

# 2024 General Election Forecasting Model

Jack Regan

```
library(dplyr)
```

Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

filter, lag

The following objects are masked from 'package:base':

intersect, setdiff, setequal, union

```
library(readr)  
library(janitor)
```

Attaching package: 'janitor'

The following objects are masked from 'package:stats':

chisq.test, fisher.test

```
library(stats)
```

```
polling_data <- read_csv("data/president_polls.csv", show_col_types = FALSE)  
glimpse(polling_data)
```

Rows: 15,971

Columns: 52

\$ poll_id	<dbl> 88806, 88806, 88836, 88836, 88817, 88817, 88~
\$ pollster_id	<dbl> 770, 770, 1895, 1895, 1741, 1741, 770, 770, ~
\$ pollster	<chr> "TIPP", "TIPP", "Quantus Insights", "Quantus~
\$ sponsor_ids	<dbl> NA, NA, 2184, 2184, NA, NA, NA, NA, NA, NA, ~
\$ sponsors	<chr> NA, NA, "TrendingPolitics", "TrendingPolitic~
\$ display_name	<chr> "TIPP Insights", "TIPP Insights", "Quantus I~
\$ pollster_rating_id	<dbl> 144, 144, 859, 859, 721, 721, 144, 144, 338,~
\$ pollster_rating_name	<chr> "TIPP Insights", "TIPP Insights", "Quantus I~
\$ numeric_grade	<dbl> 1.8, 1.8, NA, NA, NA, NA, 1.8, 1.8, 0.7, 0.7~
\$ pollscore	<dbl> -0.4, -0.4, NA, NA, NA, NA, -0.4, -0.4, 0.6,~
\$ methodology	<chr> "Online Panel", "Online Panel", "Online Pane~
\$ transparency_score	<dbl> 3.0, 3.0, 5.5, 5.5, 8.0, 8.0, 3.0, 3.0, 4.0,~
\$ state	<chr> NA, NA, "Pennsylvania", "Pennsylvania", "Flo~
\$ start_date	<chr> "10/18/24", "10/18/24", "10/17/24", "10/17/2~
\$ end_date	<chr> "10/20/24", "10/20/24", "10/20/24", "10/20/2~
\$ sponsor_candidate_id	<dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ sponsor_candidate	<chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ sponsor_candidate_party	<chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ endorsed_candidate_id	<lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ endorsed_candidate_name	<lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ endorsed_candidate_party	<lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ question_id	<dbl> 213459, 213459, 213538, 213538, 213472, 2134~
\$ sample_size	<dbl> 1244, 1244, 840, 840, 400, 400, 1254, 1254, ~
\$ population	<chr> "lv", "lv", "lv", "lv", "lv", "lv", "lv", "lv", ~
\$ subpopulation	<lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ population_full	<chr> "lv", "lv", "lv", "lv", "lv", "lv", "lv", "lv", ~
\$ tracking	<lgl> TRUE, TRUE, NA, NA, NA, NA, TRUE, TRUE, NA, ~
\$ created_at	<chr> "10/21/24 08:43", "10/21/24 08:43", "10/21/2~
\$ notes	<chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ url	<chr> "https://tippinsights.com/tipp-tracking-poll~
\$ url_article	<chr> "https://tippinsights.com/tipp-tracking-poll~
\$ url_topline	<chr> NA, NA, "https://docs.google.com/document/d/~
\$ url_crosstab	<chr> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ source	<dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
\$ internal	<lgl> NA, NA, FALSE, FALSE, FALSE, FALSE, NA, NA, ~
\$ partisan	<chr> NA, NA, "REP", "REP", NA, NA, NA, NA, "REP",~
\$ race_id	<dbl> 8914, 8914, 8872, 8872, 8778, 8778, 8914, 89~
\$ cycle	<dbl> 2024, 2024, 2024, 2024, 2024, 2024, 2024, 20~
\$ office_type	<chr> "U.S. President", "U.S. President", "U.S. Pr~
\$ seat_number	<dbl> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, ~
\$ seat_name	<lgl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~

```

$ election_date      <chr> "11/5/24", "11/5/24", "11/5/24", "11/5/24", ~
$ stage             <chr> "general", "general", "general", "general", ~
$ nationwide_batch  <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FA~
$ ranked_choice_reallocated <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FA~
$ ranked_choice_round <dbl> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, ~
$ hypothetical      <lgl> FALSE, FALSE, FALSE, FALSE, FALSE, FALSE, FA~
$ party             <chr> "DEM", "REP", "DEM", "REP", "DEM", "REP", "D~
$ answer            <chr> "Harris", "Trump", "Harris", "Trump", "Harri~
$ candidate_id      <dbl> 16661, 16651, 16661, 16651, 16661, 16651, 16~
$ candidate_name    <chr> "Kamala Harris", "Donald Trump", "Kamala Har~
$ pct               <dbl> 47.0, 48.0, 48.2, 50.3, 45.4, 54.6, 47.0, 49~

```

```

polling_data <- polling_data |>
  select(
    poll_id,
    state,
    start_date,
    end_date,
    pollster_rating_id,
    numeric_grade,
    sample_size,
    candidate_name, pct
  ) |>
  filter(candidate_name == "Kamala Harris")

```

```

mi_polling_data <- polling_data |>
  filter(state == "Michigan")
nv_polling_data <- polling_data |>
  filter(state == "Nevada")
ar_polling_data <- polling_data |>
  filter(state == "Arizona")
nm_polling_data <- polling_data |>
  filter(state == "New Mexico")
wi_polling_data <- polling_data |>
  filter(state == "Wisconsin")
pa_polling_data <- polling_data |>
  filter(state == "Pennsylvania")
nc_polling_data <- polling_data |>
  filter(state == "North Carolina")
ga_polling_data <- polling_data |>
  filter(state == "Georgia")

```

```

toss_up_states <- c("Michigan", "Nevada",
  "Arizona", "New Mexico",
  "Wisconsin", "Pennsylvania",
  "North Carolina", "Georgia"
)
polling_data |>
  filter(state %in% toss_up_states) |>
  group_by(state) |>
  summarize(count = n())

```

```

# A tibble: 8 x 2
  state      count
  <chr>      <int>
1 Arizona    112
2 Georgia    121
3 Michigan   127
4 Nevada     81
5 New Mexico  10
6 North Carolina 112
7 Pennsylvania 169
8 Wisconsin  129

```