

# Texas Political Shifts

Jason Li, Dawson Fishman, Nathan George

2024-11-15

## INTRODUCTION

The topic under study for our Exploratory Data Analysis is the political shifts of Texas in Presidential elections from 2000 to 2024. Specifically, we want to see how the trend of Texas getting more competitive changed in 2024. Using this data, we may learn what key areas of Texas may have resulted in the reversing of trends in the 2024 Presidential election.

## BACKGROUND

**IMPORTANT INFORMATION TO KNOW:** For context, Presidential elections are held every 4 years in the United States. Since 2000, 7 Presidential elections have occurred (including 2024). In each of these elections, Texas voted for the Republican candidate for the Presidency, but has become more and more competitive as Texas was shifting closer and closer to the Democratic Party from 2000 to 2020. This trend reversed in 2024 when the Republican candidate for President won Texas by nearly 14%. Additionally, it is important to know that in American politics, the two major parties are the Democratic (typically denoted with the color blue) and the Republican parties (typicalled denoted with the color red). These are not the only parties that ran presidential candidates, but for the purpose of this analysis, they will be the two parties of focus.

**DATASETS:** We used three datasets for this project. The first one is from MIT Election Lab which includes Presidential election results for every county in the U.S. from 2000-2020. After having filtered the dataset for Texas, we awaited the results of the 2024 Presidential election in Texas. After the election, we found a dataset from the Texas Secretary of State. We then joined the two datasets so that we could have a full set of data for Presidential election results from 2000 to 2024. The final dataset we used was a shapefile dataset from the Texas Department of Transportation for making the map of Texas.

**DATASET 1 (MIT ELECTION LAB):** The dataset from the MIT Election Lab has 72617 observations of 6 variables with 4572 of those observations pertaining to Texas. Some important variables included are year, state, county\_fips, county\_name, party, candidatevotes, and totalvotes. For the purpose of this analysis, all these variables are important except totalvotes because they will be used in generating visualizations.

**DATASET 2 (TEXAS SECRETARY OF STATE):** The dataset from the Texas Secretary of State has 254 observations of 7 variables. The important variables for this analysis are county\_name, REPUBLICAN, DEMOCRAT, and OTHER. The other variables are not necessary for our analysis. This dataset was then joined to DATASET 1 so that we could have the Presidential election results for all counties in Texas from 2000 to 2024.

**DATASET 3 (TEXAS DEPARTMENT OF TRANSPORTATION):** The dataset from the Texas Department of Transportation was a shapefile, so it was mainly used for mapping. We joined this file to our other datasets to create the maps.

```
# Load required libraries
library(tidyverse)
```

```

## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.1     v tibble    3.2.1
## v lubridate 1.9.3     v tidyrr    1.3.1
## v purrr    1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()   masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(ggplot2)
library(dplyr)
library(sf)

## Linking to GEOS 3.12.1, GDAL 3.8.4, PROJ 9.3.1; sf_use_s2() is TRUE

library(ggthemes)

## Warning: package 'ggthemes' was built under R version 4.4.2

library(ggpubr)

## Warning: package 'ggpubr' was built under R version 4.4.2

# Read datasets
countypres_2000_2020 <- read.csv("countypres_2000-2020.csv")
texas2024 <- read.csv("TexasResults2024.csv")

```

## Process 2000–2020 Data

```

# Filter data for Texas, group "OTHER" parties, and summarize votes
data_texas <- countypres_2000_2020 %>%
  filter(state == "TEXAS") %>%
  mutate(party = if_else(party != "DEMOCRAT" & party != "REPUBLICAN", "OTHER", party)) %>%
  group_by(year, county_name, party, totalvotes, county_fips) %>%
  summarize(candidatevotes = sum(candidatevotes), .groups = "drop")

# Convert year to factor
data_texas$year <- as.factor(data_texas$year)

```

## Prepare County FIPS Data

```

# Create a dataframe of county FIPS codes for 2020
county_fips <- data_texas %>%
  filter(year == 2020, party != "REPUBLICAN", party != "DEMOCRAT") %>%
  group_by(county_fips) %>%
  select(c("county_fips", "county_name"))

```

## Process 2024 Data

```
# Reshape 2024 data to a long format and add year
texas2024longer <- texas2024 %>%
  pivot_longer(c("REPUBLICAN", "DEMOCRAT", "OTHER"), names_to = "party", values_to = "candidatevotes") %
  mutate(year = "2024") %>%
  left_join(county_fips, by = "county_name")

# Select and organize columns for analysis
texas2024selected <- texas2024longer %>%
  select(year, county_name, party, totalvotes, candidatevotes, county_fips)
```

## Combine Datasets and Calculate Percentages

```
# Combine 2000–2020 data with 2024 data
data_texas_2000_2024 <- data_texas %>%
  full_join(texas2024selected)

## Joining with 'by = join_by(year, county_name, party, totalvotes, county_fips,
## candidatevotes)'

# Remove duplicate columns and rename
data_texas_2000_2024 <- data_texas_2000_2024 %>%
  mutate(percent = candidatevotes / totalvotes)
```

## Calculate Statewide Totals by Party

```
# Summarize total votes by party and year
state_total_votes_party <- data_texas_2000_2024 %>%
  group_by(party, year) %>%
  summarize(state_votes = sum(candidatevotes), .groups = "drop")

# Summarize total votes across all parties
state_total_votes <- state_total_votes_party %>%
  group_by(year) %>%
  summarize(total_votes = sum(state_votes))

# Ensure the 'party' factor levels are set
state_total_votes_party$party <- factor(
  state_total_votes_party$party,
  levels = c("REPUBLICAN", "DEMOCRAT", "OTHER"))
)
```

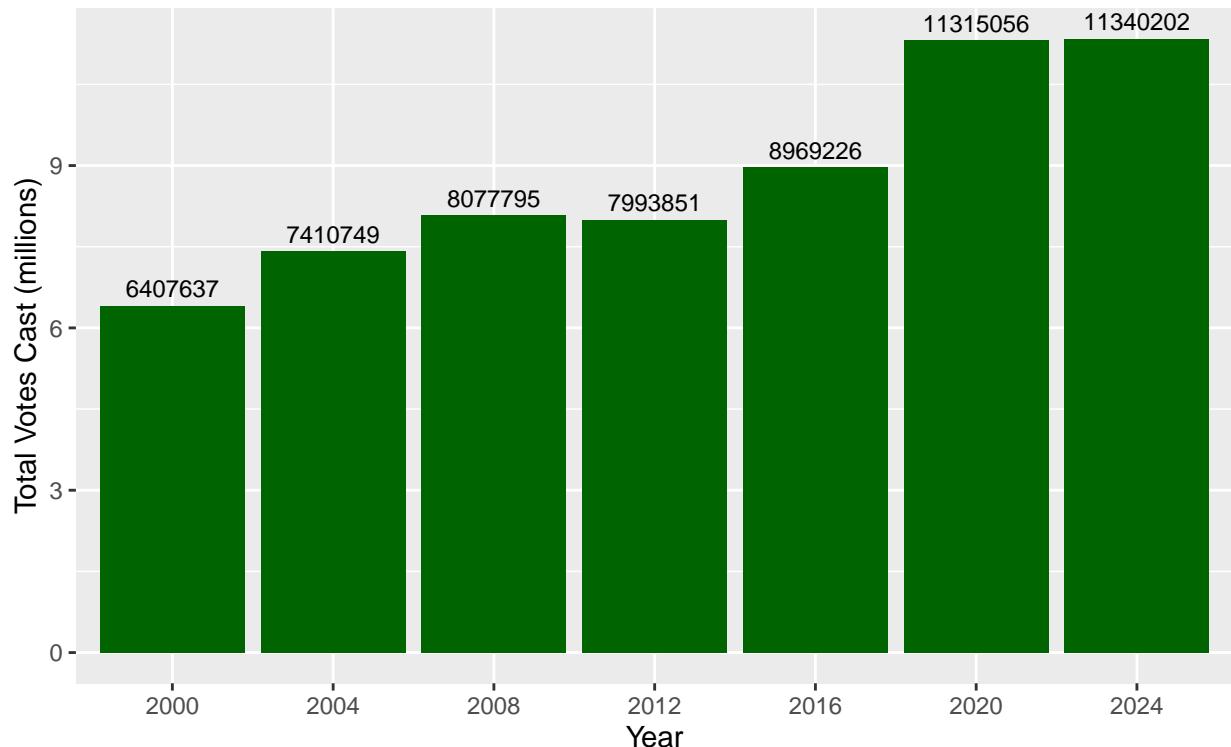
## Plot Total Votes Over Time

This plot shows the total number of votes cast in Texas Presidential elections from 2000 to 2024. We thought this may be interesting to see whether voter turnout affected the results of the race.

```
ggplot(state_total_votes, aes(x = year, y = total_votes / 1e6)) +
  geom_col(fill = "darkgreen") +
  geom_text(aes(label = total_votes), vjust = -0.5, size = 3) +
  labs(x = "Year",
       y = "Total Votes Cast (millions)",
       title = "Total Votes Cast in Texas Presidential Elections",
       subtitle = "2000–2024")
```

## Total Votes Cast in Texas Presidential Elections

2000–2024



## Calculate Competitive Counties

```
# Pivot wider to calculate differences between Republican and Democrat votes
votes_wider <- data_texas_2000_2024 %>%
  pivot_wider(values_from = c("candidatevotes", "percent"), names_from = "party") %>%
  mutate(difference = percent_PUBLICAN - percent_PUBLICAN)

# Join calculated differences back to the dataset
data_texas_2000_2024 <- data_texas_2000_2024 %>%
  right_join(votes_wider, by = c("year", "county_name")) %>%
  select(-c(county_fips.y, totalvotes.y)) %>%
  rename(county_fips = county_fips.x,
        totalvotes = totalvotes.x)

# Categorize results as Republican, Democrat, or Competitive
```

```
data_texas_2000_2024 <- data_texas_2000_2024 %>%
  mutate(result = if_else(difference > 0.05, "REPUBLICAN",
                         if_else(difference < -0.05, "DEMOCRAT", "COMPETITIVE")))
```

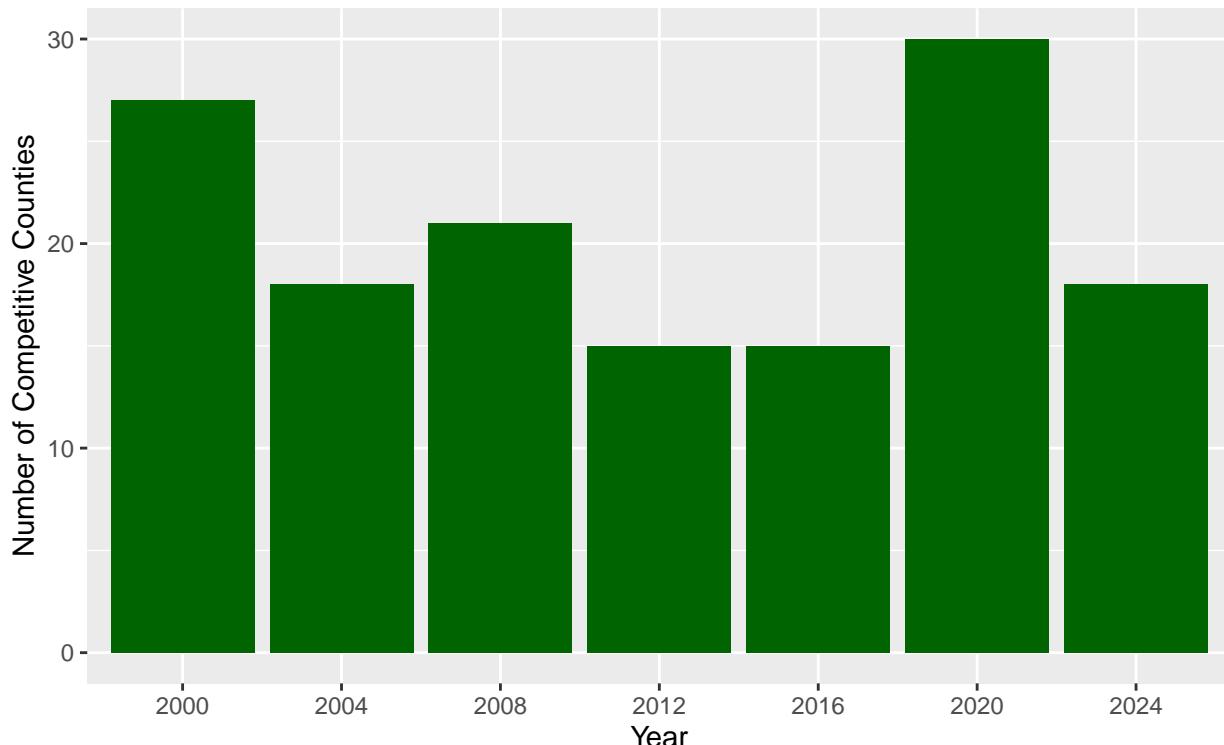
## Plot Competitive Counties Over Time

This plot shows the total number of counties in Texas in which one of the major political parties won >45% of the vote or <55% of the vote. We created this to try to see if there was a trend in Texas where the number of competitive counties corresponded to the competitiveness of the statewide race.

```
# Count competitive counties by year
results_tally <- data_texas_2000_2024 %>%
  group_by(year) %>%
  filter(result == "COMPETITIVE") %>%
  tally()

# Plot number of competitive counties
ggplot(results_tally, aes(x = year, y = n)) +
  geom_col(fill = "darkgreen") +
  labs(x = "Year",
       y = "Number of Competitive Counties",
       title = "Number of Competitive Counties in Texas Presidential Elections",
       subtitle = "2000–2024")
```

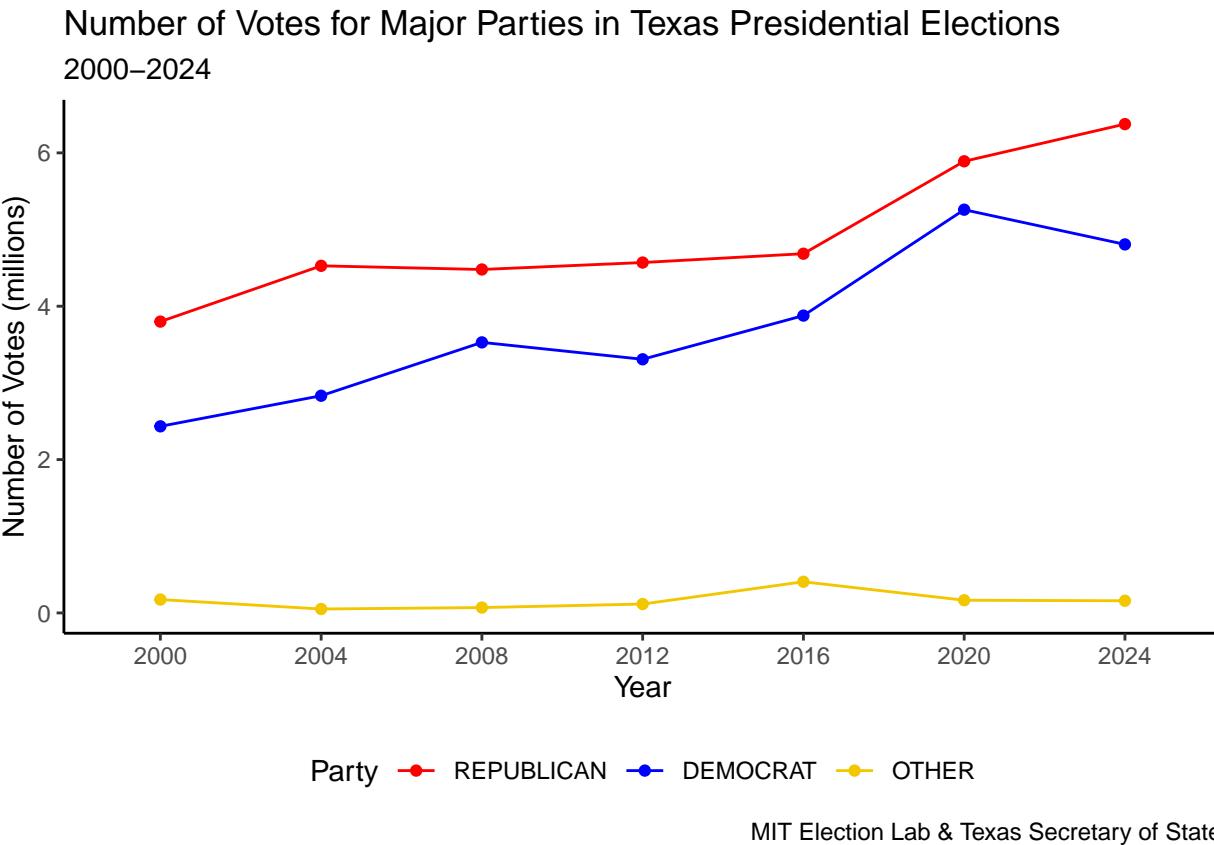
Number of Competitive Counties in Texas Presidential Elections  
2000–2024



## Plot votes by party over time

This plot shows the number of votes each party received in Texas Presidential elections from 2000 to 2024. This visualization is important to understanding the political trends in Texas' voting behavior.

```
ggplot(state_total_votes_party, aes(
  x = year,
  y = state_votes / 1e6, # Convert to millions
  color = party,
  group = party
)) +
  geom_line() +
  geom_point() +
  scale_color_manual(values = c(
    "DEMOCRAT" = "blue",
    "REPUBLICAN" = "red",
    "OTHER" = "#F2C700"
)) +
  theme_classic() +
  theme(legend.position = "bottom") +
  labs(
    x = "Year",
    y = "Number of Votes (millions)",
    color = "Party",
    title = "Number of Votes for Major Parties in Texas Presidential Elections",
    subtitle = "2000–2024",
    caption = "MIT Election Lab & Texas Secretary of State"
)
```



### Plot Competitive Margin Histogram

This plot shows the shifts in county competitiveness in Texas Presidential elections from 2000 to 2024. As seen in the visualization, concentration of non-competitive counties won by the Republican party has shifted rightward over time, meaning that these Republican counties have voted for the Republican candidate by larger margins over time. This visualization may be valuable in understanding the polarization of Texans between the two major parties.

```
ggplot(data_texas_2000_2024, aes(x = difference, fill = result)) +
  geom_histogram(binwidth = 0.025) +
  facet_wrap(~year, ncol = 2) +
  scale_fill_manual(values = c(
    "DEMOCRAT" = "blue",
    "REPUBLICAN" = "red",
    "COMPETITIVE" = "purple"
  )) +
  labs(
    x = "Republican Margin",
    y = "Number of Counties",
    fill = "Competitive Versus Non-Competitive Results",
    title = "Presidential Election Margins for Each Major Party in Texas Counties",
    subtitle = "2000–2024"
  ) +
  theme(
    legend.position = c(0.98, 0),
```

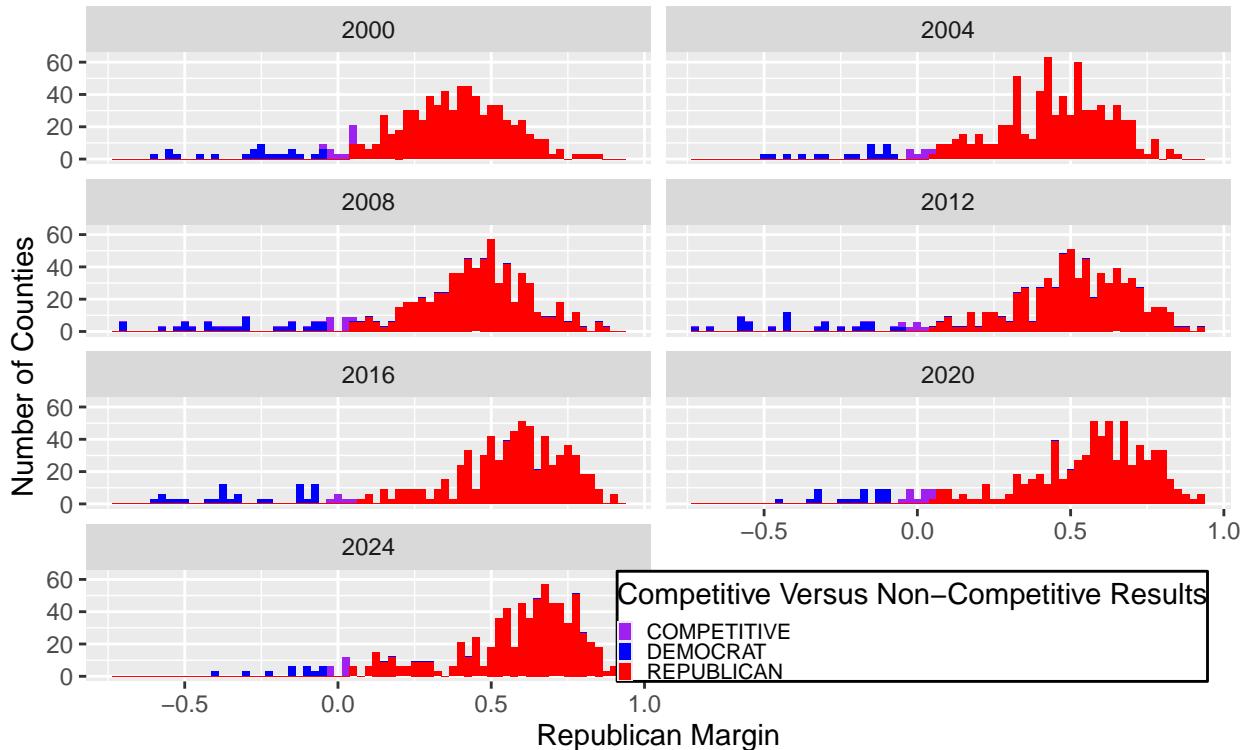
```

    legend.justification = c(1, 0),
    legend.background = element_rect(fill = "white", color = "black"),
    legend.key.size = unit(0.2, "cm"),
    legend.text = element_text(size = 8),
    legend.margin = margin(t = 3)
  )

## Warning: A numeric 'legend.position' argument in 'theme()' was deprecated in ggplot2
## 3.5.0.
## i Please use the 'legend.position.inside' argument of 'theme()' instead.
## This warning is displayed once every 8 hours.
## Call 'lifecycle::last_lifecycle_warnings()' to see where this warning was
## generated.

```

## Presidential Election Margins for Each Major Party in Texas Counties 2000–2024



## Prepare Shapefile Data and Maps

This plot shows 7 maps detailing the results of Texas Presidential elections by county from 2000 to 2024. We created this to try to see where political shifts are happening within the state.

```

# Read shapefile and join with data
texas_shape <- st_read("Texas_County_Boundaries/County.shp") %>%
  rename(county_fips = CNTY_FIPS) %>%
  mutate(county_fips = as.numeric(county_fips)) %>%

```

```

left_join(data_texas_2000_2024, by = "county_fips") %>%
st_sf()

## Reading layer 'County' from data source
##   'C:\Users\jylib\OneDrive\Documents\Case Western\ECON 216\FinalProject\Texas_County_Boundaries\County.shp'
##   using driver 'ESRI Shapefile'

## Warning in CPL_read_ogr(dsn, layer, query, as.character(options), quiet, : GDAL
## Message 1: C:\Users\jylib\OneDrive\Documents\Case Western\ECON
## 216\FinalProject\Texas_County_Boundaries\County.shp contains polygon(s) with
## rings with invalid winding order. Autocorrecting them, but that shapefile
## should be corrected using ogr2ogr for example.

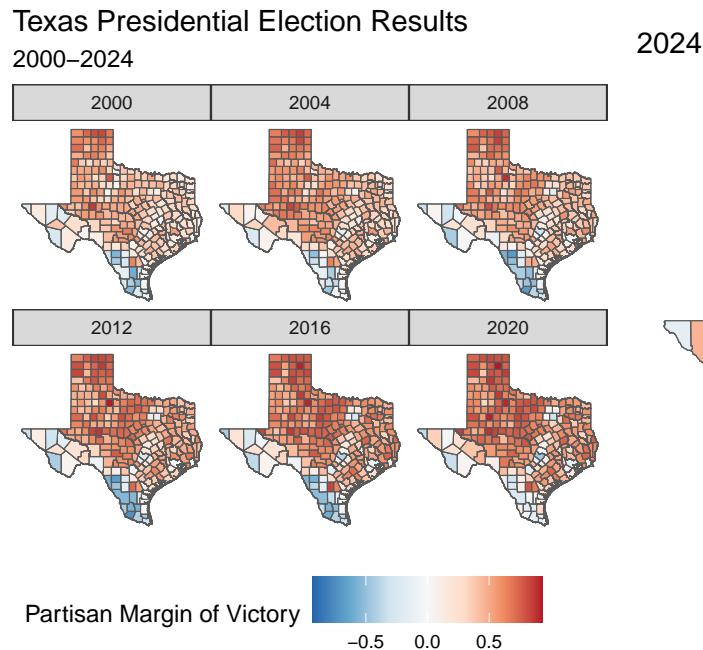
## Simple feature collection with 254 features and 12 fields
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:  xmin: -11871800 ymin: 2978934 xmax: -10409240 ymax: 4369694
## Projected CRS: WGS 84 / Pseudo-Mercator

# Create 2024 map
texas_plot_2024 <- ggplot(filter(texas_shape, year == "2024"), aes(fill = difference)) +
  geom_sf() +
  theme_map() +
  scale_fill_distiller(type = "div", palette = "RdBu", limits = c(-1, 1) * abs(max(texas_shape$difference)))
  labs(title = "2024") +
  theme(legend.position = "none")

# Create maps for 2000-2020
texas_plot_2000_2020 <- ggplot(filter(texas_shape, year != "2024"), aes(fill = difference)) +
  geom_sf() +
  facet_wrap(~year) +
  theme_map() +
  scale_fill_distiller(type = "div", palette = "RdBu", limits = c(-1, 1) * abs(max(texas_shape$difference)))
  labs(title = "Texas Presidential Election Results",
       subtitle = "2000-2024",
       fill = "Partisan Margin of Victory") +
  theme(legend.position = "bottom")

# Arrange maps side by side
ggarrange(texas_plot_2000_2020, texas_plot_2024, ncol = 2, nrow = 1)

```



## Competitive counties map

This map will highlight counties classified as “COMPETITIVE” (where the difference between the Democratic and Republican vote percentage is  $\leq 5\%$ ) for each election year. This will give us a glimpse into whether Texas has gotten more or less competitive.

```
# Prepare shapefile data with all counties classified
texas_shape_full <- texas_shape %>%
  mutate(fill_color = ifelse(result == "COMPETITIVE", "purple", "gray"))

# Create map showing competitive counties with others in gray
ggplot(texas_shape_full, aes(fill = fill_color)) +
  geom_sf() +
  facet_wrap(~year) +
  theme_map() +
  scale_fill_identity() +
  labs(
    title = "Competitive Counties in Texas Presidential Elections",
    subtitle = "2000–2024",
    fill = NULL
  )
```

Competitive Counties in Texas Presidential Elections  
2000–2024

