MGGG Virginia

State: VA

Organization: MGGG

Date Updated: 7/27/2020

1. Is all raw data available?

   Yes
   - Accessible files:
     - congressional_districts
       - Date accessed: 7/17/2020
       - Link: Common Wealth of Virginia Division of Legislative Services
       - Source: Common Wealth of Virginia Division of Legislative Services website
     - demographic_data
       - Date accessed: 7/17/2020
       - Link: IPUMS NHGIS
       - Source: National Historical Geographic Information System website
     - house_districts2011
       - Date accessed: 7/17/2020
       - Link: Common Wealth of Virginia Division of Legislative Services
       - Source: Common Wealth of Virginia Division of Legislative Services website
     - house_districts2019
       - Date accessed: 7/17/2020
       - Link: Common Wealth of Virginia Division of Legislative Services
       - Source: Common Wealth of Virginia Division of Legislative Services website
     - PGP-election-precincts
       - Date accessed: 7/17/2020
       - Link: PGP VA-gerrymander repo
       - Source: Princeton Gerrymandering Project's Github
     - senate_districts
       - Date accessed: 7/17/2020
       - Link: Common Wealth of Virginia Division of Legislative Services

- ■ Source: Common Wealth of Virginia Division of Legislative Services website
  - ○ VA_blk-level_shapefile *(used for our replication only, not listed by MGGG)*
    - ■ Date accessed: 7/17/2020
    - ■ Link: 2010 Virginia block-level shapefile, Census
    - ■ Source: Census website
  - ○ Block-level demographic data
    - ■ Date accessed: 7/17/2020
    - ■ Source: Census API

- ● Inaccessible files: N/A

2. <u>Processing steps available?</u>

Yes
- ● Description of processing steps:
  - ○ MGGG's processing steps: "Demographic data were aggregated from the block level to precincts using MGGG's proration software. Congressional, house, and senate district IDs were assigned to precincts also using this package."
  - ○ Date accessed: 7/17/2020
  - ○ Steps completed that aligned with their processing steps:
    - ■ Demographic data was aggregated to precincts using *maup* -- MGGG's proration software
    - ■ Congressional, house and senate districts IDs were assigned to precincts also using *maup*

- ● Information not in their processing steps:
  - ○ Steps to replicate the file that were not explicitly included in MGGG's processing steps:
    - ■ Setting projections of all shapefiles to that of the VA_Precincts_Elections shapefile from PGP, before setting shapefiles' geometries to *buffer(0)* for using *maup* (see code for details)
    - ■ Joining block-level demographic and election data with Census block-level shapefile *(used for replication only, not listed by MGGG)* in order to prorate block-level data to precinct-level
    - ■ Prorating voting population data from house districts to blocks
    - ■ Prorating demographic and voting population data from the block level to precincts
    - ■ Additional step not listed by MGGG: Obtaining Virginia's block-level demographic data from Census API, instead of NHGIS as listed. Doing the rest of the replication process as described. As a result, the demographic data better matched that in MGGG's VA shapefile, in terms of both numerical values and naming conventions. Using Census data, thus, is recommended in this case.

3. <u>Able to replicate joining election data and shapefiles?</u>

NA

● Precinct-shapefile comes from the Princeton Gerrymandering Project (PGP) and already contains election data. No joining of election data was necessary.

4. <u>Able to replicate joining demographic data to block-level shapefiles?</u>

Yes

● MGGG's final shapefile includes VAP population counts. However, MGGG does not specify the source of this data. These counts are included in the Commonwealth of Virginia Division of Legislative Services' state house (2011), state senate, and congressional shapefiles. Under the assumption that VAP population counts come from one of those shapefiles, VAP data was disaggregated from state house districts to block-level using *maup*.
● All demographic data joined nicely to the Census block-level shapefile *(used for replication only, not listed by MGGG)* using county, track, and block ID's.
● Additional step not listed by MGGG: Obtaining Virginia's block-level demographic data from Census API, instead of NHGIS as listed. All demographic data joined nicely to the Census block-level shapefile on GEOID. Moreover, the demographic data better matched that in MGGG's VA shapefile, in terms of both numerical values and naming conventions. Using Census data, thus, is recommended in this case.

5. <u>Able to replicate joining boundary data?</u>

No

● All state house districts (2011 and 2019) joined to blocks using *maup*
● All congressional districts joined to precincts using *maup*
● Unable to join state senate districts since the assignment of those districts was inaccurate
  ○ The assignment of state senate districts didn't match MGGG's for 2439/2439 precincts

6. <u>Successfully ran validation?</u>

Yes

● Demographic data
  ○ Using NHGIS demographic data
    ■ 99% of the precincts had a population deviation of 10 or fewer people compared to MGGG's
    ■ 81% of my precincts had a black population deviation of 20 or fewer people compared to MGGG's

- ● This may be due to the demographic data being used or the order of proration
  - ○ Using Census API demographic data
    - ■ 99% of precincts have a population deviation of 10 or fewer people
    - ■ 99% of precincts have a black population deviation of 10 or fewer people
    - ■ 99% of precincts have a black voting-age population deviation of 10 or fewer people
- ● Election boundary data
  - ○ All state house districts (2011 and 2019) matched those in MGGG's final shapefile
  - ○ All congressional districts matched those in MGGG's final shapefile
- ● No topological issues
- ● Comparing PGP's validation on organization's shapefile and replicated shapefile:
  - ○ Vote score:
    - ■ Based on the 2 reports generated by PGP's validation tool, the 2 shapefiles are identical with regard to the 2016 presidential data, with a vote score value of 0.99, suggesting that election data at the county level matched the official county state election data.
    - ■ For both reports, worst_county_vote_score deviates 11 percentage points from the ideal value of 1 at 1.11. This suggests that a county (Manassas City, upon further inspection) contains 1,700 more recorded votes than the official state election data.
    - ■ county_vote_score_dispersion is large, at 52903.85 with an ideal of 0. This indicates that there is a difference between the replicated file and MGGG's final file with regards to official state election data records at the county-level.
    - ■ County with the worst vote scores (ideal score = 1)
      - ● Manassas city - vote score: 1.11
  - ○ Area difference score:
    - ■ median_county_area_difference_score is near the ideal score of 0 at 0.002.
    - ■ Counties with the worst area difference scores (idea score = 0)
      - ● Accomack County: 0.570
      - ● Hampton city: 0.5696
      - ● Mathews County: 0.6106
      - ● Northampton County: 0.564
      - ● Poquoson city: 0.741