Analysis of a Voter Purge in Georgia

Seung Je Jung, Santosh Vempala and Eric Vigoda
Georgia Tech
10/30/2020

In October 2019, the Georgia Secretary of State's office announced their intention to remove 313,000 voters from the Georgia registered voter list[1], or about 4% of the registered voters in the state, giving three possible reasons: "National Change of Address", "No Contact" and "Returned Mail". Several lawsuits followed, and a small number of voters were reinstated.

How many voters were actually deleted from the rolls? What was their distribution across counties and precincts? These questions are increasingly important in Georgia as the margins of victory have become narrower in recent elections.

To understand the change in voter rolls and their potential effects, we performed an analysis of the voter data released by Georgia, including the list of registered voters as of December 12, 2019, and July 2, 2020; the list of proposed dropped voters (2019) and the voting data for the November 2018 gubernatorial election. We summarize our methods and findings below.

Methods.
We compared the proposed dropped voter list (DV) with the 2019 registered voter list (RV19) and the 2020 registered voter list (RV20), using registration numbers. The voters that were actually dropped and added were then mapped to the 2018 voting precincts (this involved using the shape file to match precinct names and ids, and geo-location), so that we had the number of dropped/added voters in RV20 compared to RV19 in each precinct. Using the 2018 gubernatorial voting data we calculated the percent of democratic and republican (and independent) votes in each precinct. Combining, on a per-precinct basis, the number of dropped voters with the 2018 gubernatorial voting data, we estimated the number of democrats/republicans dropped and added in each precinct, and computed the correlation between being dropped/added and being democrat/republican by precinct. Full details of the data and methods are in Appendix A.

Results.
Georgia has over 7.3 million registered voters, of which 3.93 million voted in 2018. A comparison of the registered voter lists for 2020 and 2019 shows that 277,771 registered voters, or about 3.7%, were actually dropped. Approximately 38% of these dropped voters voted in a 2012 or later election, i.e., they missed at most one presidential election. There are 2857 voting

---

[1]

precincts in Georgia with a median of 2302 registered voters per precinct (see histogram below). Using the 2018 gubernatorial voting data for each precinct, we estimated that the number of democrats dropped was 146,927 while the number of republicans dropped was 127,768. The Spearman correlation between the percent of democratic voters and the number of voters actually dropped by precinct was 0.225 (p-value < 0.00001), while for the percent of republican voters it was -0.230 (p-value < 0.00001)[2]. Similar correlations hold for % democratic voters vs % of voters dropped by precinct (0.213, p < 0.00001) and % republican voters vs % voters dropped (-0.218, p < 0.00001). Therefore, a statistically disproportionately large fraction of dropped voters were democratic. During the same period, 234,010 new voters were added, of which we estimate 119,410 democrat and 111,510 republican. From these estimates, the number of registered democrats dropped by 27,517 while the number of registered republicans dropped by 16,258, a net change of 11,259 in favor of republicans. The population of Georgia is estimated to be growing at over 1% annually. Analysis of registered voter lists show that the number of registered voters has *decreased* during the period between Dec 12, 2019 and July 2, 2020, and the dropped voters were positively correlated with being democrat.
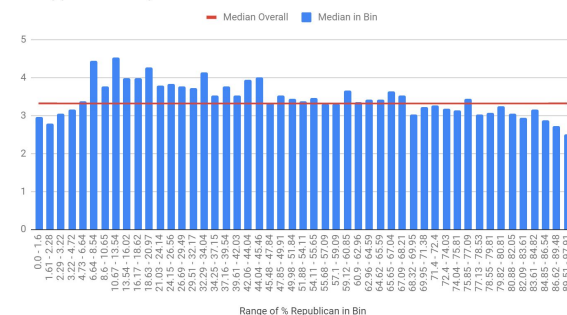

Histogram of the Number of Registered Voters per Precinct
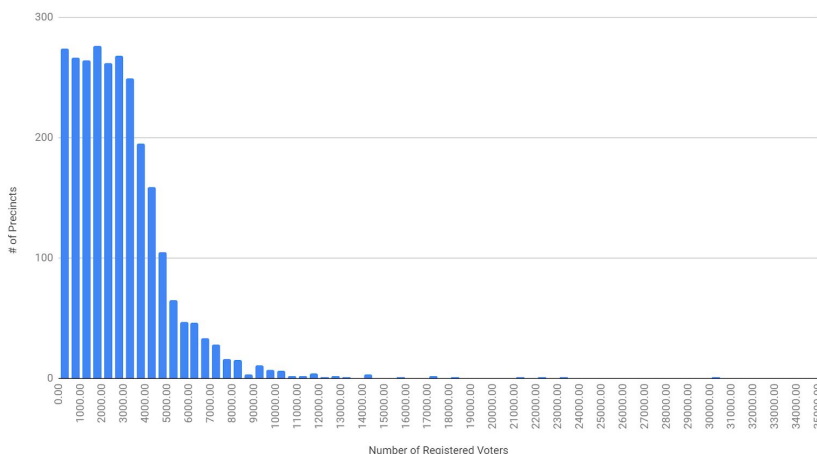
% Dropped vs % Democrat for Precincts with ~151270 Voters Per Bin
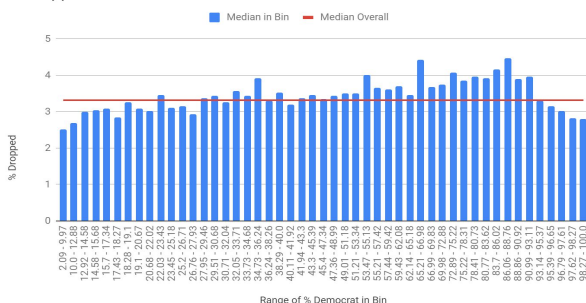
% Dropped vs % Republican for Precincts with ~151270 Voters Per Bin

---

[2] Spearman correlation looks for the strength of a monotone relationship; correlation higher than 0.2 is considered weak positive, and less than -0.2 is weak negative.

Appendix A. Detailed methods and results.

(https://github.com/santoshv/voterdata)

Data.
1. Registered voters as of December 12, 2019.
2. Registered voters as of July 2, 2020.
3. Proposed dropped voters (announced October 2019)
4. Voting data for November 2018 election of Governor

Getting the set of actually dropped voters.
Tables were created in MySQL for the registered GA voter list in 2018, registered GA voter list in 2020 and the voters proposed to be dropped in 2018, all from the GA state government. Next, a table with the counties in alphabetical order was used to create a mapping from the COUNTY_CODE field in the registered voter lists to the name of the county. The dropped voter list was merged with the registered voter list using registration number, with voters present in the dropped list but not in the registered voter list removed. Then, only voters which were present in the old list but not present in the new list were kept. These voters had their county names matched by the county mapping table made before.

Mapping dropped voter data to the 2018 election data.
The 2018 election data uses precinct names instead of precinct ids like in the registered voter lists. Therefore, we used 2018 GA precinct shapefiles from the GA government website which contained precinct names and ids to help map the ids to names. First, the precinct ids from the actually dropped voter lists were turned into pairs of (county name, precinct id) in order to uniquely identify them. The county name then had whitespace removed and was converted to lowercase, and the precinct id had whitespace removed, 0s stripped from the left and turned lower case in order to standardise the ids. This was also done for the precinct ids in the shapefile. Then, pairs were formed between identical shapefile ids and precinct ids from RV19. For select precinct ids that did not match up, geocoding was used on voter addresses from RV19 in order to create a matching. We then needed to create a matching from the shapefile precinct names to the 2018 voting data. This was done by using a modified edit distance algorithm where precincts in the 2018 voting data were all padded to have the same length, and then the edit distance was calculated between every 2018 precinct and the shapefile precinct names. These distances were treated as edges between 2018 precincts and shapefile precincts, and we ran a minimum cost matching algorithm to find the optimal assignment of 2018 precincts to shapefile precincts. This was hand reviewed and corrected. We then matched the results of the 2018 election data to the corresponding shapefile precinct, through the precinct name matching, and likewise mapped the RV19 voters using the precinct id matching.

Analysing mismatched precinct data.
Any precincts that could not be matched after hand picking were then compiled into a list of unmatched precincts, and in order to calculate the significance of these precincts, the number of actually dropped voters was summed up over all of them. It was found that there was ~8000 people who could not be matched, which is around 3% of the 277000 actually dropped voters.

Dropped voter analysis.

For each precinct, the number of republican votes, democrat votes and independent votes from 2018 were available, as well as the number of actually dropped voters in each precinct. We then calculated correlations (both Pearson correlation which quantifies for linear relationships and Spearman's correlation which quantifies monotonic relationships) and expected values to attempt to find biases in the dropped voters, as well as newly added voters (which were found by compiling a list of voters found in the new list of registered voters but not the old).

Histograms.

We first separated the actually dropped voters into three categories based on the reason they were dropped, which were National Change of Address, Returned Mail and No Contact, and then recalculated the dropped voter statistics for each precinct. Then, for each category we sorted the precincts by the % democrat voters in that precinct (calculated using democrat votes / total votes). Then, the precincts were grouped up into bins of approximately 151,270 voters (and precincts falling entirely in one bin) and the median % dropped was plotted.

% Dropped for No Contact vs % Democrat for Precincts
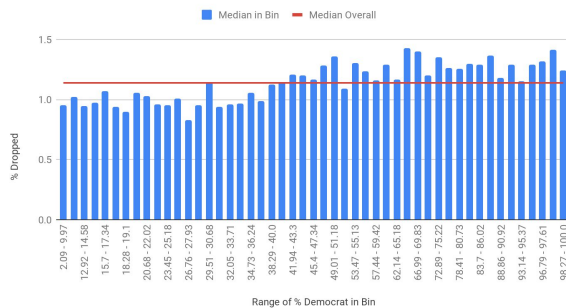
% Dropped for NCOA vs % Democrat for Precincts
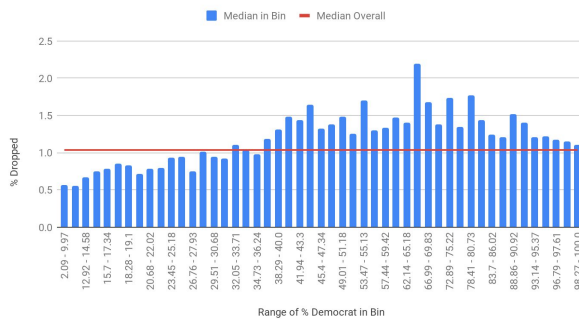
% Dropped for Returned Mail vs % Democrat for Precincts