

MGGG Texas

State: Texas

Organization: MGGG ← PGP

Date Updated: 12/3/2020

1. Is raw data available?

No

- Accessible files:
 - [PGP OpenPrecincts](#) data for Texas directly links to the MGGG GitHub [repository](#)
 - Precinct shapefile
 - [MIT Computer Science & Artificial Intelligence Lab](#)
 - Note: 2016 VTD shapefile
 - The year was not mentioned in the GitHub readme so this required downloading several [archived VTD shapefiles from Texas Capitol Data Portal](#) and merging to figure out which one it was
 - Retrieved on: 11/6/2020
 - 115th Congressional district shapefile
 - [Census](#)
 - Senate districts
 - [PLANS172](#)
 - In place from 2013-2020
 - Last updated by source on: 10/4/2019
 - Retrieved on: 11/6/2020
 - House districts
 - [PLANSH407](#)
 - Note: Not sure why MGGG used this one, the Data Portal description says Texas Latino Task Force Proposed House Plan
 - Last updated by source on: 10/7/2019
 - Retrieved on: 11/6/2020
 - Precinct-level election data
 - [Texas Capitol Data Portal](#)
 - Note: Election data and turnout results stored in separate files, has 5 races across 2012, 2014, 2016
 - [2012](#) U.S. Senate, President
 - [2014](#) U.S. Senate, Governor

- [2016](#) President
 - Retrieved on: 11/6/2020
- Files listed by MGGG that were not readily accessible
 - VTD-level demographic data
 - MGGG said demographic information was compiled by the Texas Legislative Council using 2010 Census data, but this was not publicly available on their GitHub or the TX Capitol Data Portal
 - Emailed and received directly from the Capitol Data Portal from Eddie Torres (Eddie.Torres@tlc.texas.gov)
 - Retrieved on: 11/6/2020
- Note of data documentation:
 - MGGG GitHub links to another website for downloading the final Texas elections shapefile, but did not mention which year the VTD shapefile was used. It is a direct download link, so no description was included
 - Link to Texas Legislative Council Redistricting [website](#) is broken, but the same election data could be found on the Texas Capital Data Portal

2. Processing steps available?

Yes

- Description of processing steps:
 - Very little information regarding processing steps
 - MGGG noted that shapefile topology was checked and minor corrections were made using the `check_shapefile_connectivity.py` script from their [Preprocessing Github repository](#) which is now deprecated
 - [Roundoff](#) was used to assign VTDs with U.S. congressional, state senate, and state house districts
 - MGGG has an updated proportion software called [maup](#) which can be used to perform the same aggregation functions described above
- Information not in their processing steps:
 - N/A

3. Able to replicate joining election data and shapefiles?

Yes

- Required pivoting election data so that each row has the precinct and corresponding results for various races, rather than every VTD reporting vote totals for each candidate and race (how it is initially formatted)
- Filtered out unnecessary columns before merging turnout, election, and census data into a single dataframe called `merged_voter_data`

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

Yes

- Matched to VTD-level shapefiles, all demographic fields and counts match seem to match when validated

5. Able to replicate joining boundary data?

Yes

- Was able to replicate all precinct assignments to congressional, house, and senate districts using maup.assign

6. Successfully ran validation?

Yes

- Confirmed validation by checking total population difference, 2012 presidential election difference, and 2016 presidential election difference
 - Results came back 1.0 meaning the entries matched perfectly between the validation file and MGGG final file