VEST PA 2016

State: Pennsylvania
Organization: Voting and Election Science Team (VEST)
Summary of Races included: Presidential, U.S. Senate, Attorney General, Auditor General, Treasurer
Date File Updated: 11/26/2020
Date Report Updated: 05/05/2021

RDH Criteria	Explanation

Is all raw data available?
 	
Options:
No
	<u>Accessible files:</u>
	● VEST PA 16 data file
	    ○ Accessed: 04/22/21, Source: VEST
	    ○ https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/NH5S2I
	● VEST PA 16 documentation file
	    ○ Accessed: 04/22/21, Source: VEST
	    ○ https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/NH5S2I
	● PA Precinct-Level Election Results
	    ○ Accessed: 04/02/21, Source: Open Elections
	    ○ https://github.com/openelections/openelections-data-pa
	    ○ Note: I downloaded the entire notebook and filtered down to just the 2016 results.
	● U.S. Census Bureau's 2020 Redistricting Data Program Phase 2 release
	    ○ Accessed: 04/21/21, Source: US Census
	    ○ https://www.census.gov/geo/partnerships/pvs/partnership19v2/st42_pa.html
	    ○ Note: These can only be downloaded 5 at a time, I downloaded the data for all counties and put all the counties into one folder.
	● Northampton Canvass Report
	    ○ Accessed: 05/05/21, Source: Northampton County
	    ○ https://www.northamptoncounty.org/CTYADMN/ELECTNS/Election%20Results/Archives/Nov%208,%202016%20Results%20by%20Precinct.pdf

- Susquehanna Canvass Report
    - Accessed: 05/05/21, Source: Susquehanna County
    - http://www.susqco.com/Dept/Voter/Pages/Election-Results.aspx)
    - Note: Click on "2016 General Unofficial Canvass Report"
- PA 2011-2018 Congressional District Shapefile
    - Accessed: 05/05/21, Source: Redistricting Data Hub
    - https://redistrictingdatahub.org/dataset/pennsylvania-congressional-districts-2011-to-2018/

Inaccessible files:
- Shapefiles directly from Delaware County
    - We sent an email to a contact at the Delaware County GIS office, but I have not yet received a reply.
- PA Secretary of State voter file from the 2016 election cycle
    - 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.

We did not attempt to locate county municipal boundary files because VEST does not provide enough information as to what files these are and what changes are made using these files.

Processing steps available?
 	
Options:
Yes
	Description of processing steps:
	- There is a link to VEST's full can be found using the link above. They are summarized below:
	- Election Results:
	    ○ VEST election results steps: VEST used OpenElections as the source for their election results. For Allegheny County, presidential-only election results are reported at the county-level and need to be allocated to precincts. VEST also mentions a handful of counties (Cumberland, Elk, Lancaster, Northampton, Susquehanna) where election data was corrected using canvass reports. Lastly, VEST mentions that the vote totals in their file differ from official totals due to differences in Northampton and Susquehanna counties.
	    ○ Of the election result steps, we were able to find the data, allocate the presidential only results in Allegheny County and check the canvass reports in the counties that VEST mentions. It is worth noting that the OpenElections data did entirely match VEST's in Cumberland, Elk, and Lancaster counties, so the OpenElections

data may have been corrected since VEST wrote it's documentation. In terms of the vote totals compared to official reports, there were differences for three presidential candidate totals, as well as one treasurer candidate. See below for more details on this.

- Shapefiles:
  - VEST precinct shapefile steps: VEST used the census redistricting phase 2 data as well as two shapefiles from municipalities themselves. VEST lists a number of precinct mergers and realignments that were made (using the voter file and municipal boundaries to justify these mergers) and mentions that it also used the voter file to change some precinct names.
  - Of the precinct steps, we were able to use the census as a source but not the two municipalities themselves (Delaware County and the City of Pittsburgh) because either we could not find a file or get in touch with a county official or because the file on the website matched what the census had. We were able to replicate the precinct mergers and splits that used Congressional Districts, however we were not able to make any splits or alignments that utilized the voter file or municipal files. For the precinct name changes, we were able to make an educated guess as to what VEST did by examining their final file.

Information not in their processing steps:
- To join the election results to the shapefiles there were a number of precinct mergers, joins and name changes that VEST did not mention in their documentation.

Able to replicate joining election data and shapefiles? Options: No	Ex: No, given the documentation they provided we would not be able to follow their steps to complete the join. However, we were able to join 9072 precincts with 39 election result precincts and 38 shapefile precincts unable to be matched. • How many precincts were you able to join following the documentation? • There were many (~7k) simple name changes (like adding leading zeros) that were made in order to match the election results to the VEST file and then grab the appropriate VTDST to join the election data and shapefiles. • A number of shapefiles were split according to Congressional District, and a number of both election result precincts and shapefile precincts were merged to match one another. • A list of the 39 and 38 precincts is available in the python notebook. • There were 3 no vote precincts in VEST's final file that we created as part of the replication.

| Able to replicate joining demographic data to block-level shapefiles?

N/A | There is no demographic data on the file. |

| Able to replicate joining boundary data?

N/A | There is no boundary data on the file. |

| Successfully validated election results?

Options:
No | **Election results:**
- Of the 9111 precincts, 90 of the precincts had election result differences and 9021 were exactly the same. The differences in all of the 90 precincts were explainable according to VEST's documentation.
- The 90 precincts occur across 3 counties. 5 of these precincts are in Allegheny County, where the differences are 1 vote. This is likely due to slightly different methodology in how the countywide vote totals were assigned.
- The remaining 85 precinct differences occur in Susquehanna and Northampton Counties. VEST mentions these counties as counties where they used canvass reports to correct some election results. In examining these canvass reports, VEST's numbers, and their differences from the OpenElection result number seem justified. |

Election Name	RDH Total	VEST Total	PA SOS Total	Difference (RDH & VEST)
G16PREDCLI	2,926,393	2,926,444	2,926,441	-51
G16PRERTRU	2,970,781	2,971,198	2,970,733	-417
G16PRELJOH	146,715	146,716	146,715	-1
G16PREGSTE	49,941	49,941	49,941	0
G16PRECCAS	21,572	21,572	21,572	0

G16USSDMCG	2,865,009	2,865,012	2,865,012	-3
G16USSRTOO	2,951,706	2,951,702	2,951,702	4
G16USSLCLI	235,142	235,142	235,142	0
G16ATGDSHA	3,057,013	3,057,010	3,057,010	3
G16ATGRRAF	2,891,322	2,891,325	2,891,325	-3
G16AUDDDEP	2,958,819	2,958,818	2,958,818	1
G16AUDRBRO	2,667,318	2,667,318	2,667,318	0
G16AUDLMIN	131,853	131,853	131,853	0
G16AUDGSWE	158,944	158,942	158,942	2
G16TREDTOR	2,991,405	2,991,404	2,991,404	1
G16TRERVOI	2,610,810	2,610,811	2,610,811	-1
G16TRELBAB	135,132	135,132	132,654	0
G16TREGCOM	170,276	170,275	170,275	1

- In its documentation VEST mentions that its totals differ from official results due to differences in Northampton and Susquehanna counties. As the table below illustrates, these differences occur for three presidential candidates and one candidate for treasurer. As the below table shows, this is indeed the case.

Election Name	VEST Total	PA SOS Total	Diff	VEST Nor. Sum	SOS Nor. Sum	VEST Sus. Sum	SOS Sus. Sum
G16PREDCLI	2,926,444	2,926,441	3	66,275	66,272	5,123	5,123
G16PRERTRU	2,971,198	2,970,733	465	71,736	71,736	13,356	12,891
G16PRELJOH	146,716	146,715	1	3,690	3,690	569	568
G16TRELBAB	135,132	132,654	2,478	2,494	16	697	697

Geographies:
- In attempting to join VEST's file to a file where election results were joined to shapefiles, we were able to join 9072 precincts, of these:
    - 6121 precincts w/ a difference of 0 km^2
    - 2439 precincts w/ a difference between 0 and .1 km^2
    - 325 precincts w/ a difference between .1 and .5 km^2
    - 98 precincts w/ a difference between .5 and 1 km^2
    - 59 precincts w/ a difference between 1 and 2 km^2
    - 19 precincts w/ a difference between 2 and 5 km^2
    - 11 precincts w/ a difference greater than 5 km^2

- Given the number of shapefile modifications that VEST performed, we believe that this likely explains most of the precinct shapefile differences.
- The Python notebook contains a print out of all joined precincts with a difference of greater than 1 km^2.