Ferdous Alam
Steve Hall
Sharon Wu

Data Analysis of 2020 General Election Mail Ballot Requests in Pennsylvania

Introduction

The coronavirus pandemic changed the way Americans voted in 2020. A record number of mail ballot (or absentee) votes, approximately 65 million, were cast in the 2020 general election as voters looked to avoid crowded voting locations and minimize the risk of exposure to the virus. For comparison, in the 2016 general election only 33.5 million votes were cast via mail or absentee ballot! [1)]

Heading into November 3rd, many professional forecasters and political pundits expected the Presidential outcome to be determined by the electoral board of Pennsylvania. Triggering even more scrutiny, President Trump claimed many votes by mail or absentee ballot would be fraudulent. The President went as far as saying that he would give reduced funding to the USPS so people wouldn't have the ability to cast more mail-in ballots.[2]

There were some concerns about what might lead to fraudulent voting. This was based on the premise that "Political parties, campaign organizations or other groups can perpetrate organized fraud by obtaining and voting fraudulent absentee or mail-in ballots."[3] For our analysis we focused on the state of Pennsylvania rather than all states. "State-level comparisons are limited in what they can say about vote-by-mail policies."[4]

Days following the election, our attention was consumed by the major news networks releasing updated vote totals as mail ballots were slowly counted. The official result in Pennsylvania wasn't declared until November 7th, making Joe Biden the 46th President-elect of the United States.

Despite the official vote totals suggesting Joe Biden was elected as our next President, the claims of voter fraud have not stopped. Republican Pennsylvania State Senator Doug Mastriano tweeted the image below on November 27th, which has been liked over 178k times!

Pennsylvania official votes (@11/24 8 PM):

Candidate	Election	Mail	Provision
Donald J. Trump	2,821,610	594,645	49,323
Joseph R Biden	1,408,564	1,994,597	52,847
Grand Total	**4,230,174**	**2,589,242**	**102,170**

Pennsylvania reports having mailed out 1,823,148 ballots, of which 1,462,302 were returned. Yet total mail-in votes number 2,589,242? From where did the extra 2,589,242 − 1,462,302 = 1,126,940 votes come?

Using publicly available data, we wanted to confirm the results ourselves. We will combine mail ballot request data, voter registration data and general election results by county from the state of Pennsylvania. We will examine the data for any discrepancies, or signs of fraud across party and county

lines. Ultimately, we hope to validate, or raise questions about the results of the 2020 general election in the state of Pennsylvania.

Research questions

1.) How did mail ballot application requests evolve through time? Are there any glaring discrepancies with the election outcome?
2.) What percentage of registered voters applied to vote by mail across county and party lines?
3.) How long did it take to receive and process ballots across party lines and counties?
4.) How were first time voters split across party? Are there any variations across counties?
5.) Lastly, do the mail ballot application figures jive with the overall results by party and county?

Data Description

For this analysis, we used the three datasets below:

This first dataset describes the current state of mail ballot requests in PA for the 2020 General Election. It is a snapshot in time of the current volume of ballot requests across the Commonwealth. Total registered voters by county, broken down by political party (Democratic, Republican, No Affiliation, and Other). The total in the "No Affiliation" party column includes ONLY voters with this exact designation on their record. It does not include voters who may have a blank party field, "none," "no party" or other non-affiliated designation. Overall, this dataset includes nearly 3.1MM records and details about each voter's county/district, age, political party affiliation, and application and mail dates.

https://data.pa.gov/Government-Efficiency-Citizen-Engagement/2020-General-Election-Mail-BallotRequests-Departm/mcba-yywm/data .

The second dataset includes summary election results from the Pennsylvania Department of State website. This dataset captures election returns by county, party and vote type (primarily into three categories being Election day votes, Mail in Votes, and Provisional Votes).
https://www.electionreturns.pa.gov/

We also used the current registration statistics: total registered voters by county, broken down by political party (Democratic, Republican, No Affiliation, and Other).
https://www.dos.pa.gov/VotingElections/OtherServicesEvents/VotingElectionStatistics/Pages/VotingElectionStatistics.aspx

Data Errors, Omissions and Assumptions

For the mail ballot request dataset, there are 11 variables with 6 categorical variables (see table below). The raw mail ballot request dataset (nearly 3.1mm rows) contained some clear data entry errors. We discuss those and other data assumptions below:

Index	count	unique	top	freq
County_Name	3098705	67	PHILADELPHIA	437155
Applicant_Party_Designation	3096109	202	D	1947005
Date_of_Birth	3098202	30873	07/15/1954	237
Mail_Application_Type	3098705	18	OLMAILV	1412751
Application_Approved_Date	3098670	396	09/14/2020	47591
Application_Return_Date	3098661	399	09/14/2020	47659
Ballot_Mailed_Date	3082311	105	09/28/2020	422894
Ballot_Returned_Date	2553094	67	10/20/2020	152102
State_House_District	3098650	203	182ND LEGISLATIVE DISTRICT	34189
State_Senate_District	3098650	50	1ST SENATORIAL DISTRICT	108660
Congressional_District	3098650	18	3RD CONGRESSIONAL DISTRICT	258058

1. *Date of Birth*

   There are records with the date of birth in 2020. Those records are not reasonable, also some birth dates are displayed as 1/1/1800, this is due to confidentiality reasons of the registered voters, usually this is for victims of domestic violence. In the following analysis, we will only include voters who aged from 18 to 122, (68 records with birth year before 1898 and after 2002 are excluded).

2. *Ballot Mailed Date & Ballot Returned Date*

   There are 23,305 records with counties receiving ballots earlier than voters mailing the ballots. In order to observe the potential fraud, we will conduct further investigation in the Analysis section.

3. *Application Approved Date*

   There are 102 mail ballot requests approved prior to January 1st, 2020, and 88 prior to 2019, which seemed unreasonable to include in the analysis. Additionally, there were two applications approved after November 3rd. These were also excluded from the analysis.

Analysis

What percentage of registered voters applied to vote by mail across county and party lines?

The state of Pennsylvania government website has registered voter data by county and across parties. We used this information to glean any differences in percentages of registered voters who applied for a mail ballot. Based on the anecdotal evidence, we expected a greater percentage of mail ballot requests by Democrats than Republicans. But how much higher?

At a high level, there are approximately 4.23 million registered Democratic voters in Pennsylvania, and 3.54 million Republican voters. For the 2020 general election, nearly 1.95 million Democrats applied for a mail ballot, or 46%. On the contrary, only 788 thousand Republicans applied for a mail ballot, or 22%.

The percentage of total registered voters who requested a mail ballot varies widely across counties and parties. For instance, nearly 60% of Democratic voters in Chester county applied for a mail ballot while only 26% applied for one in Fulton county. For Republican voters, slightly more than 30% of registered

voters applied for a mail ballot in Allegheny and Chester counties while only 12% applied for one in Huntingdon county.
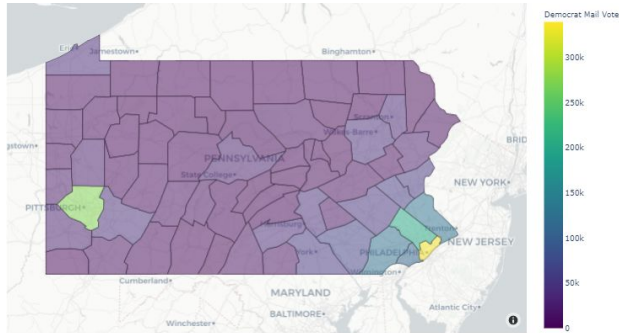
It is noteworthy that there is a positive relationship between parties for those counties that had higher and lower percentages of registered voters that requested mail ballots (see the scatterplot below). It would raise questions about the validity of results if we instead measured a negative relationship as it would be difficult to explain why a specific county would have meaningfully more or less mail ballot requests from one party but not the other.
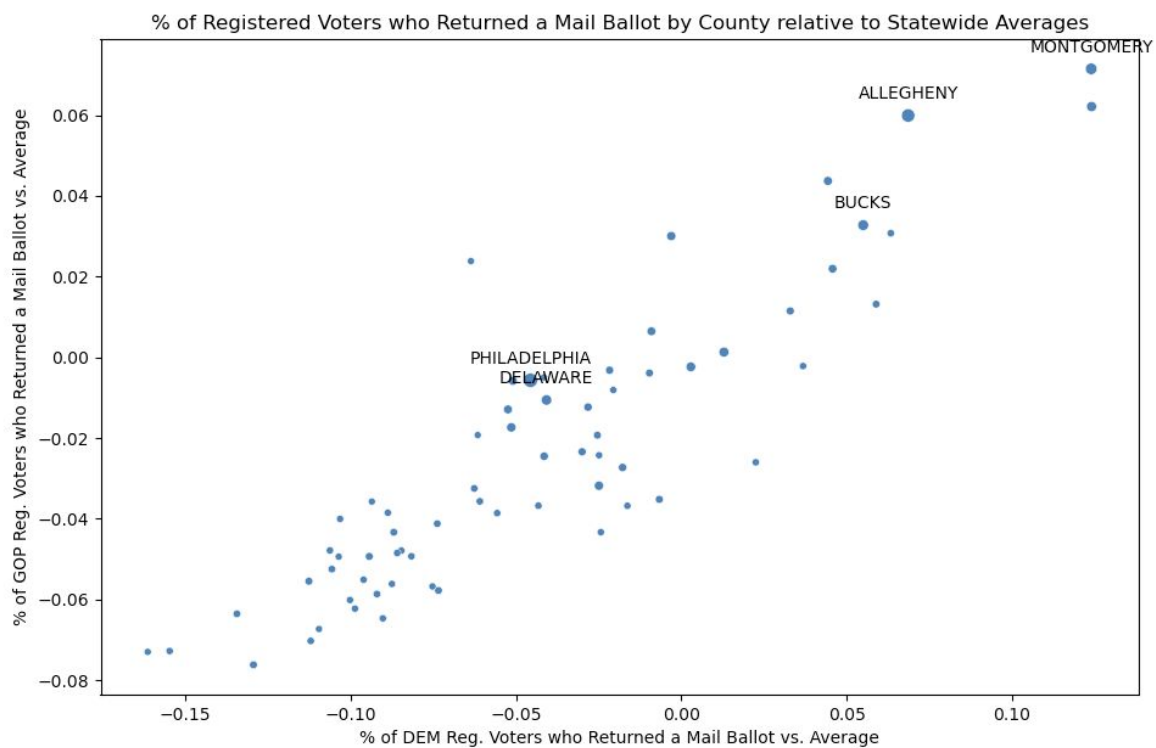
What was the breakdown of mail in votes received by party/county

Looking at the data we observe similar voting numbers across most counties. However, there was a significant number of mail in votes that were received by the Democrats in the largest counties. Out of the sixty-seven counties in Pennsylvania, the largest five counties (Philadelphia, Allegheny, Montgomery, Bucks, Chester)which are colored differently in the map here) accounted for 41% of the total returned mail in votes. Out of these, democrats received over a million votes or roughly 34% of the total mail in votes, whereas the Republican received 210 thousand votes or roughly 6.6% of the mail in votes.This meant that the margin of difference, only from mail in votes from these five counties alone was close to eight hundred thousand votes.

**Democrat mail in Votes**   **Republical Mail in votes**
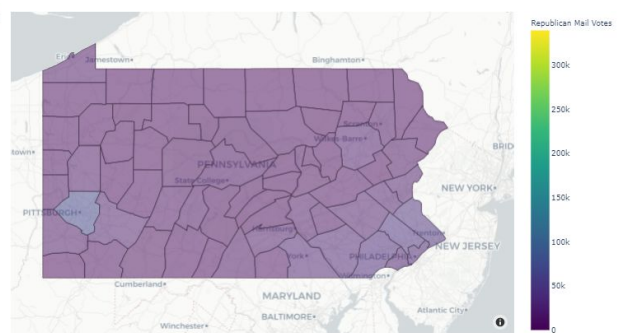
How did mail ballot requests evolve through time and across parties?

As election day approached and voting rose to the top of voters' minds, we expected mail ballot requests to increase, especially with increasing Covid-19 infection rates across most of the country. As you can see in the chart below, this was the case in Pennsylvania leading up to the general election. Interestingly, there was also a dramatic increase in mail ballot requests in May heading into the primary election. According to the metadata, this dataset only included general election mail ballot requests so we assume that voters must have applied for both a primary and general election ballot at the same time (i.e. if you need to apply for a primary ballot, why not apply for a general election ballot at the same time?).Across party affiliation, the trends are consistent with our expectation as there were more applications from Democrats than Republicans. This is consistent with the final vote totals, but we will need to dig in deeper to identify any signs of fraud.
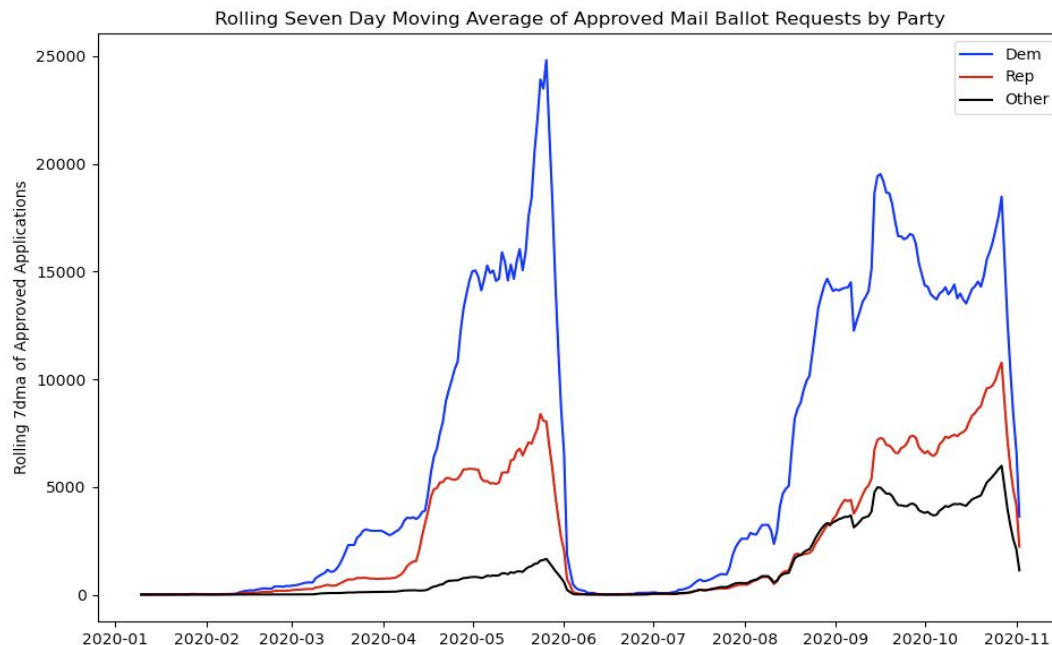
**How long did it take to process ballots across party lines and counties?**

During the 2020 primaries, coronavirus severely disrupted elections. State voting systems were overwhelmed by the safest and most secure way to vote in this pandemic, vote-by-mail. (A vote-by-mail ballot is a paper ballot, which is mailed to voters with instructions on how to mark it and how to return it. Any registered voter may ask to vote using a vote-by-mail ballot, and they do not have to be "absent" or "out of town on election day" to vote by mail). Considering this, in this section, we evaluate the time efficiency in each county by measuring how long it takes from voters mailing the ballot to the counties receiving ballots from voters.
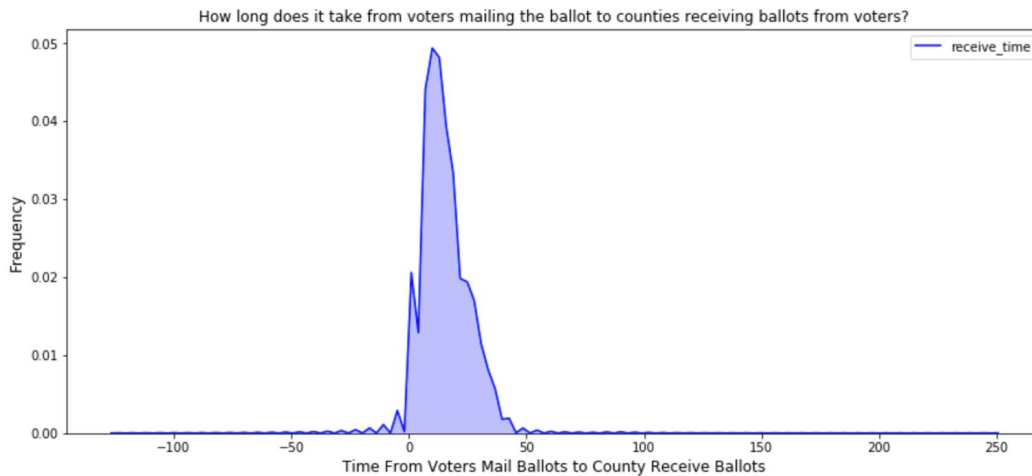
In our dataset, we noticed 23,305 records with counties receiving ballots earlier than voters mailing the ballots. In order to observe the potential fraud, we will examine those suspicious records to see whether these records came from the same county or voters are in one specific party.
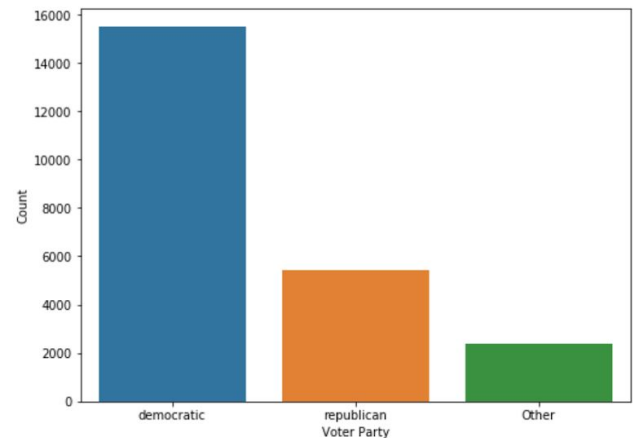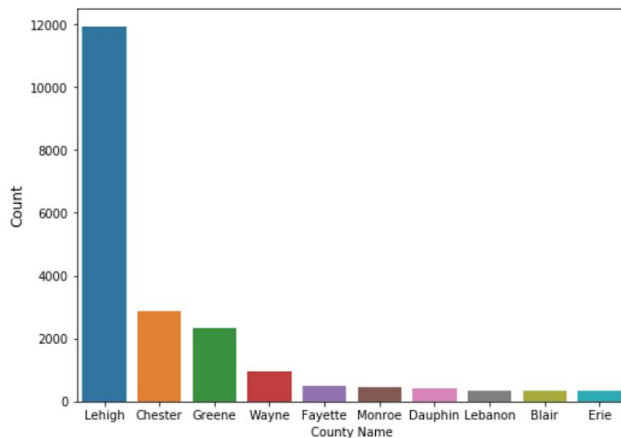

We see most of the suspicious records came from Lehigh, and in those records majority voters are democratic. Due to the limited information, it is hard for us to make any additional claims, so for the following analysis, we will remove those records.

6

How long does it take from voters mailing the ballot to counties receiving ballots from voters?
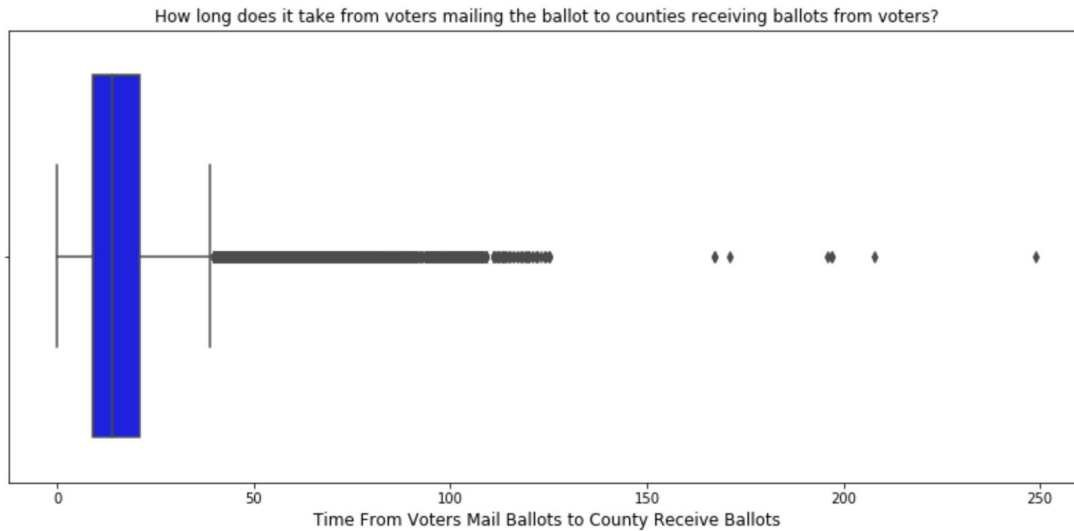
On average, we notice it takes about 16 days from voters mailing the ballot to counties receiving ballots. Overall 75% of counties need less than 3 weeks. We also see some outliers showing the county needs above 5 months to mark the ballot as received, this could happen due to the error when collecting the data.

How were first time voters split across parties and counties?

Election Day served as the culmination of an unprecedented election cycle shaped by the COVID-19 pandemic, and the boundless energy of young people who made their voices heard in the streets as well as at the ballot box. As mentioned earlier, we see some records with date of birth in 2020, and the earliest year of birth is 1800. Those records are not reasonable, in this report, we only include voters who aged from 18 to 122 (records with year of birth between 1898 and 2002).

Distribution of Voters' Political party
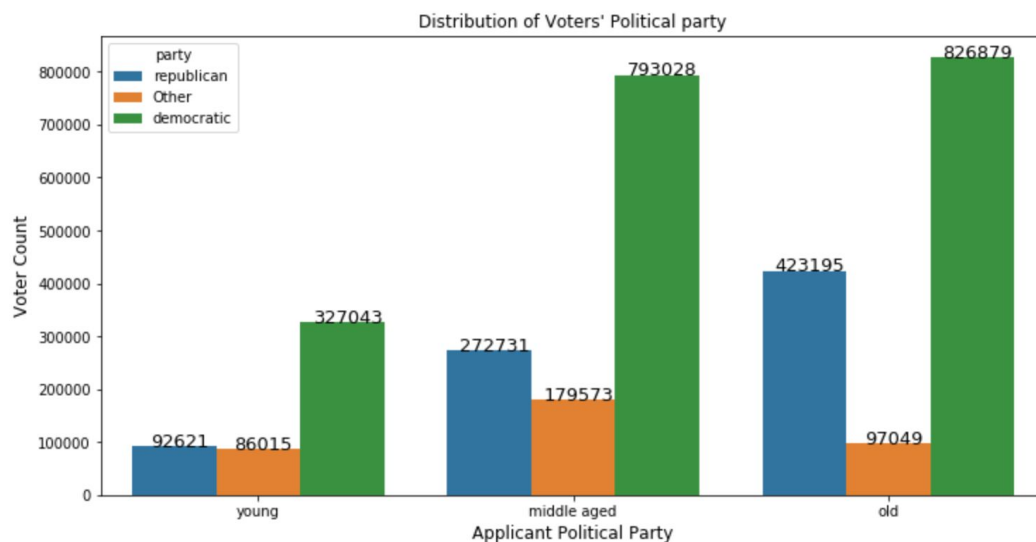
In Pennsylvania, first time voters accounted for 0.93%, and young people (ages 18-30) accounted for 16.32% of total voters. Also, young voters' participation shows an overwhelming support for Democrats.

How early would different age group submit voter registration application?
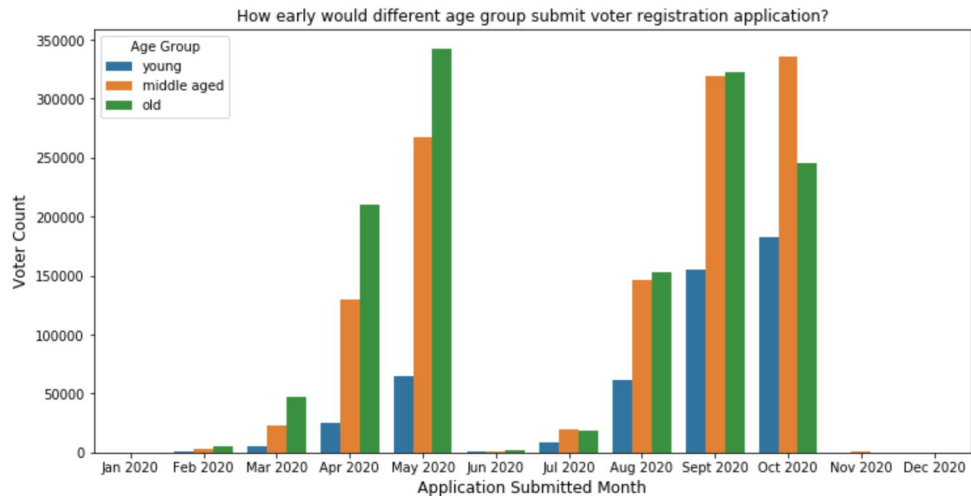
Also our analysis of data highlighted that, among all age groups, more young voters approached this electoral landscape by registering early either for primary elections or general election.

How does the mail ballot request data compare to the unofficial mail ballot results across parties and counties?

Our final discussion is a comparison of the number of mail ballots returned according to the mail ballot request dataset vs. the total number of mail-in votes (unofficial) according to the pa.gov website.

In total, there were 1.66 million mail ballots returned from Democrats according to the mail ballot request dataset. Interestingly, there were 1.99 million mail-in votes from Democrats according to the same data source (pa.gov), or 333 thousand more. On the contrary, mail ballots returned for Republicans roughly equals the number of mail-in votes counted. Finally, the number of mail ballots returned by voters who are independents or have a different political party affiliation exceeded the number of mail-in votes by 262 thousand.

As you can see, these numbers don't add up. There were more mail-in votes counted than mail ballots returned (by approx. 63k votes) according to the different data sets from the same data source. Furthermore, if we assume all of the 'Other' mail-in ballots were cast for Biden, there remains a difference of approximately 68 thousand votes that were counted for Biden. It is possible that 68 thousand registered Republicans mailed-in a vote for Biden so we cannot make any definitive conclusions without seeing and counting the individual mail-in votes.

In general, as you can see below, big blue counties counted more mail-in votes than ballots returned by registered Democrats, suggesting Republicans and other voters affiliated with other parties voted by mail for Biden, too.

Democratic Mail Ballots Returned vs. Unofficial Mail-in Vote Totals by County
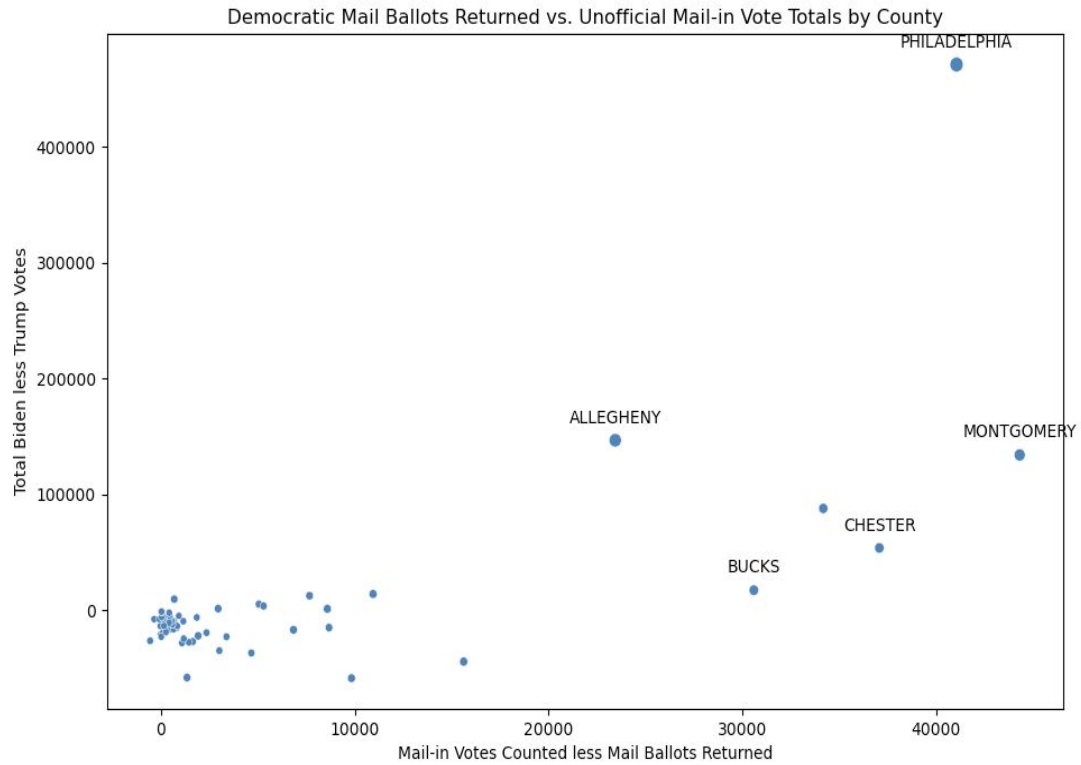
The data for Republican registered voters is more mixed as a majority of counties had higher mail ballots returned than mail-in votes counted.

Republican Mail Ballots Returned vs. Unofficial Mail-in Vote Totals by County
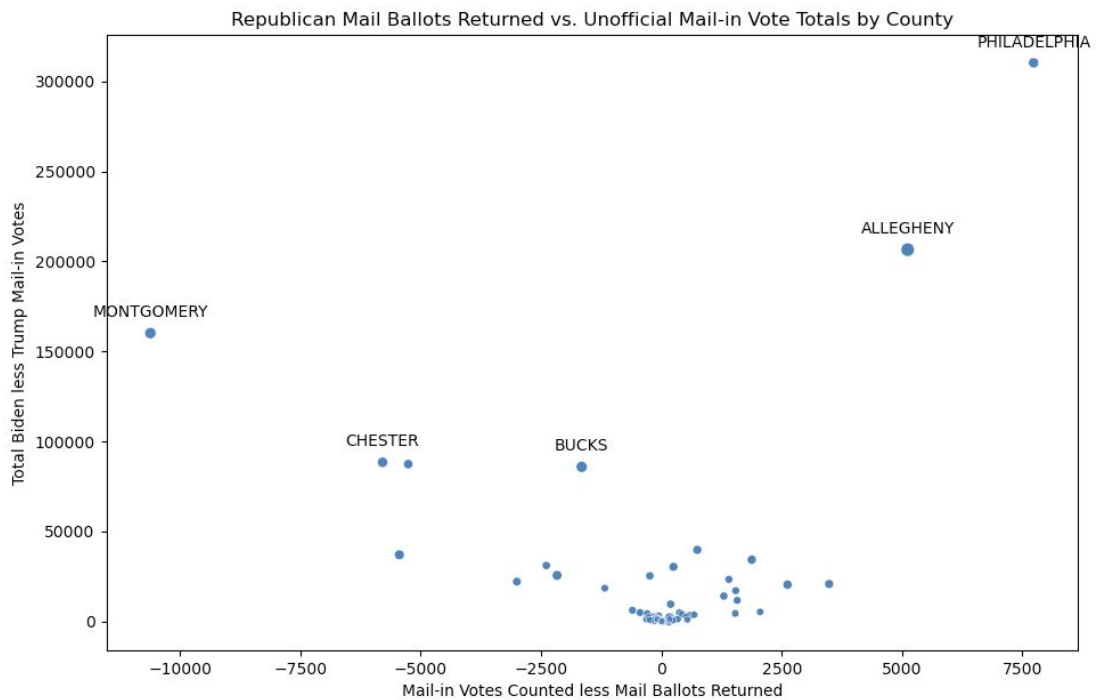
We also looked at the correlation of voting types by voting party. We converted the mail in voting data and the election day voting data as a percentage of total votes casted and looked at the correlation by party line across the sixty-seven counties.
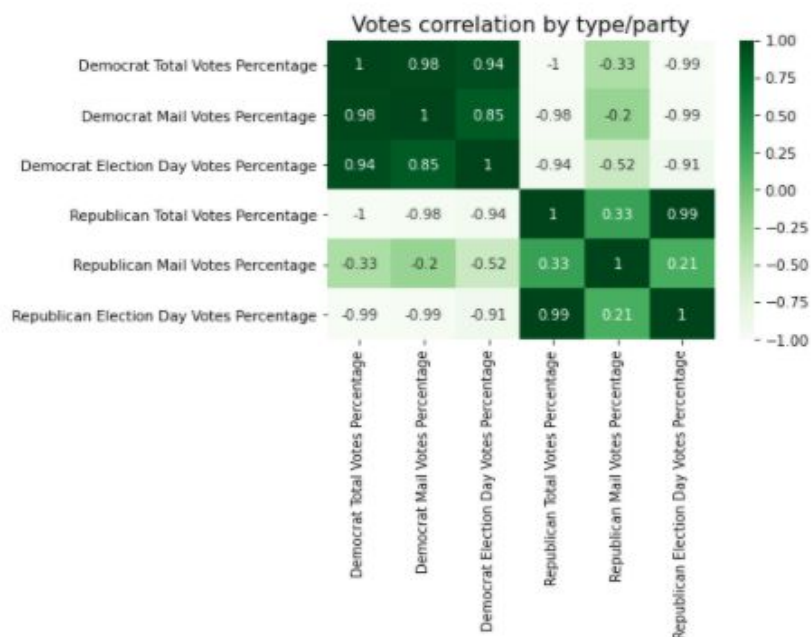
We find the highest level of correlation of (0.98) between the Democrat Total votes and Mail in votes. However, the Republican Total votes is highly correlated (with a value of 0.99) with the election day votes rather than the mail in votes. This is consistent with the significantly larger number of mail in ballots that requested by the democrats compared to the republicans. Interestingly we notice a weak correlation of negative -0.2 between the mail in votes for the two parties across counties. While there was a positive relationship between voters of both parties who returned mail in ballots, however this weak correlation suggests the relationship does not hold for the actual mail in votes that were casted. From this, we can likely infer that there might be a substantial number of registered republican Voters who voted by mail in ballots for democrats.

Conclusion

Consistent with the anecdotes of more Democrats voting via mail this election, we found that a larger percentage of registered Democrats voted by mail (46% Dem vs. 22% Rep). In counties where more than average Democrats voted by mail, we also tended to see more than average Republicans vote by mail.

Applications for mail ballots increased dramatically prior to the primary and general elections across parties, predominantly in large blue counties. The mail ballot request dataset is inconclusive with the overall "unofficial" election results by county. We discovered that more mail ballots were counted for Biden than returned by Dems and 'Other' parties combined, especially in large Democratic strongholds. We however can likely infer that more mail ballots were counted for Biden than returned by Dems and 'Other' parties combined because substantial Republican registered voters could have voted for Biden.

References

[1]Most mail and provisional ballots got counted in past U.S. elections – but many did not
https://www.pewresearch.org/fact-tank/2020/11/10/most-mail-and-provisional-ballots-got-counted-in-past-u-s-elections-but-many-did-not/

[2]Separating Fact From Fiction On Trump And The Post Office - And Why It Matters
https://www.forbes.com/sites/ebauer/2020/08/16/separating-fact-from-fiction-on-trump-and-the-post-office-and-why-it-matters/?sh=437ade933d74

[3]An Analysis of Voter Fraud in The United States, Lorraine C. Minnite, Senior Fellow, Demos.
https://www.demos.org/sites/default/files/publications/Analysis.pdf

[4]Does voting by mail increase fraud? Jonathan Auerbach and Steve Pierson Technical Report∗ Office of Science Policy American Statistical Association
https://www.amstat.org/asa/files/pdfs/POL-Vote-by-Mail.pdf