Vote Switching in the United States: Insights from ANES and GSS

1. Introduction

Voter behavior can shift over time because people's opinions, personal circumstances, and the world around them are constantly changing. Traditional one-time surveys give us a snapshot of public opinion at a certain moment, but they might miss how each person's views evolve—and how that affects election outcomes. By using panel data, where we follow the same people over multiple surveys, we can watch their beliefs and demographics change from one election to the next. In this project, I utilize American National Election Studies (ANES) and the General Social Survey (GSS) data to see if tracking these shifts gives us more prediction power predicting voter behavior than just looking at a single survey wave. Specifically, I want to find out how changes in political attitudes and demographic factors lead voters to switch parties or alter their preferences with time.

I am aiming to find some answers to following questions:

1. Comparative Predictive Power: Does incorporating panel data on changes in attitudes and demographic characteristics improve the prediction of voter behavior compared to using single cross-sectional survey data?

2. Drivers of Volatility: Which shifts—whether attitudinal (e.g., political ideology, issue positions) or demographic (e.g., income, education, age)—most strongly influence changes in voting preferences over time?

3. Impact of External Events: How do major political or economic events interact with individual-level changes to shape voter volatility, and do these external factors amplify or mitigate the impact of personal demographic and attitudinal shifts?

4. Methodological Considerations: What are the advantages and limitations of panel data approaches versus cross-sectional methods in capturing voter volatility, and how can future research leverage these findings to refine predictive models of electoral behavior?

2. Prior Work

Existing research on voter behavior has traditionally leaned on cross-sectional data, which offers a single snapshot of public opinion or party preference at one point in time. These methods can illuminate broad trends across populations but may miss the nuances of how individuals evolve in their views or party alignments (Bartels, 2002; Lewis-Beck, Norpoth, Jacoby, & Weisberg, 2008).

Many of these studies focus on demographic predictors and issue attitudes measured in a single wave of data collection, providing insights into factors like education, income, and ideology, but not necessarily capturing the trajectories that lead voters to switch or stay loyal over multiple election cycles.

More recent work highlights the importance of panel data, which follows the same respondents over multiple waves, allowing researchers to observe changes in beliefs and behaviors directly (Wooldridge, 2010). This approach helps control for unobserved individual differences that might skew results in cross-sectional studies. By tracking the same people over time, scholars can better isolate the effects of major events, shifting personal circumstances, or gradual ideological realignments. For instance, Dinas (2014) demonstrates how continuous electoral participation can reinforce or reshape party loyalties, while Sides, Tesler, and Vavreck (2018) show how evolving social identities interact with campaign events to affect vote choice. These panel-based insights highlight the potential for more accurate models of voter behavior, as they capture both the short-term shocks and the longer-term evolutions that shape people's political decisions.

3. Data

The American National Election Studies is a long-standing project that tracks how people in the United States think and vote around national elections. It's been running for several decades, so it contains a wealth of information on party preferences, attitudes toward various political issues, and core demographics. Because it surveys people close to election time, ANES is often used to spot trends in voting behavior and shifts in public opinion from one election cycle to the next.

The General Social Survey gathers information on a broad range of topics—from religion and social values to economic conditions—allowing researchers to see how these cultural and societal factors change over time. While it doesn't focus on voting and political behavior to the same extent as ANES, it still includes enough political questions to help us understand how larger social currents connect to electoral choices.

For our specific goals, we're looking for panel or longitudinal data: surveys where the same individuals are interviewed multiple times. This lets us see how a person's demographic background, political beliefs, and actual voting decisions shift (or stay the same) as the years—and elections—go by. By tracking these changes at the individual level, we can better understand why some voters switch parties while others remain loyal, and how different factors combine to shape people's electoral behavior.

4. Methodology

We began our analysis using the ANES 2016–2020 merged panel dataset, which follows the same respondents across two presidential elections. Since many survey items remained consistent in both 2016 and 2020, we were able to compare individuals' responses over time. Our first step involved separate Exploratory Data Analyses (EDA) for each election year, looking at how factors like gun control, healthcare, immigration, demographic traits, and social media usage might influence voting along a seven-point scale running from Republican to Democrat (with Independents in the middle). We also considered "thermometer" ratings—measures of how warmly (or coldly) respondents felt about different social groups and institutions, including conservative and liberal groups, religious communities, the police, and the Black Lives Matter movement.

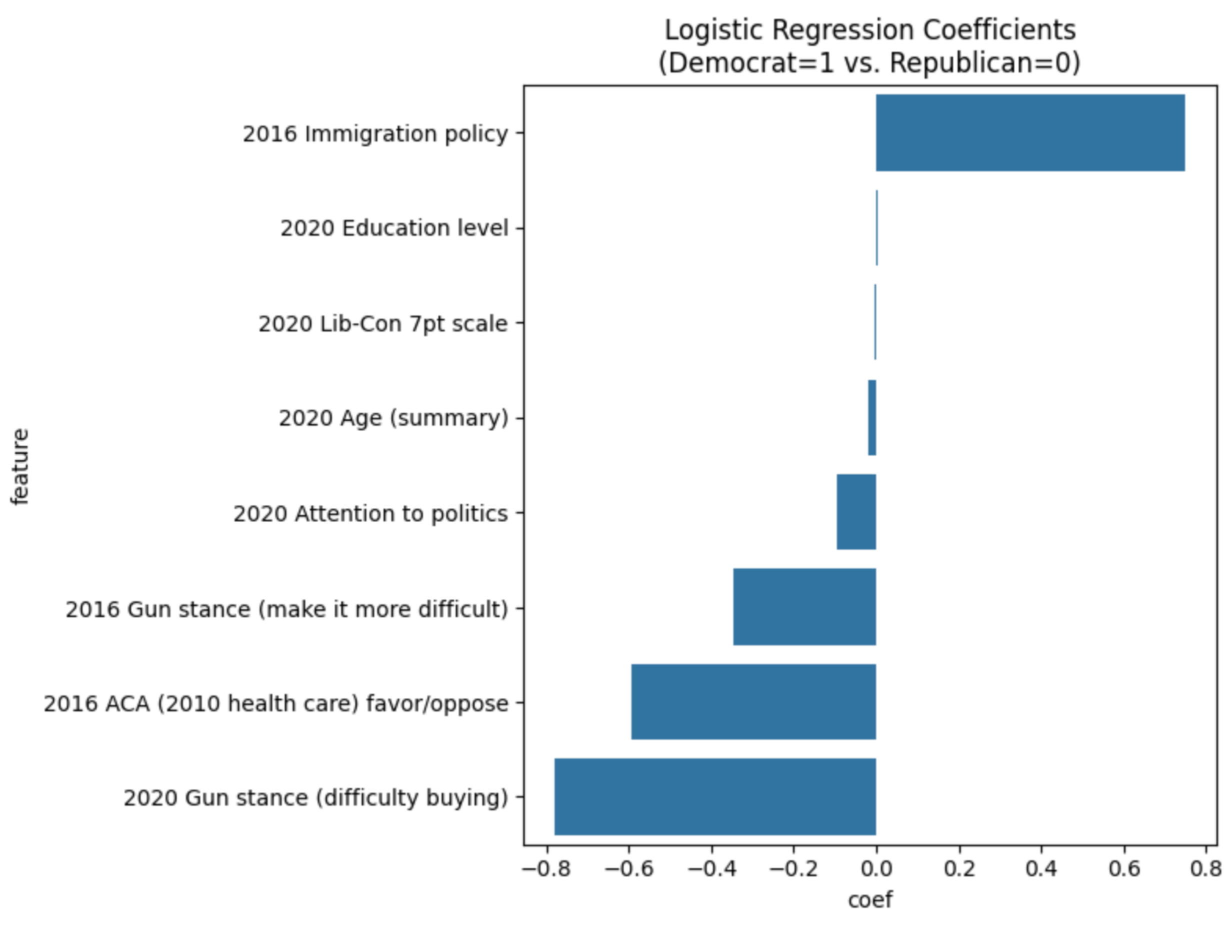

Figure showing the coefficients of some of the selected features. Visuals for EDA and more are not included in the report but you can find them on the figures folder.

When we checked whether voters switched parties between the two election cycles, previous vote surfaced as a particularly strong predictor: in most cases, people stayed consistent, but those who did change had a significant effect in close races. This finding led us to pay special attention to how much individual attitudes had shifted on key issues. Even modest changes in favorability ratings or policy views can act as early signals that someone might be on the verge of switching political alignments.

To dig deeper, we ran multiple modeling strategies. One approach used only 2016 data—such as stances on guns or healthcare—to predict the 2020 vote choice. Another focused on changes in respondents' attitudes from 2016 to 2020, testing whether these shifts tracked with a move along the seven-point scale. Finally, a third strategy combined both static positions and differences, letting us capture not just the end points but also the journey between them. By examining which variables stayed stable versus which ones shifted, we gained a clearer picture of why some voters altered their party preferences and how their evolving views ultimately influenced their electoral choices.

5. Findings

Before diving into the individual scenario results, it's important to clarify our overall approach. We ran multiple models to predict whether a respondent would vote Democrat or Republican in 2020, using data from both 2016 and 2020. Each "scenario" represents a different set of predictor variables—such as only the 2016 features, only the 2020 features, the year-to-year changes (differences), or various combinations that include an individual's 2016 vote choice. By comparing the accuracy and other metrics across these scenarios, we can see how adding or removing certain types of information affects the model's ability to correctly categorize voters. Below is a discussion of each scenario and the key takeaways from our findings.

Scenario 0: 2016 Features Only

- Feature Set: All 2016 issue stances, demographics, and thermometer scores (no previous vote included).
- Accuracy: 88.3%

This baseline model already gives decent performance, suggesting that someone's positions on issues in 2016 are fairly good predictors of how they vote later. However, the absence of prior vote data limits the model's ability to capture a person's established voting tendency.

Scenario 1: 2016 Features + 2016 Vote

- Feature Set: Same as Scenario 0, plus 16VoteBinary.
- Accuracy: 93.1%

Here, adding the 2016 vote choice boosts accuracy by nearly five percentage points. This substantial jump indicates that knowing whether someone voted Democrat or Republican in 2016 is highly predictive of their 2020 choice. The classification report shows both high precision and recall for each class, highlighting how prior vote information reduces misclassifications.

Scenario 2: 2020 Features Only

• Feature Set: All 2020 issue stances, demographics, and thermometer scores (no previous vote included).
• Accuracy: 91.6%

Compared to Scenario 0, using just 2020 features yields better results (91.6% vs. 88.3%). This makes sense: the closer the survey items are to the election in question (2020), the more relevant they become. However, we're still missing the direct indicator of someone's voting history, which, as we saw in Scenario 1, can have a major impact.

Scenario 3: 2020 Features + 2016 Vote

• Feature Set: Same as Scenario 2, plus 16VoteBinary.
• Accuracy: 94.0%

This is one of the top-performing scenarios, hitting 94% accuracy. Combining up-to-date issue stances with someone's past vote choice yields a strong predictive model. The classification report shows a healthy balance between precision and recall for both Democrats and Republicans, indicating that the model correctly identifies members of each group most of the time.

Scenario 4: Differences Only

• Feature Set: Year-to-year "difference" variables (how much a respondent changed on health, guns, immigration, etc.) but no 2016 vote or 2020 features explicitly.
• Accuracy: 91.9%

Most interesting scenario. Surprisingly, just looking at how a person's views shifted between 2016 and 2020 can produce an accuracy close to Scenario 2. This suggests that the act of changing a stance (e.g., moving from being neutral on gun control to strongly in favor) is quite informative when predicting who ends up voting Democrat or Republican. Still, it doesn't surpass the scenarios that include direct current positions or past votes.

Scenario 5: Differences + 2020 Features

• Feature Set: Combination of the differences and the 2020 features (but no 16VoteBinary).
• Accuracy: 92.2%

Adding the 2020 stances alongside the changes from 2016 leads to a slight improvement over just using differences alone. This indicates that both where someone ends up (2020 stance) and how much they changed can be valuable in predicting their eventual vote.

Scenario 6: Differences + 2016 Vote

- Feature Set: Combination of the differences and 16VoteBinary (but no 2020 features).
- Accuracy: 93.1%

Just like in previous comparisons, including the past vote again makes a substantial difference. While these changes in attitudes matter, the single biggest predictor remains how someone voted in the previous election. Accuracy here mirrors Scenario 1, suggesting that a strong part of the model's power comes from the vote history, supplemented by how much a person has shifted their views.

Scenario 7: Differences + 2020 Features + 2016 Vote

- Feature Set: Everything—difference variables, current (2020) positions, and 16VoteBinary.
- Accuracy: 94.0%

Tied for the best performance with Scenario 3, this "kitchen-sink" model uses all available information: how the respondent changed, where they stand now, and how they voted before. This result reaffirms that while new issue stances and changing views are important, the most crucial factor is often whether someone was already voting Democratic or Republican in the past. The combination of all three dimensions yields a robust prediction of the 2020 vote.

```
=== Final Summary ===
   Scenario  Accuracy
0         0  0.882530
1         1  0.930723
2         2  0.915663
3         3  0.939759
4         4  0.918675
5         5  0.921687
6         6  0.930723
7         7  0.939759
```

After exploring how well the eight scenarios performed using a full set of variables (including policy stances, demographics, and a wide range of thermometer scores), we wanted to see if these same scenarios would yield consistent patterns when we altered the types of variables fed into the model. In other words, we kept the same structural approach—eight different ways of combining 2016 features, 2020 features, and differences between years, plus the option to include or exclude the 2016 vote—but replaced our original "full" feature set with more targeted subsets of data.

Specifically, we repeated the same eight-scenario structure three times, but each time focusing on a different set of predictors:

Basic Policy Variables: We used only gun, health, immigration, and age information from both 2016 and 2020 (and their differences).

All Thermometer Ratings: We used the entire battery of thermometer questions (e.g., favorability towards various social groups).

Selected Thermometers: We narrowed our focus to thermometers specifically measuring attitudes about religion, LGBT issues, police, and the BLM movement.

=== Final Summary ===

	Scenario	Accuracy
0	0	0.851459
1	1	0.907162
2	2	0.899204
3	3	0.915119
4	4	0.907162
5	5	0.915119
6	6	0.909814
7	7	0.917772

=== Final Summary ===

	Scenario	Accuracy
0	0	0.867725
1	1	0.915344
2	2	0.902116
3	3	0.944444
4	4	0.740741
5	5	0.910053
6	6	0.928571
7	7	0.941799

=== Final Summary ===

	Scenario	Accuracy
0	0	0.8325
1	1	0.9025
2	2	0.8600
3	3	0.9150
4	4	0.7075
5	5	0.8750
6	6	0.9125
7	7	0.9150

Figure contains 8 scenarios with the order of selected variables.

Despite the shift in which specific variables we used, we saw that the core insights remained largely the same. Adding the respondent's previous vote (2016) consistently boosted accuracy across all feature subsets. Likewise, combining difference scores with 2020 positions tended to outperform using them in isolation, indicating that both where someone ended up in 2020 and how they changed since 2016 provide valuable signals. Even though the raw accuracy numbers varied depending on whether we used basic policy info, the full thermometer battery, or a smaller selection of thermometer items, the relative performance of each scenario mirrored the trends we identified in our full-variable analysis.

By replicating the eight scenarios with different feature sets, we showed that the primary drivers of voter classification remained intact regardless of whether we focused on broad policy stances or narrower social attitudes. This strongly suggests that the model's structure—particularly the inclusion of past voting behavior and the distinction between current stances and changes over time—is more critical than the exact choice of features. In practical terms, researchers or analysts can feel more confident that these scenarios hold up across various subsets of data: if 2016 vote history and differences in attitudes matter for one domain (like healthcare and guns), they likely matter for others (like religious views or attitudes toward the police and BLM).

Ultimately, this step in our research design confirms the robustness of our initial findings. It demonstrates that the eight-scenario framework is versatile enough to capture consistent predictive patterns, even when we switch out the underlying predictors.

6. Conclusion

One major thing I realized through this project is how helpful it is to look at the same people over multiple points in time, rather than just once. When we track changes in attitudes—like how someone's view on gun control or healthcare shifts from slightly supportive to strongly supportive—it does a much better job of explaining whether they'll stay with their usual party or switch in the next election. Past voting history also matters a lot: we saw that knowing how someone voted last time around was one of the biggest clues about how they'd vote later. By taking this long view, we can catch the gradual changes people go through and see how those shifts stack up to influence their eventual choices.

We also saw that big, real-world events—like a sudden crisis or a major political moment—might jolt someone's beliefs or their sense of where they fit politically. While we didn't have enough data here to fully dig into that, it makes sense that a single snapshot couldn't capture such a dramatic change. It was a bit disappointing we didn't have a large enough panel to try more advanced methods like recurrent neural networks, because those might really uncover hidden patterns in how voters move from one stance to another over time. Still, even with simpler models, we discovered that panel data—tracking the same people again and again—holds a lot more power for predicting voter behavior than just looking at one-off surveys.

References

Bartels, L. M. (2002). *Beyond the Running Tally: Partisan Bias in Political Perceptions*. Political Behavior, 24(2), 117–150. http://www.jstor.org/stable/1558352

Lewis-Beck, M. S., Jacoby, W. G., Norpoth, H., Weisberg, H. F., & Converse, P. E. (2008). *The American Voter Revisited*. University of Michigan Press. https://doi.org/10.3998/mpub.92266

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press. http://www.jstor.org/stable/j.ctt5hhcfr

Dinas, Elias. (2014). *Does Choice Bring Loyalty? Electoral Participation and the Development of Party Identification*. American Journal of Political Science. 58. 10.1111/ajps.12044.

Sides, J., Tesler, M. & Vavreck, L. (2018). *Identity Crisis: The 2016 Presidential Campaign and the Battle for the Meaning of America.* Princeton: Princeton University Press. https://doi.org/10.1515/9781400888306