Exploring the Primaries

R for Analysising Text and Data

Mid-week exercise

3 Exploring the US primary elections

This is a small exercise to explore the primary elections data set. We will use the data as downloaded from `https://github.com/houstondatavis/data-jam-august-2016` and combine the elections results with demographic data to explore aspects of the population that predict a candidate's success. The relevant files are in the `csv` folder.

Note: Write and submit the assignment as an Rmd (R markdown) file.

3.1 Choose a candidate and demographic

- Have a look at `county_facts_dictionary.csv` and select a candidate and demographic where you think that demographic might correlate with that candidate's performance. List and briefly explain your choice (1-2 lines).

3.2 Results per state

- Download the `primary_results.csv`. To do this, go to the file, then click raw, and copy the resulting url.

- Compute the fraction of votes for your candidate within his/her party **per state**. To do this, filter on the party, sum votes per candidate per state, then sum total votes per state, filter for your candidate, and divide candidate votes by total.

- Bonus: Visualize these results in a way of your choice (e.g. as a bar chart for top states, or as a coloured map)

3.3 County correlations

- Get back to the original data set

- Filter and reshape the data to have one row per county (using the fips column) with a column for each candidate.

- Download the `county_facts.csv` and create a data frame containing only the fips and the selected variable(s). Give these variables a sensible name.

- Merge the county facts with the results per county. You probably want to merge on the fips column.

- Correlate or regress your chosen variables with your candidate's vote share. Is the hypothesis confirmed?

- Bonus: Visualize the votes, demographic, and/or correlation in a way of your choice (e.g. with a scatter plot and/or regression line, or coloured map)