Election 2020 Pennsylvania Voter Data Analysis

Summary

We analyzed the Pennsylvania State Voter data (Full Voter Export FVE, PAV) and mail-in ballot data (PAM), each published by the State of Pennsylvania, and each downloaded on three different dates (October 26, 2020, November 23, 2020, and August 2, 2021). We have concluded that a number of serious irregularities compel further investigations to ensure the public's confidence in election integrity and constitutional democracy.

Our key findings are the following: (we use "at least" in this study because the related data sets were not updated on time when they were downloaded and thus are under counted.)

1 **485,284** voters were purged from the voter rolls between 11/23/2020 and 08/02/2021. Among those purged voters, at least **61,067** voted in the 2020 general election.

2 PHILADELPHIA and ALLEGHENY were the two top counties with most votes by purged voters, at **7,974** and **6,011 votes** respectively.

3   **107,001** voters were added to the voter rolls between 11/23/2020 and 08/02/2021. Among them at least **546 voters** voted in the 2020 general election.

4 Based on the data set downloaded on 11/23/2020, **2,260,952** inactive voters were on the voter rolls. Among them, **1,188,429** were born before 1988 and **798,248** of them **became active in 2020. 558, 314** of them voted in the 2020 general election.

5 Among inactive voters of all ages, at least <mark>8,480</mark> voted in the 2020 general election and <mark>**were purged**</mark> after the general election in the time frame of 11/23/2020 to 08/02/2021. PHILADELPHIA was the top county with those voters.

6 Of the **558, 314 inactive voters** voted in the 2020 general election as indicated in (4), **2,223 voters** were born before **1930.** Some of them were born as early as **1891**. PHILADELPHIA was the top county with those voters.

7 At least **12,761** mail-in ballots were missing at least one of required voter data such as date of birth (DOB), districts, application date, application approval date, and ballot send out date.

8 **4,091** voters were purged between 10/26/2020 and 11/23/2020. Yet, at least **462** of them voted in the 2020 general election.

9 **6,314** voters were added between 10/26/2020 and 11/23/2020. At least **2,004** of them voted in the 2020 election. Some of these voters were born as early as 1920.

10 There were **488,582 voters** were purged from the voter roll between 10/26/2020~08/02/2021. Of these voters, **28,321** of them voted in the 2020 general election. Voters who also voted in local elections in May 2021 were not counted because their 2020 general election voting records were overwritten by the local election of 2021. Thus voted number is lower than that in (1)

11   There were **312,453** addresses which were in the data set of 11/23/2020, but those addresses were not in the data set of 08/02/021. So they were purged after 11/23/2020.   Among those purged addresses, there were **361,823** voters and **141,240** voted in 11/03/2020 general election. Unless those residential properties were demolished or uninhabited, how were those addresses purged?

12   There were 6k duplicate voters in data set of 11/23/2020. 2k voted in the general election.

1   Who we are

The analyst of this data is an unaffiliated, independent voter, with PhD and MS degrees in multiple science and engineering disciplines. This individual has had more than 30 years of experience working as a scientist and an engineer and would like to remain anonymous.

2   Data sources and methods

PAV "Full Voter Export" data was obtained from Pennsylvania Department of State – SURE Office. It was purchased and downloaded on October 26, 2020, November 23, 2020, and August 2, 2021, with personal funds at the following PA official web site: https://www.pavoterservices.pa.gov/pages/purchasepafullvoterexport.aspx. Data can be referred to in the dictionary of the document titled "Full Voter Export" at the above web site.

PAM data refers to mail-in ballot data, which only includes partial voters' data. It was downloaded free of charge on November 23, 2020, and August 2, 2021, at the following PA official web site (please refer to the web site for data format): https://data.pa.gov/widgets/mcba-yywm.

The PAV and PAM data retrieved in November and August are presumably the updated data sets of those released and downloaded earlier. The dates of these data sets are referred to in YYYYMMDD format in this report.

We loaded the data into MySQL database and used SQL language, MS-Excel and LibreOffice-Calc, together with other Linux tools and shell scripts for data analysis. We can provide those program source codes upon request. PAV data was loaded into tables PAV20201026, PAV20201123 and PAV20210802. PAM data was loaded into tables PAM20201125 and PAM20210805.

3   Results

3.1   Irregularities in mail-in ballots data (PAM)

The PAM data sets contain fields as shown below.

County	Applicant Party Designation	DOB	Mail Application Type	Application Approved Date	Application Return Date	Ballot Mailed Date	Ballot Returned Date	State House District	State Senate District	Congressional District

Figure 1 shows key findings that reflect irregularities in the two sets of PAM. For example, columns 9 and 10 show that over 10,000 ballots returned without mail-out date were recorded by the counties (column 9). Furthermore, counties accepted as valid ballots that were missing required voter data (column 10).   Additionally, about 1,600 ballots returned had no records of the voter's date of birth

(column 8). Even if a voter was missing at least one required voter datum (listed in columns 6, 7, and 8), counties still sent out ballots (column 11).

Figure 1 Mail in ballots irregularities

1	3	4	5	6	7	8	9	10	11
							BALLOTS Returned		
Date obtained	Total Voters	Ballot send	Ballot returned	No Districts information	No Application Approval/Application Return Dates	No DOB	Missing ballot mailed date by counties	At least one of required data is missing (Col 6,7,8)	Counties mailed out ballots without one or more info of Col 6,7,8
11/25/2020	3,088,942	3,069,691	2,495,227	30	40	103	10,261	10,428	327
08/05/2021	3,079,710	3,061,799	2,648,056	58	40	1,601	11,072	12,761	2,077
difference	-9,232	-7,892	152,829	28	0	1,498	811	2,333	1,750

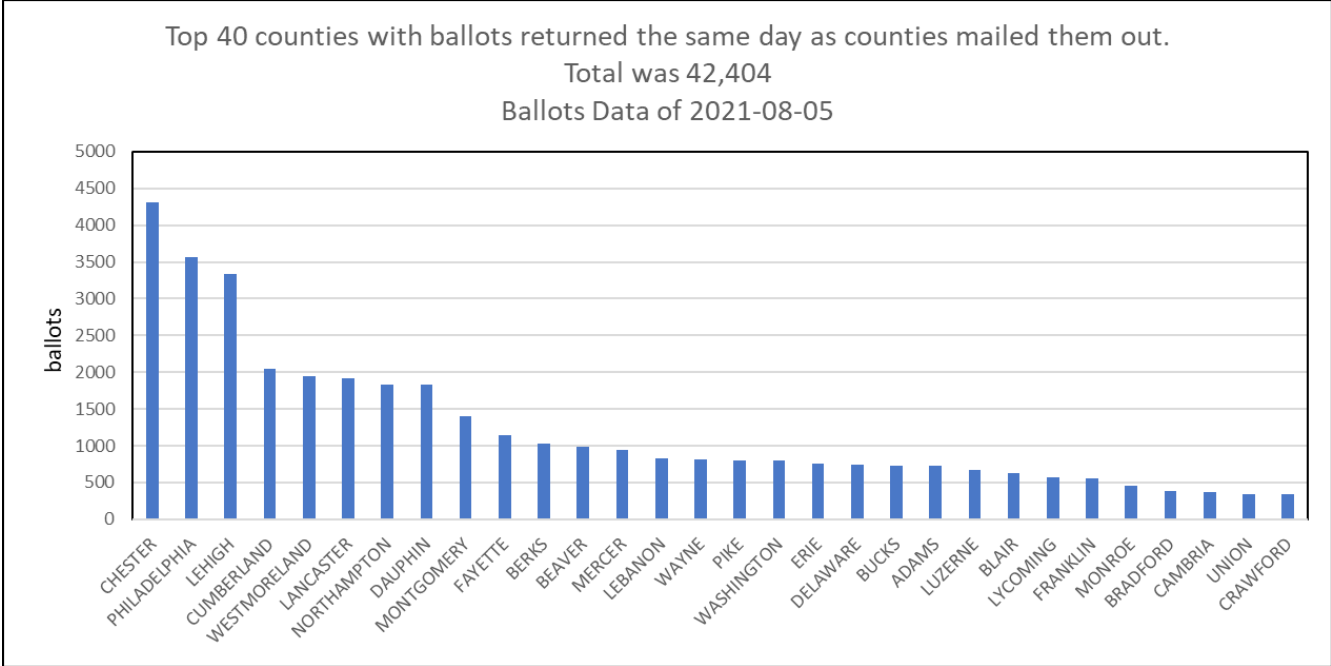*Figure 2 Top ten counties with most mail-in ballots irregularities*

Missing at least one required voter data			
Data from 2020-11-25		Data from 2021-08-05	
YORK	1,915	ALLEGHENY	2,805
BUTLER	1,755	BUTLER	2,349
ALLEGHENY	1,599	YORK	1,957
PHILADELPHIA	604	PHILADELPHIA	737
DELAWARE	499	DELAWARE	509
LEHIGH	494	LEHIGH	507
LANCASTER	454	LANCASTER	458
CENTRE	425	CENTRE	424
SOMERSET	319	SOMERSET	335
ADAMS	282	ADAMS	285

The counties with the most problematic ballots contributing to the results in column 11 of **Figure 1** are identified in Figure 2.This preliminary analysis revealed that **10,428**[1] or **12,761**[2] mail-in ballots are in question.

Based on the data downloaded on 08/05/2021, **42,404** mail-in ballots had the same return date as the send out date. Figure 3 lists the number of ballots for the top 40 counties with the same return and send out dates.[3]
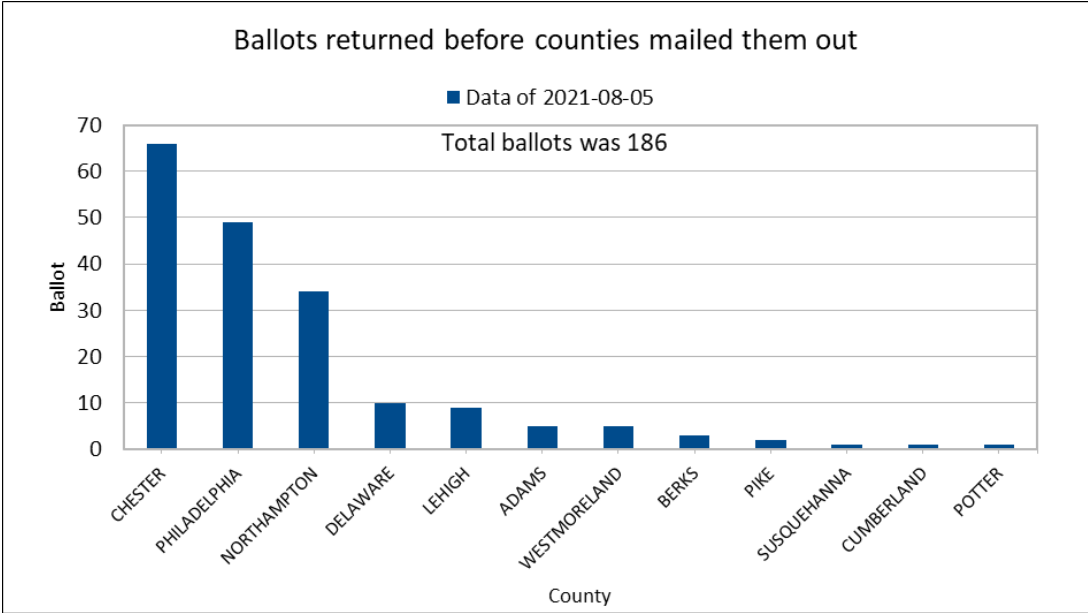
---

1 Data downloaded 11/25/2020
2 Data downloaded 08/05/2021
3 People who were monitoring the mail-in ballots data on PA government website also alleged that some of voters' ballots return dates were changed or modified after they were published and loaded into database. This data publishing website was closed to the public for several weeks after the 2020 general election when this allegation was raised. This allegation needs to be investigated.

Figure 3 Top 40 counties with most ballots with the same return and send out date.

Furthermore, based on the data downloaded on 08/02/2021, **186** ballots had a recorded return date occurring **before** the send-out date (Figure 4).

Figure 4: Ballots returned before counties sent them out

Finally, the top five counties with mail-in ballots received after 11/03/2020 are tabulated in Figure 5 The SCOTUS expressly directed that the late ballots be tallied separately.

Figure 5 Mail-in ballots received after 11/03/2020

LATEST BALLOTS			
LATE BALLOTS			
Data of 2021-11-23 total	65,026	Data of 2021-08-02 total	72,235
PHILADELPHIA	16,006	PHILADELPHIA	19,482
DELAWARE	15,525	DELAWARE	15,920
BUCKS	8,939	BUCKS	9,189
CHESTER	3,278	CHESTER	4,469
MONTGOMERY	3,098	MONTGOMERY	3,213
LATEST DATE On Record	2020-11-16	LATEST DATE Record	2021-02-23

3.2 **Voter roll shifts after 2020 general election**

We compared the voter rolls in the data sets downloaded on 11/23/2020 and on 08/02/2021.

485,284 voters were purged[4] after the 2020 general election between the time frame of 11/23/2020 and 08/02/2021. Among them, at least **61,067** voters voted in the 2020 general election. Figure 6 shows the top ten counties with most votes by purged voters in the 2020 general election.

Figure 6 Top ten counties with purged voters who voted in 2020 general election.

purged voter voted by County	
County	Num
PHILADELPHIA	7,974
ALLEGHENY	6,011
MONTGOMERY	4,649
BUCKS	3,087
LANCASTER	2,202
CHESTER	1,997
DELAWARE	1,973
YORK	1,947
LEHIGH	1,708
NORTHAMPTON	1,645

Additionally, **107,001** voters were added between 11/23/2020 and 08/02/2021. Among them, at least **546** voted in the 2020 general election.

---

4   Please be aware that the data downloaded on 11/23/2020 and 08/02/2021 may not have been up to date. Thus, the actual number of purged voters may be even higher.

Figure 7 lists the number of inactive voters born before 1988 who voted in the 2020 general election. We define "an inactive voter" as a voter not having voted 30 times in any general, midterm or local elections. Figure 7 also includes the number of voters having unusual phone numbers with only 7 digits.

Figure 7 Inactive voters born before 1988

Inactive voters voted in the 2020 general election Voters born before 1988 and did not voted for other 30 elections		
Data sets	2020-11-23	2021-08-02
total inactive voters	1,188,429	851,188
Became active in 2020	798,248	639,600
voted in 2020 general	558,314	332,298
phone has 7 digi	1,216	637
phone has 7 digi voted	489	91
*Right column did not count voters voted in May, 2021. So those numbers were well below the real votes		

Many old inactive voters who voted in the 2020 general election in the data set of 11/23/2020 but were purged from voter rolls in the 08/02/2021 data set (Figure 8).

Figure 8 Very old inactive voters who voted and purged in the data set of 11/23/2020

Birth year	Data 2020-11-03	Data 2021-08-02	Diff
1891	1	1	0
1894	1	1	0
1898	1	0	1
1900	4	2	2
1901	1	1	0
1911	1	1	0
1912	1	0	1
1913	2	2	0
1914	1	1	0
1915	4	3	1
1916	4	3	1
1917	8	4	4
1918	12	6	6
1919	18	10	8
1920	38	22	16
1921	83	39	44
1922	87	51	36
1923	149	56	93
1924	201	94	107
1925	275	133	142
1926	354	169	185
1927	427	194	233
1928	550	244	306
1929	678	323	355
1930	800	377	423

For inactive voters of all ages there were at least **8,480** voters who were purged between 11/23/2020 and 08/02/2021 yet also voted in 2020 general elections. The top ten counties of those voters are listed in Figure 9
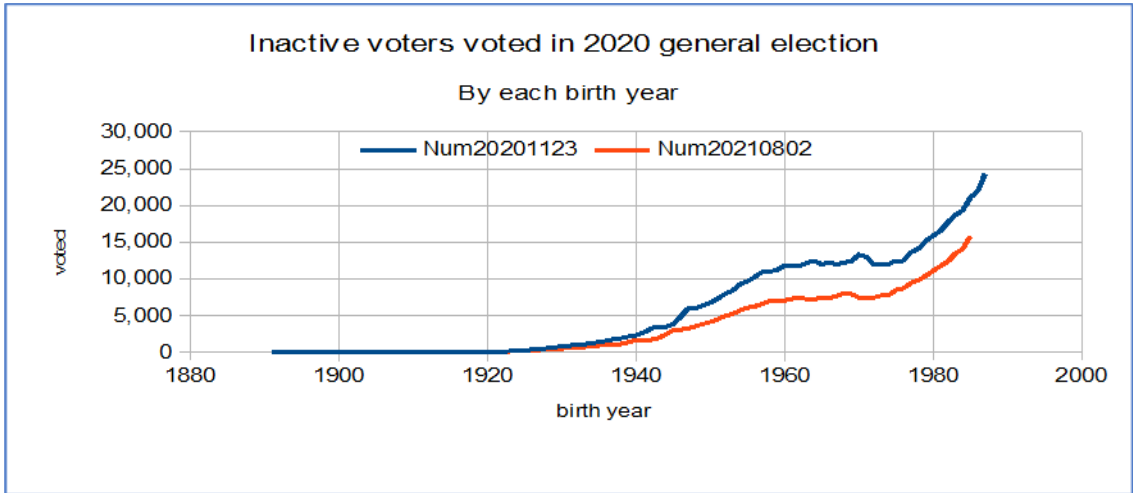
Figure 9 Top 10 counties of inactive voters who voted and later were purged between 11/23/2020 and 08/02/2021

All age voters who were inactive and voted in 2020 general election and also were purged between 11/23/2020 to 08/02/2021	
Top ten counties	purged and voted total at least 8,480
PHILADELPHIA	1938
MONTGOMERY	522
ALLEGHENY	476
BUCKS	430
YORK	420
NORTHAMPTON	306
CHESTER	294
CUMBERLAND	279
LANCASTER	220
BUTLER	216
DELAWARE	192
LEHIGH	192

Figure 10 shows the number of inactive voters by birth year. Figure 11 shows the total number of inactive voters born up to a birth year. For example, as graphed in Figure 11, 558,3145 and 332,2986 inactive voters born before 1988 voted in the 2020 general election. There appear to be many extremely aged voters with birth years before 1920.
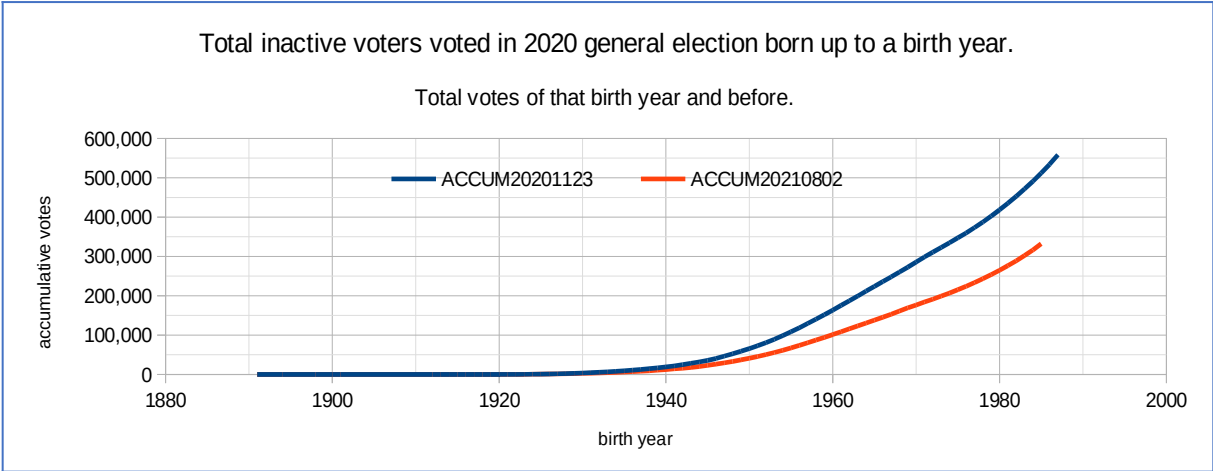
Even if a portion of those voters were subject to voting irregularities, the number of votes may still have influenced the election results. Those past inactive voters who were activated, "voted" in the 2020 election, and then purged after the election should be scrutinized.

Figure 10 Inactive voters who voted organized by birth year from two data sets

5   Sum of inactive voters born before 1988 in data set obtained 11/23/2020
6   Sum of inactive voters born before 1988 in data set obtained 08/02/2021

Figure 11 Sum of inactive voters who voted in the 2020 general election born up to a birth year

Total inactive voters voted in 2020 general election born up to a birth year.

Total votes of that birth year and before.

3.3   **Voter roll shifts around the time of 2020 general election**

A PAV data set was obtained and downloaded on Oct. 26, 2020, with data recorded before the election. A comparison of this data set with one downloaded on Nov. 23, 2020, revealed the voter roll shifts before and immediately after the election. We note that voter registration **deadline** for the 2020 general election was **Oct. 19, 2020**, and that the data downloaded on Oct. 26, 2020, and Nov. 23, 2020, may not have been completely up to date.

We found **4,091** voters were in the voter rolls of 10/26/2020 but not in those of 11/23/2020. This means that at least **4,091** voters were purged around the time of the 11/03/2020 election. Out of those purged, **at least 462** persons voted in the 2020 election. If the data downloaded on the above two dates indeed were not up to date, then the actual numbers would even be higher. Those **462** voters included some born as early as **1916**.

We also found that at least **6,314** voters were added between 10/26/2020 and 11/23/2020. Among them, a**t least 2,004** voters voted in the 2020 election. These added voters had birth years as early as **1897**. The top ten counties with voter's registration dates past the deadline is listed in Figure 12. These "voters" who voted and were then purged right afterwards should be examined carefully.

Figure 12 Past deadline registration votes

County	past deadline reg votes
MONROE	156
BUTLER	133
NORTHUMBERLAND	130
CENTRE	98
LANCASTER	90
YORK	83
WESTMORELAND	82
NORTHAMPTON	73
ALLEGHENY	71
CHESTER	70

3.4   **Voter roll shifts between 10/26/2020 and 08/02/2021**

488,684 voters were in the voter rolls of 10/26/2020 but not in those of 08/02/2021. The shift means a purge of **488,684** voters around nine months after the 11/03/2020 general election**.** Of those purged, <mark>**at least 28,327**</mark> voted in the 2020 general election, including those with birth years as early as <mark>**1823**</mark>.

Figure 13 Voters Purged from Voter Roll of 10/26/2020 between 10/26/2020~08/02/2021
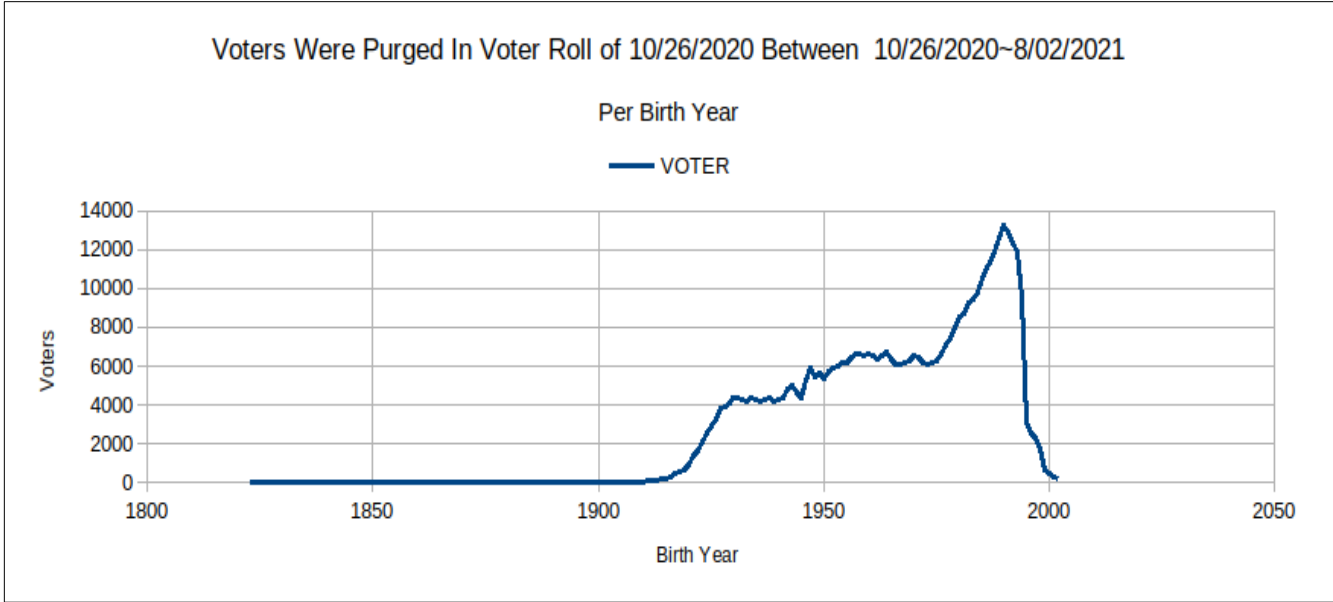

Figure 13 shows those purged voters by each birth year from the voter roll downloaded on 10/26/2020. Those voters were not in the voter roll downloaded on 08/02/2021. Therefore, these voters were purged somewhere between 10/26/2020 and 08/02/2021. The total count of purged voters is <mark>**488,582**</mark>. A large amount of those purged voters voted in the 2020 general election, at least[7] <mark>**28,321**</mark>. Some of these voters were born before 1890.

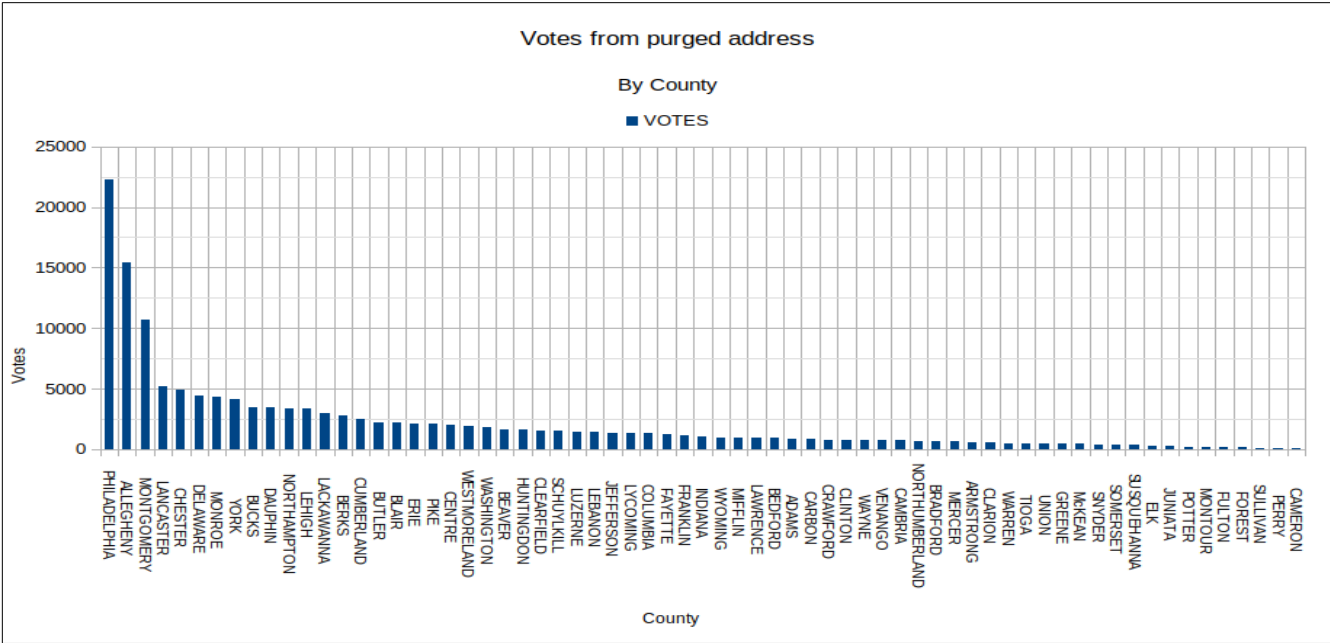**Figure 14** shows the top ten counties with most votes by purged voters as of data from 10/26/2020.

Figure 14 Top ten counties with most votes by purged voters from rolls of 10/26/2020

County	Votes
PHILADELPHIA	3,995
ALLEGHENY	3,740
MONTGOMERY	2,018
BUCKS	1,070
LANCASTER	1,057
CHESTER	1,024
DELAWARE	1,002
NORTHAMPTON	878
WESTMORELAND	747
CUMBERLAND	719

---

7   Please be aware that the data downloaded on 11/23/2020 and 08/02/2021 may not have been up to date. Thus, the actual number of purged voters may be even higher.

3.5   Votes From Purged Addresses

Figure 15: Votes from purged addresses in data set of PAV20201123

There were 312,453 addresses which were in the data set of 11/23/2020, but those addresses were not in the data set of 08/02/2021. So they were purged after 11/23/2020.   Among those purged addresses, there were 361,823 voters and 141,240 voted in the 11/03/2020 general election.   PHILADELPHIA had 22,293, ALLEGHENY had 15,433 and MONTGOMERY had 10,676.

3.6   Identical Voters

Define those properties (Last_Name,First_Name,Middle_Name,Suffix,Gender,DOB) as the voter properties. When they are the same we call them identical voters: (such as twin, triplet, ... etc). There was possibility that there were real different voters among those identical voters. For all three data sets, the numbers as shown below. They need to be visited in person to make sure they were not registered and voted by others or victims of identity theft.

data set	rows	twine	triplet	quadruplets	total multiples	twine voted	triplet voted	quadruplets voted	Total multiple voted
PAV20201026	9087869	6354	36	8	6398	549	1	0	550
PAV20201123	9090143	6394	36	8	6438	2101	8	1	2110
PAV20210802	8711598	4484	21	4	4509	1715	5	0	1720

3.7   Notes

For those who need the source code for our analytical tools and the database tables, we will provide them upon request. Please send an email to: onecitizenusa1@gmail.com.

Readers are permitted to display, use or distribute the materials in this document. The author may provide updates when further information becomes available.