

A look into Kamala Harris' polls in the upcoming 2024 US Presidential Election*

Veyasan Ragulan

November 3, 2024

The 2024 US election has been seen as a close race between the Democratic nominee Kamala Harris, and the Republican Nominee Donald Trump, with each candidate holding down key states and matching each other in battleground states such as Pennsylvania and Georgia. This paper looks at the polling as of late October 2024 to see how Kamala Harris fares as the election draws near.

```
library(tidyverse)

library(rstanarm)

bayesian_model <-
  readRDS(file = here::here("models/electionmodel.rds"))

spline_model <-
  readRDS(file = here::here("models/splinemodel.rds"))
```

1 Introduction

The 2024 US Presidential race has been closely contested since incumbent president, Joe Biden, announced his exit from the race. The race is primarily between current VP, Democratic Party candidate Kamala Harris (with running mate Time Walz), and former president, Republican Party candidate Donald Trump (with running mate J.D. Vance). Polls across the country have conflicting reports on who will come out on top, with some sources suggesting an unheard-of tie in the electoral college vote (269-269 vote distribution) (Lopez 2024).

*Code and data are available at: <https://github.com/Veyasan1/2024USElectionModel>

To get a better grasp of nationwide and state opinions going into the election, a predictive model was created in this report. The model utilises data from fivethirtyeight.com, which hosts a dataset of diverse and reputable polls taken across the country. Using this data, we (INSERT WRANGLING DONE TO DATASHEET). Then we constructed a generalized linear model, predicting (INSERT PREDICTOR), using (INSERT PREDICTORS) as predictors. The results from the model show (INSERT RESULTS FROM MODEL)

Section 2 will outline the source of this data. Section 3 covers the model and its parameters. Section 4 is where discussion will be made about the models predictions and how realistically they line up with current affairs. Finally, section 4 discusses any weakness and limitations that can be considered for another report.

2 Data

The dataset provided by fivethirtyeight contains over 15000 observations across 50 variables. Each observation is an individual poll done on the topic of the 2024 US presidency.

3 Measurement

The dataset comes from fivethirtyeight.com, a polling analysis website originally run by Nate Silvers, who is no longer affiliated with the website. Fivethirtyeight takes polls conducted on a particular topic (in this case the 2024 US presidency), and aggregates them in one dataset. Given the diverse array of polls conducted, there is the possibility of misleading polls and faulty data making its way into the aggregated dataset. This impacts the usefulness of models trained on this data, and any analysis that comes out of either the model, the dataset, or both. To combat this issue, fivethirtyeight implements its own quality control on the data being aggregated. The following breakdown of checks fivethirtyeight uses comes courtesy of their FAQ and methodology article, which is up to date as of September 24 2024 (Radcliffe and Morris 2023).

Fivethirtyeight routinely checks up on pollsters as they produce reports, whether they have an established track record, or if they are just new to data presentation. There are 2 types of checks fivethirtyeight uses: Methodology and Ethics.

Methodology refers to the scientific rigor exhibited by the pollster, when conducting the poll, and in the presentation of the results. Fivethirtyeight requires each poll to have or easily obtain the following:

- Pollster Identity
- Survey Dates
- Population Sample (Size and other distinct attributes)

- Polling Method (How was the poll conducted?)
- Sponsors (Identity and Amount)

Even if these conditions are met, fivethirtyeight may refuse to include certain polls in their data, due to flaws in methodology. Some examples include:

- Polls with inappropriate sample for target population
- Polls based upon predictive models
- Polls that implement recontact with participants (potential for bias to creep into the sample)
- Polls done by amateur/non-professional pollsters
- Polls with an “informed ballot” (when information about a candidate is given to the participant before they are asked who they would vote for)

The second check fivethirtyeight implements is ethical standard, adapted from the American Association for Public Opinion Research’s Code of Professional Ethics and Practices . Fivethirtyeight may refuse to include polls that:

- Fabricate or falsify data
- Are associated with the betting industry
- Refuse to disclose their methodology, either in their findings, or after being contacted for clarification
- Misrepresent their true purpose (being part of campaign analysis or done for a particular party or candidate)
- Disclose errors that come about from their work, and rectify them as best as possible
- Utilize methods that will give misleading results

Fivethirtyeight will conduct an initial check on pollsters, seeing if their methodology and ethics are publicly available, and are in agreement with fivethirtyeight’s checks. If unclear, fivethirtyeight will do a more thorough investigation of the pollster, and ask the organizers questions about their methodology and ethics if none were visible in the public report. If there are any violations in ethics, the severity on consequences will vary on what the offence is. If there is evidence of falsifying data or engaging in betting, the pollster is blacklisted from their data, any polls that come out of this pollster are extremely dubious and not trustworthy to build a model or report on. Otherwise, pollsters are given a chance to fix ethical issues presented by fivethirtyeight. Failure to do so results in the pollster and any of its polls being taken off the dataset, but if the pollster shows it has updated its ethics policies and applied them to recent reports, they may be allowed back in.

Partisan polls are a small subset of polls that are made with the backing or organizations affiliated with one or more political organizations. Fivethirtyeight will allow these polls into their data, but they will be marked as “Partisan”, and extra checks are put in place to ensure the polling data hasn’t been influenced unethically by the pollster or it’s affiliated organizations.

4 Model

The model contains 4 parameters. They are pollster, state, sponsors, and methodology.

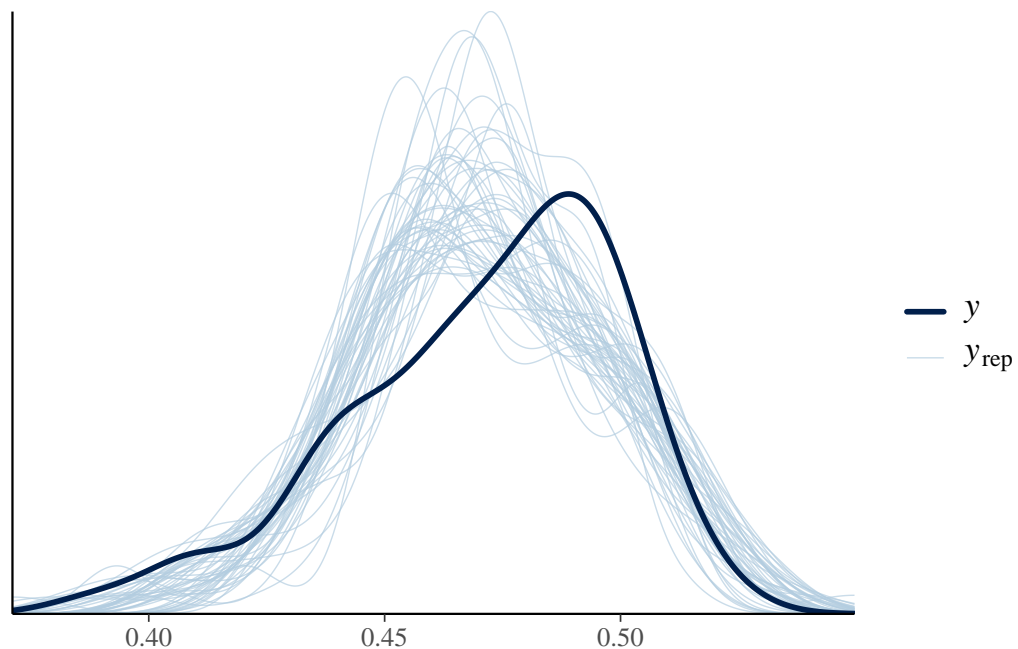
Pollster is the operator of the observed poll. Pollster provides a relatively easy way to identify trends in certain polls, depending on who is conducting them.

State tells us where the poll was conducted. Across the United States, there are states that traditionally vote Democrat (New York, California), some that traditionally vote Republican (Texas, Florida), and ‘swing states’ (Michigan, Pennsylvania). Swing states in particular are the most interesting states to watch as votes from these states tend to win a candidate the US presidency. Knowing where the polls have been taken also lets us understand if their choice was one bound by the state’s culture.

Sponsors are organizations that fund or otherwise assist the pollster in operating the poll. Sponsors carry the risk of bias with them, potentially influencing the pollster, and thus the polls to skew in a direction the sponsor favors preferably. Knowing who these sponsors are potentially lets us make informed predictions on the outcome of polls.

Methodology is the way the poll was conducted. Sometimes the composition of the poll influences people’s votes, so this is an interesting parameter to track. This also lets us see if certain methods yield better or worse predictions.

5 Results



Model Info:

```
function:      stan_glmer
family:        binomial [logit]
formula:       cbind(num_harris, sample_size - num_harris) ~ (1 | pollster) +
               (1 | state) + (1 | sponsors) + (1 | methodology)
algorithm:     sampling
sample:        4000 (posterior sample size)
priors:        see help('prior_summary')
observations:  63
groups:        sponsors (11), state (9), pollster (4), methodology (2)
```

Estimates:

	mean	sd
(Intercept)	-0.1	0.3
b[(Intercept) sponsors:Carolina_Forward]	0.0	0.1
b[(Intercept) sponsors:CATO_Institute]	0.0	0.1
b[(Intercept) sponsors:CBS_News]	0.2	0.0
b[(Intercept) sponsors:Economist]	0.0	0.0
b[(Intercept) sponsors:Institute_for_Global_Affairs]	-0.1	0.1

b[(Intercept) sponsors:Jacobin]	0.0	0.1
b[(Intercept) sponsors:Saint_Louis_University]	-0.1	0.1
b[(Intercept) sponsors:The_Philadelphia_Inquirer]	0.0	0.1
b[(Intercept) sponsors:The_Times_of_London_ _SAY24]	-0.1	0.1
b[(Intercept) sponsors:University_of_Texas]	0.0	0.1
b[(Intercept) sponsors:Yahoo_News]	0.0	0.0
b[(Intercept) state:Arizona]	0.0	0.0
b[(Intercept) state:Georgia]	0.0	0.0
b[(Intercept) state:Michigan]	0.0	0.0
b[(Intercept) state:Missouri]	0.0	0.0
b[(Intercept) state:National]	0.0	0.0
b[(Intercept) state:North_Carolina]	0.0	0.0
b[(Intercept) state:Pennsylvania]	0.0	0.0
b[(Intercept) state:Texas]	0.0	0.0
b[(Intercept) state:Wisconsin]	0.0	0.0
b[(Intercept) pollster:Siena/NYT]	0.0	0.1
b[(Intercept) pollster:YouGov]	0.0	0.1
b[(Intercept) pollster:YouGov_Blue]	0.0	0.1
b[(Intercept) pollster:YouGov/Center_for_Working_Class_Politics]	0.0	0.1
b[(Intercept) methodology:Live_Phone]	0.0	0.3
b[(Intercept) methodology:Online_Panel]	-0.1	0.3
Sigma[sponsors:(Intercept),(Intercept)]	0.0	0.0
Sigma[state:(Intercept),(Intercept)]	0.0	0.0
Sigma[pollster:(Intercept),(Intercept)]	0.0	0.1
Sigma[methodology:(Intercept),(Intercept)]	0.2	0.5
	10%	50%
(Intercept)	-0.4	-0.1
b[(Intercept) sponsors:Carolina_Forward]	-0.1	0.0
b[(Intercept) sponsors:CATO_Institute]	-0.1	0.0
b[(Intercept) sponsors:CBS_News]	0.1	0.2
b[(Intercept) sponsors:Economist]	0.0	0.0
b[(Intercept) sponsors:Institute_for_Global_Affairs]	-0.1	-0.1
b[(Intercept) sponsors:Jacobin]	-0.1	0.0
b[(Intercept) sponsors:Saint_Louis_University]	-0.2	-0.1
b[(Intercept) sponsors:The_Philadelphia_Inquirer]	-0.1	0.0
b[(Intercept) sponsors:The_Times_of_London_ _SAY24]	-0.1	-0.1
b[(Intercept) sponsors:University_of_Texas]	-0.1	0.0
b[(Intercept) sponsors:Yahoo_News]	0.0	0.0
b[(Intercept) state:Arizona]	-0.1	0.0
b[(Intercept) state:Georgia]	0.0	0.0
b[(Intercept) state:Michigan]	0.0	0.0
b[(Intercept) state:Missouri]	-0.1	0.0
b[(Intercept) state:National]	0.0	0.0

b[(Intercept) state:North_Carolina]	0.0	0.0
b[(Intercept) state:Pennsylvania]	0.0	0.0
b[(Intercept) state:Texas]	0.0	0.0
b[(Intercept) state:Wisconsin]	0.0	0.0
b[(Intercept) pollster:Siena/NYT]	-0.1	0.0
b[(Intercept) pollster:YouGov]	-0.1	0.0
b[(Intercept) pollster:YouGov_Blue]	-0.1	0.0
b[(Intercept) pollster:YouGov/Center_for_Working_Class_Politics]	-0.1	0.0
b[(Intercept) methodology:Live_Phone]	-0.3	0.0
b[(Intercept) methodology:Online_Panel]	-0.4	0.0
Sigma[sponsors:(Intercept),(Intercept)]	0.0	0.0
Sigma[state:(Intercept),(Intercept)]	0.0	0.0
Sigma[pollster:(Intercept),(Intercept)]	0.0	0.0
Sigma[methodology:(Intercept),(Intercept)]	0.0	0.1
	90%	
(Intercept)	0.2	
b[(Intercept) sponsors:Carolina_Forward]	0.1	
b[(Intercept) sponsors:CATO_Institute]	0.0	
b[(Intercept) sponsors:CBS_News]	0.2	
b[(Intercept) sponsors:Economist]	0.1	
b[(Intercept) sponsors:Institute_for_Global_Affairs]	0.0	
b[(Intercept) sponsors:Jacobin]	0.1	
b[(Intercept) sponsors:Saint_Louis_University]	0.0	
b[(Intercept) sponsors:The_Philadelphia_Inquirer]	0.1	
b[(Intercept) sponsors:The_Times_of_London_ _SAY24]	0.0	
b[(Intercept) sponsors:University_of_Texas]	0.0	
b[(Intercept) sponsors:Yahoo_News]	0.1	
b[(Intercept) state:Arizona]	0.0	
b[(Intercept) state:Georgia]	0.0	
b[(Intercept) state:Michigan]	0.0	
b[(Intercept) state:Missouri]	0.0	
b[(Intercept) state:National]	0.1	
b[(Intercept) state:North_Carolina]	0.0	
b[(Intercept) state:Pennsylvania]	0.0	
b[(Intercept) state:Texas]	0.0	
b[(Intercept) state:Wisconsin]	0.1	
b[(Intercept) pollster:Siena/NYT]	0.2	
b[(Intercept) pollster:YouGov]	0.1	
b[(Intercept) pollster:YouGov_Blue]	0.1	
b[(Intercept) pollster:YouGov/Center_for_Working_Class_Politics]	0.1	
b[(Intercept) methodology:Live_Phone]	0.4	
b[(Intercept) methodology:Online_Panel]	0.2	
Sigma[sponsors:(Intercept),(Intercept)]	0.0	

Sigma[state:(Intercept),(Intercept)]	0.0
Sigma[pollster:(Intercept),(Intercept)]	0.1
Sigma[methodology:(Intercept),(Intercept)]	0.7

Fit Diagnostics:

	mean	sd	10%	50%	90%
mean_PPD	630.2	3.3	626.0	630.2	634.3

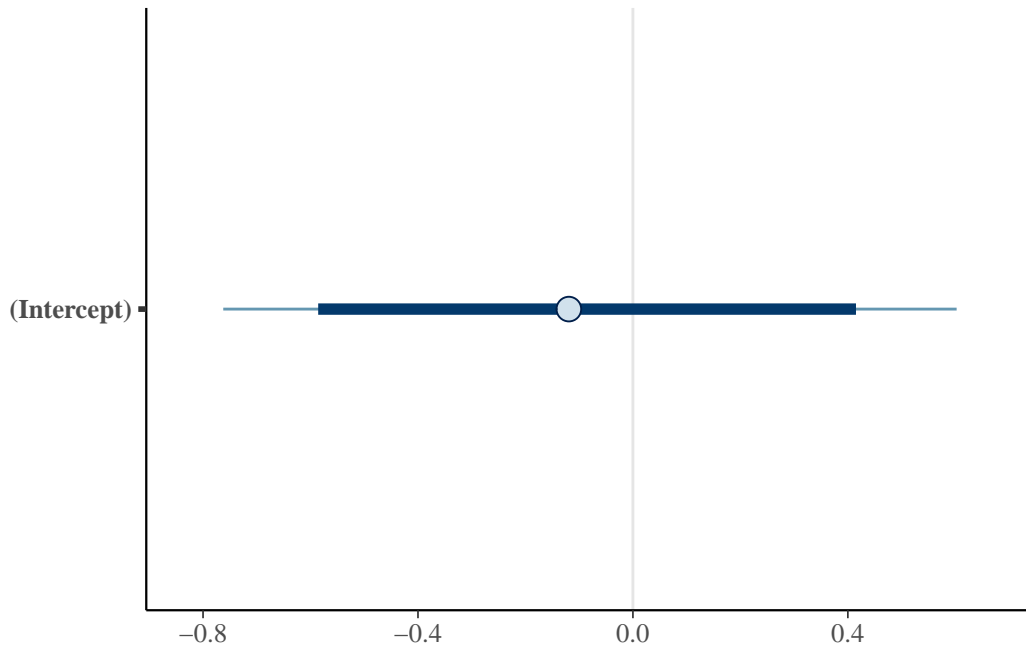
The mean_ppd is the sample average posterior predictive distribution of the outcome variable

MCMC diagnostics

	mcse	Rhat
(Intercept)	0.0	1.0
b[(Intercept) sponsors:Carolina_Forward]	0.0	1.0
b[(Intercept) sponsors:CATO_Institute]	0.0	1.0
b[(Intercept) sponsors:CBS_News]	0.0	1.0
b[(Intercept) sponsors:Economist]	0.0	1.0
b[(Intercept) sponsors:Institute_for_Global_Affairs]	0.0	1.0
b[(Intercept) sponsors:Jacobin]	0.0	1.0
b[(Intercept) sponsors:Saint_Louis_University]	0.0	1.0
b[(Intercept) sponsors:The_Philadelphia_Inquirer]	0.0	1.0
b[(Intercept) sponsors:The_Times_of_London_ _SAY24]	0.0	1.0
b[(Intercept) sponsors:University_of_Texas]	0.0	1.0
b[(Intercept) sponsors:Yahoo_News]	0.0	1.0
b[(Intercept) state:Arizona]	0.0	1.0
b[(Intercept) state:Georgia]	0.0	1.0
b[(Intercept) state:Michigan]	0.0	1.0
b[(Intercept) state:Missouri]	0.0	1.0
b[(Intercept) state:National]	0.0	1.0
b[(Intercept) state:North_Carolina]	0.0	1.0
b[(Intercept) state:Pennsylvania]	0.0	1.0
b[(Intercept) state:Texas]	0.0	1.0
b[(Intercept) state:Wisconsin]	0.0	1.0
b[(Intercept) pollster:Siena/NYT]	0.0	1.0
b[(Intercept) pollster:YouGov]	0.0	1.0
b[(Intercept) pollster:YouGov_Blue]	0.0	1.0
b[(Intercept) pollster:YouGov/Center_for_Working_Class_Politics]	0.0	1.0
b[(Intercept) methodology:Live_Phone]	0.0	1.0
b[(Intercept) methodology:Online_Panel]	0.0	1.0
Sigma[sponsors:(Intercept),(Intercept)]	0.0	1.0
Sigma[state:(Intercept),(Intercept)]	0.0	1.0
Sigma[pollster:(Intercept),(Intercept)]	0.0	1.0
Sigma[methodology:(Intercept),(Intercept)]	0.0	1.0

mean_PPD	0.1	1.0
log-posterior	0.2	1.0
	n_eff	
(Intercept)		1057
b[(Intercept) sponsors:Carolina_Forward]		3675
b[(Intercept) sponsors:CATO_Institute]		2260
b[(Intercept) sponsors:CBS_News]		1801
b[(Intercept) sponsors:Economist]		1779
b[(Intercept) sponsors:Institute_for_Global_Affairs]		2664
b[(Intercept) sponsors:Jacobin]		3099
b[(Intercept) sponsors:Saint_Louis_University]		2119
b[(Intercept) sponsors:The_Philadelphia_Inquirer]		3338
b[(Intercept) sponsors:The_Times_of_London_ _SAY24]		3061
b[(Intercept) sponsors:University_of_Texas]		2310
b[(Intercept) sponsors:Yahoo_News]		1793
b[(Intercept) state:Arizona]		3497
b[(Intercept) state:Georgia]		4801
b[(Intercept) state:Michigan]		3925
b[(Intercept) state:Missouri]		2186
b[(Intercept) state:National]		2422
b[(Intercept) state:North_Carolina]		3719
b[(Intercept) state:Pennsylvania]		3689
b[(Intercept) state:Texas]		3501
b[(Intercept) state:Wisconsin]		2260
b[(Intercept) pollster:Siena/NYT]		2528
b[(Intercept) pollster:YouGov]		1953
b[(Intercept) pollster:YouGov_Blue]		2369
b[(Intercept) pollster:YouGov/Center_for_Working_Class_Politics]		2291
b[(Intercept) methodology:Live_Phone]		1177
b[(Intercept) methodology:Online_Panel]		1128
Sigma[sponsors:(Intercept),(Intercept)]		1617
Sigma[state:(Intercept),(Intercept)]		1438
Sigma[pollster:(Intercept),(Intercept)]		1854
Sigma[methodology:(Intercept),(Intercept)]		1315
mean_PPD		3720
log-posterior		949

For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective



Model Info:

```

function:      stan_glm
family:        gaussian [identity]
formula:       pct ~ ns(end_date_num, df = 5) + pollster + state + sponsors +
               methodology
algorithm:      sampling
sample:        4000 (posterior sample size)
priors:        see help('prior_summary')
observations:  69
predictors:    28

```

Estimates:

	mean	sd	10%	50%	90%
(Intercept)	43.5	4.5	37.7	43.5	49.3
ns(end_date_num, df = 5)1	0.5	1.7	-1.6	0.5	2.7
ns(end_date_num, df = 5)2	6.0	1.7	3.8	6.0	8.1
ns(end_date_num, df = 5)3	2.6	1.3	0.9	2.6	4.2
ns(end_date_num, df = 5)4	5.5	2.5	2.3	5.5	8.8
ns(end_date_num, df = 5)5	2.6	1.3	0.9	2.6	4.2
pollsterYouGov	-0.8	3.5	-5.4	-0.8	3.7
pollsterYouGov Blue	0.8	4.1	-4.5	0.9	6.0
pollsterYouGov/Center for Working Class Politics	-0.5	4.0	-5.7	-0.5	4.5

stateGeorgia	-0.8	2.3	-3.7	-0.8	2.1
stateMichigan	1.6	1.9	-0.8	1.6	3.9
stateMissouri	-1.2	3.8	-6.1	-1.1	3.7
stateNational	2.7	1.7	0.6	2.7	4.8
stateNorth Carolina	0.8	4.1	-4.5	0.9	5.9
statePennsylvania	1.4	1.9	-1.0	1.4	3.7
stateTexas	0.2	3.7	-4.5	0.2	4.9
stateWisconsin	3.9	1.9	1.5	3.9	6.4
sponsorsCATO Institute	0.6	2.2	-2.2	0.7	3.4
sponsorsCBS News	3.2	1.9	0.7	3.3	5.8
sponsorsEconomist	-0.6	2.0	-3.1	-0.6	2.0
sponsorsInstitute for Global Affairs	-1.9	2.4	-5.0	-1.9	1.3
sponsorsJacobin	-0.6	4.0	-5.6	-0.6	4.5
sponsorsSaint Louis University	-1.2	3.9	-6.2	-1.3	3.8
sponsorsThe Philadelphia Inquirer	0.6	4.0	-4.5	0.5	5.7
sponsorsThe Times of London SAY24	-0.7	2.5	-3.9	-0.7	2.5
sponsorsUniversity of Texas	0.2	3.7	-4.5	0.1	4.9
sponsorsYahoo News	-0.9	2.0	-3.4	-0.9	1.7
methodologyOnline Panel	-0.6	4.0	-5.8	-0.6	4.5
sigma	1.8	0.2	1.6	1.8	2.1

Fit Diagnostics:

	mean	sd	10%	50%	90%
mean_PPD	47.2	0.3	46.8	47.2	47.6

The mean_ppd is the sample average posterior predictive distribution of the outcome variable

MCMC diagnostics

	mcse	Rhat	n_eff
(Intercept)	0.1	1.0	3911
ns(end_date_num, df = 5)1	0.0	1.0	4168
ns(end_date_num, df = 5)2	0.0	1.0	4307
ns(end_date_num, df = 5)3	0.0	1.0	5617
ns(end_date_num, df = 5)4	0.0	1.0	4275
ns(end_date_num, df = 5)5	0.0	1.0	5032
pollsterYouGov	0.1	1.0	3431
pollsterYouGov Blue	0.1	1.0	4413
pollsterYouGov/Center for Working Class Politics	0.1	1.0	5651
stateGeorgia	0.0	1.0	3565
stateMichigan	0.0	1.0	2129
stateMissouri	0.1	1.0	4195
stateNational	0.0	1.0	2519
stateNorth Carolina	0.1	1.0	4535

statePennsylvania	0.0	1.0	2077
stateTexas	0.1	1.0	4581
stateWisconsin	0.0	1.0	2138
sponsorsCATO Institute	0.1	1.0	1876
sponsorsCBS News	0.0	1.0	1512
sponsorsEconomist	0.1	1.0	1529
sponsorsInstitute for Global Affairs	0.0	1.0	2379
sponsorsJacobin	0.1	1.0	5423
sponsorsSaint Louis University	0.1	1.0	4066
sponsorsThe Philadelphia Inquirer	0.1	1.0	4921
sponsorsThe Times of London SAY24	0.1	1.0	2165
sponsorsUniversity of Texas	0.1	1.0	4140
sponsorsYahoo News	0.0	1.0	1631
methodologyOnline Panel	0.1	1.0	4370
sigma	0.0	1.0	3318
mean_PPD	0.0	1.0	4790
log-posterior	0.1	1.0	1278

For each parameter, mcse is Monte Carlo standard error, n_eff is a crude measure of effective

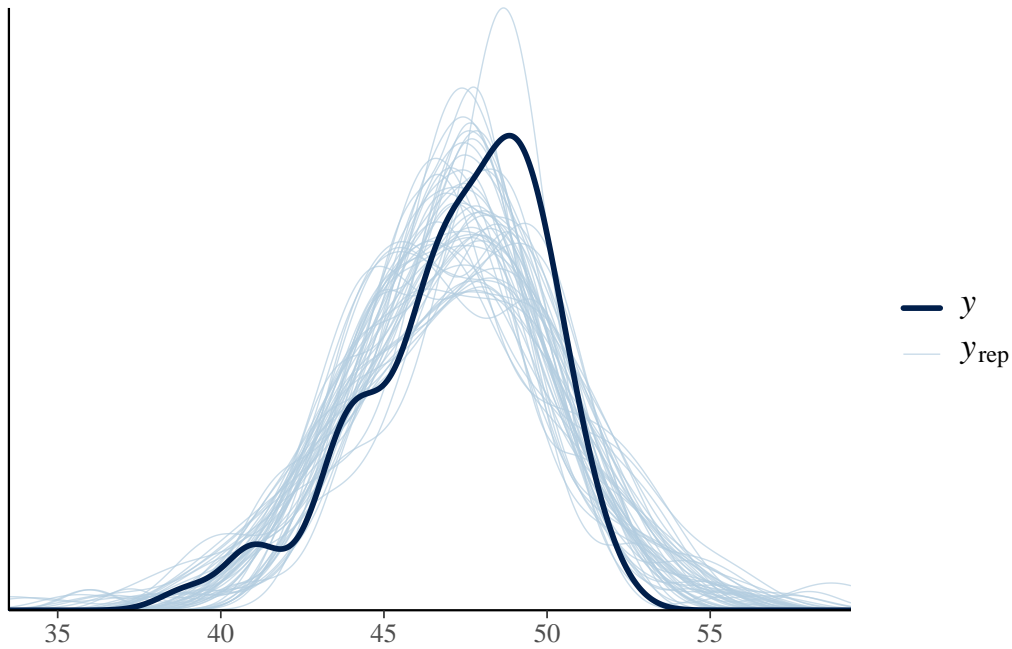


Figure 1: Street tree count of 10 most populous tree species in Toronto up to September 2024

6 Discussion

6.1 Trends from pollsters

6.2 Trends from states

6.3 Trends from sponsors

6.4 Trends from methodology

6.5 Weaknesses and next steps

7 Appendix

7.1 Pollster Analysis

7.2 Poll Blueprint

References

- Lopez, Ashley. 2024. “The race is so close. here’s what happens if there’s a 269-269 electoral college tie.” *NPR*. NPR. <https://www.npr.org/2024/10/21/nx-s1-5150099/electoral-college-tie-explainer>.
- Radcliffe, Mary, and G Elliot Morris. 2023. “538’s polls policy and FAQs.” *ABC News*. ABC News Network. <https://abcnews.go.com/538/538s-polls-policy-faqs/story?id=104489193>.