Model Evaluation

GOV 1347 Lab: Week X

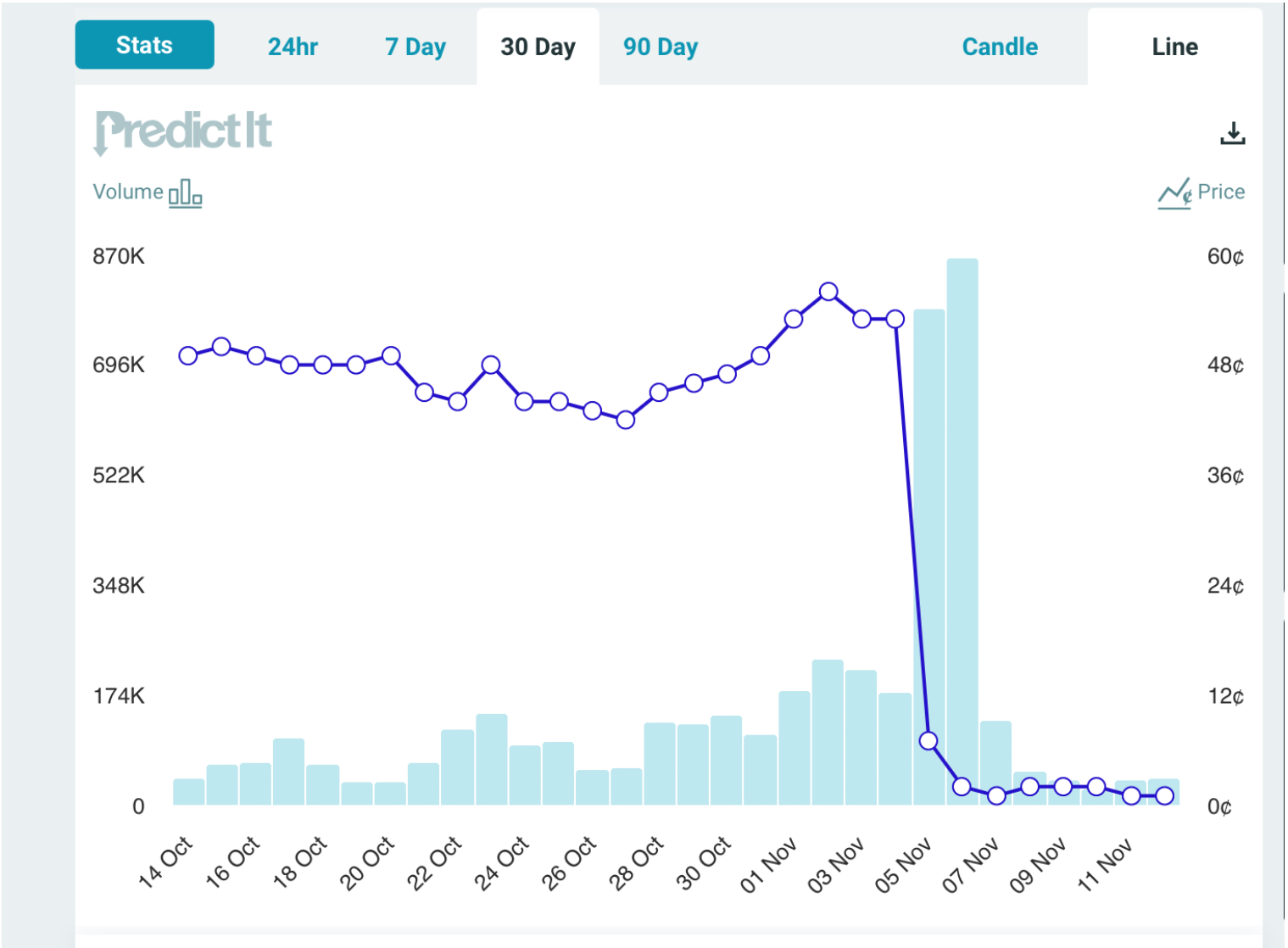
Matthew E. Dardet

Harvard University

November 13, 2024

A Penny for Your Thoughts?

I Will Not Pay You Any Pennies Because I Lost All of My Money



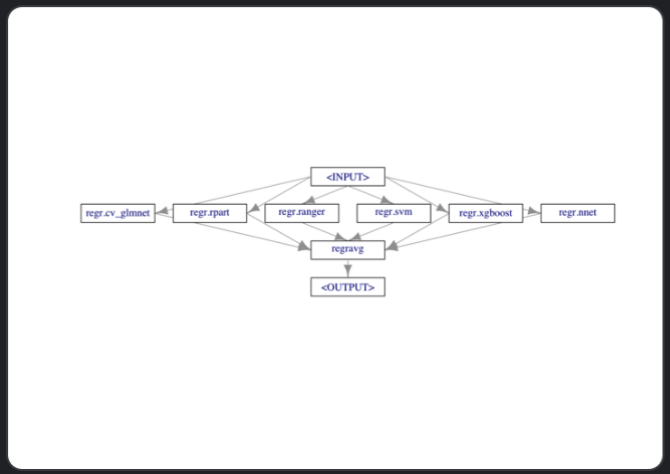
I Will Not Pay You Any Pennies Because I Lost All of My Money

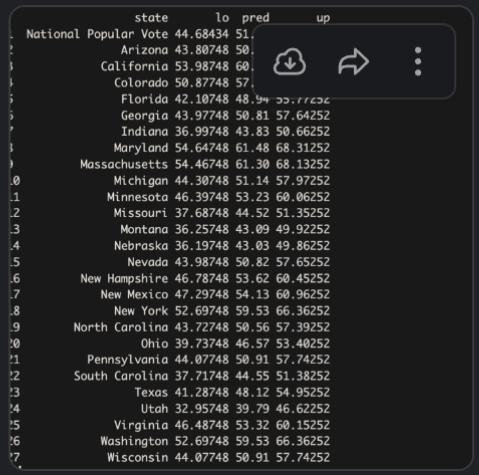
Alan Lichtman, and Alan Abramowitz? Or shall I beclown myself with erroneous predictions, thereby tarnishing my reputation and credibility amongst my students, colleagues, friends, and family? Only time shall tell...

Thank you to everyone for all the hard work so far this semester, and good luck!

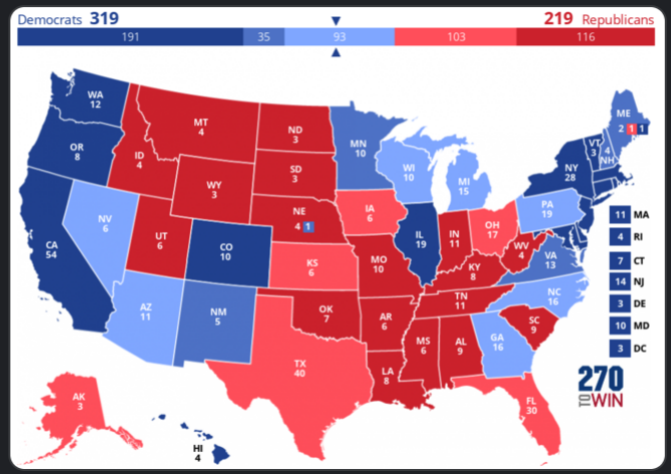
-Matthew (edited)

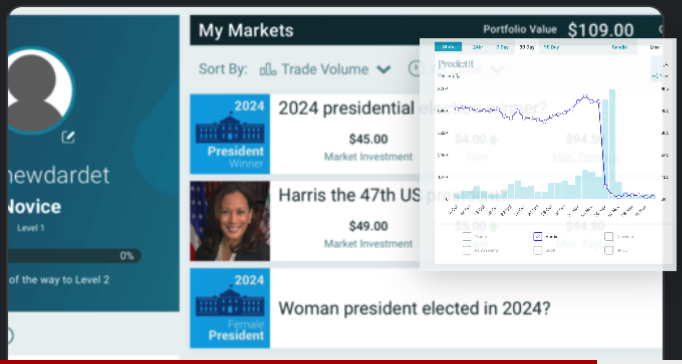
4 files





state	lo	pred	up
National Popular Vote	44.68434	51.	
Arizona	43.80748	50.	
California	53.98748	60.	
Colorado	50.87748	57.	
Florida	42.10748	48.94	55.77252
Georgia	43.97748	50.81	57.64252
Indiana	36.99748	43.83	50.66252
Maryland	54.64748	61.48	68.31252
Massachusetts	54.46748	61.30	68.13252
Michigan	44.30748	51.14	57.97252
Minnesota	46.39748	53.23	60.06252
Missouri	37.68748	44.52	51.35252
Montana	36.25748	43.09	49.92252
Nebraska	36.19748	43.03	49.86252
Nevada	43.98748	50.82	57.65252
New Hampshire	46.78748	53.62	60.45252
New Mexico	47.29748	54.13	60.96252
New York	52.69748	59.53	66.36252
North Carolina	43.72748	50.56	57.39252
Ohio	39.73748	46.57	53.40252
Pennsylvania	44.07748	50.91	57.74252
South Carolina	37.71748	44.55	51.38252
Texas	41.28748	48.12	54.95252
Utah	32.95748	39.79	46.62252
Virginia	46.48748	53.32	60.15252
Washington	52.69748	59.53	66.36252
Wisconsin	44.07748	50.91	57.74252





I Will Not Pay You Any Pennies Because I Lost All of My Money

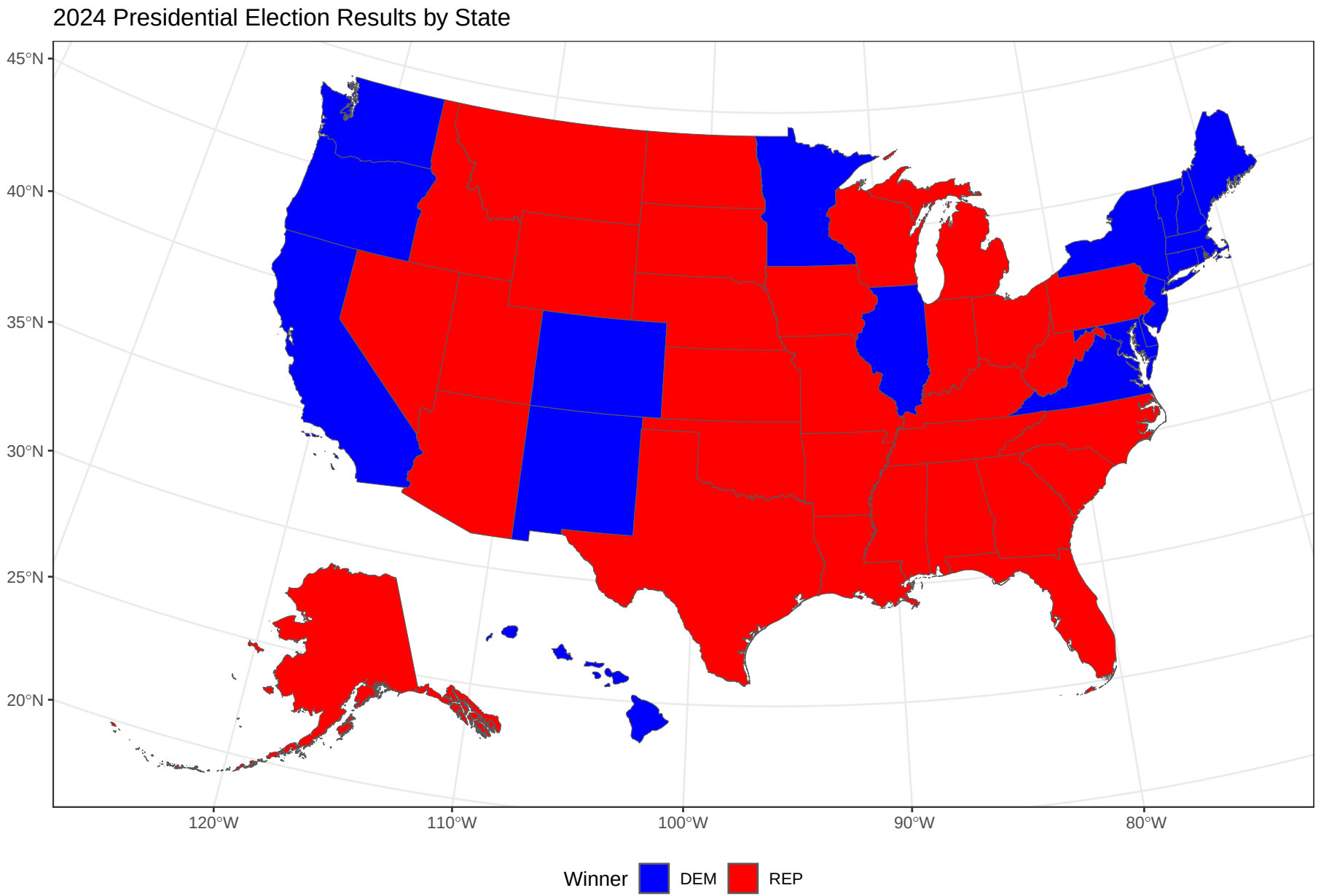
Agenda

- Visualizing Results of the 2024 Election
- Assessment of Polling Errors
- Model Evaluation

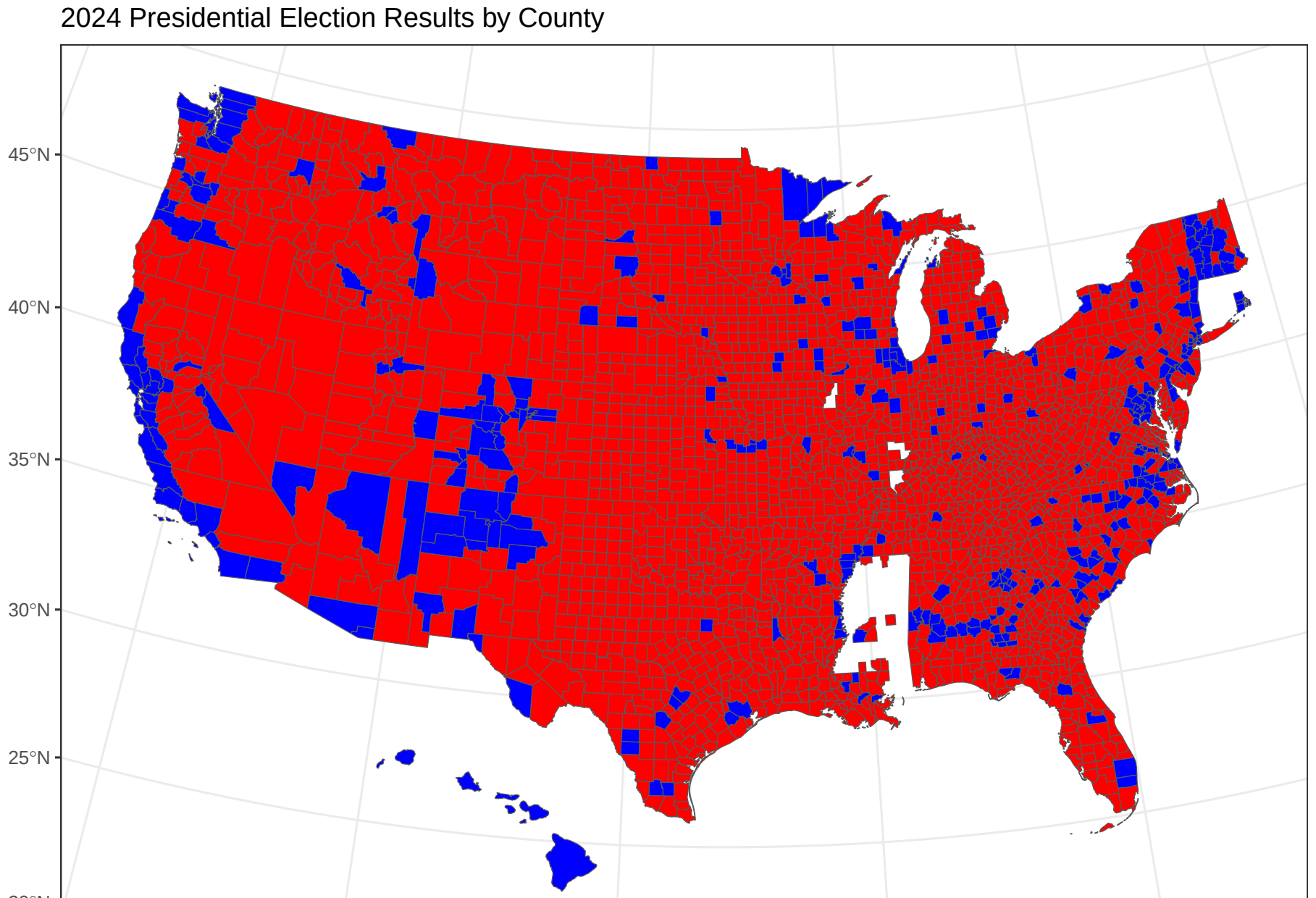
Section 1

Visualizing Results of the 2024 Election

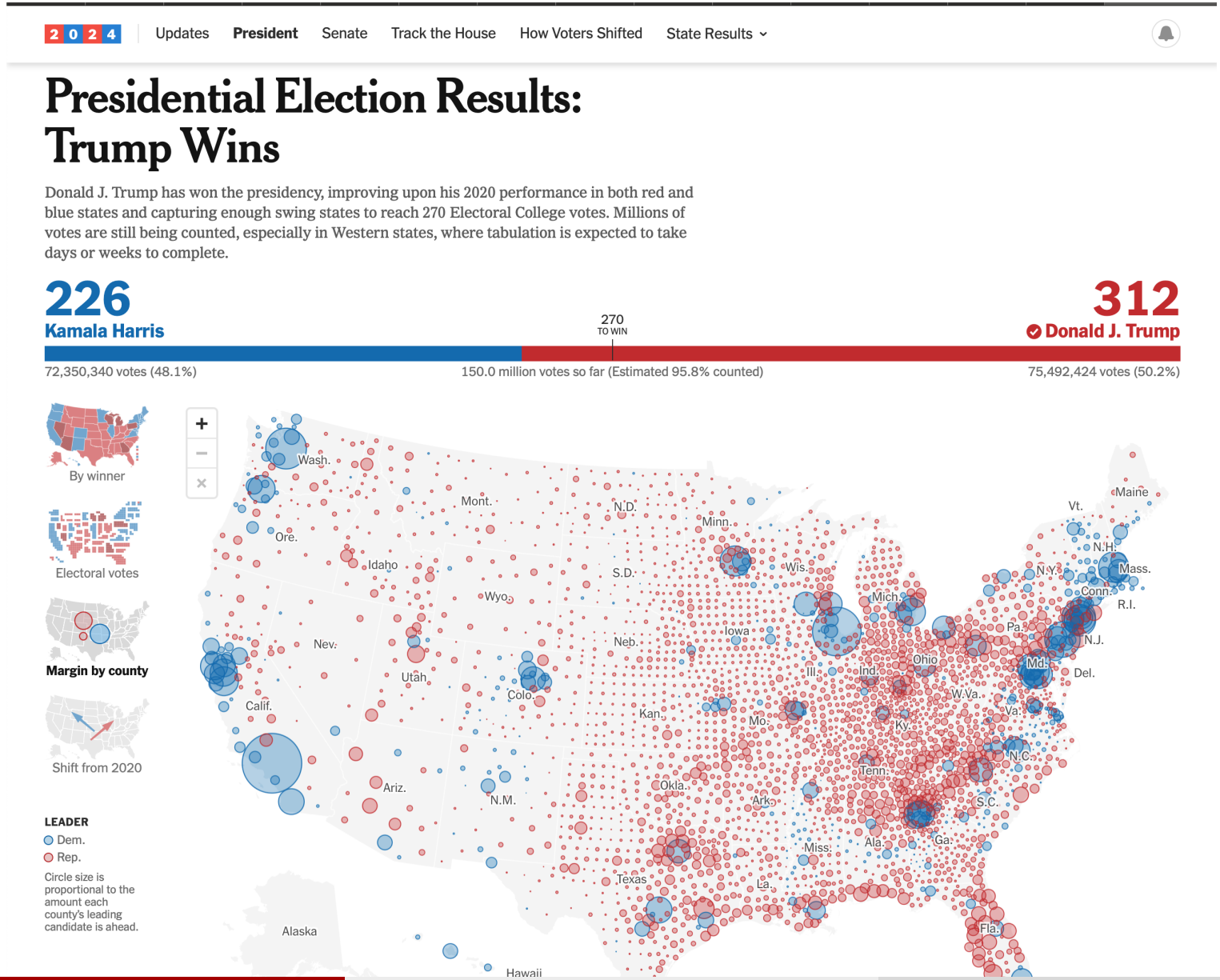
State-Level Winners



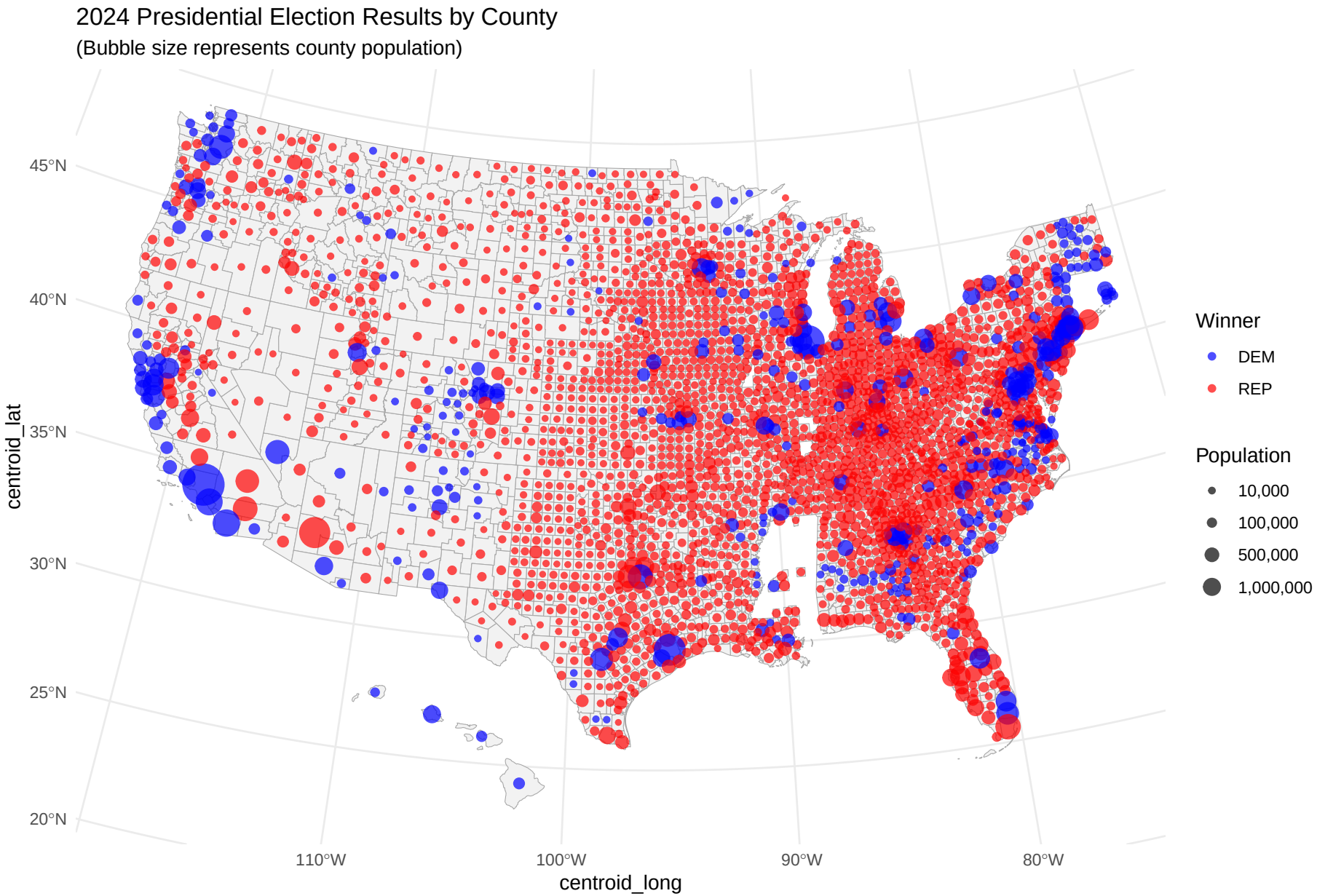
County-Level Winners



Population-Weighted County-Level Results



Population-Weighted County-Level Results



County-Level Vote Shifts from 2020 to 2024

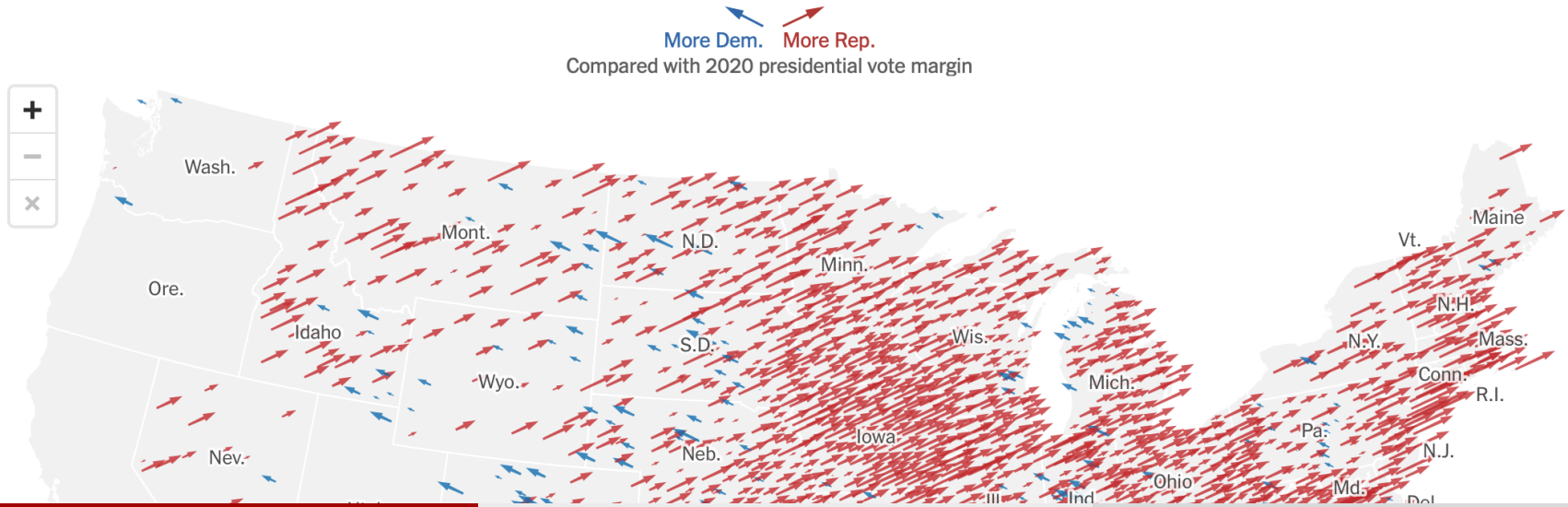
Early Results Show a Red Shift Across the U.S.

By Matthew Bloch, Keith Collins, Robert Gebeloff, Marco Hernandez, Malika Khurana and Zach Levitt

Nov. 6, 2024

Share full article

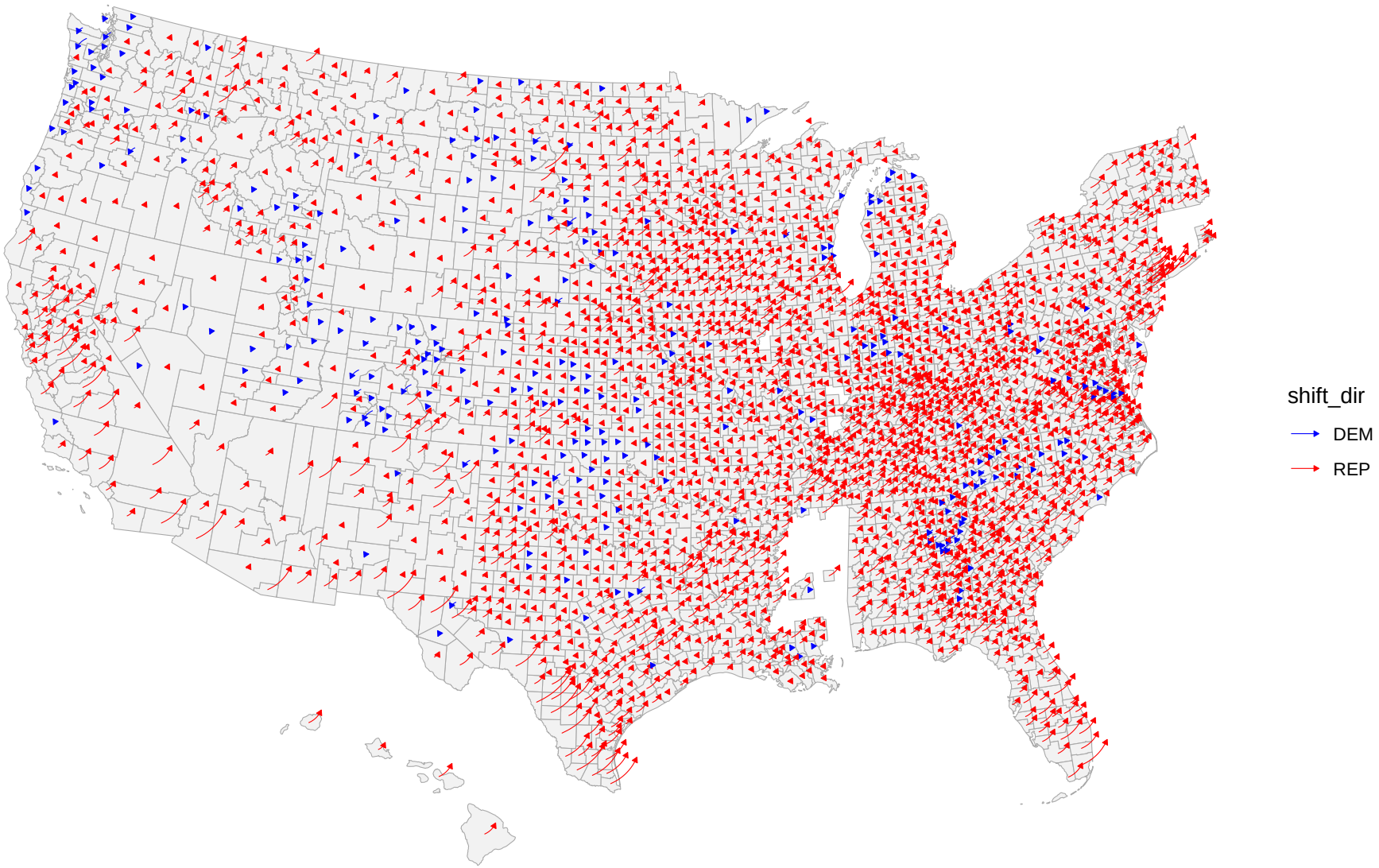
Of the counties with nearly complete results, more than 90 percent shifted in favor of former President Donald J. Trump in the 2024 presidential election, according to a Times analysis of election results reported as of early Wednesday.



County-Level Vote Shifts from 2020 to 2024

Presidential Voting Shifts by County Across the US

Democratic vs. Republican Gains

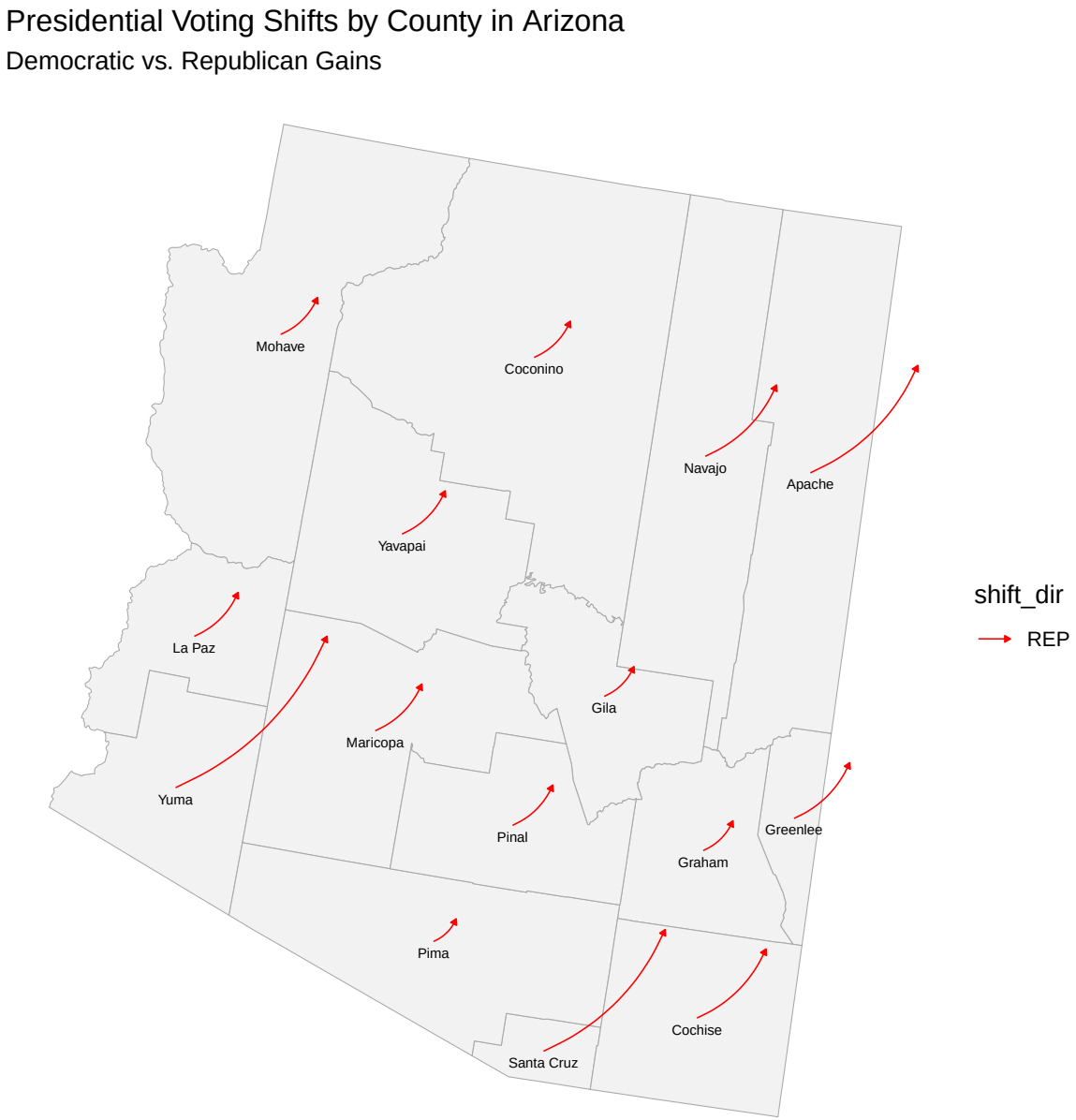


County-Level Vote Shifts from 2020 to 2024

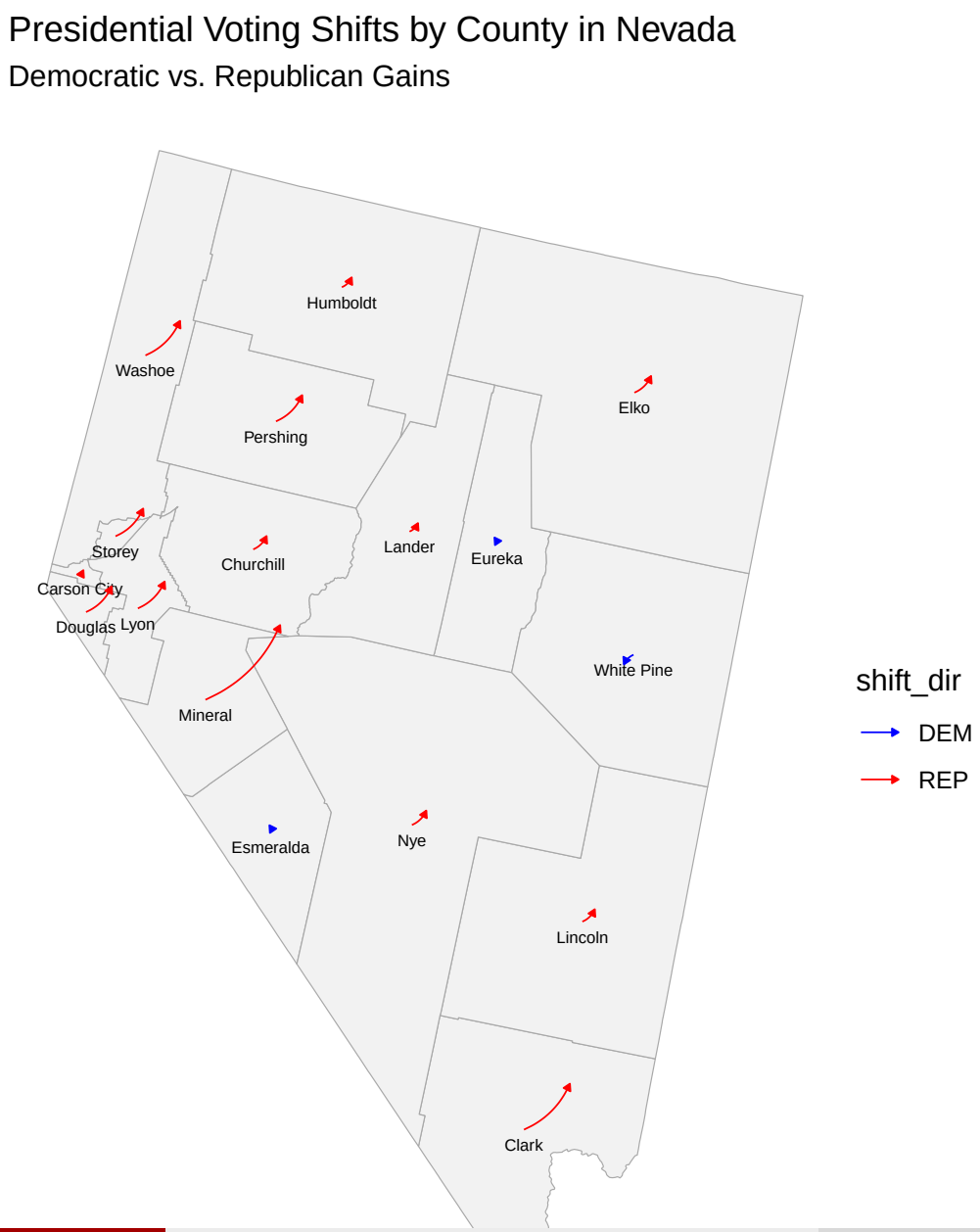
Presidential Voting Shifts by County in Pennsylvania

Democratic vs. Republican Gains

County-Level Vote Shifts from 2020 to 2024



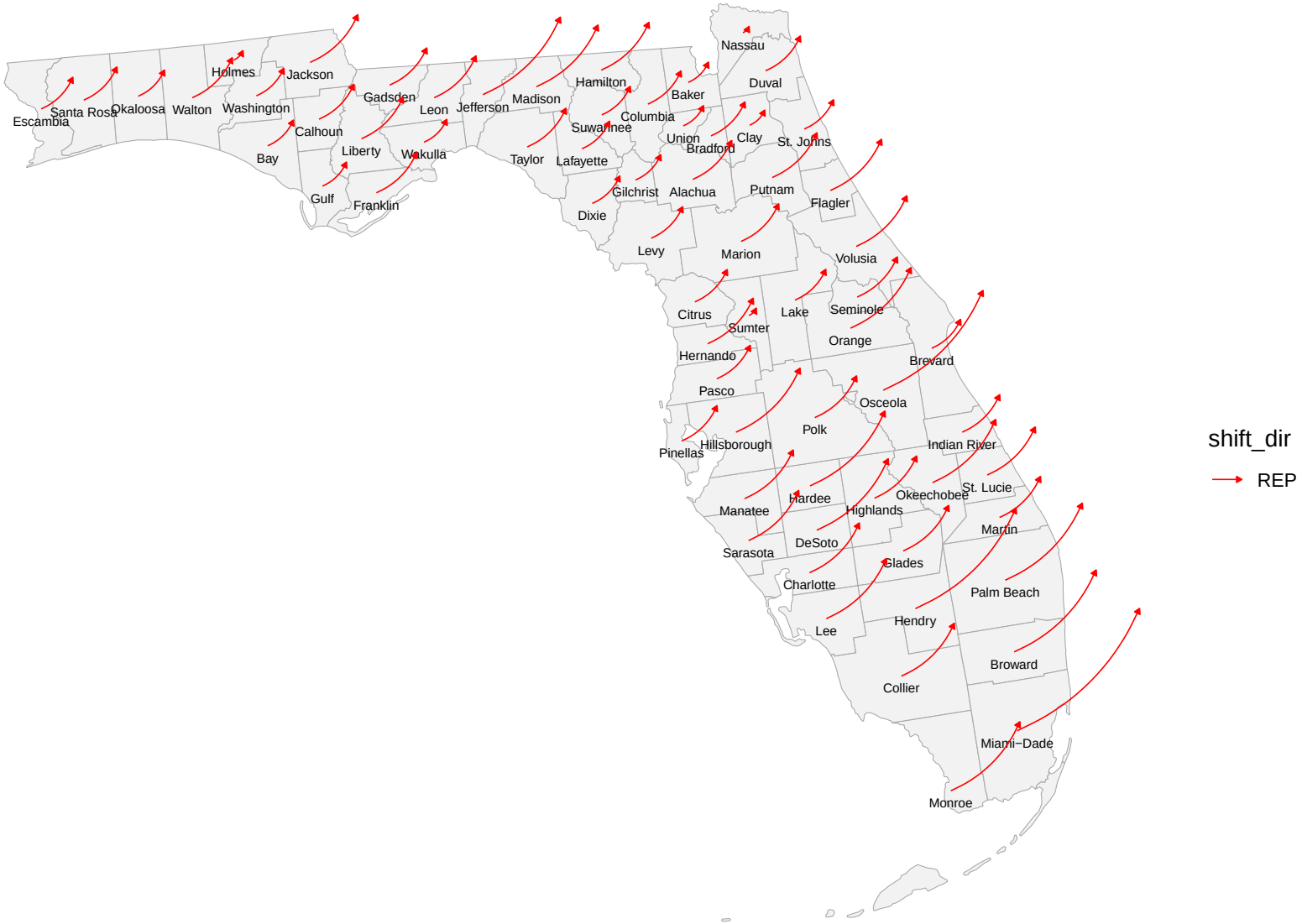
County-Level Vote Shifts from 2020 to 2024



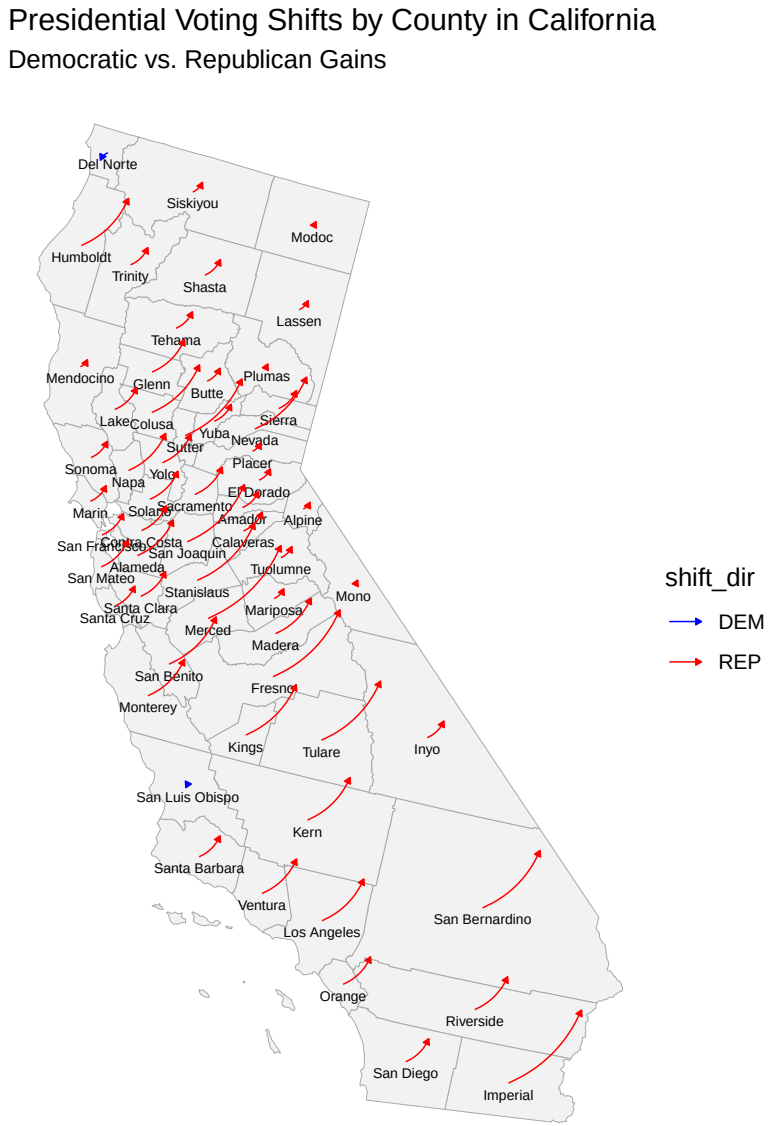
County-Level Vote Shifts from 2020 to 2024

Presidential Voting Shifts by County in Florida

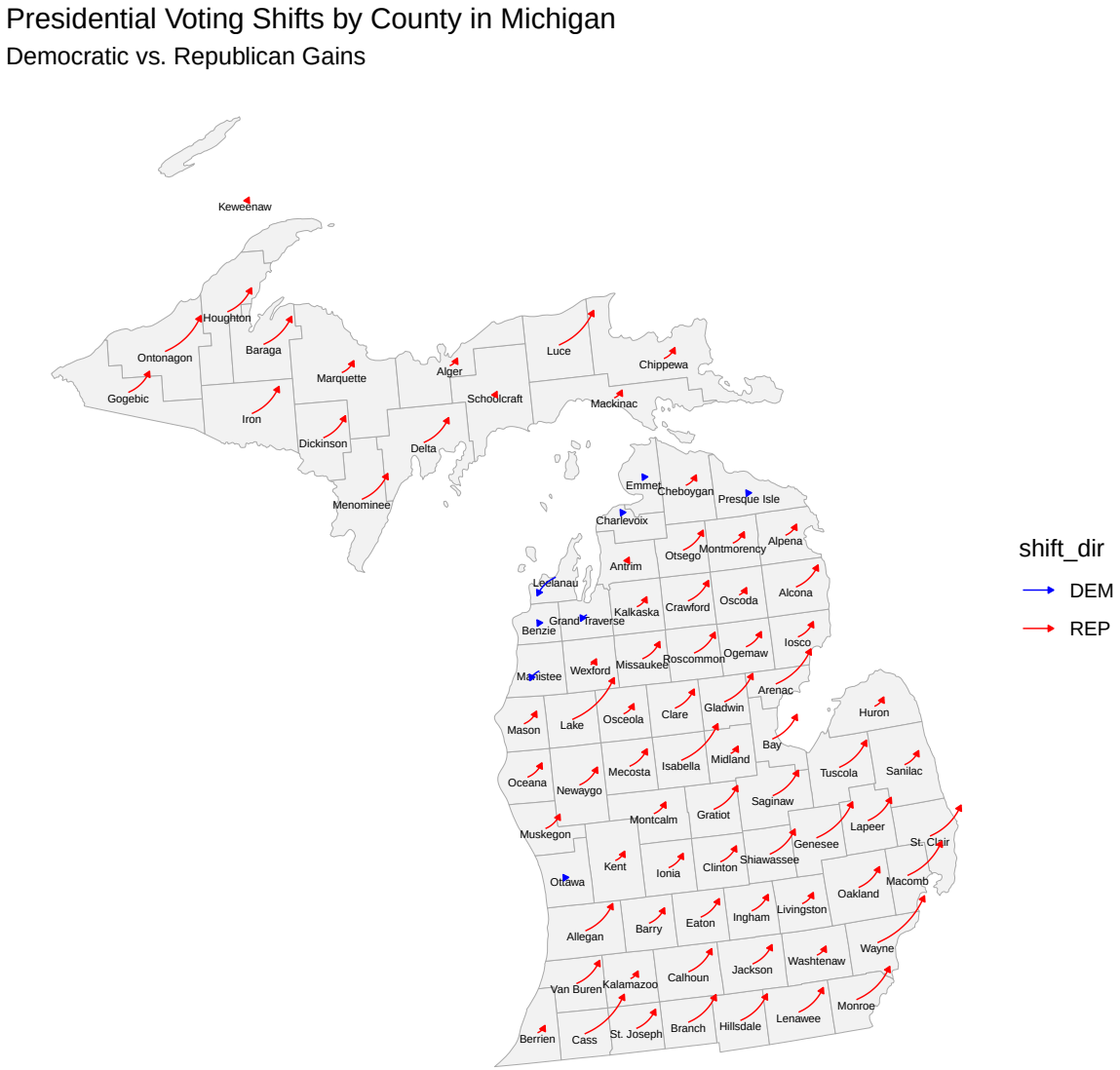
Democratic vs. Republican Gains



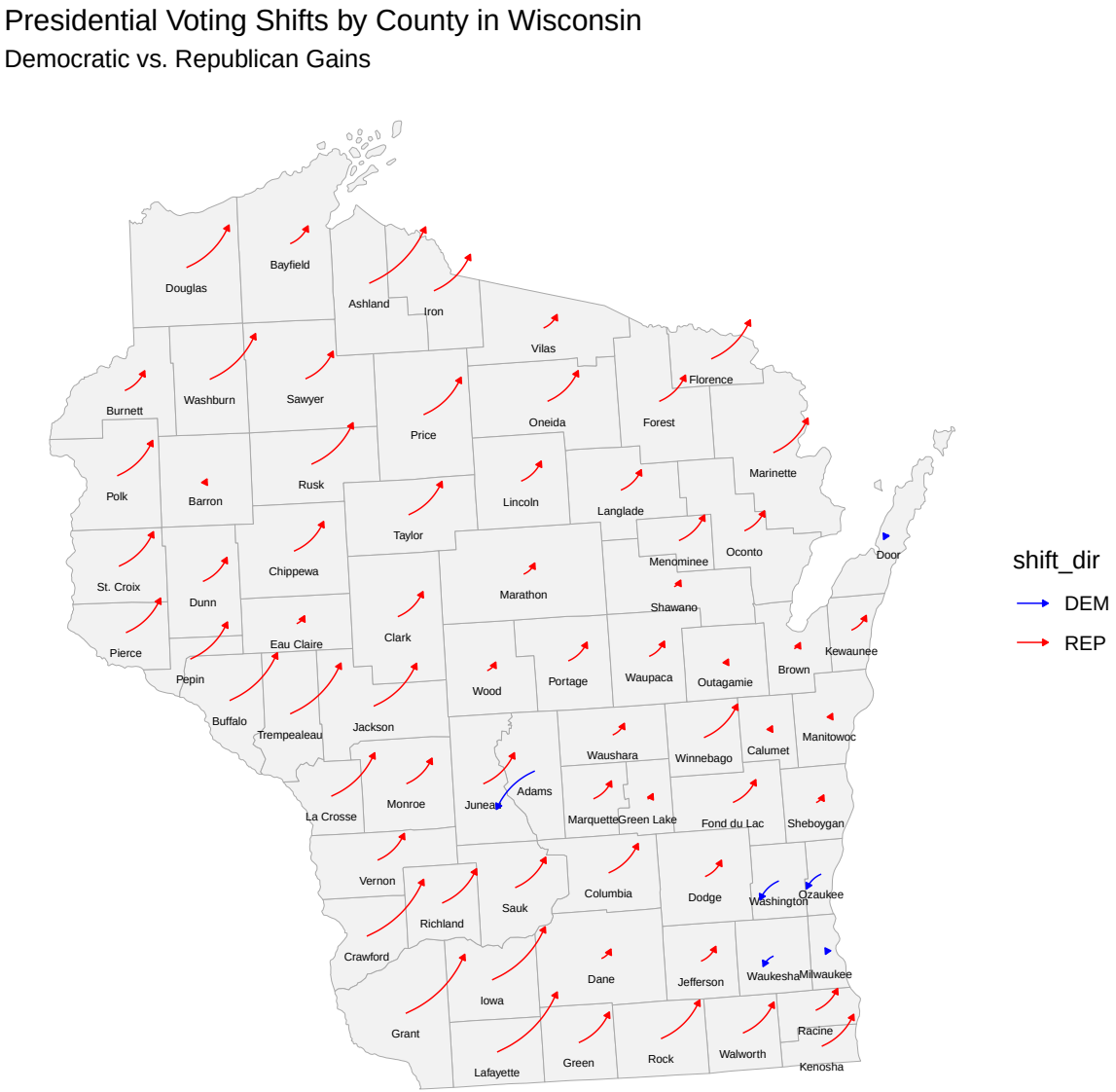
County-Level Vote Shifts from 2020 to 2024



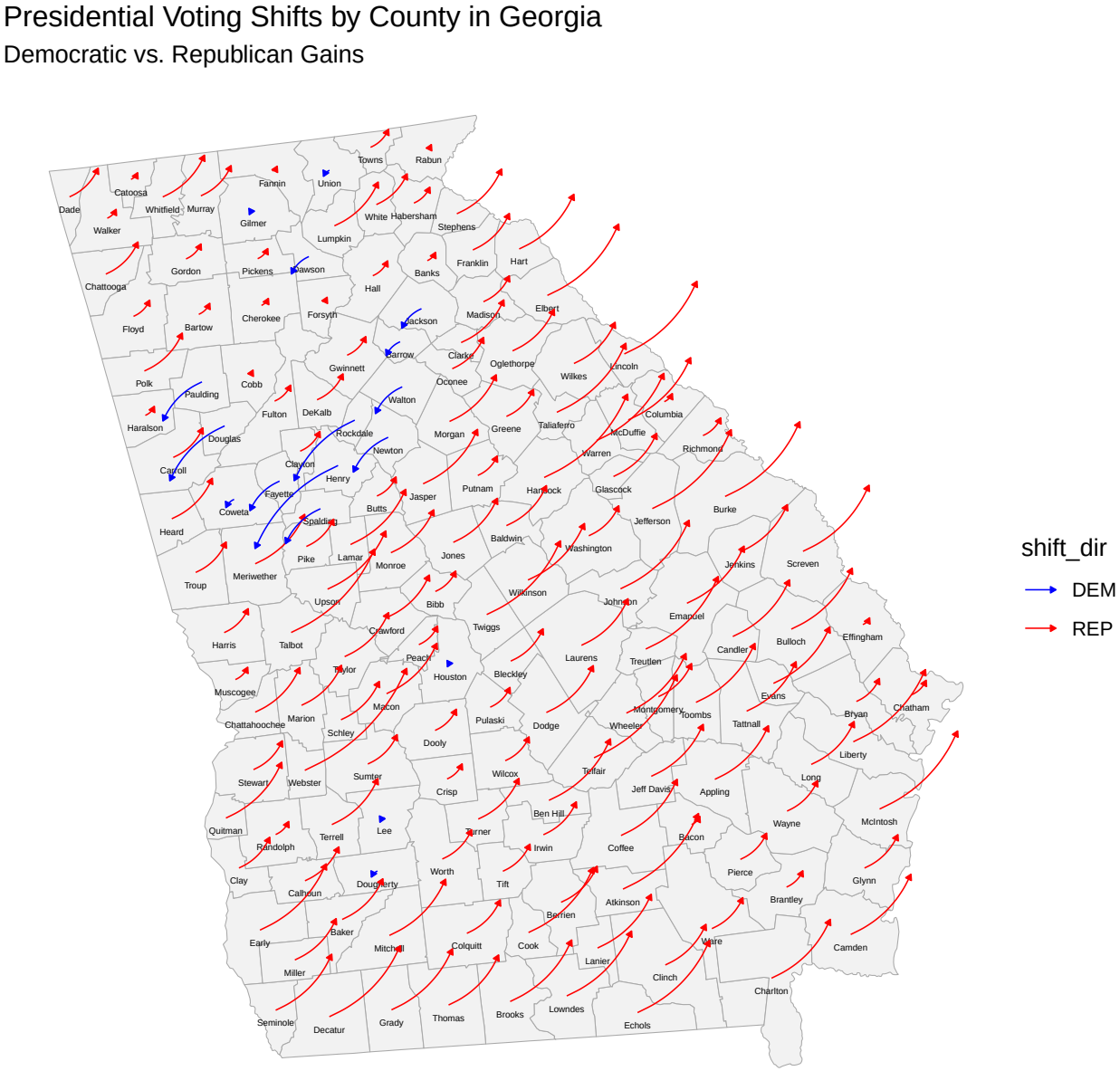
County-Level Vote Shifts from 2020 to 2024



County-Level Vote Shifts from 2020 to 2024



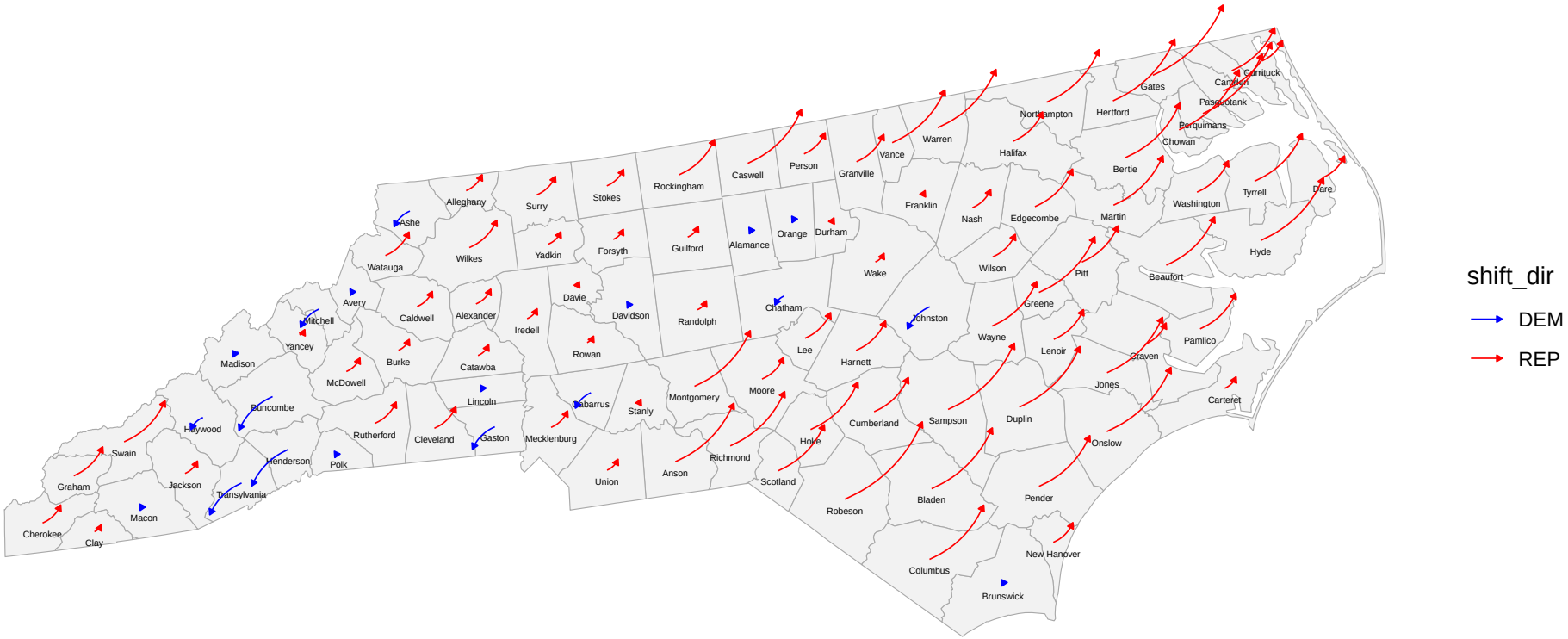
County-Level Vote Shifts from 2020 to 2024



County-Level Vote Shifts from 2020 to 2024

Presidential Voting Shifts by County in North Carolina

Democratic vs. Republican Gains



Section 2

Assessment of Polling Errors

Polling Error Counterfactuals

- In 2016 and 2020, the polls underestimated Trump vote by about 5 points.

Polling Error Counterfactuals

- In 2016 and 2020, the polls underestimated Trump vote by about 5 points.
- In 2022, the polls overestimated GOP vote in key races but were about 1 point biased overall in favor of Democrats.

Polling Error Counterfactuals

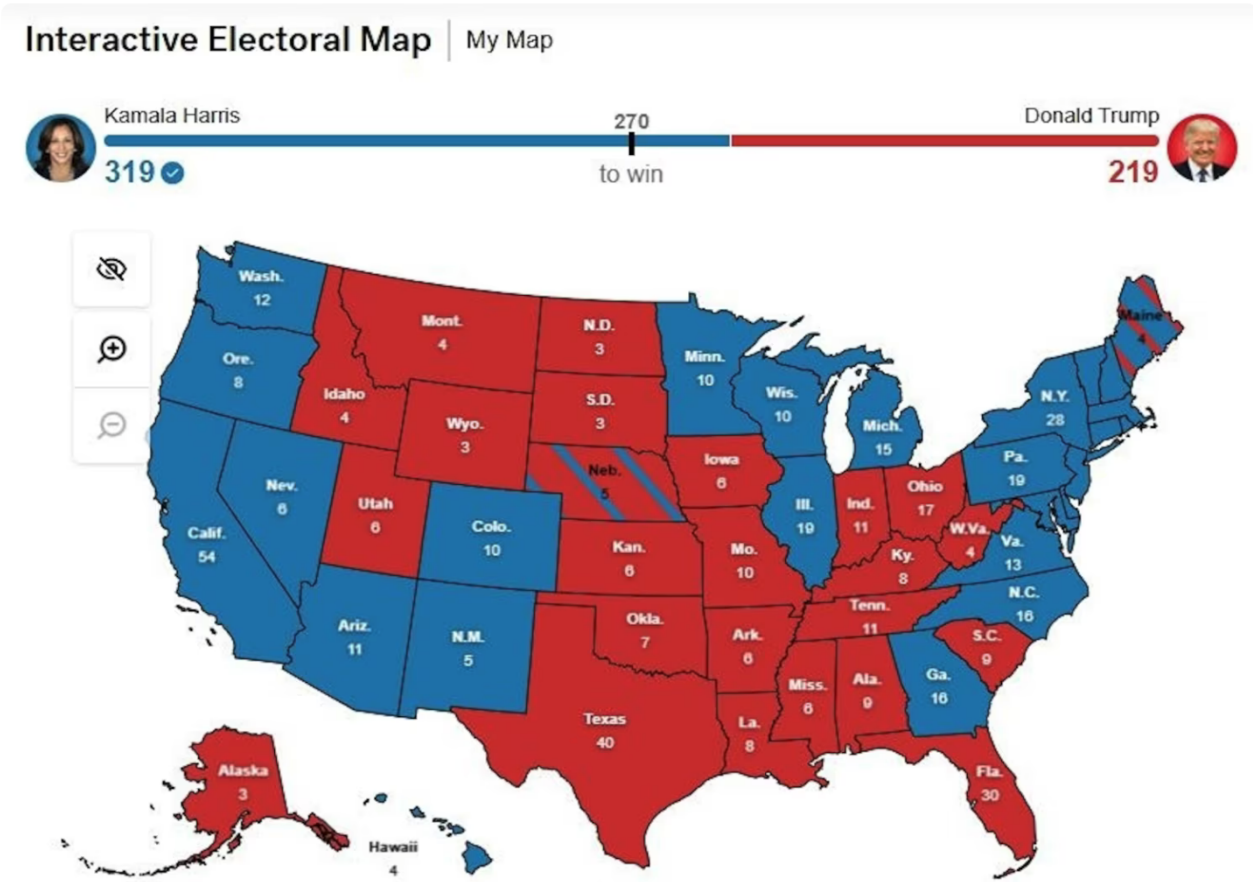
- In 2016 and 2020, the polls underestimated Trump vote by about 5 points.
- In 2022, the polls overestimated GOP vote in key races but were about 1 point biased overall in favor of Democrats.
- The night before the election day, when poll average from *FiveThirtyEight* was 48.1% Harris and 46.8% Trump.

Polling Error Counterfactuals

- In 2016 and 2020, the polls underestimated Trump vote by about 5 points.
- In 2022, the polls overestimated GOP vote in key races but were about 1 point biased overall in favor of Democrats.
- The night before the election day, when poll average from *FiveThirtyEight* was 48.1% Harris and 46.8% Trump.
- So, of course, there were two possibilities:

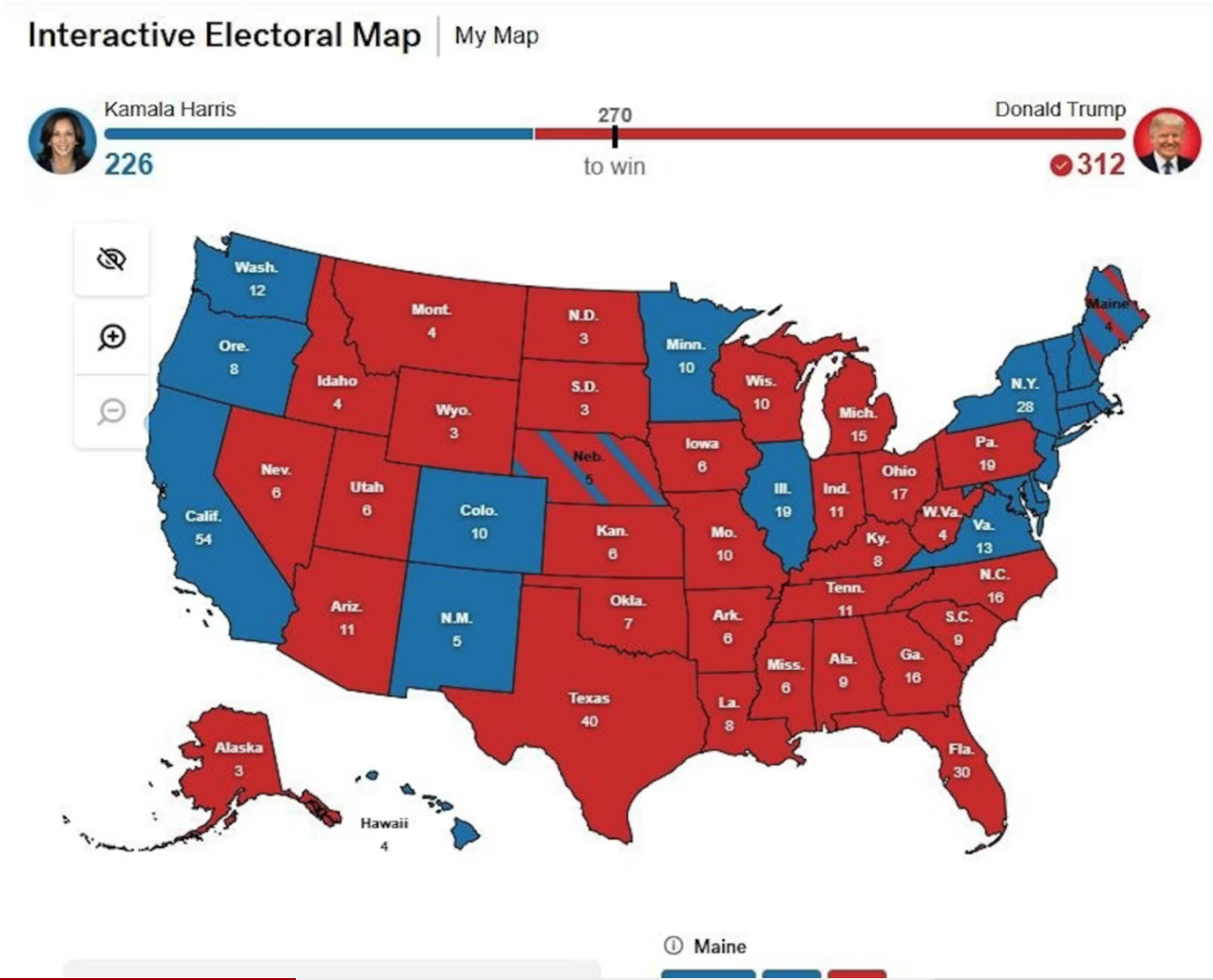
Polling Error Counterfactuals

And that's why we've been saying the race isn't necessarily going to be close just because the polls are. Trump and Harris, our model says, are both a normal polling error away from an Electoral College blowout. If we shift the polls by 4 points toward Harris, she would win the election with 319 Electoral College votes:



Polling Error Counterfactuals

Meanwhile, Trump would win with [312 electoral votes] if the polls underestimate him by the same amount:

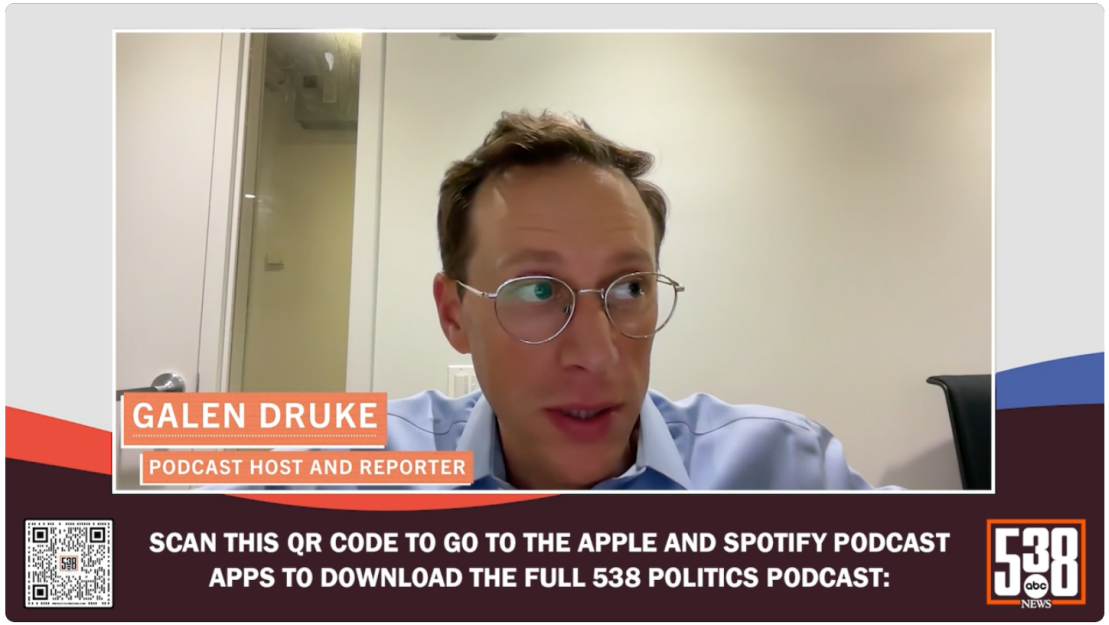


Polling Error: What Actually Happened

2024 polls were accurate but still underestimated Trump

A modest error in his favor was enough to sweep the swing states.

By G. Elliott Morris 538
November 8, 2024, 4:17 PM



How Trump won over voters

The 538 team discusses how Donald Trump won the election despite being disliked by a majority of Americans.

Here at 538, we think a big part of our jobs during election season is to explore and explain how much trust you should put in all the people telling you who's going to win. More than anyone else, given the amount of data they produce and the press and

Polling Error: What Actually Happened

- Polling error yet again underestimated Trump vote, and bigly (by about 3 points)!

Polling Error: What Actually Happened

- Polling error yet again underestimated Trump vote, and bigly (by about 3 points)!
- *FiveThirtyEight*'s fancy weighted poll averages were worse than unweighted averages from *RealClearPolitics*.

Polling Error: What Actually Happened

- Polling error yet again underestimated Trump vote, and bigly (by about 3 points)!
- *FiveThirtyEight*'s fancy weighted poll averages were worse than unweighted averages from *RealClearPolitics*.
- Open methodological and philosophical questions/conjectures from me:
  1. Does including lots of polling data diminish the quality of forecasts? How well would my model have done if I used economics and presidential approval rather than generic ballot polls and polling averages?
  2. Can statistical significance in regression be misleading as to the value of including polls/other variables in prediction models?
  3. What is the value of polls? Can they even be used for prediction, or rather for post-hoc explanation? Should these cheap, low-quality polls even be conducted?

Section 3

Model Evaluation

Mean Squared Error (MSE)

Definition:
Mean Squared Error (MSE) measures the average of the squares of the errors. It is calculated by averaging the squared differences between the predicted values and the actual values.

Formula:

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- y_i: Actual value
- $\hat{y}_i$: Predicted value
- n: Number of observations
- Penalizes large errors more due to the square term.

Root Mean Squared Error (RMSE)

Definition:
Root Mean Squared Error (RMSE) is the square root of the MSE. It is useful as it brings the error value back to the same unit as the original data.

Formula:

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

- RMSE provides a sense of the error in the same units as the response variable.

Mean Absolute Error (MAE)

Definition:
Mean Absolute Error (MAE) measures the average of the absolute errors. It calculates the average of the absolute differences between the predicted values and the actual values.

Formula:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|$$

- MAE gives an intuitive sense of the average error magnitude. It is less sensitive to outliers than MSE and RMSE.

Bias

Definition:
Bias measures the average difference between the predicted values and the actual values. It indicates whether the model systematically overpredicts or underpredicts.

Formula:

$$\text{Bias} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)$$

- y_i: Actual value
- $\hat{y}_i$: Predicted value
- A positive bias means the model tends to overpredict, while a negative bias means it underpredicts.

Bias, MSE, RMSE, and MAE

Take 15 minutes to evaluate your models across each of these metrics—Bias, MSE, RMSE, and MAE—for both the National Popular Vote (NPV) and Electoral College (EC) prediction.

Confusion Matrix and Prediction Accuracy

Confusion Matrix:
A confusion matrix is a table that summarizes the performance of a classification model by showing the counts of true positive, true negative, false positive, and false negative predictions.

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Prediction Accuracy:
Prediction accuracy is the proportion of correctly classified instances.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}$$

- Measures overall correctness of the model's predictions.

Brier Score

Definition:
The Brier Score is a measure of the accuracy of probabilistic predictions. It quantifies the mean squared difference between predicted probabilities and the actual outcomes (1 for actual positive and 0 for actual negative).

Formula:

$$\text{Brier Score} = \frac{1}{n}\sum_{i=1}^{n}(p_i - y_i)^2$$

- p_i: Predicted probability of the positive class for instance i
- y_i: Actual outcome (1 if positive, 0 if negative)

The Brier Score ranges from 0 to 1, where 0 indicates perfect predictions and 1 indicates the worst prediction.

Equivalence between Brier Score and MSE

When the model is predicting probabilities, the Brier Score is equivalent to the Mean Squared Error (MSE) for binary outcomes.

$$\text{Brier Score} = \text{MSE} \quad \text{for binary classification with probabilistic predictions}$$

- Both metrics calculate the squared differences between predicted and actual values.
- Brier Score is commonly used in probabilistic classification to evaluate the calibration of predicted probabilities.

ROC Curve (Receiver Operating Characteristic)

Definition:
The ROC curve is a graphical representation of a classification model's performance across different thresholds. It plots the True Positive Rate (TPR) against the False Positive Rate (FPR) at various threshold settings.

- **True Positive Rate (TPR)**: Also known as Sensitivity or Recall.

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

- **False Positive Rate (FPR)**:

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}}$$

The ROC curve helps visualize the trade-off between the true positive rate and false positive rate, providing insight into model performance across thresholds.

Area Under the Curve (AUC)

Definition:
The Area Under the Curve (AUC) quantifies the overall ability of a model to distinguish between positive and negative classes. It represents the probability that a randomly chosen positive instance is ranked higher than a randomly chosen negative instance.

- **AUC Value**:
  - Ranges from 0 to 1.
  - An AUC of 0.5 indicates no discriminatory power (similar to random guessing).
  - An AUC close to 1 indicates excellent discriminatory ability.

Interpretation: - Higher AUC values indicate better model performance across all classification thresholds. - An AUC of 1.0 indicates a perfect model, while an AUC of 0.5 indicates a model that performs no better than chance.

ROC Curve and AUC in Model Evaluation

- **ROC Curve**: Useful for assessing a model's performance across all classification thresholds, showing the trade-off between sensitivity and specificity.
- **AUC**: Provides a single summary metric to evaluate the overall performance of the model across thresholds.

The ROC and AUC metrics are particularly helpful in binary classification tasks, allowing for threshold-independent model evaluation.

Model Evaluation Exercise (Continued)

Transform your forecasts into binary classifications for which party would have won each state, then compare that to the actual results in a confusion matrix. Compute the Brier Score of your model and any other measures you are interested in.

Conclusion

The Post-Election Reflection Assignment is due **Monday, November 18, by 9 PM**.