

# Predicting the 2024 US Presidential Election: A Polling-Based Forecast\*

My subtitle if needed

Tianrui Fu                      Yiyue Deng

October 26, 2024

This paper forecasts the 2024 US Presidential election outcome using polling data from [insert pollster name]. By applying simple and multiple linear regression models, we analyze the effect of polling factors, including sample size, poll score, and transparency, on support percentages for key candidates. Our findings suggest significant relationships between these variables, providing an evidence-based approach to predicting election outcomes. We further explore methodological strengths and weaknesses and propose an ideal polling survey methodology. This work highlights the potential for data-driven insights into political forecasting.

## Table of contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>Data</b>	<b>2</b>
2.1	Overview . . . . .	2
2.2	Measurement . . . . .	2
2.3	Outcome variables . . . . .	2
2.4	Predictor variables . . . . .	3
<b>3</b>	<b>Model</b>	<b>5</b>
3.1	Model set-up . . . . .	5
3.1.1	Model justification . . . . .	6
<b>4</b>	<b>Results</b>	<b>6</b>

---

\*Code and data are available at: [https://github.com/RohanAlexander/starter\\_folder](https://github.com/RohanAlexander/starter_folder).

5 Discussion 9
5.1 First discussion point 9
5.2 Second discussion point 9
5.3 Third discussion point 9
5.4 Weaknesses and next steps 9
Appendix 10
A Additional data details 10
B Model details 10
B.1 Posterior predictive check 10
B.2 Diagnostics 10
References 11

1 Introduction

Overview paragraph
Estimand paragraph
Results paragraph
Why it matters paragraph
Telegraphing paragraph: The remainder of this paper is structured as follows. Section 5....

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023).... Our data (Toronto Shelter & Support Services 2024).... Following Alexander (2023), we consider...

Table 1: The example of chosen variable dataset

Table with 6 columns: pollster, pollscore, numeric\_grade, transparency\_score, end\_date, pct. Rows include Marist, Emerson, Beacon/Shaw, and AtlasIntel.

Quinnipiac	-0.5	2.8	9	2024-09-22	48.0
Marquette Law School	-1.1	3.0	10	2024-08-01	47.0
YouGov	-1.1	3.0	9	2024-08-31	49.0
Emerson	-1.1	2.9	7	2024-09-28	49.9
Marist	-0.9	2.9	7	2024-09-05	48.0
MassINC Polling Group	-0.8	2.8	7	2024-09-18	40.0

## 2.2 Measurement

Some paragraphs about how we go from a phenomena in the world to an entry in the dataset.

## 2.3 Outcome variables

Add graphs, tables and text. Use sub-sub-headings for each outcome variable or update the subheading to be singular.

Some of our data is of penguins (?@fig-bills), from[].

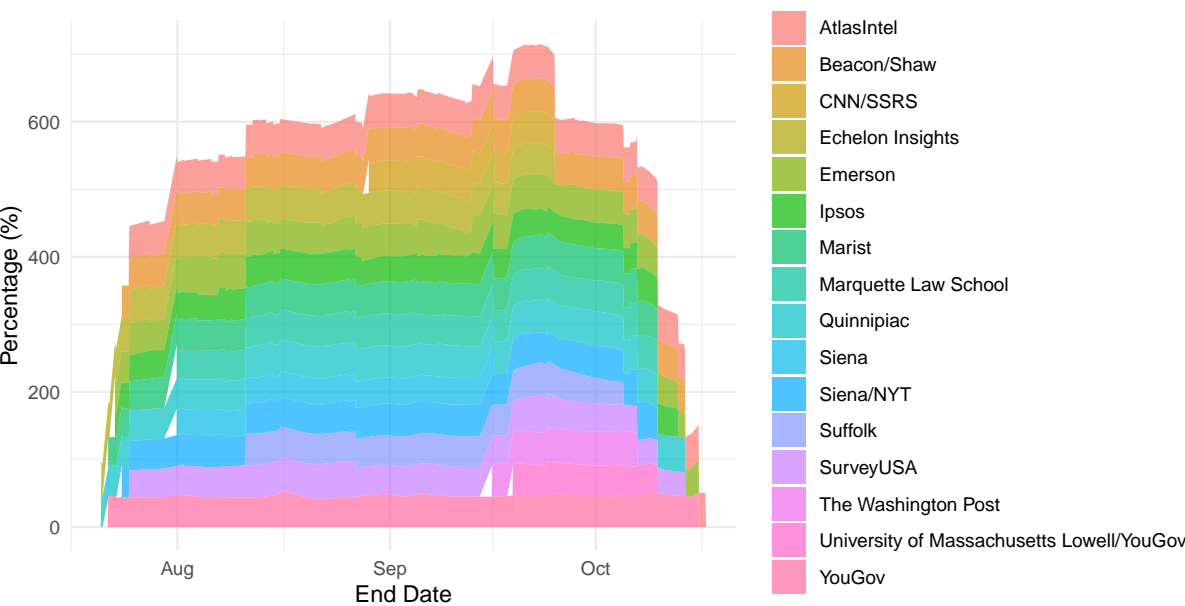


Figure 1: Percentage of Each Pollster Over Time

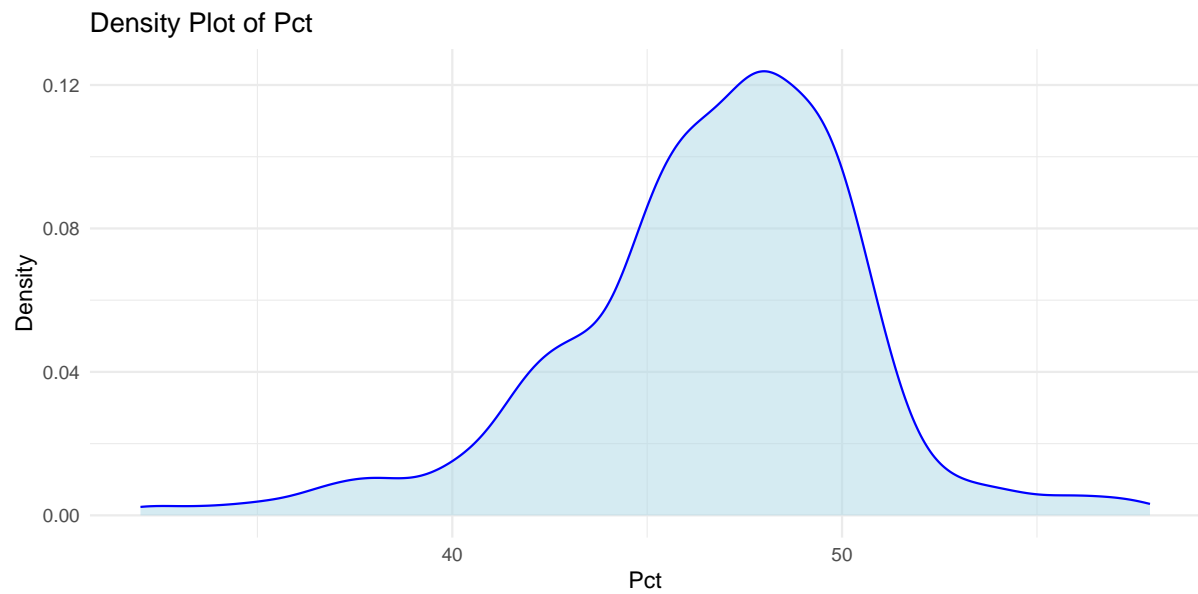


Figure 2: Density plot of pct

## 2.4 Predictor variables

Add graphs, tables and text.

Use sub-sub-headings for each outcome variable and feel free to combine a few into one if they go together naturally.

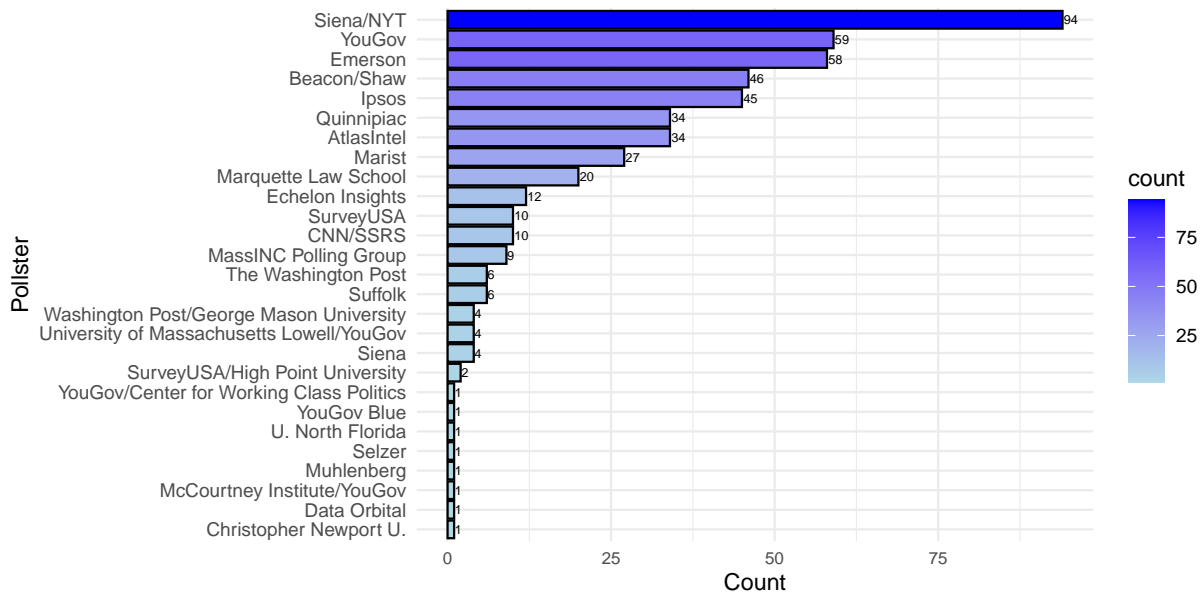


Figure 3: Frequency of each pollster

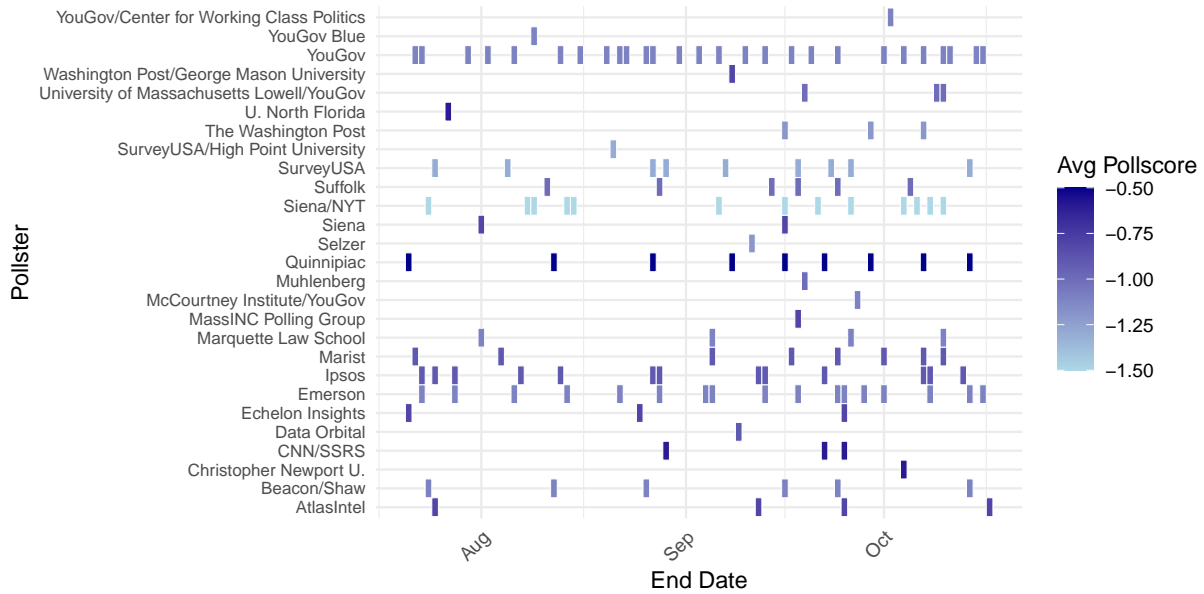


Figure 4: Average Pollscore by Pollster and End Date

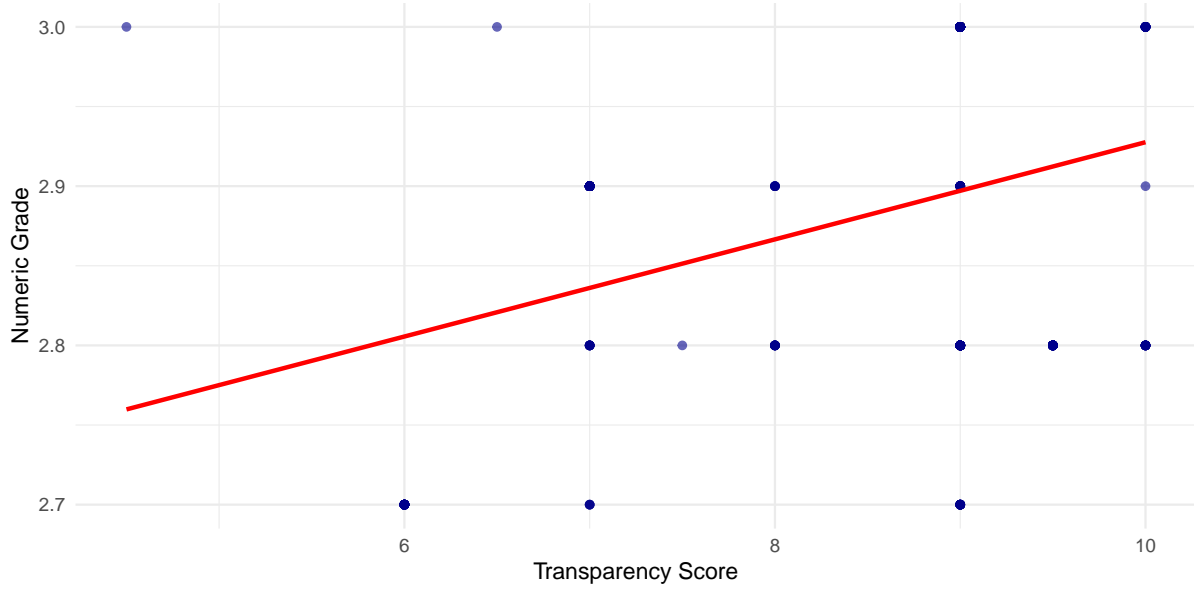


Figure 5: Scatter Plot of transparency score and numeric grade

### 3 Model

The goal of our modelling strategy is twofold. Firstly,...

Here we briefly describe the Bayesian analysis model used to investigate... Background details and diagnostics are included in [Appendix B](#).

#### 3.1 Model set-up

Define  $y_i$  as the percentage. Then  $\beta_i$  represents the numeric grade,  $\gamma_i$  represents the transparency score, and  $\delta_i$  represents the pollscore.

$$y_i | \mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \quad (1)$$

$$\mu_i = \alpha + \beta_1 \cdot \text{numeric\_grade}_i + \beta_2 \cdot \text{transparency\_score}_i + \beta_3 \cdot \text{pollscore}_i \quad (2)$$

$$+ \sum_{j=1}^N \gamma_j \cdot \text{pollster}_j + \delta_i \cdot \text{end\_date}_i \quad (3)$$

$$\alpha \sim \text{Normal}(50, 10) \quad (4)$$

$$\beta_1, \beta_2, \beta_3 \sim \text{Normal}(0, 5) \quad (5)$$

$$\gamma_j \sim \text{Normal}(0, 5) \quad (6)$$

$$\sigma \sim \text{Exponential}(1) \quad (7)$$

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

### 3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance  $\theta$ .

## 4 Results

Our results are summarized in Table [2](#).

Table 2: Explanatory models of flight time based on wing width and wing length

	Model Summary	
	Model by glm	Model by bayes
(Intercept)	-524.927*** [-773.481, -276.373]	-590.212 [-814.191, -354.733]
numeric_grade	-14.386 [-43.487, 14.716]	-3.356 [-11.443, 5.059]
transparency_score	0.903 [-0.348, 2.153]	-0.164 [-0.881, 0.569]
pollscore	-3.028 [-32.532, 26.476]	1.121 [-5.151, 7.339]
Beacon/Shaw	-2.833 [-12.913, 7.246]	0.976 [-1.509, 3.389]
Christopher Newport U.	-9.186* [-18.242, -0.130]	-5.294 [-10.778, 0.269]
CNN/SSRS	-4.208 [-10.838, 2.423]	-1.343 [-4.986, 2.341]
Data Orbital	-3.602 [-12.297, 5.093]	-0.426 [-6.114, 5.093]
Echelon Insights	-3.313 [-7.738, 1.113]	0.389 [-2.466, 3.257]
Emerson	1.010 [-6.052, 8.072]	1.561 [-0.559, 3.788]
Ipsos	-7.483** [-13.082, -1.883]	-3.927 [-6.355, -1.584]
Marist	0.162 [-4.587, 4.911]	1.111 [-0.987, 3.216]
Marquette Law School	-3.213 [-12.696, 6.270]	-0.561 [-3.603, 2.499]
MassINC Polling Group	-7.280*** [-10.601, -3.958]	-6.566 [-8.857, -4.172]
McCourtney Institute/YouGov	-1.354 [-10.616, 7.908]	-1.745 [-7.385, 3.817]
Muhlenberg	-2.040 [-10.915, 6.834]	0.196 [-5.405, 5.867]
Quinnipiac	-1.513 [-10.660, 7.633]	-0.189 [-3.441, 3.117]
Selzer	-2.952 [-15.951, 10.046]	-0.220 [-5.949, 5.492]
Siena	-9.684*** [-13.295, -6.072]	-7.344 [-10.509, -4.019]
Siena/NYT	-2.174 [-20.461, 16.113]	0.976 [-3.002, 5.133]
Suffolk	-4.753 [-10.985, 1.479]	-3.290 [-6.325, -0.369]
SurveyUSA	-5.810 [-20.672, 9.053]	-2.163 [-5.651, 1.268]
SurveyUSA/High Point University	-5.876 [-23.153, 11.401]	-0.224 [-5.003, 4.990]
The Washington Post	0.576 [-10.255, 11.407]	2.315 [-1.070, 5.912]
University of Massachusetts Lowell/YouGov	-6.065 [-13.331, 1.201]	-3.502 [-7.054, 0.165]
Washington Post/George Mason University	-8.943*** [-14.000, -3.886]	-4.783 [-8.412, -1.163]
YouGov	-2.338 [-10.874, 6.199]	-0.764 [-3.385, 1.935]
YouGov Blue	-1.128 [-11.743, 9.488]	0.335 [-5.326, 5.727]
End Date	0.030*** [0.019, 0.042]	0.032 [0.021, 0.044]
U. North Florida		1.990 [-3.834, 7.845]
YouGov/Center for Working Class Politics		-2.360 [-8.276, 3.399]
Num.Obs.	492	492
R2	0.332	0.318
R2 Adj.		0.258
AIC	2575.9	
BIC	2701.8	
Log.Lik.	-1257.932	-1264.371
ELPD		-1289.0
ELPD s.e.		25.3
LOOIC		2577.9
LOOIC s.e.		50.6
WAIC		2575.6
RMSE	3.12	3.38

+ p \num{&lt; 0.1}, \* p \num{&lt; 0.05}, \*\* p \num{&lt; 0.01}, \*\*\* p \num{&lt; 0.001}

\* This table shows the regression models with custom variable names.



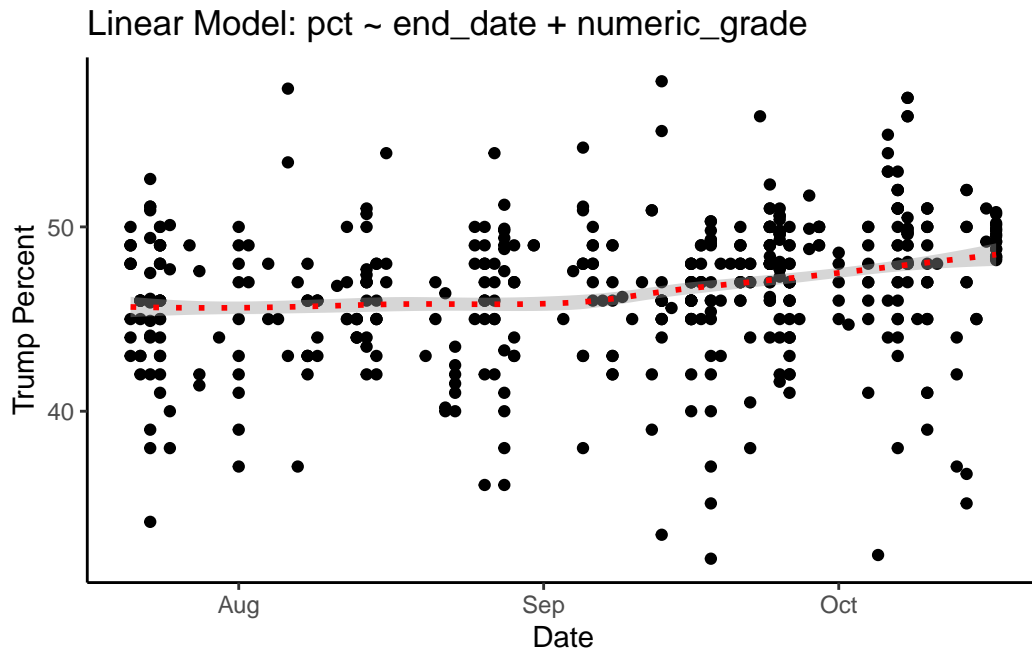


Figure 6: ddd

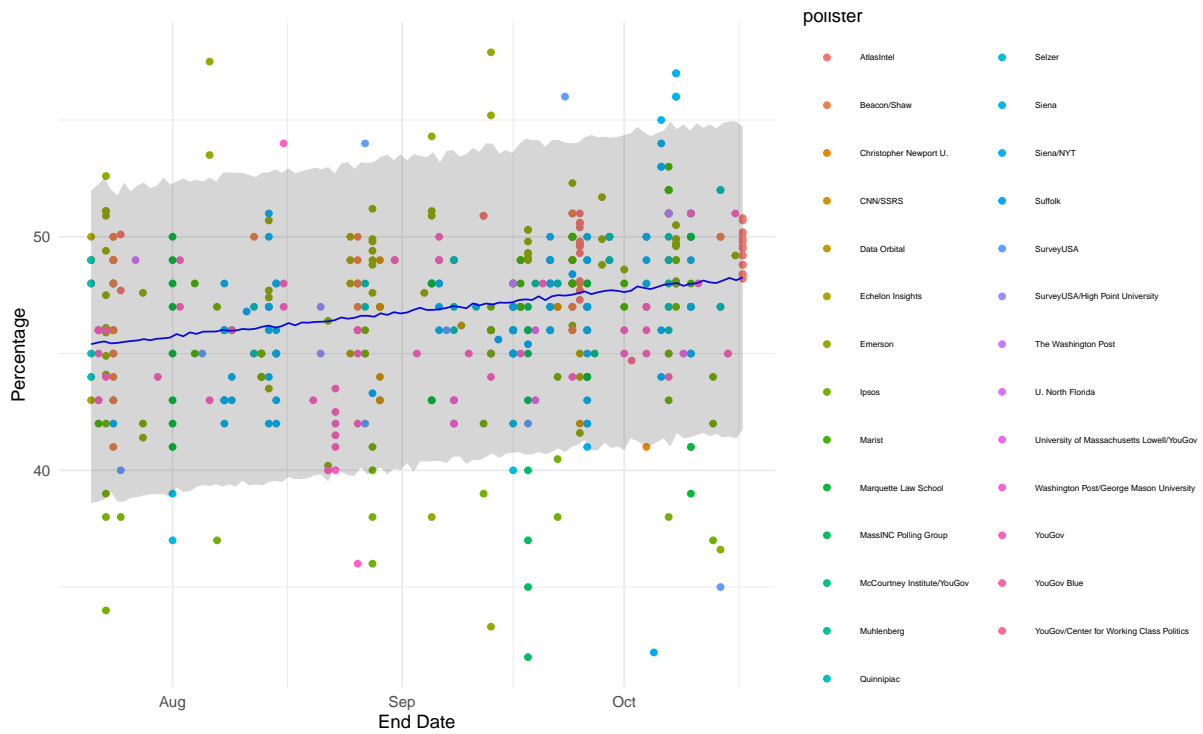


Figure 7: Poll Percentage over Time with Bayesian Fit

## **5 Discussion**

### **5.1 First discussion point**

If my paper were 10 pages, then should be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

### **5.2 Second discussion point**

Please don't use these as sub-heading labels - change them to be what your point actually is.

### **5.3 Third discussion point**

### **5.4 Weaknesses and next steps**

Weaknesses and next steps should also be included.

## Appendix

### A Additional data details

### B Model details

#### B.1 Posterior predictive check

In `?@fig-ppcheckandposteriorvsprior-1` we implement a posterior predictive check. This shows...

In `?@fig-ppcheckandposteriorvsprior-2` we compare the posterior with the prior. This shows...

Examining how the model fits, and is affected  
by, the data

#### B.2 Diagnostics

`?@fig-stanareyouokay-1` is a trace plot. It shows... This suggests...

`?@fig-stanareyouokay-2` is a Rhat plot. It shows... This suggests...

Checking the convergence of the MCMC algo-  
rithm

## References

- Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. <https://tellingstorieswithdata.com/>.
- Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. “rstanarm: Bayesian applied regression modeling via Stan.” <https://mc-stan.org/rstanarm/>.
- R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Toronto Shelter & Support Services. 2024. *Deaths of Shelter Residents*. <https://open.toronto.ca/dataset/deaths-of-shelter-residents/>.