Beyond the Rural-Urban Divide: Analyzing Electoral Change in U.S. Swing States, 2016-2020

Ashna Singh, as7051

Master's Thesis submitted in partial fulfillment
of the requirements for the degree of a
M.A. in Quantitative Methods in the Social Sciences

Columbia University
Graduate School of Arts and Sciences

December 2024

Faculty Advisor: Gregory M. Eirich Ph.D

Abstract

Swing states are decisive factors in American presidential elections. In 2016, Donald Trump won Arizona, Georgia, Michigan, Pennsylvania, and Wisconsin helping him win the presidential election. However, in 2020 Joe Biden flipped those states helping him win against Donald Trump. In 2024, Donald Trump was able to win those states, and the presidential seat back. Various variables, including racial demographics, employment statistics, religious affiliations, educational attainment, and urban-rural population distribution, have been analyzed as potential predictors of electoral outcomes in swing states. An analysis of swing states during the 2016 and 2020 elections using a multilevel model revealed that race, education, and employment in the goods-producing industry were all significant predictors of voting behavior. The relationship between rurality and the Democratic vote was a more complex, non-linear pattern. When examining the counties that flipped from Republican to Democrat between 2016 and 2020, race was an important change, but another factor, which is harder to quantify, incumbency, may have also influenced these changes. While statistical models can capture quantifiable factors like race, education, and employment, they may not fully account for more nuanced influences such as incumbency or shifting voter sentiments. Therefore, a holistic approach that combines rigorous data analysis with qualitative insights is crucial for understanding the complex dynamics of swing state elections and the evolving nature of American political behavior.

Analyzing Electoral Change in U.S. Swing States, 2016-2020

The United States presidential election is one of the biggest political events on the global stage, and with each presidential cycle comes lasting implications that affect not only the U.S. but the rest of the world as well. The scale of the presidential election prompts analysts and researchers to ask the question: can we predict the presidential election winner, and if so, how? This is a loaded and complicated question, as the answer does not lie in one place. Factors such as race, education, religion, candidate incumbency, geopolitical events, and socioeconomics (to name a few) are all important predictors of whether a voter chooses to participate in an election and how they will cast their ballot (Pew Research Center, 2024). This makes it difficult to focus on only one of these variables, as ignoring the rest means potentially overlooking crucial interactions and nuances that shape voting behavior. Moreover, it is possible to do these analyses at multiple levels such as the individual voter level, county level, state level, voting districts, and congressional districts, further complicating this area of research.

However, when examining voting patterns, many states have a history that makes them relatively easy to predict. For example, California has voted blue since 1992 and Texas has voted red since 1980 (270 To Win: 'Same Since' Electoral Maps). A handful of states flip between elections on whether they will vote majority Democrat or Republican, and these are called swing states (or battleground states). Predicting how a swing state will vote has become increasingly important as they can be decisive factors in presidential elections. Notably, in 2016 Donald Trump won Arizona, Georgia, Michigan, Pennsylvania, and Wisconsin, but Joe Biden won those states in 2020. Even within a rural or urban county exist trends which researchers have examined. For example, liberals prefer to live in more walkable communities, but conservatives prefer larger houses that are more far apart (Pew Research Center, 2014). Moreover, both liberals and conservatives are more likely to prefer to live in an area where people share their political values, however, this desire is stronger on the right (50%) than on the left (35%). Factors such as these are only increasing the political polarization in the United States.

One framework researchers have explored is the impact of a county's population size on whether it votes majority Democrat or Republican. Specifically, they look at rural versus urban counties and their voting patterns. When classifying counties as urban, the Census Bureau identifies a minimum population threshold, and as of 2020, it now has a minimum housing unit threshold (U.S. Census Bureau, 2024). After defining urban areas, what is left is classified as rural. The urban-rural political divide and increased polarization is another commonly researched topic. Since 1972, Republicans have seen a similar share of votes from both urban and rural counties, but the 2000 presidential election saw a growing divide between these counties where the largest amount of Republican votes came from rural counties, and this divide has grown wider since (Brown & Mettler, 2023). This rural shift in heavily supporting the Republican party can be seen at multiple levels of governance. Factors such as urban resentment, anti-Black racism, economic despair, and status threat have been shown to contribute to the increased Republican support among rural counties (Albrecht, 2022; Brown & Mettler, 2023; Scala et al., 2017).

A. Economic Stagnation

Employment and economic growth are widely cited as reasons that rural citizens prefer to vote with the Republican Party. In recent decades, rural areas have seen stagnant job growth while urban areas have seen a continuous increase in employment (Brown & Mettler, 2023). This has largely been due to a decline in the number of jobs in the goods-producing industries – the primary area of employment for rural populations (Albrecht, 2022). This ties into factors such as education as well. Often rural populations work in the goods industries due to low levels of education, and when those industries have fewer jobs it leads to higher unemployment levels. Even when employed, many of these jobs pay low, and declining wages leading to higher levels of poverty (Albrecht, 2022). Growing levels of economic inequality in the United States have significantly affected rural communities, while those in the top 1% have only seen substantial increases in wealth (Gould, 2019). Researchers argue that it is not an employment agenda that attracts rural populations to the Republican Party, but rather that they felt alienated from the Democratic Party when it shifted its focus from trade and financial regulation onto other issues (Brown & Mettler, 2023).

B. Education

As mentioned previously, education plays a large role in employment, especially rural employment, and the educational gap between rural and urban counties has contributed to increased political polarization (Brown & Mettler, 2023). Since 1970, urban and rural counties have seen the same level of growth in the share of the population with a four-year degree, but in recent decades, urban counties have experienced a level of growth that rural counties have not. As of 2020, 35% of people 25 years and older in urban counties hold at least a four-year degree, while in rural counties it is only 21%. Additionally, many people leave rural areas for urban counties in pursuit of higher education and rarely return to those rural areas (Albrecht, 2022). Research has shown that higher education is often associated with more progressive social values and highly educated people also tend to live in more urban and diverse areas (Iversen & Soskice, 2019). This contributes further to the growing ideological gap between rural and urban areas. Researchers also reason that rural populations may hold resentment toward highly educated urban Democrats, believing that this group is imposing their agenda onto them (Brown & Mettler, 2023).

C. Race

Perhaps the largest divide in American politics lies in that of race and ethnicity. Once upon a time, the non-Hispanic White population dominated the United States, but today that is a different story. Increased immigration and high birth rates among people of color have decreased the non-Hispanic White population from 80% in 1970 to 60% as of 2020 (Massey, 2007). Urban and rural areas saw similar percentages of people identifying as nonwhite in 1970, at 13% and 11% respectively. As of 2020, 46% of the population identifies as nonwhite, while only 25% of rural populations identify as nonwhite (U.S. Census Bureau). Positions of power are also more likely to be held by people of color now, and this was further fueled by President Barack Obama's election and time in office. For decades, notably since the passage of the Civil Rights Act, a majority of people of color have voted Democrat and this is due to the Democratic Party's focus on policies such as higher wages, fair housing, and affirmative action (Albrecht, 2022). On the other hand, many White people may feel that changing demographics are a threat to the benefits they enjoy (Chavez, 2013). The Republican party has used this to their advantage citing the Democrat's support of minorities as the main reason for poor White people's circumstances becoming worse (Albrecht, 2022). Through arguments such as these, Republicans have maintained strong support from the White working class, and thus race has become an important predictor of voting behavior. In 2016, Trump's campaign exhibited distinct racial undertones, and research indicates that this contributed to him securing a larger share of the White vote compared to previous Republican candidates in past elections (Albrecht, 2022).

D. The Rural-Urban Continuum

When it comes to identifying whether a county can be classified as rural or urban, many researchers argue that it is not that simple, and viewing the differences between counties as a dichotomy obscures a more nuanced understanding of America's political geography (Scala & Johnson, 2017). Instead, they argue that counties lie on a rural-urban continuum. Research has shown that voting patterns in the rural United States tend to be very diverse. For example, while Democratic performance has historically been relatively poor in rural America, recreational counties in swing states show strong support for Democrats (Scala et al., 2015). Recreational communities in rural areas tend to have a higher median income and higher levels of education and favor Democrats, while farming communities in rural areas tend to be the opposite and usually favor Republicans. Additionally, rural recreational counties in the past decade have experienced much higher levels of population gain compared to farming counties (Scala et al., 2015).

The United States Department of Agriculture developed a scale for not only identifying a county based on its population size but also on its proximity to other urban counties. The continuum classifies counties on a scale of 1 through 9, with 1 being counties with the highest population (most urban), and 9 being counties with the lowest population and not being adjacent to a metropolitan county (most rural). Classifying counties on this scale allows a more comprehensive understanding of county rurality and opens the door for a deeper understanding of presidential election voting patterns (see Appendix Figure 13 for complete codes).

E. Main Literature

There are two main papers which have provided the basis for my thesis. Both investigate urban-rural voting patterns during presidential elections in the United States.

I. Political Polarization Along the Rural-Urban Continuum? The geography of the Presidential Vote, 2000-2016 – Dante J. Scala & Kenneth M. Johnson

In this paper, researchers use spatial analysis to look at what might predict political attitudes during presidential elections from 2000-2016. The independent variables were rural-urban continuum, region, education, income, age, race, religious affiliation, and selected nonmetropolitan economic types. The dependent variable was the percentage of votes for Democrats during the 2012 and 2016 presidential elections in counties in the continental United States. They ran a multivariate spatial error regression model and found region, rural-urban continuum, education, age, religion, race, and income to be significant predictors of Democratic support. They concluded that the "percentage voting Democrat was greatest in large urban cores and smallest in remote nonmetropolitan counties, even holding other variables constant."

Additionally, Republicans performed better in remote nonmetropolitan farming counties with lower levels of education and higher levels of White Evangelical adherents.

II. Donald Trump and Changing Rural/Urban Voting Patterns – Don E. Albrecht

In this paper, Albrecht examined the effect of the rural vote on Donald Trump's win in the 2016 presidential election. The independent variables were position on the rural-urban continuum, employment in the goods-producing industry (measured by the proportion of the employed labor force in each county working in agriculture, logging, mining, construction, and manufacturing), race/ethnicity (proportions of residents that are non-Hispanic White, non-Hispanic Black, & Hispanic), and educational attainment (proportion of people over 25 with at least a college degree). The dependent variable was the percentage of votes for the Republican candidate (Donald Trump). The author first ran a series of correlations to examine the bivariate relationship between each of the independent variables and the dependent variable. Then he ran two OLS regression models for the 2016 presidential election weighted by county population. One model was run only with Arizona, Georgia, Michigan, Pennsylvania, and Wisconsin, and the second model was run with the rest of the states. This was repeated for the 2020 presidential election. The results showed that rural residents were more likely to vote Republican than urban residents, and as the non-Hispanic White population and percent employed in the goods-producing industries increased, the percent voting Republican also increased. The most significant variables were percent non-Hispanic White and the percentage of adults with a college degree. The relationship between percent Hispanic and the percentage voting for Trump was negative and significant, and the author noted that, "When other variables are statistically considered, the direction of the relationship reverses. This is likely because of low educational attainment levels among Hispanics and heavy dependence on employment in the goods-producing industries." In sum, the effect of continuum position became less important when compared to race & education, and greater proportions of non-Hispanic Whites and those without college education were more likely to support Trump. The author concluded that it is not rurality that is the predictor of voting patterns, but it is the demographics of the populations that make up those rural counties instead.

H. Drawing From the Above Literature

These two papers draw two different conclusions but with similar ideas. One argues that White people with lower levels of education tend to live in rural areas which then in turn vote

Republican while the other argues that rural areas tend to have White people with lower levels of education who tend to vote Republican. However, there are a few gaps in both of these papers. The Albrecht paper runs a regression without testing for spatial autocorrelation. Spatial autocorrelation is the relationship between spatial units on a map (Getis, 2009). Spatial autocorrelation coefficients can help understand the strength of spatial variables in a regression, identify spatial clusters, and help weight temporal effects. More concisely put, "Given a set S containing n geographical units, spatial autocorrelation refers to the relationship between some variable observed in each of the n localities and a measure of geographical proximity defined for all n(n–1) pairs chosen from S." (Hubert et al., 1981, p.224). Spatial autocorrelation can be tested for by using the Moran's I test which is similar to the Pearson correlation coefficient. It tests spatial randomness and a rejection of the null hypothesis indicates that spatial autocorrelation exists (Getis, 2009). Scala and Johnson tested for spatial autocorrelation and found it to be present in both the 2012 and 2016 election data, which led to the decision to use a spatial error model. Albrecht's OLS regression does not account for spatial relationships between counties meaning that there could be a spatial bias in the model's standard error (Anselin, 2005).

Neither of these two papers account for or mention testing for the possibility of collinearity or multicollinearity. Variables such as position on the rural-urban continuum, race, and education tend to be correlated. For example, rural areas are more heavily populated by non-Hispanic White people and those with lower levels of education. While the researchers do acknowledge these facts, neither of them mentions testing for collinearity to address this before running their analyses.

Research Question & Hypothesis

I want to investigate the role of rurality on voting patterns in the United States presidential elections. Specifically, I want to try to answer the question of county rurality versus population demographics. To what degree does a county influence voting patterns and to what degree is it the people who live in those counties? One key aspect of addressing this question is to use multilevel modeling. This hierarchical model would allow me to simultaneously examine the effects of individual-level characteristics such as education and race and county-level characteristics such as rural-urban classification. The use of multilevel spatial modeling depends on whether there is spatial autocorrelation present in the data or not. The dependent variable is the percentage of Democrat votes, and the independent variables are position on the rural-urban continuum, race, education, and employment in the goods-producing industry.

Methods

To create the dataset, demographic, employment, and education data were collected from the United States Census Bureau. The rural-urban continuum codes and county-level shapefiles were also collected from the Census Bureau. County-level voting data was obtained from the Harvard Dataverse. The complete datasets include demographic, employment, education, and election returns for the 2016 and 2020 presidential elections. Numeric data was converted to percentages to account for population differences across counties. All of the data was collected for both 2016 and 2020, except for the rural-urban continuum codes which were only available for 2013 and then 2023 – the 2013 codes were used for both the 2016 and 2020 elections. Only swing states were analyzed: Arizona, Colorado, Florida, Georgia, Iowa, Michigan, Minnesota, Nevada, New Hampshire, North Carolina, Ohio, Pennsylvania, Virginia, and Wisconsin.

Testing for Spatial Autocorrelation

All variables were tested for spatial autocorrelation to assess whether a spatial regression model was appropriate. Spatial autocorrelation refers to the extent to which the occurrence of an event or value in one location makes a similar event or value more or less likely in neighboring locations. Simply put, it refers to the relationship between nearby spatial units (Getis, 2010). When spatial autocorrelation is present, it violates the assumption that the measured outcomes are independent of each other. Spatial autocorrelation measures determine if adjacent spatial units share characteristics more often than would be expected by chance. The Moran's I test was used to measure how one object is similar to others surrounding it. It tests for spatial randomness, and rejecting the null hypothesis indicates that spatial autocorrelation exists (Getis, 2010). None of the Moran's I statistics were statistically significant indicating that no spatial autocorrelation is present in the data (see Appendix Figure 7). This is expected since swing states are not geographically contiguous.

Testing for Collinearity

A correlation matrix (see Appendix Figure 14) and the Variance Inflation Factor (VIF) (ran post-model) showed high levels of collinearity between measures of White and Black populations as well as high levels of collinearity between education and employment in goods-producing industries. A new variable was created by subtracting the percentage of Black people in a county from the percentage of White people in a county. The higher the value of this variable the greater the difference between White and Black populations in a given county, meaning there is a higher proportion of White people compared to Black people. To interpret this in the model, for each percentage point that the White population exceeds the Black population, the Democratic vote share is expected to increase/decrease by about x percentage points.

The employment and education variables were centered to address collinearity. Two new variables were created which reflect the difference between a county's percentage and the average percent across all counties. This created a new baseline to decrease collinearity. To interpret this in a model, a county with 1 percentage point above average employment in goods-producing industries is associated with an x percentage point higher/lower Democratic vote share, assuming average goods-producing employment. A county with 1 percentage point above average college-educated population is associated with an x percentage point higher/lower Democratic vote share, assuming average education levels. After adjusting the above variables for collinearity, all VIF values were close to 1 (see Appendix Figure 11).

Rural-Urban Linear Regression Model

First, a basic linear regression was run to help identify the effect of rural-urban differences on voting behavior. The independent variable was a county's position on the rural-urban continuum, and the dependent variable was the percentage of the county that voted Democrat. This basic model served as a point of comparison for more complex models that include additional variables. It helped assess how much of the voting pattern is explained by rurality alone versus other factors.

Multilevel Model

A multilevel model was used to examine the predictors of swing state outcomes in presidential elections. Since no spatial autocorrelation was found present, spatial weights were not included. The multilevel model was chosen to see differences at both the state and county level. The dependent variable was the percentage of the county that voted Democrat. For 2016 this meant voting for Hillary Clinton, and in 2020 voting for Joe Biden. The independent variables were the difference between the proportion of White and Black populations, proportion of the population that is Hispanic, position on the rural-urban continuum (ordinal), distance from the average percentage of the population that works in a goods-producing industry, and distance from the average proportion of the population 25 or older that has a college degree or above.

The Intraclass Correlation Coefficient (ICC) was also calculated to measure the extent to which outcomes within a group are similar, or between groups are different. The unadjusted ICC reflects the percent of variability that can be attributed to differences between states, and the unadjusted ICC reflects the percent of variability that can be attributed to differences between states after accounting for fixed effects.

Results

Figure 1

2016 Rural-Urban Linear Regression Results

Variable	Estimate (1)	Standard Error (2)	P-Value (3)
Intercept	42.25	0.94	<0.001*
Continuum Category 2	-3.17	1.41	0.025*
Continuum Category 3	-4.76	1.46	0.001*
Continuum Category 4	-6.75	1.58	<0.001*
Continuum Category 5	-3.00	2.91	0.303
Continuum Category 6	-10.04	1.29	<0.001*
Continuum Category 7	-11.08	1.45	<0.001*
Continuum Category 8	-8.63	1.76	<0.001*
Continuum Category 9	-12.20	1.66	<0.001*

* Significant value
Residual standard error: 12.61 on 1052 Df
Multiple R-squared: 0.09811, Adjusted R-squared: 0.09125
F-statistic: 14.31 on 8 and 1052 Df, p-value: < 2.2e-16

In 2016, the linear regression analysis revealed significant rural-urban differences in Democratic voting patterns. Urban counties (Category 1) show a baseline Democratic vote share of 42.25%. Moving down the rural-urban continuum towards rurality, there is a consistent negative relationship with Democratic voting, with most rural categories showing statistically significant decreases compared to urban areas. The most rural category (9) shows the strongest negative effect of -12.20%. The model explains approximately 9.8% of the variance in Democratic vote share ($R^2 = 0.098$), with Category 5 being the only non-significant predictor ($p = 0.303$).

Figure 2

2020 Rural-Urban Linear Regression Results

Variable	Estimate (1)	Standard Error (2)	P-Value (3)
Intercept	45.79	0.98	<0.001*
Continuum Category 2	-4.11	1.47	0.005*
Continuum Category 3	-5.55	1.52	<0.001*
Continuum Category 4	-8.89	1.65	<0.001*
Continuum Category 5	-3.18	3.03	0.295
Continuum Category 6	-12.70	1.34	<0.001*
Continuum Category 7	-12.82	1.52	<0.001*
Continuum Category 8	-11.61	1.83	<0.001*
Continuum Category 9	-14.09	1.73	<0.001*

* Significant value
Residual standard error: 13.15 on 1052 Df
Multiple R-squared: 0.1276, Adjusted R-squared: 0.121
F-statistic: 19.23 on 8 and 1052 Df, p-value: < 2.2e-16

In 2020, the rural-urban divide appears to intensify slightly. The baseline Democratic vote share in urban counties increases to 45.79%, while the negative effects of rurality generally strengthen. The most rural category (9) shows an even larger negative effect of -14.09% ($p < 0.001$) and categories 6 and 7 show particularly strong negative effects as well. The model's explanatory power increases slightly to 12.8% ($R^2 = 0.128$), suggesting rurality became a somewhat stronger predictor of voting patterns. Similar to 2016, Category 5 remains non-significant ($p = 0.295$), while all other categories showed significant effects.

Figure 3

2016 Election Multilevel Model Results

Random Effects

Group	Parameter	Variance	Std.Dev.
State	(Intercept)	58.51	7.649
Residual		35.87	5.989

Number of observations (counties): 1061
Number of groups (State): 14

Fixed Effects

Variable	Mean (1)	Standard Error (2)	P-Value (3)
Intercept	60.73	2.223	<0.001*
White-Black Proportion Difference	-0.392	8.415e-03	<0.001*
Hispanic Population	-0.403	0.028	<0.001*
Goods-Producing Industry Employment	-0.189	0.035	<0.001*
Education	0.571	0.028	<0.001*
Continuum Category 2	2.938	0.70	<0.001*
Continuum Category 3	1.693	0.719	0.019*
Continuum Category 4	2.655	0.805	0.001*
Continuum Category 5	3.475	1.420	0.015*
Continuum Category 6	1.744	0.708	0.014*
Continuum Category 7	0.641	0.787	0.417
Continuum Category 8	2.464	0.920	0.008*
Continuum Category 9	1.409	0.897	0.117

* Significant value

2016

Racial Composition

For every percentage increase in the difference between the proportion of White and Black populations, the Democratic vote percentage decreases by approximately 0.39% (Figure 3). This effect is statistically significant ($p < 0.001$), indicating that areas with a higher proportion of White people relative to Black people tend to have lower Democratic vote percentages. For every percentage increase in the proportion of the Hispanic population, the Democratic vote percentage decreases by approximately 0.40%. This effect is statistically significant ($p < 0.001$), indicating that areas with a higher proportion of Hispanic people tend to have lower Democratic vote percentages.

Economic & Educational Factors

A county with 1 percentage point above average employment in goods-producing industries is associated with a 0.189 percentage point lower Democratic vote share, assuming average goods-producing employment. A county with 1 percentage point above average college-educated population is associated with a 0.571 percentage point higher Democratic vote share, assuming average education levels. Both of these coefficients are also statistically significant.

Rural-Urban Continuum Code

In the basic regression models, rurality showed strong negative effects on Democratic vote share, with more rural categories having increasingly negative coefficients. However, when controlling for demographic and socioeconomic factors in the multilevel model, these effects not only became positive but also showed a non-linear pattern. Most rural categories showed positive coefficients, and categories 2-4, 6, and 8 remained statistically significant ($p < 0.05$). Category 5 became significant ($p = 0.015$), while categories 7 and 9 were not ($p = 0.417$ & $p = 0.117$, respectively). This reversal suggests that the apparent negative relationship between rurality and Democratic voting in the basic model was largely capturing the effects of other variables, particularly race, education, and employment in goods-producing industries, rather than rurality itself. The relationship between rurality and voting patterns is thus more complex than initially suggested by the basic regression model.

State Level Differences

The variance of the random intercept for state is 58.51, with a standard deviation of 7.649. This indicates substantial variability in the intercept across different states, suggesting that state-level differences are important in explaining vote percentage. The unadjusted ICC of 0.206 indicates that about 21% of the variance can be attributed to differences between states (see Appendix Figure 5). The adjusted ICC of 0.620 suggests that about 62% of the total variance in

Democratic vote percentage can be attributed to differences between states after accounting for fixed effects (race, employment, education, etc.).

Model Fit

The marginal R-squared of 0.667 indicates that about 67% of the variance in Democratic vote percentage is explained by the fixed effects alone (see Appendix Figure 5). The conditional R-squared of 0.873 indicates that about 87% of the variance in Democratic vote percentage is explained by both fixed and random effects combined, showing a strong overall model fit.

Figure 4

2020 Election Multilevel Model Results

Random Effects			
Group	Parameter	Variance	Std.Dev.
State	(Intercept)	61.30	7.829
Residual		36.03	6.003
Number of observations (counties): 1061			
Number of groups (State): 14			
Fixed Effects			
Intercept	62.63	2.269	<0.001*
White-Black Proportion Difference	-0.386	8.528e-03	<0.001*
Hispanic Population	0.316	0.028	<0.001*
Goods-Producing Industry Employment	0.036	0.035	<0.001*
Education	0.645	0.027	<0.001*
Continuum Category 2	3.103	0.699	<0.001*
Continuum Category 3	1.870	0.722	0.01*
Continuum Category 4	2.652	0.809	0.001*
Continuum Category 5	3.352	1.424	0.019*
Continuum Category 6	1.718	0.714	0.016*
Continuum Category 7	0.569	0.794	0.474
Continuum Category 8	1.844	0.922	0.046*
Continuum Category 9	0.145	0.893	0.871

* Significant value

2020

Racial Composition

For every percentage increase in the difference between the proportion of White and Black populations, the Democratic vote percentage decreases by approximately 0.39% (Figure 4). This effect is statistically significant ($p < 0.001$), indicating that areas with a higher proportion of White people relative to Black people tend to have lower Democratic vote percentages. For every percentage increase in the proportion of the Hispanic population, the Democratic vote percentage decreases by approximately 0.32%. This effect is statistically significant ($p < 0.001$), indicating that areas with a higher proportion of Hispanic people tend to have lower Democratic vote percentages.

Economic & Educational Factors

A county with 1 percentage point above average employment in goods-producing industries is associated with a 0.267 percentage point lower Democratic vote share, assuming average goods-producing employment. A county with 1 percentage point above average college-educated population is associated with a 0.645 percentage point higher Democratic vote share, assuming average education levels. Both of these coefficients are also statistically significant.

Rural-Urban Continuum Code

Most rural categories showed positive coefficients, and categories 2-4, 6, and 8 remained statistically significant ($p < 0.05$). Category 5 became significant ($p = 0.019$), while categories 7 and 9 were not ($p = 0.474$ & $p = 0.871$, respectively).

State Level Differences

The variance of the random intercept for state is 61.30, with a standard deviation of 7.829. This again indicates that there is substantial variability in the intercept across different states. The unadjusted ICC of 0.203 indicates that about 20% of the variance can be attributed to differences between states (see Appendix Figure 5). The adjusted ICC of 0.630 suggests that about 63% of the total variance in Democratic vote percentage can be attributed to differences between states after accounting for fixed effects.

Model Fit

The marginal R-squared of 0.679 indicates that about 68% of the variance in the Democratic vote percentage is explained by the fixed effects alone (see Appendix Figure 5). The conditional R-squared of 0.881 indicates that about 88% of the variance in Democratic vote percentage is explained by both fixed and random effects combined, showing a strong overall model fit.

Several notable changes emerged after adding state-level predictors to the multilevel model (see Figure 5). For 2016, the state-level variance decreased from 58.51 to 14.22, indicating that the added state-level variables explain approximately 75% of the previously unexplained variation between states. The county-level coefficients remained relatively stable, with only minor changes (ie. White-Black proportion from -0.392 to -0.393, Hispanic population from -0.403 to 0.409). The rurality effects maintained their non-linear pattern and significance levels, with categories 7 and 9 remaining non-significant. Most notably, the opposing effects of racial composition emerged at different geographic scales. While counties with higher White populations tended to vote less Democratic, states with higher White populations showed a positive relationship with Democratic voting. Similarly, the Hispanic population variable showed contrasting effects at the county and state levels. The county-level education variable showed almost identical effects, while the state-level education variable was not statistically significant ($p = 0.453$). The county-level goods-producing employment showed similar effects, while the state-level employment variable was not statistically significant ($p = 0.832$).

The unadjusted ICC decreased from 0.206 to 0.076, and the adjusted ICC decreased from 0.620 to 0.284 (see Appendix Figure 10). This decrease indicates that adding state-level predictors explained much of the clustering effect at the state level. The remaining ICC represents the unexplained similarity in voting patterns among counties in the same state after accounting for state-level demographic and economic characteristics. The marginal R-squared increased from 0.667 to 0.731 (see Appendix Figure 10). This indicates that adding state-level predictors enhanced the model's ability to explain voting patterns through fixed effects. There was a slight decrease in conditional R-squared (0.873 to 0.808). The model explained slightly less total variance, but more of the explained variance came from fixed effects (both county

Figure 5

2016 Election Multilevel Model Results with State Level Predictors

Random Effects

Group	Parameter	Variance	Std.Dev.
State	(Intercept)	14.22	3.771
Residual		35.88	5.990

Number of observations (counties): 1061
Number of groups (State): 14

Fixed Effects

Variable	Mean (1)	Standard Error (2)	P-Value (3)
Intercept	42.49	6.491	<0.001*
White-Black Proportion Difference	-0.393	8.429e-03	<0.001*
State Level White-Black Proportion Difference	0.282	0.087	0.012*
Hispanic Population	0.409	0.029	<0.001*
State Level Hispanic Population	-0.434	0.163	0.025*
Goods-Producing Industry Employment	-0.194	0.035	<0.001*
State Level Good-Producing Industry Employment	-0.078	0.358	0.832
Education	0.566	0.028	<0.001*
State Level Education	0.252	0.321	0.453
Continuum Category 2	2.948	0.694	<0.001*
Continuum Category 3	1.736	0.718	0.016*
Continuum Category 4	2.649	0.805	0.001*
Continuum Category 5	3.419	1.419	0.016*
Continuum Category 6	1.740	0.709	0.014*
Continuum Category 7	0.610	0.789	0.439
Continuum Category 8	2.453	0.921	0.008*
Continuum Category 9	1.370	0.897	0.127

* Significant value

and state-level predictors). Less of the explained variance relied on unexplained state-level differences (random effects).

After adding state-level predictors to the 2020 multilevel model, the state-level variance decreased from 61.30 to 13.90, indicating that the added state-level variables explain about 77% of the previously unexplained variation between states. The county-level coefficients remained relatively stable (ie. White-Black proportion from -0.386 to -0.387, Hispanic population from 0.316 to 0.321). The rurality effects maintained their non-linear pattern and significance levels, with categories 7 and 9 remaining non-significant. Again, the opposing effects of racial composition emerged at different geographic scales. While counties with higher White populations tended to vote less Democratic (-0.387), states with higher White populations showed a positive relationship with Democratic voting (0.297). The county-level coefficient changed from slightly positive to significantly negative, and a new state-level employment effect emerged ($p < 0.001$). County-level education remained significant, and state-level education was not statistically significant ($p = 0.259$).

The unadjusted ICC decreased from 0.203 to 0.066, and the adjusted ICC decreased from 0.630 to 0.278. Again, this decrease indicates that adding state-level predictors explained much of the clustering effect at the state level. This clustering effect and its reduction help to understand how state-level factors influence local voting patterns and the importance of considering both county and state-level characteristics in analyzing electoral behavior. The marginal R-squared increased from 0.678 to 0.761 (see Appendix Figure 10). There was a slight decrease in conditional R-squared again (0.881 to 0.828) suggesting that while the fixed effects explain more variance, the overall model fit remains strong but with less reliance on random effects.

Figure 6

2020 Election Multilevel Model Results with State Level Predictors

Random Effects			
Group	Parameter	Variance	Std.Dev.
State	(Intercept)	13.90	3.728
Residual		36.04	6.003

Number of observations (counties): 1061
Number of groups (State): 14

Fixed Effects			
Variable	Mean (1)	Standard Error (2)	P-Value (3)
Intercept	42.78	6.399	<0.001*
White-Black Proportion Difference	-0.387	8.528e-03	<0.001*
State Level White-Black Proportion Difference	0.297	8.528e-03	0.009*
Hispanic Population	0.321	0.028	<0.001*
State Level Hispanic Population	-0.304	0.161	0.0889
Goods-Producing Industry Employment	-0.271	0.036	<0.001*
State Level Goods-Producing Industry Employment	-0.049	0.359	<0.001*
Education	0.641	0.027	<0.001*
State Level Education	0.366	0.303	0.259
Continuum Category 2	3.106	0.699	<0.001*
Continuum Category 3	1.911	0.722	0.008*
Continuum Category 4	2.649	0.808	0.001*
Continuum Category 5	3.311	1.423	0.020*
Continuum Category 6	1.707	0.714	0.017*
Continuum Category 7	0.544	0.794	0.493
Continuum Category 8	1.831	0.922	0.047*
Continuum Category 9	0.108	0.893	0.904

* Significant value

Discussion

Model Results

Rural-Urban Continuum Model

The linear regression results revealed important shifts in the rural-urban voting divide between 2016 and 2020 and gave an overview of the voting patterns among rural and urban counties. The increase in baseline Democratic vote share in urban counties from 42.25% to 45.79% suggests a strengthening Democratic advantage in urban areas, possibly reflecting growing urban-rural polarization. This aligns with broader national trends of geographic sorting, where urban areas have become increasingly Democratic whereas rural areas have maintained or strengthened their Republican support.

The intensification of negative effects in more rural categories, particularly the change in Category 9 from -12.20% to -14.09%, indicates a deepening political divide between urban and rural communities (see Appendix Figure 13 for category codes). The strong negative effects in categories 6 and 7 (-12.70% and -12.82% in 2020) suggest that mid-rural counties, often characterized by smaller populations and less connection to metropolitan areas, have become strong centers of Republican support. Category 5 was non-significant in both 2016 and 2020, suggesting that medium-sized nonmetropolitan areas not adjacent to metro regions may follow distinct political patterns that don't align with the general rural-urban trend.

Base Multilevel Model

The linear regressions showed consistently negative effects of rurality on Democratic voting. However, when demographic variables (race, education, and employment) were added to the base multilevel models, these effects reversed, with most rural categories showing positive effects. This reversal suggests that the apparent negative relationship between rurality and Democratic voting in the simple linear models was not capturing the true effect of rurality itself, but rather the demographic characteristics that tend to correlate with rural areas.

The low R-squared values in the linear models compared to the much higher explanatory power of the multilevel models with demographic variables indicate that rurality alone is a weak predictor of voting behavior (see Appendix Figure 10). This finding has important implications for electoral analysis. Most importantly, using rurality as a standalone predictor may lead to misleading conclusions about voting patterns, as it masks the underlying demographic and socioeconomic factors that more directly influence political preferences. The transformation of rural effects from negative to positive when controlling for demographics suggests that the narrative of a rural-urban divide in American politics is oversimplified and that the complex interactions between demographic, economic, and geographic factors better explain voting behavior.

State Level Predictors

When adding state-level variables there were opposing coefficients between state and county-level predictors. For example, in 2016, at the county level, a higher White-Black proportion (greater White population than Black population) was associated with a lower Democratic vote share, but at the state level, a higher White-Wlack proportion was associated with higher Democratic vote share (0.282). This occurred with the Hispanic population variable as well, and in all instances, these coefficients were significant (except for the state-level Hispanic predictor in 2020). This difference demonstrates a paradox where the relationship between racial composition and voting behavior operates differently at different geographic scales: the county level and state level. The opposing effects suggest that while predominantly White counties tend to vote Republican, states with higher White populations tend to have other characteristics that are associated with more Democratic voting patterns. This could be due to urbanization, education, and various economic factors.

One explanation for this is that White voters in "Blue Wall" states vote differently than those in Southern states. The Blue Wall refers to the states that have voted Democrat in the six consecutive presidential elections from 1992 to 2012. Swing states included in the Blue Wall are Wisconsin, Pennsylvania, and Michigan, and southern swing states include Georgia, North Carolina, and Florida. Also, states with larger White populations often have higher proportions of college-educated White voters, who have shifted heavily away from Trump. The White population in these states tends to be more urban and suburban, with different voting patterns than rural White voters. The opposing Hispanic coefficients also reflect complex demographic and political patterns. At the county level, higher Hispanic populations correlate with higher Democratic support, but at the state level, higher Hispanic populations correlate with lower Democratic support. This pattern could be explained by different voting patterns among Hispanic subgroups or geographic concentration effects, where Hispanic voters in some states are more politically diverse.

The substantial reduction in state-level variance suggests that these demographic patterns at the state level capture important contextual effects that influence voting behavior. The stability of county-level coefficients indicates that local demographic and economic factors maintain their influence even when accounting for broader state-level patterns, demonstrating the complex interplay between local and state-level characteristics in shaping voting behavior.

Comparing the 2016 & 2020 Elections

Comparing the multilevel models between the 2016 and 2020 elections revealed both similarities and differences. In the base models without state predictors, the White-Black proportion effect remained notably stable, while the Hispanic population effect showed a dramatic shift from negative to positive. Education's positive influence was slightly strengthened, and goods-producing employment underwent a change from negative to positive. The state-level variance and residual variance remained relatively constant. Looking at the rural-urban continuum categories, Category 2 remained strongly positive and Category 5 showed the strongest positive effect. Categories 7 and 9 remained non-significant in both years, and all other categories show significant positive effects.

When state-level predictors were added, the county-level effects remained relatively stable, with the White-Black proportion maintaining its negative influence and the Hispanic population maintaining a positive influence but decreasing slightly. Goods-producing employment showed stronger negative effects, while education maintained its positive influence. At the state level, the White-Black proportion demonstrated consistent positive effects, while the Hispanic population's negative effect weakened and was not significant. Goods-producing employment became significant at the state level in 2020, while education remained non-significant in both years. With the addition of state-level predictors, rural-urban continuum Category 2 stayed strongly positive, and Category 5 maintained the highest coefficient. Categories 7 and 9 remained non-significant. The addition of state-level predictors substantially reduced state-level variance in both years, while residual variance remained stable, indicating the importance of state-level characteristics in explaining voting patterns.

The Hispanic population effect underwent the most dramatic transformation, shifting from negative to positive at the county level between elections, meaning they leaned Republican in 2016 but leaned Democrat in 2020. However, in 2020 they still leaned Republican at the state level. This suggests changing Hispanic voting preferences and potentially different geographic patterns between the two election cycles. There could be regional divides as well as possible educational and economic divides. Goods-producing employment showed a notable reversal, changing from leaning Republican to Democrat in the base model, but becoming strongly negative when state predictors were added in 2020. This shift possibly reflects the impact of trade policies, COVID-19's economic effects, and increased uncertainty in the manufacturing industries. The education effect also strengthened between elections, indicating a growing educational divide in voting patterns. This could be partially attributed to the influx of Gen Z voters, who have the highest college enrollment rates of any generation. This demographic shift adds to the growing pool of college-educated voters who tend to vote more Democratic, potentially amplifying the relationship between education levels and Democratic vote share at the county level.

County Flips

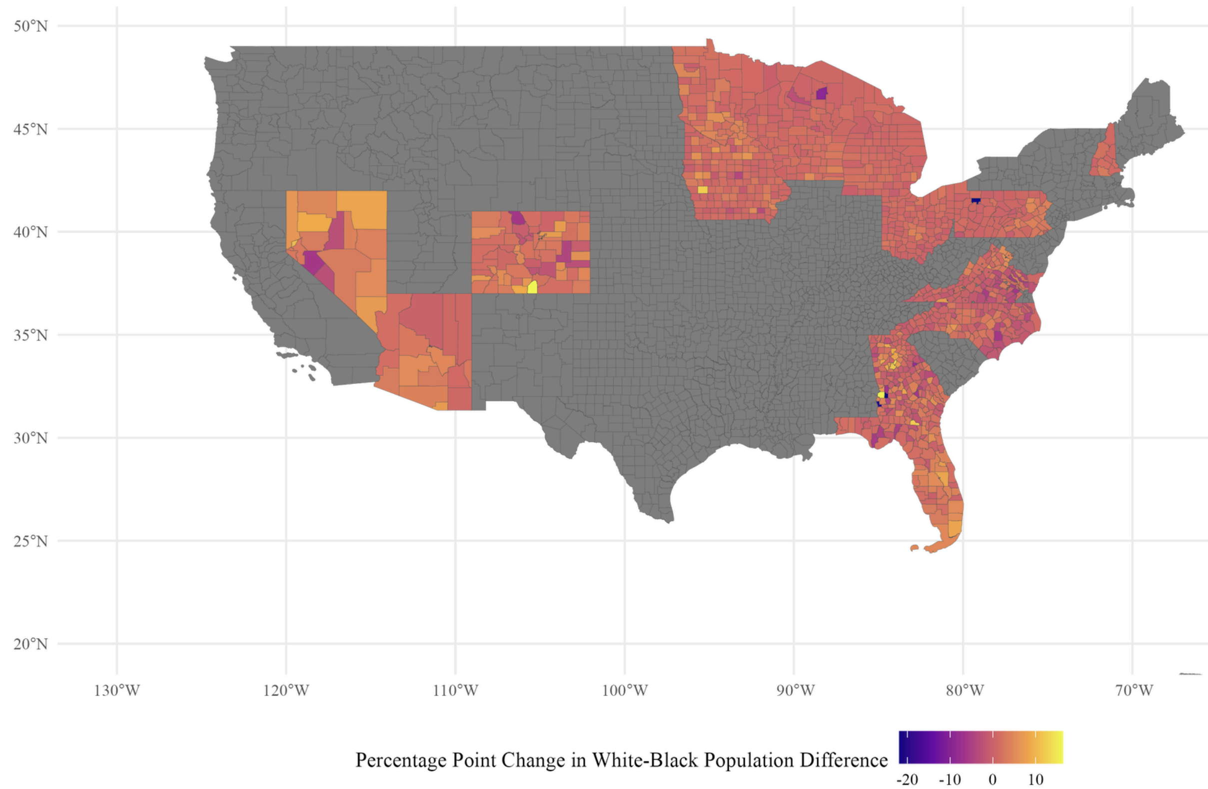
Between the 2016 and 2020 Presidential elections, there were 49 counties in the 14 examined swing states that flipped the party they supported. Only two counties in swing states went from voting majority Democrat in 2016 to majority Republican in 2020, while forty-seven counties flipped from voting majority Republican in 2016 to majority Democrat in 2020. Some notable counties include Maricopa County in Arizona, Bucks County in Pennsylvania, Cobb County in Georgia, and Kent County in Michigan.

The United States Department of Agriculture does not specify whether each rural-urban continuum code can be classified as urban, suburban, or rural. Taking a more conservative approach, assuming that categories 1 and 2 are urban, 5.06% of urban counties flipped from Republican to Democrat. Assuming that categories 3 through 5 are suburban, 6.93% of suburban counties flipped from Republican to Democrat. If we only classified category 1 as urban, 2.96% of urban counties flipped, and 9.03% of suburban counties (categories 2 through 5) flipped. This highlights a suburban shift towards the Democratic party, a key factor to Joe Biden's win.

Two significant factors to this change included a shift in the difference between the proportion of White people and Black people in a county. This difference decreased between 2016 and 2020 (Figure 7). The other shift was in education. 2020 saw a higher proportion of people over the age of 25 with a college degree or higher (Figure 8).

Figure 7

Change in White-Black Population Proportion Difference, 2016-2020

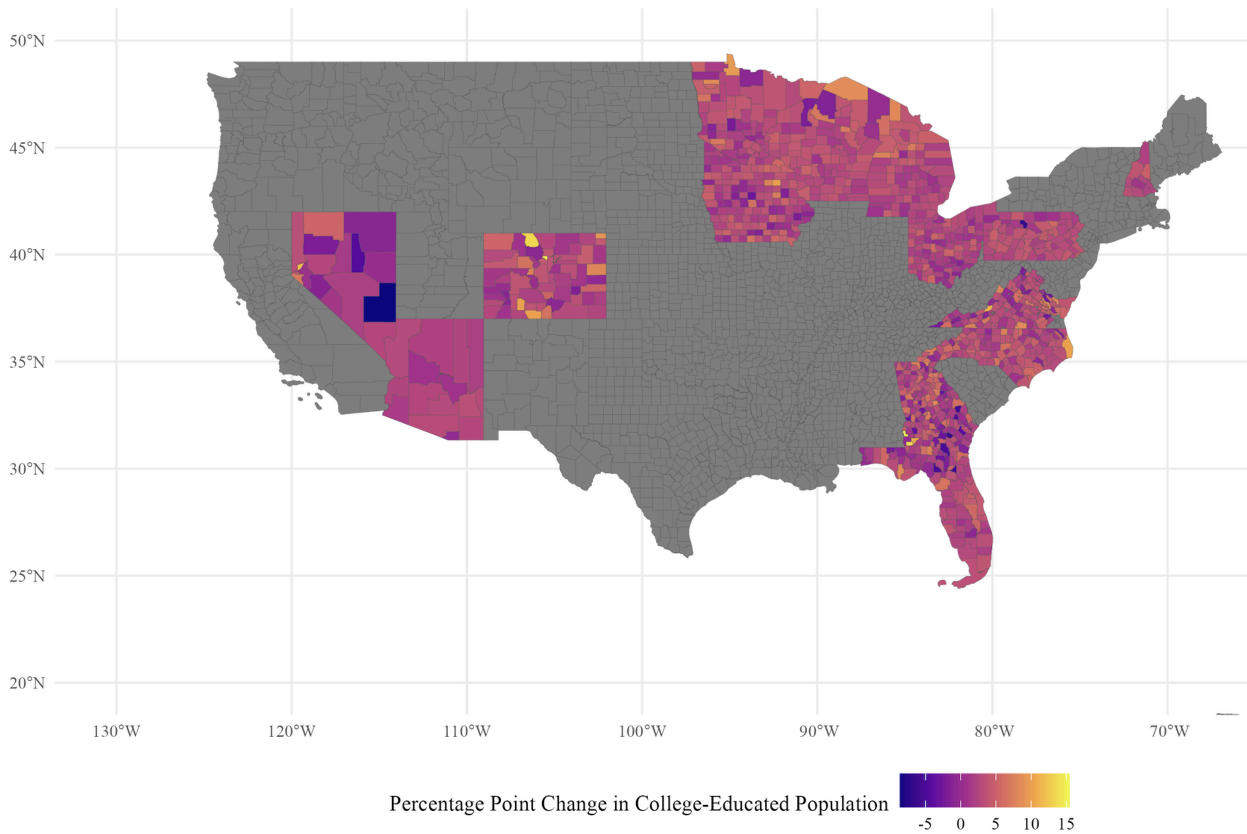


This figure illustrates the change in the difference between the proportion of the White population and the Black proportion between 2016 and 2020. It is organized by county with positive values indicating a narrowing gap between the two populations.

Within the five states that flipped from Republican to Democrat in 2020 (Arizona, Georgia, Michigan, Pennsylvania, and Wisconsin), sixteen counties flipped from Republican to Democrat. Every single one of these counties saw a decrease in the difference between the proportion of Black and White people (narrowing gap). This connects back to the significant increases seen in the coefficients for this variable between 2016 and 2020. All of these counties also saw a slight increase in their Hispanic populations. The proportion of people working in the goods producing industries both increased and decreased, however, all of these counties saw an increase in education. Again, this is consistent with the significant increases seen in the coefficients for education between 2016 and 2020.

Figure 8

Change in Population With College Degrees, 2016-2020



This figure illustrates the change in the population 25 years or older with a college degree between 2016 and 2020. It is organized by county and positive values indicate an increase in the percentage of college-educated residents from 2016 to 2020, while negative values indicate a decrease in the percentage of college-educated residents.

Another factor that may have heavily contributed to the swing state flips is candidate incumbency. In the 2016 presidential election, neither major party candidate was an incumbent president. Hillary Clinton was running as a quasi-incumbent due to her role as Secretary of State in the outgoing Obama administration and her husband's previous presidency. Donald Trump was a political outsider with no prior experience in public office. In the 2020 election, the dynamics of incumbency were even more unique. Donald Trump was the incumbent president seeking re-election, while Joe Biden, although not the sitting president, brought significant executive branch experience as the former vice president under the Obama administration. In the 2024 presidential election, Donald Trump was running as the incumbent president, seeking a second, non-consecutive term after losing in 2020. On the Democratic side, Vice President Kamala

Harris became the nominee after President Joe Biden withdrew from the race in July 2024. Harris, while not the incumbent president, held a position of quasi-incumbency similar to Hillary Clinton in 2016 and Joe Biden in 2020.

The unique incumbency situation in each election cycle led to various circumstances which can be difficult to measure. For example, in 2016 Trump represented change but in 2020 Trump as the incumbent bore the blame for national issues, contributing to his loss. In 2024, Harris, though not the president, was associated with the Biden administration's policies, making it difficult for her to distance herself from unpopular decisions. In many swing states, Biden's 2020 win against an unpopular incumbent was reversed in 2024 when he became the unpopular incumbent, indirectly affecting Harris. Additionally, the transition from Biden to Harris mid-campaign may have affected voter confidence in Democratic leadership.

2024 Presidential Election Results

On November 5, 2024, the seven swing states being closely watched for the presidential election were won by President-elect Donald Trump. This included Arizona, Georgia, Michigan, Pennsylvania, and Wisconsin which were won by Donald Trump in 2016, won by President Joe Biden in 2020, and again won in 2024 by Donald Trump. Across the country, Donald Trump saw shifts in his favor, even in historically blue states. Even in the swing cities where he lost to Vice President Kamala Harris, it was by a smaller margin than he lost to Joe Biden in 2020 (Levitt et al., 2024). Trump gained in both urban and suburban counties and continued his gains in rural counties. Hispanic-majority counties saw a substantial shift towards support for Trump, and he also improved in counties with high and low proportions of college-educated residents. The shift from Republican to Democrat support, seen in the base model for those employed in manufacturing industries, was reversed in 2024 with Trump seeing support from manufacturing counties as well as mining and farming counties.

Conclusion

This research revealed the complexity of presidential election voting patterns across swing states in the 2016 and 2020 elections. Past research has debated whether the predictor of electoral outcomes lies in the rurality of a county or its more complex demographics (ie. race, education, etc.). Initial linear regressions suggested a simple rural-urban divide, with increasingly negative effects on Democratic voting as rurality increased. However, multilevel models demonstrated that fundamental demographic and socioeconomic factors are behind this apparent geographic divide. When accounting for race, education, and employment, the relationship between rurality and voting behavior not only reversed but revealed intricate patterns at both county and state levels. The addition of state-level predictors explained substantial portions of between-state variance and highlighted how demographic effects operate differently at various geographic scales. The strengthening education effect, the shifting Hispanic vote patterns, and the variable impact of goods-producing employment suggest that voting behavior in swing states is shaped by a complex interaction of local and state-level characteristics rather than simple geographic divisions. It is important to consider multiple levels of analysis when studying voting behavior, especially when examining something as consequential as swing state outcomes. Doing so challenges commonly cited oversimplified narratives about America's political geography. The 2024 presidential election showed that these relationships continue to evolve and nuanced analytical approaches are important in understanding America's dynamic political landscape.

References

Albrecht, Don E. “Donald Trump and changing rural/urban voting patterns.” *Journal of Rural Studies*, vol. 91, 2022, https://doi.org/10.1016/j.jrurstud.2022.03.009

Anselin, Luc. 2002. Under the hood: Issues in the specification and interpretation of spatial regression models. Agricultural Economics 27:247–67.

Anselin, Luc. 2005. Exploring spatial data with GeoDa. Urbana-Champaign, IL: Spatial Analysis Lab, Department of Agricultural and Consumer economics, University of Illinois.

Brown TE, Mettler S. Sequential Polarization: The Development of the Rural-Urban Political Divide, 1976–2020. *Perspectives on Politics*. 2024;22(3):630-658. doi:10.1017/S1537592723002918

Chavez, L., 2013. The Latino Threat: Constructing Immigrants, Citizens, and the Nation. Stanford University Press.

Getis, A. (2010). Spatial Autocorrelation. In: Fischer, M., Getis, A. (eds) Handbook of Applied Spatial Analysis. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-03647-7_14

Gould, E. (2019, March 27). Decades of rising economic inequality in the U.S.: Testimony before the U.S. House of Representatives Ways and Means Committee. Economic Policy Institute. https://www.epi.org/publication/decades-of-rising-economic-inequality-in-the-u-s-testimony-before-the-u-s-house-of-representatives-ways-and-means-committee/

Iversen, Torben, and David Soskice. 2019. Democracy and Prosperity: Reinventing Capitalism Through a Turbulent Century. Princeton, NJ: Princeton University Press.

Levitt, Z., Collins, K., Gebeloff, R., Khurana, M., & Hernandez, M. (2024, November 6). See the voting groups that swung to the right in the 2024 vote. The New York Times. https://www.nytimes.com/interactive/2024/11/06/us/elections/trump-america-red-shift-victory.html

Massey, Douglas, 2007. Categorically Unequal. Russel Sage Foundation, New York.

Pew Research Center. 2014. Political polarization in the American public: How increasing ideological uniformity and partisan antipathy affect politics, compromise and everyday life. Washington, DC: Pew Research Center. Available from pewresearch.org.

Pew Research Center. 2024. Changing Partisan Coalitions in a Politically Divided Nation. Washington, DC: Pew Research Center. Available from pewresearch.org.

Scala, D. J., & Johnson, K. M. (2017). Political Polarization along the Rural-Urban Continuum? The Geography of the Presidential Vote, 2000–2016. The ANNALS of the American Academy of Political and Social Science, 672(1), 162–184. doi:10.1177/0002716217712696

Scala, Dante, Kenneth M. Johnson, and Luke T. Rogers. 2015. Red rural, blue rural? Presidential voting patterns in a changing rural America. Political Geography 48:108–18.

https://www.270towin.com/same-since-electoral-maps/

U.S. Census Bureau. (2024). Urban and rural. U.S. Department of Commerce. https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html

Appendix

Figure 9

Summary Statistics

Year	Unadjusted ICC (1)	Adjusted ICC (2)	Marginal R-Squared (3)	Conditional R-Squared (4)
2016	0.206	0.620	0.667	0.873
2020	0.203	0.630	0.678	0.881

Figure 10

Summary Statistics with State Level Predictors

Year	Unadjusted ICC (1)	Adjusted ICC (2)	Marginal R-Squared (3)	Conditional R-Squared (4)
2016	0.076	0.284	0.731	0.808
2020	0.066	0.278	0.761	0.828

Figure 11

Multicollinearity Test Using Variance Inflation Factor

Year	Variable (1)	VIF (2)	Df (3)	Adjusted VIF (4)
2016	White-Black Proportion Difference	1.10	1	1.049
	Hispanic	1.045	1	1.022
	Goods-Producing Industry Employment	1.790	1	1.338
	Education	1.986	1	1.409
	Rural-Urban Continuum	1.512	8	1.026
2020	White-Black Proportion Difference	1.118	1	1.057
	Hispanic	1.051	1	1.025
	Goods-Producing Industry Employment	1.799	1	1.341
	Education	2.001	1	1.415
	Rural-Urban Continuum	1.518	8	1.026

Figure 12

Spatial Autocorrelation Test Using Moran's I

Year	Variable (1)	Moran's I (2)	P-Value (3)
2016	White-Black Proportion Difference	0.0255	0.085
	Hispanic	0.002	0.433
	Goods-Producing Industry Employment	0.015	0.205
	Education	0.005	0.371
2020	White-Black Proportion Difference	0.027	0.073
	Hispanic	3.580e-05	0.480
	Goods-Producing Industry Employment	0.011	0.276
	Education	0.003	0.412

Figure 13

USDA Rural-Urban Continuum Codes

Category	Code
1	Metro - Counties in metro areas of 1 million population or more
2	Metro - Counties in metro areas of 250,000 to 1 million population
3	Metro - Counties in metro areas of fewer than 250,000 population
4	Nonmetro - Urban population of 20,000 or more, adjacent to a metro area
5	Nonmetro - Urban population of 20,000 or more, not adjacent to a metro area
6	Nonmetro - Urban population of 2,500 to 19,999, adjacent to a metro area
7	Nonmetro - Urban population of 2,500 to 19,999, not adjacent to a metro area
8	Nonmetro - Completely rural or less than 2,500 urban population, adjacent to a metro area
9	Nonmetro - Completely rural or less than 2,500 urban population, not adjacent to a metro area

Figure 14

Independent Variables Correlation Matrix