test post

Kelly Olmos

2024-09-09

```
# Load libraries.
## install via `install.packages("name")`
library(ggplot2)
library(ggthemes)
library(maps)
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.3     v readr     2.1.4
## v forcats   1.0.0     v stringr   1.5.0
## v lubridate 1.9.2     v tibble    3.2.1
## v purrr     1.0.2     v tidyr     1.3.0
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## x purrr::map()    masks maps::map()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(usdata)
```

```
## Warning: package 'usdata' was built under R version 4.3.3
```

```
library(knitr)
library(dplyr)

## set working directory here
# setwd("~")

####-----------------------------------------------------------#
#### Read and clean presidential popular vote.
####-----------------------------------------------------------#

# Read presidential popular vote.
d_popvote <- read_csv("popvote_1948-2020.csv")
```

```
## Rows: 38 Columns: 9
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (2): party, candidate
## dbl (3): year, pv, pv2p
```

```
## lgl (4): winner, incumbent, incumbent_party, prev_admin
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
# Subset data to most recent past election year.
d_popvote |> colnames()
```

```
## [1] "year"            "party"           "winner"          "candidate"
## [5] "pv"              "pv2p"            "incumbent"       "incumbent_party"
## [9] "prev_admin"
```

```
d_popvote |>
  filter(year == 2020) |>
  select(party, candidate, pv2p)
```

```
## # A tibble: 2 x 3
##   party      candidate         pv2p
##   <chr>      <chr>            <dbl>
## 1 democrat   Biden, Joseph R.  52.3
## 2 republican Trump, Donald J.  47.7
```

```
# Pivot data to wide format with party names as columns and two-party vote share as values.
(d_popvote_wide <- d_popvote |>
    select(year, party, pv2p) |>
    pivot_wider(names_from = party, values_from = pv2p))
```

```
## # A tibble: 19 x 3
##     year democrat republican
##    <dbl>    <dbl>      <dbl>
##  1  1948     52.3       47.7
##  2  1952     44.7       55.3
##  3  1956     42.2       57.8
##  4  1960     50.1       49.9
##  5  1964     61.3       38.7
##  6  1968     49.6       50.4
##  7  1972     38.2       61.8
##  8  1976     51.1       48.9
##  9  1980     44.8       55.2
## 10  1984     40.9       59.1
## 11  1988     46.2       53.8
## 12  1992     53.6       46.4
## 13  1996     54.8       45.2
## 14  2000     50.3       49.7
## 15  2004     48.7       51.3
## 16  2008     53.8       46.2
## 17  2012     51.9       48.1
## 18  2016     51.2       48.8
## 19  2020     52.3       47.7
```

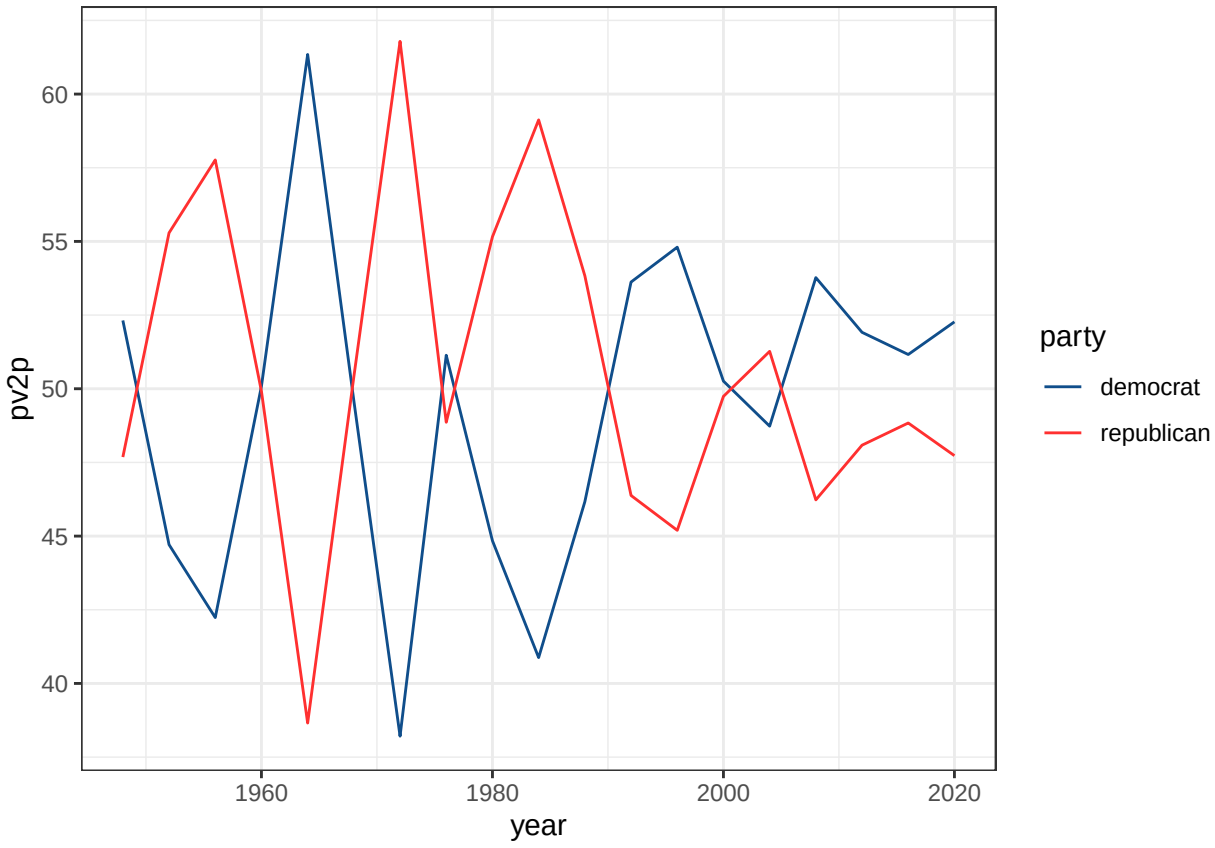
```r
# Modify winner column to show "D" if Democrats win and "R" if Republicans win.
(d_popvote_wide <- d_popvote_wide |>
    mutate(winner = case_when(democrat > republican ~ "D",
                              TRUE ~ "R")))
```

```
## # A tibble: 19 x 4
##     year democrat republican winner
##    <dbl>    <dbl>      <dbl> <chr>
##  1  1948     52.3       47.7 D
##  2  1952     44.7       55.3 R
##  3  1956     42.2       57.8 R
##  4  1960     50.1       49.9 D
##  5  1964     61.3       38.7 D
##  6  1968     49.6       50.4 R
##  7  1972     38.2       61.8 R
##  8  1976     51.1       48.9 D
##  9  1980     44.8       55.2 R
## 10  1984     40.9       59.1 R
## 11  1988     46.2       53.8 R
## 12  1992     53.6       46.4 D
## 13  1996     54.8       45.2 D
## 14  2000     50.3       49.7 D
## 15  2004     48.7       51.3 R
## 16  2008     53.8       46.2 D
## 17  2012     51.9       48.1 D
## 18  2016     51.2       48.8 D
## 19  2020     52.3       47.7 D
```

```r
# Summarize data with respect to winners.
d_popvote_wide |>
  group_by(winner) |>
  summarise(races = n())
```

```
## # A tibble: 2 x 2
##   winner races
##   <chr>  <int>
## 1 D         11
## 2 R          8
```

```r
####----------------------------------------------------------#
#### Visualize trends in national presidential popular vote.
####----------------------------------------------------------#

# Visualize the two-party presidential popular over time.

d_popvote |>
  ggplot(aes(x = year, y = pv2p, color = party)) +
  geom_line() +
  scale_color_manual(values = c("dodgerblue4", "firebrick1")) +
  theme_bw()
```




```
####-----------------------------------------------------------#
#### State-by-state map of presidential popular votes.
####-----------------------------------------------------------#

# Sequester shapefile of states from `maps` library.
states_map <- map_data("state")

# Read wide version of dataset that can be used to compare candidate votes with one another.
d_pvstate_wide <- read_csv("clean_wide_state_2pv_1948_2020.csv")
```

```
## Rows: 959 Columns: 14
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr  (1): state
## dbl (13): year, D_pv, R_pv, D_pv2p, R_pv2p, D_pv_lag1, R_pv_lag1, D_pv2p_lag...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
d_pvstate_wide
```

```
## # A tibble: 959 x 14
##     year state        D_pv  R_pv D_pv2p R_pv2p D_pv_lag1 R_pv_lag1 D_pv2p_lag1
##    <dbl> <chr>       <dbl> <dbl>  <dbl>  <dbl>     <dbl>     <dbl>       <dbl>
##  1  1948 Alabama         0  19.0      0    100        NA        NA          NA
```

```
##  2  1948 Arizona      53.8  43.8   55.1   44.9        NA        NA          NA
##  3  1948 Arkansas     61.7  21.0   74.6   25.4        NA        NA          NA
##  4  1948 California   47.6  47.1   50.2   49.8        NA        NA          NA
##  5  1948 Colorado     51.9  46.5   52.7   47.3        NA        NA          NA
##  6  1948 Connecticut  47.9  49.5   49.2   50.8        NA        NA          NA
##  7  1948 Delaware     48.8  50.0   49.4   50.6        NA        NA          NA
##  8  1948 Florida      48.8  33.6   59.2   40.8        NA        NA          NA
##  9  1948 Georgia      60.8  18.3   76.9   23.1        NA        NA          NA
## 10  1948 Idaho        50.0  47.3   51.4   48.6        NA        NA          NA
## # i 949 more rows
## # i 5 more variables: R_pv2p_lag1 <dbl>, D_pv_lag2 <dbl>, R_pv_lag2 <dbl>,
## #   D_pv2p_lag2 <dbl>, R_pv2p_lag2 <dbl>
```

```
# Merge d_pvstate_wide with state_map.
d_pvstate_wide$region <- tolower(d_pvstate_wide$state)

pv_map <- d_pvstate_wide |>
  filter(year == 2020) |>
  left_join(states_map, by = "region")

pv_map
```

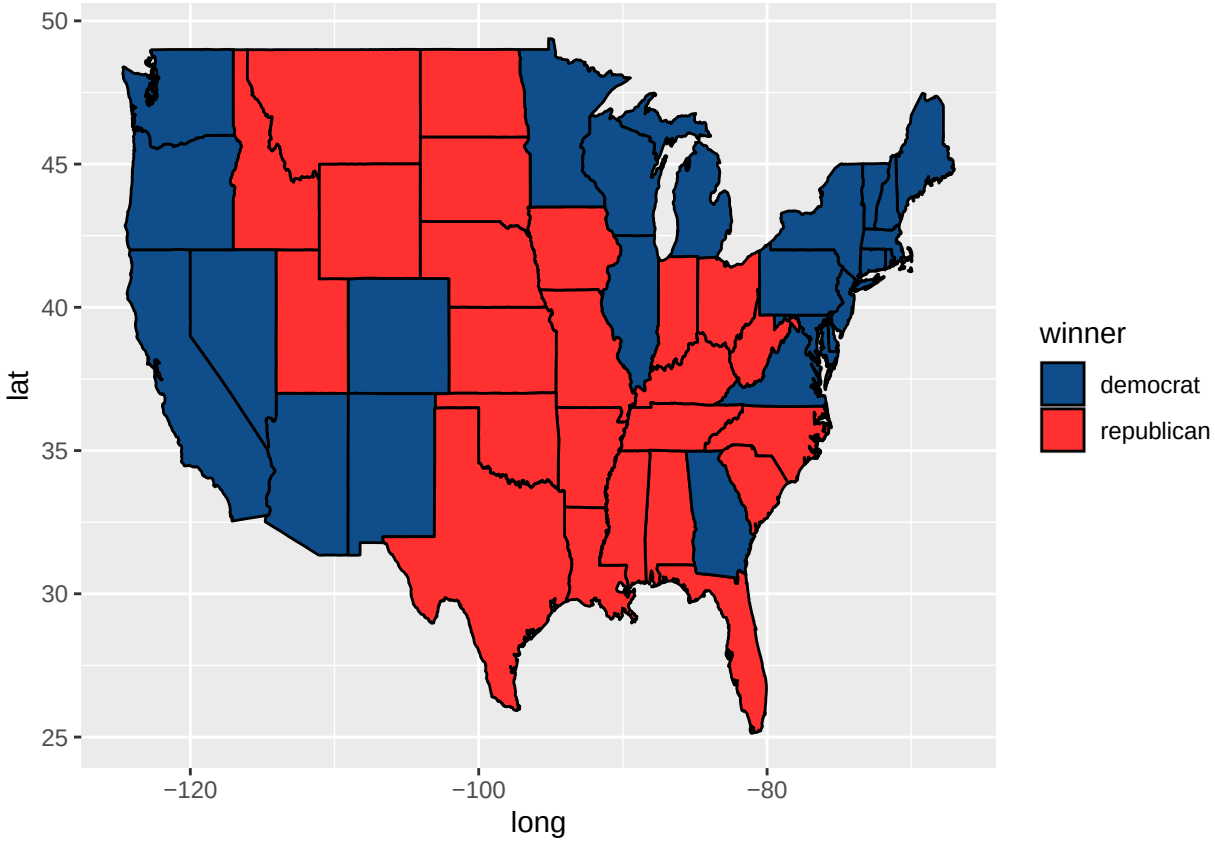
```
## # A tibble: 15,539 x 20
##     year state    D_pv  R_pv D_pv2p R_pv2p D_pv_lag1 R_pv_lag1 D_pv2p_lag1
##    <dbl> <chr>   <dbl> <dbl>  <dbl>  <dbl>     <dbl>     <dbl>       <dbl>
##  1  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  2  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  3  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  4  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  5  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  6  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  7  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  8  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  9  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
## 10  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
## # i 15,529 more rows
## # i 11 more variables: R_pv2p_lag1 <dbl>, D_pv_lag2 <dbl>, R_pv_lag2 <dbl>,
## #   D_pv2p_lag2 <dbl>, R_pv2p_lag2 <dbl>, region <chr>, long <dbl>, lat <dbl>,
## #   group <dbl>, order <int>, subregion <chr>
```

```
# Make map grid of state winners for each election year available in the dataset.

pv_win_map <- d_pvstate_wide |>
  filter(year == 2020) |>
  left_join(states_map, by = "region") |>
  mutate(winner = ifelse(R_pv > D_pv, "republican", "democrat"), state_abbr = state.abb[match(state, sta

pv_win_map
```

```
## # A tibble: 15,539 x 22
##     year state    D_pv  R_pv D_pv2p R_pv2p D_pv_lag1 R_pv_lag1 D_pv2p_lag1
##    <dbl> <chr>   <dbl> <dbl>  <dbl>  <dbl>     <dbl>     <dbl>       <dbl>
##  1  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
```

```
##  2  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  3  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  4  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  5  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  6  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  7  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  8  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
##  9  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
## 10  2020 Alabama  36.6  62.0   37.1   62.9      34.4      62.1        35.6
## # i 15,529 more rows
## # i 13 more variables: R_pv2p_lag1 <dbl>, D_pv_lag2 <dbl>, R_pv_lag2 <dbl>,
## #   D_pv2p_lag2 <dbl>, R_pv2p_lag2 <dbl>, region <chr>, long <dbl>, lat <dbl>,
## #   group <dbl>, order <int>, subregion <chr>, winner <chr>, state_abbr <chr>
```

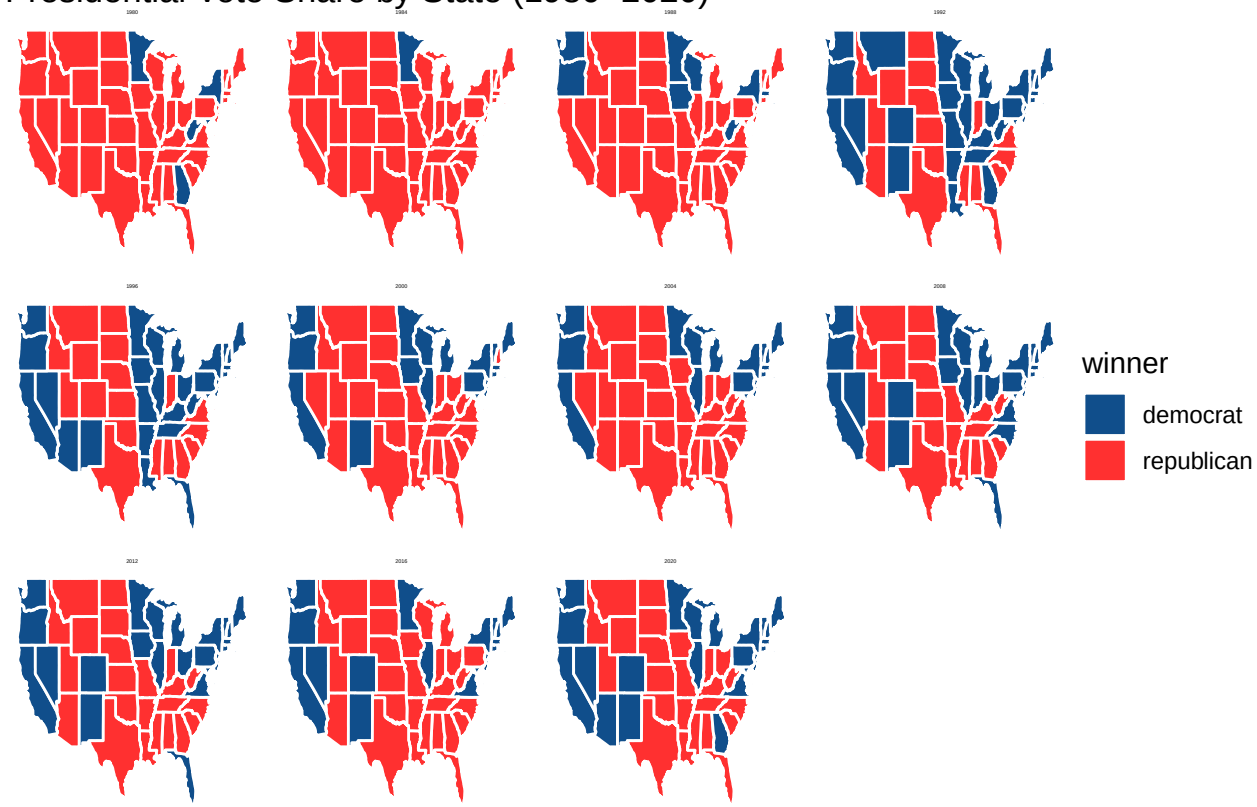
```
pv_win_map |>
  ggplot(aes(long, lat, group = group)) +
  geom_polygon(aes(fill = winner), color = "black") +
  scale_fill_manual(values = c("dodgerblue4", "firebrick1"))
```




```
d_pvstate_wide |>
  filter(year >= 1980) |>
  left_join(states_map, by = "region") |>
  mutate(winner = ifelse(R_pv > D_pv, "republican", "democrat")) |>
  ggplot(aes(long, lat, group = group)) +
  facet_wrap(facets = year ~.) +
```

```
  geom_polygon(aes(fill = winner), color = "white") +
  scale_fill_manual(values = c("dodgerblue4", "firebrick1")) +
  theme_void() +
  ggtitle("Presidential Vote Share by State (1980-2020)") +
  theme(strip.text = element_text(size = 2),
        aspect.ratio = 1)
```

```
## Warning in left_join(filter(d_pvstate_wide, year >= 1980), states_map, by = "region"): Detected an u
## i Row 1 of `x` matches multiple rows in `y`.
## i Row 1 of `y` matches multiple rows in `x`.
## i If a many-to-many relationship is expected, set `relationship =
##   "many-to-many"` to silence this warning.
```

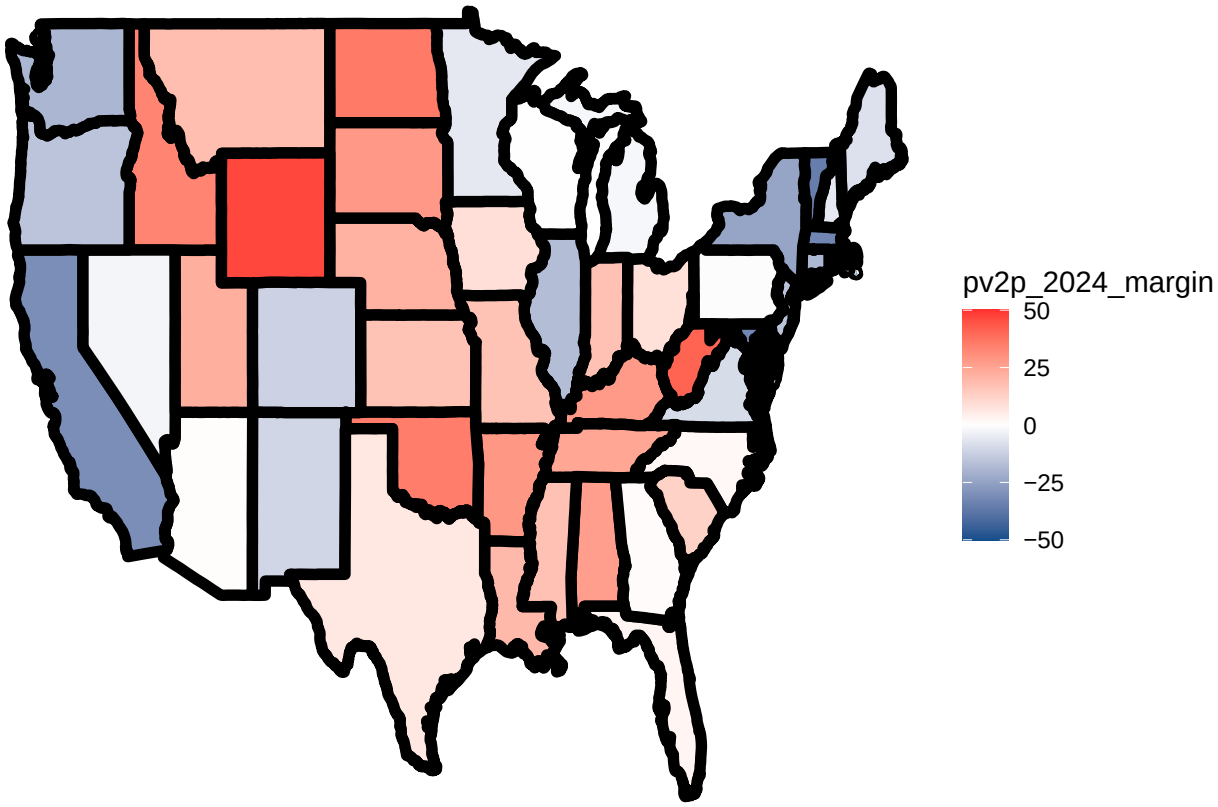



```
####-----------------------------------------------------------#
#### Forecast: simplified electoral cycle model.
####-----------------------------------------------------------#

# Create prediction (pv2p and margin) based on simplified electoral cycle model:
# vote_2024 = 3/4*vote_2020 + 1/4*vote_2016 (lag1, lag2, respectively).
d_pvstate_wide |> colnames()
```

```
##  [1] "year"         "state"        "D_pv"         "R_pv"         "D_pv2p"
##  [6] "R_pv2p"       "D_pv_lag1"    "R_pv_lag1"    "D_pv2p_lag1"  "R_pv2p_lag1"
## [11] "D_pv_lag2"    "R_pv_lag2"    "D_pv2p_lag2"  "R_pv2p_lag2"  "region"
```

```r
pv2p_2024_states <- d_pvstate_wide |>
  filter(year == 2020) |>
  group_by(state) |>
  summarize(D_pv2p_2024 = 0.75*D_pv2p + 0.25*D_pv2p_lag1,
            R_pv2p_2024 = 0.75*R_pv2p + 0.25*R_pv2p_lag1) |>
  mutate(pv2p_2024_margin = R_pv2p_2024 - D_pv2p_2024,
         region = tolower(state))

pv2p_2024_states |>
  left_join(states_map, by = "region") |>
  ggplot(aes(long, lat, group = group)) +
  geom_polygon(aes(fill = pv2p_2024_margin), color = "black", size = 2) +
  scale_fill_gradient2(high = "firebrick1",
                       mid = "white",
                       low = "dodgerblue4",
                       breaks = c(-50, -25, 0, 25, 50),
                       limits = c(-50, 50)) +
  theme_void()
```

```
## Warning: Using 'size' aesthetic for lines was deprecated in ggplot2 3.4.0.
## i Please use 'linewidth' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.
```



```r
ec <- read_csv("ec_full.csv")
```

```
## Rows: 936 Columns: 3
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (1): state
## dbl (2): electors, year
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

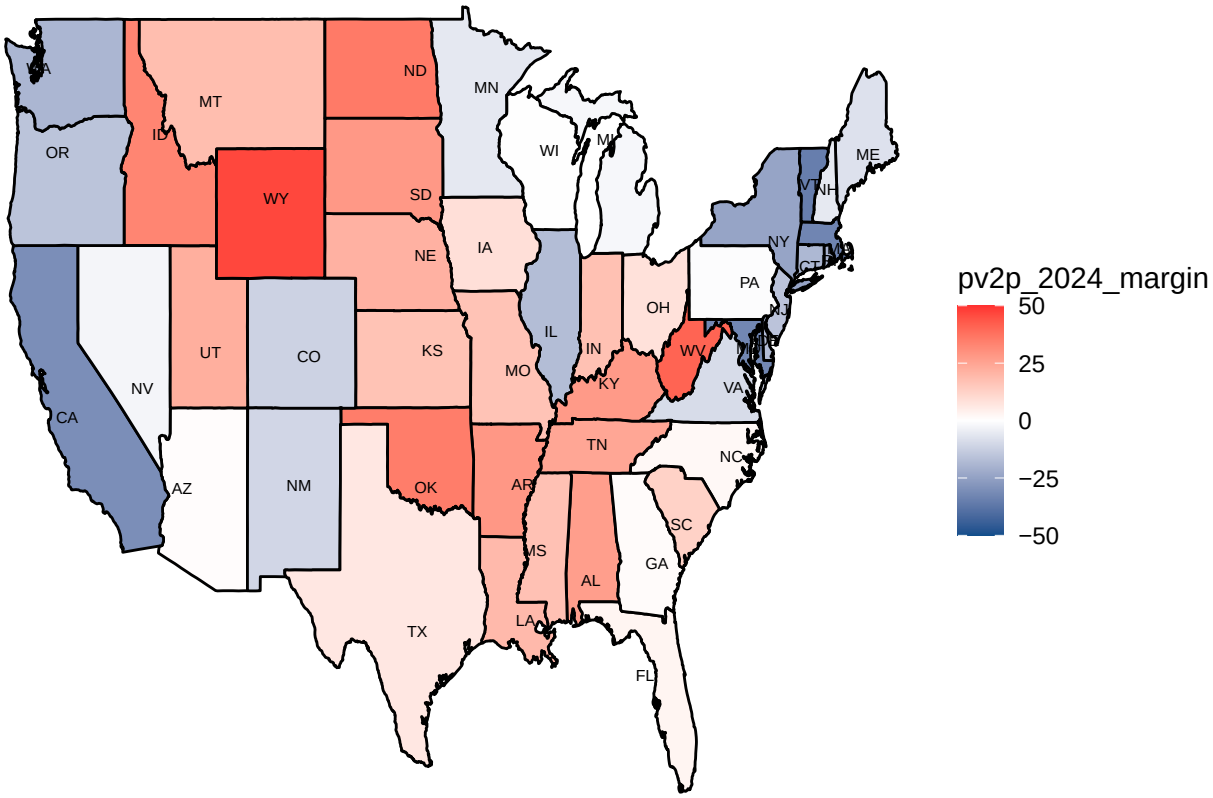
```r
pv2p_2024_states <- pv2p_2024_states |>
  mutate(year = 2024,
         winner = ifelse(R_pv2p_2024 > D_pv2p_2024, "R", "D"),
         state_abbr = state.abb[match(state, state.name)]) |>
  left_join(ec, by = c("state", "year"))

center_state_abbr <- pv2p_2024_states |>
  left_join(states_map, by = "region") |>
  group_by(state) |>
  summarize(
    long = mean(long),
    lat = mean(lat),
    state_abbr = first(state_abbr)
  )

center_state_abbr
```

```
## # A tibble: 51 x 4
##    state                 long   lat state_abbr
##    <chr>                <dbl> <dbl> <chr>
##  1 Alabama              -86.9  31.7 AL
##  2 Alaska                NA    NA   AK
##  3 Arizona             -113.   34.5 AZ
##  4 Arkansas             -91.3  34.6 AR
##  5 California          -121.   36.7 CA
##  6 Colorado            -105.   38.6 CO
##  7 Connecticut          -72.7  41.4 CT
##  8 Delaware             -75.4  39.1 DE
##  9 District Of Columbia -77.1  38.9 <NA>
## 10 Florida              -83.4  28.7 FL
## # i 41 more rows
```

```r
pv2p_2024_states |>
  left_join(states_map, by = "region") |>
  ggplot(aes(long, lat, group = group)) +
  geom_polygon(aes(fill = pv2p_2024_margin), color = "black") +
  geom_text(data = center_state_abbr,
            aes(x = long, y = lat, label = state_abbr),
            inherit.aes = FALSE,
            color = "black",
            size = 2) +
```

```
# Links/Sources/Documentation I used to place initials on the map
# (https://ggplot2.tidyverse.org/reference/geom_text.html#aesthetics)
# https://ggplot2.tidyverse.org/reference/theme.html
# https://www.rdocumentation.org/packages/usdata/versions/0.2.0/topics/state2abbr
#
  scale_fill_gradient2(high = "firebrick1",
                       mid = "white",
                       low = "dodgerblue4",
                       breaks = c(-50, -25, 0, 25, 50),
                       limits = c(-50, 50)) +
  theme_void()
```

```
## Warning: Removed 3 rows containing missing values (`geom_text()`).
```




```
pv2p_2024_states |>
  group_by(winner) |>
  summarize(electoral_votes = sum(electors))
```

```
## # A tibble: 2 x 2
##   winner electoral_votes
##   <chr>            <dbl>
## 1 D                  276
## 2 R                  262
```