U.S. Presidential Election 2024: A Predictive Model

I. Abstract

This study utilized a machine learning-based approach to predict the outcome of the 2024 US Presidential election, focusing on six key battleground swing states: Arizona, Michigan, Nevada, North Carolina, Pennsylvania, and Wisconsin. Using county-level data from previous elections in 2012, 2016, and 2020, the model evaluates the impact of demographic and socioeconomic variables, including race, age, education, income, poverty, and labor force participation, on voting patterns. The research identifies consistent trends, such as younger, diverse, and highly educated populations favoring Democratic candidates, while older, less diverse, and lower-income populations lean Republican. Data collection relied on publicly available sources, including the U.S. Census and voting records. A Random Forest classifier was used for its ability to manage nonlinear relationships and reduce overfitting through ensemble learning. Despite its strengths, the model encountered challenges such as data sparsity, variable granularity, and the unique dynamics of the 2020 election, shaped by the COVID-19 pandemic and its economic impact. Preliminary analysis highlighted correlations between certain demographic features, such as race and education, and partisan preferences, confirming several hypotheses. However, anomalies in 2012 and 2020 raised concerns about the model's reliance on historical data. Limitations include the exclusion of real-time factors like campaign strategies, economic shocks, and issue salience, which often play decisive roles in elections. Future work aims to integrate dynamic predictors such as polling data, economic indicators, and public sentiment analysis to enhance predictive accuracy. This project underscores the complexity of electoral modeling, emphasizing the interplay between structural demographic trends and evolving political landscapes. While the model provides a robust foundation, its predictive power would benefit from incorporating factors that account for both stability and volatility in voter behavior.

II. Introduction

Our group is interested in building a machine learning model that effectively predicts the outcome of the 2024 United States presidential election. Given the central role of battleground states in shaping electoral outcomes, we focused our analysis on seven key jurisdictions: Arizona, Georgia, Michigan, Nevada, North Carolina, Pennsylvania, and Wisconsin. These states are widely recognized for their decisive influence in determining the election result, as evidenced by historical trends and political analysis across diverse media platforms.

This report outlines the extensive process we employed to conceptualize and enact our project plan. Specifically, we designed a supervised machine-learning approach using demographic and economic indicators to predict the dominant political party at the county level within these states. By engaging a random forest model, we aimed to capture and analyze complex relationships in the data that influence voting behaviors. Similarly, voting in the United States has evolved into a complex intersection of demographic trends, economic conditions, and political ideologies. While federal elections dominate the political landscape, the influence of local and state-level dynamics cannot be overstated. U.S. presidential elections are conducted under an Electoral College system, where each state's weight in determining the presidency hinges on its allocation of electoral votes. Within this system, battleground states receive significant attention because of their unpredictability and potential to swing election outcomes.

Predicting voting outcomes is inherently challenging due to the multitude of factors influencing voter behavior. Demographic shifts, economic conditions, social movements, and even external events like global pandemics or economic crises can drastically alter electoral landscapes. The decentralized nature of the U.S.' elections add further complexity, as states vary in voting regulations, access to early and absentee voting, and voter identification requirements. Additionally, unforeseen variables such as last minute campaign strategies, scandals, or shifts in voter turnout can significantly impact results, making predictions even more difficult. Historical voting patterns, while informative may not always reliably predict future outcomes due to these dynamic influences.

Our project was motivated by the understanding that specific demographic and economic factors drive voting behaviors at the county level. Past studies and electoral data suggest that race, education, income levels, and historical voting patterns are among the key variables influencing election outcomes. We sought to use these variables and more in our machine learning model to make accurate county-level predictions. By integrating data from the 2012, 2016, and 2020 elections, we trained and tested a random forest model to predict the winning party for each county in the six states for the 2024 election. The findings demonstrate that our approach yielded high levels of accuracy, aligning closely with the actual 2024 election results. For each state, our analysis not only identified counties likely to vote Democratic or Republican but also provided insights into the key variables driving these outcomes. This report provides a detailed overview of our methodology, results, and the implications of using machine learning for electoral predictions. It also highlights the broader context of voting behavior in the United States, offering readers a comprehensive understanding of the factors at play and the predictive power of our model.

The United States has long been a democracy where voting serves as a cornerstone of civil engagement and representation. However, the process of voting and predicting its outcomes has become increasingly sophisticated which has been influenced by advances in technology, data analytics, and behavioral science. Our project integrates these elements into an electoral prediction framework that aims to connect the gap between historical data and contemporary voter behavior. The six battleground states chosen for our analysis represent a smaller representation of the broader United States. Arizona, a state with rapid demographic changes, has emerged as a key indicator of shifting political alliances. Similarly, Michigan, Nevada, North Carolina, Pennsylvania, and Wisconsin similarly illustrate the dynamic interplay of rural, urban, and suburban voting patterns, economic shifts, and demographic evolution. By focusing on these states, our project captures the essence of the electoral dynamics likely to shape the 2024 presidential race.

To construct our random forest model, we used a robust dataset comprising demographic and economic indicators. These included variables such as race, education levels, income distribution, population age structure, and historical voting patterns. Data preprocessing steps included handling missing values through techniques like imputation and linear interpolation, ensuring the dataset was clean and suitable for analysis. We also employed label encoding to categorize counties by their voting outcomes, streamlining the integration of categorical data into our model. The random forest model was chosen for its ability to manage large datasets with complex, non-linear relationships between variables. This method combines multiple decision trees to improve accuracy, mitigate overfitting, and enhance the interpretability of our results. Feature importance analysis within the model allowed us to identify key predictors of voting behavior. For instance, variables such as the percentage of older males, education levels, and racial identity emerged as significant determinants of county-level outcomes.

Each state presented unique challenges and insights. In Nevada, our model achieved a success rate of 94.11%, accurately predicting the outcomes for 16 out of 17 counties. The sole issue, Clark County, underscored the problems of urban voting behavior, where factors beyond demographic and economic indicators, such as voter arrival efforts, may play a pivotal role. Similarly, in Arizona, the model demonstrated a 92.86% success rate, with only one county (Cochise) misclassified. This high level of accuracy reflects the stability of voting patterns and the robustness of our model in capturing these trends. In North Carolina, the analysis highlighted the predictive power of demographic variables like the percentage of older males and educational attainment. The model's predictions aligned closely with historical trends, accurately forecasting rural counties' Republican leanings and urban areas' Democratic preferences. Despite limitations such as missing or imputed data, the results emphasized the effectiveness of our approach in identifying key patterns within the state. Michigan presented a more complex landscape due to its diverse population and economic disparities. Our model achieved a 96.39% success rate, accurately predicting outcomes for 80 out of 83 counties. Notably, urban centers like Detroit and Ann Arbor leaned Democratic, while rural areas predominantly supported Republican candidates. These findings align with broader national trends and validate the importance of demographic and economic indicators in understanding voting behavior. Pennsylvania, a decisive battleground, demonstrated the importance of capturing dynamic shifts in voter preferences. Our model accurately predicted the outcomes for 61 out of 67 counties, achieving a 91.04% success rate. Misclassified counties, such as Bucks and Lancaster, highlighted the challenges of accounting for real-time confounders like economic activity and voter turnout. Nonetheless, the analysis provided valuable insights into the factors shaping Pennsylvania's electoral landscape. In Wisconsin, the model's performance was equally notable, achieving an 87% success rate in predicting party flips for the 2024 election. By identifying 13 counties likely to shift their party alignment, the analysis underscored the model's ability to detect evolving voter patterns. These findings emphasize the potential of machine learning to enhance our understanding of electoral dynamics and inform future research in this domain.

The implications of our project extend beyond the immediate predictions for the 2024 election. By demonstrating the success of random forest models in electoral analysis, we highlight the value of integrating machine learning into political science research. This approach not only enhances the accuracy of predictions but also provides a framework for exploring the underlying drivers of voter behavior. Furthermore, our findings contribute to a broader discourse on the role of data analytics in shaping electoral strategies, policymaking, and public engagement. As the United States continues to navigate a rapidly changing political and social landscape, the ability to predict and understand voting behavior becomes increasingly vital. Our project represents a step toward harnessing the power of technology and data to inform electoral analysis, offering a blueprint for future research and applications. By bridging the gap between historical data and current voter behavior, we aim to contribute to a more detailed understanding of the variables shaping American democracy.

In conclusion, this report provides a comprehensive overview of our methodology, findings, and their broader implications. By focusing on six battleground states and employing a random forest model, we demonstrate the potential of machine learning to predict electoral outcomes with high accuracy. The insights gained from this analysis not only shed light on the dynamic factors of the 2024 presidential election but also pave the way for future innovations in electoral research. Through this work, we hope to inspire further exploration of the intersection between technology, data, and democracy, fostering a deeper understanding of the factors that define our collective political future.

III. DATA

Our group began the data collection process by individually assessing data availability online. We understood that many political, social, and economic measures were likely available through non-profits, the US Census, and other organizations. Confident in the availability of data, we attempted to survey the units of analysis to consider, the years of data available, and the uniformity of data sources across all seven states. These factors were influential in how we would craft our research design.

Following individual research, our group met for a brainstorming session. Discussing our individual findings and thoughts on how to approach the task at hand, we began to define our research project in line with the limitations that we identified. This led us to make concrete observations in three facets of the data collection process:

1) Unit of Analysis: Beginning with a concept of constructing an election-predicting model but with no concrete basis, our group asked one preliminary question: what would “one

observation” of our data entail? We explored multiple options from the individual to the state. Following discussion, we agreed on a county-level unit of analysis. We found that most existing data was collected at the county level. In addition, we felt that a county-level analysis would allow us to build the most representative predictive model. We recognized that counties within a state embody different voting practices based on factors that matter uniquely to them. With the Electoral College being the basis of the U.S. election, it was imperative to get a granular understanding of states’ voting trends in order to effectively predict the 2024 presidential outcome.

2) Time: The next consideration our group made was how time would affect our analysis. What years could we collect data from? What election years would we train our model on to predict 2024? Would we pool the data to control for time or try to account for time in building a time series model? Our group found most predictor data to be available from the early 2000’s to present. Moreover, our group found outcome data from as early as the mid-20th century. After discussion, we decided to limit the scope of our data from 2012 to 2024. This would allow us to use the outcomes of three past presidential elections (2012, 2016, and 2020) to predict the outcome of the 2024 election. As outlined below in the Challenges section of this report, though, our group is still working on the specifics for many of these questions.
3) Research Methods: Our group understood that our model would largely be built around secondary data collection methods. We aimed to use a mix of macro-level statistics (i.e., demographics, economic indicators) and poll surveys as potential predictors of the model. After further research, though, we realized that the use of poll surveys as indicators of public sentiment was impractical. Polls largely presided at the state-level, with limited data for our intended unit of analysis – the county. Moreover, polls were lacking in uniformity over states and time; there was not consistent polling done over all states nor years of interest. This led us to shift our focus to macro-level statistics as the basis of feature selection.

Feature Selection

Following this acknowledgment of limitations, our group began feature selection. Together, we brainstormed different factors we thought were influential in predicting the outcome of a presidential election. This process was designed to be more idealistic in nature; we asked, assuming the data were available, would we want to collect it? Figure 1 tabulates all the features we thought were ideal in collecting.

Variables to consider		
Race	Education level	Poverty
Gender	Foreign-born civilians	WIC/SNAP benefit usage
Economic status	Businesses present	Law enforcement spending
Age	Crime	Job market

Figure 1. A list of variables brainstormed from our group discussion.

Divvying up the variables among the seven group members, we set out to collect data for as many of these variables as possible. Each group member sought to collect data for all states in an effort to keep the data sources uniform across states.

Figure 2 shows the variables that our group was successful in collecting, with the variable type, a description, and an example of what the variable entails.

Variable	Variable type	Description	Example
Party	Categorical	County outcome for election	Democratic Party
Age x Sex	Categorical	Six age brackets by sex – proportion of total county population	Males 35-44 Percentage
Race	Categorical	Proportion of six races by total county population	Black or African American Percentage
Education	Categorical	Proportion of highest-level of schooling by total county population	Regular high school diploma Percentage
Economic status	Categorical	Proportion of civilians falling within six income brackets	$100,000-$149,999 Percentage
Labor force	Numeric	Proportion of those in labor force by total county population	In labor force Percentage
Poverty	Numeric	Proportion of those in poverty by total population	Percent in poverty

Figure 2. List of attained predictors following data collection.

Our group collected predictor data consisting of four categorical variables and two numeric variables. The data span the three election years of interest. The outcome variable is categorical, reflecting the party outcome for each county within each election year. The data sources were the U.S. Census Bureau, IPUMS, and Professor Terry Johnson’s past voting data available within the course GitHub. After a follow-up meeting discussing our data collection findings and experience, we concluded on the use of these seven variables.

Hypotheses

Finalizing the preliminary data collection process, our group reflected on how these data would be influential in predicting our intended outcome: the 2024 presidential election. Prior to beginning the data wrangling process, our group proposed twelve hypotheses indicating how we anticipate the predictors to influence the outcome.

Hypothesis #	**Description**
1	The younger a population, the more likely a population is to vote Democrat.
2	The older a population, the more likely a population is to vote Republican.
3	The larger the proportion of females in a population, the more likely the population is to vote Democrat.
4	The larger the proportion of males in a population, the more likely the population is to vote Republican.
5	The larger the proportion of Whites in a population, the more likely the population is to vote Republican.
6	The larger the proportion of non-Whites in a population, the more likely the population is to vote Democrat.
7	The larger the proportion of college graduates in a population, the more likely the population is to vote Democrat.
8	The larger the composition of lower income civilians in a population, the more likely the population is to vote Democrat.
9	The larger the composition of higher income civilians in a population, the more likely the population is to vote Republican.
10	The larger the labor force proportion in a population, the more likely the population is to vote Republican.
11	The smaller the labor force proportion in a population, the more likely the population is to vote Democrat.
12	The larger the proportion of a population is in poverty, the more likely the population is to vote Democrat.
13	The smaller the proportion of a population is in poverty, the more likely the population is to vote Democrat.

The six predictor variables of interest are fundamental for any population. Unlike industry-specific or age-specific metrics (e.g., presence of manufacturing companies, social media usage), we feel that these six variables are instrumental in predicting the outcome of a presidential election. We are confident that each of these variables will provide constructive information for the model.

Data Wrangling

All demographic variables excluding party and poverty were collected within one file, separated by year. There were a total of three demographic data files (by year), one party data file, and seven poverty data files (by state). The goal was to construct three data frames for each state by year.

To clean the demographic data, the three files were filtered by state in Python. The result was three data frames for each state, one for each year of interest (e.g., Nevada 2012, Nevada 2016, and Nevada 2020). To clean the party data, the single file was filtered by state. The data frame was then further filtered by year, resulting in three data frames for each state, one for each year of interest. Lastly, to clean the poverty data, the seven state files were filtered by year. This also resulted in three data frames for each state, one for each year of interest.

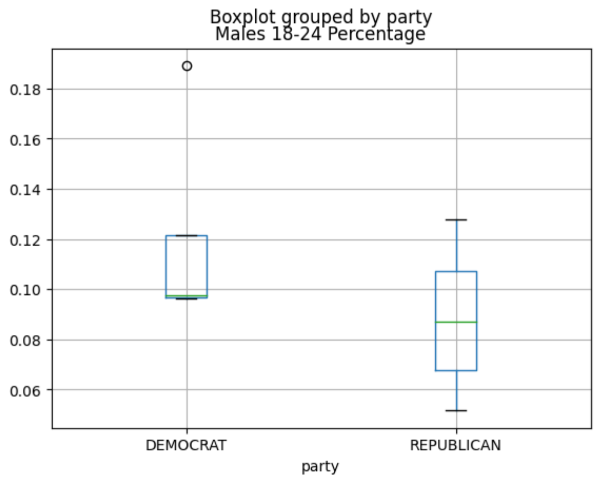
Further data wrangling was necessary in the demographic data frames. The raw data collected consisted of magnitudes of the variables of interest, including race, education, and economic status. To control for population differences, we calculated proportions of these measures by dividing the magnitudes by total population of the respective county. This resulted in new tabulations (columns) with data that we would later use for exploratory data analysis.

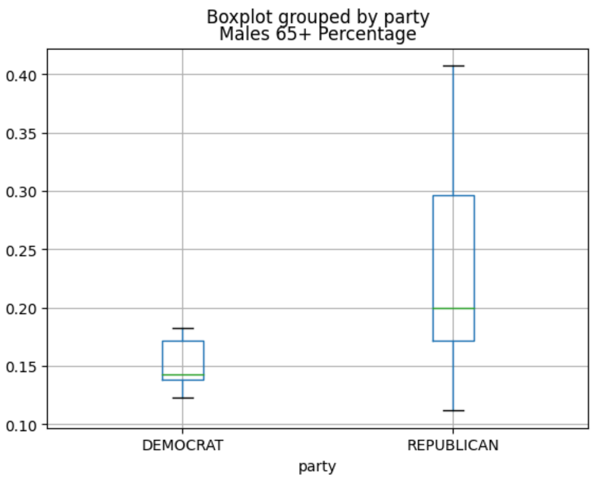
After merging the individual variables' data frames together, limited data wrangling was necessary. We had to drop state/national averages that came with many of the data sets to maintain county-level observations. There was very limited missing data. There was no need for removing a significant number of observations nor imputing values.

I. Exploratory Data Analysis

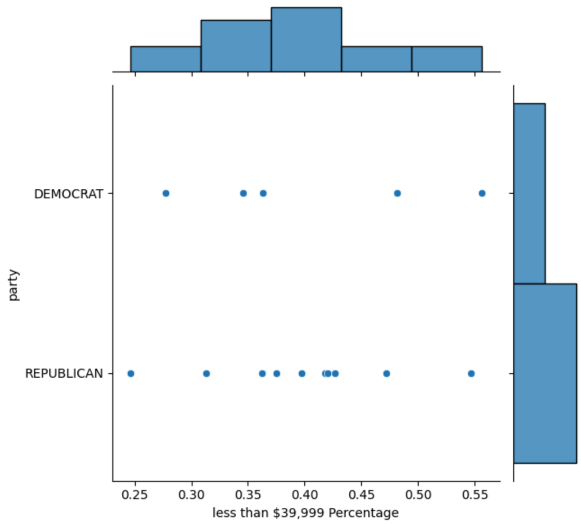
Arizona

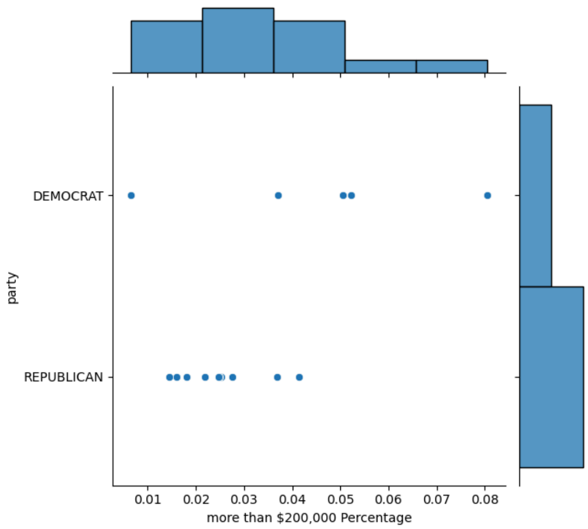
Arizona, known for its riveting sunsets and warm hospitality, is located in the Southwest. With a population of 7.3 million people, the state comprises fifteen counties, with a mix of historically Democratic and Republican-leaning areas. We emphasize that the EDA may not be particularly effective in representing the Democratic party as four counties are used to form a plot of the various quartiles.



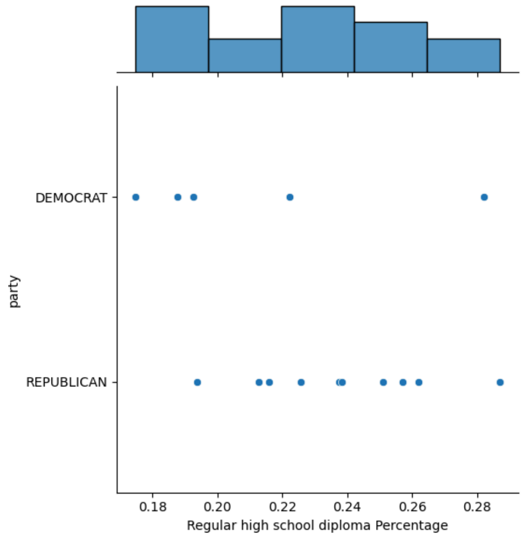


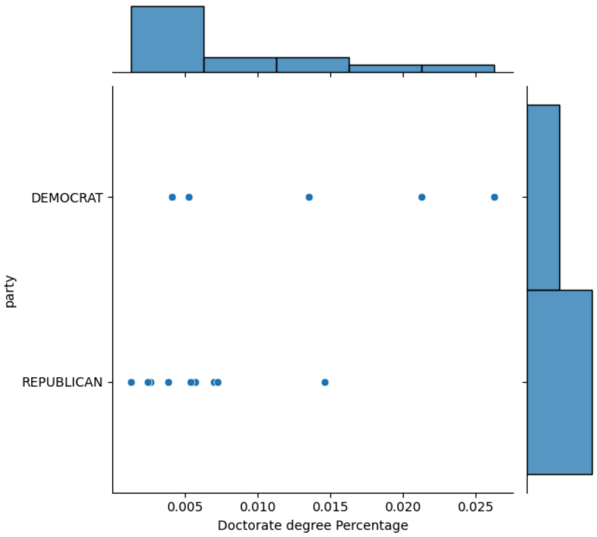
Above are two boxplots of age by party outcome in 2020. The plot on the left shows how Democratic voting counties had a greater proportion of males 18-24 than Republican counties. Conversely, the plot on the right shows how Republican counties in 2020 had higher proportions of males aged 65+ civilians than Democratic counties. This relationship coincides with Hypotheses 2 in the Hypotheses section.





Above are two joint plots depicting the relationship between income bracket and the party outcome in 2020. The plot on the right shows a relationship between counties that have a larger proportion of wealthier civilians and the party, counties with a greater proportion of wealthier civilians tend to vote democrat. However, the left plot does not demonstrate a significant relationship between counties with a greater proportion of lower income civilians and the party outcome. Counties with a higher proportion of civilians earning over $200,000 annually vote both Republican and Democrat. Thus, our preliminary analysis does not provide conclusive results for Hypothesis 8, as counties in Arizona with a greater proportion of wealthier civilians instead tended to vote Democrat. This is in accordance with Hypothesis 9.

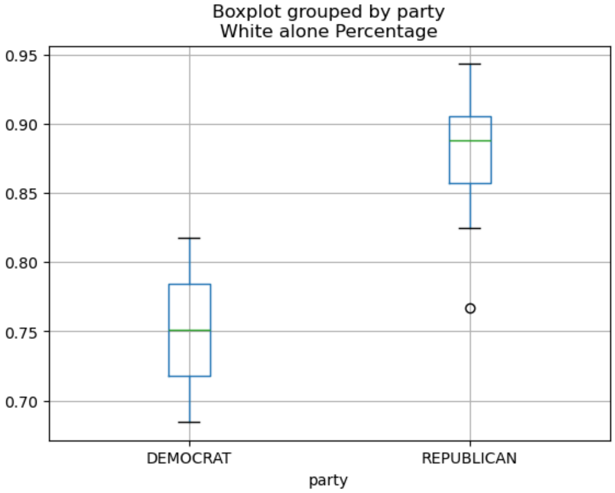


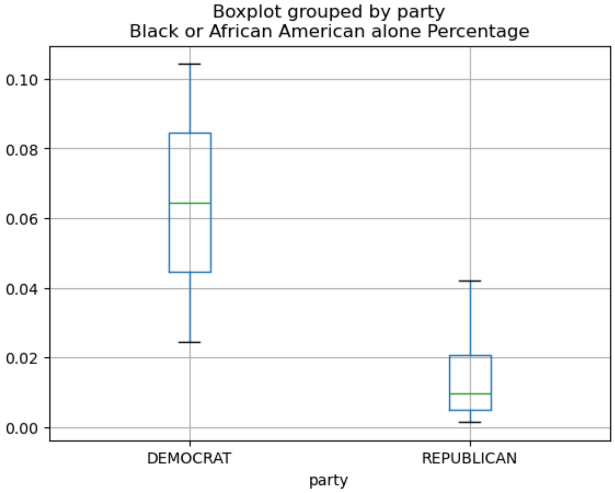


Above are two joint plots depicting the relationship between the highest level of education attained and the party outcome in 2020. With the plot on the left, we can infer that counties with a greater proportion of civilians that have a high school diploma as their highest educational attainment are more likely to vote Republican. Conversely, the joint plot on the right shows a loose but still noteworthy relationship between a higher percentage of civilians holding doctorate degrees and voting Democrat, supporting Hypothesis 7.

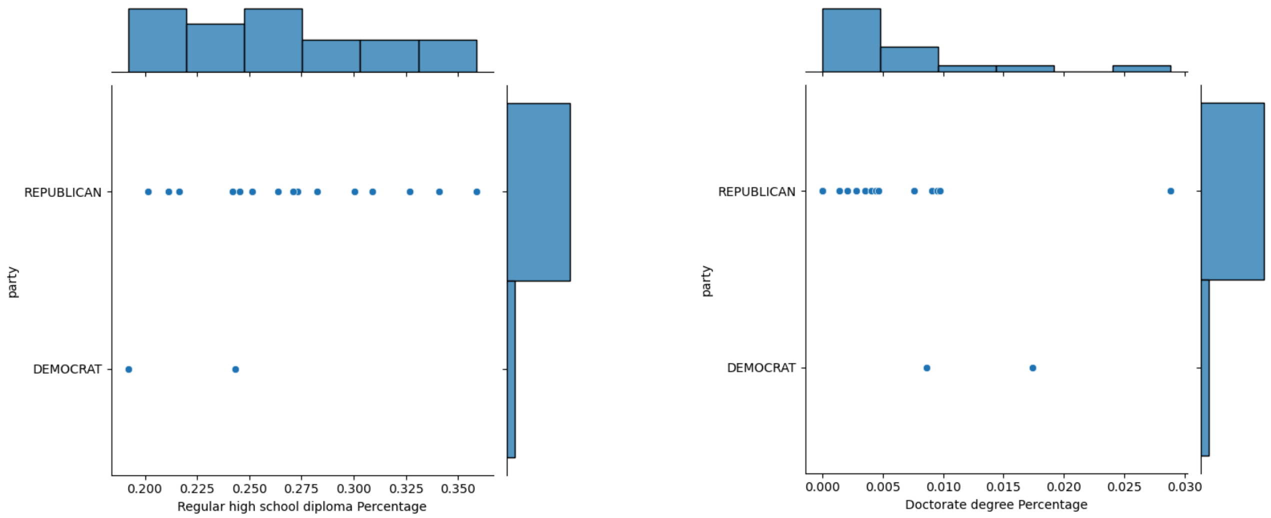
Nevada

Known for its nightlife and silver, Nevada is the seventh largest state (by area) located in Western America, bordering California and Arizona. With a population of 3.2 million people, the state comprises seventeen counties with two historically Democratic counties and fifteen historically Republican states. For data analysis purposes, it is important to note that the boxplots are not particularly informative for the Democratic party, as two counties are used to form a plot of the various quartiles.

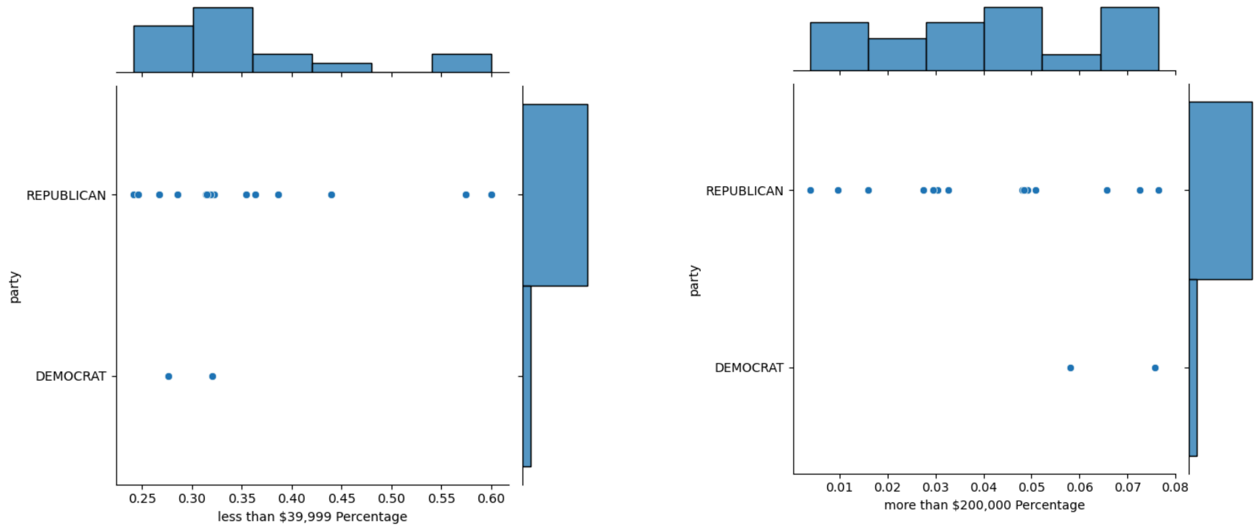




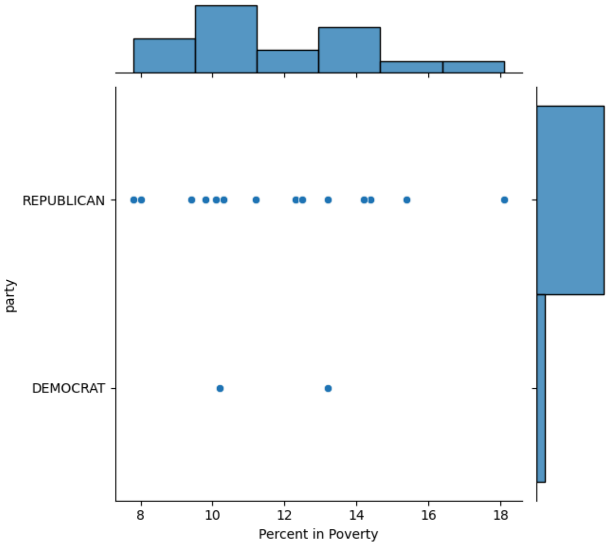
Above are two boxplots of race by party outcome in 2016. The plot on the left shows how counties with higher proportions of White civilians tended to vote Republican. Conversely, the plot on the right shows how counties with higher proportions of Black civilians tended to vote Democrat. This relationship coincides with Hypotheses 5 and 6 in the Hypotheses section.



Above are two joint plots depicting the relationship between highest level of education attainment and the party outcome in 2020. As in the case of race, the Democratic party has two counties. With the plot on the left, it can be loosely inferred that the greater the proportion of civilians that have a high school diploma as their highest educational attainment, the more likely the county is to vote Republican. Conversely, the joint plot on the right shows a loose relationship between a higher percentage of civilians having a doctorate degree with voting Democrat. This would align with Hypothesis 7.



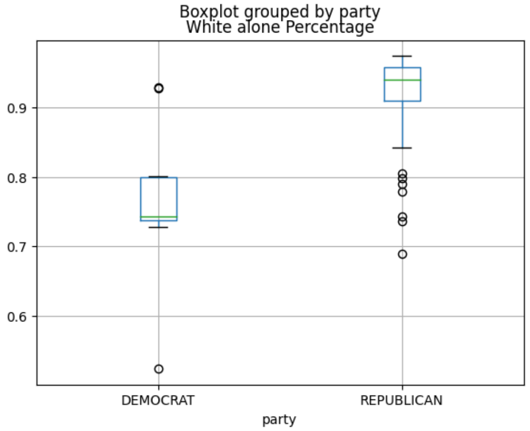
Above are two joint plots depicting the relationship between income bracket and the party outcome in 2020. The plot on the left shows a relationship between counties that tend to comprise more lower income civilians and Republican party votership. This is not in line with Hypothesis 8. Conversely, the plot on the right lacks demonstration of a significant relationship. Counties of higher proportions of civilians that earn over $200,000 vote both Republican and Democrat. The relationship is not clear.

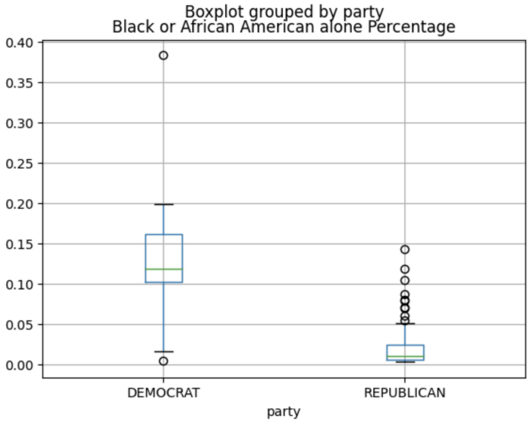


Lastly, the joint plot on the left shows the relationship between the proportion of civilians in poverty in a county and the party outcome in 2020. The relationship is unclear. As the Nevadan exploratory data analysis continues, it is imperative to account for a proper sample size. The lack of sufficient Democratic representation may result in unrepresentative results for the 2024 presidential election prediction.

Michigan

Michigan, located in the Midwest, has a population of nearly 10 million people and comprises 83 counties. Historically, Michigan has been a key swing state in presidential elections, owing largely to its industrial past and economic transition. The state has a diverse demographic composition, including a significant proportion of White, African American, and Hispanic populations, which often influences voter behavior.

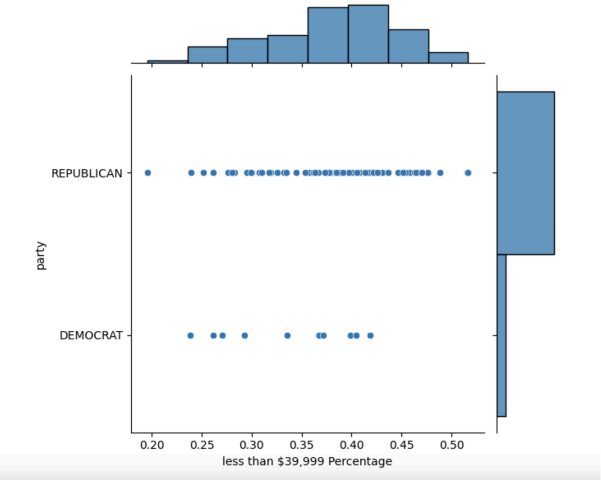


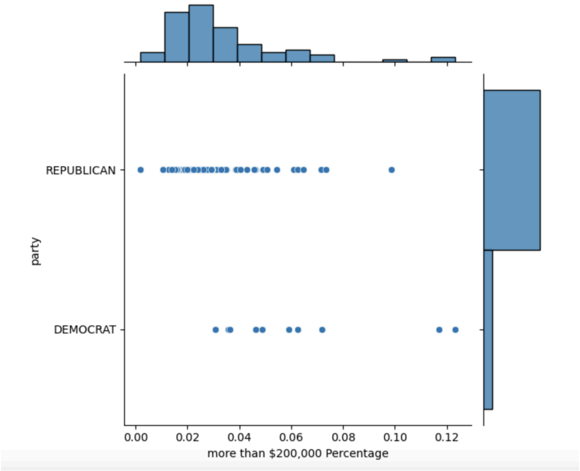


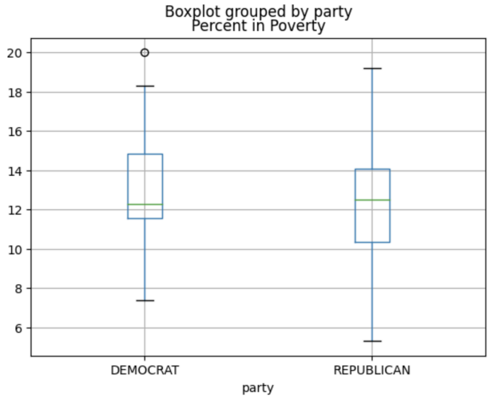
The above boxplot (*left*) shows the relationship between the percentage of the population that identifies as "White alone" and the party outcome in 2016. Counties with higher proportions of White residents tend to vote Republican. This observation aligns with the trend often seen in Midwest states, where rural, predominantly White counties tend to vote Republican, while more diverse, urban counties tend to vote Democrat.

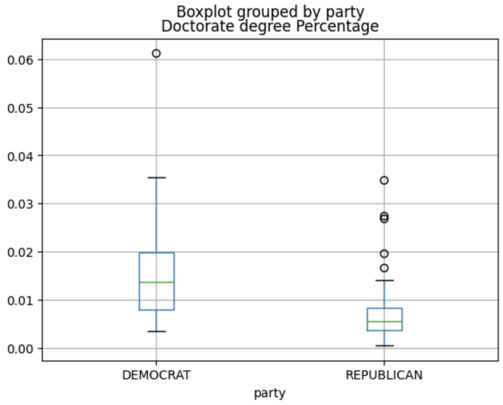
The above boxplot (*right*) for the "Black or African American alone Percentage" in Michigan (2016) illustrates a racial divide in voting behavior. Counties with a significantly higher proportion of Black residents tend to vote Democrat. In contrast, counties with much smaller Black populations tend to swing right.This pattern aligns with broader national trends where African American voters tend to support the Democratic Party, suggesting racial composition, specifically the percentage of Black residents, is a strong predictor of voting outcomes in Michigan.

The below plot (*left)* shows that in Michigan, there is no significant difference in voting patterns between counties with a higher percentage of individuals earning less than $39,999. The below plot (*right)* illustrates no difference in party outcomes with counties with a higher percentage of individuals earning over $200,000. Further statistical analysis and research is required to investigate this relationship.





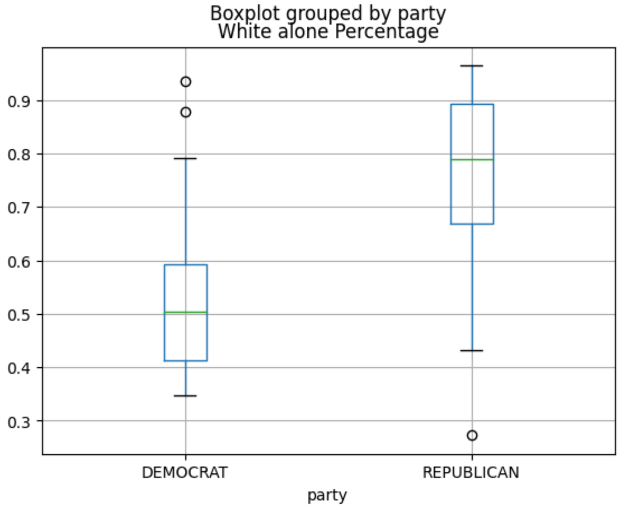


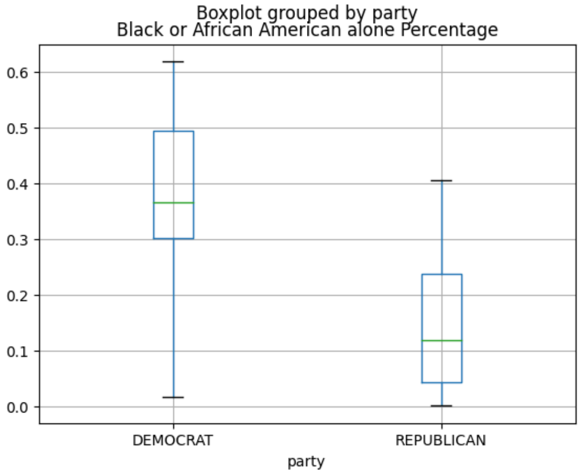


The above plot (*left)* illustrates that in Michigan, the median percentage of individuals in poverty is slightly higher for Republicans, although the difference compared to Democrats appears to be negligible. The above plot (*right*) indicates that in Michigan, counties with a higher percentage of individuals with doctoral degrees tend to vote for Democratic over Republican candidates, although the results are somewhat varied.

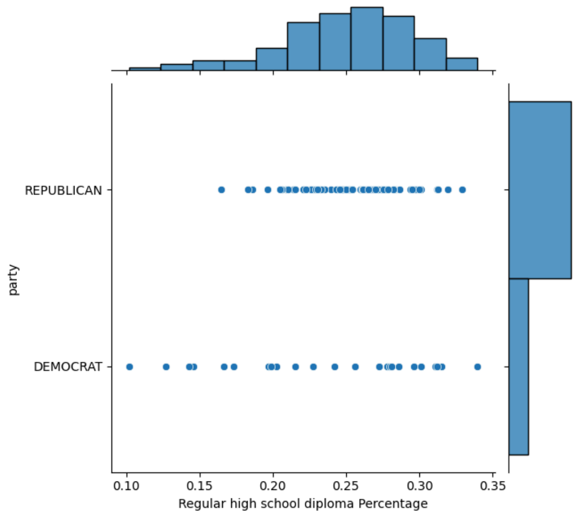
North Carolina

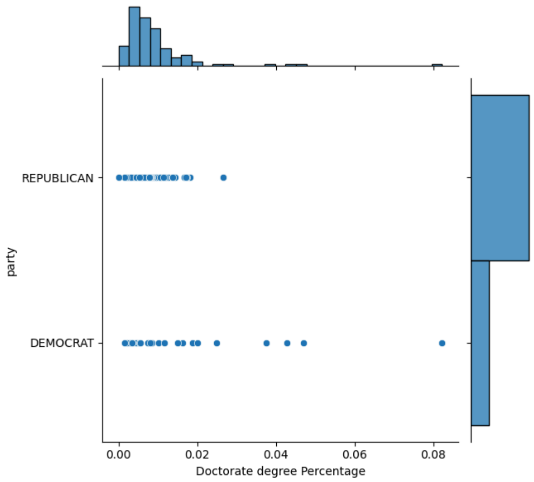
North Carolina is a southeastern state along the Atlantic Coast with a population of 10.4 million people and consists of 100 counties. Historically, North Carolina has been a battleground state in elections, fluctuating between Democratic and Republican.



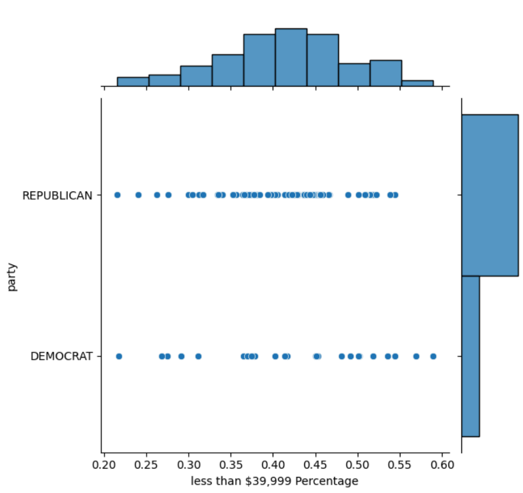


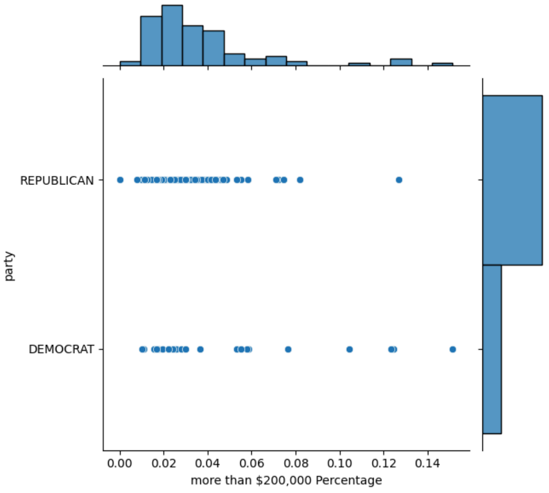
Pictured above are two boxplot graphs for the 2020 presidential election, displaying the relationship between race and party. The plot on the left shows that Democratic-voting counties tend to have a lower median white-alone percentage, whereas Republican counties have a higher concentration of white-alone populations. Similarly, the second box plot depicts that Democratic counties show a higher median of Black residents, while Republican counties exhibit much lower proportions of Black residents. This aligns with hypotheses 5 and 6.



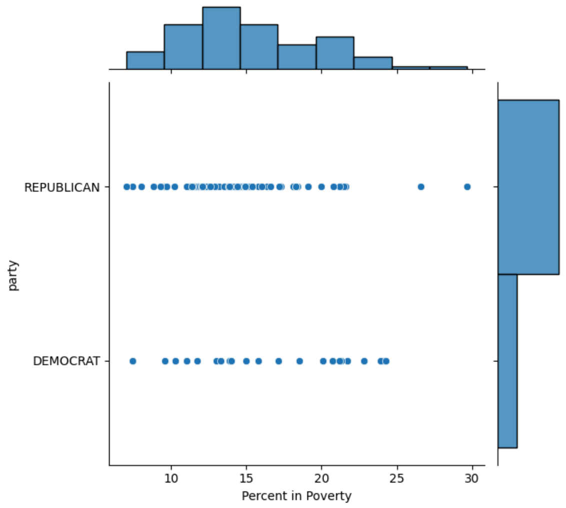


The two joint-plots above showcase the 2020 presidential election results, primarily focusing on the relationship between educational attainment level and party voting patterns. The plot on the left showcases how a higher percentage of Republicans holding only a high school diploma tend to vote Republican. The plot on the right demonstrates that counties with a greater percentage of people holding doctorate degrees tend to vote Democrat. This suggests higher education levels are associated with Democratic voting patterns, aligning with Hypothesis 7.





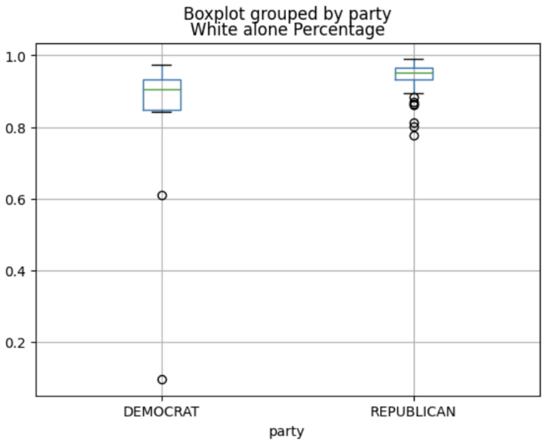
The two joint-plots display the 2020 presidential election results with a focus on the relationship between income brackets and voting patterns. The graph on the left showcases that both Democratic and Republican counties exhibit a wide range of lower-income populations, and there is no clustering or trend that would suggest a clear difference in income distribution between both parties. Similarly, the graph on the right does not have a clear concentration. As a result, these plots neither prove nor disprove Hypothesis 8.

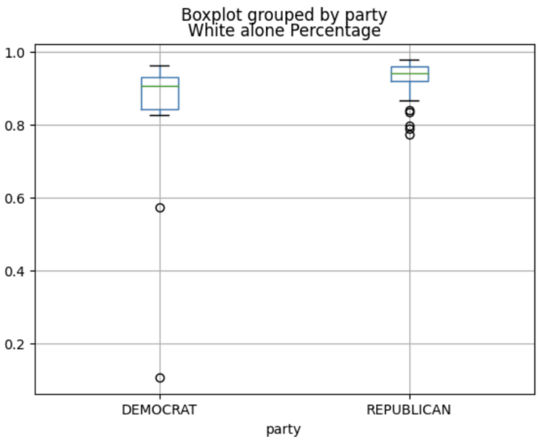


The joint plot above displays the 2020 presidential election results with a focus on the relationship between poverty and party voting patterns. The graph indicates that counties with a higher percentage of poverty tend to lean Democratic, whereas counties with a lower percentage of poverty tend to lean Republican. However, the correlation is not particularly strong. Overall, this graph supports hypothesis 11 but the distribution showcases that the data does not perfectly align with the hypothesis.

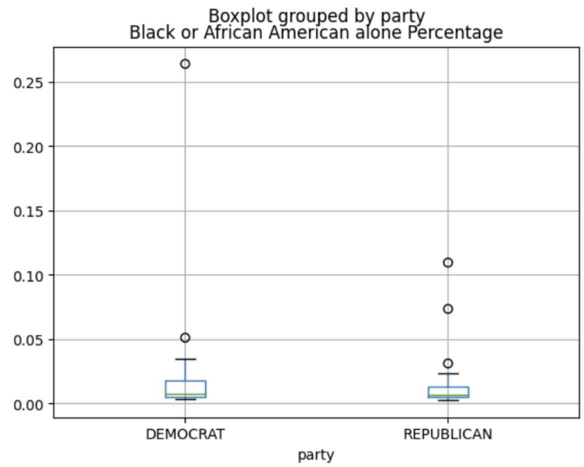
Wisconsin

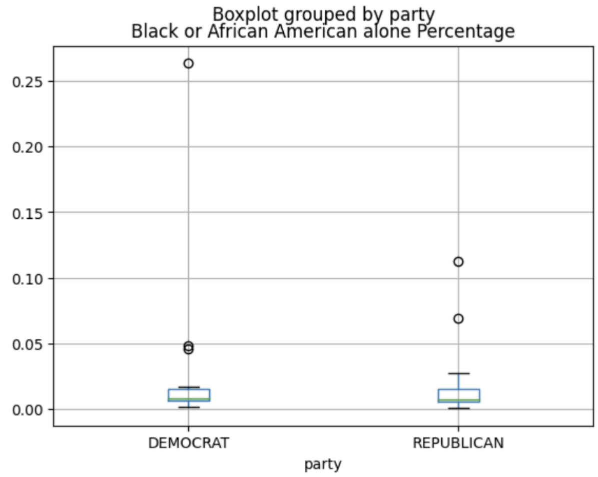
In Wisconsin, voting patterns among racial groups show a clear divide, particularly between White and Black voters. The "White alone" population in the state tends to lean Democratic, although the margin is not overwhelming, with many White voters supporting the Republican Party as well. Factors such as education level, urban versus rural residency, and income levels often play a significant role in determining political preferences among White voters.



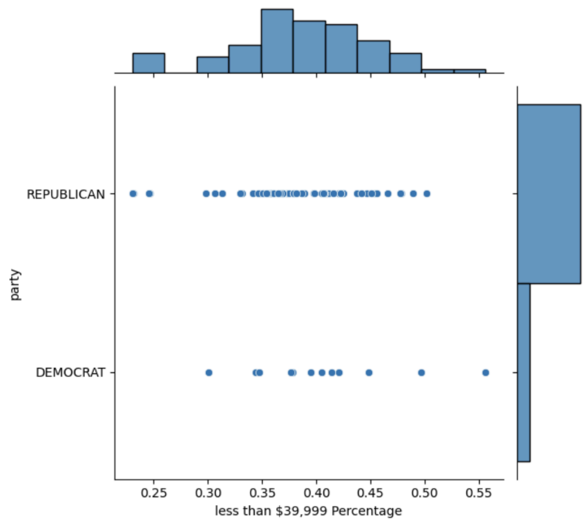


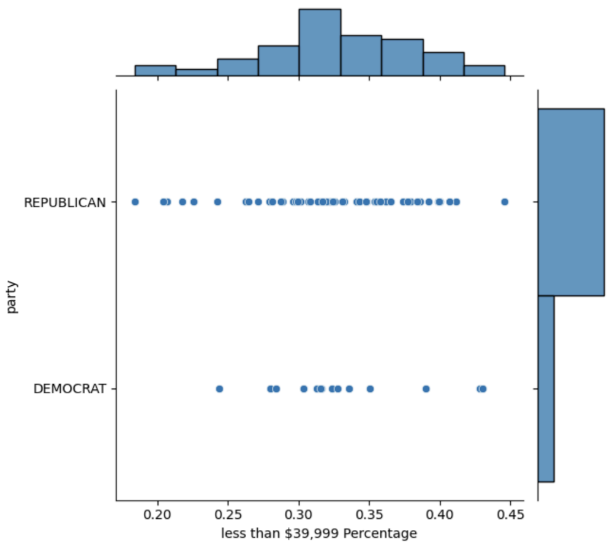
In contrast, the Black population in Wisconsin votes overwhelmingly Democratic both in 2016 and 2020. Historically, the Democratic Party has been seen as more aligned with issues of civil rights, social justice, and policies that address economic disparities, which may explain the strong Democratic loyalty among Black voters in the state. Despite making up a smaller percentage of the overall population, Black voters often play a crucial role in urban areas such as Milwaukee, where their support for Democratic candidates can be pivotal in close elections.





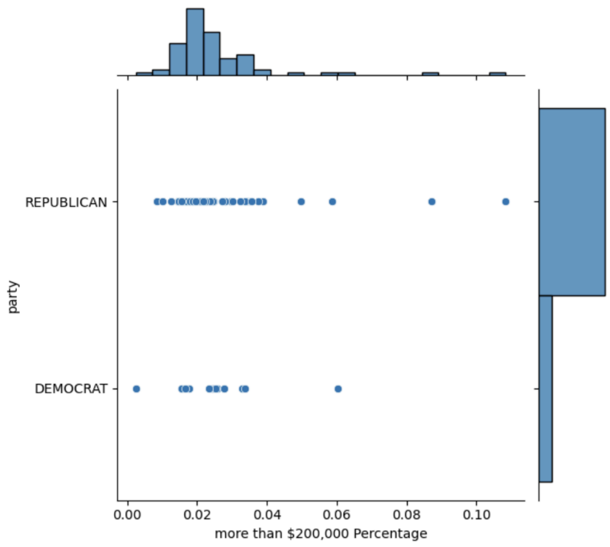
The voter data from Wisconsin in 2016 and 2020 offers valuable insights into the relationship between income levels and party affiliation, shedding light on the state's political landscape. The graphs focus on two notable income brackets: those earning less than $39,999 and those earning more than $200,000, illustrating the voter distribution for both Republican and Democratic supporters. One of the most striking features of this data is the clear income divide between the two parties. Republicans demonstrate a strong presence across both income brackets, with a particularly notable concentration in the lower-income category. In fact, Republican support in the less-than-$39,999 bracket hovers around 40-45% in both elections, suggesting a consistent and significant base among lower-income voters. This trend may reflect the resonance of Republican messaging on issues such as job creation, traditional values, or anti-establishment sentiments with this economic demographic.

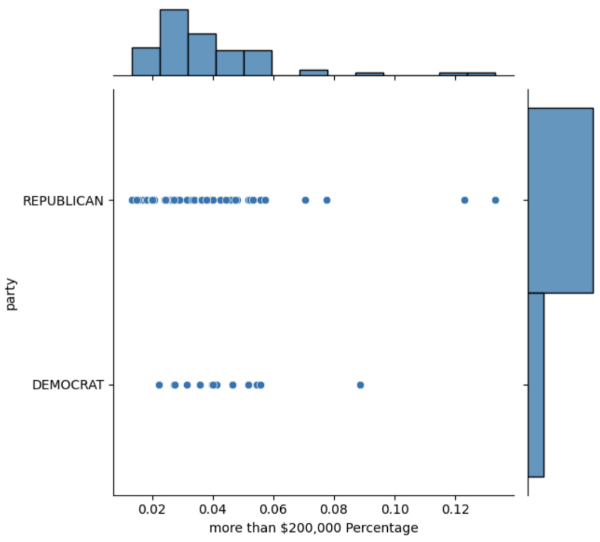




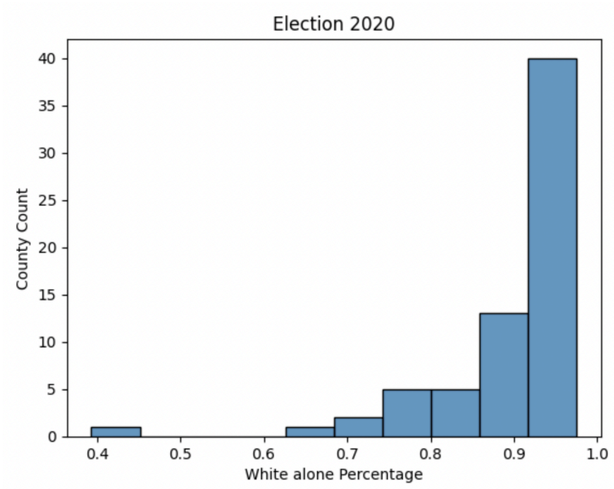
In contrast, the Democratic Party appears to have fewer voters in both income brackets, particularly in the lower-income group. This raises important questions about the party's ability to connect with and mobilize lower-income voters in Wisconsin—a demographic traditionally associated with left-leaning policies. The scarcity of Democratic voters in this bracket may be due to several factors, including challenges in voter turnout, messaging, or shifting party appeal. The persistence of this pattern across both the 2016 and 2020 elections suggests it reflects a structural, long-term issue in Wisconsin's political dynamics rather than a temporary fluctuation.

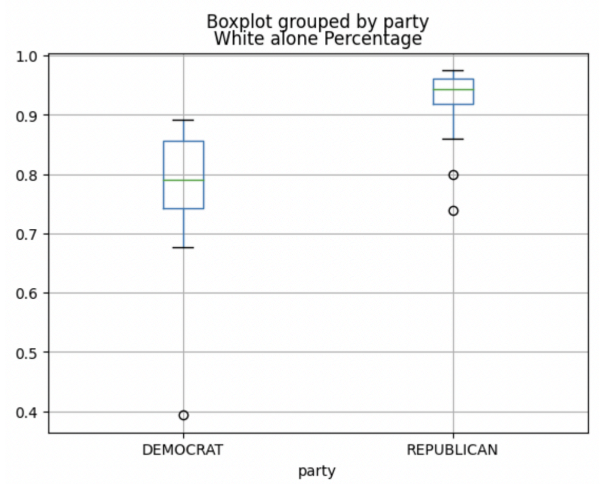
Looking at the higher-income bracket (above $200,000), a different but equally revealing picture emerges. While both parties show fewer voters in this category—a natural outcome given the smaller population size—Republicans again demonstrate a stronger presence. This suggests a fascinating economic polarization within the Republican Party in Wisconsin, as they manage to appeal to both ends of the income spectrum. By addressing diverse concerns such as tax policies that benefit higher-income individuals alongside economic opportunities for lower-income groups, Republicans have crafted a message that resonates with a broad range of voters. The consistency of these patterns between 2016 and 2020 is particularly significant. Despite major national events—such as the Trump presidency, economic shifts, and the COVID-19 pandemic—voting behavior in Wisconsin remained largely stable. This stability points to deeply entrenched political affiliations that appear resistant to short-term events or policy changes. These trends hold considerable implications for future policy-making and political strategies in Wisconsin. For Republicans, there may be pressure to maintain a delicate balance between catering to the differing needs of both lower-income and higher-income supporters. Democrats, on the other hand, face the critical challenge of broadening their appeal among lower-income voters without alienating their current base.

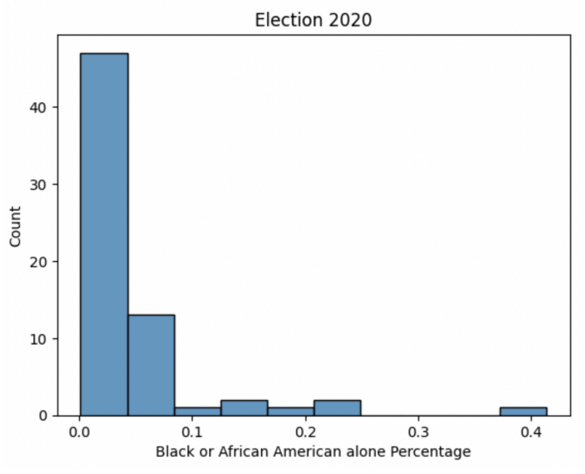


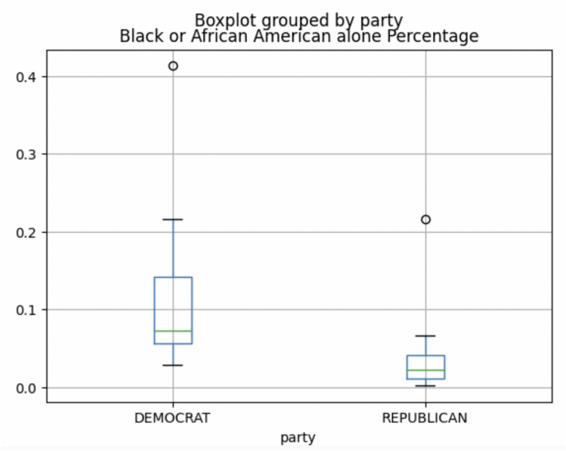


Pennsylvania

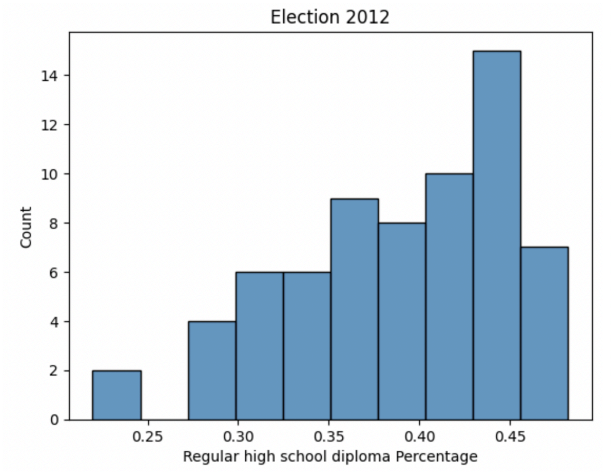


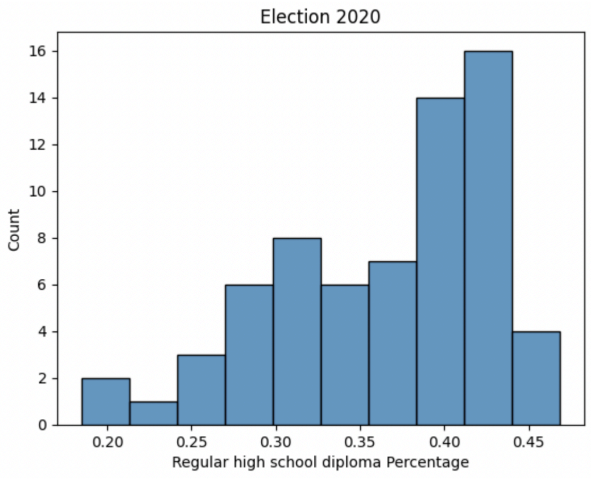


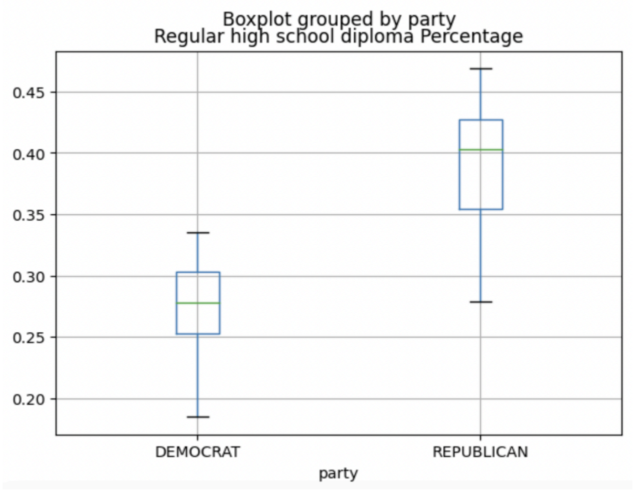


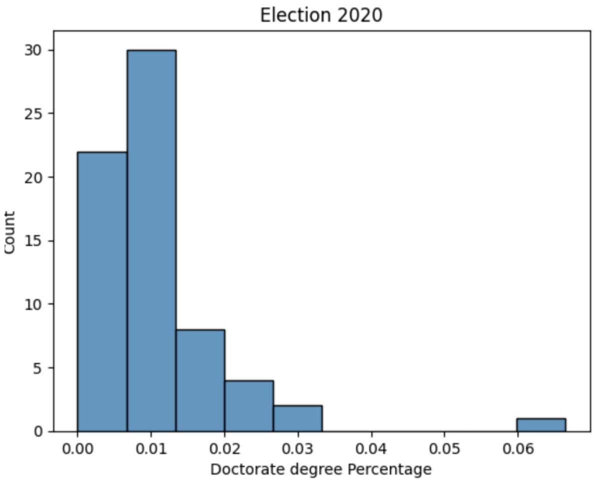


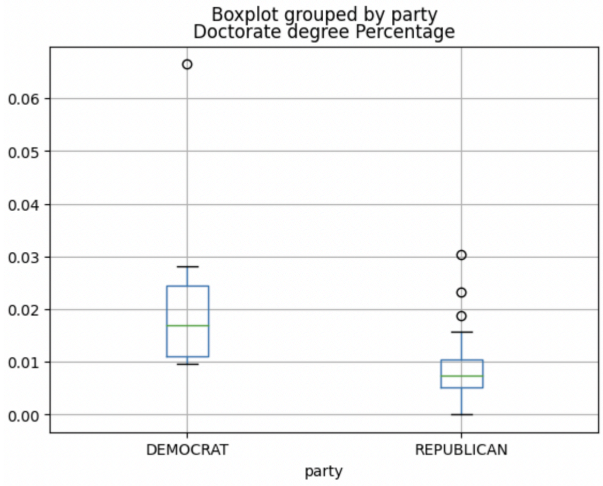
In Pennsylvania, voting patterns among white voters have remained consistent across the last three presidential elections, with a clear tendency to lean Republican in all counties. In contrast, Black voters show a stronger Democratic preference. However, there is also a significant disparity in voter participation, with white voters making up a larger share of voting individuals in Pennsylvania, reflecting their higher percentage of the state's population. While there are more white voters on average, this aligns with a greater white population making up the demographic composition of Pennsylvania. The lower voter turnout among Black voters adds to the disparity in political influence between racial groups. These disparities offer a broader context for interpreting the data.



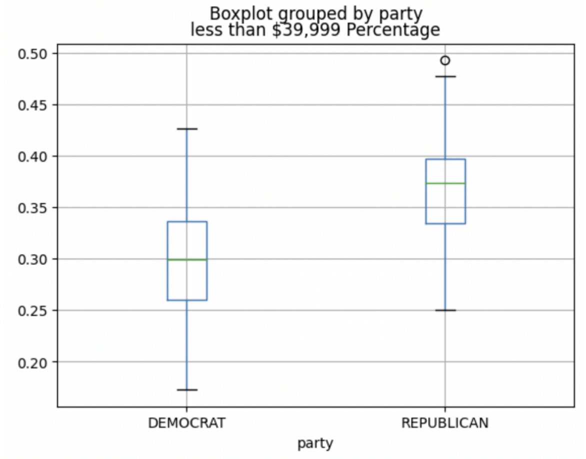


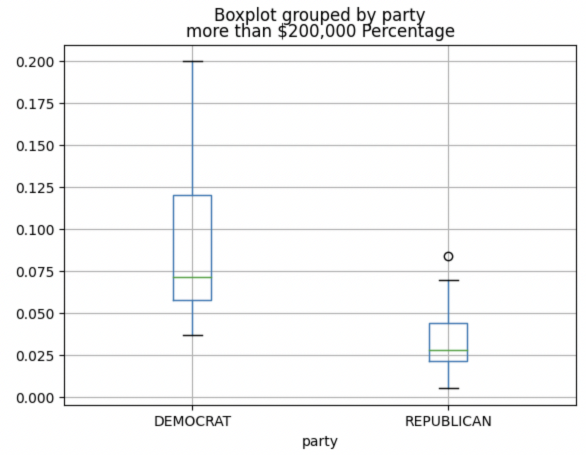






In terms of voting patterns based on levels of education, there seems to be a declining trend among those who have a regular high school diploma voting in presidential elections. Greater voter turnout was present in the 2012 presidential election, but has since decreased with each passing election. Additionally, those who hold degrees beyond a high school diploma, specifically doctorate, lean more Democratic on average. Meanwhile, those who hold only a high school diploma heavily lean Republican. This contrast highlights a growing educational divide in voting behavior, with more educated voters aligning with progressive policies and platforms, while voters who haven't pursued higher forms of education trend more conservative.





These trends also align with voters' annual income. Those with greater incomes, who earn more than $200,000 a year, vote more Democratic, while those who earn less than $40,000 vote more Republican. This reflects a broader narrative where the working class, particularly those with lower levels of education and annual income, often supports Republican candidates despite this seeming to conflict with their economic self-interest.

Challenges

As stated earlier in our data collection section, we encountered a problem in the data due to each county recording counts of individuals for each of the different variables, rather than the proportions.

To eliminate the weights based on the population, we had to convert these counts into percentages relative to the population in each county for a view that can be analyzed across each county.

Another challenge we encountered was with the years of the data we collected. We collected data from 2012, 2016, and 2020 but it is important to note that two of these years can classify as an anomaly. In 2012, the economy and market was rebounding from the recession which could represent higher economic variables. Similarly, in 2020, the economy was in the midst of a global pandemic which led to drastic abnormalities in the data. Being said, we are still in the process of deciding how to proceed with these years and if we should take these into account. Any year prior to 2010 may be too old and not a good representation of today's voting as well.

IV. METHODS

Our study leverages a machine learning model that focuses on six key battleground states – Arizona, Michigan, Nevada, North Carolina, Pennsylvania, Wisconsin – to predict the outcome of the 2024 U.S. presidential election. Each row in the data represents a county's measures in a prior election. More specifically, an observation contains information on six demographic-based variables including race, gender, education level, economic status, age, and poverty rate. Each county is represented in the data three times, one for each of the prior election cycles (2012, 2016, 2020).

We are using supervised learning in this project because we have historical voting data labeled by county as either Democrat or Republican. Supervised learning allows us to train a model using these known labels, so it learns patterns in the data - such as income levels, age distributions, education rates, and other demographic factors - and then apply what it learns to predict the voting outcome in each county for future elections. Specifically, we are performing classification using a decision tree-based model, which predicts discrete categories (Democrat or Republican) rather than a continuous value. In decision tree-based classifications, the decision tree makes predictions by asking a series of yes/no questions about the features.

Our group plans to use decision trees as the basis for our analysis. Decision trees will allow us to find patterns of voting behavior based on the variables we selected in the exploratory data analysis. All explanatory variables are numeric, therefore we hope to find splits between different variables to optimize classification of party voted for by county. With decision trees, we envision a series of decision nodes that distinguish Democratic counties from Republican counties. Outcomes are binary, therefore one terminal node will be set as Democratic and the other Republican. The goal is to find the best decision nodes that effectively split the various numeric explanatory variables at measures that distinguish counties' voting outcomes.

Our approach will be bound to "work" if the splits used to make decisions at each node for prior elections are similar to 2024 patterns. For instance, if a decision is made that counties above 0.55 proportion of White-identifying individuals compose a county vote Republican, then we would hope that the 2024 voting trends warrant a similar split. A drastic change in demographics could jeopardize predictions for the 2024 election. A successful model – accurate voting predictions for battleground states – will involve specification of decision nodes in decision trees.

A Random Forest model improves this approach by building multiple decision trees, each trained on a random subset of the data and features, and then combining their predictions. By taking a majority vote across all the trees, the Random Forest reduces errors from any individual tree and achieves a more accurate, stable prediction. This technique is particularly effective for our project because it handles complex, nonlinear relationships in the data and can work well even when features interact in intricate ways, such as how age and income together might influence voting patterns. The result is a robust prediction for each county based on learned patterns from the training data.

However, decision trees have a high possibility of overfitting, especially if the model is deep with a large number of splits. This may rely heavily on the training set and fail at generalizing to unfamiliar data with future election county demographics. To deal with this, we plan to limit the maximum depth of the tree to control the model's complexity and reduce overfitting. Another weakness can be the limited performance with categorical variables. The model may struggle with very detailed, granular income levels or the education levels categories, so we plan to group them into larger brackets which can simplify the data and help the model find meaningful splits rather than these narrow categorical splits. Finally, the model has a weakness of failing to capture trends. Over time and especially in elections, there are changes in political, economic, and social voter sentiments that are not being accounted for in the data.

If the decision trees fail to predict the election, it would bring attention to issues that we could learn from. First, voting decisions may not be entirely related to demographic data or generalizations here but rather be influenced by non-demographic factors such as current political events or party specific policies that the decision based model won't be able to capture. Secondly, if it fails then we would believe our demographic variables may not be the key drivers of voting behaviors and could be misleading features. This would call for an analysis of other influential factors that could be specific to the current election. Finally, we would learn that there was an overemphasis on historical data and that this isn't sufficient. The constantly changing political and social changes may need a model that is more adaptive and can incorporate real time inputs.

As we prepare the modell, the first step will be data preparation, where we will handle the peculiar characteristics of numeric data to ensure that each feature best fits the requirements of decision tree-based algorithms. Given the demographic variables involved in this project, the structure and representation of data will be critical to model performance and interpretability.

Variables such as race and level of education, which can have multiple categories, will be transformed into binary columns representing each category – separate columns for White, Black, Hispanic, etc., for race. It therefore gives equal representation to each category in the model, not creating an artificial hierarchy that could happen with ordinal encoding. Similarly, for binary features, such as sex, a simpler binary encoding is going to be used in order to encode the two possible values without wasting the expansion of the feature space and an incomprehensible dataset.

We will perform correlation analysis for numeric variables, like median income, poverty rate, unemployment rate, and education levels, to understand the relationships in these features. Although decision trees and random forests can handle multicollinearity, this step will show some insight into potential interactions and dependencies among the demographic and economic indicators that might influence election outcomes. Moreover, looking at correlations will help to see the underlying patterns in each state, which will enable us to interpret model results more concisely. For instance, changes in median income, education attainment, or population changes could be represented as new features, providing a dynamic perspective of evolving county-level characteristics. This helps in the capturing of any socio-economic trends or demographic shifts that might correlate with election results across the years.

V. Results

Results: Results submission, cleaned up to read as part of a paper

For our project we wanted to answer the question: “Which political party is most likely to win the majority of counties in each of the seven key battleground states (Arizona, Georgia, Michigan, Nevada, North Carolina, Pennsylvania, Wisconsin) in the 2024 U.S. presidential election based on demographic and economic indicators?”

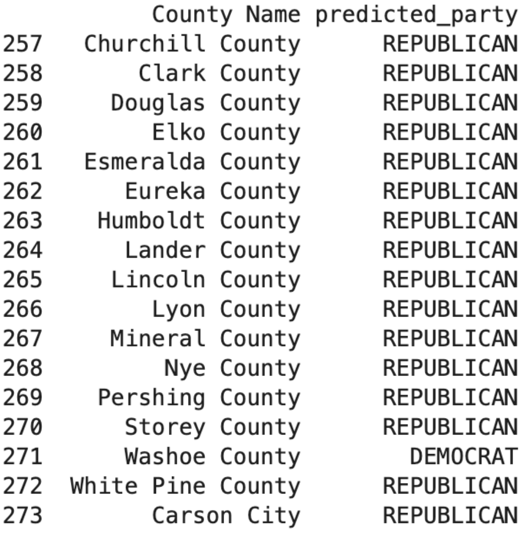
This question emphasizes the specific outcomes being predicted (party winning counties) while anchoring the scope of the analysis to the demographic and economic factors used in the model. It also highlights the practical application of the model, potentially aiding political analysts, campaign strategists, or sociopolitical researchers in understanding the influence of these factors on electoral outcomes. We employed a supervised learning classification model, starting with decision trees and expanding to Random Forests to improve accuracy and robustness. Decision trees allowed us to identify key patterns and thresholds in the data that separated counties voting Democrat from those voting Republican. Random Forests further enhanced these predictions by combining multiple decision trees to mitigate overfitting and capture complex relationships between features.

The results of our analysis closely aligned with the actual 2024 election outcomes across all seven battleground states. Using our Random Forest model, we accurately captured the voting patterns in these states by leveraging demographic and economic indicators like race, education, and poverty rate. This high level of accuracy underscores the effectiveness of our approach in predicting county-level outcomes and demonstrates the stability of key factors influencing voting behavior. The close correspondence between our predictions and the actual results highlights the robustness of our model and the relevance of historical voting trends and demographic data in forecasting electoral outcomes.

Nevada

A state with 17 counties, Nevada did not pose many problems in constructing the random forest model. We merged data from three past election cycles to train and test the model. Some observations (county data for a specific election year) contained missing data for specific measures. We imputed values of 0 if they were insignificant/common (i.e. proportion of the population identifying with the Hawaiian race). Otherwise, we removed the rows.

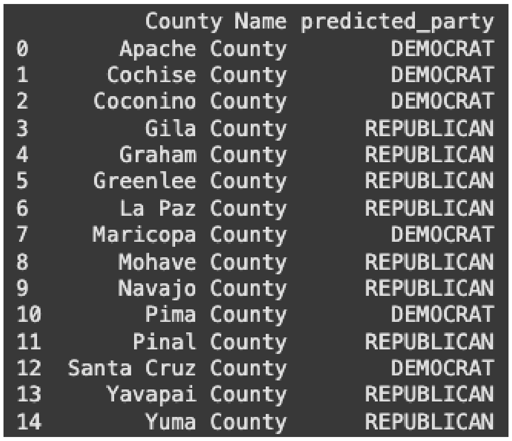
Attached are the results of the random forest model. 16/17 counties were predicted to be Republican majorities for the 2024 U.S. presidential election. In comparison with the actual results, we found that our model had a success rate of 16/17 (94.11%). Clark County, while predicted to be Republican, turned out to be Democratic. This was the only discrepancy between the model's predicted results and the actual results of the 2024 U.S. presidential election.



```
          County Name predicted_party
257   Churchill County      REPUBLICAN
258       Clark County      REPUBLICAN
259     Douglas County      REPUBLICAN
260        Elko County      REPUBLICAN
261   Esmeralda County      REPUBLICAN
262      Eureka County      REPUBLICAN
263    Humboldt County      REPUBLICAN
264      Lander County      REPUBLICAN
265     Lincoln County      REPUBLICAN
266        Lyon County      REPUBLICAN
267     Mineral County      REPUBLICAN
268         Nye County      REPUBLICAN
269    Pershing County      REPUBLICAN
270      Storey County      REPUBLICAN
271      Washoe County        DEMOCRAT
272  White Pine County      REPUBLICAN
273         Carson City      REPUBLICAN
```

Arizona

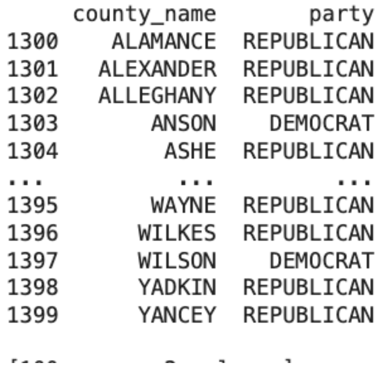
As we built the random forest model, we merged the prior 3 past elections (2012, 2016, 2020) to train the model. However, the model ran into a problem of roughly 272 incidents of NaNs being found. To mitigate this error as we split the data, we decided to fill the na's with 0 and proceeded with the random forest. As our party variables were the variable of interest to predict, we created a label encoder to represent the Democrat and Republican party for each county. After running the model, Arizona's 2024 county winning parties were predicted and were almost identical to the actual results of the election.



```
        County Name predicted_party
0     Apache County        DEMOCRAT
1    Cochise County        DEMOCRAT
2   Coconino County        DEMOCRAT
3       Gila County      REPUBLICAN
4     Graham County      REPUBLICAN
5   Greenlee County      REPUBLICAN
6     La Paz County      REPUBLICAN
7   Maricopa County        DEMOCRAT
8     Mohave County      REPUBLICAN
9     Navajo County      REPUBLICAN
10      Pima County        DEMOCRAT
11     Pinal County      REPUBLICAN
12 Santa Cruz County       DEMOCRAT
13   Yavapai County      REPUBLICAN
14      Yuma County      REPUBLICAN
```

There was only one county that differed from the prediction model to the actual outcomes which was Cochise County, predicted to be Democrat but was Republican in the 2024 presidential election. This resulted in our models having a success rate of 92.86% for Arizona (13/14).

North Carolina



```
     county_name       party
1300    ALAMANCE  REPUBLICAN
1301   ALEXANDER  REPUBLICAN
1302   ALLEGHANY  REPUBLICAN
1303       ANSON    DEMOCRAT
1304        ASHE  REPUBLICAN
...          ...         ...
1395       WAYNE  REPUBLICAN
1396      WILKES  REPUBLICAN
1397      WILSON    DEMOCRAT
1398      YADKIN  REPUBLICAN
1399      YANCEY  REPUBLICAN
```

The prediction question for the model is: How likely is each county in North Carolina to vote for the Democratic or Republican candidate in the 2024 Presidential Election, based on historical voting patterns and demographic data from 2012, 2016, 2020, and recent statistics? The key output of this code is a prediction table which provides predictions on the winning party for each county in North Carolina. The predictions align with historical trends, such as rural areas leaning Republican and urban areas leaning Democrat. The random forest feature importance analysis identifies top variables influencing predictions:

1. Males 65+ Percentage: Older populations tend to lean Republican
2. Bachelor’s Degree Percentage: Higher education levels are often associated with Democratic outcomes in urban counties.
3. Black or African American Alone Percentage: Counties with larger Black populations may show stronger Democratic preferences due to historical voting patterns.
4. Median Household Income: Counties with a higher percentage of lower-income households could vote Democrat.

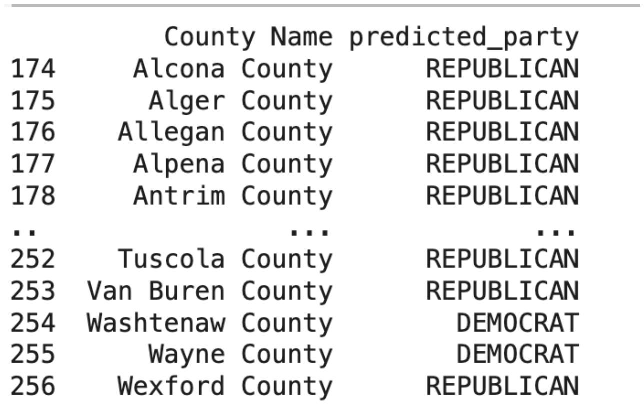
5. High School Diploma Percentage: See #2

Some limitations in our data are missing or imputed data that could introduce biases. There is also an absence of real-time confounders like polling or economic activity could limit the model's ability to adapt to current trends. The results align with the 2024 Presidential Election. Age, race, and education are significant predictors of county-level voting trends.

Michigan

To construct our random forest model, we merged data from the previous three elections - 2012, 2016, and 2020 - for training purposes. However, this process introduced over 1000 instances of missing values (NaNs) within the dataset. To mitigate this issue, linear interpolation was applied to fill gaps where possible for numeric columns. Persistently problematic rows which could not be adequately resolved through these methods were manually dropped for specific indices to ensure the dataset was fully cleaned. Remaining missing values in X_train and y_train were handled by replacing NaNs with zeros. The target variable was then converted to string type and encoded using LabelEncoder, and inconsistencies in labels were managed by mapping unexpected entries to valid options - DEMOCRAT" and "REPUBLICAN".

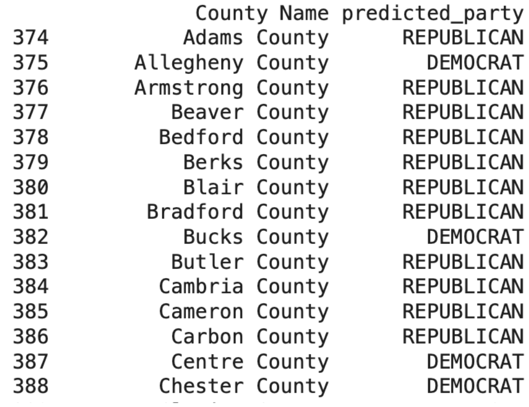
Overall, the model was accurate in predicting 80 out of 83 counties for the 2024 election. It did not predict that Marquette County and Leelanau County would go Democrat, nor did it predict that Macomb County would go Republican in 2024. However, the model was still very accurate, achieving a 96.39% success rate. As expected, urban areas, including counties surrounding Detroit, Grand Rapids, Ann Arbor, and Lansing, were predicted to vote Democratic, while more rural counties generally leaned Republican.



```
          County Name predicted_party
174     Alcona County      REPUBLICAN
175      Alger County      REPUBLICAN
176    Allegan County      REPUBLICAN
177     Alpena County      REPUBLICAN
178     Antrim County      REPUBLICAN
..                ...             ...
252    Tuscola County      REPUBLICAN
253  Van Buren County      REPUBLICAN
254  Washtenaw County        DEMOCRAT
255      Wayne County        DEMOCRAT
256    Wexford County      REPUBLICAN
```

Pennsylvania

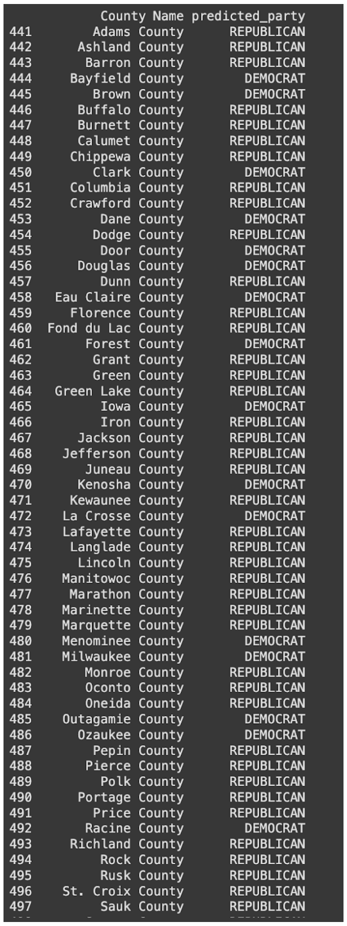
During this past presidential election, Pennsylvania emerged as a crucial battleground state with the potential to determine the next President of the United States. In 2020, after voting Republican in the 2016 election, Pennsylvania flipped blue. However, in the most recent election, the state returned to red. So, we thought it would be interesting to see how the model would predict the state’s outcome of this latest election. To construct the model, we combined data from three presidential election cycles– including 2012, 2016, and 2020– alongside county data for each of the swing states. These were the results of some of the counties in Pennsylvania:



```
          County Name predicted_party
374      Adams County      REPUBLICAN
375  Allegheny County        DEMOCRAT
376  Armstrong County      REPUBLICAN
377     Beaver County      REPUBLICAN
378    Bedford County      REPUBLICAN
379      Berks County      REPUBLICAN
380      Blair County      REPUBLICAN
381   Bradford County      REPUBLICAN
382      Bucks County        DEMOCRAT
383     Butler County      REPUBLICAN
384    Cambria County      REPUBLICAN
385    Cameron County      REPUBLICAN
386     Carbon County      REPUBLICAN
387     Centre County        DEMOCRAT
388    Chester County        DEMOCRAT
```

After running the model, it was found that 61/67 counties were accurately predicted. The remaining six counties, which the model initially predicted would vote Democrat, ultimately voted Republican (as indicated by the data from this past election cycle). These included Bucks, Cumberland, Erie, Lancaster, Monroe and Northampton counties. This meant that, for the state of Pennsylvania, the model had a 91.04% success rate.

Wisconsin



```
          County Name predicted_party
441      Adams County      REPUBLICAN
442    Ashland County      REPUBLICAN
443     Barron County      REPUBLICAN
444   Bayfield County        DEMOCRAT
445      Brown County        DEMOCRAT
446    Buffalo County      REPUBLICAN
447    Burnett County      REPUBLICAN
448    Calumet County      REPUBLICAN
449   Chippewa County      REPUBLICAN
450      Clark County        DEMOCRAT
451   Columbia County      REPUBLICAN
452   Crawford County      REPUBLICAN
453       Dane County        DEMOCRAT
454      Dodge County      REPUBLICAN
455       Door County        DEMOCRAT
456    Douglas County        DEMOCRAT
457       Dunn County      REPUBLICAN
458 Eau Claire County        DEMOCRAT
459   Florence County      REPUBLICAN
460 Fond du Lac County      REPUBLICAN
461     Forest County        DEMOCRAT
462      Grant County      REPUBLICAN
463      Green County      REPUBLICAN
464 Green Lake County      REPUBLICAN
465       Iowa County        DEMOCRAT
466       Iron County      REPUBLICAN
467    Jackson County      REPUBLICAN
468  Jefferson County      REPUBLICAN
469     Juneau County      REPUBLICAN
470    Kenosha County        DEMOCRAT
471   Kewaunee County      REPUBLICAN
472  La Crosse County        DEMOCRAT
473  Lafayette County      REPUBLICAN
474   Langlade County      REPUBLICAN
475    Lincoln County      REPUBLICAN
476  Manitowoc County      REPUBLICAN
477   Marathon County      REPUBLICAN
478  Marinette County      REPUBLICAN
479  Marquette County      REPUBLICAN
480  Menominee County        DEMOCRAT
481  Milwaukee County        DEMOCRAT
482     Monroe County      REPUBLICAN
483     Oconto County      REPUBLICAN
484     Oneida County      REPUBLICAN
485  Outagamie County        DEMOCRAT
486    Ozaukee County        DEMOCRAT
487      Pepin County      REPUBLICAN
488     Pierce County      REPUBLICAN
489       Polk County      REPUBLICAN
490    Portage County      REPUBLICAN
491      Price County      REPUBLICAN
492     Racine County        DEMOCRAT
493   Richland County      REPUBLICAN
494       Rock County      REPUBLICAN
495       Rusk County      REPUBLICAN
496  St. Croix County      REPUBLICAN
497       Sauk County      REPUBLICAN
```

The decision tree model was used to predict the dominant political party (Republican or Democrat) for Wisconsin counties in the 2024 election, with the actual 2020 outcomes serving as a baseline for comparison. The model performed very well, predicting that 13 counties would flip their party alignment in 2024, which was remarkably close to the actual 15 counties that flipped, achieving an 87% success rate for predicting flips. When applied to the 2020 outcomes, the model had a lower success rate, misclassifying several counties like Kenosha and Green, which voted Republican in 2020 but were incorrectly predicted as Democrat. The stronger performance in forecasting 2024 outcomes suggests the model was effective at identifying shifting voter patterns and adapting to recent trends. These results demonstrate its potential as a tool for predicting election outcomes, especially when combined with accurate data and fine-tuned parameters. To further enhance the model's accuracy, additional data sources such as voter demographics, economic changes, and turnout rates could be incorporated, making it even more reliable for future predictions.

VI. Conclusion - Emily, Shaveen

Two or three pages that summarize your findings for people who have read the paper. In particular, describe extensions, complications, problems, limitations that you ran into that could form the basis of future work (turn the weaknesses of your paper into results/features, rather than flaws).

To effectively predict the outcome of the 2024 U.S. presidential election, our group built a Random Forest classification mode, focusing specifically on six key battleground states: Arizona, Michigan, Nevada, North Carolina, Pennsylvania, and Wisconsin. Initially, we hypothesized that demographic and socioeconomic factors– including age, race, education level, and income distribution– would be particularly effective in offering an accurate outcome. As previously outlined, we proposed a series of hypotheses about the relationships between these variables and voting patterns. As largely confirmed by our Exploratory Data Analysis on election years 2012, 2016, and 2020, we hypothesized that younger, more diverse, and higher-income populations would favor Democratic candidates, while older, less diverse, and lower-income populations would favor Republican candidates.

Throughout the process, we faced challenges in effectively managing the substantial amount of missing data, especially in critical demographic predictor variables and even within the 2024 election results itself. Each member of our group was responsible for their own model, which made cleaning vary from state to state. In several instances, we applied linear interpolation to address missing values in numeric variables and strategically removed rows with consistently problematic gaps. In other cases, we coded for missing values (NaNs) to be replaced with zeros, particularly when the missing data were near negligible, and would not significantly impact the results. While we aimed to rectify missing data challenges, we recognize this ultimately undermined the robustness of our model. However, working with imperfect data is an inherent part of the model-making process, and we applied techniques within our skillset to minimize the impact of this imperfect data as much as possible.

Evaluation: Despite the strengths of the model, there are various natural limitations given the complexity of predicting electoral outcomes. Much of our model relies on correlating demographic and socioeconomic data with past political outcomes, and using more recent updates of this data post-2020 to predict 2024 voting behavior. However, a critical concern is whether demographic data alone is an effective predictor of elections, as variables such as age, race, and education levels tend to change very slowly over time. While these factors may highlight long-term voting trends, they do not capture the dynamic, short-term influences that drive voter behavior during individual election cycles. Additionally, the impact of COVID-19 introduces further uncertainty. The pandemic not only influenced voter turnout and preferences in the 2020 election, but may have also had lasting effects on political engagement and attitudes that are difficult to quantify. The 2020 presidential election cycle posed unique circumstances that do not align with the dynamics of previous elections or the 2024 cycle. Adding to this complexity is President Biden's abrupt drop from the 2024 election, which not only altered the political landscape, but also complicated predictions as voter behavior may diverge from patterns observed during his presidency. Elections are not determined solely by static population characteristics, but are often shaped by political dynamics, such as campaign strategies and the appeal of the individual candidate.

Additionally, the importance of specific issues often fluctuates from election year to year depending on major events, political climate, and media coverage, all of which cannot be accounted for solely by demographic predictors. For instance, Pew Research reported that in the 2024 election, the economy was the most important issue for voters overall (81%). This concern, however, varied significantly by political allegiance: 68% of Harris supporters ranked the economy as "very important," compared to 93% of Trump supporters.[1] Healthcare emerged as the second most important issue, with greater emphasis from Harris supporters (76%) than Trump supporters (55%). Supreme Court appointments were the third most important issue, with 73% of Harris supporters considering them "very important" compared to 54% of Trump supporters. Notably, the Supreme Court's decision to overturn *Roe v. Wade* in 2022 elevated abortion as a key issue. Approximately 67% of Harris Supporters considered abortion "very important," nearly double the share of Biden voters in 2020, while only 35% of Trump supporters considered the issue equally significant. These shifts demonstrate how external events and changing issue salience can significantly influence voter priorities, factors that are not captured by demographic data alone.

While our model captures larger structural trends in voter behavior, it assumes past patterns will remain stable. However, as recent elections have shown, unexpected events can disrupt these trends. The anomalies observed in 2012, when the market was still actively recovering from the lasting effects of the Great Financial Crisis, and 2020, during the unprecedented COVID-19 pandemic, highlight how external shocks can fundamentally change and distort the correlation between various socioeconomic variables and electoral results. While this approach offers clarity in stable periods, it may oversimplify voter behavior in highly volatile contexts.

However, these limitations can be viewed as opportunities for future extensions of this work. While many of these considerations are beyond the scope of this class and present challenges even for experts, integrating more dynamic, real-time predictors could enhance the model. For example, incorporating economic indicators, voter turnout rates, and survey-based measures of issue importance could allow the model to reflect changing voter concerns. Additionally, tracking public opinion trends, campaign-specific messaging, or major news events could better capture short-term shifts in political behavior. By combining static demographic data with dynamic predictors, future iterations of the model would be better equipped to account for both long-term structural trends and the real-time factors that influence elections. This approach would not only improve predictive accuracy but also provide a more nuanced understanding of the evolving political landscape.

[1] https://www.pewresearch.org/politics/2024/09/09/issues-and-the-2024-election/

Bibliography

https://www.pewresearch.org/politics/2024/09/09/issues-and-the-2024-election/