

Exercise #3 - SoDA 501

Due by class start next week - Apr 5.

Convert the text file "CentreCountyPrecinctResults2016GeneralElection.txt" (from data provided here: <http://centrecountypa.gov/Index.aspx?NID=802>) **into summary data** on precinct-level total vote, Democratic percentage share of two-party vote on statewide elections, and ballot "rolloff" on statewide elections. Definitions and example calculations are below.

Output your final summary data into "Exercise3.csv."

Do this in a replicable way. That is, provide code that, when run, reads in the input file and writes out the output file. **Provide this in an appropriate "notebook."** (Recommendations below). Use the notebook to document how your code works and to indicate what transformations of the data could be thought of as a "split," "apply," or "combine" operation.

So ... you're turning in a notebook file (probably .Rmd or .ipynb), which both runs and documents your code, and your output .csv file.

Notebooks

- For R, the primary option would be an "R Notebook" created within RStudio ... in this case you would be turning in an ".Rmd" file. (It is possible to install the "R kernel" for Jupyter and produce a Jupyter notebook that runs R. This is fine, but has no advantages that I see.)
- For Python, my preferred way is a Jupyter notebook ... in which case you would be turning in an ".ipynb" file. (The "iPython" project is now incorporated in Jupyter, so this is the current way to create an iPython notebook. I believe it is possible to create an older style iPython notebook outside of Jupyter, but I don't recommend this if you are just now starting with Notebooks.) Jupyter is installed with my recommended method of installing Python and libraries, and associated software, via "Anaconda": <https://www.anaconda.com/distribution/> .
- There are other Notebook types ... I'm aware of "Beaker" and "Zeppelin" ... which I have not yet used. These are OK, too, but I can't provide any assistance.

Definitions

Statewide elections are for President (Variable Contest == "PRESIDENTIAL ELECTORS"), U.S. Senator, Attorney General, Auditor General, and State Treasurer.

Two-party vote is the vote for the Democratic candidate plus vote for Republican candidate.

Ballot rolloff is the percentage of voters who voted for President, but did not vote for the particular office. That is: $100 * ((1 - \text{total votes cast for the office}) / (\text{total votes cast for President}))$.

Example:

So, let me work by hand Bellefonte North (`PrecNo==1`). So, all numbers below are from the first 39 rows of the input data.

Total vote:

Total vote is the `Count` column on the row where `Contest == "BALLOTS CAST - TOTAL."` this case 391.

Democratic share of two party vote:

For president, Democratic vote is 188, Republican vote is 174, so Democratic share of two party vote is $100 \times (188)/(188+174) = 51.93$.

For US Senate, $100 \times (169) / (169+185) = 47.74$.

For Attorney General, $100 \times (194) / (194 + 185) = 51.19$.

For Auditor General, $100 \times (170) / (170 + 163) = 51.05$.

For State Treasurer, $100 \times (177) / (177 + 164) = 51.91$.

Rolloff:

Total votes for president: $188 + 174 + 2 + 3 + 22 + 2 = 391$

Total votes for US Senator: $169+185+31+1 = 386$, so rolloff = $100*(1-386/391) = 1.28$ (That is 1.28% of voters who cast a vote for president, did not cast a vote for Senator).

Total votes for Attorney General = 382. Rolloff = $100*(1-382/391) = 2.30$

Total votes for Auditor General = 364. Rolloff = 6.91

Total votes for State Treasurer = 368. Rolloff = 5.89

Note: I doubt it happens in these data, but it is possible for Rolloff to be negative.

So row 1 of the output data should look something like:

PrecNo	PrecName	Tot	D2Pre	D2Sen	D2Att	D2Aud	D2Tre	ROSen	ROAtt	ROAud	ROTre
1	BELLEFONTE NORTH	391	51.93	47.74	51.19	51.05	51.91	1.28	2.30	6.91	5.89