

Forecasting the 2024 US Presidential Election: A Poll-Based Approach*

My subtitle if needed

Xinxiang Gao Ariel Xing John Zhang

October 21, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

Table of contents

1	Introduction	2
1.1	Predictor variables	3
2	Model	6
2.1	Model set-up	6
2.1.1	Model justification	7
3	Results	7
4	Discussion	7
4.1	First discussion point	7
4.2	Second discussion point	7
4.3	Third discussion point	8
4.4	Weaknesses and next steps	8
	Appendix	9
A	Pollster Methodology Overview and Evaluation: YouGov	9
A.1	Population, Frame, and Sample	9
A.2	Sample Recruitment	9
A.3	Sampling Approach and Trade-offs	9
A.4	Handling Non-Response	10

*Code and data are available at: https://github.com/xgao28/election_forecast.

A.5	Questionnaire Design	10
A.6	Evaluation of YouGov’s Methodology	11
B	Appendix: Methodology and Survey Design for 2024 U.S. Presidential Election Forecast	11
B.1	Appendix A: Methodology	11
B.1.1	1. Sampling Approach	12
B.1.2	2. Recruitment Plan	12
B.1.3	3. Survey Implementation and Design	12
B.1.4	4. Data Validation	13
B.1.5	5. Poll Aggregation and Reporting	13
B.1.6	6. Budget Allocation	14
B.2	Appendix B: Survey Questions	14
B.2.1	Survey Questions	14
B.3	Appendix C: Google Forms Link	17
C	Model details	17
C.1	Posterior predictive check	17
C.2	Diagnostics	17
	References	18

1 Introduction

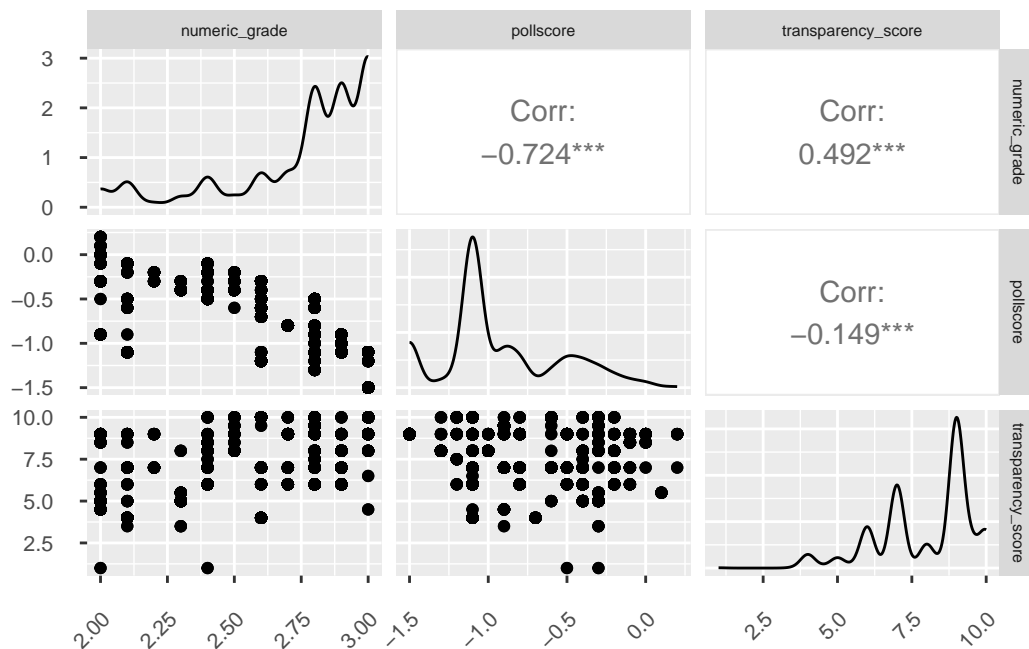
Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows.



1.1 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.

\$`Kamala Harris`

Call:

```
lm(formula = pct ~ pollscore + transparency_score + sample_size +
    state + days_towards_election, data = candidate_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-9.1312	-1.4025	0.1996	1.5883	8.9724

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	43.0745168	2.5017798	17.218	< 2e-16 ***
pollscore	0.1435381	0.3409898	0.421	0.67401
transparency_score	0.0512012	0.0888118	0.577	0.56458

sample_size	0.0018283	0.0004441	4.117	4.62e-05	***
stateArizona	2.0012881	2.5208440	0.794	0.42770	
stateCalifornia	11.2165303	2.8063596	3.997	7.58e-05	***
stateConnecticut	8.7751531	3.5022913	2.506	0.01260	*
stateFlorida	0.1941899	2.5920747	0.075	0.94032	
stateGeorgia	2.3781877	2.5270391	0.941	0.34720	
stateIndiana	-3.7895379	3.5026388	-1.082	0.27991	
stateIowa	-0.0135142	3.0237337	-0.004	0.99644	
stateMaine	8.3889379	2.7827236	3.015	0.00273	**
stateMaine CD-1	16.3981473	2.7789606	5.901	7.44e-09	***
stateMaine CD-2	2.8599431	2.7794326	1.029	0.30409	
stateMaryland	18.3676624	2.7441740	6.693	6.98e-11	***
stateMassachusetts	17.4647485	2.7880213	6.264	9.27e-10	***
stateMichigan	3.6999551	2.5211914	1.468	0.14298	
stateMinnesota	5.2694503	2.6867519	1.961	0.05051	.
stateMissouri	-1.2748379	2.8702743	-0.444	0.65716	
stateMontana	-4.7313116	2.6176165	-1.807	0.07140	.
stateNebraska	-5.8352926	3.0419293	-1.918	0.05575	.
stateNebraska CD-2	6.4530572	2.6652870	2.421	0.01589	*
stateNevada	2.8570906	2.5393374	1.125	0.26117	
stateNew Hampshire	4.1896566	2.6583662	1.576	0.11577	
stateNew Mexico	6.9754208	2.7667089	2.521	0.01206	*
stateNew York	8.3857846	2.7277860	3.074	0.00225	**
stateNorth Carolina	3.3573751	2.5239750	1.330	0.18417	
stateOhio	-0.5511072	2.6073271	-0.211	0.83270	
statePennsylvania	3.3633423	2.5131361	1.338	0.18152	
stateRhode Island	11.4457188	2.8809502	3.973	8.35e-05	***
stateSouth Carolina	-2.9907831	3.5347663	-0.846	0.39797	
stateTexas	-0.1259019	2.5941957	-0.049	0.96132	
stateVermont	24.1704929	3.0975724	7.803	4.82e-14	***
stateVirginia	3.5471294	2.5785614	1.376	0.16967	
stateWashington	4.8790596	3.0389423	1.606	0.10913	
stateWisconsin	4.6670943	2.5219969	1.851	0.06493	.
days_towards_election	-0.0115806	0.0015746	-7.355	1.01e-12	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.465 on 421 degrees of freedom

(241 observations deleted due to missingness)

Multiple R-squared: 0.7349, Adjusted R-squared: 0.7122

F-statistic: 32.42 on 36 and 421 DF, p-value: < 2.2e-16

\$`Donald Trump`

Call:

```
lm(formula = pct ~ pollscore + transparency_score + sample_size +  
    state + days_towards_election, data = candidate_data)
```

Residuals:

Min	1Q	Median	3Q	Max
-12.5097	-1.7824	0.2386	2.1316	8.0081

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	5.465e+01	1.865e+00	29.311	< 2e-16	***
pollscore	-7.098e-01	2.858e-01	-2.484	0.013167	*
transparency_score	-3.967e-01	7.264e-02	-5.461	6.03e-08	***
sample_size	8.102e-04	2.549e-04	3.179	0.001528	**
stateArizona	-5.003e+00	1.835e+00	-2.727	0.006511	**
stateArkansas	6.284e+00	3.599e+00	1.746	0.081093	.
stateCalifornia	-1.957e+01	1.942e+00	-10.076	< 2e-16	***
stateColorado	-1.290e+01	2.284e+00	-5.645	2.17e-08	***
stateConnecticut	-1.269e+01	2.842e+00	-4.463	9.02e-06	***
stateFlorida	-2.453e+00	1.909e+00	-1.285	0.199048	
stateGeorgia	-4.569e+00	1.835e+00	-2.490	0.012934	*
stateIdaho	4.140e+00	3.598e+00	1.150	0.250242	
stateIllinois	-1.340e+01	2.557e+00	-5.242	1.95e-07	***
stateIndiana	1.549e+00	2.541e+00	0.609	0.542435	
stateIowa	-3.188e+00	2.048e+00	-1.557	0.119836	
stateKansas	-1.426e-01	2.557e+00	-0.056	0.955541	
stateMaine	-1.074e+01	2.176e+00	-4.935	9.45e-07	***
stateMaine CD-1	-1.589e+01	2.314e+00	-6.868	1.16e-11	***
stateMaine CD-2	-4.067e+00	2.315e+00	-1.757	0.079246	.
stateMaryland	-2.015e+01	2.082e+00	-9.678	< 2e-16	***
stateMassachusetts	-2.257e+01	2.013e+00	-11.214	< 2e-16	***
stateMichigan	-6.110e+00	1.832e+00	-3.335	0.000885	***
stateMinnesota	-9.940e+00	1.953e+00	-5.088	4.34e-07	***
stateMissouri	1.301e+00	2.003e+00	0.650	0.516073	
stateMontana	2.565e+00	1.983e+00	1.293	0.196221	
stateNebraska	2.797e-01	2.543e+00	0.110	0.912456	
stateNebraska CD-2	-1.031e+01	2.158e+00	-4.776	2.07e-06	***
stateNevada	-5.217e+00	1.845e+00	-2.827	0.004794	**
stateNew Hampshire	-9.660e+00	1.892e+00	-5.105	3.99e-07	***
stateNew Jersey	-1.459e+01	2.839e+00	-5.140	3.32e-07	***
stateNew Mexico	-1.088e+01	2.199e+00	-4.948	8.84e-07	***

```

stateNew York      -1.458e+01  1.926e+00  -7.570  8.76e-14 ***
stateNorth Carolina -4.931e+00  1.839e+00  -2.682  0.007449 **
stateNorth Dakota   3.697e+00  3.599e+00   1.027  0.304537
stateOhio           -2.244e+00  1.895e+00  -1.184  0.236715
stateOklahoma       6.600e+00  2.540e+00   2.599  0.009500 **
stateOregon         -1.154e+01  2.860e+00  -4.034  5.92e-05 ***
statePennsylvania   -6.441e+00  1.824e+00  -3.532  0.000432 ***
stateRhode Island   -1.270e+01  2.297e+00  -5.530  4.12e-08 ***
stateSouth Carolina -1.130e+00  2.554e+00  -0.442  0.658251
stateSouth Dakota    8.777e-01  2.380e+00   0.369  0.712402
stateTennessee      -1.220e+00  2.402e+00  -0.508  0.611622
stateTexas          -3.355e+00  1.885e+00  -1.780  0.075454 .
stateUtah           -1.733e+00  2.548e+00  -0.680  0.496541
stateVermont        -2.320e+01  2.873e+00  -8.074  2.03e-15 ***
stateVirginia        -1.042e+01  1.888e+00  -5.520  4.36e-08 ***
stateWashington     -1.410e+01  2.085e+00  -6.764  2.33e-11 ***
stateWest Virginia   5.453e+00  2.840e+00   1.920  0.055180 .
stateWisconsin       -5.965e+00  1.831e+00  -3.257  0.001165 **
stateWyoming         1.685e+01  3.594e+00   4.689  3.15e-06 ***
days_towards_election -5.775e-03  6.351e-04  -9.094  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

Residual standard error: 3.106 on 962 degrees of freedom

(1092 observations deleted due to missingness)

Multiple R-squared: 0.7121, Adjusted R-squared: 0.6972

F-statistic: 47.6 on 50 and 962 DF, p-value: < 2.2e-16

2 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 [C](#).

2.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

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

$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$

$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$

$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$

$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$

$$\sigma \sim \text{Exponential}(1) \tag{6}$$

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`.

2.1.1 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 θ .

3 Results

Our results are summarized in `?@tbl-modelresults`.

4 Discussion

4.1 First discussion point

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

4.2 Second discussion point

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

4.3 Third discussion point

4.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Pollster Methodology Overview and Evaluation: YouGov

YouGov is a global public opinion and data company that conducts online surveys on a variety of topics, including politics, social issues, and consumer behavior. Founded in 2000, YouGov is known for leveraging technology to conduct large-scale online surveys, combining traditional sampling principles with advanced data analytics to measure public opinion efficiently.

A.1 Population, Frame, and Sample

- **Population:** The population refers to the group of individuals whose opinions YouGov aims to measure. For political surveys, this often includes eligible voters in a specific country (e.g., registered U.S. voters). Other surveys may focus on specific demographic groups, such as young adults or industry professionals.
- **Frame:** The frame is a list from which the sample is drawn. YouGov uses its online panel, consisting of millions of registered users worldwide. For specific surveys, the frame is the subset of panel members matching desired criteria (e.g., age, location).
- **Sample:** The sample is a subset of the population selected to participate in a survey. Political surveys often involve 1,000-3,000 respondents, weighted to match the demographic characteristics of the broader population.

A.2 Sample Recruitment

- **Recruitment Process:** YouGov recruits panel members via online advertisements, partnerships, and social media. Individuals join the panel by registering on the YouGov website and completing a demographic profile.
- **Incentives:** Panel members earn rewards through a points-based system, which can be redeemed for cash, gift cards, or other benefits.

A.3 Sampling Approach and Trade-offs

- **Sampling Method:** YouGov employs a non-probability sampling approach using quota sampling combined with statistical weighting. Respondents are selected to fill quotas based on demographics (age, gender, education, region) that align with the population.
- **Advantages of Quota Sampling:**

- **Cost-effective:** Less expensive than random sampling due to online recruitment and automation.
- **Speed:** Enables quick data collection, crucial for tracking fast-changing opinions.
- **Targeted Sampling:** Can focus on hard-to-reach populations or specific demographics.
- **Limitations of Quota Sampling:**
 - **Selection Bias:** Self-selection into the panel may introduce biases, as panel members might differ from the general population (e.g., more engaged online).
 - **Generalizability Issues:** Weighting may not fully adjust for attitudinal differences between panelists and the public.

A.4 Handling Non-Response

- **Mitigation Strategies:** YouGov reduces non-response bias with flexible survey completion times and reminder emails. Statistical weighting adjusts for demographic discrepancies caused by non-response.
- **Weighting:** Survey data are weighted to match demographic distributions (e.g., age, gender, race, education). Additional adjustments may be made for political affiliation or past voting behavior.

A.5 Questionnaire Design

- **Strengths:**
 - **Clarity:** Questions are straightforward and easy to understand, reducing measurement error.
 - **Consistency:** Surveys follow a standardized format, ensuring consistency over time, important for tracking changes in opinion.
- **Weaknesses:**
 - **Limited Depth:** Online surveys may feature shorter questionnaires to avoid fatigue, limiting topic depth.
 - **Response Options:** The design of response options (e.g., including “Don’t Know”) can influence results, potentially leading to different conclusions.

A.6 Evaluation of YouGov's Methodology

YouGov's methodology provides several strengths, such as cost, speed, and accessibility, making it suitable for political polling and market research. Online panels enable rapid data collection and targeted sampling. However, the reliance on non-probability sampling introduces biases. While weighting can mitigate some issues, it may not fully compensate for differences between panelists and the general population.

- **Strengths:**
 - **Efficient:** Cost-effective and quick data collection.
 - **Adaptable:** Can rapidly capture opinions on evolving issues.
 - **Targeted:** Capable of reaching niche demographics or regions.
- **Weaknesses:**
 - **Selection Bias:** Potential biases due to non-probability sampling.
 - **Non-Response Bias:** Some groups may be less likely to participate.
 - **Questionnaire Limitations:** Less depth compared to other survey methods.

Overall, YouGov's approach satisfies many standards for modern survey research. Although it has limitations, the insights gained are valuable when these are taken into account.

B Appendix: Methodology and Survey Design for 2024 U.S. Presidential Election Forecast

B.1 Appendix A: Methodology

This section details the **idealized methodology** used to conduct a national survey to forecast the 2024 U.S. presidential election. The methodology includes **sampling strategies, recruitment methods, data validation techniques, and poll aggregation** procedures.

B.1.1 1. Sampling Approach

- **Total Sample Size:** 10,000 respondents
 - **Distribution by State:**
 - **Minimum 100 respondents per state** (including Washington, D.C.) to ensure balanced regional representation.
 - **Oversampling in battleground states** (e.g., Georgia, Pennsylvania, Arizona) with **1,000 respondents per state** for more precise insights.
 - **Stratified Random Sampling:**
 - Stratify by **age, gender, race/ethnicity, education, income, and region**.
 - Ensure proportional representation based on **U.S. Census data** to avoid bias.
 - **Weighting Strategy:**
 - Apply **post-stratification weighting** to adjust for any sampling imbalance, aligning with national demographics.
-

B.1.2 2. Recruitment Plan

- **Recruitment Channels:**
 - Use **YouGov, Qualtrics, and MTurk** to access pre-screened online panels.
 - Complement with **social media ads** on Facebook, Instagram, and LinkedIn to reach underrepresented groups.
 - **Incentives:**
 - Offer **\$10 gift cards** or entry into sweepstakes to increase participation and engagement.
-

B.1.3 3. Survey Implementation and Design

- **Survey Platform:** Google Forms
 - **Link:** *(Insert Google Forms survey link here)*
 - Accessible to respondents via **email, social media, and direct recruitment channels**.

- **Survey Structure:**

1. **Demographics Section:** Age, gender, race/ethnicity, income, education, and state of residence.
 2. **Voting Preferences Section:** Candidate choice, likelihood of voting, and party affiliation.
 3. **Key Issues Section:** Identify priority issues (e.g., economy, healthcare, immigration).
 4. **Thank You Message:**
 - *“Thank you for completing the survey! Your input is greatly appreciated and will help provide insights into the upcoming 2024 election.”*
-

B.1.4 4. Data Validation

- **Techniques for Data Quality:**

1. **Screening Questions:** Confirm eligibility (e.g., 18+ years old, registered voter status).
 2. **Attention Checks:** Include a question like *“Please select ‘Agree’ for this item”* to verify respondents are attentive.
 3. **IP Geolocation:** Validate state residency based on reported location.
 4. **Duplicate Detection:** Identify and remove duplicate responses.
-

B.1.5 5. Poll Aggregation and Reporting

- **Poll Aggregation:**

- Combine results with data from **YouGov, Marquette, and other reliable sources** for a robust forecast.
- Use **weighted averages** to account for differences in sample size and demographics.

- **Margin of Error:**

- **National Margin of Error:** $\pm 1\%$ at the 95% confidence level.
 - **State-Level Margins:** $\pm 5\text{-}10\%$ depending on the sample size for each state.
-

B.1.6 6. Budget Allocation

Expense	Estimated Cost
Panel Provider Fees	\$60,000
Participant Incentives	\$20,000
Social Media Advertising	\$10,000
Google Forms (Platform)	Free
Data Validation & Analysis	\$8,000
Miscellaneous Expenses	\$2,000
Total	\$100,000

B.2 Appendix B: Survey Questions

Below is the full content of the survey to be implemented using Google Forms:

B.2.1 Survey Questions

B.2.1.1 Section 1: Demographics

1. What is your age?

- 18-24
- 25-34
- 35-44
- 45-54
- 55-64
- 65+

2. What is your gender?

- Male
- Female

- Non-binary/Other

3. What is your race/ethnicity? (Select all that apply)

- White
- Black or African American
- Hispanic or Latino
- Asian
- Indigenous (Native American, Alaska Native, or First Nations)
- Pacific Islander or Native Hawaiian
- Other (please specify)

4. What is your household income?

- Less than \$25,000
- \$25,000 - \$49,999
- \$50,000 - \$99,999
- \$100,000 or more

5. Which state do you currently reside in?

- (List all 50 states + Washington, D.C.)

B.2.1.2 Section 2: Voting Preferences

6. Are you a registered voter?

- Yes
- No
- Not sure

7. How likely are you to vote in the 2024 presidential election?

- Definitely will vote
- Probably will vote
- Probably will not vote
- Definitely will not vote

8. **If the 2024 election were held today, who would you vote for?**

- Kamala Harris (Democrat)
- Donald Trump (Republican)
- Other (Please specify)
- Undecided

9. **How favorable are your opinions of the following candidates?**
(Rate on a scale of 1 to 5)

- Kamala Harris
- Donald Trump
- Any third-party candidates

B.2.1.3 Section 3: Key Issues

10. **What is the most important issue for you in this election?**

- Economy
- Healthcare
- Immigration
- Climate change
- Social Security and Medicare

- Foreign policy

11. **Which candidate do you think would handle the economy better?**

- Kamala Harris
- Donald Trump
- Not sure

B.2.1.4 Thank You Message

“Thank you for completing the survey! Your input is greatly appreciated and will help provide insights into the upcoming 2024 election.”

B.3 Appendix C: Google Forms Link

Once the survey is created, insert the **Google Forms link** here for participants to access.

<https://forms.gle/2MGYeZavDsCNuWZ1A>

C Model details

C.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...

C.2 Diagnostics

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

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

References

- 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/>.