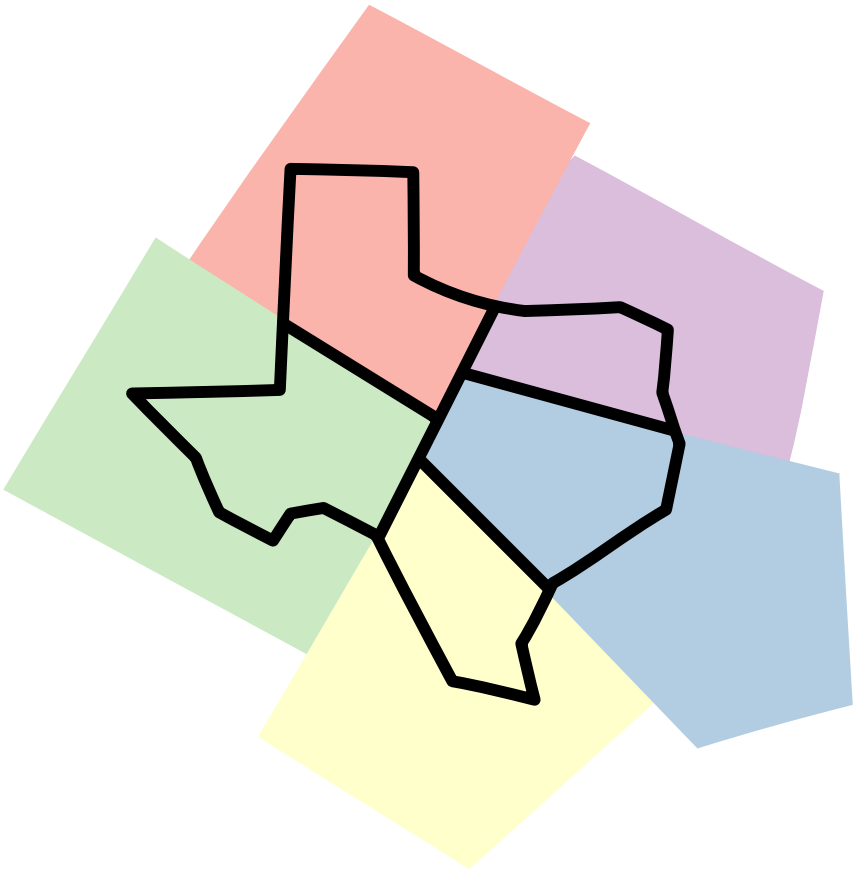
DATA FOR 2021 REDISTRICTING

IN TEXAS



DATA FOR 2021 REDISTRICTING
IN TEXAS

Published by the Texas Legislative Council
September 2021

Texas Legislative Council
Lieutenant Governor Dan Patrick, Joint Chair
Speaker Dade Phelan, Joint Chair
Jeff Archer, Executive Director

The mission of the Texas Legislative Council is
to provide professional, nonpartisan service and support
to the Texas Legislature and legislative agencies.
In every area of responsibility,
we strive for quality and efficiency.

Copies of this publication have been distributed in compliance with the state depository law (Subchapter G, Chapter 441, Government Code) and are available for public use through the Texas State Publications Depository Program at the Texas State Library and other state depository libraries. This publication can be found at https://redistricting.capitol.texas.gov/.

If you have questions or comments regarding this publication, please contact Karen White by phone at (512) 463-1155 or by email at karen.white@tlc.texas.gov.

TABLE OF CONTENTS

PREFACE

This publication, *Data for 2021 Redistricting in Texas*, has been prepared to provide members of the legislature and other interested persons with information about the data that is available through the Texas Legislative Council's redistricting computer system, including on maps and reports.

The publication describes the data that the council collects and prepares for redistricting, the sources of the data, and the methodology used by the council to link the various types of data used in the redistricting database. The data is presented in three sections: population data, geographic data, and election data.

For more information concerning redistricting data, contact the council at (512) 463-6622.

Jeff Archer
Executive Director

This page left intentionally blank.

INTRODUCTION

Concerns over the confidentiality of census responses has led the Census Bureau to introduce a new method of data privacy in the release of 2020 census data (see discussion of differential privacy below). Given changes made to census population data because of differential privacy, allocation of election data, different time periods for data collection, and different geography for data reporting, all redistricting data should be used with knowledge of its limitations. This publication explains the data provided in the Texas Legislative Council's system for redistricting in Texas and highlights those limitations, including areas in which data has been estimated. As a general rule, due to the effects of differential privacy and the allocations of data to small geographic units, data published at the smallest geographic level (census block) has the most privacy adjustments and the least precision. As larger geographic levels are compiled (census tracts and legislative districts, for example), the precision of the reported data for these larger areas improves.

This page left intentionally blank.

POPULATION DATA

The State of Texas uses population data from the 2020 federal decennial census count as the primary basis for redistricting in 2021. Under federal law, the United States Census Bureau delivers the state population totals to the president and, using a formula known as the method of equal proportions, the number of congressional seats that each state is apportioned for the next decade. In accordance with Public Law 94-171, the bureau is required to provide the states with the official census population numbers needed for redistricting, including total and voting age population by race and ethnicity for every census geographic level, by April 1, 2021. Because of complications and delays attributable to COVID-19, release of the data was postponed until August 12, 2021.

RACE AND ETHNICITY CLASSIFICATION

In previous decades, population data has been grouped into five race and ethnicity categories for Texas redistricting modeling and reporting: Black, Hispanic, Black+Hispanic, Anglo, and Other. “Black” encompassed all people identifying themselves as Black or African American on the census questionnaire, even if they also identified themselves with other racial/ethnic groups. “Hispanic” encompassed all people identifying themselves as of Hispanic, Latino, or Spanish origin, whatever their race. “Black+Hispanic” has been a combined total of all those identifying themselves as Black and all those identifying themselves as Hispanic, adjusted so that those identifying themselves as both Black and Hispanic are not counted twice. “Anglo” included all people who selected “White” as their only race and did not identify themselves as of Hispanic, Latino, or Spanish origin. “Other” encompassed everyone who does not fall into any of the four classifications described above.

In response to the growing Asian population across the state, census data categories for the coming decade from the 2020 federal census have been revised as follows:

- Anglo – As commonly used in Texas, those persons who identify their race on the census form as White only and who do not identify themselves as of Hispanic, Latino, or Spanish origin.
- Non-Anglo – A new population category developed for the Texas legislative redistricting systems that includes all persons who are not categorized as Anglo. Non-Anglo includes persons who identify their race or ethnicity on the census form in whole or part as Hispanic, Black, or Asian, and also includes American Indian or Alaska Native and Native Hawaiian or Other Pacific Islander. The sum of Anglo plus Non-Anglo in a population report equals the total population in the district.
- Asian – Those persons who identify their race on the census form in whole or part as Chinese, Filipino, Asian Indian, Vietnamese, Korean, Japanese, or Other Asian.
- Black – Those persons who identify their race on the census form in whole or part as Black or African American.
- Hispanic – Those persons who identify themselves on the census form as of Hispanic, Latino, or Spanish origin. Hispanic persons can be of any race.
- Black+Hispanic – A combined population category that includes all persons who identify their race on the census form in whole or part as Black or African American and all persons who identify themselves on the census form as of Hispanic, Latino, or Spanish origin, without counting anyone twice. The category is frequently examined for redistricting purposes in areas in which Black and Hispanic voters may form political coalitions or vote together as a bloc.

CITIZENSHIP DATA

Citizen Voting Age Population (CVAP) is established using the American Community Survey (ACS). The ACS is a sampled survey that collects data from almost 3.5 million housing unit addresses during the course of a year. The estimates are published in one-year and five-year tables. The one-year estimates are released only for geographic units with a population of 65,000 or more, while the five-year estimates are released for smaller levels of geography down to block group (and so include counties, tracts, and block groups), regardless of the number of people in the area. Because redistricting plans are built in part from geographic units with small populations, block group data from the ACS five-year tables is used to produce citizenship estimates.

Sampled data, unlike an enumeration of the total population, is subject to error that can be calculated. To allow data users to assess the precision of the estimates, ACS estimates are released with margins of error that describe how much each estimate could differ from the actual value if the data were drawn from the complete population rather than a sample.

Estimating and Reporting CVAP for Redistricting in Texas

Despite an initial decision by the U.S. Department of Commerce to include citizenship questions on the 2020 census forms, citizenship was ultimately excluded and thus is not reported with the 2020 census data. At the request of the Department of Justice, the Census Bureau publishes CVAP estimates each year from the five-year ACS data set. The ACS CVAP estimates used for Texas redistricting reports in 2021 are from data released in 2021 but collected between 2015 and 2019. Because the CVAP data derives from surveys done before the 2020 census, the geographic units for which the CVAP data are reported are from the 2010 decennial census, which often differ from sub-county 2020 decennial census geography and require approximations to be made.

Data from the annual CVAP special tabulation are reported at the block group level and cannot be accurately disaggregated to the block level, the level of detail provided for other population data used in redistricting. For Texas redistricting reports, district-level CVAP is determined by adding estimates of CVAP in block groups that are either fully contained or in which more than 50 percent of the total census population is contained within a district. As a result of the 50 percent threshold, the calculated district CVAP estimates will include some individuals who reside outside the district and exclude some individuals who live within the district.

Two redistricting reports provide more specific information about the block group population included in the redistricting system's district citizenship calculations. Red117 indicates the total population for each block group and whether the block group is contained entirely within the district. Red118 indicates the total population for block groups that are fully contained in the district, the population for those block groups that are split but are included in the district totals, and the population for those split block groups that are excluded from the analysis.

Texas redistricting systems report race and ethnicity categories from the decennial census population in summarized combinations that are slightly different from the way data are reported by the ACS. In the Texas legislative redistricting system, the "Anglo" category includes people who identify as White only and non-Hispanic. Asian, Black, and Hispanic categories include persons who identify as that plus any other category. The table below compares the Texas redistricting categories from the decennial census to the similar categories available on reports for ACS CVAP data.

ACS Categories for Estimated CVAP	Texas Redistricting Categories for 2020 Census Total and VAP	Comparison
Hispanic or Latino	Hispanic (Hispanic of any race/ethnicity)	Same
White Alone, Not Hispanic or Latino	Anglo (White Alone, Not Hispanic)	Same
Black or African American Alone, Not Hispanic or Latino Black or African American and White, Not Hispanic or Latino American Indian or Alaska Native and Black or African American, Not Hispanic or Latino	Black (Black Alone or in combination with any race/ethnicity)	Similar
Asian Alone, Not Hispanic or Latino Asian and White, Not Hispanic or Latino	Asian (Asian Alone or in combination with any race/ethnicity)	Similar
Native Hawaiian or Other Pacific Islander Alone, Not Hispanic or Latino	Included in Non-Anglo (Total minus Anglo)	Not Similar
American Indian or Alaska Native Alone, Not Hispanic or Latino American Indian or Alaska Native and White, Not Hispanic or Latino	Included in Non-Anglo (Total minus Anglo)	Not Similar
Remainder of Two or More Race Responses, Not Hispanic or Latino	Included in Non-Anglo; May be included in Black or Asian	Not Similar

DIFFERENTIAL PRIVACY

The Census Bureau's primary task is to collect accurate data about the individuals in the United States to be used for statistical purposes. In addition, the bureau is required by law to keep that data about particular individuals confidential. These mandates are not always compatible. In order to maintain confidentiality, data must not be so accurate that it risks the identification of individuals, but it must not be distorted beyond use for statistical purposes. To mitigate these competing issues, the Census Bureau for decades has used methods of disclosure avoidance, a process intended to ensure that identifiable personal information is adequately obscured while still providing useful statistics to describe the population. Previous methods of disclosure avoidance were simpler, including data swapping, a process in which some values in one geographic unit are substituted with some values in another geographic unit that has similar characteristics. Data swapping was one of the common methods of disclosure avoidance used by the Census Bureau in recent decades.

Data privacy has been a concern among computer scientists since the 1970s. Computers have become more powerful and large amounts of individual-level data have become more available, leading to a growing risk that census data can be linked to specific individuals. In light of growing academic research on data privacy, officials at the Census Bureau became interested in the possibility of implementing additional safeguards for bureau publications. Tests conducted by the bureau have indicated that threats of reconstruction of individuals' data were not just hypothetical, but a genuine risk to the bureau's ability to maintain data confidentiality. The ability of someone to reconstruct the bureau's confidential database was concerning, but a greater issue was the possibility of someone connecting auxiliary data with personally identifiable features to the reconstructed database, called a linkage attack. For example, a dataset containing demographic and location data may be linked to another dataset containing information about voting, health care, or credit card purchases using specific variables contained in each dataset such as zip code, gender, age, etc. The threat of linkage attacks was the impetus for the Census Bureau to select differential privacy as the disclosure avoidance technique to be used for the 2020 decennial census.

What Is Differential Privacy?

Differential privacy is a complex method to assess privacy risk and provide, through mathematical equations, that confidential data of respondents from surveys and other statistical tools remains unidentified. It does not change the confidential raw data, but it adds statistical "noise" to published data derived from the raw data. The added noise creates a synthetic dataset so that publication of statistics from this data makes it difficult to identify any individual in the dataset.

Differential privacy has become the preferred method of disclosure avoidance because it allows the Census Bureau to adjust the algorithms to achieve an acceptable balance between data accuracy and utility and the desired degree of safeguarding privacy. As applied by the bureau, differential privacy reportedly makes reconstructing individual data practically impossible. A major concern for redistricting is the difference of impact of differential privacy on various geographic levels and population sizes. Large geographic areas are unlikely to be significantly distorted as a result of noisy data. Small geographies, however, require additional noise to prevent reidentification of individual data, making the reported data less accurate. Small populations, including some race categories and rural areas, have more noise applied.

Effect on Data Accuracy

The Census Bureau's application of differential privacy to the data published from the 2020 decennial census affects the precision of the data used in the legislature's redistricting system. Users of the system should be aware of the ramifications of the Census Bureau's differential privacy decisions:

(1) The census population data used for redistricting is not exactly the same as the data collected by the bureau, but has been changed by the application of differential privacy algorithms. The total population, voting age population, and racial and ethnic breakdown of those populations, reported for any particular census unit, from the individual census block to the county, is not necessarily the same as what the bureau collected. The precise confidential survey data has been altered to address privacy concerns.

(2) The degree of uncertainty of the precision of the published census data available in the redistricting system is greatest at the block level. The data becomes increasingly accurate at each higher level of geography.

(3) Because block-level census data has been altered by the application of differential privacy, the allocation of certain election and voter data used in the redistricting system will be affected when the user splits a voting tabulation district (VTD) between districts. It is not possible to determine the degree of uncertainty differential privacy brings to the allocation. As discussed elsewhere in this publication, the allocation of election and voter data using census data is not a precise method in itself.

See the Appendix for more information about differential privacy as implemented for the 2020 census.

POPULATION REPORTS

Council population reports are divided into two main categories: plan and non-plan. Plan reports provide information about the population of current and proposed redistricting plans, and they can be run for a single district, several districts, or all districts in a plan. Non-plan reports present data for other political subdivisions, such as counties, cities, voting tabulation districts (VTDs), and school districts.

Some of the plan reports include the percent by which each district's total population deviates from the ideal district population, the total and voting age population, and the racial and ethnic composition of each district. The plan reports that list a district's geographic components, such as counties, cities, VTDs, or school districts, provide the percent of that component's population that is in the district. For example, for a Central Texas senate district that is composed of all of some counties and part of Bexar County, the report would list the whole counties as 100 percent and the part of Bexar County as the percent of its population that is in the district. Non-plan reports include the population totals and racial and ethnic composition for any or all counties or cities

in the state, for all school districts in the state, or for all VTDs in a county. The reports use cities as they existed at the time of the 2020 census, but use precincts, VTDs, and school districts that are updated by the council periodically throughout the decade.

Copies of reports may be found on the Capitol Data Portal (see below).

This page left intentionally blank.

GEOGRAPHIC DATA

The geographic data used in redistricting includes the census geography used as building blocks for districts, the county election precincts used to collect and report voter and election returns data, other geographic areas used for reference such as cities and school districts, and the district assignments of the residence locations of incumbent members under proposed redistricting plans.

CENSUS BUREAU GEOGRAPHIC DATA

The Texas Legislature uses census geography for modeling districts. Census geography established for a decennial census does not change during the decade, and population data is reported by these stable geographic units. Census geography traditionally included counties, census tracts, block groups, and census blocks defined by visible features such as roads and rivers to the extent possible. VTDs, groupings of census blocks that approximate county voting precincts, have been included in the geography for the 1990s redistricting, and school district boundaries were added in 2000. In 2010, the Census Bureau began to include voting precinct boundaries submitted by the states to be used for VTDs, regardless of whether they followed physical features.

TIGER

The Census Bureau has digitized maps of the entire United States into a computerized geographic database, the Topographically Integrated Geographic Encoding and Referencing (TIGER) system, which includes address files to help with census operations. The TIGER database contains all census-related map features such as roads, railroads, rivers, lakes, cities, and school districts. The file also contains related attributes (names, address ranges, and geographic relationships to other features) for each of the map features and contains geographic identification codes, such as county codes, city names, and tract and block numbers. TIGER constitutes the foundation of the redistricting geographic database from which council staff extract county, tract, block group, and block boundaries to match with the Census Bureau's population data for use in the council's redistricting application (RedAppl).

Census Redistricting Data Program

An important aspect of the council's redistricting support is the need to link the census geography used to build districts with the census population data, which is reported by census block, and with election data, which is reported by voting precinct. The Census Redistricting Data Program provides the opportunity to use VTDs as the base for this link.

Council staff work with the counties throughout the decade to build a precinct and VTD database for each election. Where county precinct boundaries follow census block boundaries, VTDs match county election precincts exactly. Where the precinct boundaries do not follow block boundaries, council staff select the block boundaries that most closely follow the precinct boundaries for the corresponding VTD boundaries.

Council staff worked with the Census Bureau to submit 2020 primary precinct boundaries, even if those boundaries were drawn on nonvisible features, for inclusion in the 2020 TIGER database. Subsequently, council staff collected county precinct changes in effect for the 2020 general election and created a 2020 general election VTD layer for use in the legislature's redistricting systems. As a result, the 2020 general election VTD layer in the legislature's systems (available for download from the capitol data portal) is a more current depiction of county precinct boundaries than the VTD layer available from the census 2020 TIGER database.

DISTRICT MODELING UNITS

RedAppl allows users to combine various units of census geography into districts. Modeling units include counties, VTDs, census tracts, block groups, blocks, and cities and census designated places.

Counties

The state's 254 counties are the largest geographic units of the state that are used as district modeling units. A number of districts in the state consist entirely of whole counties.

Voting Tabulation Districts

The VTDs in the 2021 redistricting database closely correspond to the precincts in effect for the 2020 general election. On the rare occasion that a county voting precinct is in two noncontiguous pieces, it is included as two separate VTDs in the database. For example, a noncontiguous precinct 0001 would be included in RedAppl as VTD 0001A and VTD 0001B. If a 2020 general election precinct does not match any census geography, it is consolidated into an adjacent VTD. This situation is more likely to occur when new precincts are added by a county for new subdivisions that are not included in the census 2020 TIGER block geography.

The 2020 general election VTDs in the council's redistricting geographic file are not the same as the 2020 primary VTDs in the Census Bureau's geographic file and so the population data reported for the VTDs in the Census Bureau's PL 94-171 file does not correspond to VTD population data in RedAppl. The 2020 general election VTD geography and population data that correspond to the 2020 general election VTDs are posted in the geography section of the Capitol Data Portal (https://data.capitol.texas.gov).

Census Tracts

Census tracts are statistical subdivisions of counties delineated by local committees in accordance with Census Bureau guidelines for the purpose of collecting and presenting decennial census data. They are relatively stable from decade to decade and were designed to have homogeneous population characteristics, economic status, and living conditions at the time they were established. Tract boundaries generally follow visible features and contain between 1,200 and 8,000 people, with an optimum population size of 4,000.

Census Block Groups

Census block groups are defined as a cluster of census blocks within a census tract that have the same first digit of their four-digit census block number. Block groups never cross census tract boundaries but may cross other boundaries, such as city limits, school districts, and VTD boundaries. Block groups generally contain between 600 and 3,000 people, with an optimum population size of 1,500.

Census Blocks

The census block is the smallest unit of geography for which population data is collected and reported. Census blocks are bounded either by visible features, such as roads, rivers, or shorelines, or by nonvisible features, such as county lines, city limits, school district boundaries, property lines, short line-of-sight extensions of roads, or 2020 precinct boundaries.

City Limits

City limits as of January 1, 2020, are in the 2021 redistricting database. The accuracy of city limits in the redistricting database depends to a large extent on individual cities' participation in the 2020 Boundary and Annexation Survey, a survey of all incorporated places conducted by the Census Bureau to determine the correct legal limits of these areas. City limits that serve as census block boundaries can be used as district boundaries. Cities can be used as modeling units in RedAppl, displayed on maps, and listed by district in reports. Census designated places (CDPs) are densely settled, unincorporated areas that are locally identified by name and for which the Census Bureau reports population. The boundaries of a CDP

are established by the bureau in cooperation with state and local government officials. Cities and CDPs are reported together on council population reports and in RedAppl modeling under the category of cities.

REFERENCE UNITS

In addition to the census geographic areas used for modeling, other geographic areas have been compiled for map orientation and reference purposes only, including current legislative, congressional, and State Board of Education (SBOE) districts, 2020 general election precincts, incumbent locations, road and water features, and school districts. RedAppl users also may import user-defined geography.

Current District Boundaries

Boundaries for congressional, state senate, state house, and SBOE districts in effect in 2020 are included in the geographic database.

County Election Precincts

County election precincts are the geographic units established by county commissioners courts for the purpose of election administration. Precincts can be bounded by visible or nonvisible features. Under the Texas Election Code, precincts may not contain territory from more than one congressional, state senate, state house, or SBOE district. Council staff collect precinct changes from county officials before each statewide election to ensure that the link between geographic data and election and population data in the redistricting database is accurate. Precincts for the 2020 general election are displayed in RedAppl and on maps, but data are not reported for them in the redistricting reporting system.

Incumbent Residence Locations

The council includes incumbent residence locations for members of the legislature, members of Congress elected from Texas, and members of the SBOE in the redistricting database to enable legislators to consider the effect of any proposed district boundary changes on incumbents. These locations are identified in the redistricting database as census block locations. Neither the specific address nor the precise location of a member's residence within a block is indicated in RedAppl or on reports showing the incumbents who reside in each district in a proposed plan.

Council staff obtain most of these locations from the permanent address provided on forms filed by candidates with the secretary of state. Incumbents are contacted and asked to verify that the location shown on a map is correct.

Road and Water Features

Roads (highways, county roads, city streets, etc.) and road names can be displayed in RedAppl to assist with orientation within a county or counties. Water features in the system may not be totally accurate and should be used for general reference purposes only.

School Districts

The 2020 TIGER file contains 2018-2019 school year school districts and school district names. These school district boundaries serve as census block boundaries, and population can be summed for them by block. The 2020-2021 school year school district coverage, prepared by council staff, is used as a reference layer in RedAppl, on maps, and in population data reports.

User-Defined Geography

RedAppl allows a user to import other geography as a reference overlay that may be of value to the particular user, such as a community of interest. The geography must be in a shapefile format and be in the same geographic projection as the RedAppl data. The map projection that the council uses is Texas Centric Mapping System/Lambert Conformal (TCMS/LC), North American Datum 1983 (NAD83).

GEOGRAPHY REPORTS

Generally, reports provide population or election data, or both, for districts and for various units of census and election geography. The reports that relate solely to geography include those listing components of districts, such as lists of VTDs, precincts, or school districts; plan-checking reports that are used to verify that all of the census blocks in the state are included in the redistricting plan and that all of the districts are contiguous; and compactness reports.

Compactness

Compactness is a basic redistricting principle that generally refers to the geographic shape of a district and how that district's territory is dispersed within its boundaries. Courts often refer to compactness as a traditional redistricting principle and may examine compactness when considering the extent to which a district is gerrymandered. In redistricting, compactness is frequently analyzed using methods that focus on the dispersion of the area of the district, the relation of the perimeter of the district to its area, or the dispersion of the population of the district.

Compactness Measures

There are three basic types of compactness measures included in the redistricting system: those that analyze the area, the perimeter, and the population of a district.

- **Area Dispersion:** This measure examines the relative degree to which a district's area is compact when compared with the area of a similar compact figure. It is the ratio of the area of the district to the area of the smallest convex polygon that can enclose the district (imagine a rubber band stretched around the district). A district in the shape of a square would receive a perfect score of one using this measure. This measure penalizes a district that has long "fingers" or extensions that make it less compact because it requires a larger convex polygon to enclose the entire district, yet much of that polygon is empty.
- **Perimeter:** This measure compares the relative length of the perimeter of a district to its area. It is the ratio of the area of the district to the area of a circle with the same perimeter as the district. A perimeter-to-area measure penalizes a district's compactness score whenever the boundaries are uneven or irregular: the more the boundary zigzags (for example, a river), the less compact the district under this measure.
- **Population:** This is a population-based compactness measure computed as the ratio of the population of the district to the population of all census blocks contained in the smallest convex polygon enclosing the district. The greater the population within the convex polygon that is not within the district, the less compact the district under this measure.

The measures presented in RedAppl and on reports can range from zero to one, with one indicating perfect compactness for the particular scale. The lower the score, the less compact the district.

ELECTION DATA

Election data collected by the council include election returns, voter registration (VR), Spanish surname voter registration (SSVR), turnout (TO), and Spanish surname turnout (SSTO) (as a percentage of SSVR), and precinct and candidate information for statewide election contests, selected county and city elections, and special elections for state districted offices. For redistricting in 2021, election data from the 2012 through the 2020 elections will be available in RedAppl and for maps and reports.

ELECTION RETURNS

Returns are votes that were cast for each candidate in an election, reported by precinct. County election officials submit returns for state offices to the secretary of state, as required by the Texas Election Code. Some counties submit returns for local county offices to the secretary of state along with state office returns, although they are not required to do so. The council may request precinct returns from the secretary of state and county or local election administrators for a given election.

The following election returns have been computerized for 2021 redistricting:

- **Statewide office and statewide districted returns:** 2012-2020 state Democratic and Republican primaries and runoffs and general elections are available in RedAppl and for maps and reports. Complete official election results for all years are maintained by the secretary of state. Statewide offices include president, U.S. senator, governor, lieutenant governor, attorney general, comptroller of public accounts, land commissioner, agriculture commissioner, railroad commissioner, supreme court justices, and court of criminal appeals justices; statewide districted offices include U.S. representative, SBOE, state senator, and state representative. All Texas counties are included in statewide elections unless party primaries or runoffs are not held in the county.
- **Countywide office and county districted returns:** 2012-2020 selected county elections in 22 of the most populous counties (Bell, Bexar, Brazoria, Brazos, Cameron, Collin, Dallas, Denton, El Paso, Fort Bend, Galveston, Harris, Hidalgo, Jefferson, Lubbock, McLennan, Montgomery, Nueces, Smith, Tarrant, Travis, and Williamson). Countywide offices include district judge, county judge, county court judges, district attorney, district clerk, county clerk, county treasurer, tax assessor-collector, and sheriff; county districted offices include county commissioners, justices of the peace, and constables.
- **City returns:** a limited set of selected city election contests are available for maps and reports. City elections include contests for mayor and city council members.
- **Special elections:** 2012-2020 special elections to fill vacancies for congressional and state legislative offices are available for reports and RedAppl shading only.

Precinct returns for each state and county election are entered independently by two council staff members in a double-entry process to ensure accuracy. The totals obtained from the entered precinct data for state offices are then compared with county canvass totals provided by the secretary of state or, for local contests, with totals provided by local sources. If discrepancies exist, the totals are checked and county election officials are contacted before any differences are left in the system. Most differences in totals are small and none were great enough to change the outcome of any contest.

VOTER REGISTRATION (VR)

Section 18.061, Election Code, requires the secretary of state to maintain a statewide computerized voter registration list that serves as the single system for storing and managing the official list of registered voters in the state. The voter registration (VR) totals for state and county elections in the council's elections database are

derived from this secretary of state's Statewide Voter Database. Registration data for a city or special election is derived from the registration for the most recent general or primary election prior to the date of the given election.

Chapter 15, Election Code, requires that county voter registrars maintain a suspense list containing the name of each registered voter who fails to respond within 30 days to a confirmation of residence notice that is sent when the registrar has reason to believe that a voter's current residence is different from that indicated on the registration records, or when a renewal registration certificate sent in an odd-numbered year is returned because the addressee has moved. Following the second general election that occurs after the date a voter's name is entered on the suspense list, the voter's registration is canceled by the registrar. Until the registration is canceled, a voter on the suspense list is still a registered voter. Suspense information is collected by the secretary of state, included in the Statewide Voter Database, and used to provide a count of non-suspense voters on the Red202 report. All other council election reports reflect total voter registration.

SPANISH SURNAME VOTER REGISTRATION (SSVR)

Spanish surname voter registration (SSVR), also reported in the secretary of state's Statewide Voter Database, is generated using a comparison to a Census Bureau list of Spanish surnames. The reported number of registered voters with Spanish surnames is, of course, not a precise measure of Hispanic voter registration. Some people who consider themselves Hispanic do not have surnames that are included in the Spanish surname file and will be missed by the Spanish surname matching technique. Others, who have surnames that are included in the Spanish surname file but do not consider themselves Hispanic, will nonetheless be incorrectly counted as Hispanic registered voters.

VOTER TURNOUT (TO)

Voter turnout (TO) is the number of registered voters who voted in a given election. If total turnout in an election is not provided by a county, the largest number of votes cast in any contest in that election is used as the precinct turnout total in council reports.

SPANISH SURNAME TURNOUT (SSTO)

Spanish surname turnout (SSTO) is reported in the secretary of state's Statewide Voter Database and is generated in the same manner and with the same limitations as Spanish surname voter registration (SSVR). The council only reports SSTO as a percentage of the total turnout included in the secretary of state's voter database.

PRECINCT INFORMATION

County precinct boundary changes and lists of precincts corresponding to the elections for which the council computerizes election returns are collected throughout the decade to ensure that the link between election data and geographic and population data is accurate.

CANDIDATE INFORMATION

Candidate information that the council collects for redistricting includes the candidate's name, party affiliation (for partisan state and county elections only), and identification of incumbency.

Candidates are identified from lists obtained from the secretary of state or from local election officials. In the council database, a candidate's name is entered as it appears on the ballot. If different versions of a name are used by an individual in different elections, the name is standardized so that the candidate can be tracked across elections.

Council staff identifies incumbent state officials from various published sources, as well as from canvass reports and letters of appointment from the governor. Incumbent local officials are identified in the system if that information is provided by local election officials.

ELECTION REPORTS

Election reports include election returns by candidate, voter registration (VR), Spanish surname voter registration (SSVR), turnout (TO), and Spanish surname turnout (SSTO) (as a percentage of SSVR) and are available by district, county, or VTD. Election returns, voter registration, and voter turnout are captured in the council's election database by precinct or consolidated precinct in effect for the specific election. Counties may change the boundaries of voting precincts from one election to the next. Special reports of election returns by precinct can be produced, but returns are reported by VTD for proposed redistricting plans.

Council staff create a set of VTDs that corresponds to the set of precincts in effect for each election in the election database. To allow rapid correlation of election and population data with proposed districts for purposes of analysis, election data from each election is allocated to 2020 census blocks and then summed to 2020 general election VTDs, counties, and districts. This data is available for use in RedAppl and for maps and reports produced by the council.

Allocation

For 2021 redistricting, the election returns, registration, turnout, and precinct geography from 2012 through 2020 are correlated with population from the 2020 census. Because detailed population data from earlier years is not available to be used in the allocation, the allocated data will not exactly match the precinct-level data from past elections. Used with a clear understanding of its limitations by those familiar with local circumstances, however, the data can be an effective tool for analysis of proposed districts and plans.

The allocation is accomplished through a three-step process: (1) allocating precinct-level voter registration (VR) using voting age population (VAP) from the most recent census, (2) allocating election returns using the allocated VR, and (3) allocating turnout using the allocated election returns. Using VR and returns for the subsequent allocation steps serves to control the allocated data to the precinct totals and to preserve the relationships among the data.

The first step consists of two parts, allocating SSVR using Hispanic VAP and allocating non-SSVR using non-Hispanic VAP. VR and SSVR are acquired from the secretary of state's Statewide Voter Database. The census blocks associated with a precinct are sorted in descending order by the allocation variable (in this case, non-Hispanic and Hispanic VAP), and the VR is allocated. The formula for calculating census block VR is:

$$\text{blockVR} = \frac{blockVAP}{\sum blockVAP} * totalVR$$

The result for each block is rounded to the nearest whole number and subtracted from the pool of remaining VR to be allocated. The block-level VAP is also subtracted from the total VAP for the blocks comprising the precinct (denominator). This process is repeated recursively until all of the VR for a precinct is allocated.

If a precinct either has SSVR but no Hispanic VAP or has non-SSVR but no non-Hispanic VAP, the total VAP is used for the allocation. If the blocks comprising a precinct have no VAP but have election data reported from a different year, each block is assigned a value of one and the total for the blocks is assigned the total number of census blocks. For example, a precinct with 10 census blocks would have an allocation ratio of 1/10 for the first block, with 1/9 for the next iteration, etc.

If a county reports election data as a county total, the data is allocated to census blocks in the county as discussed above as if the entire county were a single VTD.

The second step calculates block level returns by the same method as above, using the allocated VR from the first step as the allocation variable. The final step calculates block-level election turnout using the allocated returns from the second step.

Limitations of the Allocated Data

When using the allocated election data, several limitations of the data should be considered:

- Allocation based on voting age population (VAP) assumes that voters within each block within a precinct exhibited the same voting pattern, which is not likely to be the case.
- Using census data to allocate election data becomes less reliable as the date of the census data becomes progressively further from the date of the election.
- Election data for a single small geographic unit such as a census block yields a very low confidence level. As with other statistical data, confidence levels increase as the blocks are aggregated into larger units such as counties or legislative districts.
- Election data is summed to the most recent VTDs in RedAppl. For 2021, this is the 2020 general election VTDs. VTDs and precincts from other election cycles will not exactly match the current VTDs.
- The more a VTD differs from the actual precinct, the more likely the allocated data for that VTD will differ from the election data for the actual precinct. In 2021, the 2020 general election VTDs will be a very close match to the 2020 general election precincts, but as new precincts are drawn through the coming decade, the chances that precinct geography will deviate from 2020 census geography will increase.

RESOURCES FOR VIEWING AND OBTAINING DATA

CAPITOL DATA PORTAL

All the redistricting data maintained by the legislature is available from the Capitol Data Portal (https://data.capitol.texas.gov), which can be accessed from the Texas Redistricting website. Maps and data reports for redistricting plans considered by the legislature are also available.

Sections of the data portal include Elections, Geography, Redistricting 2021, and Redistricting Archive. The Elections section includes votes by VTD for general, primary, runoff, and special elections, allocated to geographic units in the Texas redistricting systems. The Geography section provides coverages of VTDs, school districts, and precincts drawn on census geography. These are the coverages used for Texas redistricting systems. Redistricting 2021 includes maps and reports for current (2020) state house, state senate, Texas congressional, and SBOE districts drawn on 2020 census geography, as well as maps and reports for all redistricting plans offered for consideration in the Texas Legislature or in the courts during the 2020 decade. Redistricting Archive (2010s) includes that same information for plans considered throughout the 2010 decade.

Data available for each redistricting plan include files that relate precincts to districts and shapefiles of counties, tracts, block groups, blocks, precincts, VTDs, and school districts. Detailed census data is reported for each district. Election returns and voter registration and turnout data are available by VTD, and additional records are added as soon as data from a new election is collected, processed, and verified.

The election and geographic data that may be obtained from this site are developed specifically for redistricting in Texas and may be used in any compatible redistricting or GIS applications.

DISTRICTVIEWER

DistrictViewer (https://dvr.capitol.texas.gov) is the legislature's public website that displays maps and associated demographic and election reports for current and proposed state senate, state house, Texas congressional, and SBOE districts. All redistricting proposals that are made public during the legislative process are made available in DistrictViewer as soon as possible after becoming public, such as by the filing of a redistricting bill or the offer of an amendment in committee or during floor consideration of a redistricting bill. DistrictViewer also includes redistricting proposals that are submitted to the legislature by committee witnesses or other interested members of the public. Redistricting plans submitted in subsequent litigation as evidentiary exhibits or proposed remedies are also included in DistrictViewer as they come to the attention of council staff. Users of DistrictViewer may overlay one plan on another to compare the two. DistrictViewer allows the user to zoom in to view the details of a proposed plan, such as street names, cities, and election precincts, with or without district boundaries. Users may also search for an address and easily access geographic, election, and demographic data and additional resources related to any public plan.

This page left intentionally blank.

APPENDIX
DIFFERENTIAL PRIVACY AND THE 2020 CENSUS

CENSUS BUREAU'S IMPLEMENTATION OF DIFFERENTIAL PRIVACY

The Census Bureau's implementation of differential privacy requires operational decisions that include setting a privacy-loss budget to measure the balance of privacy and accuracy, determining which totals will be published as invariants, and selecting the algorithm that will process the data. Each decision is a policy decision.

The Census Bureau's adoption of differential privacy has been overseen by the Data Stewardship Executive Policy Committee (DSEP), an advisory body of career professionals within the bureau. In November 2020, DSEP announced three items would be held invariant for the 2020 P.L. 94-171 redistricting data file: total population for state and state equivalents, number of group quarters by type at the block level, and total housing units at the block level. Total population for states is the data that is used for the apportionment process for congressional seats.

The Census Bureau uses a TopDown Algorithm (TDA) to implement differential privacy. The TDA begins at the largest geographic unit, the nation, and moves down the census geographic hierarchy. The TDA inputs the Census Edited File (CEF), the file of raw census responses that has imputation for empty cells and corrections for cells that are not logically possible. The privacy-loss budget is allocated six ways, one for each geographic unit along the census hierarchy (nation, state, county, tract, block group, and block). Starting at the highest level of the census geographic hierarchy, the TDA uses the national portion of the privacy-loss budget to create a histogram where the data from the CEF has noise applied. The histogram is a cross tabulation that shows counts of unique combinations of demographic attributes, such as race, voting age population (VAP), etc. for predesigned queries (e.g. how many Hispanic females are 18 years of age or older?) of each geographic unit. After the noise is applied to the national data, the TDA moves to the state level, where each state is processed individually using the level of the privacy-loss budget set for states. This process continues down for each geographic level. This process of adjusting the data from the census edited file and injecting statistical noise through the TDA cannot be reversed to return to the original numbers.

The following is from: 2020 Census State Redistricting Data (Public Law 94-171) Summary File, 2020 Census of Population and Housing, Technical Documentation, Issued June 2021, U.S. Department of Commerce, U.S. Census Bureau

---

CHAPTER 7. 2020 CENSUS: OPERATIONAL OVERVIEW AND DISCLOSURE AVOIDANCE

2020 CENSUS OPERATIONAL OVERVIEW

The 2020 Census operational design comprised a set of design decisions that drove how the 2020 Census was conducted. These design decisions were informed through research, testing, and analysis conducted from 2012 through 2018. The operational design also drove the requirements for information technology capabilities and acquisitions required to conduct the census. The 2020 Census was designed and developed in an iterative fashion, incorporating results from the various tests conducted leading up to the 2020 Census. Most of the design decisions implemented in the 2020 Census were captured in the 2020 Census Operational Plan v4.0, issued in December 2018.

Due to the COVID-19 outbreak, the U.S. Census Bureau adjusted 2020 Census operation production dates in order to protect the health and safety of the American public and Census Bureau employees, implement guidance from federal, state, and local authorities, and ensure a complete and accurate count of all communities.

2020 Census Operational Plan

The Census Bureau's 2020 Census Operational Plan (Op Plan) documents the design for conducting the 2020 Census. The Op Plan covers all operations required to execute the 2020 Census, starting with precensus address and geographic feature updates, and ending once census data products are disseminated and coverage and quality are measured. The Op Plan reflects and supports evidence-based decision-making by describing design concepts and their rationale, identifying decisions made at the time of publication, and describing significant issues and risks related to the implementation of the 2020 Census operation.

The Op Plan provides an overview of the 2020 Census and the 35 required operations, program key innovations, program tests, milestones, production dates, program-level risks, and data quality analysis to include operation descriptions, time lines, scope, recommendations (based on 2010 lessons learned), research conducted, decisions made, issues, risks, costs, and milestones.

CONFIDENTIALITY OF THE DATA

The Census Bureau has modified some data in this data release to protect confidentiality. Title 13 U.S. Code, Section 9 prohibits the Census Bureau from publishing results in which an individual's data can be identified.

Title 13 U.S. Code

Title 13 of the U.S. Code authorizes the Census Bureau to conduct surveys and censuses and mandates that any information obtained from private individuals and establishments remains confidential. Section 9 of Title 13 prohibits the Census Bureau from releasing "any publication whereby the data furnished by any particular establishment or individual under this title can be identified." Section 214 of Title 13, as modified by the Federal Sentencing Reform Act, imposes a fine of not more than $250,000 and/or imprisonment of not more than five years for publication or communication in violation of Section 9.

Disclosure Avoidance

Disclosure avoidance is the process of disguising data to protect confidentiality. A disclosure of data occurs when someone can use published statistical information to identify an individual who provided information under a pledge of confidentiality. Using disclosure avoidance, the Census Bureau modifies or removes all of the characteristics that put confidential information at risk for disclosure. Although it may appear that a table shows information about a specific individual, the Census Bureau has implemented a disclosure avoidance system based on differential privacy to disguise the original data while making sure the results are useful.

Differential Privacy

Formally, private disclosure avoidance methods, like differential privacy, are similar to a broad class of traditional disclosure avoidance methods that protect privacy through the introduction of statistical noise into the data. This noise introduces uncertainty to reduce the likelihood that a specific individual can be identified and to reduce the likelihood that sensitive attributes about that individual can be inferred.

Differential privacy differs from traditional noise-injection privacy methods insofar as the amount of noise required to protect privacy is precisely calibrated to provide provable mathematical guarantees regarding the maximum amount of privacy-loss possible from the publication of data products derived from the confidential data. This guarantee is independent of the tools and external information that a would-be attacker (present or future) could use to attempt to re-identify individuals or to infer sensitive attributes about them. This maximum bound to the privacy-loss is reflected in the privacy-loss budget for the data products and represents a mathematically rigorous, future-proof privacy guarantee to our respondents.

The TopDown Algorithm

The Disclosure Avoidance System (DAS) used to protect the Public Law 94-171 data works as follows:

(1) After the confidential Census Edited File is input into the DAS, the system's TopDown Algorithm takes an extensive series of differentially private "noisy" measurements.

(2) The algorithm uses these measurements to generate privacy-protected microdata records for the entire nation.

(3) These individual records contain every level of geography on the Census Bureau's geographic backbone based on the noisy measurements taken at each of those geographic levels and subject to the population invariants and other constraints (discussed below).

(4) These microdata records are exported into the tabulation system to generate the data products for this publication.

(5) The resulting data reflect the privacy guarantees established by the global privacy-loss budget for the 2020 Census, incorporating the greatest level of uncertainty at the census block level (where privacy risk is usually greatest), while providing increasingly accurate measures of the nation's population at each higher level of geography.

Invariants and the Privacy-Loss Budget

To meet the Census Bureau's constitutionally mandated responsibility to apportion representatives for the House of Representatives according to the actual enumerated population, the DAS departs from pure differential privacy in a few ways. The total population for each state is held invariant—used exactly as enumerated and with no noise added. Similarly, the total number of housing units in each census block and the number and type of each group quarters unit in each census block are also held invariant.

The Census Bureau's Data Stewardship Executive Policy Committee (DSEP) has not yet set the global privacy-loss budget for the 2020 Census. The Census Bureau will communicate the value of the global privacy-loss budget and its allocation across the 2020 Census Data Products to the data user community once those decisions have been made.

ADDITIONAL SOURCES

Please see the following sources for additional information about the Census Bureau's use of Differential Privacy:

- For video, please see: "Protecting Privacy with MATH" by MinutePhysics in collaboration with the Census Bureau
- For audio, please listen to the podcast episode: "Differential Privacy at the US Census" of Data Skeptic, featuring Simson Garfinkel, senior computer scientist at the Census Bureau
- For technical documentation from the Census Bureau about Differential Privacy in relation to redistricting, please see: "2020 Census State Redistricting Data (Public Law 94-171) Summary File"
- For further information, please see: "Census TopDown: Differentially Private Data, Incremental Schemas, and Consistency with Public Knowledge" by John Abowd et al. and "Chapter 2: Disclosure Avoidance in the 2020 Census" from 2020 Census Data Products: Data Needs and Privacy Considerations: Proceedings of a Workshop