

MGGG Pennsylvania

State: PA

Organization: MGGG

Date Updated: 7/17/2020

1. Is all raw data available?

No

- Accessible files:
 - Census block-level shapefile
 - Date accessed: 7/14/2020
 - Link: [2010 TIGER/Line® Shapefiles: Blocks](#)
 - Source: Census website
 - Pennsylvania 2010 Census voting tabulation district (VTD) shapefile
 - Date accessed: 7/14/2020
 - Link: [TIGER/Line Shapefile, 2012, 2010 state, Pennsylvania, 2010 Census Voting District State-based \(VTD\)](#)
 - Source: Census website
 - Election Data: 2010 (USED FOR REPLICATION, NOT USED OR LISTED BY MGGG)
 - Date accessed: 7/14/2020
 - Link: [2010 PA VTD Election data](#)
 - Source: Harvard Election Data Archive - Pennsylvania Data Files
 - Election data: 2012 - 2016 (USED FOR REPLICATION, NOT USED OR LISTED BY MGGG)
 - Date accessed: 7/14/2020
 - Link: [General, precincts shapefiles](#)
 - Source: Open Elections
 - MGGG's final PA shapefile
 - Date accessed: 7/14/2020
 - Link: [mggg-states repo](#)
 - Source: MGGG's Github

- Inaccessible files:
 - All election data
 - Election data was provided to MGGG by a private individual, and no source is listed to access
 - Congressional district ID in Governor’s counter-proposed plan shapefile
 - Congressional district ID in Turzai-Scarnati Plan shapefile
 - FiveThirtyEight Democratic favoring plan shapefile
 - FiveThirtyEight GOP favoring plan shapefile
 - FiveThirtyEight plan favoring compactness shapefile

2. Processing steps available?

Yes

- Description of processing steps:
 - MGGG’s processing steps: “Demographic data were aggregated from the census block level and precincts were assigned to districts using [MGGG's proration software](#). Election data were also prorated onto VTDs from the original precinct shapefile using the maup package.”
 - Date accessed: 7/14/2020
 - All of MGGG’s processing steps were replicated in the script
- Information not in their processing steps:
 - Steps that were done to replicate the file that were not explicitly included in MGGG’s processing steps
 - Cleaning the Census VTD shapefile by removing undefined area (north-west corner of state) in the Census VTD shapefile
 - Cleaning the Census block shapefile by removing the same blocks at the north-west corner using QGIS and file ‘block_remove.py’
 - Precinct-level election data in 2012, 2014 and 2016 was collected from Open Elections and joined to the VTDs shapefile (see below for details). This was done in an attempt to replicate the election data provided to MGGG by a private individual.

3. Able to replicate joining election data and shapefiles?

Yes

- Since MGGG had a private individual to provide their election data, the script recreated their shapefile with publicly available election data.
- For 2010 election data, VTD Election data from Harvard Election Data Archive nicely joined to the VTD shapefile on GEOID and VTD name.
- For 2012, 2014 and 2016, precinct-level election data came from Open Elections

- 2012, 2014:
 - To join the election data to VTDS, a GEOID code (containing proper county fips codes) was created, the election data was pivoted, and a script was written to manually match several hundred unmatched observations.
 - 98% of 2012 VTD election data joined.
 - 97% of 2014 VTD election data joined.
- 2016:
 - GEOIDs, or the necessary information to create a geographical ID, are not included in the Open Elections 2016 data. 2016 election data is joined with 2014 Open Elections data to attempt to create the correct GEOIDs for 2016 election data.

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

Yes

- All demographic data joined nicely to the block-level shapefile on blocks' GEOID

5. Able to replicate joining boundary data?

No

- We couldn't access the data required to join political boundary data (see #1 -- inaccessible files)

6. Successfully ran validation?

Yes

- Demographic data:
 - 99% of VTDS have a population deviation of 1 person or less
- Election result returns:
 - 2010
 - 91% of VTDS have a vote deviation of 50 or less for both parties - 2010 Senate Election
 - 2012
 - 91% of VTDS have a vote deviation of 50 or less for both parties - 2012 Presidential Election
 - 2014
 - 90% of precincts have a vote deviation of 50 or less for both parties - 2014 Gov. Election

- 2016
 - 86% of precincts have a vote deviation of 50 or less for both parties - 2016 Presidential Election
- No topological issues