

# Methodology

## The Big Picture

We use the correlation between the presidential vote on the one hand, and state legislative and congressional votes on the other, to predict how new districts will likely vote and so how biased a plan will be. Our correlations come from the last 10 years of elections, and factor in both any extra advantage incumbents might have as well as how much each state’s results might differ from others. We also allow our predictions to be imperfect by quantifying how much our method missed the actual outcomes of past elections, including the degree to which partisan tides have changed party performance from one election to the next. This enables us to generate the most accurate, data-driven, and transparent prediction we can.

## The Details

We use a Bayesian hierarchical model of district-level election returns, run for all state legislatures and congressional delegations on the elections from 2012 through 2020. Formally, the model is:

$$y_i \sim \mathcal{N}(\mathbf{X}_i\boldsymbol{\beta} + \mathbf{X}_i\boldsymbol{\beta}_{s(i)} + \mathbf{X}_i\boldsymbol{\beta}_{c(i)}, \sigma_y^2)$$
$$\begin{pmatrix} \beta_{0s} \\ \beta_{1s} \\ \beta_{2s} \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\beta_{0s}}^2 & \rho\sigma_{\beta_{0s}}\sigma_{\beta_{1s}} & \rho\sigma_{\beta_{0s}}\sigma_{\beta_{2s}} \\ \rho\sigma_{\beta_{0s}}\sigma_{\beta_{1s}} & \sigma_{\beta_{1s}}^2 & \rho\sigma_{\beta_{1s}}\sigma_{\beta_{2s}} \\ \rho\sigma_{\beta_{0s}}\sigma_{\beta_{2s}} & \rho\sigma_{\beta_{1s}}\sigma_{\beta_{2s}} & \sigma_{\beta_{2s}}^2 \end{pmatrix} \right)$$
$$\begin{pmatrix} \beta_{0c} \\ \beta_{1c} \\ \beta_{2c} \end{pmatrix} \sim \mathcal{N} \left( \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\beta_{0c}}^2 & \rho\sigma_{\beta_{0c}}\sigma_{\beta_{1c}} & \rho\sigma_{\beta_{0c}}\sigma_{\beta_{2c}} \\ \rho\sigma_{\beta_{0c}}\sigma_{\beta_{1c}} & \sigma_{\beta_{1c}}^2 & \rho\sigma_{\beta_{1c}}\sigma_{\beta_{2c}} \\ \rho\sigma_{\beta_{0c}}\sigma_{\beta_{2c}} & \rho\sigma_{\beta_{1c}}\sigma_{\beta_{2c}} & \sigma_{\beta_{2c}}^2 \end{pmatrix} \right)$$

where

- $i$  indexes district level elections
- $s$  indexes states, with  $s(i)$  denoting the state of district election  $i$
- $c$  indexes election cycles, with  $c(i)$  denoting the election cycle of district election  $i$
- $y_i$  is the Democratic share of the two-party vote in district election  $i$
- $\mathbf{X}_i$  is a matrix of covariate values for district election  $i$
- $\beta$  is a matrix of population-level intercept and slopes corresponding to covariates  $\mathbf{X}$
- $\beta_{s(i)}$  and  $\beta_{c(i)}$  are matrices of coefficients for the state and election cycle, respectively, of district election  $i$
- $\sigma_y$  is the residual population-level error term

The model includes two covariates: 1) the two-party district-level Democratic presidential vote share, averaged across 2012 and 2016 and centered around its global mean; 2) the incumbency status in district election  $i$ , coded -1 for Republican, 0 for open, and 1 for Democratic. The model allows the slope for each—as well as the corresponding intercept—to vary across both states and election cycles. Chambers accounted for minimal variation in an ANOVA test, so state legislative and congressional results were modeled together as emerging from a common distribution.

When generating predictions, PlanScore draws 1000 samples from the posterior distribution of model parameters, and uses them to calculate means and probabilities. As part of this process we sample from the covariance matrix of cycle random effects, thus allowing the uncertainty of predicting for an unknown election cycle to propagate into our predictions. This has the effect of predicting for an average election over the last 10 years, but with error bands that encompass the full range of partisan tides that actually occurred. Generally

speaking, new maps only have either the 2016 or the 2020 presidential elections results, so we must translate those results to the 2012-2020 averages used to estimate the model. For this purpose, we regressed the 2012-2020 averages separately on the 2016 and 2020 results using [congressional district data from Daily Kos](#), and then used the results for the translation. Regression coefficients and intercepts are below.

### **2020**

$$P_{avg} = 0.96 * P_{2020} + 0.01$$

### **2016**

$$P_{avg} = 0.91 * P_{2016} + 0.05$$

Table 1: PlanScore prediction model results

	Estimate	95% Credible Interval
POPULATION-LEVEL		
Intercept ( $\beta_0$ )	0.50	[0.47, 0.53]
Presidential vote ( $\beta_1$ )	0.83	[0.68, 0.97]
Incumbency ( $\beta_2$ )	0.05	[0.03, 0.07]
STATE-LEVEL		
<i>Standard Deviations</i>		
Intercept ( $\sigma_{\beta_{0s}}$ )	0.02	[0.02, 0.03]
Presidential vote ( $\sigma_{\beta_{1s}}$ )	0.10	[0.08, 0.13]
Incumbency ( $\sigma_{\beta_{2s}}$ )	0.02	[0.01, 0.02]
<i>Correlations</i>		
Intercept - Pres. vote ( $\rho\sigma_{\beta_{0s}}\sigma_{\beta_{1s}}$ )	-0.43	[-0.64, -0.18]
Intercept - Incumbency ( $\rho\sigma_{\beta_{0s}}\sigma_{\beta_{2s}}$ )	0.04	[-0.23, 0.32]
Pres. vote - Incumbency ( $\rho\sigma_{\beta_{1s}}\sigma_{\beta_{2s}}$ )	-0.69	[-0.84, -0.49]
CYCLE-LEVEL		
<i>Standard Deviations</i>		
Intercept ( $\sigma_{\beta_{0c}}$ )	0.03	[0.01, 0.07]
Presidential vote ( $\sigma_{\beta_{1c}}$ )	0.15	[0.07, 0.34]
Incumbency ( $\sigma_{\beta_{2c}}$ )	0.02	[0.01, 0.05]
<i>Correlations</i>		
Intercept - Pres. vote ( $\rho\sigma_{\beta_{0c}}\sigma_{\beta_{1c}}$ )	-0.17	[-0.80, 0.61]
Intercept - Incumbency ( $\rho\sigma_{\beta_{0c}}\sigma_{\beta_{2c}}$ )	-0.18	[-0.81, 0.60]
Pres. vote - Incumbency ( $\rho\sigma_{\beta_{1c}}\sigma_{\beta_{2c}}$ )	-0.61	[-0.97, 0.23]

Note: Model estimated in brms for R. Model based on 4 MCMC chains run for 4000 iterations each with a 2000 iteration warm-up. All model parameters converged well with  $\hat{R} < 1.01$ .