NC STATE UNIVERSITY

Analysis of Presidential Election Polling Bias

ST 533 Final Project

Morgan Schulte, Jasmine Wang, Yulun Xu

Introduction

The 2020 United States presidential election was the 59th quadrennial presidential election, held nominally on Tuesday, November 3, 2020. The Democratic ticket of former Vice President Joe Biden and U.S. Senator Kamala Harris defeated the Republican ticket of incumbent President Donald Trump and Vice President Mike Pence.

Presidential election polling and subsequent analysis have emerged as one of the most public applications of statistical methods. These spatiotemporal data are noisy and potentially biased. However, the presidential election pollings of 2016 and 2020 were not as accurate as they were before.

Objective

To study the bias in state-level Presidential election polls from the elections of 2012, 2016 and 2020, including the following 3 points:

1. Devise a method to combine the individual polls to forecast the election results in each state and each year

$$X_{it} = \sum_{j=1}^{N_t} w_{itj} P_{jt}$$

2. Whether the polling bias is constant over state and election

3. What are the effects that cause bias change over state and election

Data Description

The response: $B_{it} = \mathrm{E}(Y_{it} - X_{it})$

Where *i* represents the state, *t* represents the election

Y_{it} is polling result over the state and election

$X_{it} = \sum_{j=1}^{N_t} w_{itj} P_{jt}$ is the average of the polls

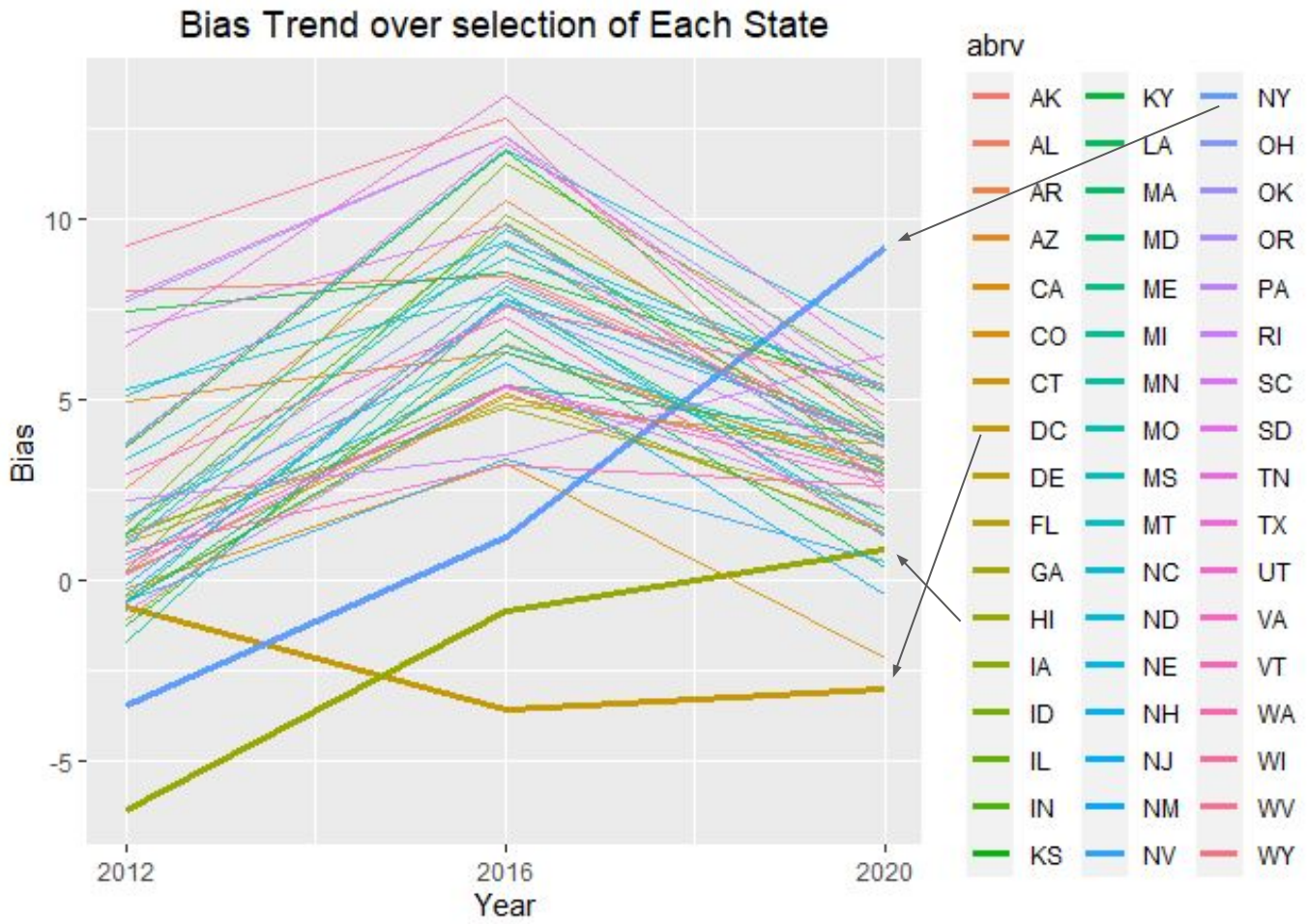
We divided our polling data to 3 categories: Pre-Debate, Debate, Post-Debate

Covariates: Population density, Sample size/Population, Percent Republican Representatives, Percent Caucasian, # of Polling resources, Previous election results, percent agricultural land use

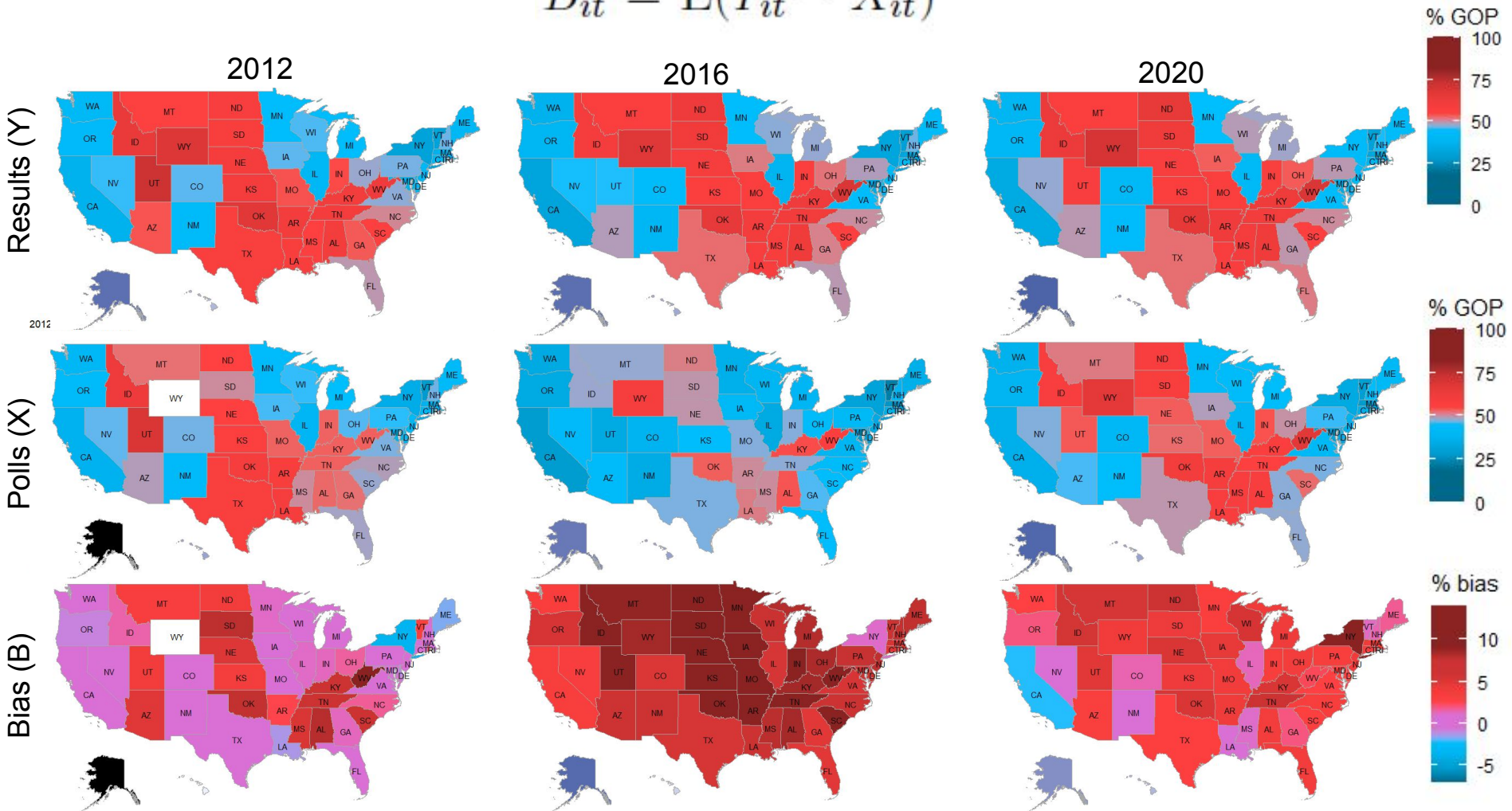
Sensitivity tests for the weights

Run	Pre-D	Debate	Post-D	Mean of $\lvert B_{it} \rvert$	$\mathrm{Var}(B_{it})$
1	0.2	0.3	0.5	5.19	19.46
2	0.02	0.25	0.7	4.93	17.23
3	0	0.15	0.9	4.68	15.26
4	0	0.05	0.95	4.61	14.80
5	0	0	1	4.55	14.37

Thus, we went with run # 5 which has the lowest bias and variance


Bias Trend over selection of Each State
Bias
10
5
0
-5
2012
2016
2020
Year
abrv
AK
AL
AR
AZ
CA
CO
CT
DC
DE
FL
GA
HI
IA
ID
IL
IN
KS
KY
LA
MA
MD
ME
MI
MN
MO
MS
MT
NC
ND
NE
NH
NJ
NM
NV
NY
OH
OK
OR
PA
RI
SC
SD
TN
TX
UT
VA
VT
WA
WI
WV
WY

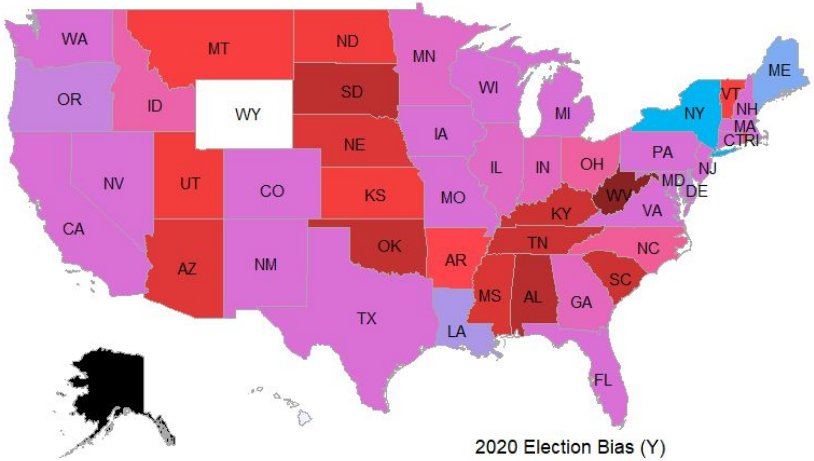
$$B_{it} = \mathrm{E}(Y_{it} - X_{it})$$



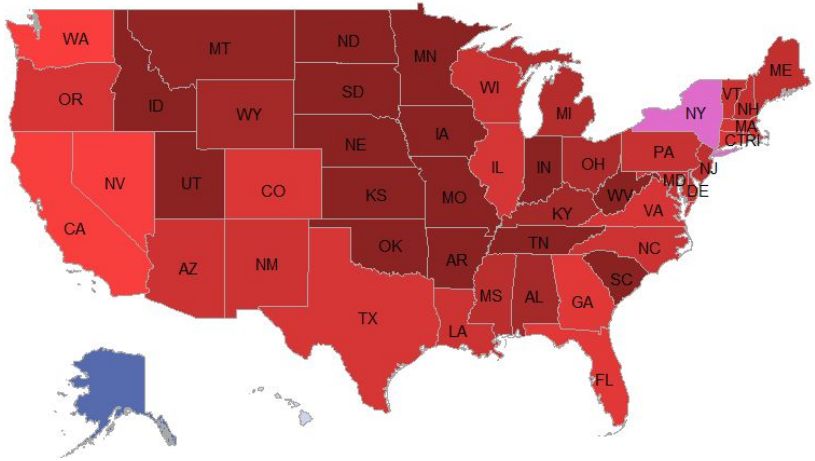
NC STATE UNIVERSITY

Response Variable | Bias

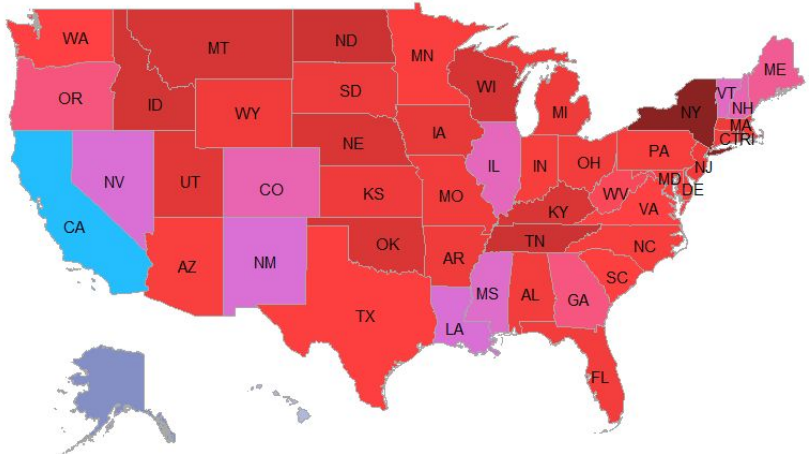
2012 Election Bias (Y)

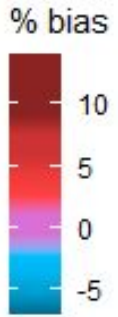


2016 Election Bias (Y)



2020 Election Bias (Y)





Spatiotemporal Models

R package CARBayesST: ST.CARar - AR(1)

$$
\begin{aligned}
\psi_{kt} &= \phi_{kt}, \\
\boldsymbol{\phi}_t|\boldsymbol{\phi}_{t-1} &\sim \mathrm{N}\left(\rho_T\boldsymbol{\phi}_{t-1}, \tau^2\mathbf{Q}(\mathbf{W},\rho_S)^{-1}\right) \qquad t = 2,\ldots,N, \\
\boldsymbol{\phi}_1 &\sim \mathrm{N}\left(\mathbf{0}, \tau^2\mathbf{Q}(\mathbf{W},\rho_S)^{-1}\right), \\
\tau^2 &\sim \text{Inverse-Gamma}(a,b), \\
\rho_S,\rho_T &\sim \text{Uniform}(0,1).
\end{aligned}
$$

ρ_T. Temporal autoregressive parameter

$\boldsymbol{\phi}_t = (\phi_{1t},\ldots,\phi_{Kt})$ Random effects for time period t

$$\mathbf{Q}(\mathbf{W},\rho_S) = \rho_S[\text{diag}(\mathbf{W1}) - \mathbf{W}] + (1-\rho_S)\mathbf{I},$$ Precision matrix

ST.CARar without Covariates

```
modelar <- ST.CARar(bias~1,family="gaussian",W=W,burnin=100000,n.sample=500000,thin=10,verbose=FALSE)
```

	Median	2.5%	97.5%	n.effective	Geweke.diag
(Intercept)	4.2640	4.0644	4.4623	66467.1	0.4
tau2	21.0755	5.2728	28.7437	639.9	1.0
nu2	0.0527	0.0032	4.4879	612.0	-0.9
rho.S	0.8994	0.6890	0.9923	1306.6	-1.2
rho.T	0.4520	0.2488	0.7987	1535.2	-1.2

Model	DIC	WAIC	LMPL	Probability
ST.CARar	-97.143	479.584	-233.154	0.218
ST.CARanova	674.637	679.035	-341.769	0.225
ST.CARlinear	798.277	797.890	-398.950	0.000

Probability of 95% interval including 0 is 0.218

ST.CARar with Covariates

```
modelar <- ST.CARar(bias~X,family="gaussian",W=W,burnin=100000,n.sample=500000,thin=10,verbose=FALSE)
```

	Median	2.5%	97.5%	n.effective	Geweke.diag
(Intercept)	4.2051	3.8838	4.5240	40000.0	0.5
XGOPrep	0.5804	-0.2420	1.4073	28274.3	0.0
Xsampop	0.1254	-0.5745	0.8384	6193.6	-0.8
Xcaucasian	0.7078	0.1606	1.2481	17627.7	1.3
Xpolln	-0.3069	-0.7133	0.0985	33660.2	0.3
Xprevele	1.1360	0.2463	2.0291	33349.1	-0.6
Xagland	0.0675	-0.3414	0.4839	30973.0	0.4
Xpopden	-0.0594	-0.5780	0.4500	35462.9	0.5
tau2	1.1370	0.0323	5.2367	1990.9	0.6
nu2	3.7398	2.3885	5.1436	2991.2	-0.7
rho.S	0.9952	0.9550	0.9999	3153.2	-0.3
rho.T	0.1527	0.0058	0.7581	8531.4	-0.3

Model	DIC	WAIC	LMPL	Probability
ST.CARar	634.843	640.204	-322.619	0.191
ST.CARanova	645.664	648.848	-324.922	0.116
ST.CARlinear	756.494	756.065	-378.210	0.0748

Space and Time Separately

CAR model

R package CARBayes, S.CARLeroux function

$$
\begin{aligned}
\psi_k &= \phi_k \\
\phi_k|\boldsymbol{\phi}_{-k}, \mathbf{W}, \tau^2, \rho &\sim \mathrm{N}\left(\frac{\rho\sum_{i=1}^{K} w_{ki}\phi_i}{\rho\sum_{i=1}^{K} w_{ki} + 1 - \rho}, \frac{\tau^2}{\rho\sum_{i=1}^{K} w_{ki} + 1 - \rho}\right)
\end{aligned}
$$

$$
\begin{aligned}
\tau^2 &\sim \text{Inverse-Gamma}(a, b) \\
\rho &\sim \text{Uniform}(0, 1).
\end{aligned}
$$

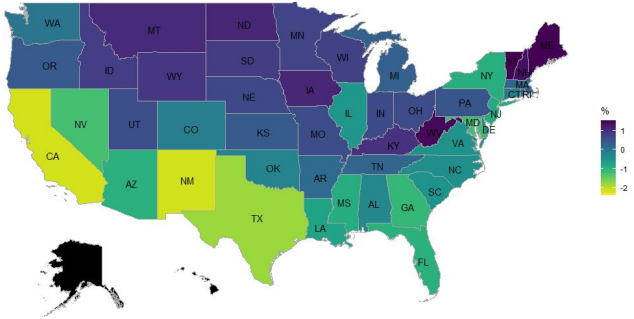
CAR Model Results of 2012 Election Bias

```
model <- S.CARleroux(Y~X, family="gaussian", W=W, burnin=20000,n.sample=200000,thin=10,verbose=FALSE)
```

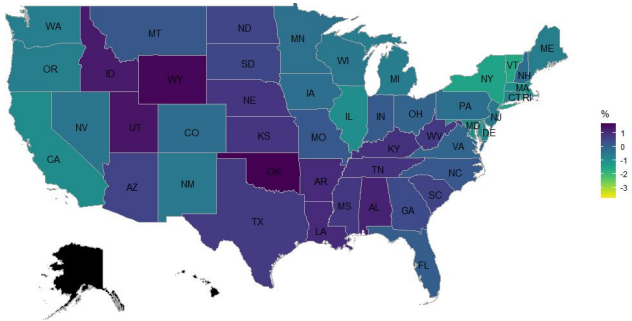
	Median	2.5%	97.5%	n.effective	Geweke.diag
(Intercept)	2.0924	1.4828	2.7014	16768.4	0.2
Xpopden	0.7473	-0.2335	1.7246	18000.0	1.0
Xrep	0.6172	-0.6874	1.9250	17593.3	-0.1
Xsamp	-0.2897	-1.0895	0.5014	17499.4	0.6
Xcaucasian	0.8113	0.0502	1.5706	17582.1	0.9
Xpolln	-0.5917	-1.3867	0.1987	18000.0	0.5
Xprevelc	1.5713	-0.0609	3.1607	18000.0	0.2
Xagland	-0.3848	-1.1107	0.3414	18000.0	0.1
nu2	4.5457	2.9955	7.2269	16448.1	0.5
tau2	0.0083	0.0021	0.0941	1692.3	-1.3
rho	0.3623	0.0173	0.9129	8442.2	0.1

significant covariates

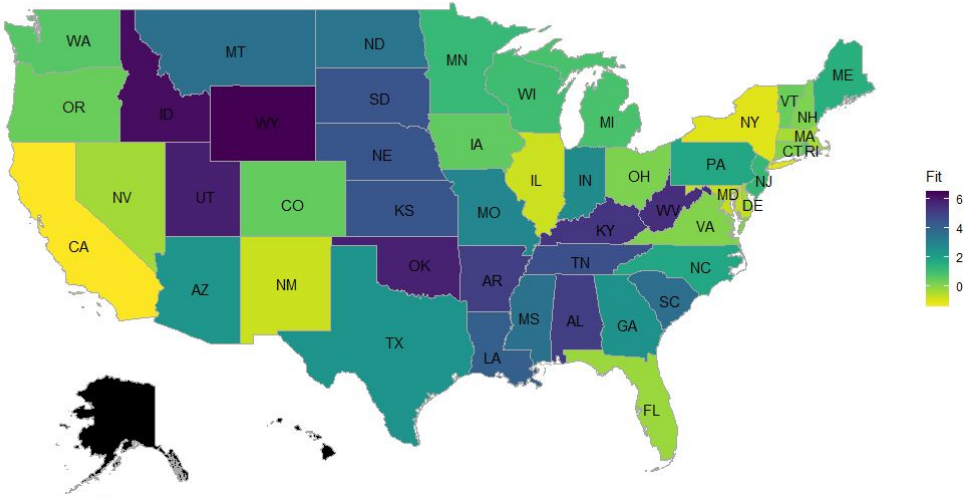
scale(% Caucasian)



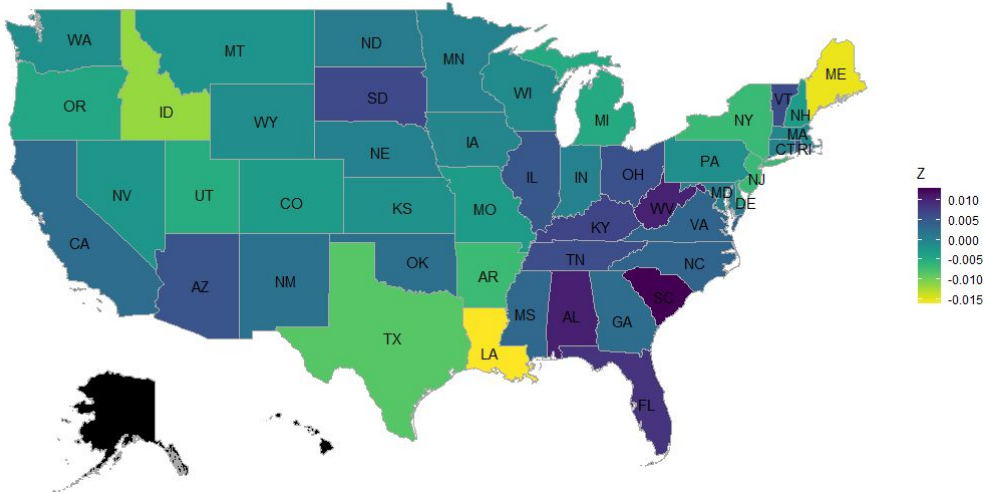
scale(Prev Election % GOP)



2012 Fit



2012 Spatial Correlation



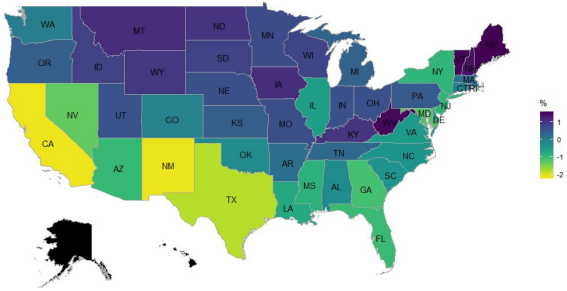
CAR Model Results of 2016 Election Bias

```
model <- S.CARleroux(Y~X, family="gaussian", W=W, burnin=20000,n.sample=500000,thin=10,verbose=FALSE)
```

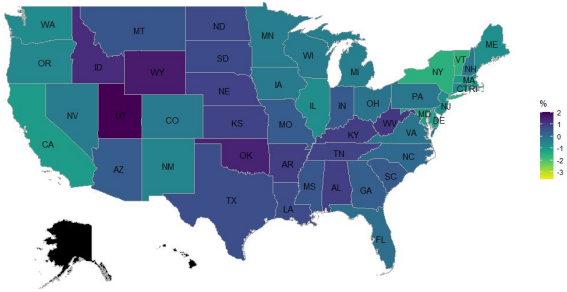
	Median	2.5%	97.5%	n.effective	Geweke.diag
(Intercept)	7.4119	6.9551	7.8682	80000.0	0.1
XGOP_rep	0.5236	-0.5703	1.6167	80000.0	-1.1
Xsampop	0.2518	-0.3591	0.8753	80000.0	-0.1
Xcaucasian	1.0389	0.3761	1.7087	80000.0	-0.1
Xpolln	-0.4868	-1.0431	0.0711	80000.0	-0.2
Xprevelc	1.5743	0.3596	2.7894	80000.0	1.0
Xagland	0.4465	-0.0600	0.9548	80000.0	1.2
Xpopden	-0.3320	-1.0088	0.3532	79049.2	0.0
nu2	2.4705	1.6531	3.9086	61882.5	-0.3
tau2	0.0085	0.0021	0.0943	3982.4	0.9
rho	0.3753	0.0176	0.9172	35957.9	2.5

significant covariates

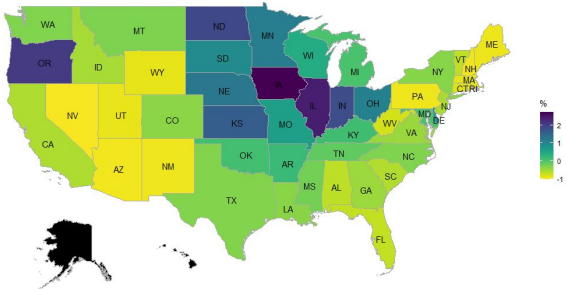
scale(% Caucasian)



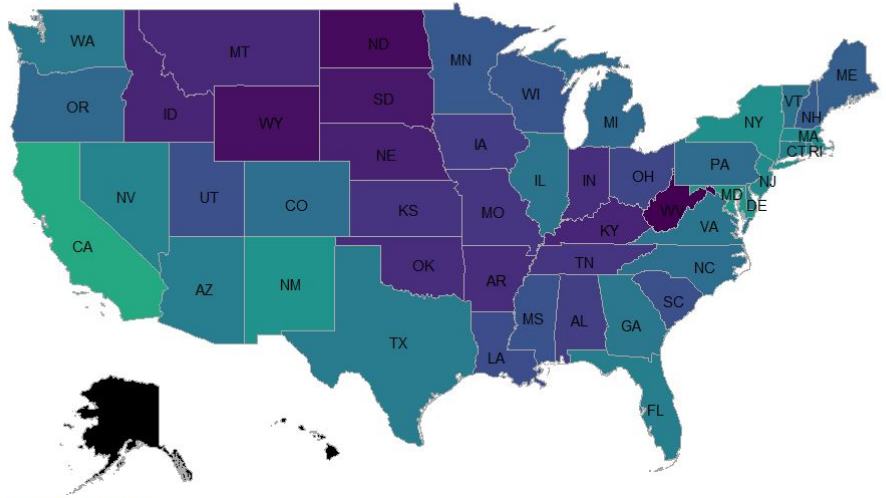
scale(Prev Election % GOP)



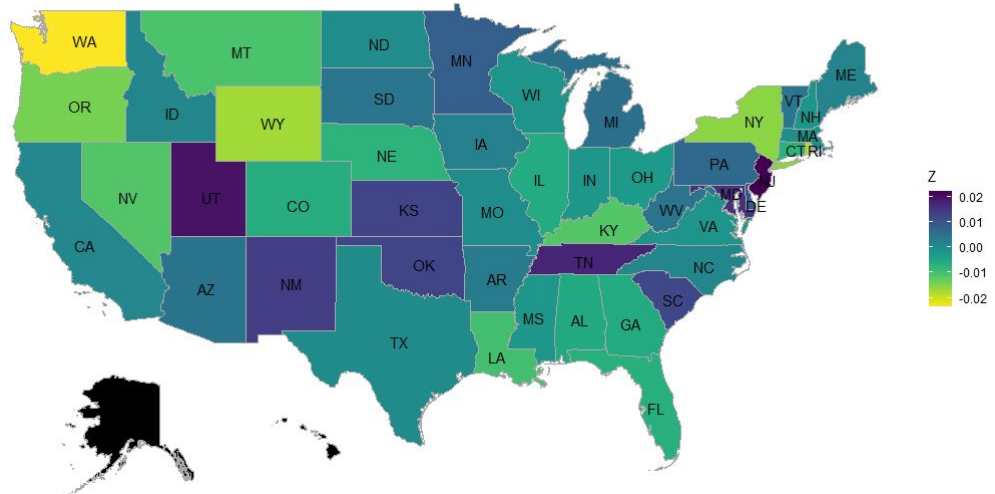
scale(% Ag Land)



2016 Fit



2016 Spatial Correlation



CAR Model Results of 2020 Election Bias

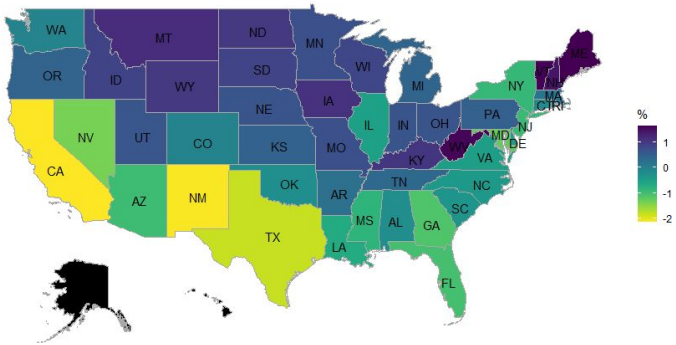
```
model <- S.CARleroux(Y~X, family="gaussian", W=W, burnin=50000,n.sample=500000,thin=10,verbose=FALSE)
```

	Median	2.5%	97.5%	n.effective	Geweke.diag
(Intercept)	3.1804	2.6535	3.6970	240000.0	0.1
XGOP_rep	0.5374	-0.7008	1.8097	36555.6	-0.8
Xsampop	-0.5797	-1.2768	0.1216	65401.2	0.0
Xcaucasian	1.0638	0.2496	1.8482	36582.6	1.5
Xpolln	0.2394	-0.3536	0.8333	121677.5	0.2
Xprevelc	-0.3699	-1.8147	1.0794	130733.1	0.0
Xagland	0.0661	-0.5053	0.6373	75519.8	-2.4
Xpopden	-0.5251	-1.2862	0.2382	91995.1	0.4
nu2	3.2801	0.0252	5.2314	952.6	0.4
tau2	0.0089	0.0022	4.7329	537.7	-0.6
rho	0.3694	0.0170	0.9154	61922.4	0.6

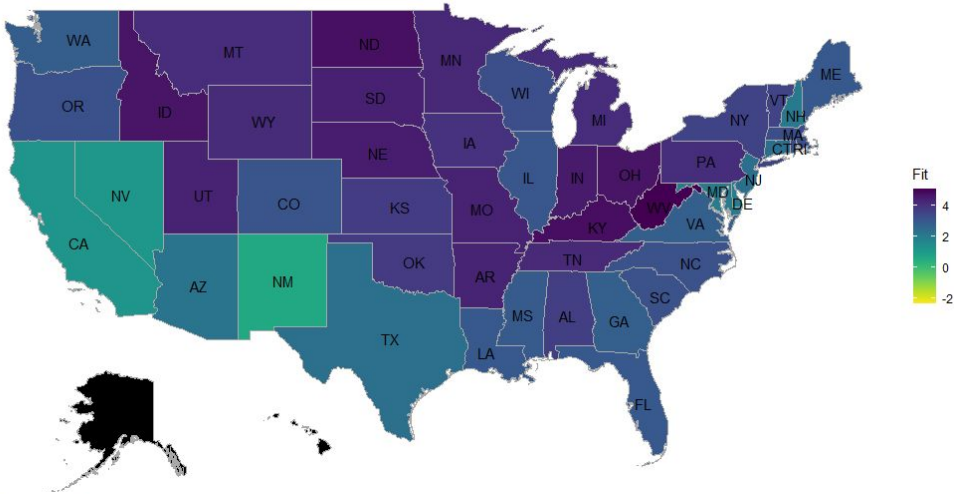
NC STATE UNIVERSITY

significant covariates

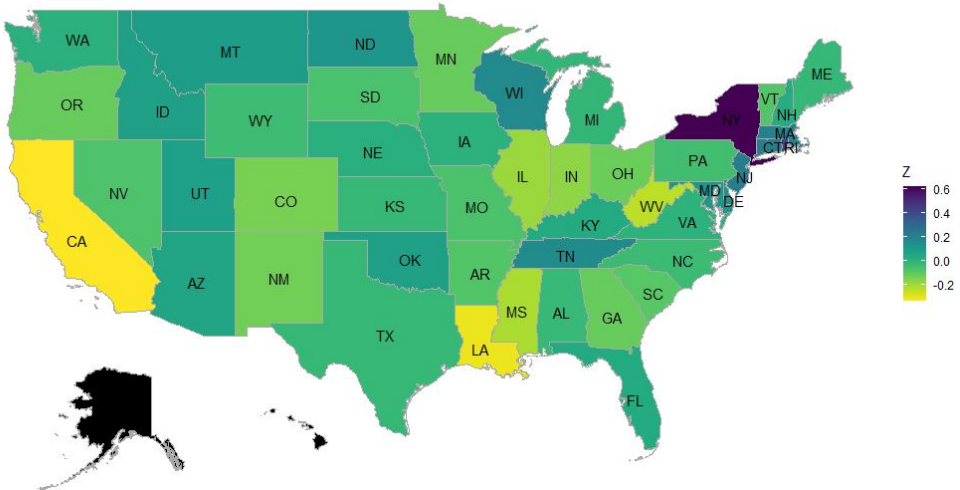
scale(% Caucasian)



2020 Fit



2020 Spatial Correlation

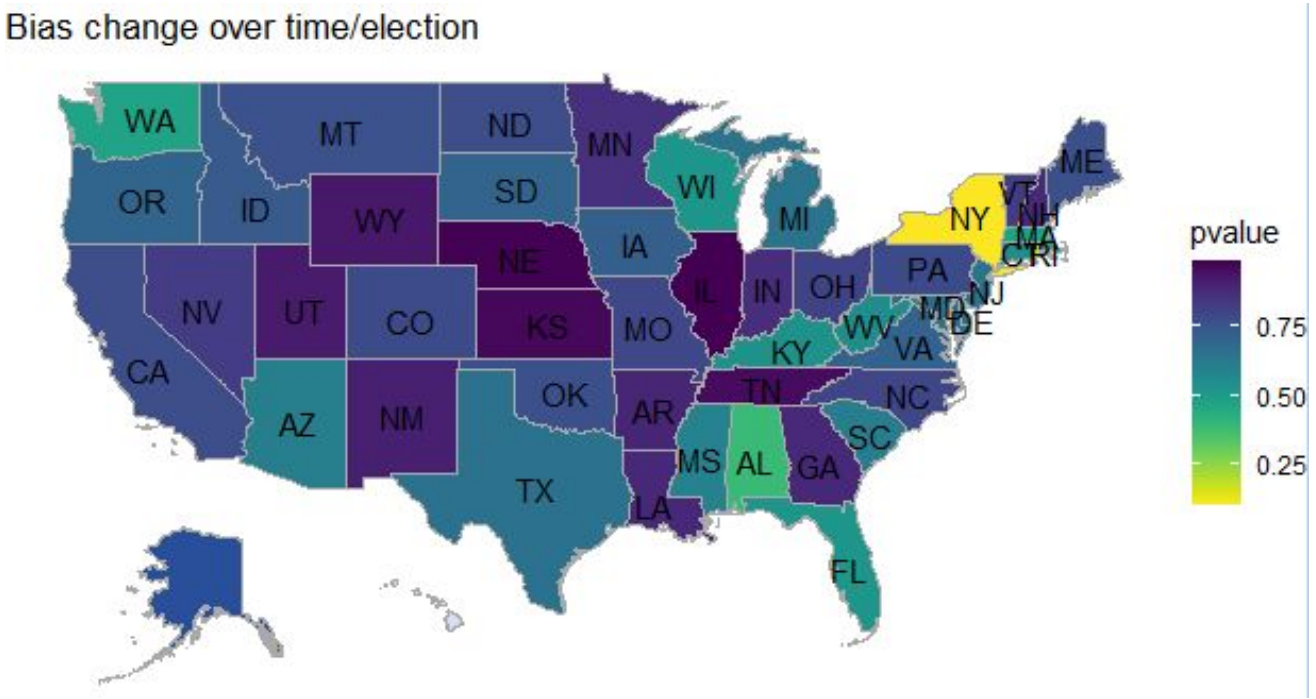


Non-spatial Linear Model

```
Call:
lm(formula = Y ~ X)

Residuals:
    Min      1Q  Median      3Q     Max
-6.7210 -2.3292 -0.5392  2.3433  7.2301

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.29057    0.25713  16.687   <2e-16 ***
Xyear        0.62745    0.46663   1.345   0.1810
Xpopden      0.10138    0.37934   0.267   0.7897
Xrep         0.47836    0.55613   0.860   0.3912
Xsamp       -0.12084    0.50416  -0.240   0.8109
Xcaucasian   0.72355    0.32276   2.242   0.0266 *
Xpolln      -0.23145    0.29580  -0.782   0.4353
Xprevelc     1.31980    0.64855   2.035   0.0438 *
Xagland      0.02641    0.28618   0.092   0.9266
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.106 on 137 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.3329,    Adjusted R-squared:  0.2939
F-statistic: 8.544 on 8 and 137 DF,  p-value: 2.14e-09
```

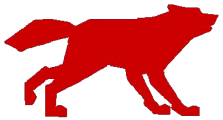
```
Call:
lm(formula = Y ~ yr)

Residuals:
    Min      1Q  Median      3Q     Max
-7.8621 -2.6600 -0.8187  2.5962  9.1913

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.2674     0.3041  14.031   <2e-16 ***
yr            0.5054     0.3057   1.653      0.1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 3.675 on 144 degrees of freedom
  (1 observation deleted due to missingness)
Multiple R-squared:  0.01863,   Adjusted R-squared:  0.01181
F-statistic: 2.734 on 1 and 144 DF,  p-value: 0.1004
```

Bias consistent over time?



Conclusion

- Spatiotemporal model shows systematic polling bias is constant over state and election
- Spatiotemporal model shows that percent Caucasian & previous election are significantly associated with the polling bias over space and time.
- Separate spatial models show percent Caucasian has an effect to the bias for all 3 elections, while previous election has effects to the bias for 2012 and 2016 elections. Percent agricultural land only has an effect in 2016 election.
- Largest bias over 3 elections is in New York. Polls failed to capture GOP support for New York in 2020 (~ +8%).
- There are spatial effects of bias for each election, however, they are not strong (rho ~ 0.37):
- Average positive bias across all years. The polls surveyed poorly and failed to capture Republican voters. More people voted for Republican than polls anticipated.
- The polls underestimated GOP supporters in general.
- Sampled surveys did not reflect GOP supports.



Thanks for your attention!