MGGG OH

State: Ohio
Organization: MGGG
Date Updated: Report: 02/03/21, File: 06/26/20

1. Is all raw data available?

No, not all of the raw data was available.

- Accessible files:
    - MGGG OH File
        - 01/27/21
        - https://github.com/mggg-states/OH-shapefiles
        - Source: mggg-states github repository
    - Ohio precinct-level elections results - MEDSL
        - 01/29/2021
        - Available from the MEDSL Dataverse, where each election must be downloaded separately
            - 2016 Precinct-Level State House & State Senate
            - 2016 Precinct-Level U.S. Presidential
            - 2016 Precinct-Level U.S. Senate
            - 2016 Precinct-Level U.S. House
        - Source: MEDSL Dataverse
        - Note: MGGG uses 2016 election results from both the Ohio Secretary of State and MEDSL. In the end we were able to write this report using just the MEDSL results
    - Demographic Data
        - 01/27/21
        - Accessed via Census API
        - Source: US Census Bureau
        - Note: MGGG appends demographic and VAP data, retrieved from IPUMS NHGIS. As in other MGGG validation report, we are assuming that this is 2010 U.S. Census data, and choose to download this data directly using the Census API.

- Inaccessible files:
    - Ohio precinct shapefile

        - 01/27/21
        - Available from the MGGG/ohio-precincts repository, unchanged since the group's report
        - Source: mggg repository, for detailed info on precinct data sources, use the link above
        - Note: Although this file was linked to by MGGG and accessible in that way, it does not list all of the sources for the precinct shapefiles. They provide the shapefile that was either received from the state or geo-referenced in the counties.zip file, but they do not provide the raw images, pdfs, or other files that were retrieved from the state.
    - Legislative Boundaries
        - 01/29/21
        - MGGG does not list a source for legislative boundaries. First, we tried to use the legislative boundaries from TIGER/LINE and discovered these were not the same as the Google District Maps on the state website. Ultimately we found the correct files on Justin Levitt's *All About Redistricting* page for OH:
            - State Senate
            - State House
            - Congressional
        - Source: Justin Levitt's *All About Redistricting*, with the links themselves linking to two open data arcGIS pages as well as one on the OH sos page.
        - Note: Thank you Justin Levitt for the files!

2. Processing steps available?

Yes

- Description of processing steps:
    - On mggg-states/OH-shapefiles: Ohio precinct shapefile with 2016 election results (accessed 02/03/21) they write: "Demographic data were aggregated from the block level using MGGG's proration software maup. Congressional, house, and senate district IDs were assigned to precincts also using this package."
- Information not in their processing steps:
    - How to resolve the "cannot reindex from a duplicate axis" error when attempting to use the maup "assign" function. We worked around this by utilizing the maup "intersections" function, but one question here is what value to assign to "area_cutoff" when running this.
    - How they matched election data to shapefiles where the precinct code did not match.
    - How they assigned the libertarian votes for the presidential election, they likely just assigned Gary Johnson votes there, but they did not specify this.
    - Where MGGG used SOS election results instead of MEDSL election results. MGGG lists both as a source, however we were able to replicate the file using only MEDSL's cleaned election results.

3. Able to replicate joining election data and shapefiles?

Yes, but not all of them as there were 8887 precincts in the MEDSL election results and only 8882 in the precinct shapefile.

- On the first join, there were 53 precincts from precinct shapefile and 58 precincts from MEDSL election results that were not matched.
- There we separate fixes for these join issues depending on the county:
    - Richland County: Joined on "PRECINCT" instead of "PRENAME"
    - Darke County: Involved appending the specific precinct name to the "PRENAME" column
    - Adams County: Has election results reported in two different locations for each of the unmatched precincts, we combined these totals and matched to the entire precinct geography.
    - Williams County: Appended some simple text to each precinct to make it work
    - Jackson County: Were unable to fix these issues
    - Mahoning County: Unable to fix these join issues
- After those fixes (which involved condensing 4 precincts w/ election results into 2, thus the slightly different totals), there were only 3 precincts from the MEDSL election results file that were not joined. Every precinct from the precinct shapefile was matched to election results. These three were: PRECINCT WELLSTON 2-B, PRECINCT MADISON II, and PRECINCT SBC 4

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

Yes, but we had slightly different totals from MGGG, see below.

5. Able to replicate joining boundary data?

Yes, we were able to join precincts to state-senate, state-house and congressional districts. We had slight differences, see below.

6. Successfully ran validation?

Yes, validation successfully ran. We were able to match all 8882 precincts in their file with our file from the original sources. This was to be expected as the shapefile uses precincts that MGGG themselves collected. Any differences are listed below:

- Election result returns:
    - Comparing precincts: of the 8882 precincts, only 55 had differences, with the largest difference between column values being 343, when a difference occurs, the average difference is 58.3 votes.
    - Comparing race totals (between MEDSL election file not matched to any shapefiles and MGGG file): of the 12 columns, 8 had identical totals and 4 had differences, with the largest difference between the sum of a column from the election file and their file being 625 votes for PRES16R, or .022% of the total votes in that column in our file.

    - Comparing race totals (between our matched file and MGGG file): of the 12 columns, 2 had identical totals and 10 had some difference with the largest difference between the sum of a column from our file and their file being 775 votes for SEN16R, or .022% of the total votes in that column in our file.
    - The differences in the election results may be explained by cases where MGGG used SOS results instead of MEDSL results, however they did not specify where this was the case.
- Shapefile differences:
    - Comparing the geometries of the shapefiles to 1 decimal place, 8,736 of the precincts matched exactly while 146 of the precincts had some differences.
    - In spot checking one of these differences, it seems like minor changes were made to the shapefiles in some instances. While this was a bit surprising, given that MGGG was the source for the shapefiles, on GitHub it appears as if they "Fix[ed] topology errors" in the file in June 26th, 2020, which may explain why there are some minor differences, although these changes were not documented.
    - The differences in the shapefiles may explain the handful of discrepancies between the boundary data.
- Boundary data differences:
    - For the state senate districts, we assigned all precincts to the same districts except for 3 out of 8,882. Checking 1 of these 3, it seemed like our assignment was correct, the other 2 we were not able to find and check.
    - For the state house districts, we assigned all precincts to the same districts except for 4 out of 8,882. Checking 2 of these 4, it seemed like our assignment was correct, the other 2 we were not able to find and check
    - For congressional districts, we assigned all precincts to the same districts except for 1 out of 8,882. We were not able to find this precinct to check.
- Census data differences:
    - The census data we matched to precincts was slightly different from MGGG's, most likely due to differences in between what maup function and parameters were used.
    - The total population difference was ~3500, which in a state of ~11.5m people is not massive.
    - Total population matched in 2609 precincts and differed in 6273 precincts.
        - 92% of precincts had a difference of less than 100 for the total population.
        - 63.6% of precincts had a difference of fewer than 10 for the total population