



# **PROJECT PROPOSAL**

## **ELECTION POLLS**

### **Phase 2: Data Cleaning & Visualization**

Section-002

Jayraj Radadiya-Team leader

Harshil Patel

Rajvi Mehta

Surbhi Patel

Krim Patel

Karan Patel

# INDEX

- 1 Data Cleaning.
- 2 Data Transformation.
- 3 Data Analysis & Visualization.
- 4 References.

## 1. DATA CLEANING

Data cleaning of 12 CSV files includes following steps:

- Missing Values
  - Most of the variable has missing values (NaN), So it has been replaced based on variable data type.
  - Categorical variable is replaced with Unknown values.
  - Replacing NaN values for Numerical column with 0.
- Dropping Irrelevant Column
  - Column with 100% values or more than 50% are dropped.
  - Column which are dropped listed below:
  - Partisan, seat number, sponsor candidate, tracking, notes, url, dem, rep, ind, Party, Politician I'd, Politician, Source, yes, No, Alternate answer.
  - After dropping useless column now total there are 33 columns with 32602 entries.
- Converting Datatype for Date containing columns
  - Three variables with Dates are: Start Date, End Date, Created at are converted from datatype object to datetime.
- Creating new CSV
  - After cleaning each CSV files using python on jupyter notebook and generating New CSV files by converting data frame to csv format.
  - Technology used : - Microsoft Excel , Power Query , Tableau.

## 2. DATA TRANSFORMATION

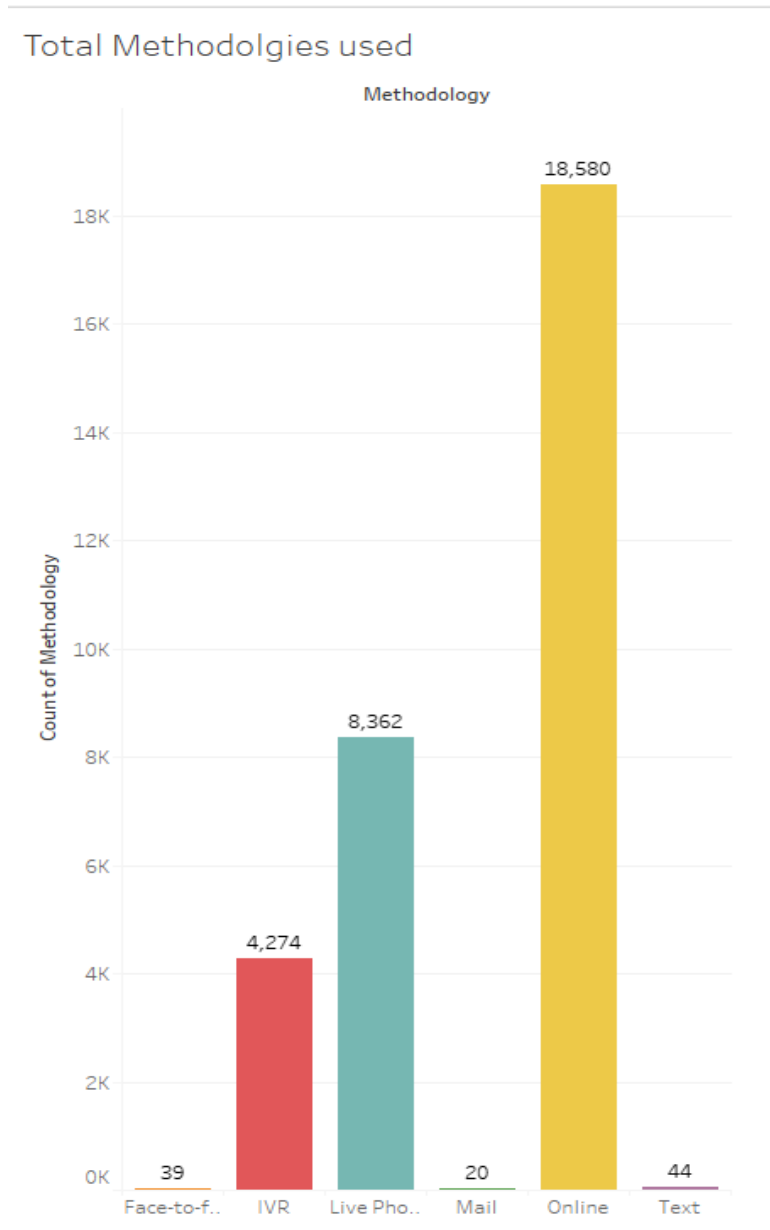
- Merging data into single DataFrame
  - After Cleaning each csv files and creating new one will get merged into single dataframe for further analysis.
- Need to do Data cleaning Again
  - After Merging all files data cleaning phase should be repeated for better visualization.
  - Repeated Steps:
  - Replacing NaN values with 0 for Numeric variable.
  - Filling Categorical variable with Unknown.
  - Converting Data Type of Date variable with datetime.
- Renaming all columns
  - All columns are renamed with valid names. Total 33 columns with dtypes: datetime64[ns](3), float64(2), int64(4), object(24)

#	Column	Non-Null Count	Dtype
0	Sr No.	32602 non-null	int64
1	Question ID	32602 non-null	int64
2	Poll ID	32602 non-null	int64
3	Race ID	32602 non-null	object
4	Cycle	32602 non-null	object
5	Pollster ID	32602 non-null	int64
6	Pollster	32602 non-null	object
7	Sponsor IDs	32602 non-null	object
8	Sponsors	32602 non-null	object
9	Display Name	32602 non-null	object
10	Pollster Rating ID	32602 non-null	float64
11	Pollster Rating Name	32602 non-null	object
12	FTE Grade	32602 non-null	object
13	Sample Size	32602 non-null	float64
14	Population	32602 non-null	object
15	Population Full	32602 non-null	object
16	Methodology	32602 non-null	object
17	Office Type	32602 non-null	object
18	Start Date	32602 non-null	datetime64[ns]
19	End Date	32602 non-null	datetime64[ns]
20	Nationwide Batch	32602 non-null	object
21	Created At	32602 non-null	datetime64[ns]
22	Stage	32602 non-null	object
23	State	32602 non-null	object
24	Seat Number	32602 non-null	object
25	Election Date	32602 non-null	object
26	Internal	32602 non-null	object
27	Ranked Choice Reallocated	32602 non-null	object
28	Answer	32602 non-null	object
29	Candidate ID	32602 non-null	object
30	Candidate Name	32602 non-null	object
31	Candidate Party	32602 non-null	object
32	Pct	32602 non-null	object

dtypes: datetime64[ns](3), float64(2), int64(4), object(24)

### 3. DATA ANALYSIS & VISUALIZATION

What medium of voting is preferred the most by the voters?

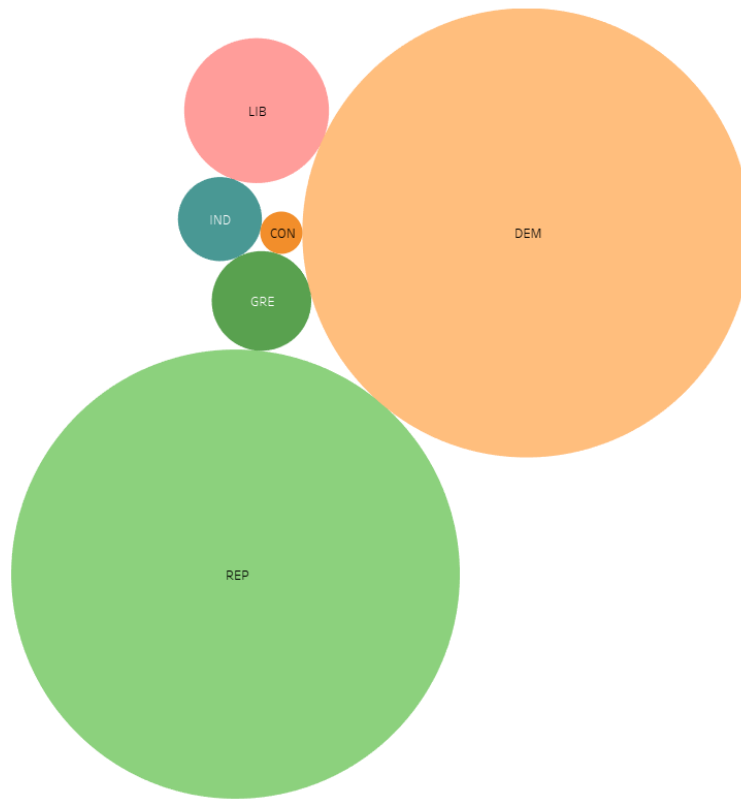


- A. Today, there are a plethora of voting technologies available, but there is a notable absence of remote voting options. But with the increasing technology, there is an increase in options open for the voters. There are six options available which are face to face, IVR (Interactive Voice Response), live phone, Mail, online and text. With the data we have when we perform analysis the highest method which is preferred by the voters is the online method. Secure, simple, and

easy online voting to modernize and simplify elections. Voting online has its own benefits like you cannot vote from your phone or online if you are in the military or a citizen of a foreign country registered in specific US states. During this time online voting comes as a solution. With this analysis as we know that online voting is highly preferred, we can make it easier, safer, and secure for the future elections.

### Which political party rules the election and has the highest number of members?

Current Parties and the count of their Party members



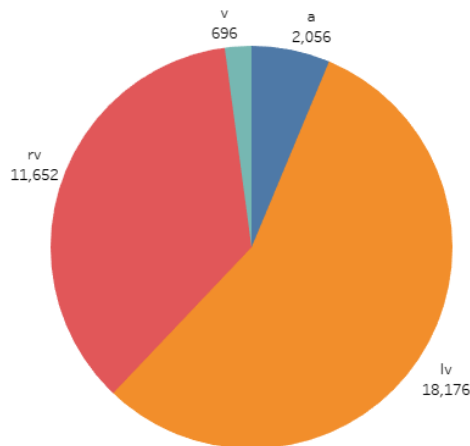
The electoral system in the U.S. is called a two-party system. There are more than 69 registered political parties, each with its own set of political beliefs. The Republicans and Democrats, on the other hand, are the only two national political parties.

There are other minor parties in the United States the Libertarian Party, the Green Party, the Constitution Party, and other minor parties have less influence than the main parties. With the graph above it is certain that both parties' rule but there is a minor difference between them with minimal numbers. Lower taxes, deregulation, more military expenditures, abortion restrictions, immigration limitations, gun rights restrictions, and labor union restrictions are all priorities for the Republic party. Americans who share the same political views as candidates from one of the third parties may instead vote for either the Democratic or Republican parties.

They do this to guarantee that their vote goes to the candidate with the best possibility of winning. This maintains the two-party system. A two-party system has the advantage of ensuring that the two major parties in power have a broad platform that represents the overall public. Because the two parties are so vast, each can accommodate a wide spectrum of political positions. This means that within any party, there may be minor differences in political perspectives on certain issues.

## How many people come out to vote?

Distribution of Population



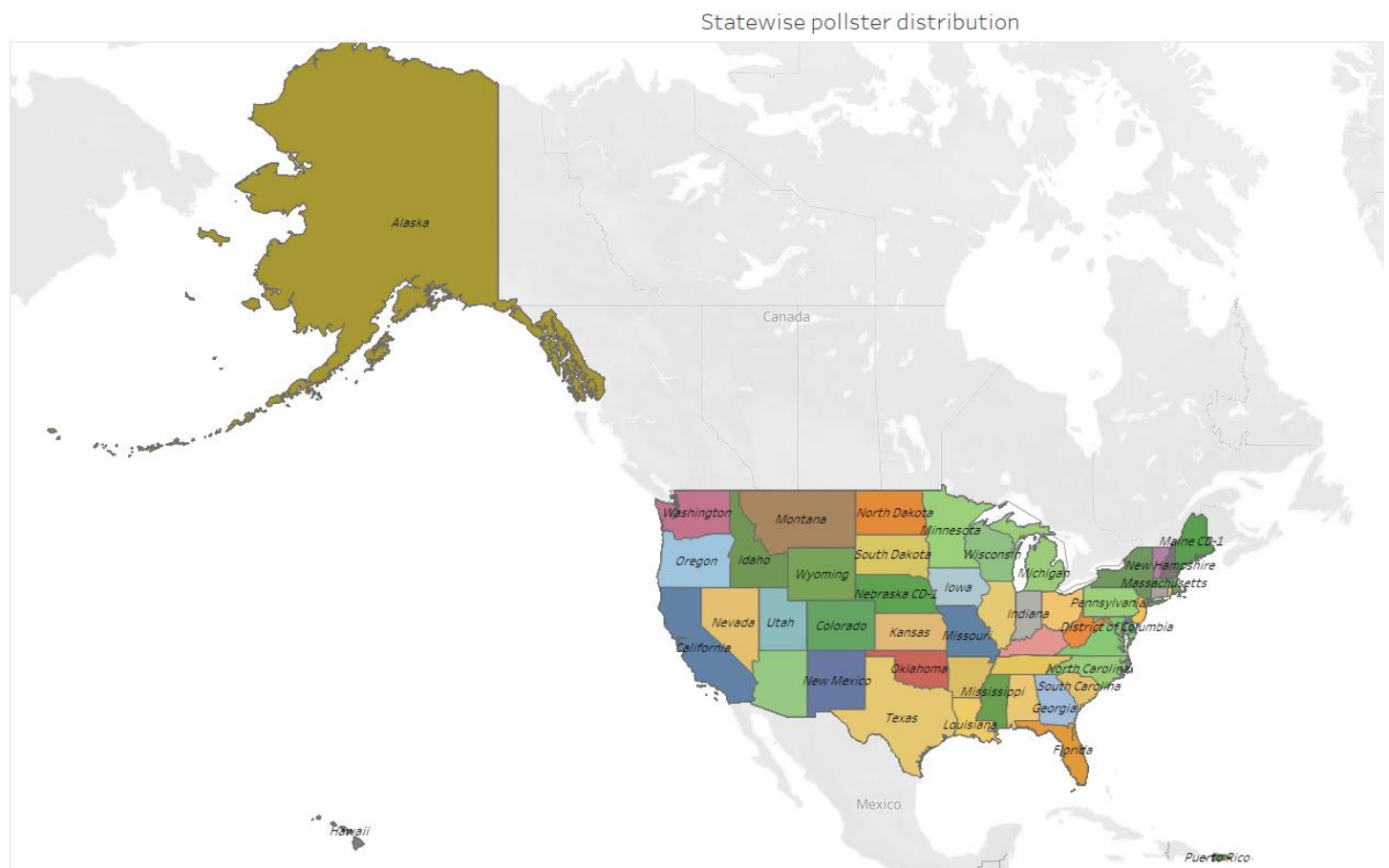
The right to vote is one of the most essential rights that American citizens have. Citizens over the age of 18 cannot be refused the right to vote based on their race, religion, gender, disability, or sexual orientation. The types of voters are divided into four major categories according to the data used, that is LV (likely voters), RV (registered voters), A(adults), V(voters). Likely voters own most of the number and voters have the least. Likely voters (LV) are those who have indicated to polling companies that they have a strong will to vote on election day.

Some characteristics of likely voters are higher levels of income, education, and occupational status associated with a keen sense of party identity and the belief that voting is a critical act. Registered voters (RV) are individuals who declare they are "registered to vote in their precinct or election district" in response to a typical poll question. This is the group whose data Gallup

most frequently publishes since it represents an estimate of Americans who are eligible to vote and who might vote if they so choose. Not all the people who have registered to vote will cast a ballot. As a result, over the years, Gallup has developed algorithms to identify likely voters, or those who the business believes are most likely to vote.

All adult citizens(A), except for a few minor exclusions, have the right to vote regardless of their money, income, gender, social standing, race, ethnicity, political attitude, or any other restriction. Voters(V) are the citizens who have the right to vote.

### The different pollster assigned to each state?

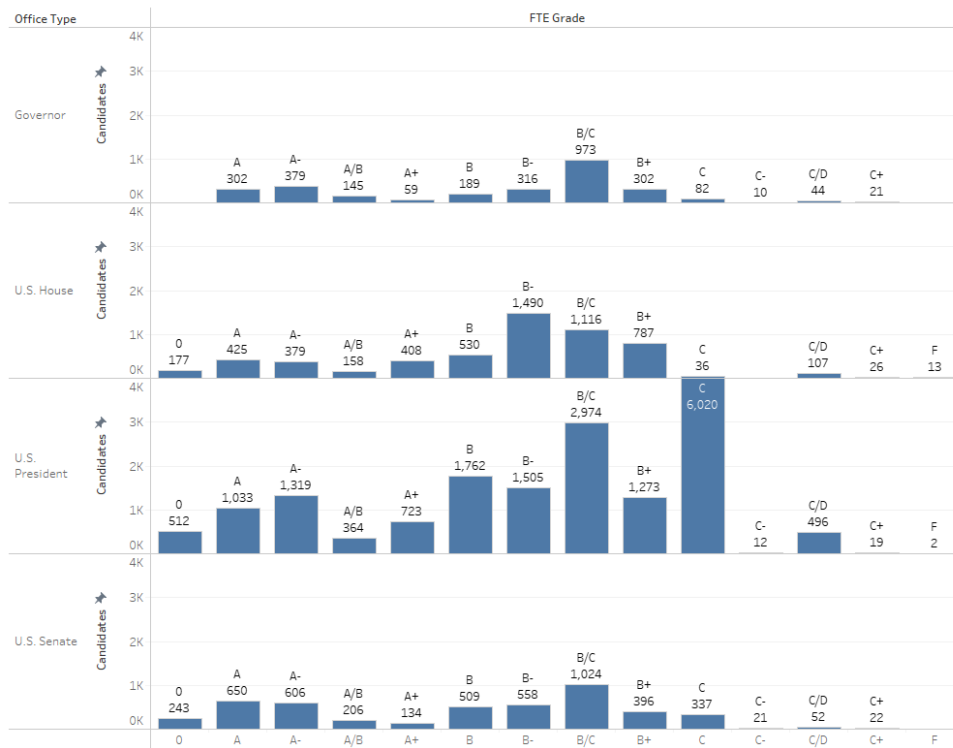


Different type of pollster is allocated to a specific state to conduct an election poll in United State of America.



## According to the different office types, which candidate has the highest/lowest fte grades?

Counted the number of candidates of different Office types based on their FTE Grade



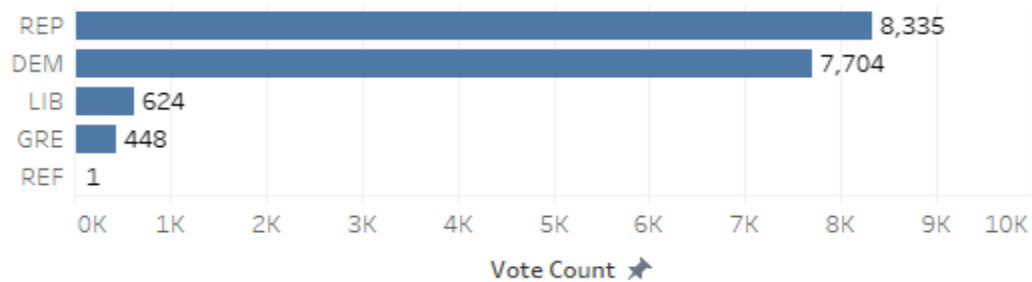
By analyzing, There are total of four office types:

- Governor
- U.S.House
- U.S.President
- U.S.Senate

U.S.President and U.S.House's office candidate receives more grades than others. There are a total of 6020 candidate of U.S.President office who achieves c grades, however very few candidates from U.S.Senate and Governor's office has c grades, which is 337 and 82 respectively. On the other hand, there are many candidates that receive c, a-,b/c, and b+ grades according to their office types. Moreover, none of the candidates receives c- and f grades of a different office. Only 21 candidates have c- grades from U.S.Senate and merely 13 candidates have f grades from the U.S.House office.

### Which party won the most vote?

Parties with their vote count

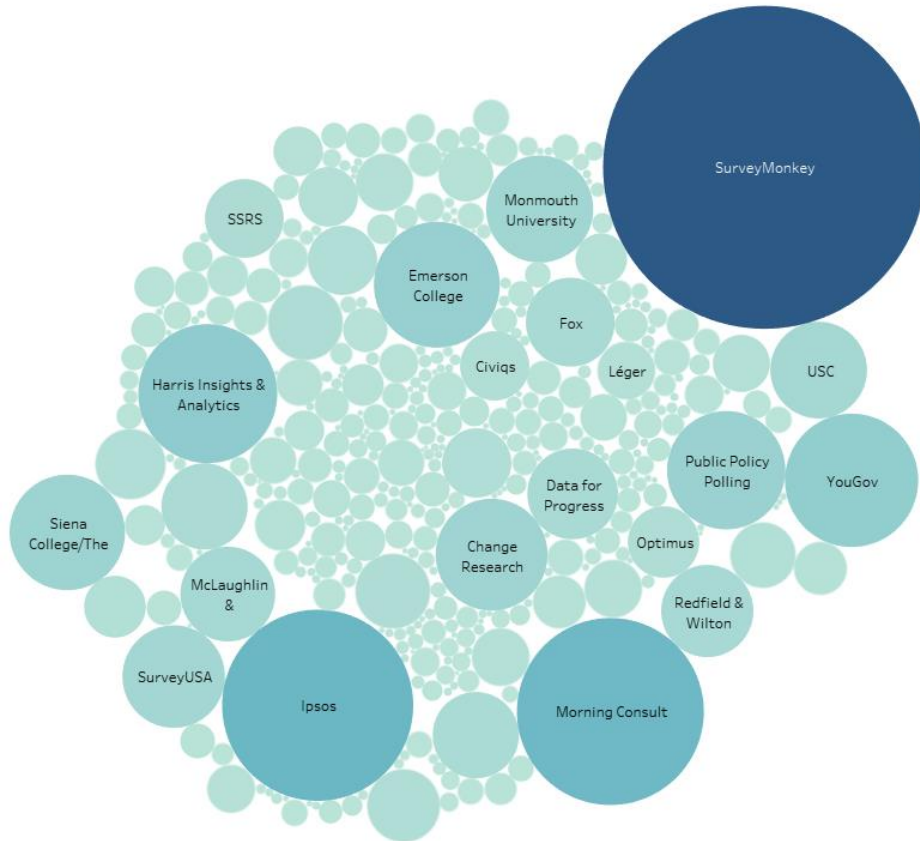


It is obvious that those parties which work hard get highest vote. But as a future perspective it will not be identified which party covers the most vote as every year, parties work very hard to win the election.

There is total 5 parties involved in election. DEM and REP has highest number of candidates. As the data describes the candidates from REP has won the highest number of seats by beating all other parties which were fighting election against them.

## Which pollster counts more winning probability for election in USA?

Major Pollster Count

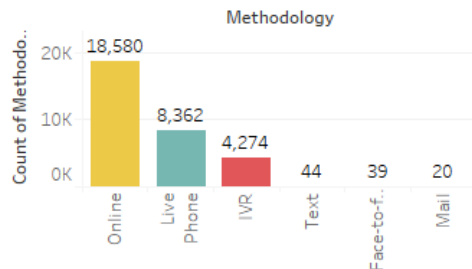


Pollster is a person who conducts an election survey to gather information based on the opinions of a sample of people. By analyzing the above graph, SurveyMonkey has the highest number of polls counting as compared to others, whereas civics countless number of the pollster. However, there are many others who conduct a survey for elections like Harris insights & analysis, public policy, YouGov, SSRS, etc.

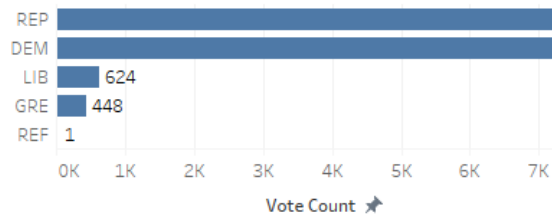
# Dashboard

## DAB103 Project Dashboard

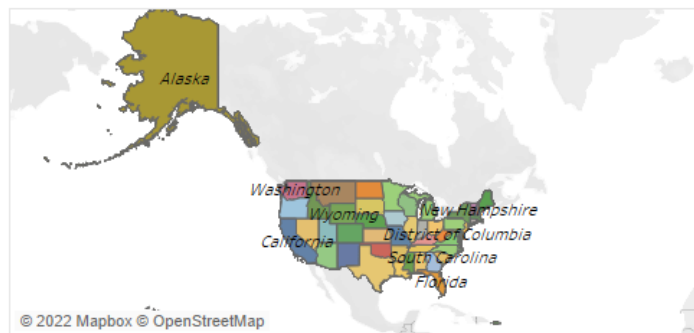
Total Methodologies used



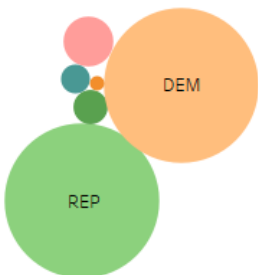
Parties with their vote count



Statewise pollster distribution

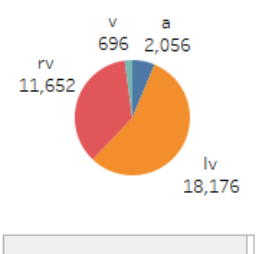


Current Parties and the count of their Party members

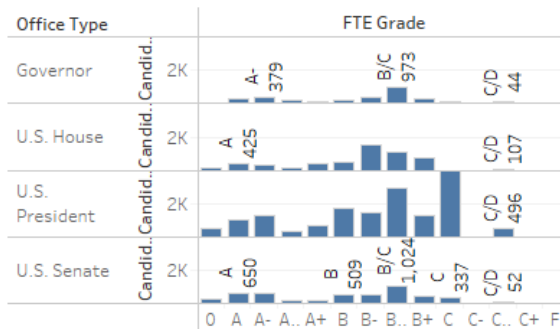


Major Pollster Count

Distribution of Population



Counted the number of candidates of different Office types based on their FTE Grade



Methodology

- Face-to-face
- IVR
- Live Phone
- Mail
- Online
- Text

Candidate Party

- CON
- DEM
- GRE
- IND
- LIB
- REP

Pollster

- 1st Tuesday Campaig..
- 20/20 Insight
- 1892 Polling
- ABC News/The Washi..
- Alaska Survey Resear..
- ALG Research
- Amber Integrated
- America First Policies
- American Research G..
- American Viewpoint
- AtlasIntel
- Auburn University at ..
- Axis Research
- AYTM
- Baldwin Wallace Univ..
- Baselice & Associates
- Basswood Research
- Battleground Connect

Population

- a
- lv
- rv
- v

The above is the dashboard of our project which included the data visualizations obtained from the data.

## 4. References

<https://news.gallup.com/poll/110287/what-difference-between-registered-voters-likely-voters.aspx>

<https://dk.usembassy.gov/da/youth-education-da/the-american-political-system/the-democrats-and-the-republicans/>