**Introduction**

While the presidential election season is in full swing, we decided to explore polling data sources that exist online. There are several individual sources that could be found online; however, the website RealClear Politics is a location that gathers, summarizes, and presents the results of the various polls in one location. It should be noted that, while this website is good of a summary view, the underlying polling data must be extracted from the various polling sources (if available) for further review and analysis. The polling sources include Emerson College, The Economist Magazine, The New York Times/Sienna College, CBS News, and many others. Some sources are free, while others incur a fee. It should be noted that the polls tend to discriminate between “Registered Voters” (RV) and “Likely Voters” (LV), and the common belief is the LV are better more indicative of election results. However, a Berkley Haas Study in 2020 reported that while the polls reached a 95% confidence level for statistical reporting, the actual election results only matched with the polls 60% of the time.

**Objectives**

Given FiveThirtyEight polling data available online, conduct an exploratory analysis of the states who had democratic and republican primaries on Super Tuesday, March 5th 2024. The states having both party primaries on that date are the following:

* Alabama
* Arkansas
* California
* Colorado
* Maine
* Massachusetts
* Minnesota
* North Carolina
* Oklahoma
* Tennessee
* Texas
* Utah
* Vermont
* Virginia

The analysis will include the following:

1. A summary of polling data for each state, including party affiliation, population surveyed, number surveyed, margin of error and methodology utilized for the polling.
2. Each state’s polling data as compared to national average of all polling data.
3. Graphic output of polling trends for each state as compared to national average.
4. Sub-categorization of polling results based upon population surveyed.

There may be additional results reported based upon the exploration of the data and review of the findings.

**Data Source**

FiveThirtyEight has a data source of the summary results of over 6300 polls dating back to 2021. The source also has links to the sub-source polling data for further analysis should we have additional opportunities to explore. There are additional sources of data on the site that could potentially be investigated for further analysis (i.e. favorability, approval, and the governors approval ratings from the states in this investigation).

**Collaboration Platform**

We use Slack, Zoom, and GitHub for communication and collaboration. Zoom serves as the platform for our meetings, where we discuss our plans and make decisions. Slack facilitates daily communication, allowing us to align on tasks and actions. Additionally, we utilize GitHub to share code and work together.

Our team has successfully leveraged these tools to communicate clearly and effectively, resulting in progress.

GitHub: <https://github.com/kohyarp/DATA607_Project3>

**Project Responsibilities and Deliverables**

1. Purpose
   1. Analysis of polling data
2. Entities
   1. People
      1. Koohyar Pooladvand: tidying, clean-up, and data frame preparation, and data analyses
      2. Anthony C: Data collection, report preparation, and analyses
      3. James Naval: Data cleanup, tidying, preparation, and analyses
      4. Victor Torres: Report preparation and analyses
   2. Sources
      1. Polling Data :
      2. Some raw CSV can be found in our GihHub: <https://github.com/kohyarp/DATA607_Project3/tree/main/Data>
      3. https://projects.fivethirtyeight.com/polls/president-general/2024/national/
   3. Platforms
      1. Relational database
   4. Functional Operations
      1. Data Identification
      2. Importing
      3. Tidy/Transformation
      4. Analytics
   5. Deliverables
      1. Presentation
      2. Analysis
3. Database Design

**Entity Relationship Diagram**

A diagram with many text boxes

Description automatically generated with medium confidence

Cited Works

1. https://www.realclearpolitics.com/epolls/2024/president/2024GeneralElection.html#!
2. Kotak, A., & Moore, D. A. (2020). Public Election Polls are 95% Confident but only 60% Accurate.
3. <https://projects.fivethirtyeight.com/polls/president-general/2024/national/>
4. https://abcnews.go.com/538/538s-polls-policy-faqs/story?id=104489193