

Data Gathering Project:

Counties Elections Dataset

Alex Takata

Feng Shen

Lihua Xiong

Data Description

- For each .csv file, the following fields exist,
 - state, county: state and county name
 - 1nd, 2nd, 3rd: the names of the top, second and third candidate (nan if no third candidate)
 - votes1, votes2, votes3: the number of votes for the top, second and third candidate
 - pct1, pct2, pct3: percentage votes of the top, second and third candidate
 - party1, party2, party3: party of the top 3 candidates e.g. D / R / O (democrats, republicans or other)
 - FIPS: only in xxxx_FIPS.csv, the FIPS of corresponding county
 - Abbreviation: only in xxxx_FIPS.csv, the state abbreviation

Dataset Example

state	county	1st	2nd	3rd	votes1	votes2	votes3	pct1	pct2	pct3	party1	party2	party3	Abbreviation	FIPS
Alabama	Autauga	Bush	Kerry	Others	15196	4758	127	75.7	23.7	0.6	R	D	O	AL	1001
Alabama	Baldwin	Bush	Kerry	Others	52971	15599	750	76.4	22.5	1.1	R	D	O	AL	1003
Alabama	Barbour	Bush	Kerry	Others	5899	4832	46	54.7	44.8	0.4	R	D	O	AL	1005
Alabama	Bibb	Bush	Kerry	Others	5472	2089	39	72	27.5	0.5	R	D	O	AL	1007
Alabama	Blount	Bush	Kerry	Others	17386	3938	180	80.9	18.3	0.8	R	D	O	AL	1009
Alabama	Bullock	Kerry	Bush	Others	3210	1494	13	68.1	31.7	0.3	D	R	O	AL	1011
Alabama	Butler	Bush	Kerry	Others	4979	3413	24	59.2	40.6	0.3	R	D	O	AL	1013
Alabama	Calhoun	Bush	Kerry	Others	29814	15083	352	65.9	33.3	0.8	R	D	O	AL	1015
Alabama	Chambers	Bush	Kerry	Others	7622	5347	63	58.5	41	0.5	R	D	O	AL	1017
Alabama	Cherokee	Bush	Kerry	Others	5923	3040	86	65.5	33.6	1	R	D	O	AL	1019
Alabama	Chilton	Bush	Kerry	Others	12829	3778	86	76.9	22.6	0.5	R	D	O	AL	1021
Alabama	Choctaw	Bush	Kerry	Others	3897	3303	27	53.9	45.7	0.4	R	D	O	AL	1023
Alabama	Clarke	Bush	Kerry	Others	6730	4627	37	59.1	40.6	0.3	R	D	O	AL	1025
Alabama	Clay	Bush	Kerry	Others	4624	1893	59	70.3	28.8	0.9	R	D	O	AL	1027
Alabama	Cleburne	Bush	Kerry	Others	4370	1391	37	75.4	24	0.6	R	D	O	AL	1029
Alabama	Coffee	Bush	Kerry	Others	13019	4480	117	73.9	25.4	0.7	R	D	O	AL	1031
Alabama	Colbert	Bush	Kerry	Others	13188	10598	149	55.1	44.3	0.6	R	D	O	AL	1033
Alabama	Conecuh	Bush	Kerry	Others	3271	2719	31	54.3	45.2	0.5	R	D	O	AL	1035
Alabama	Coosa	Bush	Kerry	Others	2905	2055	41	58.1	41.1	0.8	R	D	O	AL	1037
Alabama	Covington	Bush	Kerry	Others	11119	3423	85	76	23.4	0.6	R	D	O	AL	1039
Alabama	Crenshaw	Bush	Kerry	Others	3777	1698	25	68.7	30.9	0.5	R	D	O	AL	1041
Alabama	Cullman	Bush	Kerry	Others	26818	8045	328	76.2	22.9	0.9	R	D	O	AL	1043
Alabama	Dale	Bush	Kerry	Others	13621	4484	126	74.7	24.6	0.7	R	D	O	AL	1045
Alabama	Dallas	Kerry	Bush	Others	11175	7335	63	60.2	39.5	0.3	D	R	O	AL	1047
Alabama	DeKalb	Bush	Kerry	Others	16904	7092	139	70	29.4	0.6	R	D	O	AL	1049

Data Source

- For 2016 and 2012: Politico.
- For 2008: NYT gives results by county for the top 2 candidates. Meanwhile, Wikipedia gives results for the top 3 candidates, but some counties are missing. As a result, we scrapped data from both the NYT website and Wikipedia and then merged the two datasets to create the `general_FIPS.csv` file.
- For 2004 and 2000: Wikipedia

Notes

- There are NaNs in the dataset. For example: some states only provides results for the top 2 candidates.
- Primary Election Results by County are only available for 2016.
- In 2000, some states do not report the results by county on Wikipedia.

Add FIPS

- In order to make it easier to merge our election dataset with other datasets in the future, we added FIPS and state abbreviation to the original dataset scrapped from the web pages.
- Steps:
 - We downloaded a state-abbreviation mapping online.
 - Using the data from assignment 1, we created FIPS values for each county.
 - Finally, we merged the two supplementary datasets with the election datasets to include the “Abbreviation” and “FIPS” fields.

Top 3 factors explaining the democratic relative voting margin

Column_name	Description	R_square
HSG495213	Median value of owner-occupied housing units, 2009-2013	0.0483318
RHI125214	White alone, percent, 2014	0.0251789
RHI225214	Black or African American alone, percent, 2014	0.023545

Definition: $\text{rel_voting_margin} = (\text{democrat_votes} - \text{republican_votes}) / \text{total_votes}$

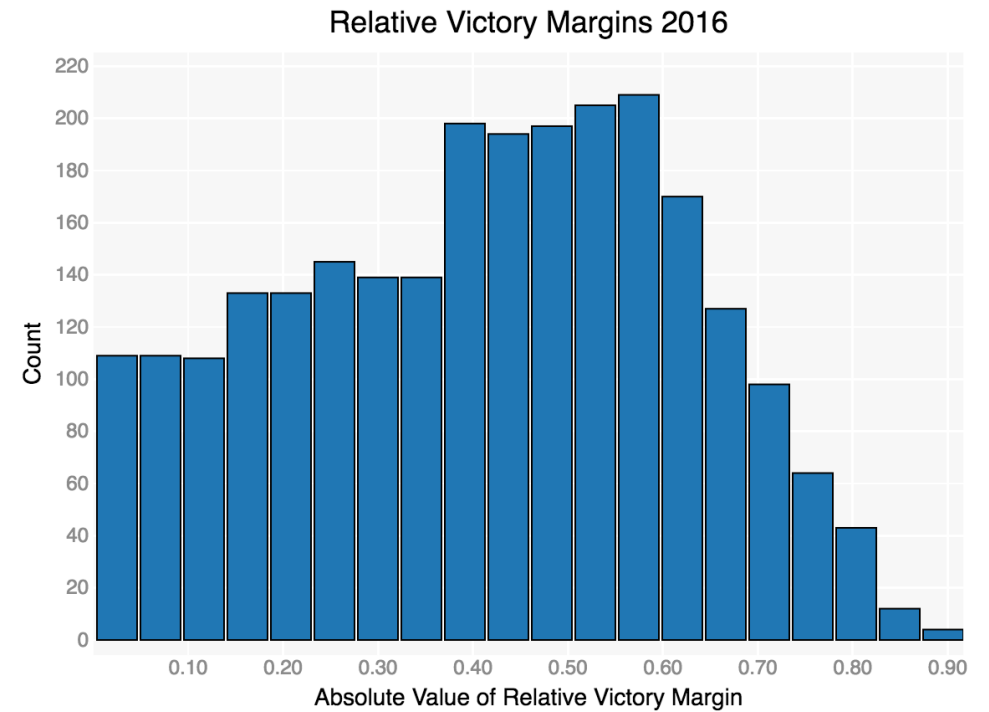
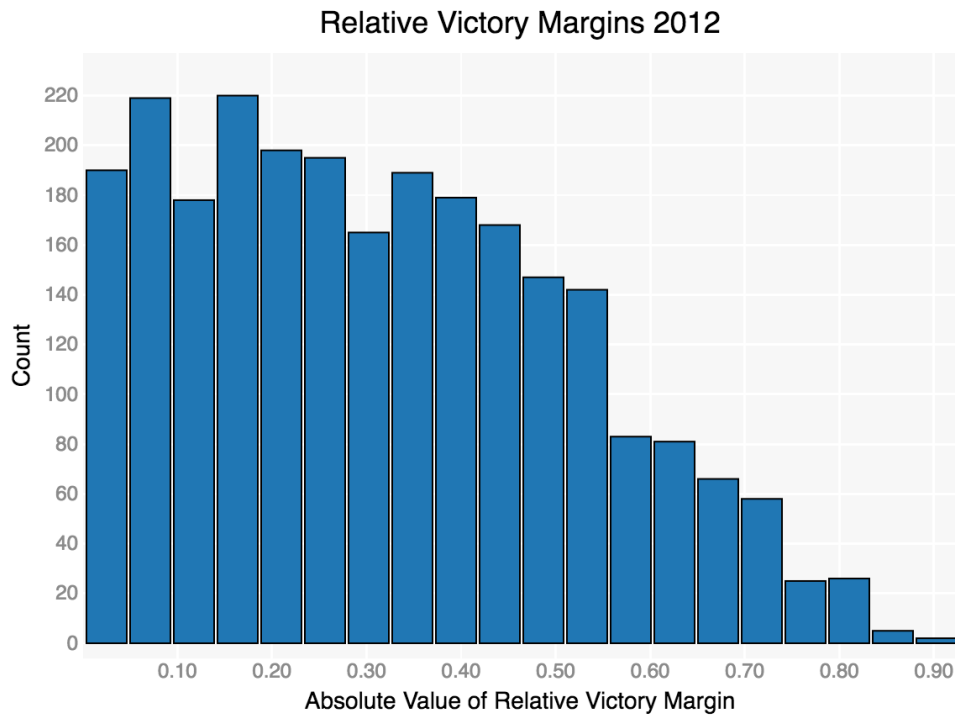
Steps:

- preprocessing the data
 - calculate rel_voting_margin for each county
 - merge dataset with `county_facts_dictionary.csv` indexed by FIPS
 - drop rows where NaN is contained
- Explain rel_voting_margin
 - Regress rel_voting_margin on each column of factor and get the r-square
 - Rank by r_square in descending order

Increased Voter Polarization 2012 vs. 2016

	2012	2016
Average Margin of Victory	30.7%	39.7%
Average Democrat Margin	21.1%	23.8%
Average Republican Margin	33.4%	42.7%
Total Democrat Counties Won	686	488
Total Republican Counties Won	2428	2625
Average Size Democrat County	99,244	137,706
Average Size Republican County	22,625	23,229

Increased Voter Polarization 2012 vs. 2016



Increased Voter Polarization 2012 vs. 2016

