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

Veyasan Ragulan

November 4, 2024

The 2024 US election has been seen as a close race between the Democratic nominess 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.

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 eletoral college vote (269-269 vote distribution) (Lopez 2024).

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 PREDICTEES) 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 it's 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.

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

2 Data

The dataset provided by fivethirtyeight contains over 15000 observations across 50 variables. Each observation is a poll conducted on the 2024 US Presidential Election. Fivethirtyeight's dataset was chosen due to it's comprehensive review of polls and pollsters, attributing grades in error and bias, as well as in transparency. Their thorough investigations ensure their dataset not only contains insightful polls, but also includes as much relevant information about the polls conducted as possible. This gives us plenty of predictors when building a model that will caluclate the percentage win of either Trump or Harris.

The dataset was read into R (R Core Team 2023), and cleaned using the R package tidyverse (Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Lucy D'Agostino McGowan and Romain François and Garrett Grolemund and Alex Hayes and Lionel Henry and Jim Hester and Max Kuhn and Thomas Lin Pedersen and Evan Miller and Stephan Milton Bache and Kirill Müller and Jeroen Ooms and David Robinson and Dana Paige Seidel and Vitalie Spinu and Kohske Takahashi and Davis Vaughan and Claus Wilke and Kara Woo and Hiroaki Yutani 2019). Due to Kamal's late entry into the race, some polls do not have any data on her specifically. This is why we set a cut-off date at 2024/07/21, the day Kamala replaced Joe Biden as the Democratic nominee.

Fivethirtyeight includes a metric called `numeric_grade`, which combines the transparency_score and transparency score. We only included polls with a `numeric_grade` at or above 2.7. This gave us the best balance between quantity and quality. Finally, we create 2 datasets, one with polls predicting a Harris win, the other a Trump win.

2.1 Predictors

2.1.1 `transparency_score`

Fivethirtyeight awards a score out of 10 to each pollster based on how informative their methodology is. Understanding the methodology of a pollster assits in understanding any errors or biases present in their findings.

Table 1: Summary of transparency_scores amongst polls predicting a Kamala Harris win. The closer the values are to 10, the more trustworthy the predictions.

Min	Mean	Median	Max
4.5	8.519821	9	10

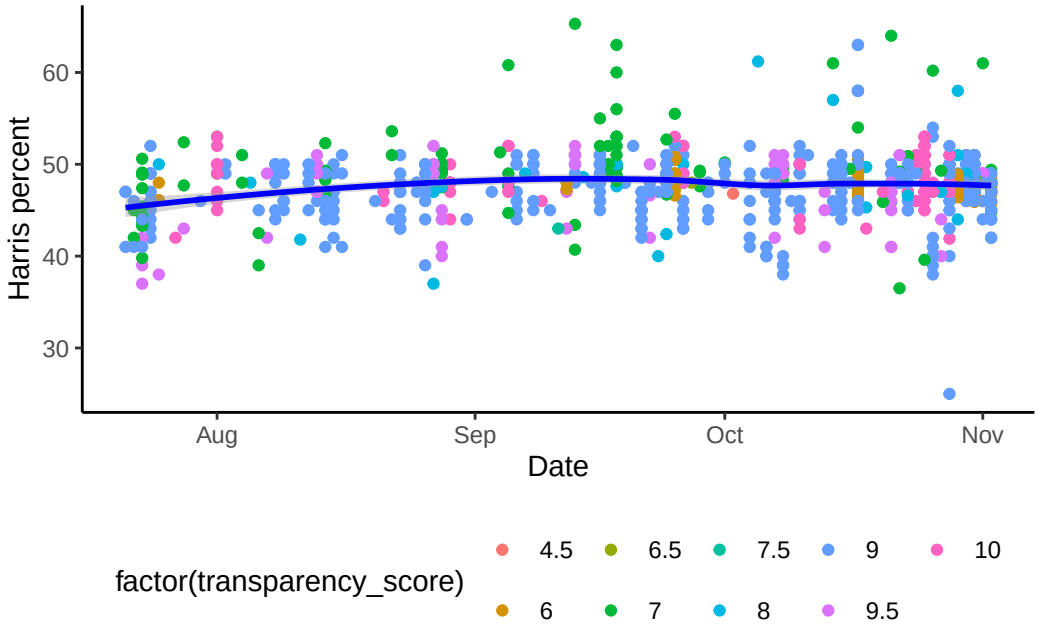


Figure 1: As election day draws close, pollsters are wrapping up their polls and posting their predictions on who will win. The polls in this graph predict Harris' win percentage, with the average in the blue line. The dots in the graph below represent individual polls conducted in the months leading up to the election. They are coloured based on their transparency_score, which is fivethirtyeight's metric for transpareny in the pollster's methodology.

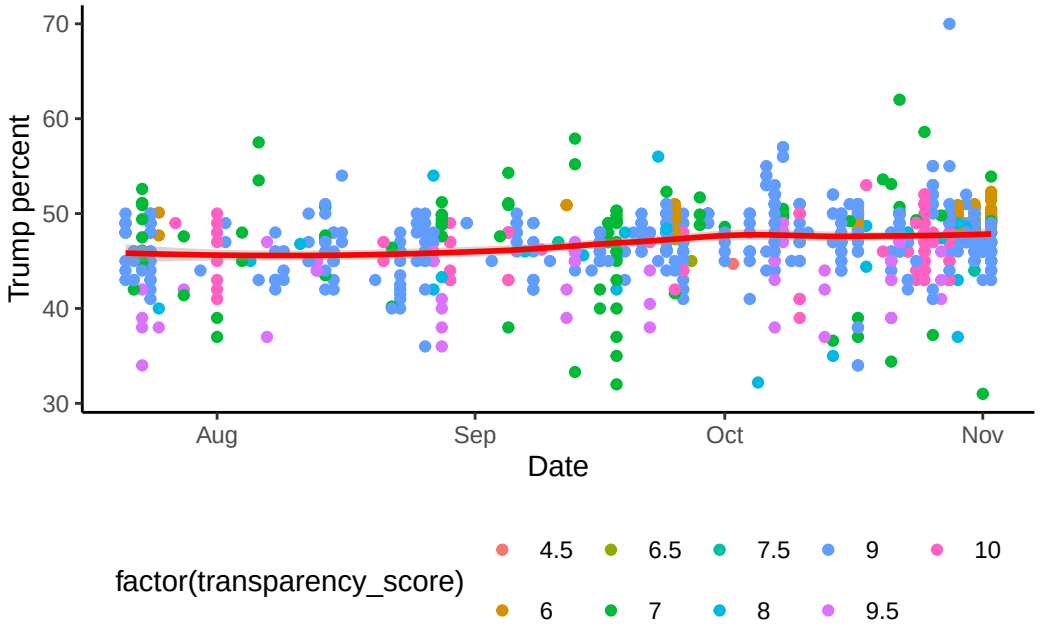


Figure 2: As election day draws close, pollsters are wrapping up their polls and posting their predictions on who will win. The polls in this graph predict Trump's win percentage, with the average in the blue line. The dots in the graph below represent individual polls conducted in the months leading up to the election. They are coloured based on their transparency_score, which is fivethirtyeight's metric for transpareny in the pollster's methodology.

Table 2: Summary of transparency_scores amongst polls predicting a Donald Trump win. The closer the values are to 10, the more trustworthy the predictions.

Min	Mean	Median	Max
4.5	8.532379	9	10

2.1.2 `numeric_grade`

A cumulative score given to pollster that combines fivethirtyeight's `pollscore` and `transparency_score`. Like `transparency_score`, we believe this metric can give us insight into potential biases in each pollster and poll, which is why it is included in our model.

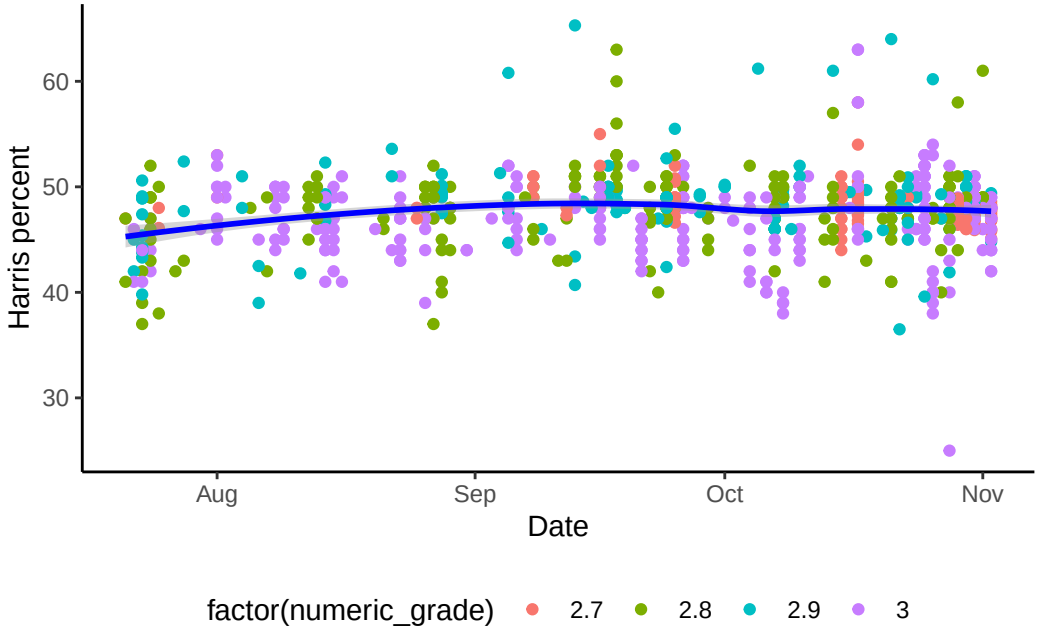


Figure 3: As election day draws close, pollsters are wrapping up their polls and posting their predictions on who will win. The polls in this graph predict Harris' win percentage, with the average in the blue line. The dots in the graph below represent individual polls conducted in the months leading up to the election. They are coloured based on their numeric_grade, which is fivethirtyeight's combined metric on transparency and bias/errors.

Table 3: Summary of numeric_grades amongst polls predicting a Kamala Harris win.

Min	Mean	Median	Max
2.7	2.882992	2.9	3

Table 4: Summary of transparency_scores amongst polls predicting a Donald Trump win. The closer the values are to 10, the more trustworthy the predictions.

Min	Mean	Median	Max
4.5	8.532379	9	10

2.1.3 `pollster`

```
# Facet by pollster
base_plot_kamala +
  geom_point() +
  geom_smooth() +
  facet_wrap(vars(pollster))
```

```
`geom_smooth()` using method = 'loess' and formula = 'y ~ x'
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: span too small.  fewer data values than degrees of freedom.
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: at 19975
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: radius 0.060025
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: all data on boundary of neighborhood. make span bigger
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at 19975
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 0.245
```

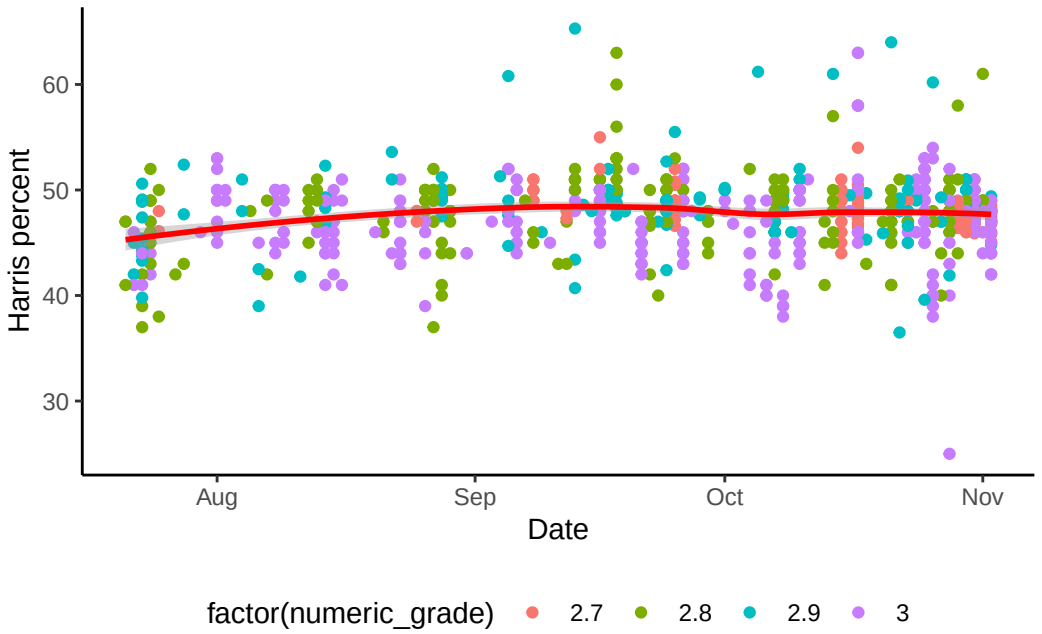


Figure 4: As election day draws close, pollsters are wrapping up their polls and posting their predictions on who will win. The polls in this graph predict Trumps' win percentage, with the average in the blue line. The dots in the graph below represent individual polls conducted in the months leading up to the election. They are coloured based on their numeric_grade, which is fivethirtyeight's combined metric on transparency and bias/errors.

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 1

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: at 20024

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: radius 0.060025

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: all data on boundary of neighborhood. make span bigger

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 0.060025

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: zero-width neighborhood. make span bigger
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: zero-width neighborhood. make span bigger

Warning: Failed to fit group -1.
Caused by error in `predLoess()`:
! NA/NaN/Inf in foreign function call (arg 5)

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at 20028

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 36.51

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 0

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 1225

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at
20028
```

```
Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
36.51

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition
number 0

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : There are other near
singularities as well. 1225

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at 19971

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 35

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 1.5664e-16

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at
19971

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius 35

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition
number 1.5664e-16

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: at 19984
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: radius 0.0484

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: all data on boundary of neighborhood. make span bigger

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at 19984

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 0.22

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 1

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 1955.4

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: zero-width neighborhood. make span bigger

Warning: Failed to fit group -1.
Caused by error in `predLoess()`:
! NA/NaN/Inf in foreign function call (arg 5)

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: span too small.  fewer data values than degrees of freedom.

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at 19981

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 36.215

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 0

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 1240.1
```

```
Warning in sqrt(sum.squares/one.delta): NaNs produced

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : span too small.  fewer
data values than degrees of freedom.

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at
19981

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
36.215

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition
number 0

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : There are other near
singularities as well. 1240.1

Warning in stats::qt(level/2 + 0.5, pred$df): NaNs produced

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: span too small.  fewer data values than degrees of freedom.

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: at 19985

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: radius 0.042025

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: all data on boundary of neighborhood. make span bigger
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at 19985

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 0.205

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 1

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: at 20026

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: radius 0.042025

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: all data on boundary of neighborhood. make span bigger

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 0.042025

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: zero-width neighborhood. make span bigger
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: zero-width neighborhood. make span bigger

Warning: Failed to fit group -1.
Caused by error in `predLoess()`:
! NA/NaN/Inf in foreign function call (arg 5)

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: span too small.  fewer data values than degrees of freedom.

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: at 19977

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: radius 0.0625
```

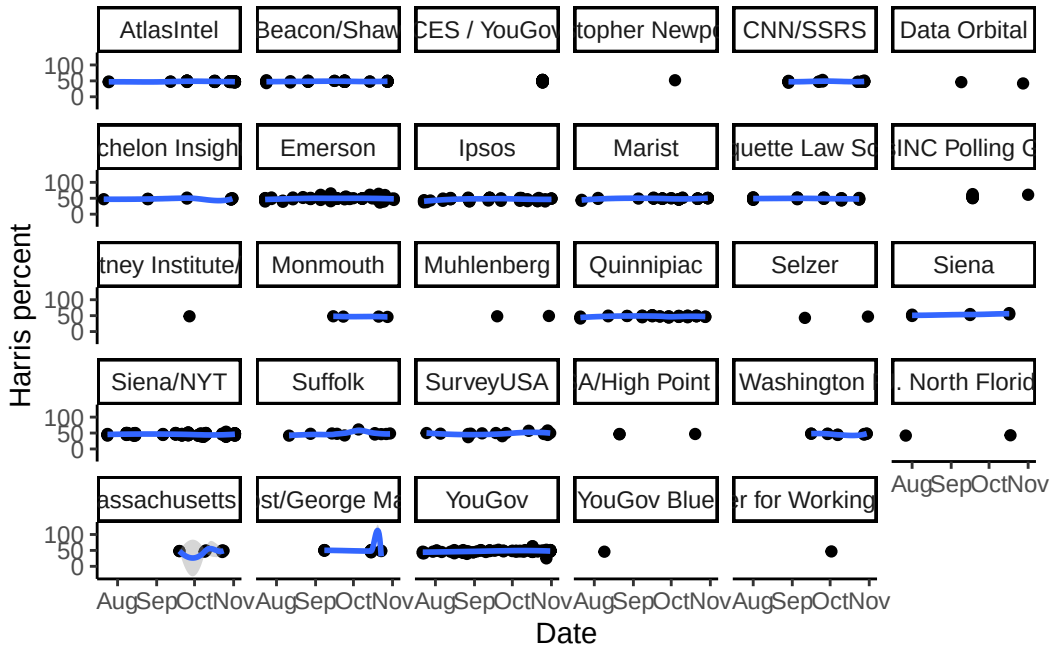
```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: all data on boundary of neighborhood. make span bigger

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at 19977

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 0.25

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 1

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: at 20027

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: radius 0.0625

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: all data on boundary of neighborhood. make span bigger

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 0.0625

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: zero-width neighborhood. make span bigger
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: zero-width neighborhood. make span bigger

Warning: Failed to fit group -1.
Caused by error in `predLoess()`:
! NA/NaN/Inf in foreign function call (arg 5)

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at 19936

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 46.385
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 0

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 985.02

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at
19936

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
46.385

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition
number 0

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : There are other near
singularities as well. 985.02

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: span too small.  fewer data values than degrees of freedom.

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: at 19956

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: radius 0.09

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: all data on boundary of neighborhood. make span bigger

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at 19956
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 0.3

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 1

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 3636.1

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: zero-width neighborhood. make span bigger

Warning: Failed to fit group -1.
Caused by error in `predLoess()`:
! NA/NaN/Inf in foreign function call (arg 5)

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at 20003

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 21

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 0

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at
20003

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius 21

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition
number 0
```

```
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: span too small.  fewer data values than degrees of freedom.

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: at 19931

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: radius 0.17223

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: all data on boundary of neighborhood. make span bigger

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at 19931

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 0.415

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 1

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: at 20014

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: radius 0.17223

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: all data on boundary of neighborhood. make span bigger

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 0.17223

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: zero-width neighborhood. make span bigger
Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: zero-width neighborhood. make span bigger
```

```
Warning: Failed to fit group -1.
Caused by error in `predLoess()`:
! NA/NaN/Inf in foreign function call (arg 5)

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: pseudoinverse used at 19974

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: neighborhood radius 37.225

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: reciprocal condition number 0

Warning in simpleLoess(y, x, w, span, degree = degree, parametric = parametric,
: There are other near singularities as well. 67.651

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : pseudoinverse used at
19974

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : neighborhood radius
37.225

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : reciprocal condition
number 0

Warning in predLoess(object$y, object$x, newx = if (is.null(newdata)) object$x
else if (is.data.frame(newdata))
as.matrix(model.frame(delete.response(terms(object)), : There are other near
singularities as well. 67.651

Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
-Inf
```



2.1.4 `states`

2.1.5 `end_date`

Every poll has a start and end date for querying its participants. Polls that are conducted closer to election day may be more reflective of American opinions as opposed to polls conducted in August or September. This is why we will be using this variable as one of our predictors in the model.

2.2 Response

2.2.1 `pct`

This the the predicted percentage each candidate will get on election day. We have assumed a two-party system for our report, as any third-party candidates would have a negligible impact on election day.

Table 5: Summary of percentage vote predictions for Kamala Harris.

Min	Mean	Median	Max
25	47.78927	48	65.3

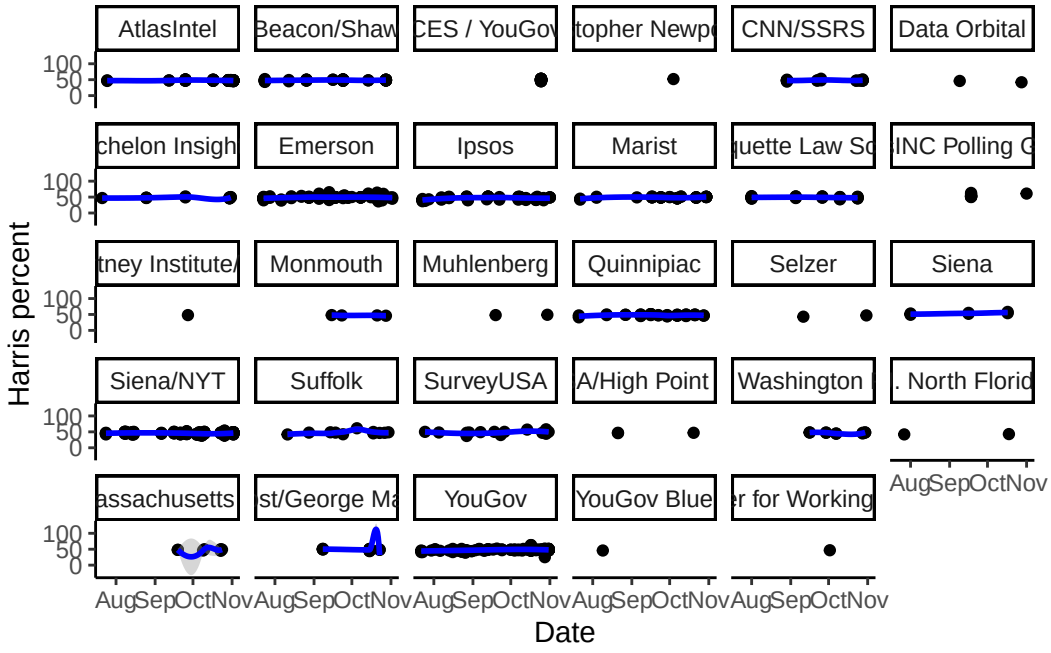


Figure 5: As election day draws close, pollsters are wrapping up their polls and posting their predictions on who will win. Each graph shows individual polls conducted in the months leading up to the election. They represent the change in prediciton for Kamala Harris winning the presidency.

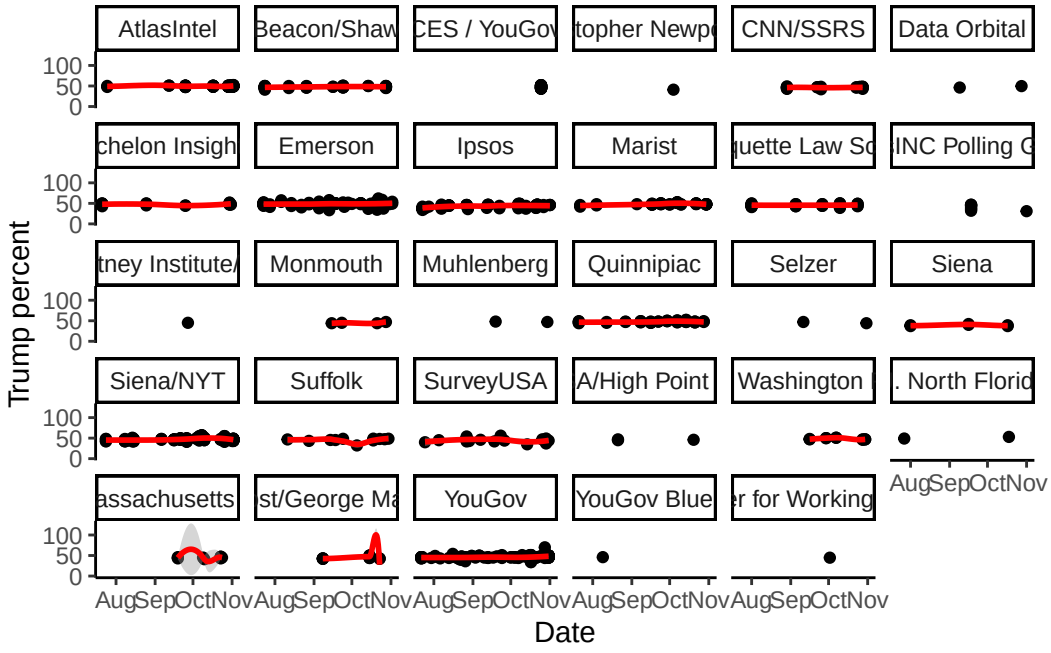


Figure 6: As election day draws close, pollsters are wrapping up their polls and posting their predictions on who will win. Each graph shows individual polls conducted in the months leading up to the election. They represent the change in prediciton for Donald Trump winning the presidency.

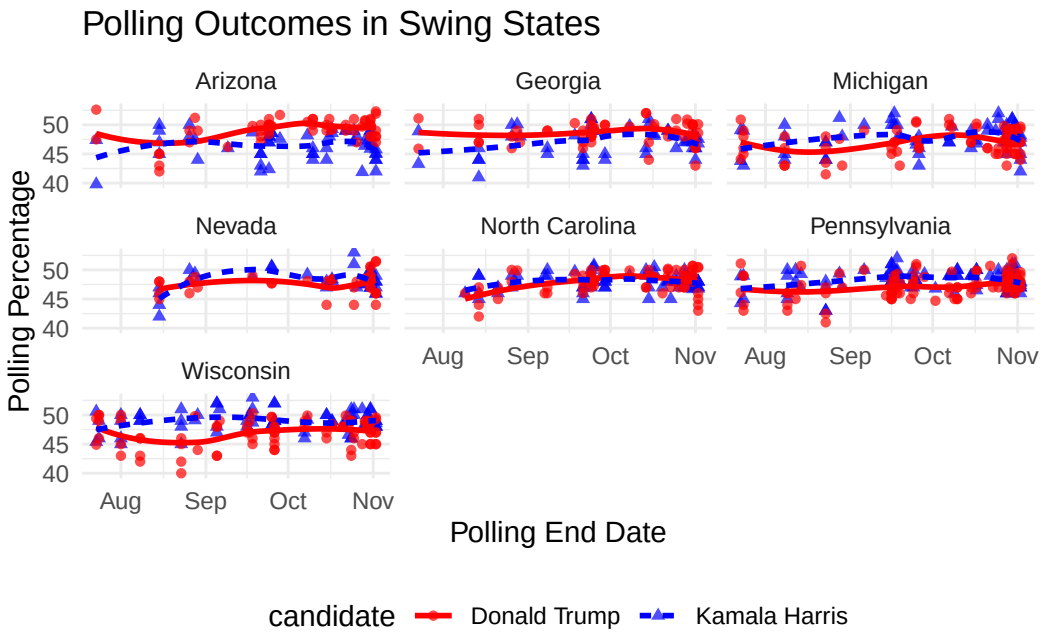


Figure 7: Battleground states are what usually determines the winner of a presidential election. This figure outlines the polling outcomes of each candidate from August to November, for each of the seven battleground states this election cycle.

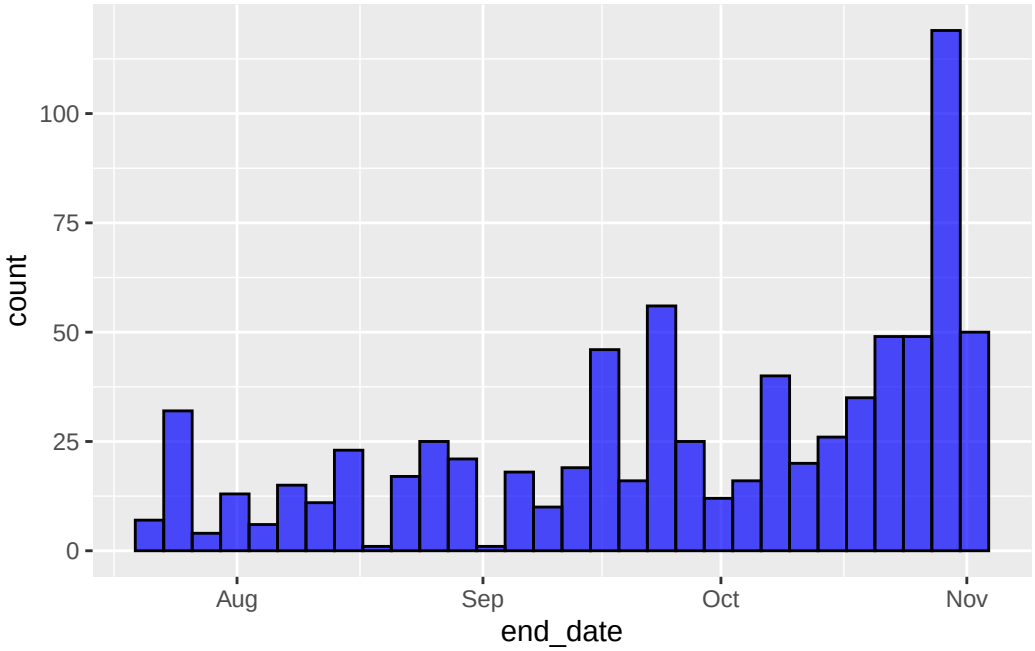


Figure 8: Polls are conducted from a start date to an end date. The closer the end date is to election day, the more insightful the results are to analysts and the general public. This distribution shows the end date of polls predicting a Kamala Harris win.

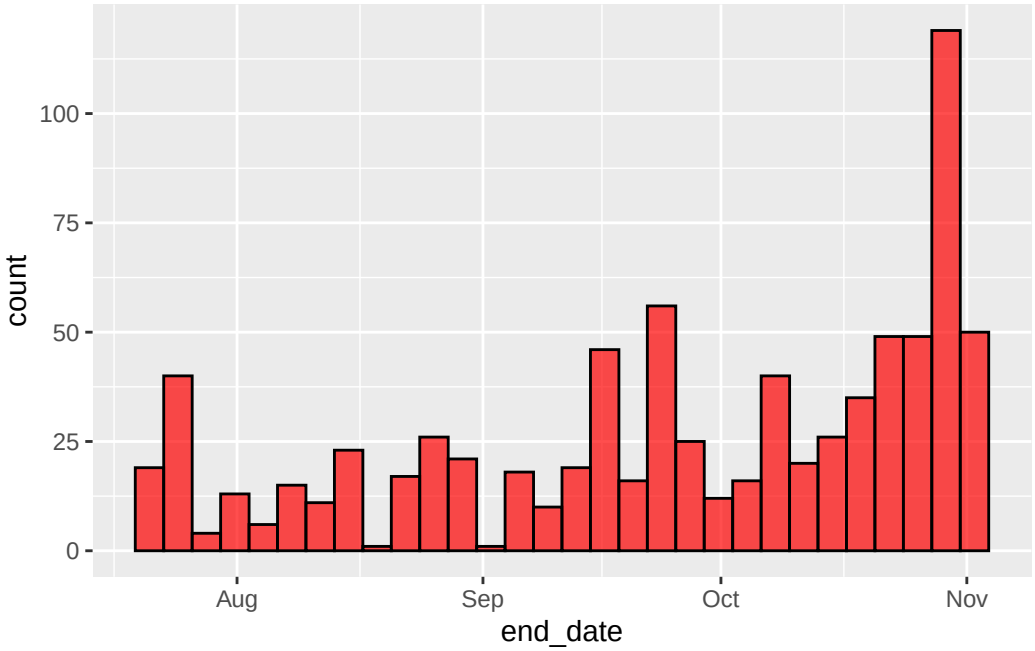


Figure 9: Polls are conducted from a start date to an end date. The closer the end date is to election day, the more insightful the results are to analysts and the general public. This distribution shows the end date of polls predicting a Donald Trump win.

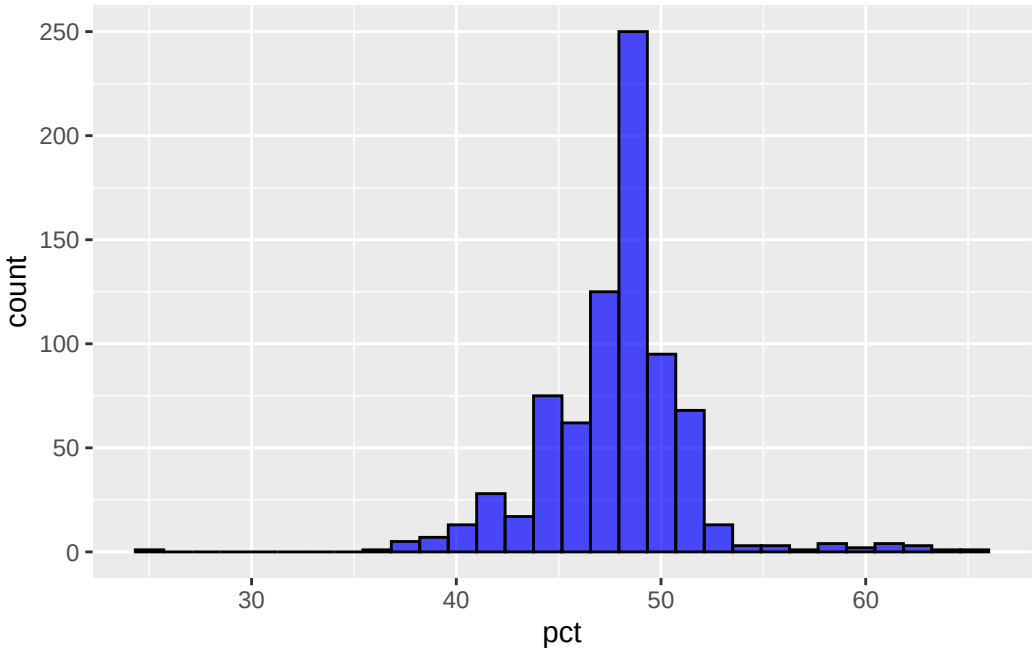


Figure 10: The distribution of predicted Harris vote percentage on election day.

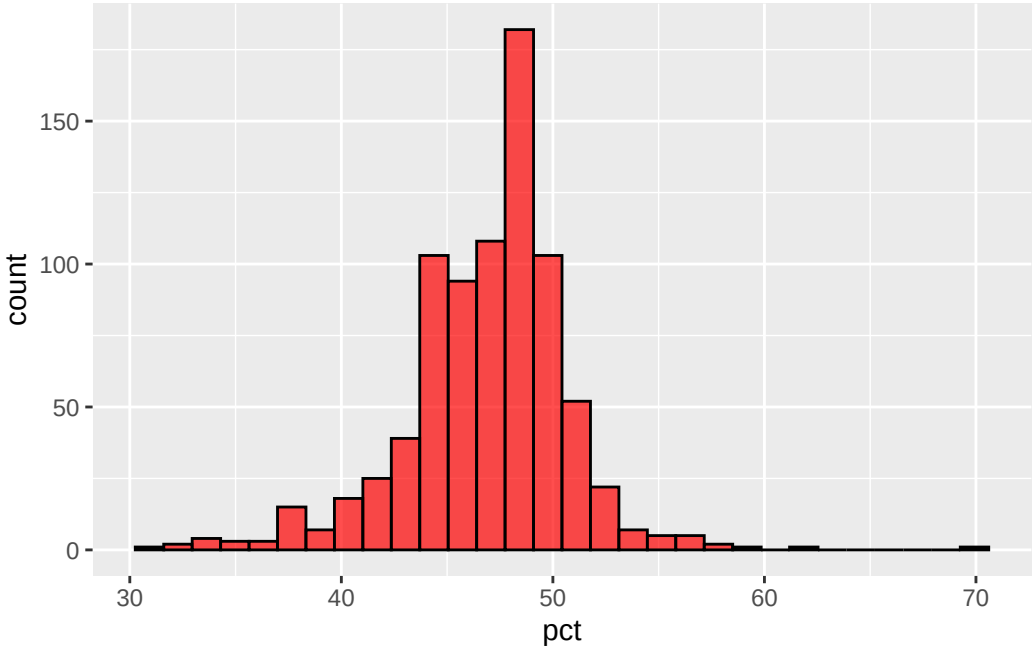


Figure 11: The distribution of predicted Trump vote percentage on election day.

Table 6: Summary of percentage vote predictions for Donald Trump.

Min	Mean	Median	Max
31	46.9788	47	70

3 Measurement

The dataset comes from fivethirtyeight.com, a statistical analysis website dedicated to US sports and elections (Ryan Best, Aaron Bycoffe 2024). Fivethirtyeight takes polls conducted on a particular topic (in this case the 2024 US presidency), and aggregates them in one dataset.

Election polls are used to gauge public opinion about candidates, political topics, voter engagement, and public opinion of the election process. Election polls also serve as key insight to political organisations representing candidates, polls can show potential weak points across demographics.

Election polls are similarly to regular polls, they involve asking a sample population a series of questions via a delivery method. The key distinction is that election polls typically try to target *voters*, or people who are most likely to vote in the election cycle being investigated. This is a key challenge to pollsters, as there are many subgroups of voters, such as people who seldom vote or never vote. According to Pew Research, a third of US citizens did not vote in the 2020 election, despite the 2020 election cycle having high voter turnout (Keeter 2024). There may also be cases when respondents of a poll reconsider whether they will vote come election day.

Election polling has exploded in recent times according to Pew Research. They suggest that double there are double the amount of active pollsters in 2022 compared to 2000 (Keeter 2024). One reason mentioned is the advent of online polling, which is simple, cheap, and can easily reach US citizens across the country, compared to traditional methods such as poll by phone and mail. As a result, while the diversity of pollsters has increased dramatically, the variance in results methodology, and ethics have also increased

Errors in election polling have been an issue in the last two cycles of presidential elections. Between the 2016 and 2020 elections, there has been a trend of pollsters not fully capturing Trump's support. For example in 2020, despite Biden winning the election, there were reports from the American Associcate for Public Opinion Research that national polls overstated his victory over Trump by 3.9 percentage points (Keeter 2024).

Fivethirty eight mitigates the influx of pollsters and the potential for errors by routinely checking up on pollsters as they produce reports. Fivethirtyeight checks every pollster, regardless of experience or prestige. 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 (Radcliffe and Morris 2023):

- 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 (Radcliffe and Morris 2023):

- 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 (Radcliffe and Morris 2023):

- 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 markets, the pollster is blacklisted from their data. 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 fivethirtyeight can reverse this decision if the pollster demonstrates satifactory ethics and methodology practices(Radcliffe and Morris 2023).

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 (Radcliffe and Morris 2023).

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, ther 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 candiate 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 potemntially 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 peoples votes, so this is an interesting parameter to track. This also lets us see if certain methods yield better or worse predictions.

5 Results

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

Hadley Wickham and Mara Averick and Jennifer Bryan and Winston Chang and Lucy D'Agostino McGowan and Romain François and Garrett Grolemund and Alex Hayes and Lionel Henry and Jim Hester and Max Kuhn and Thomas Lin Pedersen and Evan Miller and Stephan Milton Bache and Kirill Müller and Jeroen Ooms and David Robinson and Dana Paige Seidel and Vitalie Spinu and Kohske Takahashi and Davis Vaughan and Claus Wilke and Kara Woo and Hiroaki Yutani. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Keeter, Scott. 2024. "Public Opinion Polling Basics." *Public Opinion Polling Basics*. Pew Research Center. https://www.pewresearch.org/course/public-opinion-polling-basics/#how-does-polling-work.

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.

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

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.

Ryan Best, Aaron Bycoffe. 2024. "FiveThirtyEight: latest US 2024 General Election Polls." *FiveThirtyEight*. ABCNews. https://projects.fivethirtyeight.com/polls/.