Data Gathering Project

County Level Elections History Data

## Objective

* In this Data Gathering Project you will gather election results for primary and general elections at the state-county level and you will save your documented code and results in datasets/geo/elections/ to be used by the rest of the class
* You will present your results in class

## Data Description

Presidential elections data at the state-county level can be obtained from Politico’s website for the [2012](http://www.politico.com/story/2012/11/election-results-2012-by-state-083320), and [2016](http://www.politico.com/2016-election/results/map/president/). Politico also has a nice [political maps page](http://politicalmaps.org/) which goes back all the way to year 2000. With a little digging you can find county results for 2008, for example here are the [2008 results for Maine](https://www.nytimes.com/elections/2008/results/states/president/maine.html) from the NYC website. Finally, in Wiki you can find the county results for years [2004](https://en.wikipedia.org/wiki/United_States_presidential_election,_2004) and [2000](https://en.wikipedia.org/wiki/United_States_presidential_election,_2000). If you are lucky, you might be able to find county results in text format for earlier years. For example, here are the results for [Florida 1972 election](https://en.wikipedia.org/wiki/United_States_presidential_election_in_Florida,_1972). However, for the most part, county results for elections earlier than 2000, if present, are in the form of pixel images of the state. For example, the [1980 Florida election](https://en.wikipedia.org/wiki/United_States_presidential_election_in_Florida,_1980).

## Methodology and Deliverables

### Gather The Data

Scrape the 2012 and 2016 state-county election results data from Politico, the 2008 data from the NYC website, and the 2004 and 2000 data from Wiki. For each year, put your results in a elections/YYYY folder, in the form of a .csv file containing the following columns:

* 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)

For all the years since 1968 create a folder elections/YYYY/state\_pics where you save the county map image for each state that you scrape from Wiki. Document your code and save it in the elections/ directory.

### Analyze the data

There is a county\_facts.csv in the datasets/geo/ folder, whose variable descriptions are located in the county\_facts\_dictionary.csv file. These are facts mostly obtained from the 2012-2014 census. Can you find the top 3 factors explaining the relative voting margin on each county for the 2016 election? The democratic voting margin for each county is defined as:

rel\_voting\_margin = (democrat\_votes - republican\_votes)/total\_votes

Why do we use relative voting margin instead of the voting ratio democrat\_votes / republican\_votes?

### Present your methodology and results

Give a 5 minute presentation of your methodology and results in class. You are welcome to give live interactive demos using the BQplot US County map plotting tools.

### README file

Create a README\_db.txt file containing:

* The names and emails of all the teammates so you can be contacted by the next user of the dataset
* Description of the file naming convention, fields for each dataset
* Any comments on data features other users should be aware of when they use your data.