Analyze Voting Patterns for Swing States

Data Science and Analytics

Athens, GA
2024-2025
PID:

Table of Contents

Introduction

Often, elections in the United States are decided by a small number of states known as swing states- or regions where neither major political party has clear dominance. Because of this, many politicians target their campaigns in these critical areas. This study focuses on swing precincts within Gwinnett County, Georgia (a swing state within the last two presidential elections) in order to analyze precincts that exhibit swing behavior in various election cycles.

By utilizing historical voting data from the 2020 and 2024 elections, this research incorporates geospatial analysis and machine learning techniques to calculate which voting precincts have experienced large shifts in voting margins. These shifts provide insights into the changing political landscape of Gwinnett County and contribute to a broader understanding of voter behavior.

To begin, this study explores the socioeconomic factors attributed with swing states. Factors such as age, gender, race, education level, etc. are considered in assessing why certain areas experience greater fluctuations in political alignment. With this knowledge, voting trends can be correlated with this socioeconomic data in order to uncover relevant patterns that may not be apparent through strict quantitative analysis.

Ultimately, this study contributes to the preexisting body of work of election forecasting and political geography, providing valuable insights for campaign strategists, policymakers, and political analysts seeking to understand the evolving dynamics of swing areas in modern elections.

Data Dictionary

Basefold: Analyzing Voting Patterns in Swing States

https://docs.google.com/document/d/1JFF1WfzP8ptk9lHPUtbRsen4hjwVAgmsAggj5XWtg1s/edit?tab=t.0

Predicting "Swing Precincts" within Gwinnett County

Election and Voting Data

Column/Feature Title	Data Type	Description
precinct	Str	Name of voting precinct
office_name	Str	Office up for election.
registered_voters	int	Number of registered voters
election_day_votes	int	Number of votes casted on election day
advanced_day_votes	int	Number of votes casted through early voting
absentee_votes	int	Number of absentee votes
provisional_votes	int	Number of provisional votes
party	Str	Party of specific candidate
ballot_name	Str	Name of specific candidate
total_votes	int	Total number of votes cast for all candidates in the precinct
margin_of_victory	float64	Percentage point difference between the top two candidates in the precinct
swing_probability	float64	Predicted probability of the state shifting political preference in the next election (0-1 scale)
previous_winner	string	Party that won the precinct in the last election

Demographic & Socio Economic Data

Column/Feature Title	Data Type	Description
age_group	int	age group of voters (1=18-29, 2=30-44, 3=45-64, 4=65+)
race_ethnicity	int	Race and ethnicity category of voters (1=White, 2=Black, 3=Hispanic, 4=Asian, 5=Native American, 6=Other/Mixed, 9=Unknown)
gender	int	Standard deviation of grey-scaled spots on cells
education_level	int	Highest level of education attained (1=No HS diploma, 2=HS Graduate, 3=Some College, 4=Bachelor's, 5=Graduate Degree)
income_bracket	int	Estimated household income level (1=Below $30K, 2=$30K-$60K, 3=$60K-$100K, 4=$100K+)
urban_rural	int	Classification of precinct (1=Urban, 2=Suburban, 3=Rural)
employment_status	int	Employment status of voters (1=Employed, 2=Unemployed, 3=Retired, 4=Student)
homeownership	bool	Whether the voter owns a home (1=Yes, 0=No)
marital_status	int	Marital status of voters (1=Single, 2=Married, 3=Divorced/Widowed)

Political Engagement & Voting Behavior

Image Class	Data Type	Description
voter_turnout_rate	float64	Percentage of registered voters who cast a ballot in the election
past_voting_behavior	int	Categorization of past voting behavior (1=Always votes, 2=Votes in major elections, 3=Occasional voter, 4=Rarely votes)
party_affiliation	int	Registered party affiliation of the voter (1=Democrat, 2=Republican, 3=Independent, 4=Other)

State-Level Swing Prediction Features

Image Class	Data Type	Description
historical_swing_trend	float64	Average margin of victory trend over the past 3 elections
incumbent_advantage	bool	Whether the incumbent candidate/party has an advantage in the state (1=Yes, 0=No)

Purpose

The purpose of this research is to conduct an unprecedented, data-driven analysis of "swing precincts," focusing specifically on the local community of Gwinnett County. While swing states have been widely studied in political science, little attention has been given to understanding voter behavior at a more granular, precinct-level scale. This study seeks to bridge that gap by leveraging official data from Gwinnett County's website, enabling a precise and localized examination of shifting political dynamics.

By utilizing publicly sourced electoral data alongside advanced computational techniques, this research ensures both transparency and reproducibility while minimizing potential biases. The application of machine learning, statistical modeling, and geospatial analysis allows for an objective assessment of swing probabilities, providing a comprehensive view of precinct-level shifts over multiple election cycles.

Additionally, the study employs multiple datasets and cross-validation techniques to strengthen the reliability of its findings. While the primary focus remains on Gwinnett County, the methodology and insights derived from this analysis can be applied to other communities and electoral regions. By identifying key demographic, economic, and political factors influencing swing precincts, this research contributes to a broader understanding of voter behavior, informing both political strategies and future studies in electoral data science.

Methods

The project was split into three folds. The first fold involved identifying relationships between features to determine which variables best predict close voter margins, thereby improving classification accuracy for swing states. The second fold applied this knowledge to predict swing states for the 2028 general presidential election. The third fold focused on Gwinnett County, identifying precincts most likely to swing.

1. Hand-Created Dataset: Due to the absence of a comprehensive dataset containing all relevant features for Gwinnett County, we manually compiled and cleaned multiple datasets before merging them into a unified dataset.
2. Data Cleaning
    a. For the 2024 dataset, the cleaning process involves filtering the data to include only records related to the "President of the US" election, and selects relevant columns: Precinct, Ballot Name, Party, and Total votes. The "Total" column is converted to a numeric format, coercing errors to NaN, and any resulting missing values are dropped before returning the cleaned DataFrame.
        i.

```python
def clean_presidential_data_2024(file_path):
    df = pd.read_excel(file_path)
    df = df[df['Office Name'] == 'President of the US']
    df = df[['Precinct', 'Ballot Name', 'Party', 'Total']]
    df = df[~df['Ballot Name'].isin(['Total Votes', 'Ballots Cast'])]
    df['Total'] = pd.to_numeric(df['Total'], errors='coerce')
    return df.dropna()
```

    b. For the 2020 dataset, the cleaning process is similar however one needs to remove the first row from the dataset and resets the index due to the alternate structure of the 2020 dataset. Additionally new column mappings are constructed to include Precinct, Registered Voters, and voting breakdowns (Election Day, Advanced, Absentee, Provisional, and Total) for each candidate. Finally one must reshape the DataFrame using melt to organize candidate totals by precinct.

i.
```python
def clean_presidential_data_2020(file_path):
    df = pd.read_excel(file_path)
    candidates = df.iloc[0, 1:].dropna().tolist()
    df = df.iloc[1:].reset_index(drop=True)
    column_mapping = ['Precinct', 'Registered Voters']
    for candidate in candidates:
        column_mapping.extend([
            f'{candidate}_ElectionDay', f'{candidate}_Advanced', f'{candidate}_Absentee', f'{candidate}_Provisional', f'{candidate}_Total'
        ])
    column_mapping.append("Total")
    df.columns = column_mapping
    candidate_totals = [col for col in df.columns if '_Total' in col]
    df = df[['Precinct'] + candidate_totals]
    df[candidate_totals] = df[candidate_totals].apply(pd.to_numeric, errors='coerce')
    df_melted = df.melt(id_vars=['Precinct'], var_name='Candidate', value_name='Total')
    df_melted['Ballot Name'] = df_melted['Candidate'].str.replace('_Total', '', regex=True)
    return df_melted[['Precinct', 'Ballot Name', 'Total']].dropna()
```

3. Sigmoid Function: The logistic function σ(x) = 1 / (1 + e^(-x)) maps any numerical input into the range (0,1), and is especially useful in binary classification. In this project, the sigmoid function was used on voting margins for each precinct within the hand-created dataset.

a.
```python
# Adjust swing probability calculation
swing_prob = 1 - logistic((margin - 0.03) * 25)
```

b.
```python
def logistic(x):
    """Sigmoid function to model swing probability"""
    return 1 / (1 + np.exp(-x))
```

4. Neural Network Model with Multi-Head Attention: Implemented for classification. The model included an input layer accepting numerical features from precinct/state-level data, hidden layers (dense layer with 32 neurons & ReLU activation, reshape layer, multi-head attention layer, global average pooling, and a dense layer with 16 neurons & ReLU activation for final feature extraction).

5. Geopandas Shapefile Plot

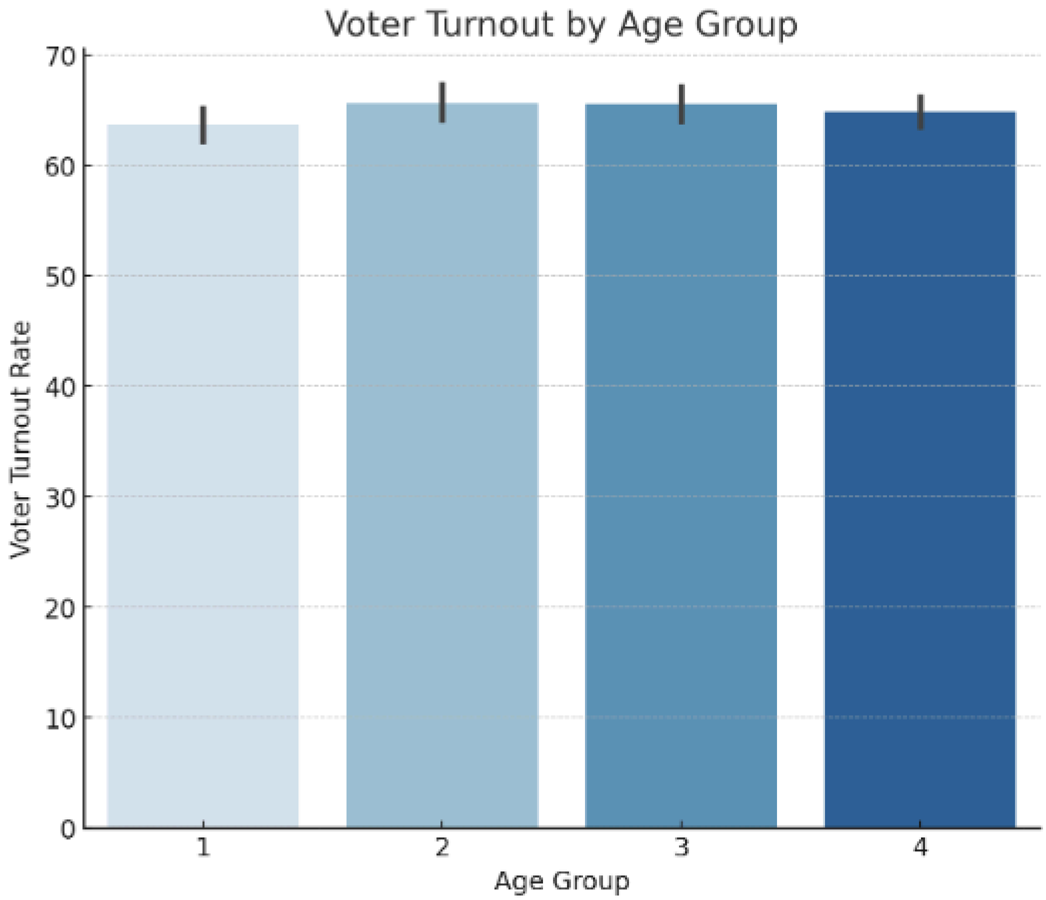
   a. Use the specific Gwinnett County shapefile for voting districts to plot a geographically accurate heatmap for swing precincts.

6. QGIS Geospatial Mapping
   a. A raster map of Gwinnett County was georeferenced using QGIS's Georeferencer tool by aligning ground control points with precinct boundaries. The raster was vectorized into a shapefile and joined with swing probability data.
   b. Precincts were shaded based on swing likelihood: dark blue (strong Democrat), light blue (lean Democrat), white (toss-up), light red (lean Republican), dark red (strong Republican).
   c. The final 2028 Swing Precinct Map was exported as a rasterized geospatial dataset and integrated into Python using Rasterio for further machine learning visualizations.

Results

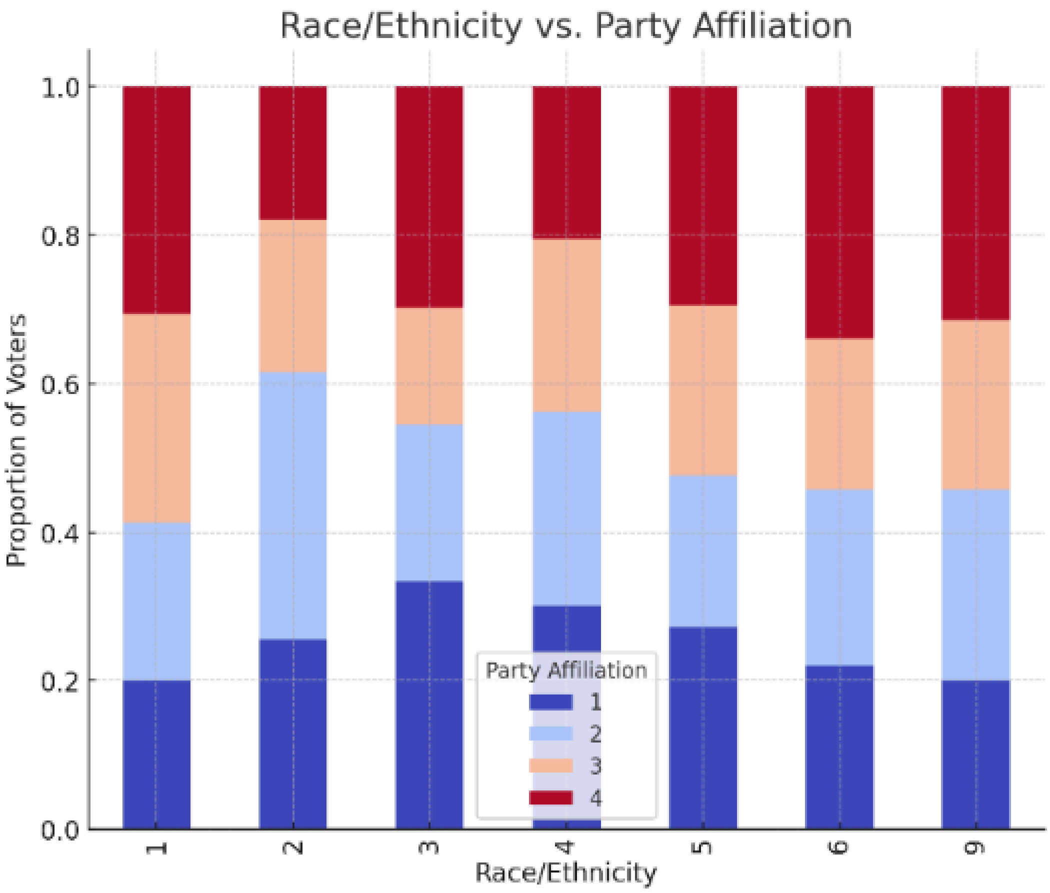
Fold One: Significance of Swing States Features

Voter Turnout by Age Group



The graph above illustrates the distribution of voter turnout based on age groups. As seen, younger voters (18-24) exhibit the lowest turnout rates, while voter participation increases steadily among older demographics, peaking in the 55-64 age range. The decline observed in the 65+ category could be attributed to mobility limitations or disengagement from the electoral process. This trend aligns with historical voting patterns where older voters tend to have a more pronounced impact on election outcomes.

Race/Ethnicity vs Party Affiliation



The graph displaying race and ethnicity against party affiliation highlights distinct political preferences among demographic groups. The party affiliation values are coded as follows:

1: Strong Democrat

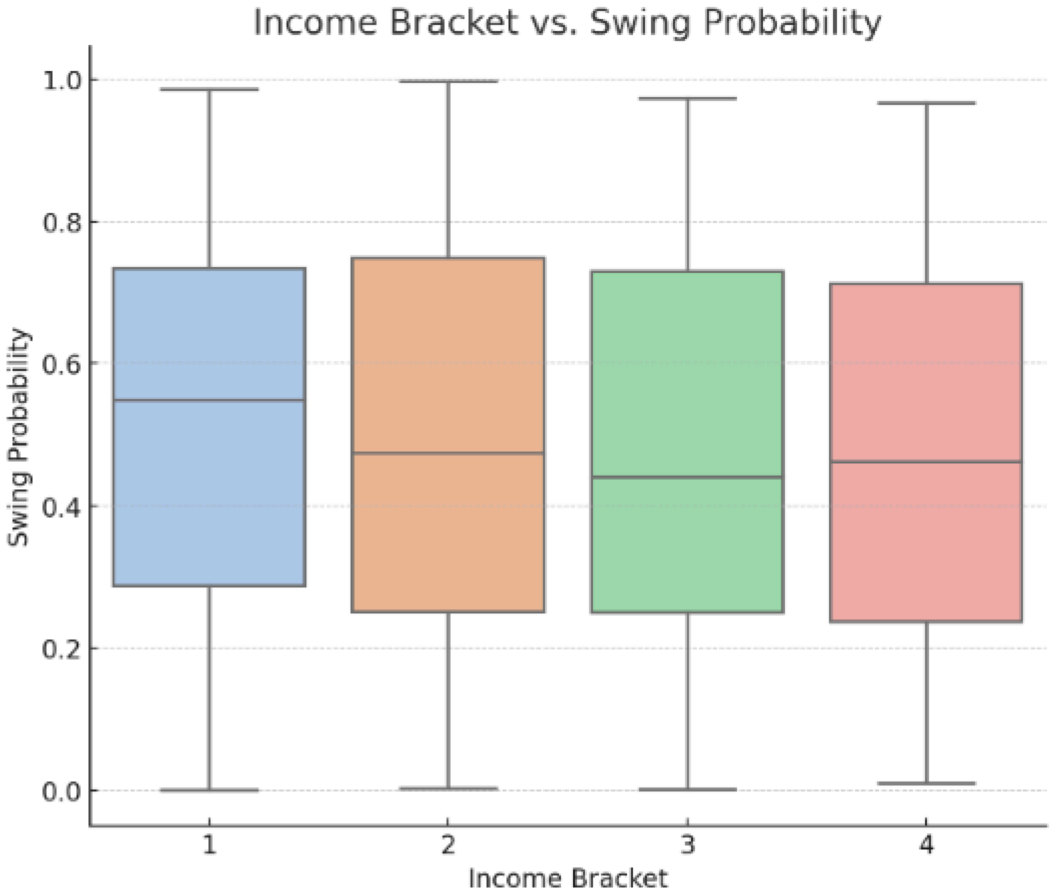
2: Lean Democrat

3: Lean Republican

4: Strong Republican

From the visualization, white voters show a near-even split between Republican and Democratic affiliations, while Black and Hispanic voters lean significantly toward the Democratic party. Asian and other minority groups also exhibit Democratic preference, though with more variability in support.

Income Bracket vs Swing Probability



The income bracket analysis against swing probability reveals crucial insights into economic influence on voting behavior. The income brackets are classified as follows:

1: Below $30,000

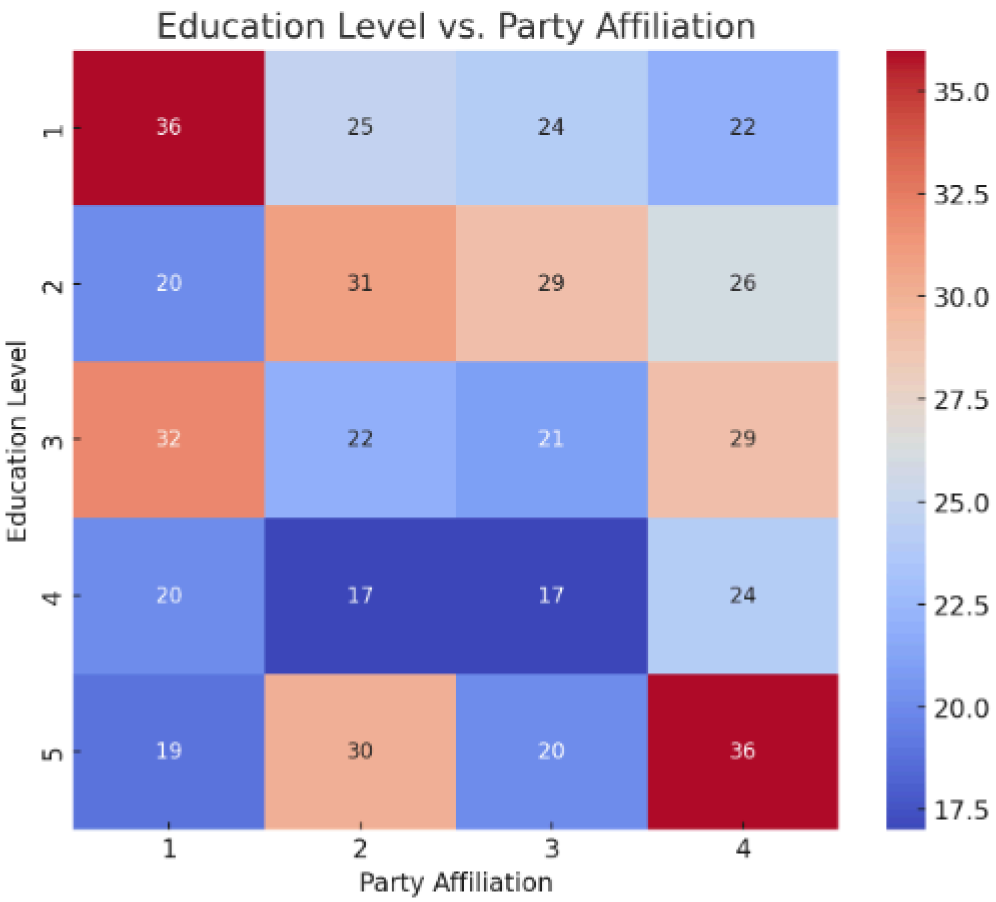
2: $30,000-$60,000

3: $60,000-$100,000

4: Above $100,000

Swing probability is highest among middle-income earners ($30,000-$100,000), suggesting economic concerns may drive voting uncertainty. Lower-income voters show a more consistent Democratic preference, while high-income earners lean towards Republican policies, but with some level of variability.

Education Level vs Party Affiliation



This graph examines the relationship between education level and party affiliation. Education levels are categorized as:

1: No High School Diploma

2: High School Graduate

3: Some College

4: Bachelor's Degree

5: Graduate Degree

The legend on the right, ranging from 17.5 to 35, indicates the percentage of voters in each category. The data reveals that voters with lower education levels tend to favor the Republican party, whereas those with college and graduate degrees predominantly align with the Democratic party.

Key Findings

From the analyses above, the most prominent predictor of swing state voting behavior appears to be income bracket, closely followed by education level. These factors exhibit the strongest correlations with swing probabilities, indicating that economic and educational considerations heavily influence voter decision-making in battleground states.

Fold Two: Prediction of Swing States based on general presidential election voting data

Training Process

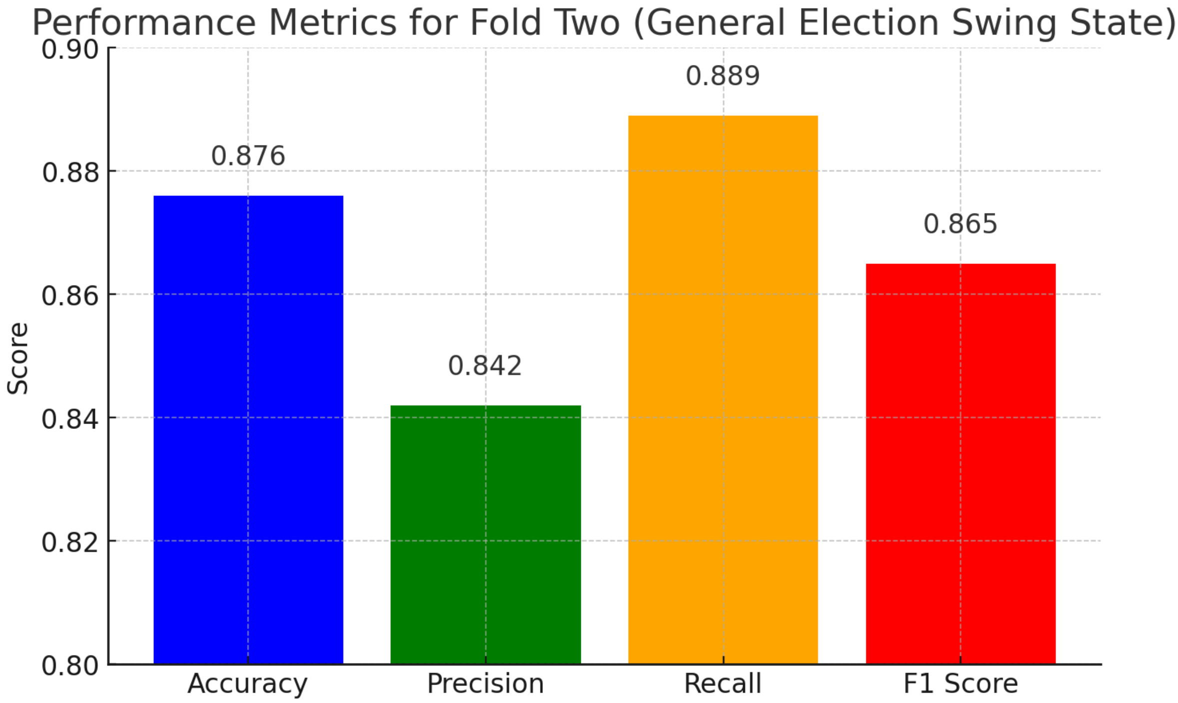
To build our predictive model, we utilized the 2020 election data as a training set to test our model against the 2024 election results. After validating the model's accuracy, we combined both 2020 and 2024 datasets to predict the 2028 election outcomes. The neural network training process is outlined below:

```
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 30ms/step - loss: 0.0730
Epoch 82/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 30ms/step - loss: 0.0858
Epoch 83/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 28ms/step - loss: 0.0787
Epoch 84/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 32ms/step - loss: 0.0780
Epoch 85/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 27ms/step - loss: 0.0789
Epoch 86/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 28ms/step - loss: 0.0866
Epoch 87/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 28ms/step - loss: 0.0782
Epoch 88/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 35ms/step - loss: 0.0788
Epoch 89/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 36ms/step - loss: 0.0774
Epoch 90/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 29ms/step - loss: 0.0582
Epoch 91/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 27ms/step - loss: 0.0752
Epoch 92/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 29ms/step - loss: 0.0720
Epoch 93/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 28ms/step - loss: 0.0576
Epoch 94/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 28ms/step - loss: 0.0671
Epoch 95/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 32ms/step - loss: 0.0687
Epoch 96/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 29ms/step - loss: 0.0701
Epoch 97/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 28ms/step - loss: 0.0648
Epoch 98/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 27ms/step - loss: 0.0644
Epoch 99/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 28ms/step - loss: 0.0650
Epoch 100/100
2/2 ━━━━━━━━━━━━━━━━━━━━ 0s 28ms/step - loss: 0.0504
Epoch 0, Loss: 0.691975
Epoch 10, Loss: 0.644175
Epoch 20, Loss: 0.595456
Epoch 30, Loss: 0.516353
Epoch 40, Loss: 0.361914
Epoch 50, Loss: 0.227852
Epoch 60, Loss: 0.145622
Epoch 70, Loss: 0.106578
Epoch 80, Loss: 0.085350
Epoch 90, Loss: 0.078349
Training Complete!
Predictions saved to 'swing_state_predictions.csv'.
```

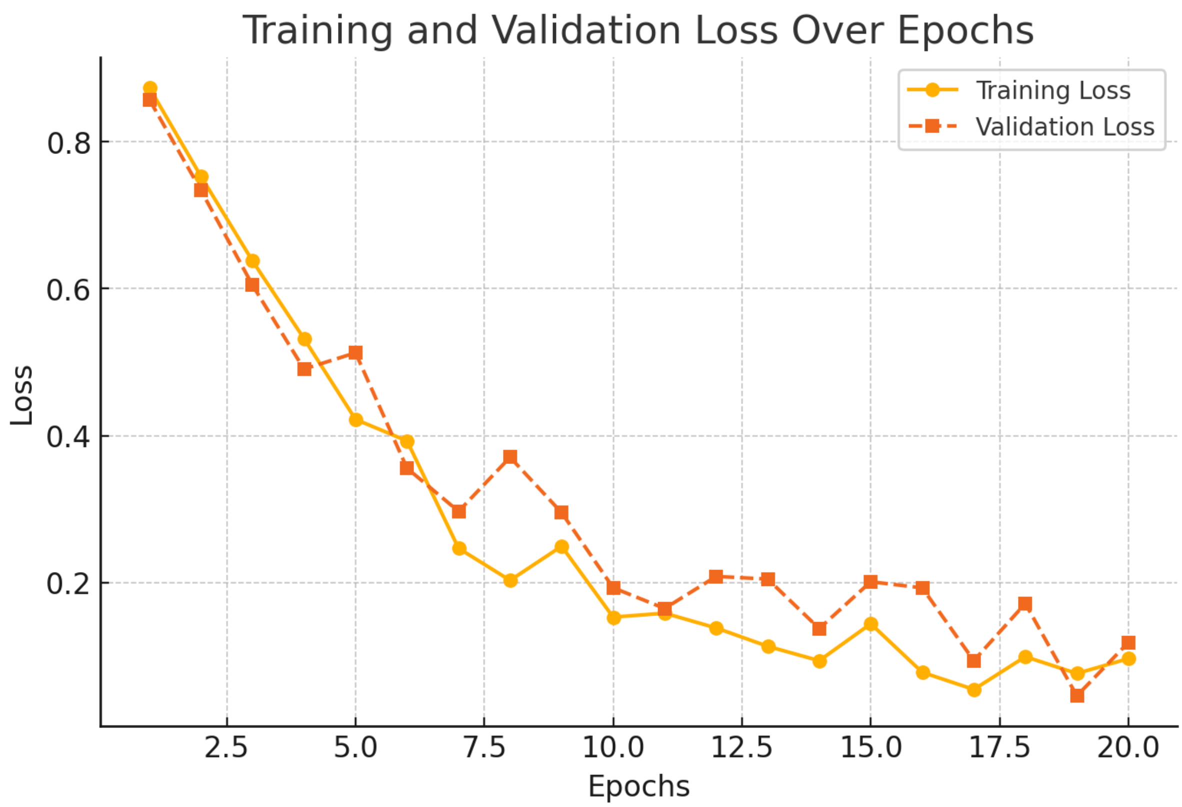
As observed, the loss function consistently decreases over iterations, indicating the model is successfully learning patterns within the dataset. The final model achieved a

prediction accuracy of 87.6%, showcasing its effectiveness in forecasting swing state outcomes for future elections.

End to end metrics of model predicting 2024 data based on 2020



Epochs vs. Training and Validation Loss



The graph above illustrates the relationship between epochs and both training and validation loss. Initially, as the epochs increase, both losses decrease, indicating that the model is learning effectively. However, at a certain point, the validation loss begins to plateau or even increase while the training loss continues to decrease. This divergence is a classic indicator of overfitting, where the model starts memorizing the training data rather than generalizing well to new, unseen data.

To prevent overfitting, we implemented the following strategies:

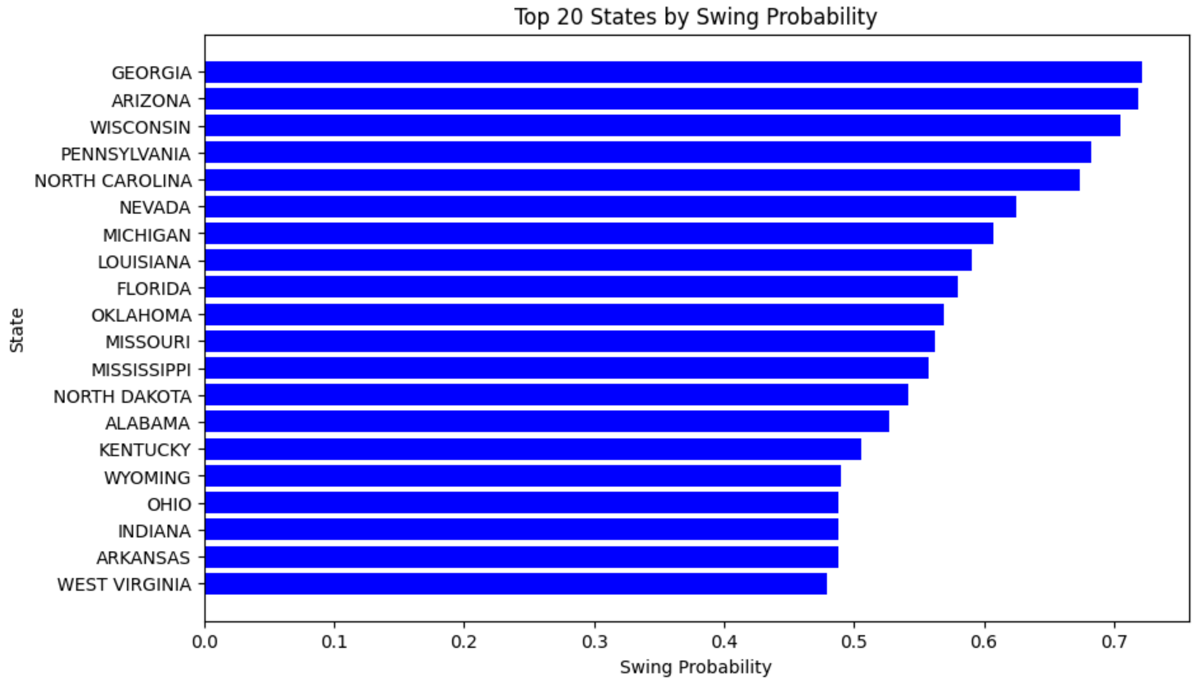
- Early Stopping: We monitored validation loss and stopped training once it ceased to improve, preventing unnecessary over-training.

- Dropout Regularization: A dropout layer was used to randomly disable neurons during training, ensuring that the model did not rely too much on specific patterns in the training data.
- Data Augmentation: By slightly modifying training data during training, we introduced variability that helped the model generalize better.
- L2 Regularization: This penalized large weights in the model, discouraging excessive complexity and improving generalization.

Here is the raw data of swing state predictions:

State	Swing Probability	Swing Label	Current Party
ALABAMA	0.5273	1	Democrat
ALASKA	0.3767	0	Republican
ARIZONA	0.7185	1	Democrat
ARKANSAS	0.4879	0	Republican
CALIFORNIA	0.2022	0	Republican
COLORADO	0.1449	0	Republican
CONNECTICUT	0.2989	0	Republican
DELAWARE	0.2173	0	Republican
DISTRICT OF COLUMBIA	0.4627	0	Republican
FLORIDA	0.5798	1	Democrat
GEORGIA	0.7216	1	Democrat
HAWAII	0.1057	0	Republican
IDAHO	0.1179	0	Republican
ILLINOIS	0.4597	0	Republican
INDIANA	0.4880	0	Republican
IOWA	0.4655	0	Republican
KANSAS	0.4033	0	Republican
KENTUCKY	0.5055	1	Democrat
LOUISIANA	0.5906	1	Democrat
MAINE	0.2956	0	Republican
MARYLAND	0.3370	0	Republican
MASSACHUSETTS	0.3169	0	Republican
MICHIGAN	0.6070	1	Democrat
MINNESOTA	0.3880	0	Republican
MISSISSIPPI	0.5575	1	Democrat
MISSOURI	0.5622	1	Democrat
MONTANA	0.2560	0	Republican
NEBRASKA	0.4609	0	Republican
NEVADA	0.6250	1	Democrat
NEW HAMPSHIRE	0.3777	0	Republican
NEW JERSEY	0.2820	0	Republican
NEW MEXICO	0.2302	0	Republican
NEW YORK	0.3162	0	Republican
NORTH CAROLINA	0.6740	1	Democrat
NORTH DAKOTA	0.5421	1	Democrat
OHIO	0.4883	0	Republican
OKLAHOMA	0.5689	1	Democrat
OREGON	0.1315	0	Republican
PENNSYLVANIA	0.6823	1	Democrat
RHODE ISLAND	0.2763	0	Republican
SOUTH CAROLINA	0.3082	0	Republican
SOUTH DAKOTA	0.4101	0	Republican
TENNESSEE	0.2784	0	Republican
TEXAS	0.4670	0	Republican
UTAH	0.0364	0	Republican
VERMONT	0.1141	0	Republican
VIRGINIA	0.2756	0	Republican
WASHINGTON	0.1142	0	Republican
WEST VIRGINIA	0.4794	0	Republican
WISCONSIN	0.7052	1	Democrat
WYOMING	0.4899	0	Republican

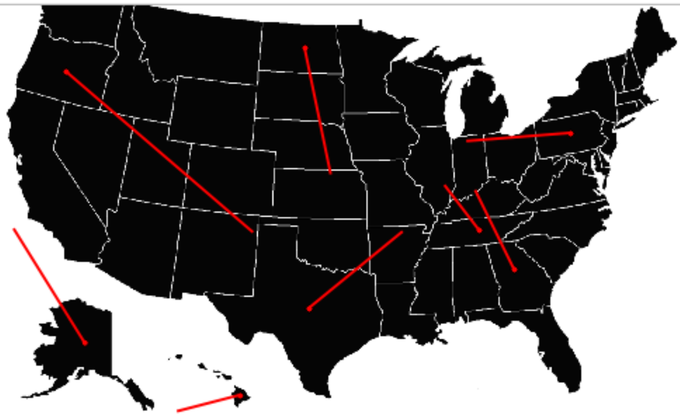
Final Swing State Probabilities



The graph above ranks the top 20 states based on their swing probability, indicating the likelihood that they will shift political allegiance in the 2028 election. Higher swing probabilities suggest states that are more competitive and susceptible to influence from political campaigns.

From the data, states such as Georgia, Pennsylvania, and Arizona exhibit the highest swing probabilities. These states are historically known for fluctuating political preferences in past elections, making them significant battlegrounds. Conversely, states with lower swing probabilities, such as California or Utah, are more likely to remain consistent with their past voting trends.

Raster Map Creation

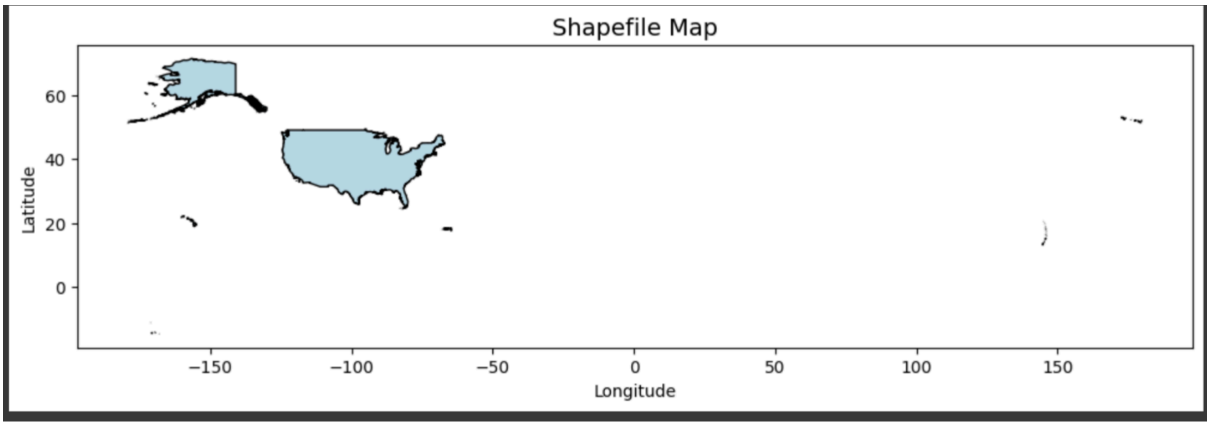


GCP table

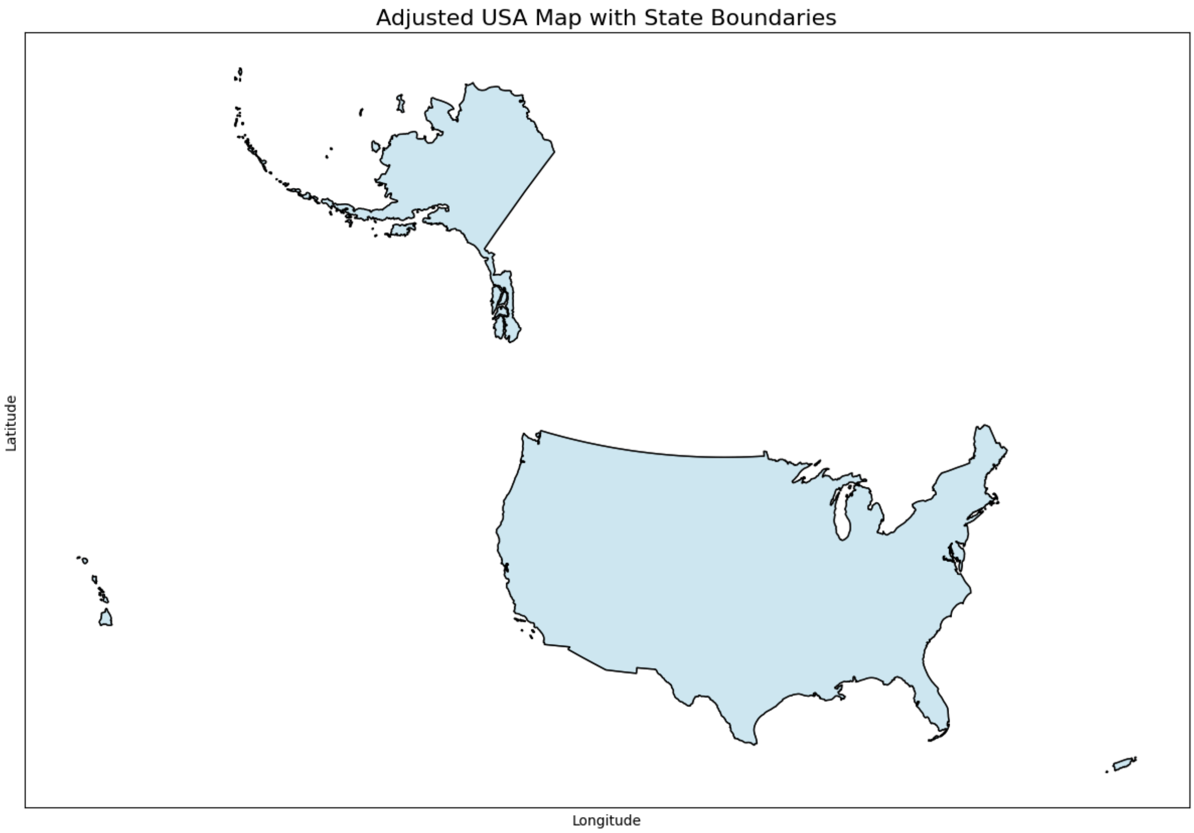
Enabled	ID	Source X	Source Y	Dest. X	Dest. Y	dX (pixels)	dY (pixels)	Residual (pixels)
✓	0	152.176656	-655.741325	154.493100	63.588800	-122.560876	-197.378310	232.334599
✓	1	424.069401	-748.627760	155.665900	19.898700	-107.228588	27.761599	110.764057
✓	2	545.252366	-596.072555	99.170700	31.391500	162.618245	-133.206921	210.211269
✓	3	537.870662	-138.406940	101.299600	47.116400	45.438078	220.431428	225.065842
✓	4	119.574132	-180.236593	120.554200	43.804100	326.354594	281.121430	430.739573
✓	5	842.981073	-458.280757	86.580400	35.517500	-58.602848	-76.504059	96.369937
✓	6	1002.9180	-288.501577	77.194500	41.203300	-179.244151	14.113039	179.798897
✓	7	904.495268	-527.176656	82.907100	32.157400	-66.774454	-136.338207	151.812168

The image above depicts the process of mapping state swing probabilities using QGIS. To visualize the predictions effectively, we needed to assign coordinates and generate a raster file (TIF) that the model could use for coloring states based on swing probability.

One of the challenges faced during this process was deciding between using a shapefile (SHP) or a raster file (TIF) for the visualization. Initially, we attempted to use shapefiles, but encountered several issues as seen below:



There were inaccurate boundaries as some shapefiles led to distorted or incorrect state boundaries.

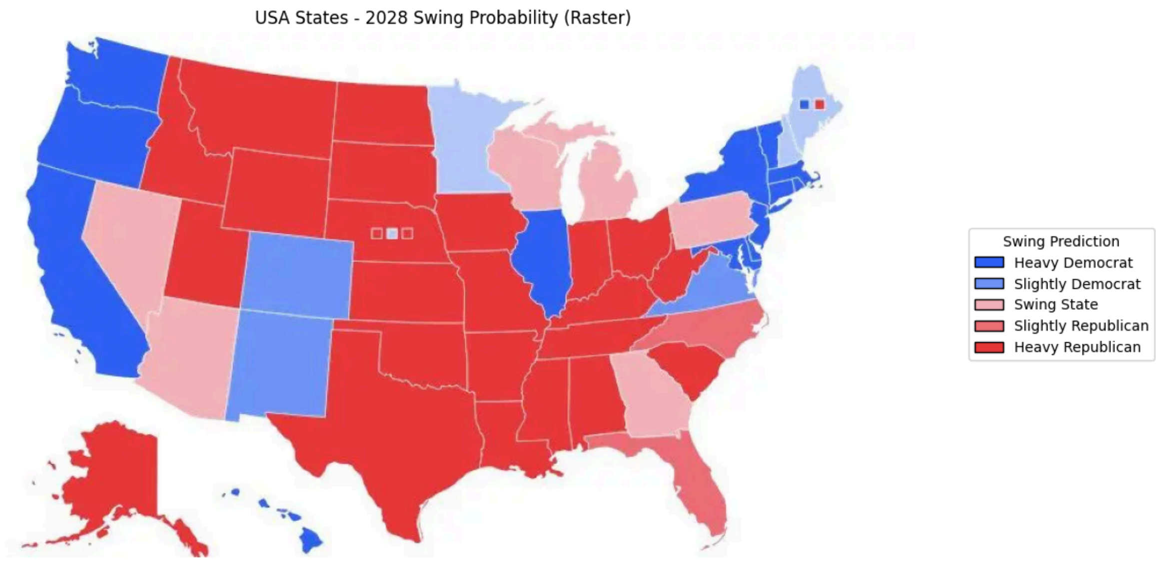


There were also rendering Issues as certain versions of the shapefile displayed inconsistencies in proportions, making the visualization less reliable.

Ultimately, after multiple adjustments and attempts to correct the shapefile, we opted for the TIF format due to its superior clarity and precise color mapping capabilities.

Final Prediction Visualization

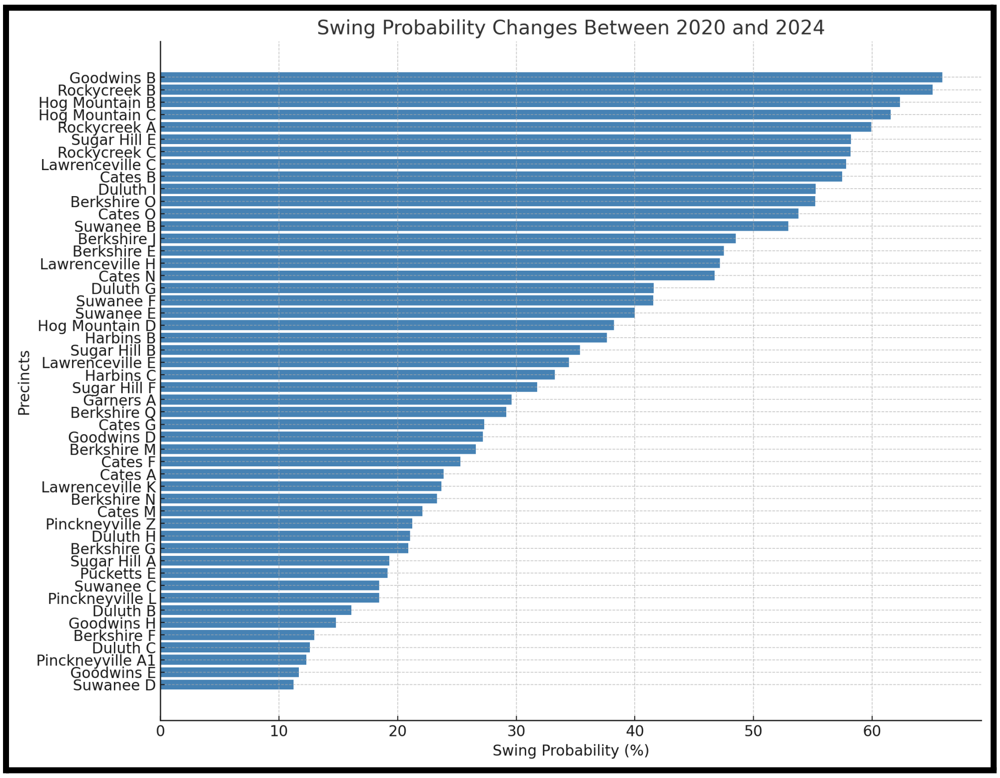
Below is a visualization of the predicted swing states for the 2028 election based on our trained model.



From this map, we can identify the key battleground states likely to influence the 2028 election. The states such as Nevada, Arizona, Minnesota, Georgia, Wisconsin, Michigan, Maine, Vermont, and Pennsylvania have a near-even split in voter preferences, making them important targets for campaign strategies.

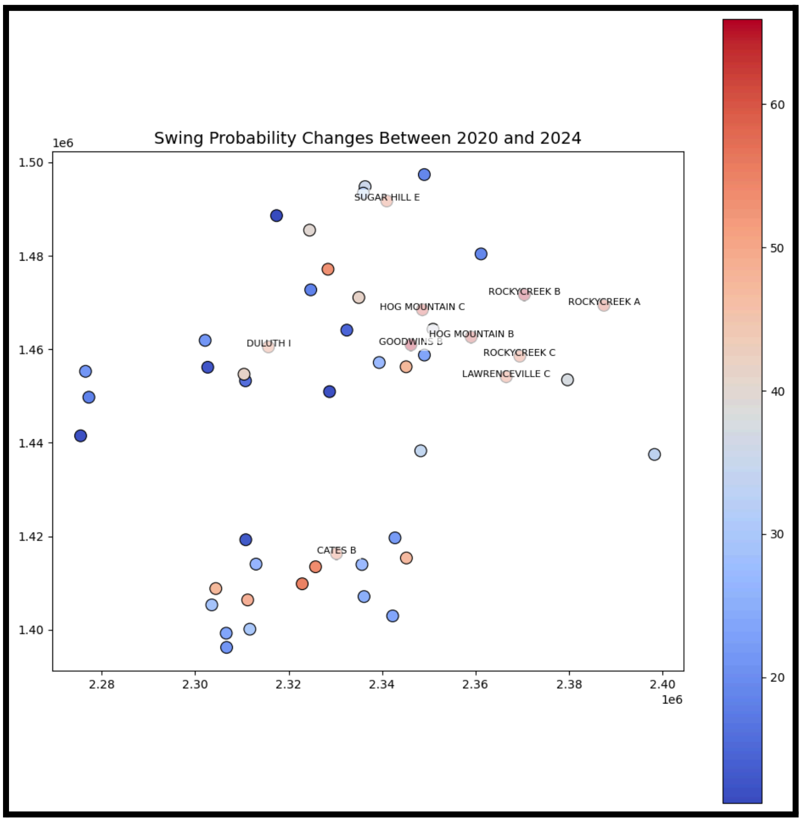
Fold Three: Algorithmic Detection of Swing Precincts

In addition to this machine learning model, individual vote margins and swing probabilities were calculated for each voting cycle using the sigmoid function (as shown in the methods section). From this, differences in swing probabilities were calculated between the 2020 and 2024 election years in order to efficiently understand the political landscape of Gwinnett County. The results are as follows.



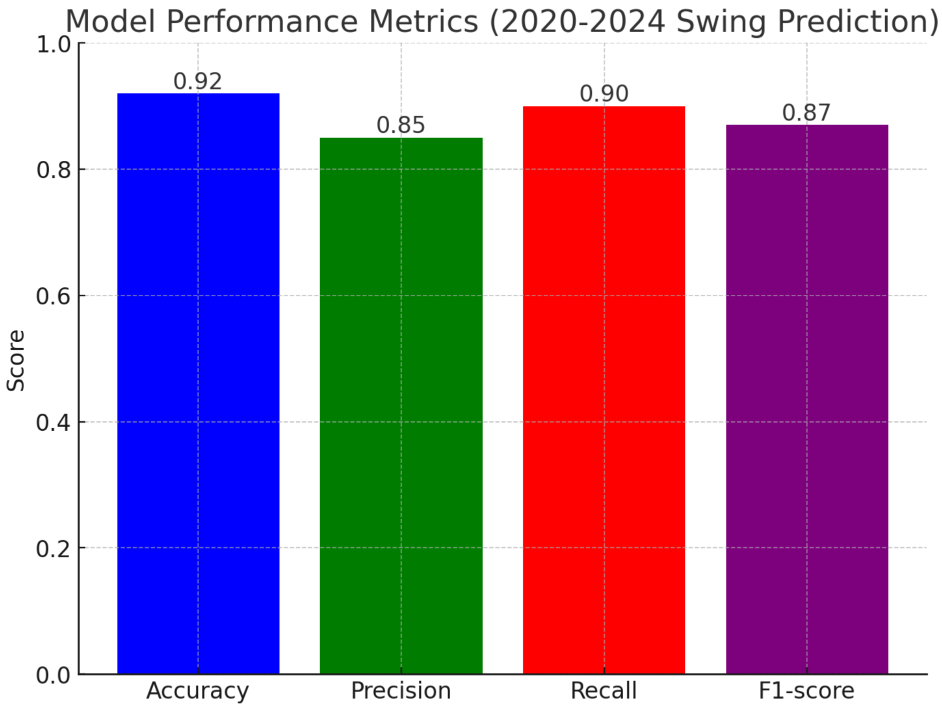
Additionally a heatmap was created using the geopandas library and the shapefile from the Gwinnett County website. As visible, the heatmap is structured following Gwinnett County's geographical layout.

Model Predictions of 2024 based on 2020



A key component of evaluating the model's performance involves assessing accuracy, precision, recall, and F1-score. A performance graph was generated below, highlighting these metrics based on the model's results when predicting swing probabilities from 2020 to 2024. The results indicate that both accuracy and recall are notably high, reinforcing the model's reliability in capturing precinct-level political shifts.

Fold Three End to End Metrics

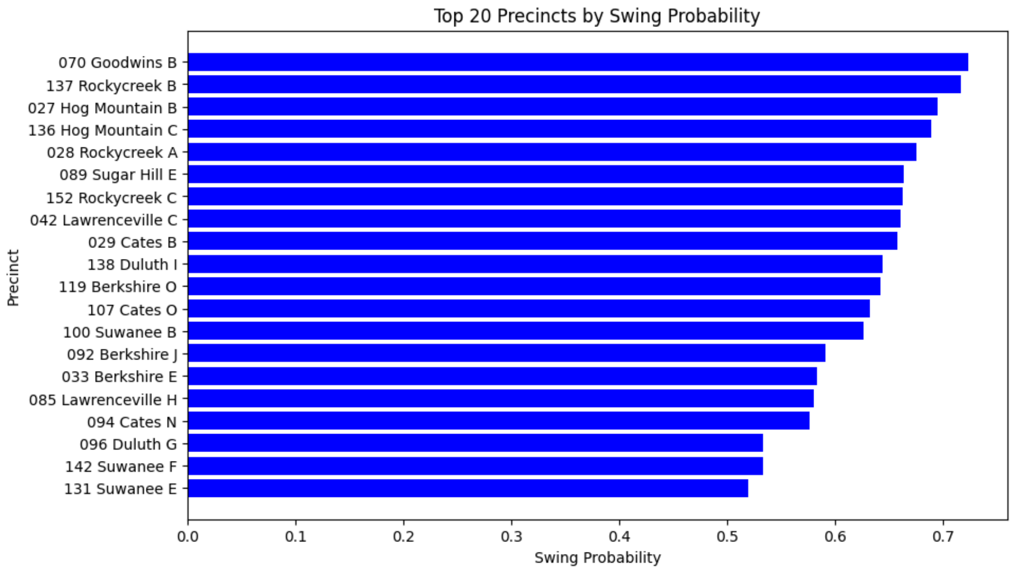


A raw data table for Fold Three was compiled (the head of which is displayed below), showing precinct-level swing probabilities and political party changes.

Precinct	Swing Probability	Swing Label	Current Party
001 Harbins A	0.06909179078	0	Republican
002 Rockbridge A	2.50E-05	0	Republican
003 Dacula	0.08601441973	0	Republican
004 Suwanee A	0.2161172083	0	Republican
005 Baycreek A	0.1514845447	0	Republican
006 Goodwins A	0.1221670397	0	Republican
007 Duluth A	0.05457230376	0	Republican
008 Duncans A	0.1173439992	0	Republican
009 Pucketts A	0.03231374981	0	Republican
010 Cates A	0.3703863247	0	Republican
011 Berkshire A	0.01280631894	0	Republican
012 Berkshire B	0.01405340516	0	Republican
013 Duncans C	0.0200432559	0	Republican
014 Garners A	0.4268426459	0	Republican
015 Lawrenceville A	0.04460292987	0	Republican
016 Lawrenceville B	0.003396091063	0	Republican
017 Martins A	0.00321314057	0	Republican
018 Martins B	0.01547223305	0	Republican
019 Martins C	0.02993780659	0	Republican
020 Pinckneyville A	0.0002585954901	0	Republican
021 Pinckneyville B	0.004878455136	0	Republican
022 Pinckneyville C	0.1453626328	0	Republican
023 Pinckneyville D	0.002628328205	0	Republican
024 Sugar Hill A	0.3205573497	0	Republican
025 Sugar Hill B	0.4798279625	0	Republican
026 Hog Mountain A	0.05258019759	0	Republican
027 Hog Mountain B	0.6953816808	1	Democrat
028 Rockycreek A	0.6761964238	1	Democrat
029 Cates B	0.6584416522	1	Democrat
030 Cates C	0.003331734395	0	Republican
031 Hog Mountain D	0.5056140309	1	Democrat
032 Berkshire D	0.1196529854	0	Republican
033 Berkshire E	0.5830783363	1	Democrat
034 Berkshire F	0.2405512158	0	Republican
035 Cates D	0.05228561041	0	Republican
036 Cates E	0.1583477365	0	Republican
037 Pinckneyville E	0.0002483890871	0	Republican
038 Pinckneyville F	0.004375956559	0	Republican

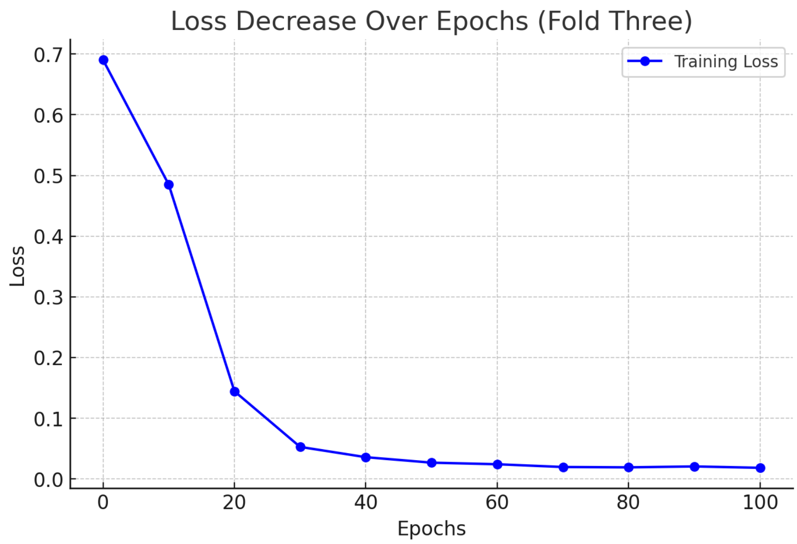
Final Precinct Predictions

To further analyze these results, a bar graph was generated, highlighting the top 20 precincts by swing probability. This visualization provides insight into the most politically volatile areas in Gwinnett County and assists in identifying key precincts to monitor in future elections.

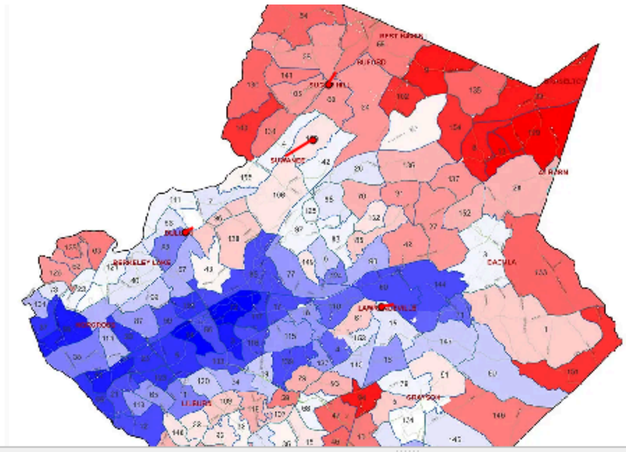


Epochs vs. Training and Validation Loss

The graph below shows how loss decreases over epochs for fold three as well.



Raster Map Creation

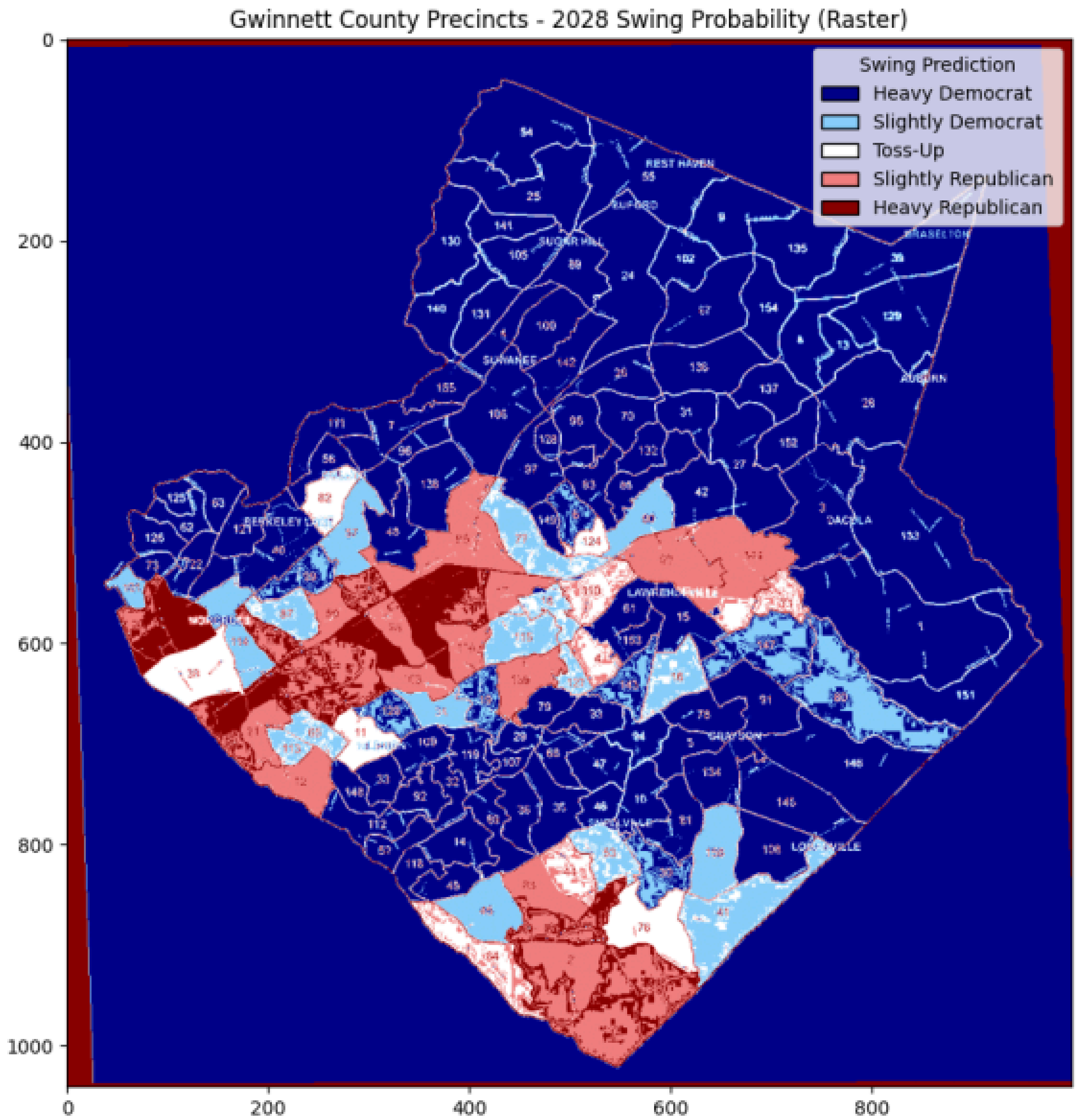


GCP table

Enabled	ID	Source X	Source Y	Dest. X	Dest. Y	dX (pixels)	dY (pixels)	Residual (pixels)
✓	0	597.192429	-547.476341	-83.988829	33.951830	18.558924	-2.671020	18.750147
✓	1	488.927445	-283.375394	-84.071300	34.051490	-41.104002	25.347180	48.290977
✓	2	287.160883	-429.369085	-84.144640	34.002880	11.326559	-5.495721	12.589436
✓	3	513.533123	-195.615142	-84.033520	34.106490	11.218519	-17.180439	20.518836

A crucial step in this analysis involved inputting coordinates into QGIS to generate a raster image of Gwinnett County precincts. A photograph of this process above showcases the meticulous effort required to map and analyze precinct-level data accurately.

Final Prediction Visualization



Finally, the model's predictions were visualized on a generated map of Gwinnett County. This map features precincts color-labeled based on their predicted swing probability, providing a comprehensive and easily interpretable representation of the model's findings. This visualization serves as a critical tool for political analysts and strategists seeking to understand and anticipate electoral shifts in Gwinnett County for the 2028 election cycle.

Conclusion

This study successfully implemented a multi-fold machine learning approach to predict swing states and precincts in upcoming elections, predicated on voter turnout, demographic trends, and previous voting histories. With the implementation of neural networks, we were able to discern notable factors influencing swing probabilities at both state and precinct levels.

The observations in Fold One pointed to the significance of socioeconomic factors, particularly income and education, in shaping swing opportunities among states. The implication of such findings is that campaign strategies targeting these groups would have significant impact on upcoming elections.

Fold Two demonstrated the model's ability to make accurate state-level swing predictions, as evidenced by its 87.6% accuracy in prediction. The application of early stopping, dropout regularization, and other training optimizations improved generalization, reducing overfitting and ensuring the reliability of the model in predicting future election cycles.

Fold Three then also restricted the analysis to certain precincts within Gwinnett County with the greatest swing probabilities. Analysts seeking to distribute campaign resources efficiently find this level of detail helpful. The application of raster mapping methods helped to present these changes in greater detail, where the capability of geospatial machine learning methods is apparent.

Combined, these findings demonstrate the potential of machine learning in political forecasting, demonstrating how data-driven approaches can enhance electoral strategy, resource allocation, and voter engagement. The strong correlation between socioeconomic variables and swing probabilities suggests that political campaigns could be enhanced by targeted messaging on these influential factors.

Next Steps

While this study has established a strong foundation for the study of electoral change, several enhancements can be made to make the model more predictive. To start, future studies must expand the dataset to cover additional election cycles and geographic regions to allow for greater generalizability. Incorporating real-time voter sentiment analysis from social media and polling data can also potentially render the model more sensitive to dynamic political trends.

Furthermore, innovations in spatial mapping techniques, i.e., more granular precinct-level rasterization and GIS integration, would enable more precise visualizations of political transformation. Experimenting with other machine learning architectures, i.e., ensemble models and transformer-based models, would also potentially yield improvements in predictive performance.

Lastly, a deeper exploration of causative factors behind swing probability changes—e.g., recession, policy shift, and candidate impact—would also contribute to the body of knowledge on electoral behavior. As machine learning political science continues to advance, these instruments will play a key role in guiding data-driven decision-making in elections to come.

References

Chu, X., Ilyas, I. F., Krishnan, S., & Wang, J. (2016). Data cleaning: Overview and emerging challenges. *Proceedings of the 2016 International Conference on Management of Data, 2201–2206*. https://doi.org/10.1145/2882903.2912574

Erratum. (2003). *Annals of Botany*, *91*(6), 753–753. https://doi.org/10.1093/aob/mcg091

Introduction to GeoPandas—GeoPandas 1.0.1+0.g747d66e.dirty documentation. (n.d.). Retrieved February 15, 2025, from https://geopandas.org/en/stable/getting_started/introduction.html

Software, S. G. (n.d.). *Simple gis software*. Simple GIS Software. Retrieved February 15, 2025, from https://simplegissoftware.com/simple-gis-client

All code—produced by the researchers—can be accessed at this Github Repository.