Election Analysis

Erma, Hiep, Jason, Mike, Blake

Project Summary: The Senate

For our final group project, we are examining senate election results, spending, demographics and changing partisan trends. The Senate is the most important legislative body in the United States. Those who hold the position of US senator are some of the most powerful people in the world. The Senate also fails to pass most legislation that comes its way. We set out to understand how much it costs to become a 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.

We are interested in how often the Senate candidate that spends the most wins the election, but also *how* the campaigns spend their funds. Some examples of how campaigns spend their money include staff salaries, advertising expenses, travel expenses, and campaign materials.

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 two solid blue/red states to test our model and make sure it is working properly
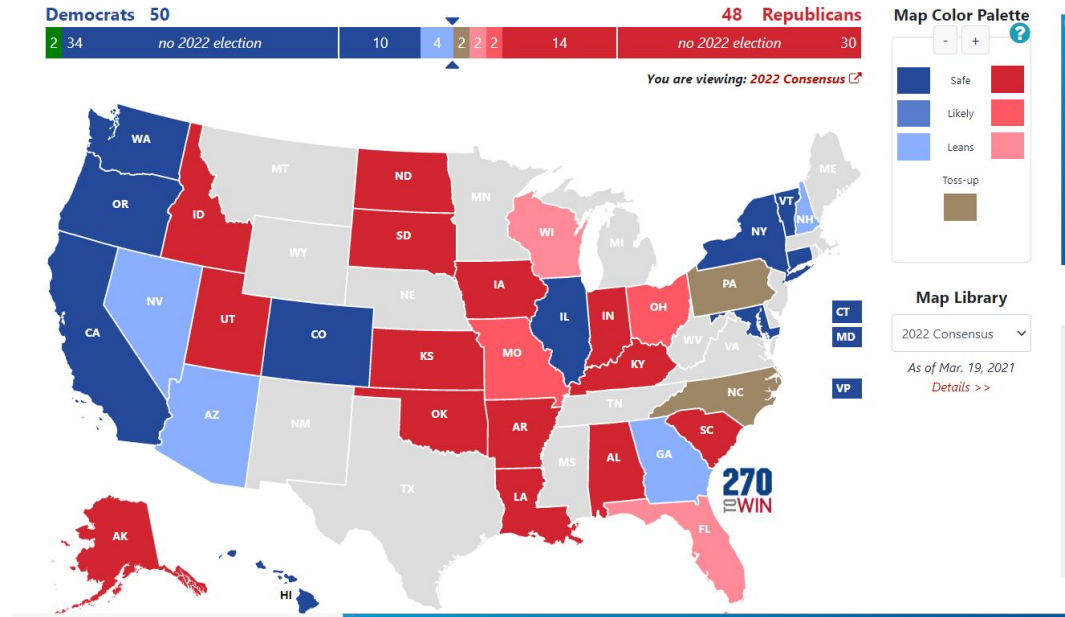
How to win senate elections

- We used 44 years of election results data, combined with 16 years of election spending data and 10 years of demographic data to better understand how campaigns can win elections.
- Sources
  - FEC reports
    - Campaigns are required to report every dollar they spend
  - Census Data
    - US Gov
  - Kaggle

Spending

- Senate spending outdoes house spending enormously


House winner vs. Senate winner Spending
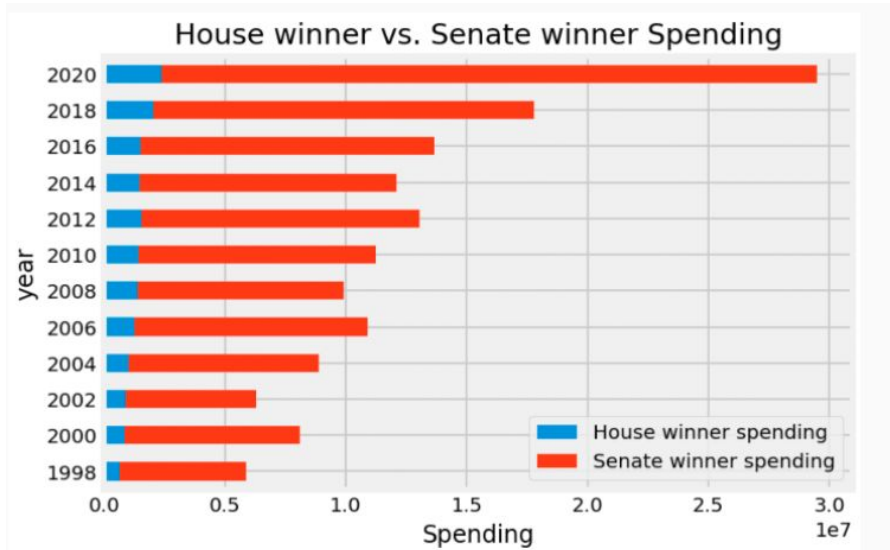
It is clearly much more expensive to become a senator than it is to be in the house of representatives.

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

Party politics – Swing States

- Swing state contributions
  - We can see texas is becoming the new battleground states in the minds of the american donors and the elite who make large contributions
  - Florida remains a very heavily invested in state

Swing States Contributions
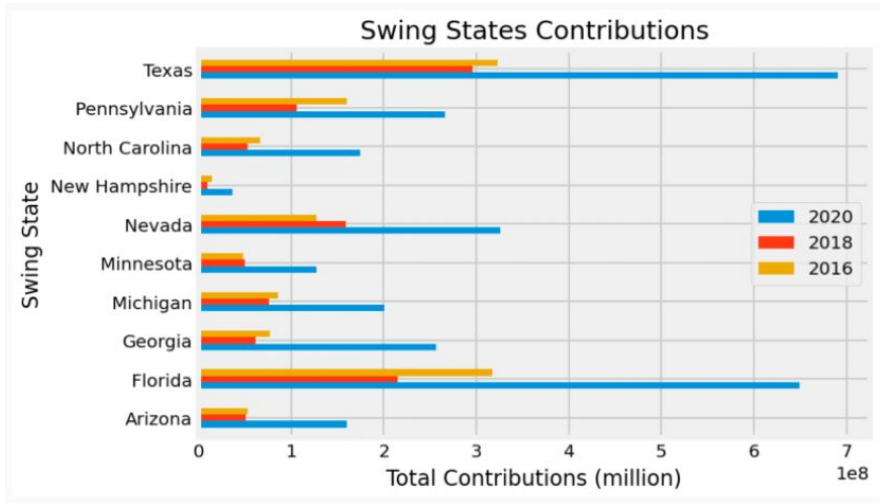
Party Politics Cont'd

Swing States Margin (2020, 2016 & 2012)

1st Bar: 2020
2nd Bar: 2016
3rd Bar: 2012
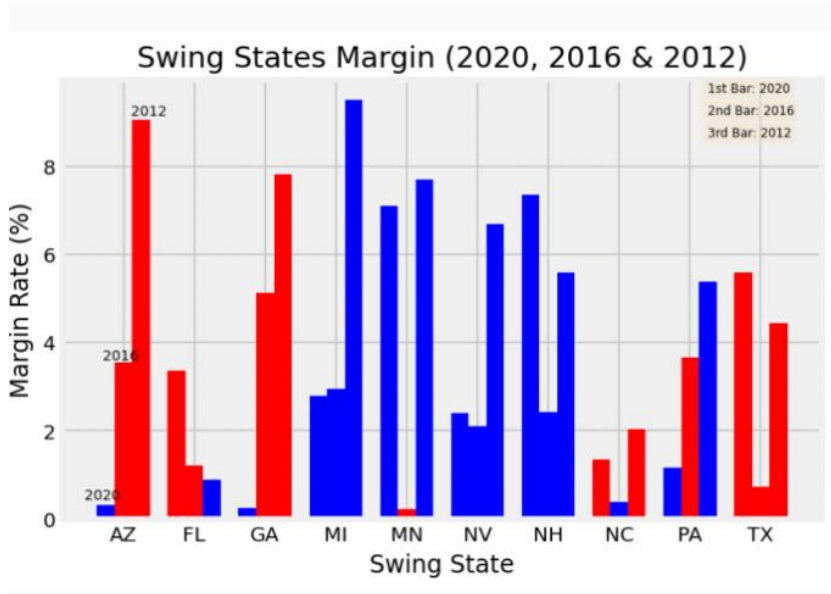
- Swing state margins
  - Florida has become much redder and the margin of victory may soon be too great for democrats to overcome in a statewide election
  - The victory in arizona was razor thin
    - This is interesting in terms of demographics, well return to this later
  - Georgia was also a razor thin margin
  - Michigan is solid blue but becoming redder as the margin of victory for democrats lessens
  - Minnesota was won by a very thin margin in 2016 by republicans
    - When we take into account swing state spending in 2016, we can understand why dems may have lost in minnesota
  - Nevada and new hampshire have a solid blue lean
  - North carolina has a republican lean, but is not unwinnable for democrats
  - Pennsylvania truly is a toss up
  - Texas appears solid red by the numbers but continue to pull in the most money
    - Could indicate bad voter rights laws
    - Also could be because of how many electoral college votes they have
    - Demographics and voter turnout data could help us understand what is happening there

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 **state demographic** 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**.

Data Analysis Phase

- For our data analysis, we relied heavily on four scripts that Hiep created. The scripts perform the following functions:
    1. First Python script: From the raw data, in CSV form, this script analyzes the data visually with graphs and charts. Each different test shows a list of command options to perform on that specific type of data, and then provides an interactive interface for user inputs. Using Python, ths allows better integration with our other Python scripts, and is far more automated with much faster overall processing time.
    2. Second Python script: With this script we went one step further, and integrated all of our tests into a single script that runs through a properly-formatted folder full of validation data. The script analyzes all of them, and puts all pertinent results in a specific Results folder for the teammates to review.
    3. Third Python script: This script deploys the machine-learning model to the website, predicting who will win, also creating the dashboard for elections.
    4. Fourth Python script: This script deploys a deep-learning neural network model to analyze the Senate dataset, also directly plots all the graphs on a web application.

Machine Learning – Hiep

1. Base on the president_dataset.csv, we found that the more population the more total votes. So we use **Linear regression** to **predict** the value of a target variable based on given **predictor** variable. Refer to Test Results

2. Use RandomForestRegressor to predict the total votes and who will win next election for Senate_dataset due to this is supervised learning where we have input variables (x) and an output variable (Y). Refer to Test Results

3. Apply deep-learning neural networks model to analyze from senate_dataset.csv. The outcome of this dataset is **totalvotes**, so we calculated the r2_score due to R2 is suitable for predicting continuous variable. A higher r-squared indicates a better fit for the model. Refer to Test Results

4. Apply deep-learning neural networks model to analyze from independent_expenditures.csv. The outcome of this dataset is **expenditure_amount**, so we also calculated the r2_score due to R2 is suitable for predicting continuous variable. A higher r-squared indicates a better fit for the model. Refer to Test Results

   Refer to Elections Results and Web App for details.

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 will be using for our database include:

- Postgres (PgAdmin) to import our data and to construct our new tables
- Integrating AWS as a server to allow for collaboration across multiple users

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.

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.
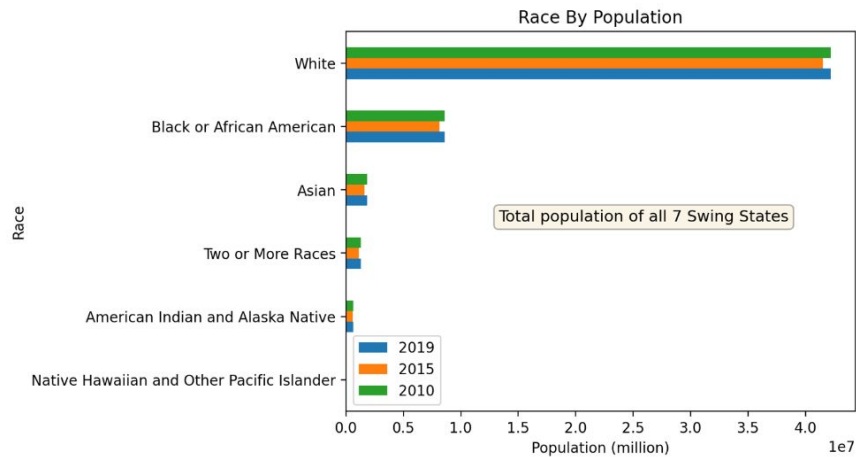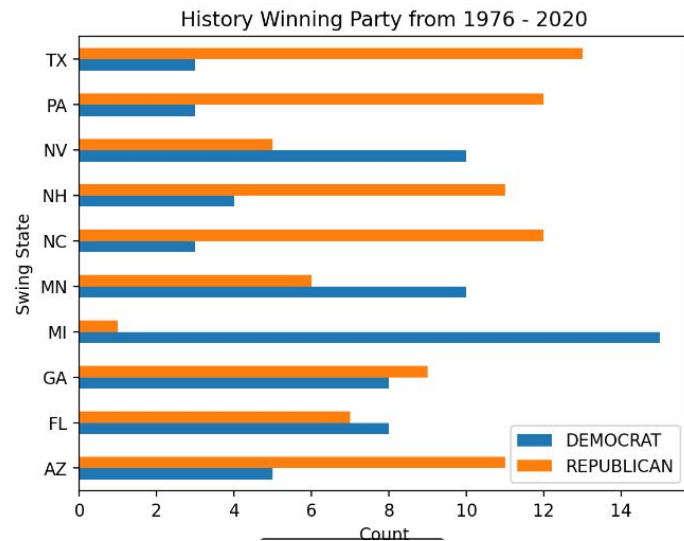
Database – Relationships
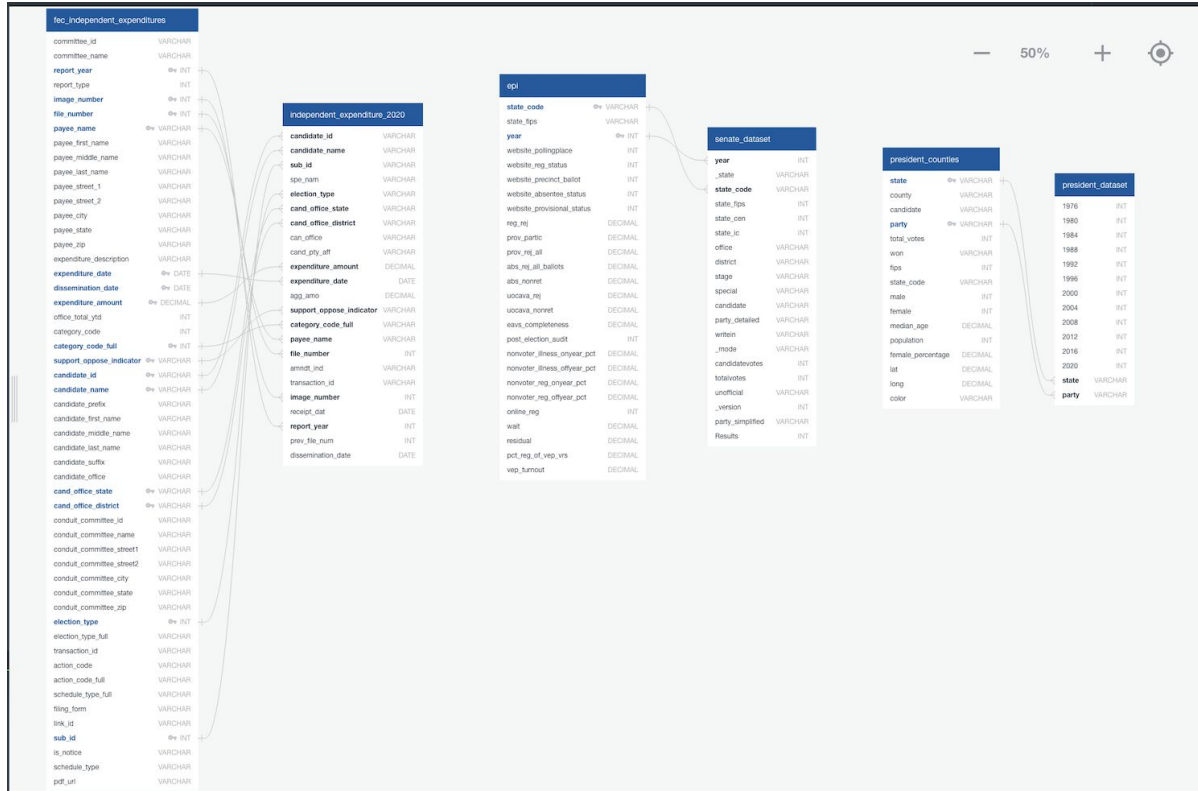
As mentioned above, the data we have found interacts with itself in many useful and interesting ways.

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

Storyboard

Storyboard

Swing States Contributions

Swing States Margin (2020, 2016 & 2012)