Election Analysis

Erma, Hiep, Jason, Mike, Blake

Project Summary: The United States Senate

For our final group project, we are examining U.S. Senate election results.

We're looking specifically at the relationship between election results and:

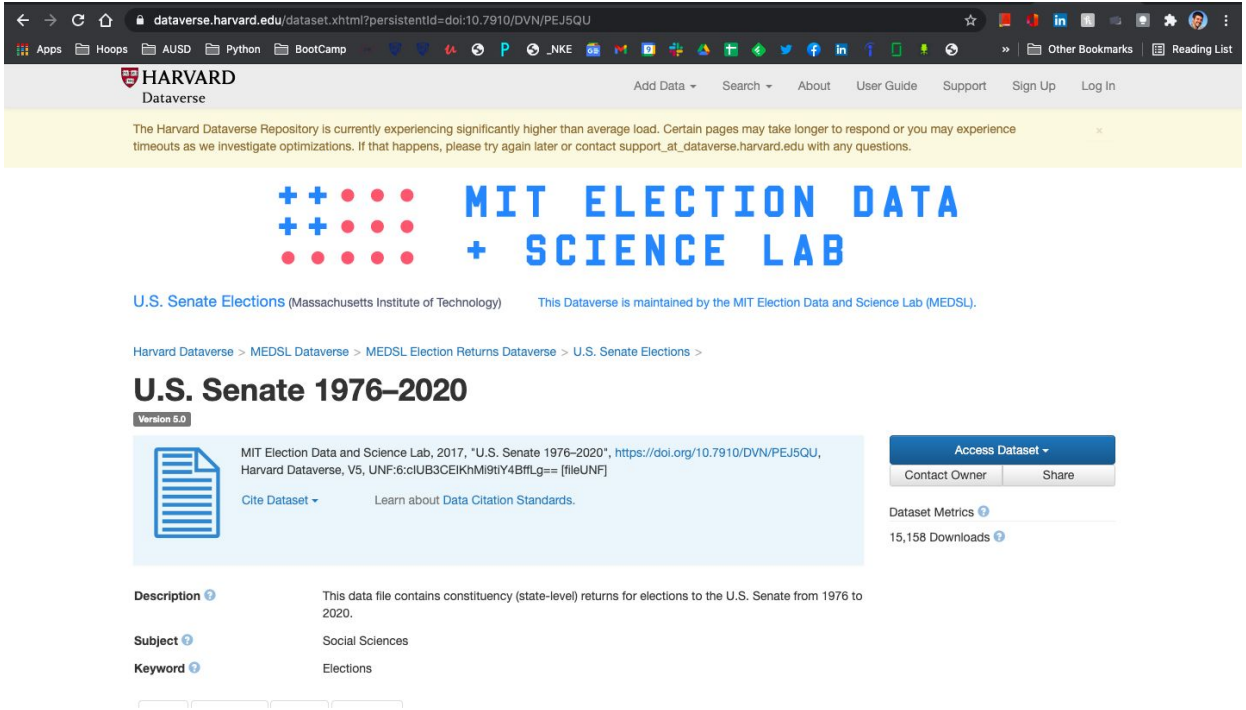
- Senate election campaign spending
- State demographics
- Changing partisan trends
- Voter turnout

Why the Senate?

- Given the standing of the U.S. in global affairs, U.S. Senators are some of the most powerful people in the world
- The Senate is arguably the most important legislative body in the United States
- Given the structure of the Senate, for example with filibuster rules that generally require 60 votes for legislation to pass, most legislation that passes in the House does not pass the Senate
- This is especially the case in today's polarized and highly partisan environment
- We set out to understand how much it costs to become a U.S. Senator. We also set out to understand what happened in 2020 in regards to the Senate and what that means for the midterm elections and beyond.

Datasets - U.S. Senate 1976-2020

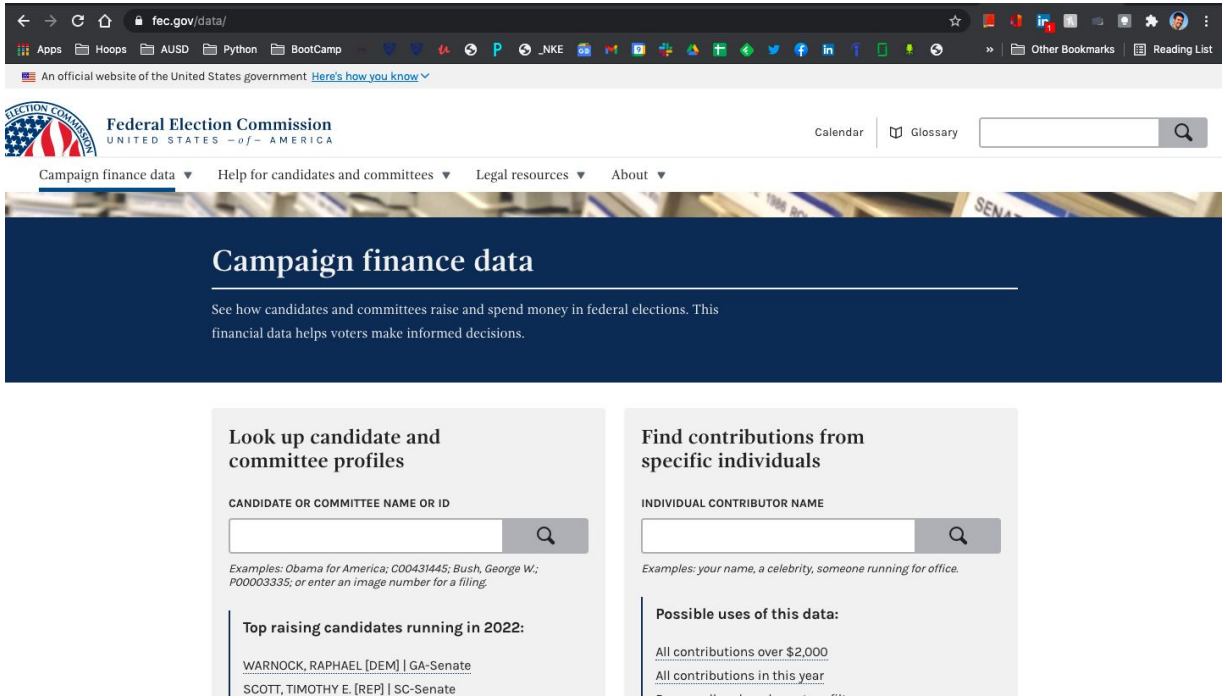
- MIT Election Data and Science Lab
- 100 U.S. Senators
  - Six-year terms
  - About 33 elections every two years
  - ~1500 elections



Source: https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/PEJ5QU

Datasets - FEC

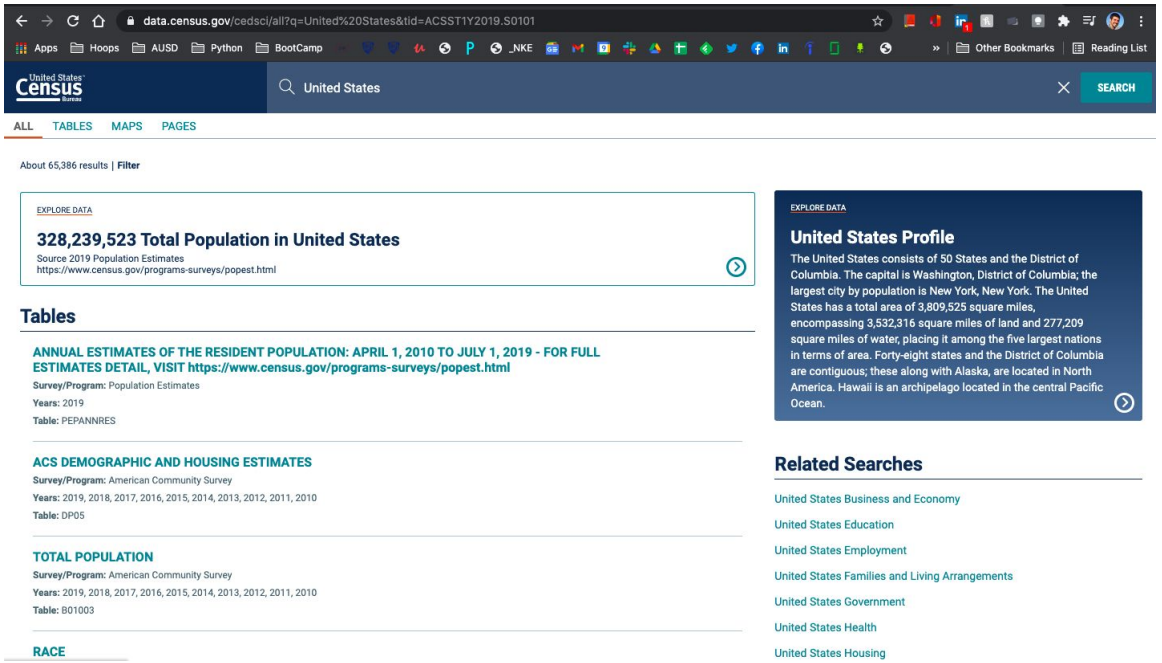
- Federal Election Commission (FEC) reports (2004-2020)
  - $14 billion spent in total on 2020 federal elections
  - Previous record <$7 billion (2016)
  - Campaigns are required to report every dollar they spend, and how it's spent
  - Correlation not necessarily causation
  - OpenSecrets.org



Source: https://www.fec.gov/data/

Datasets - U.S. Census

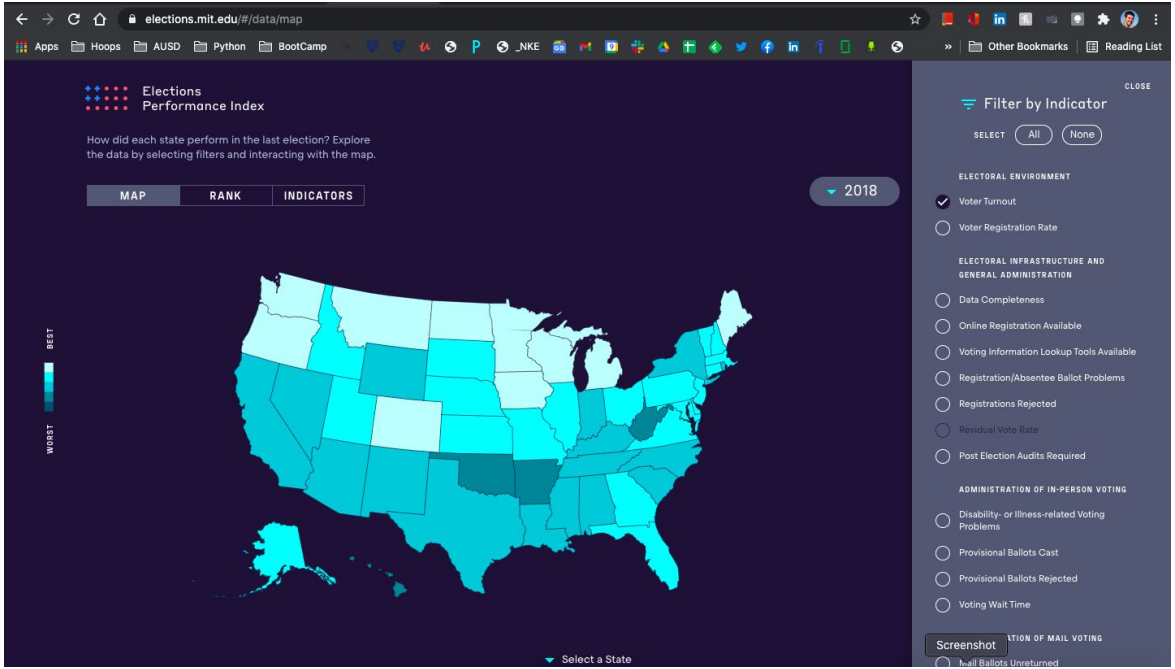
- U.S. Census every ten years, but most recent data (2020) just being released now
- Last official “Decennial Census” data is from 2010
- In 2005 the U.S. Census Bureau launched the American Community Survey, population and demographic data collected monthly and reported annually



Source: https://data.census.gov/cedsci/

Datasets - Elections Performance Index

- MIT Election Data and Science Lab
- Compares election administration **policy** and **performance** across states and over time
- Indicators include registration rates, how many ballots are rejected, wait times in line
- VEP turnout



Source: https://elections.mit.edu/#/data/map

Senate Data

Data Output

	index bigint	year bigint	state text	state_po text	state_fips bigint	state_cen bigint	state_ic bigint	office text	district text	stage text	special boolean	candidate text	party_detailed text	writein boolean	mode text	candidatevote bigint
1	0	1976	ARIZONA	AZ	4	86	61	US SENA...	statewide	gen	false	SAM STEIGER	REPUBLICAN	false	total	
2	1	1976	ARIZONA	AZ	4	86	61	US SENA...	statewide	gen	false	WM. MATHE...	INDEPENDENT	false	total	
3	2	1976	ARIZONA	AZ	4	86	61	US SENA...	statewide	gen	false	DENNIS DECO...	DEMOCRAT	false	total	
4	3	1976	ARIZONA	AZ	4	86	61	US SENA...	statewide	gen	false	ALLAN NORW...	LIBERTARIAN	false	total	
5	4	1976	ARIZONA	AZ	4	86	61	US SENA...	statewide	gen	false	BOB FIELD	INDEPENDENT	false	total	
6	5	1976	CALIFOR...	CA	6	93	71	US SENA...	statewide	gen	false	JACK MCCOY	AMERICAN INDEPE...	false	total	
7	6	1976	CALIFOR...	CA	6	93	71	US SENA...	statewide	gen	false	S. I. (SAM) HA...	REPUBLICAN	false	total	
8	7	1976	CALIFOR...	CA	6	93	71	US SENA...	statewide	gen	false	JOHN V. TUN...	DEMOCRAT	false	total	
9	8	1976	CALIFOR...	CA	6	93	71	US SENA...	statewide	gen	false	OMARI MUSA	INDEPENDENT	false	total	
10	9	1976	CALIFOR...	CA	6	93	71	US SENA...	statewide	gen	false	DAVID WALD	PEACE AND FREED...	false	total	
11	10	1976	CONNEC...	CT	9	16	1	US SENA...	statewide	gen	false	LOWELL P. WE...	REPUBLICAN	false	total	
12	11	1976	CONNEC...	CT	9	16	1	US SENA...	statewide	gen	false	SCATTER	[null]	false	total	
13	12	1976	CONNEC...	CT	9	16	1	US SENA...	statewide	gen	false	ROBERT BAR...	AMERICAN INDEPE...	false	total	
14	13	1976	CONNEC...	CT	9	16	1	US SENA...	statewide	gen	false	GLORIA SCHA...	DEMOCRAT	false	total	
15	14	1976	DELAWA...	DE	10	51	11	US SENA...	statewide	gen	false	THOMAS C. M...	DEMOCRAT	false	total	
16	15	1976	DELAWA...	DE	10	51	11	US SENA...	statewide	gen	false	WILLIAM V. R...	REPUBLICAN	false	total	
17	16	1976	DELAWA...	DE	10	51	11	US SENA...	statewide	gen	false	DONALD G. GI...	AMERICAN	false	total	
18	17	1976	DELAWA...	DE	10	51	11	US SENA...	statewide	gen	false	JOHN A. MAS...	PROHIBITION	false	total	
19	18	1976	DELAWA...	DE	10	51	11	US SENA...	statewide	gen	false	JOSEPH F. MC...	NONE	false	total	
20	19	1976	FLORIDA	FL	12	59	43	US SENA...	statewide	gen	false	LAWTON CHIL...	DEMOCRAT	false	total	
21	20	1976	FLORIDA	FL	12	59	43	US SENA...	statewide	gen	false	SCATTER	[null]	false	total	
22	21	1976	FLORIDA	FL	12	59	43	US SENA...	statewide	gen	false	JOHN GRADY	REPUBLICAN	false	total	

With this dataset alone, tracking victory since 1976, can be used to predict winners based on trends. By combining this dataset with others that help us understand where and why the nation is changing, we can build an accurate model to predict how and why different candidates can win in purple states.

FEC Data

Data Output

	index bigint	Unnamed: 0 bigint	report_year bigint	image_number double precision	file_number double precision	payee_name text	expenditure_date text	dissemination_date text	expenditure_amount double precision	category_code_full text	sup text
1	0	0	2020	2.0201e+17	1444283	Good Works Matt...	5-Oct-20	[null]	35000	Phone bankers for feder...	S
2	1	1	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	85262.15	Digital ads	S
3	2	2	2020	2.0201e+17	1466866	BERLIN ROSEN, L..	15-Oct-20	25-Oct-20	80000	Projected phone bank co...	S
4	3	3	2020	2.0201e+17	1466866	CENTURY DIREC...	19-Oct-20	25-Oct-20	3742	Mailer	S
5	4	4	2020	2.02011e+17	1470736	AB PARTNERS PBC	31-Oct-20	31-Oct-20	48093.24	Digital ads	S
6	5	5	2020	2.0201e+17	1467573	AB PARTNERS PBC	23-Oct-20	27-Oct-20	290000	Projected digital ad buy	S
7	6	6	2020	2.0201e+17	1445694	CallHub	9-Oct-20	7-Oct-20	1000	Phone bank platform	S
8	7	7	2020	2.0201e+17	1445694	CallHub	8-Oct-20	7-Oct-20	1000	Phone bank platform	S
9	8	8	2020	2.0201e+17	1445694	CallHub	7-Oct-20	7-Oct-20	500	Phone bank platform	S
10	9	9	2020	2.0201e+17	1445694	1199SEIU United ...	3-Nov-20	7-Oct-20	0	Projected staff compens...	S
11	10	10	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	120451.5	Digital ads	S
12	11	11	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	120451.5	Digital ads	S
13	12	12	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	120451.5	Digital ads	S
14	13	13	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	120451.5	Digital ads	S
15	14	14	2020	2.0201e+17	1466866	AB PARTNERS PBC	23-Oct-20	25-Oct-20	200000	Digital ads	S
16	15	15	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	85262.15	Digital ads	S
17	16	16	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	85262.15	Digital ads	S
18	17	17	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	85262.15	Digital ads	S
19	18	18	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	2700	Digital ad setup	S
20	19	19	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	168.75	Digital ad setup	S
21	20	20	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	8979	Digital ads	S
22	21	21	2020	2.0201e+17	1445692	A/B Partners	8-Sep-20	8-Oct-20	6398.25	Digital ads	S

Understanding not only how much but also on what campaigns spend their money is essential to deconstructing how the senate works, and how to better run campaigns.

Data Exploration Phase

- Our group was formed around an interest in federal election-related data as this information reveals a lot about our democracy, and similar to the hope we have for fair elections with high turnout, our group also wanted to find unbiased, robust datasets.
- Focusing on the U.S. Senate, we tracked down a dataset of **election results** data for all elections between 1976-2020.
- We also found **demographic data** for the states we are analyzing.
- The Federal Election Commission requires the reporting of all **"hard money" campaign contributions**. These contributions are those that go toward advertisements that explicitly tell voters which candidate to vote for or against (as opposed to "soft money" contributions toward party-building activities). We located multiple contribution datasets and joined them to form a record of hard money contributions to U.S. Congress races between 2004-2020.
- Finally, we chose a data set from the Elections Performance Index, a project from MIT's Election Data and Science Lab. This index "compares election administration policy and performance across the states and from one election cycle to the next." This is helpful for indicators that can be compared state-to-state and over time, such as **voter turnout**.

Database

- The database has several different interacting datasets
  - The information ranges from 1976-2020
  - Most of the data overlaps from 2000-2020
- We plan to use our database to better understand how spending influences the success of political candidates
- We have data that helps us not only understand spending, but also turnout in the several states we intend to use.
- Demographics and turnout will also be considered in relation to spending and victory

Database Tools

The tools we used to build our database were:

- Postgres (PgAdmin) to import our data and to construct our new tables
- AWS to allow collaboration across multiple users
- DBdiagram for mapping - keys were fairly easy to find, tables were huge
    - Year, parties, states, etc.

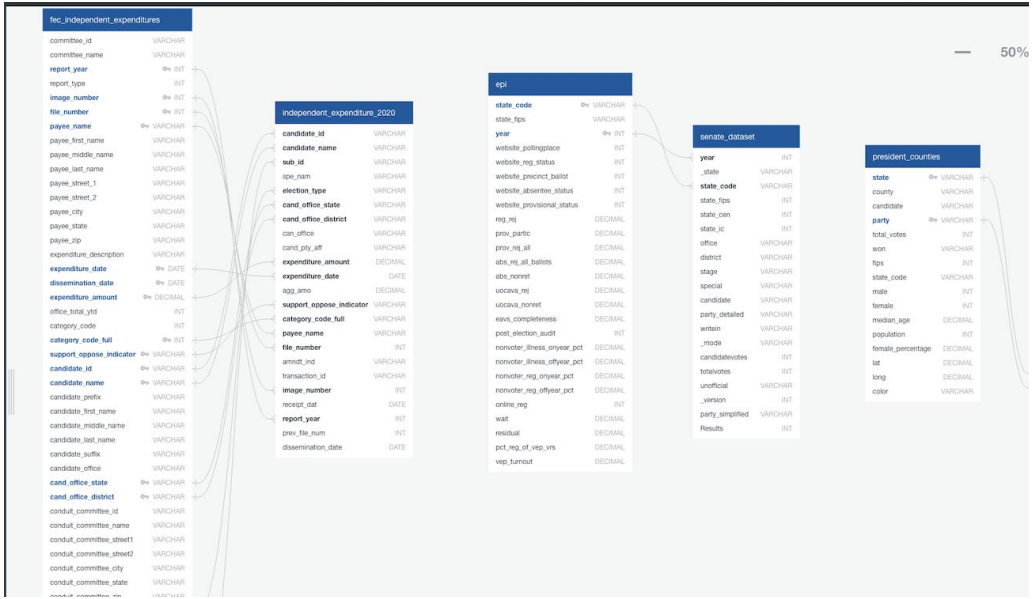


Table join

- Subqueries to help with optimization
  - Changed 30 column table to 5 columns
- Grouped on States, Years

```
--CREATE TABLE query
CREATE TABLE Support_Opposition AS (
WITH electresults AS (
    SELECT state_po,
        _year,
        "Results",
        candidate,
        party_detailed
    FROM senate_model
    )

    SELECT cand_office_state AS State,
        report_year AS Year,
        support_oppose_indicator AS S_or_O,
        SUM(expenditure_amount) AS expenditure_amount
        --MAX(CASE WHEN "Results" = 1 THEN 'Win' ELSE 'Lose' END)
    FROM fec_independent_expenditures_original AS f
    INNER JOIN electresults AS r on state_po = cand_office_state
        AND _year = report_year
    GROUP BY cand_office_state, S_or_O, year
    ORDER BY Cand_office_state, year)
```

Demographic Information

Data Output

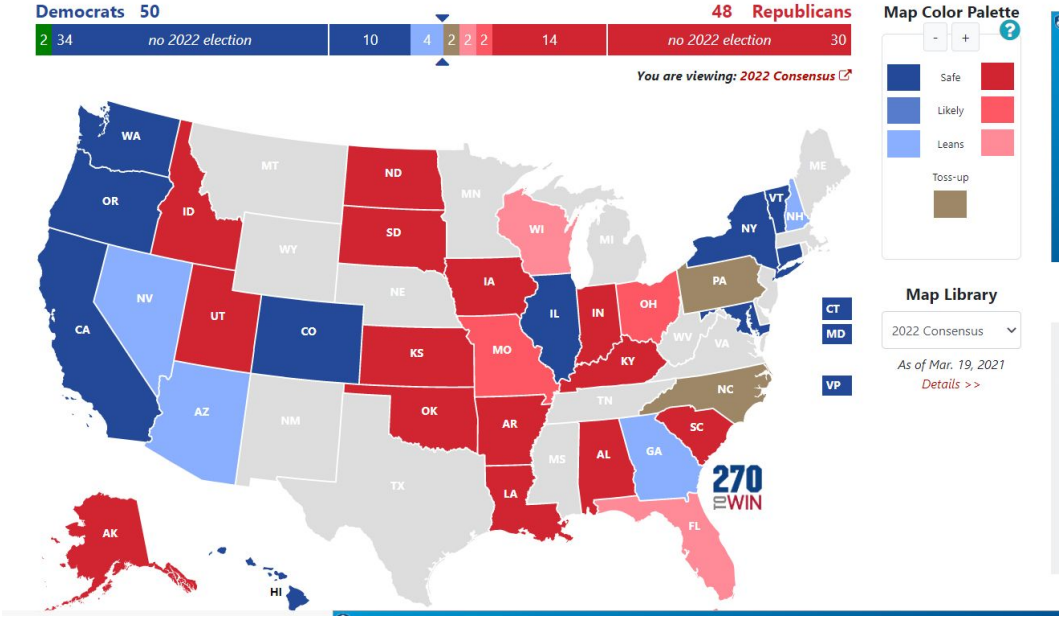
	index bigint	Year bigint	Gender text	Hispanic text	Race text	Population text	State text
1	0	2010	Male	Non Hispanic	White	1,832,998	Arizona
2	1	2010	Male	Non Hispanic	Black o...	127,328	Arizona
3	2	2010	Male	Non Hispanic	Americ...	126,085	Arizona
4	3	2010	Male	Non Hispanic	Asian	80,886	Arizona
5	4	2010	Male	Non Hispanic	Native ...	6,227	Arizona
6	5	2010	Male	Non Hispanic	Two or ...	54,380	Arizona
7	6	2010	Male	Hispanic	White	863,391	Arizona
8	7	2010	Male	Hispanic	Black o...	19,772	Arizona
9	8	2010	Male	Hispanic	Americ...	40,230	Arizona
10	9	2010	Male	Hispanic	Asian	7,985	Arizona
11	10	2010	Male	Hispanic	Native ...	2,509	Arizona
12	11	2010	Male	Hispanic	Two or ...	22,131	Arizona
13	12	2010	Female	Non Hispanic	White	1,874,603	Arizona
14	13	2010	Female	Non Hispanic	Black o...	114,756	Arizona
15	14	2010	Female	Non Hispanic	Americ...	132,838	Arizona
16	15	2010	Female	Non Hispanic	Asian	93,087	Arizona
17	16	2010	Female	Non Hispanic	Native ...	4,972	Arizona

- This data is important because it not only tells us how many people turned out to vote, but **who** voted.
- Winning elections is becoming more and more of a science, and understanding demographics and how to reach them is essential to victory.

Competitive States

- Georgia
- Pennsylvania
- Ohio
- Arizona
- Colorado
- New Hampshire
- South Carolina

We chose these states because they have a wide range of demographics, histories, and cultural differences. We hope to balance the states that lean blue with those that lean red and states from similar regions of the country with one another.
Note: We have chosen solid blue/red states to test our model and make sure it is working properly



Tableau

- We will use Tableau to create interactive dashboards for users to analyze senate data
- There will be dynamic filters which allows users to drill down to make more advanced analysis
- Add a hyperlink feature that switches back and forth between the Tableau dashboard and machine learning application
- https://public.tableau.com/profile/jason6879#!/vizhome/SenateRace_16205139382310/Dashboard1?publish=yes

Machine Learning

Link

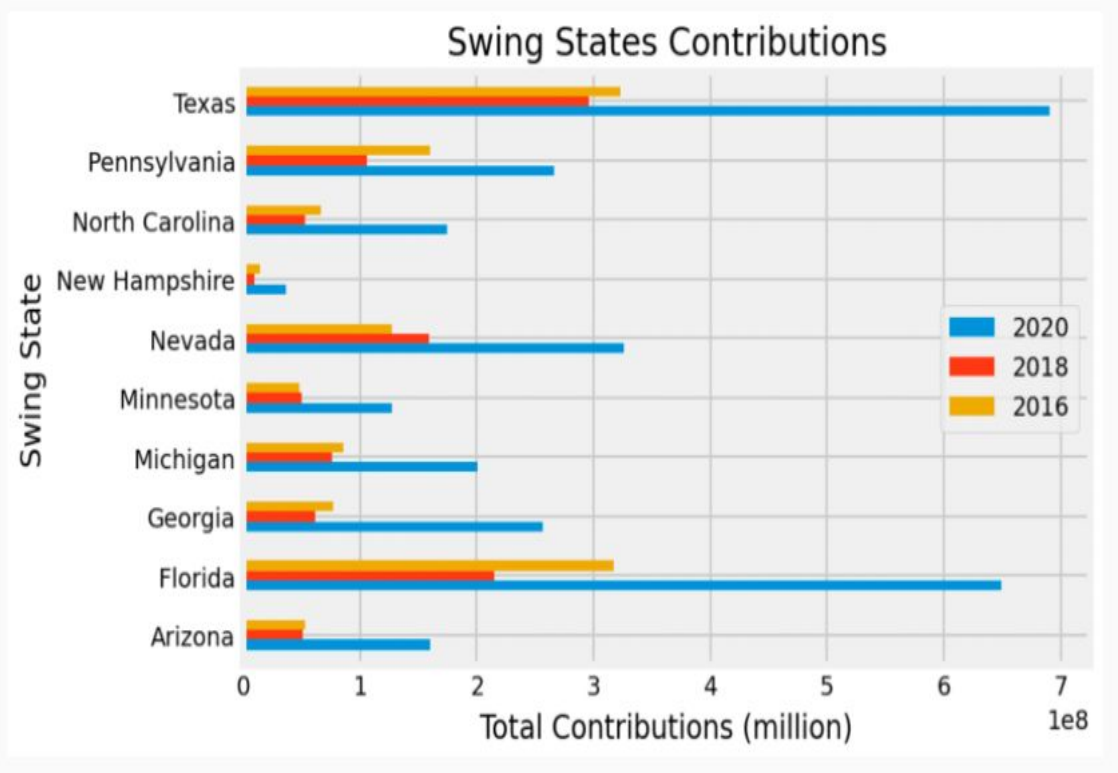
https://share.streamlit.io/hieppham8083/finalproject/main/main.py

https://predictsenate.anvil.app/

So what does this all mean?

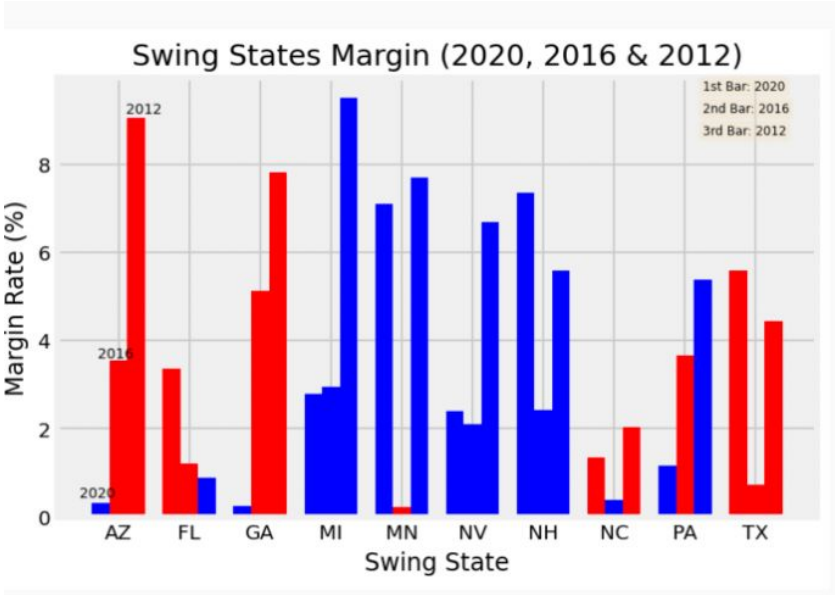
- Elections are becoming more and more expensive
  - Citizens United decision (2010)
- Battle ground states are shifting
- Demographics play a key role in winning elections
- Money talks

Party politics - Swing States



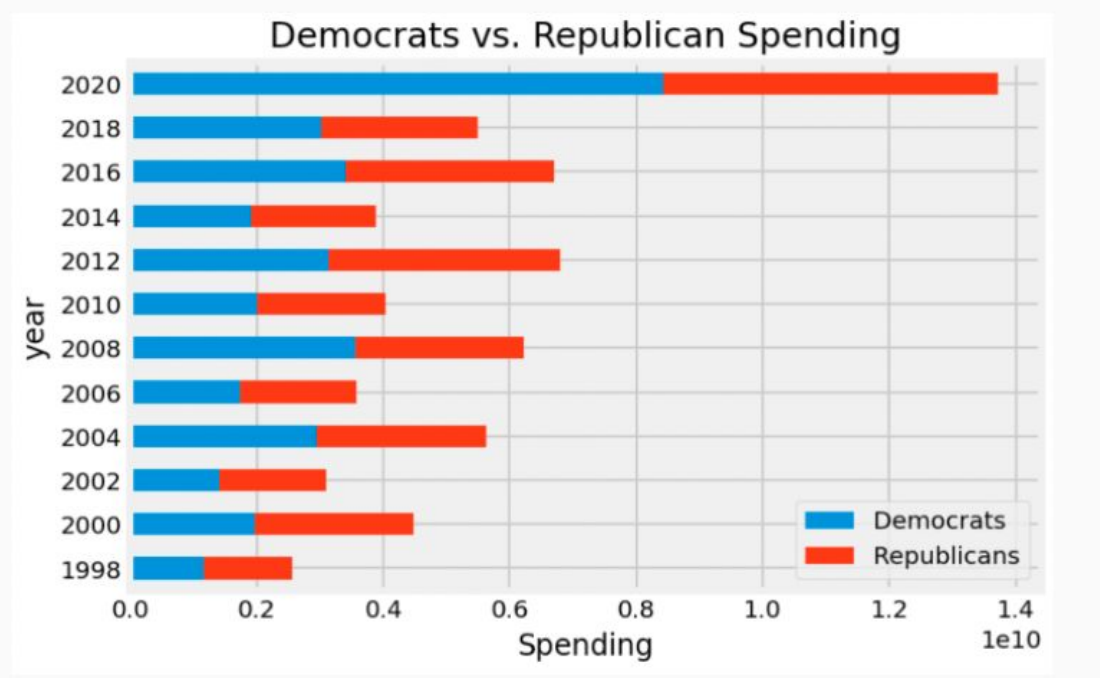
- Swing state contributions
  - We can see Texas is becoming the new battleground state in the minds of the american political donors and the elite who make large contributions
  - Florida remains a very heavily invested in state
- Spending in swing states in 2016 was much lower than average
  - Pay attention to the 2016 margin for democrats in Minnesota

Party Politics Cont'd



- Swing state margins
  - Florida has become much redder and the margin of victory may soon be too great for democrats to overcome in a statewide election
  - The Democrats victory in Arizona and Georgia was razor thin
    - demographic shifts in those states could explain this
  - North carolina has a Republican lean, but is not unwinnable for Democrats
  - Pennsylvania truly is perhaps the greatest swing state
    - But does not pull the same kind of spending as other states like florida and texas
  - Texas appears solid red by the numbers but continues to pull in the most money
    - Could indicate bad voter rights laws
    - Further analysis of demographic and voter turnout data could help us understand what is happening there

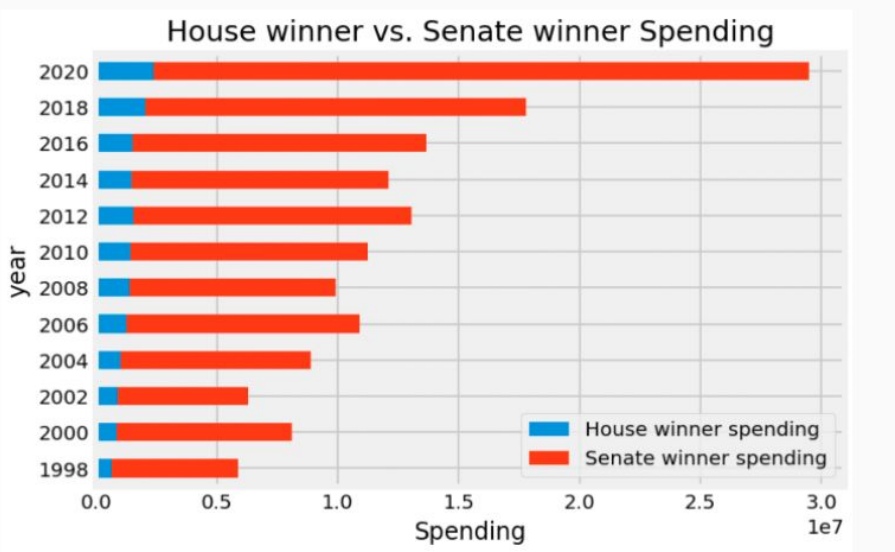
Spending by party



- Our analysis tells us that the greater spender typically wins the election in swing states.
- Democrats typically spend more than Republicans on Senate elections
    - Democrats hardly ever have a majority beyond 51 votes in the Senate.
- This can tell us that while you need a lot of money to win senate seats, the constitutional rules of the senate gives an advantage to Republicans
- Money cannot beat the fact that the senate does not proportionally represent the American People.

Spending

- Senate campaign spending outpaces House campaign spending dramatically



It is clearly much more expensive to become a U.S. Senator than it is to be elected to the House of Representatives.

This means Senators need to find donors with deeper pockets in order to win elections.

This *could* explain why the Senate fails so often to pass legislation to help the American people. Both parties need excessive amounts of money to win Senate seats and therefore need to accept money from corporate donors.

This has also gotten considerably worse over the last ten years (after the citizens united decision)

Don’t Forget to Vote every TWO years!

Living in a democracy is a privilege and a responsibility! To maintain it we all must participate.

Future Analysis : Further Questions

- Does what the campaign spend money on matter?
  - Opposition vs. Support advertising
- What portion of advertising is done online instead of traditional media?
  - As younger generations become more reliable voters, will there need to be a significant shift away from traditional commercials?
- Is the growth in spending we have seen since Citizens United exponential?
- Who is funding the senate elections and are there partisan differences between campaign contributors in the senate? Or is there relative consistency?
- What percentage of campaign funds for senators is coming from the health insurance industry?
  - Are there any senators who do not take health insurance money?

Future Analysis : Cont'd

- Is there a direct correlation between the funding of certain senators and the legislation that is passed?
  - For example
    - NRA funding and the deregulation of guns
    - Oil and gas and the investment in new pipelines
    - Defense contracting PACs and the military budget
    - Insurance industry and health care legislation
- What percentage of the population votes in swing states?
  - Does this correlate with access to voting?
  - Do less "important" states (solid red or blue) have lower turnouts by percentage?