

# Forecasting the 2024 U.S. Presidential Election\*

## Using Aggregated Polling Data to Predict the Outcome of the 2024 Presidential Race

Andy Jiang

Onon Burentuvshin

Rui Hu

October 22, 2024

The 2024 U.S. presidential election is shaping up to be a highly competitive race, with polling data playing a crucial role in gauging public sentiment and forecasting the potential outcome. This paper aims to build a predictive model using a poll-of-polls approach, aggregating data from multiple pollsters to forecast the winner of the upcoming election. The analysis incorporates polling data from various sources, filtered to include only high-quality pollsters with a numeric grade above 3, to ensure the reliability of the predictions. Additionally, a detailed examination of one selected pollster's methodology is conducted, highlighting strengths, weaknesses, and biases in the survey process. The paper concludes with an idealized polling methodology, designed within a \$100K budget, to enhance the accuracy of election forecasts. Results are presented in the context of understanding polling trends, potential biases, and the evolving landscape of voter preferences as the election approaches.

## 1 Introduction

The 2024 U.S. presidential election is shaping up to be one of the most closely watched and potentially pivotal elections in recent history. As public opinion shifts and political dynamics evolve, polling data has emerged as a critical tool for forecasting the outcome and understanding voter preferences. Polls provide a snapshot of where candidates stand at various points in time, reflecting the impact of campaigns, debates, and major events. However, the reliability and accuracy of individual polls can vary significantly based on methodology, sample size, and other factors, making it necessary to aggregate multiple sources of data for a more robust prediction.

---

\*Code and data are available at: [<https://github.com/AndyYanxunJiang/president-polls>].

This paper aims to forecast the outcome of the 2024 U.S. presidential election by leveraging a “poll-of-polls” approach, which aggregates data from various pollsters to create a more comprehensive picture of the race. By focusing on high-quality polls and conducting a detailed analysis of the polling trends for the two main candidates, Donald Trump and Kamala Harris, we seek to uncover patterns in voter support and potential factors influencing shifts in public opinion. Additionally, the paper examines the methodologies of different pollsters to assess their strengths and weaknesses, which can help identify possible biases in the aggregated data.

The analysis will further explore the geographic distribution of support across key battleground states, providing insights into the electoral landscape. Finally, the paper presents an idealized polling methodology with a \$100,000 budget, designed to improve the accuracy and representativeness of election forecasts. By integrating multiple data sources and critically evaluating polling practices, this study aims to contribute to the ongoing discussion on the role of polling in democratic processes and election forecasting.

## 2 Data

### 2.1 Overview

We use the statistical programming language R (R Core Team 2023).

### 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.

Table 1: Sample of 10 Random Observations from the Analysis Data

pollster	state	start_date	end_date	question_id	sample_size	population	party	answer	candidate_name
Siena/NYT	Pennsylvania	7/9/24	7/11/24	202982	872	rv	REP	Trump	Donald Trump
Siena/NYT	Michigan	9/21/24	9/26/24	211030	688	rv	DEM	Harris	Kamala Harris
Siena/NYT	Michigan	4/28/24	5/9/24	198699	616	lv	IND	West	Cornel West
Siena/NYT	Georgia	10/22/23	11/3/23	184836	629	lv	DEM	Biden	Joe Biden
Siena/NYT	Pennsylvania	10/22/23	11/3/23	184827	600	lv	DEM	Biden	Joe Biden
Siena/NYT	Pennsylvania	7/9/24	7/11/24	202986	872	rv	REP	Trump	Donald Trump
Siena/NYT	Georgia	10/22/23	11/3/23	185003	629	lv	DEM	Biden	Joe Biden
Siena/NYT	Georgia	10/22/23	11/3/23	184798	629	lv	DEM	Biden	Joe Biden

Siena/NYT	Pennsylvania	4/28/24	5/9/24	198703	1023	lv	REP	Trump	Donald Trump
Siena/NYT	Pennsylvania	10/7/24	10/10/24	212458	857	rv	LIB	Oliver	Chase Oliver

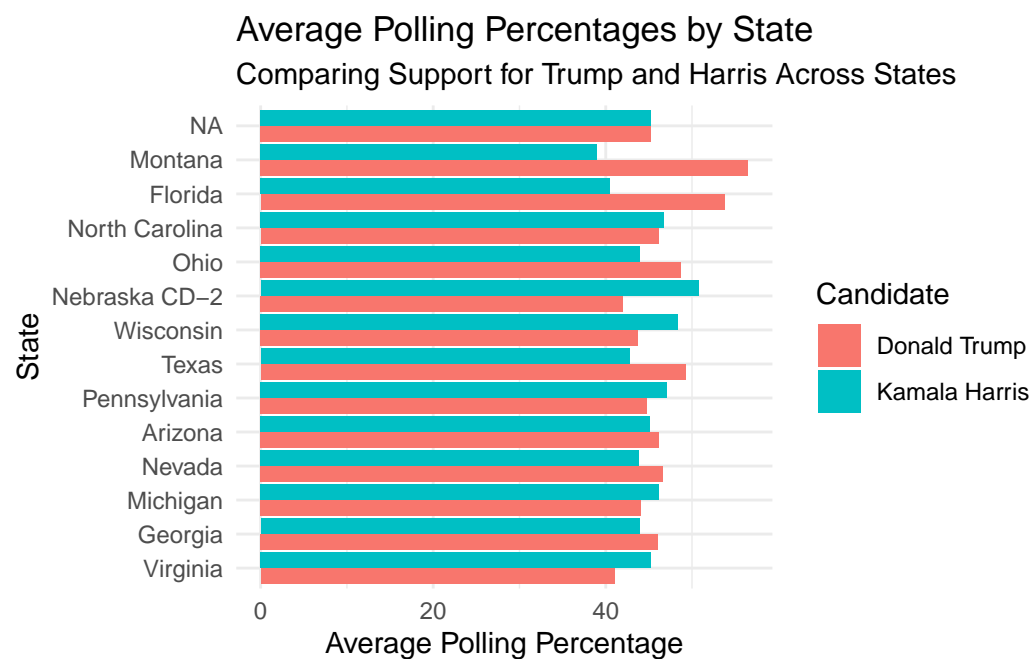
## 2.4 Predictor variables

## 3 Model

### 3.1 Model set-up

#### 3.1.1 Model justification

## 4 Results



## **5 Discussion**

### **5.1 First discussion point**

### **5.2 Second discussion point**

### **5.3 Third discussion point**

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

## **Appendix**

### **A Additional data details**

### **B Model details**

#### **B.1 Posterior predictive check**

#### **B.2 Diagnostics**

## References

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. <https://www.R-project.org/>.