

## **Data Science for Political Science Final Paper**

### **Introduction:**

Voter registration by party represents a promising predictive tool seldom invoked by analysts. With the decline of polling engendered by the contemporary political arena, experts have begun to explore viable and consistent alternatives. Thirty states register voters by party, sometimes requiring partisan identification to participate in primary contests. The following inquiry seeks to quantify the predictive significance of voter registration by party. The implications of partisan voter affiliation will be assessed in a variety of institutional settings, such as states with closed primaries, semi-closed primaries, and automated voter registration systems. Party affiliation and voting behavior share a close association, foretelling a strong parallel between voter registration trends and electoral outcomes at the state and county level.

The investigation involves scraping, classifying, and comparing statistics conveying voter registration and electoral behavior. The dataset encompasses election results and voter registration reports from six states, Arizona, Colorado, Florida, North Carolina, Oregon, and Pennsylvania, from 2004 through 2020. Between each presidential election cycle, a state's registration and voting drift will be classified as either Republican or Democrat. The two categories will then be cross-referenced. The county level figures then receive the same scrutiny. The findings of the inquiry unveiled an appreciable correlation between voter registration and election results at both the state and county level, though the state level trends proved more robust. These conclusions contain crucial knowledge informing campaign and party policy. A confident sense of regional and demographic coalition shifts contributes to a winning strategy.

### **Background:**

An accurate gauge of public sentiment enables the formulation of effective tactics by candidates and consultants. Voter registration indicates the political undercurrents influencing a particular locality. If a policy stance or issue sways the ideological slant of a county or state, the voter registrations should move in tandem. These trends, coupled with a region's economic and demographic makeup, may inform strategists of how to access voting blocs. A party may refine its rhetoric if an important coalition, such as suburban

moderates, begin to display an adverse registration trend. Likewise, candidates can weaponize these findings to maximize voter output. Contenders can allocate precious campaign resources to districts showing a favorable movement. Given its momentous potential, voter registrations electoral relevance warrants further scrutiny.

A large body of literature has expounded on the pertinence of party identification to vote choice. Party attachments reflect personal stances on leading issues. Though demographics, level of education, and issues stances influence vote choice, party affiliation holds acute relevance. Green, Palmquist, and Schickler (2004), developed statistical models demonstrating a 50% difference in vote choice between registered Republicans and Democrats regardless of other characteristics. The overwhelming support for one's own party stems from personal loyalty and interest, similar to rooting for a favorite sports team. Contrary to the vagaries of consecutive elections, party registration changes over prolonged periods. Yet, the direct effect of party registration may still be observed. Over the course of decades, the southern electorate gradually flocked from its democratic legacy towards the Republican party. These shifts accompanied a rise in Republican representation among the southern states (Green, Palmquist, & Schickler, 2004). The polarization defining contemporary political discourse amplifies the effect of party association. In the 2016 and 2020 presidential elections, partisans voted for their own party candidates at rates exceeding 90%. These phenomenon suggest a compelling, positive relationship between advances in registration and support for the party in upcoming races.

The probe will investigate the empirical relationship between voter registration advantage and electoral margins. The voter registration edge refers to the difference in overall proportion of voters enrolled in one party compared to the other. For example, if 30% of the voters in a county classify as Republicans versus 20% as Democrats, the Republicans appreciate a 10% voter registration advantage. An increase in the partisan registration advantage between cycles should forecast an augmented margin of victory. The analysis will test how often these assumptions manifest. The results will then be stratified according to institutional arrangements. The inspection will compare states with closed-primaries, Oregon and Pennsylvania, and semi-closed primaries, Arizona, Florida, North Carolina, and Colorado, in search of divergences. States with

closed primaries encourage voters to select a party. As a result, these states should have voter rolls demonstrating the disposition of the citizenry to a higher degree. States with automated voter registration systems, Colorado and Oregon, will also be isolated and scrutinized for disparate behavior. If a difference emerges, states with automated registries should exhibit a weaker trend due to the disproportionately high share of unaffiliated voters. A multi-faceted approach will facilitate the extrapolation of any findings to unreviewed states in accordance with institutional setting. Such a review should supplement future electoral predictions.

### **Data and Approach:**

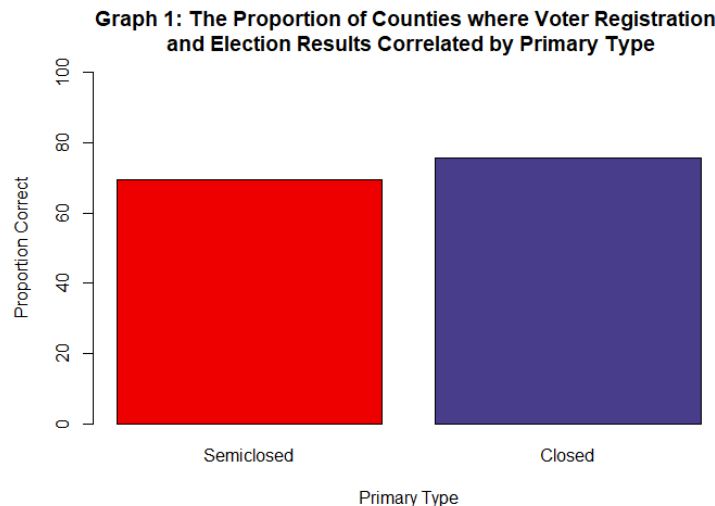
The review examines voter registration and election data from Arizona, Colorado, Florida, North Carolina, Oregon, and Pennsylvania spanning from 2004 to 2020. The voter registration figures arose directly from the secretary of state websites of the surveyed states. The election results stem from Ballotpedia, an online encyclopedia dedicated to publishing extensive election reports. Each state entails two a statewide and county level set of data. The statewide figures feature five rows corresponding to the last five presidential elections. The county level dataset has a row for each county. North Carolina, for example, contains one hundred rows. The following table lists the important variables utilized in the procedure.

<b>Names</b>	<b>Description</b>
<b>floridaregistrationandvoting</b> <b>arizonaregistrationandvoting</b> <b>coloradoregistrationandvoting</b> <b>northcarolinaregistrationandvoting</b> <b>pennsylvaniaregistrationandvoting</b> <b>oregonregistrationandvoting</b>	-a confusion matrix showing the partisan registration and election trend in each state
<b>registrationaccuracystatewide</b>	-a grouping variable that tracks the consistency of voter registration in predicting margin shift at the state level
<b>floridacountylevelaccuracy</b> <b>arizonacountylevelaccuracy</b> <b>coloradocountylevelaccuracy</b> <b>northcarolinacountylevelaccuracy</b> <b>pennsylvaniacountylevelaccuracy</b> <b>oregoncountylevelaccuracy</b>	-reports the accuracy of voter registration as a predictor for margin shift at the county level in each aforementioned state
<b>overallcountylevelaccuracy</b>	-a grouping variable amassing all county level findings to reflect overall accuracy
<b>semiclosedprimarystatewide</b> <b>closedprimarystatewide</b>	-a grouping variable tabulating the state level outcomes for voter registration and election trend matches sorted based on primary type
<b>semiclosedprimarycounty</b> <b>closedprimarycounty</b>	-a grouping variable collecting the county level correspondence between voter registration and election results stratified by primary type

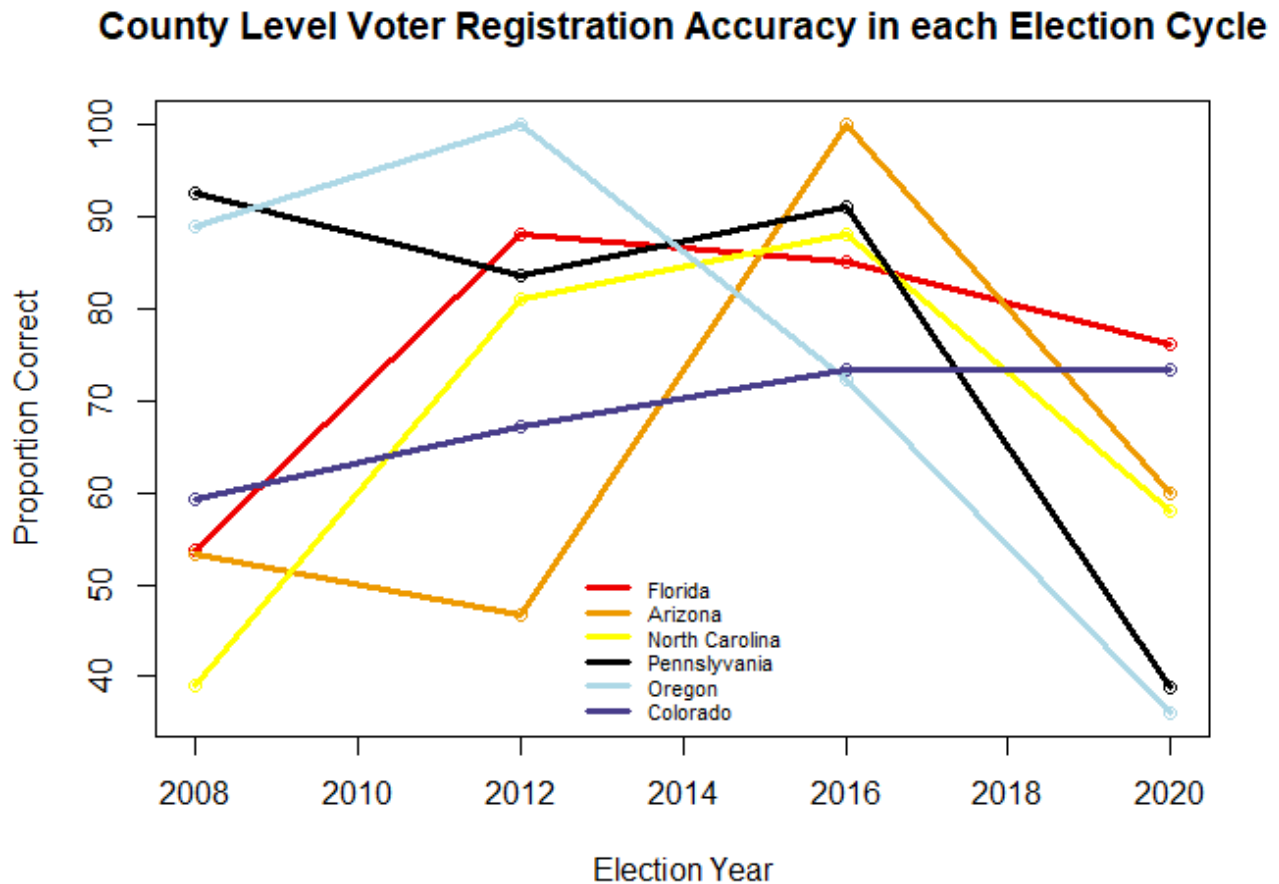
The analysis will tabulate and rank a state or county's voter registration and election trend between each election cycle. A confusion matrix will then cross reference these grades to detect matches and mismatches. After all states have been processed, overall statistics will be amassed and reported. The experiment will also sort the results on the basis of other variables. States with closed primaries will be compiled separately to discern any differences. The implementation of automated voter registration systems in certain states in recent cycles warrants exploration. A marked increase or decrease in voter registration accuracy will be noted. Though somewhat crude, the codification of states and counties will shed light on the viability of voter registration as an electoral indicator. Future studies may pursue into a more nuanced, analytical approach.

### Results:

As explained in previous sections, the probe catalogued each state and county as exhibiting a Republican or Democrat registration trend between each election cycle. The examination then assessed if the registration shift correlated with electoral trajectory. At both the state and county level, voter registration proved a potent indicator. The statewide registration trends anticipated the outcome in 79.17% of iterations, equating to 19 out of 24 cases. County level registration shifts emerged less reliable and demonstrated greater variance between states. Pennsylvania's county level data matched in 76.49% of instances, followed by Florida with 75.75%, Oregon with 74.31%, Colorado with 68.36%, Arizona with 66.67%, and lastly North Carolina with 66.50%. Collectively, the voter registrations predicted the county level election shift in 996 out of 1396 occurrences, corresponding to 71.35% of counties. These conclusions corroborate the prediction for a close association.



The experiment sorted the states based upon primary type. Pennsylvania and Oregon host closed primary elections whereas the other states conduct semi-closed primaries. As depicted in graph 1, a distinct disparity manifests. States with closed primaries overperformed states with semi-closed primaries 75.73% to 69.54% of counties, a difference of 6.19%. These results confirm theoretical expectations. By encouraging voters to select a party when registering, closed primaries yield a voter roll better reflecting the political sentiment of the citizenry.



Graph 2 chronicles the accuracy of voter registration at the county level in each election cycle from 2008 through 2020. Oregon and Colorado debuted automated voter registration systems in 2015 and 2017 respectively. Oregon voter registrations experienced a precipitous decline in veracity in 2016 and 2020. Prior to the implementation of the law, Oregon's voter rolls projected election results in 94.44% of counties. The performance would decline to just 54.17% in the next two election cycles. However, Colorado's voter registrations improved from 66.67% to 73.44% following the enactment of an automated voter registration paradigm. The general decline in accuracy between 2016 and 2020 further dilutes the significance of these

findings. These mixed results fail to affirm or rebuke the hypothesis posed for the effect of automated registration systems.

**Conclusion:**

The descriptive analysis sought to evaluate the potential of voter registration by party as a predictive tool. The performance of voter registration by party would receive further scrutiny in distinct institutional settings, including states with closed versus semi closed primary elections and automated voter registration apparatuses. The review supported the proposed utility of voter registration trends in forecasting electoral shifts. In both statewide and county level appraisals, the voter registration and election margins mostly moved in tandem. As predicted, the voter registration served as a finer bellwether in states with closed primaries than semi closed primaries. The adverse effect of automated voter registration systems received some support; however, the evidence proved insufficient to render a conclusion.

These findings have relevance in a myriad of practical applications. By providing advanced knowledge of impending electoral realignments, voter registration data can help design successful campaign strategies. Candidates can allocate time and resources towards localities where support appears to wain or shows potential for growth. Political parties may also employ such information to perfect rhetoric and messaging. If a particular policy stance induces a negative or positive registration move within a desired constituency, the party can reverse course or double down. Given the growing distrust of polling, an alternative method could prove exceedingly useful. The predictive power of voter registration by party may be further explored through a number of avenues. A subsequent review may encompass more states that collect and report partisan voter registries. Such evaluations could also grade the degree of registration shifts between cycles. A 5% move in registration likely carries greater electoral implications than a 0.5% change. Subsequent inspections should enact these refinements to garner a more nuanced understanding. Nonetheless, voter registration by party represents a meaningful metric capable of foretelling electoral outcomes.

## Works Cited

- “2022 Voter Registration Statistics.” *Colorado Secretary of State*,  
<https://www.sos.state.co.us/pubs/elections/VoterRegNumbers/VoterRegNumbers.html>.
- Ballotpedia*, [https://ballotpedia.org/Main\\_Page](https://ballotpedia.org/Main_Page).
- Green, Donald P., et al. “Chapter 8: How Partisan Attachments Structure Politics .” *Partisan Hearts and Minds: Political Parties and the Social Identities of Voters*, Yale University Press, New Haven, CT, 2004, pp. 204–229.
- “Oregon Secretary of State.” *State of Oregon: Elections - Election Statistics*,  
<https://sos.oregon.gov/elections/pages/electionsstatistics.aspx>.
- “Voter Registration - by Party Affiliation.” *Florida Department of State*,  
<https://dos.myflorida.com/elections/data-statistics/voter-registration-statistics/voter-registration-reports/voter-registration-by-party-affiliation/>.
- “Voter Registration Statistics.” *North Carolina State Board of Elections* , <https://vt.ncsbe.gov/RegStat/>.
- “Voter Registration Statistics.” *Voter Registration Statistics | Arizona Secretary of State*,  
<https://azsos.gov/elections/results-data/voter-registration-statistics>.
- “Voting & Election Statistics.” *Pennsylvania Department of State*,  
<https://www.dos.pa.gov/VotingElections/OtherServicesEvents/VotingElectionStatistics/Pages/VotingElectionStatistics.aspx>.
- GitHub Account: <https://github.com/Malicon2000/Voter-Registration-as-Election-Indicator/upload/main>