STA 9750 2024 Submission Material

mp03

Mini-Project #03: Exploring the Effect of State-Level Electoral College Vote Allocation on Presidential Elections

Do Proportional Electoral College Allocations Yield a More Representative Presidency?

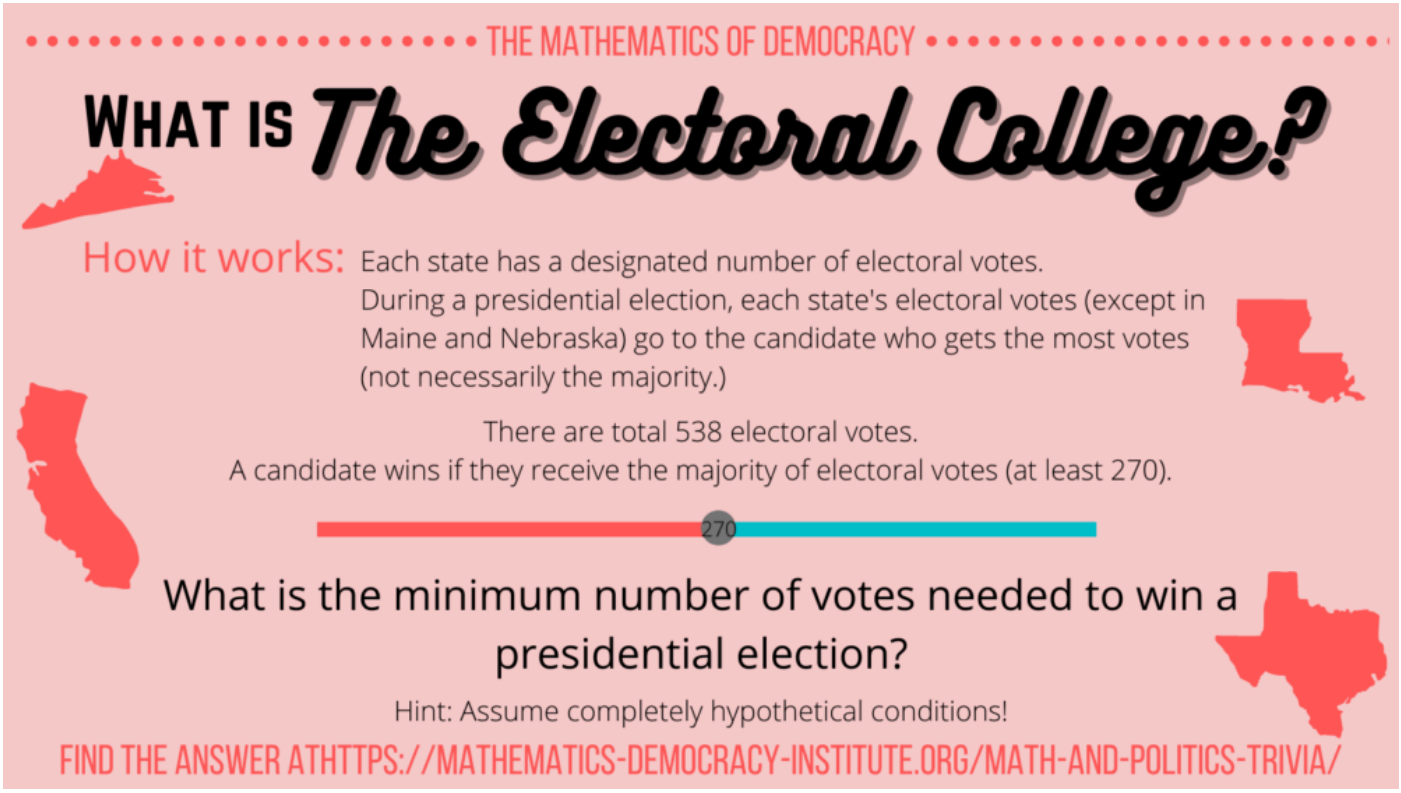
Abstact

This project aims to display a *political fact-check*, investigating the claim that the US Electoral College systemically biases election results away from the voice of the people (*vox populi).*

Background

As a refresh, the basic rules of electing the President of the United States, as outlined in the US Constitution is as follows. The selected visuals provide a clear understanding of how the electoral college allocates their votes in each state. For further background information, you may review the attached summary of the electoral college from the Institute For Mathematics and Democracy

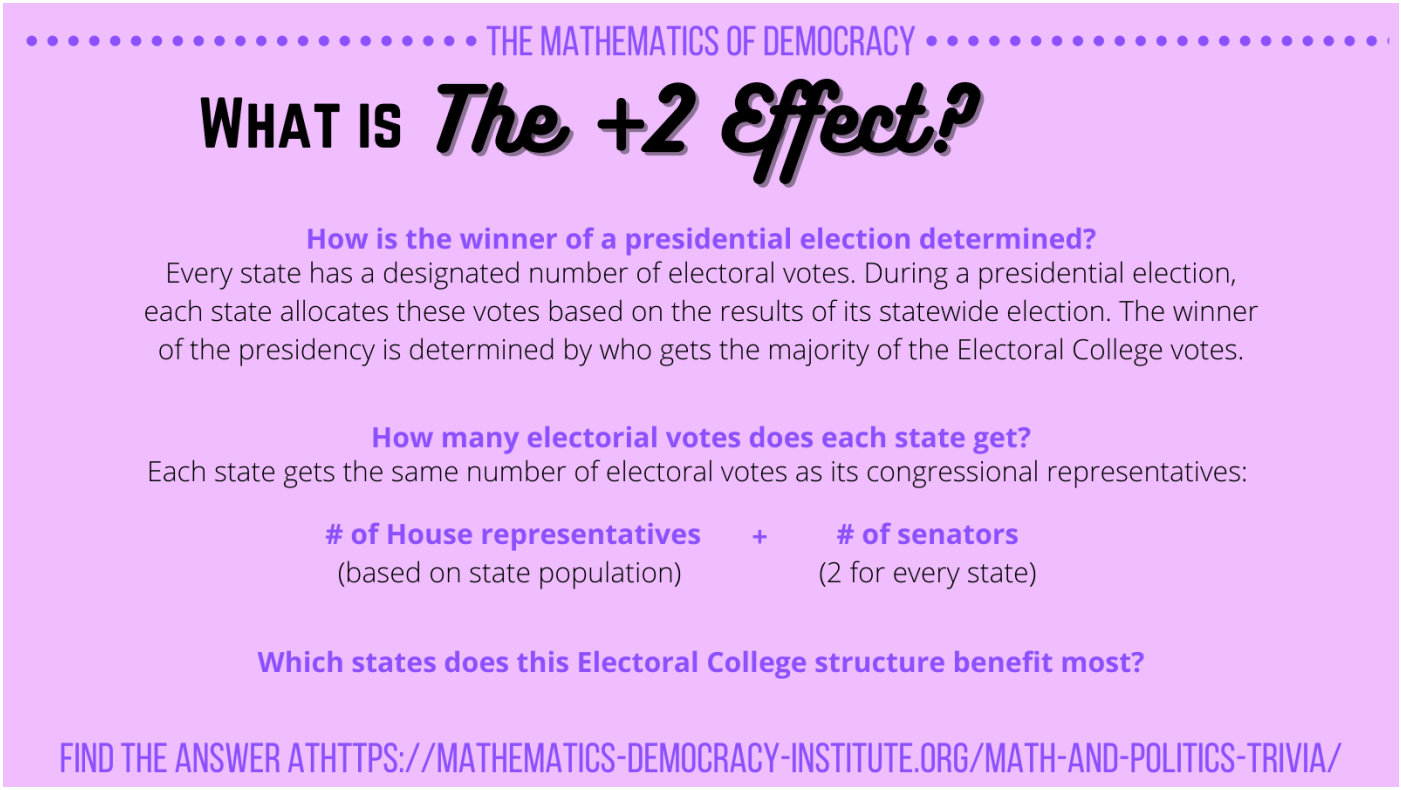
!Electoral College



Each state gets R + 2 electoral college votes, where R is the number of Representatives that state has in the US House of Representatives.

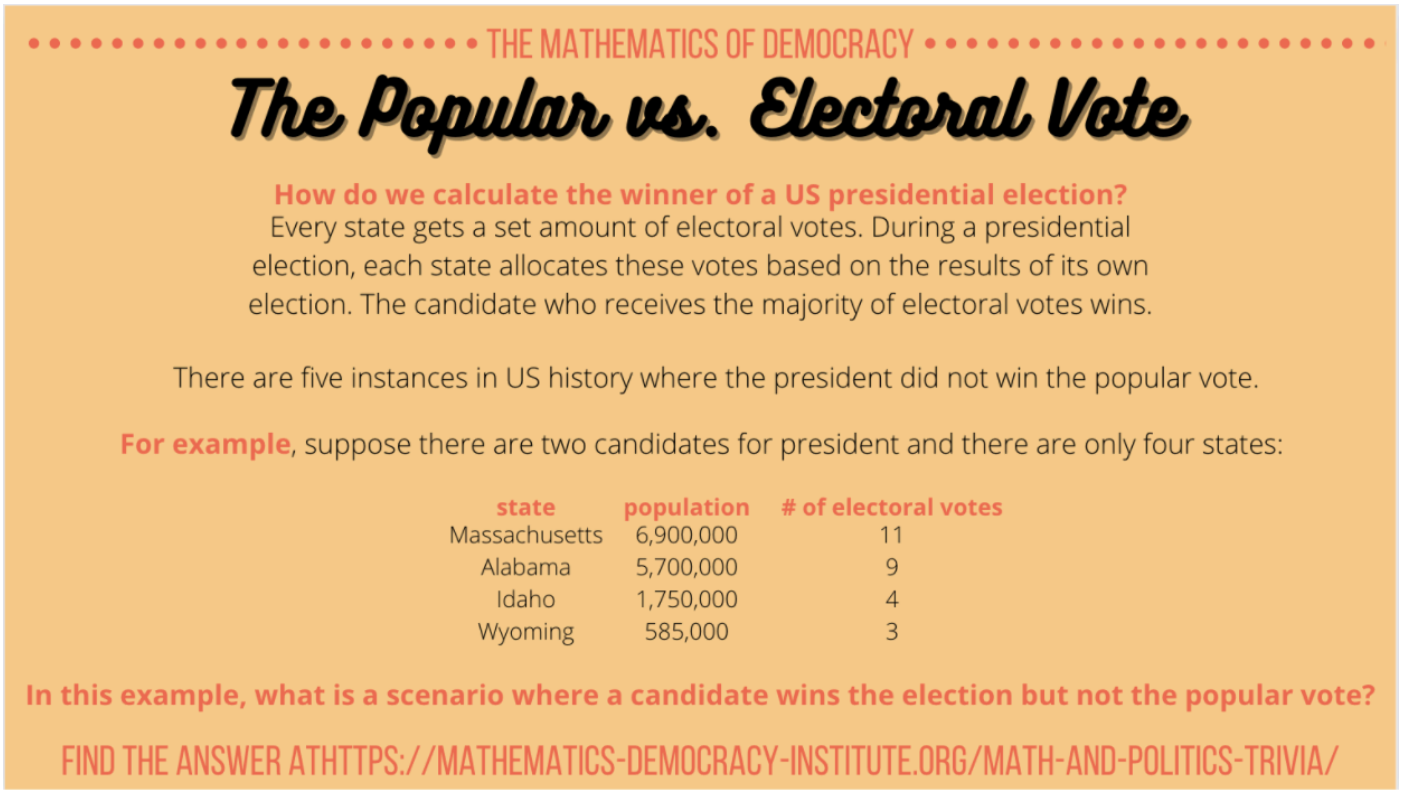
For the purposes of this project, the # of districts in a state = the # of congressional representatives

!Plus 2



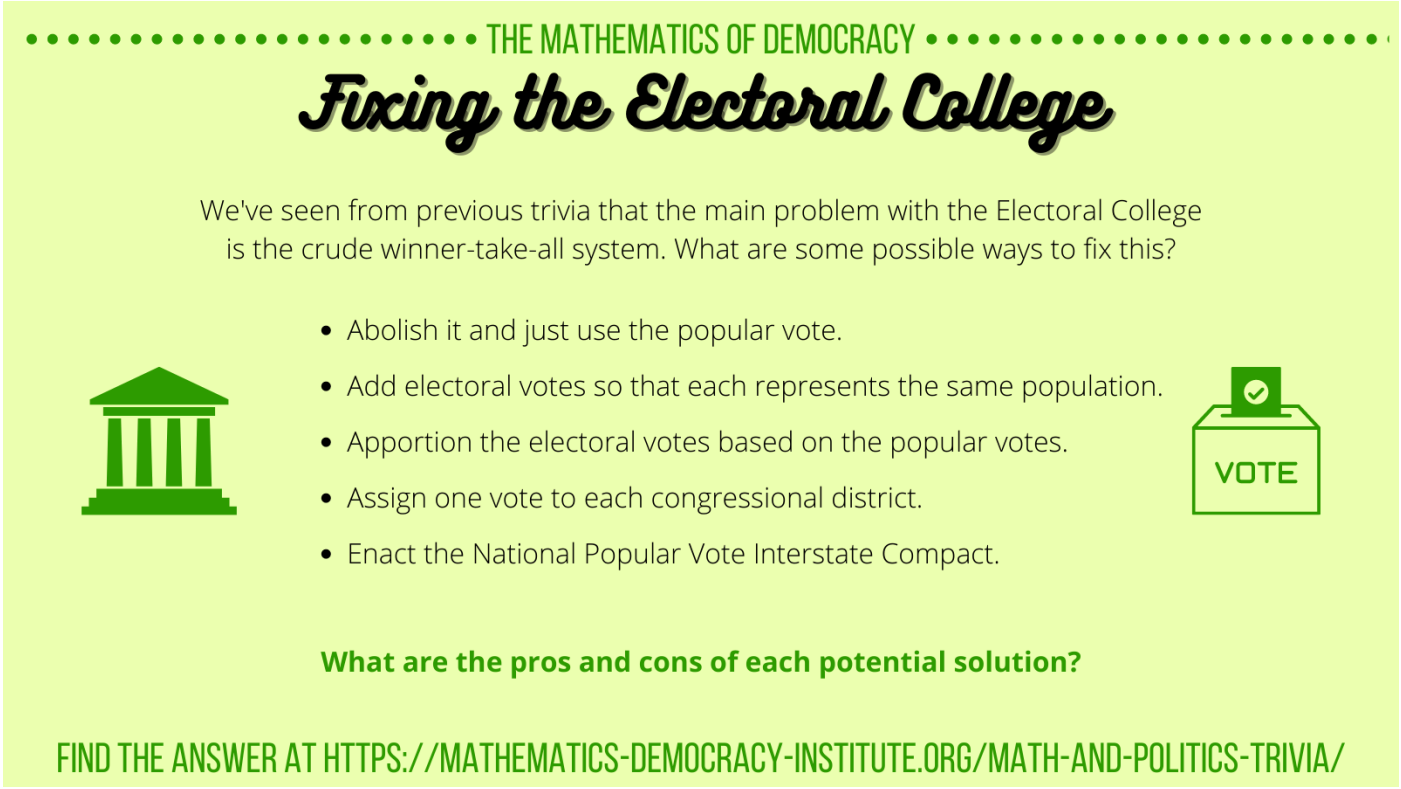
The “+2 effect” refers to the impact of the two additional electoral votes that every state receives due to its two Senate seats. While a state’s representation in the House—and therefore a portion of its electoral votes—reflects its population size, these two extra votes are the same for every state, regardless of population. As a result, individual votes in smaller states carry slightly more weight than in larger states because of this fixed “plus two” factor.

!Popular vs. Electoral Vote



This image explains the difference between the popular vote and the electoral vote in U.S. presidential elections. Each state has a set number of electoral votes, which are generally based on its population, but every state also gets two additional votes for its senators. This can create situations where the electoral vote winner does not align with the popular vote winner.

!Fixing the Electoral College



There are several proposed ideas for reforming the Electoral College, addressing criticisms of the current winner-take-all-system, used by all states other than Maine and Nebraska today.

Maine and Nebraska split Electoral College Votes (ECVs) by congressional district.

An important note is the allocation of ECVs is decided on a *state level,* and is not dictated by the Constitution. Reform is possible without constitutional amendments if states decided to change their approach.

This project will walk us through different methods of ECV allocation, and see how the outcome of the US presidential elections would have changed.

Set-Up and Initial Exploration

Data I: US House Election Votes from 1976 to 2022

This step involved downloading a dataset of the votes from all biennial congressional races in all 50 states and saving as a CSV file using a web browser. The data came from the MIT Election Data Science Lab.

Another dataset downloaded was the statewide presidential vote counts from 1976 to 2022, which may be found

The code below loads the election data into our environment.

▶ Code

▶ Code

Data II: Congressional Boundary Files 1976 to 2012

Task 1: Download Congressional Shapefiles 1976-2012

&

Task 4: Automate Zip File Extraction

The congressional district shapefiles, covering U.S. congressional districts from 1789 to 2012, are provided by Jeffrey B. Lewis, Brandon DeVine, Lincoln Pritcher, and Kenneth C. Martis. These files can be accessed on the Harvard Dataverse here.

The code below automates the download of files from the 94th to the 112th Congresses (1976-2012) and extracting the shapefiles from within the zip files.

The steps taken involve:

- Defining the function `get_congressional_shapefile()`, which automates the download and import of congressional district shapefiles for the specified range of congressional sessions (94 to 112)
- Storing each shapefile in a zip archive while the function checks if the file already exists locally. If not, file is downloaded. This avoids any redundant downloads.
- Unzipping downloaded zip file, and the shapefile within is read into R using `read_sf()`

A few additional steps were added to handle the large and complex shapefiles. The code includes a few lines involved in standardizing and simplifying the shapefiles.

The steps taken involve:

- `st_transform()` is applied to reproject each shapefile to a common Coordinate Reference System (CRS), which uses latitude and longitude coordinates, which are useful when working with geographic data
- `st_simplify()` simplifies the geometries of each shapefile. This reduces the number of intricate edges and details, making for improved performance while maintaining accuracy
- `sf_use_s2(FALSE)` is set to avoid any issues with spherical geometry when simplifying data in a projected CRS

Finally, we consolidate all shapefiles into the spatial object ALL_SHAPES.

▶ Code

Data III: Congressional Boundary Files 2014 to Present

Task 2: Download Congressional Shapefiles 2014-2022

▶ Code

Initial Exploration of Vote Count Data

Task 3: Exploration of Vote Count Data

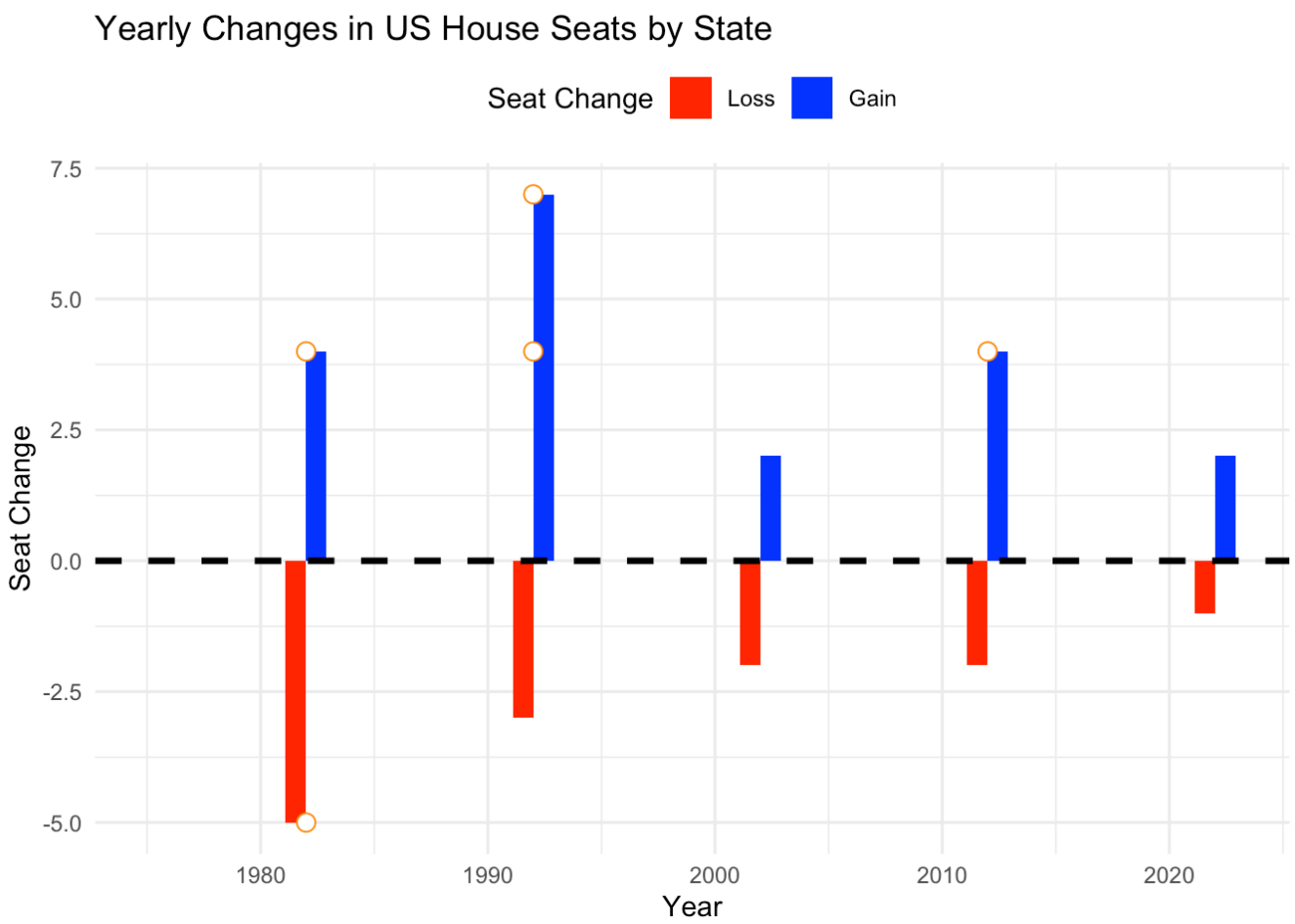
1. States Gaining and Losing House Seats (1976-2022)

For a better understanding to how the number of Representatives per state is determined, refer to this essay for more information.

> In summary,
>
> The number of US House seats per state can change every **10 years**, following the decennial census conducted by the US Census Bureau. This census determines the population distribution across states, and seats in the House of Representatives are then reapportioned based on these updated population counts.
>
> The national total number of House seats is fixed at **435**, but individual states may gain or lose seats based on shifts in the population.

First attempt: Not showing each state

▶ Code



Recognizing the need for consistent state abbreviations to enhance the clarity of visualizations, I identified that the `house_seats_by_year` table lacked state abbreviations, which would make presenting net changes by state more straightforward. To address this, I used R's built-in datasets, specifically `state.name` and `state.abb`, which contain full state names and their corresponding abbreviations.

The next few chunks of code modify the house_seats_by_year table.

```
state_abbreviations <- data.frame( state_name = state.name, state_abbreviation = state.abb
)
```

Adjusting title case in house_votes table to match the format in house_seats_by_year table, for the state column.

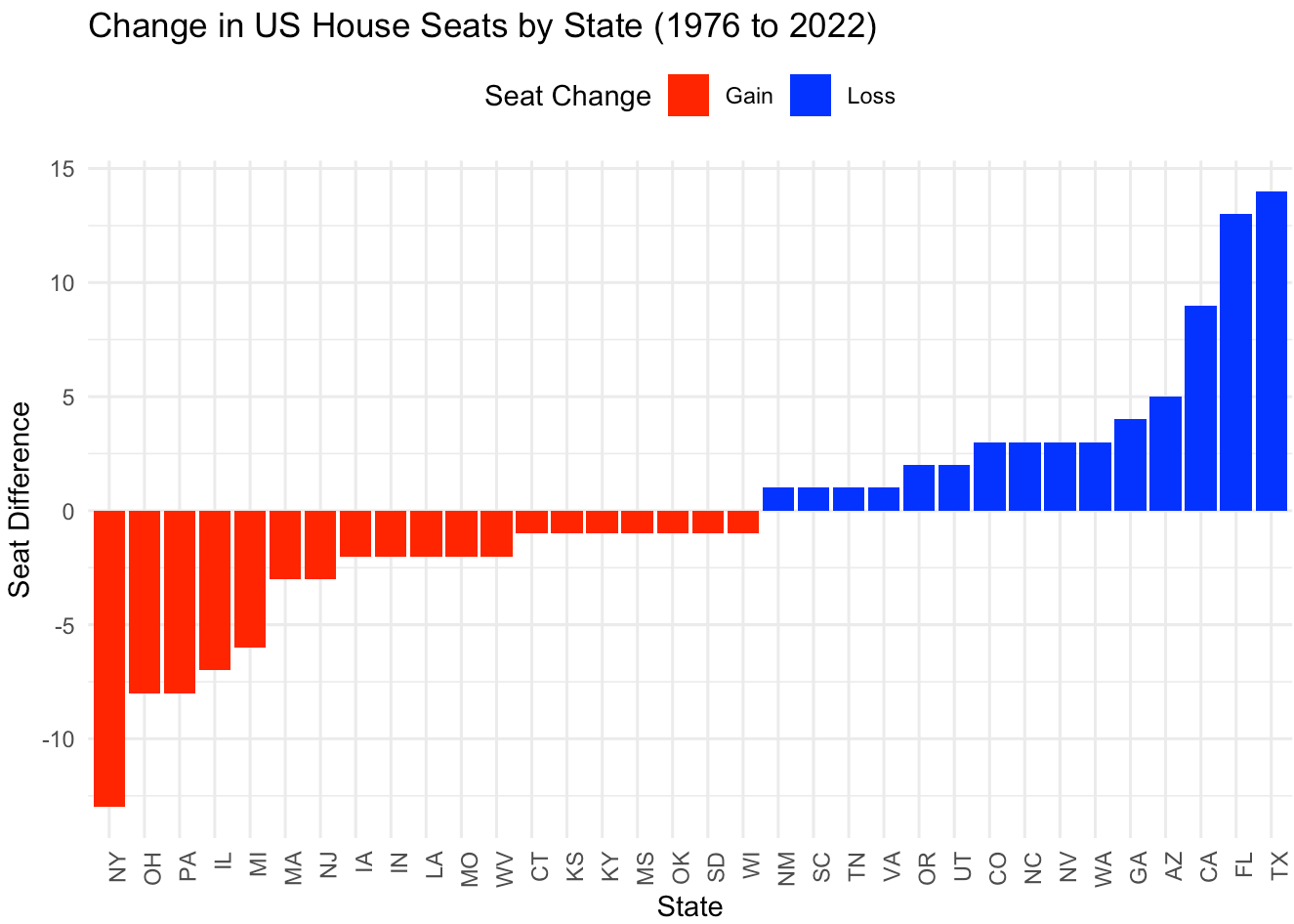
With both tables now in a consistent format, I have `house_seats_by_year` and `house_votes` tables where state names are in title case, and `house_seats_by_year` now includes a `state_abbreviation` column. These adjustments ensure that visualizations involving state abbreviations will be clean and standardized, with both tables ready for further analysis.

Now, going back to the original question, the code below analyzes changes in the US House representation by comparing the number of congressional districts in each state between 1976 and 2022.

The difference in the number of House seats in 2022 to 1976 is calculated.

The resulting data, found in the bar chart below, shows how many House seats each state gained or lost, from 1976 to 2022, excluding any states with no net change.

▶ Code



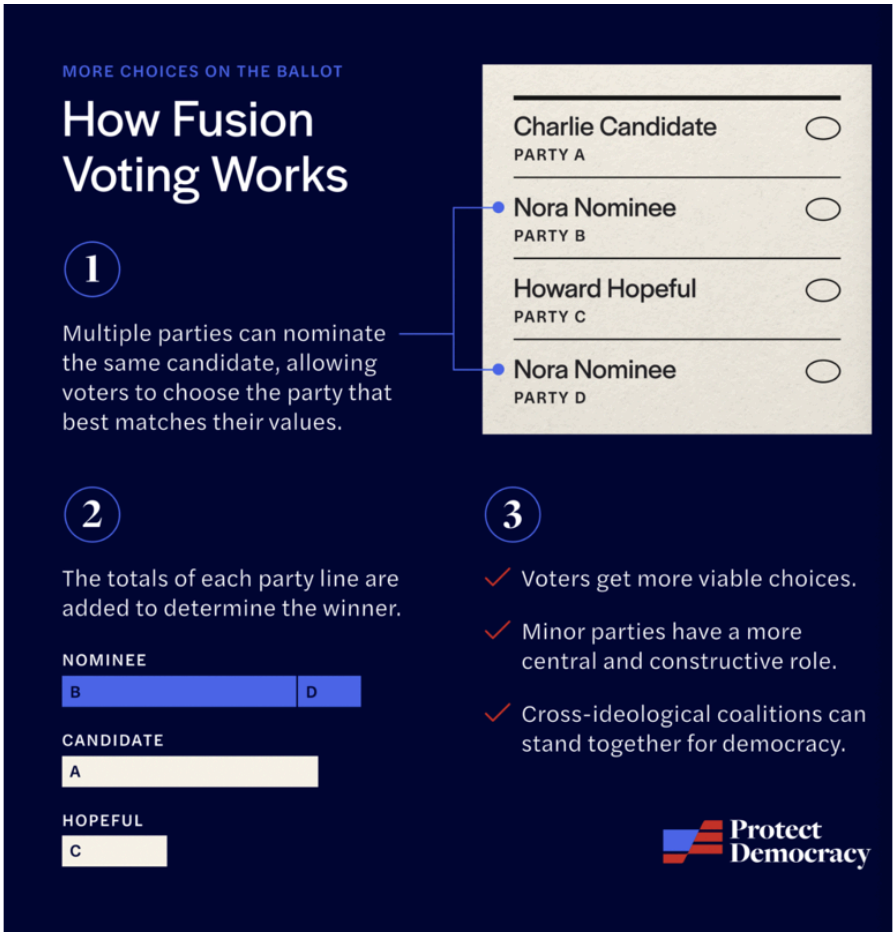
The code below highlights the top 5 states with the greatest net difference in number of House seats from 1976 to 2022.

▶ Code

Top 5 States with Seat Differences (1976 to 2022)	
State	Seat Difference
TX	14
FL	13
NY	−13
CA	9
OH	−8

As shown above, we find TX has the greatest difference, with 14 additional seats as of 2022, compared to 1976.

2. Analyzing NY's "fusion" voting system

> Please go here for a further understanding of Fusion Voting
>
> 
>

The code below involves a few steps to get us to test if there are any elections in our data where the election would have had a different outcome if the "fusion" system was not used and candidates only received the votes they received from their "major party line", instead of their total number of votes across all lines.

We limit our data to New York elections, where fusion voting occurs.

We create the data frames `major_party_votes` and `total_votes` to calculate and compare the election outcomes under different scenarios.

- `major_party_votes` (hypothetical outcome): contains the total number of votes each candidate received from their **major party line only** (Democrat or Republican) and ignores any votes they received from minor parties. [1]

- `total_votes` (actual outcome): contains the actual total number of votes each candidate received across **all party lines**, which is the real vote total used in New York State's fusion system

For each district in New York State, we check if the candidate with the highest `major_party_votes` differs from the candidate with the highest `total_votes`.

Our data frame hypothetical_winners lists any elections where the outcome would have changed if only major party line votes were counted, or our hypothetical scenario. [2]

▶ Code

```
major minor
 1419  2560
```

▶ Code

```
# A tibble: 0 × 6
# ℹ 6 variables: year <dbl>, district <dbl>, actual_winner <chr>,
#   actual_winner_party <chr>, hypothetical_winner <chr>,
#   hypothetical_winner_party <chr>
```

3. Presidential Candidate Performance Relative to Congressional Candidates

   Given this is an in-depth question, we will chunk the code into steps, for better understanding.

   (For a reviewer, I am curious if you prefer the code in the single chunk format, like I did for #2, or this step-by-step format, shown below)

   Step 1: Aggregating Votes for House and Presidential Races by Year, State, and Party

   ▶ Code

Step 2: Merge Aggregated Tables

▶ Code

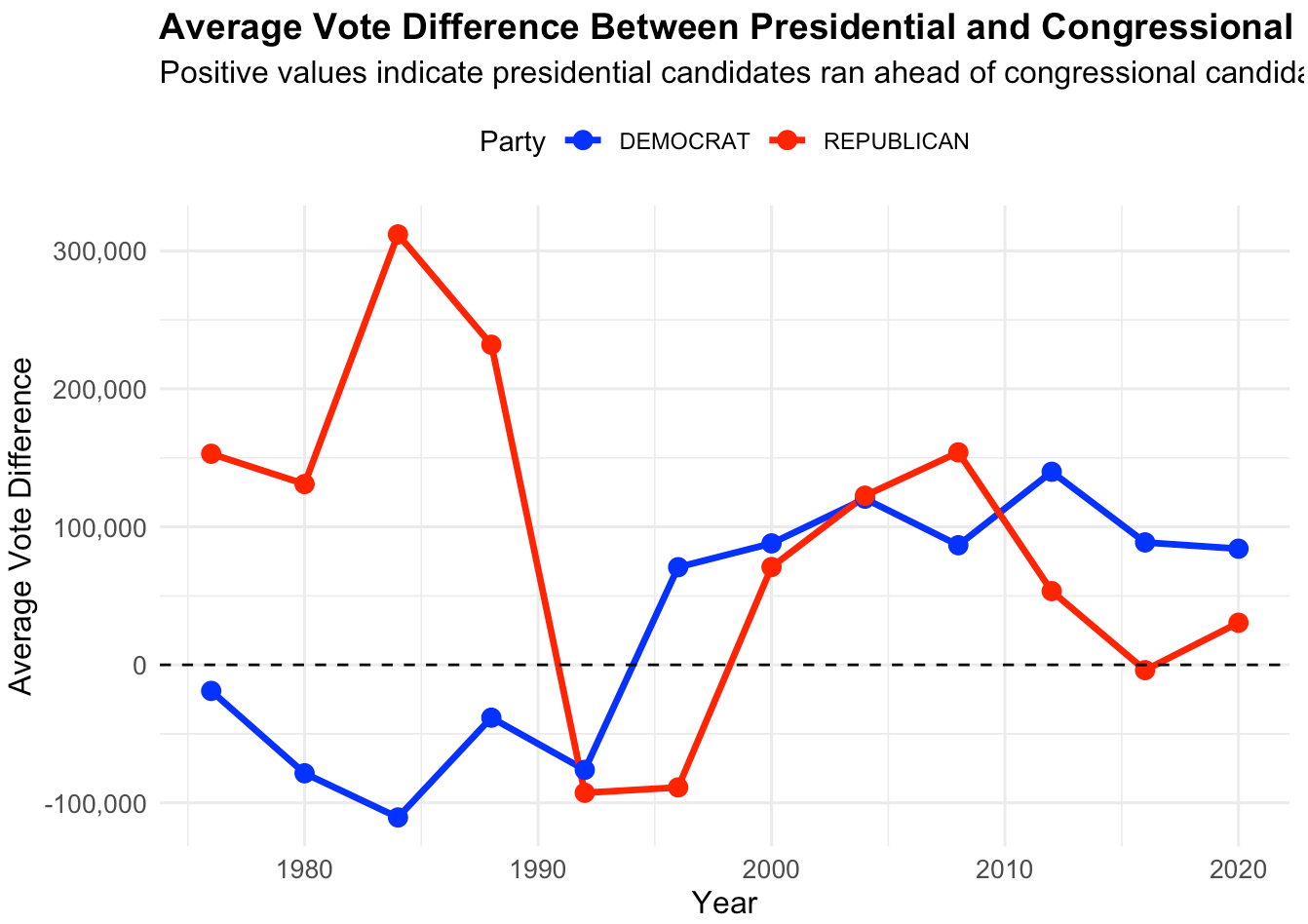
Step 3: Calculate Difference Between Presidential and House Votes

```
vote_difference = presidential_total_votes - house_total_votes
```

▶ Code

Step 4: Generate Visualization. The code below creates a time series plot showing the trend of average vote differences between presidential and congressional candidates across major parties from 1976 to 2020.

▶ Code

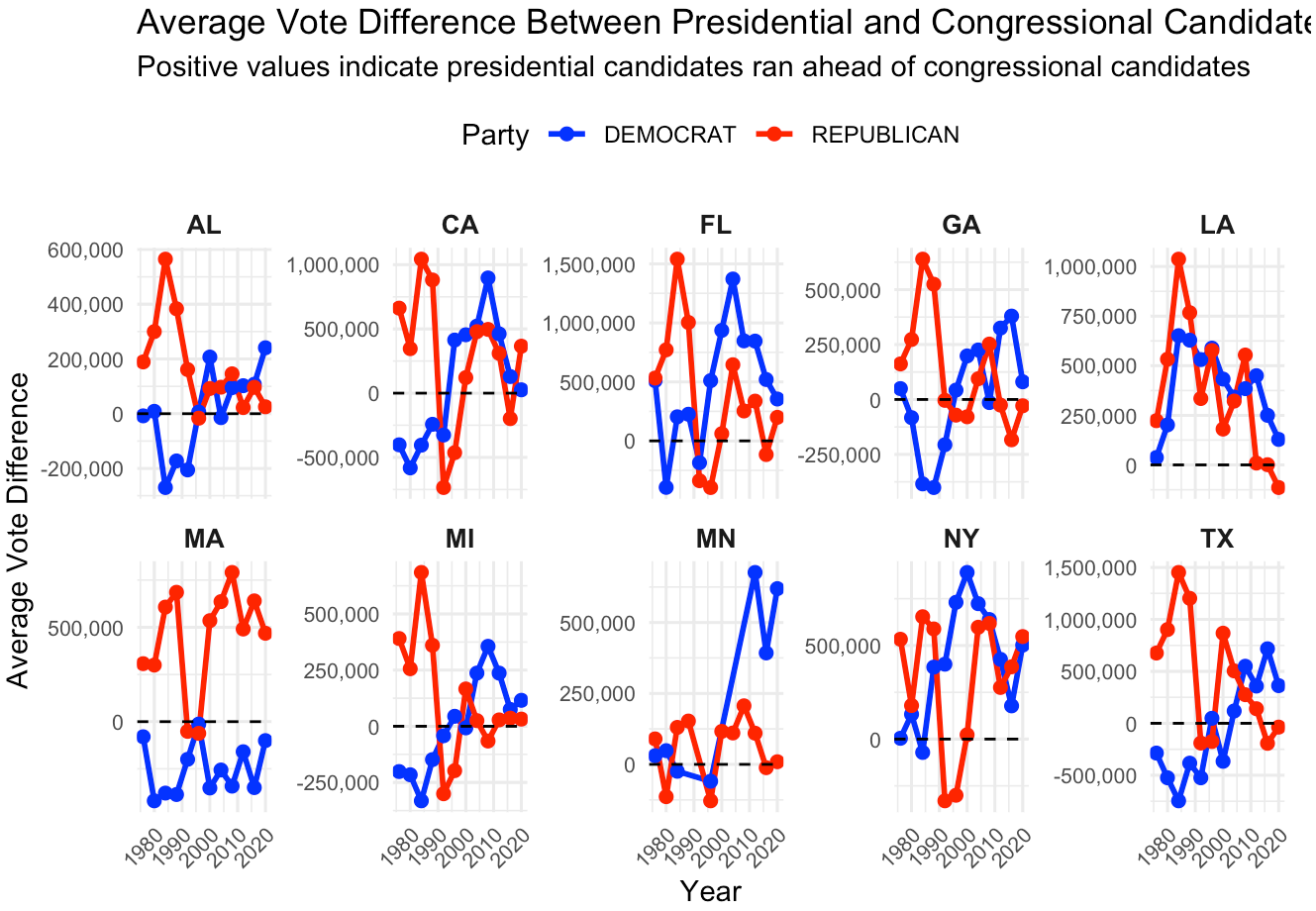


From analyzing the time series plot above, Republican presidential candidates (in red) typically received more votes than congressional candidates in the 1980's and early 1990s. Democratic presidential candidates (in blue) tended to receive fewer votes than their congressional counterparts throughout this same time period (1980s- early 1990s).

The code below aims at answering the second part to #3 - Does it differ across states or across parties? Are any presidents particularly more or less popular than their co-partisans?

This code identifies the 10 states with the largest vote differences between presidential and congressional candidates from the same party. It then creates a series of line plots, one for each state, showing whether presidential candidates received more or fewer votes than congressional candidates over time. Positive values mean presidential candidates ran ahead, while negative values mean they ran behind, making it easy to see trends for each state and party.

▶ Code



Importing and Plotting Shape File Data

Task 4: Automate Zip File Extraction

Please note, the code for Task 1, simultaneously completes both tasks 1 and 4.

The function `get_congressional_shapefile()` combines downloading the files, extracitng the shapefile, and reading into R. The function `get_congressional_shapefile()` serves the same purpose as `read_shp_from_zip()`, as provided in the sample code for Task 4.

Given the function `get_congressional_shapefile()` handles the downloading, extraction, and reading all in one step, this function is even more streamlined than using the separate read_shp_from_zip() function.

See a more detailed explanation for the combined steps taken for Tasks 1 & 4 outlined under Task 1.

Task 5: Chloropleth Visualization of the 2000 Presidential Election Electoral College Results

---

Footnotes

1. Where party = NA (no party listed), these values are excluded and considered minor party votes. ↩

2. We assume a candidate wins if they have the most votes in their district. ↩️