

ANA 515 Assignment 2

Rajas Purohit

2023-06-18

Section 1: Description of the data

Describe the dataset you are using, both in terms of the content and in terms of its format

Content:

What is this data measuring?

This dataset contains demographic, electoral and endorsement data for Senate, House and governor candidates in the 2022 primary elections (except the state of Louisiana).

How was it collected?

“Electoral, endorsement and gender data was collected by FiveThirtyEight from state election officials, organization and campaign websites, news reports, candidate interviews, social media and the Center for American Women and Politics. Race data was collected by political scientists Bernard Fraga and Hunter Rendleman using statements made by candidates, membership in racial and ethnic caucuses or organizations, racial and ethnic advocacy organizations, news reports and candidate websites.”

What kinds of research questions are you hoping to use it to answer?

It can be used to conduct demographic and endorsement analysis on candidates running for election. It was used in the following FiveThirtyEight articles:

Trump’s Endorsees Have Started Losing More. But Don’t Read Into That For 2024.(Aug. 29, 2022):

<https://fivethirtyeight.com/features/trumps-endorsees-have-started-losing-more-but-dont-read-into-that-for-2024/> (<https://fivethirtyeight.com/features/trumps-endorsees-have-started-losing-more-but-dont-read-into-that-for-2024/>)

People Of Color Make Up 41 Percent Of The U.S. But Only 28 Percent Of General-Election Candidates (Sept. 21, 2022):

<https://fivethirtyeight.com/features/2022-candidates-race-data/> (<https://fivethirtyeight.com/features/2022-candidates-race-data/>)

Progressives Took A Step Back In The 2022 Primaries — But They’re Playing The Long Game (Sept. 27, 2022):

<https://fivethirtyeight.com/features/progressives-democrats-2022/> (<https://fivethirtyeight.com/features/progressives-democrats-2022/>)

2022 Is Not Another ‘Year Of The Woman’ (Oct. 10, 2022):

<https://fivethirtyeight.com/features/2022-women-candidates-data/> (<https://fivethirtyeight.com/features/2022-women-candidates-data/>)

Format:

What type of file is it saved in?

It is a comma separated flat file.

Is it fixed width or delimited?

It is a delimited flat file, using commas as delimiters.

Section 2: Reading the data into R

We have read **dem_candidates.csv** and **rep_candidates.csv** using the tidyverse package and stored the data into dataframes

Section 3: Clean the data

We have filtered **dem_candidates.csv** using the dplyr package for candidates endorsed by AOC and Bernie Sanders

Section 4: Characteristics of the data

The dataframe **democratCandidates** has 1077 rows and 26 columns, while the dataframe **republicanCandidates** has 1599 rows and 27 columns. The names of the columns and a brief description of each are in the tables below:

common metadata:

Column	Description
Candidate	The name of the candidate.
Gender	The candidate's gender.
Race 1	What race or ethnicity (including Hispanic origin) the candidate primarily identifies as.
Race 2	What race or ethnicity (including Hispanic origin) the candidate secondarily identifies as, if any.
Race 3	What race or ethnicity (including Hispanic origin) the candidate tertiary identifies as, if any.
Incumbent	Whether the candidate was an incumbent.
Incumbent Challenger	Whether one of the candidate's opponents in the primary was an incumbent.
State	The state in which the candidate ran.
Primary Date	The date of the primary in which the candidate ran.
Office	The office for which the candidate ran.
District	For U.S. House races, the congressional district number for which the candidate ran.
Primary Votes	The raw number of votes received by the candidate in his or her primary. In states with runoffs, the results of the first primary round are used. For races that were canceled because they were uncontested, this number is 1.
Primary %	The percentage of the vote received by the candidate in his or her primary. In states with all-party primaries, this is the percentage of the Democratic or Republican vote received by the candidate. In states with runoffs, the results of the first primary round are used. For candidates who were unopposed, this number is 100.

Column	Description
Primary Outcome	Whether the candidate “Won” or “Lost” his or her primary or, if applicable, “Made runoff.”
Runoff Votes	The raw number of votes received by the candidate in his or her runoff, if applicable.
Runoff %	The percentage of the vote received by the candidate in his or her runoff, if applicable.
Runoff Outcome	Whether the candidate “Won” or “Lost” his or her runoff, if applicable.

democratCandidates specific metadata:

Column	Description
EMILY’s List	Whether the candidate was endorsed by EMILY’s List.
Justice Dems	Whether the candidate was endorsed by Justice Democrats.
Indivisible	Whether the candidate was endorsed by Indivisible.
PCCC	Whether the candidate was endorsed by the Progressive Change Campaign Committee.
Our Revolution	Whether the candidate was endorsed by Our Revolution.
Sunrise	Whether the candidate was endorsed by the Sunrise Movement.
Sanders	Whether the candidate was endorsed by Sen. Bernie Sanders.
AOC	Whether the candidate was endorsed by Rep. Alexandria Ocasio-Cortez.
Party Committee	Whether the candidate was placed on the Democratic Congressional Campaign Committee’s Red to Blue list, was endorsed by the Democratic Senatorial Campaign Committee or is an incumbent Democratic governor (the Democratic Governors Association has a policy of always supporting incumbents but not getting involved in primaries that do not feature incumbents).

republicanCandidates specific metadata:

Column	Description
2020 Election Stance	Whether the candidate “fully denied,” “raised questions” about, “fully accepted,” “accepted with reservations,” “avoided answering” questions about or had “no comment” on the legitimacy of the 2020 election. For full definitions of these categories, see here .
Trump	Whether the candidate was endorsed by former President Donald Trump.
Trump Date	The date on which the candidate was endorsed by Trump.
Club for Growth	Whether the candidate was endorsed by the Club for Growth.

Column	Description
Party Committee	Whether the candidate reached the highest tier of the National Republican Congressional Committee's Young Guns program or is an incumbent Republican senator or governor (the National Republican Senatorial Committee and Republican Governors Association had policies in 2022 of always supporting incumbents but not getting involved in primaries that do not feature incumbents).
Renew America	Whether the candidate is listed as a "Renewer" (marked as "Yes") or a "Divider" (marked as "No") by the Renew America Movement.
E-PAC	Whether the candidate was endorsed by E-PAC.
VIEW PAC	Whether the candidate was endorsed by VIEW PAC.
Maggie's List	Whether the candidate was endorsed by Maggie's List.
Winning for Women	Whether the candidate was endorsed by Winning for Women.

Section 5: Summary statistics

Summary Statistics for **democratCandidates**:

Primary Votes	Primary %	Runoff Votes
Min. : 1	Min. : NA	Min. : 5024
1st Qu.: 1	1st Qu.: NA	1st Qu.:10022
Median : 1	Median : NA	Median :22682
Mean :205	Mean :NaN	Mean :23230
3rd Qu.:341	3rd Qu.: NA	3rd Qu.:30708
Max. :997	Max. : NA	Max. :60929
NA's :924	NA's :1077	NA's :1067

Summary Statistics for **republicanCandidates**

Primary Votes	Primary %	Runoff Votes
Min. : 1	Min. : NA	Min. : 3666
1st Qu.: 1	1st Qu.: NA	1st Qu.: 13825
Median :309	Median : NA	Median : 27625
Mean :338	Mean :NaN	Mean : 49311
3rd Qu.:633	3rd Qu.: NA	3rd Qu.: 43635
Max. :997	Max. : NA	Max. :253251

Primary Votes	Primary %	Runoff Votes
NA's :1422	NA's :1599	NA's :1577