433 Final

Zahid Rangwala and Brice Turner

Election Polling

- One of the most public applications of statistical methods
- Spatiotemporal data is noisy and potentially biased
- This presentation will study the bias in state-level Presidential election polls from the elections of 2012 and 2016

Background

- With this first step, our goal is to come up with a method in which we will predict the election results in 2012 and 2016
- This will be completed by averaging the polls from 2012 and 2016 respectively and adding a weighting to each poll
- Our method up-weights polls taken closer to the election

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

How did we determine the weights?

1. Generated numbers sequentially from 0 and 3 for every poll.
2. Used those numbers as the inputs into an exponential function
3. Then scaled the weights so they sum to one.
4. Process was repeated for each state and for each year.

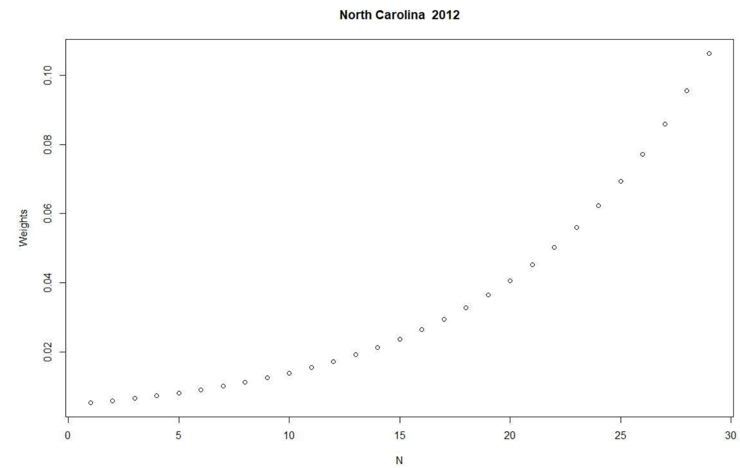
```
poll2012 <- poll2012 %>%                                  # Specify data frame
  group_by(State) %>%                       # Specify group indicator
    mutate(scale = exp(seq(0,3, length = n())))

b <- poll2012 %>% summarise_at(vars(scale), list(name = sum))
poll2012 <- left_join(poll2012 , b , by = "State")
poll2012 <- mutate(poll2012 , weight = scale/name)
```

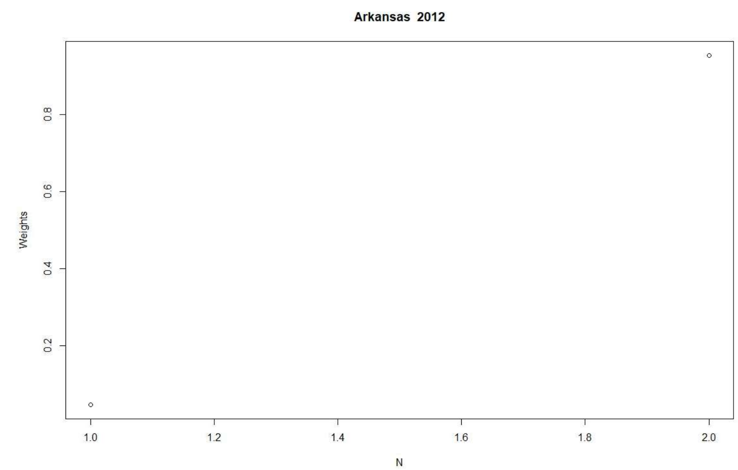
Weights-2012

Each state has its own exponential like curve.

States with more data have more pronounced curves as seen on the right

The sum of all the weights for each state sums to one



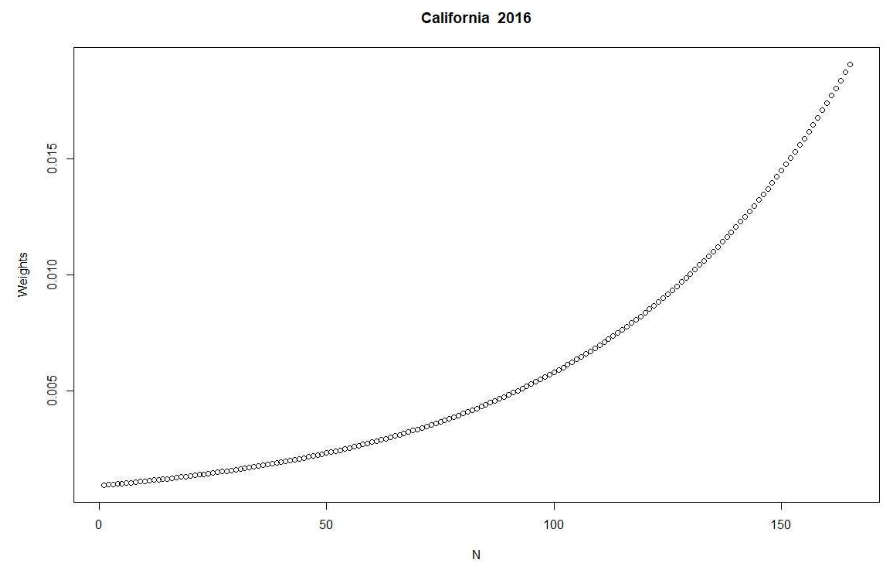


Weights-2016

The same process was used for 2016

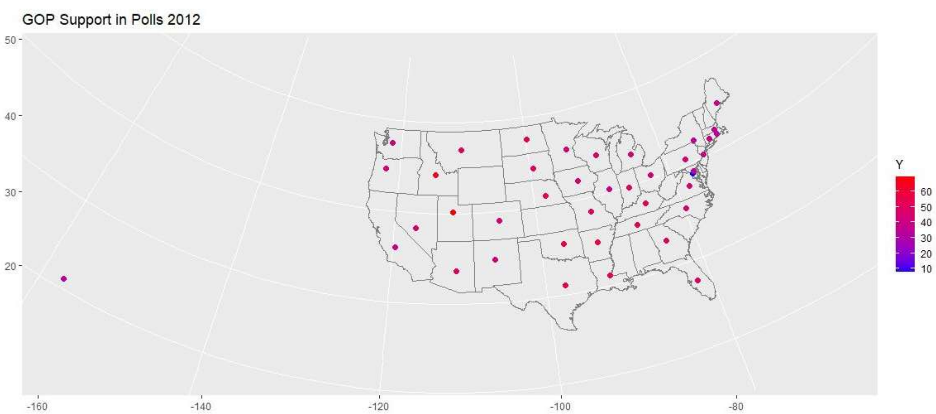
2016 in general had more polling data than 2012 , so the weighted values were able to be more accurate

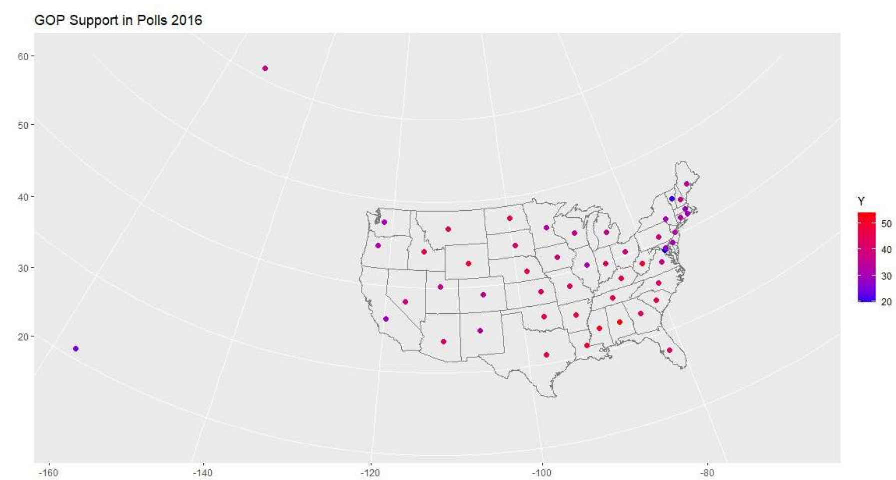
Resulting in more exponential like shapes



Polling data Visualized

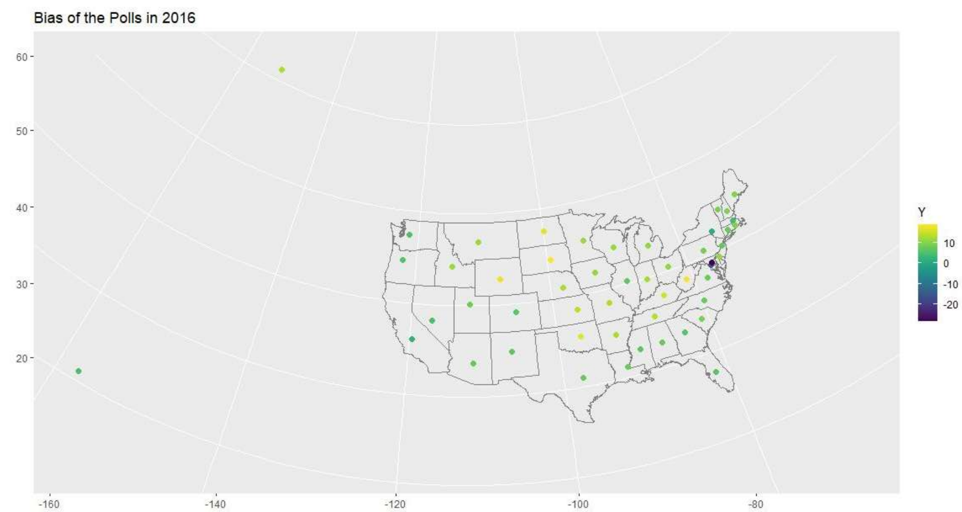
We can see how the polls were predicting the outcome of the election by calculating the weighted average of each poll for each state





Bias

Bias is calculated by subtracting the expected value of the actual support of the GOP candidate from the polling average $B_{it} = \mathrm{E}(Y_{it} - X_{it})$.



MLE-Matern

Now that we have calculated bias in 2012 and 2016 elections, we will see if this bias is significant or not using a MLE-Matern model. We will assume that the bias is constant over states and elections.

	state	X	percentGOP_Y	Z	lat	long
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Arizona	50.9	53.7	2.71	33.7	-111.
2	Arkansas	57.9	60.6	2.66	35.0	-92.4
3	California	37.6	37.1	-0.462	36.1	-120.
4	Colorado	46.7	46.1	-0.581	39.1	-105.
5	Connecticut	41.2	40.7	-0.462	41.6	-72.8
6	District of Columbia	8	7.28	-0.722	38.9	-77.0

	state	X	percentGOP_Y	Z	lat.x	long.x	lat.y	long.y
	<chr>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	Alabama	54.1	62.1	7.98	32.8	-86.8	32.8	-86.8
2	Alaska	38.1	51.3	13.2	61.4	-152.	61.4	-152.
3	Arizona	40.7	48.7	7.98	33.7	-111.	33.7	-111.
4	Arkansas	47.7	60.6	12.9	35.0	-92.4	35.0	-92.4
5	California	29.1	31.6	2.54	36.1	-120.	36.1	-120.
6	Colorado	36.9	43.3	6.31	39.1	-105.	39.1	-105.

MLE-Matern

- Using the Matern covariate, which measures statistical covariates between two points, and is solely dependant on distance, with MLE.
  - P-value of 2012 = 1.95e-07
  - P-value of 2016 = 0
- With the p-values being less than the alpha we used, .05, we can say that there is bias within both of these forecasts.

2012

```
Parameters of the mean component (trend):
  beta0   beta1
45.0916  2.5582
```

2016

```
Estimation method: maximum likelihood

Parameters of the mean component (trend):
  beta0   beta1
29.0045  1.7081
```

Systemic Bias?

-Fear of perception

-Lack of diversity in location

-Age of people being asked

-Family vs single