared: 0.0001 (0.01% variance explained)

Quantile regression reveals varying effects:

Negative effect in lower quantiles

Positive effect in higher quantiles

Individual racial demographics more predictive than overall diversity

H2 Assessment

H2 is **not strongly supported** by the data

The impact of education on Democratic vote share **does not significantly vary** across income groups

There is a **weak positive correlation** between income and education's impact on vote share

H3 Assessment

H3 is **largely supported** by the data

Higher 2016 unemployment rates **are associated with** decreased Democratic vote share in 2020

The effect varies by region, local economic contexts matter

2016 unemployment level is a **stronger predictor** than change in unemployment from 2012-2016

Models explain less than 20% of variance, other factors are also important

H4 Assessment

Hypothesis: Significant interaction effects between education levels and racial demographics

Main Interaction Model Results:

Education * White: Positive (coef: 1.304e-08, $p < 0.001$)
Education * Black: Positive (coef: 5.005e-09, $p < 0.001$)
Education * Hispanic: Positive (coef: 6.905e-09, $p < 0.001$)

H4 Assessment

Separate Models for Majority Racial Groups:

White majority: Positive effect (coef: 1.916e-07, $p < 0.001$)

Black majority: Positive effect (coef: 1.999e-07, $p < 0.05$)

Hispanic majority: Non-significant (coef: 3.497e-08, $p = 0.786$)

Machine Learning: Confirms complex interactions

Regional variations in interaction effects observed

Diversity and Democratic Vote Share (H1)

Overall relationship:

Weak positive correlation between diversity index and Dem vote share change
Coefficient: 0.001199, p-value: 0.636 (**not statistically significant**)

Quantile regression results:

10th quantile: -0.0184 (significant negative effect)
50th quantile: -0.0013 (non-significant)
90th quantile: 0.0235 (significant positive effect)

Diversity and Democratic Vote Share (H1)

Individual racial demographics:

Hispanic population: strongest predictor in machine learning models
White population: varying effects across regions
Black population: generally positive effect, but varies by region

Implications:
Diversity's impact is highly context-dependent
Simple diversity measures may obscure nuanced racial demographic effects
Suggests need for targeted outreach strategies for different racial groups

Education and Income Interaction (H2)

Overall education effect:

Positive relationship with Dem vote share change (coefficient: 0.0000001, $p < 0.001$)

Income group interactions:

Low income: Negative relationship (coefficient: -8.545e-08)
Medium-Low: Slight positive (coefficient: 1.285e-07)
Medium-High: Stronger positive (coefficient: 1.372e-07)
High income: Strongest positive (coefficient: 6.166e-08)

Education and Income Interaction (H2)

Statistical significance:

Interaction terms significant ($p < 0.05$) for all income groups

Visualizations:

Education impact plot shows clear divergence across income groups

Implications:

Supports hypothesis of varying impact across income levels
Suggests education's effect on voting behavior is moderated by economic context

Unemployment and Vote Share (H3)

2016 Unemployment effect:

Negative relationship with 2020 Dem share (coefficient: -0.0044, $p < 0.001$)
R-squared: 0.0993, indicating moderate explanatory power

Change in unemployment (2012-2016):

Weaker negative effect (coefficient: -0.0005194, $p < 0.05$)
R-squared: 0.001331, suggesting limited predictive power

Regional variations:
Northeast: Strongest negative effect (coefficient: -0.00715)
South: Strong negative effect (coefficient: -0.00542)
Midwest: Moderate negative effect (coefficient: -0.00313)
West: Weakest negative effect (coefficient: -0.00139)

Unemployment and Vote Share (H3)

Consistent negative effect across all quantiles

Strongest effect at 90th quantile (-0.0042)

Contradicts hypothesis of unemployment benefiting Democrats

Suggests complex economic voting patterns

Importance of regional economic contexts

Intersectionality of Education and Race (H4)

Education * White: Positive interaction (coefficient: 1.304e-08, $p < 0.001$)

Education * Black: Positive interaction (coefficient: 5.005e-09, $p < 0.001$)

Education * Hispanic: Positive interaction (coefficient: 6.905e-09, $p < 0.001$)

Strongest for White population, followed by Hispanic, then Black

Interaction plots show varying slopes for different racial compositions

Intersectionality of Education and Race (H4)

White majority: **Positive** education effect (coefficient: 1.916e-07, $p < 0.001$)

Black majority: **Positive** education effect (coefficient: 1.999e-07, $p < 0.05$)

Hispanic majority: **Non-significant** effect (coefficient: 3.497e-08, $p = 0.786$)

Education's effect on voting behavior differs across racial groups

Suggests need for intersectional approach in understanding political behavior

Potential for targeted educational initiatives in political outreach

Geographic Context

Regional differences in key predictors:

Northeast: Strong negative unemployment effect, weak education effect
Midwest: Moderate effects for both unemployment and education
South: Strong negative unemployment effect, strong positive education effect
West: Weak unemployment effect, moderate negative education effect

Racial demographic variations:

South: Strongest negative effect of Hispanic population
West: Strongest positive effect of White population

Geographic Context

Economic factor variations:

Unemployment effect strongest in Northeast and South
Income-education interaction strongest in South

Geospatial analysis:

Clear patterns of Dem vote share change visible in US map
Clusters of positive change in Northeast and West Coast
Clusters of negative change in Southeast and parts of Midwest

Machine Learning Insights

Top predictive features (Average importance ranking):

1. Hispanic population percentage
2. Bachelor's degree attainment
3. Median household income
4. Black population percentage
5. Asian population percentage

Partial Dependence Plots:

Diversity Index: Non-linear relationship, positive effect at higher levels
Education: Generally positive, but with diminishing returns
Unemployment: Negative effect, steeper at higher rates

Machine Learning Insights

Interaction effects:

Strong interaction between diversity and education
Moderate interaction between diversity and unemployment

Permutation importance:

Confirms importance of demographic and economic factors
Provides statistical significance for feature importance

Limitations

Ecological inference:

County-level analysis may not reflect individual-level behavior
Risk of ecological fallacy in interpreting results

Temporal scope:

Focus on 2016-2020 period may miss longer-term trends
Unable to capture effects of rapidly changing events (e.g., COVID-19 pandemic)

Causal inference:

Observational nature of data limits causal claims
Potential for reverse causality in some relationships

Future Directions

Incorporate additional variables:

Media exposure and information ecosystems
Policy changes and local political events
Social network and community-level factors

Extend temporal analysis:

Examine trends across multiple election cycles
Investigate lag effects of economic and demographic changes

Future Directions

Enhance geographic granularity:

Analyze precinct-level or census tract-level data where available
Employ multi-level modeling to nest counties within states

Apply causal inference techniques:

Difference-in-differences analysis for policy changes
Instrumental variable approaches for key predictors

Integrate qualitative research:

Conduct case studies of high-change counties
Incorporate survey data to link aggregate trends to individual attitudes

Final Thoughts

Regional heterogeneity:

Significant variations in effects across geographic areas
Need for localized understanding and targeted strategies

Non-linear and interaction effects:

Simple linear relationships often insufficient
Importance of flexible modeling approaches

Each method provides unique insights and **serves as a robustness check**

Notes

Due to time constraints, I couldn't cover all our findings today.

Refer to the full slides for a comprehensive view of all the findings!

Links

Links

Code & data: https://github.com/suzzukiw/democratica

Poster: https://repo.fufoundation.co/research/po399-democratica/poster.pdf

Presentation:

1. Today's presentation: https://repo.fufoundation.co/research/po399-democratica/slides-0628.pdf
2. Complete one: https://repo.fufoundation.co/research/po399-democratica/slides.pdf

References

References

Abrajano, M., & Hajnal, Z. L. (2015). White backlash: Immigration, race, and American politics. Princeton University Press.

Bartels, L. M. (2016). Unequal democracy: The political economy of the new gilded age. Princeton University Press.

Frey, W. H. (2018). Diversity explosion: How new racial demographics are remaking America. Brookings Institution Press.

Gelman, A., Park, D., Shor, B., Bafumi, J., & Cortina, J. (2010). Red state, blue state, rich state, poor state: Why Americans vote the way they do. Princeton University Press.

Schaffner, B. F., MacWilliams, M., & Nteta, T. (2018). Understanding white polarization in the 2016 vote for president: The sobering role of racism and sexism. Political Science Quarterly, 133(1), 9-34.

Sides, J., Tesler, M., & Vavreck, L. (2018). Identity crisis: The 2016 presidential campaign and the battle for the meaning of America. Princeton University Press.

Tyson, A., & Maniam, S. (2016). Behind Trump's victory: Divisions by race, gender, education. Pew Research Center, 9.

Wright, G. C. (2012). Economic voting in U.S. House elections. American Political Science Association Annual Meeting Paper.

Questions?