

# VEST PA 2018

---

State: Pennsylvania

Organization: VEST

Date Updated: Report Date: 04/27/2021, File Date: 11/26/2020

## 1. Is all raw data available?

No

- Accessible files:
  - VEST PA 18 final file
    - Accessed: 04/02/21
    - Online: [Harvard Dataverse Link](#)
    - Source: VEST
  - VEST PA 18 documentation file
    - Accessed: 04/02/21
    - Online: [Harvard Dataverse Link](#)
    - Source: VEST
  - File: PA Precinct-Level Election Results
    - Accessed: 04/02/21
    - Online: [Open Elections Github Link](#)
    - Source: Open Elections
  - File: U.S. Census Bureau's 2020 Redistricting Data Program Phase 2 release
    - Accessed: 04/02/21
    - Online: [PA Page](#)
    - Source: US Census Bureau
  - File: Allegheny County Shapefile
    - Accessed: 04/13/21
    - Online: [Allegheny County GIS](#)
    - Source: Allegheny County GIS Office
  - File: Cambria Canvass Report
    - Accessed: 04/27/21

- Online: [Cambria County](#)
  - File: Northumberland Canvass Report
    - Accessed: 04/27/21
    - Online: [County of Northumberland](#)
  - File: Crawford Canvass Report
    - Accessed: 04/27/21
    - Online: [Crawford County](#)
  - File: Elk Canvass Report
    - Accessed: 04/27/21
    - Online: [Elk County](#)
  - File: Forest Canvass Report
    - Accessed: 04/27/21
    - Online: [Forest County](#)
  - File: Lawrence Canvass Report
    - Accessed: 04/27/21
    - Online: [Lawrence County](#)
  - File: Lycoming Canvass Report
    - Accessed: 04/27/21
    - Online: [Lycoming County](#)
  - File: Montgomery Canvass Report
    - Accessed: 04/27/21
    - Online: [Montgomery County](#)
  - File: Susquehanna Canvass Report
    - Accessed: 04/27/21
    - Online: [Susquehanna County](#)
  - File: Juniata Canvass Report
    - Accessed: 04/27/21
    - Online: [Juniata County](#)
  - File: Montour Canvass Report
    - Accessed: 04/27/21
    - Online: [Montour County](#)
  - File: Northumberland Canvass Report
    - Accessed: 04/27/21
    - Online: [Northumberland County](#)
- Inaccessible files:
  - Delaware County Shapefile
    - Note, sent an email to a contact at the Delaware County GIS office, but I have not yet received a reply.
  - PA Secretary of State 2018 election cycle voter file
    - Note, there are voter files available from the Pennsylvania Secretary of State's office, but they are not free and they do not appear to maintain past voter files.
- Note: Many of the precinct shapefile modifications required municipality-specific files that we did not attempt to locate.

## 2. Processing steps available?

Yes

- Description of processing steps:
  - VEST's processing steps were accessed on April 2nd, 2021. In VEST's words, steps included:
    - Precinct data was corrected with canvass reports for the following counties: Berks, Blair, Bradford, Cambria, Carbon, Crawford, Elk, Forest, Franklin, Lawrence, Lycoming, Mifflin, Montgomery, Montour, Northumberland, Susquehanna. The candidate totals for Berks, Blair, Crawford, and Mifflin differ from the county totals reported by the state and therefore the statewide totals differ from the official results accordingly.
    - The shapefiles from Delaware County and the City of Pittsburgh are from the respective jurisdictions instead. Precinct numbers were corrected to match the voter file in the following locales: Allegheny (Elizabeth, Pittsburgh W12), Blair (Greenfield), Bradford (Athens), Greene (Nonongahela), Monroe (Smithfield), Montgomery (Hatfield), Northampton (Bethlehem Twp), Perry (Toboyne), Washington (New Eagle, Somerset), York (Fairview).
    - Precinct boundaries throughout the state were edited to match voter assignments in the PA Secretary of State voter file from the 2018 election cycle. While some edits reflect official updates to wards or divisions the great majority involve voters incorrectly assigned to voting districts by the counties. As such the VEST shapefile endeavors to reflect the de facto precinct boundaries and these often differ from the official voting district boundaries, in some cases quite drastically. Wherever possible edits were made using census boundaries or alternatively using the parcel shapefiles from the respective counties.
    - In certain areas voter assignments appear so erratic that it is impractical to place all voters within their assigned precinct. These areas were edited so as to place as many voters as possible within their assigned precinct without displacing a greater number from their assigned precinct. In general, municipal boundaries were retained except where significant numbers of numbers were misassigned to the wrong municipality. In cases where the odd/even split was incorrectly reversed for precinct boundary streets the official boundary was retained. All such cases involved near equal number of voters swapped between voting districts.
    - The following revisions were made to the base shapefiles to match the de facto 2018 precinct boundaries consistent with the voter file. Individual precincts are noted in cases of splits or merges. Due to the sheer number of edits boundary adjustments are noted at the borough/township level. There may be as many as

two dozen individual precincts that were revised within a given municipality.  
(Note: see VEST's documentation file for their full list of shapefile changes)

- RDH completion of processing steps:
  - We were able to locate the election result and shapefile data from the main sources that VEST describes. In terms of the alternate shapefile sources for Allegheny and Delaware County, we determined that the Allegheny file we could locate matched the Census download. We were unable to locate an alternate shapefile for Delaware County.
  - Various processing steps were required to deal with the OpenElections data. Examples of these include downloading certain counties separately, removing "Total Vote" precincts, and deleting duplicate results in certain counties. It is unclear whether VEST had to make these changes when they used the OpenElections data or whether these issues occurred with later changes of the OpenElections data.
  - In terms of utilizing the canvass reports, because VEST did not specify what they changed using the canvass reports, we first compared the source election results against VEST's file and then looked to see if the differences were explainable by the canvass reports.
  - We did not have access to the 2018 voter file, so we were unable to replicate the processing steps that required that data.
  - Although we did not have access to a voter file, we were able to complete most of the precinct number reassignment, as VEST lays out in their second step above. For the Allegheny Elizabeth reassignments, there were shapefile adjustments that made the reassignment difficult and for other counties, the data we downloaded appeared to match VEST's final file so no reassignment was needed.
  - In joining the shapefiles and election results, we merged precincts that were split by CD or special precincts where it was clear the votes had been summed in VEST's final file. In addition, one precinct in the source file (Addison) was split in VEST's file. For the purposes of validation, we merged the two precincts together in VEST's file and the source shapefile to check it.
- Information not in their processing steps:
  - How they joined election results data to the shapefile data.
  - For the precinct numbering changes, what exact precincts were changed.
  - What exactly they changed using the precinct canvass reports.

### 3. Able to replicate joining election data and shapefiles?

Yes

- There were 9153 precincts that we were able to join, 7 precincts that had only shapefile data and 5 precincts that we only had election data for. All of the election result precincts that could not be matched were in Montgomery County.
- There were initially significant differences between the shapefile and election result names. In order to join the election data to shapefiles, we used a unique ID comprised of the county FIPS and the VTDST. In order to get a VTDST for the election result file, we joined election results to VEST's final file using the vote totals as a unique ID and then pulled the relevant VTDST. The specific changes can be found in the code notebook.

4. Able to replicate joining demographic data to block-level shapefiles?

N/A

5. Able to replicate joining boundary data?

N/A

6. Successfully validated election results?

Yes

- Election results:
  - Column Totals:
    - G18GOVGGLO - VEST 27797 & Sourcefiles 27807
    - G18GOVLKRA - VEST 49238 & Sourcefiles 49238
    - G18GOVRWAG - VEST 2040233 & Sourcefiles 2039598
    - G18GOVDWOL - VEST 2895931 & Sourcefiles 2896074
    - G18USSRBAR - VEST 2135223 & Sourefiles 2134991
    - G18USSDCAS - VEST 2792693 & Sourcefiles 2792656
    - G18USSGGAL - VEST 31228 & Sourcefiles 31264
    - G18USSLKER - VEST 50927 & Sourcefiles 50965
  - Counties with Differences:
    - There are 14 counties with differences: Cambria, Carbon, Crawford, Elk, Northumberland, Bradford, Franklin, Juniata, Lawrence, Montgomery, Montour, Susquehanna, Forest, Lycoming. See below for a county specific breakdown, where the left column is the county FIPS and the right column is the difference between the source file totals and VEST's file totals.
    - G18GOVGGLO
      - 021 2.0
      - 025 -2.0
      - 039 6.0
      - 047 4.0

- G18GOVLKRA
  - 039 1.0
  - 097 -1.0
- G18GOVRWAG
  - 015 -58.0
  - 021 11.0
  - 039 288.0
  - 055 -260.0
  - 067 -8.0
  - 073 3.0
  - 091 -700.0
  - 093 -2.0
  - 097 -9.0
  - 115 100.0
- G18GOVDWOL
  - 015 -28.0
  - 021 20.0
  - 039 194.0
  - 053 -1.0
  - 067 -7.0
  - 081 3.0
  - 091 -30.0
  - 093 -4.0
  - 097 -4.0
- G18USSRBAR
  - 015 -61.0
  - 067 -9.0
  - 091 -150.0
  - 093 -2.0
  - 097 -11.0
  - 115 1.0
- G18USSDCAS
  - 015 -26.0
  - 053 -1.0
  - 067 -3.0
  - 093 -3.0
  - 097 -3.0
  - 115 -1.0
- G18USSGGAL
  - 015 33.0
  - 047 3.0
  - 091 1.0
  - 093 -1.0

- G18USSLKER
      - 015 62.0
      - 025 -2.0
      - 055 -10.0
      - 091 -12.0
    - Precinct-by-precinct differences:
      - There are 9158 total rows to check, as 1 precinct from the VEST file (Upper Mahanoy CD 09, Northumberland County) does not appear in the source election results.
        - 36 of these rows have election result differences
        - 9122 of these rows are the same
        - The max difference between any one shared column in a row is: 600
        - The average difference is: 29.4
        - There are 18 precinct results with a difference greater than 10
      - These 36 precinct differences were spread across 13 counties (the above 14 excluding Northumberland). Because all of these counties were mentioned in VEST's documentation as counties where they consulted election canvass reports, we attempted to find these canvass reports to see if they explained the differences.
        - We were not able to locate canvass reports in Bradford, Carbon, or Franklin county. However, in Franklin County the OpenElections data does appear to contain an error due to the number of votes for the particular candidate being less than the number of straight party votes for that candidate's party in the precinct.
        - We were able to locate a canvass report for Cambria, but it was not helpful for explaining who was right, as neither group's numbers matched for that particular precinct. This is likely because the document we located was unofficial.
        - In Crawford, Elk, Forest, Lawrence, Lycoming, Montgomery, and Susquehanna counties, we were able to locate canvass reports that justify VEST's values and explain the larger differences between the two files.
        - In Juniata and Montour counties, we located canvass reports that had totals more similar to the Open Elections data. The precinct result differences in these counties had a maximum of 3 votes.
        - Lastly, in looking at the canvass report for Northumberland, we were able to locate the precinct that does not appear in the OpenElections data.
  - Geographies:
    - To calculate the area differences precinct-by-precinct in km<sup>2</sup> we converted to crs 3857, calculated the area difference and then divided by 10e6.
    - Of the 9153 (the Upper Mahanoy precinct and 5 other Montgomery precincts did not join) precincts that were joined, the results were as follows:
      - 6140 precincts w/ a difference of 0 km<sup>2</sup>
      - 2510 precincts w/ a difference between 0 and .1 km<sup>2</sup>
      - 322 precincts w/ a difference between .1 and .5 km<sup>2</sup>
      - 97 precincts w/ a difference between .5 and 1 km<sup>2</sup>

- 53 precincts w/ a difference between 1 and 2 km<sup>2</sup>
- 17 precincts w/ a difference between 2 and 5 km<sup>2</sup>
- 14 precincts w/ a difference greater than 5 km<sup>2</sup>
- For the precincts, with large area differences, these seem to be as a result of precinct boundary modifications that we did not carry out, as in the example below, which is a precinct in Cambria County, where green is the precinct overlap, orange is the source file geometry and blue is the VEST file geometry

