Q: Is There Signal in the Early Vote?

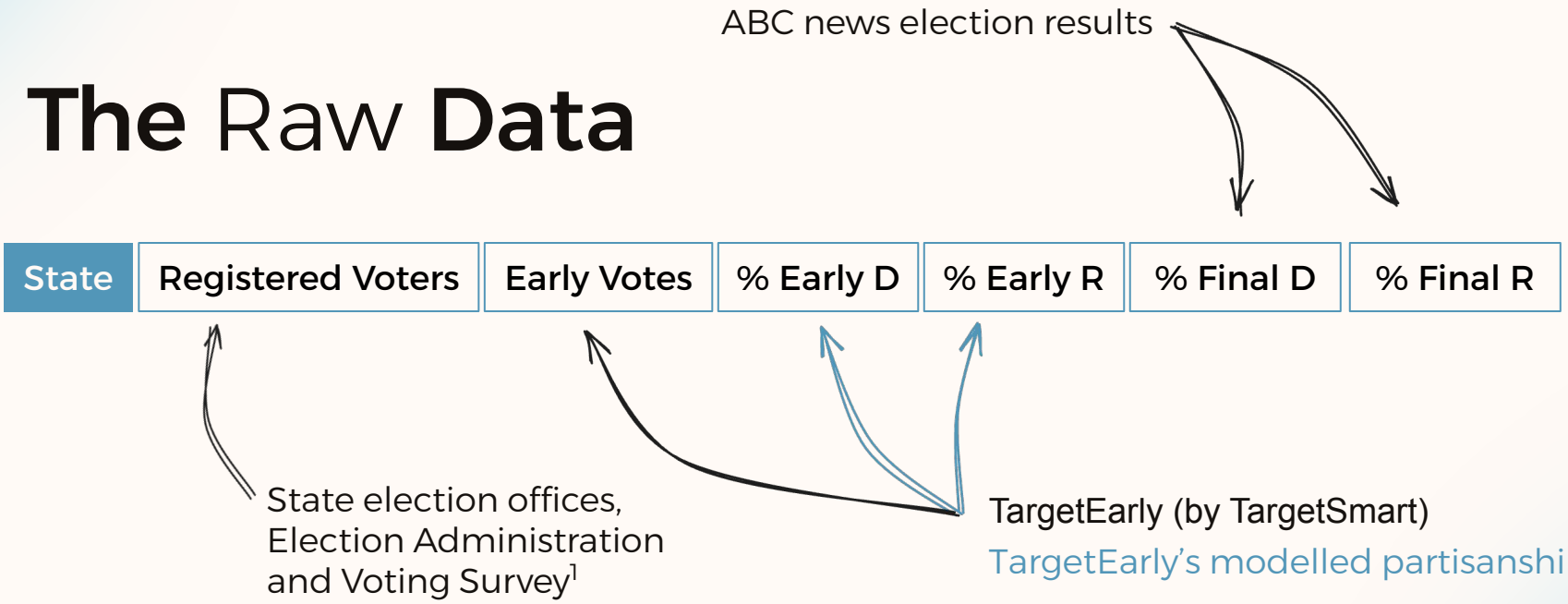
Sheline Sim
Campaigns and Elections
4 December 2024

A: Interestingly, yes!

Using modelled partisanship[1], early vote numbers[1], voter registration statistics[2], and final election numbers[2], we find a robust linear relationship between the early vote and the final election results on a per-state basis.

[1] Numbers were retrieved from TargetSmart's TargetEarly dashboard.
[2] Numbers were retrieved from the each state's election office's website, as linked to from the FEC.

The Raw **Data**



[1] Collated by the Election Assistance Commission after each federal general election.

Predicting the Election

Specifically, the democrat margin in the final election results for each state

Linear Regression

Using combinations of 2 to 4 independent variables, the best results came from using the early democrat share , early republican share , early voter turnout and previous democrat margin to predict the final democrat margin.

Accuracy
0.74[1]

Standard deviation
0.25

[1] Ranging from -inf to 1, where 1 tells us that the model can predict 100% of the variance in the dataset.

In Search Of Robustness: More Data

Limitation

50 data points is not much.

Any analysis risks ~~overfitting~~ memorizing the data and calling it correlation.

Solution

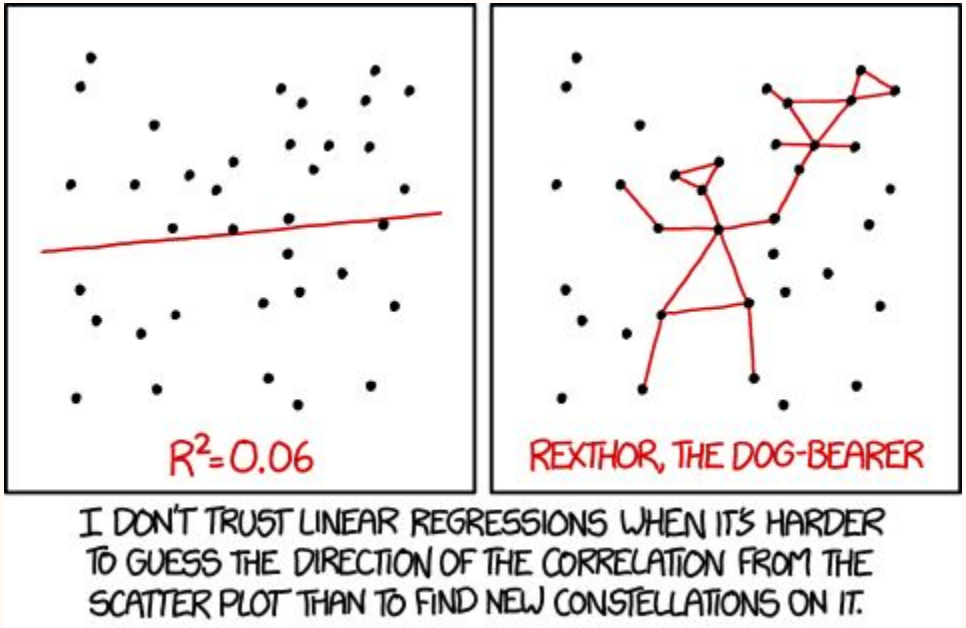
Double the size of the dataset by including data from the 2020 election.

Result

A slight bump in accuracy to **0.78** (+0.04) and a solid drop in standard deviation to **0.15** (-0.1)

This indicates a linear relationship in the data that is being captured well.

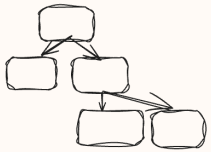
Nonlinearity?



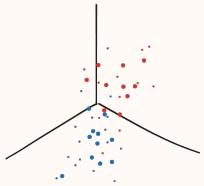
Probably not!

Modelled the same dataset using 3 different nonlinear methods.

Polynomial



Random Forest



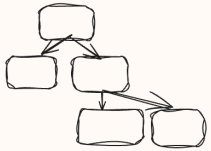
SVR

?

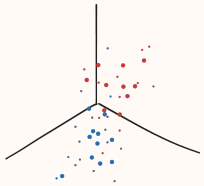
Probably not!

Modelled the same dataset using 3 different nonlinear methods.

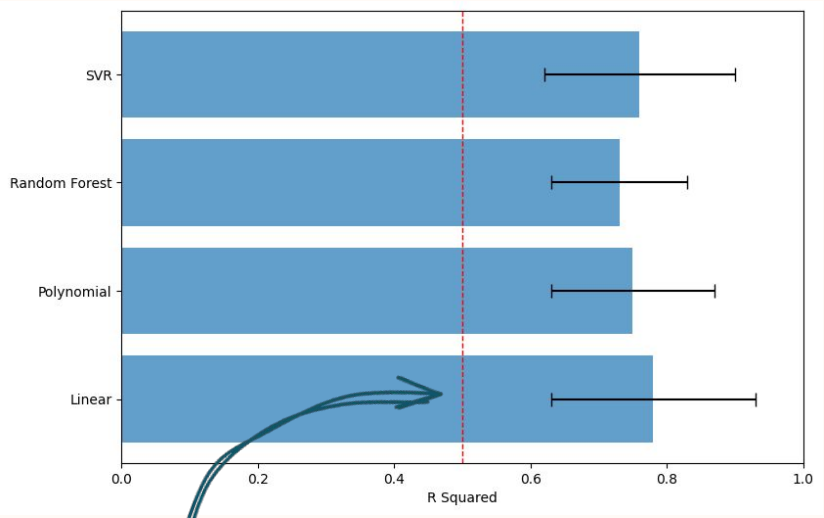
Polynomial



Random Forest



SVR



Good R squared for social phenomenon

Conclusion

Patterns are hard to observe in small samples.

Looking at all 50 states across 2 elections, **a robust linear relationship** is found in the data.

Early vote numbers can predict the final election results.

Further Analysis

Even More Data
More historical data
Precinct or county-level data

Explore **more features** or **interactions** between features

Explore **different modelling strategies**
e.g. modelling on an individual level

Thank you!

Questions?

Dataset and analysis