Election Data Handbook

Commonwealth of Pennsylvania, Department of State

Table of Contents

Preface 3

1. Data considerations and limitations 4

1.1 What is SURE? 4

1.1a The SURE system represents a “point-in-time" snapshot of the voters. 4

1.2 We use vote history to explore election trends 4

1.2a How is vote history recorded? 4

1.3 SURE data limitations. 4

1.3a Data entry is not immediate. 4

1.3b Artifacts of legislative updates can be seen in the SURE database. 5

1.3c Are all records included in the provided raw data? 5

2. Summary Election Data Analyses 5

2.1 Vote History & Voter Registration Summary 5

2.2 Mail Ballot Summary 9

2.3 Open Data Mail Ballot Requests File 13

2.4 Presidential Turnout 18

2.6 Certified Provisional Ballot Summary 19

2.6a Certified Provisional Ballots file 20

2.6b Certified Provisional Ballots – Demographic Supplemental file 24

3. Election Results & Returns 26

3.1 Election Night Results 26

3.2 Supplemental Election Night Results Dashboard 27

3.3 Bulk Election Data 27

4. Election data provided in lead-up to the election 28

4.1 Full Voter Export 28

4.2 Statewide Mail Ballot File 31

4.3 Daily Mail Ballot Report 35

Preface

This handbook accompanies the election data the Department of State (DOS) Bureau of Elections provides on its website. The data used in the files described in this handbook come from one of two sources: the Statewide Uniform Registry of Electors ("SURE") and certified election results provided by the county election offices to DOS. The source will be listed for each data file summary. An overview of the SURE system is provided along with SURE-related FAQs in Section 1 of this handbook.

If you have a data-related question that is not answered in this handbook, please check our website at https://www.pa.gov/agencies/dos/resources/voting-and-elections-resources.html. You can also reach out to DOS via email at: ra-elections@pa.gov.

Please note that this handbook does not cover the following:

- Data related to weekly voter registration. This data and explanations can be found here: https://www.pa.gov/agencies/dos/resources/voting-and-elections-resources/voting-and-election-statistics.html
- Data related to campaign finance as well as lobbying and disclosure. Links to those bulk data and reports can be found here: https://www.pa.gov/en/agencies/dos/resources/voting-and-elections-resources/voting-and-election-statistics.html
- Data provided through any official reports - such as the annual voter registration report and post-election audit reports. These reports already have a detailed explanation on the pa.gov website: https://www.pa.gov/en/agencies/dos/resources/voting-and-elections-resources/election-reports.html

Counties may provide additional data on their own websites. Please check your county's election website for more information: vote.pa.gov/county.

1. Data considerations and limitations

For more FAQ's, please go to our Fact Checking website here:
https://www.pa.gov/en/agencies/vote/elections/fact-checking-pa-related-election-claims.html

1.1 What is SURE?

Much of the data described in this handbook is maintained in the Statewide Uniform Registry of Electors (**SURE)** system database. The SURE system is the official voter registry for the Commonwealth of Pennsylvania, and all 67 counties must maintain their voter registration records in the SURE system. The Department of State (DOS) *maintains* the SURE system, which tracks voter registrations using the same system across all counties in the Commonwealth, but the 67 counties are responsible for *entering the data* into SURE. Because **counties are responsible for data quality and accuracy,** differences in county data-entry practices can cause variations.

1.1a The SURE system represents a "point-in-time" snapshot of the voters.

The data in SURE is continuously updated by counties to represent the current state of registered voters in Pennsylvania. Some updates that counties may make include adding new eligible voters to their rolls as well as removing voters who moved or died.

With counties updating information daily in SURE, data differs from day to day (even minute to minute). If you notice that two files do not match exactly, that could be because the data was pulled at two different times. Static data files providing a summary of point-in-time data in SURE give insight into election trends, without interfering with the important work our counties do to consistently monitor and update the rolls.

1.2 We use vote history to explore election trends.

Counties must record Vote History in SURE after each election. Vote history shows whether a voter voted or appeared to vote in the election. Vote history also includes the voter's method of voting (i.e., at polls, by mail, or provisionally), their party when they voted, and the precinct where they voted.

Your ballot is secret in Pennsylvania, meaning how you voted on your ballot cannot be associated with your voter record. If we want to monitor election trends like 1) how many voters within a specific legislative district voted in an election, 2) how many people voted from each political party, and 3) the average age of voters in an election, we need to use the "vote history" data associated with an individual's voter record.

1.2a How is vote history recorded?

Vote history is entered into SURE by the county. Vote history is linked to a voter's record using identifying information provided at the time of voting. For instance, when you vote in-person, you sign a poll book that has a unique identifier next to your name. When you vote by mail, vote history is awarded when the ballot is marked as returned in SURE.

1.3 SURE data limitations.

1.3a Data entry is not immediate.

Many aspects of data entry are manual, such as scanning poll books, and they take time to be completed. Counties may also wait to enter vote history into SURE until they are done with other election processes, such as reviewing provisional ballots. Further, statutory requirements prevent mail

ballots from being counted until election day. Any delays in county data entry subsequently delay DOS's ability to provide the data.

1.3b Artifacts of legislative updates can be seen in the SURE database.

The SURE system was created in 2003. Since then, SURE underwent updates to reflect legal changes, while retaining certain data practices that preceded its implementation. You may notice:

- When SURE was created, some older records were missing information, such as voters' birthdays or registration dates. These were backfilled with dates such as January 1, 1900, or January 1, 1800.
- Due to voter security concerns, some voters may not have their birthdate included in the system. These dates are then filled with a placeholder date as well.
- From 2020 onward, DOS implemented more extensive codes for counties to record mail ballot cancellation reasons to accommodate legal requirements. Counties may also change their administrative and data entry practices between elections.

1.3c Are all records included in the provided raw data?

No – protected voters are not included in the Full Voter Export, the Statewide Mail Ballot File, or the Open Data Mail Ballot Requests file. Analyses that use these data sources as the raw data will accordingly not include such excluded voters, though they may appear in protected anonymized form or in total counts in other data sources (e.g., summary files). Further, the Full Voter Export does not include cancelled voter records and only includes vote history in the last 40 elections held per county.

The Full Voter Export cannot be used to "check" the number of ballots cast in an election, nor does DOS rely on it in analyzing data for any election.

2. Summary Election Data Analyses

These analyses are provided once per election, after all counties have finished entering election-related data into SURE. This is our best estimate at 'Final' data for the election using a point-in-time database, but it is still an *estimate*.

The following summary data analysis files are provided. Please see their corresponding section for more details.

1. Vote History & Certified Voter Registration Summary
2. Mail Ballot Summary

Section 2.3 details the raw data that corresponds with the Mail Ballot Summary file, which is hosted on PA Open Data every election.

2.1 Vote History & Voter Registration Summary

Data source: This analysis uses data pulled from SURE. Vote history is used to obtain breakdowns by congressional district, state senate district, state legislative district, party, and age group.

The voter registration data used in this analysis was pulled at the voter-level for each county on the day the county certified and submitted their summary count of certified voters (Certified voters as submitted by the county are available here under the 'Voter Registration Statistics Archive'). In most cases, the analyses provided in the 'Vote History & Certified Voter Registration Summary' match the

count of certified voters submitted by the county prior to the election, but there are minor discrepancies due to point-in-time data.

Please note: There may be discrepancies between the data for the Vote History & Voter Registration Summary file and the data in the Open Data Mail Ballot Requests file due to the treatment of protected voters. For more information, refer to Section 1.3c.

File meta-data:

This file includes seven tables of data. See below for each table along with their column breakdowns. The name of each table corresponds with the sheet name in the Excel file. There is also a 'Disclaimers' sheet in the excel file, which includes some highlights of this data handbook for quick reference.

1. "By county"

Column name	Column description
County	The name of each county, in alphabetical order. There is also a 'TOTAL' row at the bottom of the table, leading to 68 rows.
Vote History	The number of unique voters with vote history in SURE by county for the election.
Registered voters	The number of unique registered voters per county.
Percent vote history/Registered voters	Represents the number of people with vote history out of the number of registered voters per county. Calculation is as follows: Vote History/Registered voters*100

2. "By vote method"

Note: This table includes vote history only, no voter registration.

Column name	Column description
County	The name of each county in alphabetical order. There is also a 'TOTAL' row at the bottom of the table, leading to 68 rows.
Absentee	The number of voters with vote history who voted via absentee ballot.
At Polls	The number of voters with vote history for voting at the polls (election day) by signing the poll book.
Mail-in Ballot (excluding absentee)	The number of voters with vote history who voted via mail-in ballot excluding absentee.
Provisional	The number of voters with vote history who voted provisionally.

3. "By congressional district"

Column name	Column description
Congressional district	The name of the congressional district associated with a voter's vote history.
Vote history	The number of unique voters with vote history in SURE by congressional district for the election.
Registered voters	The number of unique registered voters per congressional district.

Percent vote history/Registered voters	Represents the number of people with vote history out of the number of registered voters per congressional district. Calculation is as follows: Vote History/Registered voters*100

4. “By senate district”

   This is the same as the ‘By congressional district’ table but uses state senate district information instead.

5. “By state legislative district”

   This is the same as the ‘By congressional district’ table but uses state legislative district information instead.

6. “By county & party”

Column names	**Column descriptions**
County	The name of each county, in alphabetical order. There is also a ‘TOTAL’ row at the bottom of the table, leading to 68 rows.
Democratic Vote History	The number of unique voters with vote history that were registered as Democrats at the time of voting. Broken down by county.
Republican Vote History	The number of unique voters with vote history that were registered as Republicans at the time of voting. Broken down by county.
Libertarian Vote History	The number of unique voters with vote history that were registered as Libertarians at the time of voting. Broken down by county.
Other Vote History	The number of unique voters with vote history that were registered as independents (not registered as any of the 3 identified parties in this table) at the time of voting. Broken down by county.
Democratic Voter Registration	The number of unique registered Democratic voters per county.
Republican Voter Registration	The number of unique registered Republican voters per county.
Libertarian Voter Registration	The number of unique registered Libertarian voters per county.
Other Voter Registration	The number of unique registered independent voters per county.
Democratic Percent vote history/Registered voters	Represents the number of Democrats with vote history out of the number of Democrat registered voters per county. Calculation is as follows: Vote History/Registered voters*100
Republican Percent vote history/Registered voters	Represents the number of Republicans with vote history out of the number of Republican registered voters per county. Calculation is as follows: Vote History/Registered voters*100
Libertarian Percent vote history/Registered voters	Represents the number of Libertarians with vote history out of the number of Libertarian registered voters per county. Calculation is as follows: Vote History/Registered voters*100

Other Percent vote history/Registered voters	Represents the number of independents with vote history out of the number of independent registered voters per county. Calculation is as follows: Vote History/Registered voters*100

7. “By county & age”

Column name	**Column description**
County	The name of each county, in alphabetical order. There is also a ‘TOTAL’ row at the bottom of the table, leading to 68 rows.
18 to 24 Vote History	The number of unique voters with vote history between ages 18 and 24, broken down by county.
25 to 34 Vote History	The number of unique voters with vote history between ages 25 and 34, broken down by county.
35 to 44 Vote History	The number of unique voters with vote history between ages 35 and 44, broken down by county.
45 to 54 Vote History	The number of unique voters with vote history between ages 45 and 54, broken down by county.
55 to 64 Vote History	The number of unique voters with vote history between ages 55 and 64, broken down by county.
65 to 74 Vote History	The number of unique voters with vote history between ages 65 and 74, broken down by county.
75+ Vote History	The number of unique voters with vote history age 75 plus, broken down by county.
18 to 24 Voter Registration	The number of unique registered voters ages 18 to 24 per county.
25 to 34 Voter Registration	The number of unique registered voters ages 25 to 34 per county.
35 to 44 Voter Registration	The number of unique registered voters ages 35 to 44 per county.
45 to 54 Voter Registration	The number of unique registered voters ages 45 to 54 per county.
55 to 64 Voter Registration	The number of unique registered voters ages 55 to 64 per county.
65 to 74 Voter Registration	The number of unique registered voters ages 65 to 74 per county.
75+ Voter Registration	The number of unique registered voters aged 75 plus per county.
18 to 24 Percent Vote History/Registered Voters	Represents the number of voters ages 18 to 24 with vote history out of the number of registered voters ages 18 to 24 per county. Calculation is as follows: Vote History/Registered voters*100
25 to 34 Percent Vote History/Registered Voters	Represents the number of voters ages 25 to 34 with vote history out of the number of registered voters ages 25 to 34 per county. Calculation is as follows: Vote History/Registered voters*100
35 to 44 Percent Vote History/Registered Voters	Represents the number of voters ages 34 to 44 with vote history out of the number of registered voters ages 35 to 44 per county. Calculation is as follows: Vote History/Registered voters*100
45 to 54 Percent Vote History/Registered Voters	Represents the number of voters ages 45 to 54 with vote history out of the number of registered voters ages 45 to 54 per county. Calculation is as follows: Vote History/Registered voters*100
55 to 64 Percent Vote History/Registered Voters	Represents the number of voters ages 55 to 64 with vote history out of the number of registered voters ages 55 to 64 per county. Calculation is as follows: Vote History/Registered voters*100

65 to 74 Percent Vote History/Registered Voters	Represents the number of voters ages 65 to 74 with vote history out of the number of registered voters ages 65 to 74 per county. Calculation is as follows: Vote History/Registered voters*100
75+ Percent Vote History/Registered Voters	Represents the number of voters aged 75 and older with vote history out of the number of registered voters ages 75 and older per county. Calculation is as follows: Vote History/Registered voters*100

2.2 Mail Ballot Summary

Data source: This analysis uses data pulled from SURE.

When reviewing this data, please keep in mind the data considerations discussed in Section 1 of this Handbook. Variations in county data entry practices, timing of the data pulls, and assignment of vote history are particularly acute in reviewing mail ballot data, which should not be considered exact. Likewise, limitations should be kept in mind when comparing across election cycles. **See *Section 1.3b*.**

Additional notes specific to this dataset:

- The 'Open Data Mail Ballot Requests' is used for this analysis. It is a deidentified version of the Statewide Mail Ballot file. The final version of the 'Open Data Mail Ballot Requests' file uses vote history in its compilation, which differs from iterations of the file produced prior to the election.
- Since the raw data in the 'Open Data Mail Ballot Requests' file is based on a file that includes voters' personal information, protected voters are not included. Due to this, the total number of counted (i.e., recorded) mail ballots in the Mail Ballot Summary are less than the total number of mail ballot vote history counts in the Vote History & Voter Registration Summary.
- Counties can only enter one cancel code per ballot, even if there are multiple errors. Different counties may vary in what code they choose to use when this arises, or they may choose to use the catch-all category of 'CANC - OTHER'.
- In our analyses, we excluded any 'DECL – DUPLICATE APPLICATIONS' (a code that is to be used when a voter submits more than one application, and the duplicate application is declined because another valid application has been/is being processed) and filtered to only 'Approved' applications for most analyses. Reasons an application may be rejected include a voter having already submitted an approved application or a voter not being registered in that county.
- Though rare, there are times when certain codes are left NULL for a voter in our database (e.g., PrecinctCode). If this occurs, you may see a row corresponding to those NULL records and/or you may see totals that differ from the totals on the 'By county' tab. Please refer to the 'By county' tab for county totals and statewide totals.

File meta-data:

This file includes eleven tables of data. See below for each table along with their column breakdowns. The name of each table corresponds with the sheet name in the excel file, which is in quotes below. There is also a 'Disclaimers' sheet in the excel file, which includes some highlights of this data handbook.

1. "Mail ballot by county"

Column names	**Column descriptions**

County	The name of each county in alphabetical order. There is also a 'TOTAL' row at the bottom of the table, leading to 68 rows.
Total cancelled ballots	The number of mail-in and absentee ballots cancelled that were returned by a voter. This includes ballots cancelled for the following corresponding reasons in the raw data: 'CANC – NO ID', 'CANC – NO SIGNATURE', 'CANC – NO SECRECY ENVELOPE', 'CANC – RETURNED AFTER THE DEADLINE', 'CANC – VOTE CHALLENGED', 'CANC – OTHER'. This also includes corresponding PEND codes that counties did not move to cancelled (e.g., PEND – NO ID). • This data excludes administrative cancellations. • This data is provided per county. • This is provided at the ballot level. If a voter had multiple ballots cancelled with one of the codes above, all would be included in this summary. Also, if a voter had both a cancelled ballot and a recorded ballot, both would be included.
Total recorded ballots	The total number of mail-in and absentee ballots that were returned by a voter to the election's office and subsequently counted. In the raw data, this includes the following codes for 'Ballot status reason': "RECORD - BALLOT RETURNED", "RECORD - FWAB RETURNED", "FWAB OVERRIDE-OFFICIAL BALLOT RECEIVED". This table includes all recorded ballots. If a voter had both a cancelled and a recorded ballot, both would be included.
Total returned ballots	The number of mail-in and absentee ballots that were returned by a voter to the election's office. This includes both recorded ballots and cancelled ballots.
Total ballots	The number of mail-in and absentee ballots issued that could possibly be returned.
% Cancelled of total ballots	The percent of ballots cancelled out of the number of total ballots Formula: 'Total cancelled ballots'/'Total ballots'*100
% Returned of total ballots	The percent of returned ballots out of the number of total ballots Formula: 'Total returned ballots'/'Total ballots'*100
% Recorded of total ballots	The percent of recorded ballots out of the number of total ballots. Formula: 'Total recorded ballots'/'Total ballots'*100
% Cancelled of total returned	The percent of cancelled ballots out of the number of returned ballots. Formula: 'Total cancelled ballots'/'Total returned ballots'*100
% Recorded of total returned	The percent of recorded ballots out of the number of returned ballots. Formula: 'Total recorded ballots'/'Total returned ballots'*100

2. "Statewide by age"

Summary table of cancelled ballot statistics grouped by age group at the statewide level. This includes the same columns as the 'Mail ballot by county' table (above), except it drops 'County' and groups by 'Age group.' The 'Age group' column is described below:

Age group: Using the 'DateofBirth' column in the raw data, voters were organized into seven age groups – 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, 65 to 74, and 75+. Ballot counts are provided by each age group statewide.

3. "Statewide by party"

Summary table of cancelled ballot statistics grouped by political party at the statewide level. This includes the same columns as the 'Mail ballot by county' table (above), except it drops 'County' and groups by 'Party.' The 'Party' column is described below:

Party: Using the 'Party' column in the raw data, voters were organized into the following parties: Democratic, Republican, Libertarian, and Independents/Other. Ballot counts are provided by each party statewide.

4. "By county & age"

Summary table of cancelled ballot statistics grouped by age group at the county level. This includes the same columns as the 'Mail ballot by county' table, except it adds an additional column of 'Age group.' The 'Age group' column is described below:

Age group: Using the 'DateofBirth' column in the raw data, voters were organized into seven age groups – 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, 65 to 74, and 75+. Ballot counts provided by each age group per county.

5. "By county & party"

Summary table of cancelled ballot statistics grouped by political party at the county level. This includes the same columns as the 'Mail ballot by county' table, except it adds the 'Party' column. The 'Party' column is described below:

Party: Using the 'Party' column in the raw data, voters were organized into Democrats, Republicans, Libertarians, and Independents/Other. Ballot counts are provided by each party per county.

6. "By congressional district"

Summary table of cancelled ballot statistics grouped by congressional district. This includes the same columns as the 'Mail ballot by county' table, except it adds the 'Congressional District' column and drops the 'County' column. The 'Congressional district' column is described below:

Congressional District: The name of the congressional district (1-17) where the mail ballot was cast. There are circumstances where this value is NULL for a voter.

7. "By senate district"

Summary table of cancelled ballot statistics grouped by state senate district. This includes the same columns as the 'Mail ballot by county' table, except it adds the 'Senate District' column and drops the 'County' column. The 'Senate district' column is described below:

Senate District: The name of the state Senate district (1-50) where the mail ballot was cast. There are circumstances where this value is NULL for a voter.

8. “By legislative district”

Summary table of cancelled ballot statistics grouped by state legislative district. This includes the same columns as the ‘Mail ballot by county’ table, except it adds the ‘Legislative District’ column and drops the ‘County’ column. The ‘Legislative district’ column is described below:

Legislative District: The name of the state Legislative district (1-203) where the mail ballot was cast. There are circumstances where this value is NULL for a voter.

9. “Cancelled ballot statistics”

This table includes the total number of ballots returned by a voter, sorted by cancelled ballot status as recorded in SURE by each county. For related considerations, please refer to Section 1 of this handbook, and the introduction to Section 2.3.

This table includes ALL cancelled ballots per status. If a voter had multiple ballots cancelled, all would be included in this summary.

Column name	**Column description**
County	The name of each county in alphabetical order. There is also a ‘TOTAL’ row at the bottom of the table, leading to 68 rows.
CANC – NO SECRECY ENVELOPE	The number of ballots counties recorded in SURE as cancelled for being returned without a secrecy envelope.
CANC – RETURNED AFTER DEADLINE	The number of ballots counties recorded in SURE as cancelled for being returned after 8 PM on election day.
CANC – NO SIGNATURE	The number of ballots counties recorded in SURE as cancelled for not including a signature on their return envelope.
CANC - OTHER	A category that a county may choose to report a ballot was cancelled for reasons beyond what the other cancel codes indicate. Use of this code varies by county practice, but a common application is when a ballot return envelope has more than one error.
CANC – NO ID	The number of ballots counties recorded in SURE as cancelled because the voter’s identification was not verified before the statutory deadline. For more information, refer to the “Identity Verification of Absentee and Mail-in Ballot Applicants” section of the Department’s Guidance Concerning Civilian Absentee and Mail-in Ballot Procedures. https://www.pa.gov/agencies/dos/resources/voting-and-elections-resources/directives-and-guidance.html
PEND – NO ID	The PEND code that corresponds with CANC– NO ID. This code is used by counties who have provided the voter with a ballot pursuant to Pennsylvania law, but who have not yet verified their identifying information, and will be cancelled if they fail to do so by the statutory deadline. After the election, these are generally moved to the ‘CANC – NO ID’ code.
PEND - OTHER	The PEND code that corresponds with CANC – OTHER. This code is used by counties who wish to inform the voter that the county has

	a process for the voter to redress an observed error. After the election, these are generally moved to the 'CANC – OTHER' code.
CANC – VOTE CHALLENGED	The number of ballots counties recorded in SURE as cancelled because the voter's mail ballot application was successfully challenged.
PEND – NO SIGNATURE	The PEND code that corresponds with CANC – NO SIGNATURE. This code is used by counties who wish to inform the voter that the county has a process for the voter to redress the observed error. After the election, these are generally moved to the 'CANC – NO SIGNATURE' code.
PEND – NO SECRECY ENVELOPE	The PEND code that corresponds with CANC – NO SECRECY ENVELOPE. This code is used by counties who wish to inform the voter that the county has a process for the voter to redress the observed error. After the election, these are generally moved to the 'CANC – NO SECRECY ENVELOPE' code.
Total	A summary column that provides the sum of cancelled and pending ballots in each county, summed across all the CANC and PEND codes included in this table. This table only includes ballots that were returned by the voter and then subsequently cancelled. It does not include ballots cancelled for administrative error (e.g., label cancelled) or where the voter never received the ballot (e.g., ballot was returned to the election's office as undeliverable).

10. "Cancelled ballots by age"

This table is like the 'Cancelled Ballots Statistics' table but adds an additional breakdown by age group. The 'Age group' column is described below:

Age group: Using the 'DateofBirth' column in the raw data, voters were organized into seven age groups – 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, 65 to 74, and 75+. Ballot counts provided by each age group per county.

11. "Cancelled ballots by party"

This table is like the 'Cancelled Ballots Statistics' table but adds an additional breakdown by party. The 'Party' column is described below:

Party: Using the 'Party' column in the raw data, voters were organized into Democrats, Republicans, Libertarians, and Independents/Other. Ballot counts are provided by each party per county.

2.3 Open Data Mail Ballot Requests File

Hosted on PA OpenData: https://data.pa.gov/. To find the election you are interested in, enter 'Election Mail Ballot Requests' and the year of the election in the search bar.

Data source: This data is extracted from SURE and reflects activity recorded by the counties in the SURE system at the time of the data extraction. See documentation on Open Data for time and date of data extraction.

When reviewing this data, please keep in mind the data considerations discussed in Section 1 of this Handbook.

Additional notes specific to this dataset:

- Counties can enter cancellation codes without entering a ballot returned date.
- Some cancellation codes are a result of administrative processes, meaning the ballot was never mailed to the voter before it was cancelled (e.g., there was an error when the label was printed).
- The raw data used in these analyses (from an identified version of the 'Open Data Mail Ballot Requests' file) is based on a file that has voters' personal information, meaning protected voters are not included.
- Counties can only enter one cancel code per ballot, even if there are multiple errors. Different counties may vary in what code they choose to use when this situation arises, or they may choose to use the catch-all category of 'CANC - OTHER'.
- Counties may use 'PEND' codes as part of their notice and cure practice. These are usually converted to 'CANC' codes after the election. However, in situations where PEND codes remain after the election, these should be considered cancelled.
- Columns and data codes included in this file have evolved over time. For example, for past elections (e.g., 2020), cancelled ballots were not included in the file. This may make it difficult to compare data from election to election.

Type of data included in this file:
This data includes all mail ballot applications processed by counties, which includes voters on the permanent mail-in and absentee ballot lists. Multiple rows in this data may correspond to the same voter if they submitted more than one application or had a cancelled ballot(s). A deidentified voter ID has been provided to allow data users to identify when rows correspond to the same voter. This ID is randomized and cannot be used to match to SURE, the Full Voter Export, or previous iterations of the Statewide Mail Ballot File.

All application types in this file are considered a type of mail ballot. Some of the applications are considered UOCAVA (Uniformed and Overseas Citizens Absentee Voting Act) or UMOVA (Uniform Military and Overseas Voters Act) ballots. These are listed below:

• CRI - Civilian - Remote/Isolated
• CVO - Civilian Overseas
• F - Federal (Unregistered)
• M - Military
• MRI - Military - Remote/Isolated
• V - Veteran
• BV - Bedridden Veteran
• BVRI - Bedridden Veteran - Remote/Isolated
*We may not have all application types for every election.

File meta-data:

Column names	Column descriptions
CountyName	This identifies the county of registration.
Party	The voter's party as indicated on their application. The following are the possible values of this field: • DEM: Democratic • REP: Republican • LIB: Libertarian • GRN: Green • OTH: Other/Independent
DateofBirth	Voter's Month and Year of Birth in YYYY-MM format
Mail Application Type	The mail ballot application type. The following are the possible values for this field. • Alt - Alternative ballot (voters who are disabled or at least 65 years of age where the polling place may not be fully accessible) • BV - Bedridden Veteran • BVRI - Bedridden Veteran - Remote/Isolated • C - Emergency Absentee ballot application • CIV - Civilian absentee submitted via paper • CRI - Overseas voter - Remote/Isolated • CVO - Overseas voter • F - Federal voter (voter is eligible to vote in federal races only) • M - Military • MAILIN - Mail-in submitted via paper • MRI - Military - Remote/Isolated • OLMAILV - Mail-in submitted online where the applicant's ID was verified at the time of submission • OLMAILNV - Mail-In submitted online where ID verification did not occur or was not successful at the time of submission • OLREGV - Absentee submitted online where the applicant's ID was verified at the time of submission • PER - Absentee where applicant has requested permanent status • PMI - Mail-in where applicant has requested permanent status • REG - Absentee submitted via paper • V - Veteran *Note: An application is designated as NV (not verified) if the county was unable to verify the application via the ID verification process prior to approving the application.* ***The voter will be required to present proof of identification before their ballot may be counted.*** *However, even if the voter subsequently verifies their ID, the application will remain marked as NV in the SURE system. For example, a ballot application would be marked as NV if it is a new application and the provided ID was not able to be immediately verified.*

Application Request Date	The date the application was submitted by the voter.
Application Approved Date	The date the application was approved by the county.
Ballot Mailed Date	The date the county queued a ballot label to mail the ballot materials to the voter. This column may be updated when the county indicates they have mailed the ballot. If the county has not mailed the ballot or does not indicate the ballot was mailed, the date in which the ballot label was queued will remain. This data may be updated based on county data entry, not necessarily when the ballot is scanned by USPS. There can be delays between when the ballot is mailed and the entry of the Ballot Mailed Date.
Ballot Returned Date	The date the county marked the ballot as received.
State House District	The voter's state house district.
State Senate District	The voter's state senate district.
Congressional District	The voter's congressional district.
# Latitude	A generic Latitude point within the county.
# Longitude	A generic Longitude point within the county.
Perm Indicator	Indicates whether voter has requested to be on the permanent absentee list or permanent mail-in list (True = 1) or not (False = 0).
Ballot Application Disposition	Disposition of the ballot application – Approved/Declined
Ballot Application Disposition Reason	Reason for the Disposition of the Ballot Application. Counties can only choose one reason, even if there are multiple reasons to decline an application. The specific rejection reason a county chooses to enter first depends on county practices. • APPR – ID NOT VERIFIED • APPR – ID VERIFICATION PEND • APPR – ID VERIFIED • APPR - VALID APPLICATION-MAIL • DECL - AFTER DEADLINE • DECL - DECEASED • DECL – DUPLICATE APPLICATION • DECL - INCOMPLETE ADDRESS • DECL - INVALID REASON • DECL - NO SIGNATURE • DECL - NOT REGISTERED • DECL - SIGNATURE MISMATCH • DECL - UNAFFILIATED (NO NON-PARTISAN ISSUES) **This only occurs in primaries where ballots do not include non-partisan ballot questions*.
ID Verification Status	This field indicates whether the voter's ID on their mail ballot application has been verified (True) or not (False).
Ballot Status	Status assigned to the ballot. • Pending: The county needs additional information or action by the voter before the ballot can be counted. This may be because the voter has not yet returned it, or the voter needs to correct a defect with their ballot return package to ensure it meets statutory requirements before their ballot can be counted (such as the requirement that a voter sign the envelope).

	• Vote Recorded: Ballot received by county election authority and marked as recorded in SURE. • Cancelled: The county election authority cancelled the ballot. This may be because the voter has received a replacement ballot, or because their ballot could not be counted.
Ballot Status Reason	Reason for the "Ballot Status" discussed above. Counties can only choose one reason, even if there could be multiple reasons for the same disposition. The specific cancellation reason a county chooses to enter first depends on county practices. • CANC – EMAIL BALLOT UND BATCH: Ballot cannot be delivered to the email address provided by the voter. Ballots are emailed to qualifying overseas voters pursuant to federal law. • CANC – EMAIL BALLOT UNDELIVERABLE: The email sent with the ballot was returned as undeliverable. • CANC – LABEL CANCELLED: County elections office cancelled the ballot due to an administrative error on the label. • CANC – NO ID: Voter has not provided valid ID to verify their ballot. • CANC – NO SECRECY ENVELOPE: Voter did not include the secrecy envelope. • CANC – NO SIGNATURE: Voter did not sign the return envelope. • CANC – REPLACED: Voter was issued a replacement ballot. • CANC – RETURNED AFTER DEADLINE: Ballot was received after deadline. • CANC – UNDELIVERABLE: Ballot was returned to the county because it could not be delivered by USPS. • CANC – VOTE CANCELLED: The original code counties used to cancel a ballot before the above – more detailed codes – were created. Now exists as a generic category that a county may use to cancel a ballot if the reason does not fit in the more specific categories. • CANC – VOTE CHALLENGED: Ballot was not counted because of a successful challenge. • PEND – NOT YET RETURNED: Ballot that has been issued by the county to the voter but has not been returned by the voter to the county. • PEND – NO ID: A code used to alert a voter that their ballot will be cancelled if they do not verify their identification within six days of election day. • PEND – NO SIGNATURE: A code that counties may use to alert a voter that their ballot return envelope is not signed, and the county offers a way for the voter to address that. The code is changed to 'CANC - NO SIGNATURE' if the error is not corrected. • PEND – NO SECRECY ENVELOPE: A code that counties may use to alert a voter that their ballot is missing a secrecy envelope, and the county offers a way for the voter to address that. The code is changed to 'CANC - NO SECRECY ENVELOPE' if the error is not corrected. • RECORD – BALLOT RETURNED: Ballot has been returned to the county and the county has not logged an observed error that would prevent, or after the canvass, did prevent, the ballot from being counted.

Georeferenced Latitude & Longitude	A georeferenced latitude and longitude based on the mailing address that can be used to create maps.
FileID	A random ID generated to allow data users to identify rows related to the same voter. The random ID is unique to this specific data-file and cannot be used to match to SURE or any other public data files.

2.4 Presidential Turnout

Data source: This data is pulled from election results and the U.S. Census Bureau. The table is updated when the U.S. Census Bureau provides the voting age population (VAP) for the election year, which is often March of the year following the Presidential election.

DOS provides a table reporting turnout for Presidential elections going back to 1960. For every Presidential election, this table provides the vote total in the Presidential race, the certified voter registration, and an estimate of the voting age population from the U.S. Census Bureau. Three percentages are then calculated: the percent of registered voters who voted (Calculation: vote total/number of registered voters*100), the percent of the voting age population that voted (Calculation: vote total/estimate of voting age population*100), and the percent of the voting age population that was registered (Calculation: number of registered voters/estimate of voting age population*100).

This table can be found at the bottom of the page linked here: https://www.pa.gov/en/agencies/dos/resources/voting-and-elections-resources/voting-and-election-statistics.html

A screenshot of a portion of the table is provided below:

Voter Turnout - Presidential Elections

Year	Highest Vote	Office Used	Voter Registration	Percent of Registered Voters Voting	Voting Age Population (VAP*)	Percent of VAP Voting	Percent of VAP Actually Registered
2020	6,915,283	Pres.	9,090,962	76.07%	10,353,548	66.79%	87.81%
2016	6,115,402	Pres.	8,722,977	70.11%	10,018,510	61.04%	87.07%
2012	5,753,670	Pres.	8,508,015	67.63%	9,910,224	58.06%	85.85%
2008	6,010,519	Pres.	8,755,588	68.65%	9,686,275	62.05%	90.39%
2004	5,769,590	Pres.	8,366,663	68.96%	9,615,172	60.01%	87.02%
2000	4,912,185	Pres.	7,781,997	63.00%	9,358,833	52.40%	83.00%
1996	4,501,307	Pres.	6,805,612	66.10%	9,196,000	48.90%	74.01%
1992	4,959,810	Pres.	5,993,002	82.80%	9,129,000	54.30%	65.65%
1988	4,536,251	Pres.	5,875,943	77.20%	9,060,000	50.10%	64.86%
1984	4,844,903	Pres.	6,193,702	78.20%	8,975,000	54.00%	69.01%

2.6 Certified Provisional Ballot Summary

Data source: This analysis uses data pulled from SURE.

Provisional ballot data is certified by each county once all their provisional ballots have been adjudicated. The certified data includes counts of 1) whether a provisional ballot was counted, partially counted, or rejected, 2) the reason why the provisional ballot was cast, and (if applicable) 3) the reason why a provisional ballot was rejected or only partially counted. Due to the number of tables produced, the data has been separated into two data files: 1) 'Certified Provisional Ballots' and 2) 'Certified Provisional Ballots – Demographic Supplemental'. The latter file includes county-level breakdowns for age and party demographic breakdowns. The contents of each file are described further in this section.

Please keep in mind the following:

1. Age data is not available for all voters due to security concerns related to protected voters, the provisional voter not providing their birthdate, or the voters' record being older than the SURE system. Please see section 1.3b for more information.

2. Counties can enter more than one reason for why a voter cast a provisional ballot and/or why a provisional ballot was rejected. This may cause the total count in the 'Certified rejection reason' or the 'Certified provisional reason' tables to be greater than the total number of provisional ballots the county certified in the 'Certified provisional totals' table. This may vary by county.
3. Vote History and Certified Provisional Ballots are entered into different systems. As such, discrepancies may exist between the 'Certified Provisional Ballot Summary' and 'Vote History & Certified Voter Registration Summary' files.
4. Party affiliation is pulled from a voter's record. If a voter record cannot be connected to a provisional ballot (e.g., the voter is not registered), then party affiliation is left blank. Blank party affiliations are included in the 'Independent/Other' category for party breakdowns.

2.6a Certified Provisional Ballots file

File meta-data:

This file includes nine tables of data. See below for each table along with their column breakdowns. The name of each table below corresponds with the sheet name in the excel file, which is in quotes below. There is also a 'Disclaimers' sheet in the excel file, which includes some highlights of this data handbook.

1. "Certified provisional totals"

Column names	Column descriptions
County	The name of each county in alphabetical order. There is also a 'TOTAL' row at the bottom of the table, leading to 68 rows.
Counted	The number of provisional ballots the county certified to the state that they counted in full.
Partially Counted	The number of provisional ballots the county certified to the state that they partially counted (i.e., some contests but not all contests on the ballot were counted). This generally occurs when the voter cast the provisional ballot at the incorrect election district in the correct county.
Rejected	The number of provisional ballots the county certified to the state that they rejected.
Total	The total number of certified provisional ballots cast, summing the 'Counted', 'Partially Counted,' and 'Rejected' columns.
% Counted	The percent of ballots counted out of the total number of certified provisional ballots cast. Formula: 'Counted'/'Total'*100
% Partially Counted	The percent of ballots partially counted out of the total number of certified provisional ballots cast. Formula: 'Partially Counted'/'Total'*100
% Rejected	The percent of ballots rejected out of the total number of certified provisional ballots cast. Formula: 'Rejected'/'Total'*100

2. "Totals by party"

Summary table of certified provisional ballots grouped by political party at the statewide level. This includes the same columns as the 'Certified provisional totals' table (above), except it drops 'County' and groups by 'Party.' The 'Party' column is described below.

Party: Using the party column from the voter registration table in SURE, voters were organized into the following parties: Democratic, Republican, and Independent/Other. If a provisional voter was not registered in SURE, their party affiliation was returned as 'NULL' and they were grouped into the 'Independent/Other' group.

3. "Totals by age"

Summary table of certified provisional ballots grouped by age group at the statewide level. This includes the same columns as the 'Certified provisional totals' table (above), except it drops 'County' and groups by 'Age group.' The 'Age group' column is described below:

Age group: Using the 'DateofBirth' provided on the provisional ballot, voters were organized into seven age groups – 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, 65 to 74, and 75+. Ballot counts are provided by each age group statewide.

5. "Certified provisional reason"

Column names	**Column descriptions**
County	The name of each county in alphabetical order. There is also a 'TOTAL' row at the bottom of the table, leading to 68 rows.
Court order or extension	Number of provisional ballots cast due to either a court order for an individual voter or an order to extend the hours of a polling place. SURE has two different codes for court orders and extensions that were combined into one for this analysis.
Mail-in or absentee not returned	Number of provisional ballots cast because the voter requested a mail-in or absentee ballot and did not surrender the mail-in or absentee ballot at the polling place on election day. SURE has two codes, one for mail-in and the other for absentee ballots, that were combined for this analysis.
Name not on list	Number of provisional ballots cast because the voter's name did not appear on the list of registered voters at the polling place.
ID not supplied	Number of provisional ballots cast because the voter did not bring their required identification to the polling place on election day. When a voter votes at a new polling place for the first time (e.g., after registering for the first time in a county or moving to a new district within the county), they are required to provide identification.
Eligibility Challenged	Number of provisional ballots cast because the voter's identity or residence was challenged.
Sum of Provisional Reasons	The total number of reasons provided for why a provisional ballot was cast. Counties can enter more than one reason for a provisional ballot cast, meaning the total number of reasons a provisional ballot was cast may be more than the total number of certified provisional ballots.
% Court order or extension	The percent of ballots cast for a court order or extension out of the sum of provisional reasons. Formula: 'Court order or extension'/'Sum of Provisional Reasons'*100

% Mail-in or absentee not returned	The percent of ballots cast because a mail-in or absentee ballot was not returned out of the sum of provisional reasons. Formula: 'Court order or extension'/'Sum of Provisional Reasons'*100
% Name not on list	The percent of ballots cast because a voter's name did not appear on the list at the polling place out of the sum of provisional reasons. Formula: 'Court order or extension'/'Sum of Provisional Reasons'*100
% ID not supplied	The percent of ballots cast because voter did not supply ID out of the sum of provisional reasons. Formula: 'Court order or extension'/'Sum of Provisional Reasons'*100
% Eligibility Challenged	The percent of ballots cast because the eligibility of the voter was challenged out of the sum of provisional reasons. Formula: 'Court order or extension'/'Sum of Provisional Reasons'*100

5. "Provisional reason by party"

Summary table that includes the count of provisional ballots by reason a provisional ballot was cast and political party at the statewide level. This includes the same columns as the 'Certified provisional reason' table (above), except it drops 'County' and groups by 'Party.' The 'Party' column is described below.

Party: Using the party column from the voter registration table in SURE, voters were organized into the following parties: Democratic, Republican, and Independent/Other. If a provisional voter was not registered in SURE, their party affiliation was returned as 'NULL,' and they were grouped into the 'Independent/Other' group.

6. "Provisional reason by age"

Summary table that includes the count of provisional ballots by reason a provisional ballot was cast and age group at the statewide level. This includes the same columns as the 'Certified provisional reason' table (above), except it drops 'County' and groups by 'Age group.' The 'Age group' column is described below:

Age group: Using the date of birth provided on the provisional ballot, voters were organized into seven age groups – 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, 65 to 74, and 75+. Ballot counts are provided by each age group statewide.

7. "Certified rejection reason"

Column names	**Column descriptions**
County	The name of each county in alphabetical order. There is also a 'TOTAL' row at the bottom of the table, leading to 68 rows.
Duplicate Provisional Ballot Cast	Provisional ballot was rejected because voter cast more than one provisional ballot.
Incomplete Provisional Ballot Affidavit Envelope	Provisional ballot was rejected because voter did not complete all required components of the provisional ballot affidavit envelope and casting process.
Not Registered	Provisional ballot was rejected because voter was not registered in the Commonwealth.

Registered in a different county	Provisional ballot was rejected because voter was not registered in the county where they attempted to vote, but voter is registered in a different county in the Commonwealth.
Registered in different election district	Provisional ballot was rejected because voter was not registered to vote in the election district where they attempted to vote, but voter is registered in a different election district within the county. This code may be used when the provisional ballot is partially counted.
Signatures are not genuine	The county observed a disqualifying issue with the signature provided on the provisional ballot.
Voted by conventional alternative or absentee/mail-in	Provisional ballot was rejected because voter voted by mail-in or absentee ballot.
Sum of Rejection Reasons	The total number of reasons provided for why a provisional ballot was rejected. Counties can enter more than one reason for a provisional ballot being rejected, meaning the total number of reasons a provisional ballot was rejected may be more than the total number of certified provisional ballots that were rejected.
Total Provisional Ballots Cast	The 'Total' column from the 'Certified provisional totals' sheet, used for the following percent calculations.
% Duplicate Provisional Ballot Cast	The percent of total provisional ballots cast that were rejected because a duplicate provisional ballot was cast. Formula: 'Duplicate Provisional Ballot Cast'/'Total Provisional Ballots Cast'*100
% Incomplete Provisional Ballot Affidavit Envelope	The percent of total provisional ballots cast that were rejected because the voter did not complete all required components of the provisional ballot affidavit envelope. Formula: 'Incomplete Provisional Ballot Affidavit Envelope'/'Total Provisional Ballots Cast'*100
% Not Registered	The percent of total provisional ballots cast that were rejected because the voter was not registered. Formula: 'Not Registered'/'Total Provisional Ballots Cast'*100
% Registered in a different county	The percent of total provisional ballots cast that were rejected because the voter was registered in a different county. Formula: 'Registered in a different county'/'Total Provisional Ballots Cast'*100
% Registered in different election district	The percent of total provisional ballots cast that were rejected because the voter was registered in a different election district. Formula: 'Registered in different election district'/'Total Provisional Ballots Cast'*100
% Signatures are not genuine	The percent of total provisional ballots cast that were rejected because the signatures on the provisional ballot affidavit envelope were not genuine. Formula: 'Signatures are not genuine'/'Total Provisional Ballots Cast'*100
% Voted by conventional alternative or absentee/mail-in	The percent of total provisional ballots cast that were rejected because the voter voted at the polls or through an absentee or mail-in ballot.

	Formula: 'Voted by conventional alternative or absentee/mail-in'/'Total Provisional Ballots Cast'*100

8. "Rejection reason by party"

Summary table that includes the count of reasons a provisional ballot was rejected by political party at the statewide level. This includes the same columns as the 'Certified rejection reason' table (above), except it drops 'County' and groups by 'Party.' The 'Party' column is described below.

Party: Using the party column from the voter registration table in SURE, voters were organized into the following parties: Democratic, Republican, and Independent/Other. If a provisional voter was not registered in SURE, their party affiliation was returned as 'NULL,' and they were grouped into the 'Independent/Other' group.

9. "Rejection reason by age"

Summary table that includes the count of reasons a provisional ballot was rejected by age group at the statewide level. This includes the same columns as the 'Certified rejection reason' table (above), except it drops 'County' and groups by 'Age group.' The 'Age group' column is described below:

Age group: Using the date of birth provided on the provisional ballot, voters were organized into seven age groups – 18 to 24, 25 to 34, 35 to 44, 45 to 54, 55 to 64, 65 to 74, and 75+. Ballot counts are provided by each age group statewide.

2.6b Certified Provisional Ballots – Demographic Supplemental file

File meta-data:

This file includes six tables of data that breakdown the statewide age and party summary tables in the 'Certified Provisional Ballots' file (section 2.6a) by county. See below for each table along with their column breakdowns. The name of each table below corresponds with the sheet name in the excel file, which is in quotes. There is also a 'Disclaimers' sheet in the excel file, which includes some notes that are also included in this data handbook.

1. "Totals by party & county"

This table breaks down the statewide summary data in the 'Totals by party' sheet in the 'Certified Provisional Ballots' file by county. It includes the same columns. Please see the 'Totals by party' section in 2.6a for more information.

2. "Totals by age & county"

This table breaks down the statewide summary data in the 'Totals by age' sheet in the 'Certified Provisional Ballots' file by county. It includes the same columns. Please see the 'Totals by age' section in 2.6a for more information.

3. "Prov reasons by party & county"

This table breaks down the statewide summary data in the 'Provisional reason by party' sheet in the 'Certified Provisional Ballots' file by county. It includes the same columns. Please see the 'Provisional reason by party' section in 2.6a for more information.

4. “Prov reasons by age & county”

This table breaks down the statewide summary data in the ‘Provisional reason by age’ sheet in the ‘Certified Provisional Ballots’ file by county. It includes the same columns. Please see the ‘Provisional reason by age’ section in 2.6a for more information.

5. “Reject by party & county”

This table breaks down the statewide summary data in the ‘Rejection reason by party’ sheet in the ‘Certified Provisional Ballots’ file by county. It includes the same columns. Please see the ‘Rejection reason by party’ section in 2.6a for more information.

6. Reject by age & county”

This table breaks down the statewide summary data in the ‘Rejection reason by age’ sheet in the ‘Certified Provisional Ballots’ file by county. It includes the same columns. Please see the ‘Rejection reason by age’ section in 2.6a for more information.

3. Election Results & Returns

This section provides an overview of the data DOS provides related to election results. It is broken into the following sections:

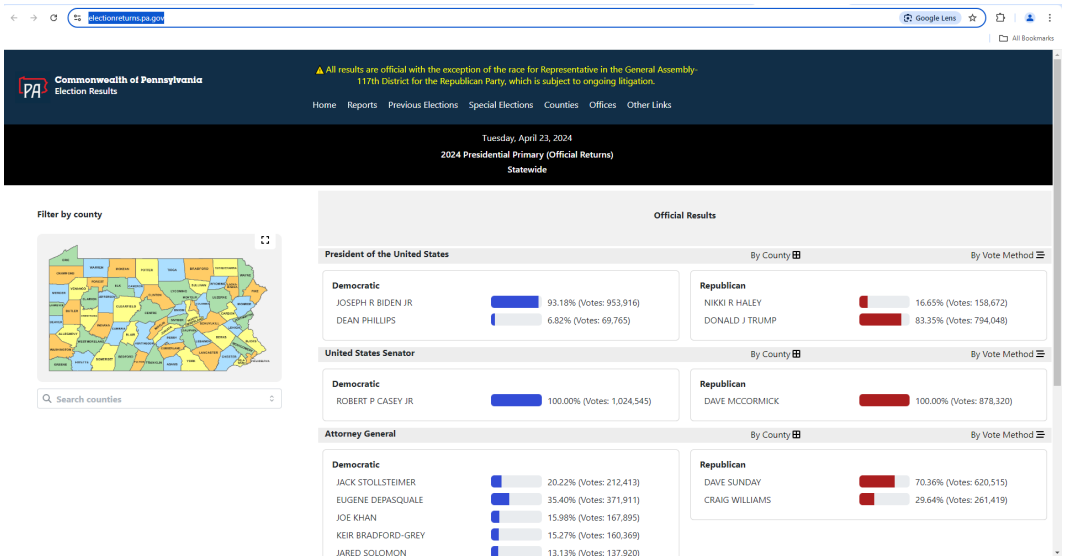
3.1 Election Night Results
3.2 Supplemental Election Night Returns Dashboard
3.3 Bulk Election Data

3.1 Election Night Results

Data source: This data is provided by the counties through text files of ballot counts submitted through our election night reporting system.

Under Pennsylvania law, counties cannot begin to record and publish results until after 8 p.m. on Election Day. Once counties start this process, election results can be found on our website: https://www.electionreturns.pa.gov/. The website is continually updated as counties provide updated returns. The election results website also includes historical election results back through 2000.

See below for a screenshot of the election results website for the Commonwealth of Pennsylvania:



As seen in the screenshot above, the election results website includes the total votes per candidate in each race and can be viewed at the county-level. Vote totals can also be broken down by vote method, which includes election day votes, mail ballot votes, and provisional ballot votes (for more information on provisional ballots, see https://www.pa.gov/en/agencies/vote/voter-support/provisional-ballot.html).

Results are not official until after they are certified, which occurs approximately three weeks after the election. The number of ballots cast can change in the days following the election, as provisional ballots are adjudicated and UOCAVA ballots (i.e., overseas military that have an extended deadline for arriving

at the county election office for counting) are received. The banner on the top of the website (the navy banner with yellow text in the screenshot above) will state whether results are official or still unofficial.

3.2 Supplemental Election Night Results Dashboard

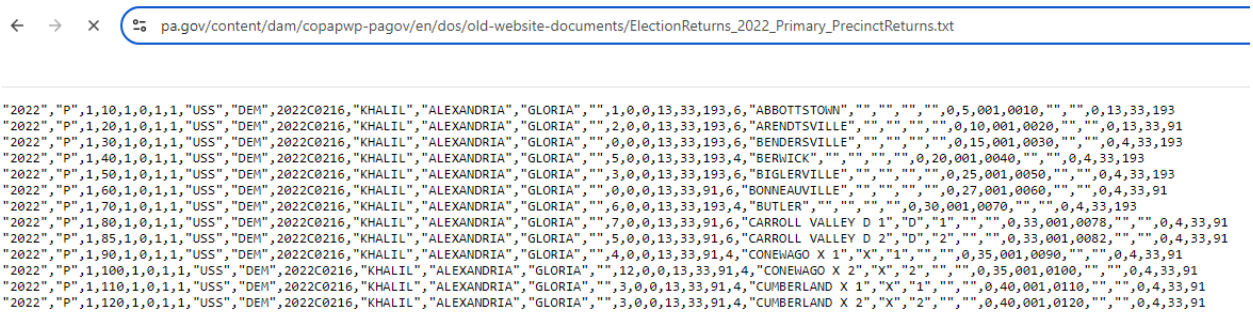
Data source: This dashboard uses both data in SURE (e.g., how many mail ballots have been received by the county according to data entry in SURE) and ballot cast data provided by each county on and after election day (e.g., mail ballots counted). Starting on election night, the supplemental dashboard shows the estimated percentage of mail ballots still uncounted for each county, calculated using the difference between the number of mail ballots recorded in SURE as "Recorded" and the number of mail ballots total reported. This dashboard is updated once daily through election returns.

Link to dashboard here, when available (after an election):
https://www.pa.gov/en/agencies/vote/elections/elections-data.html

3.3 Bulk Election Data

Data source: This data is provided by the county as part of election results certification. It is not from SURE.

For each election, DOS provides access to bulk election returns and voter registration. These files are broken down by county and precinct in both files, and race for the election returns file. Files are provided in a comma-delimited text file, and a screenshot of the first few lines of the election returns file for the 2022 Primary can be seen below:



The comma-delimited text file does not include column headers. Headers – along with their explanations – can be obtained from the election's corresponding ReadMe file. Some fields will change depending on the election (e.g., Municipal versus even-year General), so be careful to use the ReadMe file that corresponds to the bulk election data you are using.

Data for the election returns and voter registration bulk files are pulled from different data sources. The election returns data is pulled at the precinct-level, while the voter registration data includes precinct-splits for congressional and school districts (where applicable). There are circumstances where all votes are associated with one precinct but registered voters are split between two precinct-splits. For accurate turnout analysis, the registered voters will need to be combined.

Election returns and ReadMe files can be found at the link below. Please use the sort and filter options to identify the election you are interested in.

Link: https://www.pa.gov/en/agencies/dos/resources/voting-and-elections-resources/voting-and-election-statistics/bulk-election-data.html#sortCriteria=%40copapwpdate%20descending

4. Election data provided in lead-up to the election

In this section, we will discuss election related data provided on the DOS website that is available prior to the election. The sections covered include:

4.1 Full Voter Export
4.2 Statewide Mail Ballot File
4.3 Daily Mail Ballot Report

4.1 Full Voter Export

Data source: This data is pulled from SURE weekly.

The Full Voter Export is a voter-level representation of registered voters in every county. It includes voters with either "Active" or "Inactive" status. Inactive voters can still vote in an election if they provide proof of identification and residence on election day. Cancelled voters – voters removed from the voter rolls due to moving out of county, death, etc. - are not included in the Full Voter Export. Protected voters are not included in the Full Voter Export for security concerns.

The voter record includes vote history in the last 40 elections per county. As counties can have special elections, the elections provided differ by county. Even if a cancelled voter voted in one of the 40 elections included in the Full Voter Export before their record was cancelled, they will not be included in the latest Full Voter Export. The purpose of the Full Voter Export is to provide a representation of the *current* voter record, not the past. As such, counts of vote history represented in the Full Voter Export will change as voter registrations are cancelled or moved between counties. The Full Voter Export cannot be used to replicate election results.

The Full Voter Export is available upon request and with a $20 fee. Link here: https://www.pavoterservices.pa.gov/Pages/PurchasePAFullVoterExport.aspx?Langcode=en-US

Requestors are provided with the following files as part of the Full Voter Export:

- A zip file including a voter record export for each county.
- A zip file including an election mapping file for each county that maps the election names and dates to the Election 1 – Election 40 columns in the Full Voter Export (as the elections included in the Full Voter Export differ by county). There are two sets of columns related to election mapping: Election 1 Vote Method – Election 40 Vote Method (i.e., whether they voted at polls, mail-ballot, absentee, or provisional for Election 1 - 40) and Election 1 Party – Election 40 Party (i.e., the party they were registered as when/if they voted in Election 1 – 40).
- A zip file including a file per county that defines the district names, precincts, legislative districts, etc. that can be mapped to the Full Voter Export.
- A text file that provides the codes (i.e., abbreviations) for every political party included in the Full Voter Export as well as a description for each party.
- A ReadMe document for the Full Voter Export that includes column labels for the Full Voter Export, election mapping file, district zone mapping file, and political party codes.

The below table was copied from the ReadMe file and includes all variables that are included in the Full Voter Export. It is provided as a reference of the type of data that could be obtained by requesting the Full Voter Export.

Field Number	Field Description	Field Type	Description
1	ID Number	String	SURE Voter ID number. This includes a 9-digit identifier before a hyphen and two-digit county identifier following the hyphen (e.g., 012345678-01). The 9-digit identifier follows the voter from county to county, with the two-digit county identifier changing when voters move between counties.
2	Title	String	If provided by voter: MS, MR, MRS, DR, etc.
3	Last Name	String	Last name provided by voter
4	First Name	String	First name provided by voter
5	Middle Name	String	Middle name or initial provided by voter
6	Suffix	String	Suffix if provided by voter
7	Gender	String	F=Female, M=Male, U=Unknown/Other
8	DOB	Date	Date of Birth format: MM/DD/YYYY
9	Registration Date	Date	Date recorded in our system for when voter registered, in MM/DD/YYYY format. In older records, this may be filled with a placeholder date, such as '01-01-1900'
10	Voter Status	String	A=Active, I=Inactive; Both Active and Inactive voters can vote in an election. Inactive voters must show identification at polling place on election day. Cancelled voter records are not included in the Full Voter Export.
11	Status Change Date	Date	The most recent date in which voter registration or voter status was updated, in MM/DD/YYYY format.
12	Party Code	String	Registered political party provided by voters. See Political Party Code documentation for more information.
13	House Number	String	Residential/Registration Address
14	House Number Suffix	String	Residential/Registration Address
15	Street Name	String	Residential/Registration Address
16	Apartment Number	String	Residential/Registration Address

17	Address Line 2	String	Residential/Registration Address
18	City	String	Residential/Registration Address
19	State	String	Residential/Registration Address
20	Zip	String	Residential/Registration Address
21	Mail Address 1	String	Mailing Address
22	Mail Address 2	String	Mailing Address
23	Mail City	String	Mailing Address
24	Mail State	String	Mailing Address
25	Mail Zip	String	Mailing Address
26	Last Vote Date	Date	MM/DD/YYYY; The date of the last election for which the voter was assigned vote history. If voted by mail ballot, corresponds to the date of the election even if ballot was returned to the county earlier than election day.
27	Precinct Code	String	Code for the precinct where the voter votes. Use the zone mapping file to identify the names that correspond with each precinct.
28	Precinct Split ID	String	Use the zone mapping file to identify the names that correspond with each precinct split.
29	Date Last Changed	Date	The most recent date when an aspect of the voter record was updated. Though this can include voter registration updates, it also reflects mail ballot applications, submissions, etc. Provided in MM/DD/YYYY format
30	Custom Data 1	String	Legacy system ID number
31-70	Districts 1-40	String	Each field corresponds to a district from 1 to 40. *See* Zone Code and Zone Type documentation.
71	Election 1 Vote Method	String	*See* Election Map documentation for each county to identify the election that corresponds with Election 1 through Election 40. Uses abbreviations for method of voting in said election: AP=At Polls AB=Absentee MB = Mail-In Ballot P=Provisional
72	Election 1 Party	String	Voter's political party at the time of the election. *See* Political Party Code

			documentation for political party and Election Map documentation for each county to identify what election corresponds to Election 1 to Election 40.
73-150	Elections 2-40 Vote Method and Party	String	For fields 73-150, the odd numbered field will show the Vote Method for the election, and the immediately subsequent (even) field will show the party. This pattern repeats for elections 2-40.
151	Home Phone	String	Phone number provided by voter. This is not required to register and may be null.
152	County	String	This identifies the county of registration.
153	Mail Country	String	Country where voter's mail is sent.
154	Registration Method	String	Registration method will be as follows: D for DOTIA applications. This only includes initial registrations, not changes to registration; P for Paper applications; and E for Electronic applications. This includes OVR and Web API submissions.

4.2 Statewide Mail Ballot File

Data source: This data is pulled from SURE.

This is a point-in-time data file that details the processing of absentee and mail-in ballots by county election offices for each election. Each row in this file represents one mail ballot transaction, meaning there could be more than one row related to a single voter if they had ballots cancelled. This data is extracted once per day from SURE leading up to the election through certification. This file reflects mail ballot activity recorded by the county in SURE at the time it is extracted – data entered after that point will not be reflected until the following day. Due to the point-in-time nature of this file, data is published for informational purposes only.

Please note that data entry in SURE is not final on election day. Administrative practices for recording transactions in the SURE system vary by county, with some counties recording individual transactions as they occur and other counties recording batches of transactions at one time. Batch transactions may result in substantial changes to a county's data from one day to the next.

The Statewide Mail Ballot data file contains several fields in pipe delimited format. The data includes all mail ballot applications processed by counties, which includes voters on the permanent mail ballot list and federal voters. However, it does not include protected voters.

To receive the Statewide Mail Ballot data file, you must complete the Statewide Mail Ballot File Request by using this online webform or by downloading a form and emailing it to the Division of Election Security and Technology at RA-STBEST@PA.GOV. A publicly available version of this file with deidentified data is available on Open Data for every election. Once the data entry for the election is complete, the Open Data file will ultimately become the 'Open Data Mail Ballot Requests' file described earlier in this handbook.

Many of the data caveats described in Section 2.3 also apply to this file. Columns included in the Statewide Mail Ballot File are described in the table below:

Column names	Column descriptions
CountyID	4-digit identifier assigned to each county, sequentially assigned starting with 2290.
CountyName	This identifies the county of registration.
IDNumber	A unique identification number assigned to each voter in the SURE system. This number is unconnected to any personally identifiable information of the voter (i.e., this does not reflect the voter's social security number or driver's license number).
Party	The voter's party as indicated on their application. The following are possible values of this field: • D: Democratic • R: Republican • LN: Libertarian • GR: Green All other codes combined can be used to represent 'Other'/'Independent' voters.
FirstName	Voter's first name
MiddleName	Voter's middle name
LastName	Voter's last name
Suffix	Suffix to voter's name
Addr1	This is the first line of the address where the voter's mail ballot is to be sent. This may be, but is not necessarily, the address of the voter's registration.
Addr2	This is the second line of the address where the voter's mail ballot is to be sent. This may be, but is not necessarily, the address of the voter's registration.
CSZ	This is the mailing city, state and zip code of the address where the voter's mail ballot is to be sent.
Country	This is the Country of the address where the voter's mail ballot is to be sent.
DateofBirth	Voter's Month and Year of Birth in MM/YYYY format
Mail Application Type	The mail ballot application type. The following are the possible values for this field.

	• Alt - Alternative ballot (voters who are disabled or at least 65 years of age where the polling place may not be fully accessible) • BV - Bedridden Veteran • BVRI - Bedridden Veteran - Remote/Isolated • C - Emergency Absentee ballot application • CIV - Civilian absentee submitted via paper • CRI - Overseas voter - Remote/Isolated • CVO - Overseas voter • F - Federal voter (voter is eligible to vote in federal races only) • M - Military • MAILIN - Mail-in submitted via paper • MRI - Military - Remote/Isolated • OLMAILV - Mail-in submitted online where the applicant's ID was verified at the time of submission • OLMAILNV - Mail-In submitted online where ID verification did not occur or was not successful at the time of submission • OLREGV - Absentee submitted online where the applicant's ID was verified at the time of submission • OLREGNV - Absentee submitted online where ID verification did not occur or was not successful at the time of submission • PER - Absentee where applicant has requested permanent status • PMI - Mail-in where applicant has requested permanent status • REG - Absentee submitted via paper • V - Veteran *Note: An application is designated as NV (not verified) if the county was unable to verify the application via the ID verification process prior to approving the application.* ***The voter will be required to present proof of identification before their ballot may be counted.*** *However, even if the voter subsequently verifies their ID, the application will remain marked as NV in the SURE system. For example, a ballot application would be marked as NV if it is a new application and the provided ID was not able to be immediately verified.*
PrecinctCode	Identifier assigned to each precinct. This field shows the precinct of a voter's registration.
PrecinctDesc	Description/Name for the precinct of the voter's registration.
Application Request Date	The date the application was submitted by the voter.
Application Approved Date	The date the application was approved by the county.
Ballot Mailed Date	The date the county queued a ballot label to mail the ballot materials to the voter. This column may be updated when the county indicates they have mailed the ballot. If the county has not mailed the ballot or does not indicate the ballot was mailed, the date in which the ballot label was queued will remain. This data may be updated based on county data entry, not necessarily when the ballot is scanned by USPS. There can be delays between when the ballot is mailed and the entry of the Ballot Mailed Date.

Ballot Returned Date	The date the county marked the ballot as received.
State House District	The voter's state house district.
State Senate District	The voter's state senate district.
Congressional District	The voter's congressional district.
Perm Indicator	Indicates whether voter has requested to be on the permanent absentee list or permanent mail-in list (True) or not (False).
Ballot Application Disposition	Disposition of the ballot application – Approved/Declined.
Ballot Application Disposition Reason	Reason for the Disposition of the Ballot Application. Counties can only choose one reason, even if there are multiple reasons to decline an application. The specific rejection reason a county chooses to enter first depends on county practices. • APPR – ID NOT VERIFIED • APPR – ID VERIFICATION PEND • APPR – ID VERIFIED • APPR – VALID APPLICATION-MAIL • DECL – AFTER DEADLINE • DECL – DECEASED • DECL – DUPLICATE APPLICATION • DECL – INCOMPLETE ADDRESS • DECL – INVALID REASON • DECL – NO SIGNATURE • DECL – NOT REGISTERED • DECL – SIGNATURE MISMATCH • DECL – UNAFFILIATED (NO NON-PARTISAN ISSUES) **This only occurs in primaries where ballots do not include non-partisan ballot questions*.
ID Verification Status	This field indicates whether the voter's ID on their mail ballot application has been verified (True) or not (False).
Ballot Status	Status assigned to the ballot • Pending: The county needs additional information or action by the voter before the ballot can be counted. This may be because the voter has not yet returned it, or the voter needs to correct a defect with their ballot return package to ensure it meets statutory requirements before their ballot can be counted (such as the requirement that a voter sign the envelope). • Vote Recorded: Ballot received by county election authority and marked as recorded in SURE. • Cancelled: The county election authority cancelled the ballot. This may be because the voter has received a replacement ballot, or because their ballot could not be counted.
Ballot Status Reason	Reason for the "Ballot Status" discussed above. Counties can only choose one reason, even if there could be multiple reasons for the same disposition. The specific cancellation reason a county chooses to enter first depends on county practices. • CANC – EMAIL BALLOT UND BATCH: Ballot cannot be delivered to the email address provided by the voter. Ballots are emailed to qualifying overseas

	voters pursuant to federal law. • CANC – EMAIL BALLOT UNDELIVERABLE: The email sent with the ballot was returned as undeliverable. • CANC – LABEL CANCELLED: County elections office cancelled the ballot due to an administrative error on the label. • CANC – NO ID: Voter has not provided valid ID to verify their ballot. • CANC – NO SECRECY ENVELOPE: Voter did not include the secrecy envelope. • CANC – NO SIGNATURE: Voter did not sign the return envelope. • CANC – REPLACED: Voter was issued a replacement ballot. • CANC – RETURNED AFTER DEADLINE: Ballot was received after deadline. • CANC – UNDELIVERABLE: Ballot was returned to the county because it could not be delivered by USPS. • CANC – VOTE CANCELLED: The original code counties used to cancel a ballot before the above – more detailed codes – were created. Now exists as a generic category that a county may use to cancel a ballot if the reason does not fit in the more specific categories. • CANC – VOTE CHALLENGED: Ballot was not counted because of a successful challenge. • PEND – NOT YET RETURNED: Ballot that has been issued by the county to the voter but has not been returned by the voter to the county. • PEND – NO ID: A code used to alert a voter that their ballot will be cancelled if they do not verify their identification within six days of election day. • PEND – NO SIGNATURE: A code that counties may use to alert a voter that their ballot return envelope is not signed, and the county offers a way for the voter to address that. The code is changed to 'CANC - NO SIGNATURE' if the error is not corrected. • PEND – NO SECRECY ENVELOPE: A code that counties may use to alert a voter that their ballot is missing a secrecy envelope, and the county offers a way for the voter to address that. The code is changed to 'CANC - NO SECRECY ENVELOPE' if the error is not corrected. • RECORD – BALLOT RETURNED: Ballot has been returned to the county and the county has not logged an observed error that would prevent, or after the canvass, did prevent, the ballot from being counted.

4.3 Daily Mail Ballot Report

Data source: This data is pulled from SURE.

This spreadsheet provides point-in-time transactional data for informational purposes to provide a high-level overview of the processing of mail-in and absentee ballots by county election offices. This data is pulled once per day starting 4-6 weeks before the election from the SURE system, and it reflects activity recorded by the counties in the SURE system as of the time of the pull.

Please note that county election offices continuously process mail ballot applications, record mail ballots, reconcile mail ballot data, and make corrections to mail ballot data when necessary. These

activities occur in the weeks leading up to and including election day. Please note that data entry in SURE will not be final on election day, meaning these numbers do not reflect final counts.

Administrative practices for recording transactions in the SURE system vary by county (i.e., recording individual transactions as they occur versus recording batches of transactions at specific intervals). These varying practices may result in substantial changes to a county's data from one day to the next.

The spreadsheet contains three separate tabs – absentee ballots, mail-in ballots, and total ballots (meaning the total of mail-in and absentee ballots) – with each tab detailing the data by county. The data includes all mail ballot applications processed by counties.

Below is a screenshot of what this file looks like:

Information as of 04/23/2024 8:00 PM								
CountyName	Total Applications Approved	Dem Applications Approved	Rep Applications Approved	Oth Applications Approved	Total Ballots Returned	Dem Ballots Returned	Rep Ballots Returned	Oth Ballots Returned
ADAMS	7398	4297	3100	1	5737	3366	2371	0
ALLEGHENY	122186	99729	22455	2	95151	79084	16067	0
ARMSTRONG	3020	1881	1138	1	2439	1559	880	0
BEAVER	10418	7868	2550	0	8436	6574	1862	0
BEDFORD	2348	1100	1248	0	1858	900	958	0
BERKS	20671	14296	6374	1	16271	11504	4767	0
BLAIR	6841	3581	3260	0	5565	2892	2673	0
BRADFORD	2669	1394	1271	4	2146	1102	1044	0
BUCKS	63512	44871	18635	6	45419	31835	13584	0
BUTLER	13849	8773	5061	15	10851	6976	3873	2
CAMBRIA	7207	4980	2227	0	5874	4083	1791	0
CAMERON	300	124	176	0	256	107	149	0
CARBON	3402	2208	1194	0	2613	1729	884	0
CENTRE	10751	7620	3131	0	8453	6013	2440	0
CHESTER	43461	31439	12021	1	32315	23777	8538	0
CLARION	1682	942	740	0	1404	789	615	0
CLEARFIELD	3302	2015	1286	1	2556	1519	1037	0
CLINTON	1726	1071	655	0	1452	911	541	0
COLUMBIA	3004	1968	1036	0	2458	1604	854	0
CRAWFORD	4442	2782	1658	2	3500	2196	1304	0
CUMBERLAND	18753	12098	6655	0	15257	9958	5299	0
DAUPHIN	18112	13006	5103	3	14329	10408	3921	0
DELAWARE	39740	29224	10508	8	28476	21421	7055	0
ELK	1649	1046	603	0	1330	855	475	0
ERIE	19881	15221	4652	8	15600	12210	3390	0
FAYETTE	7538	5718	1820	0	5939	4625	1314	0
FOREST	275	154	121	0	239	125	114	0
FRANKLIN	7137	4100	3037	0	5774	3346	2428	0
FULTON	425	177	247	1	317	130	187	0
GREENE	1881	1369	511	1	1567	1148	419	0
HUNTINGDON	1856	995	859	2	1461	789	672	0
INDIANA	4525	2950	1575	0	3753	2486	1267	0
JEFFERSON	1832	950	879	3	1464	768	696	0
JUNIATA	825	437	388	0	677	362	315	0
LACKAWANNA	15495	12548	2947	0	12518	10342	2176	0
LANCASTER	26629	16601	10019	9	20290	12670	7620	0
LAWRENCE	5057	3428	1629	0	4157	2830	1327	0
LEBANON	8010	4634	3365	11	6310	3676	2633	1
LEHIGH	26523	19098	7189	236	20614	15057	5420	137
LUZERNE	25334	16851	6979	1504	18931	13163	5134	634

Total | Absentee | Mail-In

Link to the report: https://www.pa.gov/en/agencies/vote/elections/elections-data.html