

VEST FL 2016

State: FL

Organization: Voting and Election Science Team (VEST)

Date Updated: Report Date: 04/07/2021, File Date: 04/06/2021

1. Is all raw data available?

No

- Accessible files:
 - File: VEST Florida, 2016
 - Accessed: March 17th, 2021
 - Online: [Harvard Dataverse Link](#)
 - Source: Harvard Dataverse
 - File: VEST Florida, 2016 (Documentation)
 - Accessed: March 17th, 2021
 - Online: [Harvard Dataverse Link](#)
 - Source: Harvard Dataverse
 - File: Florida 2010 Census Blocks
 - Accessed: March 24th, 2021
 - Online: [Redistricting Data Hub Link](#)
 - Source: US Census (via Redistricting Data Hub)
 - File: Florida Precinct-level Election Results
 - Accessed: March 17th, 2021
 - Online: [Florida Division of Elections Link](#)
 - Source: Florida Division of Elections
- Inaccessible files:
 - VEST used a variety of sources to create the Florida precinct shapefile map. Given the number of counties in Florida and the difficulty in obtaining such files, we decided to validate the precinct geographies using precinct shapefile data retrieved from a public records request with the state of Florida.

- The public records request was created January 15th, 2021 and then mailed to one of our data analysts in CD format. The public records file contains all of Florida's precinct and polling place file data from 2012 through the present.

2. Processing steps available?

Yes

- Description of processing steps:
 - Documentation was accessed March 17th, 2021.
 - Election results documentation:
 - Election results from Florida Division of Elections (<http://dos.myflorida.com/elections/data-statistics/elections-data/precinct-level-election-results/>)
 - Precinct shapefile documentation from VEST:
 - Precinct shapefile from many sources.
 - The starting point was the Orlando Sentinel's precinct map for the 2016 presidential primary (<http://interactive.orlandosentinel.com/elections/2016/presidential-primary/results/dem.html>).
 - Hillsborough, Lake, Miami-Dade, Orange, Palm Beach, and Pinellas Counties were updated directly with shapefiles downloaded from county Supervisor of Elections websites.
 - Brevard, Marion, and Putnam Counties were updated with KML files from the respective county Supervisor of Elections websites, from their election result map pages.
 - Columbia, DeSoto, Leon, Polk, and Osceola Counties updates received from Supervisor of Elections websites through personal contact.
 - Highlands, Santa Rosa, Sarasota, and Volusia Counties had some precinct mergers, which were determined based on visual inspection of PDF precinct maps from the county websites.
 - Charlotte, Hendry, Holmes, Jackson, Jefferson, Levy, Madison, and Walton Counties are from VTDs released by the Census Bureau's Redistricting Data Program.
 - Hernando 99, Collier 450, and Palm Beach 8001/8002 are UOCAVA precincts with no real geography and votes were not redistributed.
 - Precinct names were adjusted to align with the formatting used in the voter registration file (e.g., some precincts have leading spaces).
- Information not in their processing steps:
 - Name changes and precinct mergers that were required to construct the precinct shapefile
 - What exactly the precinct shapefile updates entailed.
 - How the data from the Orlando Sentinel's precinct map was translated into shapefile format.

- Note: As mentioned above, this validation used data from a public records request to validate the geographies and not those described in the processing steps, due to the difficulty and complexity of gathering the files in that manner.

3. Able to replicate joining election data and shapefiles?

No

- After creating a unique ID with which to join election results to the precinct shapefiles, we were able to match 5747 precincts. There were 503 precincts shapefiles that we were not able to match election results and 226 precincts with election results that we were not able to match to a shapefile.
- Of the 503 precincts with only shapefiles, these were distributed:
 - BRO 229
 - MAN 103
 - HIL 52
 - CLM 23
 - 8 counties w/ 5-12
 - 13 counties w/ 3 or fewer
- Of the 226 precincts, these were distributed:
 - MAN 60
 - BAY 44
 - PAL 37
 - CLM 22
 - BRO 14
 - 3 counties w/ 5-12
 - 13 counties w/ 3 or fewer
- Note: For BRO and MAN the only shapefiles included in the records request file are 2012 files. There would probably be a higher match rate with more up-to-date files.
- We were unable to easily decrease the number of unmatched precincts with name changes.

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

N/A, no demographic data in the VEST file.

5. Able to replicate joining boundary data?

N/A, no boundary data in the VEST file.

6. Successfully validated election results?

No

- Election results:

Neither the column totals nor the precinct-by-precinct checks matched. Both the column totals and the precinct-by-precinct checks were run on source data that did not include shapefile data.

The differences in election result totals occurred in 5 counties: PAL, CLL, FLA, LAK, and OSC. These differences are explained below:

- In PAL, election results differ because of two things. First are the precincts 8001 and 8002, which are included in the source files, but not in VEST's file. VEST mentions in their documentation that these are Uniformed and Overseas Citizens Absentee Voting Act (UOCAVA) precincts, which do not have geography, and as a result are not included in their file. The second source of difference is due to certain precincts ["PAL01173", "PAL01189", "PAL01247", "PAL02081", "PAL02083", "PAL02097", "PAL02116", "PAL02126", "PAL04129", "PAL05003", "PAL05007", "PAL05018", "PAL05055", "PAL05112", "PAL05113", "PAL05115", "PAL05117", "PAL05119", "PAL06207", "PAL07149"] being included twice in the official Florida election results file, and as such their vote totals are twice as high as those in VESTs. This was a difficult issue to catch, because it is masked when one pivots the election results and it is unclear why these precincts appear twice in the official results, but after removing these doubly-counted votes, the PAL election totals match those of VEST. In the source election results file, every row is a result for a particular candidate in a particular race in a particular polling location. When these double counts occur, it seems as though some polling locations are being captured twice. As an example, in PAL02083 there are separate rows for Donald Trump's votes in polling location "2082, 2083, 2085, 2087, 2090, 2092, 2094" and polling location "2083" that are exactly the same.
- In CLL, election results differ due to the inclusion of precinct 450 in the sourcefiles, which VEST mentions in their processing steps.
- In FLA, election results differ due to the inclusion of precinct 999 in the sourcefiles, which VEST does not mention in their processing steps.
- In LAK, election results differ due to the inclusion of precinct 108 in the sourcefiles, which VEST does not mention in their processing steps.
- In OSC, election results differ due to the the inclusion of precinct 999 in the sourcefiles, which VEST does not mention in their processing steps.
- Note: VEST mentions that Hernando 999 is a UOCAVA precinct where votes were not distributed (similar to PAL 8001 and 8002), but it was not present in the state election results.
- After removing the double counted votes, the differences between these files are completely explained by the non-included precincts mentioned above:

- The difference between source and vest is for G16PRERTru is 129.0
 - The totals of the not included precincts for G16PRERTru are 129.0
 - The difference between source and vest is for G16PREDCli is 246.0
 - The totals of the not included precincts for G16PREDCli are 246.0
 - The difference between source and vest is for G16PRELJon is 7.0
 - The totals of the not included precincts for G16PRELJon are 7.0
 - The difference between source and vest is for G16PRECCas is 2.0
 - The totals of the not included precincts for G16PRECCas are 2.0
 - The difference between source and vest is for G16PREGSte is 4.0
 - The totals of the not included precincts for G16PREGSte are 4.0
 - The difference between source and vest is for G16PREIDeL is 0.0
 - The totals of the not included precincts for G16PREIDeL are 0.0
 - The difference between source and vest is for G16PREOth is 1.0
 - The totals of the not included precincts for G16PREOth are 1.0
 - The difference between source and vest is for G16USSRRub is 139.0
 - The totals of the not included precincts for G16USSRRub are 139.0
 - The difference between source and vest is for G16USSDMur is 156.0
 - The totals of the not included precincts for G16USSDMur are 156.0
 - The difference between source and vest is for G16USSLSta is 5.0
 - The totals of the not included precincts for G16USSLSta are 5.0
 - The difference between source and vest is for G16USSOth is 3.0
 - The totals of the not included precincts for G16USSOth are 3.0
- In terms of source file totals after removing the doubly counted votes, they sum to:
- Source file:
 - G16PRECCas: 16474
 - G16PRERTru: 4617476
 - G16PRELJon: 207014
 - G16PREDCli: 4504404
 - G16PREGSte: 64389
 - G16USSRRub: 4834738
 - G16USSDMur: 4121569
 - G16USSLSta: 196932
 - G16PREIDeL: 9107
 - G16PREOth: 77528
 - [Ballotpedia](#):
 - G16PRECCas: 16475
 - G16PRERTru: 4617886
 - G16PRELJon: 207043
 - G16PREDCli: 4504404
 - G16PREGSte: 64399
 - G16USSRRub: 4835191
 - G16USSDMur: 4122088
 - G16USSLSta: 196956

- G16PREIDeL: 9108
- G16PREoth: 77528
- While there are some small differences between these values, they are different because of results in Seminole county. After manually examining the Seminole county results, our source file value matches the original file for G16PRERTrum, so it does not appear that we (along with VEST) made a mistake, although the ballotpedia data does match the Florida DOS site, but apparently not their file.

Precinct-by-precinct:

- In comparing precinct-by-precinct, there were 5950 precincts, as 23 precincts from the sourcefiles could not be matched and 17 precincts from VEST's final file could not be matched. 6 of the 23 sourcefile precincts were the ones VEST did not include. Since the totals matched excluding these, it is very likely that the 17 remaining precincts in each files could be joined with other precincts to match one another (as none of these matched directly on their own) but we did not take the time to do this.
- In considering precinct-by-precinct election results differences, of the 5950 precincts, 29 of them had election result differences, while 5921 of them were exactly the same. While some of these differences are quite substantial (the greatest difference in one race for one precinct is 1622 votes) because the totals match, differences would likely be fixed by adding votes from non-merged precincts.
 - HAR, PUT, SAR, WAL, WAS were the counties where these differences occur and all of these counties had precincts that did not match in the above join.
- Geographies:

We compared VEST's final file against the file of precincts with both election results and geometries. There were 5746 precincts in this file. Of these 5,779 precincts, 1423 of the geographies matched between the two files, while 4323 differed based on analysis using geom_almost_equals (with decimal = 0). While we did expect there to be some differences between the two files given their different sources, that high a number of differences did seem rather high. In total the symmetric difference between the two files amounted to ~10% of the total area of each precinct file on its own. Furthermore, in plotting the symmetric difference, it seems as if much of this difference is concentrated in areas of water (see image below).

